Smoothing Splines Estimators in Functional Linear Regression with Errors-in-Variables

Hervé Cardot[1], Christophe Crambes[2], Alois Kneip[3] and Pascal Sarda[4]

E-mail addresses: [1]cardot@enesad.inra.fr, [2]Christophe.Crambes@math.ups-tlse.fr,
[3]akneip@uni-bonn.de, [4]Pascal.Sarda@math.ups-tlse.fr

July 11, 2006

Abstract The Total Least Squares method is generalized in the context of the functional linear model. A smoothing splines estimator of the functional coefficient of the model is first proposed without noise in the covariates and an asymptotic result for this estimator is obtained. Then, this estimator is adapted to the case where the covariates are noisy and an upper bound for the convergence speed is also derived. The estimation procedure is evaluated by means of simulations.

Key words Functional Linear Model, Smoothing Splines, Penalization, Errors-in-Variables, Total Least Squares.

1 Introduction

A very common problem in statistics is to explain the effects of a covariate on a response (variable of interest). While the covariate are usually considered as a vector of scalars, nowadays, in many applications (for instance in climatology, remote sensing, linguistics, ...) the data come from the observation of a continuous phenomenon over time or space: see [23] or [11] for examples. The increasing performances of measurement instruments permit henceforth to collect these data on dense grids and they can not be considered anymore as variables taking values in $\mathbb{R}^p$. This necessitated to develop for this kind of data *ad hoc* techniques which have been popularized under the name of *functional data analysis* and have been deeply studied these

[1]INRA Dijon, CESAER - ENESAD, 26, bd Docteur Petitjean, BP 87999, 21079 Dijon Cedex, France

[2]Université Paul Sabatier, Laboratoire de Statistique et Probabilités, UMR C5583, 118, route de Narbonne, 31062 Toulouse Cedex, France

[3]Statistische Abteilung, Department of Economics, Universität Bonn, Adenauerallee 24, 53113 Bonn, Germany

[4]Université Toulouse-le-Mirail, GRIMM, EA 3686, 5, allées Antonio Machado, 31058 Toulouse Cedex 9, France

last years (to get a theoretical and practical overview on functional data analysis, we refer to the books [1], [22], [23] and [11]).

Our study takes place in this framework of functional data analysis in the context of regression estimation evocated above. Thus, we consider here the case of a functional covariate while the response is scalar. To be more precise, we consider observations $(X_i, Y_i)_{i=1,\dots,n}$, where the X_i's are real functions defined on an interval I of $\mathbb{R}$ with the assumption that it is square integrable over I. As usually assumed in the literature, we then work on the separable real Hilbert space $L^2(I)$ of such functions f defined on I such that $\int_I f(t)^2 dt$ is finite. This space is endowed with its usual inner product $\langle ., . \rangle$ defined by $\langle f, g \rangle = \int_I f(t)g(t)dt$ for $f, g \in L^2(I)$, and the associated norm is noted $\|.\|_{L^2}$. Now, the model we consider to sumarize the link between covariates X_i and responses Y_i is a linear model introduced in [21] and defined by

$$Y_i = \int_I \alpha(t) X_i(t) dt + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where $\alpha \in L^2(I)$ is an unknown functional parameter and $\epsilon_i, i = 1, \dots, n$ are i.i.d. real random variables satisfying $\mathbb{E}(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma_\epsilon^2$. The functional parameter α has been estimated in various ways in the literature: see [22], [20], [3] and [4]. Here, our final goal is to deal with the problem of estimating α in the case where $X_i(t)$ is corrupted by some unobservable error.

Before going further, let us note that there can be different ways to generate the curves X_i. One possibility is a fixed design, that is, $X_1, \dots, X_n$ are fixed, non-random functions. Examples are experiments in chemical or engineering applications, where X_i corresponds to functional responses obtained under various, predetermined experimental conditions (see for instance [7]). In other applications one may assume a random design, where $X_1, \dots, X_n$ are an i.i.d. sample. In any case, $Y_1, \dots, Y_n$ are independent and the expectations always refer to the probability distribution induced by the random variables $\epsilon_1, \dots, \epsilon_n$, only. In the case of random design, they thus formally have to be interpreted as conditional expectation given $X_1, \dots, X_n$. This implies for instance that $\mathbb{E}(X_i | X_i) = 0$ and $\mathbb{E}(\epsilon_i^2 | X_i) = \sigma_\epsilon^2$.

In what precedes it is implicitly assumed that the curves X_i are observed without error (in model (1) all the errors are confined to the variable Y_i by the way of ϵ_i). Unfortunately, this assumption does not seem to be very realistic in practice, and many errors (instrument errors, human errors, ...) prevent to know $X_1, \dots, X_n$ exactly. Furthermore, it is to be noticed that in practice, the whole curves are not available, so we suppose in the following that the curves are observed in p discretization points $t_1 < \ldots < t_p$ belonging to I, that we will take equispaced in order to simplify. Taking from now on $I = [0, 1]$ in order to simplify the notations, we thus have $t_j - t_{j-1} = \frac{1}{p}$ for all $j = 2, \ldots, p$. Thus, we observe discrete noisy trajectories

$$W_i(t_j) = X_i(t_j) + \delta_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, p, \tag{2}$$

where $(\delta_{ij})_{i=1,\ldots,n, j=1,\ldots,p}$ is a sequence of independent real random variables, such that, for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$

$$\mathbb{E}(\delta_{ij}) = 0,$$

and

$$\mathbb{E}(\delta_{ij}^2) = \sigma_\delta^2.$$

The noise components δ_{ij} are not discrete realizations of continuous time "random noise" stochastic process and must be interpreted as random measurement errors at the finite discretization points (see *e.g.* [2] and [6] for similar points of view).

The problem of the *Errors-in-Variables* linear model has already been studied in many ways in the case where the covariate takes values in $\mathbb{R}$ or $\mathbb{R}^p$, that is to say when it is univariate or multivariate. For instance, the maximum likelihood method has been applied to this context (see [12]), and asymptotic results have been obtained (see for example [15]). Because this problem is strongly linked to the problem of solving linear systems

$$\mathbf{A}\boldsymbol{x} \approx \mathbf{b},$$

where $\boldsymbol{x} \in \mathbb{R}^p$ is unknown, $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A}$ is a matrix of size $n \times p$, some numerical approaches have also been proposed. One of the most famous is the *Total Least Squares* (TLS) method (see for example [17] or [27]).

Now, coming back to model (1), very few works have been done in the case of Errors-in-Variables: in a recent work [6] a two step approach is proposed which consists in first smoothing the noisy trajectories in order to get denoised curves and then build functional estimators. The point of view adopted here is quite different and deals with the extension of the TLS approach in the context of the functional linear model.

Let us describe our formal framework for Errors-in-Variables which is inspired from what is done in the literature. We introduce a discretized version of the inner product $\langle ., . \rangle$ will be denoted by $\langle ., . \rangle_p$ and defined for $f, g \in L^2(I)$ by

$$\langle f, g \rangle_p = \frac{1}{p} \sum_{j=1}^{p} f(t_j) g(t_j).$$

This approximation of $\langle ., . \rangle$ by $\langle ., . \rangle_p$ is valid only if p is large enough, so we assume this from now on. In this context of discretized curves, relation (1) then writes

$$Y_i = \frac{1}{p} \sum_{j=1}^{p} \alpha(t_j) X_i(t_j) + \epsilon_i, \quad i = 1, \ldots, n. \tag{3}$$

Finally the problem is to estimate α using data $(W_i(t_j), Y_i)_{i=1,\ldots,n, j=1,\ldots,p}$ where $W_1(t_j), \ldots, W_n(t_j)$ are noisy observations of $X_1(t_j), \ldots, X_n(t_j)$ for $j = 1, \ldots, p$. The generalization of the TLS method to the case where X_i is a functional random variable is presented in section 3. As in the multivariate case, the TLS method consists in a modification of a (penalized) least squares estimator of α for non-noisy observations: see [20] and [4] for such kind of estimators based on B-splines with two different penalties. Here, we introduce another estimator based on smoothing splines which, as far as we know, has not been studied previously in the literature. Some convergence results are also given in section 2 (in the non-noisy case) which serve as a basis for convergence results of the TLS estimator given in section 3. A more detailed study of the asymptotic behavior of the TLS estimator will be the subject of a forthcoming work. In section 4, the results of convergence for the TLS estimator are commented. Section 5 is devoted to some numerical simulations presenting an evaluation of our estimation procedure. Finally, in section 6, we give the proof of our results.

2 Estimation of α in the non-noisy case

We adopt the following matrix notations: $\mathbf{Y} = (Y_1, \ldots, Y_n)^\tau$, $\mathbf{X}_i = (X_i(t_1), \ldots, X_i(t_p))^\tau$ for all $i = 1, \ldots, n$, $\boldsymbol{\alpha} = (\alpha(t_1), \ldots, \alpha(t_p))^\tau$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\tau$. Moreover, we denote by $\mathbf{X}$ the $n \times p$ matrix with general term $X_i(t_j)$ for all $i = 1, \ldots, n$ and for all $j = 1, \ldots, p$. Using these notations, the model (3) then writes

$$\mathbf{Y} = \frac{1}{p} \mathbf{X} \boldsymbol{\alpha} + \boldsymbol{\epsilon}. \tag{4}$$

Our estimation procedure for α is motivated by the popular smoothing splines approach. Then, we want to estimate α as a smooth function, *i.e.* we assume that α is m times differentiable for some fixed $m \in \mathbb{N}^*$.

At first we briefly come back to the smoothing splines procedure in the usual univariate case. For some noisy observations z_i of a smooth function $f(t_i)$ at design points $t_1, \ldots, t_p$, an estimate $\widehat{f}$ is obtained by minimizing $\frac{1}{p}\sum_i (z_i - v(t_i))^2 + \rho \int_I v^{(m)}(t)^2 dt$ for some smoothing parameter $\rho > 0$. Minimization takes place over all functions v in an m-th order Sobolev space, that is $D^m v \in L^2(I)$. It can be shown (for an overview of results in spline theory, consider [8] and [10]) that the solution $\widehat{f}$ is in the space $NS^m(t_1, \ldots, t_p)$ of *natural splines* of order $2m$ with knots at $t_1, \ldots, t_p$. This is a p-dimensional linear functions space with $D^m v \in L^2(I)$ for any $v \in NS^m(t_1, \ldots, t_p)$, and there exist basis functions $b_1, \ldots, b_p$ such that $NS^m(t_1, \ldots, t_p) = \left\{ \sum_j \theta_j b_j \mid \theta_1, \ldots, \theta_p \in \mathbb{R} \right\}$. Different possible basis functions proposed by various authors are discussed in [10]. An important property of natural splines is that for any vector $\boldsymbol{w} = (w_1, \ldots, w_p)^\tau \in \mathbb{R}^p$, there exists a unique natural spline interpolant $s_{\boldsymbol{w}}$ with $s_{\boldsymbol{w}}(t_j) = w_j$, $j = 1, \ldots, p$. With $\mathbf{b}(t) = (b_1(t), \ldots, b_p(t))^\tau$ and $\mathbf{B}$ denoting the $p \times p$ matrix with elements $b_i(t_j)$, $s_{\boldsymbol{w}}$ is given by

$$s_{\boldsymbol{w}}(t) = \mathbf{b}(t)^\tau \left(\mathbf{B}^\tau \mathbf{B}\right)^{-1} \mathbf{B}^\tau \boldsymbol{w}. \tag{5}$$

Moreover such a spline interpolant satisfies the following fine property

$$\int_I s_{\boldsymbol{w}}^{(m)}(t)^2 dt \leq \int_I f^{(m)}(t)^2 dt \ \text{ for any other function } f$$
$$\text{with } f^{(m)} \in L^2(I) \text{ and } f(t_j) = w_j, j = 1, \ldots, p. \tag{6}$$

The inequality (6) implies that the solution $\widehat{f}$ is given by $\widehat{f} = s_{\widehat{\boldsymbol{w}}}$, where $\widehat{\boldsymbol{w}}$ is obtained by minimizing $\frac{1}{p}\sum_i (z_i - w_i)^2 + \rho \int_I s_{\boldsymbol{w}}^{(m)}(t)^2 dt$ over all vectors $\boldsymbol{w} \in \mathbb{R}^p$.

These ideas readily generalize to the problem of estimating $\boldsymbol{\alpha}$ in (4) and then the function α. An estimator $\widehat{\boldsymbol{\alpha}}^*_{FLS,X}$ may be obtained by solving the minimization problem

$$\min_{\boldsymbol{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p} \mathbf{X} \boldsymbol{a} \right\|^2 + \rho \int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt \right\}, \tag{7}$$

where $\|.\|$ stands for the usual Euclidean norm, and $\rho > 0$ is a smoothing parameter allowing a trade-off between the goodness-of-fit to the data and the smoothness of the fit. By (5), we have $\int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt = \boldsymbol{a}^\tau \mathbf{A}^*_m \boldsymbol{a}$, where $\mathbf{A}^*_m = \mathbf{B}\left(\mathbf{B}^\tau \mathbf{B}\right)^{-1} \left[\int_I \mathbf{b}^{(m)}(t) \mathbf{b}^{(m)}(t)^\tau dt\right] \left(\mathbf{B}^\tau \mathbf{B}\right)^{-1} \mathbf{B}^\tau$ is a $p \times p$ matrix. Therefore, (7) can be reformulated in the form

$$\min_{\boldsymbol{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\mathbf{Y}-\frac{1}{p}\mathbf{X}\boldsymbol{a}\right\|^2+\rho\boldsymbol{a}^\tau\mathbf{A}_m^*\boldsymbol{a}\right\}, \tag{8}$$

leading to the solution

$$\widehat{\boldsymbol{\alpha}}^*_{FLS,X}=\frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\rho\mathbf{A}_m^*\right)^{-1}\mathbf{X}^\tau\mathbf{Y}=\frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}+\rho p\mathbf{A}_m^*\right)^{-1}\mathbf{X}^\tau\mathbf{Y}.$$

However, there is a problem with this estimator which is due to the structure of the eigenvalues of $p\mathbf{A}_m^*$. These eigenvalues have been studied by many authors and a discussion of general results is given by [10]. The most precise results in our context are presented in [26]. It is shown that this matrix has exactly m zero eigenvalues $\mu_{1,p}=\ldots=\mu_{m,p}=0$, while as $p\to\infty$,

$$\sum_{j=m+1}^{p}\frac{1}{\mu_{j,p}}\longrightarrow\sum_{j=m+1}^{\infty}(\pi j)^{-2m}, \tag{9}$$

where $0<\mu_{m+1,p}<\ldots<\mu_{p,p}$ denote the $p-m$ non-zero eigenvalues of $p\mathbf{A}_m^*$. The series given in (9) converges for $m\neq 0$, so we assume this in the following.

Due to the m zero eigenvalues, existence of $\widehat{\boldsymbol{\alpha}}^*_{FLS,X}$ can only be guaranteed by introducing constraints on the structure of $\mathbf{X}$. This can, however, be avoided by introducing a minor modification of this estimator. The m-dimensional eigenspace corresponding to $\mu_{1,p}=\ldots=\mu_{m,p}=0$ is the linear vector space generated by all (discretized) polynomials of degree m, that is, E_m consists of all vectors $\boldsymbol{w}\in\mathbb{R}^p$ with $w_i=\theta_1+\sum_{j=1}^m\theta_{j+1}t_i^j$, $i=1,\ldots,p$, for some coefficients $\theta_1,\ldots,\theta_{m+1}$. Let $\mathbf{P}_m$ denote the $p\times p$ projection matrix projecting into the space E_m, and set $\mathbf{A}_m=\mathbf{P}_m+p\mathbf{A}_m^*$. Our final estimator $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ is then defined by

$$\widehat{\boldsymbol{\alpha}}_{FLS,X}=\frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y}=\frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}+\rho\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y}, \tag{10}$$

and a corresponding estimator of α is provided by $\widehat{\alpha}_{FLS,X}=s_{\widehat{\boldsymbol{\alpha}}_{FLS,X}}$. It is immediately verified that $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ is solution of the modified minimization problem

$$\min_{\boldsymbol{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\mathbf{Y}-\frac{1}{p}\mathbf{X}\boldsymbol{a}\right\|^2+\frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\}.$$

By definition, the matrix $\mathbf{A}_m$ possesses m eigenvalues equal to 1, while the remaining $p-m$ eigenvalues coincide with the eigenvalues $\mu_{m+1,p} < \ldots < \mu_{p,p}$ of $p\mathbf{A}_m^*$. Thus, by (9), we obtain $\mathrm{Tr}\left(\mathbf{A}_m^{-1}\right) \longrightarrow \sum_{j=m+1}^{\infty}(\pi j)^{-2m} + m =: C_0$ as $p \to \infty$. It follows that for any constant $C_1 > C_0$ there exists a $p_0 \in \mathbb{N}$ such that

$$\mathrm{Tr}\left(\mathbf{A}_m^{-1}\right) \leq C_1, \tag{11}$$

for all $p \geq p_0$.

We will now study the behavior of our estimator for large values of n and p. The behavior of our estimator will be evaluated with respect to the semi-norm

$$\|\boldsymbol{u}\|_\Gamma^2 = \frac{1}{p}\boldsymbol{u}^\tau \left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\right)\boldsymbol{u}.$$

It is well-known that functional linear regression belongs to the class of ill-posed problems. The semi-norm $\|.\|_\Gamma$ may be seen as a discretized version of L^2 semi-norms which are usually applied in this context. It is not possible to derive any bound for the bias by using the Euclidean norm. Suppose, for example, that all functions X_i lie in a low dimensional linear function space $\mathcal{X}$. Then any structure of α which is orthogonal to $\mathcal{X}$ cannot be identified from the data.

The regularity assumption that we will do on α follows.

$(H.1)$ For some $m \in \mathbb{N}^*$, α is m times differentiable and $\alpha^{(m)} \in L^2(I)$.

Then, let $C_2 = \int_I \alpha^{(m)}(t)^2 dt$ and $C_3^* = \int_I \alpha(t)^2 dt$. By construction of $\mathbf{P}_m$, $\mathbf{P}_m\boldsymbol{a}$ provides the best approximation (in a least squares sense) of $\boldsymbol{a}$ by (discretized) polynomials of degree m, and $\frac{1}{p}\boldsymbol{a}^\tau\mathbf{P}_m\boldsymbol{a} \leq \frac{1}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a} \longrightarrow C_3^*$ as $p \to \infty$. Let C_3 denote an arbitrary constant with $C_3^* < C_3 < \infty$. There then exists a $p_1 \in \mathbb{N}$ with $p_1 \geq p_0$ such that $\frac{1}{p}\boldsymbol{a}^\tau\mathbf{P}_m\boldsymbol{a} \leq C_3$ for all $p \geq p_1$.

As noticed before, $X_1, \ldots, X_n$ can be either fixed, non-random functions or an i.i.d. sample of random functions. In any case, expected values and variance of $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ as stated in the theorem will refer to the probability distribution induced by the random variable ϵ. In the case of random design, they stand for conditional expectation given $X_1, \ldots, X_n$.

Theorem 1 *Under assumption $(H.1)$ and the definitions of C_1, C_2, C_3, p_1, we obtain for all $n \in \mathbb{N}$, all $p \geq p_1$ and every matrix $\mathbf{X} \in \mathbb{R}^n \times \mathbb{R}^p$*

$$\|\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X}) - \boldsymbol{\alpha}\|_\Gamma^2 \leq \rho\left(\frac{1}{p}\boldsymbol{a}^\tau\mathbf{P}_m\boldsymbol{a} + C_2\right) \leq \rho\left(C_3 + C_2\right), \tag{12}$$

as well as

$$\frac{1}{p}\mathbb{E}\left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X}-\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2\right)\leq\frac{\sigma_\epsilon^2}{n\rho}C_1. \tag{13}$$

Remark When adding some additional constraint like

$$(H.2)\qquad \sup_i\sup_j|X_i(t_j)|\leq C_4<+\infty,\ \text{for all } n,p,$$

or when $(H.2)$ is almost surely satisfied in the case of a random design, then the variance can also bound the semi-norm $\|.\|_\Gamma$,

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,X}-\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|_\Gamma^2\leq\frac{C_4}{p}\mathbb{E}\left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X}-\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2\right),$$

and the theorem implies that

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,X}-\boldsymbol{\alpha}\|_\Gamma^2=O_P\left(n^{-1/2}\right),$$

if $\rho\sim n^{-1/2}$ as $n\to\infty$. This rate obviously compares favorably to existing rates in the literature.

3 Total Least Squares method for functional covariates

We address now the estimation of α from noisy covariates. At first, let us describe how the TLS method works in the case of a covariate belonging to $\mathbb{R}^p$. In that case, we have

$$Y_i=\mathbf{X}_i^\tau\boldsymbol{\alpha}+\epsilon_i,\ i=1,\dots,n,$$

and

$$\mathbf{W}_i=\mathbf{X}_i+\boldsymbol{\delta}_i,\ i=1,\dots,n,$$

where $\boldsymbol{\alpha}$, $\mathbf{X}_i$, $\mathbf{W}_i$ and $\boldsymbol{\delta}_i$ are vectors of $\mathbb{R}^p$. The TLS approach relies on the simultaneous estimation of $\boldsymbol{\alpha}$ and $\mathbf{X}_i$ by considering the minimization problem (see for example [27])

$$\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^n\left[(Y_i-\mathbf{X}_i^\tau\boldsymbol{a})^2+\ (\mathbf{X}_i-\mathbf{W}_i)^\tau(\mathbf{X}_i-\mathbf{W}_i)\right]\right\}. \tag{14}$$

The TLS algorithm solving (14) is given in [27]. In some cases, the singular values of the matrix $\mathbf{W}$ can quickly decrease to zero, and the minimization problem (14) is then *ill-conditioned*. A possible way to circumvent this problem is to introduce a regularization in (14), and the minimization problem we consider is then (see [16])

$$\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[(Y_i-\mathbf{X}_i^\tau\boldsymbol{a})^2+(\mathbf{X}_i-\mathbf{W}_i)(\mathbf{X}_i-\mathbf{W}_i)^\tau\right]\right.$$
$$\left.+\rho\boldsymbol{a}^\tau\mathbf{L}^\tau\mathbf{L}\boldsymbol{a}\right\}, \qquad (15)$$

where $\mathbf{L}$ is a $p\times p$ matrix and ρ is a regularization parameter allowing to deal with the ill-conditioning of the design matrix $\mathbf{W}^\tau\mathbf{W}$. Indeed, the TLS solution to the minimization problem (15) is given by

$$\widehat{\boldsymbol{\alpha}}_{TLS}=\left(\mathbf{W}^\tau\mathbf{W}+\rho\mathbf{L}^\tau\mathbf{L}-\sigma_k^2\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y}, \qquad (16)$$

where σ_k is the smallest non-zero singular value of the matrix $(\mathbf{W},\mathbf{Y})$ and $\mathbf{I}_p$ is the $p\times p$ identity matrix.

In our functional situation, we consider model (4) and using the same matricial notations as in section 2 we write

$$\mathbf{W}=\mathbf{X}+\boldsymbol{\delta},$$

where $\mathbf{W}$ and $\boldsymbol{\delta}$ are the $n\times p$ matrices with respective general terms $W_i(t_j)$ and δ_{ij}. So, the minimization problem we consider now is the following one: we are looking for an estimation $\widehat{\boldsymbol{\alpha}}^*_{FTLS}$ of $\boldsymbol{\alpha}$, solution of the minimization problem

$$\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\left(Y_i-\frac{1}{p}\mathbf{X}_i^\tau\boldsymbol{a}\right)^2+\frac{1}{p}\left\|\mathbf{X}_i-\mathbf{W}_i\right\|^2\right]+\frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\}, \quad (17)$$

where the matrix $\mathbf{A}_m$ is the one introduced in section 2. Now, with these notations, we have the following result.

Proposition 1 *The solution of the minimization problem (17) is given by*

$$\widehat{\boldsymbol{\alpha}}^*_{FTLS}=\frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\sigma_k^2\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y}, \qquad (18)$$

where σ_k^2 is the smallest non-zero eigenvalue of the matrix

$$\frac{1}{n}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)+\frac{\rho}{p}\begin{pmatrix}\mathbf{A}_m & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix}.$$

In equation (18), computational problems can appear due to the value of σ_k^2 which may be close to zero. Indeed, the eigenvalues of $\frac{1}{n}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)^{\tau}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)$ are known to decrease rapidly to zero, and this can of course cause numerical problems with the computation of σ_k^2. Nevertheless, we can circumvent this problem using the following result.

Proposition 2 *Suppose that for some constant $C_5 > 0$*

$$(H.3) \qquad \mathbb{E}\left(\delta_{ir}^2\delta_{is}^2\right) \leq C_5, \quad r,s = 1,\ldots,p.$$

Then, if moreover $(H.2)$ holds, we have

$$\frac{1}{np^2}\mathbf{W}^{\tau}\mathbf{W} = \frac{1}{np^2}\mathbf{X}^{\tau}\mathbf{X} + \frac{\sigma_\delta^2}{p^2}\mathbf{I}_p + \mathbf{R}, \tag{19}$$

where $\mathbf{R}$ is a matrix such that $\|\mathbf{R}\| = O_P\left(\frac{1}{n^{1/2}p}\right)$, $\|.\|$ being the usual norm of a matrix.

The last problem is that σ_δ^2 is not always known. There are several ways to estimate it. We choose to use the estimator presented in [14] and given by (as we are in the case of equispaced measurement points)

$$\widehat{\sigma}_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{6(p-2)}\sum_{j=2}^{p-1}\left[W_i(t_{j-1}) - W_i(t_j) + W_i(t_{j+1}) - W_i(t_j)\right]^2. \tag{20}$$

This leads us to change the former estimator of $\boldsymbol{\alpha}$ given by (18) and to take instead

$$\widehat{\boldsymbol{\alpha}}_{FTLS} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^{\tau}\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}\mathbf{W}^{\tau}\mathbf{Y}, \tag{21}$$

and again a corresponding estimator of α is provided by $\widehat{\alpha}_{FTLS}$.

The asymptotic behavior $\widehat{\boldsymbol{\alpha}}_{FTLS}$ is given in the following theorem.

Theorem 2 *Under assumptions $(H.1)$ - $(H.3)$, if we also assume that Y_i and δ_{ij} are independent for all $i = 1,\ldots,n$ $j = 1,\ldots,p$ and that $\mathbb{E}\left(Y_i^2\right) < +\infty$, then we have*

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^{1/2}\rho}\right). \tag{22}$$

4 Some comments

(i) In the expression (21) of the estimator of $\boldsymbol{\alpha}$, the term $-\frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p$ acts as a deregularization term. It allows us to deal with the bias introduced by the fact that we only know the matrix $\mathbf{W}$ instead of the "true" one $\mathbf{X}$.

(ii) In theorem 2, the order $\sigma_\delta/(n^{1/2}p^{1/2}\rho)$ given by relation (22) is a result in accordance with the intuition. The estimation will be improved for a high number p of discretization points and will collapse (at least in practice, see the simulations in section 5) if σ_δ becomes too high.

(iii) An immediate corollary of theorems 1 and 2 is

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \boldsymbol{\alpha}\|_\Gamma^2 = O_P\left(\frac{1}{n\rho} + \rho + \frac{\sigma_\delta^2}{np\rho^2}\right).$$

If we compare these three terms, we can see that, for p large enough (more precisely for p such that ρp goes to infinity as n goes to infinity), it remains

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \boldsymbol{\alpha}\|_\Gamma^2 = O_P\left(\frac{1}{n\rho} + \rho\right),$$

and then, for $\rho \sim n^{-1/2}$,

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \boldsymbol{\alpha}\|_\Gamma^2 = O_P\left(n^{-1/2}\right).$$

This means that, if the number of discretization points is large enough, we obtain the same upper bound for the convergence speed of the $FTLS$ estimator as the FLS estimator using the true curves $X_1, \ldots, X_n$. On the other hand, when p goes slowly to infinity (more precisely if $1/p$ goes to zero slower than ρ), then the contribution of the term $\sigma_\delta^2/(np\rho^2)$ may not be negligible anymore. In that case, if we still take $\rho \sim n^{-1/2}$, then we will have $\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \boldsymbol{\alpha}\|_\Gamma^2 = O_P(1/p) = O_P(n^{-\gamma})$ with $0 < \gamma < 1/2$.

(iv) Let us see what happens for the FLS estimator using the noisy curves $W_1, \ldots, W_n$. The estimator of $\boldsymbol{\alpha}$ is then given by

$$\widehat{\boldsymbol{\alpha}}_{FLS,W} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{W}^\tau\mathbf{Y}. \tag{23}$$

If p is large enough, a calculus analogous to the one used in the proof of theorem 2 leads us to

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,W} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^{1/2}\rho}\right),$$

that is to say we have the same upper bound of convergence speed for $\widehat{\boldsymbol{\alpha}}_{FLS,W}$ and $\widehat{\boldsymbol{\alpha}}_{FTLS}$. However, if p is not large enough (more precisely if p is negligible compared with $n^{1/2}$ and if $p\rho$ goes to zero as n goes to infinity), then we obtain

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,W} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^{3/2}\rho^2}\right),$$

which is then an upper bound bigger than the previous one. However, these results are upper bounds and we do not know if we can do better. Nevertheless, the results obtained in the simulations (see section 5) allow us to think that we improve the estimation (see last remark) using the $FTLS$ estimator instead of the FLS estimator with the noisy curves $W_1, \ldots, W_n$.

(v) Using some heuristic arguments to expand the mean quadratic error of estimation of $\boldsymbol{\alpha}$ (similarly to what is done in [2]), we can see that it is generally better to consider the $FTLS$ estimator compared to the FLS one with the variable W. More precisely, using the same notations as before, let us denote

$$\widehat{\boldsymbol{\alpha}}(\lambda) = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \lambda\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y},$$

where λ is a positive real number such that the matrix $\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\lambda\mathbf{I}_p$ is positive definite. Then we have the following result, which proof is given in section 6.

Proposition 3 *Let* $MISE(\lambda) = \mathbb{E}\left[(\widehat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha})^\tau\left(\widehat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha}\right)\right]$. *If we assume that* $\left(\mathbf{W}^\tau\mathbf{W}\right)^{-1}$ *exists and if* $\rho\|\mathbf{A}_m\|$ *is negligible compared to* $\left\|\frac{1}{np}\mathbf{W}^\tau\mathbf{W}\right\|$, *then we have*

$$\frac{\partial}{\partial\lambda}MISE(\lambda)_{|\lambda=0} < 0.$$

In other words, this result means that it is advantageous to put a deregularization term $-\lambda\mathbf{I}_p$ (with a small positive λ) in order to improve the quality of the estimation relatively to the $MISE$ criterion.

5 A simulation study

5.1 Presentation of the simulation

The aim of this simulation is to evaluate the performances of our estimator $\widehat{\boldsymbol{\alpha}}_{FTLS}$, and to compare it with $\widehat{\boldsymbol{\alpha}}_{FLS,W}$. We also compare $\widehat{\boldsymbol{\alpha}}_{FTLS}$ to

$\widehat{\boldsymbol{\alpha}}_{FLS,\widetilde{W}}$, which is given by the same formula (23) where the curve W is now replaced by a smoothed version $\widetilde{W}$. We can think that this smoothing step has a correcting effect on the noisy curve W, and then this smoothed curve $\widetilde{W}$ can be expected to be closer than W to the unknown "true" curve X. This gives us the intuition that the estimator $\widehat{\boldsymbol{\alpha}}_{FLS,\widetilde{W}}$ should be better than $\widehat{\boldsymbol{\alpha}}_{FLS,W}$. To obtain a smoothed version $\widetilde{W}$ of W, we choose to use the Nadaraya-Watson kernel estimator (see for example [19] or [24]). In the simulations, the kernel is the standard normal kernel. For the bandwidth we have tried at first a value chosen by cross validation for each curve (see [19]). We have also tried several other values applying to this cross-validated bandwidth a decreasing or increasing factor. In order to synthesize results, we only give the simulation results when X is non-random (when X is random, the simulation we have done lead to the same kind of conclusions). We have simulated $N = 100$ samples, each being composed of $n = 200$ observations $(W_i, Y_i)_{i=1,\ldots,n}$ from the model given by (1) and (2), where the fixed design curves $X_1, \ldots, X_n$ are defined on $I = [0, 1]$ by

$$X_i(t) = \begin{cases} 10 \sin(2\pi i t) \text{ if } i \text{ is even,} \\ 10 \cos(2\pi i t) \text{ if } i \text{ is odd,} \end{cases}$$

similarly to what is used for the simulation in [7]. Each sample is randomly split into a learning sample of length $n_l = 100$ (this sample is used to build the estimator) and a test sample of length $n_t = 100$ (this sample is used to see the quality of the estimator by the way of computation of error terms). We made simulations for different numbers of discretization points, $p = 50$, $p = 100$ and $p = 200$. Two functions α were considered, either $\alpha(t) = 10 \sin(2\pi t)$ or $\alpha(t) = 10 \sin^3(2\pi t^3)$. Finally, the error terms were chosen as follows: $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.2$ and $\delta_{ij} \sim \mathcal{N}(0, \sigma_\delta^2)$ for all $i = 1, \ldots, n$, $j = 1, \ldots, p$ with either $\sigma_\delta = 0.1$, $\sigma_\delta = 0.2$ or $\sigma_\delta = 0.5$. Concerning the parameters of the spline functions, the order of differentiation in the penalization is fixed to the value $m = 2$. The most important parameter to choose is the smoothing parameter value ρ (see [20]). We present in the next subsection a criterion allowing to select reasonable values and we check the effectiveness of this criterion in the simulations.

5.2 Generalized Cross Validation criteria

In the setting of the estimation of a function f by smoothing splines described in section 2, the most popular method for the selection of ρ is *generalized cross-validation* (see [28]). The first idea is to use *cross-validtion* that is to choose the ρ that yields the best prediction (in a mean squares sense) when prediction of a value is done with the remaining observations. After this, a computational simplification of the cross-validation criterion has been proposed in the literature that leads to the generalized cross-validation: see [28] p. 50. In our Functional Least Squares estimation, we can easily adapt this generalized cross-validation (*GCV*) in the following way. The GCV criterion is defined by

$$GCV_{FLS,W}(\rho) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{\left(1 - \frac{1}{n}\mathrm{Tr}(\mathbf{H}_{FLS,W}(\rho))\right)^2}, \tag{24}$$

where $\mathbf{H}_{FLS,W}(\rho)$ is the "hat matrix" given by

$$\mathbf{H}_{FLS,W}(\rho) = \frac{1}{np}\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{W}^\tau,$$

and $\widehat{\mathbf{Y}} = \mathbf{H}_{FLS,W}(\rho)\mathbf{Y}$. Then, we select the optimal parameter ρ_{GCV} as the one that minimizes the *GCV* criterion (24). The criterion (24) is a direct adaptation of the one introduced in [28] except that the "hat matrix" has been changed for our setting.

Concerning the Functional Total Least Squares estimation, although *Cross Validation* has already been studied in [25], what we want to propose here is a generalization of the *GCV* criterion (24), in the following way. The prediction of Y_i for $i = 1, \ldots, n$ is slightly different in the context of *TLS*. The estimation of the unknown $\mathbf{X}_i$, noted $\widehat{\mathbf{X}}_i$, is given by

$$\widehat{\mathbf{X}}_i = \mathbf{W}_i + \frac{Y_i - \frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau\mathbf{W}_i}{1 + \frac{1}{p}\|\widehat{\boldsymbol{\alpha}}\|^2}\,\widehat{\boldsymbol{\alpha}}, \tag{25}$$

obtained as in [13] by differentiating equation (17) with respect to $\mathbf{X}_i$. Then, we take $\widehat{Y}_i = \langle\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{X}}_i\rangle_p$ as the prediction of Y_i. Then, the proposed *GCV* criterion is given by

$$GCV_{FTLS}(\rho) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle\widehat{\boldsymbol{\alpha}}_{FTLS}, \widehat{\mathbf{X}}_i\rangle_p)^2}{\left(1 - \frac{1}{n}\mathrm{Tr}(\mathbf{H}_{FTLS}(\rho))\right)^2}, \tag{26}$$

where $\mathbf{H}_{FTLS}(\rho)$ is the "hat matrix" given by

$$\mathbf{H}_{FTLS}(\rho) = \frac{1}{np}\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau$$

Then, the optimal parameters ρ_{GCV} as obtained by minimizing the GCV criterion (26). In our simulations, these GCV criteria have been computed for ρ over a grid taking its values among $10^{-2}, 10^{-3}, \ldots, 10^{-8}$.

5.3 Results of the simulation

We use two error criteria to see the quality of the prediction. The first one is the relative mean square error of the estimator of $\boldsymbol{\alpha}$, given by

$$E_1 = \frac{\sum_{j=1}^{p} \left[\widehat{\boldsymbol{\alpha}}(t_j) - \boldsymbol{\alpha}(t_j)\right]^2}{\sum_{j=1}^{p} \boldsymbol{\alpha}(t_j)^2}, \tag{27}$$

and the second one is the mean square error of the prediction of $\mathbf{Y}$, given by

$$E_2 = \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{Y}_i - Y_i\right)^2. \tag{28}$$

These errors, evaluated on $N = 100$ simulated test samples, are given in tables 1 and 3 for the different values of p and the different functions α. We have computed the FLS estimator of α using the unknown true curves X (in order to have a reference), the observed curves W and the smooth version $\widetilde{W}$ of the observed curves W.

We can see that the $FTLS$ estimator always improves the prediction compared to FLS, W, and the improvement is really interesting when p is small with a relatively important noise level σ_δ. We can also see that the estimators $FTLS$ and $FLS, \widetilde{W}$ are quite close. $FLS, \widetilde{W}$ seems to be better when the noise level σ_δ is small whereas $FTLS$ seems to be better when this noise level becomes high. Nevertheless, it is important to note that the $FTLS$ estimator is faster to compute compared to the $FLS, \widetilde{W}$ one. Moreover, the choice of the parameter h is not evident (choosing it by cross validation is not always the best thing to do whereas it implies additional long computation times, particularly when n is large).

Moreover, it has to be noticed that the prediction is also improved when the number of discretization points increases. We can also see that the error increases between table 1 and table 3, mainly because of the shape of the second function α, which is less smooth than the first one.

Table 2 gives the estimated values of σ_δ using the estimator defined by (20) and given in [14]. We can see that we get good estimations of σ_δ, and an increasing accuracy with the number of discretization points. It also seems that the quality of the estimation is not much related to the value of σ_δ. Finally, we have plotted on figure 1 an example of the estimation of α in the case where $p = 100$ and $\sigma_\delta = 0.5$, in the case of the function $\alpha(t) = 10\sin^3(2\pi t^3)$. In order not to have too many curves on a same graphic, we chose to plot only the estimators $FTLS$, FLS, X and FLS, W. This graphic tends to confirm the values given in tables 1 and 3.

		E_1			E_2		
		$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
FLS, X	$p = 50$	0.00015	0.00014	0.00013	0.0031	0.0032	0.0032
	$p = 100$	0.00009	0.00010	0.00009	0.0027	0.0026	0.0027
	$p = 200$	0.00005	0.00006	0.00004	0.0024	0.0026	0.0025
$FTLS$	$p = 50$	0.00018	0.00061	0.00232	0.0044	0.0067	0.0180
	$p = 100$	0.00013	0.00065	0.00219	0.0040	0.0063	0.0139
	$p = 200$	0.00009	0.00057	0.00204	0.0035	0.0056	0.0091
FLS, $\widetilde{W}$	$p = 50$	0.00017	0.00080	0.00245	0.0040	0.0065	0.0209
	$p = 100$	0.00011	0.00063	0.00226	0.0036	0.0062	0.0154
	$p = 200$	0.00006	0.00056	0.00210	0.0029	0.0056	0.0112
FLS, W	$p = 50$	0.00020	0.00098	0.00366	0.0050	0.0081	0.0305
	$p = 100$	0.00015	0.00079	0.00344	0.0045	0.0072	0.0245
	$p = 200$	0.00011	0.00063	0.00329	0.0039	0.0067	0.0124

Table 1: Error E_1 on α given by $\alpha(t) = 10\sin(2\pi t)$ and error E_2 of prediction.

	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
$p = 50$	0.1141	0.2075	0.5034
$p = 100$	0.1011	0.2005	0.5005
$p = 200$	0.0999	0.1999	0.4999

Table 2: Estimated values of σ_δ according to the different values of σ_δ and the different values of p.

		E_1			E_2		
		$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
	$p = 50$	0.0508	0.0509	0.0510	0.0427	0.0426	0.0426
FLS, X	$p = 100$	0.0504	0.0504	0.0503	0.0422	0.0423	0.0424
	$p = 200$	0.0503	0.0502	0.0502	0.0414	0.0414	0.0416
	$p = 50$	0.0513	0.0526	0.0630	0.0439	0.0491	0.0830
$FTLS$	$p = 100$	0.0509	0.0522	0.0618	0.0434	0.0476	0.0762
	$p = 200$	0.0506	0.0517	0.0607	0.0429	0.0460	0.0735
	$p = 50$	0.0510	0.0525	0.0645	0.0435	0.0490	0.0851
FLS, $\widetilde{W}$	$p = 100$	0.0507	0.0520	0.0627	0.0429	0.0475	0.0790
	$p = 200$	0.0504	0.0516	0.0614	0.0422	0.0458	0.0763
	$p = 50$	0.0516	0.0530	0.0850	0.0447	0.0504	0.0960
FLS, W	$p = 100$	0.0512	0.0527	0.0822	0.0442	0.0496	0.0889
	$p = 200$	0.0508	0.0521	0.0799	0.0438	0.0488	0.0834

Table 3: Error E_1 on α given by $\alpha(t) = 10\sin^3\left(2\pi t^3\right)$ and error E_2 of prediction.

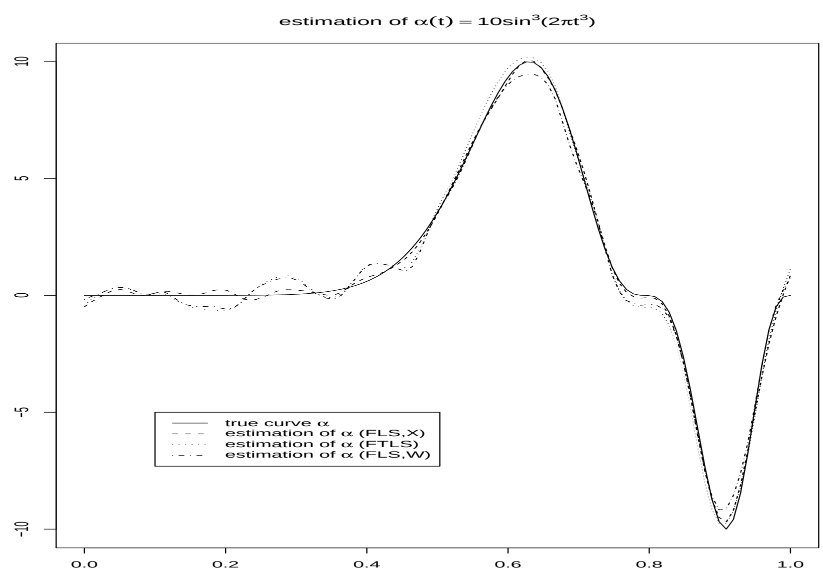

Figure 1: Estimation of α (solid line) with functional least squares using X (dashed line), functional least squares using W (dashed and dotted line) and functional total least squares (dotted line) in cases $\alpha(t) = 10\sin(2\pi t)$ and $\alpha(t) = 10\sin^3\left(2\pi t^3\right)$.

6 Proof of the results

6.1 Proof of theorem 1

First consider relation (12), and note that

$$\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right) = \frac{1}{np^2}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha}.$$

It follows that $\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)$ is solution of the minimization problem

$$\min_{\boldsymbol{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{X}\boldsymbol{a}\right\|^2 + \frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\}.$$

This implies

$$\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{X}\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)\right\|^2 + \frac{\rho}{p}\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)^\tau\mathbf{A}_m\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right) \leq \frac{\rho}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha}.$$

But definition of $\mathbf{A}_m$ and as well (6) lead to

$$\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha} = \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_I s_{\boldsymbol{\alpha}}^{(m)}(t)^2 dt \leq \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_I \alpha^{(m)}(t)^2 dt,$$

and (12) is an immediate consequence. Relation (13) follows from

$$\begin{aligned}
& \frac{1}{p}\mathbb{E}\left(\left[\widehat{\boldsymbol{\alpha}}_{FLS,X}^\tau - \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}^\tau\right)\right]\left[\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)\right]\right) \\
= \; & \frac{1}{p}\mathbb{E}\left(\frac{1}{n^2p^2}\boldsymbol{\epsilon}^\tau\mathbf{X}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{X}^\tau\boldsymbol{\epsilon}\right) \\
= \; & \frac{\sigma_\epsilon^2}{n}\mathrm{Tr}\left[\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\right] \\
\leq \; & \frac{\sigma_\epsilon^2}{n}\mathrm{Tr}\left[\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\right] \leq \frac{\sigma_\epsilon^2}{n}\mathrm{Tr}\left[(\rho\mathbf{A}_m)^{-1}\right] \leq \frac{\sigma_\epsilon^2}{n\rho}C_1.
\end{aligned}$$

This completes the proof of Theorem 1.

6.2 Proof of proposition 1

We have

$$\left(\left(\frac{\mathbf{W}}{p}, \mathbf{Y}\right) - \left(\frac{\boldsymbol{\delta}}{p}, \boldsymbol{\epsilon}\right)\right)\begin{pmatrix}\boldsymbol{\alpha} \\ -1\end{pmatrix} = 0, \qquad (29)$$

which allows us now to write the minimisation problem (17) as follows

$$\min_{\left(\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)-\left(\frac{\boldsymbol{\delta}}{p},\boldsymbol{\epsilon}\right)\right)\binom{\boldsymbol{a}}{-1}=0}\left\{\frac{1}{n}\left\|\left(\frac{\boldsymbol{\delta}}{\sqrt{p}},\boldsymbol{\epsilon}\right)\right\|_F^2+\frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\},$$

where the notation $\|.\|_F$ stands for the usual Frobenius norm, more precisely $\|\mathbf{A}\|_F^2 = \mathrm{Tr}\ (\mathbf{A}^\tau\mathbf{A})$ for every matrix $\mathbf{A}$. Then, we are led to consider the minimization problem

$$\min_{\mathbf{C}\boldsymbol{x}=\mathbf{E}\boldsymbol{x}}\left\{\frac{1}{n}\left\|\left(\frac{\boldsymbol{\delta}}{\sqrt{p}},\boldsymbol{\epsilon}\right)\right\|_F^2+\frac{\rho}{p}\boldsymbol{x}^\tau\mathbf{B}_m\boldsymbol{x}\right\}, \tag{30}$$

with $\mathbf{C}=\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)$, $\mathbf{E}=\left(\frac{\boldsymbol{\delta}}{p},\boldsymbol{\epsilon}\right)$, $\boldsymbol{x}=\binom{\boldsymbol{a}}{-1}$ and $\mathbf{B}_m=\begin{pmatrix}\mathbf{A}_m & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix}$. If we denote $\boldsymbol{\gamma}$ the $(p+1)\times(p+1)$ matrix defined by

$$\boldsymbol{\gamma}=\begin{pmatrix}\mathbf{diag}(1/\sqrt{p},\ldots,1/\sqrt{p}) & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix},$$

we have

$$\begin{aligned}\frac{1}{n}\boldsymbol{x}^\tau\boldsymbol{\gamma}^\tau\left(\frac{\boldsymbol{\delta}}{\sqrt{p}},\boldsymbol{\epsilon}\right)^\tau\left(\frac{\boldsymbol{\delta}}{\sqrt{p}},\boldsymbol{\epsilon}\right)\boldsymbol{\gamma}\boldsymbol{x} &= \frac{1}{n}\boldsymbol{x}^\tau\mathbf{E}^\tau\mathbf{E}\boldsymbol{x}=\frac{1}{n}\boldsymbol{x}^\tau\mathbf{C}^\tau\mathbf{C}\boldsymbol{x}\\ &= \frac{1}{n}\boldsymbol{x}^\tau\boldsymbol{\gamma}^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)\boldsymbol{\gamma}\boldsymbol{x},\end{aligned}$$

and then we can see that the quantity

$$\begin{aligned}&\frac{1}{n}\boldsymbol{x}^\tau\boldsymbol{\gamma}^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)\boldsymbol{\gamma}\boldsymbol{x}+\frac{\rho}{p}\boldsymbol{x}^\tau\mathbf{B}_m\boldsymbol{x}\\ =\ &\frac{1}{n}\boldsymbol{x}^\tau\boldsymbol{\gamma}^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)\boldsymbol{\gamma}\boldsymbol{x}+\boldsymbol{x}^\tau\boldsymbol{\gamma}^\tau\left(\rho\mathbf{B}_m\right)\boldsymbol{\gamma}\boldsymbol{x}\end{aligned}$$

is minimized for $\boldsymbol{x}$ eigenvector of the matrix

$$\begin{aligned}&\frac{1}{n}\boldsymbol{\gamma}^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{\sqrt{p}},\mathbf{Y}\right)\boldsymbol{\gamma}+\boldsymbol{\gamma}^\tau\left(\rho\mathbf{B}_m\right)\boldsymbol{\gamma}\\ =\ &\frac{1}{n}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)+\frac{\rho}{p}\mathbf{B}_m,\end{aligned}$$

corresponding to the smallest non-zero eigenvalue, which is denoted σ_k^2. Using the definition of this eigenvalue, we deduce that

$$\left(\frac{1}{n}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)^{\tau}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)+\frac{\rho}{p}\mathbf{B}_m\right)\widehat{\boldsymbol{x}}=\sigma_k^2\widehat{\boldsymbol{x}}.$$

This gives, keeping the p first rows,

$$\widehat{\boldsymbol{\alpha}}=\frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^{\tau}\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\sigma_k^2\mathbf{I}_p\right)^{-1}\mathbf{W}^{\tau}\mathbf{Y},$$

and the proof of the proposition 1 is now complete.

6.3 Proof of proposition 2

We can write

$$\frac{1}{n}\mathbf{W}^{\tau}\mathbf{W}=\frac{1}{n}\mathbf{X}^{\tau}\mathbf{X}+\left(\frac{1}{n}\sum_{i=1}^{n}M_{irs}\right)_{r,s=1,\ldots,p}$$

where $M_{irs}=X_i(t_r)\delta_{is}+\delta_{ir}X_i(t_s)+\delta_{ir}\delta_{is}$. Let us now study this random variable M_{irs}. First, we have

$$\begin{aligned}\mathbb{E}\left(M_{irs}\right) &= X_i(t_r)\mathbb{E}\left(\delta_{is}\right)+\mathbb{E}\left(\delta_{ir}\right)X_i(t_s)+\mathbb{E}\left(\delta_{ir}\delta_{is}\right)\\ &= \begin{cases}\sigma_\delta^2 \text{ if } r=s,\\ 0 \text{ otherwise.}\end{cases}\end{aligned}$$

On the other hand, we have with hypotheses $(H.2)$ and $(H.3)$

$$\begin{aligned}&\sup_{r,s=1,\ldots,p}\mathbb{E}\left(M_{irs}^2\right)\\ =&\sup_{r,s=1,\ldots,p}\Big\{X_i(t_r)^2\mathbb{E}\left(\delta_{is}^2\right)+\mathbb{E}\left(\delta_{ir}^2\right)X_i(t_s)^2\\ &+\mathbb{E}\left(\delta_{ir}^2\delta_{is}^2\right)+2X_i(t_r)X_i(t_s)\mathbb{E}\left(\delta_{ir}\delta_{is}\right)\Big\}\\ =&\ O(\sigma_\delta^2),\end{aligned}$$

hence we deduce

$$\sup_{r,s=1,\ldots,p}\left(\frac{1}{n}\sum_{i=1}^{n}M_{irs}\right)=\begin{cases}\sigma_\delta^2+O_P\left(\frac{\sigma_\delta}{n^{1/2}}\right) \text{ if } r=s,\\ O_P\left(\frac{\sigma_\delta}{n^{1/2}}\right) \text{ otherwise.}\end{cases}$$

We can now conclude the proof of proposition 2. If we define $\mathbf{R}$ such that

$$\left(\frac{1}{np^2}\sum_{i=1}^{n}M_{irs}\right)_{r,s=1,\ldots,p}=\frac{\sigma_\delta^2}{p^2}\mathbf{I}_p+\mathbf{R},$$

then $\sup_{r,s=1,\ldots,p} |R_{rs}| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^2}\right)$ and we get (see theorem 1.19 in [5])

$$\|\mathbf{R}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p}\right).$$

6.4 Proof of theorem 2

Let us consider the random variable $Z_{ij} = \delta_{ij}Y_i$. Using the independence between Y_i and δ_{ij}, we get $\mathbb{E}\left(Z_{ij}\right) = 0$ and

$$\sup_{j=1,\ldots,p} \mathbb{E}\left(Z_{ij}^2\right) = \sup_{j=1,\ldots,p} \mathbb{E}\left(\delta_{ij}^2\right)\mathbb{E}\left(Y_i^2\right) = O(\sigma_\delta^2),$$

from what we deduce that

$$\sup_{i=1,\ldots,n}\sup_{j=1,\ldots,p} |Z_{ij}| = O_P\left(\sigma_\delta\right).$$

Now we see that

$$\sup_{j=1,\ldots,p}\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ij}\right| = O_P\left(\frac{\sigma_\delta}{n^{1/2}}\right),$$

and then

$$\|\mathbf{V}\|^2 := \left\|\frac{1}{np}\mathbf{W}^\tau\mathbf{Y} - \frac{1}{np}\mathbf{X}^\tau\mathbf{Y}\right\|^2 = \sum_{j=1}^{p}\left(\frac{1}{np}\sum_{i=1}^{n} Z_{ij}\right)^2 = O_P\left(\frac{\sigma_\delta^2}{np}\right). \quad (31)$$

Noticing now the convergence result given in [14] of the estimator $\widehat{\sigma}_\delta^2$ of σ_δ^2, defined by (20), we have

$$\widehat{\sigma}_\delta^2 = \sigma_\delta^2 + O_P\left(\frac{1}{n^{1/2}p}\right). \quad (32)$$

Then, using this and the result (19) of the proposition 2, we can write

$$\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1} = \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m + \mathbf{R} - \frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}.$$

Now, let $\mathbf{T}$ be the $p \times p$ matrix defined by $\mathbf{T} = \mathbf{R} - \frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2}\mathbf{I}_p$. Using the result (32) and the fact that the norm of $\mathbf{I}_p$ is 1, we deduce

$$\left\|\frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2}\mathbf{I}_p\right\| = O_P\left(\frac{1}{n^{1/2}p^3}\right).$$

If we recall the order of $\|\mathbf{R}\|$ given in proposition 2, we finally obtain

$$\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}=\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m+\mathbf{T}\right)^{-1}, \qquad (33)$$

with

$$\|\mathbf{T}\|=O_P\left(\frac{\sigma_\delta}{n^{1/2}p}\right). \qquad (34)$$

Now, determining the norm of the matrix $\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}$, we get

$$\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right\|=O_P\left(\frac{1}{\rho}\right). \qquad (35)$$

Using the first inequality given in [9], we can write (with C strictly positive constant)

$$\begin{aligned}&\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m+\mathbf{T}\right)^{-1}-\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right\|\\ \leq\ & C\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right\|^2\|\mathbf{T}\|.\end{aligned}$$

Then, using relations (34) and (35), we obtain

$$\begin{aligned}&\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m+\mathbf{T}\right)^{-1}-\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right\|\\ =\ & O_P\left(\frac{\sigma_\delta}{n^{1/2}p\rho^2}\right). \qquad (36)\end{aligned}$$

Finally, if we set

$$\mathbf{S}=\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m+\mathbf{T}\right)^{-1}-\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1},$$

then we have (with relations (33) and (36))

$$\begin{gathered}\left(\tfrac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\tfrac{\rho}{p}\mathbf{A}_m-\tfrac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}=\left(\tfrac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\tfrac{\rho}{p}\mathbf{A}_m\right)^{-1}+\mathbf{S},\\ \text{with }\ \|\mathbf{S}\|=O_P\left(\tfrac{\sigma_\delta}{n^{1/2}p\rho^2}\right).\end{gathered} \qquad (37)$$

Let us now write

$$\begin{aligned}
&\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| \\
= &\left\| \left[\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m \right)^{-1} + \mathbf{S} \right] \left[\frac{1}{np}\mathbf{X}^\tau\mathbf{Y} + \mathbf{V} \right] \right. \\
&\left. \qquad - \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m \right)^{-1} \mathbf{X}^\tau\mathbf{Y} \right\| \\
\leq &\left\| \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m \right)^{-1} \right\| \|\mathbf{V}\| + \|\mathbf{S}\| \left\| \frac{1}{np}\mathbf{X}^\tau\mathbf{Y} \right\| + \|\mathbf{S}\| \, \|\mathbf{V}\| . \qquad (38)
\end{aligned}$$

The last thing to compute here is $\left\| \frac{1}{np}\mathbf{X}^\tau\mathbf{Y} \right\|$. In the same way as we have done to obtain (31), we write

$$\left\| \frac{1}{np}\mathbf{X}^\tau\mathbf{Y} \right\|^2 = \sum_{j=1}^{p} \left(\frac{1}{np} \sum_{i=1}^{n} X_i(t_j) Y_i \right)^2 .$$

Then, using the fact that $\sup_{j=1,\ldots,p} \mathbb{E}\left(X_i(t_j)Y_i\right) = O\left(1\right)$ and the fact that $\sup_{j=1,\ldots,p} \mathbb{E}\left(X_i(t_j)^2 Y_i^2\right) = O\left(1\right)$, we obtain

$$\sup_{i=1,\ldots,n} \sup_{j=1,\ldots,p} |X_i(t_j)Y_i| = O_P\left(1\right),$$

and then

$$\left\| \frac{1}{np}\mathbf{X}^\tau\mathbf{Y} \right\|^2 = O_P\left(\frac{1}{np} \right). \qquad (39)$$

Now, coming back to the inequality (38), using the results (31) and (37) as well as relations (35) and (39), we get

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^{1/2}\rho} \right) + O_P\left(\frac{\sigma_\delta}{np^{3/2}\rho^2} \right).$$

Since $\lim_{n\to+\infty} \frac{1}{n^{1/2}p\rho} = 0$, we get $\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\| = O_P\left(\frac{\sigma_\delta}{n^{1/2}p^{1/2}\rho} \right)$ and the proof of theorem 2 is complete.

6.5 Proof of proposition 3

Let us expand the $MISE(\lambda)$,

$$MISE(\lambda) = \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}(\lambda)^\tau \widehat{\boldsymbol{\alpha}}(\lambda)\right) - 2\boldsymbol{\alpha}^\tau \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}(\lambda)\right) + \boldsymbol{\alpha}^\tau\boldsymbol{\alpha},$$

to deduce, using the matricial expression of $\widehat{\boldsymbol{\alpha}}(\lambda)$

$$\frac{\partial}{\partial\lambda}MISE(\lambda)_{|\lambda=0} = 2\mathbb{E}\left[\frac{1}{n^2p^2}\mathbf{Y}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-3}\mathbf{W}^\tau\mathbf{Y} - \frac{1}{np}\boldsymbol{\alpha}^\tau\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{W}^\tau\mathbf{Y}\right]. \quad (40)$$

Now, using the fact that $\mathbf{Y} = \frac{1}{p}\mathbf{W}\boldsymbol{\alpha} - \frac{1}{p}\boldsymbol{\delta}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$

$$\begin{aligned}
&\frac{1}{np}\mathbf{Y}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1} - \boldsymbol{\alpha}^\tau \\
=\; &\frac{1}{np}\mathbf{Y}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1} - \frac{1}{np^2}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1} \\
=\; &\frac{1}{np}\left[\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right. \\
&\left.-\frac{1}{p}\boldsymbol{\alpha}^\tau\boldsymbol{\delta}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1} + \boldsymbol{\epsilon}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right] \\
&-\frac{1}{np}\left[\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\right].
\end{aligned}$$

Considering the quantity $\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1} - \left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}$, if we make an approximation at first order, we get

$$\begin{aligned}
&\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1} - \left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1} \\
\approx\; &-\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\left(\frac{\rho}{p}\mathbf{A}_m\right)\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1},
\end{aligned}$$

what gives us, coming back to relation (40)

$$
\begin{aligned}
&\frac{\partial}{\partial \lambda} MISE(\lambda)_{|\lambda=0} \\
\approx\ & 2\mathbb{E}\left[-\frac{1}{n^2p^3}\boldsymbol{\alpha}^\tau \mathbf{W}^\tau \mathbf{W}\left(\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\frac{\rho}{p}\mathbf{A}_m\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\right)\right. \\
&\left.\qquad\qquad \times\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{W}^\tau\mathbf{Y}\right] \\
&+2\mathbb{E}\left[-\frac{1}{n^2p^3}\boldsymbol{\alpha}^\tau\boldsymbol{\delta}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-3}\mathbf{W}^\tau\mathbf{Y}\right] \\
&+2\mathbb{E}\left[\frac{1}{n^2p^2}\boldsymbol{\epsilon}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-3}\mathbf{W}^\tau\mathbf{Y}\right]. \qquad (41)
\end{aligned}
$$

Using the fact that δ and ϵ are both independent from W and Y, the last two terms in relation (41) are zero, and we obtain finally

$$
\begin{aligned}
&\frac{\partial}{\partial \lambda} MISE(\lambda)_{|\lambda=0} \\
\approx\ & 2\mathbb{E}\left[-\frac{1}{n^2p^4}\boldsymbol{\alpha}^\tau \mathbf{W}^\tau \mathbf{W}\left(\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\frac{\rho}{p}\mathbf{A}_m\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\right)\right. \\
&\left.\qquad\qquad \times\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{W}^\tau\mathbf{W}\boldsymbol{\alpha}\right].
\end{aligned}
$$

This last quantity is negative, what achieves the proof of proposition 3.

References

[1] Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics, **149**, Springer-Verlag.

[2] Cardot, H. (2000). Nonparametric Estimation of Smoothed Principal Components Analysis of Sampled Noisy Functions. *Nonparametric Statistics*, **12**, 503-538.

[3] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional Linear Model. *Statistics and Probability Letters*, **45**, 11-22.

[4] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline Estimators for the Functional Linear Model. *Statistica Sinica*, **13**, 571-591.

[5] Chatelin, F. (1983). *Spectral Approximation of Linear Operators.* Academic Press, New-York.

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

[7] Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear Functional Regression: the case of a Fixed Design and Functional Response. *Canadian Journal of Statistics*, **30**, 285-300.

[8] de Boor, C. (1978). *A Practical Guide to Splines.* Springer Verlag.

[9] Demmel, J. (1992). The Componentwise Distance to the Nearest Singular Matrix. *SIAM, Journal of Matrix Analysis and Applications*, **13**, 10-19.

[10] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker.

[11] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations.* Springer-Verlag, London.

[12] Fuller, W.A. (1987). *Measurement Error Models.* Wiley, New-York.

[13] Fuller, W.A. (1997). Estimated True Values for Errors-in-Variables Models. In *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, S. Van Huffel editor, SIAM, Philadelphia, 51-57.

[14] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual Variance and Residual Pattern in Nonlinear regression. *Biometrika*, **3**, 625-633.

[15] Gleser, L.J. (1981). Estimation in a Multivariate "Errors-in-Variables" Regression Model: Large Sample Results. *The Annals of Statistics*, **9**, 24-44.

[16] Golub, G.H., Hansen, P.C and O'Leary, D.P. (1999). Tikhonov Regularization and Total Least Squares. *SIAM, Journal of Matrix Analysis and Applications*, **21**, 185-194.

[17] Golub, G.H. and Van Loan, C.F. (1980). An Analysis of the Total Least Squares Problem. *SIAM, Journal of Numerical Analysis*, **17**, 883-893.

[18] Golub, G.H. and Van Loan, C.F. (1996). *Matrix Computations*. Johns Hopkins University Press, Baltimore.

[19] Härdle, W. (1991). *Smoothing Techniques with Implementation in S*. Springer, New-York.

[20] Marx, B.D. and Eilers, P.H. (1999). Generalized Linear Regression on Sampled Signals and Curves: A *P*-Spline Approach. *Technometrics*, **41**, 1-13.

[21] Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for Functional Data Analysis. *Journal of the Royal Statistical Society, Series B*, **53**, 539-572.

[22] Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag.

[23] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis*. Springer-Verlag.

[24] Sarda, P. and Vieu, P. (2000). Kernel Regression. In *Smoothing and Regression: Approches, Computation and Application*, M.G. Schimek editor, Wiley Series in Probability and Statistics, 43-70.

[25] Sima, D.M. and Van Huffel, S. (2004). Appropriate Cross Validation for Regularized Error-in-Variables Linear Models. *Compstat 2004 Proceedings*, 1815-1822.

[26] Utreras, F. (1983). Natural Spline Functions, their Associated Eigenvalue Problem. *Numerische Mathematik*, **42**, 107-117.

[27] Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, Philadelphia.

[28] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.