# Envelope Models and Methods

*Dimension Reduction for Efficient Estimation in Multivariate Statistics*

R. Dennis Cook
School of Statistics
University of Minnesota
Minneapolis, MN 55455, U.S.A.

October 9, 2013

# Contents

# Chapter 3

# Partial Envelopes

A subset of the predictors may often be of special interest in multivariate regression. For instance, some predictors may correspond to treatments while the remaining predictors are included to account for heterogeneity among the experimental units. In such cases the columns of $\boldsymbol{\beta}$ that correspond to treatments will be of particular interest. Partial envelopes, which were proposed and studied by Su and Cook (2011), are designed to focus consideration on the coefficients corresponding to the predictors of interest. The development of partial envelopes closely follows that for envelopes given in Chapter 2, so in this chapter we emphasize connections between the two approaches.

## 3.1 Partial envelope model

Partition $\mathbf{X}$ into two sets of predictors $\mathbf{X}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{p_1}$, $p_1 + p_2 = p$, and conformably partition the columns of $\boldsymbol{\beta}$ into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Then model (2.1) can be rewritten as $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}_1\mathbf{X}_1 + \boldsymbol{\beta}_2\mathbf{X}_2 + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}_1$ corresponds to the coefficients of interest. We can now consider the $\boldsymbol{\Sigma}$-envelope for $\mathcal{B}_1 = \mathrm{span}(\boldsymbol{\beta}_1)$ – $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}_1)$ – leaving $\boldsymbol{\beta}_2$ as an unconstrained parameter. This leads to the parametric structure $\mathcal{B}_1 \subseteq \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}_1)$ and $\boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}_1}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{E}_1} + \mathbf{Q}_{\mathcal{E}_1}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{E}_1}$, where $\mathbf{P}_{\mathcal{E}_1}$ denotes the projection onto $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}_1)$, which is called the partial envelope for $\mathcal{B}_1$. This is the same as the envelope structure discussed in Chapter 2, except the enveloping is on $\mathcal{B}_1$ instead of the larger space $\mathcal{B}$. For emphasis we will henceforth refer to $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ as the full envelope. Because $\mathcal{B}_1 \subseteq \mathcal{B}$, the partial envelope is contained in the full envelope, $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}_1) \subseteq \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, which allows the partial envelope to offer gains that may not be possible with the full envelope because more information may be immaterial to the purpose of estimating $\boldsymbol{\beta}_1$.

Let $u_1 = \dim(\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1))$, let $\mathbf{\Gamma} \in \mathbb{R}^{r\times u_1}$ be a semi-orthogonal matrix for $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1)$, let $(\mathbf{\Gamma}, \mathbf{\Gamma}_0) \in \mathbb{R}^{r\times r}$ be an orthogonal matrix and let $\boldsymbol{\eta} \in \mathbb{R}^{u_1\times p_1}$ be the coordinates of $\boldsymbol{\beta}_1$ in terms of the basis matrix $\mathbf{\Gamma}$. Then the partial envelope model can be written as

$$\mathbf{Y} \;=\; \boldsymbol{\mu} + \mathbf{\Gamma}\boldsymbol{\eta}\mathbf{X}_1 + \boldsymbol{\beta}_2\mathbf{X}_2 + \varepsilon, \;\; \mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T + \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T, \tag{3.1}$$

where $\mathbf{\Omega} \in \mathbb{R}^{u_1\times u_1}$ and $\mathbf{\Omega}_0 \in \mathbb{R}^{(r-u_1)\times(r-u_1)}$.

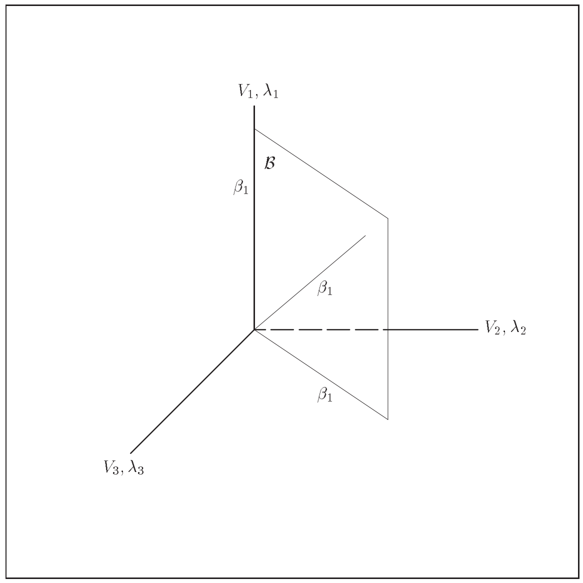


Figure 3.1: Schematic representation of a partial envelope with three possibilities for $\boldsymbol{\beta}_1$

Possibilities for the partial envelope model (3.1) are represented schematically in Figure 3.1 which shows three possibities for $\boldsymbol{\beta}_1$. The axes of the plot represent the eigenvectors $\mathbf{V}_j$, $j = 1, 2, 3$, of $\mathbf{\Sigma}$ with corresponding eigenvalues $\lambda_1 > \lambda_2 \geq \lambda_3$ for a regression having $r = 3$ responses and $p = 2$ predictors with coefficient vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, so $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathbb{R}^{3\times 2}$. The two-dimensional coefficient subspace $\mathcal{B} = \mathrm{span}(\boldsymbol{\beta})$ is depicted as a plane in the plot. If the eigenvalues $\lambda_j$ are distinct then all three eigenvectors of $\mathbf{\Sigma}$ are needed to envelope $\mathcal{B}$ and $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}) = \mathbb{R}^3$ so there is no response reduction. On the other hand if $\lambda_2 = \lambda_3$ then both edges of the plane representing $\mathcal{B}$ correspond to eigenvector of $\mathbf{\Sigma}$ and $\mathcal{B} = \mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$. The partial envelope for $\boldsymbol{\beta}_1$ depends on where it lies relative to the configuration in the plot. If $\boldsymbol{\beta}_1$ falls in $\mathrm{span}(\mathbf{V}_1)$ then $\mathcal{B}_1 = \mathrm{span}(\boldsymbol{\beta}_1) = \mathrm{span}(\mathbf{V}_1) = \mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ is one-dimensional. If $\mathrm{span}(\boldsymbol{\beta}_1) \in \mathrm{span}(\mathbf{V}_2, \mathbf{V}_3)$ then $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1) = \mathcal{B}_1$ if $\lambda_2 = \lambda_3$ and $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1) = \mathrm{span}(\mathbf{V}_2, \mathbf{V}_3)$ if $\lambda_2 \neq \lambda_3$. On the other hand, if $\boldsymbol{\beta}_1$ is the third possibility represented Figure 3.1 and $\lambda_2 \neq \lambda_3$ then $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1) = \mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}) = \mathbb{R}^3$ and there is no partial reduction.

## 3.2 Estimation

In preparation for stating the normal-theory MLE's from this model, let $\widehat{\mathbf{R}}_{1|2} \in \mathbb{R}^{p_1}$ and $\widehat{\mathbf{R}}_{\mathbf{Y}|2} \in \mathbb{R}^r$ denote the residual vectors from the multivariate linear regressions of $\mathbf{X}_1$ on $\mathbf{X}_2$ and $\mathbf{Y}$ on $\mathbf{X}_2$. Let $\mathbf{S}_{\mathbf{Y}|2} \in \mathbb{R}^{r\times r}$ denote the sample covariance matrix of $\widehat{\mathbf{R}}_{\mathbf{Y}|2}$, and let $\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}} \in \mathbb{R}^{r\times r}$ denote the sample covariance matrix of the residuals from the linear regression of $\widehat{\mathbf{R}}_{\mathbf{Y}|2}$ on $\widehat{\mathbf{R}}_{1|2}$. This sample covariance matrix is the same as the sample covariance matrix from the fit of $\mathbf{Y}$ on $\mathbf{X}$, $\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}} = \mathbf{S}_{\mathbf{Y}|\mathbf{X}}$, but we use $\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}$ throughout this chapter to emphasize the partitioned form of the envelope. Justification of this relationship is assigned as a problem a the end of this chapter. Following the general derivation in Section 2.4 of the MLE's for the full envelope model we have the following MLE for $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ given its dimension $u_1$.

$$\begin{aligned}
\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1) &= \arg\min_{\mathcal{S}\in\mathcal{G}(u_1,r)}\{\log|\mathbf{P}_{\mathcal{S}}\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\mathbf{P}_{\mathcal{S}}|_0 + \log|\mathbf{Q}_{\mathcal{S}}\mathbf{S}_{\mathbf{Y}|2}\mathbf{Q}_{\mathcal{S}}|_0\} \qquad (3.2)\\
&= \mathrm{span}\{\arg\min_{\mathbf{G}}\,(\log|\mathbf{G}^T\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\mathbf{G}| + \log|\mathbf{G}_0^T\mathbf{S}_{\mathbf{Y}|2}\mathbf{G}_0|)\}\\
&= \mathrm{span}\{\arg\min_{\mathbf{G}}\,(\log|\mathbf{G}^T\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|2}^{-1}\mathbf{G}|)\},
\end{aligned}$$

where $\min_{\mathbf{G}}$ is over $r \times u_1$ semi-orthognal marices, $(\mathbf{G}, \mathbf{G}_0) \in \mathbb{R}^{r\times r}$ is an orthogonal matrix and $\mathcal{G}(u_1, r)$ still denotes the Grassmann manifold of dimension $u_1$ in $\mathbb{R}^r$. The function to be optimized is of the general form as that given in (2.15) for the full envelope model. Following determination of $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1)$, the MLE's of the remaining parameters are as follows. The MLE $\widehat{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ is the projection onto $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ of the estimator of $\boldsymbol{\beta}_1$ from the standard model. The MLE $\widehat{\boldsymbol{\beta}}_2$ of $\boldsymbol{\beta}_2$ is the coefficient matrix from the ordinary least squares fit of the residuals $\mathbf{Y} - \bar{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_1\mathbf{X}_1$ on $\mathbf{X}_2$. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal then $\widehat{\boldsymbol{\beta}}_2$ reduces to the MLE of $\boldsymbol{\beta}_2$ from the standard model. Let $\widehat{\mathbf{\Gamma}}$ be a semi-orthogonal basis matrix for $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ and let $(\widehat{\mathbf{\Gamma}}, \widehat{\mathbf{\Gamma}}_0)$ be an orthogonal matrix. The MLE $\widehat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ is then

$$\begin{aligned}
\widehat{\mathbf{\Sigma}} &= \widehat{\mathbf{P}}_{\mathcal{E}_1}\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\widehat{\mathbf{P}}_{\mathcal{E}_1} + \widehat{\mathbf{Q}}_{\mathcal{E}_1}\mathbf{S}_{\mathbf{Y}|2}\widehat{\mathbf{Q}}_{\mathcal{E}_1} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Omega}}\widehat{\mathbf{\Gamma}}^T + \widehat{\mathbf{\Gamma}}_0\widehat{\mathbf{\Omega}}_0\widehat{\mathbf{\Gamma}}_0^T\\
\widehat{\mathbf{\Omega}} &= \widehat{\mathbf{\Gamma}}^T\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\widehat{\mathbf{\Gamma}}\\
\widehat{\mathbf{\Omega}}_0 &= \widehat{\mathbf{\Gamma}}_0^T\mathbf{S}_{\mathbf{Y}|2}\widehat{\mathbf{\Gamma}}_0,
\end{aligned}$$

where $\widehat{\mathbf{P}}_{\mathcal{E}_1}$ denotes the projection operator for $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1)$.

Comparing the MLE of $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$ in (2.15) with the MLE of $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ in (3.2) we see that the two estimators differ only by the inner product matrices: The MLE of the full envelope $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$ is constructed using $\mathbf{S}_{\mathrm{res}}$ and $\mathbf{S}_{\mathbf{Y}}$, while the MLE of the partial envelope $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ uses $\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}$ and $\mathbf{S}_{\mathbf{Y}|2}$. We see from this correspondence that $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}_1)$ is the same as the estimator of the full envelope applied in the context of the working model $\widehat{\mathbf{R}}_{\mathbf{Y}|2} = \boldsymbol{\beta}_1\widehat{\mathbf{R}}_{1|2} + \varepsilon^*$. The material and immaterial parts of $\mathbf{Y}$ in the partial envelope model can be interpreted as the material and immaterial parts of $\mathbf{Y}$ in the full envelope model $\mathbf{R}_{\mathbf{Y}|2} = \boldsymbol{\beta}_1\mathbf{R}_{1|2} + \boldsymbol{\varepsilon}$, where $\mathbf{R}_{\mathbf{Y}|2}$ and $\mathbf{R}_{1|2}$ denote the population versions of $\widehat{\mathbf{R}}_{\mathbf{Y}|2}$ and $\widehat{\mathbf{R}}_{1|2}$. In particular, $\widehat{\mathbf{P}}_{\mathcal{E}_1}\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}}\widehat{\mathbf{P}}_{\mathcal{E}_1}$ is the estimated covariance matrix for the part of $\mathbf{Y}$ that is material to the estimation of $\boldsymbol{\beta}_1$ and $\widehat{\mathbf{Q}}_{\mathcal{E}_1}\mathbf{S}_{\mathbf{Y}|2}\widehat{\mathbf{Q}}_{\mathcal{E}_1}$ is the estimated covariance matrix for the part of $\mathbf{Y}$ that is immaterial to the estimation of $\boldsymbol{\beta}_1$.

### 3.2.1 Asymptotic distribution of $\widehat{\boldsymbol{\beta}}_1$

The correspondence between partial and full envelopes carries over to asymptotic variances as well. Partition $\mathbf{\Sigma}_{\mathbf{X}} = (\mathbf{\Sigma}_{\mathbf{X}}^{(ij)})$ and $\mathbf{\Sigma}_{\mathbf{X}}^{-1} = (\mathbf{\Sigma}_{\mathbf{X}}^{-(ij)})$ according to the partitioning of $\mathbf{X}$ $(i,j = 1,2)$ and let $\mathbf{\Sigma}_{\mathbf{X}}^{(1|2)} = \mathbf{\Sigma}_{\mathbf{X}}^{(11)} - \mathbf{\Sigma}_{\mathbf{X}}^{(12)}\mathbf{\Sigma}_{\mathbf{X}}^{-(22)}\mathbf{\Sigma}_{\mathbf{X}}^{(21)}$. The matrix $\mathbf{\Sigma}_{\mathbf{X}}^{(1|2)}$ is constructed in the same way as the covariance matrix for the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ when $\mathbf{X}$ is normally distributed, although here $\mathbf{X}$ is fixed. Define

$$\mathbf{U}_{1|2} = \boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{X}}^{(1|2)}\boldsymbol{\eta}^T \otimes \mathbf{\Omega}_0^{-1} + \mathbf{\Omega} \otimes \mathbf{\Omega}_0^{-1} + \mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}_0 - 2\mathbf{I}_{u_1(r-u_1)}.$$

The limiting distribution of $\widehat{\boldsymbol{\beta}}_1$ is stated in the following proposition. (See Su and Cook, 2011, for justification and for the limiting distribution of $\widehat{\boldsymbol{\beta}}_2$.)

**Proposition 3.1** *Under the partial envelope model (3.1)* $n^{1/2}\{\mathrm{vec}(\widehat{\boldsymbol{\beta}}_1) - \mathrm{vec}(\boldsymbol{\beta}_1)\}$ *converge in distribution to a normal random vector with mean 0 and covariance matrix*

$$\begin{aligned}\mathrm{avar}\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_1)\} &= \mathbf{\Sigma}_{\mathbf{X}}^{-(1|2)} \otimes \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T + (\boldsymbol{\eta}^T \otimes \mathbf{\Gamma}_0)\mathbf{U}_{1|2}^{\dagger}(\boldsymbol{\eta} \otimes \mathbf{\Gamma}_0^T),\\ &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{1\mathbf{\Gamma}}) + \mathrm{avar}[n^{1/2}\mathrm{vec}(\mathbf{Q}_{\mathbf{\Gamma}}\widehat{\boldsymbol{\beta}}_{1\boldsymbol{\eta}})],\end{aligned}$$

*where* $\widehat{\boldsymbol{\beta}}_{1\boldsymbol{\eta}}$ *and* $\widehat{\boldsymbol{\beta}}_{1\mathbf{\Gamma}}$ *are the maximum likelihood estimators of* $\boldsymbol{\beta}_1$ *when* $\mathbf{\Gamma}$ *and* $\boldsymbol{\eta}$ *are known,* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{1\mathbf{\Gamma}})] = \mathbf{\Sigma}_{\mathbf{X}}^{-(1|2)} \otimes \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T$ *and* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\mathbf{\Gamma}}\widehat{\boldsymbol{\beta}}_{1\boldsymbol{\eta}})]$ *is defined implicitly.*

The form of the asymptotic variance of $\widehat{\boldsymbol{\beta}}_1$ given in this proposition is identical to the form of the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ from the full envelope model given in Section 2.5, although the definitions of components are different. For instance, $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}})]$ requires $\mathbf{\Sigma}_{\mathbf{X}}$, while $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_1)]$ uses $\mathbf{\Sigma}_{\mathbf{X}}^{(1|2)}$ in its place. Several additional characteristics of this proposition may be of interest. First consider regressions in which $\mathbf{\Sigma}_{\mathbf{X}}^{(12)} = 0$.

This will arise when $\mathbf{X}_1$ and $\mathbf{X}_2$ are asymptotically uncorrelated or have been chosen by design to be orthogonal. Because $\mathbf{X}$ is non-random, this condition can always be forced without an inferential cost by replacing $\mathbf{X}_1$ with $\widehat{\mathbf{R}}_{1|2}$. This replacement alters the definition of $\boldsymbol{\beta}_2$ but does not alter the parameter of interest $\boldsymbol{\beta}_1$. When $\boldsymbol{\Sigma}_{\mathbf{X}}^{(12)} = 0$, $\mathrm{avar}\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_2)\} = \boldsymbol{\Sigma}_{\mathbf{X}}^{(22)} \otimes \boldsymbol{\Sigma}^{-1}$, which is the same as the asymptotic covariance matrix for the estimator of $\boldsymbol{\beta}_2$ from the standard model. And $\mathrm{avar}\{n^{1/2}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_1)\}$ reduces to the asymptotic covariance matrix for the full envelope estimator of $\boldsymbol{\beta}_1$ in the model $\widehat{\mathbf{R}}_{\mathbf{Y}|2} = \boldsymbol{\beta}_1\mathbf{X}_1 + \boldsymbol{\varepsilon}^*$. No longer requiring that $\boldsymbol{\Sigma}_{\mathbf{X}}^{(12)} = 0$, we can carry out asymptotic inference for $\boldsymbol{\beta}_1$ based on a partial envelope by using the full envelope for $\boldsymbol{\beta}_1$ in the model $\widehat{\mathbf{R}}_{\mathbf{Y}|2} = \boldsymbol{\beta}_1\widehat{\mathbf{R}}_{1|2} + \boldsymbol{\varepsilon}^*$. If we choose $\boldsymbol{\beta}_1 = \boldsymbol{\beta}$, so $\mathbf{X}_1 = \mathbf{X}$, $\boldsymbol{\beta}_2$ is nil and $\boldsymbol{\Sigma}_{\mathbf{X}}^{(1|2)} = \boldsymbol{\Sigma}_{\mathbf{X}}$, then $\mathrm{avar}\{n^{1/2}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_1)\}$ reduces to the asymptotic covariance matrix for the full envelope estimator of $\boldsymbol{\beta}$ given in Section 2.5.

### 3.2.2 Selecting $u_1$

The methods discussed in Section 2.6 for selecting the dimension $u$ of the full envelope can be adapted straightforwardly for the selection of the dimension $u_1$ of the partial envelope.

To use a sequence of likelihood ratio tests, the hypothesis $u_1 = u_{10}$, $u_{10} < p_1$, can be tested by using the likelihood ratio statistic $\Lambda(u_{10}) = 2\{\widehat{L}_{\mathrm{sm}} - \widehat{L}(u_{10})\}$, where $\widehat{L}_{\mathrm{sm}}$ is the maximized log likelihood for the standard model given in Section 2.6 and

$$\widehat{L}(u_{10}) = -(nr/2)\{1 + \log(2\pi)\} - (n/2)\log|\widehat{\boldsymbol{\Gamma}}^T\mathbf{S}_{\mathbf{R}|2}\widehat{\boldsymbol{\Gamma}}| - (n/2)\log|\widehat{\boldsymbol{\Gamma}}^T\mathbf{S}_{\mathbf{Y}|2}^{-1}\widehat{\boldsymbol{\Gamma}}|$$

denotes the maximum value of the likelihood for the partial envelope model with $u_1 = u_{10}$. When $u_1 = p_1$, $\widehat{L}(u_{10}) = \widehat{L}_{\mathrm{sm}}$. Following standard likelihood theory, under the null hypothesis $\Lambda(u_{10})$ is distributed asymptotically as a chi-squared random variable with $p_1(r - u_{10})$ degrees of freedom. This test statistic can be used in a sequential scheme to choose $u_1$ as described in Section 2.6.

The dimension of the partial envelope could also be selected by using an information criterion:

$$\widehat{u}_1 = \arg\min_{u_1}\{-2\widehat{L}(u_1) + h(n)N_{u_1}\},$$

where $h(n)$ is equal to $\log n$ for Bayes information criterion and is equal to 2 for Akaike's information criterion, and $N_{u_1}$ is the number of real parameters in the partial envelope model,

$$N_{u_1} = r + u_1(r - u_1) + u_1p_1 + rp_2 + u_1(u_1 + 1)/2 + (r - u_1)(r - u_1 + 1)/2. \quad (3.3)$$

Subtracting $N_{u_1}$ from the number of parameters $r+pr+r(r+1)/2$ for the standard model gives the degrees of freedom for $\Lambda(u_1)$ mentioned previously.

## 3.3 Illustrations

### 3.3.1 Egyptian skulls II

We use a new dataset on Egyptian skulls to illustrate partial envelopes. These data, which are different from those considered in Section 2.10.3 but were also collated from Thomson, A. and Randall-Maciver, R. (1905), comprise observations 300 skulls. The $5 \times 1$ response vector $\mathbf{Y}$ consists of

- BH: Basibregmatic Height of Skull
- BL: Basialveolar Length of Skull
- MB: Maximal Breadth of Skull
- NH: Nasal Height of Skull
- OL: Ophryp-Occipital Length

The first four of these are the same as those used previously in Section 2.10.3. The predictors reflect the gender and epoch from which a skull was obtained. Three epochs were used: Pre-dyanstic with 60 observations per gender, 6th-12th Dynasties with 30 observations per gender and Ptolematic and Roman periods with 60 observations per gender. The multivariate linear model is thus

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_3 X_3 + \boldsymbol{\varepsilon}, \tag{3.4}$$

where $X_1 = 1$ for females and 0 for males, $X_2 = 1$ for skulls from the 6th-12th Dynasties and 0 otherwise, and $X_3 = 1$ for skulls from the Ptolematic or Roman periods and 0 otherwise. The Pre-dyanstic period is then indicated when $X_2 = 0$ and $X_3 = 0$.

The top half of Table 3.1 gives the OLS coefficient estimates along with their standard errors in parentheses. For contrast, we consider two partial envelope models

$$\begin{aligned}
\mathbf{Y} &= \boldsymbol{\alpha} + \boldsymbol{\Gamma}\eta_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_3 X_3 + \boldsymbol{\varepsilon} \\
\mathbf{Y} &= \boldsymbol{\alpha} + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\Gamma}\eta_2 X_2 + \boldsymbol{\Gamma}\eta_3 X_3 + \boldsymbol{\varepsilon},
\end{aligned}$$

Gender is the predictor of interest in the first model; the envelope estimator if its coefficient vector is shown in the second column of the lower half of Table 3.1. The coefficient of NH is the only one that exhibited a notable change. The standard errors have all been reduced, although the reduction is modest. The epoch indicators are presumed to be of interest in the second model. The envelope estimators of coefficients of the epoch indicators are shown in the third and fourth column of the lower half of Table 3.1. In this case there are appreciable chances in both the coefficients and their standard errors.

Table 3.1: Coefficients and standard errors in parentheses for fits to the Egyptian skull data. Top half: OLS analysis. Bottom half: partial envelope estimators of $\boldsymbol{\beta}_1$ and $(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ based on separate partial fits.

| $\mathbf{Y}$ | $\widehat{\boldsymbol{\beta}}_1$ | $\widehat{\boldsymbol{\beta}}_2$ | $\widehat{\boldsymbol{\beta}}_3$ |
|---|---|---|---|
| OLS fit of (3.4) | | | |
| BH | -3.68 (0.52) | 1.02 (0.71) | -1.10 (0.58) |
| BL | -3.75 (0.55) | -2.52 (0.75) | -3.06 (0.61) |
| MB | -2.97 (0.55) | 0.64 (0.75) | 2.69 (0.61) |
| NH | -2.72 (0.36) | 0.77 (0.50) | 1.45 (0.61) |
| OL | -6.20 (0.66) | -2.75 (0.90) | -0.25 (0.73) |
| Partial envelope estimators | | | |
| BH | -3.70 (0.43) | 1.97 (0.55) | -0.78 (0.46) |
| BL | -3.76 (0.46) | -1.50 (0.60) | -2.58 (0.46) |
| MB | -2.87 (0.45) | 1.43 (0.67) | 3.05 (0.53) |
| NH | -1.75 (0.27) | 1.08 (0.48) | 1.59 (0.39) |
| OL | -6.48 (0.60) | -1.08 (0.42) | 0.28 (0.36) |

### 3.3.2 Mens' urine

In this illustration we consider a set of data on the relationship between the composition of a man's urine and his weight (Smith, H. Gnanadesikan, R. and Hughes, J. B. (1962). Multivariate analysis of variance (MANOVA). *Biometrika* **18**, 22-41.) The eight responses $\mathbf{Y}$ that we use for this illustration are the concentrations of phosphate, calcium, phosphorus, creatinine and chloride in mg/ml, and boron, choline and copper in $\mu$g/ml. These responses plus two covariates, volume (ml) and $(\text{specificgravity} - 1) \times 10^3$, were measured on 45 men in four weight categories. The multivariate analysis of covariance model that Smith et al. (1962) used for these data is

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\alpha} + \boldsymbol{\beta}_1 \mathbf{X}_1 + \boldsymbol{\beta}_2 \mathbf{X}_2 + \boldsymbol{\varepsilon} \\ &= \boldsymbol{\alpha} + \boldsymbol{\beta}_{11} X_{11} + \boldsymbol{\beta}_{12} X_{12} + \boldsymbol{\beta}_{13} X_{13} + \boldsymbol{\beta}_{21} X_{21} + \boldsymbol{\beta}_{22} X_{22} + \boldsymbol{\varepsilon}, \end{aligned} \tag{3.5}$$

where $\mathbf{X}_1 = (X_{11}, X_{12}, X_{13})^T$, $X_{1k} = 1$ for category $k$ and $0$ otherwise, $k = 1, 2, 3$, and the covariates are represented by $\mathbf{X}_2 = (X_{21}, X_{22})^T$. The primary question of interest is whether the mean response varies with the weight category. In terms of the model, the question is whether or not there is evidence to indicate that elements of $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{12}, \boldsymbol{\beta}_{13})$ are non-zero. This is the kind of question for which partial envelopes are appropriate.

The assumption of normal errors tends to be most important with small data sets, as is the case here. Consequently, we first consider if coordinate-wise power transformations of the response vector might move it closer to normality. The likelihood ratio test (Cook, R. D. and Weisberg, S. (1998), ch. 13) that no transformation is required has a $p$-value of essentially 0, while the likelihood ratio test that all responses should be in the log scale has a $p$-value of $0.14$. Consequently, we use log concentrations as the responses.

Table 3.2: Ratios of coefficient estimates to their standard errors for the urine data. Left half: OLS ratios for the elements of $\boldsymbol{\beta}_1$. Right half: partial envelope ratios for the elements of $\boldsymbol{\beta}_1$.

| $\mathbf{Y}$ | $\mathbf{B}_{11}$ | $\mathbf{B}_{12}$ | $\mathbf{B}_{13}$ | $\widehat{\boldsymbol{\beta}}_{11}$ | $\widehat{\boldsymbol{\beta}}_{12}$ | $\widehat{\boldsymbol{\beta}}_{13}$ | $\widehat{\boldsymbol{\Gamma}}$ |
|---|---|---|---|---|---|---|---|
| | OLS ratios | | | Partial envelope ratios | | | |
| phosphate | -3.86 | -2.86 | -2.44 | -3.84 | -3.47 | -3.29 | -0.464 |
| calcium | -1.35 | -2.30 | -3.31 | -2.59 | -2.47 | -2.40 | -0.524 |
| phosphorus | -3.93 | -2.22 | -1.78 | -4.17 | -3.71 | -3.49 | -0.450 |
| creatinine | -3.14 | -2.41 | -2.29 | -3.69 | -3.36 | -3.20 | -0.269 |
| chloride | 4.08 | 3.26 | 1.76 | 2.95 | 2.78 | 2.68 | 0.356 |
| boron | 1.04 | -0.01 | -0.35 | 0.32 | 0.32 | 0.32 | 0.054 |
| choline | -1.21 | 1.05 | 0.73 | 0.30 | 0.30 | 0.30 | 0.028 |
| copper | -2.64 | -1.48 | -0.29 | -3.21 | -2.98 | -2.87 | 0.327 |

The dimension selection criteria AIC, BIC and LRT(0.05) all indicated that $u = 3$ for fitting the full envelope model to these data. Consequently, we would normally expect that $u_1 < 3$ when fitting the partial envelope model (3.1). As it turned out AIC, BIC and LRT(0.05) as described in Section 3.2.2 all indicated that $u_1 = 1$, so in effect only a single linear combination of the responses responds to mens' weight given the two covariates. The corresponding envelope model is

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}_{11}X_{11} + \boldsymbol{\Gamma}\boldsymbol{\eta}_{12}X_{12} + \boldsymbol{\Gamma}\boldsymbol{\eta}_{13}X_{13} + \boldsymbol{\beta}_{21}X_{21} + \boldsymbol{\beta}_{22}X_{22} + \boldsymbol{\varepsilon}. \qquad (3.6)$$

Table 3.2 shows the ratios of the estimated elements of $\boldsymbol{\beta}_1$ to their standard errors for the OLS and partial envelope (with $u_1 = 1$) fits of model (3.5). The estimated basis $\widehat{\boldsymbol{\Gamma}}$ for $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}_1)$ is shown in the last column. The results in this table suggest that in this illustration the partial envelope model has strengthened and clarified the conclusions indicated by the OLS fit. In particular, except for boron and choline, the mean responses of the three indicated weight categories all differ from weight category 4, the basal category in the parameterization of model (3.5).

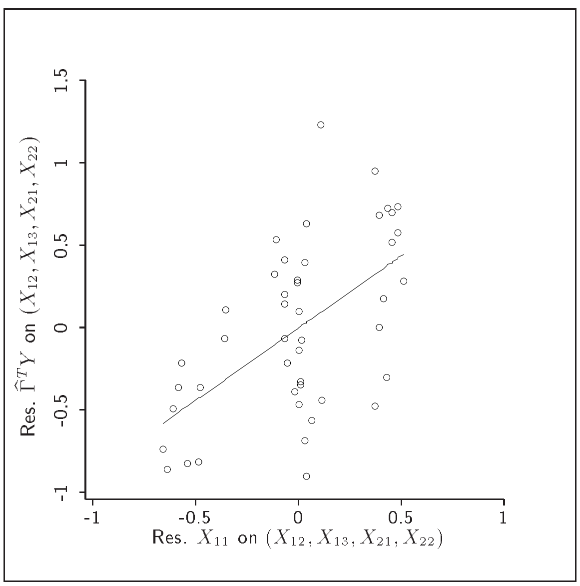


Figure 3.2: Added variable plot for $X_{11}$ in the urine data. The line is the OLS fit of the plotted data.

According to the discussion in Section 3.2 we may be able to gain insights into the regression by considering the working model $\widehat{\boldsymbol{\Gamma}}^T\widehat{\mathbf{R}}_{\mathbf{Y}|2} = \boldsymbol{\eta}_1\widehat{\mathbf{R}}_{1|2} + \text{error}$, where $\widehat{\boldsymbol{\Gamma}}^T\widehat{\mathbf{R}}_{\mathbf{Y}|2}$ is a scalar. For instance, the added variable plot in Figure 3.2 gives visual representation of the contribution of first category after accounting for the other categories and the two covariates. The line on the plot is OLS fit of the vertical axis residuals on the horizontal axis residuals. The slope of the line 0.878 is equal to $\widehat{\boldsymbol{\eta}}_{11}$ from the fit of the partial envelope model (3.6).

## 3.4 Partial envelopes for prediction

In Section 2.7 we described the prediction of the response vector at a new value $\mathbf{X}_{\text{new}}$ of $\mathbf{X}$ as $\widehat{\mathbf{Y}} = \widehat{\boldsymbol{\beta}}\mathbf{X}_{\text{new}}$, where $\widehat{\boldsymbol{\beta}}$ is the envelope estimator of $\boldsymbol{\beta}$ based on the full envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ for model (2.1). In this section we use partial envelopes to envelop $\boldsymbol{\beta}\mathbf{X}_{\text{new}}$, which leads to a new method of prediction that has the potential to yield predicted and fitted values with smaller variation than the standard predictions based on model (2.1) or the full envelope predictions described in Section 2.7.

Select an $\mathbf{A}_0 \in \mathbb{R}^{p\times(p-1)}$ so that $\mathbf{A} = (\mathbf{X}_{\text{new}}, \mathbf{A}_0) \in \mathbb{R}^{p\times p}$ has full rank. It may be helpful, although not necessary, to choose the columns of $\mathbf{A}_0$ to be orthogonal to $\mathbf{X}_{\text{new}}$. Let $\boldsymbol{\phi}_1 = \boldsymbol{\beta}\mathbf{X}_{\text{new}}$, $\boldsymbol{\phi}_2 = \boldsymbol{\beta}\mathbf{A}_0$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ and $\mathbf{Z} = \mathbf{A}^{-1}\mathbf{X} = (Z_1, \mathbf{Z}_2^T)^T$, where $\mathbf{Z}_2 \in \mathbb{R}^{p-1}$. Then we can parameterize model (2.1) as

$$\begin{aligned}
\mathbf{Y} &= \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon} \\
&= \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{A}\mathbf{A}^{-1}\mathbf{X} + \boldsymbol{\varepsilon} \\
&= \boldsymbol{\alpha} + (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)\mathbf{Z} + \boldsymbol{\varepsilon} \\
&= \boldsymbol{\alpha} + \boldsymbol{\phi}_1 Z_1 + \boldsymbol{\phi}_2\mathbf{Z}_2 + \boldsymbol{\varepsilon}
\end{aligned}$$

We can now parameterize the model in terms of a basis $\boldsymbol{\Gamma}$ for the $\boldsymbol{\Sigma}$-envelope of $\operatorname{span}(\boldsymbol{\phi}_1)$, leading to a partial envelope representation for prediction at $\mathbf{X}_{\text{new}}$:

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}Z_1 + \boldsymbol{\phi}_2\mathbf{Z}_2 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T. \tag{3.7}$$

At this point we can apply any of the methods discussed previously in this chapter.

The re-parameterization $\boldsymbol{\beta} \mapsto \boldsymbol{\phi}$ is full rank by construction so $\mathcal{B} = \operatorname{span}(\boldsymbol{\phi})$ and thus it does not change the envelope, $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\operatorname{span}(\boldsymbol{\phi}))$. However, it does change the coordinate system for moving inside $\mathcal{B}$ as represented in Figure 3.1. As $\mathbf{X}_{\text{new}}$ changes, $\boldsymbol{\phi} = \boldsymbol{\beta}\mathbf{X}_{\text{new}}$ will change locations within $\mathcal{B}$ and the discussion of Figure 3.1 in terms of $\boldsymbol{\beta}_1$ applies also to $\boldsymbol{\phi}_1$. For instance, if for some value of $\mathbf{X}_{\text{new}}$, $\operatorname{span}(\boldsymbol{\phi}_1) = \operatorname{span}(\mathbf{V}_1)$ then we can expect the partial envelope for $\boldsymbol{\phi}_1$ to yield a better prediction at $\mathbf{X}_{\text{new}}$ than that based on the standard model or on the full envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$. A practical implication of this discussion is that we may wish to do a partial envelope analysis for each prediction.

## 3.5 Pulp fibers

This dataset, which is from Johnson, R. and Wichern, D. (2007) ( *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ: Prentice Hall) and was used by Cook and Su (2011) to introduce partial envelopes, is on properties of pulp fibers and the paper made from them. There are 62 measurements on four paper properties that form the responses: breaking length, elastic modulus, stress at failure and burst strength. Three aspects of the pulp fiber are the predictors: arithmetic fiber length, long fiber fraction and fine fiber fraction. We first review the findings of Cook and Su (2011) and then turn to prediction envelopes.

Fitting the full envelope model to all three predictors, BIC and LRT(0.01) both indicated that $u = 2$. The ratios of the standard deviations of the element of $\mathbf{B}$ to those of the envelope estimator $\widehat{\boldsymbol{\beta}}$ with $u = 2$ range from 0.98 to 1.10, with an average of 1.03, so the envelope model does not bring much reduction. The reason is that the immaterial variation is small relative to the material variation.

Suppose that we are particularly interested in how the fine fiber fraction affects paper quality. Fitting the partial envelope model to the column of $\boldsymbol{\beta}$ that corresponds to fine fiber fraction. BIC and LRT(0.05) both select $u = 1$. The standard deviation ratios between the full envelope model and the partial envelope model for the four elements in the column of $\boldsymbol{\beta}$ that corresponds to fine fiber fraction range between about 66 and 7. This reduction comes about because the immaterial variation is now much larger than the material variation.

Table 3.3: Predicted values at $\mathbf{X}_{\text{new}}$ for the pulp fiber data. Envelope prediction is from the full envelope model wiith $u = 2$. Partial envelope prediction is from model (3.7) with $u = 1$. "SE fit " is the standard error for a fitted value, and "SE predict" is the standard error for a predicted value of the response vector.

| | Envelope prediction | | | Partial envelope prediction | | |
|---|---|---|---|---|---|---|
| $\mathbf{Y}$ | $\widehat{\mathbf{Y}}$ | SE fit | SE predict | $\widehat{\mathbf{Y}}$ | SE fit | SE predict |
| Breaking length | 20.112 | 1.75 | 2.61 | 20.159 | 0.26 | 2.04 |
| Elastic modulus | 6.917 | 0.45 | 0.71 | 6.910 | 0.10 | 0.56 |
| Stress at failure | 4.791 | 0.82 | 1.24 | 4.802 | 0.15 | 0.97 |
| Burst strength | 0.616 | 0.38 | 0.58 | 0.619 | 0.07 | 0.45 |

Turning to prediction, consider predicting $Y$ at $\mathbf{X}_{\text{new}} = (39.24, -0.19, 28.03)^T$, which corresponds to the tenth observation in the data. Fitting the partial envelope model (3.7), both BIC and LRT(0.01) suggest that $u = 1$. Table 3.3 gives the predictions of the elements of $\mathbf{Y}$ from the full envelope model and the partial envelope model (3.7) along with the standard errors of the fitted values and actual predictions. The standard errors from the full envelope model were determined by plugging in sample versions of the parameters in (2.24), using $n^{-1}\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}\mathbf{X})]$ for the fitted values and $n^{-1}\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}\mathbf{X})] + \boldsymbol{\Sigma}$ for the predictions. Standard errors for the fitted values predictions from the partial model were determined similarly. It can be seen from the table that fitted values change little, while the standard errors change appreciably.

## Problems

**Problem 3.1** *In reference to the discussion of Section 3.2, derive the MLE* $\widehat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\mathcal{B}_1)$.

**Problem 3.2** *Describe how the parameter count* $N_{u_1}$ *in (3.3) arises.*

**Problem 3.3** *Discuss how the code in the Matlab file predict-env2.m might be improved statistically (not computationally). Use envelopes and partial envelopes for predicting the response for the first and fourth cases in the data. Comment on the differences in the predictions for the two methods.*

**Problem 3.4** *Apply partial envelopes to the air pollution data of Chapter 2, first for* $S$ *and then for* $W$.

**Problem 3.5** *Show that* $\mathbf{S}_{\mathbf{R}_{Y|2}|\mathbf{R}_{1|2}} = \mathbf{S}_{\mathbf{Y}|\mathbf{X}}$.