Journal of Multivariate Analysis 163 (2018) 37–50

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Functional envelope for model-free sufficient dimension reduction

Xin Zhang [a,*], Chong Wang [b], Yichao Wu [c]

[a] Department of Statistics, Florida State University, Tallahassee, FL 32312, USA
[b] Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
[c] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

ARTICLE INFO

Article history:
Received 29 November 2016
Available online 7 October 2017

Keywords:
Envelope model
Functional data
Functional inverse regression
Sufficient dimension reduction

ABSTRACT

In this article, we introduce the functional envelope for sufficient dimension reduction and regression with functional and longitudinal data. Functional sufficient dimension reduction methods, especially the inverse regression estimation family of methods, usually involve solving generalized eigenvalue problems and inverting the infinite-dimensional covariance operator. With the notion of functional envelope, essentially a special type of sufficient dimension reduction subspace, we develop a generic method to circumvent the difficulties in solving the generalized eigenvalue problems and inverting the covariance directly. We derive the geometric characteristics of the functional envelope and establish the asymptotic properties of related functional envelope estimators under mild conditions. The functional envelope estimators have shown promising performance in extensive simulation studies and real data analysis.

Published by Elsevier Inc.

1. Introduction

The notion of envelope was first introduced by Cook et al. [12] in the context of sufficient dimension reduction in regression of a univariate response $Y \in \mathbb{R}$ on a multivariate predictor $X \in \mathbb{R}^p$, where the goal is to find the smallest sufficient dimension reduction subspace $\mathcal{S} \subseteq \mathbb{R}^p$ such that the conditional distribution of Y given X is the same as that of Y given the reduced predictor $P_{\mathcal{S}}X$, with $P_{\mathcal{S}}$ being the projection onto $\mathcal{S}$. While most of the standard sufficient dimension reduction methods require inversion of the sample predictor covariance matrix, the method proposed in [12] is a dimension reduction technique which does not require such an inversion and is thus applicable to a higher dimensional predictor X.

Following the notion of envelope in [12], more geometric and statistical properties of, and various estimation procedures for, envelopes were developed and investigated in the context of envelope regression models. Envelope regression was first proposed in [13], as a way of reducing the multivariate response in a multivariate linear model. It was later extended to various models and applications such as partial reduction [45], predictor reduction [10], simultaneous reduction [18], reduced-rank regression [9], generalized linear models [17], and tensor regression [40,52]. Envelope methods increase efficiency in regression coefficient estimation and improve prediction by enveloping the information in the data that is material to estimation, while excluding the information that is immaterial. The improvement in estimation and prediction can be quite substantial, as illustrated by these aforementioned studies.

The goal of this paper is to develop a class of sufficient dimension reduction techniques for functional data that require no inversion of the covariance matrix, using the idea of envelopes. To the best of our knowledge, this is the first time

* Correspondence to: Department of Statistics, Florida State University, 214 OSB 117 N. Woodward Ave. Tallahassee, FL 32306-4330, USA.
E-mail address: henry@stat.fsu.edu (X. Zhang).

https://doi.org/10.1016/j.jmva.2017.09.010
0047-259X/Published by Elsevier Inc.

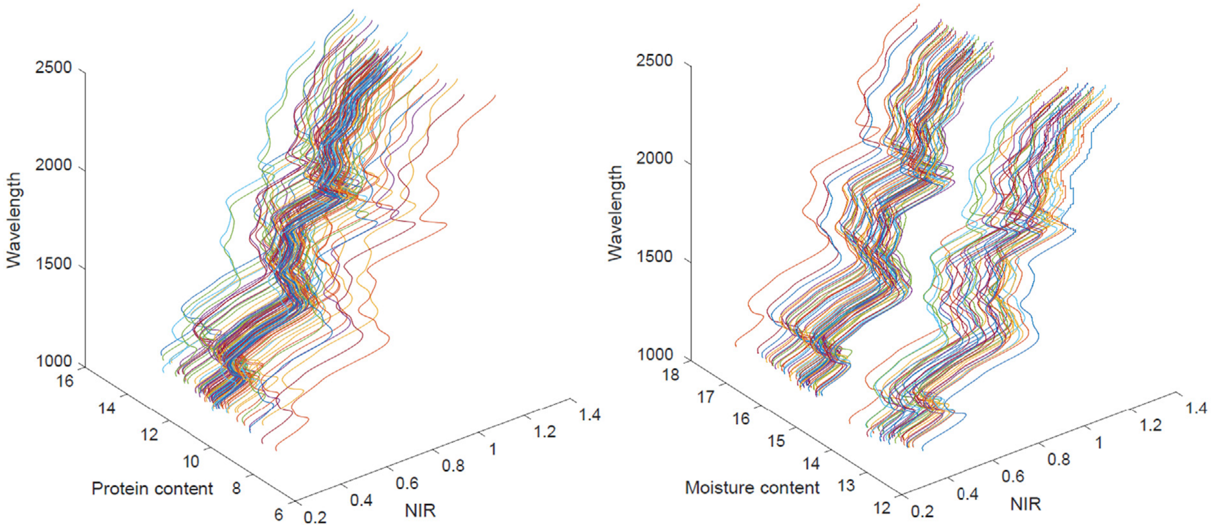

Fig. 1. Plots of the raw data from [31]: near infrared spectra (represented by the smoothed curves) of 100 wheat samples, together with their protein and moisture contents.

that envelope methodology is extended beyond the usual multivariate regression setting to functional data analysis. An important contribution of this paper is to bridge the gap between the nascent area of envelope methodology, functional data analysis, and sufficient dimension reduction. The approach here is different from many previous envelope methods, because we are developing model-free sufficient dimension reduction methods rather than focusing on a specific model. In recent years, functional sufficient dimension reduction methods, e.g., [6,26,27,30,32,37,46,47,50], especially the functional inverse regression methods, have gained interest as versatile tools for data visualization and exploratory analysis in functional regression. We propose a very generic functional envelope estimation based on the popular inverse regression class of functional sufficient dimension reduction methods. It improves essentially all the aforementioned functional SDR methods by avoiding truncation and inversion of the covariance operator of the functional predictor, and thus enriches the tactics of functional SDR estimation. The new method can also be viewed as an alternative to functional principal components in dimension reduction and regression [34,35,39,48,49]. Recent studies have revealed deep connections between envelope models and partial least squares both for a vector predictor [10] and a tensor (multi-dimensional array) predictor [52]. Our study also sheds light on the connections between functional envelopes and recent developments in functional partial least squares [21].

In functional data analysis, especially when nonparametric techniques are involved, it is well known that functional estimators suffer severely from the "curse of dimensionality" both from a theoretical and a practical point of view. See, e.g., [28] for an overview of the curse of dimensionality and related issues in functional nonparametric regression. Dimension reduction techniques such as functional principal component analysis and functional partial least squares are widely applied in recent functional data analysis studies. See [29] and [19] for excellent overviews of recent advances in functional data. Our functional envelope method is aiming to circumvent the curse of dimensionality and related issues, by finding the most effective functional dimension reduction. After efficiently reducing the infinite-dimensional functional predictor space to $\mathbb{R}^d$, where d typically is a small number (e.g., 1 or 2), standard nonparameteric or semi-parametric regression techniques can be applied directly. The proposed envelope methodology in this paper can also be combined with existing functional and high-dimensional data analysis techniques such as sparse modeling [1,50] and semi-parametric analysis [29]. Envelope reduction is similar in spirit to the functional single-index and projection pursuit methods [4,5] and provides an alternative way of pre-processing the data and eliminating redundant information as the envelope targets and models the index function and the covariance function simultaneously.

As a motivating example, we consider the wheat protein and moisture content data set from [31]. The data set consists of near infrared (NIR) spectra of $n = 100$ wheat samples with two responses: Y_1 is the protein content and Y_2 is the moisture content; the predictor $X(t)$ is the NIR absorption spectra that are measured at 351 equally spaced frequencies with a spacing of 4 nm between 1100 nm (first frequency) and 2500 nm (last frequency). Summary plots of the data can be found in Fig. 1. We consider the dimension reductions in the regression of Y_1 on $X(t)$ and in the regression of Y_2 on $X(t)$ separately. For the moisture content (Y_2), we found that the unsupervised functional PCA cannot identify the most predictive component but the supervised SDR methods such as FCS [47] and our proposed method FECS can efficiently find the important directions for improved data visualization. Plots of the response (moisture content) versus the reduced predictors by various methods can be found in Fig. 2. A more complete analysis on these data is presented in Section 5, where we further demonstrate that the FECS is more robust and effective than FCS and other alternative functional data analysis and prediction methods.