Total Least Squares for Functional Data

Christophe Crambes

Université Paul Sabatier
Laboratoire de Statistique et Probabilités
UMR C5583
118 route de Narbonne
31062 Toulouse Cedex, France
(e-mail:crambes@cict.fr)

Abstract. We are interested in the functional linear regression when the covariates are subject to errors, for instance measurement errors. The aim of this paper is to propose a procedure giving a spline estimator of the functional coefficient of the model with noisy covariates. The functional coefficient is the solution of an ill-conditioned minimization problem, so a penalization approach is used. Indeed, we present an extension of the penalized total least squares algorithm to the case where the covariates are curves. Then, this estimation procedure is evaluated by the way of simulations.
Keywords: functional linear regression, errors-in-variables, total least squares, penalization, spline functions.

1 Introduction

In many fields of applications, it is frequent to deal with the problem of the explanation of a random variable Y (response), usually scalar, using information from a random variable X (covariate), belonging to some Hilbert space E. Then, a way to formulate this problem is to consider the linear regression of Y on X that, in case of existence and unicity, allows us to write

$$Y = \mu + \langle \alpha, X \rangle + \epsilon, \tag{1}$$

where $\langle ., . \rangle$ stands for the inner product of the Hilbert space E and ϵ is a real random variable satisfying $\mathbb{E}(\epsilon) = 0$ and $\mathbb{E}(\epsilon X) = 0$. Implicitly, in (1), the variable X is supposed to be observed without error, and all errors are into the variable Y by the way of ϵ. However, in practice, this assumption seems to be quite unrealistic, for example because of instrument measurement errors. That is why it should be natural to consider that the variable X is not directly observed, but we observe instead a variable W such that

$$W = X + \delta. \tag{2}$$

In the case where E is $\mathbb{R}$ or $\mathbb{R}^p$, that is to say when X is an univariate or a multivariate random variable, this problem of *errors-in-variables* model has already been studied. Some theoretical approaches have been proposed, using the maximum likelihood method (see [Fuller, 1987]) or deconvolution techniques (see [Carroll *et al.*, 1995]). A practical point of view is given under the name of *Total Least Squares* (TLS) in [Van Huffel and Vandewalle, 1991]. However, in many fields of applications (chemistry, climatology, teledetection, linguistics, ...), the data do not belong to the frame of univariate or multivariate variables. Indeed, the data can come from the observation of continuous phenomenons (that is to say continuous functions of time, space, ...), then they are comparable to curves. These data, called *functional data* in the literature, are the object of many studies (see [Ramsay and Silverman, 1997] and [Ramsay and Silverman, 2002] for a functional data analy! sis overview). Our goal is to study the problem of *errors-in-variables* model in the framework where X is a functional random variable, in other words when E is an infinite dimension space.

In the following, we consider n couples of random variables $(X_i, Y_i)_{i=1,\dots,n}$ independant and identically distributed, with the same distribution as (X, Y), where X is a random variable taking values in some functional space E and Y belongs to $\mathbb{R}$. For sake of simplicity, we consider that E is the space $L^2(I)$ of the functions of square integrable defined on an interval I of $\mathbb{R}$. We still denote by $\langle .,. \rangle$ the usual inner product of $L^2(I)$ and by $\|.\|$ the associated norm. We rewrite (1) taking the point of view of the *functional linear regression* introduced in [Ramsay and Dalzell, 1993], hence we assume that

$$Y = \mu + \int_I \alpha(t) X(t)\, dt + \epsilon, \tag{3}$$

where $\mu \in \mathbb{R}$ and $\alpha \in L^2(I)$ are the unknown parameters of the model and ϵ is a real random variable such that $\mathbb{E}(\epsilon) = 0$ and $\mathbb{E}(\epsilon X) = 0$. We assume conditions for existence and unicity of α (see [Cardot *et al.*, 2003]). Let us remark that, if we denote by Γ_X the covariance operator of X (defined by $\Gamma_X u = \mathbb{E}\left(\langle X - \mathbb{E}(X), u\rangle (X - \mathbb{E}(X))\right)$ for all $u \in L^2(I)$) and by Δ_{XY} the cross covariance operator of X and Y (defined by $\Delta_{XY} u = \mathbb{E}\left(\langle X - \mathbb{E}(X), u\rangle Y\right)$ for all $u \in L^2(I)$), then we easily see that $\langle \Gamma_X \alpha, u\rangle = \Delta_{XY} u$ for all function $u \in L^2(I)$. One of the properties of Γ_X is that it is a nuclear operator (see [Loève, 1963] for details). So Γ_X^{-1} is not bounded and estimation of ! α is an *ill-conditioned* problem. A possibility to deal with this problem is to introduce a penalization approach (this is done in [Cardot *et al.*, 2003]), and to find μ and α as solutions of the minimization problem

$$\min_{\mu \in \mathbb{R}, \alpha \in L^2(I)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \mu - \langle \alpha, X_i \rangle\right)^2 + \rho \left\| \alpha^{(m)} \right\|^2 \right\}, \tag{4}$$

where $\alpha^{(m)}$ stands for the derivative of order m of the function α and ρ is a smoothing parameter allowing to control the regularity of the estimator of the function α.

Now coming back to our *errors-in-variables* setting, we suppose that the curve X is not directly available. In practice, the curves $X_1, \ldots, X_n$ are observed in p discretization points $t_1, \ldots, t_p \in I$ such that $t_1 \leq \ldots \leq t_p$. So, the data are

$$W(t_j) = X(t_j) + \delta(t_j), \quad j = 1, \ldots, p, \tag{5}$$

where $(\delta(t_j))_{j=1,\ldots,p}$ is a sequence of real random variables independent and identically distributed, centered and with variance σ_δ^2. We also assume that $\delta(t_j)$ and ϵ are independant for all $j = 1, \ldots, p$. These variables represent the error made on X at each measure point. The random variables W and δ give us the corresponding samples $(W_i)_{i=1,\ldots,n}$ and $(\delta_i)_{i=1,\ldots,n}$. The aim of this paper is to build an estimator of μ and α. In section 2, we generalize the TLS algorithm to our functional framework. In section 3, this estimator is evaluated by the way of simulations. Finally in section 4, we make some concluding remarks.

2 Functional Total Least Squares

The aim of this section is to adapt the *Total Least Squares* algorithm introduced in [Van Huffel, 2004] when the covariate X is of functional nature.

2.1 Total Least Squares in the multivariate case

When X is a multivariate random variable, the linear regression is written

$$Y = \mu + {}^t\mathbf{X}\boldsymbol{\alpha} + \epsilon, \tag{6}$$

where $\mathbf{X} = {}^t(X_1, \ldots, X_p)$ belongs to $\mathbb{R}^p$. We have to estimate $\mu \in \mathbb{R}$ and $\alpha \in \mathbb{R}^p$, assuming we observe Y_i and $\mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\delta}_i$ for $i = 1, \ldots, n$. We denote by $\mathbf{Y}$ the vector ${}^t(Y_1, \ldots, Y_n)$, $\boldsymbol{\epsilon}$ the vector ${}^t(\epsilon_1, \ldots, \epsilon_n)$, $\mathbf{X}$ and $\mathbf{W}$ the matrices of respective elements X_{ij} and W_{ij}. Under an hypothesis of normality for the errors (that is to say if $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\delta(t_j) \sim \mathcal{N}(0, \sigma_\delta^2)$ for all $j = 1, \ldots, p$), the likelihood function is proportional to

$$\exp\left\{ -\sum_{i=1}^{n} \left[\frac{1}{\sigma_\epsilon^2} \left(Y_i - \mu - {}^t\mathbf{X}_i\boldsymbol{\alpha}\right)^2 + \frac{1}{\sigma_\delta^2} {}^t(\mathbf{X}_i - \mathbf{W}_i)(\mathbf{X}_i - \mathbf{W}_i) \right] \right\}. \tag{7}$$

Without any more condition, the model (6) with $\mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\delta}_i$ is not identifiable and another condition needs to be imposed (see [Van Huffel, 2004]). In the following, we choose to assume that the ratio of the variances $\sigma_\epsilon^2/\sigma_\delta^2$ is known. Indeed, we can suppose that this ratio is equal to 1 (if the ratio is $\eta = \sigma_\epsilon^2/\sigma_\delta^2$, we consider the scaled variable $\widetilde{\mathbf{X}} = \sqrt{\eta}\mathbf{X}$ and then $\boldsymbol{\alpha} = \sqrt{\eta}\widetilde{\boldsymbol{\alpha}}$).

Then, the maximization of (7) comes back to the resolution of the minimization problem

$$\min_{\mu\in\mathbb{R},\boldsymbol{\alpha}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[(Y_i-\mu-\mathbf{X}_i\boldsymbol{\alpha})^2+{}^t(\mathbf{X}_i-\mathbf{W}_i)(\mathbf{X}_i-\mathbf{W}_i)\right]\right\}. \quad (8)$$

The TLS algorithm given in [Van Huffel, 2004] follows the two steps below:

- step 1: we make the singular value decomposition (SVD) of the matrix $[\,\mathbf{1}\mid\mathbf{W}\mid\mathbf{Y}\,]$, that is to say $[\,\mathbf{1}\mid\mathbf{W}\mid\mathbf{Y}\,]=\mathbf{U}\boldsymbol{\Sigma}\,{}^t\mathbf{V}$ with ${}^t\mathbf{U}\mathbf{U}=\mathbf{I}_n$ and ${}^t\mathbf{V}\mathbf{V}=\mathbf{I}_{p+2}$, where $\mathbf{I}_n$ and $\mathbf{I}_{p+2}$ are respectively the $n\times n$ and $(p+2)\times(p+2)$ identity matrices,
- step 2: if the elements of the matrix $\mathbf{V}$ are denoted by v_{jl}, then the TLS estimator of μ and $\boldsymbol{\alpha}$ is given by

$$\begin{pmatrix}\widehat{\mu}_{TLS}\\ \widehat{\boldsymbol{\alpha}}_{TLS}\end{pmatrix}=-\frac{1}{v_{p+2,p+2}}\,{}^t(v_{1,p+2},\dots v_{p+1,p+2}). \quad (9)$$

However, the problem of this algorithm is that it can not be used directly when the minimization problem (8) is *ill-conditioned* and needs a regularization. The minimization problem we consider is then (see [Golub *et al.*, 1999])

$$\min_{\mu\in\mathbb{R},\boldsymbol{\alpha}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[(Y_i-\mu-\mathbf{X}_i\boldsymbol{\alpha})^2+{}^t(\mathbf{X}_i-\mathbf{W}_i)(\mathbf{X}_i-\mathbf{W}_i)\right]+\rho\,{}^t\boldsymbol{\alpha}{}^t\mathbf{L}\mathbf{L}\boldsymbol{\alpha}\right\}, \quad (10)$$

where $\mathbf{L}$ is a $p\times p$ matrix. Using the properties of the SVD, it can be shown (see [Golub and Van Loan, 1996]) that

$$\begin{pmatrix}\widehat{\mu}_{TLS}\\ \widehat{\boldsymbol{\alpha}}_{TLS}\end{pmatrix}=\left({}^t[\,\mathbf{1}\mid\mathbf{W}\,][\,\mathbf{1}\mid\mathbf{W}\,]-\sigma_{p+2}^2\mathbf{I}_{p+1}\right)^{-1}{}^t[\,\mathbf{1}\mid\mathbf{W}\,]\mathbf{Y}, \quad (11)$$

where σ_{p+2} is the smallest singular value of the matrix $[\,\mathbf{1}\mid\mathbf{W}\mid\mathbf{Y}\,]$ and $\mathbf{I}_{p+1}$ is the $(p+1)\times(p+1)$ identity matrix. From this expression, the TLS solution to the minimization problem (10) is given by

$$\begin{pmatrix}\widehat{\mu}_{TLS}\\ \widehat{\boldsymbol{\alpha}}_{TLS}\end{pmatrix}=\left({}^t[\,\mathbf{1}\mid\mathbf{W}\,][\,\mathbf{1}\mid\mathbf{W}\,]-\lambda\mathbf{I}_{p+1}+\rho\,{}^t\mathbf{M}\mathbf{M}\right)^{-1}{}^t[\,\mathbf{1}\mid\mathbf{W}\,]\mathbf{Y}, \quad (12)$$

where $\mathbf{M}$ is the $(p+1)\times(p+1)$ matrix defined by $\mathbf{M}=\begin{pmatrix}0 & 0\dots 0\\ 0 & \\ \vdots & \mathbf{L}\\ 0 & \end{pmatrix}.$

2.2 Total Least Squares in the functional case

All that has been done in the previous paragraph can be adapted to the case where X is of functional type. The minimisation problem considered is a combination of (4) and (10), that we write

$$\min_{\mu\in\mathbb{R},\alpha\in L^2(I),X_i\in L^2(I)} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left[(Y_i-\mu-\langle\alpha,X_i\rangle)^2+\|X_i-W_i\|^2\right] + \rho\left\|\alpha^{(m)}\right\|^2\right\}. \quad (13)$$

We choose to build a spline estimator of α. We have to fix a degree $q\in\mathbb{N}$ and a number $k\in\mathbb{N}^\star$ of knots (taken equispaced) giving a subdivision of the interval I (see [de Boor, 1978] for details on spline functions). These spline functions have well-known properties, in particular, this space of spline functions is a vectorial space of dimension $k+q$. A usual basis is the set of the so-called B-spline functions, that we denote by $\mathbf{B}_{k,q} = {}^t(B_1\dots B_{k+q})$. Then, we estimate α as a linear combination of the B-spline functions, that is to say we have to find a vector $\widehat{\boldsymbol{\theta}} = {}^t(\widehat{\theta}_1\dots\widehat{\theta}_{k+q})\in\mathbb{R}^{k+q}$ such that $\widehat{\alpha} = {}^t\mathbf{B}_{k,q}\widehat{\boldsymbol{\theta}}$ with $\widehat{\mu}$ and $\widehat{\boldsymbol{\theta}}$ solutions of the minimization problem

$$\min_{\mu\in\mathbb{R},\boldsymbol{\theta}\in\mathbb{R}^{k+q},X_i\in L^2(I)} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left[\left(Y_i-\mu-\langle\,{}^t\mathbf{B}_{k,q}\widehat{\boldsymbol{\theta}},X_i\rangle\right)^2+\|X_i-W_i\|^2\right] + \rho\left\|\left({}^t\mathbf{B}_{k,q}\widehat{\boldsymbol{\theta}}\right)^{(m)}\right\|^2\right\}. \quad (14)$$

Using the work in [Cardot *et al.*, 2003] for the spline estimator of the functional coefficient and what has been done in the multivariate case (see equation (12)), it is possible to find an explicit solution to the minimization problem (14), given by

$$\begin{pmatrix}\widehat{\mu}_{FTLS}\\ \widehat{\boldsymbol{\theta}}_{FTLS}\end{pmatrix} = \frac{1}{n}(\frac{1}{n}\,{}^t\mathbf{D}\mathbf{D}-\lambda\mathbf{I}_{k+q+1}+\rho\mathbf{K})^{-1}\,{}^t\mathbf{D}\mathbf{Y}, \quad (15)$$

with

$$\mathbf{D} = \begin{pmatrix} 1 & \langle B_1,W_1\rangle & \dots & \langle B_{k+q},W_1\rangle \\ \vdots & \vdots & & \vdots \\ 1 & \langle B_1,W_n\rangle & \dots & \langle B_{k+q},W_n\rangle \end{pmatrix},$$

and

$$\mathbf{K} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & \langle B_1^{(m)}, B_1^{(m)} \rangle & \dots & \langle B_1^{(m)}, B_{k+q}^{(m)} \rangle \\ \vdots & \vdots & & \vdots \\ 0 & \langle B_{k+q}^{(m)}, B_1^{(m)} \rangle & \dots & \langle B_{k+q}^{(m)}, B_{k+q}^{(m)} \rangle \end{pmatrix}.$$

3 A simulation study

The aim of this simulation is to see the behaviour of this TLS estimator and to compare it with the spline estimator given in [Cardot *et al.*, 2003] by

$$\begin{pmatrix} \widehat{\mu}_{FLS} \\ \widehat{\boldsymbol{\theta}}_{FLS} \end{pmatrix} = \frac{1}{n} (\frac{1}{n}\, {}^t\mathbf{D}\mathbf{D} + \rho \mathbf{K})^{-1}\, {}^t\mathbf{D}\mathbf{Y}. \tag{16}$$

We choose to take

- $n = 200$: the initial sample will be splitted into a learning sample of length $n_l = 100$ (to estimate μ and α) and a test sample of length $n_t = 100$ (to see the quality of prediction),
- $p = 50$ discretization points on $I = [0, 1]$,
- X is either a standard brownian motion or an Ornstein-Uhlenbeck process on I,
- $\mu = 2$,
- $\alpha(t) = 10 \sin(2\pi t)$,
- $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.1$,
- $\delta(t_j) \sim \mathcal{N}(0, \sigma_\delta^2)$ with either $\sigma_\delta = 0.05$, $\sigma_\delta = 0.1$, or $\sigma_\delta = 0.2$.

Concerning the choice of the different parameters of the model, we have taken $k = 8$, $q = 3$ and $m = 2$. Moreover, in the functional least squares estimation, ρ is fixed by generalized cross validation (see [Wahba, 1990]). For the total least squares estimation, we have made the estimation for different values of λ and ρ among the values $10^{-2}, 10^{-3}, \dots, 10^{-10}$, and we have kept the best values for these two parameters in terms of prediction.
We have given in table 1 the mean relative errors on 50 simulations for the different models tested when X is a standard brownian motion on I and in table 2 the same errors when X is an Ornstein-Uhlenbeck process on I. The estimation of the curve X_i, noted $\widehat{X}_i$, is given by

$$\widehat{X}_i = W_i + \frac{Y_i - \widehat{\mu} - \langle \widehat{\alpha}, W_i \rangle}{1 + \|\widehat{\alpha}\|^2}\, \widehat{\alpha}, \tag{17}$$

as the generalization of $\widehat{\mathbf{X}}_i$ in the multivariate case (see [Fuller, 1997]), obtained by differentiation of equation (13) with respect to X_i. An example of the estimation of α is plotted on figure 1 in the case where X is a standard brownian motion on I with the variance noise $\sigma_\delta = 0.1$. These results show that the corrected estimator constructed with the TLS approach improves the estimation of α compared to the uncorrected estimator defined by (16).

	$\frac{(\widehat{\mu}-\mu)^2}{\mu^2}$			$\frac{\|\widehat{\alpha}-\alpha\|^2}{\|\alpha\|^2}$			$\frac{1}{n}\sum_{i=1}^{n}\left(\langle\widehat{\alpha},\widehat{X}_i\rangle-\langle\alpha,X_i\rangle\right)^2$		
	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$
FLS	0.0002	0.0009	0.0017	0.21	0.41	0.78	0.007	0.010	0.018
FTLS	0.0002	0.0009	0.0016	0.12	0.27	0.56	0.006	0.008	0.015

Table 1. Errors on μ, α and prediction - case where X is a standard brownian motion on I.

	$\frac{(\widehat{\mu}-\mu)^2}{\mu^2}$			$\frac{\|\widehat{\alpha}-\alpha\|^2}{\|\alpha\|^2}$			$\frac{1}{n}\sum_{i=1}^{n}\left(\langle\widehat{\alpha},\widehat{X}_i\rangle-\langle\alpha,X_i\rangle\right)^2$		
	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.05$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$
FLS	0.0004	0.0007	0.0017	0.07	0.19	0.39	0.006	0.011	0.021
FTLS	0.0004	0.0007	0.0015	0.02	0.11	0.26	0.005	0.010	0.019

Table 2. Errors on μ, α and prediction - case where X is an Ornstein-Uhlenbeck process on I.

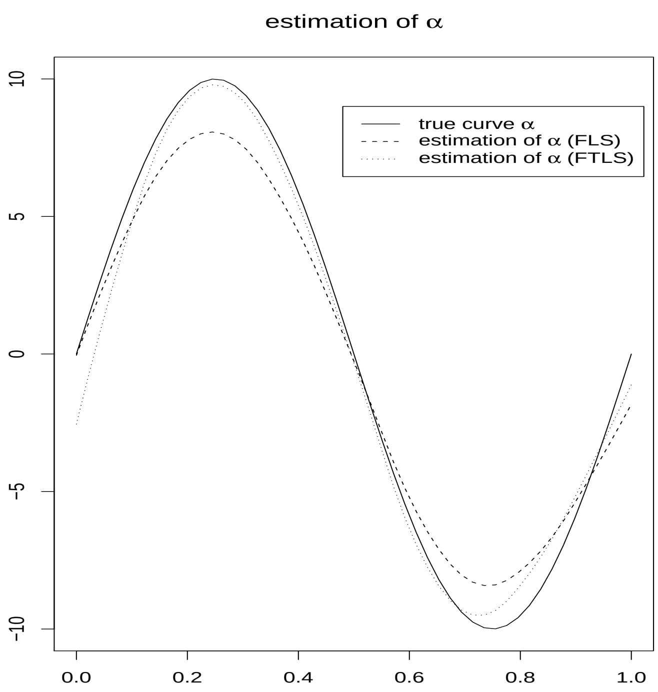

Fig. 1. Example of estimation of α (solid line) with functional least squares (dashed line) and functional total least squares (dotted line).

4 Conclusion and openings

This adaptation of the *Total Least Squares* method to the functional framework seems to give encouraging results on simulations. A theoretical work is needed to get the statistical properties of the estimator we have built. Moreover, it could also be interesting to compare this method to other ones. In particular, another idea to deal with noisy functional covariates (which is a work in progress) is to smooth the noisy curves (for instance by the way of a kernel method) and to estimate α by a procedure equivalent to a functional principal component regression used in the work of [Kneip and Utikal, 2001].

References

[Cardot *et al.*, 2003]H. Cardot, F. Ferraty, and P. Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591, 2003.

[Carroll *et al.*, 1995]R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models.* Chapman & Hall, London, 1995.

[de Boor, 1978]C. de Boor. *A Practical Guide to Splines.* Springer, New-York, 1978.

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

[Fuller, 1997]W.A. Fuller. Estimated true values for errors-in-variables models. In S. Van Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, pages 51–57, 1997.

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

[Golub *et al.*, 1999]G.H. Golub, P.C Hansen, and D.P. O'Leary. Tikhonov regularization and total least squares. *SIAM, J. Matrix Anal. Appl.*, pages 185–194, 1999.

[Kneip and Utikal, 2001]A. Kneip and K.J. Utikal. Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.*, pages 519–542, 2001.

[Loève, 1963]M. Loève. *Probability Theory,* 3^{rd} *edition.* D. Van Nostrand, New Jersey, 1963.

[Ramsay and Dalzell, 1993]J.O. Ramsay and C.J. Dalzell. Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B*, pages 539–572, 1993.

[Ramsay and Silverman, 1997]J.O. Ramsay and B.W. Silverman. *Functional Data Analysis.* Springer-Verlag, New-York, 1997.

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

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

[Van Huffel, 2004]S. Van Huffel. Total least squares and errors-in-variables modeling: Bridging the gap between statistics, computational mathematics and engineering. In J. Antoch, editor, *Compstat Proceedings in Computational Statistics*, pages 539–555, 2004.

[Wahba, 1990]G. Wahba. *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, Philadelphia, 1990.