



Nonlinear mixed-effects scalar-on-function models and variable selection

Yafeng Cheng¹ · Jian Qing Shi² · Janet Eyre³

Received: 26 February 2018 / Accepted: 2 April 2019
© The Author(s) 2019

Abstract

This paper is motivated by our collaborative research and the aim is to model clinical assessments of upper limb function after stroke using 3D-position and 4D-orientation movement data. We present a new nonlinear mixed-effects scalar-on-function regression model with a Gaussian process prior focusing on the variable selection from a large number of candidates including both scalar and function variables. A novel variable selection algorithm has been developed, namely functional least angle regression. As it is essential for this algorithm, we studied the representation of functional variables with different methods and the correlation between a scalar and a group of mixed scalar and functional variables. We also propose a new stopping rule for practical use. This algorithm is efficient and accurate for both variable selection and parameter estimation even when the number of functional variables is very large and the variables are correlated. And thus the prediction provided by the algorithm is accurate. Our comprehensive simulation study showed that the method is superior to other existing variable selection methods. When the algorithm was applied to the analysis of the movement data, the use of the nonlinear random-effect model and the function variables significantly improved the prediction accuracy for the clinical assessment.

Keywords Canonical correlation · Functional least angle regression (fLARS) · Gaussian process prior · Movement data · Scalar-on-function regression · Variable selection

1 Introduction

Stroke has emerged as a major global health problem in terms of both death and major disability. Hemiparesis, a detrimental consequence that many stroke survivors face, is the partial paralysis of one side of the body that occurs due to the

brain injury. Studies have consistently demonstrated significant, therapy-induced improvements in upper limb function can be achieved, but only with intense, repetitive and challenging practice (Langhorne et al. 2009). Limited resources, specifically lack of therapist time, are the main barriers for stroke rehabilitation. In our collaborative research, a home-based rehabilitation system based on action-video games has been developed (Serradilla et al. 2014; Shi et al. 2013). This paper focuses on one of the key parts in the system: predicting patients recovery levels for remote assessment and monitoring using their movement data. An assessment game, including 38 movements, has been designed. Patients after stroke play the assessment game at their home without any supervision by therapists. The accuracy of the devices used in the study have been tested and validated (Serradilla et al. 2014; Shi et al. 2013). The signals are generally clear with a low level of noise. The position of the upper limb with respect to the reference point over time is represented by 3-dimensional data (denoted as position data). The orientation of the upper limb with respect to the reference direction over time is represented by 4-dimensional data in the format of the quaternion (denoted as orientation data). The position

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11222-019-09871-3>) contains supplementary material, which is available to authorized users.

✉ Jian Qing Shi
j.q.shi@ncl.ac.uk
Yafeng Cheng
yafeng.cheng@mcr-bsu.cam.ac.uk
Janet Eyre
janet.eyre@ncl.ac.uk

- ¹ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
- ² School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, UK
- ³ Institute of Neurosciences, Newcastle University, Newcastle upon Tyne, UK

and orientation data for each movement are recorded and transferred to the cloud. The data is then used to estimate the recovery level of the upper limbs for patients. The recovery curve for each patient is constructed and assessed by therapists. This enables therapists to monitor patients recovery and adjust therapy accordingly.

We collected data from 70 stroke survivors without significant cognitive or visual impairment. These patients had a wide range of levels of severity in their upper limb functions when they were involved in the analysis. The video game applied in the study, with wireless controllers and movement tracking, was new to all the patients in the study. Each of the patients had up to eight assessments in a three month period. The first four assessments were arranged weekly and the following were arranged fortnightly. For each patient, we obtained one record of game data and one record of clinical assessment data in each visit except for the first one, where we only observed the clinical assessment result as the baseline for that patient. Some patients had missed a few visits or have missing data on some visits. We used the complete data only in the analysis in this paper.

The clinical assessment used in the study is called Chedoke Arm and Hand Activity Inventory, or CAHAI for short (<http://www.cahai.ca/>). This assessment contains nine tasks based on daily activities, such as dial 999 and opens a jar. It gives one overall score ranging from 9 to 63, which summarizes the recovery level of a patient. Score 63 implies that the patient upper limb function is as good as a normal person. This assessment is a fully validated measure of upper limb functional ability (Barreca et al. 2005). Figure 1 shows the values of CAHAI, where each curve represents data from one patient. Fig. 1a is for the acute patients who had a stroke for less than one month at the time of the first visit and Fig. 1b is for the chronic patients who had a stroke for more than six months. The patients are split into two groups because their recovery rates are substantially different (Langhorne et al. 2009). It is clear to see that the acute patients have an increasing trend but with different patterns, i.e. there is strong heterogeneity between acute patients, while the trend for chronic patients is flat.

We use wireless devices to track patients movements in the formats of 3-D position and 4-D orientation. The movement data can be used to build models after necessary preprocessing, including calibration, segmentation, normalization, registration, and smoothing; see the details in Shi et al. (2013). After the preprocessing, these signals can be thought as functional variables observed on dense time points. Figure 2 shows 10 samples of six functional variables collected from one movement after the preprocessing.

The statistical challenge in the project is to build a predictive regression model to estimate the recovery level (CAHAI, a scalar response) of the upper limb function of the stroke patients using the clinical information and the game signal

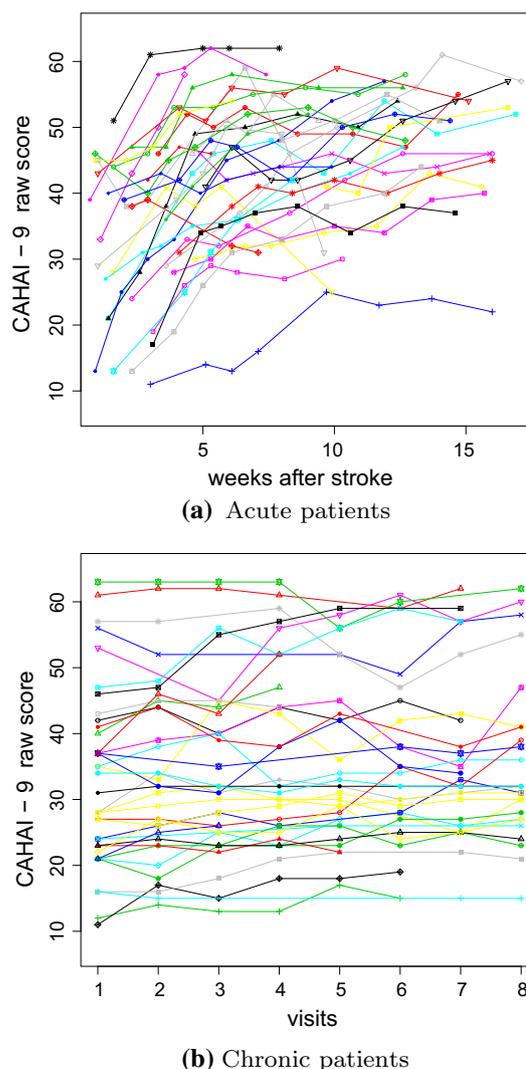


Fig. 1 Recovery level against assessment time [weeks in (a) and the visit time in (b)]

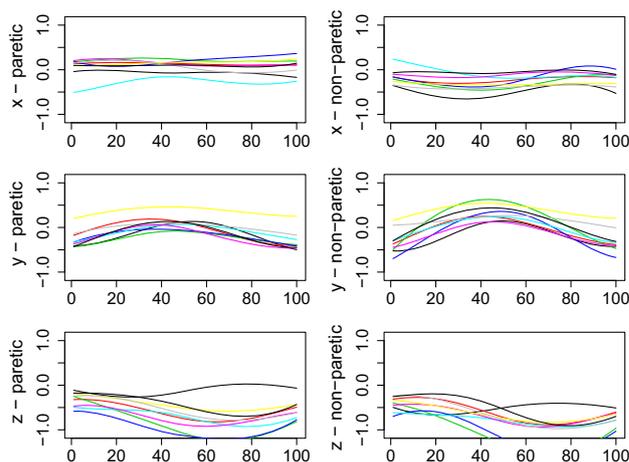


Fig. 2 10 samples of the data after preprocessing from movement Forward Roll

data. The clinical information is represented as a few scalar variables, and the game signal data can be considered as functional variables. We also use the kinematic summary statistics calculated from the signal data, representing the accuracy, synchrony, smoothness, and speed of the patients' movements; see the detail in Shi et al. (2013). The scalar variables are comprised of both kinematic summary statistics and patient-specific information such as the baseline assessment score and the number of weeks after a stroke at each assessment. Due to numerical unsuitability, we cannot use all the variables in the model. Although different varieties of variable selection techniques have been developed in the past several decades, it is still a difficult problem when it involves hundreds of mixed scalar and functional variables. Other problems need to address include the heterogeneity and nonlinearity in the data.

There are quite a few papers in the literature studied the model with a scalar response and mixed scalar and functional covariates, such as Ramsay and Silverman (2006), but mostly focused on single or small number of functional variables. Matsui and Konishi (2011), Gertheiss et al. (2013) among others proposed variable selection algorithms for functional linear regression models or functional generalized linear regression models based on the group variable selection methods with penalized likelihood such as group lasso, group SCAD, and group elastic net. Müller and Yao (2012) focused on the functional additive models, which allows multiple functional variables but did not discuss the problem of variable selection. Fan et al. (2015) proposed a variable selection algorithm for functional additive models by using group lasso. Their algorithm can handle both linear and nonlinear problems and can select from a large number of candidate variables. Collazos et al. (2016) proposed a method to provide p values in fitting the functional linear regression model with scalar response and functional covariates. They applied the method in the building of a new variable selection algorithm which has attractive theoretical properties. Most of the above algorithms use group variable selection with penalized likelihood. The computational cost becomes expensive and the estimation and selection accuracy become poor when we have a large number of mixed scalar and functional variables. We propose a new algorithm to address the problems.

Variable selection has also been discussed in other types of functional models, for example, Goldsmith et al. (2014) applied Ising prior to select informative pixel in a scalar-on-image regression model.

The problem of heterogeneity in functional regression model has been discussed by many researchers, for example, Morris and Carroll (2006), Scheipl et al. (2015), Zhu et al. (2012), Goldsmith et al. (2012), Gertheiss et al. (2013) and Cao et al. (2018). However, most of the models discussed linear mixed-effects only or target on regression with functional response. In this paper, we will introduce the nonlinear

random effect by using a Gaussian process prior. Gaussian process regression (GPR) is a Bayesian nonlinear nonparametric model using GP prior. We assume each of the patient's recovery level follows a patient-specific nonlinear model with a GP prior but share a common covariance structure for all the patients. GPR has been widely used in machine learning (Rasmussen and Williams 2006), statistics (Shi and Wang 2008; Shi and Choi 2011; Gramacy and Lian 2012; Wang and Shi 2014) and other areas.

The overall aim of this paper is to address the problems involved in a scalar-on-function regression model with a very large number of mixed scalar and functional variables, emerged from our collaborative project on estimating upper limb function. We propose a new efficient algorithm to select variables from a large number of candidates and propose to use GPR to model the nonlinear random-effect. The new variable selection method, called functional LARS or fLARS, is an extension of the least angle regression model (LARS) (Efron et al. 2003). As it is essential for this algorithm, we study the representation of functional variables with different methods and propose a new method based on Gaussian quadrature. The use of specifically designed stopping rule saves computational time as the number of candidate variables increases.

The remaining of the paper is organized as follows. The variable selection problem in a scalar-on-function model and the details of the new fLARS algorithm are discussed in Sect. 2. The idea and models to solve the problems of heterogeneity and nonlinearity are discussed in Sect. 3. The analysis of the movement data is given in Sect. 4. Finally, some concluding remarks are given in Sect. 5.

2 Variable selection in a mixed-effects scalar-on-function model

2.1 Variable selection in a scalar-on-function model

We start our discussion from the following scalar-on-function model:

$$y_{i,d} = \sum_{j=1}^J \int x_{i,d,j}(t) \beta_j(t) dt + \sum_{m=J+1}^{J+M} z_{i,d,m} \gamma_m + \epsilon_{i,d}, \tag{1}$$

where $\epsilon_{i,d} \sim N(0, \sigma^2)$. The scalar response $y_{i,d}$, in our real movement data, stands for the recovery level (CAHAI) of upper limbs for the d -th visit of the i -th patient, z_m is a scalar variable, and $x_j(t)$ is a functional variable. Note that the intercept is omitted by assuming all of the scalar variables are centered, and all of the functional variables have mean function $0(t)$. The parameters of interest are the fixed effect coefficients $\boldsymbol{\gamma}$ and functional coefficients $\beta(t)$.

The computational cost involved in variable selection mainly depends on the number of functional variables and the ways of representing the functional objects:

$$u = \int x(t)\beta(t)dt. \tag{2}$$

involved in (1).

Two of the most commonly used approximation methods are representative data points (RDP) (Leurgans et al. 1993) and basis functions (BF) (Ramsay and Silverman 2006). The former uses equally distributed dense data points to represent the functional objects, while the latter use known basis functions to represent the curves. Other methods, for example, functional principal component analysis and functional partial least squares (Reiss and Ogden 2007), are also popular. Equation (2) can be expressed by a unified formula $\int x(t)\beta(t)dt = \mathbf{x}W\tilde{C}_\beta^T$, where W could be different depending on different representing methods; see the details in Cheng (2016). Thus the problem of selecting functional variable $x(t)$ (i.e. if $\beta(t) \neq 0$) is converted to the selection of group of scalar variables involved in $\mathbf{x}W$ (i.e. if $\tilde{C}_\beta \neq 0$). This is naturally a group variable selection problem (Yuan and Lin 2006; Matsui and Konishi 2011). However, the computational cost increases and the performance deteriorates when the number of both scalar and functional variables increases.

To address the problems, we propose a new algorithm, functional least angle regression. It works efficiently for model (1) with a large number of candidate variables, either scalar or functional variables. We also propose to use Gaussian quadrature to approximate the integration. This improves the efficiency of fLARS further. This new representation method can be used as an alternative to the RDP and BF methods.

2.2 Functional least angle regression

The original LARS proposed by Efron et al. (2003) is an efficient iterative variable selection algorithm. It is able to give the outcome of other variable selection algorithms, such as lasso, with minor modifications. Later the versions in generalized linear regression and Cox proportional hazards models are proposed by Park and Hastie (2007). It is extended to select groups of variables with different dimensions by Yuan and Lin (2006). The core idea in group LARS is to do selection based on the orthogonal matrices, which are obtained by decomposing each of the groups of covariates, instead of the raw data before the selection of groups of variables. This can greatly reduce the computation time. However, the group LARS algorithm may not work in the functional case since the decomposition may fail or bring error into the estimation when the functional covariate is represented by a low-rank

matrix. We propose a new functional LARS algorithm for model (1).

The LARS algorithm can be summarized as the following. We find the most correlated candidate variable with the residual from the previous iteration, and then move the coefficient in the least square direction for the projection of the previous and newly selected variables, until a new candidate variable becomes as correlated with the current residual as the projection of the selected variables in the current model. The new variable is added to the regression equation and the process is repeated until no remain candidate variables.

There are two types of correlations required in the LARS algorithm. One is between a scalar variable, namely the residual from the last iteration, and the projection of all the selected variables; the other is between two scalar variables, namely the residual and one of the candidate variables. When we have functional variables in the candidates, with the same spirit, we can project the functional variable, or a group of mixed scalar and functional variables to a one-dimensional vector and use the projection to calculate the Pearson's correlation with the residual variable as in the LARS algorithm. The idea is similar to that used in the canonical correlation analysis (CCA). A functional version, functional canonical correlation analysis (FCCA), is given in Ramsay and Silverman (2006), where it is used to measure the correlation between two functional variables. We will modify FCCA to fit our algorithm. Moreover, we propose a new method to efficiently and accurately calculate the projection of functional variables by using Gaussian quadrature.

2.2.1 The notation and fLARS algorithm

By centering both response and covariates and omitting the intercept in (1), we have

$$y = \sum_{j=1}^J \int x_j(t)\beta_j(t)dt + \sum_{m=J+1}^{J+M} z_m\gamma_m + \epsilon, \tag{3}$$

where all the notations are the same as before. We define A as the set of indices of the selected variables and A^c as the set for the remaining candidate variables. Suppose that the residual obtained from the previous iteration is $r^{(k)}$, where k is the index of the current iteration.

The algorithm starts with $k = 1$, $r^{(1)} = y$, $\beta(t) = \mathbf{0}$, $\gamma = \mathbf{0}$, $A = \emptyset$ and $A^c = \{1, \dots, J, \dots, M + J\}$. The first selection is based on the correlation between $r^{(1)}$ and $(\mathbf{x}(t), \mathbf{z})$. The variable that has the largest squared correlation with $r^{(1)}$ is selected. After the first variable is selected, we carry out the following steps:

1. Define the direction $u^{(k)}$ to move in by projecting the selected variables to the current residual:

$$u^{(k)} = \frac{\sum_j \int x_j(t)\beta_j^{(k)}(t)dt + \sum_m z_m\gamma_m}{sd(\sum_j \int x_j(t)\beta_j^{(k)}(t)dt + \sum_m z_m\gamma_m)}, \tag{4}$$

where $j, m \in A$. The direction of the parameters is estimated in this step.

- For each remaining variable in A^c , depending on the type of the variable, we compute α_l using either

$$\begin{aligned} \text{Cor}(u^{(k)}, r^{(k)} - \alpha_l u^{(k)})^2 &= \rho(x_l(t), r^{(k)} - \alpha_l u^{(k)})^2 \text{ or} \\ \text{Cor}(u^{(k)}, r^{(k)} - \alpha_l u^{(k)})^2 &= \text{Cor}(z_l, r^{(k)} - \alpha_l u^{(k)})^2, \end{aligned} \tag{5}$$

for $l \in A^c$, where Cor is the Person’s correlation and ρ is the modified FCCA. α_l can be calculated from solving a quadratic function of α_l . The variable with the smallest positive α_l is selected into the regression equation. Denote the index of that variable as l^* and move l^* from A^c to the set A . The distance to move in the direction $u^{(k)}$ is $\alpha_{l^*}^{(k)}$.

- The new residual for next iteration is:

$$r^{(k+1)} = r^{(k)} - \alpha_{l^*}^{(k)} u^{(k)}. \tag{6}$$

The coefficient of a variable up to the K -th iteration is the sum of all the coefficients for that variable calculated up to and including the current iteration.

Like LARS, fLARS can be modified to remove variables selected in the previous iterations if they become less important. Such a modification on LARS leads to a faster way to get lasso solution than the original shooting algorithm. We apply a similar logic here. We measure the contribution of a variable by calculating the variance of its projection, e.g., $\text{Var}(\int x(t)\beta(t)dt)$. The variable is removed when the total variance is reduced since it indicates little contribution to the total variance. This criterion can be transformed to the following two conditions:

$$\begin{aligned} \text{Var}\left(\int x_j(t)\beta_j^{(k)}(t)\right) &< \max\left(\text{Var}\left(\int x_j(t)\beta_j^{(k^*)}(t)\right)\right) \\ \text{Var}\left(\int x_j(t)\beta_j^{(k)}(t)\right) &\leq \kappa \text{Var}(y), \end{aligned}$$

for $k^* \in \{1, \dots, k-1\}$. Here κ is a threshold. In our simulation study, 5% is used. Different thresholds can be tested in the real situation by methods such as k -fold cross validation. The first condition is to avoid removing newly selected variables, since the contribution of the newly selected variables is usually small.

2.2.2 Representation of a functional object using Gaussian quadrature

In addition to RDP and BF, we propose to use Gaussian quadrature (GQ) to approximate functional object, or the integration in (2). Since GQ uses a small set of discrete data points it can be thought of as an extension of the RDP method. The advantage of using GQ method is its efficiency compared to the original RDP method. Depending on the number of points used, the calculation could also be faster than that using BF while giving similar estimation accuracy.

The basic formula of GQ is $\int_{-1}^1 f(t)dt \approx \sum_{q=1}^Q w_q f(t_q)$, where Q is the number of abscissas and the integration interval $[-1, 1]$ is specific to some GQ, for example, Gauss–Legendre or Chebyshev–Gauss. Other GQ solutions may have different polynomial functions and intervals for integration and therefore different weights and abscissae. We use Gauss–Legendre in this paper. By using Gaussian quadrature, the integration (2) can be written as:

$$\begin{aligned} u &= \frac{1}{2} \int_{-1}^1 x(t)\beta(t)dt = \frac{1}{2} \int_{-1}^1 f(t)dt \\ &= \frac{1}{2} \sum_{q=1}^Q w_q f(t_q) = \mathbf{x}W_{GQ}\tilde{\beta}_{GQ}, \end{aligned}$$

where W_{GQ} is a diagonal matrix with diagonal entries having weights w_q ’s at the points closest to the abscissas, and 0 everywhere else, similarly for $\tilde{\beta}_{GQ}$. The accuracy of the integration can be improved by increasing the number of abscissas Q .

Gaussian quadrature uses prefixed abscissas. This method can be improved if the abscissas are chosen based on the information of the functional variables, or even based on the relationship between the functional variable and the response variable.

2.2.3 Modified functional canonical correlation

For model (3), we need to consider the correlation between two scalar variables, between one scalar variable and one functional variable, and between one scalar variable and a group of mixed scalar and functional variables. The correlation between two scalar variables is just the Person’s correlation. The correlation between one scalar variable and one functional variable can be obtained by a modified FCCA. Original FCCA has been studied by several researchers, for example, Leurgans et al. (1993) used RDP representation with constraints for smoothness on ‘curve data’; Ramsay and Silverman (2006) applied a roughness penalty with BF representation; and He et al. (2010) combined functional principal component analysis and canonical correlation analysis for a function-on-scalar regression model.

Let us denote the scalar variable as y and a functional variable as $x(t)$. With a roughness penalty controlling the smoothness and a ridge penalty controlling the numerical stability (Simon and Tibshirani 2012), we can define the canonical correlation between them as

$$\rho(x(t), y) = \text{Max}_{b(t), a} \frac{\text{Cov}(\int x(t)b(t)dt, ay)}{\sqrt{[\text{Var}(\int x(t)b(t)dt) + \text{PEN}][\text{Var}(ay)]}}, \tag{7}$$

where PEN is the penalty function, defined as $\text{PEN} = \lambda_1 \int [b''(t)]^2 dt + \lambda_2 \int [b(t)]^2 dt$ with tuning parameters λ_1 and λ_2 . From Sect. 2.1, by using any of RDP, BF or GQ to represent the integration in (2), we have

$$\begin{aligned} \int x(t)b(t)dt &\approx xW\tilde{C}_b, \\ \int [b''(t)]^2 dt &\approx \tilde{C}_b^T W_1 \tilde{C}_b, \\ \int [b(t)]^2 dt &\approx \tilde{C}_b^T W_2 \tilde{C}_b. \end{aligned}$$

Assuming the tuning parameters are known, we have the following solution of the optimization:

squared correlation: $\rho^2 = \frac{V_{x,y}^T P_{x,x}^{-1} V_{x,y}}{V_y}$ (8)

coefficients for $x(t)$: $\tilde{C}_b = \frac{P_{x,x}^{-1} V_{x,y}}{\rho \|y\|_2}$, (9)

coefficients for y : $a = 1/\text{sd}(y)$ (10)

where $V_{x,y} = (\mathbf{x}W)^T y$, $P_{x,x} = W^T \mathbf{x}^T \mathbf{x}W + \lambda_1 W_1 + \lambda_2 W_2$ and $V_y = y^T y$.

The correlation between one scalar variable and a group of mixed scalar and functional variables can be easily extended from Eqs. (8) and (9) by replacing matrices $P_{x,x}$ and $V_{x,y}$ with block matrices. For $P_{x,x}$, the penalty functions are applied on the diagonal blocks related to the functional variables.

The value of the tuning parameters greatly affects the outcome. We used generalized cross-validation (GCV) and cross-validation for λ_1 and λ_2 , respectively.

In the k -th iteration, we obtain one coefficient based on (9) with respect to the current residual $r^{(k)}$. Simultaneously we get the distance $\alpha^{(k)}$ to move on the direction unit vector from (6). The regression coefficient $\tilde{\beta}^{(k)}$ in the k -th iteration is

$$\tilde{\beta}^{(k)} = \alpha^{(k)} \frac{P_{x,x}^{-1} V_{x,r^{(k)}}}{\rho \|r^{(k)}\|_2} \Big/ \text{sd} \left(\frac{P_{x,x}^{-1} V_{x,r^{(k)}}}{\rho \|r^{(k)}\|_2} \right).$$

When the fLARS stops at iteration K , we can have an estimation of the the final regression coefficient:

$$\hat{\beta} = \sum_{k=1}^K \tilde{\beta}^{(k)}. \tag{11}$$

2.2.4 The stopping rule

Practically, we can always use cross-validation to find the optimal stopping point, but it is very time-consuming. Mallows's C_p -type criteria have been used in the LARS and group LARS. Other measures, including Akaike information criterion (AIC), Bayesian information criterion (BIC) and adjusted R^2 coefficient, can also be used. However, these criteria cannot be used in the fLARS algorithm as the degrees of freedom is not correlated with the number of variables in the model. We propose a new stopping rule in this section.

Intuitively, the algorithm can stop when the newly selected variables can explain little variation in the current residual and the remaining variables are not informative with respect to the current residual. We build our stopping rule based on these two aspects.

The variation explained by the current variables in iteration k is reflected by $\alpha^{(k)} u^{(k)}$ from (6). As the direction vectors $u^{(k)}$ is a centred normalized vector, $\alpha^{(k)}$ can be written as: $\alpha^{(k)} \text{sd}(u^{(k)}) = \text{sd}(\alpha^{(k)} u^{(k)})$. Thus $\alpha^{(k)}$ can represent the amount of variation explained in the k -th iteration. A small $\alpha^{(k)}$ indicates that the amount of variation in the current residual explained by the selected variables is small.

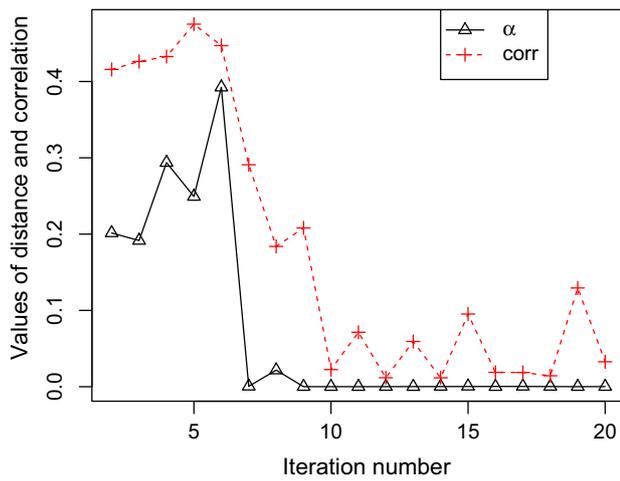
The level of informativeness of the remaining variables can be represented by the correlations. In each iteration, we selected the variable that is most correlated with the current residual. A small correlation ρ_k indicates the newly selected variable is not informative, and thus the remaining candidates are even less informative.

We combine the above two quantities and define 'correlation times distance' (CD) as:

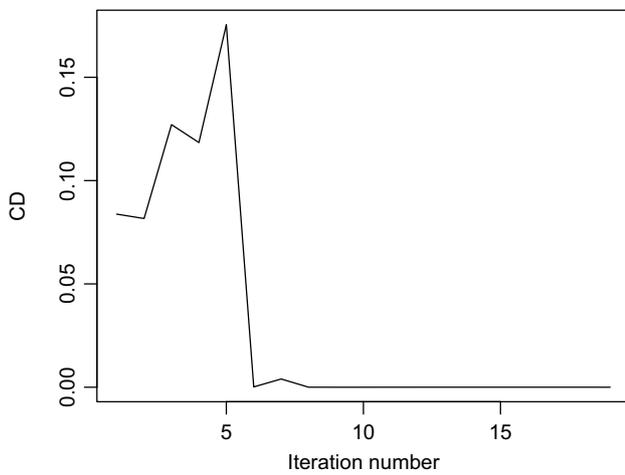
$$\text{CD}_k = \rho_{k-1} \times \alpha_k,$$

which is calculated after the k -th iteration and a new variable is selected. The algorithm stops when CD_k reduces below a threshold. One could simply use the minimum in practice. Such stopping point normally appears after a sudden dip of CD.

Figure 3 illustrates the changes of α and the correlations against the iteration number in an example data set from the simulation study. The plots are drawn based on a model with 100 candidate variables and six true variables. Based on the plot, the distance α reduces to almost 0 after the sixth iteration, and the correlation starts to reduce markedly. The first six selections include all six true variables. And a similar conclusion can be drawn from Fig. 3b.



(a) Change of α and ρ^* .



(b) Change of 'CD'.

Fig. 3 Illustration of the stopping rule based on a simulated data with 100 candidate variables and six true variables

2.3 Simulation study

We consider three scenarios:

- S1 has 12 candidate variables: seven functional and five scalar variables. The signal-noise-ratios (snr) tested are 10 and 2 respectively. To make the simulation more realistic, we introduced correlation between those variables, so that all the variables are correlated, and a few of the variables in the true model are highly correlated with a few that are not in the model.
- S2 has 100 candidate variables: 50 functional and 50 scalar variables. The signal-noise-ratios tested are 10 and 2 respectively.
- S3 has 12 candidate variables: seven functional and five scalar variables. The signal-noise-ratio tested is 2. We added a proportion of nonlinearity generated from the

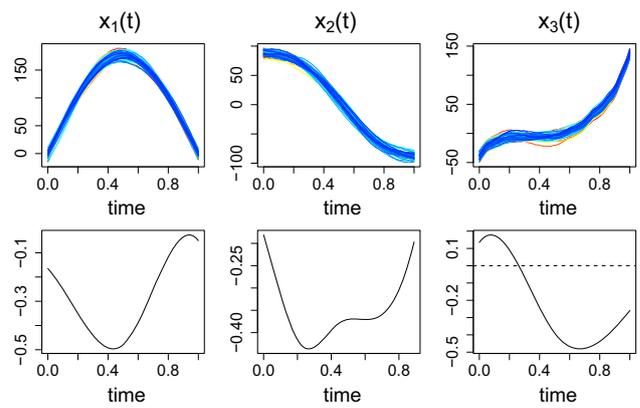


Fig. 4 The true values of the functional coefficients

sine function on one of the scalar variables in the true model. The standard deviation of the nonlinear part is 50% of that from the linear part.

As an illustrative example, one set of functional variables and the corresponding functional coefficients are shown in Fig. 4.

For Scenario 1 and 2, fLARS with three representation methods are tested, where we use 100 equal-distanced points in RDP method, 18 basis functions in BF method with order 6, and 18 points Gauss–Legendre quadrature in GQ method. In all cases, the smoothing parameter is found from 41 candidate values. As comparison, we also consider the group lasso method with roughness penalty (GLP), using 18 B-spline basis functions, 40 candidate smoothing parameters and 15 candidate hyper parameters for the L_1 penalty, and the one without roughness penalty (GLB), using 9 B-spline basis functions and 40 candidate hyper parameters for the L_1 penalty. For Scenario 3, we compare the fLARS algorithm with the model similar to Eq. (12) from next section: fit a Gaussian process model first and use the residual to fit the fLARS algorithm (denoted by GP + flars). This is to investigate the performance of the algorithm when a nonlinear relationship exists in the data. We choose such hyper-parameter setting to ensure that the performance of the models is close to their optimal. Some details of the GLP is included in the supplementary material.

The results for three scenarios are summarized in Table 1 based on 360 replications. The prediction accuracy is represented by root-mean-square-error (RMSE) between the prediction and its simulated true observation. The selection accuracy is jointly represented by the number of true positive (True +) selections and false positive (False +) selections. The standard deviation of the metrics is also included. The computational time for the fLARS is the one after the algorithm reaches the optimal stopping point for one replication,

Table 1 Summary of the simulation study in both scenarios

	RMSE (SD)	True + (SD)	False + (SD)	Time (s)
Scenario 1 (snr = 10)				
Flars (RDP)	0.058 (0.009)	5.933 (0.075)	0.084 (0.278)	1.83
Flars (BF)	0.060 (0.033)	5.888 (0.278)	0.095 (0.294)	0.34
Flars (GQ)	0.063 (0.015)	5.916 (0.105)	0.101 (0.302)	0.22
GLB	0.115 (0.024)	6 (0.000)	1.581 (0.685)	0.37
GLP	0.059 (0.008)	6 (0.000)	3.235 (1.050)	51.98
Scenario 1 (snr = 2)				
Flars (RDP)	0.310 (0.077)	5.051 (1.472)	0.128 (0.376)	1.306
Flars (BF)	0.299 (0.074)	5.300 (1.308)	0.168 (0.494)	0.260
Flars (GQ)	0.313 (0.081)	5.066 (1.530)	0.242 (0.653)	0.160
GLB	0.279 (0.035)	6.000 (0.000)	2.560 (0.877)	0.290
GLP	0.275 (0.036)	6.000 (0.000)	4.568 (1.226)	42.881
Scenario 2 (snr = 10)				
Flars (RDP)	0.062 (0.029)	2.987 (0.260)	3.132 (0.355)	8.894
Flars (BF)	0.064 (0.038)	2.965 (0.331)	3.210 (0.434)	3.027
Flars (GQ)	0.071 (0.031)	2.989 (0.285)	3.461 (0.520)	1.938
GLB	0.364 (0.071)	0.286 (0.876)	18.129 (8.585)	3.205
GLP	0.101 (0.026)	3.307 (0.538)	42.981 (4.719)	280.785
Scenario 2 (snr = 2)				
Flars (RDP)	0.348 (0.096)	1.819 (1.134)	2.768 (0.909)	7.558
Flars (BF)	0.342 (0.085)	2.040 (1.063)	2.992 (1.081)	2.532
Flars (GQ)	0.356 (0.092)	1.935 (1.143)	3.263 (1.372)	1.699
GLB	0.459 (0.061)	0.014 (0.192)	14.635 (6.974)	3.914
GLP	0.380 (0.064)	3.305 (0.840)	43.523 (7.884)	566.115
Scenario 3				
GP + flars (RDP)	0.430 (0.198)	3.970 (1.198)	0.028 (0.229)	2.139
GP + flars (BF)	0.433 (0.198)	4.135 (1.141)	0.180 (0.779)	0.423
GP + flars (GQ)	0.437 (0.196)	3.970 (1.217)	0.152 (0.603)	0.272
Flars (RDP)	0.456 (0.078)	4.046 (1.602)	0.091 (0.380)	2.702
Flars (BF)	0.454 (0.077)	4.360 (1.588)	0.404 (1.111)	0.515
Flars (GQ)	0.464 (0.081)	4.008 (1.554)	0.251 (0.769)	0.323

and for the group lasso versions are the time that the best tuning parameter(s) are selected.

The prediction accuracy, in terms of RMSE, from the fLARS algorithm is similar to that from the GLP method, while in most of the cases, the prediction accuracy from GLB method is the lowest compare to the other tested algorithms. When a nonlinear term is included, GP + flars (RDP) shows slightly better performance than the others.

The fLARS algorithm has a similar performance on the number of true positive selection compared to that from GLB and GLP. However, the number of false positive from fLARS algorithm is much smaller than that from GLB and GLP. When a nonlinear term is included, GP + flars (RDP) shows similar performance on the selection accuracy.

fLARS with RDP is slower compare to fLARS with other representation methods due to the high dimensionality of the

design matrices. fLARS with GQ method performed slightly worse due to the error introduced by Gaussian quadrature, but it has the fastest speed and can be improved by increasing the number of abscissas. GLB generally has similar speed like that from fLARS, while GLP is very slow due to the addition hyper-parameter to be selected by cross-validation.

It is worth noting that the space of the hyper-parameters where the grid-search/cross-validation is carried out has a noticeable impact on the corresponding model performance. fLARS uses CV on the hyper-parameter for the ridge penalty and GCV on hyper-parameter for the roughness penalty. The ridge penalty is helpful in reducing numerical uncertainty while has little impact on the final results. The GCV is efficient so that we can search through many candidate values with low cost. We use similar candidate smoothing parameters to set the searching space for GLP.

3 A nonlinear mixed-effects model

To address the problems of heterogeneity and nonlinearity, we propose to use the following model.

$$y_{i,d} = \sum_{j=1}^J \int x_{i,d,j}(t) \beta_j(t) dt + \sum_{m=J+1}^{J+M} z_{i,d,m} \gamma_m + g(\phi_{i,d}) + \epsilon_{i,d}$$

$$g(\phi_i) \sim GP(0, \kappa(\phi_{i,d}, \phi_{i,d'}; \theta)) \quad \epsilon_{i,d} \sim N(0, \sigma^2), \tag{12}$$

where the notations are the same as defined in (1) with $g(\phi_i)$ to be a nonlinear function. We use a non-parametric Bayesian approach with GP prior to fit $g(\phi_i)$, with hyper-parameter θ . Note that the intercept is omitted by assuming all of the variables are centred as before. As argued in Shi et al. (2012), Bayesian nonparametric GPR model is a good candidate to model nonlinear random-effect.

A general form can be written as $y = f(\mathbf{z}, \mathbf{x}(t)) + g(\phi) + \epsilon$, where $f(\cdot)$ is the fixed-effects part and $g(\cdot)$ is the random-effect part.

By ignoring the random effect part, the parameters in the fixed-effects part are estimated from fLARS algorithm as a by product; see the discussion around (11) in Sect. 2.2.3. Once the parameters in the fixed-effects are estimated, we can estimate the hyper-parameters involved in the GP random-effect part:

$$r = y - \hat{f}(\mathbf{z}, \mathbf{x}(t)) = g(\phi) + \epsilon;$$

$$g(\phi) \sim GP(0, \kappa(\phi, \phi'; \theta)),$$

where $\hat{f}(\mathbf{z}, \mathbf{x}(t))$ is the fitted value, and $\kappa(\cdot, \cdot; \theta)$ is a covariance kernel with hyper-parameters θ . Taking squared exponential kernel as an example, the covariance between sample ϕ and ϕ' is

$$\kappa(\phi, \phi'; \theta) = v_0 \exp\left(-\frac{1}{2} \sum_{h=1}^H v_h (\phi - \phi')^2\right),$$

with the hyper-parameters $\theta = (v_0, v_1, \dots, v_H)$.

Many different types of covariance kernels can be used. They are designed to fit in different situations; see details in Rasmussen and Williams (2006) and Shi and Choi (2011). We use the empirical Bayesian method to estimate the hyper-parameters.

The fitted value of $g(\phi)$ can be calculated once the hyper-parameters are estimated from the data collected from all subjects. Suppose data with total D visits are recorded for a particular subject, the $D \times D$ covariance matrix of $g(\phi)$ is denoted by \mathbf{c} , each element calculated from a covariance kernel with the estimated hyper-parameters. The mean and

the variance are given by

$$\hat{\mathbf{g}} = \mathbf{c}(\mathbf{c} + \sigma^2 \mathbf{I})^{-1} \mathbf{r} \quad \text{and} \quad \text{Var}(\hat{\mathbf{g}}) = \sigma^2 \mathbf{c}(\mathbf{c} + \sigma^2 \mathbf{I})^{-1}.$$

The above one-step method usually provides good results. More accurate estimates can be calculated by repeating the following iterative procedure until convergence.

1. Let $\tilde{y} = y - \hat{g}(\phi) = f(\mathbf{z}, \mathbf{x}(t)) + \epsilon$. Given the estimation of θ and $\hat{g}(\phi)$, this is a fixed-effects scalar-on-function regression model, we can estimate all the parameters using any methods discussed in Sect. 4.
2. Let $r = y - \hat{f}(\mathbf{z}, \mathbf{x}(t)) = g(\phi) + \epsilon$. Given the estimate of $\beta(t)$ and γ , we can update the estimation of $\hat{\theta}$ and calculate the fitted value of $\hat{g}(\phi)$ as discussed above.

The prediction for the fixed-effects part is calculated by using the estimated coefficients from fLARS. If we want to calculate the prediction of the random-effect part at a new point for a subject where we have already recorded data, i.e. forecast their future recovery level, the predictive mean and variance are given by (see e.g. Shi and Choi 2011):

$$\mathbb{E}(y^*) = \mathbf{c}^{*T} (\mathbf{c} + \sigma \mathbf{I})^{-1} \mathbf{r}, \quad r = y - \hat{f}$$

$$\text{Var}(y^*) = \kappa(\phi^*, \phi^*) - \mathbf{c}^{*T} (\mathbf{c} + \sigma^2 \mathbf{I})^{-1} \mathbf{c}^* + \sigma^2,$$

where ϕ^* is the covariate corresponding to the new data point, \mathbf{c} is the covariance matrix of $g(\phi)$ calculated from the D observed data points, \mathbf{c}^* is a $D \times 1$ vector with elements $\kappa(\phi^*, \phi_d)$, i.e., the covariance of $g(\phi)$ between the new data points and the observed data points.

For a new subject or patient, we can use the prediction calculated from the fixed-effects part and update it once we record data for the subject. An alternative way is to calculate the random-effect part using the following way: $\hat{y}^* = \sum_{i=1}^N w_i \hat{y}_i^*$, where \hat{y}_i^* is the prediction as if the new data point for the i -th subject, w_i is the weight which takes larger values if the new subject has similar conditions to the i -th subject (see Shi and Wang 2008).

4 Real data analysis

For the movement data, we first remove some movements due to the low rate of completion. We also remove some irrelevant variables, for example, all the 4-D orientation variables for the movements of outstretching, since we are only interested in the trajectory of 3-D position for this type of movements. We model the acute patients and chronic patients separately due to complete different recovery curve for those two groups. The rehabilitation for acute patients is the main interest of the project since the eventual recovery level for

each stroke survivor depends mainly on the performance in the first six months. For acute patients, there are 173 samples from 34 patients with 42 functional variables from 5 movements and 70 kinematic scalar variables from 17 movements; for chronic patients, there are 181 samples from 36 patients with 30 functional variables from 5 movements and 71 kinematic scalar variables from 17 movements. We also include the baseline measurement and the time since stroke or visit time and in the candidate variables.

Table 2 show the result by using model (1). We report the results using fLARS with RDP representation method. More specifically, each functional variable is represented in 100 dimension matrices and their coefficient are represented by basis functions. The results using other representation methods are almost the same. For example, the GQ representation method is computationally more efficient but gives slightly less accurate results. The detailed discussion can be found in Cheng (2016). The selected variables and their meanings are listed in Table 3. As a comparison, group lasso methods, GLP and GLB, are also considered, taking the same settings as specified in Sect. 2.3. All the methods selected baseline CAHAI and time in the results. For acute patients, four functional variable and six scalar variables from the movements, are selected using fLARS with RDP representation. This is based on the stopping rule proposed in the previous section. Six and seven variables from movement data are selected using group lasso without and with roughness penalty, respectively. However, none of the functional variables are selected. As pointed out in Yuan and Lin (2006), the dimension of the groups of variables will affect the selection using group lasso. When there are mixed scalar and functional variables, group lasso may fail since it may lack normalization between two types of variables. A more detailed discussion can be found in Chapter 5 of Cheng (2016).

Table 2 reports the average predictive root mean squared errors (RMSE) based on 400 replications of fivefold random cross-validation (sixfold for chronic patient data). The table shows that the model using the variables selected by fLARS outperform the one based on group lasso. The result for chronic patients is similar to that for acute patients but the improvement is not as significant as the one for the latter. This is because the change for each chronic patient is relatively flat and depends heavily on the initial value, and thus the contribution from other covariates is small. We can see that the group lasso based algorithms select no functional variables. This is an intrinsic drawback of group lasso like based functional variable selection algorithms when they are applied to problems with mixed scalar and functional variables.

Many different models have been investigated for the movement data. Among them, we found the Model (12) provides the best results for acute patients. In the nonlinear random effects part by GPR, we simply used the number

Table 2 Summary of variable selection from movement data for acute and chronic patient data

	RMSE	Functional	Scalar
Acute			
Flars (RDP)	6.267	4	6
GLB	6.886	0	6
GLP	6.844	0	7
Chronic			
Flars (RDP)	3.546	4	6
GLB	3.857	0	10
GLP	3.857	0	10

of weeks since stroke (for acute patients) or visit time (for chronic patients) as the covariate. The variables selected by fLARS or other methods are used in the fixed-effects part.

To show the performance of the models, we use random k -fold cross-validation to calculate the root mean squared error (RMSE) of predictions. In each replication, one- k th patients are randomly selected as a test group. The data of the other patients are used to train the model. The trained model is used to predict the recovery level at each visit for each patient in the test group. In other words, we can provide predictions for a patient without using any observed CAHAI from him/her except for the baseline CAHAI. The RMSE is calculated between the predictions and the observed values. The results presented in Table 4 are the average of RMSE based on 400 replications. As a comparison, we also report the results by using the fixed-effects models with the variables selected by fLARS (denoted by FE-flars), GLB (denoted by FE-GLB), GLP (denoted by FE-GLP). We can see the models ME-flars outperforms the others for both acute and chronic patient data. For acute patient data, the difference between the prediction performances of FE-flars and ME-flars are quite large. This indicates that the heterogeneity cannot be ignored for the acute patient and a nonlinear GPR random effects model can capture the patient-specific recovery rate. For chronic patient data, this difference is small. This indicates that heterogeneity among chronic patients is little.

The movement data discussed in this section has a complex structure. Many problems worth a further study, for example, a longitudinal study over the time after stroke for each patient using a function-on-function regression model and a study of missing data problems.

5 Conclusion and discussion

In this paper, we proposed a new variable selection algorithm, fLARS, for linear regression with scalar response and mixed functional and scalar covariates, motivated by the analysis

Table 3 Variables selected by fLARS for both acute and chronic patients with corresponding movement names

Patient type	Type	Variable name	Variable or movement meanings
Acute and Chronic	Scalar	base	Baseline CAHAI
	Scalar	rom_NP_LA37	LA37: mid-line to pronated
Acute only	Scalar	nwps	Number of weeks since stroke
	Scalar	sp_P_LA10	LA10:chopping
	Scalar	sp_P_LA05	LA05: forward roll
	Scalar	rom_NP_LA37	LA37: mid-line to pronated
	Scalar	rom_NP_LA03	LA03:arms outstretched
	Functional	LA09_lx	LA09: sawing
	Functional	LA10_rx	LA10: chopping
	Functional	LA19_rx	LA19: low to high crossover
	Functional	LA28_rqy	LA28: wrist mid-line to pronated
Chronic only	Scalar	visits	Number of visits
	Scalar	rom_P_LA21	LA21: to horizontal, one on top of the other
	Scalar	rom_P_LA41	LA41: arcs over head
	Scalar	rom_P_LA35	LA35: alternate chopping
	Functional	LA09_rx	LA09: sawing
	Functional	LA05_lx	LA05: forward roll
	Functional	LA03_lz	LA03: arms outstretched
	Functional	LA07_rx	LA07: low to high

In the variable names, LAXx stands for the movement index, sp stands for average speed, rom stands for range of maximum reach out of the movement, lx, rx and lz are left or right hand side of the x or z axis, and rqy is one of the quaternion axis from right hand side. The descriptions of the movement names explain the posture of the hands and the movements of the arms

Table 4 Model comparison using prediction RMSE based on 400 replications of random k-fold cross-validation, where k is 5 and 6 for acute and chronic patients respectively

		FE-GLB	FE-GLP	FE-flars	ME-flars
Acute	RMSE	6.886	6.844	6.267	5.653
	SD	0.201	0.202	0.212	0.179
Chronic	RMSE	3.857	3.857	3.546	3.448
	SD	0.108	0.108	0.091	0.089

of the movement data. A nonlinear mixed-effects model is proposed and applied to the real data. The application in the movement data shows that the functional variables in the model improve the performance of the prediction accuracy and the non-linear random effect from GP also improves the model performance by capturing the heterogeneity beyond the baseline difference between patients with multiple covariates. The movement data, especially the subset from the acute patients, are also suitable for the function-on-function regression model, as suggested by one of the referees. This is one of the further research directions using the data set that has been taking place. We found a slightly better outcome using function-on-function regression models than that of using scalar-on-function models.

The proposed fLARS algorithm is efficient and accurate. The correlation measure used in the algorithm is from a modified functional canonical correlation analysis. It gives a correlation and a projection simultaneously. Due to the dependency of the tuning parameters, conventional stopping rules fail in this algorithm. We proposed a new stopping rule. The simulation studies and the real data analysis show that the performance of this new algorithm together with the new stopping rule performs consistently well. The integration involved in the calculation for functional objects is carried out by three different ways: conventional RDP and BF and a new method based on Gaussian Quadrature. Compare to the conventional methods, the new method turns out to be comparable in accuracy and better in efficiency. Further research is justified to define the optimal representative data points for functional variables. In addition, as canonical correlation analysis is one of many correlation measures in the literature, there is potential to apply others, such as kernel canonical correlation, in the algorithm to capture non-linearity or further improve the efficiency of the calculation.

fLARS is an efficient algorithm to replace lasso or related algorithms when the latter is inefficient for problems involving a large number of mixed scalar and functional variables. Asymptotic theory of the selection procedure for fLARS is similar to LARS which can be found in e.g. Efron et al. (2003). More specifically, because the modified LARS can

produce lasso solutions, the asymptotic properties of LARS are the same as those of lasso. We proposed modification for fLARS under the same logic in Sect. 2.2.1 as that of LARS. Thus, we suggest there is a link between the modified fLARS and functional lasso, and therefore the asymptotic properties can be shared between the two. However, further research on the link and the asymptotic properties for functional lasso with different model settings are necessary.

An R package, named as fLARS, has been developed and is available in R CRAN.

Acknowledgements This publication presents independent research commissioned by the Health Innovation Challenge Fund (Grant No. HICF 1010 020), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Barreca, S.R., Stratford, P.W., Lambert, C.L., Masters, L.M., Streiner, D.L.: Test-retest reliability, validity, and sensitivity of the chedoke arm and hand activity inventory: a new measure of upper-limb function for survivors of stroke. *Arch. Phys. Med. Rehabil.* **86**, 1616–1622 (2005)
- Cao, C., Shi, J.Q., Lee, Y.: Robust functional regression model for marginal mean and subject-specific inferences. *Stat. Methods Med. Res.* **27**(11), 3236–3254 (2018)
- Cheng, Y.: Functional regression analysis and variable selection for motion data. Ph.D. thesis Newcastle University, UK (2016)
- Collazos, J.A., Dias, R., Zambom, A.Z.: Consistent variable selection for functional regression models. *J. Multivar. Anal.* **146**, 63–71 (2016)
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* **32**, 407–499 (2003)
- Fan, Y., James, G.M., Radchenko, P.: Functional additive regression. *Ann. Stat.* **43**(5), 2296–2325 (2015)
- Gertheiss, J., Maity, A., Staicu, A.M.: Variable selection in generalized functional linear models. *Stat* **2**, 86–101 (2013)
- Gertheiss, J., Goldsmith, J., Crainiceanu, C., Greven, S.: Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* **14**(3), 447–461 (2013)
- Goldsmith, J., Crainiceanu, C.M., Caffo, B., Reich, D.: Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Ser. C Appl. Stat.* **61**(3), 453–469 (2012)
- Goldsmith, J., Huang, L., Crainiceanu, C.M.: Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comput. Graph. Stat.* **23**, 46–64 (2014)
- Gramacy, R.B., Lian, H.: Gaussian process single-index models as emulators for computer experiments. *Technometrics* **54**, 30–41 (2012)
- He, G., Müller, H.G., Wang, J.L., Yang, W.: Functional linear regression via canonical analysis. *Bernoulli* **16**, 705–729 (2010)
- Langhorne, P., Coupar, F., Pollock, A.: Motor recovery after stroke: a systematic review. *Lancet Neurol.* **8**, 741–754 (2009)
- Leurgans, S.E., Moyeed, R.A., Silverman, B.W.: Canonical correlation analysis when the data are curves. *J. R. Stat. Soc. Ser. B* **55**, 725–740 (1993)
- Matsui, H., Konishi, S.: Variable selection for functional regression models via the ℓ_1 regularization. *Comput. Stat. Data Anal.* **55**, 3304–3310 (2011)
- Morris, J.S., Carroll, R.J.: Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B* **68**, 179–199 (2006)
- Müller, H.G., Yao, F.: Functional additive models. *J. Am. Stat. Assoc.* **103**(484), 1534–1544 (2012)
- Park, M.Y., Hastie, T.: L1-regularization path algorithm for generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69**, 659–677 (2007)
- Ramsay, J.O., Silverman, B.W.: *Functional Data Analysis*. Wiley Online Library, Hoboken (2006)
- Rasmussen, C.E., Williams, C.K.L.: *Gaussian Processes for Machine Learning*. The MIT Press, ISBN 0-262-18253-X (2006)
- Reiss, P.T., Ogden, R.T.: Functional principal component regression and functional partial least squares. *J. Am. Stat. Assoc.* **102**, 984–996 (2007)
- Scheipl, F., Staicu, A.M., Greven, S.: Functional additive mixed models. *J. Comput. Graph. Stat.* **24**, 477–501 (2015)
- Serradilla, J., Shi, J., Cheng, Y., Morgan, G., Lambden, C., Eyre, J.: Automatic assessment of upper limb function during play of the action video game, circus challenge: validity and sensitivity to change. *SeGAH* **2014**, 1–7 (2014)
- Shi, J., Cheng, Y., Serradilla, J., Morgan, G., Lambden, C., Ford, G.A., Price, C., Rodgers, H., Cassidy, T., Rochester, L.: Evaluating functional ability of upper limbs after stroke using video game data. In: *International Conference on Brain and Health Informatics*, pp. 181–192 (2013)
- Shi, J., Wang, B.: Curve prediction and clustering with mixtures of Gaussian process functional regression models. *Stat. Comput.* **18**, 267–283 (2008)
- Shi, J., Wang, B., Will, E., West, R.: Mixed-effects Gaussian process functional regression models with application to dose–response curve prediction. *Stat. Med.* **31**, 3165–3177 (2012)
- Shi, J.Q., Choi, T.: *Gaussian Process Regression Analysis for Functional Data*. CRC Press, Boca Raton (2011)
- Simon, N., Tibshirani, R.: Standardization and the group lasso penalty. *Stat. Sin.* **22**, 983–1001 (2012)
- Wang, B., Shi, J.Q.: Generalized Gaussian process regression model for non-Gaussian functional data. *J. Am. Stat. Assoc.* **109**, 1123–1133 (2014)
- Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* **68**, 49–67 (2006)
- Zhu, H., Brown, P.J., Morris, J.S.: Robust, adaptive functional regression in functional mixed model framework. *J. Am. Stat. Assoc.* **106**, 1167–1179 (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.