Working paper

2023-07

Statistics and Econometrics
ISSN 2387-0303

Penalized function-on-function partial least-squares regression

Harold A. Hernández-Roig, M. Carmen Aguilera-Morillo, Ana M. Aguilera, and Cristian Preda

uc3m | Universidad Carlos III de Madrid

Penalized function-on-function partial least-squares regression

Harold A. Hernández-Roig [*] M. Carmen Aguilera-Morillo [†]
Ana M. Aguilera [‡] Cristian Preda [§]

July 5, 2023

Abstract

This paper deals with the "function-on-function" or "fully functional" linear regression problem. We address the problem by proposing a novel penalized Function-on-Function Partial Least-Squares (pFFPLS) approach that imposes smoothness on the PLS weights. Our proposal introduces an appropriate finite-dimensional functional space with an associated set of bases on which to represent the data and controls smoothness with a roughness penalty operator. Penalizing the PLS weights imposes smoothness on the resulting coefficient function, improving its interpretability. In a simulation study, we demonstrate the advantages of pFFPLS compared to non-penalized FFPLS. Our comparisons indicate a higher accuracy of pFFPLS when predicting the response and estimating the true coefficient function from which the data were generated. We also illustrate the advantages of our proposal with two case studies involving two well-known datasets from the functional data analysis literature. In the first one, we predict log precipitation curves from the yearly temperature profiles recorded in 35 weather stations in Canada. In the second case study, we predict the hip angle profiles during a gait cycle of children from their corresponding knee angle profiles.

Keywords: Functional Data Analysis, Partial Least Squares, Function-on-Function Regression, Roughness Penalties.

[*]Department of Statistics, Universidad Carlos III de Madrid, Madrid, Spain, and uc3m-Santander Big Data Institute, Madrid, Spain. Email: haroldantonio.hernandez@uc3m.es.

[†]Department of Applied Statistics and Operational Research, and Quality, Universitat Politècnica de València, Valencia, Spain. Email: mdagumor@eio.upv.es.

[‡]Department of Statistics and Operations Research, Institute of Mathematics, Universidad de Granada, Granada, Spain. Email: aaguiler@ugr.es.

[§]UMR CNRS 8524—Laboratoire Paul Painlevé, Université de Lille, Lille, France. Email: cristian.preda@univ-lille.fr.

1 Introduction

Functional Data Analysis (FDA) covers the study of curves, surfaces, and objects naturally appearing as functions. The rise of FDA is closely tied to the advancement of computational methods and the increasing capacity to collect large amounts of data. These data can be high-dimensional and have very complex structures for which classical multivariate techniques are inappropriate. That could be the case with environmental data, biomedical images like functional magnetic resonance images (fMRI) and electroencephalographs (EEG), time series, and spectroscopy curves, among many others. The books by [Ramsay and Silverman, 2005, Ferraty and Vieu, 2006, Ramsay et al., 2009, Horváth and Kokoszka, 2012, Kokoszka and Reimherr, 2017] provide a very thorough overview of FDA techniques with plenty of examples and case studies, usually involving R [R Core Team, 2023]. Wang et al. [Wang et al., 2016] provide a more compact review of the most common FDA methods in the literature.

This paper deals with the "function-on-function" or "fully functional" linear model:

$$Y(q) = \int_{\mathcal{D}_X} X(p)\beta(q,p)dp + \varepsilon(q), \quad q \in \mathcal{D}_Y, \tag{1}$$

where the response $Y(q),\ q \in \mathcal{D}_Y$, and the predictor $X(p),\ p \in \mathcal{D}_X$, are mean-centered functional random variables observed over potentially different domains $\mathcal{D}_Y \subset \mathbb{R}$ and $\mathcal{D}_X \subset \mathbb{R}$, respectively. According to this formulation, at any point $q \in \mathcal{D}_Y$, the value of $Y(q)$ depends on the entire trajectory of X. Model (1) can be seen as the direct extension of the multivariate linear regression in Euclidean space to linear regression in the space $\mathcal{L}_2$ of square-integrable functions. The goal is to estimate the parameter/coefficient function β, a surface with a domain in $\mathcal{D}_X \times \mathcal{D}_Y$.

We introduce a novel approach called penalized Function-on-Function Partial Least Squares (pFFPLS) that builds upon the basis representation of the data and the partial least-squares weights, and incorporates a penalty term to control the smoothness of the weight functions used in the algorithm. By doing so, the smoothness of the estimated coefficient function is influenced. The pFFPLS method can be viewed as an extension of the scalar-on-function model proposed in [Aguilera et al., 2016]. The main motivation behind this approach is that the existing literature on FFPLS primarily focuses on enhancing the model's predictive accuracy while often neglecting the interpretability of $\hat{\beta}$. In contrast, our strategy involves fixing the number of bases to a sufficiently large value and imposing smoothness by using penalties.

The fully functional linear model (1) was first studied in [Ramsay and Dalzell, 1991]. The books [Ramsay and Silverman, 2005, Ramsay et al., 2009] provide case studies on fitting this model via penalized least squares when the data are observed over a dense set of nodes. The idea is to reduce the infinite dimension of the coefficient function by translating the problem to a lower-dimensional space spanned by a fixed set of basis functions, such as B-splines. An alternative data-driven strategy is considering Principal Components Analysis (PCA), as demonstrated in [Yao et al., 2005]. In this approach, the lower-dimensional space is constructed from the eigenfunctions of the associated

autocovariance operator. Ivanescu et al. [Ivanescu et al., 2015] propose a penalized solution that exploits the connection to mixed models, allowing for realistic situations involving multiple predictors or sparsely-sampled functions.

The first FFPLS approach is due to [Preda and Schiltz, 2011]. Their proposal establishes a connection between FFPLS and its multivariate PLS counterpart [Wold et al., 1983], assuming that the functional data can be represented using a set of basis functions. The FFPLS approach follows a similar concept to the scalar-on-function setting proposed by [Preda and Saporta, 2005, Aguilera et al., 2010]. Since the publication of Preda and Jang's paper in 2011, only a few studies have tackled the function-on-function problem from a PLS perspective. For instance, Beyaztas and Shang [Beyaztas and Shang, 2020] demonstrate the advantages of FFPLS compared to several of the aforementioned approaches, such as those presented by [Ivanescu et al., 2015, Ramsay and Silverman, 2005]. They also extend FFPLS to the case of multiple functional predictors. In a subsequent paper [Beyaztas and Shang, 2021], the authors introduce the possibility of considering interaction terms within the FFPLS framework. An alternative to the basis expansion approach is presented in [Zhou, 2021], where the author proposes a fast implementation for FFPLS regression, similar to the "alternative PLS" method for scalar-on-function regression introduced by [Delaigle and Hall, 2012].

The rest of the paper is organized as follows. Section 2 introduces the theoretical framework supporting the FFPLS algorithm, focusing on the basis representation approach proposed by [Preda and Schiltz, 2011]. Section 3 presents our proposed pFFPLS method. In particular, Proposition 3.1 summarizes the main contribution of this paper. Section 4 presents a simulation study comparing pFFPLS with FFPLS. The performance and effectiveness of pFFPLS are evaluated regarding predictive accuracy and interpretability of the resulting coefficient function. Section 5 showcases two real-data applications. First, in Subsection 5.1, we predict log precipitation based on temperature profiles from 35 weather stations in Canada. Second, in Subsection 5.2, we focus on predicting hip angle profiles from knee angle profiles using data on 39 children undergoing a gait cycle. These case studies are representative examples in the function-on-function literature, and the necessary data can be accessed through the R package `fda` [Ramsay, 2023]. Finally, Section 6 presents our conclusions and discusses potential future research directions.

2 Function-on-function PLS regression

The FFPLS regression consists of building PLS components as linear functionals of X:

$$t = \int_{\mathcal{D}_X} X(p)\omega(p)dp,$$

obtained by maximizing the squared covariance between $Y(q)$ and $X(p)$, also known as Tucker's criterion:

$$\max_{\|\omega\|_{\mathcal{L}_2(\mathcal{D}_X)}=\|\nu\|_{\mathcal{L}_2(\mathcal{D}_Y)}=1} \operatorname{cov}^2\left(\int_{\mathcal{D}_X} X(p)\,\omega(p)\,dp, \int_{\mathcal{D}_Y} Y(q)\,\nu(q)\,dq\right). \tag{2}$$

The PLS algorithm is an iterative procedure. Define the Escoufier operators associated with X and Y:

$$\begin{aligned}\mathcal{W}^X Z &= \int_{\mathcal{D}_X} \mathbb{E}(X(p)Z)X(p)dp,\\ \mathcal{W}^Y Z &= \int_{\mathcal{D}_Y} \mathbb{E}(Y(q)Z)Y(q)dq,\end{aligned}$$

for any random variable $Z \in \mathcal{L}_2(\Omega)$ with finite second moment. Then, following [Preda and Schiltz, 2011], the first PLS component t_1 is given by the eigenvector associated with the largest eigenvalue of the operator $\mathcal{W}^X\mathcal{W}^Y$:

$$\mathcal{W}^X\mathcal{W}^Y t_1 = \lambda_{\max} t_1.$$

The first PLS weight ω_1 associated with t_1 is estimated by:

$$\omega_1(p) = \frac{\int_{\mathcal{D}_Y} \mathbb{E}\left(X(p)Y(q)\right) dq}{\sqrt{\int_{\mathcal{D}_X}\left[\int_{\mathcal{D}_Y} \mathbb{E}\left(X(p)Y(q)\right) dq\right]^2 dp}}, \quad p \in \mathcal{D}_X,$$

such that $t_1 = \int_{\mathcal{D}_X} X(p)\omega_1(p)dp$.

The first step is completed with the ordinary linear regression of $X(p)$ and $Y(q)$ on t_1, resulting in the coefficients:

$$\begin{aligned}\varrho_1(p) &= \mathbb{E}(X(p)t_1)/\mathbb{E}(t_1^2),\\ \delta_1(q) &= \mathbb{E}(Y(q)t_1)/\mathbb{E}(t_1^2),\end{aligned}$$

with residuals $X_1(p)$ and $Y_1(q)$:

$$\begin{aligned}X_1(p) &= X(p) - \varrho_1(p)t_1, \quad p \in \mathcal{D}_X,\\ Y_1(q) &= Y(q) - \delta_1(q)t_1, \quad q \in \mathcal{D}_Y.\end{aligned}$$

Letting $X_0(p) = X(p)$ and $Y_0(q) = Y(q)$, the h-th PLS component ($h \geq 1$) is given by the eigenvector associated with the largest eigenvalue of the operator $\mathcal{W}^{X_{h-1}}\mathcal{W}^{Y_{h-1}}$:

$$\mathcal{W}^{X_{h-1}}\mathcal{W}^{Y_{h-1}} t_h = \lambda_{\max} t_h,$$

with weight ω_h:

$$\omega_h(p) = \frac{\int_{\mathcal{D}_Y} \mathbb{E}\left(X_{h-1}(p)Y_{h-1}(q)\right) dq}{\sqrt{\int_{\mathcal{D}_X}\left[\int_{\mathcal{D}_Y} \mathbb{E}\left(X_{h-1}(p)Y_{h-1}(q)\right) dq\right]^2 dp}}, \quad p \in \mathcal{D}_X,$$

such that $t_h = \int_{\mathcal{D}_X} X_{h-1}(p)\omega_h(p)dp$. The h-th PLS step is completed with the linear regression of $X_{h-1}(p)$ and $Y_{h-1}(q)$ on t_h:

$$\begin{aligned}X_h(p) &= X_{h-1}(p) - \varrho_h(p)t_h, \quad p \in \mathcal{D}_X,\\ Y_h(q) &= Y_{h-1}(q) - \delta_h(q)t_h, \quad q \in \mathcal{D}_Y,\end{aligned}$$

where $\varrho_h(p) = \left(\mathbb{E}(X_{h-1}(p)t_h)/\mathbb{E}(t_h^2)\right)$ and $\delta_h(q) = \left(\mathbb{E}(Y_{h-1}(q)t_h)/\mathbb{E}(t_h^2)\right)$. This procedure is known as deflation and is intended to remove variability already explained from X_{h-1} and Y_{h-1} by subtracting the best predictions we can make of them using the estimated PLS components.

This approach is asymmetric because the same component t_h relative to the weight ω_h is used to deflate both the predictor and response variables. This is a widespread technique, sometimes called "PLS2" or "regression mode" in multivariate settings (see, for example, [Lê Cao et al., 2008] and references therein). The alternative is the symmetric PLS, also called "Mode A" in multivariate settings, that uses u_h relative to the weight ν_h to deflate the response variable Y_{h-1}. In this case, u_h is the eigenvector associated with the largest eigenvalue of the operator $\mathcal{W}^{Y_{h-1}}\mathcal{W}^{X_{h-1}}$:

$$\mathcal{W}^{Y_{h-1}}\mathcal{W}^{X_{h-1}}u_h = \rho_{\max}u_h.$$

The corresponding weight ν_h is computed as:

$$\nu_h(q) = \frac{\int_{\mathcal{D}_X}\mathbb{E}\left(X_{h-1}(p)Y_{h-1}(q)\right)dp}{\sqrt{\int_{\mathcal{D}_Y}\left[\int_{\mathcal{D}_X}\mathbb{E}\left(X_{h-1}(p)Y_{h-1}(q)\right)dp\right]^2 dq}}, \quad q \in \mathcal{D}_Y,$$

such that the corresponding PLS component is $u_h = \int_{\mathcal{D}_Y} Y_{h-1}(q)\nu_h(q)dq$.

The properties of this function-on-function PLS approach are summarized in the following proposition by [Preda and Schiltz, 2011]:

Proposition 2.1. *For any $h \geq 1$:*

1. $\{t_h\}_{h\geq 1}$ *forms an orthogonal system in* $\mathcal{L}_2\left(\mathcal{D}_X\right)$,
2. $Y(q) = \delta_1(q)t_1 + \delta_2(q)t_2 + \ldots + \delta_h(q)t_h + Y_h(q), \quad q \in \mathcal{D}_Y$,
3. $X(p) = \varrho_1(p)t_1 + \varrho_2(p)t_2 + \ldots + \varrho_h(p)t_h + X_h(p), \quad p \in \mathcal{D}_X$,
4. $\mathbb{E}[Y_h(q)t_j] = 0, \quad \forall q \in \mathcal{D}_Y,\ \forall j = 1,\ldots,h$,
5. $\mathbb{E}[X_h(p)t_j] = 0, \quad \forall p \in \mathcal{D}_X,\ \forall j = 1,\ldots,h$.

Note that every component can be written in terms of the original X as $t_h = \langle X, \omega_h^*\rangle$, where $\omega_h^* \in \text{span}\{\omega_1,\ldots,\omega_h\},\ h = 1,\ldots,H$. Considering the decomposition of $Y(q)$ established in Proposition 2.1, we can approximate the response using H components:

$$\begin{aligned}
\hat{Y}_H(q) &= \delta_1(q)\int_{\mathcal{D}_X} X(p)\omega_1^*(p)dp + \ldots + \delta_H(q)\int_{\mathcal{D}_X} X(p)\omega_H^*(p)dp, \\
&= \int_{\mathcal{D}_X} X(p)\left\{\delta_1(q)\omega_1^*(p) + \ldots + \delta_H(q)\omega_H^*(p)\right\}dp, \\
&= \int_{\mathcal{D}_X} X(p)\hat{\beta}_H(q,p)dp,
\end{aligned}$$

where $\hat{\beta}_H(q,p)$ is the approximation to the parameter function $\beta(q,p)$ at points $(q,p) \in \mathcal{D}_Y \times \mathcal{D}_X$ using H components. The coefficient function β can also be represented using the series:

$$\beta = \sum_h \delta_h \phi_h,$$

where the functions ϕ_h are defined recursively:

$$\begin{aligned} \phi_1 &= \omega_1, \\ \phi_h &= \omega_h - \langle \varrho_1, \omega_h \rangle \phi_1 - \ldots - \langle \varrho_{h-1}, \omega_h \rangle \phi_{h-1}, \ \forall h \geq 2. \end{aligned}$$

2.1 The basis representation approach

Let $\boldsymbol{\phi}(p) = (\phi_1(p), \ldots, \phi_K(p))^T$, be a K-dimensional vector of basis functions in $\mathcal{L}_2(\mathcal{D}_X)$. Analogously, let $\boldsymbol{\psi} = (\psi_1(q), \ldots, \psi_L(q))^T$, be an L-dimensional vector of basis functions in $\mathcal{L}_2(\mathcal{D}_Y)$. We define the $K \times K$ matrix $\mathbf{R}_{\boldsymbol{\phi}} = \int_{\mathcal{D}_X} \boldsymbol{\phi}(p)\boldsymbol{\phi}^T(p)dp$ and the $L \times L$ matrix $\mathbf{R}_{\boldsymbol{\psi}} = \int_{\mathcal{D}_Y} \boldsymbol{\psi}(q)\boldsymbol{\psi}^T(q)dq$ of inner products between the basis functions. Let $\mathbf{R}_{\boldsymbol{\phi}}^{1/2}$ and $\mathbf{R}_{\boldsymbol{\psi}}^{1/2}$ be the square roots of $\mathbf{R}_{\boldsymbol{\phi}}$ and $\mathbf{R}_{\boldsymbol{\psi}}$, respectively. The square roots are taken in the sense that $\mathbf{R}_{\boldsymbol{\phi}} = \mathbf{R}_{\boldsymbol{\phi}}^{1/2}(\mathbf{R}_{\boldsymbol{\phi}}^{1/2})^T$ and $\mathbf{R}_{\boldsymbol{\psi}} = \mathbf{R}_{\boldsymbol{\psi}}^{1/2}(\mathbf{R}_{\boldsymbol{\psi}}^{1/2})^T$.

Assume that the functional predictor and response can be represented using a basis expansion of the form:

$$X(p) = \sum_{k=1}^{K} \alpha_k \phi_k(p) = \boldsymbol{\alpha}^T \boldsymbol{\phi}(p) \quad p \in \mathcal{D}_X, \tag{3}$$

$$Y(q) = \sum_{l=1}^{L} \gamma_l \psi_l(q) = \boldsymbol{\gamma}^T \boldsymbol{\psi}(q), \quad q \in \mathcal{D}_Y, \tag{4}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)^T$, the random vectors of coefficients for X and Y respectively. As a consequence, the weight functions can also be expressed in terms of the same bases:

$$\omega(p) = \sum_{k=1}^{K} w_k \phi_k(p) = \mathbf{w}^T \boldsymbol{\phi}(p), \quad p \in \mathcal{D}_X, \tag{5}$$

$$\nu(q) = \sum_{l=1}^{L} v_l \psi_l(q) = \mathbf{v}^T \boldsymbol{\psi}(q), \quad q \in \mathcal{D}_Y, \tag{6}$$

with $\mathbf{w} = (w_1, \ldots, w_K)^T$ and $\mathbf{v} = (v_1, \ldots, v_L)^T$ the corresponding vectors of basis coefficients.

The following proposition by [Preda and Schiltz, 2011] provides an equivalence between the FFPLS and its multivariate counterpart.

Proposition 2.2. *Assume $X(p)$ and $Y(q)$ allow a basis representation as in* (3)-(4) *and define:*

$$\boldsymbol{\Lambda} = \mathbf{R}_{\boldsymbol{\phi}}^{1/2}\boldsymbol{\alpha} \quad \textit{and} \quad \boldsymbol{\Pi} = \mathbf{R}_{\boldsymbol{\psi}}^{1/2}\boldsymbol{\gamma}. \tag{7}$$

i) The PLS regression of $Y(q)$, $q \in \mathcal{D}_Y$, on $X(p)$, $p \in \mathcal{D}_X$, is equivalent to the PLS regression of $\mathbf{\Pi}$ on $\mathbf{\Lambda}$ in the sense that at each step h of the PLS algorithm, $1 \leq h \leq K$, we have the same PLS component t_h for both regressions.

ii) If $\mathbf{\Theta}$ is the $L \times K$ matrix of regression coefficients of $\mathbf{\Pi}$ on $\mathbf{\Lambda}$ obtained with the PLS regression at step h, $1 \leq h \leq K$. i.e.:

$$\mathbf{\Pi} = \mathbf{\Theta}\mathbf{\Lambda} + \boldsymbol{\varepsilon},$$

then the FFPLS approximation of the regression coefficient function $\beta(q,p)$ from (1) is:

$$\hat{\beta}(q,p) = \sum_{i=1}^{L}\sum_{j=1}^{K} \psi_i(q)\mathbf{B}_{i,j}\phi_j(p), \quad (q,p) \in \mathcal{D}_Y \times \mathcal{D}_X,$$

where $\mathbf{B} = \mathbf{R}_{\boldsymbol{\psi}}^{-1/2}\mathbf{\Theta}\mathbf{R}_{\boldsymbol{\phi}}^{-1/2}$.

The proposition can be verified through an induction on h, using the approximations $\mathcal{W}^X \approx \mathcal{W}^{\mathbf{R}_{\boldsymbol{\phi}}^{1/2}\boldsymbol{\alpha}}$ and $\mathcal{W}^Y \approx \mathcal{W}^{\mathbf{R}_{\boldsymbol{\psi}}^{1/2}\boldsymbol{\gamma}}$. The procedure is similar to the one presented in [Aguilera et al., 2010] for the scalar-on-function setting. Additionally, the paper [Beyaztas and Shang, 2021] provides proof of this proposition in the context of function-on-multiple-functions regression.

At each step $h \geq 1$ of the FFPLS algorithm, the vector of basis coefficients of the weight function associated with the h-th component t_h is given by $\mathbf{w}_h = [\mathbf{R}_{\boldsymbol{\phi}}^{-1/2}]^T\tilde{\mathbf{w}}_h$, with $\tilde{\mathbf{w}}_h$ being the eigenvector associated with the largest eigenvalue of the eigenproblem:

$$[\mathbf{R}_{\boldsymbol{\phi}}^{1/2}]^T\mathbf{\Sigma}_{h-1}\mathbf{R}_{\boldsymbol{\psi}}^{1/2}\frac{\tilde{\mathbf{v}}_h\tilde{\mathbf{v}}_h^T}{\tilde{\mathbf{v}}_h^T\tilde{\mathbf{v}}_h}[\mathbf{R}_{\boldsymbol{\psi}}^{1/2}]^T\mathbf{\Sigma}_{h-1}^T\mathbf{R}_{\boldsymbol{\phi}}^{1/2}\tilde{\mathbf{w}}_h = \mu\tilde{\mathbf{w}}_h, \quad \tilde{\mathbf{w}}_h \in \mathbb{R}^K,\ \langle\tilde{\mathbf{w}}_h,\tilde{\mathbf{w}}_h\rangle_{\mathbb{R}^K} = 1,$$

where $\mathbf{\Sigma}_{h-1}$ is the $K \times L$ covariance matrix between the vectors of basis coefficients $\boldsymbol{\alpha}_{h-1}$ and $\boldsymbol{\gamma}_{h-1}$ of X_{h-1} and Y_{h-1}, respectively. Analogously, the vector of coefficients for the weight ν_h is $\mathbf{v}_h = [\mathbf{R}_{\boldsymbol{\psi}}^{-1/2}]^T\tilde{\mathbf{v}}_h$, where $\tilde{\mathbf{v}}_h$ is the eigenvector associated with the largest eigenvalue of the problem:

$$[\mathbf{R}_{\boldsymbol{\psi}}^{1/2}]^T\mathbf{\Sigma}_{h-1}^T\mathbf{R}_{\boldsymbol{\phi}}^{1/2}\frac{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^T}{\tilde{\mathbf{w}}_h^T\tilde{\mathbf{w}}_h}[\mathbf{R}_{\boldsymbol{\phi}}^{1/2}]^T\mathbf{\Sigma}_{h-1}\mathbf{R}_{\boldsymbol{\psi}}^{1/2}\tilde{\mathbf{v}}_h = \mu\tilde{\mathbf{v}}_h, \quad \tilde{\mathbf{v}}_h \in \mathbb{R}^L,\ \langle\tilde{\mathbf{v}}_h,\tilde{\mathbf{v}}_h\rangle_{\mathbb{R}^L} = 1.$$

Notice that $[\mathbf{R}_{\boldsymbol{\phi}}^{1/2}]^T\mathbf{\Sigma}_{h-1}\mathbf{R}_{\boldsymbol{\psi}}^{1/2} = \mathbf{R}_{\boldsymbol{\phi}}^{1/2}\mathbf{\Sigma}_{h-1}\mathbf{R}_{\boldsymbol{\psi}}^{1/2}$ is the cross-covariance matrix of $\mathbf{\Pi}_{h-1} = \mathbf{R}_{\boldsymbol{\psi}}^{1/2}\boldsymbol{\gamma}_{h-1}$ and $\mathbf{\Delta}_{h-1} = \mathbf{R}_{\boldsymbol{\phi}}^{1/2}\boldsymbol{\alpha}_{h-1}$.

3 Penalized approach

In FFPLS, the smoothness of the estimated β is influenced by the level of noise present in the observed functional data and the number of basis functions employed in the representation. Generally, a small number of basis functions tends to yield smooth

estimates, whereas a large number can result in rough estimates. The estimation of a smooth β is crucial for the accurate interpretation of the model.

In this section, we propose incorporating a penalty term into the FFPLS algorithm. Our emphasis is on controlling the roughness of the weights ω and ν, which directly influence the smoothness of the estimated β. The present approach can be viewed as an extension of the penalized solution for the scalar-on-function model introduced in [Aguilera et al., 2016].

Consider the norm $\|\cdot\|_\lambda$ associated with the inner product $\langle\cdot,\cdot\rangle_\lambda$:

$$\langle f, g\rangle_\lambda = \int_{\mathcal{D}} f(p)g(p)dp + \lambda\mathcal{P}(f,g), \quad \forall p \in \mathcal{D},$$

where $\mathcal{P}$ is a bilinear form that acts as a global roughness penalty on $f, g \in \mathcal{L}_2(\mathcal{D})$, with penalty parameter $\lambda > 0$. Then, the optimization problem in (2) can be rewritten as:

$$\max_{\omega\in\mathcal{L}_2(\mathcal{D}_X),\, \nu\in\mathcal{L}_2(\mathcal{D}_Y)} \frac{\operatorname{cov}^2\left(\int_{\mathcal{D}_X} X(p)\,\omega(p)\,dp, \int_{\mathcal{D}_Y} Y(q)\,\nu(q)\,dq\right)}{\|\omega\|^2_{\lambda_X}\|\nu\|^2_{\lambda_Y}}. \tag{8}$$

For any function $f \in \mathcal{L}_2(\mathcal{D})$, $\mathcal{D} \subset \mathbb{R}$, we can quantify its roughness by considering its d-order derivative $D^d f(t)$, $d \geq 1$. In this line, O'Sullivan considers the penalties of the form [O'Sullivan, 1986]:

$$\mathcal{P}_d(f,f) = \int_{\mathcal{D}} [D^d f(p)][D^d f(p)]dp, \quad p \in \mathcal{D},\ d \geq 1.$$

A common discretization of $\mathcal{P}_d$ is $\boldsymbol{P}^d = (\Delta^d)^T\Delta^d$, with Δ^d the matrix of d-order differences between adjacent basis coefficients. For example, the second differences ($d = 2$) for the coefficients of ω can be calculated as:

$$\Delta^2 w_j = \Delta(\Delta w_j) = \Delta(w_j - w_{j-1}) = w_j - 2w_{j-1} + w_{j-2}, \quad j = 3, \ldots, K.$$

This discretization is quite popular, mainly when using B-splines [Eilers and Marx, 1996, Eilers et al., 2015]. Other penalties can be useful depending on the nature of the data and the type of basis functions used to represent them. For example, in [Ramsay and Silverman, 2005], the authors usually rely on the "harmonic acceleration operator" for the case of periodic data:

$$\mathcal{P}(f,f) = \int_{\mathcal{D}} [Lf(p)][Lf(p)]dp, \quad p \in \mathcal{D},$$

where

$$Lf = \left(\frac{2\pi}{T}\right)^2 Df + D^3 f, \quad \mathcal{D} = [0,T] \subset \mathbb{R}.$$

Assume that ω and ν allow a meaningful basis representation as in equations (5)-(6), then:

$$\begin{aligned} \langle\omega,\omega\rangle_{\lambda_X} &= \mathbf{w}^T\mathbf{R}_{\boldsymbol{\phi}}\mathbf{w} + \lambda_X\mathbf{w}^T\boldsymbol{P}_X\mathbf{w} \\ \langle\nu,\nu\rangle_{\lambda_Y} &= \mathbf{v}^T\mathbf{R}_{\boldsymbol{\psi}}\mathbf{v} + \lambda_Y\mathbf{v}^T\boldsymbol{P}_Y\mathbf{v}, \end{aligned}$$

where $\boldsymbol{P}_X$ ($\boldsymbol{P}_Y$) is a $K \times K$ ($L \times L$) positive semi-definite matrix representing the discretization of the penalty operator $\mathcal{P}$ applied to ω (ν). Then, we can summarize the main contribution of this paper, the pFFPLS, in the following proposition.

Proposition 3.1. *Assume Y and X admit the expansion given in equations (3)-(4). Then, for any $\lambda_X > 0$ and $\lambda_Y > 0$:*

i) The pFFPLS regression of Y on X is equivalent to the PLS regression of the finite random vector $\boldsymbol{\gamma}$ on the random vector $\boldsymbol{\alpha}$ using the metrics:

$$M_1 = (\mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}})^T(\mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}}) \text{ and } M_2 = (\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}})^T(\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}),$$

respectively, in the space of $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$. The matrices $\mathbf{L}_{\lambda_X}$ and $\mathbf{L}_{\lambda_Y}$ are such that $\mathbf{L}_{\lambda_X}\mathbf{L}_{\lambda_X}^T = \mathbf{R}_{\boldsymbol{\phi}} + \lambda_X\mathbf{P}_X$ and $\mathbf{L}_{\lambda_Y}\mathbf{L}_{\lambda_Y}^T = \mathbf{R}_{\boldsymbol{\psi}} + \lambda_Y\mathbf{P}_Y$.

The equivalence is in the sense that at each step h of the functional and multivariate algorithms, $1 \leq h \leq K$, we obtain the same component t_h.

ii) If $\boldsymbol{\Theta}_2$ is the $L \times K$ matrix of regression coefficients of $\boldsymbol{\Pi}_2 = \mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}}\boldsymbol{\gamma}$ on $\boldsymbol{\Lambda}_2 = \mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha}$, obtained via multivariate PLS regression at step h, $1 \leq h \leq K$, i.e.:

$$\boldsymbol{\Pi}_2 = \boldsymbol{\Theta}_2\boldsymbol{\Lambda}_2 + \boldsymbol{\varepsilon},$$

then, the pFFPLS approximation at step h of the coefficient function $\beta(q,p)$ from (1) is given by:

$$\hat{\beta}(q,p) = \sum_{i=1}^{L}\sum_{j=1}^{K}\mathbf{B}_{i,j}\psi_i(q)\phi_j(p), \quad (q,p) \in \mathcal{D}_Y \times \mathcal{D}_X,$$

where $\mathbf{B} = \mathbf{L}_{\lambda_X}^{-1}\boldsymbol{\Theta}_2^T\mathbf{L}_{\lambda_Y}\mathbf{R}_{\boldsymbol{\psi}}^{-1}$.

Proof 3.1.

Let us consider the basis representation of Y and X as in equations (3)-(4). Consequently, let us also consider the basis expansion of ω and ν as in equations (5)-(6). Then, problem (8) can be rewritten as follows:

$$\begin{array}{c}\max \\ \mathbf{w} \in \mathbb{R}^K \\ \mathbf{v} \in \mathbb{R}^L\end{array} \quad \frac{\mathbf{w}^T\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}\mathbf{R}_{\boldsymbol{\psi}}\mathbf{v}\mathbf{v}^T\mathbf{R}_{\boldsymbol{\psi}}\boldsymbol{\Sigma}^T\mathbf{R}_{\boldsymbol{\phi}}\mathbf{w}}{(\mathbf{w}^T\mathbf{R}_{\boldsymbol{\phi}}\mathbf{w} + \lambda_X\mathbf{w}^T\mathbf{P}_X\mathbf{w})(\mathbf{v}^T\mathbf{R}_{\boldsymbol{\psi}}\mathbf{v} + \lambda_Y\mathbf{v}^T\mathbf{P}_Y\mathbf{v})}, \tag{9}$$

where $\mathbf{P}_X$ and $\mathbf{P}_Y$ are penalty matrices; $\boldsymbol{\Sigma}$ is a $K \times L$ matrix with entries $\boldsymbol{\Sigma}_{k,l} = \text{cov}(\alpha_k, \gamma_l)$, for $k = 1, \ldots, K$, $l = 1, \ldots, L$; and $\mathbf{w} = (w_1, \ldots, w_K)^T$, $\mathbf{v} = (v_1, \ldots, v_L)^T$ are the vectors of basis coefficients of ω and ν, respectively.

Let us assume the decomposition:

$$\begin{aligned}\mathbf{L}_{\lambda_X}\mathbf{L}_{\lambda_X}^T &= \mathbf{R}_{\boldsymbol{\phi}} + \lambda_X\mathbf{P}_X, \\ \mathbf{L}_{\lambda_Y}\mathbf{L}_{\lambda_Y}^T &= \mathbf{R}_{\boldsymbol{\psi}} + \lambda_Y\mathbf{P}_Y,\end{aligned}$$

where $\mathbf{L}_{\lambda_X}$ and $\mathbf{L}_{\lambda_Y}$ are the $K \times K$ and $L \times L$ square root matrices of $\mathbf{R}_{\boldsymbol{\phi}} + \lambda_X \mathbf{P}_X$ and $\mathbf{R}_{\boldsymbol{\psi}} + \lambda_Y \mathbf{P}_Y$, respectively. Then, problem (9) can be rewritten as:

$$\begin{array}{c} \max \\ \mathbf{w} \in \mathbb{R}^K \\ \mathbf{v} \in \mathbb{R}^L \end{array} \quad \frac{\mathbf{w}^T \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\Sigma} \mathbf{R}_{\boldsymbol{\psi}} \mathbf{v} \mathbf{v}^T \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}^T \mathbf{R}_{\boldsymbol{\phi}} \mathbf{w}}{[\mathbf{w}^T (\mathbf{L}_{\lambda_X} \mathbf{L}_{\lambda_X}^T) \mathbf{w}][\mathbf{v}^T (\mathbf{L}_{\lambda_Y} \mathbf{L}_{\lambda_Y}^T) \mathbf{v}]}. \tag{10}$$

Defining $\tilde{\mathbf{w}} = \mathbf{L}_{\lambda_X}^T \mathbf{w}$ (i.e. $\mathbf{w} = (\mathbf{L}_{\lambda_X}^{-1})^T \tilde{\mathbf{w}}$) and $\tilde{\mathbf{v}} = \mathbf{L}_{\lambda_Y}^T \mathbf{v}$ (i.e. $\mathbf{v} = (\mathbf{L}_{\lambda_Y}^{-1})^T \tilde{\mathbf{v}}$), problem (10) can be rewritten as:

$$\begin{array}{c} \max \\ \tilde{\mathbf{w}} \in \mathbb{R}^K \\ \tilde{\mathbf{v}} \in \mathbb{R}^L \end{array} \quad \frac{\tilde{\mathbf{w}}^T \mathbf{L}_{\lambda_X}^{-1} \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\Sigma} \mathbf{R}_{\boldsymbol{\psi}} (\mathbf{L}_{\lambda_Y}^{-1})^T \tilde{\mathbf{v}} \tilde{\mathbf{v}}^T \mathbf{L}_{\lambda_Y}^{-1} \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}^T \mathbf{R}_{\boldsymbol{\phi}} (\mathbf{L}_{\lambda_X}^{-1})^T \tilde{\mathbf{w}}}{(\tilde{\mathbf{w}}^T \tilde{\mathbf{w}})(\tilde{\mathbf{v}}^T \tilde{\mathbf{v}})}.$$

For a fixed $\tilde{\mathbf{v}}$, the solution $\tilde{\mathbf{w}}$ is the eigenvector associated with the largest eigenvalue of the eigenproblem:

$$\begin{gathered} \mathbf{L}_{\lambda_X}^{-1} \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\Sigma} \mathbf{R}_{\boldsymbol{\psi}} (\mathbf{L}_{\lambda_Y}^{-1})^T \frac{\tilde{\mathbf{v}} \tilde{\mathbf{v}}^T}{\tilde{\mathbf{v}}^T \tilde{\mathbf{v}}} \mathbf{L}_{\lambda_Y}^{-1} \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}^T \mathbf{R}_{\boldsymbol{\phi}} (\mathbf{L}_{\lambda_X}^{-1})^T \tilde{\mathbf{w}} = \mu \tilde{\mathbf{w}}, \\ \tilde{\mathbf{w}} \in \mathbb{R}^K, \langle \tilde{\mathbf{w}}, \tilde{\mathbf{w}} \rangle_{\mathbb{R}^K} = 1. \end{gathered}$$

Analogously, for a fixed $\tilde{\mathbf{w}}$, the solution $\tilde{\mathbf{v}}$ is the eigenvector associated with the largest eigenvalue of the problem:

$$\begin{gathered} (\mathbf{L}_{\lambda_Y}^{-1})^T \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}^T \mathbf{R}_{\boldsymbol{\phi}} \mathbf{L}_{\lambda_X}^{-1} \frac{\tilde{\mathbf{w}} \tilde{\mathbf{w}}^T}{\tilde{\mathbf{w}}^T \tilde{\mathbf{w}}} (\mathbf{L}_{\lambda_X}^{-1})^T \mathbf{R}_{\boldsymbol{\phi}}^T \boldsymbol{\Sigma} \mathbf{R}_{\boldsymbol{\psi}}^T \mathbf{L}_{\lambda_Y}^{-1} \tilde{\mathbf{v}} = \mu \tilde{\mathbf{v}}, \\ \tilde{\mathbf{v}} \in \mathbb{R}^L, \langle \tilde{\mathbf{v}}, \tilde{\mathbf{v}} \rangle_{\mathbb{R}^K} = 1. \end{gathered}$$

The first pFFPLS step is completed by computing the residuals X_1 and Y_1 of the ordinary linear regression of $X_0 = X$ and $Y_0 = Y$ on $t_1 = \mathbf{w}^T \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\alpha} = \tilde{\mathbf{w}}^T (\mathbf{L}_{\lambda_X}^{-1})^T \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\alpha}$.

In general, the coefficients of the weight function that defines the h-th component ($h > 1$) is the solution to the eigenproblem:

$$\begin{gathered} \mathbf{L}_{\lambda_X}^{-1} \mathbf{R}_{\boldsymbol{\phi}} \boldsymbol{\Sigma}_{h-1} \mathbf{R}_{\boldsymbol{\psi}} (\mathbf{L}_{\lambda_Y}^{-1})^T \mathbf{L}_{\lambda_Y}^{-1} \frac{\tilde{\mathbf{v}}_h \tilde{\mathbf{v}}_h^T}{\tilde{\mathbf{v}}_h^T \tilde{\mathbf{v}}_h} \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}_{h-1}^T \mathbf{R}_{\boldsymbol{\phi}} (\mathbf{L}_{\lambda_X}^{-1}) \tilde{\mathbf{w}}_h = \lambda \tilde{\mathbf{w}}_h, \\ \tilde{\mathbf{w}}_h \in \mathbb{R}^K, \langle \tilde{\mathbf{w}}_h, \tilde{\mathbf{w}}_h \rangle_{\mathbb{R}^K} = 1, \end{gathered}$$

where $\boldsymbol{\Sigma}_{h-1}$ is the $K \times L$ covariance matrix between the vectors of basis coefficients $\boldsymbol{\alpha}_{h-1}$ and $\boldsymbol{\gamma}_{h-1}$ of X_{h-1} and Y_{h-1}, respectively. Analogously, the vector of coefficients for the weight ν_h is $\mathbf{v}_h = (\mathbf{L}_{\lambda_Y}^{-1})^T \tilde{\mathbf{v}}_h$, where $\tilde{\mathbf{v}}_h$ is the eigenvector associated with the largest eigenvalue solving:

$$\begin{gathered} (\mathbf{L}_{\lambda_Y}^{-1})^T \mathbf{R}_{\boldsymbol{\psi}} \boldsymbol{\Sigma}_{h-1}^T \mathbf{R}_{\boldsymbol{\phi}} \mathbf{L}_{\lambda_X}^{-1} \frac{\tilde{\mathbf{w}}_h \tilde{\mathbf{w}}_h^T}{\tilde{\mathbf{w}}_h^T \tilde{\mathbf{w}}_h} (\mathbf{L}_{\lambda_X}^{-1})^T \mathbf{R}_{\boldsymbol{\phi}}^T \boldsymbol{\Sigma}_{h-1} \mathbf{R}_{\boldsymbol{\psi}}^T \mathbf{L}_{\lambda_Y}^{-1} \tilde{\mathbf{v}}_h = \mu \tilde{\mathbf{v}}_h, \\ \tilde{\mathbf{v}}_h \in \mathbb{R}^L, \langle \tilde{\mathbf{v}}_h, \tilde{\mathbf{v}}_h \rangle_{\mathbb{R}^K} = 1. \end{gathered}$$

Notice that $\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}_{h-1}\mathbf{R}_{\boldsymbol{\psi}}(\mathbf{L}_{\lambda_Y}^{-1})^T$ is the cross-covariance matrix of $\boldsymbol{\Pi}_{h-1} = \mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}}\boldsymbol{\gamma}_{h-1}$ and $\boldsymbol{\Delta}_{h-1} = \mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha}_{h-1}$. Also, the h-th component is computed through:

$$t_h = \mathbf{w}_h^T\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha}_{h-1} = \tilde{\mathbf{w}}_h^T(\mathbf{L}_{\lambda_X}^{-1})^T\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha}_{h-1}.$$

By analogy to the FFPLS by [Preda and Schiltz, 2011], our penalized approach is reduced to a multivariate PLS regression of $\boldsymbol{\Pi}_2 = \mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}}\boldsymbol{\gamma}$ on $\boldsymbol{\Delta}_2 = \mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha}$. In other words, we can say that pFFPLS is reduced to a multivariate PLS of $\boldsymbol{\gamma}$ on $\boldsymbol{\alpha}$ with the metrics $M_1 = (\mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}})^T(\mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}})$ and $M_2 = (\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}})^T(\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}})$ in the space of $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$, respectively.

The approximation of the coefficient function β in terms of the multivariate PLS coefficient $\boldsymbol{\Theta}_2$ comes from the following reasoning:

$$\begin{aligned}
\boldsymbol{\Pi}_2 &= \boldsymbol{\Theta}_2\boldsymbol{\Lambda}_2 + \boldsymbol{\varepsilon} \\
\mathbf{L}_{\lambda_Y}^{-1}\mathbf{R}_{\boldsymbol{\psi}}\boldsymbol{\gamma} &= \boldsymbol{\Theta}_2\mathbf{L}_{\lambda_X}^{-1}\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\
\boldsymbol{\gamma} &= \left(\mathbf{R}_{\boldsymbol{\psi}}^{-1}\mathbf{L}_{\lambda_Y}\boldsymbol{\Theta}_2\mathbf{L}_{\lambda_X}^{-1}\right)\mathbf{R}_{\boldsymbol{\phi}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\
\boldsymbol{\gamma} &= \mathbf{B}^T\,\mathbf{R}_{\boldsymbol{\phi}}\,\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\
\boldsymbol{\gamma} &= \mathbf{B}^T\left(\int_{\mathcal{D}_X}\boldsymbol{\phi}(p)\boldsymbol{\phi}^T(p)dp\right)\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\
\boldsymbol{\gamma} &= \int_{\mathcal{D}_X}\mathbf{B}^T\,\boldsymbol{\phi}(p)\,\boldsymbol{\phi}^T(p)\,\boldsymbol{\alpha}\,dp + \boldsymbol{\varepsilon} \\
\boldsymbol{\psi}^T(q)\boldsymbol{\gamma} &= \int_{\mathcal{D}_X}\boldsymbol{\psi}^T(q)\,\mathbf{B}^T\,\boldsymbol{\phi}(p)\;\;\boldsymbol{\phi}^T(p)\,\boldsymbol{\alpha}\,dp + \boldsymbol{\varepsilon} \\
Y(q) &= \int_{\mathcal{D}_X}\beta(q,p)\,X(p)\,dp + \boldsymbol{\varepsilon} \quad \square
\end{aligned}$$

3.1 Sample estimation

Consider a sample of n centered functional data x_i from X, and y_i from Y, $i = 1, \ldots, n$. The functional predictors x_i are observed at a common set of m_x discrete observation points $p_j \in \mathcal{D}_X$, where $j = 1, \ldots, m_x$. The functional responses y_i are observed at a common set of m_y discrete observation points $q_k \in \mathcal{D}_Y$, where $k = 1, \ldots, m_y$. The data for all samples are then organized into an $n \times m_x$ matrix $\mathbf{X}$ with entries $x_{ij} = x_i(p_j)$, and an $n \times m_y$ matrix $\mathbf{Y}$ with entries $y_{ik} = y_i(q_k)$, for $i = 1, \ldots, n$; $j = 1, \ldots, m_x$; and $k = 1, \ldots, m_y$.

Let $\phi_1, \ldots, \phi_K$ be a collection of basis functions defined on $\mathcal{D}_X$, and let $\boldsymbol{\Phi}$ be the $m_x \times K$ matrix with entries $(\phi_k(p_j))$:

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_1(p_1) & \cdots & \phi_K(p_1) \\ \vdots & \ddots & \vdots \\ \phi_1(p_{m_x}) & \cdots & \phi_K(p_{m_x}) \end{bmatrix}.$$

Also, let $\psi_1, \ldots, \psi_L$ be a collection of basis functions defined on $\mathcal{D}_Y$, and let $\mathbf{\Psi}$ be the $m_y \times L$ matrix with entries $(\psi_l(q_k))$:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1(q_1) & \cdots & \psi_L(q_1) \\ \vdots & \ddots & \vdots \\ \psi_1(q_{m_y}) & \cdots & \psi_L(q_{m_y}) \end{bmatrix}.$$

The pFFPLS algorithm consists of the following steps:

1. Use the least square criterion to estimate the vectors of coefficients for each sample path using sets of K and L basis functions for X and Y, respectively:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_i &= \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\boldsymbol{x}_i \\ \hat{\boldsymbol{\gamma}}_i &= \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^T\boldsymbol{y}_i, \end{aligned}$$

   for each sample $i = 1, \ldots, n$, and $\boldsymbol{x}_i \in \mathbb{R}^{m_x}$, $\boldsymbol{y}_i \in \mathbb{R}^{m_y}$ the transposes of the first rows of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

2. Let $\mathbf{A}$ be the $n \times K$ matrix of basis coefficients for X and $\mathbf{G}$ be the $n \times L$ matrix for Y. In both cases, rows correspond to samples, and columns correspond to the number of basis functions. Then, pFFPLS is reduced to the multivariate PLS of:

$$\mathbf{G}\,\mathbf{R}_{\boldsymbol{\psi}}\,\mathbf{L}_{\lambda_Y}^{-1} \;\text{ on }\; \mathbf{A}\,\mathbf{R}_{\boldsymbol{\phi}}\,\mathbf{L}_{\lambda_X}^{-1}. \tag{11}$$

3. The $K \times L$ coefficient matrix $\mathbf{\Theta}_2^T$ relates the multivariate response and predictor in equation (11) through:

$$\mathbf{G}\,\mathbf{R}_{\boldsymbol{\psi}}\,\mathbf{L}_{\lambda_Y}^{-1} = \mathbf{A}\,\mathbf{R}_{\boldsymbol{\phi}}\,\mathbf{L}_{\lambda_X}^{-1}\,\mathbf{\Theta}_2^T + \mathbf{E}, \tag{12}$$

   where $\mathbf{E}$ is an $n \times L$ matrix of residuals. If we define $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ to be the vectors of basis functions for X and Y, respectively, and let the $K \times L$ matrix $\mathbf{B} = \mathbf{L}_{\lambda_X}^{-1}\mathbf{\Theta}_2^T\mathbf{L}_{\lambda_Y}\mathbf{R}_{\boldsymbol{\psi}}^{-1}$, then equation (12) can be rewritten as:

$$\begin{aligned} \mathbf{G} &= \mathbf{A}\,\mathbf{R}_{\boldsymbol{\phi}}\left(\mathbf{L}_{\lambda_X}^{-1}\,\mathbf{\Theta}_2^T\,\mathbf{L}_{\lambda_Y}\,\mathbf{R}_{\boldsymbol{\psi}}^{-1}\right) + \mathbf{E}, \\ \mathbf{G} &= \mathbf{A}\,\mathbf{R}_{\boldsymbol{\phi}}\,\mathbf{B} + \mathbf{E}, \\ \mathbf{G}\,\boldsymbol{\psi}(q) &= \int_{\mathcal{D}_X} \mathbf{A}\,\boldsymbol{\phi}(p)\,\boldsymbol{\phi}^T(p)\,\mathbf{B}\,\boldsymbol{\psi}(q)\,dp + \mathbf{E}, \quad q \in \mathcal{D}_Y. \end{aligned}$$

   From this equation, we obtain our estimation of the coefficient function:

$$\beta(q,p) = \boldsymbol{\phi}^T(p)\,\mathbf{B}\,\boldsymbol{\psi}(q), \quad (q,p) \in \mathcal{D}_Y \times \mathcal{D}_X.$$

3.2 Parameters selection

The performance of the pFFPLS approach is determined by three key features: the number of bases employed for data representation (K, L), the values of the penalty parameters λ_X and λ_Y, and the number of components used to summarize the data ($H \geq 1$). When using a small number of bases, the estimates tend to be smoother compared to using a larger number. Selecting appropriate values for the penalty parameters is a common consideration in all penalized problems, and it is typically addressed through cross-validation (CV) procedures. Determining the optimal number of components is a typical step in PLS and Principal Components techniques and can be approached using empirical methods such as the "elbow rule."

In this paper, we employ κ-fold CV to select the optimal values for the smoothing parameters. We construct a two-dimensional grid with various potential values for the pair (λ_X, λ_Y). The data is then divided into $2 \leq \kappa \leq n$ subsets (folds), where one fold is used for testing while the remaining $\kappa - 1$ folds are utilized for model fitting. Let $S_k,\ k = 1, \ldots, \kappa$, denote the k-th fold. For each (λ_X, λ_Y) in the grid and for every $1 \leq h \leq H$, we can estimate the test Integrated Mean Square Error (IMSE) for the fitted model:

$$\text{IMSE}_k(\lambda_X, \lambda_Y, h) = \frac{1}{\text{length}(\mathcal{D}_Y)} \int_{\mathcal{D}_Y} \frac{1}{|S_k|} \sum_{i \in S_k} \left(\hat{y}_i^{[-k]}(q) - y_i(q) \right)^2 \, dq.$$

In the formula, $|S_k|$ represents the number of statistical units in the fold. The term $\hat{y}_i^{[-k]}(q)$ denotes the predicted response function for observation $i \in S_k$ using a model fitted with all other observations $i \notin S_k$. The estimated κ-fold cross-validated error (CVE) for a specific $(\lambda_X, \lambda_Y, h)$ triplet is computed as the average:

$$\text{CVE}(\lambda_X, \lambda_Y, h) = \frac{1}{\kappa} \sum_{k=1}^{\kappa} \text{IMSE}_k(\lambda_X, \lambda_Y, h).$$

Then, for each h, the optimal pair $(\lambda_X^*, \lambda_Y^*)$ is determined by minimizing the CVE:

$$(\lambda_X^*, \lambda_Y^*) = \arg \min_{\lambda_X, \lambda_Y} \text{CVE}(\lambda_X, \lambda_Y, h).$$

By plotting h against the $\text{CVE}(\lambda_X^*, \lambda_Y^*, h)$, we can identify the best number of components H based on the curve's shape. The ideal choice for H is where the curve exhibits an "elbow" shape, transitioning from a steep slope to a flatter region. In general, the goal is to achieve a low CVE with a minimal number of components.

4 Simulation study

This section considers a simulation scenario similar to the one presented in [Preda and Schiltz, 2011]. We generate a sample of $n = 100$ observations of a functional

predictor:

$$x_i(p) = \sum_{k=1}^{K} \alpha_{i,k}\phi_k(p) = \boldsymbol{\alpha}_i^T \boldsymbol{\phi}(p), \quad p \in [0,1], \quad i = 1, \ldots, 100, \tag{13}$$

where $\{\alpha_{i,k}\}_{i,k}$ are independent and identically distributed (iid) uniform random variables in the interval $[-1,1]$, and $\{\phi_k\}_{k=1}^K$ is a collection of K cubic B-splines basis functions defined on $[0,1]$.

If we let $\mu_X(p)$ be the mean of the generated predictor curves and β_{true} a fixed coefficient function (a surface in this case), then we compute the response values y_i^ϵ by numerical approximation of the integral:

$$\begin{aligned} y_i^\epsilon(q) &= \int_0^1 \left(x_i(p) - \mu_X(p)\right) \beta_{\text{true}}(q,p)dp + \epsilon_{i,q}, \\ &= y_i(q) + \epsilon_{i,q}, \quad q \in [0,1], \quad i = 1, \ldots, 100, \end{aligned}$$

where $y_i(q)$ is the "clean" response and $y_i^\epsilon(q)$ adds iid error terms $\epsilon_{i,q} \sim N(0, \sigma_q^2)$, for any $q \in [0,1]$. The variance of the errors σ_q^2 is chosen such that the true model's squared correlation function $R^2(q)$ is approximately 0.9 for any $q \in [0,1]$. We use the definition of $R^2(q)$ provided in [Ramsay and Silverman, 2005]:

$$R^2(q) = 1 - \frac{\sum_i \{y_i^\epsilon(q) - y_i(q)\}^2}{\sum_i \{y_i(q) - \overline{y}(q)\}^2}.$$

The numerical approximation of $y_i^\epsilon(q)$ is obtained by assuming the response curves can be represented using a set $\{\psi_l\}_{l=1}^L$ of L cubic B-spline basis functions defined on $[0,1]$.

We consider two different coefficient functions β_{true}. The first function is a smooth and symmetric surface:

$$\beta_1(q,p) = (q-p)^2, \quad (q,p) \in [0,1]^2,$$

as in [Preda and Schiltz, 2011]. The second function is a smooth "monkey saddle":

$$\beta_2(q,p) = (4p-2)^3 - 3(4p-2)(4q-2)^2, \quad (q,p) \in [0,1]^2.$$

In our simulation, we consider two different settings based on the number of basis functions used to represent the predictor and response functions. In the first setting, we use a small number of basis functions, specifically $K = 7$ for the predictor and $L = 5$ for the response, following the setup used in [Preda and Schiltz, 2011]. We use more basis functions in the second setting, $K = L = 40$, for both the predictor and response functions. Figure 1 shows a sample of 100 simulated data using β_2 and $K = L = 40$.

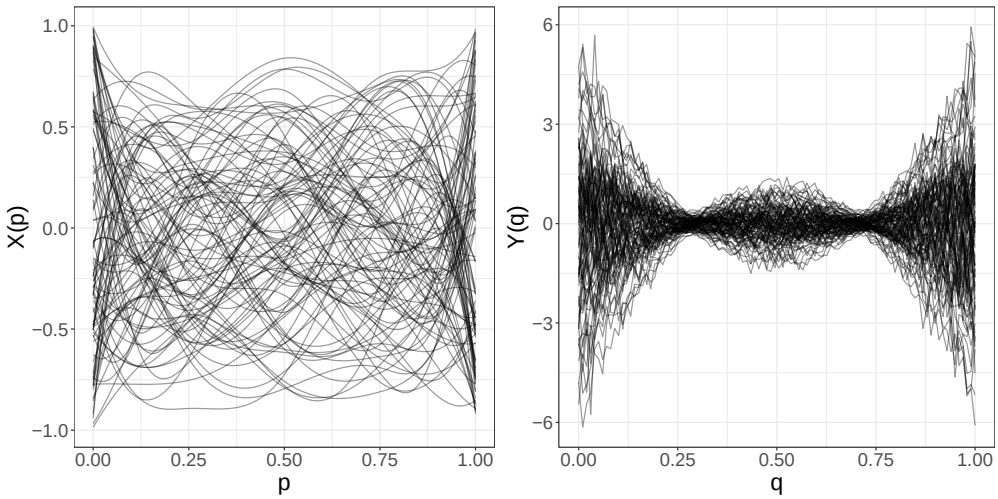


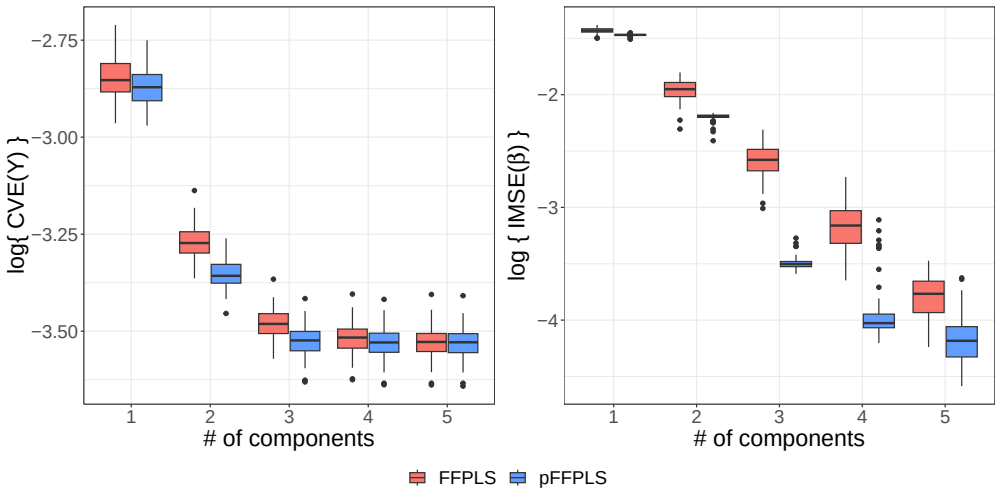
Figure 1: A sample of 100 simulated data pairs (X, Y) using β_2 and $K = L = 40$.

In our comparison between pFFPLS and FFPLS, we use a 5-fold CVE relative to the response variable Y and an IMSE relative to the coefficient function β_{true}, i.e.:

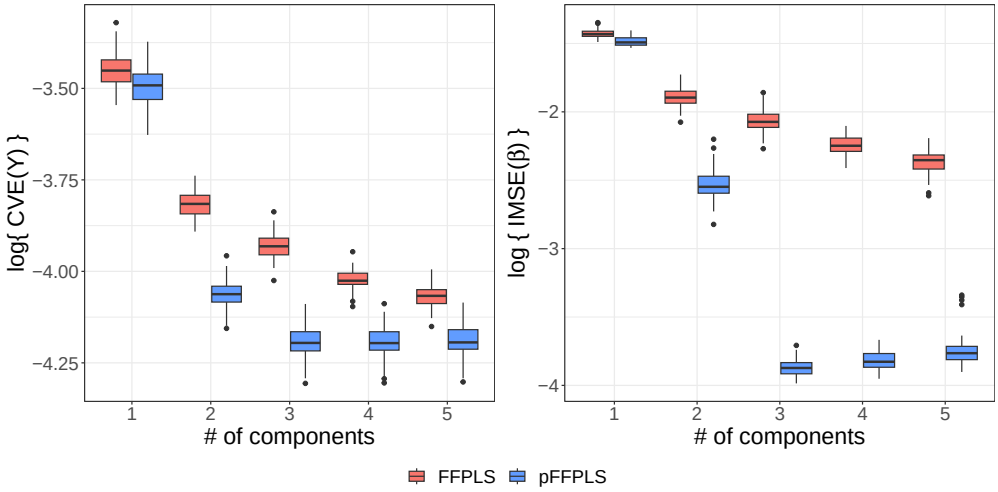
$$\text{IMSE}(\beta) = \frac{1}{\text{area}(\mathcal{D}_Y \times \mathcal{D}_X)} \int_{\mathcal{D}_Y} \int_{\mathcal{D}_X} \left(\hat{\beta}(q,p) - \beta_{\text{true}}(q,p)\right)^2 dp\, dq,$$

where $\mathcal{D}_Y$ and $\mathcal{D}_X$ represent the domains of the response and predictor functions, respectively. The IMSE is computed using all the available samples. The CVE and IMSE are calculated using the best pair of penalties $(\lambda_X^*, \lambda_Y^*)$ chosen from a 10×10 grid of possible values. The grid spans a range from 10^{-6} to 10^{12}, with the values generated in logarithmic scale (10^a) where a is a sequence of 10 equally spaced integers in the range $[-6, 12]$. To ensure robustness and variability estimation, the simulations are repeated 100 times.

Figure 2 displays the resulting CVEs and IMSEs for β_1 on a logarithmic scale. The boxplots clearly illustrate the superior performance of pFFPLS compared to FFPLS. For $K = 7, L = 5$, Figure 2a shows that both methods exhibit a similar distribution of CVEs, particularly when more than three components are used. However, pFFPLS consistently achieves smaller IMSEs. The differences become more pronounced when a larger number of bases is employed ($K = L = 40$), as depicted in Figure 2b. Applying the elbow rule with the CVEs, three components appear sufficient for pFFPLS to perform well. On the other hand, FFPLS requires more than three components when $K = L = 40$.



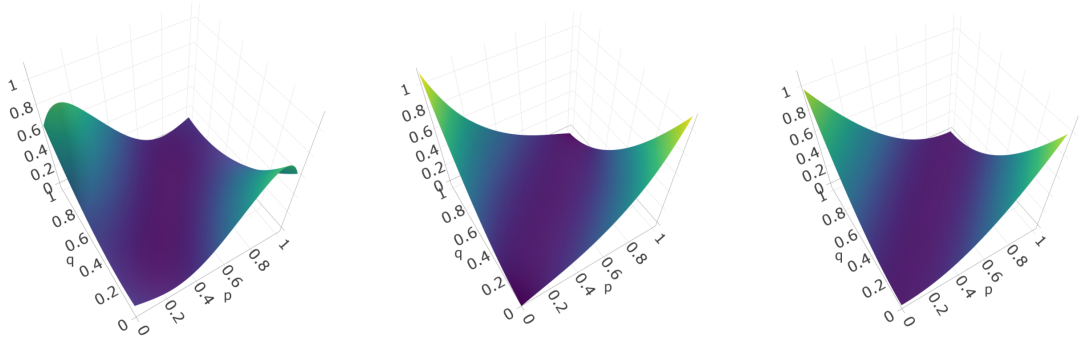
(a) Setting 1: $K = 7,\ L = 5$.



(b) Setting 2: $K = L = 40$.

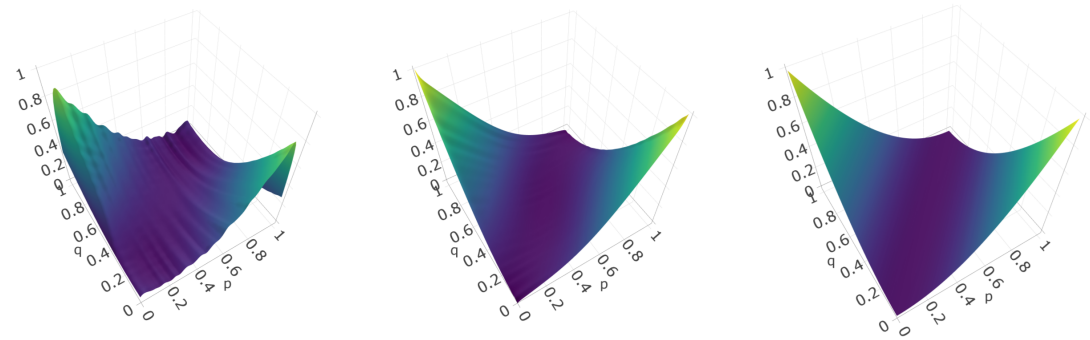
Figure 2: Comparison of the CVE and IMSE(β_1) when predicting the response variable Y and the true coefficient function β_1, respectively. Plot (a) corresponds to the first setting $K = 7, L = 5$, while plot (b) corresponds to the second setting $K = L = 40$.

Figures 3 and 4 showcase the means of the estimated $\hat{\beta}_1$ for both methods compared to the true β_1. The estimates provided by pFFPLS closely resemble the shape of β_1, but FFPLS fails to reproduce its shape. For example, under the first setting, $K = 7,\ L = 5$ in Figure 3, FFPLS over smooth $\hat{\beta}_1$ when $(q, p) \to (0, 1)$ and $(q, p) \to (1, 0)$. Under the second setting, $K = L = 40$ in Figure 4, the estimates for FFPLS are not smooth along the q-axis.



(a) FFPLS (b) pFFPLS (c) True β_1

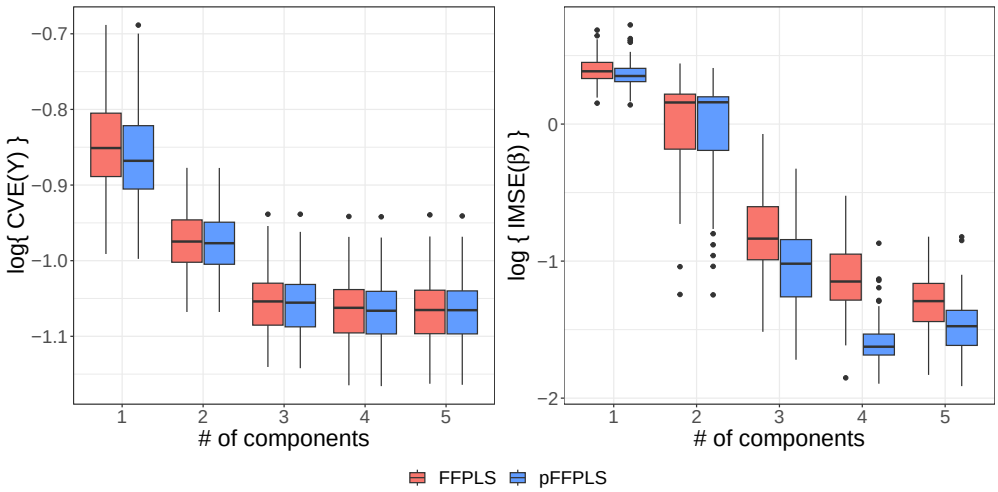
Figure 3: Mean of the estimated coefficient functions $\hat{\beta}_1$ for FFPLS (a) versus pFFPLS (b) using three components under the first setting $K = 7$, $L = 5$. The estimates are compared with the true β_1 in panel (c).



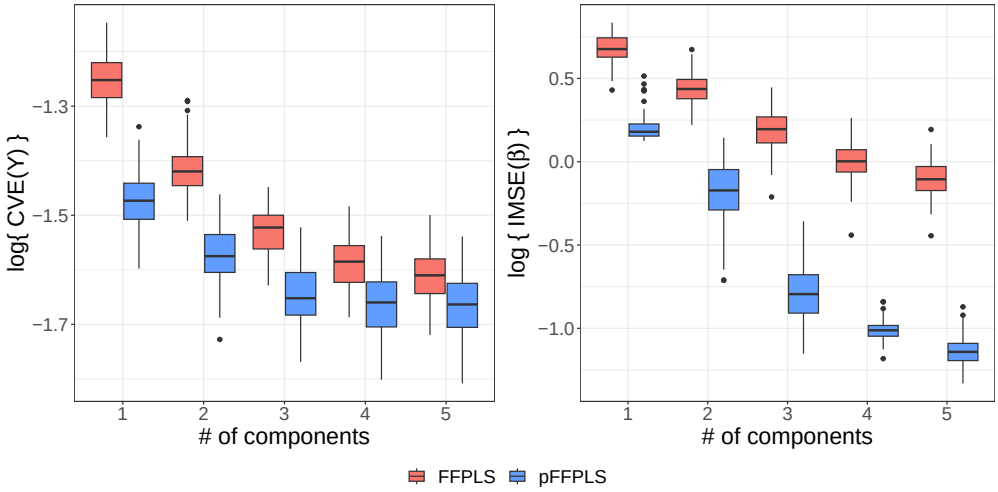
(a) FFPLS (b) pFFPLS (c) True β_1

Figure 4: Mean of the estimated coefficient functions $\hat{\beta}_1$ for FFPLS (a) versus pFFPLS (b) using three components under the second setting $K = L = 40$. The estimates are compared with the true β_1 in panel (c).

Figure 5 compares the CVEs and IMSEs for β_2 on a logarithmic scale. We observe a similar pattern to the comparisons against the true β_1. pFFPLS consistently exhibits the lowest errors, particularly for a larger number of bases $K = L = 40$. When considering the CVEs, three components appear sufficient for pFFPLS to achieve good performance. On the other hand, FFPLS requires more than three components under the second setting.



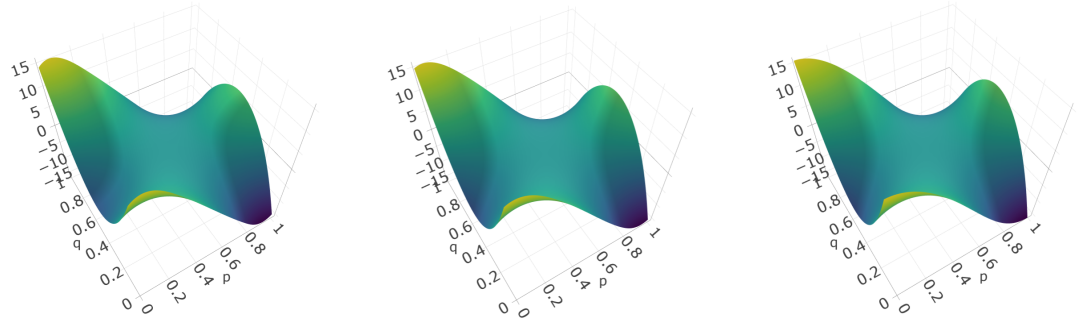
(a) Setting 1: $K = 7,\ L = 5$.



(b) Setting 2: $K = L = 40$.

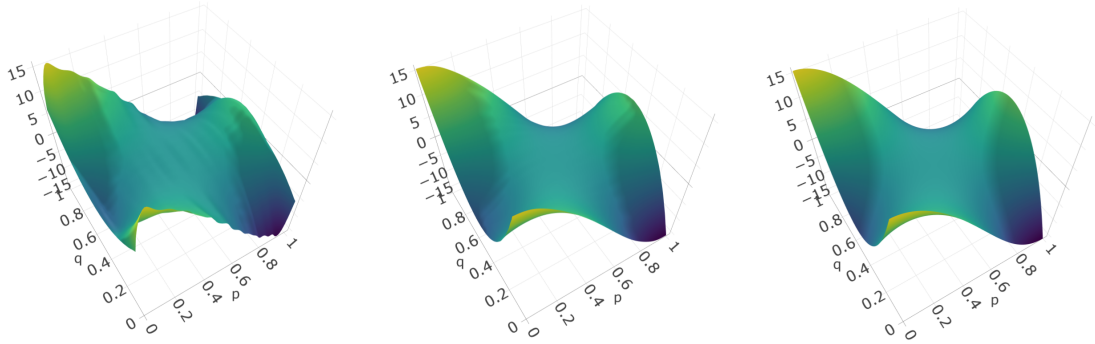
Figure 5: Comparison of the CVE and IMSE(β_2) when predicting the response variable Y and the true coefficient function β_2, respectively. Plot (a) corresponds to the first setting $K = 7, L = 5$, while plot (b) corresponds to the second setting $K = L = 40$.

Figure 6 shows no substantial differences in the estimated $\hat{\beta}_2$ for both methods when $K = 7,\ L = 5$. However, when $K = L = 40$, the estimates for FFPLS are rough, particularly along the q-axis, as depicted in Figure 7. On the other hand, pFFPLS provides very good estimates, no matter the number of bases.



(a) FFPLS (b) pFFPLS (c) True β_1

Figure 6: Mean of the estimated coefficient functions $\hat{\beta}_2$ for FFPLS (a) versus pFFPLS (b) using three components under the first setting $K = 7$, $L = 5$. The estimates are compared with the true β_2 in panel (c).



(a) FFPLS (b) pFFPLS (c) True β_1

Figure 7: Mean of the estimated coefficient functions $\hat{\beta}_2$ for FFPLS (a) versus pFFPLS (b) using three components under the second setting $K = L = 40$. The estimates are compared with the true β_2 in panel (c).

5 Case studies

5.1 Predicting log precipitation from temperature in Canada

In this subsection, we examine the performance of our pFFPLS approach using the Canadian weather dataset, described in [Ramsay and Silverman, 2005]. The dataset is accessible in the R package `fda` [Ramsay, 2023] and contains information on daily temperature and precipitation at 35 locations in Canada, averaged over the years 1960 to 1994. Our objective is to predict the complete log daily precipitation profile $y_i(q)$, where q ranges from 1 to 365, for each weather station $i = 1, \ldots, 35$, using its corresponding yearly temperature profile $x_i(p)$, with p ranging from 1 to 365.

The Canadian weather dataset is preprocessed according to the methodology outlined in Chapter 16 of [Ramsay and Silverman, 2005]. Specifically, we use 65 Fourier basis functions and a roughness penalty smoother to remove localized variations. The penalty parameters for the Fourier series expansion are determined through generalized cross-validation, resulting in values of approximately 10^6 for X (temperature) and $10^{1.5}$ for Y (log precipitation). Chapter 5 of [Ramsay et al., 2009] shows the implementation details for this procedure in R. Figure 8 displays the preprocessed temperature (X) and log precipitation (Y) profiles for the 35 weather stations in Canada.

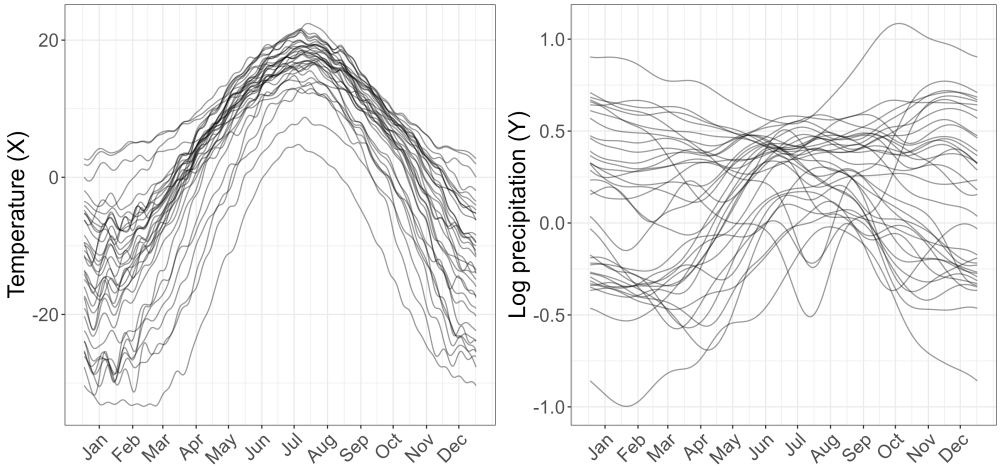


Figure 8: Preprocessed temperature (X) and log precipitation (Y) profiles for the 35 Canadian weather stations.

The function-on-function model described in Ramsay and Silverman's book explores the effect of the number of bases used to represent the data. In summary, they show that using $K = L = 65$ Fourier basis to fit the model (using the preprocessed data) produces rougher estimates of the coefficient function, compared to the resulting $\hat{\beta}$ when the number of bases is truncated to $K = L = 7$. Using pFFPLS, one can skip the tuning of K and L by fixing them to be "large" enough and enforcing smoothness through applying a roughness penalty, as described in Section 3. In this illustration, we fix $K = L = 65$ and let $\boldsymbol{P}$ be the penalty matrix associated with the harmonic acceleration operator.

Figure 9 presents the root of the CVEs obtained by repeating 30 times a 5-fold cross-validation. We observe that just three components are needed. No differences are observed between FFPLS and pFFPLS because the CV procedure selects small penalty values of λ_X and λ_Y, making both methods equivalent. We also consider a second scenario for pFFPLS in which we fix the penalties to be large. pFFPLS (with $\lambda_X = \lambda_Y = 10^{12}$) yields a higher CVE, but this loss in predictive performance is offset by the gain in interpretability of the estimated coefficient function.

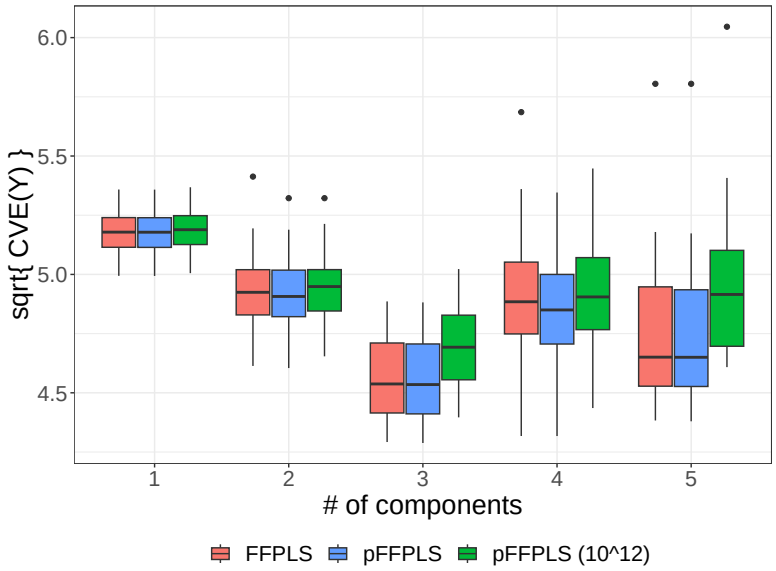
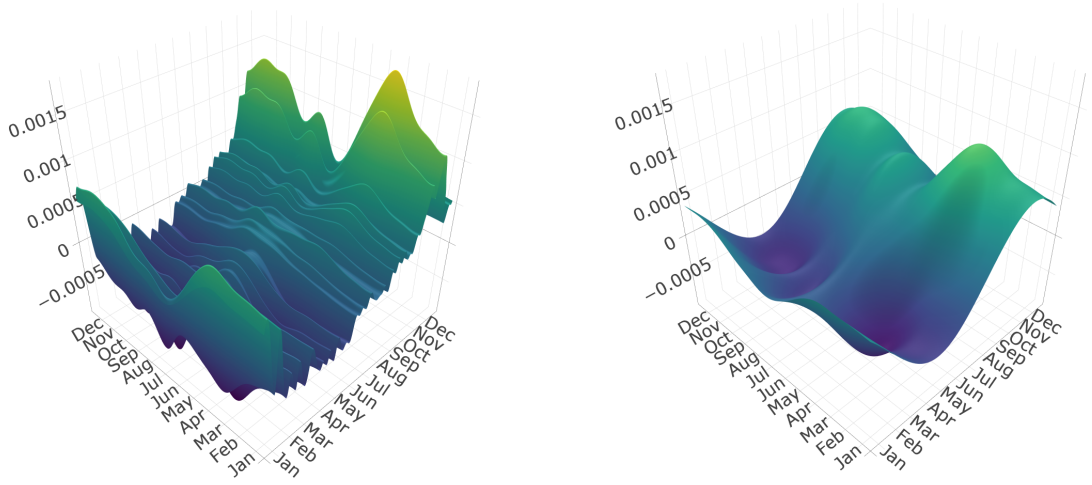


Figure 9: Repeated 5-fold cross-validated errors (showing roots) when predicting the yearly log precipitation Y of each station from its temperature profile X.

Figure 10 illustrates the estimated coefficient functions $\hat{\beta}$ for FFPLS (Figure 10a) and pFFPLS with the fixed penalties $\lambda_X = \lambda_Y = 10^{12}$ (Figure 10b) using three components. The penalized model, implemented in pFFPLS ($\lambda_X = \lambda_Y = 10^{12}$), provides smooth estimates of the coefficient function even when many bases are used. Furthermore, Figure 11 presents a contour plot of the 3D $\hat{\beta}$ for the penalized model in pFFPLS ($\lambda_X = \lambda_Y = 10^{12}$). The $\hat{\beta}(q, p)$ values have been multiplied by 10^4 for easier interpretation. From the plot, several important associations can be observed. Firstly, temperatures from February to June show a negative association with log precipitation throughout the year. This negative association is particularly strong in the period of March to April. Conversely, a strong positive association exists between temperatures in September to November and log precipitation throughout the year. The highest peak for this positive association is also observed from March to April.



(a) FFPLS (b) pFFPLS ($\lambda_X = \lambda_Y = 10^{12}$)

Figure 10: Estimated $\hat{\beta}$ surfaces for the log precipitation versus temperature model using the Canada weather stations. Both PLS models use three components.

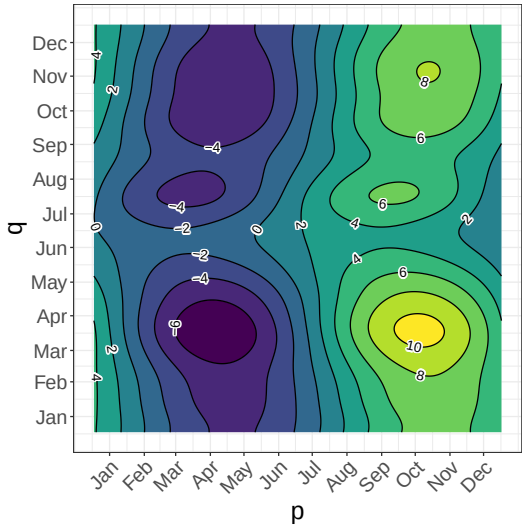


Figure 11: Estimated $\hat{\beta}$ using pFFPLS ($\lambda_X = \lambda_Y = 10^{12}$) with three components. The values over the contour lines were multiplied by 10^4 to facilitate the interpretation. The horizontal p-axis represents the argument of the temperature profiles (X). The vertical q-axis represents the argument of the log precipitations profiles (Y).

5.2 Hip and knee angles in gait cycle

We now focus on a dataset from the Motion Analysis Laboratory at Children's Hospital, San Diego, originally studied in [Olshen et al., 1989]. The dataset is freely available under the command `gait` in the R package `fda` [Ramsay, 2023] and has been widely analyzed in [Ramsay and Silverman, 2005]. It consists of the angles formed by the hip and knee of each of 39 children over each child's gait cycle, as depicted in Figure 12. For each subject, 20 records for each curve are observed on the same time interval $[0; 20]$. We aim to predict the hip angle profiles from the knee angle profiles.

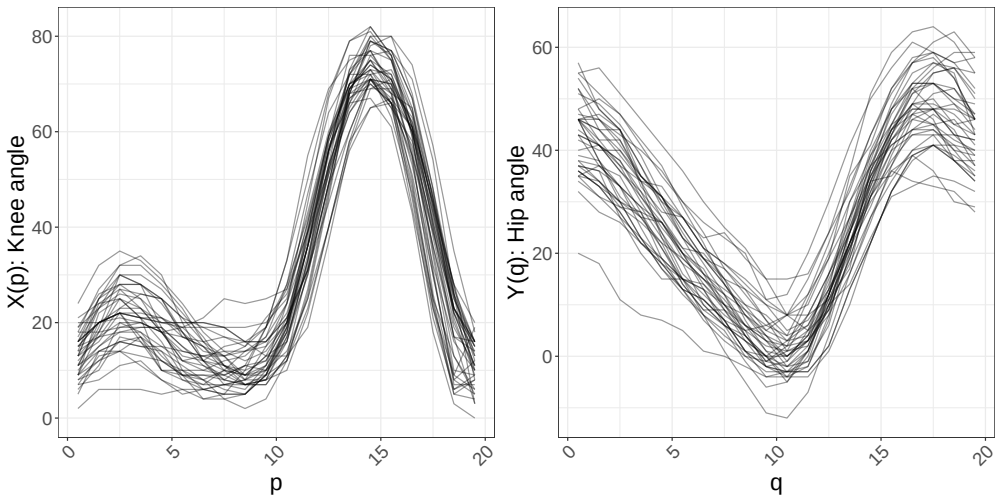


Figure 12: Knee angles (X) and hip angles (Y) over a time period that covers the gait cycle of 39 children.

The response and predictor curves correspond to periodic curves. The cycles begin and end at the point where the heel of the limb under observation strikes the ground. Therefore, it makes sense to consider Fourier bases to model this data. For this illustration, we set many bases, $K = L = 41$, considering that the curves are recorded in just 20 instants of time. Again, the harmonic accelerator operator is a feasible choice for penalizing the PLS weights.

Figure 13 shows the 30-times-repeated 5-fold CVEs (in square roots) of FFPLS versus pFFPLS. Our penalized approach provides smaller errors, particularly when more components are added. Following the elbow rule, three components are enough for both PLS approaches to obtain good predictions.

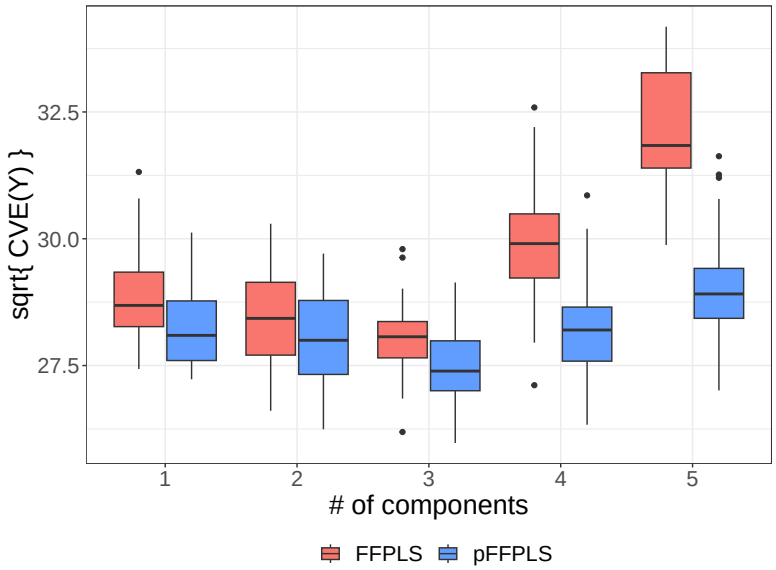
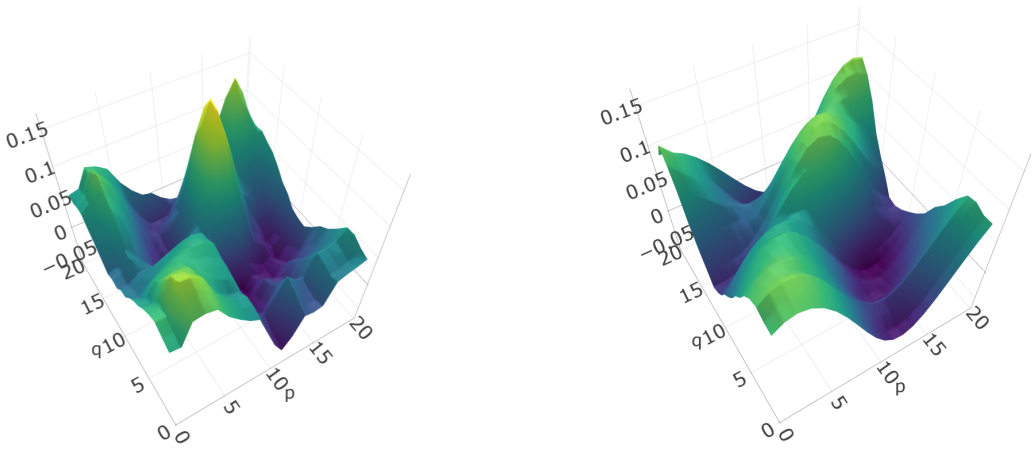


Figure 13: Repeated 5-fold cross-validated errors (showing roots) when predicting the hip angle profile Y of each subject from its knee angle profile X.

Figures 14a and 14b depict the estimated $\hat{\beta}$ for FFPLS and pFFPLS, respectively. Although both surfaces present a similar shape, it is clear that pFFPLS smooths out the peaks and the rough areas provided by FFPLS. A closer look at the contour plot of the smooth $\hat{\beta}(q,p)$ in Figure 15 indicates that it is approximately symmetrical with respect to the $q = p$ diagonal. The strongest positive association between hip angle and knee angle is observed over the diagonal $q = p$, particularly at the beginning and end of the gait cycle, i.e., $(q,p) \in [0;5] \times [0;5]$ and $(q,p) \in [12.5;20] \times [12.5;20]$. On the other hand, the strongest negative association between hip angle and knee angle is observed below the $q = p$ diagonal, mainly around $(q,p) \in [0;10] \times [10;18]$.



(a) FFPLS (b) pFFPLS

Figure 14: Estimated $\hat{\beta}$ surfaces for the hip angle versus knee angle model using the gait cycle data. The coefficient functions correspond to models with three components.

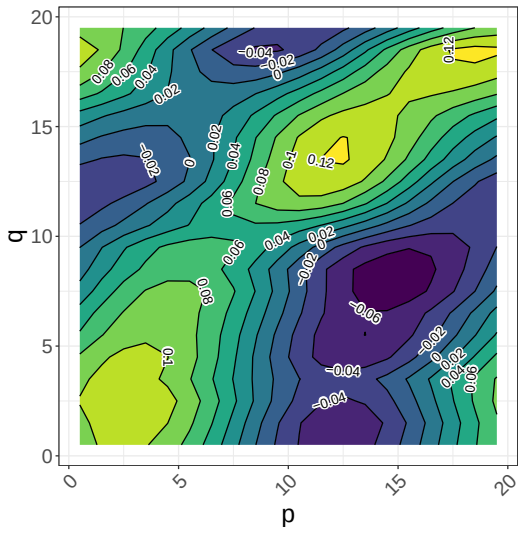


Figure 15: Estimated $\hat{\beta}$ using the pFFPLS with three components. The horizontal p-axis represents the argument of the knee angle profiles (X). The vertical q-axis represents the argument of the hip angle profiles (Y).

6 Conclusion and discussion

This paper introduced pFFPLS, a penalized partial least squares approach to address the function-on-function regression problem. pFFPLS is an iterative algorithm that primarily

operates on the coefficients obtained from a basis representation of the data and the PLS weights. The main contribution of our approach to the current state-of-the-art FFPLS methods is the incorporation of smoothness by penalizing the norm of the PLS weights, even when using a large number of basis functions. We showed that the selection of the number of basis functions is not a relevant issue in pFFPLS since the degree of smoothness in the estimated coefficient function is controlled by means of the penalty term introduced in the estimation of the PLS weights.

Our simulation study demonstrated that pFFPLS outperformed the non-penalized FFPLS method. pFFPLS consistently provided more accurate estimations of the functional response variable Y and the coefficient function β. More precisely, we observed that FFPLS over smoothed the estimated $\hat{\beta}$ when the number of bases was small and that it provided rough estimates when this number was larger. Given the importance of smoothness in interpreting the resulting model, we considered pFFPLS the superior alternative.

Environmental data has been a common application domain in the function-on-function regression literature. Our first case study illustrated our proposed method by predicting log precipitation based on temperature profiles throughout the year from 35 Canadian weather stations, as in [Ramsay and Silverman, 2005]. In terms of predictive accuracy, both FFPLS and pFFPLS proved equivalent because the cross-validated procedure selects very small values of the penalties. However, with close-to-zero or zero penalties, interpretability was compromised. To overcome this, we fixed a large value for the penalties that slightly increased the CVE but introduced a remarkable improvement in the smoothness and interpretability of the coefficient function.

In a second case study, we illustrated the advantages of pFFPLS compared to FFPLS in predicting hip angle profiles from knee angle profiles using data on the gait cycle of 39 children. Our penalized approach proved superior in predictive accuracy when the penalties were selected by cross-validation. Also, the resulting coefficient function for pFFPLS was smoother, facilitating its interpretation.

It is important to note that both FFPLS and pFFPLS, as presented in this paper, are asymmetric PLS algorithms. This asymmetry arises from the fact that only the component $t = \int_{\mathcal{D}_X} X(p)\omega(p)dp$ is used to deflate both X and Y. As a result, our emphasis in pFFPLS is primarily on penalizing the weights ω associated with X rather than the weights ν associated with Y. This approach aligns with the standard procedure followed in many scalar-on-function PLS studies [Aguilera et al., 2010, Aguilera et al., 2016], where the aim is to avoid restricting the deflation to the rank of the scalar response. However, in the function-on-function regression setting, the concern of rank limitation is less relevant, as both $X(p)$ and $Y(q)$ are typically observed at a comparable number of nodes in $\mathcal{D}_X$ and $\mathcal{D}_Y$, respectively. It is worth mentioning that developing a symmetric version of pFFPLS is an open area of research that we intend to explore in the near future, as it might enhance the interpretability of the resulting coefficient function.

In the simulation and case studies, FFPLS and pFFPLS were implemented using a least-squares basis representation of the data without incorporating any penalties for estimating the representation coefficients. This approach was chosen to avoid adding

unnecessary complexity and computational burden to the algorithms. However, it is indeed an interesting avenue for future research to explore the impact of incorporating prior smoothing through penalties in estimating the curves in FFPLS compared to our smoothing inside the pFFPLS algorithm.

Extending pFFPLS to handle multiple predictors is another intriguing possibility, as demonstrated in [Beyaztas and Shang, 2020, Beyaztas and Shang, 2021]. Since pFFPLS and the multivariate versions proposed by Beyaztas and Shang are all based on the FFPLS framework originally introduced in [Preda and Schiltz, 2011], the extension should be relatively straightforward. This extension would allow for modeling complex relationships involving multiple predictors and interaction terms. It would be a valuable direction for future research to explore the performance and interpretability of pFFPLS in a multiple-predictors setting.

Acknowledgements

This research was partly funded by Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación, Spain, grant numbers: PID2019-104901RB-I00, PID2020-113961GB-I00, and PDC2022-133359-I00.

References

[Aguilera et al., 2016] Aguilera, A. M., Aguilera-Morillo, M. C., and Preda, C. (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems*, 154:80–92.

[Aguilera et al., 2010] Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression. *Chemometrics and Intelligent Laboratory Systems*, 104(2):289–305.

[Beyaztas and Shang, 2020] Beyaztas, U. and Shang, H. L. (2020). On function-on-function regression: partial least squares approach. *Environmental and Ecological Statistics*, 27(1):95–114.

[Beyaztas and Shang, 2021] Beyaztas, U. and Shang, H. L. (2021). A partial least squares approach for function-on-function interaction regression. *Computational Statistics*, 36(2):911–939.

[Delaigle and Hall, 2012] Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1):322–352.

[Eilers et al., 2015] Eilers, P. H., D., M. B., and Maria, D. (2015). Twenty Years of P-Splines. *SORT-Statistics and Operations Research Transactions*, 39(2):149—-186.

[Eilers and Marx, 1996] Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

[Ferraty and Vieu, 2006] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Series in Statistics. Springer New York.

[Horváth and Kokoszka, 2012] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Science+Business Media, Inc., New York, NY.

[Ivanescu et al., 2015] Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015). Penalized function-on-function regression. *Computational Statistics*, 30(2):539–568.

[Kokoszka and Reimherr, 2017] Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press Taylor & Francis Group.

[Lê Cao et al., 2008] Lê Cao, K. A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology*, 7(1).

[Olshen et al., 1989] Olshen, R. A., Biden, E. N., Wyatt, M. P., and Sutherland, D. H. (1989). Gait Analysis and the Bootstrap. *The Annals of Statistics*, 17(4):1306–1310.

[O'Sullivan, 1986] O'Sullivan, F. (1986). A stastical perspective on ill-posed inverse problems. *Statistical Science*, 1:505–527.

[Preda and Saporta, 2005] Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 48(1):149–158.

[Preda and Schiltz, 2011] Preda, C. and Schiltz, J. (2011). Functional PLS regression with functional response: the basis expansion approach. In Fagerberg, J., Mowery, D. C., and Nelson, R. R., editors, *Proceedings ASMDA 2011: 14th Applied Stochastic Models and Data Analysis Conference, June 2011*, pages 1126–1133. Università di Roma La Spienza, Roma.

[R Core Team, 2023] R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Ramsay, 2023] Ramsay, J. (2023). *fda: Functional Data Analysis*. R package version 6.1.4.

[Ramsay et al., 2009] Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer New York, New York, NY.

[Ramsay and Dalzell, 1991] Ramsay, J. O. and Dalzell, C. J. (1991). Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572.

[Ramsay and Silverman, 2005] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer New York, New York, NY, second edi edition.

[Wang et al., 2016] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295.

[Wold et al., 1983] Wold, S., Martens, H., and Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In Ruhe, A. and Ka strHm, B., editors, *Proceedings of the Conference Matrix Pencils, March 1982, Lecture Notes in Mathematics*, pages 286–293. Springer, Heidelberg.

[Yao et al., 2005] Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

[Zhou, 2021] Zhou, Z. (2021). Fast implementation of partial least squares for function-on-function regression. *Journal of Multivariate Analysis*, 185:104769.