

Renvlp: An R Package for Efficient Estimation in Multivariate Analysis Using Envelope Models

Minji Lee
University of Florida

Zhihua Su
University of Florida

Abstract

The envelope models can achieve substantial efficiency gains in multivariate analysis. The only software that implements envelope models so far is the `matlab` toolbox `envlp` (Cook, Su, and Yang 2015b). Estimation in `envlp` requires optimization over a Grassmann manifold, which is slow in sizable problems. This article introduces the R package `Renvlp` which uses a non-Grassmann algorithm for envelope estimation. This algorithm is much faster and more accurate than the estimation algorithms that involve manifold optimization. Besides the envelope models in the `matlab` toolbox `envlp`, `Renvlp` also implements the latest developments in envelope methodology. The models supported by `Renvlp` include response envelope model, partial envelopes model, predictor envelope model, simultaneous envelope model, heteroscedastic envelope model, groupwise envelope model, scaled response envelope model, scaled predictor envelope, envelope model in logistic regression, envelope model in poisson regression, weighted response envelope estimation, weighted partial envelope estimation, and weighted predictor envelope estimation. For each model, `Renvlp` provides model fitting and inference functions for bootstrapping, cross validation, prediction and hypothesis testing. Examples are provided for illustration.

Keywords: dimension reduction, envelope model, multivariate linear regression, generalized linear models, R.

1. Introduction

The envelope model proposed by Cook, Li, and Chiaromonte (2010) has the potential to gain efficiency in estimation and improve prediction in multivariate analysis. Efficiency gains are achieved by identifying the immaterial information and accounting for it in subsequent analysis. In particular, massive gains can be obtained when the immaterial information is substantially more variable than the material information. After the development in Cook *et al.* (2010), many advances have been taken place in this area (Su and Cook 2011, 2013; Cook, Helland, and Su 2013; Cook and Su 2013; Cook and Zhang 2015a,b; Cook, Forzani, and Zhang 2015a; Li, Yang, Nobel, and Shen 2016; Su, Zhu, Chen, and Yang 2016; Khare, Pal, and Su 2017; Li and Zhang 2017; Zhang and Li 2017; Park, Su, and Zhu 2017).

The only software that implements the envelope models is the `envlp` toolbox developed in `matlab` (Cook *et al.* 2015b). `envlp` is well-suited for moderate problem, but it can be quite slow in sizable problems. This is because the estimation algorithm in `envlp` involves optimization over Grassmann manifolds, which is challenging in high dimensional problems.

The goal of this article is to introduce the R (R Core Team 2017) package `Renvlp` that imple-

ments thirteen different envelope models including some latest developments in the area, using an algorithm (Cook, Forzani, and Su 2016) which avoids manifold optimization in envelope estimation. Compared with `matlab` toolbox **envlp**, we make the following several improvements. First, **Renvlp** is more comprehensive and contains many recently developed envelope models. Not only it implements the response envelope model (Cook *et al.* 2010), partial envelope model (Su and Cook 2011), predictor envelope model (Cook *et al.* 2013), heteroscedastic envelope model (Su and Cook 2013), and scaled response envelope model (Cook and Su 2013) which are in **envlp**, but also it incorporates the latest developments including the scaled predictor envelope model (Cook and Su 2016), simultaneous envelope model (Cook and Zhang 2015b), envelope model in logistic regression, envelope model in poisson regression (Cook and Zhang 2015a), groupwise envelope model (Park *et al.* 2017), and weighted envelope model (Eck and Cook 2017). Second, **Renvlp** adopts the non-Grassmann algorithm (Cook *et al.* 2016) for envelope estimation. As a result, **Renvlp** is much faster than **envlp** and is capable to handle larger datasets. Furthermore, the non-Grassmann manifold algorithm (Cook *et al.* 2016) is developed only for one type of objective function, in order to make this algorithm available in broader context, we extend the applicability of the algorithm such that it can be used for general envelope estimation. At last, the R implementation of envelope estimation is more accessible to the statistical community than the `matlab` implementation.

The remainder of this paper is organized as follows. We provide a brief overview of envelope models supported by **Renvlp** in Section 2. In Section 3, we introduce the estimation algorithm and investigate its performance with numerical experiments. Section 4 contains a detailed description of the structure of **Renvlp**. Section 5 illustrates the use of the package with an example. Section 6 concludes with a short discussion.

The following notations will be used in our discussion: We use $\mathbf{P}_{(\cdot)}$ to denote a projection operator onto the subspace indicated by its arguments and $\mathbf{Q}_{(\cdot)} = \mathbf{I} - \mathbf{P}_{(\cdot)}$. If \mathbf{A} is a matrix, $\text{span}(\mathbf{A})$ indicates the subspace spanned by the column vectors of \mathbf{A} . The symbol $\|\cdot\|$ denotes the spectral norm of a matrix, and $\mathbf{A} \sim \mathbf{B}$ means that \mathbf{A} and \mathbf{B} have the same distribution.

2. Envelope Models

2.1. Response envelope model

The response envelope model aims to reduce the standard error of the estimated regression coefficients in the context of multivariate linear regression. It is first introduced in Cook *et al.* (2010) under the model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon} \quad (1)$$

where $\mathbf{Y} \in \mathbb{R}^r$ is the response vector, $\mathbf{X} \in \mathbb{R}^p$ is the non-stochastic predictor vector, and the error vector $\boldsymbol{\epsilon}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma}$. We assume that $\boldsymbol{\Sigma} > 0$. The intercept $\boldsymbol{\mu} \in \mathbb{R}^r$ and the regression coefficient matrix $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ are unknown parameters.

The response envelope model achieves the efficiency gains by identifying part of the response variables that are invariant to the changes in \mathbf{X} . More specifically, let \mathcal{S} denote a subspace of \mathbb{R}^r . We assume that \mathbf{Y} can be decomposed into a material part $\mathbf{P}_{\mathcal{S}}\mathbf{Y}$ and an immaterial part $\mathbf{Q}_{\mathcal{S}}\mathbf{Y}$ such that (i) $\mathbf{Q}_{\mathcal{S}}\mathbf{Y}|\mathbf{X} \sim \mathbf{Q}_{\mathcal{S}}\mathbf{Y}$ and (ii) $\text{COV}(\mathbf{Q}_{\mathcal{S}}\mathbf{Y}, \mathbf{P}_{\mathcal{S}}\mathbf{Y}|\mathbf{X}) = 0$. Conditions (i) and (ii) indicate that $\mathbf{Q}_{\mathcal{S}}\mathbf{Y}$ does not carry any information about $\boldsymbol{\beta}$ directly or indirectly

through its conditional covariance with $\mathbf{P}_S \mathbf{Y}$. Under model (1), (i) and (ii) are equivalent to (i)' $\text{span}(\boldsymbol{\beta}) \subseteq \mathcal{S}$ and (ii)' $\boldsymbol{\Sigma} = \mathbf{P}_S \boldsymbol{\Sigma} \mathbf{P}_S + \mathbf{Q}_S \boldsymbol{\Sigma} \mathbf{Q}_S$, where $\mathbf{P}_S \boldsymbol{\Sigma} \mathbf{P}_S = \text{VAR}(\mathbf{P}_S \mathbf{Y})$ represents the variation of the material part $\mathbf{P}_S \mathbf{Y}$ and $\mathbf{Q}_S \boldsymbol{\Sigma} \mathbf{Q}_S = \text{VAR}(\mathbf{Q}_S \mathbf{Y})$ represents the variation of the immaterial part $\mathbf{Q}_S \mathbf{Y}$. The $\boldsymbol{\Sigma}$ -envelope of $\boldsymbol{\beta}$ is defined as the intersection of all $\mathcal{S} \subseteq \mathbb{R}^r$ that satisfies (i) and (ii), and is denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, or \mathcal{E} for short. The dimension of the envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is denoted by u , and $0 \leq u \leq r$.

Let $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ be an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, and let $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ be a completion of $\boldsymbol{\Gamma}$ such that $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0) \in \mathbb{R}^{r \times r}$ is an orthogonal matrix. Under the envelope parameterization, (1) can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{\eta} \mathbf{X} + \boldsymbol{\epsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^\top, \quad (2)$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma} \boldsymbol{\eta}$. The matrix $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ carries the coordinates of $\boldsymbol{\beta}$ with respect to the basis $\boldsymbol{\Gamma}$, and $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ carry the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$. We call model (2) the response envelope model, and (1) the standard model. The number of parameters under the standard model is $pr + r(r+1)/2$, and the number of parameters under the envelope model is $pu + r(r+1)/2$. When $u = r$, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \mathbb{R}^r$ and the envelope model degenerates to the standard model.

The estimation of parameters will be discussed in more details in Section 3. Cook *et al.* (2010) has shown that the response envelope estimator of $\boldsymbol{\beta}$ is always as efficient as or more efficient than the standard estimator of $\boldsymbol{\beta}$ asymptotically. Substantial efficiency gains can be achieved when the variation of the immaterial part is large relative to the variation of the material part, i.e. $\|\boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^\top\| \gg \|\boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^\top\|$.

2.2. Partial envelope model

The partial envelope model (Su and Cook 2011) focuses on the efficient estimation of the coefficients of main interest in the multivariate linear regression (1). For example, in a clinical study, the predictor of interest is the presence or absent of the drug under study, while demographic characteristics of the patients are also measured as covariates to reduce variability. We then partition \mathbf{X} 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 main interest and \mathbf{X}_2 contains other predictors. We can partition the columns of $\boldsymbol{\beta}$ accordingly into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Then (1) can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}_1 \mathbf{X}_1 + \boldsymbol{\beta}_2 \mathbf{X}_2 + \boldsymbol{\epsilon}, \quad (3)$$

where $\boldsymbol{\beta}_1 \in \mathbb{R}^{r \times p_1}$ contains the coefficients of main interest, and $\boldsymbol{\beta}_2 \in \mathbb{R}^{r \times p_2}$ contains the coefficients for \mathbf{X}_2 .

Now instead of imposing an envelope structure to $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ as in (2), we consider imposing the envelope structure on $\boldsymbol{\beta}_1$ and leaving $\boldsymbol{\beta}_2$ unconstrained. More specifically, we consider $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1)$, the $\boldsymbol{\Sigma}$ -envelope of $\boldsymbol{\beta}_1$, and call it the partial envelope of $\boldsymbol{\beta}_1$. If $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1)$ appears in subscripts, we denote it by \mathcal{E}_1 . Let u_1 be the dimension of the partial envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1)$. Since $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}_1) \subseteq \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, we have $u_1 \leq u$. The response \mathbf{Y} can be decomposed into the immaterial part $\mathbf{Q}_{\mathcal{E}_1} \mathbf{Y}$ and the material part $\mathbf{P}_{\mathcal{E}_1} \mathbf{Y}$, depending on whether they carry information on $\boldsymbol{\beta}_1$. The material and immaterial parts satisfy (i) $\mathbf{Q}_{\mathcal{E}_1} \mathbf{Y} \mid (\mathbf{X}_1, \mathbf{X}_2) \sim \mathbf{Q}_{\mathcal{E}_1} \mathbf{Y} \mid \mathbf{X}_2$ and (ii) $\text{COV}(\mathbf{Q}_{\mathcal{E}_1} \mathbf{Y}, \mathbf{P}_{\mathcal{E}_1} \mathbf{Y} \mid \mathbf{X}) = 0$. Together, (i) and (ii) indicate that the immaterial part $\mathbf{Q}_{\mathcal{E}_1} \mathbf{Y}$ carries no information on $\boldsymbol{\beta}_1$ directly or indirectly. Under the context of (1), conditions (i) and

(ii) are equivalent to $\text{span}(\boldsymbol{\beta}_1) \subseteq \mathcal{E}_\Sigma(\boldsymbol{\beta}_1)$ and $\boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}_1} \boldsymbol{\Sigma} \mathbf{P}_{\mathcal{E}_1} + \mathbf{Q}_{\mathcal{E}_1} \boldsymbol{\Sigma} \mathbf{Q}_{\mathcal{E}_1}$. Since the immaterial variation with respect to $\boldsymbol{\beta}_1$ is larger than or at least equal to the immaterial information with respect to $\boldsymbol{\beta}$: $\text{VAR}(\mathbf{Q}_{\mathcal{E}_1} \mathbf{Y}) = \mathbf{Q}_{\mathcal{E}_1} \boldsymbol{\Sigma} \mathbf{Q}_{\mathcal{E}_1} \geq \mathbf{Q}_{\mathcal{E}} \boldsymbol{\Sigma} \mathbf{Q}_{\mathcal{E}} = \text{VAR}(\mathbf{Q}_{\mathcal{E}} \mathbf{Y})$, we usually achieve more efficiency gains in estimating $\boldsymbol{\beta}_1$ using partial envelope than using response envelope.

Let $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u_1}$ be an orthonormal basis of $\mathcal{E}_\Sigma(\boldsymbol{\beta}_1)$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u_1)}$ be its completion. The coordinate version of the partial envelope model is

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{\eta} \mathbf{X}_1 + \boldsymbol{\beta}_2 \mathbf{X}_2 + \boldsymbol{\epsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^\top, \quad (4)$$

where $\boldsymbol{\beta}_1 = \boldsymbol{\Gamma} \boldsymbol{\eta}$. The matrix $\boldsymbol{\eta} \in \mathbb{R}^{u_1 \times p_1}$ carries the coordinates of $\boldsymbol{\beta}_1$ with respect to $\boldsymbol{\Gamma}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u_1 \times u_1}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u_1) \times (r-u_1)}$ are positive definite, and they carry the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$. We call (4) the partial envelope model. When $u_1 = 0$, we have $\boldsymbol{\beta}_1 = 0$ and the partial envelope model reduces to the multivariate linear regression with \mathbf{Y} being the responses and \mathbf{X}_2 being the predictors. When $u_1 = r$, the partial envelope model degenerates to the standard model (1) with \mathbf{Y} being the responses and both \mathbf{X}_1 and \mathbf{X}_2 being the predictors.

2.3. Heteroscedastic envelope model

The heteroscedastic envelope model (Su and Cook 2013) is derived to incorporate heteroscedastic error structure in the context of estimating multivariate means for different groups.

The standard model for multivariate mean estimation for p groups can be written as

$$\mathbf{Y}_{(i)j} = \boldsymbol{\mu} + \boldsymbol{\beta}_{(i)} + \boldsymbol{\epsilon}_{(i)j}, \quad i = 1, \dots, p, \quad j = 1, \dots, n_{(i)}, \quad (5)$$

where the subscript (i) indicates the i th group, $\mathbf{Y}_{(i)j} \in \mathbb{R}^r$ is the j th observation vector in the i th group, $\boldsymbol{\mu} \in \mathbb{R}^r$ is the grand mean over all the observations, $\boldsymbol{\beta}_{(i)} \in \mathbb{R}^r$ is the difference between the mean of the i th group and the grand mean, and the error vector $\boldsymbol{\epsilon}_{(i)j}$ has mean 0 and covariance $\boldsymbol{\Sigma}_{(i)}$. The i th group has $n_{(i)}$ observations and we assume that $\sum_{i=1}^p n_{(i)} \boldsymbol{\beta}_{(i)} = 0$. If $\boldsymbol{\Sigma}_{(1)} = \dots = \boldsymbol{\Sigma}_{(p)}$, (5) reduces to (1) with \mathbf{X} being p group indicators.

Let $\mathbf{B} = (\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(p)}) \in \mathbb{R}^{r \times p}$ and \mathcal{M} be a collection of covariance matrices, i.e. $\mathcal{M} = \{\boldsymbol{\Sigma}_{(i)} : i = 1, \dots, p\}$. We define the heteroscedastic envelope, denoted by $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$ or \mathcal{E} , to be the smallest subspace that contains each $\boldsymbol{\beta}_{(i)}$ and decomposes every matrix in \mathcal{M} , i.e., $\boldsymbol{\beta}_{(i)} \in \mathcal{E}_{\mathcal{M}}(\mathbf{B})$ and $\boldsymbol{\Sigma}_{(i)} = \mathbf{P}_{\mathcal{E}} \boldsymbol{\Sigma}_{(i)} \mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}} \boldsymbol{\Sigma}_{(i)} \mathbf{Q}_{\mathcal{E}}$ for $i = 1, \dots, p$. Let $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ be an orthonormal basis of $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$, where u is the dimension of $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$, and let $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ be its completion. Under this parametrization, (5) can be written as

$$\mathbf{Y}_{(i)j} = \boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{\eta}_{(i)} + \boldsymbol{\epsilon}_{(i)j}, \quad \boldsymbol{\Sigma}_{(i)} = \boldsymbol{\Gamma} \boldsymbol{\Omega}_{1(i)} \boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^\top, \quad (6)$$

where $\boldsymbol{\beta}_{(i)} = \boldsymbol{\Gamma} \boldsymbol{\eta}_{(i)}$, $\boldsymbol{\eta}_{(i)} \in \mathbb{R}^u$ contains the coordinates of $\boldsymbol{\beta}_{(i)}$ with respect to the basis $\boldsymbol{\Gamma}$, $\boldsymbol{\Omega}_{1(i)} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are both positive definite, and $\sum_{i=1}^p n_{(i)} \boldsymbol{\eta}_{(i)} = 0$. We call (6) the heteroscedastic envelope model. When $u = r$, the heteroscedastic envelope model degenerates to model (5).

2.4. Groupwise envelope model

The groupwise envelope model (Park *et al.* 2017) is designed to accommodate both distinct regression coefficients and distinct error structures for different groups in the context of (1).

Suppose that each group has different regression coefficients and error covariance matrix, the multivariate linear regression model can be formulated as

$$\mathbf{Y}_{(i)j} = \boldsymbol{\mu}_{(i)} + \boldsymbol{\beta}_{(i)}\mathbf{X}_{(i)j} + \boldsymbol{\epsilon}_{(i)j}, \quad i = 1, \dots, I, \quad j = 1, \dots, n_{(i)}, \quad (7)$$

where subscripts (i) indicate the i th group, $\mathbf{Y}_{(i)j} \in \mathbb{R}^r$ is the j th response vector in the i th group, $\boldsymbol{\mu}_{(i)} \in \mathbb{R}^r$ is the intercept of the i th group, $\mathbf{X}_{(i)j} \in \mathbb{R}^p$ is the j th covariate vector in the i th group, $\boldsymbol{\beta}_{(i)} \in \mathbb{R}^{r \times p}$ is the regression coefficient matrix for the i th group, and the error vector $\boldsymbol{\epsilon}_{(i)j}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma}_{(i)}$. The sample size of the i th group is denoted as $n_{(i)}$ and the total sample size is $n = \sum_{i=1}^I n_{(i)}$. While (5) aims at estimating the mean for different groups and does not involve any predictors, (7) studies the relationship between the response vector and the predictors in different groups.

Let $\mathbf{B} = (\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(I)}) \in \mathbb{R}^{r \times pI}$ and \mathcal{M} be a collection of covariance matrices, i.e. $\mathcal{M} = \{\boldsymbol{\Sigma}_{(i)} : i = 1, \dots, p\}$. The groupwise envelope, denoted by $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$ or \mathcal{E} , is the smallest subspace that contains each $\text{span}(\boldsymbol{\beta}_{(i)})$ and decomposes every matrix in \mathcal{M} , i.e., $\text{span}(\boldsymbol{\beta}_{(i)}) \subseteq \mathcal{E}_{\mathcal{M}}(\mathbf{B})$ and $\boldsymbol{\Sigma}_{(i)} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Sigma}_{(i)}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Sigma}_{(i)}\mathbf{Q}_{\mathcal{E}}$, for $i = 1, \dots, I$. The dimension of $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$ is u ($0 \leq u \leq r$). Take $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ to be an orthonormal basis of $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$, and choose $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ such that $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ is an orthogonal matrix. Then (7) can be written as

$$\mathbf{Y}_{(i)j} = \boldsymbol{\mu}_{(i)} + \boldsymbol{\Gamma}\boldsymbol{\eta}_{(i)}\mathbf{X}_{(i)j} + \boldsymbol{\epsilon}_{(i)j}, \quad \boldsymbol{\Sigma}_{(i)} = \boldsymbol{\Gamma}\boldsymbol{\Omega}_{(i)}\boldsymbol{\Gamma}^{\top} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\top}, \quad (8)$$

where $\boldsymbol{\beta}_{(i)} = \boldsymbol{\Gamma}\boldsymbol{\eta}_{(i)}$, $\boldsymbol{\eta}_{(i)} \in \mathbb{R}^{u \times p}$ carries the coordinates of $\boldsymbol{\beta}_{(i)}$ with respect to $\boldsymbol{\Gamma}$, and both $\boldsymbol{\Omega}_{(i)} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are positive definite. We call (8) the groupwise envelope model. When $I = 1$, the groupwise envelope model reduces to the response envelope model (2). By accounting for the common characteristics in the response vector across groups and connecting the material information from all groups via $\mathcal{E}_{\mathcal{M}}(\mathbf{B})$, the groupwise envelope model achieves more efficiency gains than fitting a separate response envelope model for each group.

2.5. Scaled response envelope model

The scaled response envelope model (Cook and Su 2013) is a scale-invariant version of the response envelope model. Similar to principal component analysis or partial least squares, the response envelope model is not scale invariant. The scaled response envelope model considers the scaling in the model building and is invariant under rescaling of the responses. Moreover, it has the potential to provide efficiency gains when the response envelope model degenerates to the standard model.

Let $1, \lambda_2, \dots, \lambda_r$ be the scaling parameters with $\lambda_i > 0$ for $i = 2, \dots, r$, and let $\boldsymbol{\Lambda}$ be a diagonal matrix $\boldsymbol{\Lambda} = \text{diag}\{1, \lambda_2, \dots, \lambda_r\} \in \mathbb{R}^{r \times r}$. The first scaling parameter is set to 1 for identifiability. Given $\boldsymbol{\Lambda}$, we assume that the scaled response $\mathbf{Y}_N = \boldsymbol{\Lambda}^{-1}\mathbf{Y}$ and the predictors \mathbf{X} follows the response envelope model. More specifically, the scaled regression coefficients $\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta}$ and error covariance matrix $\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}$ follow the following two conditions: (a) $\text{span}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta}) \subseteq \mathcal{E}_{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta})$, and (b) $\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}\mathbf{Q}_{\mathcal{E}}$, where \mathcal{E} is short for $\mathcal{E}_{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta})$ when it appears in subscripts. These two conditions are exactly the same as conditions (i) and (ii) in response envelope (Section 2.1) after scaling. Let u be the dimension of $\mathcal{E}_{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta})$, $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ be an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta})$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (r-u)}$ be a completion of $\boldsymbol{\Gamma}$. Under this parameterization, (1) can be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\top}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\top}\boldsymbol{\Lambda}, \quad (9)$$

where $\boldsymbol{\beta} = \mathbf{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$, both $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are positive definite. We call (9) the scaled response envelope model. When $u \geq r - (r - 1)/p$, the scaled response envelope model degenerates to the standard model (1). Note that when $u \neq r$, the scaled envelope model is not nested within the standard model. Therefore the likelihood ratio testing (LRT) is not suitable for selecting the dimension u .

2.6. Predictor envelope model

The predictor envelope model (Cook *et al.* 2013) is derived for dimension reduction of the predictors in the context of (1). It offers efficiency gains in the estimation of the regression coefficients, and also has a better prediction performance than the standard model. Furthermore, it has a close connection to partial least squares. To describe the predictor envelope model, we reformulate the linear regression model (1) to be consistent with the convention used in (Cook *et al.* 2013)

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}^\top (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \boldsymbol{\epsilon}, \quad (10)$$

where $\mathbf{Y} \in \mathbb{R}^r$ is the response vector, and $\mathbf{X} \in \mathbb{R}^p$ is the stochastic predictor vector with mean $\boldsymbol{\mu}_{\mathbf{X}}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{X}}$. The error vector $\boldsymbol{\epsilon}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}$. The regression coefficients are contained in $\boldsymbol{\beta} \in \mathbb{R}^{p \times r}$, and $\boldsymbol{\mu} \in \mathbb{R}^r$ is the intercept. Here \mathbf{Y} can be univariate response or multivariate response vector.

Let \mathcal{S} denote a subspace of \mathbb{R}^p . We decompose \mathbf{X} into a material part $\mathbf{P}_{\mathcal{S}}\mathbf{X}$ and an immaterial part $\mathbf{Q}_{\mathcal{S}}\mathbf{X}$ such that the following two conditions are satisfied (i) $\text{COV}(\mathbf{Y}, \mathbf{Q}_{\mathcal{S}}\mathbf{X} | \mathbf{P}_{\mathcal{S}}\mathbf{X}) = 0$ and (ii) $\text{COV}(\mathbf{P}_{\mathcal{S}}\mathbf{X}, \mathbf{Q}_{\mathcal{S}}\mathbf{X}) = 0$. These two conditions imply that $\mathbf{Q}_{\mathcal{S}}\mathbf{X}$ does not affect the distribution of \mathbf{Y} directly or through the association with $\mathbf{P}_{\mathcal{S}}\mathbf{X}$. Conditions (i) and (ii) are equivalent to (i)' $\text{span}(\boldsymbol{\beta}) \subseteq \mathcal{S}$ and (ii)' $\boldsymbol{\Sigma}_{\mathbf{X}} = \mathbf{P}_{\mathcal{S}}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{P}_{\mathcal{S}} + \mathbf{Q}_{\mathcal{S}}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{Q}_{\mathcal{S}}$. And the intersection of all \mathcal{S} that satisfies (i)' and (ii)' is the $\boldsymbol{\Sigma}_{\mathbf{X}}$ -envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$ or \mathcal{E} . Its dimension is denoted by u . Let $\boldsymbol{\Gamma} \in \mathbb{R}^{p \times u}$ be an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p \times (p-u)}$ be its completion. Under conditions (i)' and (ii)', (10) can be formulated as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\eta}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}^\top (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^\top, \quad (11)$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma} \boldsymbol{\Omega}^{-1} \boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^{u \times r}$, and $\boldsymbol{\Omega}^{-1} \boldsymbol{\eta}$ carries the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u) \times (p-u)}$ are positive definite. We call (11) the predictor envelope model. When $u = p$, the predictor envelope model reduces to the linear regression model (10). When $u = 0$, $\boldsymbol{\beta} = 0$ and the covariance between \mathbf{X} and \mathbf{Y} is zero. Cook *et al.* (2013) shows that the predictor envelope model is at least as efficient as (10) asymptotically. Substantial efficiency gains can be achieved if $\|\boldsymbol{\Omega}\| > \|\boldsymbol{\Omega}_0\|$.

A connection between partial least squares and the predictor envelope model is also established: In population, partial least squares and the predictor envelope model are estimating the same reduction $\mathbf{P}_{\mathcal{E}}\mathbf{X}$, but they use different sample algorithms. A popular sample algorithm for partial least squares is SIMPLS algorithm (de Jong 1993), which is a moment-based iterative algorithm; while the predictor envelope model optimizes an objective function based on normal likelihood. Both the SIMPLS estimator and predictor envelope estimator are \sqrt{n} -consistent. But the predictor envelope estimator usually has a better performance in prediction than the SIMPLS estimator (Cook *et al.* 2013).

2.7. Simultaneous envelope model

The simultaneous envelope model (Cook and Zhang 2015b) achieves efficiency gains by performing dimension reduction for both predictors and responses.

Under model (10), we assume that $r > 1$. Let $d \leq \min(r, p)$ denote the rank of β . We consider the following two envelopes:

1. Predictor envelope $\mathcal{E}_{\Sigma_{\mathbf{X}}}(\beta)$ with dimension d_X , where $d \leq d_X \leq p$.
2. Response envelope $\mathcal{E}_{\Sigma_{\mathbf{Y}|\mathbf{X}}}(\beta^\top)$ with dimension d_Y , where $d \leq d_Y \leq r$.

By imposing the predictor envelope and the response envelope simultaneously, we expect to obtain more efficiency gains in the estimation of β than using either the predictor envelope or response envelope alone.

Let $\Phi \in \mathbb{R}^{p \times d_X}$ be an orthonormal basis of $\mathcal{E}_{\Sigma_{\mathbf{X}}}(\beta)$, $\Gamma \in \mathbb{R}^{r \times d_Y}$ be an orthonormal basis of $\mathcal{E}_{\Sigma_{\mathbf{Y}|\mathbf{X}}}(\beta^\top)$, and $\Phi_0 \in \mathbb{R}^{p \times (p-d_X)}$ and $\Gamma_0 \in \mathbb{R}^{r \times (r-d_Y)}$ be completions of Φ and Γ respectively. Under this parameterization, (10) can be written as

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu} + \Gamma \boldsymbol{\eta}^\top \Phi^\top (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \boldsymbol{\epsilon}, \\ \Sigma_{\mathbf{X}} &= \Phi \boldsymbol{\Delta} \Phi^\top + \Phi_0 \boldsymbol{\Delta}_0 \Phi_0^\top, \\ \Sigma_{\mathbf{Y}|\mathbf{X}} &= \Gamma \boldsymbol{\Omega} \Gamma^\top + \Gamma_0 \boldsymbol{\Omega}_0 \Gamma_0^\top, \end{aligned} \quad (12)$$

where $\beta = \Phi \boldsymbol{\eta} \Gamma^\top$, $\boldsymbol{\eta} \in \mathbb{R}^{d_X \times d_Y}$, the matrices $\boldsymbol{\Delta} \in \mathbb{R}^{d_X \times d_X}$, $\boldsymbol{\Delta}_0 \in \mathbb{R}^{(p-d_X) \times (p-d_X)}$, $\boldsymbol{\Omega} \in \mathbb{R}^{d_Y \times d_Y}$, and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-d_Y) \times (r-d_Y)}$ are positive definite. We call (12) the simultaneous envelope model. When $d_X = p$, the simultaneous envelope model reduces to the response envelope model (2). When $d_Y = r$, it reduces to the predictor envelope model (11). It degenerates to the linear regression model (10) when $d_X = p$ and $d_Y = r$.

2.8. Scaled predictor envelope model

Like the scaled response envelope model, the scaled predictor envelope model (Cook and Su 2016) gives a scaled-invariant version of the predictor envelope model.

The scaled predictor envelope model allows different predictors to share the same scaling parameters. Without loss of generality, suppose that the first p_1 predictors have scaling parameters 1, the next p_2 predictors have scaling parameters λ_1, \dots , and the last p_q predictors have scaling parameters λ_{q-1} , where $\sum_{i=1}^q p_i = p$. Then we construct $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times p}$ to be a diagonal matrix with repeated diagonal elements $1, \dots, 1, \lambda_1, \dots, \lambda_1, \dots, \lambda_{q-1}, \dots, \lambda_{q-1}$, where $\lambda_1, \dots, \lambda_{q-1}$ are $q-1$ positive numbers. Given $\boldsymbol{\Lambda}$, \mathbf{Y} and the scaled predictor $\mathbf{X}_N = \boldsymbol{\Lambda}^{-1} \mathbf{X}$ follows the predictor envelope model (11). Equivalently, β and $\Sigma_{\mathbf{X}}$ satisfy (i) $\text{span}(\boldsymbol{\Lambda} \beta) \subseteq \mathcal{E}_{\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda} \beta)$ and (ii) $\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1} = \mathbf{P}_{\mathcal{E}} \boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1} \mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}} \boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1} \mathbf{Q}_{\mathcal{E}}$, where \mathcal{E} is short for $\mathcal{E}_{\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda} \beta)$ when it appears in subscripts.

Suppose that the dimension of $\mathcal{E}_{\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda} \beta)$ is u . Let $\Gamma \in \mathbb{R}^{p \times u}$ be an orthonormal basis of $\mathcal{E}_{\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1}}(\boldsymbol{\Lambda} \beta)$, and $\Gamma_0 \in \mathbb{R}^{p \times (p-u)}$ be its completion. Then (i) implies that there exists a matrix $\boldsymbol{\eta} \in \mathbb{R}^{u \times r}$ such that $\boldsymbol{\Lambda} \beta = \Gamma \boldsymbol{\eta}$, and $\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1}$ can be written as $\boldsymbol{\Lambda}^{-1} \Sigma_{\mathbf{X}} \boldsymbol{\Lambda}^{-1} = \Gamma \boldsymbol{\Omega} \Gamma^\top + \Gamma_0 \boldsymbol{\Omega}_0 \Gamma_0^\top$, where $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u) \times (p-u)}$ are both positive definite matrices. Using this parameterization, (10) is written as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\eta}^\top \Gamma^\top \boldsymbol{\Lambda}^{-1} (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \boldsymbol{\epsilon}, \quad \Sigma_{\mathbf{X}} = \boldsymbol{\Lambda} \Gamma \boldsymbol{\Omega} \Gamma^\top \boldsymbol{\Lambda} + \boldsymbol{\Lambda} \Gamma_0 \boldsymbol{\Omega}_0 \Gamma_0^\top \boldsymbol{\Lambda} \quad (13)$$

where $\boldsymbol{\beta} = \boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta}$. We call (13) the scaled predictor envelope model. When $u \geq p - (q - 1)/r$, (13) reduces to the standard linear regression (10). The scaled predictor envelope model is not nested within (10) if $u \neq p$, hence LRT cannot be applied for dimension selection.

2.9. Envelope model in generalized linear models

The envelope model can also be applied to generalized linear models (GLM) to achieve efficient estimation (Cook and Zhang 2015a). 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. We assume that $\mathbf{X} \in \mathbb{R}^p$ follows a multivariate normal distribution $N(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$. The canonical link function is $\theta(\boldsymbol{\mu}, \boldsymbol{\beta}) = \boldsymbol{\mu} + \boldsymbol{\beta}^\top \mathbf{X}$, where $\theta(\boldsymbol{\mu}, \boldsymbol{\beta})$ is a smooth and monotonic function of $\mathbb{E}(Y|\mathbf{X}, \theta)$.

The conditional log likelihood is denoted by $\mathcal{C}(\theta) := \log f(y|\theta) = y\theta - b(\theta) + c(y)$, where $\theta = \boldsymbol{\mu} + \boldsymbol{\beta}^\top \mathbf{X}$. For instance, $\mathcal{C}(\theta) = y\theta - \exp(\theta)$ in Poisson regression and $\mathcal{C}(\theta) = y\theta - \log(1 + \exp(\theta))$ in logistic regression. The standard estimator of $\boldsymbol{\beta}$ can be obtained by Fisher scoring.

To construct the envelope model under GLM, Cook and Zhang (2015a) considers the $\boldsymbol{\Sigma}_{\mathbf{X}}$ -envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$. Let u denote its dimension, $\boldsymbol{\Gamma} \in \mathbb{R}^{p \times u}$ be an orthonormal basis for $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\boldsymbol{\beta})$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{p \times (p-u)}$ be its completion. Then $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ have the structure $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^u$ contains the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$; and $\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^\top$, where $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u) \times (p-u)}$ are positive definite matrices. Under the envelope parameterization, GLM can be written as

$$\log f(y|\theta) = y\theta - b(\theta) + c(y), \quad \theta = \boldsymbol{\mu} + \boldsymbol{\eta}^\top \boldsymbol{\Gamma}^\top \mathbf{X}, \quad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^\top. \quad (14)$$

When $u = p$, (14) reduces to the standard GLM. Cook and Zhang (2015a) shows that the envelope estimator is asymptotically at least as efficient as the standard GLM estimator. The objective function of envelope estimation in this context is more complicated than other envelope models. We will discuss the estimation in more details in Section 3.

2.10. Weighted envelope estimation

With all the preceding envelope models, the dimension of the envelope subspace u must be decided before applying the envelope model to the data. The selection of u can be tricky sometimes, and it brings an extra source of variation of the envelope estimator. The weighted envelope estimation (Eck and Cook 2017) considers a weighted estimator of $\boldsymbol{\beta}$ from all envelope models with $u = 1, \dots, r$. It is derived under the context of the response envelope model (2), but the same idea can be applied to all the envelope models. Specifically, the weighted envelope estimator is defined as

$$\hat{\boldsymbol{\beta}}_w = \sum_{j=1}^r w_j \hat{\boldsymbol{\beta}}_j, \quad (15)$$

where the weights w_j are positive numbers for $j = 1, \dots, r$, and $\sum_{j=1}^r w_j = 1$. We can compute w_j based on the Bayesian Information Criterion (BIC). The BIC value for the envelope model with $u = j$ is $b_j = -2l(\hat{\boldsymbol{\beta}}_j) + k(j) \log(n)$, where $l(\cdot)$ is the log likelihood function, $k(j)$

is the number of parameters in the envelope model with dimension j and n is the sample size. Then the weight for the envelope model with $u = j$ is

$$w_j = \frac{\exp(-b_j)}{\sum_{k=1}^r \exp(-b_k)}.$$

Eck and Cook (2017) shows that the weighted envelope estimator (15) is a \sqrt{n} -consistent estimator of β . The variation of the weighted envelope estimator is evaluated by residual bootstrap. We have implemented the weighted envelopes estimator for the response envelope model, partial envelope model as well as predictor envelope model in **Renvlp**. The users do not need to pick u before using these estimators.

3. Algorithm

3.1. Algorithm for the estimation of the envelope subspace

In this subsection, we discuss the algorithm to estimate the envelope subspace implemented in **Renvlp**. In all envelope models, if the envelope subspace is known, the envelope estimator is easy to obtain by applying the standard method on the reduced response vector or / and reduced predictor vector. Let $\mathbf{\Gamma} \in \mathbb{R}^{a \times b}$ ($a > b$) be an orthonormal basis of the envelope subspace \mathcal{E} . The optimization of the envelope subspace takes three forms:

$$\hat{\mathcal{E}} = \arg \min_{\text{span}(\mathbf{\Gamma}) \in \mathcal{G}(a \times b)} \log |\mathbf{\Gamma}^\top \mathbf{M} \mathbf{\Gamma}| + \log |\mathbf{\Gamma}^\top (\mathbf{M} + \mathbf{U})^{-1} \mathbf{\Gamma}|, \quad (16)$$

$$\hat{\mathcal{E}} = \arg \min_{\text{span}(\mathbf{\Gamma}) \in \mathcal{G}(a \times b)} \log |\mathbf{\Gamma}^\top \mathbf{V}^{-1} \mathbf{\Gamma}| + \sum_{i=1}^p \frac{n_{(i)}}{n} \log |\mathbf{\Gamma}^\top \mathbf{M}_{(i)} \mathbf{\Gamma}|, \quad (17)$$

$$\hat{\mathcal{E}} = \arg \min_{\text{span}(\mathbf{\Gamma}) \in \mathcal{G}(a \times b)} -C_n(\alpha, \mathbf{\Gamma} \boldsymbol{\eta}_{\mathbf{\Gamma}}) + \frac{n}{2} \{ \log |\mathbf{\Gamma}^\top \mathbf{S}_{\mathbf{X}} \mathbf{\Gamma}| + \log |\mathbf{\Gamma}^\top \mathbf{S}_{\mathbf{X}}^{-1} \mathbf{\Gamma}| + \log |\mathbf{S}_{\mathbf{X}}| \}, \quad (18)$$

where $\mathcal{G}(a \times b)$ denotes an a by b Grassmann manifold, \mathbf{M} , $\mathbf{M} + \mathbf{U}$, $\mathbf{M}_{(i)}$ ($i = 1, \dots, p$) and \mathbf{V} are positive definite matrices, $\mathbf{S}_{\mathbf{X}}$ denotes the sample covariance matrix of \mathbf{X} and it is positive definite when $n > p$, \mathbf{X}_i is the i -th sample of \mathbf{X} , and $C_n(\mu, \mathbf{\Gamma} \boldsymbol{\eta}_{\mathbf{\Gamma}}) = \sum_{i=1}^n \mathcal{C}(\mu + \boldsymbol{\eta}_{\mathbf{\Gamma}}^\top \mathbf{\Gamma}^\top \mathbf{X}_i)$ is the conditional log likelihood function in (14) with μ fixed and $\boldsymbol{\eta}$ treated as a function of $\mathbf{\Gamma}$ (denoted by $\boldsymbol{\eta}_{\mathbf{\Gamma}}$).

The objective function in (17) is used in the heteroscedastic envelope model and groupwise envelope model; the objective function in (18) is used in the envelope model in GLM; and the objective function in (16) is used in all other envelope models. In different envelope models, \mathbf{M} , \mathbf{U} and other matrices take different forms. For example, in response envelope model (c.f. Section 2.1), $\mathbf{M} = \mathbf{S}_{\text{res}}$ and $\mathbf{M} + \mathbf{U} = \mathbf{S}_{\mathbf{Y}}$, where \mathbf{S}_{res} is the sample covariance matrix of the residuals from the standard linear regression of \mathbf{Y} on \mathbf{X} and $\mathbf{S}_{\mathbf{Y}}$ is the sample covariance matrix of \mathbf{Y} . In predictor envelope model (c.f. Section 2.6), $\mathbf{M} = \mathbf{S}_{\mathbf{X}} - \mathbf{S}_{\mathbf{X}\mathbf{Y}} \mathbf{S}_{\mathbf{Y}}^{-1} \mathbf{S}_{\mathbf{Y}\mathbf{X}}$ and $\mathbf{M} + \mathbf{U} = \mathbf{S}_{\mathbf{X}}$, where $\mathbf{S}_{\mathbf{X}\mathbf{Y}}$ is the sample covariance of \mathbf{X} and \mathbf{Y} .

The optimization problems in (16), (17) and (18) all involve Grassmann manifold optimization, which can be slow in sizable problems. To resolve this issue, Cook *et al.* (2016) introduced a re-parametrization of the parameters in the context of (16) that converts the Grassmann

manifold optimization to an unconstrained matrix optimization. We first review this re-parametrization with (16) and then generalize this technique to the objective functions in (17) and (18).

Let $\mathbf{\Gamma} \in \mathbb{R}^{a \times b}$ be an orthonormal basis of the envelope subspace \mathcal{E} . Without loss of generality, we assume the matrix $\mathbf{\Gamma}_1$ formed by the first b rows of $\mathbf{\Gamma}$ is nonsingular. Let $\mathbf{\Gamma}_2$ be the $(a - b) \times b$ matrix formed by the rest of the rows. Then $\mathbf{\Gamma}$ can be partitioned as

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_b \\ \mathbf{\Gamma}_2 \mathbf{\Gamma}_1^{-1} \end{pmatrix} \mathbf{\Gamma}_1 := \begin{pmatrix} \mathbf{I}_b \\ \mathbf{A} \end{pmatrix} \mathbf{\Gamma}_1 := \mathbf{G}_\mathbf{A} \mathbf{\Gamma}_1. \quad (19)$$

Note that \mathbf{A} depends on $\mathbf{\Gamma}$ only through $\text{span}(\mathbf{\Gamma}) = \mathcal{E}$. The re-parameterization in (19) builds a one-to-one correspondence between \mathcal{E} and \mathbf{A} : If we know about \mathbf{A} , then $\mathcal{E} = \text{span}(\mathbf{G}_\mathbf{A})$. If \mathcal{E} is given, we can take any of its orthonormal basis and go through the steps in (19) to get \mathbf{A} . Under this re-parametrization, the optimization in (16) is converted to

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A} \in \mathbb{R}^{(a-b) \times b}} \log |\mathbf{G}_\mathbf{A}^\top \mathbf{M} \mathbf{G}_\mathbf{A}| + \log |\mathbf{G}_\mathbf{A}^\top (\mathbf{M} + \mathbf{U})^{-1} \mathbf{G}_\mathbf{A}| - 2 \log |\mathbf{G}_\mathbf{A}^\top \mathbf{G}_\mathbf{A}|. \quad (20)$$

Note that the optimization in (20) does not involve manifold optimization and is unconstrained. When $(r - u)u$ is small, the minimization in (20) can be carried out directly using standard optimization tools. Otherwise, the optimization can be performed by the blockwise coordinate descent algorithm, treating each row of \mathbf{A} as a block. The details are included in Cook *et al.* (2016).

Cook *et al.* (2016) suggested four possible initial values of \mathbf{A} computed from the eigenvectors of \mathbf{M} or $\mathbf{M} + \mathbf{U}$, and showed the asymptotic property of these initial values.

Now we apply the re-parametrization to (17) and (18). After some calculations, the optimization in (17) can be converted to

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A} \in \mathbb{R}^{(a-b) \times b}} \log |\mathbf{G}_\mathbf{A}^\top \mathbf{V}^{-1} \mathbf{G}_\mathbf{A}| + \sum_{i=1}^p \frac{n_{(i)}}{n} \log |\mathbf{G}_\mathbf{A}^\top \mathbf{M}_{(i)} \mathbf{G}_\mathbf{A}| - 2 \log |\mathbf{G}_\mathbf{A}^\top \mathbf{G}_\mathbf{A}|, \quad (21)$$

and the optimization in (18) can be converted to

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A} \in \mathbb{R}^{(a-b) \times b}} -C_n(\mu, \mathbf{G}_\mathbf{A} \boldsymbol{\eta}^*) + \frac{n}{2} (\log |\mathbf{G}_\mathbf{A}^\top \mathbf{S}_\mathbf{X} \mathbf{G}_\mathbf{A}| + \log |\mathbf{G}_\mathbf{A}^\top \mathbf{S}_\mathbf{X}^{-1} \mathbf{G}_\mathbf{A}| - 2 \log |\mathbf{G}_\mathbf{A}^\top \mathbf{G}_\mathbf{A}|), \quad (22)$$

where $\boldsymbol{\eta}^* \in \mathbb{R}^b$ carries the coordinates of $\boldsymbol{\beta}$ with respect to $\mathbf{G}_\mathbf{A}$, i.e., $\boldsymbol{\beta} = \mathbf{G}_\mathbf{A} \boldsymbol{\eta}^*$, and $C_n(\mu, \mathbf{G}_\mathbf{A} \boldsymbol{\eta}^*) = \sum_{i=1}^n \mathcal{C}(\mu + \boldsymbol{\eta}^{*T} \mathbf{G}_\mathbf{A}^\top \mathbf{X}_i)$. Under Poisson regression, $C_n(\mu, \mathbf{G}_\mathbf{A} \boldsymbol{\eta}^*) = \sum_{i=1}^n [Y_i(\mu + \boldsymbol{\eta}^{*T} \mathbf{G}_\mathbf{A}^\top \mathbf{X}_i) - \exp(\mu + \boldsymbol{\eta}^{*T} \mathbf{G}_\mathbf{A}^\top \mathbf{X}_i)]$.

The optimization of (21) and (22) can be performed similarly as (20) using blockwise coordinate descent. The initial values for solving (21) can be computed from the eigenvectors of \mathbf{V} and $\sum_{i=1}^p n_{(i)} \mathbf{M}_{(i)} / n$ using the same way as computing the initial value for (20). To compute the initial value for (22), we first fit the standard GLM to obtain $\tilde{\mu}$ and $\tilde{\boldsymbol{\beta}}$ (e.g. use the Fisher scoring method implemented in R package `stats`). Then a \sqrt{n} -consistent initial value of \mathbf{A} can be obtained by solving (20) with $\mathbf{U} = \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^\top$ and \mathbf{M} being the estimated asymptotic variance of $\tilde{\boldsymbol{\beta}}$ (Cook and Zhang 2015a). Specifically, the asymptotic variance of $\tilde{\boldsymbol{\beta}}$ can be estimated by

$\mathbf{M} = \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}(W)}) \mathcal{C}''(\tilde{\theta}_i) (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}(W)})^\top / n$, where \mathcal{C}'' denotes the second derivative of \mathcal{C} , $\tilde{\theta}_i = \tilde{\mu} + \tilde{\boldsymbol{\beta}}^\top \mathbf{X}_i$, $W_i = n\mathcal{C}''(\tilde{\theta}_i) / \sum_{i=1}^n \mathcal{C}''(\tilde{\theta}_i)$ and $\boldsymbol{\mu}_{\mathbf{X}(W)} = \sum_{i=1}^n W_i \mathbf{X}_i / n$.

3.2. Numerical demonstration

In this subsection, we compare our non-Grassmann estimation algorithm with the envelope coordinate descent (ECD) algorithm (Cook and Zhang 2017) and the one-direction-at-a-time (1D) algorithm (Cook and Zhang 2016). The 1D algorithm estimates each column of the orthonormal basis of \mathcal{E} at a time. The ECD algorithm is developed under the framework of the 1D algorithm. It utilizes an approximation in estimating each column and turns out to be much faster than the 1D algorithm. Both the ECD and 1D algorithms yield \sqrt{n} -consistent estimators. In the simulation studies, the non-Grassmann, ECD and 1D algorithms are all implemented in R. We compare the computing time and estimation accuracy among these algorithms. The estimation accuracy is measured by the principal angle between two subspaces. Its calculation is described in Knyazev and Argentati (2002). The convergence criterion is met when the relative change of the objective function is less than 10^{-3} or the number of iteration reaches 100. The computing time is recorded on a Macbook with 1.3GHz dual-core Intel Core i5 and RAM 4GB.

The first simulation focuses on the computing time. We adopted the simulation settings in Cook and Zhang (2017) and estimated the envelope subspace with the objective function (16) given \mathbf{M} and \mathbf{U} . We fixed $r = 200$ and considered three scenarios. For each scenario, we considered $u = 5$ and $u = 10$, and generated the matrices \mathbf{M} and \mathbf{U} as follows

$$\mathbf{M} = \begin{cases} \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^\top, & \text{for Scenario I,} \\ \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + 0.01\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^\top, & \text{for Scenario II,} \\ 0.01\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^\top, & \text{for Scenario III,} \end{cases}$$

$$\mathbf{U} = \boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^\top, \quad \text{for all scenarios,}$$

where $\mathbf{A} \in \mathbb{R}^{u \times u}$, $\mathbf{B} \in \mathbb{R}^{(r-u) \times (r-u)}$ and $\mathbf{C} \in \mathbb{R}^{u \times u}$ were matrices containing independent uniform $(0, 1)$ variates, and $\boldsymbol{\Omega} = \mathbf{A}\mathbf{A}^\top$, $\boldsymbol{\Omega}_0 = \mathbf{B}\mathbf{B}^\top$ and $\boldsymbol{\Phi} = \mathbf{C}\mathbf{C}^\top$. The matrix $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ was obtained by normalizing an $r \times r$ matrix of independent uniform $(0, 1)$ variates. After \mathbf{M} was generated, we added $0.0001\mathbf{I}_r$ to \mathbf{M} to assure that \mathbf{M} is positive definite.

For each u in each scenario, 50 replications were generated. The results were summarized in Table 1, where each cell was the averaged computing time with the standard deviation in parentheses. The angle between the true envelope subspace and the estimated envelope subspace was less than 10^{-5} for all algorithms in each replication, so it is not reported in the table. Under this setting, the non-Grassmann algorithm is about five to ten times faster than the ECD algorithm. The ECD algorithm is faster than the 1D algorithm in Scenario I and about the same as the 1D algorithm in Scenarios II and III.

The second simulation focuses on estimation accuracy. The simulation settings are the same as in Cook *et al.* (2016), which is under the context of the response envelope model. We set $n = 250$, $r = 100$, $p = 100$ and varied u from 1 to 90. The elements in \mathbf{X} were generated as independent $N(0, 20^2)$. The matrix $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ was obtained by normalizing an $r \times r$ matrix of independent uniform $(0, 1)$ variates, and the elements in $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ were generated as independent uniform $(0, 10)$ variates. The intercept and the coefficients were $\boldsymbol{\mu} = 0$ and $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$. First, we considered the setting where the immaterial part of \mathbf{Y} is more variant

Scenarios	u	non-Grassmann	ECD	1D
Scenario I	5	0.31 (0.017)	1.67 (0.082)	4.24 (0.340)
	10	0.33 (0.017)	3.53 (0.262)	9.33 (0.396)
Scenario II	5	0.30 (0.020)	1.82 (0.048)	1.21 (0.028)
	10	0.32 (0.017)	3.62 (0.070)	2.35 (0.033)
Scenario III	5	0.31 (0.016)	1.80 (0.032)	1.21 (0.021)
	10	0.33 (0.016)	3.36 (0.043)	2.32 (0.033)

Table 1: Computing time (in seconds) for each algorithm.

than the material part of \mathbf{Y} and set the error covariance matrix as $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^\top + \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^\top$, where $\mathbf{\Omega} = \mathbf{A}\mathbf{A}^\top$, $\mathbf{\Omega}_0 = 25\mathbf{C}\mathbf{C}^\top$. The elements in $\mathbf{A} \in \mathbb{R}^{u \times u}$ and $\mathbf{C} \in \mathbb{R}^{(r-u) \times (r-u)}$ were independent $N(0, 1)$ variates. We generated 50 replications and recorded the angle between the true envelope subspace and the estimated subspace as well as the computing time for all three algorithms. The results are summarized in Table 2. The ECD algorithm is faster and has a smaller angle when $u \leq 50$. However, as u increases, the advantage of the non-Grassmann algorithm becomes more apparent. The 1D algorithm is inferior than the non-Grassmann algorithm in both accuracy and computational efficiency when $u > 10$.

We repeated the simulation with the setting that the material part of \mathbf{Y} is more variant than the immaterial part of \mathbf{Y} . The error covariance matrix was generated exactly the same as in the last setting except that $\mathbf{\Omega} = 25\mathbf{A}\mathbf{A}^\top$, $\mathbf{\Omega}_0 = \mathbf{C}\mathbf{C}^\top$. In this setting, estimating the envelope subspace is more challenging. The results are summarized in Table 3. Although the non-Grassmann algorithm and the ECD algorithm have about the same accuracy time when u is not large, the non-Grassmann algorithm is much faster than the ECD algorithm. The 1D algorithm is inferior than the competitors, both in terms of accuracy and computational efficiency.

4. Structure of **Renvlp**

The **Renvlp** package implements all the envelope models discussed in Section 2 based on the non-Grassmann algorithm described in Section 3. The envelope models supported by **Renvlp** along with the corresponding R function names are presented in Table 4. For example, the function **env** implements the response envelope model and the function **genv** implements the groupwise envelope model. Besides these main functions, **Renvlp** includes auxiliary tools that select the dimension of the envelope subspace, and inference tools which perform bootstrapping, prediction and hypothesis testing under different envelope models. The **Renvlp** package depends on R package **Rsolnp** (Ghalanos and Theussl 2015) for the estimation of the scaled response envelope model and scaled predictor envelope model.

4.1. Main functions

Given the predictors and responses, the main functions **Function name** fit the corresponding envelope model with the specified envelope dimension u . The corresponding **Function name** for different envelope model is given in Table 4. The outputs of the main functions are the envelope estimators as well as a few important statistics. All the main functions have the same basic structure for inputs and outputs, but they vary slightly according to different models.

u	non-Grassmann	ECD	1D
1	0.67 (0.26)	0.66 (0.26)	0.67 (0.27)
5	3.04 (0.74)	2.18 (0.28)	2.17 (0.29)
10	3.94 (0.78)	2.93 (0.30)	2.88 (0.31)
20	4.79 (0.72)	3.76 (0.55)	16.86 (30.51)
30	5.91 (0.92)	4.33 (0.39)	21.09 (34.17)
40	6.44 (0.81)	4.88 (0.67)	21.27 (33.38)
50	7.12 (1.10)	7.01 (11.99)	27.68 (36.49)
60	7.60 (1.48)	9.27 (16.65)	39.63 (39.22)
70	8.46 (1.78)	9.57 (16.26)	48.71 (38.73)
80	8.84 (2.50)	17.20 (27.02)	41.49 (33.13)
90	9.72 (6.69)	27.39 (35.03)	33.00 (23.04)

u	non-Grassmann	ECD	1D
1	0.09 (0.01)	0.09 (0.01)	0.28 (0.09)
5	2.36 (2.49)	0.43 (0.02)	1.45 (0.25)
10	3.38 (2.77)	0.79 (0.02)	2.69 (0.30)
20	4.49 (3.23)	1.35 (0.03)	4.98 (0.41)
30	4.45 (3.61)	1.77 (0.03)	6.91 (0.52)
40	6.09 (3.76)	2.04 (0.04)	8.27 (0.60)
50	5.11 (3.33)	2.28 (0.05)	9.40 (0.70)
60	4.95 (2.84)	2.40 (0.07)	10.24 (0.77)
70	3.68 (1.90)	2.50 (0.13)	11.10 (0.80)
80	2.79 (1.35)	2.66 (0.16)	11.66 (0.75)
90	1.39 (0.47)	3.23 (0.32)	11.74 (0.99)

Table 2: Upper panel: Angles between the true envelope subspace and the estimated envelope subspace. Lower panel: Computing time (in seconds). Each cell contains the average value from 50 replications with the standard deviation in parentheses.

u	non-Grassmann	ECD	1D
1	0.30 (0.06)	0.30 (0.06)	0.30 (0.06)
5	0.77 (0.06)	0.77 (0.07)	0.77 (0.07)
10	0.90 (0.08)	0.90 (0.08)	2.26 (5.47)
20	1.09 (0.08)	1.09 (0.08)	4.49 (6.71)
30	1.24 (0.10)	1.24 (0.11)	13.87 (19.58)
40	1.33 (0.10)	1.33 (0.11)	13.57 (15.80)
50	1.49 (0.14)	1.50 (0.15)	24.55 (32.16)
60	1.57 (0.19)	1.59 (0.22)	16.32 (26.51)
70	1.56 (0.17)	1.60 (0.30)	11.06 (21.20)
80	1.54 (0.19)	8.61 (24.16)	12.18 (25.78)
90	1.31 (0.21)	15.53 (32.69)	12.59 (28.22)

u	non-Grassmann	ECD	1D
1	0.07 (0.02)	0.07 (0.01)	0.27 (0.01)
5	0.18 (0.03)	0.35 (0.20)	1.36 (0.25)
10	0.28 (0.06)	0.82 (0.33)	2.71 (0.38)
20	0.38 (0.11)	2.12 (0.66)	5.17 (0.43)
30	0.43 (0.19)	4.91 (0.96)	7.03 (0.32)
40	0.55 (0.19)	7.85 (0.87)	8.52 (0.47)
50	0.61 (0.28)	9.66 (0.80)	9.66 (0.48)
60	0.74 (0.30)	11.18 (0.69)	10.82 (0.64)
70	0.69 (0.33)	12.15 (0.82)	11.58 (0.77)
80	0.59 (0.30)	13.02 (0.80)	11.91 (0.75)
90	0.43 (0.17)	14.53 (0.82)	12.13 (0.74)

Table 3: Upper panel: Angles between the true envelope subspace and the estimated envelope subspace. Lower panel: Computing time (in seconds). Each cell contains the average value from 50 replications with the standard deviation in parentheses.

Function name	Model
<code>env</code>	Response envelope model (Section 2.1)
<code>penv</code>	Partial envelope model (Section 2.2)
<code>henv</code>	Heteroscedastic envelope model (Section 2.3)
<code>genv</code>	Groupwise envelope model (Section 2.4)
<code>senv</code>	Scaled response envelope model (Section 2.5)
<code>xenv</code>	Predictor envelope model (Section 2.6)
<code>stenv</code>	Simultaneous envelope model (Section 2.7)
<code>sxenv</code>	Scaled predictor envelope model (Section 2.8)
<code>logit.env</code>	Envelope model in logistic regression (Section 2.9)
<code>pois.env</code>	Envelope model in poisson regression (Section 2.9)
<code>weighted.env</code>	Weighted response envelope estimation (Section 2.10)
<code>weighted.penv</code>	Weighted partial envelope estimation (Section 2.10)
<code>weighted.xenv</code>	Weighted predictor envelope estimation (Section 2.10)

Table 4: A list of models supported by **Renvlp**

For instance, `weighted.env` does not require the users to specify u ; and `stenv` requires the users to specify the dimensions for both the predictor envelope $\mathcal{E}_{\Sigma_X}(\beta)$ and the response envelope $\mathcal{E}_{\Sigma_{Y|X}}(\beta^\top)$. We take `env` as an example to explain the usage of the function, since it presents the basic structure of the main functions. The syntax of `env` is

```
env(X, Y, u, asy = TRUE, init = NULL)
```

The required inputs are `X`, `Y` and `u`, where `X` contains the predictors, `Y` contains the multivariate responses and `u` is the dimension of the envelope subspace. The selection of the dimension is included in Section 4.2. The optional inputs are `asy` and `init`: `asy` is the flag for computing the asymptotic variance of the envelope estimator. The default setting is `TRUE`. If only the envelope estimators are needed, the flag can be set to `asy = FALSE`. `init` is the user-specified initial value of `Gamma`. If no starting value is provided, `env` will use the starting value suggested in Cook *et al.* (2016).

The output of the main function is a list which consists of the envelope estimators and a few important statistics. The envelope estimators include `beta`, `Sigma`, `Gamma`, `Gamma0`, `eta`, `Omega`, `Omega0` and `mu`, which are estimators of β , Σ , Γ , Γ_0 , η , Ω , Ω_0 and μ in (2). The statistics in the output list are `loglik`, the maximized log likelihood; `covMatrix`, the asymptotic covariance of the vectorized `beta`; `asySE`, the asymptotic standard error for elements in `beta`; `ratio`, the asymptotic standard error ratio of the standard multivariate linear regression estimator over the envelope estimator and `n`, the number of observations in the dataset. An illustration is given in Section 5.

4.2. Selection of the dimension

The functions `u.Function name` choose the dimension of the envelope subspace u based on AIC, BIC and likelihood ratio testing (LRT) with specified significance level. For example, `u.env` selects u under the response envelope model, and `u.henv` selects u under the heteroscedastic envelope model. We again use the example of the response envelope model. The syntax of `u.env` is

```
u.env(X, Y, alpha = 0.01)
```

The required inputs are `X` and `Y`, which are the predictors and responses. The optional input `alpha` is the significance level for LRT. The default value is 0.01.

The output of `u.Function name` includes `u.aic`, `u.bic` and `u.lrt`, which are dimensions selected by AIC, BIC and LRT; as well as `loglik.seq`, `aic.seq` and `bic.seq`, which are sequences of log likelihood, AIC value and BIC for $u = 0, \dots, r$.

4.3. Inference functions

Renvlp supports inference tools including bootstrapping, prediction and hypothesis testing.

The function `boot.Function name` computes the standard errors of the envelope estimator of β by residual bootstrap. For the response envelope model, the syntax of `boot.env` is

```
boot.env(X, Y, u, B)
```

The inputs \mathbf{X} , \mathbf{Y} , \mathbf{u} and B are predictors, responses, dimension of the envelope subspace, and the number of bootstrap samples. The output is a matrix with the same size as $\hat{\boldsymbol{\beta}}$ containing the standard error for each element in the estimated $\hat{\boldsymbol{\beta}}$.

Prediction under the envelope models is performed by `pred.Function name`. For example,

```
pred.env(m, Xnew)
```

performs prediction under the response envelope model. The input m is the output from the main function `env` and the input \mathbf{X}_{new} is the value of the predictor. The outputs include `value`, which is the fitted or the predicted value evaluated at \mathbf{X}_{new} ; `covMatrix.estm` and `SE.estm`, which are the covariance matrix of the fitted value and standard error of each element in the fitted value; `covMatrix.pred` and `SE.pred`, which are the covariance matrix of the predicted value and standard error of each element in the predicted value. The covariance matrix and standard errors are estimated with normality assumption.

The function `cv.Function name` provides the prediction errors using m -fold cross validation. If \mathbf{Y}_{pred} is the predicted value and \mathbf{Y}_o is the true value, then the prediction error computed in `cv.Function name` is the square root of $(\mathbf{Y}_{\text{pred}} - \mathbf{Y}_o)^\top (\mathbf{Y}_{\text{pred}} - \mathbf{Y}_o)$. For the response envelope model, the syntax of `cv.env` is

```
cv.env(X, Y, u, m, nperm)
```

The inputs of this function \mathbf{X} , \mathbf{Y} , \mathbf{u} , m and $nperm$ are predictors, responses, dimension of the envelope subspace, the number of folds, and the number of permutations. The data is divided into m parts randomly and each part is in turn used for testing while the rest $m - 1$ parts are used for training. This process is repeated for $nperm$ times and the output is the average prediction error.

The function `testcoef.Function name` tests for the hypotheses

$$H_0 : \mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} = \mathbf{A} \quad \text{versus} \quad H_a : \mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} \neq \mathbf{A}, \quad (23)$$

where \mathbf{L} , \mathbf{R} and \mathbf{A} are constant matrices. The test statistics is $\text{vec}(\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} - \mathbf{A})\hat{\mathbf{V}}^{-1}\text{vec}(\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} - \mathbf{A})^\top$, where $\hat{\boldsymbol{\beta}}$ is the envelope estimator, vec is the vector operator that vectorize a matrix into a vector columnwise, and $\hat{\mathbf{V}}$ is the estimated asymptotic covariance of $\text{vec}(\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} - \mathbf{A})$. Suppose \mathbf{A} is a d_1 by d_2 matrix. The reference distribution is chi-squared distribution with degrees of freedom $d_1 * d_2$. To test hypothesis (23) under the envelope model, we use

```
testcoef.env(m, L, R, A)
```

The input m is the output from the main function `env`, and the inputs L , R and A correspond to the matrices \mathbf{L} , \mathbf{R} and \mathbf{A} in (23). The output of `testcoef.env` includes the test statistic, the degrees of freedom of the reference chi-squared distribution, the p -value of the test, and $\hat{\mathbf{V}}$, the estimated covariance matrix of $\text{vec}(\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{R} - \mathbf{A})$.

5. Illustration

This section illustrates the usage of the main functions, dimension selection functions and inference functions with the response envelope model. The Berkeley guidance study (Tuddenham and Snyder 1953) contains height measurement for 39 boys and 54 girls born in 1928–1929 in Berkeley, CA. For simplicity, we take heights at ages 13 and 14 to be the bivariate response $\mathbf{Y} = (Y_1, Y_2)^\top$, and gender to be the predictor X . The predictor takes value 1 or 0 to indicate boys or girls. Therefore $\beta = E(\mathbf{Y} \mid X = 1) - E(\mathbf{Y} \mid X = 0)$ indicate the height difference between boys and girls. The data is included in the package. We first load the data and select the dimension of the envelope subspace with `u.env`.

```
library("Renvlp")
data("Berkeley")
X <- Berkeley[, 1]
Y <- Berkeley[, c(22, 24)]
u <- u.env(X, Y)
u

## $u.aic
## [1] 2
##
## $u.bic
## [1] 1
##
## $u.lrt
## [1] 1
##
## $loglik.seq
## [1] -547.0461 -506.6899 -505.0067
##
## $aic.seq
## [1] 1104.092 1025.380 1024.013
##
## $bic.seq
## [1] 1116.755 1040.575 1041.741
```

From the output of `u.env`, BIC and LRT agree that the dimension of the envelope subspace is 1 while AIC picks the dimension to be 2. AIC tends to overfit the model when the sample size is small. So we fit the envelope model with $u = 1$.

```
m <- env(X, Y, 1)
m$beta

##           [,1]
## [1,] -2.149607
## [2,]  2.134949

m$Gamma

##           [,1]
```

```
## [1,] -0.7095217
## [2,]  0.7046835

m$Gamma0

##           [,1]
## [1,]  0.7046835
## [2,]  0.7095217
```

The envelope estimator of β is $(-1.72, 1.66)^\top$, indicating that girls are 1.72 cm taller than boys on average at age 13 but boys 1.66 cm are taller than girls on average at age 14. By the construction of the envelope model (2), $\Gamma_0^\top \mathbf{Y} \approx 0.71(Y_1 + Y_2)$ is immaterial to the variation in X and $\Gamma^\top \mathbf{Y} \approx -0.71(Y_1 - Y_2)$ is material to the variation in X . This indicates that the sum of the height $Y_1 + Y_2$ does not differ between boys and girls at these ages. It is the difference of the height $Y_1 - Y_2$ that carries the gender information. We also fit the standard model to the data. When $u = 2$, the envelope model reduces to the standard model. So we can simply use the `env` function with `u = 2`.

```
m2 <- env(X, Y, 2)
m2$beta

##           [,1]
## V22  0.7844729
## V24  5.0891738
```

The standard estimator indicates that the boys are slightly taller than girls at age 13 and this height difference is more pronounced at age 14. This shows a discrepancy between the envelope model `m` and the standard model `m2` on whether boys or girls are taller at age 13. To compare the two estimators, we calculate the standard errors of each estimator.

```
m$asySE / sqrt(93)

##           [,1]
## [1,]  0.1878946
## [2,]  0.1866617

m2$asySE / sqrt(93)

##           [,1]
## [1,]  1.595197
## [2,]  1.606984
```

To obtain the standard errors of each estimator, we use the asymptotic standard deviation `asySE` divided by the square root of the sample size 93. Notice that the standard errors of the standard estimator are 7.4 and 8.0 times as large as that of the envelope estimator,

which indicates the envelope estimator is much more efficient than the standard estimator. `asySE` is calculated by assuming normality. The sample distribution of boys and girls both resembles normal, so the estimates of the standard errors should be reliable. Nevertheless, we also estimated the standard errors using bootstrap.

```
bootse1 <- boot.env(X, Y, 1, 200)
bootse2 <- boot.env(X, Y, 2, 200)
bootse1

##           [,1]
## [1,] 0.1659043
## [2,] 0.1752482

bootse2

##           [,1]
## [1,] 1.668975
## [2,] 1.693617
```

The bootstrap standard errors are very close to the standard errors derived from asymptotic standard errors. By investigating the standard errors, the height difference at age 13 under the standard model is not significant, while the envelope model detects that the girls are significantly taller than the boys. This is a result of the efficiency gains from the envelope model. We also perform a hypothesis testing to confirm that the first element in β is 0, i.e.,

$$H_0 : (1, 0)^\top \beta = 0 \quad \text{versus} \quad H_a : (1, 0)^\top \beta \neq 0,$$

both under the standard model and the envelope model.

```
L <- matrix(c(1, 0), 1, 2)
R <- as.matrix(1)
A <- as.matrix(0)
hres <- testcoef.env(m, L, R, A)
hres2 <- testcoef.env(m2, L, R, A)
hres

## $chisqStatistic
##           [,1]
## [1,] 130.885
##
## $dof
## [1] 1
##
## $pValue
##           [,1]
## [1,] 2.623771e-30
##
```

```
## $covMatrix
##          [,1]
## [1,] 0.03530436

hres2

## $chisqStatistic
##          [,1]
## [1,] 0.2418397
##
## $dof
## [1] 1
##
## $pValue
##          [,1]
## [1,] 0.6228806
##
## $covMatrix
##          [,1]
## [1,] 2.544652
```

The p values indicate that the height difference is significant under the envelope model and not significant under the standard model, which is consistent with our preceding discussion. At last, suppose we want to predict heights at ages 13 and 14 for a boy, we can use `pred.env`.

```
pres <- pred.env(m, 1)
pres

## $value
##          [,1]
## [1,] 158.6604
## [2,] 166.2257
##
## $covMatrix.estm
##          [,1]      [,2]
## [1,] 0.6778970 0.5963893
## [2,] 0.5963893 0.6860584
##
## $SE.estm
## [1] 0.8233450 0.8282864
##
## $covMatrix.pred
##          [,1]      [,2]
## [1,] 60.43901 59.19872
## [2,] 59.19872 61.24913
```

```
##
## $SE.pred
## [1] 7.774253 7.826182
```

The heights of a boy at ages 13 and 14 are predicted to be 158.66cm and 166.23cm, with prediction standard error 7.77cm and 7.83cm.

6. Conclusion

The R package **Renvlp** aims to provide accurate and computational efficient estimations for the envelope models. It includes the fundamental envelope models as well as the recent developments. The structure of the package is straightforward: for a particular envelope model, it contains the main function that fits the model, the dimension selection tools and the inference functions. So it is easy for the developers to include new models or extend the functionality of the package. The package will be uploaded to the Comprehensive R Archive Network (CRAN) and under constant maintenance.

Acknowledgements

The authors would like to sincerely thank Yeonhee Park for allowing us to modify the group-wise envelope model codes for use in the **Renvlp** package. This work is supported by the Graduate School Fellowship (GSF) at the University of Florida and a grant from the National Science Foundation DMS 1407460.

References

- Cook RD, Forzani L, Su Z (2016). “A Note on Fast Envelope Estimation.” *Journal of Multivariate Analysis*, **150**, 42–54. doi:10.1016/j.jmva.2016.05.006.
- Cook RD, Forzani L, Zhang X (2015a). “Envelopes and reduced-rank regression.” *Biometrika*, **102**(2), 439–456. doi:10.1093/biomet/asv001.
- Cook RD, Helland IS, Su Z (2013). “Envelopes and Partial Least Squares Regression.” *Journal of the Royal Statistical Society: Series B*, **75**, 851–877. doi:10.1111/rssb.12018.
- Cook RD, Li B, Chiaromonte F (2010). “Envelope Models for Parsimonious and Efficient Multivariate Linear Regression.” *Statistica Sinica*, **20**(3), 927–960.
- Cook RD, Su Z (2013). “Scaled Envelopes: Scale Invariant and Efficient Estimation in Multivariate Linear Regression.” *Biometrika*, **100**(4), 939–954. doi:10.1093/biomet/ast026.
- Cook RD, Su Z (2016). “Scaled Predictor Envelopes and Partial Least-Squares Regression.” *Technometrics*, **58**(2), 155–165. doi:10.1080/00401706.2015.1017611.
- Cook RD, Su Z, Yang Y (2015b). “envlp: A MATLAB Toolbox for Computing Envelope Estimators in Multivariate Analysis.” *Journal of Statistical Software, Articles*, **62**(8), 1–20. doi:10.18637/jss.v062.i08.

- Cook RD, Zhang X (2015a). “Foundations for Envelope Models and Methods.” *Journal of the American Statistical Association*, **110**(510), 599–611. doi:10.1080/01621459.2014.983235.
- Cook RD, Zhang X (2015b). “Simultaneous Envelopes for Multivariate Linear Regression.” *Technometrics*, **57**, 11–25. doi:10.1080/00401706.2013.872700.
- Cook RD, Zhang X (2016). “Algorithms for Envelope Estimation.” *Journal of Computational and Graphical Statistics*, **25**(1), 284–300. doi:10.1080/10618600.2015.1029577.
- Cook RD, Zhang X (2017). “Fast Envelope Algorithms.” *Statistica Sinica*, **To appear**. doi:10.5705/ss.202016.0037.
- de Jong S (1993). “SIMPLS: An Alternative Approach to Partial Least Squares Regression.” *Chemometrics and Intelligent Laboratory Systems*, **18**(3), 251–263. doi:10.1016/0169-7439(93)85002-X.
- Eck DJ, Cook RD (2017). “Weighted envelope estimation to handel variability in model selection.” *Biometrika*, **To appear**. doi:10.1093/biomet/asx035.
- Ghalanos A, Theussl S (2015). *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method*. R package version 1.16., URL <https://CRAN.R-project.org/package=Rsolnp>.
- Khare K, Pal S, Su Z (2017). “A Bayesian Approach for Envelope Models.” *The Annals of Statistics*, **45**(1), 196–222. doi:10.1214/16-AOS1449.
- Knyazev AV, Argentati ME (2002). “Principal Angles between Subspaces in an A-based Scalar Product: Algorithms and Perturbation Estimates.” *SIAM Journal on Scientific Computing*, **23**(6), 2008–2040. doi:10.1137/S1064827500377332.
- Li G, Yang D, Nobel AB, Shen H (2016). “Supervised singular value decomposition and its asymptotic properties.” *Journal of Multivariate Analysis*, **146**, 7–17. doi:10.1016/j.jmva.2015.02.016.
- Li L, Zhang X (2017). “Parsimonious Tensor Response Regression.” *Journal of the American Statistical Association*, pp. 1–16. doi:10.1080/01621459.2016.1193022.
- Park Y, Su Z, Zhu H (2017). “Groupwise Envelope Models for Imaging Genetic Analysis.” *Biometrics*, **To appear**. doi:10.1111/biom.12689.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Su Z, Cook RD (2011). “Partial Envelopes for Efficient Estimation in Multivariate Linear Regression.” *Biometrika*, **98**, 133–146. doi:10.1093/biomet/asq063.
- Su Z, Cook RD (2013). “Estimation of Multivariate Means With Heteroscedastic Errors Using Envelope Models.” *Statistica Sinica*, **23**, 213–230. doi:10.5705/ss.2010.240.
- Su Z, Zhu G, Chen X, Yang Y (2016). “Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression.” *Biometrika*, **103**(3), 579–593. doi:10.1093/biomet/asw036.

Tuddenham RD, Snyder MM (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.

Zhang X, Li L (2017). “Tensor Envelope Partial Least-Squares Regression.” *Technometrics*, pp. 1–11. doi:[10.1080/00401706.2016.1272495](https://doi.org/10.1080/00401706.2016.1272495).

Affiliation:

Minji Lee, Zhihua Su
Department of Statistics
University of Florida
102 Griffin-Floyd Hall
Gainesville, FL, United States of America
E-mail: mlee9@ufl.edu, zihuasu@stat.ufl.edu