

Statistica Sinica Preprint No: SS-2023-0418

Title	Residual-based Alternative Partial Least Squares for Generalized Functional Linear Models
Manuscript ID	SS-2023-0418
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0418
Complete List of Authors	Yue Wang, Xiao Wang, Joseph G. Ibrahim and Hongtu Zhu
Corresponding Authors	Yue Wang
E-mails	yue.2.wang@cuanschutz.edu

Residual-based Alternative Partial Least Squares for Generalized Functional Linear Models

Yue Wang¹, Xiao Wang², Joseph G. Ibrahim³, and Hongtu Zhu³

¹*Department of Biostatistics and Informatics, Colorado School of Public Health*

²*Department of Statistics, Purdue University*

³*Department of Biostatistics, University of North Carolina at Chapel Hill*

Abstract: Many biomedical studies collect high-dimensional medical imaging data to identify biomarkers for the detection, diagnosis, and treatment of human diseases. Consequently, it is crucial to develop accurate models that can predict a wide range of clinical outcomes (both discrete and continuous) based on imaging data. By treating imaging predictors as functional data, we propose a residual-based alternative partial least squares (RAPLS) model for a broad class of generalized functional linear models that incorporate both functional and scalar covariates. Our RAPLS method extends the alternative partial least squares (APLS) algorithm iteratively to accommodate additional scalar covariates and non-continuous outcomes. We establish the convergence rate of the RAPLS estimator for the unknown slope function and, with an additional calibration step, we prove the asymptotic normality and efficiency of the calibrated RAPLS estimator for the scalar parameters. The effectiveness of the RAPLS algorithm is demonstrated through multiple simulation studies and an application predicting

Alzheimer's disease progression using neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

Key words and phrases: Alzheimer's Disease; dimension reduction; functional data; high-dimensional data.

1. Introduction

An important class of problems in medical imaging research is to identify imaging patterns associated with clinical outcomes of interest. Suppose that we observe a sample of n *i.i.d* subjects $\{y_i, x_i(\cdot), \mathbf{z}_i\}_{i=1}^n$ from the joint distribution of $(y, x(\cdot), \mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^q$ represents q -dimensional covariates, $y \in \mathbb{R}$ is a continuous or discrete outcome variable of interest, and $x(\cdot)$ represents the functional imaging data, observed at a set of grid points in a nondegenerate and compact space $\mathcal{S} \subset \mathbb{R}^K$ for some positive integer K . Functional regression models are useful tools to examine how these imaging predictors impact the outcome while adjusting for scalar covariates. An important class of functional regression models is the generalized functional linear model (GFLM, Müller and Stadtmüller (2005)). Specifically, for some underlying true parameters α_0^* , $\boldsymbol{\alpha}^*$ and $b^*(\cdot)$, let $\eta_i^* = \alpha_0^* + \mathbf{z}_i^\top \boldsymbol{\alpha}^* + \int_{\mathcal{S}} x_i(s) b^*(s) ds$ denote the true linear predictor for the i -th subject. Given η_i^* , GFLM specifies the distribution of y_i within the exponential family;

that is,

$$p(y_i, \eta_i^*) = h(y_i) \exp\{\eta_i^* T(y_i) - A(\eta_i^*)\} \quad (1.1)$$

for some $h(\cdot)$, $T(\cdot)$ and $A(\cdot)$. A special case of the GFLM (1.1) is the partially functional linear model (PFLM) where $y_i = \alpha_0^* + \mathbf{z}_i^T \boldsymbol{\alpha}^* + \int_{\mathcal{S}} x_i(s) b^*(s) ds + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2)$. When no scalar covariates \mathbf{z}_i exist, PFLM reduces to the simple functional linear model (SFLM), which have been extensively studied in the literature; see Ramsay and Silverman (2005); Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Morris (2015); Wang and Ruppert (2015); Kong et al. (2016) and the references therein. Extensions of (1.1) for predicting survival and longitudinal outcomes have also been developed in the literature (Gellar et al., 2015; Lee et al., 2015; Qu et al., 2016; Li and Luo, 2017; Wang et al., 2020).

Estimating the functional parameter $b^*(\cdot)$ in (1.1) requires dimensionality reduction due to the infinite dimension of $b^*(\cdot)$. The general approach is to approximate both $x_i(\cdot)$ and $b^*(\cdot)$ using a set of orthonormal basis functions, reducing the estimation of $b^*(\cdot)$ to estimating the coefficients associated with each basis function. Functional principal component analysis (FPCA) is a key technique for constructing data-driven basis functions; for an extensive review of FPCA, see Besse and Ramsay (1986); Ramsay and Dalzell (1991); Boente and Fraiman (2000); James et al. (2000) and Cai

and Hall (2006). The top principal components (PCs), which correspond to the largest eigenvalues and explain the most data variation, are then used to simplify GFLM into the classical generalized linear model (GLM). However, using principal components (PCs) in regression problems has two notable limitations. First, functional principal component analysis (FPCA) does not incorporate information from the outcome, meaning that the top PCs may not capture the relationship between the functional predictor and the outcome. This can result in suboptimal prediction or estimation accuracy (Cook, 2007). Second, when tail PCs are important for improving prediction, FPCA requires a large sample size to accurately estimate these PCs, which are associated with smaller eigenvalues (Jung and Maron, 2009). This poses a challenge in many medical imaging studies, where sample sizes are often constrained due to budgetary limitations.

To address the limitations of FPCA, partial least squares (PLS) methods have been developed for functional regression models. Unlike FPCA, PLS incorporates information from both the covariates and the outcome when constructing basis functions, eliminating the need to compute empirical eigenfunctions. Early PLS methods were designed for simple functional linear models (SFLMs, Preda and Saporta, 2005a), using an iterative procedure to estimate PLS components by maximizing the covariance be-

tween the outcome and a linear form of the functional predictor. Later, Delaigle and Hall (2012) introduced an alternative approach, known as alternative partial least squares (APLS), which provides a different set of basis functions that span the same space as the PLS basis but are computationally easier to obtain. Since then, theoretical advancements have also been made in functional PLS (FPLS) for SFLMs (Preda and Saporta, 2005a,b; Preda et al., 2005; Escabias et al., 2007; Reiss and Ogden, 2007; Krämer et al., 2008; Aguilera et al., 2010; Delaigle and Hall, 2012; Aguilera et al., 2016; Febrero-Bande et al., 2017). However, a key limitation of these PLS methods is their inherent design for linear relationships, making them less adaptable to nonlinear models. While extensions of PLS for nonlinear models exist, most involve repeatedly fitting GLMs, obtaining residuals, and using these residuals to construct PLS components (Bastien et al., 2005). Consequently, in nonlinear models, PLS components no longer maximize the covariance between the outcome and covariates, which undermines one of the central advantages of PLS in linear settings and leads to poorer estimation and prediction.

We propose a functional partial least squares (FPLS) method, called residual-based alternative partial least squares (RAPLS), for estimating the generalized functional linear model (GFLM) in (1.1), which fundamentally

differs from existing nonlinear PLS methods. RAPLS extends the alternative partial least squares (APLS) procedure from Delaigle and Hall (2012) to nonlinear models by iteratively fitting reweighted functional linear models. This approach adapts the iteratively reweighted least squares (IRLS) method (Green, 1984) to functional regression models and leverages the computational efficiency of the APLS approach. We also realize the gap that the existing theory for PLS has primarily been developed in linear settings, but theoretical justifications for nonlinear FPLS procedures are scarce. We bridge this gap by establishing the theoretical properties of the proposed RAPLS algorithm. Specifically, we establish the convergence rate of the RAPLS estimate for $b^*(\cdot)$, forming the foundation for proving the consistency of RAPLS estimates. We allow the number of components used in the RAPLS algorithm to diverge with the sample size, reflecting the infinite-dimensional nature of functional data. Moreover, we develop a calibrated estimator for the scalar covariate that is asymptotically normal and efficient. Finally, we compare the finite-sample performance of the RAPLS algorithm with multiple existing methods using simulated data sets and an application focused on predicting the progression of Alzheimer's Disease (AD).

Throughout the paper, for any vector $\mathbf{v} \in \mathbb{R}^d$, we use v_j to denote the j -

th element of \mathbf{v} for $j = 1, \dots, d$. For any matrix $\mathbf{M} \in \mathbb{R}^{n \times d}$, let \mathbf{m}_j and m_{ij} denote the j -th column and (i, j) -th entry of \mathbf{M} , respectively for $i = 1, \dots, n$ and $j = 1, \dots, d$. For ease of notation, we use a single notation $\|\cdot\|$ to denote the ℓ_2 -norm for vectors, matrices, functions, operators, and kernels. Specifically, let $\|\mathbf{v}\|^2 = \sum_{j=1}^d v_j^2$ and $\|\mathbf{M}\| = \sup_{\|\mathbf{v}\|=1} \|\mathbf{M}\mathbf{v}\|$. For any square-integrable function $f(\cdot)$, we let $\|f\|^2 = \int f^2(s)ds$. For any positive semi-definite kernel function $K(\cdot, \cdot)$, we let $\|K\| = \sup_{\|f\|=1} \|K(f)\|$. Also, we denote $\|K\|_F^2 = \iint K(s, t)^2 ds dt$. We use \mathbf{I}_d and $\mathbf{1}_d$ to denote the $d \times d$ identity matrix and the d -dimensional vector with all ones, respectively.

2. Population-level RAPLS for GFLMs

In this section, we introduce the population-level RAPLS for GFLM introduced in (1.1). Recall that the true linear predictor is $\eta^* = \alpha_0^* + \mathbf{z}^\top \boldsymbol{\alpha}^* + \int_{\mathcal{S}} x(s)b^*(s)ds$. Without loss of generality, we assume that $\mathbb{E}\{x(s)\} = 0$ for all $s \in \mathcal{S}$ and $\mathbb{E}(\mathbf{z}) = \mathbf{0}$. For any random variable $w \in \mathbb{R}$, define $m(\mathbf{z}, w) = w - \mathbf{z}^\top \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1} \mathbb{E}(\mathbf{z}w)$, where $\mathbf{z}^{\otimes 2} = \mathbf{z}\mathbf{z}^\top$. Since $\mathbb{E}(\mathbf{z}) = \mathbf{0}$, applying the function $m(\mathbf{z}, \cdot)$ to both sides of η^* yields

$$m(\mathbf{z}, \eta^*) = \int_{\mathcal{S}} m(\mathbf{z}, x(s)) b^*(s) ds. \quad (2.1)$$

One can view $m(\mathbf{z}, \eta^*)$ and $m(\mathbf{z}, x(s))$ as the *residual* of η^* and $x(s)$, respectively, after removing the effect of \mathbf{z} . We also define the covariance kernel of

the residual process $\{m(\mathbf{z}, x(s))\}_{s \in \mathcal{S}}$ by $\mathcal{C}(s, t) = \text{cov}(m(\mathbf{z}, x(s)), m(\mathbf{z}, x(t)))$ for any $s, t \in \mathcal{S}$. Note that \mathcal{C} does not depend on any specific realization of \mathbf{z} .

Consider first an oracular scenario where the true linear predictor η^* is known. According to Preda and Saporta (2005b), the FPLS basis functions for (2.1), denoted by $\{\rho_k(\cdot)\}_{k \geq 1}$, can be obtained by sequentially maximizing

$$\text{cov} \left(m(\mathbf{z}, \eta^*) - g_{k-1}(m(\mathbf{z}, x)), \int_{\mathcal{S}} m(\mathbf{z}, x(s)) \rho_k(s) ds \right) \quad (2.2)$$

subject to

$$\iint_{\mathcal{S}^2} \rho_j(s) \mathcal{C}(s, t) \rho_k(t) ds dt = 0 \quad \text{for } 1 \leq j \leq k-1 \quad \text{and} \quad \iint_{\mathcal{S}^2} \rho_k(s) \mathcal{C}(s, t) \rho_k(t) ds dt = 1,$$

where $g_k(m(\mathbf{z}, x)) = \int_{\mathcal{S}} m(\mathbf{z}, x(s)) \tilde{b}_k^*(s) ds$, and $\tilde{b}_k^*(s)$ is the orthogonal projection of $b^*(s)$ onto the space spanned by $\rho_1(\cdot), \dots, \rho_k(\cdot)$ under the ℓ_2 -norm.

According to (2.2), the first FPLS basis function $\rho_1(\cdot)$ can be obtained by maximizing $\iint_{\mathcal{S}^2} \rho_1(s) \mathcal{C}(s, t) b^*(t) ds dt$ subject to $\iint_{\mathcal{S}^2} \rho_1(s) \mathcal{C}(s, t) \rho_1(t) ds dt = 1$. Some functional calculus yields that $\rho_1(t)$ is proportional to $\mathcal{C}(b^*)(t)$, where \mathcal{C} is an operator that maps $b(t)$ to $\int_{\mathcal{S}} \mathcal{C}(s, t) b(s) ds$. Indeed, this result can be extended to all FPLS basis functions obtained by maximizing (2.2).

Lemma 1. *Given $\{\rho_i(\cdot)\}_{i=1}^j$, $\rho_{j+1}(\cdot)$ is unique up to sign change and de-*

terminated by

$$\rho_{j+1}(s) = c_0 \left[\mathcal{C} \left(b^*(s) - \sum_{l=1}^j \left\{ \int_{\mathcal{S}} b^*(t) \rho_l(t) dt \right\} \rho_l(s) \right) + \sum_{l=1}^j c_l \rho_l(s) \right],$$

where $\{\rho_j(\cdot)\}_{j \geq 1}$ are defined in (2.2), c_0 is uniquely defined up to a sign change, and c_l are obtained by solving the linear system of j equations

$$\iint_{\mathcal{S}^2} \rho_l(s) \rho_{j+1}(s) \mathcal{C}(s, t) ds dt = 0, \quad l = 1, \dots, j.$$

The proof is similar to that of Theorem 3.1 in Delaigle and Hall (2012), which is omitted. An immediate observation is that the space spanned by $\rho_1(s), \dots, \rho_p(s)$ is the same as that spanned by $\mathcal{C}(b^*), \dots, \mathcal{C}^p(b^*)$, where for $k > 1$, $\mathcal{C}^k(b^*) = \int_{\mathcal{S}} \mathcal{C}^{k-1}(b^*)(t) \mathcal{C}(s, t) dt$. We call $\mathcal{C}^j(b^*)$ the j -th population-level residual-based APLS (RAPLS) basis function for GFLM (1.1) for $j \geq 1$.

The next lemma, similar to Theorem 3.2 in Delaigle and Hall (2012), gives conditions under which any square-integrable function $b(\cdot)$ can be written in a linear form of the RAPLS basis functions $\{\mathcal{C}^j(b)\}_{j=1}^{\infty}$.

Lemma 2. *If $\mathcal{C}(s, t)$ is positive definite, then any square-integrable function $b(\cdot)$ can be written as $b = \sum_{j=1}^{\infty} \gamma_j \mathcal{C}^j(b)$ for some constants $\{\gamma_j\}_{j=1}^{\infty}$, which converges in terms of the ℓ_2 -norm.*

Letting $K(s, t) = \text{cov}(x(s), x(t))$, one can verify that $\mathcal{C}(s, t) = K(s, t) - \mathbb{E}\{\mathbf{z}^{\top} x(s)\} \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1} \mathbb{E}\{\mathbf{z} x(t)\}$. Thus, the positive definiteness of $\mathcal{C}(s, t)$

can be guaranteed if $K(s, t)$ is positive definite, and $x(s)$ and \mathbf{z} are not perfectly collinear, i.e., there do not exist $\nu_1(s), \dots, \nu_{q_z}(s)$ such that $x(s) = \sum_{l=1}^{q_z} z_l \nu_l(s)$ for any $s \in \mathcal{S}$. Lemma 2 indicates that in practice, we can truncate $\sum_{j=1}^{\infty} \gamma_j \mathcal{C}^j(b^*)$ to a finite number of components to approximate b^* . More specifically, for an arbitrary positive integer p , we can define $b_p^* = \sum_{j=1}^p \gamma_j^* \mathcal{C}^j(b^*)$ as the optimal p -th order RAPLS approximation of b^* . Thus, Lemma 2 guarantees that $\|b - b_p^*\| \rightarrow 0$ as $p \rightarrow \infty$, if $\mathcal{C}(s, t)$ is positive definite. This serves as the basis for the theoretical analyses in Section 4. The “optimal” coefficients $\gamma_1^*, \dots, \gamma_p^*$ minimize

$$\omega_p(\gamma_1, \dots, \gamma_p) = \mathbb{E} \left\{ m(\mathbf{z}, \eta^*) - \sum_{j=1}^p \gamma_j \int_{\mathcal{S}} m(\mathbf{z}, x(s)) \mathcal{C}^j(b^*)(s) ds \right\}^2.$$

Letting $\boldsymbol{\gamma}^* = (\gamma_1^*, \dots, \gamma_p^*)^\top$, we obtain $\boldsymbol{\gamma}^* = \mathbf{H}^{*-1} \boldsymbol{\beta}^*$, where $\mathbf{H}^* = (h_{jk}^*)_{j,k=1,\dots,p}$ and $\boldsymbol{\beta}^* = (\beta_j^*)_{j=1,\dots,p}$ with

$$h_{jk}^* = \int_{\mathcal{S}} \mathcal{C}^{j+1}(b^*)(s) \mathcal{C}^k(b^*)(s) ds \quad \text{and} \quad \beta_j^* = \int_{\mathcal{S}} \mathcal{C}(b^*)(s) \mathcal{C}^j(b^*)(s) ds. \quad (2.3)$$

3. Empirical RAPLS for GFLMs

We start with the PFLM $y_i = \mathbf{z}_i^\top \boldsymbol{\alpha}^* + \int_{\mathcal{S}} x_i(s) b^*(s) ds + \epsilon_i$ to gain some intuitions; here, we assumed $\mathbb{E}(y_i) = 0$ such that $\alpha_0 = 0$. The population RAPLS algorithm in Section 2 depends on two critical quantities: the residual covariance kernel $\mathcal{C}(s, t)$ and the true linear predictor η^* . Since $\mathbb{E}(\mathbf{z}_i) = \mathbf{0}$

and $\mathbb{E}\{x_i(s)\} = 0$ for all $s \in \mathcal{S}$, one can easily see that $\mathbb{E}\{m(\mathbf{z}_i, x_i(s))\} = 0$ for all $s \in \mathcal{S}$. Thus, $\mathcal{C}(s, t)$ can be naturally estimated by its empirical counterpart $\hat{\mathcal{C}}(s, t) = n^{-1} \sum_{i=1}^n m(\mathbf{z}_i, x_i(s))m(\mathbf{z}_i, x_i(t))$. While the true linear predictor η^* remains unknown in practice, it can be approximated by leveraging the relationship between the outcome y and η^* . For PFLM, since $\mathbb{E}(y | \eta^*) = \eta^*$, it is natural to use the outcome y to approximate η^* . In particular, since

$$\mathcal{C}(b^*)(s) = \int_{\mathcal{S}} \mathcal{C}(s, t)b^*(t)dt = \mathbb{E}\{m(\mathbf{z}, x(s))m(\mathbf{z}, y)\}, \quad (3.1)$$

a natural estimate of $\mathcal{C}(b^*)(s)$ is $\hat{\mathcal{C}}(b)(s) = n^{-1}\mathbf{X}(s)^\top \mathbf{M}_Z \mathbf{y}$, where $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\mathbf{X}(s) = (x_1(s), \dots, x_n(s))^\top$, $\mathbf{M}_Z = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top$ with $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top$. Subsequently, for $j \geq 1$, we have $\hat{\mathcal{C}}^{j+1}(b) = \int_{\mathcal{S}} \hat{\mathcal{C}}^j(b)(t)\hat{\mathcal{C}}(s, t)dt$. With these estimated basis functions, we estimate h_{jk}^* and β_j^* , respectively, with

$$\hat{h}_{jk} = \int_{\mathcal{S}} \hat{\mathcal{C}}^{j+1}(b)(s)\hat{\mathcal{C}}^k(b)(s)ds \quad \text{and} \quad \hat{\beta}_j = \int_{\mathcal{S}} \hat{\mathcal{C}}(b)(s)\hat{\mathcal{C}}^{j+1}(b)(s)ds.$$

Then, we calculate $\hat{\boldsymbol{\gamma}} = \hat{\mathbf{H}}^{-1}\hat{\boldsymbol{\beta}}$, where $\hat{\mathbf{H}} = (\hat{h}_{jk})$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_j)$. This leads to the RAPLS estimate of $b_p(s)$:

$$\hat{b}_p^*(s) = \sum_{j=1}^p \hat{\gamma}_j \hat{\mathcal{C}}^j(b)(s), \quad (3.2)$$

where $\hat{\gamma}_j$ is the j -th entry of $\hat{\gamma}$. Given $\hat{b}_p(s)$, we obtain a plug-in estimator of α^* :

$$\hat{\alpha}_p = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \left(\mathbf{y} - \int_{\mathcal{S}} \mathbf{X}(s) \hat{b}_p(s) ds \right). \quad (3.3)$$

Next, we extend this RAPLS procedure to nonlinear GFLMs where $\mathbb{E}(y | \eta^*) \neq \eta^*$. Our idea is to iteratively approximate the model (1.1) with a sequence of PFLMs. Let $\alpha_0^{(m)}$, $\alpha^{(m)}$, and $b^{(m)}(\cdot)$ denote the estimation of α_0^* , α^* , and $b^*(\cdot)$ at the m -th iteration, respectively. We approximate η_i^* with $\eta_i^{(m)} = \alpha_0^{(m)} + \mathbf{z}_i^\top \alpha^{(m)} + \int_{\mathcal{S}} x_i(s) b^{(m)}(s) ds$. Denote by

$$r(y, \eta) = \frac{\partial}{\partial \eta} \log p(y; \eta) = T(y) - \dot{A}(\eta), \quad w(\eta) = -\mathbb{E} \left\{ \frac{\partial^2}{\partial \eta^2} \log p(y; \eta) \mid x(\cdot), \mathbf{z} \right\} = \ddot{A}(\eta)$$

the score function and Fisher information with respect to η , respectively, where $\dot{A}(\eta) = dA(\eta)/d\eta$ and $\ddot{A}(\eta) = d^2 A(\eta)/d\eta^2$. For ease of notation, let $r_i^{(m)} = r(y_i, \eta_i^{(m)})$ and $w_i^{(m)} = w(\eta_i^{(m)})$. Motivated by the iteratively reweighted least squares (IRLS, Green (1984)), we define the pseudo-response at the m -th iteration

$$\tilde{y}_i^{(m)} = \eta_i^{(m)} + \{w_i^{(m)}\}^{-1} r_i^{(m)} \quad \text{for } i = 1, \dots, n.$$

Then, we can obtain $b^{(m+1)}(s)$ by applying the empirical RAPLS algorithm, which was presented above, to the PFLM $\tilde{y}_i^{(m)} = \alpha_0 + \mathbf{z}_i^\top \alpha + \int_{\mathcal{S}} x_i(s) b(s) ds + \epsilon_i^{(m)}$. We then obtain $\alpha_0^{(m+1)}$ and $\alpha^{(m+1)}$ by solving the score equation

$$\sum_{i=1}^n \mathbf{z}_i r(y_i, \alpha_0 + \mathbf{z}_i^\top \alpha + \int_{\mathcal{S}} x_i(s) b^{(m+1)}(s) ds) = \mathbf{0}. \quad (3.4)$$

We iterate this process until $|\alpha_0^{(m+1)} - \alpha_0^{(m)}| + \|\alpha^{(m+1)} - \alpha^{(m)}\| + \|b^{(m+1)}(s) - b^{(m)}(s)\| \leq \varrho$ for some pre-specified small ϱ , say 10^{-4} .

Remark 1. A similar FPLS algorithm for the functional joint model (FJM), called FJM-FPLS, has been used in our earlier work Wang et al. (2020). However, the population-level algorithm is missing in Wang et al. (2020), which hinders the theoretical justification of FJM-FPLS. In the next section, we will take advantage of the population-level RAPLS algorithm in Section 2 to study the asymptotic properties of the empirical RAPLS algorithm.

4. Theoretical Properties

In this section, we establish the asymptotic properties of the estimators for GFLM in the previous section.

Throughout the section, we consider the high-dimensional regime that the number of basis functions $p = p(n) \rightarrow \infty$ as $n \rightarrow \infty$. We make the following additional assumptions.

(A1) $\|b^*\| + \mathbb{E}(\|x\|^4) < \infty$ and $\lambda_{\min}(\mathbb{E}(\mathbf{z}^{\otimes 2})) > 0$, where $\lambda_{\min}(M)$ denotes the smallest eigenvalue of any symmetric matrix M .

(A2) $\|\mathcal{C}\| < 1$ and $p = O(n^{1/2})$ as $n \rightarrow \infty$.

Assumption (A1) is a standard regularity condition. The condition $\|\mathcal{C}\| < 1$ in Assumption (A2) holds by scaling the functional covariates $x(s)$ such that $\mathbb{E}\|x\|^4 < 1$. To see this, note that

$$\|K\|^2 = \iint_{S^2} \{\mathbb{E}x(s)x(t)\}^2 dsdt \leq \mathbb{E}\|x\|^4 < 1.$$

Then, since $\mathcal{C}(s, t) = K(s, t) - \mathbb{E}\{\mathbf{z}^\top x(s)\} \{\mathbb{E}(\mathbf{z}^{\otimes 2})\}^{-1} \mathbb{E}\{\mathbf{z}x(t)\}$, we get $\|\mathcal{C}\| \leq \|K\| < 1$.

Recall that $\alpha_0^{(m)}$, $\boldsymbol{\alpha}^{(m)}$, and $b^{(m)}(\cdot)$, respectively, denote the m -th RAPLS iterate of α_0^* , $\boldsymbol{\alpha}_p^*$, and $b^*(\cdot)$ under the GFLM (1.1) for $m \geq 0$. The following result shows that with deterministic initial values $\alpha_0^{(0)}$, $\boldsymbol{\alpha}^{(0)}$, $b^{(0)}(\cdot)$, the first-step iteration $b^{(1)}$ is not necessarily a consistent estimator of b_p^* .

Proposition 1. *Suppose Assumptions (A1) and (A2) hold. Then, as $n \rightarrow \infty$, we have $\|b^{(1)} - b_p^*\| = O_p(\lambda_p^{-3})$, where $\lambda_p = \lambda_{\min}(\mathbf{H}^*)$ is the smallest eigenvalue of \mathbf{H}^* , which is defined in (2.3).*

As $p \rightarrow \infty$ along with n , λ_p may converge to 0, indicating that $b^{(1)}$ may not be a consistent estimator. Thus, Proposition 1 necessitates the use of better initial values to guarantee the consistency of the RAPLS estimates. Letting $\alpha_{0,n}^{(0)}$, $\boldsymbol{\alpha}_n^{(0)}$, $b_n^{(0)}(\cdot)$ denote data-driven initial values with the convergence rate τ_n , we introduce the following assumption.

(A3) As $n \rightarrow \infty$, $|\alpha_{0,n}^{(0)} - \alpha_0^*| + \|\boldsymbol{\alpha}_n^{(0)} - \boldsymbol{\alpha}^*\| + \|b_n^{(0)} - b^*\| = O_p(\tau_n)$ and

$$\|b_p^* - b^*\| = O(\lambda_p^{-2}\tau_n^2), \text{ where } \tau_n = o(1) \text{ as } n \rightarrow \infty.$$

Initial values that satisfy Assumption (A3) may be obtained from existing methods. For example, Müller and Stadtmüller (2005) established \sqrt{n} -consistency for their functional estimator, constructed using arbitrary deterministic basis expansion. For PFLM, Kong et al. (2016) used FPCA and penalized approaches, and established consistency of their estimators where the convergence rate depends on the eigenstructure of the covariance kernel. Additionally, Lv et al. (2023) constructed an estimator within a reproducing kernel Hilbert space (RKHS) and established its convergence rate as a function of the decay rate of the eigenvalues of the “orthonormalized” covariance kernel. The theorems below require $\lambda_p^{-2}\tau_n = O(1)$ or $\lambda_p^{-2}\tau_n^2 = o(n^{-1/4})$. These two assumptions are generally weak as λ_p usually converges to 0 at a slow rate. Thus, all the above methods can be used to obtain initial values that satisfy the required assumptions.

While Assumption (A3) seems a strong assumption, we show in the following result that our RAPLS iterates can potentially achieve faster convergence rates than the initial values. Furthermore, in finite-sample applications, we will illustrate that in Sections 5 and 6, the RAPLS algorithm with simple deterministic initial values, such as $\alpha_0^{(0)} = 0$, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$, and $b^{(0)}(\cdot) \equiv 0$, or random initial values, can still outperform existing methods

in terms of estimation and prediction accuracy.

Theorem 1. *Suppose Assumptions (A1)–(A3) hold. If $\lambda_p^{-2}\tau_n = O(1)$ as $n \rightarrow \infty$, then for each $m \geq 1$, we have*

$$\|b^{(m)} - b^*\| = O_p(\lambda_p^{-2}\tau_n^2). \quad (4.1)$$

Eq. (4.1) implies that if $\lambda_p^{-2}\tau_n = O(1)$, then all RAPLS iterates are consistent estimators with the convergence rate $O_p(\tau_n)$. In particular, if $\lambda_p^{-2}\tau_n = o(1)$, $b^{(m)}$ converges to b^* at a faster rate than the initial value $b_n^{(0)}$ for all $m \geq 1$. When λ_p is a constant, Theorem 1 simplifies to $\|b^{(m)} - b^*\| = O_p(\tau_n^2)$. Since $\tau_n = o(1)$, this indicates $b^{(m)}$ has a much faster convergence rate than the initial values. This special situation occurs if there exists a constant M_λ such that either (i) $\theta_k = 0$ for $k > M_\lambda$, or (ii) $b_k^* = 0$ for $k > M_\lambda$. Scenario (i) arises if $x_i(\cdot)$ belongs to the M_λ -dimensional space $\mathcal{S}(M_\lambda)$ for each $i = 1, \dots, n$, while scenario (ii) holds if $b^*(\cdot)$ belongs to $\mathcal{S}(M_\lambda)$.

It is unfortunate that the “plug-in” estimators in (3.3) and (3.4) are not asymptotically normal nor efficient. This issue is common in semi-parametric inference of the low-dimensional parameter in the presence of high-dimensional nuisance parameters, where the plug-in approach causes a potential bias and fails to be efficient (Chernozhukov et al., 2018). To address this issue, we develop a calibration procedure. First, recall that $w(\eta)$ is

the information function with respect to the linear predictor η for the GFLM (1.1). We introduce a new functional operator $K_w : f \rightarrow \int K_w(s, t)f(s)ds$, where $K_w(s, t) = \mathbb{E}[w^*x(s)x(t)]$ with $w^* = w(\eta^*)$. Since $w^* > 0$, $K_w(s, t)$ represents a generalized covariance kernel, and thus is always positive definite. Then, we define $\zeta_{ik} = z_{ik} - \int_{\mathcal{S}} x_i(s)K_w^{-1}(\mathbb{E}\{w_i^*x_i(s)z_{ik}\})ds$ for $i = 1, \dots, n$ and $k = 1, \dots, q$, where K_w^{-1} is the inverse operator of K_w .

One can check that

$$\begin{aligned}\mathbb{E}\{w_i^*x_i(s)\zeta_{ik}\} &= \mathbb{E}\{w_i^*x_i(s)z_{ik}\} - \mathbb{E}[w_i^*x_i(s) \int_{\mathcal{S}} x_i(t)K_w^{-1}(\mathbb{E}\{w_i^*x_i(t)z_{ik}\})dt] \\ &= \mathbb{E}\{w_i^*x_i(s)z_{ik}\} - \int_{\mathcal{S}} \mathbb{E}\{w_i^*x_i(s)x_i(t)\}K_w^{-1}(\mathbb{E}\{w_i^*x_i(t)z_{ik}\})dt \\ &= \mathbb{E}\{w_i^*x_i(s)z_{ik}\} - K_w(K_w^{-1}(\mathbb{E}\{w_i^*x_i(s)z_{ik}\})) \\ &= 0.\end{aligned}$$

for each i, l and $s \in \mathcal{S}$. For ease of notation, denote $\theta_k^*(s) = K_w^{-1}(\mathbb{E}\{w_i^*x_i(s)z_{ik}\})$ for $k = 1, \dots, q$, and Let $\hat{\theta}_k(s)$ being an estimate of $\theta_k^*(s)$ that will be discussed below. We further calculate $\hat{\zeta}_{ik} = z_{ik} - \int_{\mathcal{S}} x_i(s)\hat{\theta}_k(s)ds$ for $i = 1, \dots, n$ and $k = 1, \dots, q$. Finally, the calibrated estimator of α is defined as

$$\hat{\alpha}_p^{\text{cal}} = \underset{\alpha}{\operatorname{argmax}} n^{-1} \sum_{i=1}^n \log p \left(y_i, \int_{\mathcal{S}} x_i(s) \left\{ \hat{b}_p(s) + \hat{\alpha}_p^{\text{T}} \hat{\theta}(s) \right\} ds + \hat{\alpha}_{0,p} + \hat{\zeta}_i^{\text{T}} \alpha \right), \quad (4.2)$$

where $\widehat{\boldsymbol{\zeta}}_i = (\widehat{\zeta}_{i1}, \dots, \widehat{\zeta}_{iq})^\top$, $\widehat{\boldsymbol{\theta}}(s) = (\widehat{\theta}_1(s), \dots, \widehat{\theta}_q(s))^\top$, and $\widehat{b}_p(s)$, $\widehat{\alpha}_{0,p}$, and $\widehat{\boldsymbol{\alpha}}_p$ are, respectively, the p -th order RAPLS estimates of $b^*(\cdot)$, α_0^* and $\boldsymbol{\alpha}^*$ at convergence.

We next discuss how to obtain each $\widehat{\theta}_k(\cdot)$. Motivated by the fact that $\mathbb{E}[w_i^* \zeta_{ik} x_i(s)] = 0$, we develop a three-step estimation procedure for each ζ_{ik} .

Step 1: We estimate w_i^* with $\widehat{w}_i = w(\widehat{\eta}_i)$ respectively, where $\widehat{\eta}_i = \widehat{\alpha}_{0,p} + \mathbf{z}_i^\top \widehat{\boldsymbol{\alpha}}_p + \int_{\mathcal{S}} x_i(s) \widehat{b}_p(s) ds$ with $\widehat{\alpha}_{0,p}$, $\widehat{\boldsymbol{\alpha}}_p$, and $\widehat{b}_p(s)$ obtained from the RAPLS algorithm.

Step 2: With a set of deterministic orthonormal basis functions $\{\pi_j(\cdot)\}_{j \geq 1}$, we expand $x_i(s)$ and $\theta(s)$, respectively, as

$$x_i(s) = \sum_{j=1}^{\infty} U_{ij} \pi_j(s) \text{ and } \theta_k^*(s) = \sum_{j=1}^{\infty} \theta_{kj}^* \pi_j(s),$$

where $U_{ij} = \int x_i(s) \pi_j(s) ds$ and $\theta_{kj}^* = \int \theta_k^*(s) \pi_j(s) ds$ for $k = 1, \dots, q$ and all i and j . With a pre-selected truncation parameter s_n , we obtain the weighted least squares estimator of $\{\theta_{kj}^*\}$:

$$\{\widehat{\theta}_{k1}, \dots, \widehat{\theta}_{ks_n}\} = \operatorname{argmin}_{\theta_{k1}, \dots, \theta_{ks_n}} \sum_{i=1}^n \widehat{w}_i (z_{ik} - \sum_{j=1}^{s_n} U_{ij} \theta_{kj})^2 \text{ for } k = 1, \dots, q.$$

Step 3: We obtain $\widehat{\theta}_k(s) = \sum_{j=1}^{s_n} \widehat{\theta}_{kj} \pi_j(s)$ and $\widehat{\boldsymbol{\zeta}}_i = (\widehat{\zeta}_{i1}, \dots, \widehat{\zeta}_{iq})^\top$, where $\widehat{\zeta}_{ik} = z_{ik} - \int_{\mathcal{S}} x_i(s) \widehat{\theta}_k(s) ds$ for $i = 1, \dots, n$ and $k = 1, \dots, q$.

Plugging $\widehat{\boldsymbol{\zeta}}_i$ into (4.2), we obtain the calibrated estimator $\widehat{\boldsymbol{\alpha}}_p^{\text{cal}}$. The

following result establishes the asymptotic normality of $\hat{\alpha}_p^{\text{cal}}$.

Theorem 2. *Suppose Assumptions (A1)–(A3) hold. If $\tau_n^2 \lambda_p^{-2} = o(n^{-1/4})$, $\|\theta_k^*(s) - \sum_{j=1}^{s_n} \theta_{kj}^* \pi_j(s)\|^2 = O(s_n^{1-2b})$, where $s_n = Cn^a$ for some constant C with $1/\{2(2b-1)\} < a \leq 1/4$, then we have as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\alpha}_p^{\text{cal}} - \alpha^*) \xrightarrow{d} N(0, \Sigma_\zeta^{-1}), \quad (4.3)$$

where $\Sigma_\zeta = \mathbb{E}(w_i^* \zeta_i \zeta_i^\top)$.

Theorem 2 imposes a slightly different condition on τ_n compared to Theorem 1, though neither condition is necessarily stronger than the other. Specifically, when $\tau_n = o(n^{-1/4})$, the condition $\tau_n \lambda_p^{-2} = O(1)$ from Theorem 1 implies the condition in Theorem 2. Conversely, when $\tau_n \gtrsim n^{-1/4}$, the condition $\tau_n^2 \lambda_p^{-2} = o(n^{-1/4})$ from Theorem 2 implies the condition in Theorem 1.

The condition $\|\theta_k^*(s) - \sum_{j=1}^{s_n} \theta_{kj}^* \pi_j(s)\|^2 = O(s_n^{1-2b})$ requires the s_n -term approximation of $\theta_k^*(s)$ based on the basis functions $\{\pi_j(\cdot)\}$ to be sufficiently accurate. Since $s_n = Cn^a$ and $1/\{2(1-2b)\} < a$, we can derive that

$$\|\theta_k^*(s) - \sum_{j=1}^{s_n} \theta_{kj}^* \pi_j(s)\| = O(n^{a(1-2b)/2}) = o(n^{-1/4}),$$

which is a relatively slow convergence rate.

Theorem 2 also implies the semi-parametric efficiency of the calibrated

estimator $\hat{\alpha}_p^{\text{cal}}$. In practice, Σ_ζ can be estimated by $\hat{\Sigma}_\zeta = n^{-1} \sum_{i=1}^n \hat{w}_i \hat{\zeta}_i \hat{\zeta}_i^\top$, where \hat{w}_i and $\hat{\zeta}_i$ are, respectively, estimated in Step 1 and Step 3.

5. Simulation Studies

We examined the finite-sample performance of the proposed RAPLS method in two settings: a partially functional linear model (PFLM) and a functional Poisson model (FPM). We compared RAPLS with two existing methods: the functional principal component regression (FPCR, Hall and Horowitz, 2007), and an existing method of partial least squares for generalized linear models (plsRglm, Meyer et al., 2010). As a key procedure in FPCR, FPCA was performed using the R package `fdapace`; plsRglm was performed using the R package `plsRglm`. AIC was used to determine the optimal number of components for all the methods.

In both settings, each curve $x_i(t)$ was generated in a way similar to that in Yuan and Cai (2010). More specifically, for $t \in [0, 1]$ and $k = 1, \dots, 50$, let $\phi_k(t) = \sqrt{2} \cos(k\pi t)$. For $i = 1, \dots, n$, we generated $x_i(t)$ according to $x_i(t) = \sum_{k=1}^{50} k^{-1/4} \xi_{ik} \phi_k(t)$, where $\xi_{ik} \stackrel{i.i.d}{\sim} N(0, 1)$. These curves were evaluated at 900 equally spaced points on $[0, 1]$. Simple algebra yields that $\text{Cov}(x_i(s), x_i(t)) = \sum_{k=1}^{50} k^{-1/2} \phi_k(s) \phi_k(t)$, indicating that the eigenvalues are $1, 1/\sqrt{2}, \dots, 1/\sqrt{50}$, and the eigenfunctions are $\phi_1(\cdot), \dots, \phi_{50}(\cdot)$. Finally,

5.1 Setting I: PFLM

the scalar covariate z_i was generated from a normal distribution with mean 0 and variance $\xi_{i5}^2/5$, imposing correlations between z_i and $x_i(\cdot)$.

5.1 Setting I: PFLM

In this setting, given $\{x_i(\cdot), z_i\}_{i=1,\dots,n}$, the outcome y_i was generated according to $y_i = 0.5 + \alpha^* z_i + \int_0^1 x_i(s) b^*(s) ds + \epsilon_i$, where ϵ_i follows a normal distribution with mean 0 and variance 0.8. We considered $\alpha^* = 1$ and two scenarios of $b^*(\cdot)$. In the first scenario, $b^*(\cdot)$ was in the span of the top 25 eigenfunctions as $b^*(s) = \sum_{k=1}^{25} (-1)^k \phi_k(s)$ for $s \in [0, 1]$. In the second scenario, $b^*(\cdot)$ was in the span of the tail 25 eigenfunctions as $b^*(s) = \sum_{k=26}^{50} (-1)^k \phi_k(s)$ for $s \in [0, 1]$. Theoretically, $b^*(\cdot)$ in the first scenario is easier to estimate, as top eigenfunctions are easier to estimate than tail eigenfunctions.

We generated 500 independent data sets for $n = 100, 200$ and 500. For each replication, we implemented RAPLS, FPCR, and plsRglm, and obtained the estimates of $b^*(\cdot)$. We implemented the proposed calibration procedure to estimate the scalar parameter α^* . To account for both bias and variance, we reported mean squared errors for all these estimates. Specifically, we calculated $\int_0^1 \{\hat{b}(s) - b^*(s)\}^2 ds$ and $(\hat{\alpha} - \alpha^*)^2$ for all estimators, and reported their averages over 500 replications as $\text{MSE}(b)$ and $\text{MSE}(\alpha)$,

5.1 Setting I: PFLM

respectively.

The results for $\text{MSE}(b)$ are presented in Table 1. In both settings, the proposed RAPLS method consistently outperforms FPCR and plsRglm in estimating the functional parameter b^* , highlighting the effectiveness of RAPLS in the PFLM context. While plsRglm is also based on partial least squares, it performs the worst in nearly all settings, except for Scenario II with $n = 100$. This is because, unlike RAPLS, which efficiently handles the correlation between scalar and functional covariates when constructing PLS basis functions, plsRglm simply combines the scalar and functional covariates to calculate the PLS components. As a result, the extracted plsRglm components fail to effectively represent the relationship between the functional covariates and the outcome. FPCR, on the other hand, performs worst in Scenario II with $n = 100$, underscoring a major limitation of FPCR/FPCA: when the true functional parameter is aligned with tail eigenvectors, FPCA struggles to estimate these eigenvectors in small sample sizes, leading to poor functional parameter estimation. Notably, RAPLS demonstrates similar performance across both scenarios of $b^*(\cdot)$, highlighting its robustness regardless of the alignment between $b^*(\cdot)$ and the eigenfunctions.

We also compared RAPLS, FPCR, and plsRglm estimators of α^* with

5.2 Setting II: FPM

n	Scenario	RAPLS	FPCR	plsRglm
100	I	0.834 (0.28)	2.109 (1.06)	4.440 (1.18)
200	I	0.263 (0.06)	0.701 (0.52)	2.487 (0.65)
500	I	0.089 (0.02)	0.344 (0.31)	1.545 (0.35)
100	II	0.820 (0.27)	2.853 (1.68)	1.884 (0.48)
200	II	0.265 (0.06)	1.137 (1.03)	1.272 (0.35)
500	II	0.089 (0.02)	0.725 (0.87)	1.169 (0.25)

Table 1: Simulation results for the PFLM: $\text{MSE}(b)$ and the standard error over 500 independent data sets for $n = 100, 200, 500$: Scenario I corresponds to $b^*(s) = \sum_{k=1}^{25} (-1)^k \phi_k(s)$, while Scenario II corresponds to $b^*(s) = \sum_{k=26}^{50} (-1)^k \phi_k(s)$.

results presented in Table 2. The proposed calibrated estimator achieves the best estimation accuracy in all scenarios.

5.2 Setting II: FPM

For $i = 1, \dots, n$, the outcome y_i was generated from a Poisson distribution with parameter $\exp(\eta_i)$, where $\eta_i = 0.5 + \alpha^* z_i + \int_0^1 x_i(s) b^*(s) ds$. Like Setting I, we considered $\alpha^* = 1$ and two scenarios of $b^*(\cdot)$. However, the choices of $b^*(\cdot)$ in Setting I were too large in scale, which led to infinite values in y_i .

5.2 Setting II: FPM

n	Scenario	RAPLS	FPCR	plsRglm
100	I	0.0099 (0.016)	0.0143 (0.021)	0.0109 (0.018)
200	I	0.0027 (0.003)	0.0034 (0.004)	0.0028 (0.003)
500	I	0.0008 (0.001)	0.0009 (0.001)	0.0009 (0.001)
100	II	0.0099 (0.016)	0.0152 (0.023)	0.0114 (0.019)
200	II	0.0027 (0.003)	0.0043 (0.007)	0.0029 (0.003)
500	II	0.0008 (0.001)	0.0010 (0.001)	0.0009 (0.001)

Table 2: Simulation results for the PFLM: $\text{MSE}(\alpha)$ over 500 independent data sets for $n = 100, 200, 500$: Scenario I corresponds to $b^*(s) = \sum_{k=1}^{25} (-1)^k \phi_k(s)$, while Scenario II corresponds to $b^*(s) = \sum_{k=26}^{50} (-1)^k \phi_k(s)$.

Thus, we rescaled each $b^*(\cdot)$ as $b^*(s) = \frac{2}{3} \sum_{k=1}^{25} (-1)^k \phi_k(s)$ for the first scenario and $b^*(s) = \frac{2}{3} \sum_{k=26}^{50} (-1)^k \phi_k(s)$ for the second scenario. To illustrate the robustness of the proposed RAPLS method regarding the initial values, we considered both deterministic and random initial values for RAPLS. For deterministic initial values, we set $b^{(0)}(\cdot) \equiv 0$, and then obtained $\alpha^{(0)}$ fitting the Poisson regression function with y_i as the outcome and z_i as the covariate. For random initial values, $b^{(0)}(\cdot)$ was generated from a Gaussian process with zero mean and the kernel $K(s, t) = e^{-10(s-t)^2}$ for each replica-

5.2 Setting II: FPM

tion. Then, $\alpha^{(0)}$ was obtained by fitting the Poisson regression model where y_i is the outcome, z_i is the covariate, and $\int_0^1 x_i(s)b^{(0)}(s)ds$ is the offset term. This approach is referred to as RAPLS-random hereafter.

Like Setting I, we generated 500 independent data sets for $n = 100, 200$ and 500, and for each setting, we reported results for RAPLS, RAPLS-random, FPCR, and plsRglm. The results for $MSE(b)$ are displayed in Table 4. Like Table 1, plsRglm show the worst performance in all cases except

n	Scenario	RAPLS-random	RAPLS	FPCR	plsRglm
100	I	1.369 (0.433)	1.371 (0.442)	2.235 (0.891)	4.434 (1.55)
200	I	0.379 (0.189)	0.373 (0.179)	0.447 (0.358)	3.250 (1.04)
500	I	0.098 (0.037)	0.096 (0.036)	0.167 (0.189)	2.233 (0.631)
100	II	1.279 (0.409)	1.275 (0.402)	2.603 (1.19)	2.028 (0.613)
200	II	0.303 (0.103)	0.298 (0.1)	0.726 (0.601)	0.786 (0.265)
500	II	0.094 (0.027)	0.089 (0.027)	0.396 (0.51)	0.433 (0.136)

Table 3: Simulation results for the FPM: $MSE(b)$ and its standard error over 500 independent data sets for $n = 100, 200, 500$: Scenario I corresponds to $b^*(s) = \frac{2}{3} \sum_{k=1}^{25} (-1)^k \phi_k(s)$, while Scenario II corresponds to $b^*(s) = \frac{2}{3} \sum_{k=26}^{50} (-1)^k \phi_k(s)$.

for Scenario II with $n = 100$, and RAPLS and RAPLS-random outperform

FPCR and plsRglm in all settings. FPCR performs better in Scenario I than Scenario II, because $b^*(\cdot)$ in Scenario II aligns with tail eigenfunctions, which require large sample sizes to estimate. RAPLS and RAPLS-random show similar results, demonstrating that the proposed RAPLS algorithm is robust to initial values. Results for $\text{MSE}(\alpha)$ show very similar patterns: both RAPLS methods outperform FPCR and plsRglm, while RAPLS and RAPLS-random show similar results. The estimation accuracy for the FPM is generally lower for all methods compared to the PFLM, likely due to the non-linear nature of the Poisson model.

6. Real data analysis

We applied the proposed RAPLS algorithm to a sample of patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study to analyze the progression of AD based on brain images. ADNI currently has 4 phases: ADNI1, ADNI-GO, ADNI2, and ADNI3, and its primary goal is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), and neuropsychological assessments can be used to measure the progression of AD. Participants were assessed at multiple visits. At each visit, various clinical measures, brain images, and neuropsychological assessments were collected. Detailed information about ADNI can be found

n	Scenario	RAPLS-random	RAPLS	FPCR	plsRglm
100	I	0.227 (0.478)	0.231 (0.482)	0.368 (0.707)	0.535 (0.265)
200	I	0.175 (0.369)	0.172 (0.366)	0.276 (0.576)	0.372 (0.357)
500	I	0.037 (0.079)	0.034 (0.075)	0.053 (0.103)	0.2 (0.283)
100	II	0.229 (0.478)	0.232 (0.482)	0.368 (0.708)	0.544 (0.256)
200	II	0.171 (0.370)	0.168 (0.367)	0.276 (0.576)	0.325 (0.362)
500	II	0.035 (0.073)	0.033 (0.069)	0.052 (0.095)	0.2 (0.282)

Table 4: Simulation results for the FPM: $\text{MSE}(b)$ over 500 independent data sets for $n = 100, 200, 500$: Scenario I corresponds to $b^*(s) = \frac{2}{3} \sum_{k=1}^{25} (-1)^k \phi_k(s)$, while Scenario II corresponds to $b^*(s) = \frac{2}{3} \sum_{k=26}^{50} (-1)^k \phi_k(s)$.

at the official website <http://www.adni-info.org>.

The stage of late mild cognitive impairment (LMCI) is considered a critical transitional stage between the normal stage and AD. However, it is unclear what brain regions drive the transition from LMCI to AD patients. PET neuroimaging directly measures the regional use of glucose with a lower glucose metabolic rate indicating less intensive neuronal activity, which has been proven as an important alternative to MRI images for AD diagnosis. We selected 302 subjects from ADNI1 without missing

data in the covariates of interest. Among the 302 subjects, 95 subjects were diagnosed with AD before the study's completion and the remaining 207 individuals were diagnosed with LMCI; 107 were female, while 195 were male. The majority of the subjects were right-handed (283) compared to only 19 left-handed people. The subjects' ages span from 55 to 89 with the median being 75.

We considered the following functional logistic model:

$$\text{logit}(\text{pr}\{y_i = 1 \mid \mathbf{z}_i, x_i(s)\}) = \alpha + \mathbf{z}_i^T \boldsymbol{\beta} + \int_{\mathcal{S}} x_i(s) b(s) ds \quad \text{for } i = 1, \dots, 236, \quad (6.1)$$

where $\text{logit}(x) = \log(x) - \log(1 - x)$, $y_i = 1$ indicates AD, and $y_i = 0$ indicates MCI. The covariates \mathbf{z}_i includes gender (1=Male; 0=Female), handedness (1=Right; 0=Left), and age. The functional predictor $x_i(s)$ is the PET imaging data measured on $160 \times 160 \times 96$ voxels. The PET images underwent four preprocessing steps, which are introduced in detail in the supplementary document. We also removed the regions outside the skull, and around 900,000 voxels remained.

Our primary analysis aims to identify brain regions that drive the AD progression from LMCI. We estimated $b(\cdot)$ in (6.1) using the proposed RAPLS. The optimal number of RAPLS basis components, determined by AIC, was 11. Fig. 1 presents three selected slices of the negative regions

of both estimates of $b(s)$, where decreased glucose metabolism is associated with the progression to AD. This finding aligns with current biological understanding. Reduced glucose metabolism, is a well-documented feature in Alzheimer's Disease, particularly in regions such as the parietal lobe, temporal lobe, and posterior cingulate cortex (Sanabria-Diaz et al., 2013). These areas are often implicated in cognitive functions like memory, attention, and spatial orientation, which are commonly affected in AD patients (Wilson et al., 2012). While the possibility of noise in the data should be acknowledged, the findings provide meaningful insights into the neurodegenerative processes in AD.

As a secondary analysis, we compared the predictive performance of RAPLS using a leave-one-out cross-validation (LOOCV) procedure to minimize variation in model evaluation. In addition, we included existing methods in the comparison, such as FPCR, plsRglm, linear discriminant analysis (LDA), and random forest (RF) models. Since plsRglm, LDA, and RF require substantial memory and computational time to handle images with 900,000 voxels, we facilitated the comparison by subsampling the images at 1,000 equally spaced voxels across the entire domain. The dimension-reduced images were used for all downstream analyses. Additionally, because LDA suffers from singular covariance matrices in high-dimensional

settings, we applied LDA to the top 50 principal components (PCs) of the reduced imaging data, which explained about 90% of the total variation. Based on the primary analysis, the optimal number of basis components was set to 11 for RAPLS. For FPCR and plsRglm, 8 and 14 basis components were selected based on AIC, respectively. The RF models were fitted using the R package `randomForest` (Liaw et al., 2002). The prediction accuracy for RAPLS, FPCR, plsRglm, LDA, and RF was 68.5%, 66.6%, 64.2%, 63.9%, and 66.9%, respectively. Notably, RAPLS achieved the highest prediction accuracy among the methods. Like the simulation results, RAPLS outperformed plsRglm, another method based on partial least squares. This suggests that the IRLS-based construction of PLS basis functions in RAPLS is more suitable for nonlinear models compared to the GLM-based construction used in plsRglm in terms of prediction.

We should acknowledge that the distinction between LMCI and early-stage AD is often subtle and difficult to define. Clinically, the transition from LMCI to early AD is marked by a gradual progression of symptoms, with considerable overlap in cognitive decline, memory impairment, and other neurodegenerative indicators. As a result, the biological and clinical boundaries between these two groups are inherently blurred. This overlap makes it challenging to develop models that can consistently and accurately

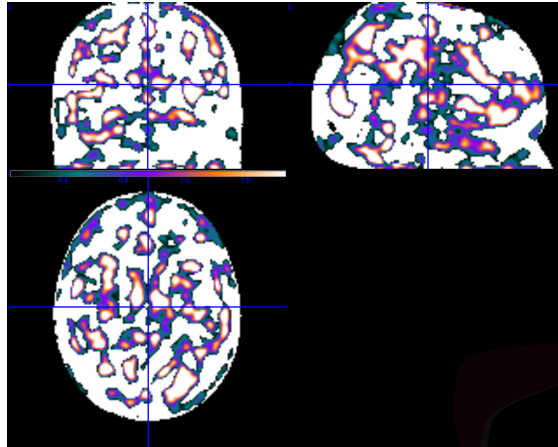


Figure 1: ADNI results: the negative regions of estimated $b(s)$ using the entire cohort: RAPLS with 10 basis functions. From left to right, each coefficient image is displayed with 3 slices in its transverse, coronal, and sagittal view located at $\{80, 80, 48\}$, respectively.

predict the correct classification. Consequently, as observed in prior studies, prediction accuracies tend to be low, which is consistent with existing studies with even larger sample sizes (Nozadi et al., 2018).

7. Discussions

This paper introduces a residual-based alternative partial least squares (RAPLS) method for parameter estimation in a class of generalized functional linear models (GFLM). The key idea behind RAPLS is the integration of the iteratively reweighted least squares (IRLS) and the alternative partial least squares (APLS), enabling accurate approximation of nonlinear

functional models through a sequence of functional linear models. RAPLS demonstrates clear advantages in both estimation and prediction, particularly when the underlying function parameter is closely aligned with the tail eigenfunctions.

Identifiability is a general challenge in functional regression models. For example, in a simple functional linear model, $y = \alpha + \int x(s)b(s)ds + \epsilon$, the function $b(\cdot)$ is not identifiable if $x(\cdot)$ lies within a low-dimensional space \mathcal{S} , but $b(\cdot)$ does not. Specifically, any part of $b(\cdot)$ that lies outside of \mathcal{S} will remain unidentifiable. This challenge persists when dimensionality reduction techniques are applied to estimate $b(s)$. For instance, when using the top p RAPLS basis functions, the estimated parameter is $b_p(\cdot)$, the projection of $b(\cdot)$ onto the space spanned by the top p RAPLS basis functions. In this case, the difference $b - b_p$ remains unidentifiable. This issue is particularly concerning if $\|b - b_p\|$ does not converge to 0 as $p \rightarrow \infty$. Fortunately, RAPLS overcomes this issue, as Lemma 2 guarantees that $\|b - b_p\|$ diminishes as more basis functions are used. A similar result holds for FPCR when eigenfunctions are employed as basis functions, provided that the covariance kernel of the functional covariates is positive definite.

This work is motivated by brain imaging applications where the images are regularly spaced. However, we acknowledge that in many other

applications, such as those involving longitudinal designs, functional data may be sparse or measured irregularly. While RAPLS is methodologically applicable to such data, we anticipate that some pre-smoothing will be necessary in these cases. Evaluating the performance of RAPLS on irregularly observed functional data will be left as a future research direction.

Given the connectivity structures of the brain, the functional parameter associated with brain images is typically assumed to be smooth across voxels. In other contexts, achieving sparsity in the estimated coefficient function may be desirable for greater interpretability. To induce sparsity in the coefficient function, one potential approach is to introduce sparsity into the RAPLS basis functions. In non-functional partial least squares (PLS), sparsity can be achieved by adding constraints to the iterative algorithms. For example, consider the linear model $Y = X\beta + \epsilon$. The first “sparse” PLS basis function can be constructed by solving:

$$\max \mathbf{w}^\top X^\top Y Y^\top X \mathbf{w}, \quad \text{subject to } \|\mathbf{w}\|_2 = 1, \|\mathbf{w}\|_1 < \lambda,$$

where λ controls the level of sparsity. Extending this idea to RAPLS presents an exciting opportunity for future research.

Supplementary Materials

Online supplementary material includes additional simulations and theoretical results, proofs of the main theorems, and supporting information for the real data application.

Acknowledgements

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at <http://adni.loni.usc.edu/wpcontent/uploads/howtoapply/ADNIAcknowledgementList.pdf>. Dr. Zhu was partially supported by the National Institutes of Health (NIH) grants 1R01AR082684, 1OT2OD038045-01, and the National Institute on Aging (NIA) of the National Institutes of Health (NIH) grants U01AG079847, 1R01AG085581, RF1AG082938, and R01AR082684.

REFERENCES

References

- Aguilera, A., M. Aguilera-Morillo, and C. Preda (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems* 154, 80–92.
- Aguilera, A. M., M. Escabias, C. Preda, and G. Saporta (2010). Using basis expansions for estimating functional PLS regression: applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems* 104(2), 289–305.
- Bastien, P., V. E. Vinzi, and M. Tenenhaus (2005). Pls generalised linear regression. *Computational Statistics & data analysis* 48(1), 17–46.
- Besse, P. and J. O. Ramsay (1986). Principal components analysis of sampled functions. *Psychometrika* 51, 285–311.
- Boente, G. and R. Fraiman (2000). Kernel-based functional principal components. *Statistics Probability Letters* 48, 335 – 345.
- Cai, T. and P. Hall (2006). Prediction in functional linear regression. *The Annals of Statistics* 34, 2159–2179.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1), C1–C68.
- Cook, R. D. (2007). Fisher Lecture: Dimension Reduction in Regression. *Statistical Science* 22(1), 1 – 26.

REFERENCES

- Delaigle, A. and P. Hall (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics* 34, 2159–2179.
- Escabias, M., A. M. Aguilera, and M. J. Valderrama (2007). Functional PLS logit regression model. *Computational Statistics & Data Analysis* 51(10), 4891–4902.
- Febrero-Bande, M., P. Galeano, and W. González-Manteiga (2017). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review* 85(1), 61–83.
- Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementation*. Springer, New York.
- Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling* 15, 256–278.
- Green, P. J. (1984). Pls regression on a stochastic process. *Journal of the Royal Statistical Society. Series B* 46, 149–192.
- Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* 35(1), 70–91.
- Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.
- James, G., T. Hastie, and C. Sugar (2000). Principal component models for sparse functional data. *Biometrika* 87, 587–602.

REFERENCES

- Jung, S. and J. S. Marron (2009). PCA consistency in high dimensional, low sample size context. *The Annals of Statistics* 37(6B), 4104–4130.
- Kong, D., A.-M. Staicu, and A. Maity (2016). Classical testing in functional linear models. *Journal of nonparametric statistics* 28(4), 813–838.
- Kong, D., K. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika* 103(1), 147–159.
- Krämer, N., A.-L. Boulesteix, and G. Tutz (2008). Penalized partial least squares with applications to b-spline transformations and functional data. *Chemometrics and Intelligent Laboratory Systems* 94(1), 60–69.
- Lee, E., H. Zhu, D. Kong, Y. Wang, K. S. Giovanello, and J. G. Ibrahim (2015). Bflerm: A bayesian functional linear Cox regression model for predicting time to conversion to alzheimer’s disease. *The Annals of Applied Statistics* 9(4), 2153–2178.
- Li, K. and S. Luo (2017). Functional joint model for longitudinal and time-to-event data: an application to alzheimer’s disease. *Statistics in Medicine* 36(22), 3560–3572.
- Liaw, A., M. Wiener, et al. (2002). Classification and regression by randomforest. *R News* 2(3), 18–22.
- Lv, S., X. He, and J. Wang (2023). Kernel-based estimation for partially functional linear model: Minimax rates and randomized sketches. *Journal of Machine Learning Research* 24(55), 1–38.

REFERENCES

- Meyer, N., M. Maumy-Bertrand, and F. Bertrand (2010). Comparaison de la régression pls et de la régression logistique pls: application aux données d'allélotypage. *Journal de la Société Française de Statistique* 151, 1–18.
- Morris, J. S. (2015). Functional regression. *Annual Reviews of Statistics and its Application* 2, 321–359.
- Müller, H.-G. and U. Stadtmüller (2005). Generalized functional linear models. *The Annals of Statistics*, 774–805.
- Nozadi, S. H., S. Kadoury, A. D. N. Initiative, et al. (2018). Classification of alzheimer's and mci patients from semantically parcelled pet images: A comparison between av45 and fdg-pet. *International journal of biomedical imaging* 2018(1), 1247430.
- Preda, C. and G. Saporta (2005a). Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis* 49, 99–108.
- Preda, C. and G. Saporta (2005b). PLS regression on a stochastic process. *Computational Statistics & Data Analysis* 49, 149–158.
- Preda, C., G. Saporta, and C. Leveder (2005). PLS classification of functional data. *Computational Statistics & Data Analysis* 49, 223–235.
- Qu, S., J.-L. Wang, and X. Wang (2016). Optimal estimation for the functional Cox model. *The Annals of Statistics* 44, 1708–1738.
- Ramsay, J. O. and C. J. Dalzell (1991). Some tools for functional data analysis (with discussion).

REFERENCES

- Journal of the Royal Statistical Society.* 53, 539–572.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. New York: Springer-Verlag.
- Reiss, P. T. and R. T. Ogden (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* 102(479), 984–996.
- Sanabria-Diaz, G., E. Martinez-Montes, L. Melie-Garcia, and A. D. N. Initiative (2013). Glucose metabolism during resting state reveals abnormal brain networks organization in the alzheimer’s disease and mild cognitive impairment. *PloS one* 8(7), e68860.
- Wang, X. and D. Ruppert (2015). Optimal prediction in an additive functional model. *Statistica Sinica* 25(2), 567–589.
- Wang, Y., J. G. Ibrahim, and H. Zhu (2020). Partial least squares for functional joint models with applications to the alzheimer’s disease neuroimaging initiative study. *Biometrics* 76(4), 1109–1119.
- Wilson, R. S., E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hize, and D. A. Bennett (2012). The natural history of cognitive decline in alzheimer’s disease. *Psychology and aging* 27(4), 1008.
- Yuan, M. and T. Cai (2010). A reproducing kernel hilbert space approach to functional linear regression. *The Annals of Statistics* 38, 3412–3444.

REFERENCES

E-mail: yue.2.wang@cuanschutz.edu

Xiao Wang

E-mail: wangxiao@purdue.edu

Joseph G. Ibrahim

E-mail: ibrahim@bios.unc.edu

Hongtu Zhu

E-mail: htzhu@email.unc.edu