J Syst Sci Complex (2026) 39(1): 284–308

Envelope Inverse Regression for Dimension Reduction: A Review and New Perspectives*

ZENG Jing · WANG Ning · ZHANG Xin

DOI: 10.1007/s11424-026-5097-8
Received: 22 March 2025 / Revised: 15 May 2025

Abstract In this note, the authors revisit the envelope dimension reduction, which was first introduced for estimating a sufficient dimension reduction subspace without inverting the sample covariance. Motivated by the recent developments in envelope methods and algorithms, the authors refresh the envelope inverse regression as a flexible alternative to the existing inverse regression methods in dimension reduction. The authors discuss the versatility of the envelope approach and demonstrate the advantages of the envelope dimension reduction through simulation studies.

Keywords Dimension selection, envelope model, subspace estimation, sufficient dimension reduction.

1 Introduction

Sufficient dimension reduction (SDR) has been known as a powerful tool for data compression, visualization, and improved regression analysis[1, 2]. The fundamental objective of SDR is to seek a lower-dimensional representation of the predictor $\boldsymbol{X} \in \mathbb{R}^p$ that retains all the information about the conditional distribution of the response variable Y given $\boldsymbol{X}$. In linear SDR, the target is the smallest subspace $\mathcal{S} \subseteq \mathbb{R}^p$ such that

$$Y \perp\!\!\!\perp \boldsymbol{X} \mid \boldsymbol{P}_{\mathcal{S}}\boldsymbol{X}, \tag{1}$$

where $\boldsymbol{P}_{\mathcal{S}}$ denotes the projection matrix onto the subspace $\mathcal{S}$. Any subspace $\mathcal{S}$ satisfying (1) is referred to as a *dimension reduction subspace*. Equation (1) is equivalent to stating that the

ZENG Jing · WANG Ning
International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230026*, China; Center of Statistics and Data Science, Beijing Normal University at Zhuhai, Zhuhai* 519087*, China.* Email: zengjxl@ustc.edu.cn; ningwangbnu@bnu.edu.cn.
ZHANG Xin (Corresponding author)
Department of Statistics, Florida State University, Tallahassee FL 32306*, USA.* Email: henry@stat.fsu.edu.
*ZENG Jing's research was supported by the National Natural Science Foundation of China under Grant No. 12301365. WANG Ning's research was supported by the National Natural Science Foundation of China under Grant No. 2241200071 and Guangdong Basic and Applied Basic Research Foundation under Grant No. 2023A1515110001.
◇ *This paper was recommended for publication by Guest Editor LI Yingxing.*
†ZENG Jing and WANG Ning contributed equally to this work.

conditional distribution of $Y \mid \boldsymbol{X}$ is the same as $Y \mid \boldsymbol{P}_{\mathcal{S}}\boldsymbol{X}$, that is, projecting the predictor $\boldsymbol{X}$ onto $\mathcal{S}$ does not result in any loss of information regarding the conditional distribution of Y in regression. The intersection of all dimension reduction subspaces, provided it is a dimension reduction subspace itself, is called the *central subspace* (*CS*), denoted by $\mathcal{S}_{Y|\boldsymbol{X}}$.

A successful line of research has been devoted to the efficient estimation of (a part of) $\mathcal{S}_{Y|\boldsymbol{X}}$. [3] developed a non-parametric method, called minimum average variance estimation (MAVE). They formulated a minimization problem by approximating the mean function through a local linear expansion. By solving the minimization problem, they estimated the directions and the non-parametric link function simultaneously. Other non-parametric methods include [4] and [5]. [6] proposed a semi-parametric method by constructing a class of influence functions, based on which the estimators are formed. Another notable family of SDR methods is based on inverse regression–estimating (a portion of) the central subspace by non-parametric estimates of the conditional mean, covariance, and higher moments of $\boldsymbol{X} \mid Y$. Traditional inverse regression methods include sliced inverse regression [SIR; 53], sliced averaged variance estimation [SAVE; 54], directional regression [DR; 55], among others. Moreover, some dimension reduction notions and methodologies are developed specifically for handling categorical responses. [7] proposed the difference of covariances (DOC) method for dimension reduction in regression with binary responses. [8] proposed the notion of central discriminant subspace (CDS) for the dimension reduction in discriminant analysis. [9] coined the maximum separation subspace (MASES) in the context of regression and discriminant analysis with categorical response, which shares the same nice properties of CS. There are also some works devoting to the dimension reduction for functional data[10–13]. The SDR technique has also been used in model checking[14–17]. See [18–21] for review articles of the existing developments.

The traditional inverse regression methods require the sample size n to be larger than the dimensionality of the predictors p, as they need the sample covariance matrix to be invertible. When $n \leq p$ or n is not substantially larger than p, the inversion of the sample covariance becomes infeasible or unstable. Consequently, classical inverse regression-based SDR methods do not scale well to moderate or high-dimensional data. To enhance scalability, various sparse SDR methods have been developed, demonstrating favorable computational and theoretical properties, see [22–27], to name a few.

[28] introduced the envelope subspace to avoid the inversion of the sample covariance in inverse regression estimators. Their approach constructs the envelope via Krylov matrices, allowing the matrix inversion to be replaced by a power transformation of the sample covariance. This adjustment enables the estimation of $\mathcal{S}_{Y|\boldsymbol{X}}$ even in settings where n does not substantially exceed p. More properties and characterizations of the envelope were later established in [29] in the context of multivariate linear regression. Since then, various envelope models have been developed. In particular, extensive research has advanced in multivariate linear regression[30–37]. The envelope methodology has later been extended to other statistical models, including finite mixture models and clustering[38, 39], and other regression problems such as generalized linear models[40] and quantile regression[41]. Under these specific models, envelope has been shown to yield substantial efficiency gains in parameter estimation and prediction. In

addition to continuous data, related envelope models have also been developed for categorical response[42, 43], tensor data[44, 45], functional data[46, 47], and time series data[48]. In linear discriminant analysis, [42, 49] demonstrated envelope methods' advantages in classification and estimation. [40] provided the methodological foundations for constructing envelopes to improve existing multivariate statistical estimation procedures. However, these envelope methods do not scale well to high-dimensional data. Some works are devoted to developing sparse envelope methods, which allow for a diverging p. In multivariate linear regression, [50] proposed a sparse envelope estimator which achieves the response variable selection while preserving the efficiency gains offered by the envelope model. [51] developed a sparse envelope estimator for predictor reduction and variable selection. Nevertheless, their approaches assume that p does not diverge too quickly with n. [52] addressed the limitation of dimensionality by proposing a scalable estimation procedure based on a principal components regression formulation. In this paper, we focus on the model-free envelope, and argue that it may serve as a flexible and potentially useful alternative in SDR problems.

The rest of the paper is organized as follows. In Section 2, we provide a review of inverse regression-based SDR methods and the work of [28], where the envelope inverse regression subspace was introduced. Section 3 proposes the integration of the two types of modern envelope algorithms into the envelope inverse regression subspace estimation. This section also covers the statistical properties of envelope estimators, envelope dimension selection, and some extensions of envelope inverse regression. In Sections 4 and 5, we present numerical results based on simulated data and a real example. A brief discussion is provided in Section 6.

2 Inverse Regression Subspaces and Envelopes

2.1 Inverse Regression SDR Methods

Inverse regression methods are among the most widely used SDR methods, primarily due to their simplicity and computational efficiency. Let $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X})$ denote the covariance matrix of the predictors, and let $F_{\boldsymbol{X}|Y}$ represent the cumulative distribution function of the conditional distribution $\boldsymbol{X} \mid Y$. In inverse regression, a kernel matrix $\boldsymbol{U} \equiv \boldsymbol{U}(F_{\boldsymbol{X}|Y})$ is constructed such that

$$\boldsymbol{\Sigma}^{-1}\,\mathrm{span}(\boldsymbol{U}) \subseteq \mathcal{S}_{Y|\boldsymbol{X}} \tag{2}$$

under some mild conditions. The subspace $\boldsymbol{\Sigma}^{-1}\,\mathrm{span}(\boldsymbol{U})$ is referred to as the *inverse regression subspace*. The kernel matrix $\boldsymbol{U}$ is a function of the inverse regression distribution $F_{\boldsymbol{X}|Y}$, which explains the origin of the term "inverse regression". When sample estimates of $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{U}$ are available, the inverse regression subspace can be recovered through the spectral decomposition of the sample estimate of $\boldsymbol{\Sigma}^{-1}\boldsymbol{U}$. In the following, we review several representative inverse regression methods. To avoid notation proliferation, we use the general notation $\boldsymbol{U}$ to denote the kernel matrix across different methods.

Sliced inverse regression (SIR), introduced by [53], is the first inverse regression method. In SIR, the kernel matrix $\boldsymbol{U} = \mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}$ captures the subspace spanned by the conditional mean vectors $\mathrm{E}(\boldsymbol{X} \mid Y)$. To connect the inverse regression subspace in SIR with $\mathcal{S}_{Y|\boldsymbol{X}}$, a well-known linearity condition (LC) is required, which we state as follows:

(LC) Let $\boldsymbol{\eta}$ denote a basis matrix of $\mathcal{S}_{Y|\boldsymbol{X}}$, assume that $\mathrm{E}(\boldsymbol{X} \mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})$ is a linear function of $\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}$.

However, it is known that SIR loses effectiveness when the regression model contains a symmetric component since the kernel matrix in SIR only describes the variation in the conditional mean. [54] proposed the sliced averaged variance estimation (SAVE) method to capture more information contained in the conditional variance $\mathrm{Cov}(\boldsymbol{X} \mid Y)$. In SAVE, the kernel matrix is $\boldsymbol{U} = \mathrm{E}\{\mathrm{Cov}(\boldsymbol{X} \mid Y) - \mathrm{Cov}(\boldsymbol{X})\}^2$. By taking into account the conditional variance, SAVE is able to discover the directions missed by SIR when the regression model contains a symmetric component. In addition to LC, the connection between SAVE directions and $\mathcal{S}_{Y|\boldsymbol{X}}$ relies on the constant variance condition (CVC), stated in the following:

(CVC) Let $\boldsymbol{\eta}$ denote a basis matrix of $\mathcal{S}_{Y|\boldsymbol{X}}$, assume that $\mathrm{Cov}(\boldsymbol{X} \mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})$ is a non-stochastic matrix.

It has been widely accepted that SIR and SAVE have acknowledged limitations: SIR falls short when the model exhibits a symmetric structure, and SAVE becomes inefficient in seeking the monotone trend, especially when the sample size is limited. To mitigate these restrictions, [55] proposed the directional regression (DR) method to capture the information in both conditional mean and variance, integrating the strengths of SIR and SAVE. The inclusion (2) holds for DR method under LC and CVC.

When Y is continuous, all of SIR, SAVE, and DR demand the non-parametric estimation, e.g., $\mathrm{E}(\boldsymbol{X} \mid Y)$ and $\mathrm{Cov}(\boldsymbol{X} \mid Y)$. [56] proposed a cumulative slicing scheme, converting the conditional moments to unconditional moments, avoiding the non-parametric estimation. In parallel to SIR, SAVE, and DR methods, [56] proposed cumulative mean estimation (CUME), cumulative variance estimation (CUVE), and cumulative directional regression (CUDR) methods. We only review CUME and CUVE since CUDR is constructed quite similarly. The kernel matrix in CUME method is developed by converting the conditional mean $\mathrm{E}(\boldsymbol{X} \mid Y)$ in SIR to the unconditional mean $\mathrm{E}\{\boldsymbol{X}I(Y \le y)\}$ for some point y on the support of Y. [56] have proven that (2) holds under LC. Following the ideas in CUME, the CUVE method replaces the conditional variance $\mathrm{Cov}(\boldsymbol{X} \mid Y)$ in SAVE with the unconditional variance $\mathrm{Cov}\{\boldsymbol{X}I(Y \le y)\}$. It has been shown that under LC and CVC, (2) holds.

While the above-mentioned methods are all efficient in certain situations, they cannot be applied to multivariate response data directly. [57] proposed a high-dimensional inverse regression method, inspired by the martingale difference divergence matrix [MDDM; 58], a non-linear metric measuring the conditional mean dependence between two random objects. Their proposal adapts well to the multivariate response. It has been shown in [57] that the MDDM of $\boldsymbol{X} \mid \boldsymbol{Y}$ is the kernel matrix, and (2) holds under LC. For easy reference, we summarize all above-mentioned kernel matrices in Table 1.

Table 1 Kernel matrices in different inverse regression SDR methods. The random objects $\boldsymbol{X}'$, $\boldsymbol{Y}'$, and Y' denote the independent copies of $\boldsymbol{X}$, $\boldsymbol{Y}$, and Y, respectively

Method	Kernel matrix $\boldsymbol{U}$
SIR	$\mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}$
SAVE	$\mathrm{E}\{\mathrm{Cov}(\boldsymbol{X} \mid Y) - \mathrm{Cov}(\boldsymbol{X})\}^2$
DR	$\mathrm{E}\{\boldsymbol{\Lambda}(Y, Y') - 2\boldsymbol{\Sigma}\}^2$, where $\boldsymbol{\Lambda}(Y, Y') = \mathrm{E}\{(\boldsymbol{X} - \boldsymbol{X}')(\boldsymbol{X} - \boldsymbol{X}')^{\mathrm{T}} \mid Y, Y'\}$
CUME	$\mathrm{E}\{\boldsymbol{m}(Y')\boldsymbol{m}^{\mathrm{T}}(Y')\omega(Y')\}$, where $\boldsymbol{m}(Y') = \mathrm{E}\{\boldsymbol{X}I(Y \le Y')\}$ and $\omega(\cdot)$ is a nonnegative weight function
CUVE	$\mathrm{E}\{\boldsymbol{T}^2(Y')\}$, where $\boldsymbol{T}(Y') = \mathrm{Cov}\{\boldsymbol{X}I(Y \le Y')\} - F(Y')\boldsymbol{\Sigma}$ and $F(Y') = \mathrm{E}\{I(Y \le Y')\}$
MDDM	$\mathrm{MDDM}(\boldsymbol{X} \mid \boldsymbol{Y})$, where $\mathrm{MDDM}(\boldsymbol{X} \mid \boldsymbol{Y}) = -\mathrm{E}[\{\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\}\{\boldsymbol{X}' - \mathrm{E}(\boldsymbol{X}')\}^{\mathrm{T}}\|\boldsymbol{Y} - \boldsymbol{Y}'\|]$ and $\|\cdot\|$ denotes the Euclidean norm

We remark that the inverse regression subspaces of SIR, CUME, and MDDM, which are known as the first-order inverse regression methods because they focus on the conditional means, are the same. Similarly, the second-order inverse regression methods, including SAVE, DR, and CUVE, share the same population target. Moreover, as indicated by (2), the inverse regression methods are estimating a part of $\mathcal{S}_{Y|\boldsymbol{X}}$ under LC and CVC. However, in this note, all envelope methods we reviewed do not rely on the connection between $\boldsymbol{\Sigma}^{-1}\mathrm{span}(\boldsymbol{U})$ and $\mathcal{S}_{Y|\boldsymbol{X}}$. Without further explanation, we only focus on the inverse subspace $\boldsymbol{\Sigma}^{-1}\mathrm{span}(\boldsymbol{U})$ in the rest of the note.

2.2 Envelope in Sufficient Dimension Reduction

The envelope is a dimension reduction concept that extends beyond the dimension reduction subspace by incorporating additional structural constraints. While SDR is primarily developed in a model-free context and focuses on reducing the dimensionality of predictors, the envelope framework applies to both predictor and response reduction in both model-free and model-based settings. The concept of the envelope was first introduced by [28] and has since been further developed in both model-based and model-free contexts. Along with the conceptual development of envelopes, a variety of algorithms are also proposed for estimating the envelope. We first restate the formal definitions of the reducing subspace and the envelope in the following.

Definition 2.1 (Reducing subspace) A subspace $\mathcal{R} \subseteq \mathbb{R}^p$ is a reducing subspace of $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ if $\mathcal{R}$ decomposes $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \boldsymbol{P}_{\mathcal{R}}\boldsymbol{\Sigma}\boldsymbol{P}_{\mathcal{R}} + \boldsymbol{Q}_{\mathcal{R}}\boldsymbol{\Sigma}\boldsymbol{Q}_{\mathcal{R}}$, where $\boldsymbol{P}_{\mathcal{R}}$ is the projection matrix onto $\mathcal{R}$ and $\boldsymbol{Q}_{\mathcal{R}} = \boldsymbol{I}_p - \boldsymbol{P}_{\mathcal{R}}$ is the projection onto the orthogonal complement of $\mathcal{R}$. If $\mathcal{R}$ is a reducing subspace of $\boldsymbol{\Sigma}$, we say $\mathcal{R}$ reduces $\boldsymbol{\Sigma}$.

Definition 2.2 (Envelope) For some subspace $\mathcal{S} \subseteq \mathrm{span}(\boldsymbol{\Sigma})$, the $\boldsymbol{\Sigma}$-envelope of $\mathcal{S}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S})$, is the intersection of all reducing subspaces of $\boldsymbol{\Sigma}$ that contain $\mathcal{S}$. When $\mathcal{S}$ is represented as $\mathcal{S} = \mathrm{span}(\boldsymbol{U})$ for some matrix $\boldsymbol{U}$, the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S})$ can be equivalently expressed as $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}) \equiv \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$.

The $\boldsymbol{\Sigma}$-envelope of $\mathcal{S}_{Y|\boldsymbol{X}}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\boldsymbol{X}})$, improves the estimation efficiency in SDR

by imposing the additional independence/uncorrelatedness assumption between two orthogonal projections of $\boldsymbol{X}$. Let $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ be a basis matrix of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\boldsymbol{X}})$, where u is the dimension of the envelope. When $\boldsymbol{X}$ is from a normal distribution, according to Definition 2.2, it is implied that:

$$\text{(i) } Y \perp\!\!\!\perp \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{X} \mid \boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{X}, \quad \text{(ii) } \boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{X}, \tag{3}$$

where $\boldsymbol{P}_{\boldsymbol{\Gamma}}$ denotes the projection matrix onto the subspace span($\boldsymbol{\Gamma}$) and $\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{\Gamma}}$. Condition (i) in (3) indicates that $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\boldsymbol{X}})$ inherits the properties of the dimension reduction subspace, given that $\mathcal{S}_{Y|\boldsymbol{X}} \subseteq \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\boldsymbol{X}})$. Moreover, Condition (ii) in (3) is implied by the fact that $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{S}_{Y|\boldsymbol{X}})$ is a reducing subspace of $\boldsymbol{\Sigma}$. Specifically, according to the definition of the reducing subspace, $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \mathrm{Cov}(\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{X}, \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{X}) = 0$. When $\boldsymbol{X}$ is normally distributed, the uncorrelatedness between $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{X}$ and $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{X}$ implies independence. For general $\boldsymbol{X}$, the independence in Condition (ii) is replaced by uncorrelatedness. Similar to the position of the inverse regression subspace in SDR, the $\boldsymbol{\Sigma}$-envelope of the inverse regression subspace, spanned by $\boldsymbol{\Sigma}^{-1}\boldsymbol{U}$, would be a more constructive and estimable target in envelope dimension reduction.

In its first appearance, the $\boldsymbol{\Sigma}$-envelope of the inverse regression subspace, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$, is introduced by [28] to circumvent the inversion of the sample covariance in estimating the inverse regression subspace. According to the definition of envelope, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$ encapsulates $\boldsymbol{\Sigma}^{-1}\boldsymbol{U}$. Therefore,

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{U} = \boldsymbol{P}_{\boldsymbol{\Gamma}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}^{-1}\boldsymbol{U} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{U}, \tag{4}$$

where $\boldsymbol{P}_{\boldsymbol{\Gamma}(\boldsymbol{\Sigma})} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}$ is the projection onto span($\boldsymbol{\Gamma}$) with respect to the $\boldsymbol{\Sigma}$ inner product. According to (4), by projecting the inverse regression subspace $\boldsymbol{\Sigma}^{-1}$span($\boldsymbol{U}$) onto the larger envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$, the inversion of $\boldsymbol{\Sigma}$ is replaced by that of $\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Gamma}$, whose sample-level estimation has no computational issue as long as the dimension of the envelope u is small compared to the sample size n. In fact, the matrix inversion $\boldsymbol{\Sigma}^{-1}$ in the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$ can be omitted due to the results in [29], Proposition 2.4, i.e.,

$$\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U}).$$

In the following writing, we use the notation $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ to denote the $\boldsymbol{\Sigma}$-envelope of the inverse regression subspace, which we refer to as the *envelope inverse regression subspace*, unless otherwise specified.

[28] proposed an estimation approach of the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ by constructing the Krylov matrices, which are obtained by iteratively multiplying the kernel matrix $\boldsymbol{U}$ by the covariance $\boldsymbol{\Sigma}$:

$$\boldsymbol{K}_t = (\boldsymbol{U}, \boldsymbol{\Sigma}\boldsymbol{U}, \cdots, \boldsymbol{\Sigma}^{t-1}\boldsymbol{U}), \quad t = 1, 2, \cdots.$$

It has been confirmed in [28], Theorem 1, that the span of $\boldsymbol{K}_t$ will coincide with the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ at some point as u increases. Specifically, there exists an integer $\widetilde{t}$ such that span($\boldsymbol{K}_t$) is strictly increasing until $t = \widetilde{t}$, and settles upon $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ thereafter:

$$\mathrm{span}(\boldsymbol{K}_1) \subset \cdots \subset \mathrm{span}(\boldsymbol{K}_{\widetilde{t}}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U}) = \mathrm{span}(\boldsymbol{K}_{\widetilde{t}+1}) = \cdots, \tag{5}$$

where $\subset$ denotes the strict inclusion. Since the inverse regression subspace is contained in the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$, it is indicated by (5) that by iteratively transforming the kernel matrix $\boldsymbol{U}$ by $\boldsymbol{\Sigma}$, the increasing subspace sequence $\text{span}(\boldsymbol{K}_t)$ will capture the inverse regression subspace $\boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{U})$ at some point $t^* \leq \widetilde{t}$, that is, $\boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{U}) \subseteq \text{span}(\boldsymbol{K}_{t^*})$. Thus, similar to (4), the inverse regression subspace can be recovered with $\boldsymbol{K}_{t^*}$ through:

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{U} = \boldsymbol{K}_{t^*}(\boldsymbol{K}_{t^*}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{K}_{t^*})^{-1}\boldsymbol{K}_{t^*}^{\mathrm{T}}\boldsymbol{U}. \tag{6}$$

Let $\boldsymbol{B}_t = \boldsymbol{K}_t(\boldsymbol{K}_t^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{K}_t)^{-1}\boldsymbol{K}_t^{\mathrm{T}}\boldsymbol{U}$. To better track the iteration and locate the point t^*, [28] proposed the matrix $\boldsymbol{\Delta}_t = \boldsymbol{Q}_{\boldsymbol{B}_t(\boldsymbol{\Sigma})}\boldsymbol{B}_{t+1}$ for $t = 1, 2, \cdots$, to quantify the discrepancy between the successive two matrices $\boldsymbol{B}_t$ and $\boldsymbol{B}_{t+1}$. When $u \geq t^*$, $\text{span}(\boldsymbol{B}_t)$ coincides with $\text{span}(\boldsymbol{B}_{t+1})$ and $\boldsymbol{\Delta}_t = \boldsymbol{0}$. Otherwise, there is a non-empty overlap between $\text{span}(\boldsymbol{B}_t)$ and $\text{span}(\boldsymbol{B}_{t+1})$ such that $\boldsymbol{\Delta}_t \neq \boldsymbol{0}$. By tracking $\boldsymbol{\Delta}_t$, the point t^* can be easily located.

It is worth noting that though [28] introduced the envelope and developed the corresponding estimation method, the focus of [28] is still on the estimation of the inverse regression subspace. However, the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ itself can be regarded as the target subspace. In addition, it has been argued in the literature that while the construction of Krylov matrices is straightforward and computationally easy, it is less efficient compared with methods developed later[31]. In the next section, we move our focus to more efficient approaches for estimating the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ and review two envelope algorithms, which are based on the manifold optimization and sparse principal components.

3 Method

3.1 Envelope Estimation

In the following, we review two envelope estimation algorithms. The first algorithm is based on solving a manifold optimization problem. The second algorithm is the non-iterative envelope component estimation (NIECE) algorithm, recently proposed by [52]. The NIECE algorithm has non-sparse and sparse versions, used for handling low- and high-dimensional data.

The manifold-based algorithm is based on a generic population-level objective function $J(\boldsymbol{\Gamma})$. Without loss of generality, assume the kernel matrix $\boldsymbol{U}$ is a semi-positive definite matrix. Otherwise, we can always replace $\boldsymbol{U}$ by $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$, which spans the same subspace as $\boldsymbol{U}$. The objective function $J(\boldsymbol{\Gamma})$ is defined as $J(\boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^{\mathrm{T}}(\boldsymbol{\Sigma}+\boldsymbol{U})^{-1}\boldsymbol{\Gamma}|$, where $\boldsymbol{\Gamma} \in \mathcal{G}_{p,u}$ and $\mathcal{G}_{p,u}$ denote the (p,u)-Grassmann manifold composed of all u-dimensional basis matrices in $\mathbb{R}^p$ space. Let $\widetilde{\boldsymbol{\Gamma}} = \text{argmin}_{\boldsymbol{\Gamma}\in\mathcal{G}_{p,u}} J(\boldsymbol{\Gamma})$, it has been shown that

$$\text{span}(\widetilde{\boldsymbol{\Gamma}}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U}).$$

Therefore, one sample estimate of $\boldsymbol{\Gamma}$ is obtained by minimizing the sample version of the objective $J(\boldsymbol{\Gamma})$, denoted by $J_n(\boldsymbol{\Gamma})$. Let $\boldsymbol{X}_i$, $i = 1, \cdots, n$, denote the *i.i.d.* samples of $\boldsymbol{X}$, and $\mathbb{X}_n = (\boldsymbol{X}_1 - \overline{\boldsymbol{X}}, \cdots, \boldsymbol{X}_n - \overline{\boldsymbol{X}})^{\mathrm{T}} \in \mathbb{R}^{n\times p}$ denote the (centered) data matrix, where $\overline{\boldsymbol{X}} = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i$ is the overall sample mean of $\boldsymbol{X}$. Define the sample covariance $\widehat{\boldsymbol{\Sigma}} = n^{-1}\mathbb{X}_n^{\mathrm{T}}\mathbb{X}_n$.

Let $\widehat{\boldsymbol{U}}$ denote the sample estimate of the kernel matrix $\boldsymbol{U}$. The sample objective function $J_n(\boldsymbol{\Gamma})$ is given as follows,

$$J_n(\boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}} + \widehat{\boldsymbol{U}})^{-1}\boldsymbol{\Gamma}|. \tag{7}$$

Many off-the-shelf packages can be readily used for (approximately) solving the optimization problem (7). For instance, [59] provided a ready-to-use R package **TRES**, including the full Grassmann algorithm for solving the optimization problem (7) exactly, and other faster algorithms solving (7) approximately, including the one-dimensional (1D) algorithm[36] and the envelope coordinate descent (ECD) algorithm[60]. Though the 1D and ECD algorithms only solve (7) approximately, they are shown to provide $\sqrt{n}$-consistent estimators with much cheaper computational cost.

However, the manifold optimization-based algorithm does not scale well with high-dimensional data, which is accessed easily nowadays during the big-data era. [52] timely proposed a novel envelope estimation NIECE to fill the gap from both computational and theoretical perspectives. NIECE is known to be the first high-dimensional envelope method with statistical guarantees. Computationally, NIECE method overcomes the computational bottleneck of the existing FG optimization algorithms in high dimensions. Theoretically, [52] established the non-asymptotic analysis of envelope subspace estimation error. The NIECE method is motivated by the connection between envelope and principal components. Let $\boldsymbol{v}_j \in \mathbb{R}^p$ denote the j-th eigenvector of $\boldsymbol{\Sigma}$. For simplicity, it is assumed that the first $d \leq n$ eigenvalues of $\boldsymbol{\Sigma}$ are distinct and the span of the first d eigenvectors contains $\mathrm{span}(\boldsymbol{U})$. Then, the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ can be expressed in terms of $\boldsymbol{v}_j$ as follows,

$$\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U}) = \sum_{j=1, \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{U}\boldsymbol{v}_j \neq 0}^{d} \mathrm{span}(\boldsymbol{v}_j).$$

Given the envelope dimension u, there are exactly u eigenvectors $\boldsymbol{v}_j$ such that $\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{U}\boldsymbol{v}_j \neq 0$. The main idea of the NIECE algorithm is to estimate the sparse eigenvectors $\boldsymbol{v}_j$ such that $\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{U}\boldsymbol{v}_j \neq 0$. [52] defined the envelope score as $\phi_j = \boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{U}\boldsymbol{v}_j$, $j = 1, \cdots, d$, which is used to choose vectors in the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$. There are two versions of the NIECE algorithm, one without the sparsity and one with the sparsity, which differ by the estimation of eigenvectors. In the sample level, once the estimated eigenvectors $\widehat{\boldsymbol{v}}_j$'s are obtained, the sample version of the envelope score $\widehat{\phi}_j$ is estimated as $\widehat{\phi}_j = \widehat{\boldsymbol{v}}_j^{\mathrm{T}}\widehat{\boldsymbol{U}}\widehat{\boldsymbol{v}}_j$. The subspace spanned by $\widehat{\boldsymbol{v}}_j$'s corresponding to the largest u envelope scores $\widehat{\phi}_j$'s is the estimator of the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$. For the low-dimensional NIECE, $\boldsymbol{v}_j$ is estimated through the eigen-decomposition of the sample covariance $\widehat{\boldsymbol{\Sigma}}$. For the high-dimensional NIECE, the penalized matrix decomposition (PMD) method[61] is adopted for estimating the sparse eigenvector $\widehat{\boldsymbol{v}}_j$. In addition, to avoid handling the large-size matrix $\widehat{\boldsymbol{\Sigma}}$ in the high dimensions directly, the PMD$(\cdot, L_1)$ algorithm is applied to the data matrix $\mathbb{X}_n$ for obtaining the sparse right singular vectors of $\mathbb{X}_n$. We summarize the sparse NIECE algorithm in Algorithm 1.

Algorithm 1 Sparse NIECE algorithm

Input: Data matrix $\mathbb{X}_n = (\boldsymbol{X}_1 - \overline{\boldsymbol{X}}, \cdots, \boldsymbol{X}_n - \overline{\boldsymbol{X}})^{\mathrm{T}}$, sample estimate $\widehat{\boldsymbol{U}}$, number of principal components d, envelope dimension u, where $0 \leq u \leq d \leq p$.

Step 1 (Estimate sparse eigenvectors): Obtain d sparse right singular vectors of $\mathbb{X}_n$ by adopting the PMD$(\cdot, L_1)$ algorithm on $\mathbb{X}_n$. Let $\widehat{\boldsymbol{v}}_j$, $j = 1, \cdots, d$, denote the j-th right singular vector.

Step 2 (Calculate the envelope scores): Compute the envelop scores $\widehat{\phi}_j \equiv \widehat{\boldsymbol{v}}_j^{\mathrm{T}} \widehat{\boldsymbol{U}} \widehat{\boldsymbol{v}}_j$ for $j = 1, \cdots, d$, and organize them in descending order $\widehat{\phi}_{(1)} \geq \cdots \geq \widehat{\phi}_{(d)}$ and define $\widehat{\boldsymbol{v}}_{(j)}$ such that $\widehat{\phi}_{(j)} \equiv \widehat{\boldsymbol{v}}_{(j)}^{\mathrm{T}} \widehat{\boldsymbol{U}} \widehat{\boldsymbol{v}}_{(j)}$

Output: The estimated envelope is $\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{U}) = \mathrm{span}(\widehat{\boldsymbol{v}}_{(1)}, \cdots, \widehat{\boldsymbol{v}}_{(u)})$.

3.2 Statistical Properties

In this section, we provide some new theoretical results for the envelope inverse regression methods. As a demonstration purpose, we use the envelope-SIR problem as an example. We extend the asymptotic results from the FG algorithm and the non-asymptotic results from the high-dimensional NIECE algorithm to the specific envelope-SIR problem.

In envelope-SIR, the goal is to estimate the envelope $\mathcal{S}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$, where $\boldsymbol{U}$ is the kernel matrix in SIR. When Y is continuous, the kernel matrix is often recovered by $\boldsymbol{U} \equiv \mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid Y)\} = \mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid \widetilde{Y})\}$, where $\widetilde{Y}$ is the discretized version of Y. The discretized response $\widetilde{Y}$ is constructed by first dividing the range of Y into H disjoint intervals and letting $\widetilde{Y} = h$ if the corresponding Y is in the h-th interval. Let $(\boldsymbol{X}_i, Y_i)$, $i = 1, \cdots, n$, denote the *i.i.d.* samples of $(\boldsymbol{X}, Y)$. Moreover, let $\widehat{\boldsymbol{U}}$ denote the sample estimator of $\boldsymbol{U}$, that is $\widehat{\boldsymbol{U}} = \sum_{h=1}^{H} (n_h/n)(\widehat{\boldsymbol{\mu}}_h - \overline{\boldsymbol{X}})(\widehat{\boldsymbol{\mu}}_h - \overline{\boldsymbol{X}})^{\mathrm{T}}$, where $n_h = \sum_{i=1}^{n} I(\widetilde{Y}_i = h)$ is the sample size in slice h and $\widehat{\boldsymbol{\mu}}_h = n_h^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i I(\widetilde{Y}_i = h)$ is the sample mean of $\boldsymbol{X}$ in slice h with $I(\cdot)$ being the indicator function. In the low-dimensional setting, it has been shown in [36], Proposition 3, that the solution from the manifold-based algorithm is guaranteed to have $\sqrt{n}$-consistency once the estimators $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$ are $\sqrt{n}$-consistent. The $\sqrt{n}$-consistency of the estimator from the envelope-SIR is given in the following.

Theorem 3.1 *Let $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$ be the sample estimators of the covariance matrix $\boldsymbol{\Sigma}$ and the kernel matrix in SIR $\boldsymbol{U}$, and let $\widehat{\boldsymbol{\Gamma}}$ be the solution of* (7)*. The estimator $\boldsymbol{P}_{\widehat{\boldsymbol{\Gamma}}}$ is $\sqrt{n}$-consistent for the projection onto $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$.*

In the high-dimensional setting, [52] established the non-asymptotic error bounds for the sparse NIECE algorithm. Their results provide insights on how the estimation error of the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ interplays with the envelope dimension, the estimation error of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$, the eigen-gap of $\boldsymbol{\Sigma}$, and the envelope score gap. Once the non-asymptotic max-norm bounds for $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$ are given, the generic error bound results in [52] can be easily extended to the envelope-SIR problem. In high dimensions, [25] derived the non-asymptotic error bound for another widely used estimator for the kernel matrix $\boldsymbol{U}$. Let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ denote the ordered response and $\boldsymbol{X}_{(i)}$ denote the predictor associated with $Y_{(i)}$. According to the identity that $\boldsymbol{U} \equiv \mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid Y)\} = \boldsymbol{\Sigma} - \mathrm{E}\{\mathrm{Cov}(\boldsymbol{X} \mid Y)\}$, the kernel matrix $\boldsymbol{U}$ can be estimated by $\widehat{\boldsymbol{U}} = \widehat{\boldsymbol{\Sigma}} - \widehat{\mathrm{E}}\{\mathrm{Cov}(\boldsymbol{X} \mid Y)\}$, where

$$\widehat{\mathrm{E}}\{\mathrm{Cov}(\boldsymbol{X} \mid Y)\} = \frac{1}{n}\sum_{i=1}^{n/2}\{\boldsymbol{X}_{(2i)} - \boldsymbol{X}_{(2i-1)}\}\{\boldsymbol{X}_{(2i)} - \boldsymbol{X}_{(2i-1)}\}^{\mathrm{T}}.$$

To establish the non-asymptotic results for $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$, the following two assumptions are needed:

(A1) Suppose that $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are i.i.d. sub-Gaussian random vectors, that is, for some constant $\kappa > 0$ and any $\boldsymbol{a} \in \mathbb{R}^p$, $\mathrm{E}\{\exp(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}_i)\} \le \exp(\kappa^2\|\boldsymbol{a}\|_2^2/2)$, $i = 1, \cdots, n$.

(A2) Let $B > 0$ and $\mathcal{C}(B)$ be the collection of all the n-points partitions $-B \le y_{(1)} \le \cdots \le y_{(n)} \le B$ on the interval $[-B, B]$. For any fixed $B > 0$,

$$\lim_{n\to\infty} \frac{1}{n^{1/4}} \sup_{\mathcal{C}(B)} \sum_{i=1}^{n-1} \|\boldsymbol{m}(y_{(i+1)}) - \boldsymbol{m}(y_{(i)})\|_\infty = 0,$$

where $\|\boldsymbol{a}\|_\infty = \max_j |a_j|$.

Moreover, assume that the first d eigenvalues of $\boldsymbol{\Sigma}$ are distinct. Define the minimal eigen-gap of the first d eigenvalues of $\boldsymbol{\Sigma}$ as $\Delta = \min_{j=1,\cdots,d}\{\lambda_j(\boldsymbol{\Sigma}) - \lambda_{j+1}(\boldsymbol{\Sigma})\}$, and the sparsity in the true eigenvectors $\boldsymbol{v}_k$ as $\tau_0 = \max_{1\le k\le d}\|\boldsymbol{v}_k\|_1$. With the sorted envelope scores $\phi_{(1)} \ge \cdots \ge \phi_{(u)} > \phi_{(u+1)} = \cdots = \phi_{(d)} = 0$, the envelope score gap is defined as $\Delta_{\boldsymbol{U}} = \phi_{(u)} - \phi_{(u+1)} = \boldsymbol{v}_{(u)}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{v}_{(u)}$. In addition, let $\nu = \|\boldsymbol{U}\|_{\mathrm{op}}$ denote the largest eigenvalue of $\boldsymbol{U}$. The estimation error of the NIECE estimator in envelope-SIR is given as follows.

Theorem 3.2 *Under Assumptions* (A1) *and* (A2)*, if* $\tau > \tau_0$*,* $\varepsilon \ge c_0 n^{-1/2}$*, and* $\Delta_{\boldsymbol{U}} > \sqrt{8c_1\nu^2\tau^2\varepsilon} + (2c_1\nu + 1)\tau^2\varepsilon$ *for some constants* $c_0, c_1 > 0$*, there exist some constants* $C, C' > 0$ *that do not depend on* n*,* p*, or* ε *such that*

$$\|\boldsymbol{P}_{\widehat{\mathcal{E}}} - \boldsymbol{P}_{\mathcal{E}}\|_F \le C\varepsilon\tau^2,$$

with probability at least $1 - Cp^2\exp(-C'n\varepsilon^2)$*, where* $\|\cdot\|_F$ *denotes the Frobenius norm of a matrix. Moreover, if* $\log p = o(n)$*, and* $\Delta_{\boldsymbol{U}}$ *and* τ *satisfy* $\max\{\tau^2\sqrt{\log p}, \Delta_{\boldsymbol{U}}^{-2}\tau^2\sqrt{\log p}\} = o(\sqrt{n})$*, we have that* $\|\boldsymbol{P}_{\widehat{\mathcal{E}}} - \boldsymbol{P}_{\mathcal{E}}\|_F^2 \to 0$ *in probability as* $n, p \to \infty$*.*

In fact, the theoretical results of the manifold-based algorithm and sparse NIECE algorithm are fairly generic in the sense that they can be readily applied in any context as long as some decent properties of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$ are verified. Following the same idea, the results in Theorems 3.1 and 3.2 can be easily extended to other envelope SDR methods. For instance, we can obtain the similar asymptotic and non-asymptotic results for envelope-SAVE, envelope-DR, envelope-CUME, envelope-CUVE, and envelope-MDDM, once we derive the $\sqrt{n}$-consistency or the error bound of $\widehat{\boldsymbol{U}}$ for each method-specific kernel matrix.

3.3 Envelope Dimension Selection

In the development of envelope estimation procedures and algorithms, it is typically assumed that the envelope dimension is known. However, in real applications, the envelope dimension is usually unknown, and the accurate selection of the envelope dimension is a theoretically challenging but crucial issue. [32] proposed a likelihood-ratio test for dimension selection, [35]

developed a sequential asymptotic χ^2-test, and [62] derived the selection consistency of BIC under the scaled envelope model. Nevertheless, these dimension selection methods all rely on the linear model assumption and normality assumption. The dimension selection consistency of these methods becomes unclear when the model or distributional assumption is violated. To address the limitation of model assumptions, [63] proposed a selection approach without requiring distributional or model assumptions in the low-dimensional setting, which is shown to enjoy the selection consistency.

In the low-dimension setting, we select the dimension of the envelope inverse regression subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ following the procedure of [63], which relies on the manifold-based algorithm. Recall that the kernel matrix $\boldsymbol{U}$ is semi-positive definite and can be replaced by $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$ otherwise. Let $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$ denote the sample estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{U}$. For each $k = 1, \cdots, p$, define the estimator $\widehat{\boldsymbol{\Gamma}}_k$ as $\widehat{\boldsymbol{\Gamma}}_k = \mathrm{argmin}_{\boldsymbol{\Gamma} \in \mathcal{G}_{p,k}} J_n(\boldsymbol{\Gamma})$, where $J_n(\boldsymbol{\Gamma})$ is the sample objective function defined in (7). Then, the information criterion is defined as

$$\mathcal{I}_n(k) = J_n(\widehat{\boldsymbol{\Gamma}}_k) + \frac{Ck \log n}{n}, \quad k = 0, 1, \cdots, p,$$

where C is a positive constant and $\mathcal{I}_n(0) = 0$. The selected envelope dimension is given by $\widehat{u} = \mathrm{argmin}_{0 \le k \le p} \mathcal{I}_n(k)$. It has been shown that the criterion $\mathcal{I}_n(k)$ is highly related to other likelihood-based BIC[29–31]. However, $\mathcal{I}_n(k)$ is a totally likelihood-free function. In the high-dimensional setting, we treat the envelope dimension as a tuning parameter and estimate it via data-driven approaches such as cross-validation. For instance, for a candidate dimension k, we first obtain an estimator $\widehat{\boldsymbol{\Gamma}}_k$ of the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ using either the manifold-based algorithm or the NIECE algorithm. We then reduce the predictors $\boldsymbol{X}$ to $\widehat{\boldsymbol{\Gamma}}_k^{\mathrm{T}} \boldsymbol{X}$, and evaluate the association between response Y and the reduced predictors $\widehat{\boldsymbol{\Gamma}}_k^{\mathrm{T}} \boldsymbol{X}$ using some criterion such as the distance covariance[64]. The envelope dimension is selected as the k yielding the best criterion value (e.g., the largest distance covariance).

3.4 Extensions and Generalization

In this section, we provide some extensions and generalizations of envelope inverse regression and envelope algorithms, including other versions of kernel matrices, variations of envelope algorithms, extensions of the envelope to tensor data, and semi-supervised settings.

Hybrid kernel matrices. In the literature, it has been shown that hybrid methods which combine inverse regression approaches can uncover more comprehensive data structures than solely using one method. For example, [53] proposed the SIR-II method, which combines SIR and SAVE, while [65] studied combinations of SIR and Principal Hessian Directions (PHD) methods[66]. Additionally, [67] systematically explored convex combinations of SIR and SAVE with the kernel matrix defined as:

$$\boldsymbol{U} = (1 - a)\mathrm{Cov}\{\mathrm{E}(\boldsymbol{X} \mid Y)\} + a\mathrm{E}\{\mathrm{Cov}(\boldsymbol{X} \mid Y) - \mathrm{Cov}(\boldsymbol{X})\}, \tag{8}$$

where $0 \le a \le 1$. By taking the kernel matrix as $\boldsymbol{U}$ defined in (8), the estimate of envelope $\mathcal{S}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$ potentially captures more comprehensive dimension reduction structures than SIR and SAVE, inheriting the advantages of the hybrid SDR methods.

Highly correlated predictors. When the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$ itself is the target parameter, the envelope algorithm can be applied to an already obtained estimator of the inverse regression subspace $\boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{U})$. This is motivated by the fact that the envelope structure enhances the efficiency of parameter estimation, particularly when predictors exhibit strong correlations. As shown in (3), the envelope takes into account the interaction between the material part $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{X}$ and the immaterial part $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{X}$. In contrast to SDR, which ignores the interaction among covariates, the envelope model accounts for these correlations, leading to substantial gains in parameter estimation and statistical inference. In the context of multivariate linear regression, [31] demonstrated that the envelope model provides predictive improvements when predictors are strongly correlated. Thus, the envelope can be viewed as a more efficient alternative to SDR methods when the highly correlated covariates are present.

Sparse kernel matrices. In the high-dimensional setting, the sparsity assumption is often imposed on the inverse regression subspace $\boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{U})$ for improving the estimation efficiency. In the literature, various sparse SDR methods have been developed for estimating $\boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{U})$, such as [24–26], and [27], to name a few. Let $\boldsymbol{\beta} \in \mathbb{R}^{p\times d}$ denote the basis matrix of the inverse regression subspace such that $\text{span}(\boldsymbol{\beta}) = \boldsymbol{\Sigma}^{-1}\,\text{span}(\boldsymbol{U})$. Then, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})$. Therefore, once the sparse estimator $\widehat{\boldsymbol{\beta}}$ is available, we can implement the sparse NIECE algorithm with the inputs $\widehat{\boldsymbol{U}} = \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}$ to construct the estimator of envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})$.

Variations of envelope algorithms. NIECE is a generic algorithm for estimating envelopes, including the envelope inverse regression subspace. For instance, by replacing $\widehat{\boldsymbol{U}}$ in the NIECE algorithm with the hybrid kernel matrix or sparse kernel matrix, we can obtain the corresponding envelope estimators. The NIECE algorithm is also flexible in that the estimation of eigenvectors $\boldsymbol{v}_j$ can be achieved through any sparse principal component method. In addition to the PMD algorithm, methods such as [68, 69], and [70] can also be used to estimate the sparse eigenvectors of $\boldsymbol{\Sigma}$. The finite-sample properties of NIECE can be easily inherited as long as the consistency results for these sparse principal component methods are available.

Extension to tensor data. In recent decades, the envelope model has also been extended to tensor data, which is a multi-dimensional array. Two special cases of tensor are vector and matrix, which are the first-order tensor and the second-order tensor, respectively. Tensor data has been prevalent in diverse areas. For example, in neuroimaging studies, electroencephalography (EEG) generates two-dimensional tensor (matrix) data, while anatomical magnetic resonance imaging (MRI) produces three-dimensional tensor data, where the voxels represent brain locations that are spatially associated. More background on the tensor data and tensor operations can be found in [71].

[44] and [45] extended the envelope model to tensor-valued responses and predictors, respectively, leading to more parsimonious models with improved interpretability. We take the tensor response model in [44] as an example and demonstrate how the envelope algorithms in this paper can be applied. Let $\mathcal{Y} \in \mathbb{R}^{r_1\times\cdots\times r_m}$ be an m-way tensor response, $\boldsymbol{X} \in \mathbb{R}^p$ be a vector predictor. The tensor regression model assumes that

$$\mathcal{Y} = \mathcal{B}\overline{\times}_{(m+1)}\boldsymbol{X} + \boldsymbol{\varepsilon}, \tag{9}$$

where $\mathcal{B} \in \mathbb{R}^{r_1 \times \cdots \times r_m \times p}$ is an $(m+1)$-way tensor coefficient, $\overline{\times}_{(m+1)}$ denotes the $(m+1)$-mode vector product, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{r_1 \times \cdots \times r_m}$ is assumed to be independent of $\boldsymbol{X}$ with mean zero. The variance of $\boldsymbol{\varepsilon}$ is also assumed to enjoy the separable Kronecker covariance structure such that $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{\Sigma}_1$ with $\boldsymbol{\Sigma}_k > 0$, $k = 1, \cdots, m$, where $\otimes$ denotes the Kronecker product. Let $\boldsymbol{B}_{(k)} \in \mathbb{R}^{r_k \times (\prod_{j \neq k} r_j \times p)}$ denote the mode-k matricization of $\mathcal{B}$, mapping the tensor $\mathcal{B}$ into a matrix along mode k. According to [44], the tensor envelope for $\mathcal{B}$ in (9) can be expressed as the Kronecker product of a series of envelopes as follows:

$$\mathcal{E}_{\boldsymbol{\Sigma}_m}(\boldsymbol{B}_{(m)}) \otimes \cdots \otimes \mathcal{E}_{\boldsymbol{\Sigma}_1}(\boldsymbol{B}_{(1)}).$$

Therefore, the estimation of the tensor envelope of $\mathcal{B}$ can be approached by estimating each individual envelope $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\boldsymbol{B}_{(k)})$, where the kernel matrix is $\boldsymbol{B}_{(k)}$. By providing the consistent estimators of $\boldsymbol{B}_{(k)}$ and the eigenvectors of $\boldsymbol{\Sigma}_k$ to the manifold-based algorithm or the NIECE algorithm, we obtain the estimators of $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\boldsymbol{B}_{(k)})$, $k = 1, \cdots, m$, respectively, which together give the estimator of the tensor envelope of $\mathcal{B}$.

Extension to semi-supervised setting. We also discuss how the efficiency of the envelope model can be significantly improved in semi-supervised learning tasks. Semi-supervised learning leverages both labeled and unlabeled data, often leading to performance improvements when labeled data is scarce or expensive to obtain. For instance, in healthcare, medical data is abundant, but labeling requires domain expertise, and in recommendation systems, user interaction data is plentiful, but labeled feedback is sparse. Similarly, in autonomous driving, vast amounts of sensor data are generated, but labeling data, like identifying pedestrians and vehicles, can be costly. For many statistical approaches, the challenges brought by semi-supervised learning is that only labeled data is used for estimation. To utilize unlabeled data, pseudo-labeling techniques are often employed, which involve multiple rounds of iteration and can introduce computational overhead and bias. However, the envelope model can benefit from unlabeled data in semi-supervised learning without the extra cost introduced by the pseudo-label generation process.

Consider n_L labeled data $(\boldsymbol{X}_i^{(L)}, Y_i)$, $i = 1, \cdots, n_L$, which are independent copies of the random vector $(\boldsymbol{X}, Y)$, and n_U unlabeled predictor data $\boldsymbol{X}_i^{(U)}$, $i = 1, \cdots, n_U$, which are independent copies of $\boldsymbol{X}$. Assume that the labeled data and unlabeled data are independent. As discussed in Subsection 3.1, in manifold-based and NIECE algorithms, the covariance matrix $\boldsymbol{\Sigma}$ and kernel matrix $\boldsymbol{U}$ are separately estimated for estimating the envelope inverse regression subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{U})$. While the estimation of kernel matrix $\boldsymbol{U} \equiv \boldsymbol{U}(F_{\boldsymbol{X}|Y})$ requires the full information from the labeled data, the estimation of covariance matrix $\boldsymbol{\Sigma}$ only requires the predictor data, which can be improved by integrating unlabeled predictor data. Specifically, we estimate $\boldsymbol{U}$ using the labeled data $(\boldsymbol{X}_i^{(L)}, Y_i)$, $i = 1, \cdots, n_L$ and estimate $\boldsymbol{\Sigma}$ using both the labeled and unlabeled data such that $\widehat{\boldsymbol{\Sigma}} = (n_L + n_U)^{-1}(\sum_{i=1}^{n_L} \boldsymbol{X}_i^{(L)} + \sum_{i=1}^{n_U} \boldsymbol{X}_i^{(U)})$. In Section 4, we present empirical results of envelope estimation in semi-supervised learning, where we will see the efficiency improvement.

4 Simulation Studies

4.1 The Envelope Subspace Equals the Central Subspace

In this section, we provide some investigation into the empirical performance of the envelope inverse regression methods in the low-dimensional setting, where the envelope subspace is equivalent to the central subspace. Let $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_u) \in \mathbb{R}^{p\times u}$ be an orthogonal matrix and $\boldsymbol{\Gamma}_0$ be the orthogonal complement of $\boldsymbol{\Gamma}$ such that $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ is a full-rank orthogonal matrix. To generate $\boldsymbol{\Gamma}$, we construct a p-by-u matrix with each element being generated from the standard normal distribution, followed by orthogonalization. The predictor $\boldsymbol{X}$ is generated from $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$ for some matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u)\times(p-u)}$. Specifically, let $\boldsymbol{\Omega} = \boldsymbol{O}\boldsymbol{D}\boldsymbol{O}^{\mathrm{T}}$ and $\boldsymbol{\Omega}_0 = \boldsymbol{O}_0\boldsymbol{D}_0\boldsymbol{O}_0^{\mathrm{T}}$, where $\boldsymbol{O} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{O}_0 \in \mathbb{R}^{(p-u)\times(p-u)}$ are orthogonal matrices, $\boldsymbol{D} \in \mathbb{R}^{u\times u}$ is a diagonal matrix with diagonal elements being u evenly spaced numbers between 2 and 1.5, and $\boldsymbol{D} \in \mathbb{R}^{(p-u)\times(p-u)}$ is a diagonal matrix with diagonal elements being $p-u$ evenly spaced numbers between 0.1 and 1. We then standardize $\boldsymbol{\Sigma} = \alpha\boldsymbol{\Sigma}/\|\boldsymbol{\Sigma}\|_F$, where the constant α (now the new Frobenius norm of $\boldsymbol{\Sigma}$) is used to make the models neither too simple nor too difficult. We consider the following six models, where the envelope subspace is equivalent to the central subspace:

(M1) **(Linear)** $Y = \boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X} + \varepsilon$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 5$.

(M2) **(Logistic)** $Y \sim \text{Bernoulli}(p)$, where $p = \exp(2\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X})/\{1 + \exp(2\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X})\}$ and $\|\boldsymbol{\Sigma}\|_F = 2$.

(M3) **(Classical SDR model)** $Y = (\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X})/(0.5 + (1.5 + \boldsymbol{\gamma}_2^{\mathrm{T}}\boldsymbol{X})^2) + \varepsilon$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 5$.

(M4) **(Classical SDR model)** $Y = \text{sign}(\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X} + \varepsilon_1) \cdot \log(|\boldsymbol{\gamma}_2^{\mathrm{T}}\boldsymbol{X} + 5 + \varepsilon_2|)$, where $\varepsilon_1 \sim N(0, 0.5)$ and $\varepsilon_2 \sim N(0, 0.5)$ are independent, and $\|\boldsymbol{\Sigma}\|_F = 2$.

(M5) **(Non-linear)** $Y = (\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X})^2 \cdot \sin(\boldsymbol{\gamma}_2^{\mathrm{T}}\boldsymbol{X}) + \varepsilon$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 5$.

(M6) **(Heteroskedasticity)** $Y = (\boldsymbol{\gamma}_1^{\mathrm{T}}\boldsymbol{X}) \cdot \exp(\boldsymbol{\gamma}_2^{\mathrm{T}}\boldsymbol{X} + \varepsilon)$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 5$.

In all models except for Model (M2), the noise term is independent of $\boldsymbol{X}$. The envelope dimension $u = 1$ for Models (M1) and (M2), and $u = 2$ for Models (M3)–(M6). In this section and Subsection 4.2, we set $p = 10$ and $n = 100$ or $n = 400$. We consider SIR, SAVE, CUME, and CUVE as the basic inverse regression methods and construct the corresponding envelope subspaces using the manifold-based algorithm. The envelope inverse regression methods implemented by the manifold-based algorithm are referred to as Env-SIR, Env-SAVE, Env-CUME, and Env-CUVE, respectively.

In Table 2, we present the estimation error of the envelope span($\boldsymbol{\Gamma}$), measured by $\|\boldsymbol{P}_{\widehat{\boldsymbol{\Gamma}}} - \boldsymbol{P}_{\boldsymbol{\Gamma}}\|_F$. When $n = 100$, the classic inverse regression methods fail to give an accurate estimation for the central subspace. Although their performances improve when n increases to 400, the estimation error remains large. As a comparison, the envelope inverse regression methods gain performance improvement by employing the envelope structure. Moreover, for these low-dimensional simulation examples with $p = 10$, we observe that the envelope methods using

the manifold-based algorithm perform better than those using the NIECE algorithm. This is expected because the NIECE algorithm only returns eigen-spaces of the sample covariance, as a way to speed up computation and avoid local optima issues in higher dimensional settings (see Subsection 4.3 for demonstrations).

Table 2 Means (and standard errors) of envelope estimation errors under Models (M1)–(M6) based on 100 replicates. The envelope subspaces obtained by the NIECE algorithm are generally less accurate than the FG algorithm, but still significantly improve over classical inverse regression methods

Method	M1		M2		M3	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$	$n = 100$	$n = 400$
SIR	0.35 (0.01)	0.17 (0.004)	0.78 (0.02)	0.44 (0.01)	1.50 (0.01)	1.01 (0.02)
Env-SIR	0.26 (0.01)	0.13 (0.003)	0.28 (0.01)	0.14 (0.003)	0.52 (0.01)	0.25 (0.01)
SAVE	1.41 (0.01)	0.21 (0.01)	1.07 (0.02)	0.49 (0.01)	1.93 (0.01)	1.59 (0.02)
Env-SAVE	0.26 (0.01)	0.12 (0.003)	0.28 (0.01)	0.14 (0.003)	1.16 (0.04)	0.29 (0.01)
CUME	0.42 (0.01)	0.20 (0.003)	0.78 (0.02)	0.44 (0.01)	1.44 (0.02)	0.99 (0.02)
Env-CUME	0.24 (0.01)	0.12 (0.003)	0.28 (0.01)	0.14 (0.003)	0.58 (0.02)	0.31 (0.01)
CUVE	1.16 (0.02)	0.57 (0.01)	1.37 (0.01)	1.17 (0.02)	1.82 (0.01)	1.45 (0.02)
Env-CUVE	1.32 (0.03)	0.27 (0.01)	1.39 (0.01)	1.14 (0.04)	1.97 (0.01)	1.56 (0.03)
Method	M4		M5		M6	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$	$n = 100$	$n = 400$
SIR	0.98 (0.02)	0.49 (0.01)	1.68 (0.01)	1.50 (0.01)	0.89 (0.01)	0.45 (0.01)
Env-SIR	0.43 (0.01)	0.21 (0.01)	0.67 (0.03)	0.29 (0.01)	0.42 (0.01)	0.21 (0.004)
SAVE	1.97 (0.002)	1.05 (0.02)	1.78 (0.01)	1.43 (0.02)	1.97 (0.01)	0.60 (0.01)
Env-SAVE	0.78 (0.04)	0.23 (0.01)	0.89 (0.04)	0.29 (0.01)	0.66 (0.03)	0.20 (0.004)
CUME	0.95 (0.02)	0.51 (0.01)	1.65 (0.01)	1.48 (0.01)	0.84 (0.01)	0.44 (0.01)
Env-CUME	0.42 (0.01)	0.21 (0.004)	0.81 (0.04)	0.53 (0.03)	0.39 (0.01)	0.20 (0.004)
CUVE	1.81 (0.01)	1.18 (0.02)	1.82 (0.01)	1.55 (0.01)	1.75 (0.01)	1.11 (0.02)
Env-CUVE	1.98 (0.01)	1.42 (0.04)	1.98 (0.07)	1.71 (0.03)	1.91 (0.02)	1.21 (0.04)

We further demonstrate that as the correlation among predictors strengthens, the performance disparity between the proposed methods and the inverse regression methods becomes more pronounced. To illustrate this, we consider Model (M4) with $n = 400$, where SIR and CUME perform reasonably well compared to the envelope approaches. To make the correlation of the predictor stronger, we adjust the $\boldsymbol{D}$ matrix to $c\boldsymbol{D}$, where c takes value from $\{0.5, 0.75, 1, 2, 3, 4, 5\}$. A larger c shifts the covariance away from the isotropic structure, thereby increasing the correlation between predictors. The subspace estimation error is shown in Figure 1. As the correlation intensifies, the envelope approaches exhibit superior performance, while the inverse regression methods SIR and CUME deteriorate. Notably, when the correlation is strong (e.g., $c = 5$), SIR and CUME fail to provide meaningful subspace estimations. Strong predictor correlations generally pose greater challenges for inverse regression methods. In contrast, the envelope method leverages the covariance structure effectively and can even benefit from increased correlation. Even under weak correlation, the envelope approach outperforms SIR and CUME. These simulation results underscore the practical advantage of the

proposed envelope approach as a more stable alternative to inverse regression methods, particularly in scenarios with strongly correlated predictors.

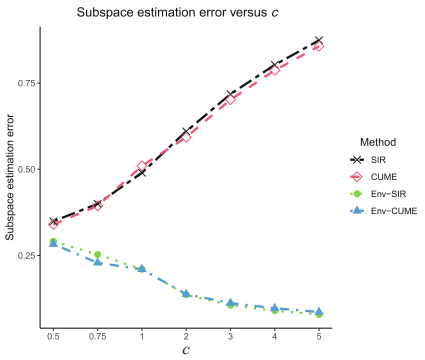

Figure 1 Envelope estimation error versus c, representing the strength of correlation. The standard errors for all the methods are smaller than 0.02

4.2 The Envelope Subspace Properly Contains the Central Subspace

Although the envelope subspace always exists, it may not equal the central subspace but find a larger one. We use two additional simulation studies to demonstrate that the proposed envelope approach still enhances the subspace estimation. We set the envelope dimension $u = 3$, while fixing the structural dimension of the central subspace at 1. We consider the following two examples, where the error term is independent of $\boldsymbol{X}$:

(M7) $Y = \{(\boldsymbol{\gamma}_1^{\mathrm{T}} + \boldsymbol{\gamma}_2^{\mathrm{T}} + \boldsymbol{\gamma}_3^{\mathrm{T}})\boldsymbol{X}\}^2 \cdot \sin\{(\boldsymbol{\gamma}_1^{\mathrm{T}} + \boldsymbol{\gamma}_2^{\mathrm{T}} + \boldsymbol{\gamma}_3^{\mathrm{T}})\boldsymbol{X}\} + \varepsilon$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 2$.

(M8) $Y = \{(\boldsymbol{\gamma}_1^{\mathrm{T}} + \boldsymbol{\gamma}_2^{\mathrm{T}} + \boldsymbol{\gamma}_3^{\mathrm{T}})\boldsymbol{X}\} \cdot \exp\{(\boldsymbol{\gamma}_1^{\mathrm{T}} + \boldsymbol{\gamma}_2^{\mathrm{T}} + \boldsymbol{\gamma}_3^{\mathrm{T}})\boldsymbol{X} + \varepsilon\}$, where $\varepsilon \sim N(0, 0.5)$ and $\|\boldsymbol{\Sigma}\|_F = 5$.

In these two models, the subspaces captured by the envelope and SDR methods are different. The envelope subspace is $\mathrm{span}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$ and the central subspace is $\mathrm{span}(\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 + \boldsymbol{\gamma}_3)$. Hence, a direct comparison of the subspace estimation error is unfair. Note that the envelope subspace contains the central subspace in population. We use the following envelope-assisted method to estimate the central subspace. After obtaining the basis matrix $\widehat{\boldsymbol{\Gamma}}$ of the envelope subspace, we redo inverse regression methods on $(Y, \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\boldsymbol{X})$ to obtain a basis matrix $\widehat{\boldsymbol{\eta}}$. Then we let $\widetilde{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\eta}}$, which is the envelope-assisted estimator of the basis of the central subspace.

From Table 3, the first moment-based inverse regression methods, SIR and CUME fail to work well in most cases, especially when $n = 100$. The envelope-assisted approaches can significantly improve the results. Notably, while SAVE occasionally yields large estimation errors, the envelope-assisted approach effectively mitigates this issue. Although the envelope approach identifies a larger subspace, potentially introducing redundant predictor information, it excels in capturing the true dimension reduction directions. Although the dimension considered p is not large, inverse regression methods need a large number of observations to capture the data information and provide an accurate estimation of the central subspace. As a comparison, the envelope approach requires fewer observations and can provide a more accurate estimation of

the envelope subspace. By projecting the origin predictor onto the envelope subspace, the dimension of the predictor and the model complexity can be reduced significantly, and irrelevant information that may be harmful to the inverse regression methods is removed, which assists inverse regression methods in finding the central subspace. These simulation results demonstrate that, prior to applying inverse regression methods, it is advantageous to first employ the envelope approach to obtain a low-dimensional subspace containing the central subspace. Subsequently, the inverse regression method can be applied to the response and the projected predictor.

Table 3 Means (and standard errors) of central space estimation errors under Models (M7) and (M8) based on 100 replicates

Method	M7		M8	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$
SIR	1.10 (0.02)	0.71 (0.02)	0.47 (0.02)	0.20 (0.01)
Env-SIR	0.56 (0.02)	0.25 (0.01)	0.28 (0.01)	0.13 (0.003)
SAVE	1.21 (0.02)	0.54 (0.02)	1.41 (0.01)	0.21 (0.01)
Env-SAVE	0.54 (0.03)	0.19 (0.01)	0.48 (0.03)	0.13 (0.004)
CUME	1.08 (0.02)	0.80 (0.02)	0.62 (0.02)	0.32 (0.01)
Env-CUME	0.64 (0.03)	0.36 (0.01)	0.35 (0.01)	0.18 (0.01)
CUVE	1.25 (0.02)	0.66 (0.02)	0.97 (0.02)	0.45 (0.01)
Env-CUVE	1.38 (0.01)	0.67 (0.03)	1.06 (0.03)	0.38 (0.01)

In practice, the envelope dimension u must be specified in the algorithm. To evaluate the robustness of the proposed method against misspecification of u, we examine the method's performance under model (M8) and vary values of u from 2 to 6. The results are presented in Figure 2. It can be seen that even with substantial misspecification (e.g., $u = 6$), the envelope-based methods (Env-SIR and Env-CUME) consistently outperform their standard counterparts (SIR and CUME). Moreover, when u is close to the true value, the performance degradation of the envelope methods is marginal. These findings highlight the robustness and practical effectiveness of the envelope framework in downstream applications.

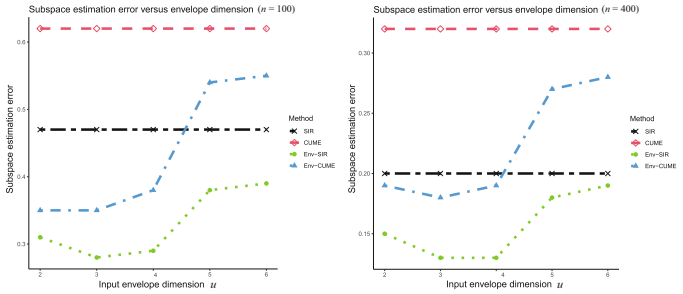

Figure 2 Simulation results for (M8) with misspecified envelope dimension

4.3 High-Dimensional Simulations Studies

In this section, we study the high-dimensional performance of the proposed envelope SDR method using the sparse NIECE algorithm in Algorithm 1. We consider two high-dimensional

simulation settings, which are similar to Models (M3) and (M4), but with a larger dimension $p = 500$. Different from the previous simulation studies, we consider a sparse basis matrix $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_s^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{p\times d}$, where $\boldsymbol{\Gamma}_s \in \mathbb{R}^{s\times d}$ are randomly generated basis matrices satisfying $\boldsymbol{\Gamma}_s^{\mathrm{T}}\boldsymbol{\Gamma}_s = \boldsymbol{I}_s$. We assume that only the first $s = 10$ predictors are involved in the simulation data generation and assume that they are sampled from $N(\mathbf{0}, \boldsymbol{\Sigma}_s)$, where $\boldsymbol{\Sigma}_s$ has the same structure as previous $\boldsymbol{\Sigma}$ except that it is a $s \times s$ matrix. For the other predictors, we generated them from $N(0, 0.5)$ independently. The other data generating procedures are exactly the same as Models (M3) and (M4) and we denote the new models as Models (M3*) and (M4*). We consider sample size $n = 100$ or $n = 400$.

For those two high dimensional settings, the conventional SIR and CUME are not directly applicable because they require an estimation for the inverse of the covariance matrix. To show the performance and the adaptability of the envelope approach, we report the results of SIR and CUME with the knowledge of the true important predictors, namely, SIR and CUME are constructed based on $(y, \boldsymbol{X}_s)$, where $\boldsymbol{X}_s \in \mathbb{R}^s$ denotes the sub-vector comprising the first s elements of $\boldsymbol{X}$. We denote them as SIR* and CUME* for distinction. Besides those two competitors, we also included two high-dimensional inverse regression methods, SEAS-SIR[27] and Lasso-SIR[24]. The subspace estimation errors are reported in Table 4. The proposed method performs much better than SEAS-SIR and Lasso-SIR by using the envelope covariance information. Although SIR* and CUME* use the information of the true active set, the proposed envelope methods ESIR and ECUME perform slightly well than them when $n = 100$ and perform much better when $n = 400$. These two examples demonstrate that the proposed envelope approach scales well with high dimensions.

Table 4 High-dimensional setting: Means (and standard errors) of envelope estimation errors under Models (M3*) and (M4*) based on 100 replicates, where $p = 500$

Method	M3*		M4*	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$
SIR*	1.54 (0.02)	1.04 (0.02)	0.84 (0.02)	0.47 (0.01)
CUME*	1.46 (0.01)	1.04 (0.02)	0.84 (0.02)	0.49 (0.01)
SEAS-SIR	1.34 (0.02)	1.02 (0.02)	1.08 (0.02)	1.09 (0.03)
Lasso-SIR	1.41 (0.00)	1.41 (0.00)	1.41 (0.00)	1.41 (0.00)
Env-SIR	1.00 (0.02)	0.48 (0.01)	0.79 (0.02)	0.38 (0.01)
Env-CUME	1.09 (0.03)	0.48 (0.01)	0.79 (0.02)	0.38 (0.01)

4.4 Semi-Supervised Case

In this section, we study the semi-supervised case, where additional unsupervised predictors are observed. We use two relatively challenging Models (M5) and (M6) to show the adaptability and nice performance of the proposed envelope SDR method. We set the size of labeled samples $n_L = 100$ and generate additional n_U unlabeled data from $N(\mathbf{0}, \boldsymbol{\Sigma})$. Intuitively, the unlabeled samples enhance the estimation of the $\boldsymbol{\Sigma}$ matrix of the envelope and further improve the envelope subspace estimation accuracy.

In Table 5, we show the subspace estimation accuracy for different envelope inverse regression methods. Please see also Figure 3 for the visualization of Table 5. We vary n_U over 0 (no additional unlabeled data), 100, 500, and 2500. The results for SIR and CUME are almost unchanged (CUME becomes slightly worse for Model (M6)) although the sample covariance matrix for the predictor is calculated using additional samples. For envelope SIR and envelope CUME, with the increase of n_U, the subspace estimation accuracy improves. However, this improvement may have a bottleneck because the estimation of the kernel matrix has not been improved.

Table 5 Semi-supervised study: Means (and standard errors) of envelope estimation errors under Models (M5) and (M6) based on 100 replicates. Env-SIR1 and Env-CUME1 are results of FG algorithm, and Env-SIR2 and Env-CUME2 are results of NIECE algorithm

Method	M5				M6			
	$n_1 = 0$	$n_1 = 100$	$n_1 = 500$	$n_1 = 2500$	$n_1 = 0$	$n_1 = 100$	$n_1 = 500$	$n_1 = 2500$
SIR	1.68 (0.01)	1.68 (0.01)	1.67 (0.01)	1.66 (0.01)	0.89 (0.01)	0.88 (0.01)	0.88 (0.01)	0.89 (0.01)
CUME	1.65 (0.01)	1.65 (0.01)	1.66 (0.01)	1.68 (0.01)	0.84 (0.01)	1.05 (0.01)	1.15 (0.01)	1.15 (0.01)
Env-SIR1	0.67 (0.03)	0.54 (0.03)	0.48 (0.03)	0.37 (0.02)	0.42 (0.01)	0.35 (0.01)	0.30 (0.01)	0.29 (0.01)
Env-CUME1	0.81 (0.04)	0.69 (0.03)	0.67 (0.04)	0.67 (0.04)	0.39 (0.01)	0.36 (0.01)	0.35 (0.01)	0.35 (0.01)
Env-SIR2	0.74 (0.03)	0.53 (0.03)	0.36 (0.03)	0.17 (0.03)	0.62 (0.02)	0.39 (0.01)	0.26 (0.01)	0.12 (0.003)
Env-CUME2	0.79 (0.04)	0.53 (0.03)	0.37 (0.04)	0.21 (0.03)	0.63 (0.02)	0.39 (0.01)	0.26 (0.01)	0.12 (0.003)

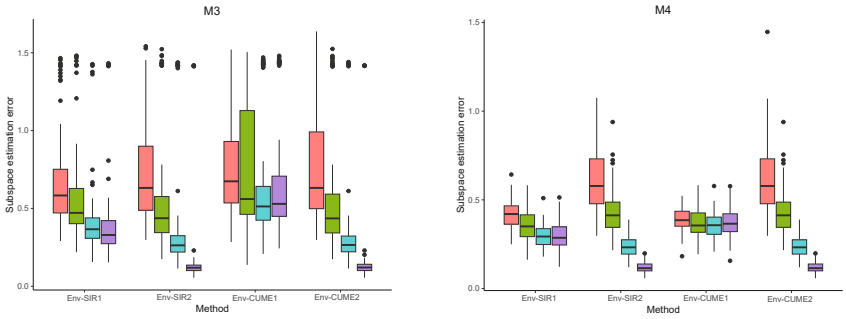

Figure 3 Semi-supervised study: Boxplots of envelope estimation errors under Models (M5) and (M6). Env-SIR1 and Env-CUME1 are results of FG algorithm, and Env-SIR2 and Env-CUME2 are results of NIECE algorithm. For each method, from left to right, the four boxes represent the results with $n_u = 0, 100, 500, 2500$, respectively.

5 Real Data Analysis

We study the meat property data[72], which includes near-infrared (NIR) spectroscopy measurements along with water, fat, and protein compositions for $n = 103$ meat samples. Following [52], we extract spectral measurements at every two wavelengths between 850 nm and 1050 nm, resulting in $p = 100$ predictors. Among the 103 samples, 49 are beef and 54 are pork. We randomly split the data 100 times into a training set of size $n_{\text{train}} = 52$ and a testing set of size

$n_{\text{test}} = 51$. We vary the envelope dimension u over $4, \cdots, 12$ and record the classification error rates for each method. Since $p > n_{\text{train}}$, traditional inverse regression methods such as SIR and CUME are not directly applicable. However, with the sparse NIECE algorithm, the envelope inverse regression methods perform well on this data. We compare our method to some relevant high-dimensional methods: Lasso-penalized logistic regression (PMLE), LassoSIR[24], and SEAS-SIR[27]. For envelope inverse regression methods, we report the results of envelope SIR (Env-SIR) and envelope CUME (Env-CUME), which have been demonstrated to have strong performance in earlier simulations. We also include the downstream procedure considered in Subsection 4.2, which has been shown to perform well even when the envelope subspace is larger than the central subspace. We denote them as Env-SIR* and Env-CUME* as distinction.

To evaluate the performance of the dimension reduction methods, we fit a logistic regression model for classification. The classification errors are reported in Figure 4. The results of Env-SIR and Env-CUME are nearly identical. When the envelope dimension $u \geq 4$, both methods outperform PMLE, and when $u \geq 6$, they achieve remarkably low classification errors (below 1%). These results are comparable to those of the state-of-the-art inverse regression methods, SEAS-SIR (1.0% error) and LassoSIR (0.84% error). Notably, the downstream approaches Env-SIR* and Env-CUME* further enhance the performance of Env-SIR and Env-CUME, respectively. When the envelope dimension $u = 12$, they achieve the smallest classification error (0.25%) among all the competitors. This result demonstrate again that the envelope approach can handle high-dimensional data well and assists the conventional SDR methods to find the central subspace more accurately.

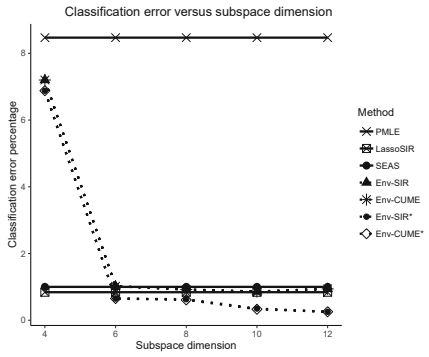

Figure 4 Classification error percentage on the meat data set. For SDR method, the central subspace dimension is 1 because it is a binary classification problem

6 Discussion

In this note, we have reviewed the envelope model in sufficient dimension reduction. The enveloped dimension reduction was initially introduced to circumvent the inversion of the sample covariance matrix in inverse regression subspace estimation. It was later recognized to improve the efficiency of model-based regression methods, particularly when covariates are highly correlated. Our simulation study and real data applications further demonstrate the advan-

tages of the envelope dimension reduction in assisting SDR methods in capturing the central subspace either directly or through downstream analysis. However, the manifold-based optimization for the envelope estimation is non-convex and known to be computationally intensive, which requires carefully designed algorithms to accelerate computation. Non-iterative envelope component estimation (NIECE) algorithm provides a flexible and scalable generalization, yet it remains unclear the efficiency gains of NIECE approach in theory. We regard this as a promising future research direction.

The envelope inverse regression methods discussed in this note primarily focus on inverse regression approaches. While many other SDR methods exist, integrating the strengths of envelope techniques with these alternatives presents a promising research direction. For example, extending envelope methods to semi-parametric SDR estimators[3, 6, 73] could be a future direction. Additionally, developing envelope approaches for estimating the central mean subspace[74] also worth further investigation.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Dennis Cook R, *Regression Graphics: Ideas for Studying Regressions Through Graphics*, Volume 318, John Wiley & Sons, Hoboken, 1998.

[2] Li B, *Sufficient Dimension Reduction: Methods and Applications with R*, Chapman and Hall/CRC, Boca Raton, 2018.

[3] Xia Y C, Tong H, Li W K, et al., An adaptive estimation of dimension reduction space, *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), 2002, **64**(3): 363–410.

[4] Xia Y C, A constructive approach to the estimation of dimension reduction directions, *The Annals of Statistics*, 2007, **35**(6): 2654–2690.

[5] Yin X R and Dennis Cook R, Direction estimation in single-index regressions, *Biometrika*, 2005, **92**(2): 371–384.

[6] Ma Y Y and Zhu L P, A semiparametric approach to dimension reduction, *Journal of the American Statistical Association*, 2012, **107**(497): 168–179.

[7] Dennis Cook R and Lee H, Dimension reduction in binary response regression, *Journal of the American Statistical Association*, 1999, **94**(448): 1187–1200.

[8] Dennis Cook R and Yin X R, Theory & methods: Special invited paper: Dimension reduction and visualization in discriminant analysis (with discussion), *Australian & New Zealand Journal of Statistics*, 2001, **43**(2): 147–199.

[9] Zhang X, Mai Q, and Zou H, The maximum separation subspace in sufficient dimension reduction with categorical response, *Journal of Machine Learning Research*, 2020, **21**(29): 1–36.

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

[11] Lian H and Li G R, Series expansion for functional sufficient dimension reduction, *Journal of Multivariate Analysis*, 2014, **124**: 150–165.

[12] Yao F, Lei E, and Wu Y, Effective dimension reduction for sparse functional data, *Biometrika*, 2015, **102**(2): 421–437.

[13] Qu W X, Liang B T, and Wang G C, Projective resampling functional sliced inverse regression, *Journal of Systems Science & Complexity*, 2025, **38**(5): 2185–2203.

[14] Guo X, Wang T, and Zhu L X, Model checking for parametric single-index models: A dimension reduction model-adaptive approach, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2016, **78**(5): 1013–1035.

[15] Zhu X H, Guo X, and Zhu L X, An adaptive-to-model test for partially parametric single-index models, *Statistics and Computing*, 2017, **27**: 1193–1204.

[16] Zhu X H and Zhu L X, Dimension reduction-based significance testing in nonparametric regression, *Electronic Journal of Statistics*, 2018, **12**(1): 1468–1506.

[17] Xia L L, Du J, and Zhang Z Z, A nonparametric model checking test for functional linear composite quantile regression models, *Journal of Systems Science & Complexity*, 2024, **37**(4): 1714–1737.

[18] Ma Y Y and Zhu L P, A review on dimension reduction, *International Statistical Review*, 2013, **81**(1): 134–150.

[19] Dennis Cook R, Principal components, sufficient dimension reduction, and envelopes, *Annual Review of Statistics and Its Application*, 2018, **5**(1): 533–559.

[20] Li L, Meggie Wen X R, and Yu Z, A selective overview of sparse sufficient dimension reduction, *Statistical Theory and Related Fields*, 2020, **4**(2): 121–133.

[21] Li L, Shao X F, and Yu Z, A slicing-free perspective to sufficient dimension reduction: Selective review and recent developments, *International Statistical Review*, 2024, **92**(3): 355–382.

[22] Wang Q and Yin X R, A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse mave, *Computational Statistics & Data Analysis*, 2008, **52**(9): 4512–4520.

[23] Cai Z B, Xia Y C, and Hang W Q, An outer-product-of-gradient approach to dimension reduction and its application to classification in high dimensional space, *Journal of the American Statistical Association*, 2023, **118**(543): 1671–1681.

[24] Lin Q, Zhao Z G, and Liu J S, Sparse sliced inverse regression via lasso, *Journal of the American Statistical Association*, 2019, **114**(528): 1726–1739.

[25] Tan K M, Wang Z R, Zhang T, et al., A convex formulation for high-dimensional sparse sliced inverse regression, *Biometrika*, 2018, **105**(4): 769–782.

[26] Tan K, Shi L, and Yu Z, Sparse sir: Optimal rates and adaptive estimation, *The Annals of Statistics*, 2020, **48**(1): 64–85.

[27] Zeng J, Mai Q, and Zhang X, Subspace estimation with automatic dimension and variable selection in sufficient dimension reduction, *Journal of the American Statistical Association*, 2024, **119**(545): 343–355.

[28] Dennis Cook R, Li B, and Chiaromonte F, Dimension reduction in regression without matrix inversion, *Biometrika*, 2007, **94**(3): 569–584.

[29] Dennis Cook R, Li B, and Chiaromonte F, Envelope models for parsimonious and efficient

multivariate linear regression, *Statistica Sinica*, 2010, **20**(3): 927–960.

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

[31] Dennis Cook R, Helland I S, and Su Z, Envelopes and partial least squares regression, *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), 2013, **75**(5): 851–877.

[32] Schott J R, On the likelihood ratio test for envelope models in multivariate linear regression, *Biometrika*, 2013, **100**(2): 531–537.

[33] Su Z H and Dennis Cook R, Estimation of multivariate means with heteroscedastic errors using envelope models, *Statistica Sinica*, 2013, **23**(1): 213–230.

[34] Su Z H and Dennis Cook R, Inner envelopes: Efficient estimation in multivariate linear regression, *Biometrika*, 2012, **99**(3): 687–702.

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

[36] Dennis Cook R and Zhang X, Algorithms for envelope estimation, *Journal of Computational and Graphical Statistics*, 2016, **25**(1): 284–300.

[37] Zhang X, Lee C E, and Shao X F, Envelopes in multivariate regression models with nonlinearity and heteroscedasticity, *Biometrika*, 2020, **107**(4): 965–981.

[38] Wang W J, Zhang X, and Mai Q, Model-based clustering with envelopes, *Electronic Journal of Statistics*, 2020, **14**(1): 82–109.

[39] Deng K and Zhang X, Tensor envelope mixture model for simultaneous clustering and multiway dimension reduction, *Biometrics*, 2022, **78**(3): 1067–1079.

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

[41] Ding S S, Su Z H, Zhu G Y, et al., Envelope quantile regression, *Statistica Sinica*, 2021, **31**(1): 79–105.

[42] Zhang X and Mai Q, Efficient integration of sufficient dimension reduction and prediction in discriminant analysis, *Technometrics*, 2019, **61**(2): 259–272.

[43] Wang N, Wang W J, and Zhang X, Parsimonious tensor discriminant analysis, *Statistica Sinica*, 2024, **34**(1): 157–180.

[44] Li L X and Zhang X, Parsimonious tensor response regression, *Journal of the American Statistical Association*, 2017, **112**(519): 1131–1146.

[45] Zhang X and Li L X, Tensor envelope partial least-squares regression, *Technometrics*, 2017, **59**(4): 426–436.

[46] Zhang X, Wang C, and Wu Y C, Functional envelope for model-free sufficient dimension reduction, *Journal of Multivariate Analysis*, 2018, **163**: 37–50.

[47] Su Z H, Li B, and Cook D, Envelope model for function-on-function linear regression, *Journal of Computational and Graphical Statistics*, 2023, **32**(4): 1624–1635.

[48] Wang L and Ding S S, Vector autoregression and envelope model, *Stat.*, 2018, **7**(1): e203.

[49] Soale A N and Dong Y X, Envelope dimension reduction with application to binary classification, *Journal of Systems Science & Complexity*, 2026, **39**(1): 79–87.

[50] Su Z, Zhu G, Chen X, et al., Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression, *Biometrika*, 2016, **103**(3): 579–593.

[51] Zhu G Y and Su Z H, Envelope-based sparse partial least squares, *The Annals of Statistics*, 2020, **48**(1): 161–182.

[52] Zhang X, Deng K, and Mai Q, Envelopes and principal component regression, *Electronic Journal of Statistics*, 2023, **17**(2): 2447–2484.

[53] Li K C, Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, 1991, **86**(414): 316–327.

[54] Dennis Cook R and Weisberg S, Discussion of sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, 1991, **86**(414): 328–332.

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

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

[57] Mai Q, Shao X F, Wang R M, et al., Slicing-free inverse regression in high-dimensional sufficient dimension reduction, *Statistica Sinica*, 2025, **35**(1): 1–23.

[58] Lee C E and Shao X, Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series, *Journal of the American Statistical Association*, 2018, **113**(521): 216–229.

[59] Zeng J, Wang W J, and Zhang X, Tres: An *r* package for tensor regression and envelope algorithms, *Journal of Statistical Software*, 2021, **99**: 1–31.

[60] Dennis Cook R and Zhang X, Fast envelope algorithms, *Statistica Sinica*, 2018, **28**(3): 1179–1197.

[61] Witten D M, Tibshirani R, and Hastie T, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, *Biostatistics*, 2009, **10**(3): 515–534.

[62] Dennis Cook R and Su Z H, Scaled envelopes: Scale-invariant and efficient estimation in multivariate linear regression, *Biometrika*, 2013, **100**(4): 939–954.

[63] Zhang X and Mai Q, Model-free envelope dimension selection, *Electronic Journal of Statistics*, 2018, **12**(2): 2193–2216.

[64] Székely J G, Rizzo M L, and Bakirov N K, Measuring and testing dependence by correlation of distances, *The Annals of Statistics*, 2007, **35**(6): 2769–2794.

[65] Ye Z S and Weiss R E, Using the bootstrap to select one of a new class of dimension reduction methods, *Journal of the American Statistical Association*, 2003, **98**(464): 968–979.

[66] Li K C, On principal hessian directions for data visualization and dimension reduction: Another application of Stein's lemma, *Journal of the American Statistical Association*, 1992, **87**(420): 1025–1039.

[67] Zhu L X, Ohtaki M, and Li Y X, On hybrid methods of inverse regression-based algorithms, *Computational Statistics & Data Analysis*, 2007, **51**(5): 2621–2635.

[68] Zou H, Hastie T, and Tibshirani R, Sparse principal component analysis, *Journal of Computational and Graphical Statistics*, 2006, **15**(2): 265–286.

[69] Vu V Q and Lei J, Minimax sparse principal subspace estimation in high dimensions, *The Annals of Statistics*, 2013, **41**(6): 2905–2947.

[70] Cai T T, Ma Z M, and Wu Y H, Sparse PCA: Optimal rates and adaptive estimation, *The Annals of Statistics*, 2013, **41**(6): 3074–3110.

[71] Kolda T G and Bader B W, Tensor decompositions and applications, *SIAM Review*, 2009, **51**(3): 455–500.

[72] Sæbø S, Almøy T, Aarøe J, et al., St-pls: A multi-directional nearest shrunken centroid type classifier via pls, *Journal of Chemometrics: A Journal of the Chemometrics Society*, 2008, **22**(1): 54–62.

[73] Xia Y C, A multiple-index model and dimension reduction, *Journal of the American Statistical Association*, 2008, **103**(484): 1631–1640.

[74] Dennis Cook R and Li B, Dimension reduction for conditional mean in regression, *The Annals of Statistics*, 2002, **30**(2): 455–474.

[75] Saracco J M, An asymptotic theory for sliced inverse regression, *Communications in Statistics-Theory and Methods*, 1997, **26**(9): 2141–2171.

[76] Ravikumar P, Wainwright M J, Raskutti G, et al., High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence, *Electronic Journal of Statistics*, 2011, **5**: 935–980.

Appendix Proofs of Theorems

A.1 Proof of Theorem 3.1

By the central limit theorem, it can be easily shown that the sample estimate $\widehat{\boldsymbol{\Sigma}}$ is an $\sqrt{n}$-consistent estimator of $\boldsymbol{\Sigma}$. Additionally, by Theorem 1 in [75], $\widehat{\boldsymbol{U}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{U}$. By Proposition 3 in [36], given the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{U}}$, the estimator $\boldsymbol{P}_{\widehat{\boldsymbol{\Gamma}}}$ is $\sqrt{n}$-consistent.

A.2 Proof of Theorem 3.2

Under the sub-Gaussianity assumption of $\boldsymbol{X}$, the error bound of $\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max}$ can be obtained directly from [76], Lemma 1, that is,

$$\Pr(\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max}\leq\varepsilon)\geq 1-Cp^2\exp(-C'n\varepsilon^2),\quad 0<\varepsilon\leq C'.$$

The error bound of $\|\widehat{\boldsymbol{U}}-\boldsymbol{U}\|_{\max}$ has been confirmed in [25]. It has been established that under Assumptions (A1) and (A2), when $\varepsilon\geq c_0n^{-1/2}$ for some constant $c_0>0$,

$$\Pr(\|\widehat{\boldsymbol{U}}-\boldsymbol{U}\|_{\max}\leq\varepsilon)\geq 1-Cp^4\exp(-C'n\varepsilon^2).$$

Therefore, by applying Theorem 4 in [52], we have that with probability at least $1-Cp^4\exp(-C'n\varepsilon^2)$,

$$\|\boldsymbol{P}_{\widehat{\mathcal{E}}}-\boldsymbol{P}_{\mathcal{E}}\|_F^2\leq C\varepsilon\tau^2.$$

Take $\varepsilon=c\sqrt{\log p/n}$ for some constant $c>0$. When $\max\{\tau^2\sqrt{\log p},\Delta_{\boldsymbol{U}}^{-2}\tau^2\sqrt{\log p}=o(\sqrt{n})\}$, we have that

$$\Delta_{\boldsymbol{U}}^{-1}\{\sqrt{8c_1\nu^2\tau^2\varepsilon}+(2c_1\nu+1)\tau^2\varepsilon\}\to 0\quad\text{and}\quad\|\boldsymbol{P}_{\widehat{\mathcal{E}}}-\boldsymbol{P}_{\mathcal{E}}\|_F^2\leq C\tau^2\sqrt{\log p/n}\to 0.$$

This completes the proof.