

BEST ESTIMATION OF FUNCTIONAL LINEAR MODELS

GIACOMO ALETTI, CATERINA MAY, AND CHIARA TOMMASI

ABSTRACT. Observations which are realizations from some continuous process are frequent in sciences, engineering, economics, and other fields. We consider linear models, with possible random effects, where the responses are random functions in a suitable Sobolev space. The processes can not be observed directly. With smoothing procedures from the original data, both the response curves and their derivatives can be reconstructed, even separately. From both these samples of functions, just one sample of representatives is obtained to estimate the vector of functional parameters. We hence get a strong functional version of the Gauss-Markov theorem.

Keywords: functional data analysis; Sobolev spaces; linear models; repeated measurements; Gauss-Markov theorem; Riesz representation theorem; best linear unbiased estimator; experimental designs; optimal design.

1. INTRODUCTION

Observations which are realizations from some continuous process are ubiquitous in many fields like sciences, engineering, economics and other fields. For this reason, the interest for statistical modeling of functional data is increasing, with applications in many areas. Reference monographs on functional data analysis are, for instance, the books of [19] and [10], and the book of [8] for the non-parametric approach. They cover topics like data representation, smoothing and registration; regression models; classification, discrimination and principal component analysis; derivatives and principal differential analysis; and more other.

Regression models with functional variables can cover different situations: it may be the case of functional responses, or functional predictors, or both. In the present paper linear models with functional response and multivariate (or univariate) regressor are considered. We consider the case of repeated measurements, which may be particularly useful in the context of experimental designs, but all the theoretical results proved remain valid in the standard case. Focus of the work is the best estimation of the functional coefficients of the regressors.

The use of derivatives is very important for exploratory analysis of functional data as well as for inference and prediction methodologies. High quality derivative information can be provided, for instance, by reconstructing the functions with spline smoothing procedures. Recent developments on estimation of derivatives are contained in the works of [20] and in [17]. See also [3], who have obtained derivatives in the context of survival analysis, and [9] who have estimated derivatives in a non-parametric model.

In the literature the usual space for functional data is L^2 , and the data are used to reconstruct curve functions or derivatives. The novelty of the present work is that the curves are random elements of a suitable Sobolev space. The heuristic

justification for this choice is that the data may provide information on both curve functions and their derivatives. Of course, if we take into consideration the whole information, about curves and derivatives, we may improve our estimates. Curves and derivatives are actually reconstructed from a set of observed values, because the response processes cannot be observed directly. Two situations may occur: the sample of functions are reconstructed by a smoothing procedure of the data, and derivatives are then obtained by differentiation. At our knowledge, this is the most common method adopted in functional data analysis.

However, the sample of functions and the sample of derivatives may be obtained separately. For instance, different smoothing techniques may be used to obtain the functions and the derivatives. Another possibility is when two sets of data are available, which are suitable to estimate functions and derivatives, respectively. For example, in the case of a motion process, data concerning positions and data about velocities may be observed.

In this paper we propose a new method which incorporates the information provided by both the sample of functions and the sample of derivatives. We show that the full information of the sample of functions and the sample of derivatives is not a weighted mean between, e.g., the functions and the integrated derivatives. From the two samples of reconstructed functions and derivatives just one sample of representatives is obtained. In addition, even if the sample of functions and the sample of derivatives belong to L^2 , the information carried by them on the regression parameters is a sample of function in H^1 , which implies that this new sample cannot be obtained from a linear combination involving the original sample of functions. We use this sample to estimate the functional parameters. Once this method is found, the optimal results may appear as a straightforward extension of the well-known classical case, although the proof requires much technical effort.

A new version of the Gauss-Markov theorem is hence proved in the proper infinite-dimensional space (H^1), showing that our new sample carries all the relevant information on the parameters. More precisely, we prove that any linear operator on H^1 applied to the OLS estimator on the reconstructed sample has minimum variance. In this sense, the OLS estimator is H^1 -strong BLUE. As a particular case, the OLS estimator is also the best linear unbiased estimator in H^1 . In practice this H^1 -BLUE cannot be computed explicitly, since the representative is only implicitly defined in infinite dimensions. Nevertheless, practitioners are used to represent functions in a basis expansion which is truncated at some order. This means to work in a finite-dimensional subspace \mathcal{S} of H^1 . We prove that the OLS estimator in \mathcal{S} is the projection of the H^1 -BLUE and it is itself the H^1 -BLUE in \mathcal{S} . Moreover, the OLS estimator in \mathcal{S} can be practically computed, since the projection of the representative depends explicitly on the sample of functions and derivatives.

As a consequence of the results proved, a rigorous generalization of the theory of optimal design of experiments in infinite-dimensional spaces is presented (see also [1]). An application to an ergonomic data-set shows the advantages of this theory in the real world.

The paper is organized as follows. Section 2 contains the model description and the new theoretical results offered in this work. Section 3 presents some considerations which are fundamental from a practical point of view. Section 4 is focused on the experimental context, and optimal designs for the model estimation are derived.

Section 5 is a summary together with some final remarks. Some additional results and proofs of theorems are deferred to Appendix A.

2. STRONG H^1 -BLUE IN FUNCTIONAL LINEAR MODELS

Let us consider a regression model where the response y is a random function which depends linearly on a vectorial (or scalar) known variable \mathbf{x} through a functional coefficient, which needs to be estimated. In particular, we assume that \mathbf{x} is an experimental condition and that n experiments are performed in batches at n (not necessarily different) experimental conditions $\mathbf{x}_1, \dots, \mathbf{x}_n$. The i -th experiment is formed by $r \geq 1$ trials (repetitions) at the same \mathbf{x}_i . Therefore, the following random effect model is considered:

$$(1) \quad y_{ij}(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) + \alpha_i(t) + \varepsilon_{ij}(t) \quad i = 1, \dots, n; \quad j = 1, \dots, r,$$

where $y_{ij}(t)$ denotes the response curve of the j -th observation at the i -th experiment; $\mathbf{f}(\mathbf{x}_i)$ is a p -dimensional vector of known functions; $\boldsymbol{\beta}(t)$ is an unknown p -dimensional functional vector; $\alpha_i(t)$ is a zero-mean process which denotes the random effect due to the i -th experiment and takes into account the correlation among the r repetitions; $\varepsilon_{ij}(t)$ is a zero-mean error process.

An example for the model (1) is provided in Section 4.1; if $r = 1$ this model reduces to the functional response model described, for instance, in [10].

In a real world setting, the functions $y_{ij}(t)$ are not directly observed. By a smoothing procedure from the original data, the investigator can reconstruct both the functions and their derivatives, obtaining $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$, respectively. Hence we can assume that the model for the reconstructed functional data is

$$(2) \quad \begin{cases} y_{ij}^{(f)}(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) + \alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t) \\ y_{ij}^{(d)}(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t) + \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t) \end{cases} \quad i = 1, \dots, n; \quad j = 1, \dots, r,$$

where

- (i) the n bivariate vectors $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$ are zero-mean independent and identically distributed couples of processes such that $E(\|\alpha_i^{(f)}(t)\|_{L^2}^2 + \|\alpha_i^{(d)}(t)\|_{L^2}^2) < \infty$;
- (ii) all the $n \times r$ couples $(\varepsilon_{ij}^{(f)}(t), \varepsilon_{ij}^{(d)}(t))$ are zero-mean identically distributed processes, each process being independent of all the other processes, with $E(\|\varepsilon_{ij}^{(f)}(t)\|_{L^2}^2 + \|\varepsilon_{ij}^{(d)}(t)\|_{L^2}^2) < \infty$.

Not that the investigator might reconstruct each function $y_{ij}^{(f)}(t)$ and its derivative $y_{ij}^{(d)}(t)$ separately. In this case, the right-hand term of the second equation in (2) is not the derivative of the right-hand term of the first equation. The particular case when $y_{ij}^{(d)}(t)$ is obtained by differentiation $y_{ij}^{(f)}(t)$ is the most simple situation in model (2).

Let us consider an estimator $\hat{\boldsymbol{\beta}}(t)$ of $\boldsymbol{\beta}(t)$, formed by p random functions in the Sobolev space $H^1 = H^1(\tau)$. Recall that a function $g(t)$ is in H^1 if $g(t)$ and its derivative $g'(t)$ belong to L^2 . Moreover, H^1 is an Hilbert space with inner product

$$(3) \quad \begin{aligned} \langle g_1(t), g_2(t) \rangle_{H^1} &= \langle g_1(t), g_2(t) \rangle_{L^2} + \langle g_1'(t), g_2'(t) \rangle_{L^2} \\ &= \int g_1(t)g_2(t)dt + \int g_1'(t)g_2'(t)dt, \quad g_1(t), g_2(t) \in H^1. \end{aligned}$$

Definition 2.1. We define the H^1 -generalized covariance matrix $\Sigma_{\hat{\beta}}$ of an unbiased estimator $\hat{\beta}(t)$ as the $p \times p$ matrix whose (l_1, l_2) -th element is

$$(4) \quad E\langle \hat{\beta}_{l_1}(t) - \beta_{l_1}(t), \hat{\beta}_{l_2}(t) - \beta_{l_2}(t) \rangle_{H^1}.$$

This global notion of covariance has been used also in [15, Definition 2], in the context of predicting georeferenced functional data. These authors have found a BLUE estimator for the drift of their underlying process, which can be seen as an example of the results given in this paper.

Definition 2.2. In analogy with classical settings, we define the H^1 -functional best linear unbiased estimator (H^1 -BLUE) as the estimator with minimal (in the sense of Loewner Partial Order) H^1 -generalized covariance matrix (4), in the class of the linear unbiased estimators of $\beta(t)$.

It is simple to show that an equivalent definition of H^1 -BLUE. In fact a H^1 -BLUE minimizes the quantity

$$E\left(\left\langle \sum_{i=1}^p \alpha_i (\hat{\beta}_i(t) - \beta_i(t)), \sum_{i=1}^p \alpha_i (\hat{\beta}_i(t) - \beta_i(t)) \right\rangle_{H^1}\right)$$

for any choice of $(\alpha_1, \dots, \alpha_p)$, in the class of the linear unbiased estimators $\hat{\beta}(t)$ of $\beta(t)$. In other words, the H^1 -BLUE minimizes the H^1 -variance of any linear combination of its components. A stronger request is the following.

Definition 2.3. We define the H^1 -strong functional best linear unbiased estimator (H^1 -SBLUE) as the estimator with minimal variance,

$$E\left(\left\langle O(\hat{\beta}_i(t) - \beta_i(t)), O(\hat{\beta}_i(t) - \beta_i(t)) \right\rangle_{H^1}\right)$$

for any choice of linear operator $O : (H^1)^p \rightarrow H^1$, in the class of the linear unbiased estimators $\hat{\beta}(t)$ of $\beta(t)$.

Given a couple $(y^{(f)}(t), y^{(d)}(t)) \in L^2 \times L^2$, it may be defined a linear continuous operator on H^1 as follows

$$\phi(h) = \langle y^{(f)}, h \rangle_{L^2} + \langle y^{(d)}, h' \rangle_{L^2}, \quad \forall h \in H^1.$$

From the Riesz representation theorem, there exists a unique $\tilde{y} \in H^1$ such that

$$(5) \quad \langle \tilde{y}, h \rangle_{H^1} = \langle y^{(f)}, h \rangle_{L^2} + \langle y^{(d)}, h' \rangle_{L^2}, \quad \forall h \in H^1.$$

Definition 2.4. The unique element $\tilde{y} \in H^1$ defined in (5) is called the *Riesz representative* of the couple $(y^{(f)}(t), y^{(d)}(t)) \in L^2 \times L^2$.

This definition will be useful to provide a nice expression for the functional OLS estimator $\hat{\beta}(t)$. Actually the Riesz representative synthesizes, in some sense, in H^1 the information of both $y^{(f)}(t)$ and $y^{(d)}(t)$.

The *functional OLS estimator* for the model (2) is

$$\begin{aligned} \hat{\beta}(t) &= \arg \min_{\beta(t)} \left(\sum_{j=1}^r \sum_{i=1}^n \|y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \beta(t)\|_{L^2}^2 + \sum_{j=1}^r \sum_{i=1}^n \|y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \beta'(t)\|_{L^2}^2 \right) \\ &= \arg \min_{\beta(t)} \sum_{j=1}^r \sum_{i=1}^n \left(\|y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \beta(t)\|_{L^2}^2 + \|y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \beta'(t)\|_{L^2}^2 \right) \end{aligned}$$

The quantity

$$\|y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{L^2}^2 + \|y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t)\|_{L^2}^2$$

resembles

$$\|y_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{H^1}^2,$$

because $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$ reconstruct $y_{ij}(t)$ and its derivative function, respectively. The functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ minimizes, in this sense, the sum of the H^1 -norm of the unobservable residuals $y_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)$.

Theorem 2.5. *Given a model as in (2),*

a) *the functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ can be computed by*

$$(6) \quad \hat{\boldsymbol{\beta}}(t) = (F^T F)^{-1} F^T \bar{\mathbf{y}}(t),$$

where $\bar{\mathbf{y}}(t) = (\bar{y}_1(t), \dots, \bar{y}_n(t))$ is a vector, whose component i -th is the mean of the Riesz representatives of the replications:

$$\bar{y}_i(t) = \frac{\sum_{j=1}^r \tilde{y}_{ij}(t)}{r},$$

and $F = [\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n)]^T$ is the $n \times p$ design matrix.

b) *The estimator $\hat{\boldsymbol{\beta}}(t)$ is unbiased and its generalized covariance matrix is proportional to $(F^T F)^{-1}$.*

Proof of Theorem 2.5. Part a). We consider the sum of square residuals:

$$\begin{aligned} S(\boldsymbol{\beta}(t)) &= \sum_{j=1}^r \sum_{i=1}^n \left(\|y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{L^2}^2 + \|y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t)\|_{L^2}^2 \right) \\ &= \sum_{j=1}^r \sum_{i=1}^n \left(\langle y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t), y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) \rangle_{L^2} \right. \\ &\quad \left. + \langle y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t), y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t) \rangle_{L^2} \right) \end{aligned}$$

The Gâteaux derivative of $S(\cdot)$ at $\boldsymbol{\beta}(t)$ in the direction of $\mathbf{g}(t) \in (H^1)^p$ is

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{S(\boldsymbol{\beta}(t) + h\mathbf{g}(t)) - S(\boldsymbol{\beta}(t))}{h} &= 2 \left(\sum_{j=1}^r \sum_{i=1}^n \left(\langle y_{ij}^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t), \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}(t) \rangle_{L^2} \right. \right. \\ &\quad \left. \left. + \langle y_{ij}^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t), \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}'(t) \rangle_{L^2} \right) \right) \\ (7) \quad &= 2r \left(\langle F^T \bar{\mathbf{y}}^{(f)}(t) - F^T F \boldsymbol{\beta}(t), \mathbf{g}(t) \rangle_{(L^2)^p} \right. \\ &\quad \left. + \langle F^T \bar{\mathbf{y}}^{(d)}(t) - F^T F \boldsymbol{\beta}'(t), \mathbf{g}'(t) \rangle_{(L^2)^p} \right), \end{aligned}$$

where $\bar{\mathbf{y}}^{(f)}(t)$ and $\bar{\mathbf{y}}^{(d)}(t)$ are two $n \times 1$ vectors whose i -th elements are

$$\bar{\mathbf{y}}_i^{(f)}(t) = \frac{\sum_{j=1}^r y_{ij}^{(f)}(t)}{r}, \quad \bar{\mathbf{y}}_i^{(d)}(t) = \frac{\sum_{j=1}^r y_{ij}^{(d)}(t)}{r}.$$

Developing the right-hand side of (7), we have that the Gâteaux derivative is

$$\begin{aligned}
&= 2r \left(\left(\langle F^T \bar{\mathbf{y}}^{(f)}(t), \mathbf{g}(t) \rangle_{(L^2)^p} + \langle F^T \bar{\mathbf{y}}^{(d)}(t), \mathbf{g}'(t) \rangle_{(L^2)^p} \right) \right. \\
&\quad \left. - \left(\langle F^T F \boldsymbol{\beta}(t), \mathbf{g}(t) \rangle_{(L^2)^p} + \langle F^T F \boldsymbol{\beta}'(t), \mathbf{g}'(t) \rangle_{(L^2)^p} \right) \right) \\
(8) \quad &= 2r \left(\langle F^T \bar{\mathbf{y}}(t), \mathbf{g}(t) \rangle_{(H^1)^p} - \langle F^T F \boldsymbol{\beta}(t), \mathbf{g}(t) \rangle_{(H^1)^p} \right),
\end{aligned}$$

where $\bar{\mathbf{y}}(t)$ is a $n \times 1$ vector whose i -th element is the Riesz representative of $(\bar{\mathbf{y}}_i^{(f)}(t), \bar{\mathbf{y}}_i^{(d)}(t))$.

The Gâteaux derivative (8) is equal to 0 for any $\mathbf{g}(t) \in (H^1)^p$ if and only if $\hat{\boldsymbol{\beta}}(t)$ is given by the following equation:

$$F^T F \hat{\boldsymbol{\beta}}(t) = F^T \bar{\mathbf{y}}(t),$$

which proves the first statement of the theorem.

Part b) Definition (2.4) and model (2) imply that, for any $h(t) \in H^1$,

$$\begin{aligned}
\langle E(\tilde{y}_{ij}(t)), h(t) \rangle_{H^1} &= E(\langle y_{ij}^{(f)}(t), h(t) \rangle_{L^2}) + E(\langle y_{ij}^{(d)}(t), h'(t) \rangle_{L^2}) \\
&= \langle \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t), h(t) \rangle_{H^1},
\end{aligned}$$

then $E(\bar{\mathbf{y}}(t)) = F \boldsymbol{\beta}(t)$, and hence $\hat{\boldsymbol{\beta}}(t)$ is unbiased. Moreover,

$$(9) \quad \bar{y}_i(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) = \tilde{\alpha}_i(t) + \frac{\sum_{j=1}^r \tilde{\varepsilon}_{ij}(t)}{r}, \quad i = 1, \dots, n$$

where $\tilde{\alpha}_i(t)$ and $\tilde{\varepsilon}_{ij}(t)$ denote the Riesz representatives of $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$ and $(\varepsilon_{ij}^{(f)}(t), \varepsilon_{ij}^{(d)}(t))$, respectively. From the hypothesis (i) and (ii) in the model (2), the left-hand side quantities in (9) are zero-mean i.i.d. processes, for $i = 1, \dots, n$. Therefore, the generalized covariance matrix of $\bar{\mathbf{y}}(t)$ is $\sigma^2 I_n$, where $\sigma^2 = E(\|\bar{y}_i(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{H^1}^2)$. Hence, the generalized covariance matrix of $\hat{\boldsymbol{\beta}}(t)$ is $\Sigma_{\hat{\boldsymbol{\beta}}} = \sigma^2 (F^T F)^{-1}$. \square

The functional OLS estimator obtained in Theorem 2.5 by means of the Riesz representatives is also the best linear unbiased estimator in the Sobolev space, as stated in the next theorem. The proof is deferred to Appendix A.

Theorem 2.6. *The functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ for the model (2) is a H^1 -functional **SBLUE**.*

Remark 2.7. The theory and the results presented in this work may be generalized to other Sobolev spaces. The extension to H^m , $m \geq 2$, is straightforward. Moreover, as in Bayesian context, the investigator might have a different *a priori* consideration of $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$. Thus, different weights $p^{(f)}$ and $p^{(d)}$ may be used for curve functions and derivatives, respectively. Another interesting generalization might be to add positive weight functions $w^{(f)}(t)$ and $w^{(d)}(t)$ on τ , when, for instance, distinct zones of τ are considered to have different relevance. Therefore, the inner product given in (3) may be extended to

$$\langle g_1(t), g_2(t) \rangle_H = p^{(f)} \int_{\tau} g_1(t) g_2(t) w^{(f)}(t) dt + p^{(d)} \int_{\tau} g_1(t) g_2(t) w^{(d)}(t) dt.$$

When $p^{(f)} = p^{(d)} = 1$ and $w^{(f)}(t) = w^{(d)}(t) = w(t)$, the Hilbert space is called weighted Sobolev space, see [12]. These generalizations, which have different impacts on applications, will be object of future works.

3. PRACTICAL CONSIDERATIONS

In a real world context, we work with a finite dimensional subspace \mathcal{S} of H^1 . Let $S = \{w_1(t), \dots, w_N(t)\}$ be a base of \mathcal{S} . Without loss of generality, we may assume that $\langle w_h(t), w_k(t) \rangle_{H^1} = \delta_h^k$, where

$$\delta_h^k = \begin{cases} 1 & \text{if } h = k; \\ 0 & \text{if } h \neq k; \end{cases}$$

is the Kronecker delta symbol, since a Gram-Schmidt orthonormalization procedure may be always applied. More precisely, given any base $\tilde{S} = \{\tilde{w}_1(t), \dots, \tilde{w}_N(t)\}$ in H_1 , the corresponding orthonormal base is given by:

for $k = 1$, define $w_1(t) = \frac{\tilde{w}_1(t)}{\|\tilde{w}_1(t)\|_{H^1}}$,

for $k \geq 2$, let $\hat{w}_k(t) = \tilde{w}_k(t) - \sum_{h=1}^{k-1} \langle \tilde{w}_k(t), w_h(t) \rangle_{H^1} w_h(t)$, and $w_k(t) = \frac{\hat{w}_k(t)}{\|\hat{w}_k(t)\|_{H^1}}$.

With this orthonormalized base, the projection $\tilde{y}(t)_{\mathcal{S}}$ on \mathcal{S} of the Riesz representative $\tilde{y}(t)$ of the couple $(y^{(f)}(t), y^{(d)}(t))$ is given by

$$\begin{aligned} \tilde{y}(t)_{\mathcal{S}} &= \sum_{k=1}^N \langle \tilde{y}(t), w_k(t) \rangle_{H^1} \cdot w_k(t) \\ (10) \quad &= \sum_{k=1}^N \left(\langle y^{(f)}(t), w_k(t) \rangle_{L^2} + \langle y^{(d)}(t), w'_k(t) \rangle_{L^2} \right) w_k(t), \end{aligned}$$

where the last equality comes from the definition (5) of the Riesz representative. Now, if $\mathbf{m}_l = (m_{l,1}, \dots, m_{l,n})^T$ is the l -th row of $(F^T F)^{-1} F^T$, then

$$\begin{aligned} \langle \hat{\beta}_l(t), w_k(t) \rangle_{H^1} &= \sum_{i=1}^n \langle m_{l,i} \bar{y}_i(t), w_k(t) \rangle_{H^1} \\ &= \sum_{i=1}^n m_{l,i} \langle \bar{y}_i(t), w_k(t) \rangle_{H^1}, \quad \text{for any } k = 1, \dots, N, \end{aligned}$$

$$\hat{\beta}_l(t)_{\mathcal{S}} = \mathbf{m}_l^T \bar{\mathbf{y}}(t)_{\mathcal{S}},$$

hence $\hat{\beta}(t)_{\mathcal{S}} = (F^T F)^{-1} F^T \bar{\mathbf{y}}(t)_{\mathcal{S}}$.

Let us note that, even if the Riesz representative (5) is implicitly defined, its projection on \mathcal{S} can be easily computed by (10). From a practical point of view, the statistician can work with the data $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$ projected on a finite linear subspace \mathcal{S} and the corresponding OLS estimator $\hat{\beta}(t)_{\mathcal{S}}$ is the projection on \mathcal{S} of the H^1 -SBLUE estimator $\hat{\beta}(t)$ obtained in the Section 2. As a consequence of the Theorem 2.6, $\hat{\beta}(t)_{\mathcal{S}}$ is H^1 -SBLUE in \mathcal{S} , since it is unbiased and the projection is linear. For the projection, it is crucial to take a base of \mathcal{S} which is orthonormal in H^1 .

It is straightforward to prove that the estimator (6) becomes

$$\hat{\beta}(t) = (F^T F)^{-1} F^T \mathbf{y}^{(f)}(t),$$

in two cases: when we do not take into consideration $y^{(d)}$, or when $y^{(d)} = (y^{(f)})'$. In both the cases, from the results obtained in the Section 2, $\hat{\beta}$ is an L^2 -BLUE. Up to our knowledge, this is the most common situation considered in the literature.

4. OPTIMAL DESIGNS OF EXPERIMENTS IN FUNCTIONAL MODELS

In this section, we assume to work in an experimental setup. Therefore, \mathbf{x}_i , with $i = 1, \dots, n$, are not observed auxiliary variables; they can be freely chosen by an experimenter on the design space \mathcal{X} . The set of experimental conditions $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is called an exact design. A more general definition is that of continuous design, as a probability measure ξ with support on \mathcal{X} (see, for instance, [11]). The choice of ξ may be done with the aim of obtaining accurate estimates of the model parameters.

From Theorem 2.6, $\hat{\beta}(t)$ given in (6) is the H^1 -BLUE for the model (2). This optimal estimator can be further improved by a “clever” choice of the design. In analogy with the criteria proposed in the finite-dimensional theory (see for instance, [6, 18, 23]) we define a functional optimal design as a design which minimizes an appropriate convex function of the generalized covariance matrix $\Sigma_{\hat{\beta}}$ (see Definition 2.1). For instance, a *functional D-optimum* design is a design ξ_D^* which minimizes $\det(\Sigma_{\hat{\beta}})$. Part b) of Theorem 2.5 proves that $\Sigma_{\hat{\beta}} \propto (F^T F)^{-1}$. From the definition of continuous design

$$F^T F \propto \int_{\mathcal{X}} \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T d\xi(\mathbf{x}),$$

and hence ξ_D^* maximizes $\Phi_D(\xi) = \det(\int_{\mathcal{X}} \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T d\xi(\mathbf{x}))$.

4.1. An example: the ergonomic data. Herein, we study the performance of the design proposed in [21]. In detail, to forecast the motion of drivers, some data are to be collected on the motion of a single subject to different locations within a test car. An experimental design is given by the choice of these locations (the different experimental conditions) and by the number of times that the experiment has to be replicated at each location. The response curve $y(t)$ is the angle formed at the right elbow between the upper and the lower arm, which is measured by a motion capture equipment. In the design used by [21], 3 motion curves were observed at 20 different locations, spread around the glove compartment, gear shift, the central instrument panel and an overhead panel. These data are available from Faraway’s website. In [21], 3 different models are compared to predict the motion given the coordinates $\mathbf{x} = (x, y, z)$. From this comparison, the following quadratic model seems to be adequate:

$$y_{ij}(t) = \mathbf{f}(\mathbf{x}_i)^T \beta(t) + \varepsilon_{ij}(t), \quad i = 1, \dots, 20, \quad j = 1, 2, 3,$$

where

$$\mathbf{f}(\mathbf{x}_i)^T = (1, x_i, y_i, z_i, x_i y_i, x_i z_i, y_i z_i, x_i^2, y_i^2, z_i^2).$$

Let us denote by ξ_O the exact design used in [21] to collect observations $y_{ij}(t)$. This design has 20 different locations at which three trials are repeated, therefore the total number of observations is 60.

We standardize support points (coordinates) of ξ_O so that they belong to $\mathcal{X} = [-1; 1]^3$. For this experimental domain, [2, Section 11.5] provides both the continuous D-optimum design, say ξ_D^* , and exact D-optimum designs. Since ξ_D^* has

support on the points of the 3^m factorial (with $m = 3$), [2] proposes to search an exact D-optimum design (for a small n) over the points of the 3^m factorial. To be consistent with the design used in [21], we consider the exact design with $n = 20$ locations, some of which may be repeated. At each location, an experiment is performed in batches of three trials. When a location is replicated more times, several experiments (each one formed by three trials) is performed at this location. Thus, again the total number of observation is 60. Let us call this design ξ_{ED}^1 .

In addition, we search an exact D-optimum design over the whole experimental domain $\mathcal{X} = [-1; 1]^3$. This exact design has 9 different locations which are not points of the 3^m factorial, 5 different points of the 3^m factorial which are not replicated and 3 different points of the 3^m factorial which are replicated twice. Three observations are taken at each location. Let this exact design be denoted by ξ_{ED}^2 .

Finally, in a grid of 51 equally spaced points in $[-1, 1]$, an exact D-optimum design over the 51^m factorial has been find with Matlab 2014b routine `rowexch`, finding ξ_{ED}^3 .

In practice all designs are exact. Therefore, the continuous D-optimum design ξ_D^* is used only as a benchmark, in order to measure the goodness of a design ξ with respect to it. As a measure of goodness the D-efficiency is used:

$$\text{Eff}_D(\xi) = \left(\frac{|\Phi_D(\xi)|}{|\Phi_D(\xi_D^*)|} \right)^{1/p},$$

where $p = 10$ is the dimension of the functional vector $\beta(t)$. The D-efficiency $\text{Eff}_D(\xi)$ is proportional to the design size, thus a design ξ with D-efficiency equal to 0,5 needs the double of observations of ξ_D^* to get the same precision in the estimates.

The following table lists the D-efficiencies of ξ_{ED}^1 , ξ_{ED}^2 , ξ_{ED}^3 and ξ_O .

Design	D-efficiency
ξ_O	0.3396
ξ_{ED}^1	0.9779
ξ_{ED}^2	0.9789
ξ_{ED}^3	0.9788

TABLE 1. D-efficiencies of the designs ξ_O , ξ_{ED}^1 , ξ_{ED}^2 and ξ_{ED}^3 .

From Table 4.1 we have that the exact D-optimum designs ξ_{ED}^2 and ξ_{ED}^3 are almost equivalent. The D-efficiency of ξ_O is instead very low, only equal to 0.3396. This example clearly shows the importance of choosing the experimental design according to an optimality criterion.

5. SUMMARY

Functional data are suitably modeled in separable Hilbert spaces (see [10] and [4]) and L^2 is usually sufficient to handle the majority of the techniques proposed in the literature of functional data analysis.

Differently, we consider proper Sobolev spaces, since we guess that the data may provide information on both curve functions and their derivatives. The classical theory for linear regression models is extended to this context by means of the the sample of Riesz representatives. Roughly speaking, the Riesz representatives are

“quantities” which incorporates both functions and derivatives information in a non trivial way.

Using the sample of Riesz representatives, we provide a strong, generalized version of the well known Gauss-Markov theorem for the functional linear regression models considered. Despite the complexity of the problem we obtain an elegant and simple solution, through the use of the Riesz representatives which always belong to a Sobolev space.

Interesting studies on experimental design with functional observations are presented in the recent literature (see, for instance, [14], [13] and [24]). In the present work a rigorous theoretical support is provided to apply D-optimal designs to linear models with functional responses (similar models are considered, for instance, in [5], [21], [22]). As a matter of future work, we will study also different optimality criteria in the context of functional data models. In addition, we intend to develop our theory also in the functional setting considered by [7], where the design region is a subset of some functional space and the response is a scalar quantity.

APPENDIX A. PROOF OF THEOREM 2.6

The OLS $\hat{\beta}(t)$ is a linear map which associates an element $\hat{\beta}(t)$ in $(H^1)^p$ to any nr -tuple $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$ in $(L^2 \times L^2)^{nr}$. In what follows, we show that it is the “best” among all the linear unbiased operators $\mathbf{C} : \text{Dom}(\mathbf{C}) \subseteq (L^2 \times L^2)^{nr} \rightarrow (H^1)^p$.

The model (2) may be written in the following vectorial form:

$$(11) \quad \begin{cases} \mathbf{y}^{(f)}(t) = (F \otimes \mathbf{1}_r)\beta(t) + (\alpha^{(f)}(t) \otimes \mathbf{1}_r) + \varepsilon^{(f)}(t) \\ \mathbf{y}^{(d)}(t) = (F \otimes \mathbf{1}_r)\beta'(t) + (\alpha^{(d)}(t) \otimes \mathbf{1}_r) + \varepsilon^{(d)}(t) \end{cases}$$

where $\mathbf{1}_r$ is the column vector of length r with all components equal to 1. Let

$$\bar{\mathbf{y}}^{(k)}(t) = \left(y_{11}^{(k)}(t), \dots, y_{1r}^{(k)}(t), y_{21}^{(k)}(t), \dots, y_{2r}^{(k)}(t), \dots, y_{n1}^{(k)}(t), \dots, y_{nr}^{(k)}(t) \right)^T$$

be a $nr \times 1$ block vector, with $k = 1, 2$. Given any couple of $nr \times 1$ block vectors $(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t))$, we may define the following n dimensional vector

$$(12) \quad \bar{\mathbf{y}}^{(12)}(t) = \left(\bar{y}_1^{(12)}(t), \dots, \bar{y}_n^{(12)}(t) \right)^T,$$

where

$$\bar{y}_i^{(12)}(t) = \frac{\sum_{j=1}^r \tilde{y}_{ij}^{(12)}(t)}{r}$$

and $\tilde{y}_{ij}^{(12)}(t)$ is the Riesz representative of $(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t))$ as in (5).

Now we can introduce the following linear operator

$$(13) \quad \mathbf{D}(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)) = \mathbf{C}(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)) - (F^T F)^{-1} F^T \bar{\mathbf{y}}^{(12)}(t).$$

Hence,

$$(14) \quad \begin{aligned} \mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) &= \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) - (F^T F)^{-1} F^T \bar{\mathbf{y}}(t) \\ &= \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) - \hat{\beta}(t) \end{aligned}$$

and

$$\mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) = \mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) + \hat{\beta}(t).$$

The thesis follows immediately if we prove that $O(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)))$ and $O(\widehat{\beta}(t))$ are uncorrelated.

Since both \mathbf{C} and $\widehat{\beta}(t)$ are unbiased, $E(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))) = \mathbf{0}$, and thus we have to prove that

$$(15) \quad E\langle O(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))), O(\widehat{\beta}(t) - \beta(t)) \rangle_{H^1} = 0,$$

for any choice of linear operator $O : (H^1)^p \rightarrow H^1$.

The proof of equality (15) is developed in five steps.

First step. *The goal of this step is to prove that \mathbf{D} applied to the deterministic part of the model $((F \otimes \mathbf{1}_r)\beta(t), (F \otimes \mathbf{1}_r)\beta'(t))$ is identically null. As a consequence,*

$$(16) \quad \mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) = \mathbf{D}(\alpha^{(f)}(t) \otimes \mathbf{1}_r + \varepsilon^{(f)}(t), \alpha^{(d)}(t) \otimes \mathbf{1}_r + \varepsilon^{(d)}(t)).$$

Proof

From the linearity of \mathbf{C} and the zero-mean hypothesis (i) and (ii), we have that

$$\begin{aligned} E(\mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))) &= E(\mathbf{C}((F \otimes \mathbf{1}_r)\beta(t) + (\alpha^{(f)}(t) \otimes \mathbf{1}_r) + \varepsilon^{(f)}(t), \\ &\quad (F \otimes \mathbf{1}_r)\beta'(t) + (\alpha^{(d)}(t) \otimes \mathbf{1}_r) + \varepsilon^{(d)}(t))) \\ &= \mathbf{C}((F \otimes \mathbf{1}_r)\beta(t), (F \otimes \mathbf{1}_r)\beta'(t)). \end{aligned}$$

Since $E(\mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))) = \beta(t)$ we have that

$$(17) \quad \mathbf{C}((F \otimes \mathbf{1}_r)\beta(t), (F \otimes \mathbf{1}_r)\beta'(t)) = \beta(t)$$

In addition, from the definition (12) if

$$\mathbf{y}^{(1)}(t) = F\beta(t) \otimes \mathbf{1}_r \quad \text{and} \quad \mathbf{y}^{(2)}(t) = F\beta'(t) \otimes \mathbf{1}_r$$

then

$$(18) \quad \bar{\mathbf{y}}^{(12)}(t) = F\beta(t).$$

Combining (13), (17) and (18) gives

$$(19) \quad \mathbf{D}((F \otimes \mathbf{1}_r)\beta(t), (F \otimes \mathbf{1}_r)\beta'(t)) = \mathbf{0},$$

and hence (16).

Second step. *Representation of the linear operator D_l .*

For the linearity of the l -th component of \mathbf{D} with respect to the bivariate observations $(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t))$:

$$(20) \quad D_l(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)) = \sum_{i,j} D_{l,ij} (y_{ij}^{(1)}(t), y_{ij}^{(2)}(t)),$$

where, for any $i = 1, \dots, n$ and $j = 1, \dots, r$, $D_{l,ij}$ is linear. The domain of $D_{l,ij}$ is contained in $L^2(\mathbf{R}^2)$. Let $(\Psi_k = (\Psi_{k,1}, \Psi_{k,2}))_k$ a suitable base of $L^2(\mathbf{R}^2)$ that will

be specified in the fourth step, and $(\phi_h)_h$ be an orthonormal base of H^1 . With this notation

$$(21) \quad D_{l,ij}(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t)) = \sum_{k,h} [\langle \Psi_{k,1}, y_{ij}^{(1)}(t) \rangle_{L^2} + \langle \Psi_{k,2}, y_{ij}^{(2)}(t) \rangle_{L^2}] d_{l,ij}^{k,h} \phi_h(t),$$

where

$$d_{l,ij}^{k,h} = \langle D_{l,ij}(\Psi_k)(t), \phi_h(t) \rangle_{H^1}.$$

Third step. *Proof of*

$$(22) \quad \sum_{i=1}^n \sum_{j=1}^r m_{l_2,i} d_{l_1,ij}^{k,h} = 0, \quad k, h, l_1, l_2,$$

where $\mathbf{m}_{l_2} = (m_{l_2,1}, \dots, m_{l_2,n})^T$ is the l_2 -th row of $(F^T F)^{-1} F^T$.

Let $\mathbf{g}^{(l_2)}(t) \in (H^1)^p$ be the null vector except for the l_2 -th component which is $g \in H^1$, and let $\mathbf{h}(t) = (F^T F)^{-1} \mathbf{g}^{(l_2)}(t) \in (H^1)^p$. Setting $\beta(t) = \mathbf{h}(t)$ in (19),

$$\begin{aligned} 0 &= D_{l_1}((F \otimes \mathbf{1}_r) \mathbf{h}(t), (F \otimes \mathbf{1}_r) \mathbf{h}'(t)) \\ &= D_{l_1}((F \mathbf{h}(t)) \otimes \mathbf{1}_r, (F \mathbf{h}'(t)) \otimes \mathbf{1}_r) \\ &= D_{l_1}(F(F^T F)^{-1} \mathbf{g}^{(l_2)}(t) \otimes \mathbf{1}_r, F(F^T F)^{-1} \mathbf{g}^{(l_2)'}(t) \otimes \mathbf{1}_r) \\ &= D_{l_1}(g(t) \mathbf{m}_{l_2} \otimes \mathbf{1}_r, g'(t) \mathbf{m}_{l_2} \otimes \mathbf{1}_r) \\ &= \sum_{i=1}^n \sum_{j=1}^r D_{l_1,ij}(g(t) m_{l_2,i}, g'(t) m_{l_2,i}) \\ (23) \quad &= \sum_{i,j} \sum_{k,h} (\langle \Psi_{k,1}, m_{l_2,i} g \rangle_{L^2} + \langle \Psi_{k,2}, m_{l_2,i} g' \rangle_{L^2}) d_{l_1,ij}^{k,h} \phi_h(t), \end{aligned}$$

where the last equality is due to (21).

From the Riesz representation theorem we have that, for any $\Psi_k(t) = (\Psi_{k,1}(t), \Psi_{k,2}(t))$, there exists $\tilde{\psi}_k(t) \in H^1$ such that

$$(24) \quad \langle \tilde{\psi}_k, h \rangle_{H^1} = \langle \Psi_{k,1}, h \rangle_{L^2} + \langle \Psi_{k,2}, h' \rangle_{L^2}, \quad h \in H^1.$$

From (24), equality (23) becomes

$$\begin{aligned} 0 &= \sum_{i,j} \sum_{k,h} (\langle \Psi_{k,1}, m_{l_2,i} g \rangle_{L^2} + \langle \Psi_{k,2}, m_{l_2,i} g' \rangle_{L^2}) d_{l_1,ij}^{k,h} \phi_h(t) \\ &= \sum_{i,j} \sum_{k,h} m_{l_2,i} \langle \tilde{\psi}_k, g \rangle_{H^1} d_{l_1,ij}^{k,h} \phi_h(t) \\ &= \sum_{k,h} \left(\sum_{i,j} m_{l_2,i} d_{l_1,ij}^{k,h} \right) \langle g, \tilde{\psi}_k \rangle_{H^1} \phi_h(t). \end{aligned}$$

The arbitrary choice of g implies (22).

Fourth step. *Karhunen–Loève representation of the noise process and definition of the base $(\Psi_k)_k$.*

For a given (i, j) , the couple $(\alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t), \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t))$ is a process in $L^2(\mathbb{R}^2)$. Let $\mathbf{R}(s, t) = \sum_k \lambda_k \Psi_k(s) \Psi_k(t)^T$ be the spectral representation of the

covariance matrix, which implies $\lambda_k \geq 0$, and we assume, without loss of generality, that the sequence $\{\Psi_k(t), k = 1, 2, \dots\}$ forms a orthonormal base (by completing it, defining $\lambda_k = 0$ when needed). Note that $\mathbf{R}(s, t)$ does not depend on i and j , since, from the hypothesis (i) and (ii) in the model (2)

$$(\alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t), \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t)), \quad i = 1, \dots, n; \quad j = 1, \dots, r$$

are identically distributed. From Karhunen–Loève Theorem (see, e.g., [16]), there exists an array of zero-mean unit variance random variables $\{X_{ij,k}; i = 1, \dots, n; j = 1, \dots, r; k = 1, 2, \dots\}$ such that

$$(25) \quad (\alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t), \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t)) = \sum_k \sqrt{\lambda_k} X_{ij,k} \Psi_k(t).$$

From (24) the Riesz representative of the noise process (25) is $\sum_k \sqrt{\lambda_k} X_{ij,k} \tilde{\psi}_k(t)$ and we can define the following means of replications,

$$(26) \quad \bar{\varepsilon}_i = \sum_{j=1}^r \frac{\sum_k \sqrt{\lambda_k} X_{ij,k} \tilde{\psi}_k(t)}{r}, \quad i = 1, \dots, n,$$

which will be useful in the next fifth step.

Finally, from Karhunen–Loève Theorem, we may obtain the following relations which will be useful in the fifth step of the proof. From (25) we have that for any i, j and k ,

$$(27) \quad \langle \Psi_{k,1}, \alpha_i^{(f)} + \varepsilon_{ij}^{(f)} \rangle_{L^2} + \langle \Psi_{k,2}, \alpha_i^{(d)} + \varepsilon_{ij}^{(d)} \rangle_{L^2} = \sqrt{\lambda_k} X_{ij,k}.$$

In addition, the independence assumptions in the hypothesis (i) and (ii) ensure that $X_{i_1 j_1, k_1}$ and $X_{i_2 j_2, k_2}$ are independent if $i_1 \neq i_2$. For the same observation (i.e. $(i_1, j_1) = (i_2, j_2)$), the Karhunen–Loève representation gives $E(X_{i_1 j_1, k_1} X_{i_1 j_1, k_2}) = \delta_{k_1}^{k_2}$. Finally, for different replications of the same experiment (i.e. $i_1 = i_2$ but $j_1 \neq j_2$) the identically distributed bivariate process $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$ yields to a correlation which does not depend on the experiment $i_1 = i_2$ neither on the replications $j_1 \neq j_2$: $E(X_{i_1 j_1, k_1} X_{i_1 j_2, k_2}) = \rho(k_1, k_2)$. Summing up, the independence assumptions given in the hypothesis (i) and (ii) imply

$$E(X_{i_1 j_1, k_1} X_{i_2 j_2, k_2}) = \delta_{i_1}^{i_2} (\delta_{j_1}^{j_2} \delta_{k_1}^{k_2} + (1 - \delta_{j_1}^{j_2}) \rho(k_1, k_2)),$$

and hence

$$(28) \quad E(X_{i_1 j_1, k_1} \sum_{j_2} X_{i_2 j_2, k_2}) = \delta_{i_1}^{i_2} (\delta_{k_1}^{k_2} + (r - 1) \rho(k_1, k_2)).$$

Fifth step. *Proof of (15):*

$$E \langle \mathbf{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))), \mathbf{O}(\hat{\beta}(t) - \beta(t)) \rangle_{H^1} = 0,$$

for any choice of linear operator $\mathbf{O} : (H^1)^p \rightarrow H^1$.

From the definitions given in the part a of Theorem 2.5 and from Equations (5), (24) and (26) we have that, for any $h \in H^1$

$$\begin{aligned}
\langle \bar{y}_i - \mathbf{f}(\mathbf{x}_i)^T \beta, h \rangle_{H^1} &= \frac{1}{r} \sum_{j=1}^r \left(\langle y_{ij}^{(f)} - \mathbf{f}(\mathbf{x}_i)^T \beta, h \rangle_{L^2} + \langle y_{ij}^{(d)} - \mathbf{f}(\mathbf{x}_i)^T \beta', h' \rangle_{L^2} \right) \\
&= \frac{1}{r} \sum_{j=1}^r \left(\langle \alpha_i^{(f)} + \varepsilon_{ij}^{(f)}, h \rangle_{L^2} + \langle \alpha_i^{(d)} + \varepsilon_{ij}^{(d)}, h' \rangle_{L^2} \right) \\
&= \frac{1}{r} \sum_{j=1}^r \left(\left\langle \sum_k \sqrt{\lambda_k} \mathbf{X}_{ij,k} \Psi_{k,1}, h \right\rangle_{L^2} + \left\langle \sum_k \sqrt{\lambda_k} \mathbf{X}_{ij,k} \Psi_{k,2}, h' \right\rangle_{L^2} \right) \\
&= \left\langle \sum_{j=1}^r \frac{\sum_k \sqrt{\lambda_k} \mathbf{X}_{ij,k} \tilde{\psi}_k}{r}, h \right\rangle_{H^1} = \langle \bar{\varepsilon}_i, h \rangle_{H^1},
\end{aligned}$$

which proves that

$$\bar{\mathbf{y}}(t) = F\beta(t) + \bar{\varepsilon}(t),$$

where $\bar{\varepsilon}(t) = (\bar{\varepsilon}_1(t), \dots, \bar{\varepsilon}_n(t))^T$.

From this last result and from (6), $\hat{\beta}(t) - \beta(t) = (F^T F)^{-1} F^T \bar{\varepsilon}$, and hence

$$\begin{aligned}
(29) \quad E \langle O(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))), O(\hat{\beta}(t) - \beta(t)) \rangle_{H^1} \\
= E \langle O(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))), O((F^T F)^{-1} F^T \bar{\varepsilon}(t)) \rangle_{H^1} \\
= E \left\langle O(\mathbf{D}(\alpha^{(f)}(t) \otimes \mathbf{1}_r + \varepsilon^{(f)}(t), \alpha^{(d)}(t) \otimes \mathbf{1}_r + \varepsilon^{(d)}(t))), \right. \\
\left. O((F^T F)^{-1} F^T \bar{\varepsilon}(t)) \right\rangle_{H^1},
\end{aligned}$$

where the last equality is a consequence of (16).

From the linearity of the operator $O : (H^1)^p \rightarrow H^1$, we have that

$$O(b_1(t), \dots, b_p(t)) = \sum_{l=1}^p O(\underbrace{0, \dots, 0}_{l-1 \text{ times}}, b_l(t), \underbrace{0, \dots, 0}_{p-l \text{ times}}).$$

Since $b_l(t) = \sum_g \langle b_l(t), \phi_g(t) \rangle_{H^1} \phi_g(t) = \sum_g b_l^g \phi_g(t)$, where $b_l^g = \langle b_l(t), \phi_g(t) \rangle_{H^1}$, we have

$$O(b_1(t), \dots, b_p(t)) = \sum_{l,g} b_l^g O(\underbrace{0, \dots, 0}_{l-1 \text{ times}}, \phi_g(t), \underbrace{0, \dots, 0}_{p-l \text{ times}}).$$

Setting

$$O_l^{g,h} = \langle O(\underbrace{0, \dots, 0}_{l-1 \text{ times}}, \phi_g(t), \underbrace{0, \dots, 0}_{p-l \text{ times}}), \phi_h(t) \rangle_{H^1},$$

then

$$O(b_1(t), \dots, b_p(t)) = \sum_{l,g,h} b_l^g O_l^{g,h} \phi_h(t).$$

Hence, from Equations (29), (20) and (21), the thesis (15) becomes

$$\begin{aligned}
E \left\langle \sum_{l,g,h} \left(\sum_{i,j,k} (\langle \Psi_{k,1}, \alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t) \rangle_{L^2} + \langle \Psi_{k,2}, \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t) \rangle_{L^2}) d_{l_1,ij}^{k,g} \right) O_l^{g,h} \phi_h(t), \right. \\
\left. \sum_{l,g,h} \left(\langle \bar{\varepsilon}(t)^T \mathbf{m}_l, \phi_g(t) \rangle_{H^1} \right) O_l^{g,h} \phi_h(t) \right\rangle_{H^1} = 0,
\end{aligned}$$

From (26), (27) and (28), the left-hand side of the last equation becomes

$$\begin{aligned}
& E \left\langle \sum_{l,g,h} O_l^{g,h} \phi_h(t) \right. \\
& \quad \left(\sum_{i,j,k} (\langle \Psi_{k,1}, \alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t) \rangle_{L^2} + \langle \Psi_{k,2}, \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t) \rangle_{L^2}) d_{l,ij}^{k,g} \right) , \\
& \quad \left. \sum_{l,g,h} O_l^{g,h} \phi_h(t) \langle \bar{\varepsilon}(t)^T \mathbf{m}_{l_2}, \phi_g(t) \rangle_{H^1} \right\rangle_{H^1} \\
&= E \left(\sum_{l_1, l_2, g_1, g_2, h_1, h_2} O_{l_1}^{g_1, h_1} O_{l_2}^{g_2, h_2} \langle \phi_{h_1}(t), \phi_{h_2}(t) \rangle_{H^1} \right. \\
& \quad \left. \sum_{i_1, j_1, k_1} \sqrt{\lambda_{k_1}} X_{i_1 j_1, k_1} d_{l_1, i_1 j_1}^{k_1, g_1} \sum_{i_2, j_2, k_2} \frac{\sqrt{\lambda_{k_2}} X_{i_2 j_2, k_2} \langle \tilde{\psi}_{k_2}(t), \phi_{g_2}(t) \rangle_{H^1}}{r} m_{l_2, i_2} \right) \\
&= \sum_{l_1, l_2, g_1, g_2, h_1, h_2} O_{l_1}^{g_1, h_1} O_{l_2}^{g_2, h_2} \langle \phi_{h_1}(t), \phi_{h_2}(t) \rangle_{H^1} \\
& \quad \sum_{i_1, i_2, j_1} \sum_{k_1, k_2} \sqrt{\lambda_{k_1}} \sqrt{\lambda_{k_2}} d_{l_1, i_1 j_1}^{k_1, g_1} m_{l_2, i_2} E \left(X_{i_1 j_1, k_1} \frac{\sum_{j_2} X_{i_2 j_2, k_2}}{r} \right) \langle \tilde{\psi}_{k_2}(t), \phi_{g_2}(t) \rangle_{H^1} \\
&= \sum_{l_1, l_2, g_1, g_2, h_1, h_2} O_{l_1}^{g_1, h_1} O_{l_2}^{g_2, h_2} \langle \phi_{h_1}(t), \phi_{h_2}(t) \rangle_{H^1} \\
& \quad \sum_{i_1, i_2, j_1} \sum_{k_1, k_2} \sqrt{\lambda_{k_1}} \sqrt{\lambda_{k_2}} d_{l_1, i_1 j_1}^{k_1, g_1} m_{l_2, i_2} \delta_{i_1}^{i_2} \frac{\delta_{k_1}^{k_2} + (r-1)\rho(k_1, k_2)}{r} \langle \tilde{\psi}_{k_2}(t), \phi_{g_2}(t) \rangle_{H^1} \\
&= \sum_{l_1, l_2, g_1, g_2, h_1, h_2} O_{l_1}^{g_1, h_1} O_{l_2}^{g_2, h_2} \langle \phi_{h_1}(t), \phi_{h_2}(t) \rangle_{H^1} \\
& \quad \sum_{k_1, k_2} \sqrt{\lambda_{k_1}} \sqrt{\lambda_{k_2}} \left(\sum_{i,j} d_{l_1, i j}^{k_1, g_1} m_{l_2, i} \right) \frac{\delta_{k_1}^{k_2} + (r-1)\rho(k_1, k_2)}{r} \langle \tilde{\psi}_{k_2}(t), \phi_{g_2}(t) \rangle_{H^1} \\
&= 0,
\end{aligned}$$

the last equality being a consequence of (22).

REFERENCES

- [1] G. Aletti, C. May, and C. Tommasi. Optimal designs for linear models with functional responses. In E. G. Bongiorno, E. Salinelli, A. Goia, and P. Vieu, editors, *Contributions in Infinite-Dimensional Statistics and Related Topics*, pages 19–24. Società Editrice Esculapio, 2014.
- [2] A. C. Atkinson, A. N. Donev, and R. D. Tobias. *Optimum experimental designs, with SAS*, volume 34 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2007.
- [3] S. Baraldo, F. Ieva, A. M. Paganoni, and V. Vitelli. Outcome prediction for heart failure telemonitoring via generalized linear models with functional covariates. *Scand. J. Stat.*, 40(3):403–416, 2013.
- [4] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. Theory and applications.
- [5] J.-M. Chiou, H.-G. Müller, and J.-L. Wang. Functional response models. *Statist. Sinica*, 14(3):675–693, 2004.
- [6] V. V. Fedorov. *Theory of optimal experiments*. Academic Press, New York-London, 1972. Translated from the Russian and edited by W. J. Studden and E. M. Klimko, Probability and Mathematical Statistics, No. 12.

- [7] V. V. Fedorov and P. Hackl. *Model-oriented design of experiments*, volume 125 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1997.
- [8] F. Ferraty and P. Vieu. *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York, 2006. Theory and practice.
- [9] P. Hall, H.-G. Müller, and F. Yao. Estimation of functional derivatives. *Ann. Statist.*, 37(6A):3307–3329, 2009.
- [10] L. Horváth and P. Kokoszka. *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York, 2012.
- [11] J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, 2:849–879, 1974.
- [12] A. Kufner. *Weighted Sobolev spaces*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1985. Translated from the Czech.
- [13] C. J. Marley. *Screening experiments using supersaturated designs with application to industry*. PhD thesis, University of Southampton, 2011.
- [14] C. J. Marley and D. C. Woods. A comparison of design and model selection methods for supersaturated experiments. *Comput. Statist. Data Anal.*, 54(12):3158–3167, 2010.
- [15] A. Menafoglio and P. Secchi. A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electron. J. Stat.*, 7:2209–2240, 2013.
- [16] G. Perrin, C. Soize, D. Duhamel, and C. Funfschilling. Karhunen-Loève expansion revisited for vector-valued random fields: scaling, errors and optimal basis. *J. Comput. Phys.*, 242:607–622, 2013.
- [17] D. Pigoli and L. Sangalli. Wavelets in functional data analysis: Estimation of multidimensional curves and their derivatives. *Computational Statistics and Data Analysis*, 56(6):1482–1498, 2012. cited By 4.
- [18] F. Pukelsheim. *Optimal design of experiments*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1993. A Wiley-Interscience Publication.
- [19] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.
- [20] L. Sangalli, P. b. Secchi, S. Vantini, and A. Veneziani. Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 58(3):285–306, 2009. cited By 13.
- [21] Q. Shen and J. Faraway. An F test for linear models with functional responses. *Statist. Sinica*, 14(4):1239–1257, 2004.
- [22] Q. Shen and H. Xu. Diagnostics for linear models with functional responses. *Technometrics*, 49(1):26–33, 2007.
- [23] S. D. Silvey. *Optimal design*. Chapman & Hall, London-New York, 1980. An introduction to the theory for parameter estimation, Monographs on Applied Probability and Statistics.
- [24] D. C. Woods, C. J. Marley, and S. M. Lewis. Designed experiments for semi-parametric models and functional data with a case-study in tribology. In *World Statistics Congress, Hong Kong*, 2013.

ADAMSS CENTER & DEPARTMENT OF MATHEMATICS, UNIVERSITÀ DEGLI STUDI DI MILANO,
20131 MILANO, ITALY

E-mail address: `giacomo.aletti@unimi.it`

UNIVERSITY OF EASTERN PIEDMONT, ITALY

E-mail address: `caterina.may@unipmn.it`

UNIVERSITÀ DEGLI STUDI DI MILANO, ITALY

E-mail address: `chiara.tommasi@unimi.it`