A Review of Envelope Models

Minji Lee [*] and Zhihua Su [†]

October 1, 2019

Summary

The envelope model was first introduced as a parsimonious version of multivariate linear regression. It uses dimension reduction techniques to remove immaterial variation in the data and has the potential to gain efficiency in estimation and improve prediction. Many advances have taken place since its introduction, and the envelope model has been applied to many contexts in multivariate analysis, including partial least squares, generalized linear models, Bayesian analysis, variable selection, and quantile regression, among others. This article serves as a review of the envelope model and its developments for those who are new to the area.

Key words: dimension reduction; envelope model; multivariate linear regression; partial least squares; generalized linear models.

[*]Department of Statistics, University of Florida, Gainesville, Florida, USA
Email: mlee9@ufl.edu

[†]Department of Statistics, University of Florida, Gainesville, Florida, USA
Email: zhihuasu@stat.ufl.edu

1 Introduction

The envelope model aims to achieve efficient estimation in multivariate analysis. It was first introduced in the context of multivariate linear regression (Cook et al., 2010), and has been applied to many other contexts including partial least squares, matrix-variate or tensor-variate regression, variable selection, Bayesian linear regression, generalized linear regression, and quantile regression, among others. The structure of the envelope model has been extended and enriched to remove constraints on the data structure and further improve the efficiency in estimation (e.g. partial envelope model, inner envelope model, scaled envelope model, heteroscedastic envelope model and groupwise envelope model). All these developments are scattered in a number of research papers in different journals. Cook (2018) is a textbook that gives a comprehensive review of the envelope model. However, there is currently no short review of the envelope model designed for people who have little or no acquaintance with the area of envelope models. The goal of this article is to provide a quick introduction to the envelope model and a brief overview of the developments in the area aimed at non-experts. This article also include some important new developments, such as the envelope quantile regression, that appeared after the publication of Cook (2018).

The goal of the envelope model is to improve the estimation efficiency of standard multivariate analysis methods, sometimes equivalent to taking many more observations. It achieves efficiency gains by using sufficient dimension reduction techniques to remove immaterial information. With the development of measurement technologies, the volume of data has increased rapidly, and many datasets contain information that is immaterial to the goal at hand. For example, to assess the effectiveness and side effects of a new drug in clinical trials, many clinical variables of patients are recorded before and after taking the drug. While some clinical variables respond to the effect of the drug, other clinical variables, or more generally, some linear combinations of all clinical variables do not change after taking the drug. These linear combinations are called the immaterial part of the data. The immaterial part provides no information on the estimation of the parameter of interest, e.g. the effect of the drug; however, it brings extraneous variation in the estimation. The envelope model identifies and removes the immaterial part, so that the subsequent analysis is based on the material part only, and is therefore more efficient.

We first use an example to illustrate the estimation efficiency gains obtained by the envelope model and its working mechanism in the context of multivariate linear regression. The Berkeley guidance data (Tuddenham and Snyder, 1953) include height measurements for 39 boys and 54 girls born in 1928–1929 in Berkeley, CA. For illustration purposes, the heights at ages 13 and 14 are taken to be the bivariate responses $(Y_1, Y_2)^T$. The predictor X is the sex indicator, which takes value 0 or 1 to denote girls or boys, respectively. Then $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = \mathrm{E}(\mathbf{Y} \mid X = 1) - \mathrm{E}(\mathbf{Y} \mid X = 0)$ describes the height difference between boys and girls. We first fit the standard multivariate linear regression model to the data. The left panel of Figure 1 shows the standard inference on β_1. The projection path for a representative point 'X' is marked as A. The two curves at the bottom present the projection distribution of the two groups (boys and girls) onto the Y_1 axis. The two curves are barely distinguishable, which suggests that a large sample size is required to detect the difference under the standard model. To estimate the standard deviation of the estimator of $\boldsymbol{\beta}$, the bootstrap standard deviation based on the residual bootstrap (the bootstrap method that involves resampling the residuals) with 200 replications is calculated. The bootstrap standard deviations for the two elements in the standard estimator of $\boldsymbol{\beta}$ are 1.80 and 1.81. The inference under the envelope model is described in the right panel of Figure 1. The envelope model exploits the feature that the distribution of the linear combination of $\mathbf{Y}$ marked as $\boldsymbol{\Gamma}_0^T\mathbf{Y}$ does not depend on

X, where $\mathbf{\Gamma}_0 \in \mathbb{R}^2$ is a vector that indicates the direction of the linear combination. So a point 'X' is first projected onto the direction that is orthogonal to $\mathbf{\Gamma}_0$, marked as $\mathbf{\Gamma}$, to remove the immaterial information, and is then projected onto the horizontal axis. The two segments in the projection path are marked as B_1 and B_2. The two distribution curves are well separated, which indicates the efficiency gains from the envelope estimation. The bootstrap standard deviations for the two elements in the envelope estimator of $\boldsymbol{\beta}$ are 0.19 and 0.19. In other words, under the standard inference, we would need a sample size that is about $(1.80/0.19)^2 = 90$ times the original sample size to achieve the same accuracy as we have here. The direction $\mathbf{\Gamma}$ is estimated as $(1, -1)^T$ and the direction $\mathbf{\Gamma}_0$ is estimated as $(1, 1)^T$ (estimation is discussed in Section 2.2), which means that the average height at ages 13 and 14, i.e. $(Y_1 + Y_2)/2$, is immaterial (or independent) to the changes in X, while the height difference $Y_1 - Y_2$ is material (or dependent) to the changes in X. This example shows that the mechanism of the efficiency gains in the envelope model is to identify the immaterial part of $\mathbf{Y}$, and carry out the subsequent analysis only on the material part. From the envelope inference (see the right panel of Figure 1), it is easy to tell that girls are taller than boys on age 13, and the height difference is significant. However, standard inference suggests that the height difference is insignificant. Because of the efficiency gains in estimation, the envelope model can detect weak signals which the standard model fails to detect. A formal definition of the envelope model and more details regarding this example are given in Section 2.1.

The rest of the article is organized as follows. Section 2 introduces the first envelope model (Cook et al., 2010), and describes a general framework of model formulation, estimation, selection of envelope dimension and inference under the envelope model. Section 3 surveys the developments in the envelope model since its introduction. Its applications in partial least squares, matrix-variate or tensor-variate regression, variable selection, Bayesian linear regression, general-

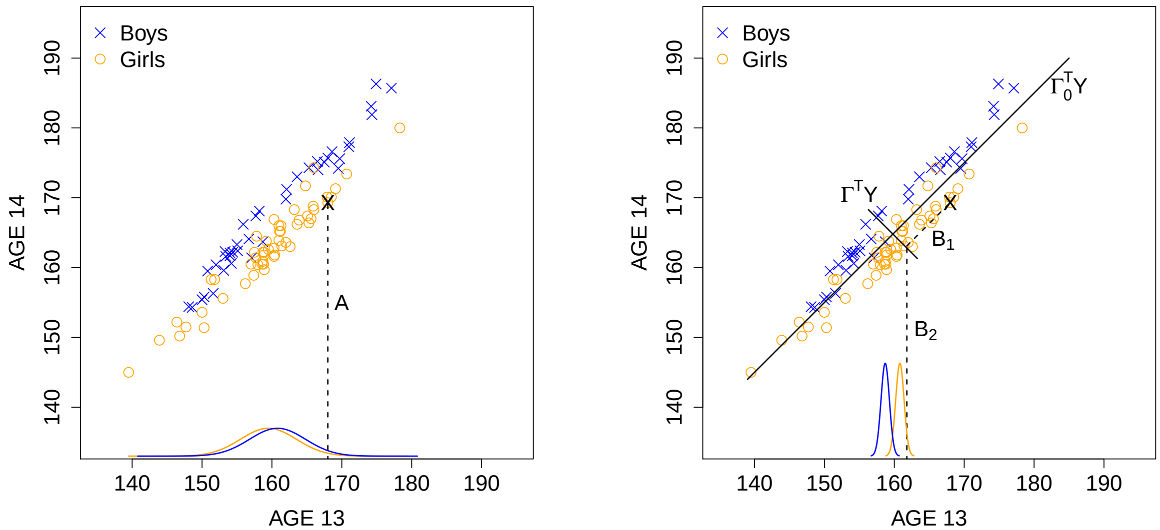

Figure 1: Height of boys and girls at ages 13 and 14. Left panel: Illustration of inference under the standard linear regression model. Right panel: Illustration of inference under the envelope model.

ized linear models, and quantile regression are elaborated in Sections 3.1 – 3.6. Envelope models in other contexts of multivariate analysis are discussed in Section 3.7. Developments in the structure of the envelope model are reviewed in Section 3.8. Software for envelope models is presented in Section 3.9. Further research directions are discussed in Section 4.

2 The First Envelope Model

2.1 Formulation

The envelope model was first introduced by (Cook et al., 2010) under the classical multivariate linear regression model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^r$ is the response vector, $\mathbf{X} \in \mathbb{R}^p$ is the predictor vector, which can be fixed or random, and $\boldsymbol{\varepsilon}$ is the error vector that has mean 0 and positive definite covariance matrix $\boldsymbol{\Sigma}$. The intercept $\boldsymbol{\mu} \in \mathbb{R}^r$, the regression coefficient matrix $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ and the error covariance matrix $\boldsymbol{\Sigma}$ are unknown parameters. The standard method to estimate $\boldsymbol{\beta}$ is to perform an independent linear regression for each element in $\mathbf{Y}$ on $\mathbf{X}$. The relationship among the response variables is not used.

The envelope model seeks to achieve efficient estimation of $\boldsymbol{\beta}$ by performing dimension reduction on $\mathbf{Y}$. Let $\mathcal{S}$ be a d-dimensional subspace of $\mathbb{R}^r$ ($d \leq r$). Let $\mathbf{G} \in \mathbb{R}^{r \times d}$ be an orthonormal basis of $\mathcal{S}$ and $\mathbf{G}_0 \in \mathbb{R}^{(r-d) \times d}$ be an orthonormal basis of $\mathcal{S}^{\perp}$, the orthogonal complement of $\mathcal{S}$. Then $\mathbf{G}^T\mathbf{Y}$ forms a reduction of $\mathbf{Y}$. The envelope model imposes the following two conditions on $\mathbf{G}^T\mathbf{Y}$ and $\mathbf{G}_0^T\mathbf{Y}$: (i) $\mathbf{G}_0^T\mathbf{Y}|\mathbf{X} \sim \mathbf{G}_0^T\mathbf{Y}$ and (ii) $\mathrm{cov}(\mathbf{G}^T\mathbf{Y}, \mathbf{G}_0^T\mathbf{Y}|\mathbf{X}) = 0$, where $\sim$ means equal in distribution. Condition (i) indicates that the distribution of $\mathbf{G}_0^T\mathbf{Y}$ does not depend on $\mathbf{X}$ and thus does not carry information on $\boldsymbol{\beta}$. Condition (ii) indicates that $\mathbf{G}_0^T\mathbf{Y}$ does not carry information about $\boldsymbol{\beta}$ indirectly through its correlation with $\mathbf{G}^T\mathbf{Y}$. Therefore, all the material information of $\boldsymbol{\beta}$ is contained in the reduction $\mathbf{G}^T\mathbf{Y}$, and $\mathbf{G}_0^T\mathbf{Y}$ only contains the immaterial information. Under model (1), Cook et al. (2010) showed that (i) and (ii) hold if and only if (a) span($\boldsymbol{\beta}$) $\subseteq \mathcal{S}$ and (b) $\boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{S}}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{S}} + \mathbf{Q}_{\mathcal{S}}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{S}}$, where $\mathbf{P}$ represents a projection operator and $\mathbf{Q} = \mathbf{I} - \mathbf{P}$. When $\mathcal{S}$ satisfies (b), it is called a reducing subspace of $\boldsymbol{\Sigma}$ (Conway, 1990). The $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ is defined as the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathrm{span}(\boldsymbol{\beta})$. It is denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, or $\mathcal{E}$ for short. In other words, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is the smallest subspace that satisfies conditions (a) and (b) (or equivalently, conditions (i) and (ii)). Let u ($0 \leq u \leq r$) denote the dimension of $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. Let $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ denote orthonormal basis matrices for $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})^{\perp}$, respectively. Then $\boldsymbol{\Gamma}^T\mathbf{Y}$ represents the minimal reduction of $\mathbf{Y}$ that satisfies conditions (i) and (ii). When (a) and (b) hold with $\mathcal{S} = \mathcal{E}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$, model (1) can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad \mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T + \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T, \tag{2}$$

where $\boldsymbol{\beta} = \mathbf{\Gamma}\boldsymbol{\eta}$. The matrix $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ carries the coordinates of $\boldsymbol{\beta}$ with respect to $\mathbf{\Gamma}$. The matrices $\mathbf{\Omega} \in \mathbb{R}^{u \times u}$ and $\mathbf{\Omega}_0 \in \mathbb{R}^{(r-u)\times(r-u)}$ are coordinates of $\mathbf{\Sigma}$ with respect to $\mathbf{\Gamma}$ and $\mathbf{\Gamma}_0$, respectively. Typically, $\boldsymbol{\beta}$ is the parameter of main interest, and $\boldsymbol{\eta}$, $\mathbf{\Omega}$, $\mathbf{\Omega}_0$ and $\mathcal{E}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$ are constituent parameters. We call (2) the envelope model, and call (1) the standard model hereafter. The response $\mathbf{Y}$ can be decomposed to the material part $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ and immaterial part $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$. Under the envelope model (2), $\boldsymbol{\beta}$ is only related to the material part, and the covariance matrix can be decomposed as the sum of the variance of the material part $\mathrm{var}(\mathbf{P}_{\mathcal{E}}\mathbf{Y}) = \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T$ and the variance of the immaterial part $\mathrm{var}(\mathbf{Q}_{\mathcal{E}}\mathbf{Y}) = \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T$. Cook et al. (2010) showed that the envelope estimator of $\boldsymbol{\beta}$ is asymptotically more efficient than or as efficient as the standard estimator. The efficiency gains are substantial especially when the immaterial variation is larger than the material variation, or in other words, $\|\mathrm{var}(\mathbf{Q}_{\mathcal{E}}\mathbf{Y})\| = \|\mathbf{\Omega}_0\| \geq \|\mathbf{\Omega}\| = \|\mathrm{var}(\mathbf{P}_{\mathcal{E}}\mathbf{Y})\|$, where $\|\cdot\|$ denotes the spectral norm of a matrix. When $u = r$, $\mathcal{E}_{\mathbf{\Sigma}}(\boldsymbol{\beta}) = \mathbb{R}^r$, and the envelope model reduces to the standard model.

In the height example in Section 1, we have $\|\widehat{\mathbf{\Omega}}\| = 1.56$ and $\|\widehat{\mathbf{\Omega}}_0\| = 118.7$. Recall that the bootstrap standard deviations for the two elements in the standard estimator of $\boldsymbol{\beta}$ are 1.80 and 1.81, and 0.19 and 0.19 for the two elements in the envelope estimator of $\boldsymbol{\beta}$. This confirms that we can achieve substantial efficiency gains when the immaterial variation is larger than the material variation. On the other hand, if we take the height measurements at ages 17 and 18 as the bivariate responses $(Y_1, Y_2)^T$ and fit the envelope model, we have $\|\widehat{\mathbf{\Omega}}\| = 79.5$ and $\|\widehat{\mathbf{\Omega}}_0\| = 0.156$. In this case, the variation of the immaterial part is smaller than that of the material, as shown in Figure 2.

The bootstrap standard deviations are 1.36 and 1.37 for the two elements in the standard estimator of $\boldsymbol{\beta}$, and 1.30 and 1.37 for the two elements in the envelope estimator of $\boldsymbol{\beta}$. In this case, the envelope model offers limited efficiency gains in the estimation of $\boldsymbol{\beta}$. As indicated in the right panel, the project of a point 'X' under the envelope model is very close to that under the standard model, indicating that the envelope inference and the standard inference yield similar results.

More illustration examples can be found in Section 2 of Cook (2018).

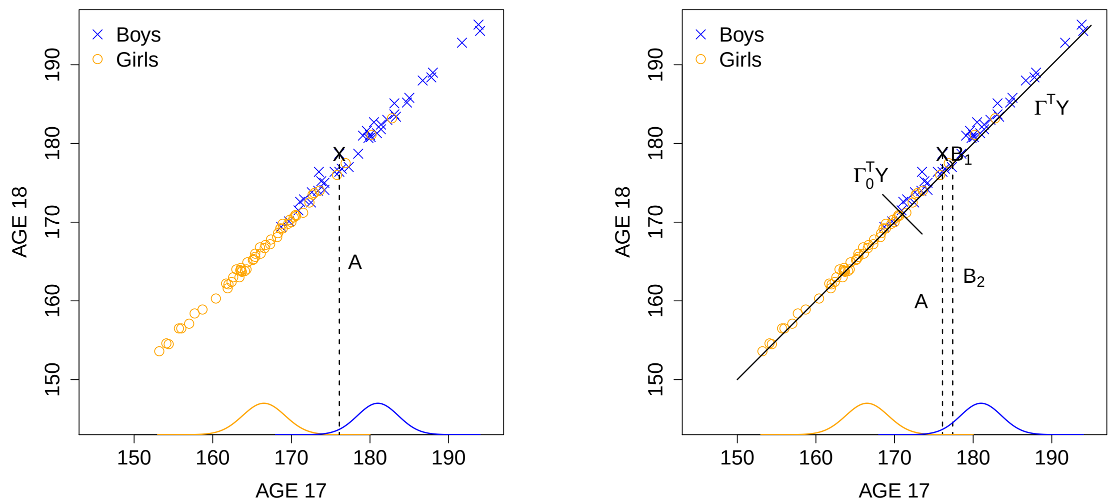

Figure 2: Height of boys and girls at ages 17 and 18. Left panel: Illustration of inference under the standard model. Right panel: Illustration of inference under the envelope model.

2.2 Estimation

The parameters involved in the envelope model (2) include the dimension of the envelope subspace u and the model parameters $\boldsymbol{\mu}$, $\boldsymbol{\eta}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_0$ and $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. In this section, we assume that u is known and will discuss the selection of u in Section 2.3. Cook et al. (2010) used the normal log likelihood (of $\mathbf{Y}$ given $\mathbf{X}$) as the objective function for estimation, i.e.,

$$(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Omega}}_0, \widehat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})) = \underset{\boldsymbol{\mu},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})}{\arg\max} L_u(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}))$$

where

$$\begin{aligned} L_u &= -\frac{nr}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}(\mathbf{Y}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i)^T(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)^{-1}(\mathbf{Y}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i). \end{aligned}$$

Since $\boldsymbol{\eta}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ all depend on $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, we first fix $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and estimate the other parameters by maximizing the normal likelihood. After some calculations, the estimators of the other parameters can be written as explicit functions of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} - \mathbf{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\text{ols}}\bar{\mathbf{X}}, \quad \hat{\boldsymbol{\eta}} = \boldsymbol{\Gamma}^T\widehat{\boldsymbol{\beta}}_{\text{ols}} = \boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{-1}, \quad \widehat{\boldsymbol{\Omega}} = \boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}|\mathbf{X}}\boldsymbol{\Gamma} \quad \text{and} \quad \widehat{\boldsymbol{\Omega}}_0 = \boldsymbol{\Gamma}_0^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}}\boldsymbol{\Gamma}_0,$$

where $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ are the sample means of $\mathbf{X}$ and $\mathbf{Y}$ respectively, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$ are the sample covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$ respectively, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}$ is the sample covariance matrix of $\mathbf{Y}$ and $\mathbf{X}$, $\widehat{\boldsymbol{\beta}}_{\text{ols}} = \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{-1}$ is the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$, and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}|\mathbf{X}} = \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}^T$ is the sample covariance matrix of the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$. Substituting these estimators into L_u, we obtain the objective function for estimation of $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ as

$$\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \underset{\text{span}(\mathbf{H})\in\mathcal{G}(r,u)}{\arg\min} \log|\mathbf{H}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}|\mathbf{X}}\mathbf{H}| + \log|\mathbf{H}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}}^{-1}\mathbf{H}|, \tag{3}$$

where $\mathcal{G}(r,u)$ denotes the $r \times u$ Grassmann manifold. An $r \times u$ Grassmann manifold $\mathcal{G}(r,u)$ is the set of all u-dimensional subspaces of an r-dimensional space. Once $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$ is obtained, $\widehat{\mathbf{\Gamma}}$ and $\widehat{\mathbf{\Gamma}}_0$ can be taken as any orthonormal bases of $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$ and $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\boldsymbol{\beta})^{\perp}$. Then the envelope estimators of the constituent parameters are

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \bar{\mathbf{Y}} - \mathbf{P}_{\widehat{\mathbf{\Gamma}}}\widehat{\boldsymbol{\beta}}_{\text{ols}}\bar{\mathbf{X}}, & \hat{\boldsymbol{\eta}} &= \widehat{\mathbf{\Gamma}}^T\widehat{\boldsymbol{\beta}}_{\text{ols}}, \\ \widehat{\mathbf{\Omega}} &= \widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{\Sigma}}_{\mathbf{Y}|\mathbf{X}}\widehat{\mathbf{\Gamma}}, & \widehat{\mathbf{\Omega}}_0 &= \widehat{\mathbf{\Gamma}}_0^T\widehat{\mathbf{\Sigma}}_{\mathbf{Y}}\widehat{\mathbf{\Gamma}}_0. \end{aligned}$$

And the envelope estimators of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{\Sigma}}$ are

$$\widehat{\boldsymbol{\beta}} = \widehat{\mathbf{\Gamma}}\hat{\boldsymbol{\eta}} = \mathbf{P}_{\widehat{\mathcal{E}}}\widehat{\boldsymbol{\beta}}_{\text{ols}}, \quad \text{and} \quad \widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Omega}}\widehat{\mathbf{\Gamma}}^T + \widehat{\mathbf{\Gamma}}_0\widehat{\mathbf{\Omega}}_0\widehat{\mathbf{\Gamma}}_0^T.$$

Note that the envelope model (2) does not rely on normality. The normal log likelihood is used as an objective function to obtain the envelope estimator. Section 1.9 of Cook (2018) points out that if the errors have finite fourth moments, and the maximum leverage tends to 0 as $n \to \infty$, then the envelope estimator is $\sqrt{n}$-consistent. Numerical experiments show that under a mild or moderate departure from normality, the envelope estimator has about the same efficiency as under normal errors (Su and Cook, 2011, 2013; Park et al., 2017). The $\sqrt{n}$-consistency of the inner envelope estimator (c.f. Section 3.8) is established in Su and Cook (2012), and the $\sqrt{n}$-consistency for other envelope extensions can be established similarly.

The optimization in (3) is discussed in several works. Cook and Zhang (2016) developed the one-direction-at-a-time (1D) algorithm, which estimates the columns of the basis of $\mathcal{E}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$ sequentially. Under the 1D algorithm framework, the envelope coordinate descent (ECD) algorithm was recently derived in Cook and Zhang (2018). The ECD algorithm utilizes an approximation in estimating each column and turns out to be faster than the 1D algorithm. Cook et al. (2016) proposed a non-Grassmann estimation algorithm. The key idea is based on a reparametrization of (2). Instead of taking $\mathbf{\Gamma}$ to be an arbitrary orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}}(\boldsymbol{\beta})$, it is required that u rows in $\mathbf{\Gamma}$ form an identity matrix. The other $(r-u)u$ parameters in $\mathbf{\Gamma}$ are unconstrained. Thus the Grassmann manifold optimization in (3) can be converted to an unconstrained matrix optimization. All these algorithms yield $\sqrt{n}$-consistent estimators. Since the optimization in (3) is non-convex, the choice of starting values was investigated in Cook et al. (2016). When r is large, Cook and Zhang (2018) proposed a screening algorithm, called envelope component screening (ECS), that can reduce the dimension r to a manageable dimension d prior to the application of an optimization algorithm.

2.3 Selection of u

The dimension of the envelope subspace, u, is a model selection parameter. Likelihood ratio testing (LRT) and information criteria are the most common tools for the selection of u. LRT is based on sequential hypothesis testing, starting with $u = 0$ and using a common significance level α. The estimate $\hat{u}$ of u is chosen as the first hypothesized value that is not rejected. Among the information criteria, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) are the most prevalent. AIC or BIC values for all possible candidates of u are computed. The candidate that corresponds to the smallest AIC or BIC is selected as the envelope dimension. Su and Cook (2013) compared the numerical performance of model selection criteria, and found that LRT is stable when the sample size is small, but asymptotically the error probability is equal to the significance level. AIC is likely to overestimate u, and works well for larger u. BIC is consistent, and works well when the sample size is not too small. Under the normal error, Cook and Su (2013) proves that BIC is selection consistent, while AIC selects a model that at least contains the true model. However, theoretical properties for LRT, AIC and BIC under nonnormal errors remain unknown. Nonparametric methods such as cross validation can also be used to choose u. Recently Zhang and Mai (2018) proposed two non-likelihood based approaches, called full Grassmannian (FG) and 1D selections. Their selection methods are based on Grassmann optimization and 1D algorithms, and can be applied to many contexts that arise in envelope models. They established consistency of both methods. To avoid the selection of u, Eck and Cook (2017) constructed a weighted envelope estimator, which is a weighted average of all envelope estimators with $u = 1, \ldots, r$. The weights are calculated based on the BIC values for each model. Eck and Cook (2017) showed that the bootstrap sample variance of the weighted envelope estimator is a $\sqrt{n}$-consistent estimator of the asymptotic variance of the envelope estimator when u is known.

2.4 Inference

The main issue in inference under the envelope model is the estimation of the variance of the envelope estimator. The estimation procedure of the envelope model (2) takes two steps. The first is to estimate u and the second step is to estimate the other parameters when u is fixed. Therefore the variance of the envelope estimator comes from two sources, the variability in the model selection and the variability in the estimation of parameters in the selected model. As mentioned earlier, if u is known, the envelope estimator is asymptotically at least as efficient as the standard estimator (the OLS estimator). In this case, the variance of the envelope estimator can be estimated through its asymptotic variance or bootstrapping. The estimation of u adds additional variability to the estimator. When the variation of the immaterial part is small compared to that of the material part and the dimension of the envelope model is large relative to r (i.e. u is close to r), the envelope estimator could be less efficient than the standard estimator. However, if the immaterial part has variation that is larger than that of the material part, the envelope estimator is still expected to provide efficiency gains. Up to now, there are no theoretical results on the efficiency gains of the envelope model considering both the variability due to the selection of u and the variability due to estimation of the parameters in the selected model. An investigation on such theory would provide a clearer guideline on when to use the envelope models.

3 Advances in Envelope Models

Although the formulation of the first envelope model is based on dimension reduction of the response vector $\mathbf{Y}$, the construction of envelope subspaces is flexible and can be based on dimension reduction of the predictor vector $\mathbf{X}$ or other objects. It does not rely on the linear model or any parametric model either. In this section, we describe developments and extensions of the envelope model. Estimation and inference under the envelope models in this section are similar to those described in Sections 2.2, 2.3 and 2.4 unless otherwise discussed.

3.1 Predictor Envelope Model and Partial Least Squares

Partial Least Squares (Wold, 1966, PLS) is an alternative to OLS for estimating the regression coefficients in linear regression. It is the dominant method in chemometrics and is now widely used in many applied scientific disciplines. It is known that PLS often has superior prediction performances compared to OLS, especially in high-dimensional settings. Historically, PLS has been defined in terms of the iterative algorithms NIPALS and SIMPLS. As a result, knowledge about its theoretical properties is limited. Cook et al. (2013) developed a link between the predictor envelope and PLS, allowing PLS to be studied in a traditional likelihood-based framework.

The predictor envelope model is derived under the linear regression model (1), but $\mathbf{Y}$ can be univariate or multivariate, and $\mathbf{X}$ is a random vector with mean $\boldsymbol{\mu}_\mathbf{X}$ and covariance matrix $\boldsymbol{\Sigma}_\mathbf{X}$. The predictor envelope model is constructed based on a dimension reduction of the predictor vector $\mathbf{X}$. It considers the $\boldsymbol{\Sigma}_\mathbf{X}$-envelope of $\mathrm{span}(\boldsymbol{\beta}^T)$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\boldsymbol{\beta}^T)$. The envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\boldsymbol{\beta}^T)$ divides $\mathbf{X}$ into two parts, the material part $\mathbf{P}_\mathcal{E}\mathbf{X}$ and the immaterial part $\mathbf{Q}_\mathcal{E}\mathbf{X}$, such that they satisfy two conditions (i) $\mathbf{Q}_\mathcal{E}\mathbf{X}$ is uncorrelated with $\mathbf{P}_\mathcal{E}\mathbf{X}$ and (ii) $\mathbf{Y}$ is uncorrelated with $\mathbf{Q}_\mathcal{E}\mathbf{X}$ given $\mathbf{P}_\mathcal{E}\mathbf{X}$. Conditions (i) and (ii) are equivalent to that $(\mathbf{Y}, \mathbf{P}_\mathcal{E}\mathbf{X})$ is uncorrelated with $\mathbf{Q}_\mathcal{E}\mathbf{X}$, and therefore $\mathbf{Q}_\mathcal{E}\mathbf{X}$ is effectively linearly immaterial to the regression. Let u denote the dimension of $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\boldsymbol{\beta}^T)$, and $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ denote an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\boldsymbol{\beta}^T)$. The predictor envelope model is formulated as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\mathbf{X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Sigma}_\mathbf{X} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{4}$$

where $\boldsymbol{\beta}^T = \boldsymbol{\Gamma}\boldsymbol{\eta}$ and $\boldsymbol{\eta} \in \mathbb{R}^{u\times r}$ carries the coordinates of $\boldsymbol{\beta}^T$ with respect to $\boldsymbol{\Gamma}$. The matrix $\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T = \mathrm{var}(\mathbf{P}_\mathcal{E}\mathbf{X})$ presents the variation of the material part, $\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T = \mathrm{var}(\mathbf{Q}_\mathcal{E}\mathbf{X})$ presents the variation of the immaterial part, where $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p\times(p-u)}$ is an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_\mathbf{X}}(\mathcal{B}^T)^\perp$ and $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u)\times(p-u)}$ are positive definite matrices. Compared to the envelope model (2), the predictor envelope model (4) seeks a dimension reduction on $\mathbf{X}$ instead of on $\mathbf{Y}$. Note that in the context of linear regression, the response envelope model (2) requires the material part and immaterial part to be conditionally uncorrelated, which is also required by other envelope models that perform dimension reduction on $\mathbf{Y}$ in regression setting, such as in matrix or tensor variate envelope model (c.f. Section 3.2). In contrast, marginal uncorrelation between the material part and immaterial part is required by the predictor envelope model or other envelope models that perform dimension reduction on $\mathbf{X}$, such as envelope model in generalized linear model (c.f. Section 3.5) or envelope quantile regression (c.f. Section 3.6).

On the other hand, PLS operates by first reducing the predictors to a few linear combinations, $\mathbf{X} \mapsto \mathbf{W}^T\mathbf{X}$, where $\mathbf{W} \in \mathbb{R}^{p\times d}$ is a semi-orthogonal matrix, $d \leq p$ and d is called number of components. A popular algorithm is SIMPLS (De Jong, 1993), which constructs an estimator $\widehat{\mathbf{W}}_{\mathrm{PLS}}$ of $\mathbf{W}$ sequentially as follows: Set $\widehat{\mathbf{w}}_1$ to be the eigenvector of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}}^T$ corresponding to its largest eigenvalue. Let $\widehat{\mathbf{W}}_k = (\widehat{\mathbf{w}}_1, \ldots, \widehat{\mathbf{w}}_k)$, $k = 1, \ldots, d-1$. Given $\widehat{\mathbf{W}}_k$,

$$\widehat{\mathbf{w}}_{k+1} = \arg\max_{\mathbf{w}} \mathbf{w}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}}^T\mathbf{w}, \quad \text{subject to } \mathbf{w}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}\widehat{\mathbf{W}}_k = 0 \text{ and } \mathbf{w}^T\mathbf{w} = 1. \tag{5}$$

Then $\widehat{\mathbf{W}}_{\mathrm{PLS}} = \widehat{\mathbf{W}}_d$. Once we have $\widehat{\mathbf{W}}_{\mathrm{PLS}}$, we fit the reduced model $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\eta}^T(\widehat{\mathbf{W}}_{\mathrm{PLS}}^T\mathbf{X}) + \boldsymbol{\varepsilon}$, and obtain the OLS estimator of the coefficients $\boldsymbol{\eta} \in \mathbb{R}^{d\times r}$: $\hat{\boldsymbol{\eta}} = (\widehat{\mathbf{W}}_{\mathrm{PLS}}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}\widehat{\mathbf{W}}_{\mathrm{PLS}})^{-1}\widehat{\mathbf{W}}_{\mathrm{PLS}}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{XY}}$. The PLS estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}} = (\widehat{\mathbf{W}}_{\mathrm{PLS}}\hat{\boldsymbol{\eta}})^T = \widehat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\widehat{\mathbf{W}}_{\mathrm{PLS}}(\widehat{\mathbf{W}}_{\mathrm{PLS}}^T\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}\widehat{\mathbf{W}}_{\mathrm{PLS}})^{-1}\widehat{\mathbf{W}}_{\mathrm{PLS}}^T = \widehat{\boldsymbol{\beta}}_{\mathrm{ols}}\mathbf{P}_{\widehat{\mathbf{W}}_{\mathrm{PLS}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}})}^T$, where $\mathbf{P}_{\widehat{\mathbf{W}}_{\mathrm{PLS}}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}})}^T$ denotes the projection matrix onto $\mathrm{span}(\widehat{\mathbf{W}}_{\mathrm{PLS}})$ in the $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}$ inner product. Other variants use different inner products in the constraints or different objective function. For example, NIPALS (Wold, 1975) modifies the length constraint $\mathbf{w}^T\mathbf{w} = 1$ to $\mathbf{w}^T(\mathbf{I}_p - \mathbf{P}_{\widehat{\mathbf{W}}_k})\mathbf{w} = 1$ (Cook, 2018, Section 4.2.1).

Cook et al. (2013) showed that $\mathrm{span}(\widehat{\mathbf{W}}_{\mathrm{PLS}})$ is also a $\sqrt{n}$-consistent estimator of $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}^T)$ when the number of PLS components d equals u, the dimension of $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}^T)$. Thus there is a close connection between SIMPLS and the envelope model: The envelope and SIMPLS testimators, $\widehat{\boldsymbol{\beta}}_{\mathrm{env}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, are based on the same population construct $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}^T)$, but differ in their estimation methods. While SIMPLS uses the iterative algorithm (5), the predictor envelope model uses the likelihood-based estimation procedure discussed in Section 2.2. As a result, $\widehat{\boldsymbol{\beta}}_{\mathrm{env}}$ typically dominates $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ in both estimation and prediction accuracy, especially when the variation of the immaterial part is large. While the envelope estimator typically dominates the PLS estimator, there are scalable versions of the PLS algorithm available for big data applications (Schwartz et al., 2010; Zeng and Li, 2014; Tabei et al., 2016; Cook and Forzani, 2018) and their theoretical properties are explored in Cook and Forzani (2018, 2019). The development of scalable envelope estimator serviceable to big data is an interesting future research direction.

The predictor envelope model provides an avenue to study PLS and extend the scope of PLS. The PLS estimator is not invariant or equivariant to scale transformations of $\mathbf{X}$, which tends to limit its scope to applications where the predictors are measured in the same or similar units. Cook and Su (2016) derived the scaled predictor envelope model that incorporates predictor scaling into the model formulation. In addition to being a scale-invariant method, the scaled predictor envelope model can offer efficiency gains beyond those given by PLS, and further reduce prediction errors. Sparse PLS (Chun and Keleş, 2010; Chung and Keleş, 2010; Huang et al., 2004; Lê Cao et al., 2008; Lee et al., 2011) has been derived and studied in statistics, chemometrics and genetics to perform variable selection in PLS. Zhu and Su (2019) derived an envelope-based sparse PLS estimator for both linear regression and generalized linear models, and showed that the envelope-based sparse PLS estimator has better prediction performance than existing sparse PLS estimators especially when the immaterial part has larger variation than the material part. They also established the oracle property and asymptotic normality of the estimator and derived the convergence rate in high-dimensional setting.

3.2 Envelope Model with Matrix or Tensor Variate Response

In many applications such as neuroimaging, social networks or signal processing, the response or predictor can be a matrix or a tensor instead of a vector. Li and Zhang (2017) and Ding and Cook (2018) extended the envelope model (2) to matrix-valued and tensor-valued responses or / and predictors.

Ding and Cook (2018) considered the linear regression model with both matrix-valued predictor and matrix-valued response

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}_1 \mathbf{X} \boldsymbol{\beta}_2^T + \boldsymbol{\varepsilon}, \tag{6}$$

where $\mathbf{Y} \in \mathbb{R}^{r_1 \times r_2}$ and $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$ denote the response and predictor, $\boldsymbol{\mu} \in \mathbb{R}^{r_1 \times r_2}$ is the intercept, and $\boldsymbol{\beta}_1 \in \mathbb{R}^{r_1 \times p_1}$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{r_2 \times p_2}$ are the row and column coefficient matrices. For identifiability, $\boldsymbol{\beta}_2$ is defined to have Frobenius norm 1. The error $\boldsymbol{\varepsilon}$ has mean 0 and a separable Kronecker covariance structure (Hoff, 2011; Fosdick and Hoff, 2014), i.e. $\mathrm{cov}[\,\mathrm{vec}(\boldsymbol{\varepsilon})] = \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$, where $\otimes$ denotes the Kronecker product, the vec operator stacks a matrix into a vector columnwise, and $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{r_1 \times r_1}$ and $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{r_2 \times r_2}$ are positive definite matrices.

Under the matrix regression model (6), Ding and Cook (2018) assumed that the envelope structure can be assumed on both rows and columns of $\mathbf{Y}$. More specifically, let $\mathcal{S}_1 \subseteq \mathbb{R}^{r_1}$ and $\mathcal{S}_2 \subseteq \mathbb{R}^{r_2}$ be two subspaces such that they satisfies conditions (a) $\mathbf{Q}_{\mathcal{S}_1}\mathbf{Y} \mid \mathbf{X} \sim \mathbf{Q}_{\mathcal{S}_1}\mathbf{Y}$, (b) $\mathrm{cov}_c(\mathbf{P}_{\mathcal{S}_1}\mathbf{Y}, \mathbf{Q}_{\mathcal{S}_1}\mathbf{Y} \mid \mathbf{X}) = 0$, (c) $\mathbf{Y}\mathbf{Q}_{\mathcal{S}_2} \mid \mathbf{X} \sim \mathbf{Y}\mathbf{Q}_{\mathcal{S}_2}$ and (d) $\mathrm{cov}_r(\mathbf{Y}\mathbf{P}_{\mathcal{S}_2}, \mathbf{Y}\mathbf{Q}_{\mathcal{S}_2} \mid \mathbf{X}) = 0$, where cov_c and cov_r denote the row covariance and column covariance. For two matrix-value variables $\mathbf{U} \in \mathbb{R}^{r_1 \times r_2}$ and $\mathbf{V} \in \mathbb{R}^{r_1 \times r_2}$, $\mathrm{cov}_c(\mathbf{U}, \mathbf{V}) = \mathrm{E}\{[\mathbf{U} - \mathrm{E}(\mathbf{U})][\mathbf{V} - \mathrm{E}(\mathbf{V})]^T\}$ and $\mathrm{cov}_r(\mathbf{U}, \mathbf{V}) = \mathrm{E}\{[\mathbf{U} - \mathrm{E}(\mathbf{U})]^T[\mathbf{V} - \mathrm{E}(\mathbf{V})]\}$. The smallest subspace that satisfies conditions (a) and (b) is the $\boldsymbol{\Sigma}_1$-envelope of $\mathrm{span}(\boldsymbol{\beta}_1)$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_1}(\boldsymbol{\beta}_1)$. Its dimension is denoted as u_1

$(0 \le u_1 \le r_1)$, and $\mathbf{\Gamma}_1 \in \mathbb{R}^{r_1 \times u_1}$ denotes an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}_1}(\boldsymbol{\beta}_1)$. Similarly, the smallest subspace that satisfies conditions (c) and (d) is the $\mathbf{\Sigma}_2$-envelope of $\mathrm{span}(\boldsymbol{\beta}_2)$, denoted by $\mathcal{E}_{\mathbf{\Sigma}_2}(\boldsymbol{\beta}_2)$. We use u_2 $(0 \le u_2 \le r_2)$ to denote its dimension and $\mathbf{\Gamma}_2 \in \mathbb{R}^{r_2 \times u_2}$ to denote an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}_2}(\boldsymbol{\beta}_2)$. Then the matrix regression model (6) can be parametrized as

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{\Gamma}_1 \boldsymbol{\eta}_1 \mathbf{X} \boldsymbol{\eta}_2^T \mathbf{\Gamma}_2^T + \boldsymbol{\varepsilon}, \qquad \mathbf{\Sigma}_i = \mathbf{\Gamma}_i \mathbf{\Omega}_i \mathbf{\Gamma}_i^T + \mathbf{\Gamma}_{i0} \mathbf{\Omega}_{i0} \mathbf{\Gamma}_{i0}^T \quad \text{for} \ \ i = 1, 2, \tag{7}$$

where $\boldsymbol{\beta}_i = \mathbf{\Gamma}_i \boldsymbol{\eta}_i$, and $\boldsymbol{\eta}_i \in \mathbb{R}^{u_i \times p_i}$ carries the coordinates of $\boldsymbol{\beta}_i$ with respect to $\mathbf{\Gamma}_i$. The matrix $\mathbf{\Gamma}_{i0} \in \mathbb{R}^{r_i \times (r_i - u_i)}$ is an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}_i}(\boldsymbol{\beta}_i)^{\perp}$, and $\mathbf{\Omega}_i \in \mathbb{R}^{u_i \times u_i}$ and $\mathbf{\Omega}_{i0} \in \mathbb{R}^{(r_i - u_i) \times (r_i - u_i)}$ carry the coordinates of $\mathbf{\Sigma}_i$ with respect to $\mathbf{\Gamma}_i$ and $\mathbf{\Gamma}_{i0}$. Ding and Cook (2018) proved that the envelope estimators of $\boldsymbol{\beta}_i$ and $\mathbf{\Sigma}_i$ derived from model (7) are at least as efficient as the standard estimator from model (6) asymptotically. Furthermore, they developed the sparse matrix variate regression that accommodates sparsity structures in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ under the envelope model (7).

Li and Zhang (2017) adopted the envelope model to the tensor response linear model where the response is an mth order tensor and the predictor is a vector,

$$\mathbf{Y} = \mathbf{B} \bar{\times}_{(m+1)} \mathbf{X} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\mathbf{Y} \in \mathbb{R}^{r_1 \times \cdots \times r_m}$ is an mth order tensor response, $\mathbf{X} \in \mathbb{R}^p$ is a predictor vector, $\mathbf{B} \in \mathbb{R}^{r_1 \times \cdots \times r_m \times p}$ is an $(m+1)$th order tensor and $\bar{\times}_{(m+1)}$ is the $(m+1)$-mode vector product. The error $\boldsymbol{\varepsilon} \in \mathbb{R}^{r_1 \times \cdots \times r_m}$ has mean zero and a separable Kronecker covariance structure, i.e. $\mathrm{cov}\{\mathrm{vec}(\boldsymbol{\varepsilon})\} = \mathbf{\Sigma}_m \otimes \cdots \otimes \mathbf{\Sigma}_1$, where $\mathbf{\Sigma}_k \in \mathbb{R}^{r_k \times r_k}$ is a positive definite matrix, for $k = 1, \ldots, m$. For more details on tensor notations and algebra, Kolda and Bader (2009) provided a complete review on tensor decomposition and applications.

Li and Zhang (2017) extended conditions (i) and (ii) in Section 2.1 to the tensor response $\mathbf{Y}$. Let $\mathcal{S}_k$ denote a subspace of $\mathbb{R}^{r_k}$, $k = 1, \ldots, m$, and $\times_k$ denote k-mode product. The tensor response $\mathbf{Y}$ is assumed to satisfy the conditions (i*) $\mathbf{Y} \times_k \mathbf{Q}_{\mathcal{S}_k} \mid \mathbf{X} \sim \mathbf{Y} \times_k \mathbf{Q}_{\mathcal{S}_k}$ and (ii*) $\mathbf{Y} \times_k \mathbf{Q}_{\mathcal{S}_k} \perp\!\!\!\perp \mathbf{Y} \times_k \mathbf{P}_{\mathcal{S}_k} \mid \mathbf{X}$, where $\perp\!\!\!\perp$ denotes independence. These two conditions imply that $\mathbf{Y} \times_k \mathbf{Q}_{\mathcal{S}_k}$ does not depend on $\mathbf{X}$ or is affected by $\mathbf{Y} \times_k \mathbf{P}_{\mathcal{S}_k}$. Under tensor response linear model (8), conditions (i*) and (ii*) are equivalent to conditions (a*) $\mathrm{span}(\mathbf{B}_{(k)}) \subseteq \mathcal{S}_k$, where $\mathbf{B}_{(k)}$ denotes mode-k matricization and (b*) $\boldsymbol{\Sigma}_k = \mathbf{P}_{\mathcal{S}_k} \boldsymbol{\Sigma}_k \mathbf{P}_{\mathcal{S}_k} + \mathbf{Q}_{\mathcal{S}_k} \boldsymbol{\Sigma}_k \mathbf{Q}_{\mathcal{S}_k}$. For a fixed k, the intersection of all subspaces that satisfy conditions (a*) and (b*) is the $\boldsymbol{\Sigma}_k$-envelope of $\mathbf{B}_{(k)}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\mathbf{B}_{(k)})$. The tensor envelope $\mathcal{T}_{\boldsymbol{\Sigma}}(\mathbf{B})$ is defined as $\mathcal{T}_{\boldsymbol{\Sigma}}(\mathbf{B}) = \mathcal{E}_{\boldsymbol{\Sigma}_m}(\mathbf{B}_{(m)}) \otimes \cdots \otimes \mathcal{E}_{\boldsymbol{\Sigma}_1}(\mathbf{B}_{(1)})$. Here $\mathcal{S}_1 \otimes \mathcal{S}_2$ means $\mathbf{P}_{\mathcal{S}_1 \otimes \mathcal{S}_2} = \mathbf{P}_{\mathcal{S}_1} \otimes \mathbf{P}_{\mathcal{S}_2}$ for any subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$. Let u_k denote the dimension of $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\mathbf{B}_{(k)})$, and $\boldsymbol{\Gamma}_k \in \mathbb{R}^{r_k \times u_k}$ denote an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\mathbf{B}_{(k)})$, for $k = 1, \ldots, m$. Then the coefficients $\mathbf{B}$ and the covariance matrices $\boldsymbol{\Sigma}_k$ in model (8) satisfies

$$\begin{aligned} \mathbf{B} &= \boldsymbol{\Theta} \times_1 \boldsymbol{\Gamma}_1 \times_2 \cdots \times_m \boldsymbol{\Gamma}_m \quad \text{for some } \boldsymbol{\Theta} \in \mathbb{R}^{u_1 \times \cdots \times u_m \times p} \qquad (9) \\ \boldsymbol{\Sigma}_k &= \boldsymbol{\Gamma}_k \boldsymbol{\Omega}_k \boldsymbol{\Gamma}_k^T + \boldsymbol{\Gamma}_{0k} \boldsymbol{\Omega}_{0k} \boldsymbol{\Gamma}_{0k}^T, \quad k = 1, \ldots, m, \end{aligned}$$

where the decomposition of $\mathbf{B}$ is called the Tucker decomposition, and $\boldsymbol{\Theta}$ is called the core tensor. Li and Zhang (2017) established the $\sqrt{n}$-consistency of the estimator of $\mathbf{B}$ under the envelope structure (9) and showed that the envelope estimator is at least as efficient as the standard estimator asymptotically.

Zhang and Li (2017) considered a tensor linear regression model where the response is a vector and the predictor is an mth order tensor. The envelope structure (9) is imposed on the coefficients of the predictor tensor and the corresponding covariance matrices. Based on the connection between the predictor envelope model and PLS (c.f. Section 3.1), this leads to a formulation of the tensor envelope PLS regression.

3.3 Sparse Envelope Model

The sparse envelope model (Su et al., 2016) is motivated when some response variables are invariant to the changes in $\mathbf{X}$ and have coefficients zero under multivariate linear regression (1). The sparse envelope model performs response variable selection and at the same time preserves the efficiency obtained from the envelope model (2). A response variable is called an inactive response if the corresponding row in $\boldsymbol{\Gamma}$ is zero, otherwise, it is called an active response. Without loss of generality, the response vector can be written as $\mathbf{Y} = (\mathbf{Y}_{\mathcal{A}}^T, \mathbf{Y}_{\mathcal{I}}^T)^T$, where $\mathbf{Y}_{\mathcal{A}} \in \mathbb{R}^{r_{\mathcal{A}}}$ contains all active responses, $\mathbf{Y}_{\mathcal{I}} \in \mathbb{R}^{r_{\mathcal{I}}}$ contains all inactive responses, $r_{\mathcal{A}}$ and $r_{\mathcal{I}}$ denote the number of active and inactive responses, and $r_{\mathcal{A}} + r_{\mathcal{I}} = r$. Then the sparse envelope model is formulated as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{\mathcal{A}} \\ 0 \end{pmatrix}, \tag{10}$$

where $\boldsymbol{\Gamma}_{\mathcal{A}} \in \mathbb{R}^{r_{\mathcal{A}} \times u}$ is a semi-orthogonal matrix, and $\boldsymbol{\beta}$ has the structure $\boldsymbol{\beta} = ((\boldsymbol{\Gamma}_{\mathcal{A}}\boldsymbol{\eta})^T, 0)^T$. So the inactive response $\mathbf{Y}_{\mathcal{I}}$ has regression coefficients zero and the active response $\mathbf{Y}_{\mathcal{A}}$ has regression coefficients $\boldsymbol{\Gamma}_{\mathcal{A}}\boldsymbol{\eta}$. To induce sparsity, Su et al. (2016) added a row-wise group-lasso penalty on the objective function (3)

$$\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \underset{\operatorname{span}(\mathbf{H}) \in \mathcal{G}(r,u)}{\arg\min} \log|\mathbf{H}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}|\mathbf{X}} \mathbf{H}| + \log|\mathbf{H}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}}^{-1} \mathbf{H}| + \sum_{i=1}^{r} \lambda_i \|\mathbf{h}_i\|, \tag{11}$$

where λ_i's are the tuning parameters and $\mathbf{h}_i$ is the ith row of $\mathbf{H}$. The group-lasso penalty depends on $\mathbf{H}$ only through $\mathrm{span}(\mathbf{H})$ and is invariant to an orthogonal transformation of $\mathbf{H}$. This penalty is often used in dimension reduction literature (Chen and Huang, 2012; Chen et al., 2010) to induce sparsity on certain coordinates of a subspace. Su et al. (2016) established the oracle property and asymptotic distribution of the sparse envelope estimator, as well as its convergence rate in high-dimensional settings. They also pointed out that after variable selection, the inactive response variables should be kept in the model to improve the efficiency in estimation. This is a major difference between response variable selection and predictor variable selection, where the inactive predictors are eliminated from the model once identified.

3.4 Bayesian Envelope Model

Bayesian analysis allows investigators to combine prior information and current data to make better decisions. The literature addressing envelope models from a Bayesian point of view is rather scarce. Since the envelope subspace is a parameter on a Grassmann manifold, to specify a prior for the envelope subspace that respects the manifold structure is the key issue in the development of a Bayesian approach. Khare et al. (2017) derived a Bayesian approach by considering an alternative parameterization of the envelope model (2). Instead of taking $\mathbf{\Gamma}$ to be an arbitrary orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$, $\mathbf{\Gamma}$ is chosen to be the orthonormal basis that makes $\mathbf{\Omega}$ a diagonal matrix with descending diagonal elements. In addition, it is also required that the maximum entry (in absolute value) in each column of $\mathbf{\Gamma}$ is positive. Then $\mathbf{\Gamma}$ becomes the unique representing basis for $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$. Note that $\mathbf{\Gamma}$ is defined on an $r \times u$ Stiefel manifold, where an $r \times u$ Stiefel manifold is the set of all $r \times u$ semi-orthogonal matrices. Similar requirements are imposed on $\mathbf{\Gamma}_0$. Let $\mathbf{O} = (\mathbf{\Gamma}, \mathbf{\Gamma}_0)$. The matrix Bingham distribution is used for the prior distribution of $\mathbf{O}$, i.e.

$$\pi(\mathbf{O}) \propto \exp\left[-(1/2)\,\mathrm{tr}(\mathbf{D}^{-1}\mathbf{O}^T\mathbf{G}\mathbf{O})\right],$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{G}$ is a positive definite or positive semi-definite matrix. The prior information about $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$ can be incorporated by carefully adjusting the hyperparameters $\mathbf{D}$ and $\mathbf{G}$. The prior distribution for $\boldsymbol{\eta}$ is assumed to be matrix normal. The diagonal elements of $\mathbf{\Omega}$ (or $\mathbf{\Omega}_0$) are a priori distributed as order statistics of u (or $r-u$) independent and identical inverse-Gamma random variates. A flat improper prior is assumed for $\boldsymbol{\mu}$. Khare et al. (2017) proved the propriety of the posterior distribution. A block Gibbs sampler is derived to sample from the posterior distribution, and the Harris ergodicity of the Gibbs sampler is established. The Deviance Information Criterion (DIC) (Gelman et al., 2013) is used to determine the dimension of the envelope subspace u.

This approach has the attractive feature of being able to incorporate the manifold structure of $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$. The techniques in this approach can also be applied to Bayesian sufficient dimension reduction or other areas where the parameter has a manifold structure. Unlike frequentist approaches that use bootstrapping or asymptotic distribution to approximate the standard deviation of the estimator $\widehat{\boldsymbol{\beta}}$ (see Section 2.4), Bayesian envelope directly provides estimation uncertainty by posterior credible interval. The Bayesian envelope model can also handle situations where the sampler size is smaller than the number of responses. However, Bayesian envelope model depends on the formulation of the envelope model (2) to choose hyper-parameters, such that conjugacy is obtained in all parameters. The application of this approach to more complex contexts such as the generalized linear models or quantile regression may require a different construction of the prior distributions.

3.5 Envelope Model Beyond Linear Regression

Cook and Zhang (2015a) introduced a model-free framework to incorporate the envelope structure to any consistent estimation procedure. A guideline for construction of an envelope estimator for an existing estimation procedure is as follows. Let $\boldsymbol{\theta} = (\boldsymbol{\psi}^T, \boldsymbol{\phi}^T)^T \in \mathbb{R}^m$ be the parameter vector, where $\boldsymbol{\phi} \in \mathbb{R}^p$ is the parameter vector of interest and $\boldsymbol{\psi} \in \mathbb{R}^{m-p}$ is the nuisance parameter vector. Let $\boldsymbol{\theta}_t = (\boldsymbol{\psi}_t^T, \boldsymbol{\phi}_t^T)^T$ denote the true value of $\boldsymbol{\theta}$, and $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\psi}}^T, \widehat{\boldsymbol{\phi}}^T)^T$ be an estimator of $\boldsymbol{\theta}$. Suppose that $\widehat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator, in other words, $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t)$ converges to a multivariate normal vector with mean 0 and positive definite covariance matrix $\mathbf{V}(\boldsymbol{\theta}_t) \in \mathbb{R}^{m \times m}$ in distribution as $n \to \infty$. Then the envelope subspace for the parameter $\boldsymbol{\phi}$ is taken to be $\mathcal{E}_{\mathbf{V}(\boldsymbol{\phi}_t)}(\boldsymbol{\phi}_t)$, where $\mathbf{V}(\boldsymbol{\phi}_t)$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\phi}}$, which corresponds to the $p \times p$ lower right block of $\mathbf{V}(\boldsymbol{\theta}_t)$. Let u denote the dimension of $\mathcal{E}_{\mathbf{V}(\boldsymbol{\phi}_t)}(\boldsymbol{\phi}_t)$, and $\boldsymbol{\Gamma}$ be any orthonormal basis for $\mathcal{E}_{\mathbf{V}(\boldsymbol{\phi}_t)}(\boldsymbol{\phi}_t)$. The envelope subspace can be estimated as

$$\widehat{\mathcal{E}}_{\mathbf{V}(\boldsymbol{\phi}_t)}(\boldsymbol{\phi}_t) = \underset{\mathrm{span}(\mathbf{H}) \in \mathcal{G}(m,u)}{\arg\min} \log|\mathbf{H}^T \widehat{\mathbf{M}} \mathbf{H}| + \log|\mathbf{H}^T (\widehat{\mathbf{M}} + \widehat{\mathbf{U}})^{-1} \mathbf{H}|, \tag{12}$$

where $\widehat{\mathbf{M}} = \widehat{\mathbf{V}}(\boldsymbol{\phi}_t)$ is a $\sqrt{n}$-consistent estimator of $\mathbf{V}(\boldsymbol{\phi}_t)$ and $\widehat{\mathbf{U}} = \widehat{\boldsymbol{\phi}}\widehat{\boldsymbol{\phi}}^T$. Once we obtain the estimated envelope subspace from solving (12), the envelope estimator of $\boldsymbol{\phi}$ is $\widehat{\boldsymbol{\phi}}_{\mathrm{env}} = \mathbf{P}_{\widehat{\mathcal{E}}} \widehat{\boldsymbol{\phi}}$, where $\mathbf{P}_{\widehat{\mathcal{E}}}$ is the projection matrix onto the estimated envelope subspace.

The preceding procedure gives a general algorithm to obtain an envelope model, without any parametric model. It can be shown that if $\mathbf{M} = \mathbf{V}(\boldsymbol{\phi}_t)$ and $\mathbf{U} = \boldsymbol{\phi}\boldsymbol{\phi}^T$, then $\mathcal{E}_{\mathbf{V}(\boldsymbol{\phi}_t)}(\boldsymbol{\phi}_t)$ is the minima of the objective function in (12). But if a parametric model is known, the objective function can also be derived based on the likelihood function. Cook and Zhang (2015a) showed that the envelope model can be derived for generalized linear model (GLM) with canonical link, weighted least squares and Cox regression model.

We describe the application of the envelope model to GLM as an example. Let Y be a random variable that belongs to an exponential family. For simplicity, we restrict attention to the natural exponential family, which only has the natural parameter. Let f denote the probability mass function or density function of Y: $f(y|\theta) = \exp\{y\theta - b(\theta) + c(y)\}$, where θ is the natural parameter, $b(\cdot)$ is the cumulant function and $c(\cdot)$ is some specific function. The predictor vector $\mathbf{X} \in \mathbb{R}^p$ is assumed to follow a multivariate normal distribution $N(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$. The canonical link function is $\theta(\mu, \boldsymbol{\beta}) = \mu + \boldsymbol{\beta}^T\mathbf{X}$, where $\theta(\mu, \boldsymbol{\beta})$ is a smooth and monotonic function of $E(Y|\mathbf{X}, \theta)$. The conditional log likelihood is $\log f(y|\theta) = y\theta - b(\theta) + c(y)$. The standard estimator of $\boldsymbol{\beta}$ can be obtained by Fisher scoring. To construct the envelope model under GLM, Cook and Zhang (2015a) considered the $\boldsymbol{\Sigma}_{\mathbf{X}}$-envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$. Let u denote its dimension, $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p\times(p-u)}$ be an orthonormal basis for $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$ and $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})^{\perp}$. Then under the envelope parameterization, GLM can be written as

$$\log f(y|\theta) = y\theta - b(\theta) + c(y), \quad \theta = \mu + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\mathbf{X}, \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{13}$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^u$ contains the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$, and $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u)\times(p-u)}$ are positive definite matrices that carry the coordinates of $\boldsymbol{\Sigma}_{\mathbf{X}}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$. Cook and Zhang (2015a) used the joint distribution of Y and $\mathbf{X}$ as objective function to estimate the model parameters and derive the envelope estimator. They showed that the envelope estimator is asymptotically at least as efficient as the standard GLM estimator.

Although the framework in Cook and Zhang (2015a) is generally applicable to any consistent estimation procedure, the objective function (12) is rooted from normal likelihood with constant variance for the errors. If the data has heteroscedastic error structure or the data distribution is very different from normal, it can take a large sample size for the envelope estimator to get close to its population parameter. Moreover, the dimension u can be underestimated, leading to a loss of material information and thus a biased envelope estimator. In such cases, it is better to develop an envelope estimator based on the specific parametric structure of the model, which is normally more efficient than the estimator produced from the general procedure. Besides the GLM example, other examples on using the specific parametric structure to develop an envelope estimator can be found in Cook et al. (2015a); Rekabdarkolaee et al. (2019); Forzani and Su (2019).

3.6 Envelope Quantile Regression

Ding et al. (2019) derived the envelope quantile regression that advances the envelope model to distribution-free settings and creates a different estimation and inference scheme for the envelope model. Quantile regression (Koenker, 2005) considers the relationship between the conditional quantile of the response and the predictors at different quantile levels. It does not impose distributional assumptions on error terms, thus it is able to incorporate heterogeneous errors and is robust to outliers. Let $Q_Y(\tau \mid \mathbf{X} = \mathbf{x}) = \inf\{y : F_Y(y \mid \mathbf{X} = \mathbf{x}) \geq \tau\}$ denote the τ-th conditional quantile of Y given $\mathbf{X} = \mathbf{x}$, for $0 \leq \tau \leq 1$. A linear quantile regression model is given by

$$Q_Y(\tau \mid \mathbf{X} = \mathbf{x}) = \mu_\tau + \boldsymbol{\beta}_\tau^T \mathbf{X}, \tag{14}$$

where $\mu_\tau \in \mathbb{R}$ is the intercept and $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ contains the regression coefficients. The predictor vector $\mathbf{X}$ is assumed to be random with mean $\boldsymbol{\mu}_\mathbf{X}$ and covariance matrix $\boldsymbol{\Sigma}_\mathbf{X}$.

The envelope quantile regression assumes that some linear combinations of the predictors does not affect the conditional quantile of the response. More specifically, let $\mathcal{S}_\tau$ be a subspace of $\mathbb{R}^p$, the envelope quantile regression assumes that (i) $Q_Y(\tau \mid \mathbf{X}) = Q_Y(\tau \mid \mathbf{P}_{\mathcal{S}_\tau}\mathbf{X})$ and (ii) $\mathrm{cov}(\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}, \mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}) = 0$. Assumptions (i) and (ii) indicate that $Q_Y(\tau \mid \mathbf{X})$ depends on $\mathbf{X}$ only through $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$, and $\mathbf{Q}_{\mathcal{S}_\tau}\mathbf{X}$ does not carry information on $Q_Y(\tau \mid \mathbf{X})$ through its correlation with $\mathbf{P}_{\mathcal{S}_\tau}\mathbf{X}$. The intersection of all such $\mathcal{S}_\tau$ is the $\mathbf{\Sigma_X}$-envelope of $\boldsymbol{\beta}_\tau$, denoted by $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. Let u_τ ($0 \leq u_\tau \leq p$) denote the dimension of $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$, and let $\mathbf{\Phi}_\tau \in \mathbb{R}^{p \times u_\tau}$ be an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. Then the envelope quantile regression is formulated as

$$Q_Y(\tau \mid \mathbf{X} = \mathbf{x}) = \mu_\tau + \boldsymbol{\eta}_\tau^T \mathbf{\Phi}_\tau^T \mathbf{X}, \qquad \mathbf{\Sigma_X} = \mathbf{\Phi}_\tau \mathbf{\Omega}_\tau \mathbf{\Phi}_\tau^T + \mathbf{\Phi}_{0\tau} \mathbf{\Omega}_{0\tau} \mathbf{\Phi}_{0\tau}^T, \tag{15}$$

where $\boldsymbol{\beta}_\tau = \mathbf{\Phi}_\tau \boldsymbol{\eta}_\tau$, and $\boldsymbol{\eta}_\tau \in \mathbb{R}^{u_\tau}$ carries the coordinates of $\boldsymbol{\beta}_\tau$ with respect to $\mathbf{\Phi}_\tau$. The matrix $\mathbf{\Phi}_{0\tau} \in \mathbb{R}^{p \times (p-u_\tau)}$ is an orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)^\perp$. Matrices $\mathbf{\Omega}_\tau \in \mathbb{R}^{u_\tau \times u_\tau}$ and $\mathbf{\Omega}_{0\tau} \in \mathbb{R}^{(p-u_\tau)\times(p-u_\tau)}$ are positive definite, and carry the coordinates of $\mathbf{\Sigma_X}$ with respect to $\mathbf{\Phi}_\tau$ and $\mathbf{\Phi}_{0\tau}$.

Since the envelope quantile regression does not impose any assumptions on the error distribution, the estimation of the parameters cannot be performed using a likelihood-based method like in Section 2.2. Its estimation procedure is different from any other envelope models in the literature. The robust cross validation (Oh et al., 2004, RCV) is applied to select u_τ, the dimension of the envelope $\widehat{\mathcal{E}}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta}_\tau)$. With fixed u_τ, Ding et al. (2019) proposed to use the generalized method of moments (GMM) for estimation. To use GMM, we first construct the estimating equations

$$\mathbf{g} = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (1, \mathbf{X}_i)^T [I(Y_i < \mu_\tau + \boldsymbol{\eta}_\tau^T \mathbf{\Phi}_\tau^T \mathbf{X}_i) - \tau] \\ \mathrm{vech}(\mathbf{\Phi}_\tau \mathbf{\Omega}_\tau \mathbf{\Phi}_\tau^T + \mathbf{\Phi}_{0\tau} \mathbf{\Omega}_{0\tau} \mathbf{\Phi}_{0\tau}^T) - \mathrm{vech}(\mathbf{S_X}) \\ \boldsymbol{\mu}_\mathbf{X} - \bar{\mathbf{X}} \end{pmatrix} = o_p(n^{-1/2}). \tag{16}$$

where $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$ is a random sample of $(\mathbf{X}, Y)$, $I(\cdot)$ is an indicator function, vech denotes the vector-half operator that stretches the lower triangle of a symmetric matrix into a vector, and $\mathbf{S}_{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}})^T$ is the sample covariance matrix of $\mathbf{X}$ given $\boldsymbol{\mu}_{\mathbf{X}}$. The estimating equation in g_1 is based on a property of the quantile regression, that the standard estimators of μ_τ and $\boldsymbol{\beta}_\tau$ from the quantile regression model (14) satisfy $n^{-1}\sum_{i=1}^{n}(1, \mathbf{X}_i)^T[I(Y_i < \mu_\tau + \boldsymbol{\beta}_\tau^T\mathbf{X}_i) - \tau] = o_p(n^{-1/2})$ (Koenker, 2005; Wang and Wang, 2009). The estimating equations in g_2 and g_3 are based on the first and second moment conditions of $\mathbf{X}$. Then the GMM estimators are obtained by solving

$$(\widehat{\mu}_\tau, \widehat{\boldsymbol{\eta}}_\tau, \widehat{\boldsymbol{\Omega}}_\tau, \widehat{\boldsymbol{\Omega}}_{0\tau}, \widehat{\mathcal{E}}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}_\tau)) = \underset{\boldsymbol{\mu}_\tau, \boldsymbol{\eta}_\tau, \boldsymbol{\Omega}_\tau, \boldsymbol{\Omega}_{0\tau}, \mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}_\tau)}{\arg\min} \mathbf{g}^T \widehat{\boldsymbol{\Delta}} \mathbf{g}, \tag{17}$$

where $\widehat{\boldsymbol{\Delta}}$ is a $\sqrt{n}$-consistent estimator of $[\mathrm{E}(\mathbf{g}\mathbf{g}^T)]^{-1}$, $\mathbf{g}$ is defined in (16), and the minimization is over the parameters in $\mathbf{g}$, i.e. $\boldsymbol{\mu}_\tau, \boldsymbol{\eta}_\tau, \boldsymbol{\Omega}_\tau, \boldsymbol{\Omega}_{0\tau}$, and $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}$. Note that unlike the other envelope models, the objective function in (17) is non-smooth. The Nelder-Mead method (Nelder and Mead, 1965) is applied to solve the optimization problem. Once we obtain $\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}_\tau)$, $\widehat{\boldsymbol{\Phi}}_\tau$ is taken to be an orthonormal basis of $\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta}_\tau)$. The envelope quantile regression estimator of $\boldsymbol{\beta}_\tau$ is $\widehat{\boldsymbol{\beta}}_\tau = \widehat{\boldsymbol{\Phi}}_\tau \hat{\boldsymbol{\eta}}_\tau$. Ding et al. (2019) showed that the envelope quantile regression estimator is at least as efficient as the standard quantile regression estimator asymptotically.

The estimation method and theoretical development in envelope quantile regression are completely different from the likelihood-based envelope models. The envelope quantile regression provides a new estimation procedure for envelope models with GMM. The study of asymptotic properties involves rather different techniques which handle both non-smooth objective function and over-parameterization. The envelope model theory is advanced to the distribution-free settings. Using similar techniques, the envelope model can be applied to other quantile regression settings, such as censored quantile regression (Koenker and Geling, 2001; Peng and Huang, 2008) and partially linear quantile regression (Lee, 2003; Sherwood and Wang, 2016), for survival and other complex data analysis. It can also be applied in expectile regression (Newey and Powell, 1987) to achieve efficient estimation or improve prediction performance.

3.7 Envelope Model in Other Contexts of Multivariate Analysis

In this section, we review developments of the envelope model in other contexts of multivariate analysis. Reduced-rank regression (Anderson, 1951) imposes a rank constraint on the regression coefficients in the multivariate linear regression (1), which reduces the number of parameters and improves the efficiency in estimation. Cook et al. (2015a) introduced the envelope structure to the reduced rank regression, which removes the immaterial variation in the response vector and further improves the efficiency gains. Cook and Zhang (2015b) derived the simultaneous envelope model which performs dimension reduction on both $\mathbf{X}$ and $\mathbf{Y}$ to achieve further efficiency gains than the response envelope model or predictor envelope model. Forzani and Su (2019) applied the envelope model to elliptical multivariate linear regression to improve estimation efficiency gains. This models allows for heteroscedastic errors without requiring any groupings of the data. Li et al. (2016) made a connection between supervised singular value decomposition and the envelope model. Zhang et al. (2018) introduced the envelope model for sufficient dimension reduction in functional data. Zhang and Mai (2019) constructed an envelope discriminant subspace to improve prediction performance in discriminant analysis and classification. Up to now, all envelope models require that observations are independent to each other. Under the multivariate spatial regression model, Rekabdarkolaee et al. (2019) derived a spatial envelope model that allows for dependent observations. Wang and Ding (2018) developed the envelope models for time series data in the context of vector autoregression model. The use of the envelope model in the aster model is an example of its application in biology. Darwinian fitness is the total number of offsprings of a plant or an animal, and the aster model (Geyer et al., 2007) is a statistical model derived to model the distributions of Darwinian fitness. The envelope model is applied by Eck et al. (2017) in the aster model for variance reduction in life history analyses.

3.8 Extensions in the Structure of Envelope Model

In this section, we review developments in the structure of the envelope model. These developments either make the envelope model adaptive to more flexible data structure or achieve more efficiency gains.

Under the linear regression model (1), Su and Cook (2011) introduced the partial envelope model that can achieve further efficiency gains than the envelope model in the estimation of the coefficients for predictors of special interest. Suppose that $\mathbf{X}$ can be partitioned into $\mathbf{X}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{p_2}$ $(p_1 + p_2 = p)$, where $\mathbf{X}_1$ contains the predictors of special interest and $\mathbf{X}_2$ contains other predictors. For example, in a clinical trial, the predictor of interest is the presence or absence of the drug under study, while demographical characteristics of the patients are also measured as covariates to reduce variability. We can partition the columns of $\boldsymbol{\beta}$ accordingly into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_1 \in \mathbb{R}^{r \times p_1}$ contains the coefficients of $\mathbf{X}_1$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{r \times p_2}$ contains the coefficients of $\mathbf{X}_2$. Su and Cook (2011) considered the envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1)$, and imposed the envelope structure only on $\boldsymbol{\beta}_1$. Since $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1) \subseteq \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, the partial envelope model is able to further reduce the dimension of the material part compared to the envelope model, making it more efficient.

The envelope model (2) requires that the errors have constant variance across the sample. However, if the data consist of measurements from several groups, the covariance matrices in the different groups can be different. Sometimes the regression coefficients can differ among the groups. The heteroscedastic envelope model (Su and Cook, 2013) was developed to incorporate heteroscedastic error structure in the context of estimating multivariate means for different groups. The groupwise envelope model (Park et al., 2017) was proposed to accommodate both distinct regression coefficients and distinct error structures for different groups in the linear regression model (1). It is motivated by the analysis on the associations between genetic variants and brain imaging phenotypes of Alzheimer patients. Measurements are taken on 749 Alzheimer patients on volumes of 93 brain regions and 1071 single nucleotide polymorphisms (SNPs) on 40 candidate genes. It is known that the brain structures of males and females are different. Numerical analysis shows that by considering the heteroscedasticity of the groups, the groupwise envelope model gives a less biased and more efficient estimation compared to the envelope model (2).

Like principal component analysis or partial least squares, the envelope model (2) is not scale invariant or equivariant. The scaled response envelope model (Cook and Su, 2013) includes the scales as model parameters, which gives a scale-invariant envelope estimator. Let $\mathbf{\Lambda} \in \mathbb{R}^{r\times r}$ be a diagonal scaling matrix with positive diagonal elements $1, \lambda_2, \ldots, \lambda_r$. The scaled envelope model assumes that the scaled response $\mathbf{\Lambda}^{-1}\mathbf{Y}$ and $\mathbf{X}$ follows the envelope model (2), i.e.

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{\Lambda\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad \mathbf{\Sigma} = \mathbf{\Lambda\Gamma\Omega\Gamma}^T\mathbf{\Lambda} + \mathbf{\Lambda\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T\mathbf{\Lambda}.$$

Cook and Su (2013) showed that the scaled envelope estimator is at least as efficient as the OLS estimator asymptotically, and it has the potential to provide efficiency gains when the envelope model (2) degenerates to the standard model.

Su and Cook (2012) proposed a different envelope construction than (2) to achieve efficient estimation. Instead of considering the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathrm{span}(\boldsymbol{\beta})$ (i.e. $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$), it considers the largest reducing subspace of $\boldsymbol{\Sigma}$ that is contained in $\mathrm{span}(\boldsymbol{\beta})$, called the inner $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ and denoted by $\mathcal{IE}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. Let $\mathbf{P}_{\mathcal{IE}}$ denote the projection matrix onto the inner envelope subspace, $\mathbf{Q}_{\mathcal{IE}} = \mathbf{I}_r - \mathbf{P}_{\mathcal{IE}}$, and $\dim\{\mathcal{IE}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})\} = d$. Then $\boldsymbol{\beta}$ can be decomposed as $\boldsymbol{\beta} = \mathbf{P}_{\mathcal{IE}}\boldsymbol{\beta} + \mathbf{Q}_{\mathcal{IE}}\boldsymbol{\beta}$. When $\mathrm{span}(\boldsymbol{\beta}) = \mathbb{R}^r$, $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\xi}_1 + \boldsymbol{\Gamma}_0\boldsymbol{\xi}_2$, where $\boldsymbol{\xi}_1 \in \mathbb{R}^{d\times p}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{(p-d)\times p}$. When $\mathrm{span}(\boldsymbol{\beta}) \subset \mathbb{R}^r$, to further reduce the number of parameters in the model, Su and Cook (2012) used the parameterization $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}_1 + \boldsymbol{\Gamma}_0\mathbf{B}\boldsymbol{\eta}_2$, where $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times d}$ and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r\times(r-d)}$ denote an orthonormal basis of $\mathcal{IE}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and $\mathcal{IE}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})^\perp$, $\mathbf{B} \in \mathbb{R}^{(r-d)\times(p-d)}$ is a semi-orthogonal matrix such that $\boldsymbol{\Gamma}_0\mathbf{B}$ denotes an orthonormal basis of $\mathrm{span}(\mathbf{Q}_{\mathcal{IE}}\boldsymbol{\beta})$, and $\boldsymbol{\eta}_1 \in \mathbb{R}^{d\times p}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{(p-d)\times p}$ carries the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0\mathbf{B}$. Then the inner envelope model can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + (\boldsymbol{\Gamma}\boldsymbol{\eta}_1 + \boldsymbol{\Gamma}_0\mathbf{B}\boldsymbol{\eta}_2)\mathbf{X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T,$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are positive definite matrices. We expect to achieve an efficient estimation on $\mathbf{P}_{\mathcal{IE}}\boldsymbol{\beta}$, especially when $\|\boldsymbol{\Omega}\| \leq \|\boldsymbol{\Omega}_0\|$. We expect to have about the same efficiency on $\mathbf{Q}_{\mathcal{IE}}\boldsymbol{\beta}$ as the standard analysis. Then overall, the inner envelope estimator of $\boldsymbol{\beta}$ is more efficient than the standard estimator. Su and Cook (2012) showed that the inner envelope estimator is at least as efficient as the standard estimator asymptotically. Furthermore, because the inner envelope model has a different mechanism for efficient estimation from the envelope model (2), it can offer efficiency gains in the cases when the envelope model cannot achieve any efficiency gains.

3.9 Software

Software to implement the envelope model includes R package `Renvlp` and Matlab toolbox `envlp` (Cook et al., 2015b). The optimization in the `envlp` toolbox depends on a Grassmann and Stiefel manifold optimization Matlab toolbox `sg_min` by Lippert. The `Renvlp` package uses the non-Grassmann estimation algorithm (Cook et al., 2016), which makes the computation faster. Both packages implement a variety of envelope models and several different inference tools. R package `TRES` implements envelope models in tensor regression. A complete list of software for envelope models can be found in `http://users.stat.umn.edu/~rdcook/envelopes/`.

4 Future Research

Almost all advances in envelope models are based on a linear model, while nonlinear regression has wide applications in face recognition, speech recognition, meteorology and longitudinal data analysis. Recently there are some interesting new developments in nonlinear sufficient dimension reduction (Lee et al., 2013; Li and Song, 2017). Some of the tools or techniques may be applicable to the derivation of a nonlinear envelope model. Another direction is on multivariate generalized linear model. Multivariate generalized linear model arises in genetics or medical sciences (Li and Wong, 2010; Schaid et al., 2019; Lu and Yang, 2012), where the response variables are discrete and correlated. The envelope model (2) requires a continuous response vector and cannot be adopted to such context. Incorporating the envelope structure in multivariate generalized linear model potentially has important applications in applied sciences. Up to now, the Bayesian approach is only derived for the first envelope model (2). While the Bayesian approach offers a different perspective from the frequentist approach, it would be beneficial to further explore Bayesian approaches for envelope models in other contexts.

Acknowledgements

We are grateful to the editor and two referees for their insightful suggestion and comments that helped us improve the paper. This work was supported in part by grant DMS-1407460 from the US National Science Foundation, grant 632688 from Simons Foundation and the Graduate School Fellowship (GSF) at the University of Florida.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22:327–351.

Chen, L. and Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *Journal of the American Statistical Association*, 107(500):1533–1545.

Chen, X., Zou, C., and Cook, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *The Annals of Statistics*, 38(6):3696–3723.

Chun, H. and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25.

Chung, D. and Keleş, S. (2010). Sparse partial least squares classification for high dimensional data. *Statistical Applications in Genetics and Molecular Biology*, 9:1–29.

Conway, J. (1990). *A Course in Functional Analysis*. Graduate Texts in Mathematics. New York: Springer.

Cook, R. D. (2018). *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics*. Hoboken, NJ: John Wiley & Sons.

Cook, R. D. and Forzani, L. (2018). Big data and partial least-squares prediction. *Canadian Journal of Statistics*, 46(1):62–78.

Cook, R. D. and Forzani, L. (2019). Partial least squares prediction in high-dimensional regression. *The Annals of Statistics*, 47(2):884–908.

Cook, R. D., Forzani, L., and Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis*, 150:42–54.

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

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

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

Cook, R. D. and Su, Z. (2013). Scaled envelopes: scale invariant and efficient estimation in multivariate linear regression. *Biometrika*, 100(4):939–954.

Cook, R. D. and Su, Z. (2016). Scaled predictor envelopes and partial least-squares regression. *Technometrics*, 58(2):155–165.

Cook, R. D., Su, Z., and Yang, Y. (2015b). envlp: A matlab toolbox for computing envelope estimators in multivariate analysis. *Journal of Statistical Software*, 62(8):1–20.

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

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

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

Cook, R. D. and Zhang, X. (2018). Fast envelope algorithms. *Statistica Sinica*, 28:1179–1197.

De Jong, S. (1993). Simpls: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263.

Ding, S. and Cook, R. D. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B*, 80(2):387–408.

Ding, S., Su, Z., Zhu, G., and Wang, L. (2019). Envelope quantile regression. *Statistica Sinica*, To appear.

Eck, D. J. and Cook, R. D. (2017). Weighted envelope estimation to handle variability in model selection. *Biometrika*, 104(3):743–749.

Eck, D. J., Geyer, C. J., and Cook, R. D. (2017). An application of envelope methodology and aster models. *arXiv preprint arXiv:1701.07910*.

Forzani, L. and Su, Z. (2019). Envelope for elliptical multivariate linear regression. *Statistica Sinica*, To appear.

Fosdick, B. K. and Hoff, P. D. (2014). Separable factor analysis with applications to mortality data. *The Annals of Applied Statistics*, 8(1):120–147.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. New York: Chapman and Hall/CRC.

Geyer, C. J., Wagenius, S., and Shaw, R. G. (2007). Aster models for life history analysis. *Biometrika*, 94(2):415–426.

Hoff, P. D. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Huang, X., Pan, W., Park, S., Han, X., Miller, L. W., and Hall, J. (2004). Modeling the relationship between lvad support time and gene expression changes in the human heart by penalized partial least squares. *Bioinformatics*, 20(6):888–894.

Khare, K., Pal, S., and Su, Z. (2017). A bayesian approach for envelope models. *The Annals of Statistics*, 45(1):196–222.

Koenker, R. (2005). *Quantile regression*. Cambridge (England): Cambridge University Press.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, 96(454):458–468.

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

Lê Cao, K.-A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A sparse pls for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology*, 7(1):1–32.

Lee, D., Lee, W., Lee, Y., and Pawitan, Y. (2011). Sparse partial least-squares regression and its applications to high-throughput data analysis. *Chemometrics and Intelligent Laboratory Systems*, 109(1):1–8.

Lee, K.-Y., Li, B., and Chiaromonte, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *The Annals of Statistics*, 41(1):221–249.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19(1):1–31.

Li, B. and Song, J. (2017). Nonlinear sufficient dimension reduction for functional data. *The Annals of Statistics*, 45(3):1059–1095.

Li, G., Yang, D., Nobel, A. B., and Shen, H. (2016). Supervised singular value decomposition and its asymptotic properties. *Journal of Multivariate Analysis*, 146:7–17.

Li, J. and Wong, W. K. (2010). Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics*, 12(1):143–155.

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

Lu, M. and Yang, W. (2012). Multivariate logistic regression analysis of complex survey data with application to brfss data. *Journal of Data Science*, 10(2):157–173.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–47.

Oh, H.-S., Nychka, D., Brown, T., and Charbonneau, P. (2004). Period analysis of variable stars by robust smoothing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(1):15–30.

Park, Y., Su, Z., and Zhu, H. (2017). Groupwise envelope models for imaging genetic analysis. *Biometrics*, 73(4):1243–1253.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.

Rekabdarkolaee, H. M., Wang, Q., Naji, Z., and Fuentes, M. (2019). New parsimonious multivariate spatial model: Spatial envelope. *Statistica Sinica*, To appear.

Schaid, D. J., Tong, X., Batzler, A., Sinnwell, J. P., Qing, J., and Biernacka, J. M. (2019). Multivariate generalized linear model for genetic pleiotropy. *Biostatistics*, 20(1):111–128.

Schwartz, W. R., Guo, H., and Davis, L. S. (2010). A robust and scalable approach to face identification. In *European Conference on Computer Vision*, pages 476–489. Springer.

Sherwood, B. and Wang, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics*, 44(1):288–317.

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

Su, Z. and Cook, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika*, 99(3):687–702.

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

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

Tabei, Y., Saigo, H., Yamanishi, Y., and Puglisi, S. J. (2016). Scalable partial least squares regression on grammar-compressed data matrices. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1875–1884, New York, NY, USA. ACM.

Tuddenham, R. D. and Snyder, M. M. (1953). Physical growth of california boys and girls from birth to eighteen years. *Publications in child development. University of California, Berkeley*, 1(2):183–364.

Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, 104(487):1117–1128.

Wang, L. and Ding, S. (2018). Vector autoregression and envelope model. *Stat*, 7(1).

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *In Multivariate Analysis*, 59:391–420.

Wold, H. (1975). Path models with latent variables: The nipals approach. In *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, pages 307–357. New York: Academic Press.

Zeng, X.-Q. and Li, G.-Z. (2014). Incremental partial least squares analysis of big streaming data. *Pattern Recogn.*, 47(11):3726–3735.

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

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

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

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

Zhu, G. and Su, Z. (2019). Envelope-based sparse partial least squares. *The Annals of Statistics*, To appear.