Functional linear regression via canonical analysis

GUOZHONG HE[1], HANS-GEORG MÜLLER[2,*], JANE-LING WANG[2,**]
and WENJING YANG[2,†]

[1]*University of California, San Francisco, Institute for Health and Aging and California Diabetes Program, California Department of Public Health P.O. Box 997377, MS 7211, Sacramento, CA 95899-7377, USA.*
E-mail: ghe@dhs.ca.gov
[2]*Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA.*
E-mail: *mueller@wald.ucdavis.edu; **wang@wald.ucdavis.edu; †wyang@wald.ucdavis.edu

We study regression models for the situation where both dependent and independent variables are square-integrable stochastic processes. Questions concerning the definition and existence of the corresponding functional linear regression models and some basic properties are explored for this situation. We derive a representation of the regression parameter function in terms of the canonical components of the processes involved. This representation establishes a connection between functional regression and functional canonical analysis and suggests alternative approaches for the implementation of functional linear regression analysis. A specific procedure for the estimation of the regression parameter function using canonical expansions is proposed and compared with an established functional principal component regression approach. As an example of an application, we present an analysis of mortality data for cohorts of medflies, obtained in experimental studies of aging and longevity.

Keywords: canonical components; covariance operator; functional data analysis; functional linear model; longitudinal data; parameter function; stochastic process

1. Introduction

With the advancement of modern technology, data sets which contain repeated measurements obtained on a dense grid are becoming ubiquitous. Such data can be viewed as a sample of curves or functions and are referred to as functional data. We consider here the extension of the linear regression model to the case of functional data. In this extension, both predictors and responses are random functions rather than random vectors. It is well known (Ramsay and Dalzell (1991); Ramsay and Silverman (2005)) that the traditional linear regression model for multivariate data, defined as

$$\mathbf{Y} = \boldsymbol{\alpha}_0 + \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon},\tag{1}$$

may be extended to the functional setting by postulating the model, for $s \in T_1, t \in T_2$,

$$Y(t) = \alpha_0(t) + \int_{T_1} X(s)\beta_0(s, t)\,\mathrm{d}s + \varepsilon(t).\tag{2}$$

Writing all vectors as row vectors in the classical model (1), $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are random vectors in $\mathbf{R}^{p_2}$, $\mathbf{X}$ is a random vector in $\mathbf{R}^{p_1}$, and $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are, respectively, $1 \times p_2$ and $p_1 \times p_2$ matrices containing the regression parameters. The vector $\boldsymbol{\varepsilon}$ has the usual interpretation of an error vector, with $E[\boldsymbol{\varepsilon}] = 0$ and $\text{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, I denoting the identity matrix. In the functional model (2), random vectors $\mathbf{X}, \mathbf{Y}$ and $\boldsymbol{\varepsilon}$ in (1) are replaced by random functions defined on the intervals T_1 and T_2. The extension of the classical linear model (1) to the functional linear model (2) is obtained by replacing the matrix operation on the right-hand side of (1) with an integral operator in (2). In the original approach of Ramsay and Dalzell (1991), a penalized least-squares approach using L-splines was adopted and applied to a study in temperature-precipitation patterns, based on data from Canadian weather stations.

The functional regression model (2) for the case of scalar responses has attracted much recent interest (Cardot and Sarda (2005); Müller and Stadtmüller (2005); Hall and Horowitz (2007)), while the case of functional responses has been much less thoroughly investigated (Ramsay and Dalzell (1991); Yao, Müller and Wang (2005b)). Discussions on various approaches and estimation procedures can be found in the insightful monograph of Ramsay and Silverman (2005). In this paper, we propose an alternative approach to predict $Y(\cdot)$ from $X(\cdot)$, by adopting a novel canonical representation of the regression parameter function $\beta_0(s, t)$. Several distinctive features of functional linear models emerge in the development of this canonical expansion approach.

It is well known that in the classical multivariate linear model, the regression slope parameter matrix is uniquely determined by $\boldsymbol{\beta}_0 = \text{cov}(\mathbf{X})^{-1} \text{cov}(\mathbf{X}, \mathbf{Y})$, as long as the covariance matrix $\text{cov}(\mathbf{X})$ is invertible. In contrast, the corresponding parameter function $\beta_0(\cdot, \cdot)$, appearing in (2), is typically not identifiable. This identifiability issue is discussed in Section 2. It relates to the compactness of the covariance operator of the process X which makes it non-invertible. In Section 2, we demonstrate how restriction to a subspace allows this problem to be circumvented. Under suitable restrictions, the components of model (2) are then well defined.

Utilizing the canonical decomposition in Theorem 3.3 below leads to an alternative approach to estimating the parameter function $\beta_0(\cdot, \cdot)$. The canonical decomposition links Y and X through their functional canonical correlation structure. The corresponding canonical components form a bridge between canonical analysis and linear regression modeling. Canonical components provide a decomposition of the structure of the dependency between Y and X and lead to a natural expansion of the regression parameter function $\beta_0(\cdot, \cdot)$, thus aiding in its interpretation. The canonical regression decomposition also suggests a new family of estimation procedures for functional regression analysis. We refer to this methodology as *functional canonical regression analysis*. Classical canonical correlation analysis (CCA) was introduced by Hotelling (1936) and was connected to function spaces by Hannan (1961). Substantial extensions and connections to reproducing kernel Hilbert spaces were recently developed in Eubank and Hsing (2008); for other recent developments see Cupidon *et al.* (2007).

Canonical correlation is known not to work particularly well for very high-dimensional multivariate data, as it involves an inverse problem. Leurgans, Moyeed and Silverman (1993) tackled the difficult problem of extending CCA to the case of infinite-dimensional functional data and discussed the precarious regularization issues which are faced; He, Müller and Wang (2003, 2004) further explored various aspects and proposed practically feasible regularization procedures for functional CCA. While CCA for functional data is worthwhile, but difficult to implement and interpret, the canonical approach to functional regression is here found to compare

favorably with the well established principal-component-based regression approach in an example of an application (Section 5). This demonstrates a potentially important new role for canonical decompositions in functional regression analysis. The functional linear model (2) includes the varying coefficient linear model studied in Hoover *et al.* (1998) and Fan and Zhang (2000) as a special case, where $\beta(s, t) = \beta(t)\delta_t(s)$; here, $\delta_t(\cdot)$ is a delta function centered at t and $\beta(t)$ is the varying coefficient function. Other forms of functional regression models with vector-valued predictors and functional responses were considered by Faraway (1997), Shi, Weiss and Taylor (1996), Rice and Wu (2000), Chiou, Müller and Wang (2003) and Ritz and Streibig (2009).

The paper is organized as follows. Functional canonical analysis and functional linear models for L_2-processes are introduced in Section 2. Sufficient conditions for the existence of functional normal equations are given in Proposition 2.2. The canonical regression decomposition and its properties are the theme of Section 3. In Section 4, we propose a novel estimation technique to obtain regression parameter function estimates based on functional canonical components. The regression parameter function is the basic model component of interest in functional linear models, in analogy to the parameter vector in classical linear models. The proposed estimation method, based on a canonical regression decomposition, is contrasted with an established functional regression method based on a principal component decomposition. These methods utilize a dimension reduction step to regularize the solution of the inverse problems posed by both functional regression and functional canonical analysis. As a selection criterion for tuning parameters, such as bandwidths or numbers of canonical components, we use minimization of prediction error via leave-one-curve-out cross-validation (Rice and Silverman (1991)). The proposed estimation procedures are applied to mortality data obtained for cohorts of medflies (Section 5). Our goal in this application is to predict a random trajectory of mortality for a female cohort of flies from the trajectory of mortality for a male cohort which was raised in the same cage. We find that the proposed functional canonical regression method gains an advantage over functional principal component regression in terms of prediction error.

Additional results on canonical regression decompositions and properties of functional regression operators are compiled in Section 6. All proofs are collected in Section 7.

2. Functional linear regression and the functional normal equation

In this section, we explore the formal setting as well as identifiability issues for functional linear regression models. Both response and predictor functions are considered to come from a sample of pairs of random curves. A basic assumption is that all random curves or functions are square-integrable stochastic processes. Consider a measure μ on a real index set T and let $L_2(T)$ be the class of real-valued functions such that $\| f \|^2 = \int_T |f|^2 \, d\mu < \infty$. This is a Hilbert space with the inner product $\langle f, g \rangle = \int_T fg \, d\mu$ and we write $f = g$ if $\int_T (f - g)^2 \, d\mu = 0$. The index set T can be a set of time points, such as $T = \{1, 2, \ldots, k\}$, a compact interval $T = [a, b]$ or even a rectangle formed by two intervals S_1 and S_2, $T = S_1 \times S_2$. We focus on index sets T that are either compact real intervals or compact rectangles in $\mathbf{R}^2$ and consider μ to be the Lebesgue measure on $\mathbf{R}^1$ or $\mathbf{R}^2$. Extensions to other index sets T and other measures are self-evident. An L_2-process

is a stochastic process $X = \{X(t), t \in T\}$, $X \in L_2(T)$, with $E[\|X\|^2] < \infty$, $E[X(t)^2] < \infty$ for all $t \in T$.

Let $X \in L_2(T_1)$ and $Y \in L_2(T_2)$.

Definition 2.1. *Processes (X, Y) are subject to a functional linear model if*

$$Y(t) = \alpha_0(t) + \int_{T_1} X(s)\beta_0(s, t)\,\mathrm{d}s + \varepsilon(t), \qquad t \in T_2, \tag{3}$$

where $\beta_0 \in L_2(T_1 \times T_2)$ is the parameter function, $\varepsilon \in L_2(T_2)$ is a random error process with $E[\varepsilon(t)] = 0$ for $t \in T_1$, and ε and X are uncorrelated, in the sense that $E[X(t)\varepsilon(s)] = 0$ for all $s, t \in T_1$.

Without loss of generality, we assume from now on that all processes considered have zero mean functions, $EX(t) = 0$ and $EY(s) = 0$ for all t, s. We define the regression integral operator $\mathcal{L}_X : L_2(T_1 \times T_2) \to L_2(T_2)$ by

$$(\mathcal{L}_X\beta)(t) = \int_{T_1} X(s)\beta(s, t)\,\mathrm{d}s \qquad \text{for any } \beta \in L_2(T_1 \times T_2).$$

Equation (3) can then be rewritten as

$$Y = \mathcal{L}_X\beta_0 + \varepsilon. \tag{4}$$

Denote the auto- and cross-covariance functions of X and Y by

$$r_{XX}(s, t) = \mathrm{cov}[X(s), X(t)], \qquad s, t \in T_1,$$
$$r_{YY}(s, t) = \mathrm{cov}[Y(s), Y(t)], \qquad s, t \in T_2, \quad \text{and}$$
$$r_{XY}(s, t) = \mathrm{cov}[X(s), Y(t)], \qquad s \in T_1, t \in T_2.$$

The autocovariance operator of X is the integral operator $R_{XX} : L_2(T_1) \to L_2(T_1)$, defined by

$$(R_{XX}u)(s) = \int_{T_1} r_{XX}(s, t)u(t)\,\mathrm{d}t, \qquad u \in L_2(T_1).$$

Replacing r_{XX} by r_{YY}, r_{XY}, we analogously define operators $R_{YY} : L_2(T_2) \to L_2(T_2)$ and $R_{XY} : L_2(T_2) \to L_2(T_1)$, similarly R_{YX}. Then R_{XX} and R_{YY} are compact, self-adjoint and non-negative definite operators, and R_{XY} and R_{YX} are compact operators (Conway (1985)). We refer to He *et al.* (2003) for a discussion of various properties of these operators.

Another linear operator of interest is the integral operator $\Gamma_{XX} : L_2(T_1 \times T_2) \to L_2(T_1 \times T_2)$,

$$(\Gamma_{XX}\beta)(s, t) = \int_{T_1} r_{XX}(s, w)\beta(w, t)\,\mathrm{d}w. \tag{5}$$

The operator equation

$$r_{XY} = \Gamma_{XX}\beta, \qquad \beta \in L_2(T_1 \times T_2) \tag{6}$$

is a direct extension of the least-squares normal equation and may be referred to as the functional population normal equation.

Proposition 2.2. *The following statements are equivalent for a function $\beta_0 \in L_2(T_1 \times T_2)$:*

    (a) β_0 *satisfies the linear model* (4);
    (b) β_0 *is a solution of the functional normal equation* (6);
    (c) β_0 *minimizes* $E\|Y - \mathcal{L}_X\beta\|^2$ *among all* $\beta \in L_2(T_1 \times T_2)$.

The proof is found Section 7. In the infinite-dimensional case, the operator Γ_{XX} is a Hilbert–Schmidt operator in the Hilbert space L_2, according to Proposition 6.6 below. A problem we face is that it is known from functional analysis that a bounded inverse does not exist for such operators. A consequence is that the parameter function β_0 in (3), (4) is not identifiable without additional constraints. In a situation where the inverse of the covariance matrix does not exist in the multivariate case, a unique solution of the normal equation always exists within the column space of cov(**X**) and this solution then minimizes $E\|Y - \mathcal{L}_X\beta\|^2$ on that space. Our idea to get around the non-invertibility issue in the functional infinite-dimensional case is to extend this approach for the non-invertible multivariate case to the functional case. Indeed, as is demonstrated in Theorem 2.3 below, under the additional Condition (C1), the solution of (6) exists in the subspace defined by the range of Γ_{XX}. This unique solution indeed minimizes $E\|Y - \mathcal{L}_X\beta\|^2$.

We will make use of the Karhunen–Loève decompositions (Ash and Gardner (1975)) for L_2-processes X and Y,

$$X(s) = \sum_{m=1}^{\infty} \xi_m \theta_m(s), \qquad s \in T_1 \quad \text{and} \quad Y(t) = \sum_{j=1}^{\infty} \zeta_j \varphi_j(t), \qquad t \in T_2, \qquad (7)$$

with random variables ξ_m, ζ_j, m, $j \geq 1$, and orthonormal families of L_2-functions $\{\theta_m\}_{m \geq 1}$ and $\{\varphi_j\}_{j \geq 1}$. Here, $E\xi_m = E\zeta_j = 0$, $E\xi_m\xi_p = \lambda_{Xm}\delta_{mp}$, $E\zeta_j\zeta_p = \lambda_{Yj}\delta_{jp}$ and $\{(\lambda_{Xm}, \theta_m)\}$, $\{(\lambda_{Yj}, \varphi_j)\}$ are the eigenvalues and eigenfunctions of the covariance operators R_{XX} and R_{YY}, respectively, with $\sum_m \lambda_{Xm} < \infty$, $\sum_j \lambda_{Yj} < \infty$. Note that δ_{mj} is the Kronecker symbol with $\delta_{mj} = 1$ for $m = j$, $\delta_{mj} = 0$ for $m \neq j$.

We consider a subset of L_2 on which inverses of the operator Γ_{XX} can be defined. As a Hilbert–Schmidt operator, Γ_{XX} is compact and therefore not invertible on L_2. According to Conway (1985), page 50, the range of Γ_{XX},

$$G_{XX} = \{\Gamma_{XX}h : h \in L_2(T_1 \times T_2)\},$$

is characterized by

$$G_{XX} = \left\{ g \in L_2(T_1 \times T_2) : \sum_{m,j=1}^{\infty} \lambda_{Xm}^{-2} |\langle g, \theta_m\varphi_j \rangle|^2 < \infty, \, g \perp \ker(\Gamma_{XX}) \right\}, \qquad (8)$$

where $\ker(\Gamma_{XX}) = \{h : \Gamma_{XX}h = 0\}$. Defining

$$G_{XX}^{-1} = \left\{ h \in L_2(T_1 \times T_2): h = \sum_{m,j=1}^{\infty} \lambda_{Xm}^{-1} \langle g, \theta_m \varphi_j \rangle \theta_m \varphi_j, g \in G_{XX} \right\},$$

we find that Γ_{XX} is a one-to-one mapping from the vector space $G_{XX}^{-1} \subset L_2(T_1 \times T_2)$ onto the vector space G_{XX}. Thus, restricting Γ_{XX} to a subdomain defined by the subspace G_{XX}^{-1}, we can define its inverse for $g \in G_{XX}$ as

$$\Gamma_{XX}^{-1} g = \sum_{m,j=1}^{\infty} \lambda_{Xm}^{-1} \langle g, \theta_m \varphi_j \rangle \theta_m \varphi_j. \tag{9}$$

Γ_{XX}^{-1} then satisfies the usual properties of an inverse, in the sense that $\Gamma_{XX}\Gamma_{XX}^{-1} g = g$ for all $g \in G_{XX}$, and $\Gamma_{XX}^{-1}\Gamma_{XX}h = h$ for all $h \in G_{XX}^{-1}$.

The following Condition (C1) for processes (X, Y) is of interest.

Condition (C1). *The L_2-processes X and Y with Karhunen–Loève decompositions* (7) *satisfy*

$$\sum_{m,j=1}^{\infty} \left\{ \frac{E[\xi_m \zeta_j]}{\lambda_{Xm}} \right\}^2 < \infty.$$

If (C1) is satisfied, then the solution to the non-invertibility problem as outlined above is viable in the functional case, as demonstrated by the following basic result on functional linear models.

Theorem 2.3 (Basic theorem for functional linear models). *A unique solution of the linear model* (4) *exists in* $\ker(\Gamma_{XX})^{\perp}$ *if and only if X and Y satisfy Condition* (C1). *In this case, the unique solution is of the form*

$$\beta_0^*(t, s) = (\Gamma_{XX}^{-1} r_{XY})(t, s). \tag{10}$$

As a consequence of Proposition 2.2, solutions of the functional linear model (4), solutions of the functional population normal equation (6) and minimizers of $E\|Y - \mathcal{L}_X \beta\|^2$ are all equivalent and allow the usual projection interpretation.

Proposition 2.4. *Assume X and Y satisfy Condition* (C1). *The following are then equivalent*:

(a) *the set of all solutions of the functional linear model* (4);
(b) *the set of all solutions of the population normal equation* (6);
(c) *the set of all minimizers of $E\|Y - \mathcal{L}_X \beta\|^2$ for $\beta \in L_2(T_1 \times T_2)$*;
(d) *the set $\beta_0^* + \ker(\Gamma_{XX}) = \{\beta_0^* + h | h \in L_2(T_1 \times T_2), \Gamma_{XX}h = 0\}$.*

It is well known that in a finite-dimensional situation, the linear model (6) always has a unique solution in the column space of Γ_{XX}, which may be obtained by using a generalized inverse of

the matrix Γ_{XX}. However, in the infinite-dimensional case, such a solution does not always exist. The following example demonstrates that a pair of L_2-processes does not necessarily satisfy Condition (C1). In this case, the linear model (6) does not have a solution.

Example 2.5. Assume processes X and Y have Karhunen–Loève expansions (7), where the random variables ξ_m, ζ_j satisfy

$$\lambda_{Xm} = E[\xi_m^2] = \frac{1}{m^2}, \qquad \lambda_{Yj} = E[\zeta_j^2] = \frac{1}{j^2} \tag{11}$$

and let

$$E[\xi_m \zeta_j] = \frac{1}{(m+1)^2 (j+1)^2} \qquad \text{for } m, j \geq 1. \tag{12}$$

As shown in He *et al.* (2003), (11) and (12) can be satisfied by a pair of L_2-processes with appropriate operators R_{XX}, R_{YY} and R_{XY}. Then

$$\sum_{m,j=1}^{\infty} \left\{ \frac{E[\xi_m \zeta_j]}{\lambda_{Xm}} \right\}^2 = \lim_{n \to \infty} \sum_{m,j=1}^{n} \left[\frac{m}{(m+1)(j+1)} \right]^4$$

$$= \lim_{n \to \infty} \sum_{m=1}^{n} \left[\frac{m}{(m+1)} \right]^4 \sum_{j=1}^{\infty} \frac{1}{(j+1)^4} = \infty$$

and, therefore, Condition (C1) is not satisfied.

3. Canonical regression analysis

Canonical analysis is a time-honored tool for studying the dependency between the components of a pair of random vectors or stochastic processes; for multivariate stationary time series, its utility was established in the work of Brillinger (1985). In this section, we demonstrate that functional canonical decomposition provides a useful tool to represent functional linear models. The definition of functional canonical correlation for L_2-processes is as follows.

Definition 3.1. *The first canonical correlation ρ_1 and weight functions u_1 and v_1 for L_2-processes X and Y are defined as*

$$\rho_1 = \sup_{u \in L_2(T_1), v \in L_2(T_2)} \mathrm{cov}(\langle u, X \rangle, \langle v, Y \rangle) = \mathrm{cov}(\langle u_1, X \rangle, \langle v_1, Y \rangle), \tag{13}$$

where u and v are subject to

$$\mathrm{var}(\langle u_j, X \rangle) = 1, \qquad \mathrm{var}(\langle v_j, Y \rangle) = 1 \tag{14}$$

for $j = 1$. The kth canonical correlation ρ_k and weight functions u_k, v_k for processes X and Y for $k > 1$ are defined as

$$\rho_k = \sup_{u \in L_2(T_1), v \in L_2(T_2)} \text{cov}(\langle u, X \rangle, \langle v, Y \rangle) = \text{cov}(\langle u_k, X \rangle, \langle v_k, Y \rangle),$$

where u and v are subject to (14) *for $j = k$ and*

$$\text{cov}(\langle u_k, X \rangle, \langle u_j, X \rangle) = 0, \qquad \text{cov}(\langle v_k, Y \rangle, \langle v_j, Y \rangle) = 0$$

for $j = 1, \ldots, k - 1$. We refer to $U_k = \langle u_k, X \rangle$ and $V_k = \langle v_k, Y \rangle$ as the kth canonical variates and to $(\rho_k, u_k, v_k, U_k, V_k)$ as the kth canonical components.

It has been shown in He *et al.* (2003) that canonical correlations do not exist for all L_2-processes, but that Condition (C2) below is sufficient for the existence of canonical correlations and weight functions. We remark that Condition (C2) implies Condition (C1).

Condition (C2). *Let X and Y be L_2-processes, with Karhunen–Loève decompositions* (7) *satisfying*

$$\sum_{m,j=1}^{\infty} \left\{ \frac{E[\xi_m \zeta_j]}{\lambda_{Xm} \lambda_{Yj}^{1/2}} \right\}^2 < \infty.$$

The proposed functional canonical regression analysis exploits features of functional principal components and of functional canonical analysis. In functional principal component analysis, one studies the structure of an L_2-process via its decomposition into the eigenfunctions of its autocovariance operator, the Karhunen–Loève decomposition (Rice and Silverman (1991)). In functional canonical analysis, the relation between a pair of L_2-processes is analyzed by decomposing the processes into their canonical components. The idea of canonical regression analysis is to expand the regression parameter function in terms of functional canonical components for predictor and response processes. The canonical regression decomposition (Theorem 3.3) below provides insights into the structure of the regression parameter functions and not only aids in the understanding of functional linear models, but also leads to promising estimation procedures for functional regression analysis. The details of these estimation procedures will be discussed in Section 4. We demonstrate in Section 5 that these estimates can lead to competitive prediction errors in a finite-sample situation.

We now state two key results. The first of these (Theorem 3.2) provides the canonical decomposition of the cross-covariance function of processes X and Y. This result plays a central role in the solution of the population normal equation (6). This solution is referred to as *canonical regression decomposition* and it leads to an explicit representation of the underlying regression parameter function $\beta_0^*(\cdot, \cdot)$ of the functional linear model (4). The decomposition is in terms of functional canonical correlations ρ_j and canonical weight functions u_j and v_j. Given a predictor process $X(t)$, we obtain, as a consequence, an explicit representation for $E(Y(t)|X) = (\mathcal{L}_X \beta_0^*)(t)$, where $\mathcal{L}_X$ is as in (4). For the following main results, we refer to the definitions of ρ_j, u_j, v_j, U_j, V_j in Definition 3.1. All proofs are found in Section 7.

Theorem 3.2 (Canonical decomposition of cross-covariance function). *Assume that L_2-processes X and Y satisfy Condition (C2). The cross-covariance function r_{XY} then allows the following representation in terms of canonical correlations ρ_j and weight functions u_j and v_j:*

$$r_{XY}(s, t) = \sum_{m=1}^{\infty} \rho_m R_{XX} u_m(s) R_{YY} v_m(t). \tag{15}$$

Theorem 3.3 (Canonical regression decomposition). *Assume that the L_2-processes X and Y satisfy Condition (C2). One then obtains, for the regression parameter function $\beta_0^*(\cdot, \cdot)$ (10), the following explicit solution:*

$$\beta_0^*(s, t) = \sum_{m=1}^{\infty} \rho_m u_m(s) R_{YY} v_m(t). \tag{16}$$

To obtain the predicted value of the response process Y, we use the linear predictor

$$Y^*(t) = E(Y(t)|X) = (\mathcal{L}_X \beta_0^*)(t) = \sum_{m=1}^{\infty} \rho_m U_m R_{YY} v_m(t). \tag{17}$$

This canonical regression decomposition leads to approximations of the regression parameter function β_0^* and the predicted process $Y^*(t) = \mathcal{L}_X \beta_0^*$ via a finitely truncated version of the canonical expansions (16) and (17). The following result provides approximation errors incurred from finite truncation. Thus, we have a vehicle to achieve practically feasible estimation of β_0^* and associated predictions Y^* (Section 4).

Theorem 3.4. *For $K \geq 1$, let $\beta_K^*(s, t) = \sum_{k=1}^{K} \rho_k u_k(s) R_{YY} v_k(t)$ be the finitely truncated version of the canonical regression decomposition (16) for β_0^* and define $Y_K^*(t) = (\mathcal{L}_X \beta_K^*)(t)$. Then,*

$$Y_K^*(t) = \sum_{k=1}^{K} \rho_k U_k R_{YY} v_k(t) \tag{18}$$

with $E[Y_K^*] = 0$. Moreover,

$$E\|Y^* - Y_K^*\|^2 = \sum_{k=K+1}^{\infty} \rho_k^2 \|R_{YY} v_k\|^2 \to 0 \qquad as\ K \to \infty$$

and

$$E\|Y - Y_K^*\|^2 = E\|Y\|^2 - E\|\mathcal{L}_X \beta_K^*\|^2 = \text{trace}(R_{YY}) - \sum_{k=1}^{K} \rho_k^2 \|R_{YY} v_k\|^2. \tag{19}$$

In finite-sample implementations, to be explored in the next two sections, truncation as in (18) is a practical necessity; this requires a choice of suitable truncation parameters.

4. Estimation procedures

4.1. Preliminaries

Estimating the regression parameter function and obtaining fitted processes from the linear model (2) based on a sample of curves is central to the implementation of functional linear models. In practice, data are observed at discrete time points and we temporarily assume, for simplicity, that the N_x time points are the same for all observed predictor curves and are equidistantly spaced over the domain of the data. Analogous assumptions are made for the N_y time points where the response curves are sampled. Thus, the original observations are (X_i, Y_i), $i = 1, \ldots, n$, where X_i is an N_x-dimensional vector sampled at time points s_j, and Y_i is an N_y-dimensional vector sampled at time points t_j. We assume that N_x and N_y are both large. Without going into any analytical details, we compare the finite-sample behavior of two functional regression methods, one of which utilizes the canonical decomposition for regression and the other a well established direct principal component approach to implement functional linear regression.

The proposed practical version of functional regression analysis through functional canonical regression analysis (FCR) is discussed in Section 4.2. This method is compared with a more standard functional linear regression implementation that is based on principal components and referred to as *functional principal regression* (FPR), in Section 4.3. For the choice of the smoothing parameters for the various smoothing steps, we adopt leave-one-curve-out cross-validation (Rice and Silverman (1991)). Smoothing is implemented by local linear fitting for functions and surfaces (Fan and Gijbels (1996)), minimizing locally weighted least squares.

In a pre-processing step, all observed process data are centered by subtracting the cross-sectional means $X_i(s_j) - \frac{1}{n} \sum_{l=1}^{n} X_l(s_j)$, and analogously for Y_i. If the data are not sampled on the same grid for different individuals, a smoothing step may be added before the cross-sectional average is obtained. As in the previous sections, we use in the following the notation X, Y, X_i, Y_i to denote centered processes and trajectories.

When employing the Karhunen–Loève decomposition (7), we approximate observed centered processes by the fitted versions

$$\hat{X}_i(s) = \sum_{l=1}^{L} \hat{\xi}_{il} \hat{\theta}_l(s), \qquad \hat{Y}_i(t) = \sum_{l=1}^{L} \hat{\zeta}_{il} \hat{\varphi}_l(t), \tag{20}$$

where $\{\hat{\theta}_l(s)\}_{l=1}^{L}$ and $\{\hat{\varphi}_l(t)\}_{l=1}^{L}$ are the estimated first L smoothed eigenfunctions for the random processes X and Y, respectively, with the corresponding estimated eigenscores $\{\hat{\xi}_{il}\}_{l=1}^{L}$ and $\{\hat{\zeta}_{il}\}_{l=1}^{L}$ for the ith subject. We obtain these estimates as described in Yao *et al.* (2005a). Related estimation approaches, such as those of Rice and Silverman (1991) or Ramsay and Silverman (2005), could alternatively be used.

4.2. Functional canonical regression (FCR)

To obtain functional canonical correlations and the corresponding weight functions as needed for FCR, we adopt one of the methods proposed in He *et al.* (2004). In preliminary studies,

we determined that the eigenbase method as described there yielded the best performance for regression applications, with the Fourier base method a close second. Adopting the eigenbase method, the implementation of FCR is as follows:

(i) Starting with the eigenscore estimates as in (20), estimated raw functional canonical correlations $\tilde{\rho}_l$ and L-dimensional weight vectors $\hat{\mathbf{u}}_l$, $\hat{\mathbf{v}}_l$, $l = 1, \ldots, L$, are obtained by applying conventional numerical procedures of multivariate canonical analysis to the estimated eigenscore vectors $(\hat{\xi}_{i1}, \ldots, \hat{\xi}_{iL})'$ and $(\hat{\zeta}_{i1}, \ldots, \hat{\zeta}_{iL})'$. This works empirically well for moderately sized values of L, as typically obtained from automatic selectors.

(ii) Smooth weight function estimates $\tilde{u}_l(t)$, $\tilde{v}_l(t)$ are then obtained as

$$\tilde{u}_l(t) = \hat{\mathbf{u}}_l \hat{\boldsymbol{\theta}}(t), \qquad \tilde{v}_l(t) = \hat{\mathbf{v}}_l \hat{\boldsymbol{\varphi}}(t),$$

where $\hat{\boldsymbol{\theta}}(t) = (\hat{\theta}_1(t), \ldots, \hat{\theta}_L(t))'$, $\hat{\boldsymbol{\varphi}}(t) = (\hat{\varphi}_1(t), \ldots, \hat{\varphi}_L(t))'$.

(iii) The estimated regression parameter function $\hat{\beta}$ is obtained according to (16) by

$$\hat{\beta}(s, t) = \sum_{l=1}^{L} \tilde{\rho}_l \tilde{\mathbf{u}}_l(s) \int_{T_2} \hat{r}_{YY}(s, t) \tilde{v}_l(s) \, \mathrm{d}s,$$

where $\hat{r}_{YY}$ is an estimate of the covariance function of Y, obtained by two-dimensional smoothing of the empirical autocovariances of Y. This estimate is obtained as described in Yao *et al.* (2005a). Since the data are regularly sampled, the above integrals are easily obtained by the approximations $\sum_{j=1}^{m_y} \hat{r}_{YY}(t_j, t) \tilde{v}_l(t_j)(t_j - t_{j-1}), l = 1, \ldots, L$, with t_0 defined analogously to s_0 in (22) below.

(iv) Fitted/predicted processes

$$\hat{Y}_i(t) = \int_{T_1} \hat{\beta}(s, t) X_i(s) \, \mathrm{d}s \qquad \text{for } i = 1, \ldots, n, \tag{21}$$

are obtained, where the integral is again evaluated numerically by

$$\hat{Y}_i(t) = \sum_{j=1}^{N_x} \hat{\beta}(s_j, t) X_i(s_j)(s_j - s_{j-1}). \tag{22}$$

Here, s_0 is chosen such that $s_1 - s_0 = s_2 - s_1$.

This procedure depends on two tuning parameters, a bandwidth h for the smoothing steps (which are defined in detail, e.g., in Yao *et al.* (2005a)) and the number of canonical components L that are included. These tuning parameters may be determined by leave-one-out cross-validation (Rice and Silverman (1991)) as follows. With $\alpha = (h, L)$, the ith leave-one-out estimate for β is

$$\hat{\beta}_\alpha^{(-i)} = \sum_{l=1}^{L} \breve{\rho}_l^{(-i)} \tilde{u}_{h,l}^{(-i)}(s) \hat{R}_{YY}^{(-i)}(\tilde{v})_{h,l}^{(-i)}(t) \qquad \text{for } i = 1, \ldots, n, \tag{23}$$

where $\check{\rho}_l^{(-i)}$ is the lth canonical correlation, and $\tilde{u}_{h,l}^{(-i)}$ and $\hat{R}_{YY}^{(-i)}\tilde{v}_{h,l}^{(-i)}$ are the lth weight function untransformed and transformed with the covariance operator, respectively, all obtained while leaving out the data for the ith subject. Computation of these estimates follows steps (iii) and (iv) above, using tuning parameter $\alpha = (h, L)$, and omitting the ith pair of observed curves (X_i, Y_i). The average leave-one-out squared prediction error is then

$$PE_\alpha = \frac{1}{n} \sum_{i=1}^n \int_{T_2} \left(Y_i(t) - \int_{T_1} X_i(s)\hat{\beta}_\alpha^{(-i)}(s, t)\, ds \right)^2 dt. \tag{24}$$

The cross-validation procedure then selects the tuning parameter that minimizes the approximate average prediction error,

$$\hat{\alpha} = \arg\min_\alpha \tilde{PE}_\alpha,$$

where $\tilde{PE}_\alpha$ is obtained by replacing the integrals on the right-hand side of (24) by sums of the type (22).

4.3. Functional principal component regression (FPR)

Yao *et al.* (2005b) considered an implementation of functional linear regression whereby one uses functional principal component analysis for predictor and response functions separately, followed by simple linear regressions of the response principal component scores on the predictor scores. We adopt this approach as FPR.

Briefly, defining $\sigma_{mp} = E(\xi_m \zeta_p)$, this approach is based on representations

$$\beta(s, t) = \sum_{m=1}^\infty \sum_{p=1}^\infty \frac{\sigma_{mp}}{\lambda_{Xm}} \theta_m(s)\varphi_p(t)$$

of the regression parameter function $\beta(s, t)$, where

$$\sigma_{mp} = \int_{T_2} \int_{T_1} \theta_m(s) r_{XY}(s, t)\varphi_p(t)\, ds\, dt \tag{25}$$

for all m and p.

For estimation, one first obtains a smooth estimate $\hat{r}_{XY}$ of the cross-covariance r_{XY} by smoothing sample cross-covariances, for example, by the method described in Yao *et al.* (2005b). This leads to estimates $\hat{\sigma}_{mp}$ of σ_{mp}, $1 \le m, p \le L$, by plugging in estimates $\hat{r}_{XY}$ for r_{XY} and $\hat{\theta}_l, \hat{\varphi}_l$ for eigenfunctions θ_l, φ_l (as described in Section 4.2), in combination with approximating the integrals in (25) by appropriate sums. One may then use these estimates in conjunction with estimates $\hat{\lambda}_{Xm}$ of eigenvalues λ_{Xm} to arrive at the estimate $\hat{\beta}$ of the regression parameter function $\beta(s, t)$ given by

$$\hat{\beta}(s, t) = \sum_{m=1}^L \sum_{p=1}^L \frac{\hat{\sigma}_{mp}}{\hat{\lambda}_{Xm}} \hat{\theta}_m(s)\hat{\varphi}_p(t).$$

For further details about numerical implementations, we refer to Yao *et al.* (2005b).

# 5. Application to medfly mortality data

In this section, we present an application to age-at-death data that were collected for cohorts of male and female medflies in a biodemographic study of survival and mortality patterns of cohorts of male and female Mediterranean fruit flies (*Ceratitis capitata*; for details, see Carey *et al.* (2002)). A point of interest in this study is the relation of mortality trajectories between male and female medflies which were raised in the same cage. One specifically desires to quantify the influence of male survival on female survival. This is of interest because female survival determines the number of eggs laid and thus reproductive success of these flies. We use a sub-sample of the data generated by this experiment, comprising 46 cages of medflies, to address these questions. Each cage contains both a male and a female cohort, consisting each of approximately 4000 male and 4000 female medflies. These flies were raised in the shared cage from the time of eclosion. For each cohort, the number of flies alive at the beginning of each day was recorded, simply by counting the dead flies on each day; we confined the analysis to the first 40 days. The observed processes $X_i(t)$ and $Y_i(t)$, $t = 1, \ldots, 40$, $i = 1, \ldots, 46$, are the estimated random hazard functions for male and female cohorts, respectively. All deaths are fully observed so that censoring is not an issue. In a pre-processing step, cohort-specific hazard functions were estimated nonparametrically from the lifetable data, implementing the transformation approach described in Müller *et al.* (1997a).

A functional linear model was used to study the specific influence of male mortality on female mortality for flies that were raised in the same cage, with the hazard function of males as predictor process and that of females as response process. We applied both the proposed regression via canonical representation (FCR) and the more conventional functional regression based on principal components (FPR), implementing the estimation procedures described in the previous section. Tuning parameters were selected by cross-validation. Table 1 lists the average squared prediction error (PE) (24) obtained by the leave-one-out technique. For this application, the FCR procedure is seen to perform about 20% better than FPR in terms of PE.

The estimated regression parameter surface $\hat{\beta}(s, t)$ that is obtained for the FCR regression when choosing the cross-validated values for h and L, as given in Table 1, is shown in Figure 1. The shape of the regression surface indicates that female mortality at later ages is very clearly affected by male mortality throughout male lifespan, while female mortality at very early ages is not much influenced by male mortality. The effect of male mortality on female mortality is periodically elevated, as evidenced by the bumps visible in the surface. The particularly influential predictive periods are male mortality around days 10 and 20, which then has a particularly large influence on female mortality around days 15 and 25, that is, about five days later, and,

Table 1. Results for medfly data, comparing functional canonical regression (FCR) and functional principal component regression (FPR) with regard to average leave-one-out squared prediction error (PE) (24); values for bandwidth h and number L of components as chosen by cross-validation are also shown

|	h	L	PE
FCR	1.92	3	0.0100
FPR	1.65	3	0.0121
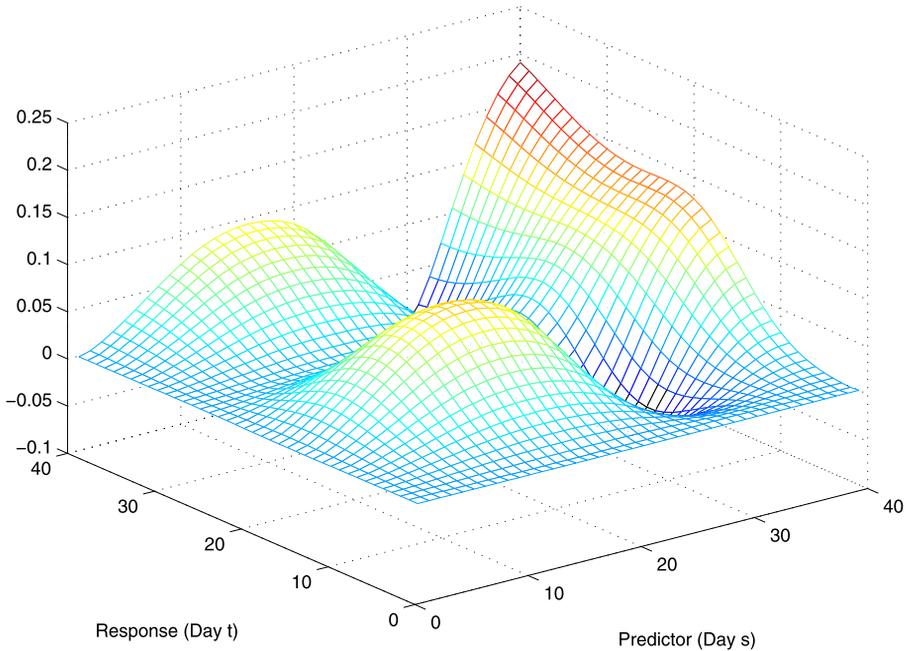
Figure 1. Estimated regression parameter surface obtained by functional canonical regression for the medfly study.

again, around days 35 and 40, judging from the locations of the peaks in the surface of $\hat{\beta}(s, t)$. In contrast, enhanced male mortality around day 30 leads to lessened female mortality throughout, while enhanced male mortality at age 40 is associated with higher older-age female mortality. These observations point to the existence of periodic waves of mortality, first affecting males and subsequently females. While some of the waves of increased male mortality tend to be associated with subsequently increased female mortality, others are associated with subsequently decreased female mortality.

These waves of mortality might be related to the so-called "vulnerable periods" that are characterized by locally heightened mortality (Müller *et al.* (1997b)). One such vulnerable period occurs around ages 10 and 20, and the analysis suggests that heightened male mortality during these phases is indicative of heightened female mortality. In contrast, heightened male mortality during a non-vulnerable period such as the time around 30 days seems to be associated with lower female mortality. A word of caution is in order as no inference methods are available to establish that the bumps observed in $\hat{\beta}(s, t)$ are real, so one cannot exclude the possibility that these bumps are enhanced by random fluctuations in the data.

Examples of observed, as well as predicted, female mortality trajectories for three randomly selected pairs of cohorts (male and female flies raised in the same cages) are displayed in Figure 2. The predicted female trajectories were constructed by applying both regression methods (FCR and FPR) with the leave-one-out technique. The prediction of an individual response tra-
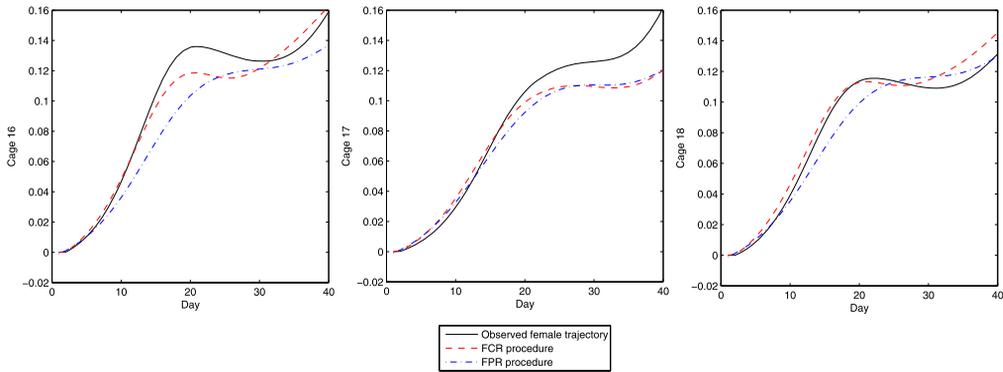
Figure 2. Functional regression of female (response) on male (predictor) medfly trajectories quantifying mortality in the form of cohort hazard functions for three cages of flies. Shown are actually observed female trajectories (solid) that are not used in the prediction, as well as the predictions for these trajectories obtained through estimation procedures based on functional principal component regression (FPR) (dash–dot) and on functional canonical regression (FCR) (dashed).

jectory from a predictor trajectory cannot, of course, be expected to be very close to the actually observed response trajectory, due to the extra random variation that is a large inherent component of response variability; this is analogous to the situation of predicting an individual response in the well-known simple linear regression case. Nevertheless, overall, FCR predictions are found to be closer to the target.

We note the presence of a "shoulder" at around day 20 for the three female mortality curves. This "shoulder" is related to the wave phenomenon visible in $\hat{\beta}(s, t)$ as discussed above and corresponds to a phase of elevated female mortality. The functional regression method based on FCR correctly predicts the shoulder effect and its overall shape in female mortality. At the rightmost points, for ages near 40 days, the variability of the mortality trajectories becomes large, posing extra difficulties for prediction in the right tail of the trajectories.

6. Additional results

Theorems 6.3 and 6.4 in this section provide a functional analog to the sums-of-squares decomposition of classical regression analysis. In addition, we provide two results characterizing the regression operators $\mathcal{L}_X$. We begin with two auxiliary results which are taken from He *et al.* (2003). The first of these characterizes the correlation operator between processes X and Y.

Lemma 6.1. *Assume that the L_2-processes X and Y satisfy Condition (C2). The correlation operator $R_{XX}^{-1/2} R_{XY} R_{YY}^{-1/2}$ can then be extended continuously to a Hilbert–Schmidt operator R on $L_2(T_2)$ to $L_2(T_1)$. Hence, $R_0 = R^* R$ is also a Hilbert–Schmidt operator with a countable number of non-zero eigenvalues and eigenfunctions $\{(\lambda_m, q_m)\}, m \geq 1, \lambda_1 \geq \lambda_2 \geq \cdots, p_m = Rq_m / \sqrt{\lambda_m}$. Then:*

(a) $\rho_m = \sqrt{\lambda_m}$, $u_m = R_{XX}^{-1/2} p_m$, $v_m = R_{YY}^{-1/2} q_m$ and both u_m and v_m are L_2-functions;
(b) $\mathrm{corr}(U_m, U_j) = \langle u_m, R_{XX} u_j \rangle = \langle p_m, p_j \rangle = \delta_{mj}$;
(c) $\mathrm{corr}(V_m, V_j) = \langle v_m, R_{XX} v_j \rangle = \langle q_m, q_j \rangle = \delta_{mj}$;
(d) $\mathrm{corr}(U_m, V_j) = \langle u_m, R_{XX} v_j \rangle = \langle p_m, R q_j \rangle = \rho_m \delta_{mj}$.

One of the main results in He *et al.* (2003) reveals that the L_2-processes X and Y can be expressed as sums of uncorrelated component functions and the correlation between the mth components of the expansion is the mth corresponding functional canonical correlation between the two processes.

Lemma 6.2 (Canonical decomposition). *Assume L_2-processes X and Y satisfy Condition* (C2). *There then exists a decomposition*:
  (a)

$$X = X_{c,K} + X_{c,K}^\perp, \qquad Y = Y_{c,K} + Y_{c,K}^\perp,$$

where

$$X_{c,K} = \sum_{j=1}^{K} U_j R_{XX} u_j, \qquad X_{c,K}^\perp = X - X_{c,K},$$

$$Y_{c,K} = \sum_{j=1}^{K} V_j R_{YY} v_j, \qquad Y_{c,K}^\perp = Y - Y_{c,K}.$$

The index K stands for canonical decomposition with K components, and U_j, V_j, u_j, v_j are as in Definition 3.1. *Here, (X, Y) and $(X_{c,K}, Y_{c,K})$ share the same first K canonical components, and $(X_{c,K}, Y_{c,K})$ and $(X_{c,K}^\perp, Y_{c,K}^\perp)$ are uncorrelated, that is,*

$$\mathrm{corr}(X_{c,K}, X_{c,K}^\perp) = 0, \qquad \mathrm{corr}(Y_{c,K}, Y_{c,K}^\perp) = 0,$$

$$\mathrm{corr}(X_{c,K}, Y_{c,K}^\perp) = 0, \qquad \mathrm{corr}(Y_{c,K}, X_{c,K}^\perp) = 0.$$

  (b) *Let $K \to \infty$ and $X_{c,\infty} = \sum_{m=1}^{\infty} U_m R_{XX} u_m$, $Y_{c,\infty} = \sum_{m=1}^{\infty} V_m R_{YY} v_m$. Then*

$$X = X_{c,\infty} + X_{c,\infty}^\perp, \qquad Y = Y_{c,\infty} + X_{c,\infty}^\perp,$$

where $X_{c,\infty}^\perp = X - X_{c,\infty}$, $Y_{c,\infty}^\perp = Y - Y_{c,\infty}$. Here, $(X_{c,\infty}, Y_{c,\infty})$ and (X, Y) share the same canonical components, $\mathrm{corr}(X_{c,\infty}^\perp, Y_{c,\infty}^\perp) = 0$, and $(X_{c,\infty}^\perp, Y_{c,\infty}^\perp)$ and $(X_{c,\infty}, Y_{c,\infty})$ are uncorrelated. Moreover, $X_{c,\infty}^\perp = 0$ if $\{p_m, m \geq 1\}$ forms a basis of the closure of the domain of R_{XX} and $Y_{c,\infty}^\perp = 0$ if $\{q_m, m \geq 1\}$ forms a basis of the closure of the domain of R_{YY}.

Since the covariance operators of L_2-processes are non-negative self-adjoint, they can be ordered as follows. The definitions of Y^*, Y_K^*, $Y_{c,\infty}$ are in (17), (18) and Lemma 6.2(b), respectively.

Theorem 6.3. *For $K \geq 1$, $R_{Y_K^* Y_K^*} \leq R_{Y^* Y^*} \leq R_{Y_{c,\infty} Y_{c,\infty}} \leq R_{YY}$.*

In multiple regression analysis, the ordering of the operators in Theorem 6.3 is related to the ordering of regression models in terms of a notion analogous to the regression sum of squares (SSR). The canonical regression decomposition provides information about the model in terms of its canonical components. Our next result describes the canonical correlations between observed and fitted processes. This provides an extension of the coefficient of multiple determination, $R^2 = \text{corr}(Y, \hat{Y})$, an important quantity in classical multiple regression analysis, to the functional case; compare also Yao *et al.* (2005b).

Theorem 6.4. *Assume that L_2-processes X and Y satisfy Condition (C2). The canonical correlations and weight functions for the pair of observed and fitted response processes (Y, Y^*) are then $\{(\rho_m, v_m, v_m/\rho_m); m \geq 1\}$ and the corresponding K-component (or ∞-component) canonical decomposition for Y^*, as defined in Lemma 6.2 for $K \geq 1$ and denoted here by $Y_{c,K}^*$ (or $Y_{c,\infty}^*$), is equivalent to the process Y_K^* or Y^* given in Theorem 3.4, that is,*

$$Y_{c,K}^* = Y_K^* = \sum_{m=1}^{K} \rho_m U_m R_{YY} v_m, \qquad K \geq 1, \qquad Y_{c,\infty}^* = Y^* = \sum_{m=1}^{\infty} \rho_m U_m R_{YY} v_m. \qquad (26)$$

We note that if Y is a scalar, then $R^2 = \rho_1$, and for a functional response Y, R^2 is replaced by the set $\{\rho_m, m \geq 1\}$.

The following two results serve to characterize the regression operator $\mathcal{L}_X$ defined in (4). They are used in the proofs provided in the following section.

Proposition 6.5. *The adjoint operator of $\mathcal{L}_X$ is $\mathcal{L}_X^* : L_2(T_2) \to L_2(T_1 \times T_2)$, where*

$$(\mathcal{L}_X^* z)(s, t) = X(s)z(t) \qquad \text{for } z \in L_2(T_2).$$

We have the following relation between the correlation operator Γ_{XX} defined in (5) and the regression operator $\mathcal{L}_X$.

Proposition 6.6. *The operator Γ_{XX} is a self-adjoint non-negative Hilbert–Schmidt operator and satisfies $\Gamma_{XX} = E[\mathcal{L}_X^* \mathcal{L}_X]$.*

7. Proofs

In this section, we provide sketches of proofs and some auxiliary results. We use tensor notation to define an operator $\theta \otimes \varphi : H \to H$,

$$(\theta \otimes \varphi)(h) = \langle h, \theta \rangle \varphi \qquad \text{for } h \in H.$$

Proof of Proposition 2.2. To prove (a) $\Rightarrow$ (b), we multiply equation (4) by X on both sides and take expected values to obtain $E(XY) = E(X\mathcal{L}_X \beta_0) + E(X\varepsilon)$. Equation (6) then follows from $E(XY) = r_{XY}$, $E(X\mathcal{L}_X \beta_0) = \Gamma_{XX}\beta_0$ (by Propositions 6.5 and 6.6) and $E(X\varepsilon) = 0$.

For (b) $\Rightarrow$ (c), let β_0 be a solution of equation (6). For any $\beta \in L_2(T_1 \times T_2)$, we then have $E\|Y - \mathcal{L}_X\beta\|^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 + E\|\mathcal{L}_X(\beta_0 - \beta)\|^2 + 2E[\langle Y - \mathcal{L}_X\beta_0, \mathcal{L}_X(\beta_0 - \beta)\rangle]$. Since

$$E\langle Y - \mathcal{L}_X\beta_0, \mathcal{L}_X(\beta_0 - \beta)\rangle$$
$$= E\langle \mathcal{L}_X^* Y - \mathcal{L}_X^* \mathcal{L}_X \beta_0, \beta_0 - \beta\rangle$$
$$= \langle E[\mathcal{L}_X^* Y] - E[\mathcal{L}_X^* \mathcal{L}_X \beta_0], \beta_0 - \beta\rangle = \langle r_{XY} - \Gamma_{XX}\beta_0, \beta_0 - \beta\rangle = 0,$$

by Proposition 6.6, we then have

$$E\|Y - \mathcal{L}_X\beta\|^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 + E\|\mathcal{L}_X(\beta_0 - \beta)\|^2 \geq E\|Y - \mathcal{L}_X\beta_0\|^2,$$

which implies that β_0 is indeed a minimizer of $E\|Y - \mathcal{L}_X\beta\|^2$.

For (c) $\Rightarrow$ (a), let

$$d^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 = \min_{\beta \in L_2(T_1 \times T_2)} E\|Y - \mathcal{L}_X\beta\|^2.$$

Then, for any $\beta \in L_2(T_1 \times T_2)$, $a \in \mathbf{R}$,

$$d^2 = E\|Y - \mathcal{L}_X\beta_0\|^2 \leq E\|Y - \mathcal{L}_X(\beta_0 + a\beta)\|^2$$
$$= E\|Y - \mathcal{L}_X\beta_0\|^2 - 2E\langle Y - \mathcal{L}_X\beta_0, \mathcal{L}_X(a\beta)\rangle + E\|\mathcal{L}_X(a\beta)\|^2$$
$$= d^2 - 2a\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle + a^2 E\|\mathcal{L}_X\beta\|^2.$$

Choosing $a = \langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle / E\|\mathcal{L}_X\beta\|^2$, it follows that $|\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle|^2 / E\|\mathcal{L}_X\beta\|^2 \leq 0$ and $\langle E[X(Y - \mathcal{L}_X\beta_0)], \beta\rangle = 0$. Since β is arbitrary, $E[X(Y - \mathcal{L}_X\beta_0)] = 0$ and therefore β_0 satisfies the functional linear model (4). $\qquad\square$

Proof of Theorem 2.3. Note, first, that $r_{XY}(s, t) = \sum_{m,j} E[\xi_m \zeta_j]\theta_m(s)\varphi_j(t)$. Thus, Condition (C1) is equivalent to $r_{XY} \in G_{XX}$. Suppose that a unique solution of (4) exists in $\ker(\Gamma_{XX})^\perp$. This solution is then also a solution of (6), by Proposition 2.2(b). Therefore, $r_{XY} \in G_{XX}$, which implies (C1). On the other hand, if (C1) holds, then $r_{XY} \in G_{XX}$, which implies that $\Gamma_{XX}^{-1} r_{XY} = \sum_m \lambda_{Xm}^{-1} \langle r_{XY}, \theta_m \varphi_j \rangle \theta_m \varphi_j$ is a solution of (6), is in $\ker(\Gamma_{XX})^\perp$ and, therefore, is the unique solution in $\ker(\Gamma_{XX})^\perp$ and also the unique solution of (4) in $\ker(\Gamma_{XX})^\perp$. $\qquad\square$

Proof of Proposition 2.4. The equivalence of (a), (b) and (c) follows from Proposition 2.2 and (d) $\Rightarrow$ (b) is a consequence of Theorem 2.3. We now prove (b) $\Rightarrow$ (d). Let β_0 be a solution of (6). Proposition 2.2 and Theorem 2.3 imply that both β_0 and β_0^* minimize $E\|Y - \mathcal{L}_X\beta\|^2$ for $\beta \in L_2(T_1 \times T_2)$. Hence, $E\|Y - \mathcal{L}_X\beta_0\|^2 = E\|Y - \mathcal{L}_X\beta_0^*\|^2 + E\|\mathcal{L}_X(\beta_0^* - \beta_0)\|^2 + 2E\langle Y - \mathcal{L}_X\beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_0)\rangle$, which, by Proposition 6.6, implies that $2E\langle \mathcal{L}_X^*(Y - \mathcal{L}_X\beta_0^*), \beta_0^* - \beta_0\rangle = 2\langle r_{XY} - \Gamma_{XX}\beta_0^*, \beta_0^* - \beta_0\rangle = 0$. Therefore, $E\|\mathcal{L}_X(\beta_0^* - \beta_0)\|^2 = \|\Gamma_{XX}^{1/2}(\beta_0^* - \beta_0)\|^2 = 0$. It follows that $\beta_0^* - \beta_0 \in \ker(\Gamma_{XX})$, or $\beta_0 = \beta_0^* + h$, for an $h \in \ker(\Gamma_{XX})$. $\qquad\square$

Proof of Theorem 3.2. According to Lemma 6.2(b), Condition (C2) guarantees the existence of the canonical components and canonical decomposition of X and Y. Moreover,

$$r_{XY}(s,t) = E[X(s)Y(t)] = E\big[\big(X_{c,\infty}(s) + X_{c,\infty}^{\perp}(s)\big)\big(Y_{c,\infty}(t) + Y_{c,\infty}^{\perp}(t)\big)\big]$$

$$= E[X_{c,\infty}(s)Y_{c,\infty}(t)] = E\left[\sum_{m=1}^{\infty} U_m R_{XX} u_m(s) \sum_{m=1}^{\infty} V_m R_{YY} v_m(t)\right]$$

$$= \sum_{m,j=1}^{\infty} E[U_m V_j] R_{XX} u_m(s) R_{YY} v_m(t) = \sum_{m=1}^{\infty} \rho_m R_{XX} u_m(s) R_{YY} v_m(t).$$

We now show that the exchange of the expectation with the summation above is valid. From Lemma 6.1(b), for any $K > 0$ and the spectral decomposition $R_{XX} = \sum_m \lambda_{Xm} \theta_m \otimes \theta_m$,

$$\sum_{m=1}^{K} E\|U_m R_{XX} u_m\|^2 = \sum_{m=1}^{K} E[U_m^2] \|R_{XX}^{1/2} p_m\|^2 = \sum_{m=1}^{K} \langle p_m, R_{XX} p_m \rangle$$

$$= \sum_{m=1}^{K} \sum_{j=1}^{\infty} \lambda_{Xj} \langle p_m, \theta_j \rangle^2 = \sum_{j=1}^{\infty} \lambda_{Xj} \left(\sum_{m=1}^{K} \langle p_m, \theta_j \rangle^2\right)$$

$$\leq \sum_{j=1}^{\infty} \lambda_{Xj} \|\theta_j\|^2 = \sum_{j=1}^{\infty} \lambda_{Xj} < \infty,$$

where the inequality follows from the fact that $\sum_{m=1}^{K} \langle p_m, \theta_j \rangle^2$ is the square length of the projection of θ_j onto the linear subspace spanned by $\{p_1, \ldots, p_K\}$. Similarly, we can show that for any $K > 0$,

$$\sum_{m=1}^{K} E\|V_m R_{YY} v_m\|^2 < \sum_{j=1}^{\infty} \lambda_{Yj} < \infty.$$ □

Proof of Theorem 3.3. Note that Condition (C2) implies Condition (C1). Hence, from Theorem 2.3, $\beta_0^* = \Gamma_{XX}^{-1} r_{XY}$ exists and is unique in $\ker(\Gamma_{XX})^{\perp}$. We can show (16) by applying Γ_{XX}^{-1} to both sides of (6), exchanging the order of summation and integration. To establish (17), it remains to show that

$$\sum_{m=1}^{\infty} \|\rho_m u_m R_{YY} v_m\|^2 < \infty, \tag{27}$$

where $u_m R_{YY} v_m(s,t) = u_m(s) R_{YY} v_m(t)$ in $L_2(T_1 \times T_2)$. Note that

$$\rho_m u_m = \rho_m R_{XX}^{1/2} p_m = R_{XX}^{1/2} R q_m = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_{Xj}}} \langle R q_m, \theta_j \rangle \theta_j,$$

where the operator $R = R_{XX}^{1/2} R_{XY} R_{YY}^{1/2}$ is defined in Lemma 6.1 and can be written as $R = \sum_{k,\ell} r_{k\ell} \varphi_k \otimes \theta_\ell$, with $r_{km} = E[\xi_k \zeta_\ell]/\sqrt{\lambda_{Xk}\lambda_{Y\ell}}$, using the Karhunen–Loève expansion (7). Then,

$$Rq_m = \sum_{k,\ell} r_{k\ell} \langle \varphi_k, q_m \rangle \theta_\ell, \qquad \langle Rq_m, \theta_j \rangle = \sum_k r_{kj} \langle \varphi_k, q_m \rangle$$

and, therefore,

$$\sum_m \| \rho_m u_m R_{YY} v_m \|^2$$

$$\leq \sum_m \| \rho_m u_m \|^2 \| R_{YY} v_m \|^2 = \sum_m \left[\sum_j \frac{1}{\lambda_{Xj}} (\langle Rq_m, \theta_j \rangle)^2 \right] \| R_{YY} v_m \|^2$$

$$= \sum_m \left[\sum_j \frac{1}{\lambda_{Xj}} \left\{ \sum_k r_{kj} \langle \varphi_k, q_m \rangle \right\}^2 \right] \| R_{YY} v_m \|^2$$

$$\leq \sum_m \left[\sum_j \frac{1}{\lambda_{Xj}} \sum_k r_{kj}^2 \sum_\ell \langle \varphi_\ell, q_m \rangle^2 \right] \| R_{YY} v_m \|^2$$

$$= \left[\sum_j \frac{1}{\lambda_{Xj}} \sum_k r_{kj}^2 \right] \sum_m \left[\sum_\ell \langle \varphi_\ell, q_m \rangle^2 \right] \| R_{YY} v_m \|^2$$

$$= \sum_{j,k} \frac{r_{kj}^2}{\lambda_{Xj}} \sum_m \| R_{YY} v_m \|^2 \qquad \text{as } \| q_m \| = 1.$$

Note that by (C2), the first sum on the right-hand side is bounded. For the second sum,

$$\sum_m \| R_{YY} v_m \|^2 = \sum_m \| R_{YY}^{1/2} q_m \|^2 = \sum_m \langle q_m, R_{YY} q_m \rangle = \sum_m \sum_j \lambda_{Yj} \langle q_m, \varphi_j \rangle^2$$

$$= \sum_j \lambda_{Yj} \sum_m \langle q_m, \varphi_j \rangle^2 \leq \sum_j \lambda_{Yj} \| \varphi_j \|^2 \leq \sum_j \lambda_{Yj} < \infty,$$

which implies (27). $\square$

Proof of Theorem 3.4. Observing

$$Y_K^* = \mathcal{L}_X \beta_K^* = \sum_{m=1}^K \rho_m \mathcal{L}_X(u_m) R_{YY} v_m$$

$$= \sum_{m=1}^K \rho_m \langle u_m, X \rangle R_{YY} v_m = \sum_{m=1}^K \rho_m U_m R_{YY} v_m,$$

$$E\|Y^* - Y_K^*\|^2 = E\left\|\sum_{m=K+1}^{\infty} \rho_m U_m R_{YY} v_m\right\|^2 = \sum_{m=K+1}^{\infty} \rho_m \|R_{YY} v_m\|^2 \quad \text{and}$$

$$E\|\mathcal{L}_X \beta_K^*\|^2 = E\left\|\sum_{m=1}^{\infty} \rho_m U_m R_{YY} v_m\right\|^2$$

$$= \sum_{m,j=1}^{\infty} \rho_m \rho_j E[U_m U_j]\langle R_{YY} v_m, R_{YY} v_j\rangle = \sum_{m=1}^{\infty} \rho_m^2 \|R_{YY} v_m\|^2 < \infty,$$

we infer that $E\|Y^* - Y_K^*\|^2 \to 0$ as $K \to \infty$. From $E[U_m] = 0$, for $m \geq 1$, we have $E[Y_K^*] = 0$ and, moreover,

$$E\|Y - Y_K^*\|^2 = E\|(Y - \mathcal{L}_X \beta_0^*) + \mathcal{L}_X(\beta_0^* - \beta_K^*)\|^2$$

$$= E\|Y - \mathcal{L}_X \beta_0^*\|^2 + E\|\mathcal{L}_X(\beta_0^* - \beta_K^*)\|^2 + 2E\langle Y - \mathcal{L}_X \beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_K^*)\rangle.$$

Since $E\|Y - \mathcal{L}_X \beta_0^*\|^2 = \text{trace}(R_{YY}) - E\|\mathcal{L}_X \beta_0^*\|^2$ and as β_0^* is the solution of the normal equation (6), we obtain $E\langle Y - \mathcal{L}_X \beta_0^*, \mathcal{L}_X(\beta_0^* - \beta_K^*)\rangle = E\langle \mathcal{L}_X^*(Y - \mathcal{L}_X \beta_0^*), \beta_0^* - \beta_K^*\rangle = 0$. Likewise,

$$E\|\mathcal{L}_X(\beta_0^* - \beta_K^*)\|^2 = \sum_{m=K+1}^{\infty} \rho_m^2 \|R_{YY} v_m\|^2,$$

implying (19). □

Proof of Theorem 6.3. From (17), (18) for any $K \geq 1$,

$$R_{Y^*Y^*} - R_{Y_K^* Y_K^*} = R_{YY}^{1/2}\left[\sum_{m=K+1}^{\infty} \rho_m^2 q_m \otimes q_m\right] R_{YY}^{1/2} = R_{YY}^{1/2} R_{K+1}^* R_{K+1} R_{YY}^{1/2},$$

where $R_{K+1} = \text{Proj}_{\text{span}\{q_m, m \geq K+1\}} R$ and hence, $R_{Y^*Y^*} - R_{Y_K^* Y_K^*} \geq 0$. Note that

$$r_{Y_{c,\infty} Y_{c,\infty}}(s, t) = E[Y_{c,\infty}(s) Y_{c,\infty}(t)] = \sum_{m,j=1}^{\infty} E[V_m V_j] R_{YY} v_m(s) R_{YY} v_j(t)$$

$$= \sum_{m=1}^{\infty} R_{YY} v_m(s) R_{YY} v_j(t) = \sum_{m=1}^{\infty} R_{YY}^{1/2}(q_m)(s) R_{YY}^{1/2}(q_m)(t),$$

implying that

$$R_{Y_{c,\infty} Y_{c,\infty}} - R_{Y^*Y^*} = R_{YY}^{1/2}\left[\sum_{m=1}^{\infty}(1 - \rho_m^2) q_m \otimes q_m\right] R_{YY}^{1/2} \geq 0.$$

Finally, from Lemma 6.2(b), we have $Y = Y_{c,\infty} - Y_{c,\infty}^{\perp}$, therefore $r_{YY} = r_{Y_{c,\infty} Y_{c,\infty}} + r_{Y_{c,\infty}^{\perp} Y_{c,\infty}^{\perp}}$. This leads to $r_{YY} - r_{Y_{c,\infty} Y_{c,\infty}} = r_{Y_{c,\infty}^{\perp} Y_{c,\infty}^{\perp}}$ and $R_{YY} - R_{Y_{c,\infty} Y_{c,\infty}} = R_{Y_{c,\infty}^{\perp} Y_{c,\infty}^{\perp}} \geq 0$. $\qquad\square$

We need the following auxiliary result to prove Theorem 6.3. We call two L_2-processes X and Y *uncorrelated* if and only if $E[\langle u, X\rangle\langle v, Y\rangle] = 0$ for all L_2-functions u and v.

Lemma 7.1. $Y_{c,\infty}^{\perp}$ *and* Y^* *are uncorrelated.*

Proof. For any $\tilde{u}, \tilde{v} \in L_2(T_2)$, write $\tilde{v} = \tilde{v}_1 + \tilde{v}_2$, with $R_{YY}^{1/2}\tilde{v}_1 \in \text{span}\{q_m; m \geq 1\}$, which is equivalent to $\tilde{v}_1 \in \text{span}\{v_m; m \geq 1\}$ and $R_{YY}^{1/2}\tilde{v}_2 \in \text{span}\{q_m; m \geq 1\}^{\perp}$. Then

$$\langle \tilde{v}_2, Y^*\rangle = \sum_{m=1}^{\infty} \rho_m U_m \langle \tilde{v}_2, R_{YY} v_m\rangle = \sum_{m=1}^{\infty} \rho_m U_m \langle R_{YY}^{1/2}\tilde{v}_2, q_m\rangle = 0.$$

With $\tilde{v}_1 = \sum_m a_m v_m$, write $\langle \tilde{v}, Y^*\rangle = \langle \tilde{v}_1, Y^*\rangle = \sum_{m,j} a_m \rho_j U_j \langle v_m, R_{YY} v_j\rangle = \sum_m a_m \rho_m U_m$. Furthermore, from Lemma 6.2(b), $E[U_m \langle \tilde{u}, Y_{c,\infty}^{\perp}\rangle] = 0$ for all $m \geq 1$. We conclude that $E[\langle \tilde{u}, Y_{c,\infty}^{\perp}\rangle\langle \tilde{v}, Y^*\rangle] = 0$. $\qquad\square$

Proof of Theorem 6.4. Calculating the covariance operators for (Y, Y^*),

$$r_{Y^*Y^*}(s,t) = E[Y^*(s)Y^*(t)] = \sum_{m,j} \rho_m \rho_j E[U_m U_j] R_{YY} u_m(s) R_{YY} v_j(t)$$

$$= \sum_m \rho_m^2 R_{YY} u_m(s) R_{YY} v_m(t) = \sum_m \rho_m^2 R_{YY}^{1/2} q_m(s) R_{YY}^{1/2} q_m(t)$$

so that

$$R_{Y^*Y^*} = \sum_m \rho_m^2 R_{YY}^{1/2}[q_m \otimes q_m]R_{YY}^{1/2} = R_{YY}^{1/2}\left[\sum_m \rho_m^2 q_m \otimes q_m\right]R_{YY}^{1/2} = R_{YY}^{1/2} R_0 R_{YY}^{1/2}.$$

Now, from Lemmas 6.2 and 7.1,

$$r_{YY^*}(s,t) = E[Y(s)Y^*(t)] = E\left[\left(Y_{c,\infty}(s) + Y_{c,\infty}^{\perp}(s)\right)Y^*(t)\right]$$

$$= E[Y_{c,\infty}(s)Y^*(t)] = E\left[\sum_m V_m R_{YY} v_m(s) \sum_j \rho_j U_j R_{YY} v_j(t)\right]$$

$$= \sum_{m,j} E[V_m U_j \rho_j R_{YY} v_m(s) R_{YY} v_j(t)]$$

$$= \sum_m \rho_m^2 R_{YY} v_m(s) R_{YY} v_j(t) = r_{Y^*Y^*}(s,t).$$

Hence, $R_{YY^*} = R_{Y^*Y^*}$. The correlation operator for (Y, Y^*) is $\tilde{R} = R_{YY}^{-1/2} R_{YY^*} R_{Y^*Y^*}^{-1/2} = R_{YY}^{-1/2} R_{Y^*Y^*}^{1/2}$ with $\tilde{R}\tilde{R}^* = R_{YY}^{-1/2} R_{Y^*Y^*} R_{YY}^{-1/2} = R_0$. Hence, $\tilde{\rho}_m = \rho_m$, $\tilde{p}_m = q_m$ and $\tilde{q}_m =$

$\tilde{R}^* \tilde{p}_m / \tilde{\rho}_m = R_{Y^*Y^*}^{1/2} R_{YY}^{-1/2} q_m / \rho_m = R_{Y^*Y^*}^{1/2} v_m / \rho_m$. Moreover, $\tilde{u}_m = R_{YY}^{-1/2} \tilde{p}_m = R_{YY}^{-1/2} q_m = v_m$ and $\tilde{v}_m = R_{Y^*Y^*}^{-1/2} \tilde{q}_m = R_{Y^*Y^*}^{-1/2} R_{Y^*Y^*}^{1/2} v_m / \rho_m = v_m / \rho_m$. Note that $Y_{c,\infty}^* = \sum_m \tilde{V}_m R_{Y^*Y^*} \tilde{v}_m$ with

$$\tilde{V}_m = \langle \tilde{v}_m, Y^* \rangle = \left\langle v_m / \rho_m, \sum_j \rho_j U_j R_{YY} v_j \right\rangle = \sum_j U_j \langle v_m, R_{YY} v_j \rangle = U_m,$$

$$R_{Y^*Y^*} \tilde{v}_m = R_{YY}^{1/2} R_0 R_{YY}^{1/2} v_m / \rho_m = R_{YY}^{1/2} R_0 q_m / \rho_m = \rho_m R_{YY}^{1/2} q_m = \rho_m R_{YY} v_m.$$

Substituting into the equation on the left-hand side of (26), one obtains the equation on the right-hand side of (26). □

Proof of Proposition 6.5. From the definition, $\mathcal{L}_X^*$ must satisfy $\langle \mathcal{L}_X \beta, z \rangle = \langle \beta, \mathcal{L}_X^* z \rangle$ for $\beta \in L_2(T_1 \times T_2)$ and $z \in L_2(T_2)$. Note that $\langle \mathcal{L}_X \beta, z \rangle = \int_{T_2} (\mathcal{L}_X \beta)(t) z(t) \, dt = \int_{T_2} \int_{T_1} X(s) \beta(s,t) z(t) \, ds \, dt$ and $\langle \beta, \mathcal{L}_X^* z \rangle = \int \int_{T_1 \times T_2} \beta(s,t) (\mathcal{L}_X^* z)(s,t) \, ds \, dt$. For the differences, we obtain $\int \int \beta(s,t) [X(s) z(t) - (\mathcal{L}_X^* z)(s,t)] \, ds \, dt = 0$ for arbitrary $\beta \in L_2(T_1 \times T_2)$ and $z \in L_2(T_2)$. This implies that $(\mathcal{L}_X^* z)(s,t) = X(s) z(t)$. □

Proof of Proposition 6.6. By Proposition 6.5, $\Gamma_{XX} = E[\mathcal{L}_X^* \mathcal{L}_X]$. Since the integral operator Γ_{XX} has the L_2-integral kernel r_{XX}, it is a Hilbert–Schmidt operator (Conway (1985)). Moreover, for $\beta_1, \beta_2 \in L_2(T_1 \times T_2)$,

$$\langle \Gamma_{XX} \beta_1, \beta_2 \rangle = \int \int (\Gamma_{XX} \beta_1)(s,t) \beta_2(s,t) \, ds \, dt = \int \int \int r_{XX}(s,w) \beta_1(w,t) \beta_2(s,t) \, dw \, ds \, dt,$$

$$\langle \beta_1, \Gamma_{XX} \beta_2 \rangle = \int \int \beta_1(s,t) (\Gamma_{XX} \beta_2(s,t)) \, ds \, dt = \int \int \int \beta_1(w,t) r_{XX}(s,w) \beta_2(s,t) \, dw \, ds \, dt,$$

implying that Γ_{XX} is self-adjoint. Furthermore, Γ_{XX} is non-negative definite because, for arbitrary $\beta \in L_2(T_1 \times T_2)$,

$$\langle \Gamma_{XX} \beta, \beta \rangle = \int \int \int E[X(s) X(w)] \beta(w,t) \beta(s,t) \, dw \, ds \, dt$$

$$= E\left[\int (\mathcal{L}_X \beta)(t)(\mathcal{L}_X \beta)(t) \, dt \right] = E\|\mathcal{L}_X \beta\|^2 \geq 0.$$ □

Acknowledgements

References

Ash, R.B. and Gardner, M.F. (1975). *Topics in Stochastic Processes*. New York: Academic Press. MR0448463

Brillinger, D.R. (1985). *Time Series: Data Analysis and Theory*, 2nd ed. Holden Day: San Francisco. MR0595684

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.* **92** 24–41. MR2102242

Carey, J.R., Liedo, P., Orozco, D. and Vaupel, J.W. (2002). Slowing of mortality rates at older ages in large medfly cohorts. *Science* **258** 457–461.

Chiou, J.M., Müller, H.G. and Wang, J.L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc. Ser. B* **65** 405–423. MR1983755

Conway, J.B. (1985). *A Course in Functional Analysis*. New York: Springer. MR0768926

Cupidon, J., Gilliam, D.S., Eubank, R. and Ruymgaart, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli* **13** 1179–1194. MR2364231

Eubank, R. and Hsing, T. (2008). Canonical correlation for stochastic processes. *Stochastic Process. Appl.* **118** 1634–1661. MR2442373

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, J. and Zhang, J.T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. Roy. Statist. Soc. Ser. B* **62** 303–322.

Faraway, J.J. (1997). Regression analysis for a functional response. *Technometrics* **39** 254–262. MR1462586

Hall, P. and Horowitz, J. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

Hannan, E.J. (1961). The general theory of canonical correlation and its relation to functional analysis. *J. Aust. Math. Soc.* **2** 229–242. MR0166869

He, G., Müller, H.G. and Wang, J.L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.* **85** 54–77. MR1978177

He, G., Müller, H.G. and Wang, J.L. (2004). Methods of canonical analysis for functional data. *J. Statist. Plann. Inference* **122** 141–159. MR2057919

Hoover, D.R., Rice, J., Wu, C.O. and Yang, L. (1998). Nonparametric smoothing estimates of time varying coefficient models with longitudinal data. *Biometrika* **85** 809–740. MR1666699

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.

Leurgans, S.E., Moyeed, R.A. and Silverman, B.W. (1993). Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. Ser. B* **55** 725–740. MR1223939

Müller, H.G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159

Müller, H.G., Wang, J.L. and Capra, W.B. (1997a). From lifetables to hazard rates: The transformation approach. *Biometrika* **84** 881–892. MR1624976

Müller, H.G., Wang, J.L., Capra, W.B., Liedo, P. and Carey, J.R. (1997b). Early mortality surge in protein-deprived females causes reversal of sex differential of life expectancy in Mediterranean fruit flies. *Proc. Natl. Acad. Sci. USA* **94** 2762–2765.

Ramsay, J. and Dalzell, C.J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B* **53** 539–572. MR1125714

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. New York: Springer. MR2168993

Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

Rice, J. and Wu, C.O. (2000). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259. MR1833314

Ritz, C. and Streibig, J.C. (2009). Functional regression analysis of fluorescence curves. *Biometrics* **65** 609–617.

Shi, M., Weiss, R.E. and Taylor, J.M.G. (1996). An analysis of paediatric CD4 counts for Acquired Immune Deficiency Syndrome using flexible random curves. *Appl. Statist.* **45** 151–163.

Yao, F., Müller, H.G., Wang, J.L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

Yao, F., Müller, H.G. and Wang, J.L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106