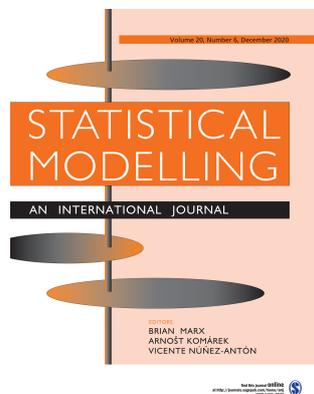


# Multi-class classification of biomechanical data: A functional LDA approach based on multi-class penalized functional PLS

- M. Carmen Aguilera-Morillo, Ana María Aguilera del Pino
- Multi-class classification of biomechanical data: A functional LDA approach based on multi-class penalized functional PLS
- *Statistical Modelling*, 2020; 20(6): 592-616
- DOI: [10.1177/1471082X17871157](https://doi.org/10.1177/1471082X17871157)



# Multi-class classification of biomechanical data: A functional LDA approach based on multi-class penalized functional PLS

M. Carmen Aguilera-Morillo<sup>1</sup> and Ana M. Aguilera<sup>2</sup>

<sup>1</sup>Department of Statistics, Escuela Politécnica Superior and UC3M-BS Santander Big Data Institute, Universidad Carlos III de Madrid, Madrid, Spain.

<sup>2</sup>Department of Statistics and O. R. and IEMath-GR, Facultad de Ciencias, Universidad de Granada, Granada, Spain.

**Abstract:** A functional linear discriminant analysis approach to classify a set of kinematic data (human movement curves of individuals performing different physical activities) is performed. Kinematic data, usually collected in linear acceleration or angular rotation format, can be identified with functions in a continuous domain (time, percentage of gait cycle, etc.). Since kinematic curves are measured in the same sample of individuals performing different activities, they are a clear example of functional data with repeated measures. On the other hand, the sample curves are observed with noise. Then, a roughness penalty might be necessary in order to provide a smooth estimation of the discriminant functions, which would make them more interpretable. Moreover, because of the infinite dimension of functional data, a reduction dimension technique should be considered. To solve these problems, we propose a multi-class approach for penalized functional partial least squares (FPLS) regression. Then linear discriminant analysis (LDA) will be performed on the estimated FPLS components. This methodology is motivated by two case studies. The first study considers the linear acceleration recorded every two seconds in 30 subjects, related to three different activities (walking, climbing stairs and down stairs). The second study works with the triaxial angular rotation, for each joint, in 51 children when they completed a cycle walking under three conditions (walking, carrying a backpack and pulling a trolley). A simulation study is also developed for comparing the performance of the proposed functional LDA with respect to the corresponding multivariate and non-penalized approaches.

**Key words:** functional data, linear discriminant analysis, multi-class classification, PLS regression, P-spline penalty

Received December 2018; revised April 2019; accepted July 2019

## 1 Introduction

Biomechanical gait data are commonly used to differentiate between several gait pathologies or different physical activities. Different statistical methodologies for

---

Address for correspondence: M. Carmen Aguilera-Morillo, Department of Statistics, Escuela Politécnica Superior, Universidad Carlos III de Madrid, Avda. de la Universidad 30, 28911, Madrid, Spain.

E-mail: mariacarmen.aguilera@uc3m.es

classification of gait data have been developed in the literature with this aim. Gait data are usually collected at different points in a continuous scale, such as time, so that they are generated from an underlying time-varying function. In the applications that motivate this research, acceleration curves are measured at real time every two seconds, and rotation curves are measured in terms of the percentage of gait cycle. In spite of the continuous nature of kinematics curves, in many of the biomechanical studies the statistical analysis is made from a sample of summary measures for each curve such as minima, maxima, area under the curve, angles at heel strike, range of motion or the timing of specific events (Sadeghi et al., 1997; Sadeghi, 2003; Schmidt and Docherty, 2010; Orantes-González et al., 2015).

Currently, data science tools are being developed to build accurate classification and prediction models for gait research by using all the available information. One of them is functional data analysis (FDA) that begins by reconstructing the true functional form of each sample curve from discrete sampling points that could be different for different subjects. This way all the available information about the sample curves is exploited to explain the movement in the statistical analysis. A detailed description of the most common FDA methodologies, and interesting FDA applications with R and Matlab, are described in the pioneer books by Ramsay and Silverman (1997), Ramsay and Silverman (2002) and Ramsay et al. (2009). A complete study of nonparametric FDA techniques can be seen in Ferraty and Vieu (2006). Statistical inference for functional data has been studied in Horvath and Kokoszka (2012). A more recent book is by Kokoszka and Reimherr (2017), where new methods for FDA are collected (functional time series, spatial functional data, between others).

Biomechanical data are usually obtained over a number of discrete time points (snapshots) and assumed to be generated by some underlying smooth function. Because of this, the steps typically used when FDA is applied to the analysis of biomechanical data are: basis expansion representation, curve registration (time normalization or landmark registration) and functional principal component analysis (FPCA) (dimension reduction and variability explanation in terms of uncorrelated scalar variables). The first step consists of obtaining smooth representations of the sample curves in terms of a basis of functions (Fourier, B-splines, wavelets, etc.). In the second step, registration is used to reduce phase variability between curves while preserving the individual curve's shape and amplitude. The effect of different registration methods on cyclical kinematic data is analysed in Crane et al. (2010). In the third step, FPCA is successfully applied to explain the ways of variation in biomechanics of sports injuries and coordination in race-walking, jumping and running (Ryan et al., 2006; Harrison et al., 2007; Donoghue et al., 2008; Donà et al., 2009). On the other hand, multivariate principal component analysis (PCA) on the raw discrete-time observations of the biomechanics curves has also been applied to provide biomechanics interpretation of the principal component scores and to discriminate between healthy and sick subjects (Deluzio et al., 1997; Daffertshofer et al., 2004; Milovanović and Popović, 2012).

This work is motivated by two different sets of biomechanical data: the human activity data and the gait data. The first dataset is based on the human activity

recognition collected by Anguita et al. (2013). The original experiments were carried out with a group of 30 volunteers within an age bracket of 19–48 years. Each person performed three different activities wearing a smartphone on the waist (walking, walking upstairs and walking downstairs). Using its embedded accelerometer and gyroscope, they captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz every two seconds. The aim is to classify the curves of acceleration according to the categorical variable given by the three mentioned physical activities. The gait dataset comes from a wide experimental study developed in the biomechanics laboratories of the Sport and Health Institute of the University of Granada (iMUDS). In order to collect the data, 26 reflective markers were placed on the skin of a total of 51 children between 8 and 11 years old. The data were recorded by a 3D motion capture system meanwhile each subject completed a cycle walking over the platform in three conditions (walking, carrying a backpack that weighs 20% of the subject's weight and pulling a trolley that weighs 20% of the subject's weight). For each subject, the 3-axial angular rotation were registered for each join (ankle, foot progress, hip, knee, pelvis, thorax), and all conditions, in 101 equidistant points of the complete gait cycle. Now, the aim is to classify the rotation curves according to the three different load types.

Both case studies considered in this article are clear examples of functional data with repeated measures because the functional variables of interest (angular rotation and linear acceleration) are measured repeatedly on the same sample individuals under different experimental conditions. Therefore, the aim of this article is to classify a set of functional data with repeated measures (kinematic curves) according to a categorical variable with more than two categories (measurement conditions). In this article, an extension of linear discriminant analysis (LDA) to the case of functional data with repeated measurements will be considered. Alternative classification approaches based on logit regression were developed and successfully applied in different areas as environment, medicine, marketing and so on (James, 2002; Escabias et al., 2004, 2005, 2007; Delaigle and Hall, 2012a; Aguilera-Morillo and Aguilera, 2013; Escabias et al., 2014). An extension to functional data of nonparametric Bayes classifiers based on simple density ratios was proposed in (Dai et al., 2017) for the case of binary classification.

LDA is a popular and consolidate methodology for classification and dimension reduction (Fisher, 1936). So, LDA provides a sequence of linear combinations of the original predictor variables (linear discriminants) that maximize the between-class variance relative to the within-class variance. An important problem is that LDA overfits the data when we have a large number of highly correlated predictor variables because the within-class covariance matrix is difficult to estimate. This is the case of FDA that works with a large number of discrete-time observations for each sample curve. A way to solve this problem is to introduce some kind of regularization in the estimation of the covariance matrix (Friedman, 1989; Frank and Friedman, 1989; Hastie et al., 1995). A general review of regularized techniques in discriminant analysis was developed in Mkhadri et al. (1997). A Bayesian approach for Fisher's discriminant analysis of stochastic process was introduced in Shin (2008). An alternative solution, for the case of irregularly sampled curves, is based on using

as predictors a spline basis expansion of the discrete-time observations and assuming Gaussian distribution of the vector of basis coefficients with common covariance matrix for all classes, by analogy with LDA (James and Hastie, 2001). This functional LDA approach was used to assessment of embryonic growth (Bottomley et al., 2009). Other way of solution consists of projecting the predictor vectors (multivariate analysis) or curves (FDA) onto a finite dimensional basis so that LDA, or any other discrimination procedures, can then be used taking the resulting basis coefficients as predictors. A common approach is based on the representation of the original predictors by a small set of orthogonal scalar variables given by PCA or partial least squares (PLS) regression. Both approaches are commonly used in chemometrics to discriminate between essential properties of substances from discrete measurements of the near-infrared (NIR) spectra. PCA was applied in Sato (1994) to classify NIR spectral data of vegetable oils. PLS regression was performed in Li et al. (2000) to discriminate edible fats and oils by Fourier transform NIR spectroscopy. Multivariate classification techniques such as LDA, quadratic discriminant analysis (QDA) and logit regression were applied in Indahl et al. (1999). In the functional data context, reduction dimension techniques and regularization can be jointly used to get good classification rates and accurate interpretation of the results. The functional LDA approach presented in this article is inspired in this research line.

Functional PCA and functional PLS regression were introduced as natural extensions of their multivariate counterparts to solve the problems of high dimension and multicollinearity associated with the scalar-on-function linear model (Deville, 1974; Dauxois et al., 1982; Ocaña et al., 1999, 2007; Preda and Saporta, 2005; Aguilera et al., 2016). Both methodologies were compared on different simulated datasets concluding that they have similar forecasting performance, but the estimated parameter function provided by functional PLS regression is more accurate and needs fewer components (Reiss and Ogden, 2007; Aguilera et al., 2010; Delaigle and Hall, 2012b,c). In addition, the performance of PCA for discrimination may not be optimal, because PCA only identifies gross variability, and then it is not capable to distinguish the between-groups and within-groups variability. In that sense, several authors proposed to use PLS regression for dimension reduction, as for example Barker and Rayens (2003), Pérez-Enciso and Tenenhaus (2003) in the multivariate context and Preda et al. (2007) in the functional context.

From a methodological point of view, the originality of our work lies, on the one hand, in proposing a penalized spline estimation of the PLS components to solve the problem of lack of smoothness, and on the other hand, in considering a multi-class approach for PLS regression based on the split-up variation to solve the problem of repeated measures. Then, classic LDA is carried out on the estimated PLS components. To the best of our knowledge, it is the first approach in the literature for functional linear discriminant analysis (FLDA) with repeated measures based on functional PLS. Then, in order to classify gait curves according to the kind of activity and take into account both, the lack of smoothness in the data and the correlation between repeated measurements, a new multi-class FLDA based on penalized functional PLS regression for repeated measures is proposed in Section 2. The classification results of the considered multi-class functional LDA-PLS approach on simulated data and both

biomechanical datasets are presented and discussed in Sections 3 and 4. Conclusions are finally included in Section 5.

## 2 Multi-class classification of functional data

Let us consider a categorical response variable  $Y = \{1, \dots, K\}$ , with  $K$  ( $K > 2$ ) categories and a functional predictor  $X = \{X(t) : t \in T\}$  (continuous and second order stochastic process whose sample paths belong to the Hilbert space of square-integrable functions  $L_2(T)$ ). Then, the aim is to predict  $Y$  from  $X(t)$ . That means to classify the sample curves according to the groups defined by the categories of  $Y$ .

The aim of FLDA is to find linear combinations

$$\Phi_i(X) = \alpha_i + \int_T X(t)\beta^i(t)dt, \quad \{i = 1, \dots, K - 1\}, \quad (2.1)$$

so that the between-class variance is maximized with respect to the total variance

$$\max_{\beta} \frac{V(E[\Phi(X)|Y])}{V(\Phi(X))}.$$

Due to the high dimension of  $X(t)$  (the number of observed variables is usually larger than the number of sample observations), the estimation of the discriminant functions is an ill-posed problem because the standard estimate for the within-class covariance matrix is singular so that it is impossible to apply the usual classification rule. This problem is solved in this article by using a dimension reduction approach based on functional PLS regression Preda et al. (2007). Then, the multi-class classification problem from repeated measures is solved by using the following two-step algorithm:

- Step 1. Dimension reduction by penalized and non-penalized multi-class functional PLS regression (for repeated measures) of the random vector  $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_{K-1})$  on the functional predictor  $X(t)$ , with  $\{\tilde{Y}_i \in \{0, 1\} : i = 1, \dots, K - 1\}$  being the dummy variables associated with the categorical response  $Y$ .
- Step 2. Functional LDA of  $Y$  on a reduced set of functional PLS components estimated in the above step.

### 2.1 Multi-class functional PLS regression

In order to reduce the dimension, PLS regression of the vector  $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_{K-1})$  on the functional predictor  $X(t)$ , with  $\{\tilde{Y}_i \in \{0, 1\} : i = 1, \dots, K - 1\}$  is performed. In that context, the non-penalized PLS components are generalized linear combinations,

of type  $t = \int_T X(t) w(t) dt$ , estimated by solving the following maximization problem

$$\{w, c\} = \operatorname{argmax} \operatorname{Cov}^2 \left( \int_T X(t) w(t) dt, \sum_{i=1}^{K-1} c_i \tilde{Y}_i \right) \quad (2.2)$$

$$\|w\| = 1$$

$$\|c\| = 1$$

with  $\|\cdot\|$  representing the usual norms in the spaces  $L_2[T]$  and  $\mathbb{R}^{K-1}$  where the component weights belong to, respectively, and  $c$  being the weight vector of  $\tilde{Y}$ .

From Proposition 1 in Preda and Saporta (2005), it can be seen that the solution to this maximization problem is reached for  $w$  and  $c$ , the eigenvectors associated with the largest eigenvalue of  $U_X$  and  $U_{\tilde{Y}}$ , respectively, defined by  $U_X = C_{X\tilde{Y}} \circ C_{\tilde{Y}X}$  and  $U_{\tilde{Y}} = C_{\tilde{Y}X} \circ C_{X\tilde{Y}}$ , with  $C_{\tilde{Y}X}$  and  $C_{X\tilde{Y}}$  being the cross-covariance operators.

Then, the first PLS component,  $t_1 = \int_T X(t) w_1(t) dt$ , is estimated in terms of the associated weight function,  $w_1(t)$ , which is solution to the following problem

$$C_{X\tilde{Y}} \circ C_{\tilde{Y}X}(w_1) = \lambda_{\max} w_1.$$

PLS regression is an iterative algorithm and the first PLS step is completed by ordinary linear regression of  $X_0(t) = X(t)$  and  $\tilde{Y}_0 = \tilde{Y}$  on  $t_1$ , obtaining the residuals,  $X_1(t)$  and  $\tilde{Y}_1$ , necessary to the following step

$$X_1(t) = X_0(t) - p_1(t)t_1 \quad t \in T$$

$$\tilde{Y}_{1,i} = \tilde{Y}_{0,i} - c_{1,i}t_1 \quad i = 1, \dots, K - 1,$$

where  $p_1(t) = (\mathbb{E}(X_0(t)t_1)/\mathbb{E}(t_1^2))$  and  $c_1 = (\mathbb{E}(Y_0t_1)/\mathbb{E}(t_1^2))$ .

In general, let us consider the  $b$ th PLS component,  $t_b = \int_T X_{b-1}(t) w_b(t) dt$ , which is estimated by the following problem

$$C_{X\tilde{Y}}^{b-1} \circ C_{\tilde{Y}X}^{b-1}(w_b) = \lambda_{\max} w_b,$$

where  $C_{X\tilde{Y}}^{b-1}$  and  $C_{\tilde{Y}X}^{b-1}$  are the cross covariance operators of  $X_{b-1}(t)$  and  $\tilde{Y}_{b-1}$ , respectively. The  $b$ th PLS step is completed by ordinary linear regression of  $X_{b-1}(t)$  and  $\tilde{Y}_{b-1}$  on  $t_b$ , obtaining the residuals  $X_b(t)$  and  $\tilde{Y}_b$

$$X_b(t) = X_{b-1}(t) - p_b(t)t_b \quad t \in T$$

$$\tilde{Y}_{b,i} = \tilde{Y}_{b-1,i} - c_{b,i}t_b \quad i = 1, \dots, K - 1,$$

where  $p_b(t) = (\mathbb{E}(X_{b-1}(t)t_b)/\mathbb{E}(t_b^2))$  and  $c_b = (\mathbb{E}(Y_{b-1}t_b)/\mathbb{E}(t_b^2))$ .

*A basis representation approach*

Let us consider now the basis representation of the functional predictor and the weight function, that is,

$$X(t) = \sum_{j=1}^p \alpha_j \phi_j(t) \quad \text{and} \quad w(t) = \sum_{j=1}^p w_j \phi_j(t).$$

As a consequence, the maximization problem in Equation (2.2) can be written as follows:

$$\{w, c\} = \operatorname{argmax} \frac{w' \Psi \Sigma_{\alpha \tilde{Y}} c c' \Sigma'_{\alpha \tilde{Y}} \Psi w}{w' \Psi w + c' c},$$

with  $w = (w_1, \dots, w_p)'$  being the vector of basis coefficients of  $w(t)$  and  $\Sigma_{\alpha \tilde{Y}} = (\sigma_{ji})_{p \times K-1}$  being the cross-covariance matrix between  $\alpha$  (the vector of basis coefficients of  $X(t)$ ) and  $\tilde{Y}$ , where  $\sigma_{ji} = E[\alpha_j \tilde{Y}_i]$ , with  $\{j = 1, \dots, p; i = 1, \dots, K - 1\}$ .

The cross-covariance operators expressed in terms of the basis expansion of the functional predictor  $X(t)$  are given by

$$\begin{aligned} C_{\tilde{Y}X} : L^2(T) &\rightarrow \mathbb{R}^{K-1} \\ f = \sum_{j=1}^p f_j \phi_j(t) &\rightarrow C_{\tilde{Y}X}(f) = \Sigma'_{\alpha \tilde{Y}} \Psi f \\ C_{X\tilde{Y}} : \mathbb{R}^{K-1} &\rightarrow L^2(T) \\ \mathbf{x} \rightarrow f(t) = \sum_{j=1}^{K-1} f_j \phi_j(t) &: \mathbf{f} = \Sigma_{\alpha \tilde{Y}} \mathbf{x}, \end{aligned}$$

where  $f = (f_1, \dots, f_p)'$  is the vector of basis coefficients of the function  $f$ .

Then, at the first PLS step, the weight function  $w_1(t)$  is given by the eigenfunction associated with the largest eigenvalue of  $U_X$  so that

$$\Sigma_{\alpha \tilde{Y}} \Sigma'_{\alpha \tilde{Y}} \Psi w_1 = \lambda w_1, \tag{2.3}$$

and the weight vector  $c_1$  is given by the eigenvector associated with the largest eigenvalue of  $U_{\tilde{Y}}$  so that

$$\Sigma'_{\alpha \tilde{Y}} \Psi \Sigma_{\alpha \tilde{Y}} c_1 = \lambda c_1.$$

Let us consider now the decomposition  $\Psi = (\Psi^{1/2}) (\Psi^{1/2})'$ . Then,

$$w_1' \Psi w_1 = w_1' (\Psi^{1/2}) (\Psi^{1/2})' w_1 = \tilde{w}_1' \tilde{w}_1,$$

with  $\tilde{\boldsymbol{w}}_1 = (\boldsymbol{\Psi}^{1/2})' \boldsymbol{w}_1$  ( $\boldsymbol{w}_1 = (\boldsymbol{\Psi}^{-1/2})' \tilde{\boldsymbol{w}}_1$ ). This way, Equation (2.3) can be expressed as follows

$$(\boldsymbol{\Psi}^{1/2})' \boldsymbol{\Sigma}_{\alpha \tilde{\boldsymbol{Y}}} \boldsymbol{\Sigma}'_{\alpha \tilde{\boldsymbol{Y}}} \boldsymbol{\Psi}^{1/2} \tilde{\boldsymbol{w}}_1 = \lambda \tilde{\boldsymbol{w}}_1. \tag{2.4}$$

At the  $h$ th PLS step, the  $h$ th PLS component is estimated in terms of the associated weight function by solving to the following problem

$$(\boldsymbol{\Psi}^{1/2})' \boldsymbol{\Sigma}_{\alpha_{b-1} \tilde{\boldsymbol{Y}}_{b-1}} \boldsymbol{\Sigma}'_{\alpha_{b-1} \tilde{\boldsymbol{Y}}_{b-1}} \boldsymbol{\Psi}^{1/2} \tilde{\boldsymbol{w}}_b = \lambda \tilde{\boldsymbol{w}}_b, \tag{2.5}$$

where  $\boldsymbol{\Sigma}_{\alpha_{b-1} \tilde{\boldsymbol{Y}}_{b-1}}$  is the cross-covariance matrix between  $\boldsymbol{\alpha}_{b-1}$  (the vector of basis coefficients of  $X_{b-1}(t)$ ) and  $\tilde{\boldsymbol{Y}}_{b-1}$ , and  $\tilde{\boldsymbol{w}}_b = (\boldsymbol{\Psi}^{1/2})' \boldsymbol{w}_b$  ( $\boldsymbol{w}_b = (\boldsymbol{\Psi}^{-1/2})' \tilde{\boldsymbol{w}}_b$ ).

In general, by considering Equations (2.4) and (2.5) and taking into account that  $(\boldsymbol{\Psi}^{1/2})' \boldsymbol{\Sigma}_{\alpha \tilde{\boldsymbol{Y}}}$  is the cross-covariance matrix between  $\tilde{\boldsymbol{Y}}$  and  $(\boldsymbol{\Psi}^{1/2})' \boldsymbol{\alpha}$ , it can be concluded that FPLS with multiple response is equivalent to ordinary PLS of  $\tilde{\boldsymbol{Y}}$  on the random vector  $(\boldsymbol{\Psi}^{1/2})' \boldsymbol{\alpha}$ .

## 2.2 Penalized multi-class FPLS regression

As the biomechanical data analysed in this article are not smooth enough, a penalized approach for multi-class functional partial least squares (FPLS) regression is proposed.

Let us consider now the roughness penalty function

$$Pen_d(w) = \int_T [D^d(w)(t)]^2 dt,$$

with  $d = 2$ , as a measure of roughness of  $w$  (O’Sullivan, 1986). By considering the basis expansion of  $w(t)$ , the penalty function can be written as  $Pen_d(w) = \boldsymbol{w}' \boldsymbol{P}_d \boldsymbol{w}$ , where  $\boldsymbol{w} = (w_1, \dots, w_p)'$  is the vector of basis coefficients of  $w(t)$  and  $\boldsymbol{P}_d$  the matrix of the cross inner products of the  $d$ -order derivatives of the basis functions. As alternative, working in any context where regression on B-splines is useful, it can be considered  $\boldsymbol{P}_d = (\boldsymbol{\Delta}^d)' \boldsymbol{\Delta}^d$ , with  $\boldsymbol{\Delta}^d$  the matrix of  $d$ -order differences between the adjacent basis coefficients (Eilers and Marx, 1996).

In this section the PLS components are estimated by solving the following problem

$$\{w, c\} = \underset{\|w\|_\lambda = 1}{\underset{\|c\| = 1}{\operatorname{argmax}}} \operatorname{Cov}^2 \left( \int_T X(t) w(t) dt, \sum_{i=1}^{K-1} c_i \tilde{Y}_i \right),$$

with  $\|\cdot\|$  being the norm in the space  $\mathbb{R}^{K-1}$  and  $\|\cdot\|_\lambda$  being the norm associated with an inner product defined by  $\langle \mathbf{f}, \mathbf{g} \rangle_\lambda = \langle \mathbf{f}, \mathbf{g} \rangle + \lambda(\mathbf{f}'\mathbf{P}_d\mathbf{g})$ , with  $\mathbf{f} = (f_1, \dots, f_p)'$  and  $\mathbf{g} = (g_1, \dots, g_p)'$  being the vectors of basis coefficients of  $f(t)$  and  $g(t)$ , respectively.

Then, the  $h$ th PLS component ( $h > 1$ ),  $t_h = \int_T X_{b-1}(t)w_b(t)dt$ , is estimated by solving the problem

$$\{w, c\} = \underset{\substack{\|w\|_\lambda = 1 \\ \|c\| = 1}}{\operatorname{argmax}} \operatorname{Cov}^2 \left( \int_T X_{b-1}(t)w(t)dt, \sum_{i=1}^{K-1} c_i \tilde{Y}_{b-1,i} \right). \tag{2.6}$$

By assuming the basis representation of the functional predictor and the weight function, the problem in (2.6) can be written as follows

$$\{w, c\} = \underset{\operatorname{argmax}}{\frac{w'\Psi\Sigma_{\alpha_{b-1}\tilde{Y}_{b-1}}cc'\Sigma'_{\alpha_{b-1}\tilde{Y}_{b-1}}\Psi w}{w'(\Psi + \lambda\mathbf{P}_d)w + c'c}}, \tag{2.7}$$

with  $w = (w_1, \dots, w_p)'$  being the vector of basis coefficients of  $w(t)$ ,  $\Sigma_{\alpha_{b-1}\tilde{Y}_{b-1}} = (\sigma_{ji})_{p \times K-1}$  being the cross-covariance matrix between  $\alpha_{b-1}$  and  $\tilde{Y}_{b-1}$ , where  $\sigma_{ji} = E[\alpha_{b-1,j}\tilde{Y}_i]$ , with  $\{j = 1 \dots, p; i = 1, \dots, K - 1\}$ , and  $\lambda$  and  $\mathbf{P}_d$  the smoothing parameter and the penalty matrix, respectively.

Assuming the decomposition  $\mathbf{L}\mathbf{L}' = \Psi + \lambda\mathbf{P}_d$  and defining  $\tilde{w} = \mathbf{L}'w$  ( $w = (\mathbf{L}^{-1})'\tilde{w}$ ) the problem in (2.7) can be expressed as follows

$$\{w, c\} = \underset{\operatorname{argmax}}{\frac{\tilde{w}'\mathbf{L}^{-1}\Psi\Sigma_{\alpha\tilde{Y}}cc'\Sigma'_{\alpha\tilde{Y}}\Psi(\mathbf{L}^{-1})'\tilde{w}}{\tilde{w}'\tilde{w} + c'c}}.$$

By analogy with the non-penalized approach, the associated eigenproblem is

$$\mathbf{L}^{-1}\Psi\Sigma_{\alpha\tilde{Y}}\Psi'(\mathbf{L}^{-1})'\tilde{w} = \lambda\tilde{w}, \quad \tilde{w} \in \mathbb{R}^p. \tag{2.8}$$

In general, the  $h$ th PLS component,  $t_h = \int_T X(t)w_b(t)dt$ , is defined by  $w_b = (\mathbf{L}^{-1})'\tilde{w}_b$ , which is the vector of basis coefficients of  $w_b(t)$ , with  $\tilde{w}_b$  being the eigenvector related to the largest eigenvalue of the problem in (2.8).

Finally, taking into account that  $\mathbf{L}^{-1}\Psi\Sigma_{\alpha\tilde{Y}}$  is the cross-covariance matrix between  $\tilde{Y}$  and  $\mathbf{L}^{-1}\Psi\alpha$ , the penalized approach is reduced to a classical PLS of  $\tilde{Y}$  on the random vector  $\mathbf{L}^{-1}\Psi\alpha$ .

### 2.3 The problem of repeated measures

In previous sections, penalized and non-penalized approaches for the multi-class FPLS regression have been proposed. However, these approaches are based on ordinary PLS, and then the between-subject and the within-subject variations are not studied by separate.

As solution to the problem of repeated measures, in this section a multi-class approach for penalized and non-penalized FPLS regression based on the split-up variation is proposed. In the same spirit as in de Nord and Theobald (2005), Westerhuis et al. (2010) and Liquet et al. (2012), this approach first decomposes the variability in the data matrix and then applies the multi-class PLS regression on the within-subject variation matrix.

For a more detailed explanation of the multi-class approach, let us consider  $\{x_i(t) : t \in T, i = 1, \dots, n\}$  a sample of the functional variable  $X(t)$  and  $\{y_1, y_2, \dots, y_n\}$  a random sample of  $Y$  associated with it, with  $y_i = \{1, 2, \dots, K\}$ . Henceforth, let  $\tilde{Y}$  denote a  $N \times (K - 1)$  matrix with zeros and ones, comprising the values associated with the dummy variables obtained from  $Y$ .

Then, the sample estimation of the multi-class FPLS described in sections 2.1 (non-penalized) and 2.2 (penalized) is as follows:

- Non-penalized multi-class FPLS: multi-class PLS of  $\tilde{Y}$  on the matrix  $\mathbf{X}_{(N \times p)}^1 = \mathbf{A}\Psi^{1/2}$
- Penalized multi-class FPLS: multi-class PLS of  $\tilde{Y}$  on the matrix  $\mathbf{X}_{(N \times p)}^2 = \mathbf{A}\Psi(\mathbf{L}^{-1})'$ , so that  $\mathbf{L}\mathbf{L}' = \Psi + \lambda\mathbf{P}_d$ ,

with  $\mathbf{A}$  being the matrix of basis coefficients of the functional variable  $X$ , and  $\Psi$  and  $\mathbf{P}_d$  as in previous sections.

In practice, functional data are observed in a finite set of points. Because of this, matrix  $\mathbf{A}$  must be estimated. By assuming that the underline process to biomechanical data is smooth, B-spline basis functions are considered, and the basis coefficients are estimated by the least squares criterion. This approach is called regression splines. Additionally, for data observed with noise, a penalized estimation of the basis coefficients in terms of P-splines can be considered (see Eilers and Marx, 1996; Aguilera and Aguilera-Morillo, 2013, for more details).

#### *Split-up variation*

In general, let us consider a data matrix  $\mathbf{X}_{N \times p}$  containing the information related to  $n$  subjects and each of the  $K$  possible stimulus ( $N = n \times K$ ), with  $p$  being the number of basis functions (or variables in the multivariate setting). In that sense, let  $x_{ik}^j$  denote the value of the  $j$ th column of  $X$  for the  $i$ th subject under the  $k$ th stimulus.

Following the mixed-effect models philosophy, matrix  $\mathbf{X}$  can be decomposed into an offset term, a between-subject and a within-subject part so that

$$\mathbf{X} = \mathbf{X}_{\text{offset}} + \mathbf{X}_{\text{between-subject}} + \mathbf{X}_{\text{within-subject}}.$$

But mixed models rely on assumptions such as Gaussian distribution of random effects. As an alternative, a split-up variation approach, which does not require the above-mentioned assumptions, is considered in this article. Exactly, the offset, between-subject and within-subject variation matrices ( $\mathbf{X}_o$ ,  $\mathbf{X}_b$  and  $\mathbf{X}_w$ , respectively) can be obtained as follows:  $\mathbf{X}_o = \mathbf{1}_{(N \times 1)} \mathbf{x}'_{\cdot(1 \times p)}$ , where  $\mathbf{x}_{\cdot} = (x_{\cdot}^1, \dots, x_{\cdot}^p)'$  and  $x_{\cdot}^j = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^n x_{ik}^j$ .  $\mathbf{X}_b$  is obtained by concatenating the matrices  $\mathbf{1}_{(K \times 1)} \mathbf{x}'_{bi(1 \times p)}$  for each subject into  $\mathbf{X}_b$ , where  $\mathbf{x}_{bi} = (x_{i1}^1 - x_{\cdot}^1, \dots, x_{i1}^p - x_{\cdot}^p)'$  and  $x_{i1}^j = \frac{1}{K} \sum_{k=1}^K x_{ik}^j$ .  $\mathbf{X}_w = \mathbf{X} - \mathbf{X}_b$ , with  $\mathbf{X}_i$  obtained by concatenating the matrices  $\mathbf{1}_{(K \times 1)} \mathbf{x}'_{i(1 \times p)}$  for each subject into  $\mathbf{X}_i$ , with  $\mathbf{x}_i = (x_i^1, \dots, x_i^p)'$ .

#### Multi-class approach for FPLS

Once the split-up variation has been carried out, penalized and non-penalized multi-class FPLS regression is performed on the within-subject variation matrix. In that sense, the penalized and non-penalized multi-class approaches for FPLS regression are given by

- Non-penalized multi-class FPLS: multi-class PLS of  $\tilde{\mathbf{Y}}$  on the matrix  $\mathbf{X}_w^1$
- Penalized multi-class FPLS: multi-class PLS of  $\tilde{\mathbf{Y}}$  on the matrix  $\mathbf{X}_w^2$ ,

where  $\mathbf{X}_w^1$  and  $\mathbf{X}_w^2$  are the within-subject variation matrices from  $\mathbf{X}^1$  and  $\mathbf{X}^2$ , respectively.

## 2.4 Functional LDA on a reduced set of functional PLS components

Let  $\mathbf{T}$  be a  $n \times q$  matrix comprising the columns of the first  $q$  PLS component scores (by variability order). By considering the sample estimation of the penalized and non-penalized multi-class approaches for FPLS regression proposed in Section 2.3, the sample estimation of  $\mathbf{T}$  is as follows

- Non-penalized multi-class FPLS:  $\mathbf{T} = \mathbf{A}\Psi^{1/2}\mathbf{V}$
- Penalized multi-class FPLS:  $\mathbf{T} = \mathbf{A}\Psi(\mathbf{L}^{-1})'\mathbf{V}$ ,

with  $\mathbf{V}$  being the matrix comprising the columns of the eigenvectors  $\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_q$  associated with the  $t_1, \dots, t_q$  PLS components.

Once the PLS components have been estimated, LDA of the original sample values of  $Y$  ( $y_1, \dots, y_n$ ) on the matrix  $\mathbf{T}$  is carried out. Then, linear combinations such as  $\Phi_i(X) = \alpha_i + \mathbf{T}\beta_{PLS}^i$ ,  $\{i = 1, \dots, K - 1\}$  are estimated, so that the between-class variance is maximized with respect to the within-class variance, with  $\beta_{PLS}^i = (\beta_{PLS_1}^i, \dots, \beta_{PLS_q}^i)'$  being the vector of discriminant coefficients.

But the aim is to estimate the discriminant functions  $\beta^i(t)$ ,  $i = 1, \dots, K - 1$  included in the linear combinations given in equation (2.1). To this end, let us

consider the basis representation of the sample curves and the discriminant functions ( $\beta^i(t) = \sum_{j=1}^p \beta_j^i \phi_j(t)$ ) and, as a consequence, equation (2.1) can be rewritten as

$$\Phi_i(X) = \alpha_i + \int_0^T X(t)\beta^i(t)dt = \alpha_i + \mathbf{A}\Psi\beta^i,$$

with  $\beta^i$  being the vector of basis coefficients of the  $i$ th discriminant function.

Finally, taking into account the sample estimation of  $\mathbf{T}$ , we can conclude that

- Non-penalized multi-class FPLS:  $\beta^i = (\Psi^{-1/2})'\mathbf{V}\beta_{PLS}^i, i = 1, \dots, K - 1$
- Penalized multi-class FPLS:  $\beta^i = (\mathbf{L}^{-1})'\mathbf{V}\beta_{PLS}^i, i = 1, \dots, K - 1.$

Once the random vector of discriminant functions  $\beta(t) = (\beta^1(t), \dots, \beta^{K-1}(t))'$  is obtained, the classification is then performed in the transformed space based on some distance metric, such as Euclidean distance.

Then given a new sample observation  $x_0(t)$ , it is classified to

$$\operatorname{argmin}_k d\left(\int_T x_0(t)\beta(t), \int_T \bar{x}^k(t)\beta(t)\right), \tag{2.9}$$

where  $\bar{x}^k(t) = \frac{1}{m_k} \sum_{x(t) \in C_k} x(t)$  is the mean curve related to the  $k$ th class and  $m_k$  the number of sample observations in the  $k$ th class, with  $k = 1, \dots, K$ .

By assuming the basis representations of  $x_0(t)$ ,  $\beta(t)$  and  $\bar{x}^k(t)$

$$x_0(t) = \sum_{j=1}^p a_{0j} \phi_j(t), \quad \beta^i(t) = \sum_{j=1}^p \beta_j^i \phi_j(t), \quad \bar{x}^k(t) = \sum_{j=1}^p \bar{a}_j^k \phi_j(t),$$

problem (2.9) can be rewritten as follows

$$\operatorname{argmin}_k d\left(\mathbf{a}'_0 \Psi \mathbf{B}, \bar{\mathbf{a}}^k \Psi \mathbf{B}\right),$$

with  $\mathbf{a}_{0_{p \times 1}}$  being the vector of basis coefficients of  $x_0(t)$ ,  $\bar{\mathbf{a}}^k_{p \times 1}$  being the vector of basis coefficients of  $\bar{x}^k(t)$ ,  $\Psi_{p \times p}$  being the inner product matrix,  $\mathbf{B}_{p \times (K-1)}$  the matrix with columns the vectors of basis coefficients of the  $K - 1$  discriminant functions (i.e.,  $\beta^1, \dots, \beta^{K-1}$ ). Finally, the Euclidean distance between two vectors  $\mathbf{V}^1 = \mathbf{a}'_0 \Psi \mathbf{B}$  and  $\mathbf{V}^2 = \bar{\mathbf{a}}^k \Psi \mathbf{B}$  is computed as  $d(\mathbf{V}^1, \mathbf{V}^2) = \sqrt{\sum_{i=1}^{K-1} (\mathbf{V}_i^1 - \mathbf{V}_i^2)^2}$ .

The smoothing parameter  $\lambda$  and the number of PLS components used in the LDA approach based on Penalized multi-class FPLS have been selected by leave-one-out cross-validation on a training sample, selecting the combination of both that minimizes the misclassification rate obtained from LDA on the PLS components.

In the non-penalized approach, only the number of PLS components is required, and it is chosen by leave-one-out cross-validation on a training sample, minimizing the misclassification rate obtained from functional LDA.

All the analysis in Sections 3 and 4 were carried out in a computer with an Intel Core i5 processor at 3 GHz with 8Gb of RAM, running R-project version 3.5.1 (R Development Core Team, 2008). The two multi-class functional PLS approaches were reduced to multi-class PLS regression (in the multivariate context), and then they were implemented in R using the function *spls* available in the *mixOmics* R package developed by Rohart et al. (2017).

Authors are preparing an R package that, among other things, includes the code used in the manuscript. For readers wishing to use the software, it will be available as soon as possible at the R repository website. However, you can contact the corresponding author if you need more information regarding the code.

### 3 Simulation study

In order to test the performance of the proposed methods, a simulation study has been carried out. Inspired by Cuevas et al. (2004), let us consider 120 sample curves, 40 for each of the 3 classes, according to the following process

$$X_i^k(t) = m_k(t) + \epsilon_i^k(t), \quad t \in [0, 1], \quad i = 1, \dots, 40, \quad k = 1, 2, 3,$$

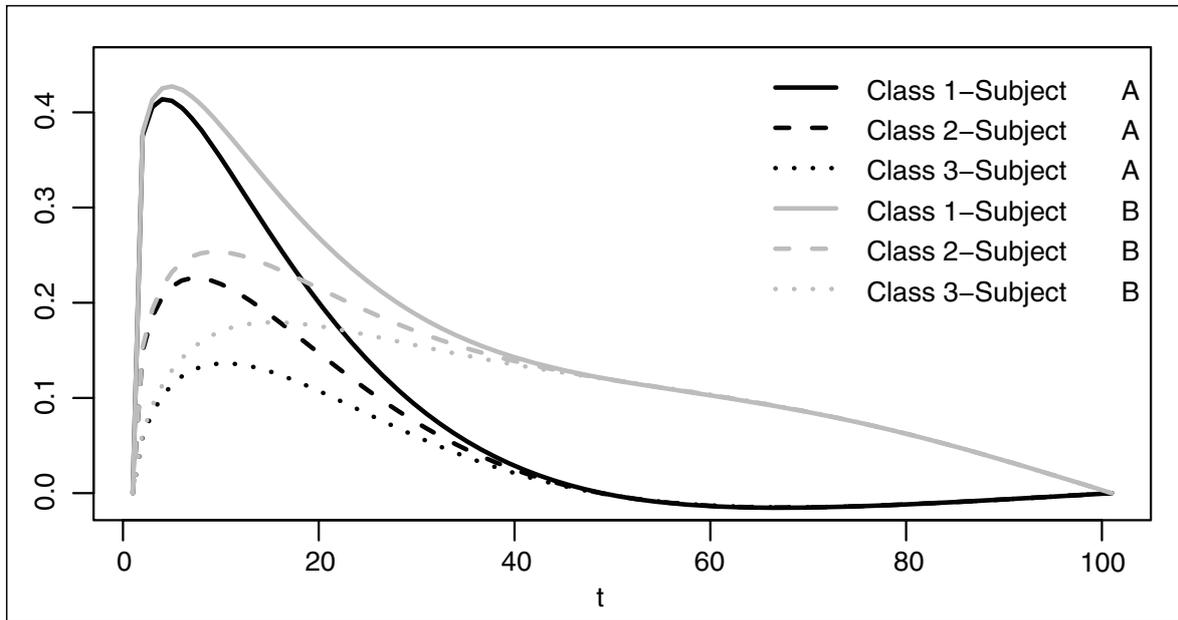
where  $m_i(t) = t^{(k/5)}(1-t)^{(6-k/5)}$ . The process  $X_i^k(t)$  was generated in discretized version  $X_i^k(t_j)$ , for  $j = 1, \dots, 101$ , with  $t_1, \dots, t_{101}$  being equidistant points in  $[0, 1]$ . Finally,  $\epsilon_i^k(t_j)$  are i.i.d. random variables  $N(0, \sigma_\epsilon = 0.2)$ .

But our work is focused on the presence of repeated measurements in the  $k$  classes, and then an additional term must be considered in order to represent the subject effect. In that sense, and inspired by Durbán et al. (2005), a specific-subject term has been included in the above process so that

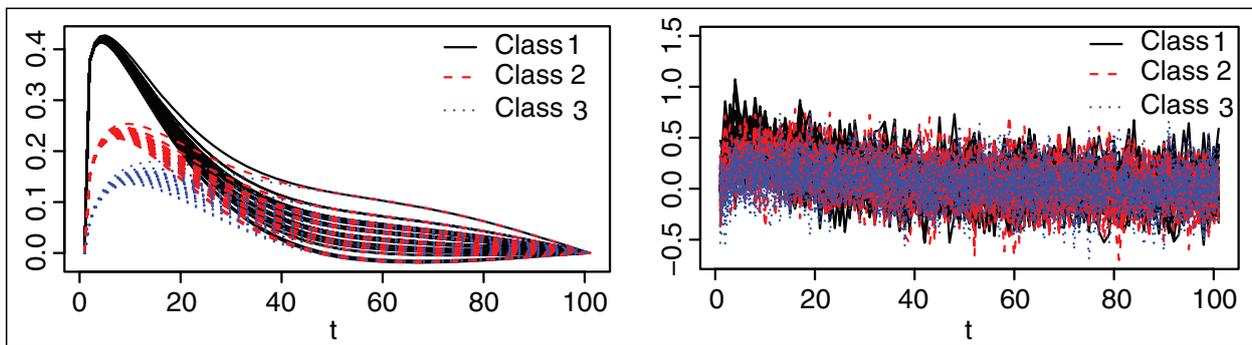
$$X_i^k(t) = m_k(t) + a_i \sin(t\pi) + \epsilon_i^k(t), \quad t \in [0, 1], \quad i = 1, \dots, 40, \quad k = 1, 2, 3,$$

with  $a_i \sim N(\mu_i, \sigma_s = 0.02)$  and  $\mu_i \sim \text{Uniform}(0, 0.05)$ . The  $\sin(t\pi)$  function has been considered in order to simulate curves taking the same value at the beginning and at the end of the observation interval (as in the gait data study). Of the 120 simulated sample curves, 90 (30 for each class) were considered as training sample and the remaining 30 (10 for each class) as test sample.

The sample curves from two different subjects, and for each of the three classes, are shown in Figure 1. As we can see, there is a clear dependency structure between the three observations of the same subject, and the between-subjects variation is not only given by a magnitude effect, but also by a shape variation. This is a realistic simulation of the performance of kinematic data since it makes no sense to assume that the curves of the different subjects are parallel.



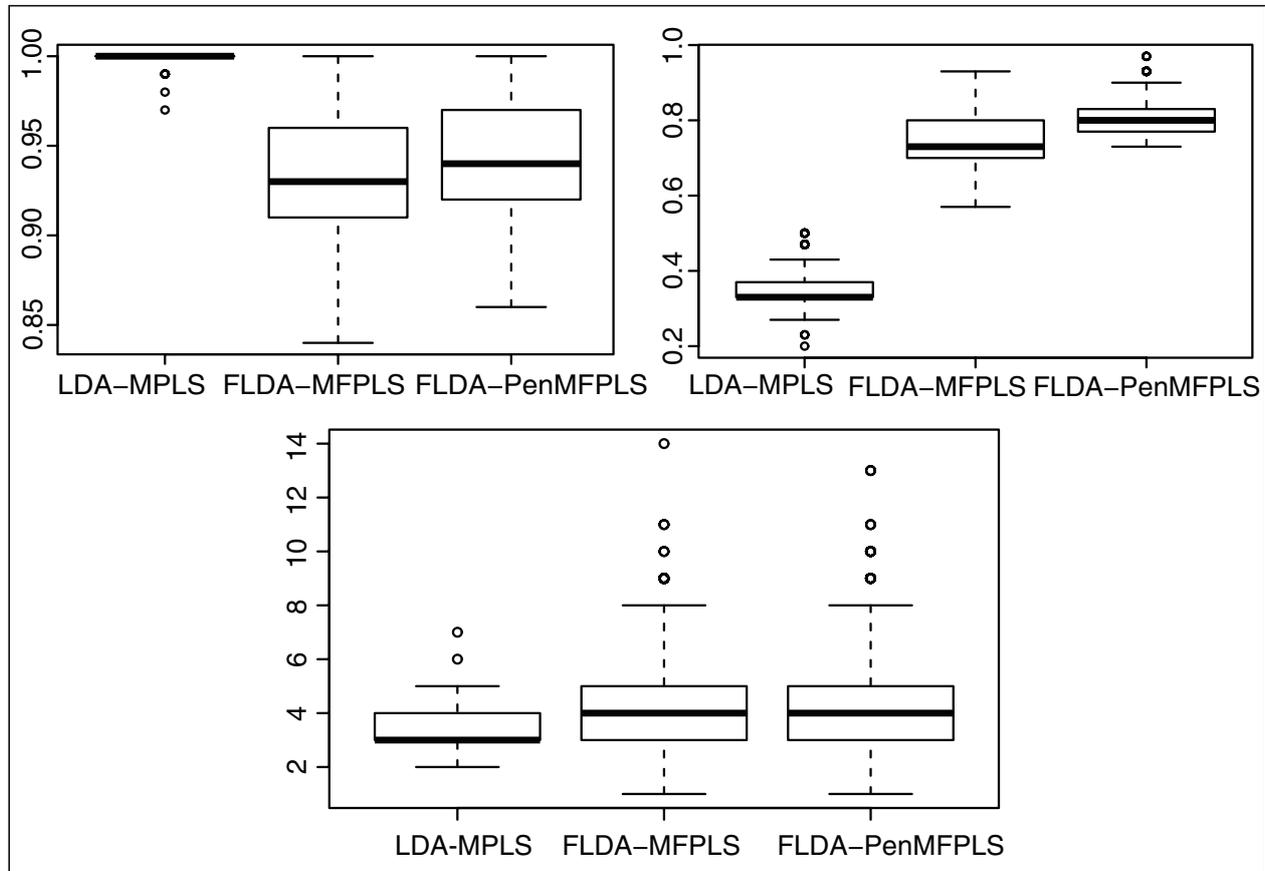
**Figure 1** Simulation study. Sample paths related to classes 1, 2 and 3 (solid, dashed and dotted line, respectively) for two subjects A and B, black and grey lines, respectively



**Figure 2** Simulation study. Smooth and noisy sample paths (left and right panel, respectively) related to classes 1, 2 and 3 (solid line, dashed line and dotted line, respectively) for 30 subjects

In Figure 2, all the sample curves, with and without noise, are displayed highlighting the three classes by different line types. In order to fit the regression splines, a cubic B-spline basis defined on 15 basis knots has been considered.

The simulation scheme was run 500 times and the main results are summarized in Figure 3. As we can see, the two functional approaches tend to select a larger number of PLS components than the multivariate version, as covariates in the LDA. But there is not a great difference between them. Regarding the correct classification rates (CCR), there is a clear overfitting in the multivariate approach, which provides CCR close to 1 in the cross-validation. Between the two proposed functional approaches,



**Figure 3** Simulation study. Box plots showing to the number of PLS components used in the LDA (bottom panel), the correct classification rates from the cross-validation on the training sample and from the classification of the test sample (top-left and top-right panels, respectively). Experiments run 500 times

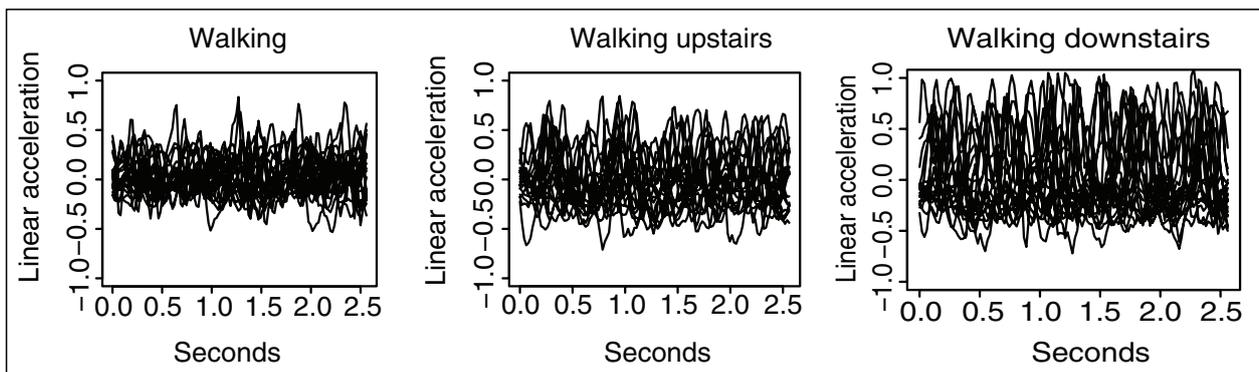
the one based on penalized multi-class functional PLS regression provides the best performance, with the highest CCR and the minimum variability.

#### 4 Case studies

In this section, the results from the analysis of two real kinematic datasets are summarized. In general, kinematic data from inertial sensors placed on human body is usually collected in the format of linear acceleration or angular rotation. Linear acceleration or angular rotation represent the trajectory of the human movement, which can be identified with a function in time and frequency domain. In that sense, data from kinematic studies can be considered as a functional dataset, and then functional data techniques can be applied to analyse this type of data. In addition, it is important to remark the presence of repeated measures in kinematic studies. So, both the between-subject and within-subject variations must be taken into account.

#### 4.1 Human activity data

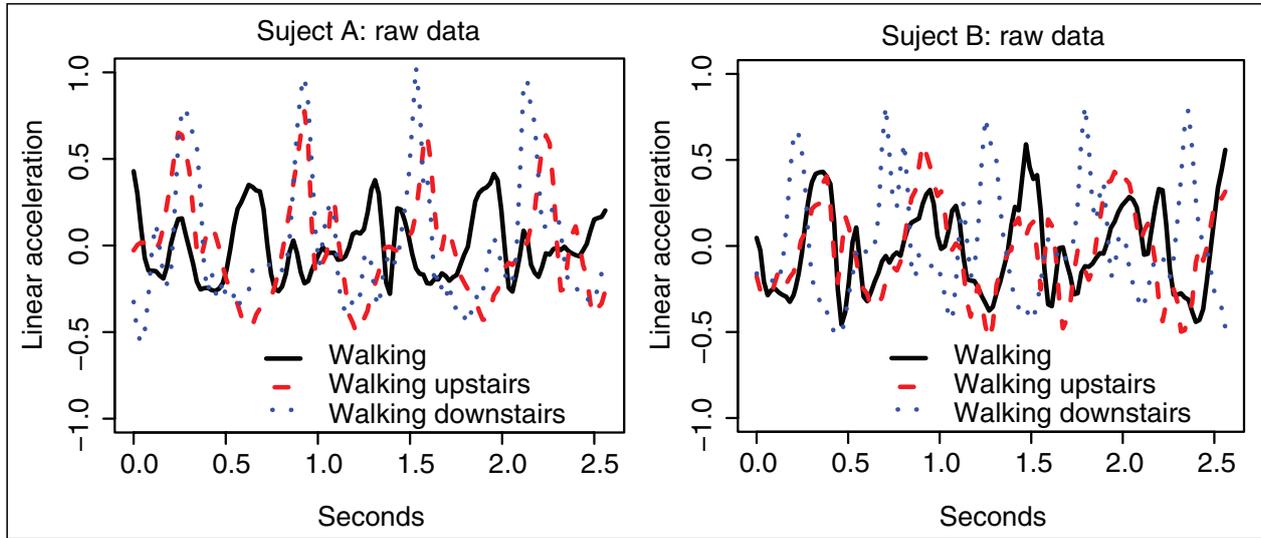
The human activity dataset is part of a wider experiment, focused on the human activity recognition, carried out by Anguita et al. (2013). In this article, let us consider the linear acceleration (metre per second squared), measured on axis X and recorded in 128 equidistant knots at the interval  $[0, 2.56]$  seconds, and related to three different stimulus (walking, walking upstairs and walking downstairs). A total of 30 subjects participated in this experiment. Then, a total of 90 sample paths are considered (one sample path per stimulus and subject). Subject 23 was removed from the sample for being an outlier. The remaining sample observations were distributed into training and test sample, with observations (repeated measures) related to 20 and 9 subjects, respectively. The raw data (displayed by stimulus) are shown in Figure 4. In order to visualize the variability of the spectra between stimulus, in Figure 5 the spectra related to the three activities have been overlapped for two subjects A and B, left and right panel, respectively.



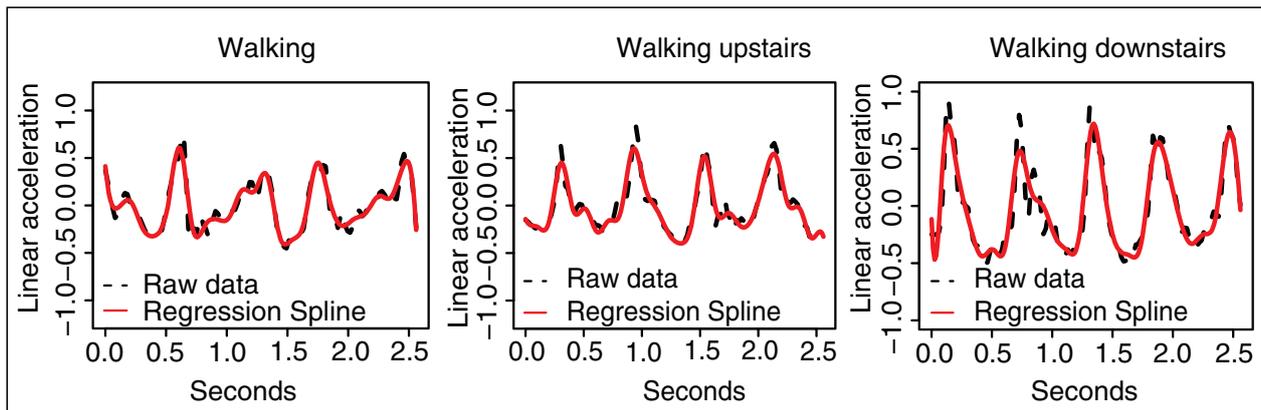
**Figure 4** Human activity data. Raw data. Sample paths displayed separately by stimulus: walking, walking upstairs and walking downstairs, from left to right, respectively

The first step in the analysis was to approximate the sample curves by mean of a basis representation and using a cubic B-spline basis defined on 25 equidistant knots. In Figure 6 the raw data together with the regression spline can be seen. Once the sample curves have been approximated, the aim is the human activity pattern recognition. To this end FLDA based on a multi-class approach for functional PLS regression (FLDA-MFPLS) was carried out to classify the sample curves according to the stimulus which produced them. It could be assumed that the raw data, which were collected by a smartphone on the waist of the subjects, are affected by some error or noise. In that sense, and aiming to avoid a possible lack of smoothness in the estimation of the discriminant functions, the penalized version of FLDA-MFPLS (FLDA-PenMFPLS), proposed in section 2.3, was also considered.

The discriminant functions estimated from penalized and non-penalized functional LDA can be seen in Figure 7. The necessity of smoothing in the estimation of the discriminant functions is obvious, with penalized functional LDA providing the smoothest and the most interpretable functions.



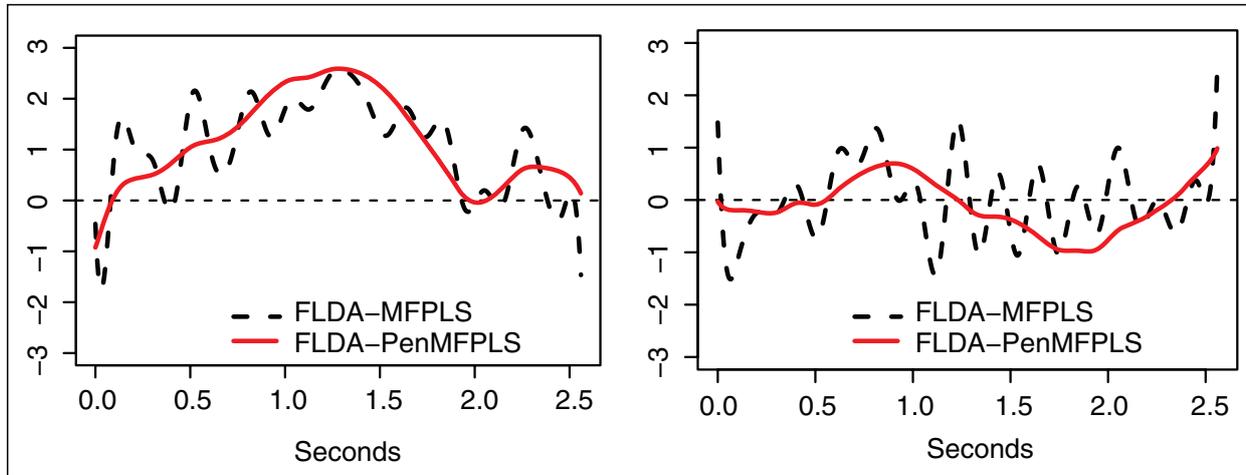
**Figure 5** Human activity data. Raw data. Sample paths related to walking (solid line), walking upstairs (dashed line) and walking downstairs (dotted line) for two subjects A and B, left and right panel, respectively



**Figure 6** Human activity data. A sample path for each stimulus (dashed line) together with the regression spline (solid line) estimated with a cubic B-splines basis defined on 25 equidistant knots

In order to evaluate the ability of penalized and non-penalized approaches for human activity pattern recognition, Table 1 summarizes information related to the CCR based on a leave-one-out cross-validation on the training sample ( $CCR_{training}$ ), and based on the test sample classification ( $CCR_{test}$ ). In this table, a multivariate version of LDA based on multi-class PLS regression (LDA-MPLS) is also considered for comparison purpose.

It can be seen that functional LDA performs better than the multivariate approach. Besides, between the two functional approaches (penalized and non-penalized), the one based on penalized multi-class functional PLS (FLDA-PenMFPLS) achieves the highest CCR in both, training and test samples. A considerable difference between



**Figure 7** Human activity data. Discriminant functions estimated by functional LDA based on non-penalized (dashed line) and penalized (solid line) multi-class functional PLS regression, using 4 and 7 PLS components, respectively

**Table 1** Human activity data. Correct classification rate based on a leave-one-out cross-validation on the training sample ( $CCR_{training}$ ) and based on the test sample classification ( $CCR_{test}$ ), and the number of PLS components ( $\#PLScp's$ ) used in the functional LDA based on penalized and non-penalized multi-class functional PLS regression (FLDA-PenMFPLS and FLDA-MFPLS, respectively).  $\lambda$  denotes the smoothing parameter of the penalized approach. A multivariate version of LDA based on multi-class PLS regression (LDA-MPLS) is also considered for comparison purpose

	LDA-MPLS	FLDA-MFPLS	FLDA-PenMFPLS
$CCR_{training}$ (%)	83	95	93
$CCR_{test}$ (%)	56	67	78
$\#PLScp's$	5	4	7
$\lambda$	–	–	0.41

the CCR in training and test samples highlights a possible overfitting mainly in the multivariate and in the non-penalized functional approaches. In the penalized approach this fact is not so remarkable, providing then the best classification rate.

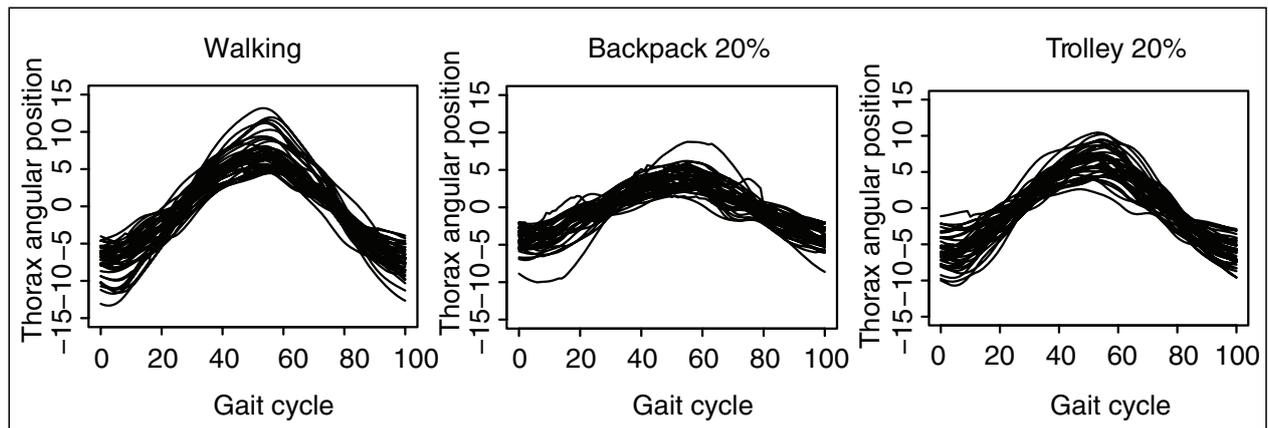
## 4.2 Gait data

The gait dataset comes from a wide experimental study developed in the biomechanics laboratories of the iMUDS. A total of 51 participants (25 boys and 28 girls) between 8 and 11 years old were involved in this study. In order to collect the data, 26 reflective markers were placed on the children's skin. The kinematics data was recorded by a 3D motion capture system (Qualisys AB, Göteborg, Sweden). Each subject completed a cycle walking over the platform in three conditions (walking, carrying a backpack that weighs 20% of the subject's weight and pulling a trolley that weighs 20% of the subject's weight). For each subject, the 3-axial angular rotation were registered for each joint (ankle, foot progress, hip, knee, pelvis, thorax) in all conditions. Finally, the angular rotation curve represents the observation for the subject's gait cycle for

each joint, axis direction and experimental condition. Interested readers could request access to the data from Jose M. Heredia-Jiménez and Eva Orantes-González from iMUDS.

In this section, only a part of the above-described experimental dataset has been considered. Exactly, we are interested on the thorax angular position (radians) measured on axis Z. This is a clear example of functional data with repeated measures in the sense that three curves, concerning the three experimental conditions, are available for each subject. In addition, each sample curve was recorded in 101 equidistant points of the gait cycle. So, the interval  $[0, 100]$  represents the percentage of gait cycle completed by the subject.

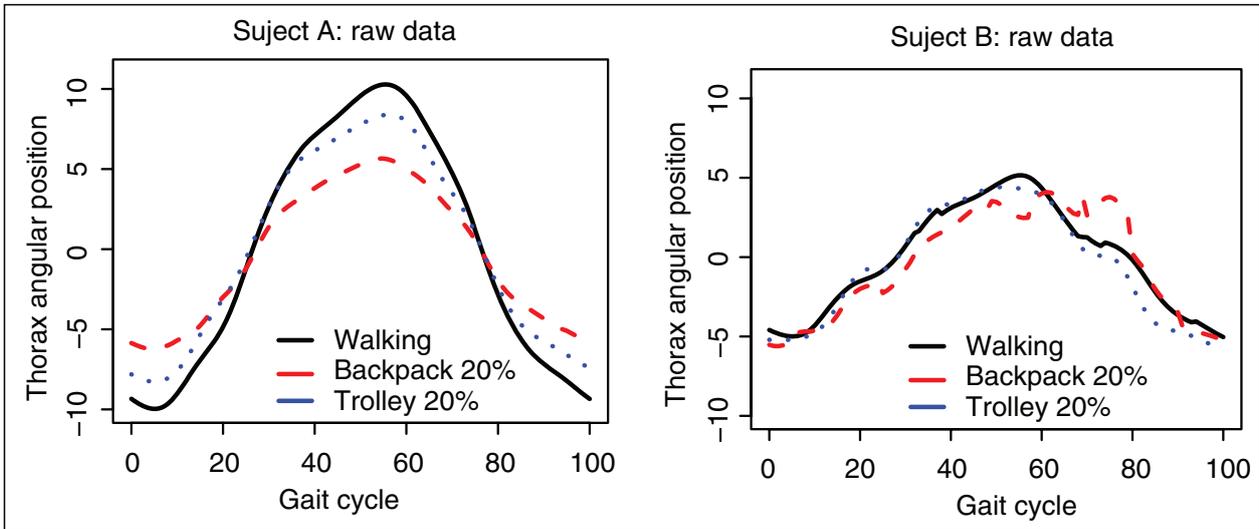
Subject 10 was removed from the sample for being an outlier. The remaining sample observations were distributed into training and test samples, with observations (repeated measures) related to 39 and 11 subjects, respectively. The raw data (displayed by experimental conditions) are shown in Figure 8.



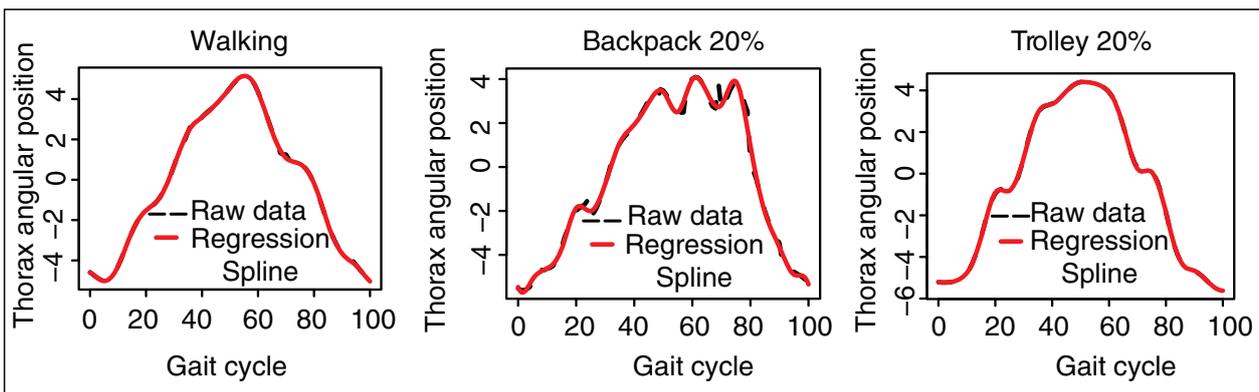
**Figure 8** Gait data. Raw data. Sample paths displayed separately by type of activity: walking, carrying a backpack and pulling a trolley, both weighing 20% of the children's weight, from left to right, respectively

In order to visualize the within subjects and the between subjects variability, in Figure 9 the spectra related to the three gait events have been overlapped for two subjects A and B, left and right panel, respectively. It can be seen that walking and walking with a trolley provide sample paths with a more similar shape than walking with a backpack. Also it is interesting to highlight the between subjects variability (differences in scale and shape).

The sample curves have been approximated by mean of a basis representation using a cubic B-spline basis defined on 20 equidistant knots. In Figure 10 the raw data and the regression splines have been displayed. Once the sample curves have been approximated, the aim is the classification of the thorax angular rotation curves. In that sense, the discriminant functions estimated by penalized and non-penalized functional LDA can be seen in Figure 11. Once again the penalized approach provides the smoothest functions, providing an intuitive interpretation in relation with detecting periods in which the function is positive or negative.



**Figure 9** Gait data. Raw data. Sample paths related to walking (solid line), carrying a backpack (dashed line) and pulling a trolley (dotted line), both weighing 20% of the children’s weight and walking downstairs from two sample individuals A and B, left and right panel, respectively

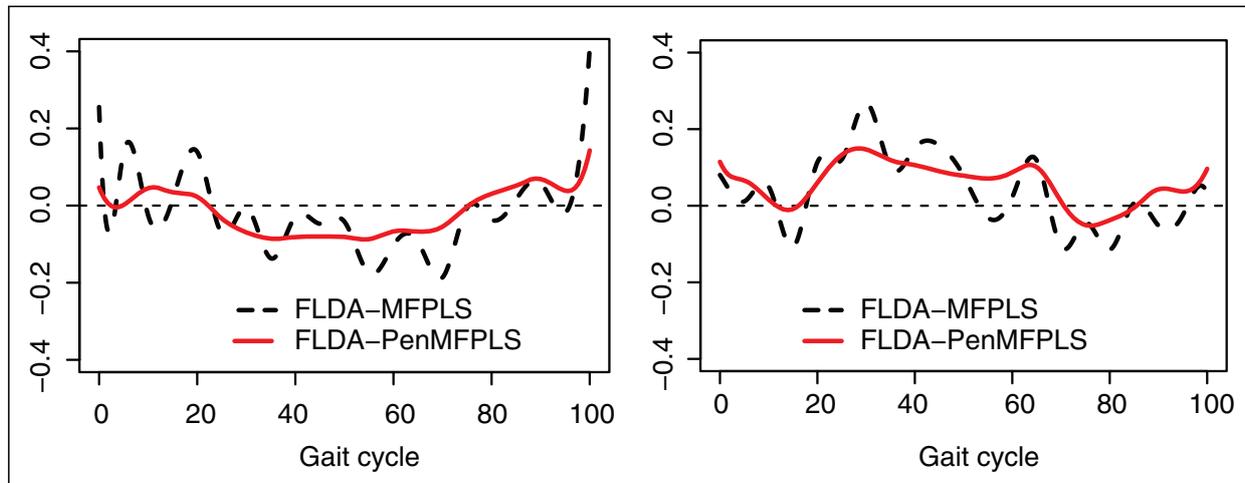


**Figure 10** Gait data. A sample path for each stimulus (dashed line) together with the regression spline (solid line) estimated with a cubic B-splines basis defined on 20 equidistant knots

The classification performance of the penalized and non-penalized approaches is shown in Table 2. It obvious that functional LDA performs better than the multivariate approach. Between the two functional approaches, the one based on penalized multi-class functional PLS (FLDA-PenMFPLS) achieves the best classification performance (the highest CCR on the test sample).

## 5 Conclusions

In this work, a methodological solution to the problem of multi-class classification of functional data with repeated measures has been proposed. FLDA has been



**Figure 11** Gait data. Discriminant functions estimated by functional LDA based on non-penalized (dashed line) and penalized (solid line) multi-class functional PLS regression, using 5 and 3 PLS components, respectively

**Table 2** Gait data. Correct classification rate based on leave-one-out cross-validation on the training sample ( $CCR_{training}$ ) and based on the test sample classification ( $CCR_{test}$ ), and the number of PLS components ( $\#PLScp's$ ) used in the functional LDA based on penalized and non-penalized multi-class functional PLS regression (FLDA-PenMFPLS and FLDA-MFPLS, respectively).  $\lambda$  denotes the smoothing parameter of the penalized approach. A multivariate version of LDA based on multi-class PLS regression (LDA-MPLS) is also considered for comparison purpose

	LDA-MPLS	FLDA-MFPLS	FLDA-PenMFPLS
$CCR_{training}(\%)$	83	87	82
$CCR_{test}(\%)$	33	61	85
$\#PLScp's$	3	5	3
$\lambda$	–	–	2.87

considered as a classifier, by solving the problem of infinite dimension of the functional data by means of a novel approach of functional PLS regression for repeated measures described in Section 2. This work has been motivated by two real problems related to kinematic data. In both cases, data is affected by some noise, and then some type of penalization must be considered in the estimation of the discriminant functions. To solve this problem in the case of functional data with repeated measures, a multi-class approach for penalized functional PLS, that introduces a P-spline penalty in the definition of the inner product in the PLS algorithm, is proposed in Sections 2.2 and 2.3.

In Sections 3 and 4, the performance of the two functional approaches (penalized and non-penalized, FLDA-PenMFPLS and FLDA-MFPLS, respectively) are compared with a multivariate version of LDA based on multi-class PLS regression (LDA-MPLS). As we can see in Figure 3 and Tables 1 and 2, from a classification point of view, functional approaches perform better than the multivariate version, increasing the CCR on a test sample in a very significant way. Regarding the two functional

approaches, the one based on penalized multi-class functional PLS achieves the best classification ability, with a difference in the CCR of more than 10% in both case studies.

The discriminant functions estimated from penalized and non-penalized functional LDA are shown in Figures 7 and 11. It can be seen that the penalized version (FLDA-PenMFPLS) provides smoother functions, which are more interpretable in the sense of detecting periods in which the function is positive or negative.

Finally, it can be concluded that LDA based on a multi-class approach for penalized functional PLS is the most competitive method from both classification and estimation point of view.

## Acknowledgements

The authors would like to thank Jose M. Heredia-Jiménez, Eva Orantes-González and Manuel Escabias for providing us the gait data from iMUDS. In addition, the authors thank the reviewers, the associate editor and the editor for their contribution to the improvement of this work.

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## Funding

This research has been supported by the research projects MTM2017-88708-P, Spanish Ministry of Economy and Competitiveness (also supported by the FEDER program) and IJCI-2017-34038, Agencia Estatal de Investigación, Ministerio de Ciencia, Innovación y Universidades.

## References

- Aguilera AM and Aguilera-Morillo MC (2013) Comparative study of different B-spline approaches for functional data. *Mathematical and Computer Modelling*, **58**, 1568–79.
- Aguilera AM, Aguilera-Morillo MC and Preda C (2016) Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems*, **154**, 80–92.
- Aguilera AM, Escabias M, Preda C and Saporta G (2010) Using basis expansion for estimating functional PLS regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, **104**, 289–305.
- Aguilera-Morillo MC and Aguilera AM (2013) P-spline estimation of functional classification methods for improving

- the quality in the food industry. *Communications in Statistics: Simulation and Computation*, **44**, 2513–34.
- Anguita D, Ghio A, Oneto L, Parra X and Reyes-Ortiz JL (2013) A Public Domain Dataset for Human Activity Recognition Using Smartphones. In *Proceedings of the 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013*, Bruges, Belgium, 437–42.
- Barker M and Rayens W (2003) Partial least squares for discrimination. *Journal of Chemometrics*, **17**, 166–73.
- Bottomley C, Daemen C, Mukri F, Papageorghiou AT, Kirk E, Pexsters A, De Moor B, Timmerman D and Bourne T (2009) Functional linear discriminant analysis: A new longitudinal approach to the assessment of embryonic growth. *Human Reproduction*, **24**, 278–83.
- Crane EA, Cassidy RB, Rothman ED and Gerstner G (2010) Effect of registration on cyclical kinematic data. *Journal of Biomechanics*, **43**, 2444–47.
- Cuevas A, Febrero M and Fraiman R (2004) An ANOVA test for functional data. *Computational Statistics and Data Analysis*, **47**, 111–22.
- Daffertshofer A, Lamoth CJC, Meijer OG and Beek PJ (2004) PCA in studying coordination and variability: A tutorial. *Clinical Biomechanics*, **19**, 415–28.
- Dai X, Müller HG and Yao F (2017) Optimal Bayes classifiers for functional data and density ratios. *Biometrika*, **79**, 545–60.
- Dauxois J, Pousse A and Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136–54.
- de Nord OE and Theobald EH (2005) Multilevel component analysis and multilevel PLS of chemical process data. *Journal of Chemometrics*, **19**, 301–07.
- Delaigle A and Hall P (2012a) Componentwise classification and clustering of functional data. *Biometrika*, **99**, 299–313.
- (2012b) Methodology and theory for partial least squares applied to functional data. *Annals of Statistics*, **40**, 322–52.
- (2012c) Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B*, **74**, 267–86.
- Deluzio K, Wyss U, Zee B, Costigan P and Serbie C (1997) Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. *Human Movement Science*, **16**, 201–17.
- Deville J (1974) Methodes statistiques et numeriques de l'analyse harmonique. *Annales De L'insee*, **123**, 3–101.
- Dona G, Preatoni E, Claudio Cobelli C, Rodano R and Harrison AJ (2009) Application of functional principal component analysis in race walking: An emerging methodology, sports biomechanics. *Sports Biomechanics*, **8**, 284–301.
- Donoghue OA, Harrison AJ, Coffey N and Hayes K (2008) Functional data analysis of running kinematics in chronic Achilles tendon injury. *Medicine and Science in Sport and Exercise*, **40**, 1323–35.
- Durban M, Harezlak J, Wand MP and Carroll RJ (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–67.
- Eilers PHC and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121.
- Escabias M, Aguilera AM and Aguilera-Morillo MC (2014) Functional PCA and base-line logit models. *Journal of Classification*, **31**, 296–324.
- Escabias M, Aguilera AM and Valderrama MJ (2004) Principal component estimation of functional logistic regression: Discussion of two different approaches. *Nonparametric Statistics*, **16**, 365–84.
- (2005) Modeling environmental data by functional principal component logistic regression. *Environmetrics*, **16**, 95–107.
- (2007) Functional PLS logit regression model. *Computational Statistics and Data Analysis*, **51**, 4891–4902.

- Ferraty F and Vieu P (2006) *Nonparametric Functional Data Analysis. Theory and Practice*. Berlin: Springer-Verlag.
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179–88.
- Frank I and Friedman JH (1989) Classification: Oldtimers and newcomers. *Journal of Chemometrics*, 3, 463–75.
- Friedman JH (1989) Regularized discriminant analysis. *Journal of the American Statistical Association*, 84, 165–75.
- Harrison AJ, Ryan W and Hayes K (2007) Functional data analysis of joint coordination in the development of vertical jump performance. *Sports Biomechanics*, 6, 196–211.
- Hastie T, Buja A and Tibshirani R (1995) Penalized discriminant analysis. *The Annals of Statistics*, 23, 73–102.
- Horvath L and Kokoszka P (2012) *Inference for Functional Data with Applications*. Berlin: Springer-Verlag.
- Indahl UG, Sahni NS, Kirkhus B and Naes T (1999) Multivariate strategies for classification based on NIR-spectra—with application to mayonnaise. *Chemometrics and Intelligent Laboratory Systems*, 49, 19–31.
- James, G. M. (2002). Generalized Linear Models with Functional Predictors. *Journal of the Royal Statistical Society. Series B*, 64, 411–32.
- James GM and Hastie TJ (2001) Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B*, 63, 533–50.
- Kokoszka P and Reimherr M (2017) *Introduction to Functional Data Analysis*. London: Chapman and Hall/CRC.
- Li H, van de Voort FR, Ismail AA, Sedman J, Cox R, Simard C and Buijs H (2000) Discrimination of edible oil products and quantitative determination of their iodine value by Fourier transform near-infrared spectroscopy. *Journal of the American Oil Chemists' Society*, 77, 29–36.
- Liquet B, Lê Cao KA, Hocini H and Thiebaut R (2012) A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics*, 13, 325.
- Milovanovic I and Popovic DB (2012) Principal component analysis of gait kinematics data in acute and chronic stroke patients. *Computational and Mathematical Methods in Medicine*, 2012, 1–8.
- Mkhadri A, Celeux G and Nasroallah A (1997) Regularization in discriminant analysis: An overview. *Computational Statistics and Data Analysis*, 23, 403–23.
- Ocaña FA, Aguilera AM and Escabias M (2007) Computational considerations in functional principal component analysis. *Computational Statistics*, 22, 449–65.
- Ocaña FA, Aguilera AM and Valderrama MJ (1999) Functional principal components analysis by choice of norm. *Journal of Multivariate Analysis*, 71, 262–76.
- Orantes-González E, Heredia-Jimenez J and Soto-Hermoso VM (2015) The effect of school trolley load on spatiotemporal gait parameters of children. *Gait and Posture*, 42, 390–93.
- O'Sullivan F (1986) A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1, 505–27.
- Perez-Enciso M and Tenenhaus M (2003) Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics*, 112, 581–92.
- Preda C and Saporta G (2005) PLS regression on a stochastic process. *Computational Statistics and Data Analysis*, 48, 149–58.
- Preda C, Saporta G and Leveder C (2007) PLS classification of functional data. *Computational Statistics*, 22, 223–35.
- R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL <http://www.R-project.org> (last accessed 29 August 2019).
- Ramsay JO and Silverman BW (1997) *Functional Data Analysis*. Berlin: Springer-Verlag.
- (2002) *Applied Functional Data Analysis: Methods and Case Studies*. Berlin: Springer-Verlag.

- Ramsay JO, Hooker G and Graves S (2009) *Functional Data Analysis with R and MATLAB*. Berlin: Springer-Verlag.
- Reiss PT and Ogden RT (2007) Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102**, 984–96.
- Rohart F, Gautier B, Singh A and Lê Cao KA (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, **13**, 1–19.
- Ryan W, Harrison AJ and Hayes K (2006) Functional data analysis in biomechanics: A case study of knee joint vertical jump kinematics. *Sports Biomechanics*, **5**, 121–38.
- Sadeghi H (2003) Local or global asymmetry in gait of people without impairments. *Gait and Posture*, **17**, 197–204.
- Sadeghi H, Allard P and Duhaime M (1997). Functional gait asymmetry in able-bodied subjects. *Human Movement Science*, **16**, 243–58.
- Sato T (1994). Application of principal-component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *Journal of the American Oil Chemists' Society*, **71**, 293–98.
- Schmidt J and Docherty S (2010) Comparison of the posture of school children carrying backpacks versus pulling them on trolleys. *Clinical Chiropractic*, **13**, 253–60.
- Shin H (2008) An extension of Fisher's discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, **99**, 1191–1216.
- Westerhuis JA, van Velzen EJ, Hoefsloot HC and Smilde AK (2010) Multivariate paired data analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics*, **6**, 119–28.