Statistica Sinica **36** (2026), 1-19
doi:https://doi.org/10.5705/ss.202023.0418

RESIDUAL-BASED ALTERNATIVE PARTIAL LEAST SQUARES FOR GENERALIZED FUNCTIONAL LINEAR MODELS

Yue Wang*[1], Xiao Wang[2], Joseph G. Ibrahim[3] and Hongtu Zhu[3]

[1] *Colorado School of Public Health,* [2] *Purdue University*
and [3] *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 *independent and identically distributed (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

*Corresponding author. E-mail: yue.2.wang@cuanshutz.edu