Functional envelope for model-free sufficient dimension reduction

Xin Zhang*, Chong Wang†, and Yichao Wu‡

Florida State University and North Carolina State University

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.

Key Words: Envelope model; functional data; functional inverse regression; sufficient dimension reduction.

*Xin Zhang is Assistant Professor, Department of Statistics, Florida State University, Tallahassee, FL 32312, USA; Email: henry@stat.fsu.edu

†Chong Wang is Graduate Student, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA; Email: cwang19@ncsu.edu

‡Yichao Wu is Associate Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA; Email: wu@stat.ncsu.edu

1 Introduction

The notion of envelopes was first introduced by Cook et al. (2007) in the context of sufficient dimension reduction in regression of a univariate response $Y \in \mathbb{R}^1$ 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 by Cook et al. (2007) is a dimension reduction technique without the need for such inversion of sample covariance matrix and is thus applicable to higher dimensional predictor X.

Following the notion of envelopes in Cook et al. (2007), more geometric and statistical properties and various estimation procedures of envelopes are further developed and investigated in the context of envelope regression models. Envelope regression was first proposed by Cook et al. (2010), as a way of reducing the multivariate response in a multivariate linear model, and was later extended to various models and applications such as partial reduction (Su and Cook, 2011), predictor reduction (Cook et al., 2013), simultaneous reduction (Cook and Zhang, 2015b), reduced-rank regression (Cook et al., 2015), generalized linear models (Cook and Zhang, 2015a), tensor regression (Li and Zhang, 2016), among others. 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 many recent studies.

The goal of this paper is to develop a class of sufficient dimension reduction techniques for functional data that require no inversion of covariance matrix, using the idea of envelopes. To the best of our knowledge, this is the first paper that extends the envelope methodology beyond the usual multivariate regression settings to functional data analysis. An important contribution of this paper is to bridge the gap between the nascent area of envelope methodology and the

well known 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 other than focusing on a specific model. In recent years, functional sufficient dimension reduction methods (Ferré and Yao, 2003; Ferré et al., 2005; Jiang et al., 2014; Wang et al., 2015; Yao et al., 2015, 2016; Chen et al., 2015; Li and Song, 2016; Lee and Shao, 2016, for example), especially the functional inverse regression methods, have gained increasing interest as versatile tools for data visualization and exploratory analysis in functional regressions. 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 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 (Yao et al., 2005a,b; Li, 2011; Li et al., 2013; Li and Guan, 2014). Recent studies have reveal profound connections between envelope models and partial least squares for multivariate (vector) predictor (Cook et al., 2013) and for tensor (multi-dimensional array) predictor (Zhang and Li, 2016). Our study will also shed light on the connections between functional envelopes and recent developments of functional partial least squares (Delaigle et al., 2012, e.g.).

In functional data analysis, especially when non-parametric techniques are involved, it is well known functional estimators suffer severely from the "curse-of-dimensionality" in both theoretical and practical aspects. See, for example, Geenens et al. (2011) for an overview of the curse-of-dimensionality and related issues in function non-parametric 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 Goia and Vieu (2016) and Cuevas (2014) 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

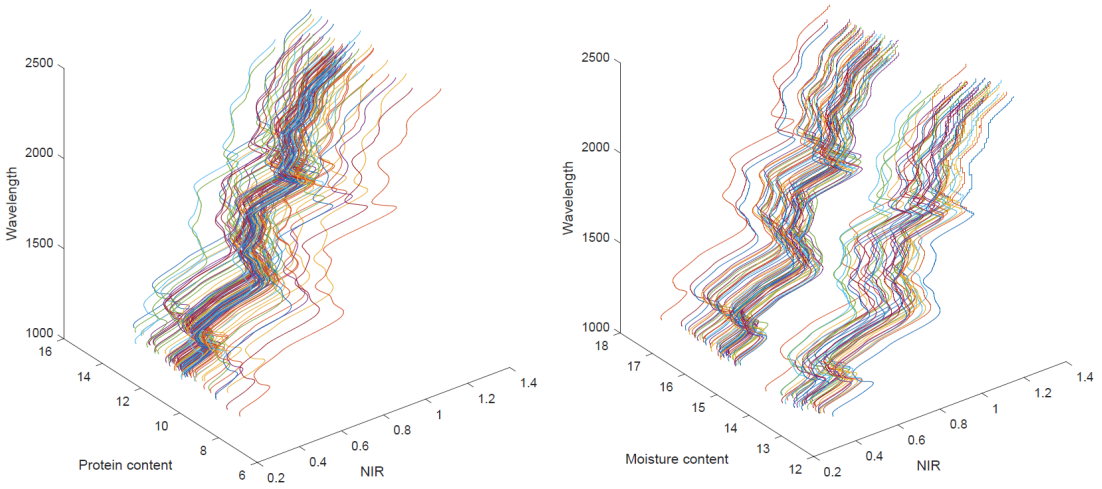

Figure 1.1: Plots of the raw data from Kalivas (1997): 100 wheat samples' near infrared spectra (represented by the smoothed curves) and their protein and moisture contents.

the infinite dimensional functional predictor space to $\mathbb{R}^d$, where d typically can be a small number 1 or 2, standard non-parameteric 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 (Aneiros and Vieu, 2016; Yao et al., 2016, e.g.) and semi-parametric analysis (Goia and Vieu, 2016, e.g.). The envelope reduction behaves similar in spirit to the functional single-index and projection pursuit methods (Chen et al., 2011a,b) 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 Kalivas (1997). 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; predictor $X(t)$ is the NIR absorption spectra that are measured at 351 equally spaced frequencies with a spacing of 4nm between 1100nm (first frequency) and 2500nm (last frequency). Summary plots of the data can be found in Figure 1.1. 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

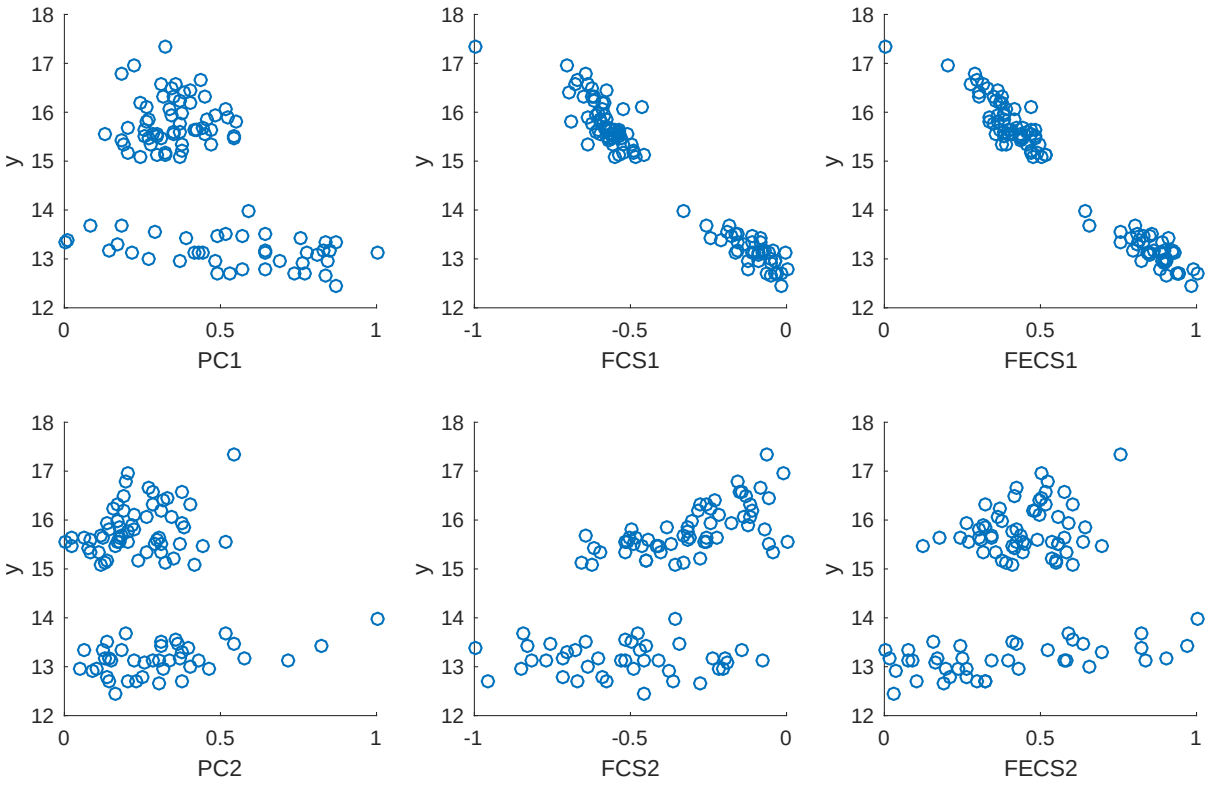

Figure 1.2: Moisture content (y-axis) versus the six dimension reduction directions (x-axes): the first two principal components (PC1 and PC2 on the left column of plots); the first two directions from the functional cumulative slicing estimator (FCS1 and FCS2 on the middle column of plots); the first two directions from the functional envelope cumulative slicing estimator (FECS1 and FECS2 on the middle column of plots).

found that the unsupervised functional PCA can not identify the most predictive component but the supervised SDR methods such as FCS (Yao et al., 2015) and our proposed method FECS can efficiently find the important directions which further visualize the data better. Plots of the response (moisture content) versus the reduced predictors by various methods can be found in Figure 1.2. A more complete analysis on this data is presented in Section 5, where we further demonstrate the FECS is more robust and effective than FCS and other alternative functional data analysis and prediction methods.

2 Functional envelope

2.1 Sufficient dimension reduction in functional data

In functional data analysis, we consider the problem of a scalar response variable $Y \in \mathbb{R}^1$ and a functional random variable $X(t)$, where t is an index defined on a closed and compact interval $\mathcal{T}$. See, for example, Silverman and Ramsay (2005) for some background on functional data analysis. Let X be defined on the real separable Hilbert space $\mathcal{H} \equiv L^2(\mathcal{T})$ with inner product $\langle f, g \rangle = \int_{\mathcal{T}} f(t)g(t)dt$ and norm $\|f\|_{\mathcal{H}} = \langle f, f \rangle^{1/2}$. Statistical analysis typically focuses on the collection of all bounded linear operators from $\mathcal{H}$ to $\mathcal{H}$, which is denoted as $\mathcal{B}(\mathcal{H}) \equiv \mathcal{B}(\mathcal{H}, \mathcal{H})$, where the vector operations are defined point-wise. Sufficient dimension reduction (SDR) in regression of Y on $X(t)$ seeks for the set of linear functions $\eta_1(t), \ldots, \eta_m(t)$ such that Y is independent of $X(t)$ given the m sufficient variables $\langle \eta_1, X \rangle, \ldots, \langle \eta_m, X \rangle$. Let $\mathrm{span}(\eta_1, \ldots, \eta_m)$ be the subspace spanned by all possible linear combinations of the functions $\eta_1, \ldots, \eta_m$. Then it is called a sufficient dimension reduction subspace. Sufficient dimension reduction subspaces are not unique. Therefore we seek for the central subspace (Cook, 1998). The central subspace of Y on X, denoted by $\mathcal{S}_{Y|X}$, is defined as the intersection of all possible sufficient dimension reduction subspaces that is also a sufficient dimension reduction subspace itself. By definition, the central subspace, assuming that it exists throughout our exposition, is unique and is the smallest sufficient dimension reduction subspace.

We assume that the central subspace $\mathcal{S}_{Y|X}$ has dimension d, $1 \leq d < \infty$, and thus can be expressed as $\mathcal{S}_{Y|X} = \mathrm{span}(\beta_1, \ldots, \beta_d)$ for some linearly independent index functions $\beta_1(t), \ldots, \beta_d(t)$. Then we can write

$$Y \perp\!\!\!\perp X \mid \langle \beta_1, X \rangle, \ldots, \langle \beta_d, X \rangle, \tag{2.1}$$

which implies that Y is independent of the (infinite dimensional) functional random variable X, given the d-dimensional projected random variables $\langle \beta_j, X \rangle \in \mathbb{R}^1$, $j = 1, \ldots, d$. Especially, the above statement (2.1) includes a broad class of semi-parametric index models as follows,

$$Y = g(\langle \beta_1, X \rangle, \ldots, \langle \beta_d, X \rangle; \epsilon), \tag{2.2}$$

where $g: \mathbb{R}^{d+1} \mapsto \mathbb{R}^1$ is an unknown link function and the error process ϵ has zero mean, finite variance $\sigma^2 > 0$, and is independent of X. Although the basis functions $\beta_1, \ldots, \beta_d$ are not unique, the span of them is the unique central subspace, which is the target of most sufficient dimension reduction methods as we briefly review in the following.

We assume $X(t)$ is centered and has finite fourth moment: $\mathrm{E}\{X(t)\} = 0$ for all $t \in \mathcal{T}$ and $\int_{\mathcal{T}} \mathrm{E}\{X^4(t)\}dt < \infty$. Let $\Sigma \equiv \Sigma(s,t) = \mathrm{E}\{X(s)X(t)\}$ be the covariance operator. Most of existing sufficient dimension reduction methods estimate directions in the central subspace sequentially as a generalized eigenvalue problem,

$$\Sigma v_i = \lambda_i \Lambda v_i, \tag{2.3}$$

where Λ is called the kernel of a sufficient dimension reduction method and in functional data analysis, $\Lambda = \Lambda(s,t)$ and $(\Lambda v_i)(t) = \int_{\mathcal{T}} \Lambda(s,t)v_i(s)ds$. Perhaps the most popular functional sufficient dimension reduction methods are inverse regression type estimators, see, for example, Ferré and Yao (2003); Ferré et al. (2005); Jiang et al. (2014); Yao et al. (2015). Those methods all fall into the aforementioned generalized eigenvalue problem framework, where they aim at the same kernel and proposed various different non-parametric or semi-parametric estimations. The kernel is defined as follows,

$$\Lambda(s,t) = \mathrm{E}[\mathrm{E}\{X(s) \mid Y\}\mathrm{E}\{X(t) \mid Y\}] \equiv \mathrm{var}\{\mathrm{E}(X \mid Y)\}. \tag{2.4}$$

Such sufficient dimension reduction methods typically assume the linearity condition: for any function $b \in \mathcal{H}$, the conditional expectation $\mathrm{E}(\langle b, X\rangle \mid \langle \beta_1, X\rangle, \ldots, \langle \beta_d, X\rangle)$ is a linear function of $\langle \beta_1, X\rangle, \ldots, \langle \beta_d, X\rangle$. Then under the linearity condition, $\mathrm{span}(\Lambda) \subseteq \Sigma\mathcal{S}_{Y|X}$. We further assume the coverage condition to ensure equality $\mathrm{span}(\Lambda) = \Sigma\mathcal{S}_{Y|X}$. See Cook and Ni (2005); Li and Wang (2007); Cook and Zhang (2014) for more discussions on the linearity and coverage conditions. Under these commonly used linearity and coverage conditions, the generalized eigenvalue problem (2.3) only has d nonzero λ_i's and the corresponding d eigen-vectors or eigen-functions $v_1, \ldots, v_d$ will span the central subspace as $\mathcal{S}_{Y|X} = \mathrm{span}(\beta_1, \ldots, \beta_d) = \mathrm{span}(v_1, \ldots, v_d)$. This also implies that the rank of Λ is also d, and that

$\mathcal{S}_{Y|X} = \text{span}(\Sigma^{-1}\Lambda)$ provided that Σ^{-1} is well-defined. The central subspace can thus be recovered as $\mathcal{S}_{Y|X} = \text{span}(\Sigma^{-1}\Lambda)$ under appropriate assumptions on Σ to make $\Sigma^{-1}\Lambda$ well-defined (e.g. Assumption 3 in Yao et al. (2015)). If the dimension of the central subspace d is known, the central subspace is estimated as the span of the first d (right) eigenvectors of $\widehat{\Sigma}^{-1}\widehat{\Lambda}$, with truncated covariance estimator $\widehat{\Sigma}$ and various sample estimators for $\widehat{\Lambda}$ that varied from method to method. Our proposed functional envelope approach extends all these methods in the same fashion, regardless of various different estimation procedures for $\widehat{\Lambda}$. While there are many different methods of estimating $\widehat{\Lambda}$, our envelope estimation framework provides a generic method as an alternative to the generalized eigenvalue problem and thus can fit with any consistent estimator $\widehat{\Lambda}$. For illustration, we use $\widehat{\Lambda}$ from Yao et al. (2015), because the functional cumulative slicing method in Yao et al. (2015) and the original cumulative slicing method proposed by Zhu et al. (2010) for multivariate $X \in \mathbb{R}^p$, to avoid the selection of the number of slices in sliced inverse regression type methods (Li, 1991; Ferré et al., 2005). We give a brief review of the estimation procedure for the functional cumulative slicing (FCS) in Section 3.1. Additionally, by avoiding truncating and inverting $\widehat{\Sigma}$, our method does not require any assumptions to make $\Sigma^{-1}\Lambda$ well-defined.

2.2 Definition of functional envelopes

The key concept in this paper is the functional envelope for sufficient dimension reduction. Envelope and its basic properties were first proposed and studied in the classical multivariate set-up of sufficient dimension reduction (Cook et al., 2007) and multivariate linear regression Cook et al. (2010). We define the functional envelope in this section as a generalization of the classical envelopes to functional data analysis.

First of all, we review the definition of reducing subspace in the following. The notion of reducing subspace is crucial for the developments of envelopes and arises commonly in functional analysis, see for example, Conway (1990).

Definition 1. *Let $\mathcal{R} \subseteq \mathcal{H}$ be a subspace of $\mathcal{H}$, and let $M \in \mathcal{B}(\mathcal{H})$ be a bounded linear operator.*

If $M\mathcal{R} \subseteq \mathcal{R}$, then we call $\mathcal{R}$ a invariant subspace of M. If in addition $M\mathcal{R}^{\perp} \subseteq \mathcal{R}^{\perp}$, where $\mathcal{R}^{\perp}$ is the orthogonal complement of $\mathcal{R}$, then $\mathcal{R}$ is a reducing subspace of M.

The next proposition illustrates a basic property of reducing subspace, which is the key to our development of functional envelopes.

Proposition 1. *The subspace $\mathcal{R}$ is a reducing subspace of M if and only if M can be written in the form*

$$M = P_{\mathcal{R}} M P_{\mathcal{R}} + Q_{\mathcal{R}} M Q_{\mathcal{R}}, \tag{2.5}$$

where $P_{\mathcal{R}} = P_{\mathcal{R}}(s,t) \in \mathcal{B}(\mathcal{H})$ and $Q_{\mathcal{R}}(s,t) \in \mathcal{B}(\mathcal{H})$ are projections onto $\mathcal{R}$ and $\mathcal{R}^{\perp}$.

For a bounded linear operator $M \in \mathcal{B}(\mathcal{H})$, we define the M-envelope of a subspace $\mathcal{S} \subseteq \mathcal{H}$ as following. This definition of functional envelope is a direct generalization of Definition 2.1 of Cook et al. (2010) from the Euclidean space to Hilbert space, and is the key concept for the developments in this paper.

Definition 2. *The M-envelope of $\mathcal{S}$, denoted as $\mathcal{E}_M(\mathcal{S})$, is the intersection of all reducing subspace of M that contains $\mathcal{S}$.*

The functional envelope $\mathcal{E}_M(\mathcal{S})$ always exists, since $\mathcal{H}$ is a reducing subspace of M that contains $\mathcal{S}$. Becasue of Proposition 1, the intersection of any two reducing subspace of M is still a reducing subspace of M. Therefore, the functional envelope $\mathcal{E}_M(\mathcal{S})$, by its construction, is guarenteed to be unique and is indeed the smallest reducing subspace of M that contains $\mathcal{S}$.

Remark 1. (Existence of the functional envelope) First, recall that we assume the existence of the central subspace $\mathcal{S}_{Y|X}$ throughout our exposition. However, it is possible that the central subspace does not exist, since it is possible that the intersection of some sufficient dimension reduction subspaces is no longer a sufficient dimension reduction subspace. In such cases, since the generalized eigenvalue problem (2.3) is still valid and meaningful, the envelope $\mathcal{E}_{\Sigma}(\Lambda)$ is also valid and preserves relevant information in the generalized eigenvalue problem. Secondly, the inversion Σ^{-1} may not exist and well-defined even when the central subspace exists. This

makes the envelope method even more appealing, comparing to the traditional functional inverse regression methods that involve Σ^{-1}. Henceforth, we will assuming existence of the central subspace and the covariance inversion, although the envelope methodology is still applicable without such assumptions.

Under the assumption that $X(t)$ is centered and has finite fourth moment, Σ has a spectral decomposition that $\Sigma(s,t) = \sum_{j=1}^{\infty} \theta_j \phi_j(s)\phi_j(t)$, where the eigenfunctions ϕ_j's form a complete orthonormal basis in $\mathcal{H}$ and the eigenvalues θ_j's satisfy the following conditions

$$\theta_1 > \theta_2 > \cdots > 0, \quad \sum_{j=1}^{\infty} \theta_j < \infty. \tag{2.6}$$

Such an assumption on distinct eigenvalues is commonly used in the literature (Yao et al., 2015, e.g.), for proving theoretical results and dealing with identification issues of eigenfunctions. However, it is worth mentioning that we can easily relax such a condition and still preserve our theoretical results in Theorems 3 and 4 because the notion of envelope is based on reducing subspaces, which are more general than eigenvectors or eigenfunctions. The nonzero eigenvalue assumption in (2.6) simplifies some technical proofs but is not required for envelope construction. Cook and Zhang (2017, Proposition 3) offered some insights and a detailed discussion on how the zero eigenvalue affects the dimension and construction of envelopes. Specifically, let A and A_0 be the basis matrices of the nonzero eigenspace and zero eigenspace, then the envelope in the following Proposition 2 can be written as $\mathcal{E}_{\Sigma}(\mathrm{span}(\Lambda)) = A\mathcal{E}_{A^T\Sigma A}(\mathrm{span}(A^T\Lambda A))$. Then we can focus on the envelope of $\mathcal{E}_{A^T\Sigma A}(\mathrm{span}(A^T\Lambda A))$, where $A^T\Sigma A$ automatically satisfies the nonzero eigenvalue condition. We assume this condition (2.6) for ease of interpretation and technical proofs. To gain more intuition about the functional envelope, we have the following property.

Proposition 2. $\mathcal{E}_{\Sigma}(\mathcal{S}_{Y|X}) = \mathcal{E}_{\Sigma}(\mathrm{span}(\Sigma^{-1}\Lambda)) = \mathcal{E}_{\Sigma}(\mathrm{span}(\Lambda)) = \oplus_{j=1}^{\infty}\mathrm{span}\{(\phi_j \otimes \phi_j)\Lambda\}$, *where $\oplus$ is the direct sum of subspaces and $\phi_j \otimes \phi_j$ is the rank-one projection operator onto the j-th eigenspace* $\mathrm{span}(\phi_j)$, $j = 1, 2, \ldots$.

The above result suggests that the envelope is the sum of all eigenspaces of Σ that are not

orthogonal to $\mathrm{span}(\Lambda)$. In other words, this means $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \oplus_{j\in\mathcal{J}}\mathrm{span}(\phi_j)$, where $\mathcal{J} = \{j \mid \langle\phi_j, \Lambda\phi_j\rangle \neq 0, j = 1, 2, \dots\}$ is the index set of the eigenvectors that are not orthogonal to $\mathrm{span}(\Lambda)$. This will connect our methodology closely with the functional principal components of Σ. Simply put, we are not selecting principal components from Σ, but instead, we are selecting eigenfunctions of Σ that intersect with $\mathrm{span}(\Lambda)$, which is equivalent to intersecting with the central subspace $\mathcal{S}_{Y|X}$. In the case where assumption (2.6) is violated, i.e. eigenvalues are not strictly decreasing so that some eigenvalues may have multiplicity greater than one, we can simply replace the rank-one projections $\phi_j \otimes \phi_j$ in Proposition 2 with projections onto eigensubspace that have dimension possibly greater than one due to the existence of common eigenvalues.

2.3 Functional envelope for sufficient dimension reduction

In functional sufficient dimension reduction literature, there is a key assumption that $\Sigma^{-1}\Lambda$ is well-defined in $\mathcal{H}$ because inversion of the operator Σ may not exist. Here we adopt the idea of dimension reduction without inverting Σ in Cook et al. (2007), and propose a class of dimension reduction methods for functional data without inversion of Σ.

Our goal is to estimate the central subspace $\mathcal{S}_{Y|X}$. However, instead of targeting at $\mathcal{S}_{Y|X} = \mathrm{span}(\Sigma^{-1}\Lambda)$, we consider aiming at the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$, which is the smallest reducing subspace of Σ that contains the central subspace $\mathcal{S}_{Y|X}$. By targeting at this larger dimension reduction subspace, as $\mathcal{S}_{Y|X} \subseteq \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$, we may avoid the inversion of Σ and provide a more robust estimation procedure inspired by the following property. Suppose the dimension of the envelope is $u \equiv \dim\{\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})\}$. Then $u \geq d = \dim\{\mathcal{S}_{Y|X}\}$ because the envelope contains the central subspace. Let $\gamma_1(t), \dots, \gamma_u(t)$ be an arbitrary set of linearly independent functions that spans the envelope, i.e. $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathrm{span}(\gamma_1, \dots, \gamma_u)$. Then we have the following two statements,

$$Y \perp\!\!\!\perp X \mid \langle\gamma_1, X\rangle, \dots, \langle\gamma_u, X\rangle; \tag{2.7}$$

$$\langle\alpha_0, \Sigma\alpha\rangle = 0, \quad \text{for any } \alpha \in \mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) \text{ and } \alpha_0 \in \mathcal{E}_\Sigma^\perp(\mathcal{S}_{Y|X}). \tag{2.8}$$

The first statement (2.7) implies that the envelope is a functional sufficient dimension reduction subspace; the second statement (2.8) further implies that any functional component of X in the envelope is uncorrelated with any functional component of X in the orthogonal complement of the envelope: in other words, $\langle \alpha, X \rangle$ is uncorrelated with $\langle \alpha_0, X \rangle$. The statement (2.7) is due to the fact that $\mathcal{S}_{Y|X} \subseteq \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ and the statement (2.8) holds because $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ is a reducing subspace of Σ (c.f. Proposition 1). The two statements together guarantee that the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ contains all the sufficient information in the regression, and moreover there is no leakage of information from the envelope via correlation in X.

An important advantage of targeting on the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ rather than on the central subspace $\mathcal{S}_{Y|X}$ is due to (2.8). Although the central subspace has the smallest dimensionality, the estimation of the central subspace often becomes unstable in presence of high correlation or co-linearity among predictors. For example, it is likely to happen, especially in functional data, that there exists a component $\beta \in \mathcal{S}_{Y|X}$ and another component $\beta_0 \in \mathcal{S}_{Y|X}^{\perp}$ such that $\langle \beta, X \rangle$ and $\langle \beta_0, X \rangle$ are highly correlated. Then the estimation of $\mathcal{S}_{Y|X}$ will be extremely difficult because it is hard to distinguish β from β_0 in practice. On the other hand, the estimation of the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ can be more stable because it targets at a subspace that possesses the property of (2.8) and in addition it requires no inversion of Σ as we will see in Section 3.

The following proposition is a constructive property of the functional envelope that motivates our estimation procedure in the next section.

Theorem 1. *For the sequence of subspaces defined as* $\mathcal{S}_k = \mathrm{span}(\Lambda, \Sigma\Lambda, \ldots, \Sigma^{k-1}\Lambda), \quad k = 1, 2, \ldots,$ *there exists an integer* K *such that*

$$\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_K = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathcal{S}_{K+1} = \mathcal{S}_{K+2} = \ldots. \tag{2.9}$$

If $d = u$*, then* $K = 1$*; if* $d < u$ *and there are* q *distinct eigenspaces of* Σ *not orthogonal to* $\mathcal{S}_{Y|X}$*, then* $K \leq q$*.*

This proposition is analogous to Theorem 1 of Cook et al. (2007) in the multivariate case. It indicates that the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ is a dimension reduction subspace that can be recovered by

subspace $\mathcal{S}_k$ with any k that $k \geq K$. This also suggest that selection of K is not a crucial task, overestimating K will not fail the estimation procedure of the envelope. In some applications such as functional partial least squares, $\Lambda(s,t)$ is replaced by an one-dimensional curve $\beta(t)$ and the series of subspaces $\mathcal{S}_k$, $k = 1, 2, \ldots$, becomes a Krylov sequence.

Since the dimension of the central subspace is assumed to be a fixed number $d = \dim\{\mathcal{S}_{Y|X}\} = \dim\{\text{span}(\Sigma^{-1}\Lambda)\}$, the rank of the kernel matrix Λ thus equals to d. Recall that Λ has rank d. We let $V_d = (\nu_1(t), \ldots, \nu_d(t))$ denote the d nonzero eigenvectors of Λ. We then have the following result to facilitate estimation of $\mathcal{S}_k$ in Theorem 1.

Theorem 2. *For any* $k = 1, 2, \ldots$, *let* $R_k = (V_d, \Sigma V_d, \ldots, \Sigma^{k-1} V_d)$. *Then* $\text{span}(R_k) = \text{span}(\Lambda, \Sigma\Lambda, \ldots, \Sigma^{k-1}\Lambda) = \mathcal{S}_k$.

Since $\mathcal{S}_k = \text{span}(R_k)$, for the estimation procedure described in the next section, we will focus on estimating $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ from spectral decomposition of R_k for some integer $k \geq K$ from (2.9). Recalling that $\dim\{\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})\} = u \geq d = \dim\{\mathcal{S}_{Y|X}\}$, we want to clarify that the number K is similar in spirit to the number of slices in sliced inverse regression (SIR; Li, 1991). While the dimension d is a critical hyper-parameter in SIR method, the number of slices is not that crucial but has to be no less than $d + 1$. From Theorem 1, we know that the size of the sequence of subspaces $\mathcal{S}_k$ will stop increasing after at most q steps, where q is the number of distinct eigenspaces of Σ not orthogonal to $\mathcal{S}_{Y|X}$ as remarked in Theorem 1.

Remark 2. (Connections to the functional partial least squares method). Analogous to the findings in Cook et al. (2013) that the popular partial least squares algorithm by De Jong (1993, SIMPLS) is essentially targeting at the multivariate predictor envelope, our results in Theorem 1 establish a connection between the functional partial least squares algorithm in Delaigle et al. (2012, APLS) and the functional envelope. One straightforward implication of Theorem 1 is that the APLS algorithm in Delaigle et al. (2012) is exactly targeting at the functional envelope $\mathcal{E}_\Sigma(\Lambda_{\text{PLS}})$, where the matrix $\Lambda_{\text{PLS}} = \text{cov}(XY)\text{cov}^T(XY)$ for the partial least squares regression model.

3 Estimation procedure and consistency

3.1 Estimation of FCS

The first step of the estimation procedure is to obtain a sample estimator of $\widehat{\Sigma}$ and $\widehat{\Lambda}$ for the generalized eigenvalue problem (2.3). The covariance operator can be the standard sample covariance for functional data. While there are many different methods of estimating $\widehat{\Lambda}$, our envelope estimation framework provides a generic method as an alternative to the generalized eigenvalue problem and thus can fit with any consistent estimator $\widehat{\Lambda}$. For illustration, we use $\widehat{\Lambda}$ from Yao et al. (2015), because the functional cumulative slicing method in Yao et al. (2015) and Zhu et al. (2010) avoids the selection of the number of slices in sliced inverse regression type methods (Li, 1991; Ferré et al., 2005). Details on obtaining the functional operator $\widehat{\Lambda}$ can be found in the orginal articles (e.g., Yao et al., 2015).

For completely observed (or fully observed at regular time points for all i.i.d. samples) functional data, $\widehat{\Sigma}(s,t) = n^{-1}\sum_{i=1}^{n} X_i(s)X_i(t)$ and $\widehat{\Lambda}(s,t) = n^{-1}\sum_{i=1}^{n} \widehat{m}(s,Y_i)\widehat{m}(t,Y_i)w(Y_i)$ where $\widehat{m}(t,\widetilde{y}) = n^{-1}\sum_{i=1}^{n} X_i(t)I(Y_i \le \widetilde{y})$ is the sample estimator for function $m(t,\widetilde{y}) = \mathrm{E}\{X(t)I(Y \le \widetilde{y})\}$ and $w(\cdot)$ is a given nonnegative weight function. We will use constant weights $w(\cdot) = 1$ for all our numerical studies for simplicity, as is also suggested in the original articles Zhu et al. (2010) and Yao et al. (2015).

For sparsely and irregularly observed functional data, Yao et al. proposed the following estimation for $\widehat{m}(t,\widetilde{y})$ and $\widehat{\Sigma}$ from local linear estimators (see Fan and Gijbels (1996); Yao et al. (2005a, 2015) for more details). Suppose X_i is observed in form of (T_{ij}, U_{ij}), $i = 1,\ldots,n$, $j = 1,\ldots,N_i$ and $U_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$ is the possibly contaminated observations with i.i.d. mean zero (unobservable) measurement error ε_{ij}. For $\widehat{m}(t,\widetilde{y})$, Yao et al. (2015) suggested to use the minimizer $\widehat{a}_0$ from

$$\min_{(a_0,a_1)} \sum_{i=1}^{n}\sum_{j=1}^{N_i} \{U_{ij}I(Y_i \le \widetilde{y}) - a_0 - a_1(T_{ij}-t)\}^2 K_1(\frac{T_{ij}-t}{h_n}),$$

where K_1 is a nonnegative and symmetric univariate kernel density and h_n is the bandwidth. Then $\widehat{\Lambda}(s,t) = n^{-1}\sum_{i=1}^{n} \widehat{m}(s,Y_i)\widehat{m}(t,Y_i)w(Y_i)$ is estimated in the same way as in the com-

petely observed functional data scenario. For $\widehat{\Sigma}(s,t)$, Yao et al. (2015, 2005a) suggested to use the minimizer $\widehat{b_0}$ from

$$\min_{(b_0,b_1,b_2)} \sum_{i=1}^{n} \sum_{j\neq l}^{N_i} \{U_{ij}U_{il} - b_0 - b_1(T_{ij}-s) - b_2(T_{il}-t)\}^2 K_2(\frac{T_{ij}-s}{h_n}, \frac{T_{il}-t}{h_n}),$$

where K_2 is a nonnegative bivariate kernel density and h_n is the bandwidth. The bandwidth can be chosen by cross-validation, and can be different in estimating $\widehat{m}$ and $\widehat{\Sigma}$. But for simplicity, we abuse the notation a bit and use the same h_n to denote the bandwidth. The asymptotic convergence of $\widehat{\Sigma}$ and $\widehat{\Lambda}$ has already been well-studied in Yao et al. (2015), which is summarized in the following Lemma.

The following conditions are commonly used regularity conditions for sparse functional data. Let interval $\mathcal{T} = [a,b]$ and then $\mathcal{T}^\delta = [a-\delta, b+\delta]$ for some $\delta > 0$. Density functions of time variable T is $f_1(t)$ and of bivariate time variables $(T_1,T_2)^T$ is $f_2(t,s)$

C1. The number of time points N_i's are independent and identically distributed as a positive discrete random variable N_n, where $\mathrm{E}(N_n) < \infty$, $\Pr(N_n \geq 2) > 0$ and $\Pr(N_n \leq M_n) = 1$ for some constant sequence M_n that is allowed to go to infinity as $n \to \infty$. Moreover, (T_{ij}, U_{ij}), $j \in \mathrm{J}_i$, are independent of N_i for $J_i \subseteq \{1,\ldots,N_i\}$.

C2. For nonnegative integers ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 = 2$, $\partial^2\Sigma(s,t)/(\partial s^{\ell_1}\partial t^{\ell_2})$ is continuous on $\mathcal{T}^\delta \times \mathcal{T}^\delta$ and $\partial^2 m(t,\widetilde{y})/\partial t^2$ is bounded and continuous for all $t \in \mathcal{T}$ and $\widetilde{y} \in \mathbb{R}$.

C3. For nonnegative integers ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 = 1$, $\partial f_2(s,t)/(\partial s^{\ell_1}\partial t^{\ell_2})$ is continuous on $\mathcal{T}^\delta \times \mathcal{T}^\delta$ and $\partial f_1(t)/\partial t$ is continuous on $\mathcal{T}^\delta$.

C4. Bandwidth $h_n \to 0$ and $nh_n^3/\log n \to \infty$ (univariate kernel) and $nh_n^2 \to \infty$ (bivariate kernel).

C5. The kernel functions are nonnegative with compact supports, bounded, and of order $(0,2)$ (univariate kernel) and $\{(0,0)^T, 2\}$ (bivariate kernel), respectively.

Lemma 1. *Under the regularity conditions C1-5, we have* $\|\widehat{\Sigma} - \Sigma\|_{\mathcal{H}} = O_p(n^{-1/2}h_n^{-1/2} + h_n^2)$ *and* $\|\widehat{\Lambda} - \Lambda\|_{\mathcal{H}} = O_p(n^{-1/2}h_n^{-1/2} + h_n^2)$.

3.2 Estimating the Σ-envelope of the central subspace

After obtaining $\widehat{\Sigma}$ and $\widehat{\Lambda}$, (Yao et al., 2015) (and most of the other functional SDR methods) truncate $\widehat{\Sigma}$ by keeping only its first s_n eigenvalues $\widehat{\theta}_j$ and eigenfunctions $\widehat{\phi}_j$, $j = 1, \ldots, s_n$, where s_n diverges with sample size n and is the adaptive number of components. Then use $\widehat{\Sigma}_{s_n} = \sum_{j=1}^{s_n} \widehat{\theta}_j \widehat{\phi}_j \otimes \widehat{\phi}_j$ and $\widehat{\Sigma}_{s_n}^{-1} = \sum_{j=1}^{s_n} \widehat{\theta}_j^{-1} \widehat{\phi}_j \otimes \widehat{\phi}_j$ for further estimating the central subspace from $\widehat{\Sigma}_{s_n}^{-1}\widehat{\Lambda}$. Instead of calculating the right d eigenfunctions of $\widehat{\Sigma}_{s_n}^{-1}\widehat{\Lambda}$ with a truncated $\widehat{\Sigma}_{s_n}$, we first calculate the leading d eigenvectors of $\widehat{\Lambda}$, denoted as $\widehat{V}_d = (\widehat{v}_1, \ldots, \widehat{v}_d)$. Then, in order to estimate the envelope, we compute the eigenvectors of $\widehat{R}_K$, which is defined as

$$\widehat{R}_K = (\widehat{V}_d, \widehat{\Sigma}\widehat{V}_d, \ldots, \widehat{\Sigma}^{K-1}\widehat{V}_d), \tag{3.1}$$

where no truncation of the covariance operator $\widehat{\Sigma}$ is required and the number K is defined in Theorem 1. The last step of the estimation procedure is to obtain a linearly independent functional basis for the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$. This can be easily done by eigen-decomposing $\widehat{R}_K$. The first u eigenfunction of $\widehat{R}_K$ will span a subspace that is consistent for the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$.

Our asymptotic results thus concern the consistency of $P_{\mathcal{E}}$, the projection onto the envelope $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$, and its estimate $\widehat{P}_\gamma$.

Theorem 3. *Under the regularity conditions C1-5, we have* $\|\widehat{P}_\gamma - P_{\mathcal{E}}\|_{\mathcal{H}} = O_p(n^{-1/2}h_n^{-1/2} + h_n^2)$.

3.3 Estimating the central subspace

If $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathcal{S}_{Y|X}$, then the estimation of envelope will also give an estimate of the central subspace. However, in many situations it is more likely that $\mathcal{S}_{Y|X} \subset \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$. Then the estimated u-dimensional envelope may be used as a stand-alone method for dimension reduction

because the envelope is, after all, a sufficient dimension reduction subspace. Alternatively, if the central subspace is the ultimate goal, one could also use the envelope as an upper bound of the central subspace and apply the following refining procedure to get an estimate of the central subspace.

Let $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_u$ be the first u eigenfunctions of $\widehat{R}_K$, and let $\widehat{P}_\gamma$ and $\widehat{Q}_\gamma$ be the projection onto $\mathrm{span}(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_u)$ and its orthogonal subspace, respectively. Then the envelope estimator for Σ and Λ is $\widehat{\Sigma}_{\mathrm{env}} = \widehat{P}_\gamma \widehat{\Sigma} \widehat{P}_\gamma + \widehat{Q}_\gamma \widehat{\Sigma} \widehat{Q}_\gamma$ and $\widehat{\Lambda}_{\mathrm{env}} = \widehat{P}_\gamma \widehat{\Lambda} \widehat{P}_\gamma$, respectively. Then the central subspace can be estimated from the d left eigen-functions of $(\widehat{P}_\gamma \widehat{\Sigma} \widehat{P}_\gamma)^\dagger \widehat{P}_\gamma \widehat{\Lambda} \widehat{P}_\gamma$, where $(\widehat{P}_\gamma \widehat{\Sigma} \widehat{P}_\gamma)^\dagger$ is the generalized inverse of the rank-u operator $\widehat{P}_\gamma \widehat{\Sigma} \widehat{P}_\gamma$. Equivalently, the central subspace can be estimated as $\mathrm{span}\{(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_u)\widehat{\Psi}_d\}$ where $\widehat{\Psi}_d = (\widehat{\psi}_1, \ldots, \widehat{\psi}_d) \in \mathbb{R}^{u \times d}$ is the coordinate matrix of the central subspace for Y on $Z \in \mathbb{R}^u$ with $Z_j = \langle \widehat{\gamma}_j, X \rangle$, $j = 1, \ldots, u$. The estimation of $\widehat{\Psi}_d$ can be achieved by any standard dimension reduction methods (Li, 1991; Cook and Ni, 2005; Zhu et al., 2010, e.g.).

Our asymptotic results concerns the consistency of $P_{\mathcal{S}}$, the projection onto the central subspace $\mathcal{S}_{Y|X}$, and its estimate $\widehat{P}_\beta$.

Theorem 4. *Under the regularity conditions C1-5, we then have* $\|\widehat{P}_\beta - P_{\mathcal{S}}\|_{\mathcal{H}} = O_p(n^{-1/2} h_n^{-1/2} + h_n^2)$.

3.4 Dimension selection

There are many ways to select the dimension d of the central subspace, including but not limited to sequential asymptotic or permutation tests (Schott, 1994; Cook et al., 2004, 2007; Zeng, 2008, e.g.), information criteria (Zhu et al., 2012; Ma and Zhang, 2015; Zhu et al., 2016, e.g.), plots (Luo and Li, 2016, e.g.), and cross-validations. Some of these methods in the literature, for determining the structural dimension d, are very generic and can be directly applied to our context of functional SDR (Zhu et al., 2016; Luo and Li, 2016, e.g.). Instead of developing a new method to select dimension in functional SDR (Li and Hsing, 2010, e.g.), we will be using cross-validation prediction error, which is arguably the most straightforward and intuitive cri-

terion, to select the dimension (d, u) simultaneously. We illustrate the empirical performances of cross-validation in our numerical studies in the following Section 4.3.

4 Simulations

4.1 Estimation comparison

In this section, we compare the functional envelope cumulative slicing (FECS) and the functional cumulative slicing (FCS) estimation of the central subspace. Recall from the beginning of Section 3.2 that the FCS truncates $\widehat{\Sigma}$ by keeping only its first s_n eigenvalues $\widehat{\theta}_j$ and eigenfunctions $\widehat{\phi}_j$, $j = 1, \ldots, s_n$, where s_n diverges with sample size n and is the adaptive number of components. Then use $\widehat{\Sigma}_{s_n} = \sum_{j=1}^{s_n} \widehat{\theta}_j \widehat{\phi}_j \otimes \widehat{\phi}_j$ and $\widehat{\Sigma}_{s_n}^{-1} = \sum_{j=1}^{s_n} \widehat{\theta}_j^{-1} \widehat{\phi}_j \otimes \widehat{\phi}_j$ for further estimating the central subspace from $\widehat{\Sigma}_{s_n}^{-1} \widehat{\Lambda}$. In the simulations studies, we investigates the effect of s_n on the performance of FCS and use a fixed number $s_n = 5$, 10, 20 and 30 instead of data dependent tuning parameter s_n that diverges with the sample size.

We use $\|\widehat{P}_\beta - P_{\mathcal{S}}\|_{\mathcal{H}}$ as the criterion for estimation performance of the two methods. We consider the following four models,

$$
\begin{aligned}
\text{Model} \quad \text{I}: \quad Y &= \langle \beta_1, X \rangle + \epsilon, \\
\text{Model} \quad \text{II}: \quad Y &= \arctan(\pi \langle \beta_1, X \rangle) + \epsilon, \\
\text{Model} \quad \text{III}: \quad Y &= \langle \beta_1, X \rangle + \exp(\langle \beta_2, X \rangle / 10) + \epsilon, \\
\text{Model} \quad \text{IV}: \quad Y &= \frac{1.5 \langle \beta_1, X \rangle}{0.5 + (1.5 + \langle \beta_2, X \rangle)^2} + 0.2\epsilon,
\end{aligned}
$$

where the dimension of the central subspace is $d = 1$ for Models (I) & (II) and $d = 2$ for Model (III) & (IV). The four models were chosen to illustrate four different types of models: (I) simple linear; (II) single-index non-linear; (III) additive model of linear and non-linear components; (IV) non-linear and non-additive model, which is a classical and widely used simulation model in sufficient dimension reduction literature (Li, 1991, e.g.). To mimic our spectroscopic data examples in the next section, we generated the functional predictor $X(t)$ from the regular, evenly-spaced, 100 grid points on the interval $t \in [0, 10]$, $X(t) = \sum_{j=1}^{100} \xi_j \phi_j(t)$ with $\phi_j(t) =$

$\sin(\pi jt/5)/\sqrt{5}$ or $\cos(\pi jt/5)/\sqrt{5}$ and $\xi_j \overset{i.i.d.}{\sim} N(0, \theta_j)$ for eigenvalue $\theta_j > 0$, $j = 1, \ldots, 100$. We consider the following three scenarios (also graphically illustrated in Figure 4.1) for the eigenvalues of the covariance operator $\Sigma(s, t)$.

- Scenario (a). We constructed eigenvalues that decay slowly, so that we can compare **robustness** of functional dimension reduction methods. The 100 eigenvalues are evenly spaced from 0.01 to 1, that means eigenvalues are $0.01k$ for $k = 1, \ldots, 100$.

- Scenario (b). We constructed eigenvalues that decay quickly after few large and close eigenvalues, so that we can compare **efficiency** of functional dimension reduction methods. The first six eigenvalues linearly decrease from 2.15 to 2.1 $k = 1, ..., 6$ and the remaining eigenvalues are $k^{-1.25}$ for $k = 7, ..., 100$;

- Scenario (c). We constructed the first ten eigenvalues as $2.0, 1.95, \ldots, 1.55$, and the remaining eigenvalues as $10k^{-1}$ for $k = 11, ..., 100$. We construct this scenario to be extremely **in favor of the FCS** estimator with truncated $\widehat{\Sigma}_{s_n}$ using the first ten functional principal components. We let the first ten eigenvalues well-separated and we also let the central subspace lies within the first ten eigenspace.

We use $\beta_1 = C_1\phi_5$ and $\beta_2 = C_2\phi_6$ such that the envelope is the central subspace $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathcal{S}_{Y|X}$ and thus $u = d$. Different normalizing constants C_1 and C_2 were used for each models such that the variance of $\langle \beta_i, X \rangle$, $\arctan(\pi\langle \beta_1, X \rangle)$ and $\exp(\langle \beta_2, X \rangle)$ are all close to 2.0 in Model I, II and III. For Model IV, the variances of both $\langle \beta_1, X \rangle$ and $(1.5 + \langle \beta_2, X \rangle)^2$ were controlled to be approximately 1.0. Therefore, we can directly compare FECS for estimating the envelope with the FCS that estimates the central subspace. For the envelope estimator, we simply use $K = u$ for the number of terms in R_K, this will guarentee the coverage of the envelope and the central subspace.

We simulated 100 data sets for each simulation settings, with $n = 100$ and 400 and summarized the results in Table 1. It is observed that the proposed FECS very competitively. It delivers the best performance for both Scenarios (a) and (b). Even for Scenario (c) which is

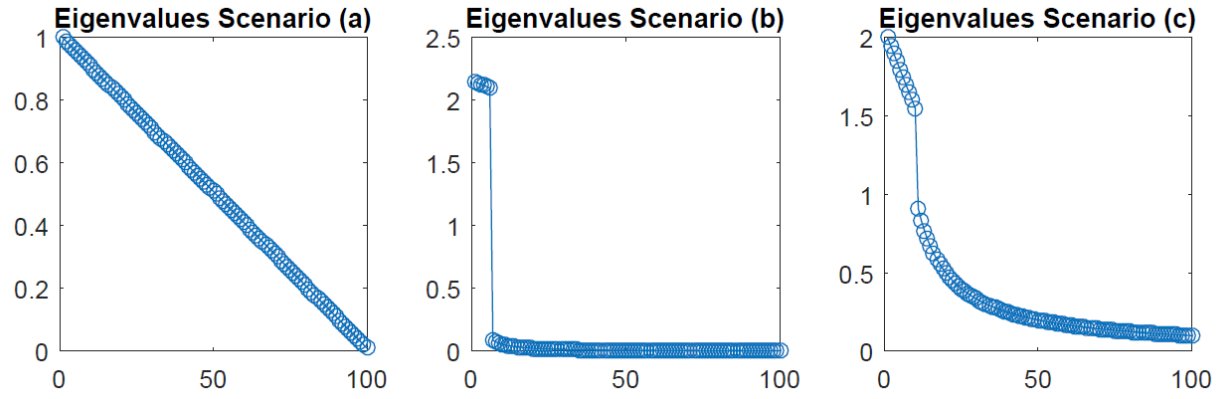

Figure 4.1: Three scenarios about the 100 eigenvalues of Σ.

especially designed to be in favor of FCS, the FECS's performance is very close to the best performer FCS with $s_n = 10$.

4.2 Prediction comparison

In this section, we compare the prediction performances of the FECS and the FCS estimators. For every simulated data set, we evaluate the prediction performance on an independent and identically generated testing data set, where we evaluate the relative prediction error as the criterion for prediction performance of the two methods. The relative prediction error is evaluated at the non-extrapolated values, and is defined as $\sum_{i=1}^{n_e}(\widehat{Y}_i - Y_i)^2/n_e/\widehat{\sigma}^2$, where n_e denotes the number of non-extrapolated Y_i's and $\widehat{\sigma}^2$ is the estimated variance of Y from testing data. To get predicted value $\widehat{Y}_i$, we used the Gaussian kernel smoothing with optimal bandwidth selection from Bowman and Azzalini (1997).

We used the same four models and three covariance operators as in Section 4.2. However, we changed the central subspace functional by letting $\beta_1 = C_1 \sum_{j=1}^{100} b_j \phi_j$ with $b_j = 1$ for $j = 1, 2, 3$, and $b_j = 4(j-2)^{-3}$ for $j = 4, \ldots, 100$, and keeping $\beta_2 = C_2\phi_6$ same as in Section 4.1. Normalizing constants C_1 and C_2 are chosen in the same way as in Section 4.1. That means now $\mathcal{S}_{Y|X} \subset \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ and $u > d$. In fact, this is an extreme case where the true population envelope dimension $u = 100$ equals the number of grid points $t \in \mathcal{T}$. Thus the

Model	n	FECS	FCS				S.E.$\leq$
			$s_n = 5$	$s_n = 10$	$s_n = 20$	$s_n = 30$	
(I-a)	100	**0.67**	0.93	0.87	0.78	0.73	0.01
	400	**0.40**	0.90	0.79	0.63	0.53	0.01
(II-a)	100	**0.72**	0.93	0.88	0.80	0.77	0.01
	400	**0.45**	0.90	0.80	0.64	0.56	0.01
(III-a)	100	**0.83**	0.95	0.92	0.87	0.85	0.01
	400	**0.67**	0.92	0.85	0.77	0.73	0.01
(IV-a)	100	**0.78**	0.94	0.89	0.83	0.81	0.01
	400	**0.53**	0.90	0.80	0.67	0.60	0.01
(I-b)	100	**0.23**	0.39	0.74	0.99	1.05	0.02
	400	**0.11**	0.41	0.71	1.01	1.09	0.02
(II-b)	100	**0.27**	0.41	0.78	1.00	1.05	0.02
	400	**0.14**	0.41	0.72	1.01	1.08	0.02
(III-b)	100	**0.51**	0.58	0.82	0.99	1.02	0.01
	400	**0.36**	0.47	0.77	0.99	1.04	0.01
(IV-b)	100	**0.36**	0.42	0.77	0.98	1.02	0.01
	400	**0.18**	0.36	0.66	0.96	1.02	0.01
(I-c)	100	**0.50**	0.73	0.52	0.54	0.65	0.01
	400	**0.28**	0.68	**0.28**	0.34	0.43	0.02
(II-c)	100	**0.54**	0.74	0.55	0.61	0.72	0.01
	400	**0.30**	0.69	**0.30**	0.38	0.48	0.02
(III-c)	100	**0.71**	0.81	**0.71**	0.74	0.80	0.01
	400	0.52	0.75	**0.49**	0.60	0.67	0.01
(IV-c)	100	0.62	0.76	**0.59**	0.68	0.78	0.01
	400	0.37	0.71	**0.32**	0.42	0.55	0.01

Table 1: Estimation Comparison. Averaged $\|\widehat{P}_\beta - P_{\mathcal{S}}\|_{\mathcal{H}}$ over 100 simulated data sets. We highlighted the best performance in bold. The last column "S.E.$\leq$" gives the largest standard error (S.E.) among all the five estimators (FECS, FCS with four different s_n values).

envelope estimator is essentially a finite sample approximation of the true envelope. However, as long as $u > d$, we can still estimate the central subspace at the right dimension and make prediction. We use 10-fold cross-validation to choose u and d for the FECS estimator, under the constraint that $u \geq d$. We use the true central subspace dimension d for the FCS estimator. Therefore, the simulation set-up is in favor of the FCS method. The results are summarized in Table 2 with the FECS delivering the best performance for all three eigenvalue scenarios. During the review process, one reviewer pointed out that the performance of FCS in Table 2 seems to keep getting better as s_n increases for some cases in eigen scenario (a) and concerned that it will beat the performance of FECS. While revising, we tried FCS with high s_n. The results confirmed that the performance of FCS will eventually deteriorate as s_n increases and FECS is indeed performing better than FCS. Yet to save space, we choose not to include the extended results here.

When prediction is the primary goal, kernel non-parametric regression techniques combined with functional PCA is widely applied (Bosq, 2000; Ferraty and Vieu, 2002, 2006; Ferraty et al., 2010, e.g.). We used a nonparametric functional PCA method that is implemented in the PACE (Principal Analysis by Conditional Expectation; `http://www.stat.ucdavis.edu/PACE/`) Matlab package to estimate eigenfunctions, where the number of eigenfunctions is chosen by one-curve-leave-out cross-validation procedures (Yao et al., 2005a). Then a multivariate kernel regression with Gaussian kernel on the eigenfunctions are fitted. The results are summarized in Table 2, where FPCA method was dominated by our FECS estimator but outperformed FCS in some model settings.

4.3 Dimension selection

As an illustration, we select (d, u) simultaneously based on the same 10-fold cross-validation selection procedure described in Section 4.2: we consider pairs of (d, u) satisfying $d \leq u$ and choose the pair with the smallest cross-validation prediction error. As an illustration, we take the classical sufficient dimension reduction model, model (IV) in the previous sections,

Model	n	FECS	FPCA (s_n)	FCS				S.E$\leq$
				$s_n = 5$	$s_n = 10$	$s_n = 20$	$s_n = 30$	
(I-a)	100	**0.44**	0.54 (12.7)	0.86	0.75	0.62	0.54	0.01
	400	**0.20**	0.45 (21.1)	0.74	0.59	0.41	0.32	0.01
(II-a)	100	**0.63**	0.65 (11.1)	0.91	0.82	0.73	0.68	0.01
	400	**0.38**	0.57 (18.8)	0.81	0.70	0.55	0.47	0.01
(III-a)	100	**0.50**	0.61 (13.6)	0.89	0.79	0.67	0.60	0.01
	400	**0.29**	0.53 (21.6)	0.79	0.67	0.49	0.40	0.01
(IV-a)	100	**0.66**	0.74 (11.0)	0.92	0.85	0.78	0.74	0.01
	400	**0.40**	0.64 (18.6)	0.84	0.75	0.65	0.60	0.01
(I-b)	100	**0.17**	0.32 (7.0)	0.30	0.29	0.49	0.59	0.02
	400	**0.09**	0.19 (7.0)	0.21	0.18	0.38	0.50	0.02
(II-b)	100	**0.37**	0.43 (6.8)	0.49	0.51	0.70	0.78	0.01
	400	**0.24**	0.30 (6.7)	0.36	0.35	0.56	0.66	0.01
(III-b)	100	**0.26**	0.33 (6.8)	0.36	0.35	0.53	0.63	0.01
	400	**0.19**	0.21 (6.8)	0.30	0.25	0.44	0.57	0.01
(IV-b)	100	**0.29**	0.54 (6.9)	0.55	0.52	0.61	0.67	0.01
	400	**0.15**	0.36 (6.6)	0.53	0.50	0.58	0.64	0.01
(I-c)	100	**0.26**	0.49 (11.4)	0.52	0.31	0.29	0.31	0.02
	400	**0.11**	0.34 (11.8)	0.35	0.13	0.13	0.14	0.02
(II-c)	100	**0.45**	0.59 (9.9)	0.65	0.47	0.48	0.52	0.01
	400	**0.27**	0.43 (10.2)	0.49	0.28	0.29	0.31	0.01
(III-c)	100	**0.33**	0.50 (12.1)	0.61	0.39	0.37	0.37	0.02
	400	**0.21**	0.36 (12.3)	0.50	0.23	0.22	0.23	0.01
(IV-c)	100	**0.46**	0.73 (10.0)	0.70	0.57	0.57	0.59	0.01
	400	**0.23**	0.57 (10.2)	0.60	0.49	0.49	0.50	0.01

Table 2: Prediction Performance. Averaged $n^{-1}\sum_{i=1}^{n}\|\widehat{Y}_i - Y_i\|$ for 100 training-testing data sets pairs. For every simulated data set, we evaluate the prediction performance on an independent and identically generated testing data set of size $10n$, where we evaluate the relative prediction error as the criterion for prediction performance of the two methods. FECS using 10-fold CV. FPCA is the functional PCA combined with kernel non-parametric regression prediction, where the average number of selected principal components is also included in the parenthesis. The last column "S.E.$\leq$" gives the largest standard error (S.E.) among all the five estimators (FECS, FCS with four different s_n values).

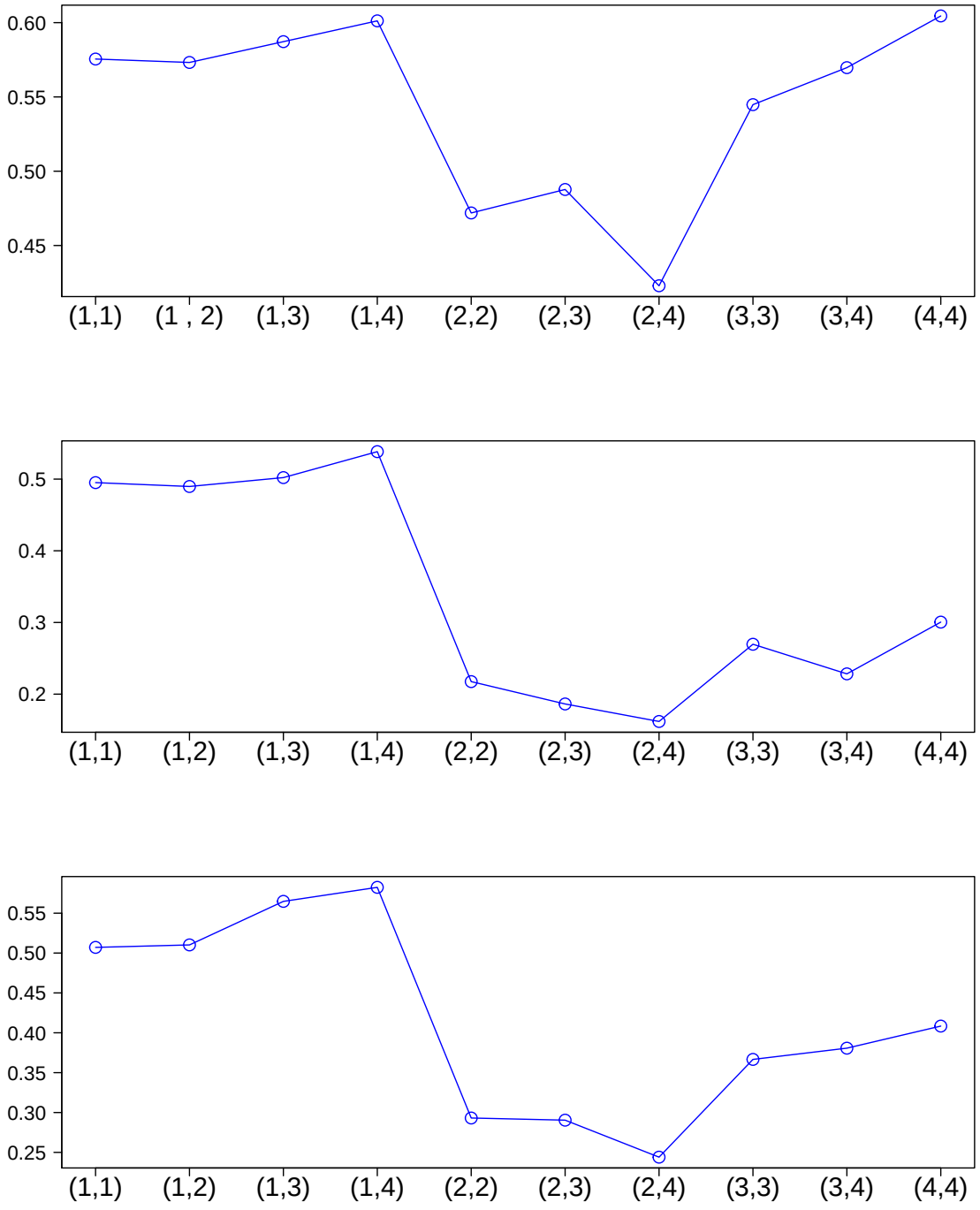

Figure 4.2: Averaged 10-fold cross-validation prediction errors for various dimensions, (d, u), in model (IV) with $n = 400$. From top to bottom, the three figures correspond to eigenvalue settings (a)–(c), respectively, in Figure 4.1.

$n = 400$	$d = 1$				$d = 2$			$d = 3$		$d = 4$	$d = \widehat{d}$
	$u = 1$	$u = 2$	$u = 3$	$u = 4$	$u = 2$	$u = 3$	$u = 4$	$u = 3$	$u = 4$	$u = 4$	
(IV-a)	0	1	2	1	20	9	63	0	2	2	92
(IV-b)	0	0	0	0	13	15	58	1	11	2	86
(IV-c)	0	0	0	0	22	22	52	1	1	2	96

Table 3: Illustration of Dimension Selection.

and we considered all three eigenvalue scenarios (i.e. Figure 4.1). We focus on the selection of the dimension d, which is more crucial than the envelope dimension u. We use the more challenging setting in Section 4.2, where the envelope structure dimension is $u = p = 100$ so that the envelope dimension is only a finite sample approximation, but the central subspace has the true dimension $d = 2$. For 100 replicate data sets with sample size $n = 400$, we have the dimension selection results summarized in Table 4, where it is clear that the dimension d can be correctly selected as we introducing the envelope dimension $u \geq d$. The envelope dimension in such case is acting like a tuning parameter that helps reducing the variability in the sample estimation procedure. Furthermore, Figure 4.2 summarized the averaged prediction performance for various dimensions. Again, we can see that the central subspace dimension d is crucial: underestimated dimension, $\widehat{d} = 1$, will always lead to poor prediction performance and overestimated dimension, $\widehat{d} = 3$ or 4, sometimes cause a drastic increase in prediction error (top panel) and sometimes only cause a small increase (middle and bottom panels); meanwhile, for each dimension d from 1 to 4, the relative prediction performance is not sensitive to the choices of envelope dimensions.

5 Real data

We consider the data example introduced at the end of Section 1, where Y_1 is the protein content and Y_2 is the moisture content; predictor $X(t)$ is the NIR absorption spectra that are measured at 351 equally spaced frequencies with a spacing of 4nm between 1100nm (first frequency) and 2500nm (last frequency). We first look at the prediction performance of the FECS estimator

		$d=1$	$d=2$	$d=3$	$d=4$
Freqency	Moisture	0	69	31	0
	Protein	7	29	64	0
Most frequent u	Moisture	NA	3	4	NA
	Protein	2	3	4	NA

Table 4: Selecting d and u . First two rows are the freqency of selected dimensions d based on prediction performance of FECS on 100 data splits, bottom two row indicates the most frequently selected u for each d, where "NA" indicates that the corresponding dimension d is not likely to be selected.

with various (d, u) combinations where $1 \leq d \leq u$. We constructed 100 data splits, each with 90 training samples and 10 testing samples, and the freqency of the selected dimensions are summarized in Table 4. The first functional principal component will cover more than 95% of the total variation, the first two PCs will cover more than 99%. Therefore, we also include the comparison with functional PCA in this data set with only the first two components. For the FCS method, we find $d = 2$ has the best predictive dimension for moisture and $d = 3$ is the best predictive dimension for protein. Overall prediction performances of each methods are summarized in Table 5. FECS is clearly the most robust and reliable dimension reduction method. In addition, we also compared with the functional kernel non-parametric regression (FKR) estimators (Ferraty and Vieu, 2002, 2006; Ferraty et al., 2010) in terms of prediction but not dimension reduction. From the results in Table 5, comparing to our FECS prediction, FKR had slightly better prediction for the protein content but much worse prediction for the moisture content.

We next plotted the first two dimension reduction directions of each methods in Figure 5.1 for protein content and in Figure 1.2 for moisture content, where we used $s_n = 5$ for the FCS and the optimal $u = 3$ for FECS. For both the protein content and the moisture content, FCS and FECS have similar findings. The correlation between the first directions of the two methods are 0.99 (protein) and 0.97 (moisture). For the second directions, FECS essentially find the direction that lies within the first two principal components. For predicting the moisture content, the functional PCA is clearly not effective. Therefore FECS agreed more with the

	FECS	PCA	FKR	FCS			S.E. $\leq$
				$s_n = 5$	$s_n = 10$	$s_n = 20$	
Moisture	0.18	0.78	0.61	0.16	0.39	0.45	0.02
Protein	0.68	0.70	0.62	0.81	0.82	0.82	0.02
Combined	0.86	1.48	1.23	0.97	1.21	1.27	0.04

Table 5: Prediction performance of each methods from 100 random data splits at testing/training ratio one to nine. The FECS use ten-fold cross-validation selected dimension (d, u) from the training set. The PCA use the first two components. The FCS use $d = 2$ for the moisture data and $d = 3$ for the protein data, and all s_n =5, 10 or 20 are reported in the table.

FCS and worked really well. Then in the protein data, the functional PCA are very effective. Correspondingly, FECS was similar to functional PCA in terms of prediction and is better than FCS.

Acknowledgment

We thank the editor, an associate editor, and two reviewers for constructive comments that greatly improved this manuscript. Wu is supported by NSF grant DMS-1055210. Zhang is supported by NSF grant DMS-1613154.

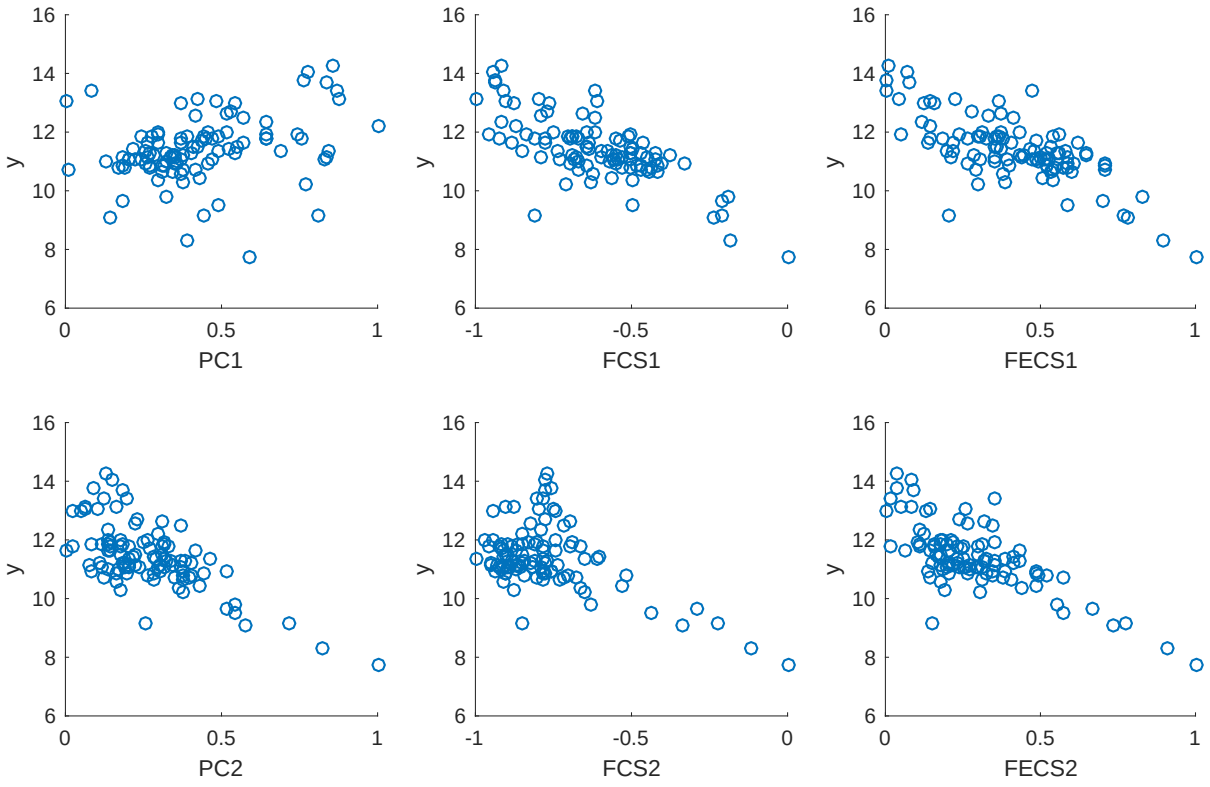

Figure 5.1: Protein content (y-axis) versus the six dimension reduction directions (x-axes): the first two principal components (PC1 and PC2 on the left column of plots); the first two directions from the functional cumulative slicing estimator (FCS1 and FCS2 on the middle column of plots); the first two directions from the functional envelope cumulative slicing estimator (FECS1 and FECS2 on the middle column of plots).

Appendix

A Proof of Proposition 1

Proof. The proof is analogous to the proof of Proposition 2.1 in Cook et al. (2010), for a $p \times p$ matrix M and its reducing subspace $\mathcal{R} \subseteq \mathbb{R}^p$, and is thus omitted.

□

B Proof of Proposition 2

Proof. From the definition of reducing subspace, every eigenspace of Σ is a reducing subspace of Σ. Moreover, due to the orthogonality of eigenspace, any reducing subspace of Σ can be writen in the form of $\oplus_{j\in\mathcal{J}}\text{span}(\phi_j) = \oplus_{j\in\mathcal{J}}\text{span}(\phi_j \otimes \phi_j)$ for some index set $\mathcal{J}$. Then by the definition of functional envelope, $\mathcal{E}_\Sigma(\text{span}(\Lambda))$ is the direct sum of all such subspaces that is not orthogonal to $\text{span}(\Lambda)$. Hence, we proved $\mathcal{E}_\Sigma(\text{span}(\Lambda)) = \oplus_{j=1}^{\infty}\text{span}\{(\phi_j \otimes \phi_j)\Lambda\}$, where $\text{span}\{(\phi_j \otimes \phi_j)\Lambda\} = \text{span}(\phi_j)$ if $\langle\phi_j, \Lambda\phi_j\rangle \neq 0$ and $\text{span}\{(\phi_j \otimes \phi_j)\Lambda\} = 0$ if $\langle\phi_j, \Lambda\phi_j\rangle = 0$. Use the same logic, we can get $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathcal{E}_\Sigma(\text{span}(\Sigma^{-1}\Lambda)) = \oplus_{j=1}^{\infty}\text{span}\{(\phi_j \otimes \phi_j)\Sigma^{-1}\Lambda\}$. Since Σ and Σ^{-1} share the same eigenvectors, $\text{span}\{(\phi_j \otimes \phi_j)\Sigma^{-1}\Lambda\} = \text{span}\{(\phi_j \otimes \phi_j)\Lambda\}$ for all $j = 1, 2, \ldots$. Therefore, $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \oplus_{j=1}^{\infty}\text{span}\{(\phi_j \otimes \phi_j)\Lambda\} = \mathcal{E}_\Sigma(\text{span}(\Lambda))$.

□

C Proof of Theorems 1 and 2

Proof. We prove Theorem 2 first. From the definition of V_d, $\text{span}(V_d) = \text{span}(\Lambda)$ and thus $\text{span}(\Sigma^t V_d) = \Sigma^t\text{span}(V_d) = \text{span}(\Sigma^t\Lambda)$ for all $t = 0, 1, \ldots$. Therefore, $\text{span}(R_k) = \text{span}(\Lambda, \Sigma\Lambda, \ldots, \Sigma^{k-1}\Lambda) = \mathcal{S}_k$ for any $k = 1, 2, \ldots$. This completes the proof of Theorem 2.

Next, to prove Theorem 1 based on the results from Theorem 2, it is sufficient to show the following two statements: (I) there exists an integer K such that $\text{span}(R_k) \subseteq \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ for $k < K$ and $\text{span}(R_k) = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ for $k \geq K$; (II) if $\text{span}(R_k) = \text{span}(R_{k+1})$ for some k,

then $\mathrm{span}(R_k) = \mathrm{span}(R_j)$ for all $j > k$. The statement (II) is needed to guarentee the strict increase of the sequence of subspaces, $\mathcal{S}_k \subset \mathcal{S}_{k+1}$, until we reach $k = K$. The proof follows the same logic as the proof of Theorem 1 in Cook et al. (2007) for multivariate case, and is a generalization of Cook et al. (2007).

Proof of statement (I). From Proposition 2, we know that $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \oplus_{j\in\mathcal{J}}\mathrm{span}(\phi_j)$, where $\mathcal{J} = \{j \mid \langle\phi_j, \Lambda\phi_j\rangle \neq 0\}$ is the index set of the eigenvectors that are not orthogonal to $\mathrm{span}(\Lambda)$. The dimension of the envelope, u, is hence equal to the size of the set $\mathcal{J}$. We re-arrange those u eigenvectors as $\widetilde{\phi}_1, \ldots, \widetilde{\phi}_u$ and re-arrange the distinct eigenvalues $\widetilde{\lambda}_1 > \cdots > \widetilde{\lambda}_q$, where $q \leq u$, and the corresponding projection matrices $\widetilde{P}_1, \ldots, \widetilde{P}_q$. Then $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \oplus_{j=1}^{u}\mathrm{span}(\widetilde{\phi}_j) = \oplus_{l=1}^{q}\mathrm{span}(\widetilde{P}_l)$, the projection onto $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ is $\sum_{l=1}^{q}\widetilde{P}_l$. Let $M_l = \widetilde{P}_l V_d$, then because $\mathrm{span}(V_d) \subseteq \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$, we have $V_d = \sum_{l=1}^{q}\widetilde{P}_l V_d = \sum_{l=1}^{q} M_l$. For any number $m = 0, 1, \ldots$, we have the following equalities

$$\Sigma^m V_d = \Sigma^m(\sum_{l=1}^{q}\widetilde{P}_l V_d) = \sum_{l=1}^{q}(\Sigma^m\widetilde{P}_l V_d) = \sum_{l=1}^{q}\widetilde{\lambda}_l^m\widetilde{P}_l V_d = \sum_{l=1}^{q}\widetilde{\lambda}_l^m M_l,$$

where the second to last equality is because that $\widetilde{P}_l$ is projection onto eigenfunctions of Σ, we have $\widetilde{P}_l\Sigma = \Sigma\widetilde{P}_l = \widetilde{\lambda}_l\widetilde{P}_l$ for any $l = 1, \ldots, q$, and thus, $\widetilde{P}_l\Sigma^m = \Sigma^m\widetilde{P}_l = \widetilde{\lambda}_l^m\widetilde{P}_l$. The operator R_k can therefore be expressed as

$$R_k = \left(V_d, \Sigma V_d, \ldots, \Sigma^{k-1}V_d\right) = \left(\sum_{l=1}^{q} M_l, \sum_{l=1}^{q}\widetilde{\lambda}_l M_l, \ldots, \sum_{l=1}^{q}\widetilde{\lambda}_l^{k-1} M_l\right),$$

which can be further re-expressed as matrix product $R_k = (M_1, \ldots, M_q) \cdot H_k$, where H_k is a $q \times k$ matrix with element $[H_k]_{ij} = \widetilde{\lambda}_i^{j-1}$, $i = 1, \ldots, q$, and $j = 1, \ldots, k$. It then follows that $\mathrm{span}(R_k) \subseteq \mathrm{span}(M_1, \ldots, M_u) = \mathrm{span}(\widetilde{P}_1, \ldots, \widetilde{P}_u) = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ for any k. Recall that the q eigenvalues are distint eigenvalues, thus by applying the well-know properties of the Vandermonde matrix, on H_k, we have $\det(H_k) \neq 0$ for $k < q$ and $\det(H_k) = 0$ for $k \geq q$. Therefore, exists an integer K, $K \leq q$, such that $\mathrm{span}(R_k) \subseteq \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ for $k < K$ and $\mathrm{span}(R_k) = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ for $k \geq K$.

Proof of statement (II). It is sufficient to show the following: if, for some k, $\mathrm{span}(\Sigma^k V_d) \subseteq$

$\mathrm{span}(R_k)$ then $\mathrm{span}(\Sigma^m V_d) \subseteq \mathrm{span}(R_k)$ for all $m > k$. The rest of proof follows from the proof of Theorem 1 in Cook et al. (2007) for multivariate case, and is thus omitted.

Finally, we have already shown the generic case of $K \leq q$ in the proof of statement (I). Now for the special case of $d = u$, it is clear that $\mathcal{S}_{Y|X} = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ in this case. Therefore $\mathrm{span}(\Lambda) = \Sigma\mathcal{S}_{Y|X} = \Sigma\mathcal{E}_\Sigma(\mathcal{S}_{Y|X}) = \mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$, where the last equality is because $\mathcal{E}_\Sigma(\mathcal{S}_{Y|X})$ is a reducing subspace of Σ. So we have $K = 1$ in (2.9).

□

D Proof of Theorems 3 and 4

Proof. The consistency of $\|\widehat{\Sigma}-\Sigma\|_{\mathcal{H}} = O_p(n^{-1/2}h_n^{-1/2}+h_n^2)$ and $\|\widehat{\Lambda}-\Lambda\|_{\mathcal{H}} = O_p(n^{-1/2}h_n^{-1/2}+h_n^2)$ can be found in Yao et al. (2015). Then the estimation procedure directly implies the same rate of convergence for $\widehat{P}_\gamma$ and $\widehat{P}_\beta$ since they are obtained from eigen-decompositions of matrix $\widehat{R}_k$, which consists of matrices $\widehat{\Sigma}$ and eigenfunctions of $\widehat{\Lambda}$.

□

References

Aneiros, G. and Vieu, P. (2016). Sparse nonparametric model for regression with functional covariate. *Journal of Nonparametric Statistics*, 28(4):839–859. 1

Bosq, D. (2000). Stochastic processes and random variables in function spaces. In *Linear Processes in Function Spaces*, pages 15–42. Springer. 4.2

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations: The kernel approach with S-Plus illustrations*. OUP Oxford. 4.2

Chen, D., Hall, P., Müller, H.-G., et al. (2011a). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, 39(3):1720–1747. 1

Chen, K., Chen, K., Müller, H.-G., and Wang, J.-L. (2011b). Stringing high-dimensional data for functional analysis. *Journal of the American Statistical Association*, 106(493):275–284. 1

Chen, T.-L., Huang, S.-Y., Ma, Y., Tu, I., et al. (2015). Functional inverse regression in an enlarged dimension reduction space. *arXiv preprint arXiv:1503.03673*. 1

Conway, J. (1990). *A Course in Functional Analysis. Second edition.* Springer, New York. 2.2

Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*, volume 318. John Wiley & Sons. 2.1

Cook, R. D., Forzani, L., and Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika*, 102(2):439–456. 1

Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(5):851–877. 1, 2.3

Cook, R. D., Li, B., and Chiaromonte, F. (2007). Dimension reduction in regression without matrix inversion. *Biometrika*, 94(3):569–584. 1, 2.2, 2.3, 2.3, 3.4, C

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica*, 20(3):927–960. 1, 2.2, 2.2, A

Cook, R. D., Li, B., et al. (2004). Determining the dimension of iterative hessian transformation. *The Annals of Statistics*, 32(6):2501–2531. 3.4

Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, 100(470):410. 2.1, 3.3

Cook, R. D. and Zhang, X. (2014). Fused estimators of the central subspace in sufficient dimension reduction. *Journal of the American Statistical Association*, 109(506):815–827. 2.1

Cook, R. D. and Zhang, X. (2015a). Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510):599–611. 1

Cook, R. D. and Zhang, X. (2015b). Simultaneous envelopes for multivariate linear regression. *Technometrics*, 57(1):11–25. 1

Cook, R. D. and Zhang, X. (2017). Fast envelope algorithms. *Statistica Sinica*, (just-accepted). 2.2

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23. 1

De Jong, S. (1993). Simpls: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3):251–263. 2.3

Delaigle, A., Hall, P., et al. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1):322–352. 1, 2.3

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press. 3.1

Ferraty, F., KEILEGOM, I. V., and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*, 37(2):286–306. 4.2, 5

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17(4):545–564. 4.2, 5

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media. 4.2, 5

Ferré, L., Yao, A., et al. (2005). Smoothed functional inverse regression. *Statistica Sinica*, 15(3):665. 1, 2.1, 2.1, 3.1

Ferré, L. and Yao, A.-F. (2003). Functional sliced inverse regression analysis. *Statistics*, 37(6):475–488. 1, 2.1

Geenens, G. et al. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Statistics Surveys*, 5:30–43. 1

Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, (146):1–6. 1

Jiang, C.-R., Yu, W., Wang, J.-L., et al. (2014). Inverse regression for longitudinal data. *The Annals of Statistics*, 42(2):563–591. 1, 2.1

Kalivas, J. H. (1997). Two data sets of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 37(2):255–259. 1.1, 1

Lee, C. E. and Shao, X. (2016). Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. *Journal of the American Statistical Association*, (just-accepted). 1

Li, B. and Song, J. (2016). Nonlinear sufficient dimension reduction for functional data. *The Annals of Statistics*, (just-accepted). 1

Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102(479):997–1008. 2.1

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327. 2.1, 2.3, 3.1, 3.3, 4.1

Li, L. and Zhang, X. (2016). Parsimonious tensor response regression. *Journal of the American Statistical Association*, (just-accepted). 1

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, 98(2):355–370. 1

Li, Y. and Guan, Y. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association*, 109(507):1205–1215. 1

Li, Y. and Hsing, T. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *The Annals of Statistics*, 38(5):3028–3062. 3.4

Li, Y., Wang, N., and Carroll, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108(504):1284–1294. 1

Luo, W. and Li, B. (2016). Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, (just-accepted). 3.4

Ma, Y. and Zhang, X. (2015). A validated information criterion to determine the structural dimension in dimension reduction models. *Biometrika*, page asv004. 3.4

Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression. *Journal of the American Statistical Association*, 89(425):141–148. 3.4

Silverman, B. and Ramsay, J. (2005). *Functional Data Analysis*. Springer. 2.1

Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, 98(1):133–146. 1

Wang, G., Zhou, Y., Feng, X.-N., and Zhang, B. (2015). The hybrid method of fsir and fsave for functional effective dimension reduction. *Computational Statistics & Data Analysis*, 91:64–77. 1

Yao, F., Lei, E., and Wu, Y. (2015). Effective dimension reduction for sparse functional data. *Biometrika*, 102(2):421–437. 1, 1, 2.1, 2.1, 2.2, 3.1, 3.2, D

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590. 1, 3.1, 4.2

Yao, F., Müller, H.-G., Wang, J.-L., et al. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903. 1

Yao, F., Wu, Y., and Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *Test*, 25(1):1–22. 1, 1

Zeng, P. (2008). Determining the dimension of the central subspace and central mean subspace. *Biometrika*, 95(2):469–479. 3.4

Zhang, X. and Li, L. (2016). Tensor envelope partial least squares regression. *Technometrics*, (just-accepted). 1

Zhu, L., Miao, B., and Peng, H. (2012). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*. 3.4

Zhu, L.-P., Zhu, L.-X., and Feng, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association*, 105(492):1455–1466. 2.1, 3.1, 3.3

Zhu, X., Wang, T., and Zhu, L. (2016). Dimensionality determination: a thresholding double ridge ratio criterion. *arXiv preprint arXiv:1608.04457*. 3.4