Asymptotics in Statistics and Probability, pp. 1–14
M.L. Puri (Ed.)
2000 VSP

EXTENDING CORRELATION AND REGRESSION FROM MULTIVARIATE TO FUNCTIONAL DATA *

G. HE [1], H. G. MÜLLER [2] and J. L. WANG [2]

[1] *Biometrics Unit, California Department of Fish and Game, Sacramento, CA 95814*
[2] *Division of Statistics, University of California, Davis, CA 95616-8705*

ABSTRACT

In this paper, we discuss concepts and methods of functional data analysis. The focus is on the case where a pair of random functions is sampled per experimental unit. We discuss the quantification of the dependency between pairs of random functions by means of functional canonical analysis and linear modeling for L_2-processes. Basic concepts of multivariate analysis are extended to the domain of functional data analysis and the conditions under which such an extension is feasible are discussed. Our main results demonstrate how basic properties of canonical correlation and linear regression known from multivariate statistical analysis can be restated for functional data, if appropriate conditions are satisfied.

1. INTRODUCTION

In many experiments, the observations consist of a sample of random functions or curves. From the functional data analysis point of view, each curve corresponds to one observation. This is an extension of multivariate data analysis where observations consist of vectors of finite dimension. In multivariate analysis, it is customary to use the spectral decomposition of the covariance matrix of a random vector for principal components analysis, the singular value decomposition of the cross-covariance matrix between a pair of random vectors for canonical correlations, and least squares for the multivariate linear model (Anderson, 1984).

*This research was supported in part by NSF grants DMS 98-03627 and DMS 99-71602.

While these notions are fundamental in multivariate data analysis in order to study dependency and linear relations between random vectors, their extensions to functional data are not so obvious and require tools from functional analysis. A basic problem in infinite-dimensional space is the inversion of linear operators. Attempts to extend concepts of multivariate data analysis to functional data analysis include the work of Dauxois, Pousse and Romain (1982), who described the extension of classical multivariate principal components analysis to the functional context and investigated properties of the maximum variance criterion, as well as Rice and Silverman (1991), who proposed a method to study the covariance structure of a sample of random curves by estimating its eigenvalues and smoothed eigenfunctions. Ramsay and Dalzell (1991) provided tools for studying principal components and linear models for a sample of random curves, using L-splines, derived from a linear differential operator. Leurgans, Moyeed, and Silverman (1993) applied smoothing splines with a special penalty term to functional canonical correlation analysis. Ramsay and Silverman (1997) provide an overview on current methods for functional analysis, with emphasis on various versions of smoothing splines and base function expansions.

By using the Karhunen-Loève decomposition (Ash and Gardner, 1985), it is relatively straightforward to extend multivariate principal components analysis from random vectors to random functions. However, as will be seen in Section 2, such a straightforward extension from multivariate data analysis to functional data analysis does not exist for canonical correlation and regression. One main obstacle for this extension involves the inversion of covariance operators. In contrast to the situation in the multivariate case, this inversion is not feasible in infinite dimensional Hilbert spaces. In this sense, canonical correlation is an inverse problem (He, 1999). In this paper, we derive properties of canonical analysis and linear modeling for random processes, extending properties which are well known for the multivariate case. In Section 2, notation and basic assumptions as well as canonical functional linear models are introduced for square integrable stochastic processes. The main result for functional canonical analysis is Theorem 3.1 of Section 3. Issues regarding the properties of functional linear models are discussed in Section 4, and the main results are Theorems 4.3 and 4.4.

2. PRELIMINARIES

We introduce here notations and assumptions.

In multivariate analysis, the canonical correlation between two random vectors, $\boldsymbol{X} \in \mathbf{R}^{\mathbf{p}}$, $\boldsymbol{Y} \in \mathbf{R}^{\mathbf{q}}$, can be defined as follows:

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 3 (169-248)

For $1 \leqslant k \leqslant \min(p, q)$, the kth canonical correlation ρ_k, and weight vectors u_k and v_k, satisfy

$$\rho_k = \sup_{\mathbf{u} \in \mathbf{R}^\mathbf{p},\ \mathbf{v} \in \mathbf{R}^\mathbf{q}} \mathrm{Corr}(\boldsymbol{u}^T \boldsymbol{X}, \boldsymbol{v}^T \boldsymbol{Y}) = \mathrm{Corr}(\boldsymbol{u}_k^T \boldsymbol{X}, \boldsymbol{v}_k^T \boldsymbol{Y}),$$

and, in addition, for $k > 1$, the requirement that the canonical variable pair $(\boldsymbol{u}_k^T \boldsymbol{X}, \boldsymbol{v}_k^T \boldsymbol{Y})$ is uncorrelated to all the previous $k-1$ canonical variable pairs.

A multivariate linear regression can be defined as

$$\boldsymbol{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}_0^T \boldsymbol{X} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{\epsilon} \in \mathbf{R}^\mathbf{q}$, with $E[\boldsymbol{\epsilon}] = 0$, and $\boldsymbol{\beta}_0 \in \mathbf{R}^{\mathbf{p}\times\mathbf{q}}$ is the parameter matrix.

We will discuss here how to extend canonical correlation and linear regression from finite dimensions to the infinite dimensional case. The random processes we consider here will be assumed to be square integrable, *i.e.*, to be in the L_2 space of square integrable functions. For a stochastic process with support T, on a probability space Ω, $X(t) = \{X(t, \omega); \omega \in \Omega, t \in T\}$, it holds that $X \in L_2(T)$, if $E \int_T X^2(t)dt < \infty$. For convenience, we will always assume that T and T_1, T_2 below are compact intervals. We note that $L_2(T)$ is a Hilbert space if equipped with the inner product $\langle f, g\rangle = \int_T f(t)g(t)dt$, for $f, g \in L_2(T)$, where dt is the Lebesgue measure. The results can be easily extended to cover more general measures μ and scalar products in spaces $L_2(T; \mu)$.

Functional Canonical Correlation Analysis

Extending the concept of canonical correlation to L_2-processes, we define:

Let $X \in L_2(T_1)$, $Y \in L_2(T_2)$ be L_2-processes. For $k \geqslant 1$, the kth canonical correlation and weight functions, ρ_k, u_k, v_k, satisfy

$$\rho_k = \sup_{u \in L_2(T_1), v \in L_2(T_2)} \mathrm{Corr}(\langle u, X\rangle, \langle v, Y\rangle) = \mathrm{Corr}(\langle u_k, X\rangle, \langle v_k, Y\rangle), \tag{2.2}$$

where the canonical variates are

$$U_k = \langle u_k, X\rangle, V_k = \langle v_k, Y\rangle,$$

and in addition for $k > 1$, the canonical variable pair (U_k, V_k) is uncorrelated with all previous $(k-1)$ canonical variable pairs.

In multivariate analysis, the maximization of (2.2) is equivalent to solving for the eigenvalues of the *cross-correlation matrix*,

$$R = R_{XX}^{-1/2} R_{XY} R_{YY}^{-1/2}, \tag{2.3}$$

where R_{XX} and R_{YY}, the covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively, are assumed to be invertible, and R_{XY} is the cross-covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Y}$ (Anderson, 1984). In the infinite-dimensional case, covariance matrices are

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 4 (248-312)

generalized to *covariance operators*. Specifically, the covariance operator $R_{XX}\colon L_2(T_1) \to L_2(T_1)$, is given by

$$R_{XX}u(s) = \int_{T_1} r_{XX}(s,t)u(t)dt, \quad u \in L_2(T_1). \tag{2.4}$$

where

$$r_{XX}(s,t) = \mathrm{Cov}\big[X(s), X(t)\big], \quad s,t \in T_1,$$

is the covariance function of process X. Similarly we can define covariance functions

$$r_{YY}(s,t) = \mathrm{cov}\big[Y(s), Y(t)\big], \quad s,t \in T_2,$$

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

The covariance operators$R_{YY}\colon L_2(T_2) \to L_2(T_2)$, $R_{XY}\colon L_2(T_2) \to L_2(T_1)$, and $R_{YX}\colon L_2(T_1) \to L_2(T_2)$ are defined in complete analogy to R_{XX}. We note that we do not assume stationarity of the processes as in Brillinger (1975, ch. 10) and that we are aiming at quantifying the dependency between pairs of processes. Thus our target is different from that of a body of work, exemplified by Roussas (1990), where the impact of dependency in a sequence of low-dimensional data on nonparametric regression and related approaches is considered.

One of the basic problems which sets infinite-dimensional data analysis apart from multivariate statistical analysis is that the covariance operators are not invertible. The reason is that a covariance operator of an L_2-process is a compact operator, which is not invertible in an infinite dimensional Hilbert space. One can indeed show that canonical weight functions do not necessarily exist for a given L_2-process (He, 1999). It is therefore of interest to provide a sufficient condition for the existence of canonical correlations and canonical weight functions for L_2-processes. This is the purpose of Condition 2.1 below.

Using the Karhunen-Loève decomposition, X and Y may be expanded as

$$\begin{aligned} X(s) &= E\big[X(s)\big] + \sum_{i=1}^{\infty} \xi_i\theta_i(s), \quad s \in T_1, \\ Y(t) &= E\big[Y(t)\big] + \sum_{i=1}^{\infty} \zeta_i\phi_i(t), t \in T_2, \end{aligned} \tag{2.5}$$

with a sequence of uncorrelated random variables ξ_i with $E(\xi_i) = 0$, and a sequence of uncorrelated random variables ζ_i with $E(\zeta_i) = 0$. Here, $\lambda_{Xi} = E[\xi_i^2]$, $\lambda_{Yi} = E[\zeta_i^2]$, $\sum_{i=1}^{\infty} \lambda_{X_i} < \infty$, $\Sigma_{i=1}^{\infty}\lambda_{Y_i} < \infty$, and $\{(\lambda_i, \theta_i)\}$, $\{(\zeta_j, \phi_j)\}$ are the eigenvalues and eigenfunctions of the covariance operators

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 5 (312-393)

R_{XX} and R_{YY}. The following condition refers to expansion (2.5) and, as will be seen in Theorem 3.1 below, ensures that functional canonical correlation is well defined.

Condition 2.1. L_2-processes X and Y satisfy

$$(a) \quad \sum_{i,j=1}^{\infty} \frac{E^2[\xi_i \zeta_i]}{\lambda_{Xi}^2 \lambda_{Yj}} < \infty, \quad or \quad (b) \quad \sum_{i,j=1}^{\infty} \frac{E^2[\xi_i \zeta_i]}{\lambda_{Xi} \lambda_{Yj}^2} < \infty. \tag{2.6}$$

Functional Linear Regression Model

Now we define a functional linear model for L_2-processes as follows (compare Ramsay and Dalzell, 1991).

Consider L_2-processes $X \in L_2(T_1)$, $Y \in L_2(T_2)$. The functional linear regression model is defined as

$$Y(t) = \alpha(t) + \int_{T_1} X(s)\beta_0(s,t)ds + \epsilon(t), \tag{2.7}$$

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

By assuming, without loss of generality, that $EX(t) = 0$ and $EY(s) = 0$, for all t, s, one may simplify the linear model (2.7) to

$$Y(t) = \int_{T_1} X(s)\beta_0(s,t)ds + \epsilon(t). \tag{2.8}$$

Throughout the rest of the paper, unless stated otherwise, all L_2-processes are assumed to be centered, *i.e.*, to have zero mean functions.

Define a random integral operator $\mathcal{L}_X$: $L_2(T_1 \times T_2) \to L_2(T_2)$ by

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

It is easy to see that the adjoint operator of $\mathcal{L}_X$ is $\mathcal{L}_X^*$: $L_2(T_2) \to L_2(T_1 \times T_2)$, defined by

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

Note that (2.8) can be rewritten as

$$Y(t) = (\mathcal{L}_X\beta_0)(t) + \epsilon(t). \tag{2.9}$$

In multivariate analysis, we seek the solution of a linear regression model (2.1) by finding the parameter matrix $\beta_0^* \in \mathbf{R}^{\mathbf{p}\times\mathbf{q}}$ which minimizes the squared distance $E\|Y - \beta X\|^2$. When the covariance matrix of X is invertible, by classical least squares theory (see, *e.g.*, Anderson, 1984), the unique minimizer can be found as

$$\beta_0^* = R_{XX}^{-1} R_{XY}. \tag{2.10}$$

For the functional linear model (2.9), we seek a parameter function β_0^* such that

$$\beta_0^* = \arg\min_{\beta \in L_2(T_1 \times T_2)} E\|Y - \mathcal{L}_X \beta\|^2. \tag{2.11}$$

However, (2.10) cannot be simply extended to the functional setting. In fact, the solution to (2.11) is not unique, and there is the problem of the non-existence of the inverse of the covariance R_{XX} in the infinite dimensional case. Under the following condition, which refers to the Karhunen-Loève expansion (2.5), this problem can be alleviated by a suitable generalized inverse, as we demonstrate in Theorems 4.3 and 4.4 below.

Condition 2.2. *L_2-processes X and Y with the expansion (2.5) satisfy*

$$\sum_{i,j=1}^{\infty} \frac{E^2[\xi_i \zeta_i]}{\lambda_{Xi}^2} < \infty. \tag{2.12}$$

3. PROPERTIES OF FUNCTIONAL CANONICAL CORRELATION

In this section, we prove a main result, Theorem 3.1, which shows that basic properties of canonical correlation as known for multivariate analysis can be extended to the infinite dimensional case, if Condition 2.1 holds.

Let H_1 and H_2 be two Hilbert spaces, and let A be a linear operator from the subspace $D(A) \subset H_1$ of H_1, called the *domain* of A, into H_2. The image $R(A) = A(D(A)) = \{Ah : h \in D(A)\}$ is called the *range* of A. For a compact self-adjoint operator A on H, $D(A^{-1}) = \{f \in H : \forall h \in H, \exists! f$, s.t. $Af = h\}$. Then, referring to Karhunen-Loève expansion (2.5), the domain of operators $R_{XX}^{-1/2}$, $R_{YY}^{-1/2}$ can be defined as

$$D(R_{XX}^{-1/2}) = \Big\{u \in L_2(T_1) : \sum_{i=1}^{\infty} \lambda_{Xi}^{-1} \big|\langle u, \theta_i\rangle\big|^2 < \infty \\ \text{and} \quad R_{XX} u \neq 0 \text{ if } u \neq 0\Big\},$$

$$D(R_{YY}^{-1/2}) = \Big\{v \in L_2(T_2) : \sum_{i=1}^{\infty} \lambda_{Yi}^{-1} \big|\langle v, \phi_i\rangle\big|^2 < \infty \\ \text{and} \quad R_{YY} v \neq 0 \text{ if } v \neq 0\Big\},$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 7 (460-537)

respectively (Conway, 1985).

THEOREM 3.1. *Assume the L_2-processes X and Y satisfy Condition 2.1. Then all canonical correlations are well defined. Specifically, let (λ_i, q_i), $i \geqslant 1$ be the ith non-zero eigenvalue and orthonormal eigenvector of R^*R, and let $p_i = Rq_i/\sqrt{\lambda_i}$. Then for $I, j \geqslant 1$,*
(a) $p_i \in D(R_{XX}^{-1/2})$, $q_i \in D(R_{YY}^{-1/2})$;
(b) $\rho_i = \sqrt{\lambda_i}$, $u_i = R_{XX}^{-1/2} p_i$, and $v_i = R_{YY}^{-1/2} q_i$;
(c) $\mathrm{Corr}(U_i, U_j) = \langle u_i, R_{XX} u_j\rangle = \langle p_i, p_j\rangle = \delta_{ij}$;
(d) $\mathrm{Corr}(V_i, V_j) = \langle v_i, R_{YY} v_j\rangle = \langle q_i, q_j\rangle = \delta_{ij}$;
(e) $\mathrm{Corr}(U_i, V_j) = \langle u_i, R_{XY} v_j\rangle = \langle p_i, Rq_j\rangle = \rho_i \delta_{ij}$.

Note that Condition 2.1 allows the extension of the correlation operator R to a Hilbert-Schmidt operator on $L_2(T_2)$. Therefore, the operator R^*R has a countable sequence of eigenvalues which converges to zero. Before proceeding with the proof of Theorem 3.1, we establish results regarding the maximization properties of functional canonical correlation. We call the smallest closed subspace containing a subspace H the closure of H, denoted as $\overline{H}$. To simplify notations, let

$$H_1 = D(R_{XX}^{-1/2}), \qquad H_2 = D(R_{YY}^{-1/2}).$$

PROPOSITION 3.2. *Let $\Pi = \{\pi \subset L_2(T_2), \dim(\pi^\perp \cap \overline{H}_2) = k\}$ be a collection of subspaces in $L_2(T_2)$ with co-dimensions equal to k. Then*
(a) $\lambda_{k+1} = \sup_{q \in \mathrm{span}\{q_1,\ldots,q_k\}^\perp \cap \overline{H}_2} \dfrac{\|Rq\|^2}{\|q\|^2}$, *and*
(b) $\lambda_{k+1} = \inf\limits_{\pi \in \Pi} \sup\limits_{q \in \pi} \dfrac{\|Rq\|^2}{\|q\|^2}$.

Proof. Let $q \in \mathrm{span}\{q_1, \ldots, q_k\}^\perp$. Then

$$\begin{aligned}
\|Rq\|^2 &= \langle Rq, Rq\rangle = \langle q, R^*Rq\rangle \\
&= \left\langle \sum_{i=1}^{\infty} \langle q, q_i\rangle q_i, \sum_{i=1}^{\infty} \lambda_i \langle q, q_i\rangle q_i \right\rangle \\
&= \sum_{i>k} \lambda_i \langle q, q_i\rangle^2 \leqslant \sum_{i>k} \lambda_{k+1} \langle q, q_i\rangle^2 \\
&\leqslant \lambda_{k+1} \|q\|^2.
\end{aligned}$$

This implies

$$\lambda_{k+1} \geqslant \sup_{q \in \mathrm{span}\{q_1,\ldots,q_k\}^\perp} \frac{\|Rq\|^2}{\|q\|^2}.$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 8 (537-612)

Since $\text{span}\{q_1, \ldots, q_k\}^\perp \in \Pi$, then

$$\lambda_{k+1} \geqslant \inf_{\pi \in \Pi} \sup_{q \in \pi} \frac{\|Rq\|^2}{\|q\|^2}.$$

This proves one direction. To prove the other direction, let $\pi \in \Pi$ be arbitrary, and let $q \in \text{span}\{q_1, \ldots, q_{k+1}\} \cap \pi$. Then

$$\begin{aligned}\|Rq\|^2 &= \sum_{i=1}^{k+1} \lambda_i \langle q_i, q\rangle^2 \\ &\geqslant \lambda_{k+1} \sum_{i=1}^{k+1} \langle q_i, q\rangle^2 = \lambda_{k+1}\|q\|^2.\end{aligned}$$

This implies

$$\lambda_{k+1} \leqslant \sup_{q \in \pi} \frac{\|Rq\|^2}{\|q\|^2}.$$

Since π is arbitrary, we immediately have the following two inequalities:

$$\lambda_{k+1} \leqslant \sup_{q \in \text{span}\{q_1,\ldots,q_k\}^\perp} \frac{\|Rq\|^2}{\|q\|^2},$$

and

$$\lambda_{k+1} \leqslant \inf_{\pi \in \Pi} \sup_{q \in \pi} \frac{\|Rq\|^2}{\|q\|^2}.$$

□

Proof of Theorem 3.1. We omit the proof of (a) which can be found in He (1999).

It is easy to see that the definition of canonical correlations is equivalent to

$$\rho_k = \sup_{u \in L_2(T_1), v \in L_2(T_2)} \langle u, R_{XY} v\rangle = \langle u_k, R_{XY} v_k\rangle, \tag{3.1}$$

where u and v are subject to

$$\langle u, R_{XX} u\rangle = 1, \quad \text{and} \quad \langle v, R_{YY} v\rangle = 1, \tag{3.2}$$

and, in addition, for $k > 1$, to the requirement that

$$(U_k, V_k) \text{ is uncorelated with } (U_i, V_i), \text{ for } i = 1, \ldots, k-1. \tag{3.3}$$

For $i, j \geqslant 1$, we denote $\tilde{\rho}_i = \sqrt{\lambda_i}$, $\tilde{u}_i = R_{XX}^{-1/2} p_i$, $\tilde{v}_i = R_{YY}^{-1/2} q_i$, $\tilde{U}_i = \langle \tilde{u}_i, X\rangle$, and $\tilde{V}_i = \langle \tilde{v}_i, Y\rangle$ and prove that $(\tilde{\rho}_i, \tilde{u}_i, \tilde{v}_i, \tilde{U}_i, \tilde{V}_i)$ satisfy (3.1)–(3.3). We find

$$E[\tilde{U}_i \tilde{U}_j] = \langle \tilde{u}_i, R_{XX} \tilde{u}_j\rangle = \langle R_{XX}^{1/2} \tilde{u}_i, R_{XX}^{1/2} \tilde{u}_j\rangle = \langle p_i, p_j\rangle = \delta_{ij},$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 9 (612-692)

$$E[\tilde{V}_i \tilde{V}_j] = \langle \tilde{v}_i, R_{YY} \tilde{v}_j \rangle = \langle R_{YY}^{1/2} \tilde{v}_i, R_{YY}^{1/2} \tilde{v}_j \rangle = \langle q_i, q_j \rangle = \delta_{ij}, \quad \text{and}$$

$$E[\tilde{U}_i \tilde{V}_j] = \langle \tilde{u}_i, R_{XY} \tilde{v}_j \rangle = \langle R_{XX}^{-1/2} p_i, R_{XY} R_{YY}^{-1/2} q_j \rangle = \langle p_i, R q_j \rangle = \rho_i \delta_{ij},$$

so that the constraints of (3.2) and (3.3) are satisfied. Next, we show that $(\tilde{u}_1, \tilde{v}_1)$ is indeed the first pair of canonical weight functions for X and Y. For any $u \in H_1, v \in H_2$, that satisfy (3.2), let $p = R_{XX}^{1/2} u$, $q = R_{YY}^{1/2} v$. Then

$$1 = \langle u, R_{XX} u \rangle = \langle R_{XX}^{1/2} u, R_{XX}^{1/2} u \rangle = \|p\|^2,$$

and also $\|q\| = 1$. Hence,

$$\begin{aligned} \left|\langle u, R_{XY} v \rangle\right| &= \left|\langle R_{XX}^{-1/2} p, R_{XY} R_{YY}^{-1/2} q \rangle\right| \\ &= \left|\langle p, Rq \rangle\right| \\ &\leqslant \|p\| \, \|Rq\| \text{ (Cauchys inequality)} \\ &\leqslant \sqrt{\lambda_1} \|q_1\| \text{ (Proposition 3.2(a), for } k = 1) \\ &= \sqrt{\lambda_1} = \tilde{\rho}_1 = \langle p_1, R q_1 \rangle = \langle \tilde{u}_1, R_{XY} \tilde{v}_1 \rangle. \end{aligned}$$

We have

$$\sup_{u \in \overline{H}_1, v \in \overline{H}_2} \langle u, R_{XY} v \rangle \leqslant \langle \tilde{u}_1, R_{XY} \tilde{v}_1 \rangle. \tag{3.4}$$

Since $R(R_{XY}) \subseteq \overline{H}_1$, (3.4) is equivalent to

$$\sup_{u \in L_2(T_1), v \in L_2(T_2)} \langle u, R_{XY} v \rangle \leqslant \langle \tilde{u}_1, R_{XY} \tilde{v}_1 \rangle.$$

We conclude that

$$\sup_{u \in L_2(T_1), v \in L_2(T_2)} \langle u, R_{XY} v \rangle = \langle \tilde{u}_1, R_{XY} \tilde{v}_1 \rangle.$$

This implies that $(\tilde{\rho}_1, \tilde{u}_1, \tilde{v}_1)$ satisfy (3.1), and therefore, $(\tilde{\rho}_1, \tilde{u}_1, \tilde{v}_1) = (\rho_1, u_1, v_1)$. Now, for $k > 1$, and $u \in H_1, v \in H_2$ that satisfy (3.2) and (3.3), let $p = R_{XX}^{1/2} u$, $q = R_{YY}^{1/2} v$. Then, for $I < k$, we have again that $\|p\| = \|q\| = 1$. Furthermore,

$$\langle p, p_i \rangle = \langle u, R_{XX} u_i \rangle = \mathrm{Corr}(\langle u, X \rangle, \langle u_i, X \rangle) = 0,$$

and analogously $\langle q, q_i \rangle = 0$. Hence, $p \in \operatorname{span}\{p_1, \ldots, p_{k-1}\}^\perp \cap H_1$, and $q \in \operatorname{span}\{q_1, \ldots, q_{k-1}\}^\perp \cap H_2$. Then, again by using Proposition 3.2(a), we have

$$\begin{aligned} \left|\langle u, R_{XY} v \rangle\right| &= \left|\langle q, Rp \rangle\right| \leqslant \|p\| \, \|Rq\| \leqslant \sqrt{\lambda_k} \|q\| = \tilde{\rho}_k \\ &= \langle p_k, R q_k \rangle = \langle \tilde{u}_k, R_{XY} \tilde{v}_k \rangle, \end{aligned}$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 10 (692-769)

and

$$\sup_{u \in \overline{H}_1, v \in \overline{H}_2} \langle u, R_{XY} v \rangle \leqslant \langle \tilde{u}_k, R_{XY} \tilde{v}_k \rangle.$$

Using the same arguments as for $(\tilde{u}_1, \tilde{v}_1)$, we have

$$\sup_{u \in L_2(T_1), v \in L_2(T_2)} \langle u, R_{XY} v \rangle = \langle p_k, R q_k \rangle = \langle \tilde{u}_k, R_{XY} \tilde{v}_k \rangle.$$

This shows that $(\tilde{\rho}_k, \tilde{u}_k, \tilde{v}_k)$ satisfy (3.1) under the constraints (3.2) and (3.3), and hence that these are the kth canonical correlation and canonical weight functions.

Parts (c)–(e) follow immediately from (b) and constraints (3.2) and (3.3). □

Another characterization of functional canonical correlation is as follows.

PROPOSITION 3.3. *Let*

$$\Pi_1 = \{\pi \in H_1, \operatorname{codim} \pi = k\}, \quad \textit{and} \quad \Pi_2 = \{\pi \in H_2, \operatorname{codim} \pi = k\}.$$

Then

$$\inf_{\pi_1 \in \Pi_1, \pi_2 \in \Pi_2} \sup_{u \in \pi_1, v \in \pi_2} \operatorname{Corr}\big(\langle u, X \rangle, \langle v, Y \rangle\big) = \rho_{k+1}.$$

Proof. The proof follows immediately from Theorem 3.1 and Proposition 3.2 by using the fact that $p = R_{XX}^{1/2} u$, $q = R_{YY}^{1/2} v$, and $|\operatorname{Corr}(\langle u, X \rangle, \langle v, Y \rangle)| = \|Rq\|/\|q\|$. □

4. THE FUNCTIONAL LINEAR REGRESSION MODEL

In this section, we explore properties of functional linear regression and investigate the extension of corresponding results known for linear regression in multivariate analysis, where the dependent variable is a finite-dimensional vector..

PROPOSITION 4.1. *Let* β_0 *be a solution of the linear regression model (2.9). Then*

$$\beta_0 \in \underset{\beta \in L_2(T_1 \times T_2)}{\arg\min}\, E\|Y - \mathcal{L}_X \beta\|^2.$$

Proof. According to the usual model assumptions listed after (2.7), X and ϵ are uncorrelated, which implies that for any $\beta \in L_2(T_1 \times T_2)$, it holds that $E[\langle \mathcal{L}_X \beta, \epsilon \rangle] = 0$. Then

$$\begin{aligned} E\|Y - \mathcal{L}_X \beta\|^2 &= E\big\|(\mathcal{L}_X \beta_0 + \epsilon) - \mathcal{L}_X \beta\big\|^2 \\ &= E\|\mathcal{L}_X \beta_0 - \mathcal{L}_X \beta\|^2 + E\|\epsilon\|^2 + 2E\big[\langle \mathcal{L}_X \beta_0 - \mathcal{L}_X \beta, \epsilon \rangle\big] \end{aligned}$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 11 (769-850)

$$= E\|\mathcal{L}_X\beta_0 - \mathcal{L}_X\beta\|^2 + E\|Y - \mathcal{L}_X\beta_0\|^2$$
$$\geqslant E\|Y - \mathcal{L}_X\beta_0\|^2.$$

□

Motivated by the form of the least squares solution (2.10) for the multivariate linear model, we define a linear integral operator $\Gamma_{XX}\colon L_2(T_1 \times T_2) \to L_2(T_1 \times T_2)$ as

$$(\Gamma_{XX}\beta)(s,t) = \int_{T_1} r_{XX}(s,w)\beta(w,t)dw.$$

It is easy to see that $\Gamma_{XX} = E[\mathcal{L}_X^*\mathcal{L}_X]$. Moreover, $\Gamma_{XX}|_{L_2(T_1)} = R_{XX}$, where $\Gamma_{XX}|_{L_2(T_1)}$ is the restriction of Γ_{XX} to $L_2(T_1)$, and $L_2(T_1)$ is treated as an embedded subspace of $L_2(T_1 \times T_2)$. Furthermore, Γ_{XX} is a self-adjoint nonnegative Hilbert-Schmidt operator. Denote the range of the operator R_{XX} as $\mathrm{R}(R_{XX})$, where $\mathrm{R}(R_{XX}) = \mathrm{span}\{\theta_i | i \geqslant 1\}$, and $\{\theta_i\}$ is the eigenbasis for X used in the Karhunen-Loève decomposition (2.5). Then, the range of Γ_{XX} is $\mathrm{R}(\Gamma_{XX}) = \mathrm{R}(R_{XX}) \times L_2(T_2)$.

PROPOSITION 4.2. *Let $\beta \in L_2(T_1 \times T_2)$. Then*

$$\beta \in \underset{\beta \in L_2(T_1 \times T_2)}{\arg\min}\ E\|Y - \mathcal{L}_X\beta\|^2 \text{ if and only if } \mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta = \mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta_0,$$

where $\mathrm{P}_{\mathrm{R}(\Gamma_{XX})}$ is the projection from $L_2(T_1 \times T_2)$ to $\mathrm{R}(\Gamma_{XX})$.

Proof. From the proof of Proposition 4.1, β is a minimizer of $E\|Y - \mathcal{L}_X\beta\|^2$ if and only if $E\|\mathcal{L}_X\beta - \mathcal{L}_X\beta_0\|^2 = 0$. Observing that

$$\begin{aligned} E\|\mathcal{L}_X\beta - \mathcal{L}_X\beta_0\|^2 &= E\big[\langle \mathcal{L}_X(\beta - \beta_0), \mathcal{L}_X(\beta - \beta_0)\rangle\big] \\ &= \big\langle \beta - \beta_0, E\big[(\mathcal{L}_X^*\mathcal{L}_X)(\beta - \beta_0)\big]\big\rangle \\ &= \big\langle \beta - \beta_0, \Gamma_{XX}(\beta - \beta_0)\big\rangle \\ &= \big\|\Gamma_{XX}^{1/2}(\beta - \beta_0)\big\|^2, \end{aligned}$$

β is a minimizer if and only if $\|\Gamma_{XX}^{1/2}(\beta - \beta_0)\|^2 = 0$. This is equivalent to $\Gamma_{XX}^{1/2}\beta = \Gamma_{XX}^{1/2}\beta_0$, which in turn is equivalent to $\mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta = \mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta_0$. □

Applying $\mathcal{L}_X^*$ to both sides of (2.9), and taking expectations, one obtains

$$E(\mathcal{L}_X^*Y)(s,t) = E(\mathcal{L}_X^*\mathcal{L}_X\beta_0)(s,t) + E(\mathcal{L}_X^*\varepsilon)(s,t),$$

where

$$E(\mathcal{L}_X^*Y)(s,t) = E\big[X(s)Y(t)\big] = r_{XY}(s,t),$$

$$E(\mathcal{L}_X^*\mathcal{L}_X\beta_0)(s,t) = E\int X(s)X(w)\beta_0(w,t)dw = \Gamma_{XX}\beta_0(s,t),$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 12 (850-926)

and

$$E(\mathcal{L}_X^* \varepsilon) = E[X\varepsilon] = 0.$$

Hence, $r_{XY} = \Gamma_{XX}\beta_0$. Accordingly, we refer to

$$r_{XY} = \Gamma_{XX}\beta, \quad \text{for } \beta \in L_2(T_1 \times T_2), \tag{4.1}$$

as the *functional normal equation.*

THEOREM 4.3. *Let X and Y be L_2-processes with the expansion (2.5) which satisfy Condition 2.2. Then,*

(a) $\beta_0^* = \Gamma_{XX}^{-1} r_{XY}$ exists and is the unique solution of (4.1) in $R(\Gamma_{XX})$;

(b) β_0^* has the representation

$$\beta_0^*(s,t) = \sum_{i,j=1}^{\infty} \frac{E[\xi_i \zeta_i]}{\lambda_{Xi}} \theta_i(s)\phi_j(t);$$

(c) The set of the solutions of (4.1) is:

$$\beta_0^* + \ker(\Gamma_{XX}) := \left\{\beta_0^* + h | h \in \ker(\Gamma_{XX})\right\},$$

where ker(Γ_{XX}) is the kernel space of Γ_{XX}, i.e., ker$(\Gamma_{XX}) = \{h \in L(T_1 \times T_2)$: $\Gamma_{XX} h = 0\}$.

Proof. Using the Karhunen-Loève representation (2.5), expand the cross-covariance function as

$$r_{XY}(s,t) = \sum_{i,j=1}^{\infty} E[\xi_i \zeta_j]\theta_i(s)\phi_j(t).$$

The proof for (I) and (ii) follows from Conway (1985). To prove (iii), first note that $\beta_0^* + \ker(\Gamma_{XX})$ are the solutions of (4.1). On the other hand, let $\beta_0 \in L_2(T_1 \times T_2)$ be a solution of (4.1). Then $\mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta_0$ must also be a solution for (4.1). Note that $\mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta_0 \in \mathrm{R}(\Gamma_{XX})$. From the uniqueness in part (I), $\mathrm{P}_{\mathrm{R}(\Gamma_{XX})}\beta_0 = \beta_0^*$. Therefore, $\beta_0 = \beta_0^* + h$, with $h \in \ker(\Gamma_{XX})$. □

Combining Proposition 4.2 with Theorem 4.3, we immediately obtain the following result, which provides a characterization of the set arg $\min_\beta E\|Y - \mathcal{L}_X\beta\|^2$.

THEOREM 4.4. *Assume condition 2.4 holds for X and Y. Then*

$$\arg\min_\beta E\|Y - \mathcal{L}_X\beta\|^2 = \beta_0^* + \ker(\Gamma_{XX}),$$

and this coincides with the set of solutions for the functional normal equation (4.1).

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 13 (926-1002)

The following result provides yet another characterization of functional least squares. It extends the corresponding result from the multivariate linear model to the functional linear model.

PROPOSITION 4.5. $\min_{\beta \in L_2(T_1 \times T_2)} E\|Y - \mathcal{L}_X\beta\|^2 = \operatorname{tr}(R_{YY}) - E\|\mathcal{L}_X\beta_0\|^2.$

Proof.

$$\begin{aligned}\min_{\beta \in L_2(T_1 \times T_2)} E\|Y - \mathcal{L}_X\beta\|^2 &= E\|Y - \mathcal{L}_X\beta_0\|^2 \\ &= E\|Y\|^2 + E\|\mathcal{L}_X\beta_0\|^2 - 2E\big[\langle Y, \mathcal{L}_X\beta_0\rangle\big],\end{aligned}$$

where

$$E\|Y\|^2 = \int r_{YY}(t,t)dt = \operatorname{tr}(R_{YY}),$$

and

$$\begin{aligned}E\big[\langle Y, \mathcal{L}_X\beta_0\rangle\big] &= E\big[\langle \mathcal{L}_X\beta_0 + \epsilon, \mathcal{L}_X\beta_0\rangle\big] \\ &= E\|\mathcal{L}_X\beta_0\|^2 + \langle E\mathcal{L}_X^*\epsilon, \beta_0\rangle.\end{aligned}$$

From the assumption for the functional linear model (2.9),

$$E\mathcal{L}_X^*\epsilon = E[X\epsilon] = 0,$$

which completes the proof. □

The following theorem demonstrates an important property of functional linear models: Any minimizer, $\arg\min_\beta E\|Y - \mathcal{L}_X\beta\|^2$, maximizes the correlation between the response and predictor functions. Again, the analogous result is well known for the multivariate linear model.

THEOREM 4.6. *Assume Condition 2.2 holds. Then*

$$\arg\min_\beta E\|Y - \mathcal{L}_X\beta\|^2 \subseteq \arg\max_\beta \frac{E^2[\langle Y, \mathcal{L}_X\beta\rangle]}{E\|Y\|^2 E\|\mathcal{L}_X\beta\|^2}. \tag{4.2}$$

Proof. Observing that

$$\begin{aligned}\big\{E\big[\langle Y, \mathcal{L}_X\beta\rangle\big]\big\}^2 &= \big\langle E[\mathcal{L}_X^* Y], \beta\big\rangle^2 = \langle r_{XY}, \beta\rangle^2 \\ &= \langle \Gamma_{XX}^{-1/2} r_{XY}, \Gamma_{XX}^{1/2}\beta\rangle^2 \leqslant \|\Gamma_{XX}^{-1/2} r_{XY}\|^2 \|\Gamma_{XX}^{1/2}\beta\|^2\end{aligned}$$

and

$$\begin{aligned}E\|\mathcal{L}_X\beta\|^2 &= E\big[\langle \mathcal{L}_X\beta, \mathcal{L}_X\beta\rangle\big] = \big\langle \beta, E[\mathcal{L}_X^*\mathcal{L}_X\beta]\big\rangle \\ &= \langle \beta, \Gamma_{XX}\beta\rangle = \|\Gamma_{XX}^{1/2}\beta\|^2,\end{aligned}$$

VSP(ROUS) 2000/06/14 Prn:23/06/2000; 8:43 F:ROUS12.tex; VTeX/AS p. 14 (1002-1087)

one obtains

$$\frac{E^2[< Y, \mathcal{L}_X\beta >]}{E\|Y\|^2 E\|\mathcal{L}_X\beta\|^2} \leqslant \|\Gamma_{XX}^{-1/2} r_{XY}\|^2 / E\|Y\|^2, \quad \text{for any } \beta \in L_2(T_1 \times T_2).$$

Equality holds if and only if $\Gamma_{XX}^{-1/2} r_{XY} = c\Gamma_{XX}^{1/2}\beta$, for some constant c. That is, $\beta = c\Gamma_{XX}^{-1} r_{XY} = c\beta_0$, which proves that β_0 satisfies (4.2). From Theorem 4.4, any $\beta \in \arg\min_\beta E\|Y - \mathcal{L}_X\beta\|^2$ is of the form $\beta = \beta_0 + h$, where $h \in \ker(\Gamma_{XX})$. Therefore, $\mathcal{L}_X\beta = \mathcal{L}_X\beta_0$, and so (4.2) holds for β as well. □

REFERENCES

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition. John Wiley & Sons, New York.

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

Brillinger, D. R. (1975). *Time Series. Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

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

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivar. Anal.* **12**, 136–154.

He, G. (1999). *Functional Canonical Analysis and Linear Modeling*. Ph.D. Dissertation, University of California, Davis.

Leurgans, S. E., Moyeed, R. A. and Silverman, B. W. (1993). Canonical correlation analysis when the data are curves. *J. Royal Statist. Soc.* **B55**, 725–740.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis (with discussion). *J. Royal Statist. Soc.* **B53**, 539–572.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Royal Statist, Soc*. **B53**, 233–243.

Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stoch. Processes Appl.* **36**, 107–116.