Received: 19 May 2019 | Revised: 24 June 2019 | Accepted: 27 July 2019

DOI: 10.1002/wics.1484

OVERVIEW

WIREs COMPUTATIONAL STATISTICS WILEY

Envelope methods

R. Dennis Cook

School of Statistics, University of Minnesota, Minneapolis, Minnesota

Correspondence
R. Dennis Cook, School of Statistics, University of Minnesota, Minneapolis, MN.
Email: dennis@stat.umn.edu

Abstract

An envelope is a relatively new construct for decreasing estimative and predictive variation relative to standard methods in multivariate statistics, sometimes by amounts equivalent to increasing the sample size many times over. Essentially a form of targeted dimension reduction that is descendent from sufficient dimension reduction, an envelope inherits its underlying philosophy from Fisher's notion of sufficient statistics. The initial development of envelope methods took place largely in the context of the multivariate linear model, resulting in response envelopes for response reduction, predictor envelopes for predictor reduction, simultaneous envelopes for response and predictor reduction and partial envelopes for specialized considerations, each demonstrating a potential for substantial reduction in estimative variation. These advances demonstrated that there are close connections between envelopes and some standard multivariate methods like partial least squares regression and canonical correlations. Subsequently, envelope methodology has been adapted and extended to diverse areas, including envelopes for regressions with matrix and tensor-valued responses, envelopes for spatial statistics, quantile envelopes for quantile regression, Bayesian response envelopes. Sparse versions of response and predictor envelopes have also been developed. More generally, there is also envelope methodology for reducing the variation in any asymptotically normal vector-valued estimator. These advances have opened a new chapter in multivariate statistics, allowing variation that is material to the goals of an analysis to be separated effectively from immaterial variation that serves only to confound estimation and prediction.

This article is categorized under:

Statistical Models > Multivariate Models
Statistical and Graphical Methods of Data Analysis > Multivariate Analysis
Statistical and Graphical Methods of Data Analysis > Dimension Reduction

KEYWORDS

canonical correlation, Inner envelopes, partial least squares regression, predictor envelopes, reduced-rank regression, response envelopes, sufficient dimension reduction

1 | INTRODUCTION

The goal of envelope methodology is to reduce estimative and predictive variation relative to standard multivariate statistical methods. Essentially a form of targeted dimension reduction, envelopes separate information in the data that is material to goals of the analysis from that which is immaterial, and in doing so guard against extraneous variation that could otherwise confound estimation and prediction. While the amount of variance reduction can vary from application to application, studies over the past several years have demonstrated repeatedly that envelope methodology frequently produces substantial gains relative to standard methods, sometimes by amounts equivalent to increasing the sample size many times over. The philosophy of dimension reduction that underlies envelopes follows an historical trend that began with Fisher's seminal 1922 paper, "On the mathematical foundations of theoretical statistics," which initiated the modern era of Statistics.

Viewed broadly, dimension reduction is a huge area. It touches on all of the applied sciences, is a leitmotiv of statistics and is an active research area today. The general notion of reducing data was in the statistical lexicon in 1922, but the community then lacked adequate foundations for effectively advancing complex ideas. After discussing the importance of theoretical foundations for the development of statistical methodology, Fisher (1922) set a fundamental goal of Statistics to be, using contemporary vernacular, dimension reduction without loss of information:

> Since the number of independent facts supplied in the data is usually far greater than the number of facts sought, much of the information supplied by any actual sample is irrelevant. It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data.

According to Fisher, we should start the statistical process of extracting relevant information from a sample $\mathcal{D}=\{Y_1,\ldots,Y_n\}$ by specifying the underlying model up to parameters $\boldsymbol{\theta}\in\Theta$. Then a statistic $t(\mathcal{D})$ is said to be sufficient for $\boldsymbol{\theta}$ if the distribution of $\mathcal{D}$ given t does not depend on $\boldsymbol{\theta}$: $\mathcal{D}\mid(t,\boldsymbol{\theta}=\boldsymbol{\theta}_1)\sim\mathcal{D}\mid(t,\boldsymbol{\theta}=\boldsymbol{\theta}_2)\ \forall\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta$, where ~ means identically distributed. This statement is a formal expression of the idea that t captures all the information about $\boldsymbol{\theta}$ that is contained in $\mathcal{D}$. In particular, if we know t, then we can replace $\mathcal{D}$ with $t(\mathcal{D})$ without loss of information on $\boldsymbol{\theta}$. This was seen as a brilliant idea at the time and considerable effort was devoted to the study of sufficiency for the next 40 years. But by the 1970's Fisher's sufficiency had generally fallen out of favor as a paradigm for guiding methodological studies because of its dependence on a known model and the increasingly complex nature of models. Nevertheless, Fisher's idea, which Cox and Mayo (2010) called "reduction by sufficiency," has a clear legacy in dimension reduction, a legacy that leads to envelopes.

1.1 | Sufficient reductions: Fisher's sufficiency legacy

Consider a pair of vectors, $\mathbf{X}\in\mathbb{R}^p$ and $\mathbf{Y}\in\mathbb{R}^r$, at least one of which is stochastic. We wish to reduce the dimension of $\mathbf{X}$ to facilitate subsequent study. The notation $A \perp\!\!\!\perp B \mid C$ means that the random variables A and B are independent given any value for C, while $A \perp\!\!\!\perp B$ indicates marginal independence. The following is a general definition of dimension reduction that is guided by Fisher's notion of sufficiency.

Definition 1 (Cook, 2007) *A reduction* $R:\mathbb{R}^p\to\mathbb{R}^q$ $q\le p$ *of* $\mathbf{X}$ *is sufficient for* $\mathbf{Y}$ *if at least one of the following hold*

1. $\mathbf{X}\mid(\mathbf{Y}=\mathbf{y}_1, R(\mathbf{X}))\sim\mathbf{X}\mid(\mathbf{Y}=\mathbf{y}_2, R(\mathbf{X})),\ \forall\mathbf{y}_1,\mathbf{y}_2$ *in the space of* $\mathbf{Y}$,
2. $\mathbf{Y}\mid\mathbf{X}\sim\mathbf{Y}\mid R(\mathbf{X})$,
3. $\mathbf{Y}\perp\!\!\!\perp\mathbf{X}\mid R(\mathbf{X})$.

Statement 1 requires $\mathbf{X}$ to be random, but not $\mathbf{Y}$. If we think of $\mathbf{X}$ as the total data (previously denoted as $\mathcal{D}$) and $\mathbf{Y}$ as a parameter vector (previously denoted as $\boldsymbol{\theta}$) then the statement reduces to the requirement for a sufficient statistic; that is, $R(\mathbf{X})$ is then a sufficient statistic. Statement 2 is the classical regression context where the predictors $\mathbf{X}$ are fixed. Here, only $\mathbf{Y}$ needs to be random. Statement 3 requires a joint distribution. In each statement, $R(\mathbf{X})$ is relevant information in Fisher's sense. The three statements are equivalent if $\mathbf{X}$ and $\mathbf{Y}$ have a joint distribution and then we have considerable flexibility in pursuing a reduction. Sufficient dimension reduction is concerned with regressions as in Statement 2, but historically has pursued methodology via Statement 1. These three statements and related constructions form a foundation for modern dimension reduction.

1.2 | Sufficient dimension reduction

Sufficient dimension reduction (SDR) was developed primarily for reducing the dimension of the stochastic predictors $\mathbf{X}\in\mathbb{R}^p$ in univariate regression without requiring a parsimonious model, which stands in contrast to Fisher's original setup for sufficient statistics. It is based on Definition 1, restricting the reduction form to be linear, $R(\mathbf{X}) = \mathbf{\Gamma}^T\mathbf{X}$, where $\mathbf{\Gamma}$ is a $p \times q$ matrix. We then pursue the fewest linear combinations $\mathbf{\Gamma}^T\mathbf{X}$ of $\mathbf{X}$ so that $Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{\Gamma}^T\mathbf{X}$, without assuming a model for the conditional distribution of Y given $\mathbf{X}$. Of course there are assumptions that facilitate the development of methodology, but a Fisherian model is not required.

The statement $Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{\Gamma}^T\mathbf{X}$ holds if and only if $Y \perp\!\!\!\perp \mathbf{X} \mid (\mathbf{\Gamma}\mathbf{A})^T\mathbf{X}$ for any full rank matrix $\mathbf{A}$. Consequently, we will not be able to estimate $\mathbf{\Gamma}$ but only span($\mathbf{\Gamma}$), the subspace of $\mathbb{R}^p$ spanned by the columns of $\mathbf{\Gamma}$. For that reason, some prefer to write the condition in the form $Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{P}_{\mathcal{S}}\mathbf{X}$, where $\mathcal{S}$ is a subspace of $\mathbb{R}^p$, $\mathbf{P}_{\mathcal{S}}$ denotes the projection onto $\mathcal{S}$ and, for future reference, $\mathbf{Q}_{\mathcal{S}} = \boldsymbol{I} - \mathbf{P}_{\mathcal{S}}$ denotes the projection onto the orthogonal complement of $\mathcal{S}$. The intersection of all subspaces for which this holds is called the central subspace (Cook, 1998) and symbolized as $\mathcal{S}_{Y|\mathbf{X}}$. The central subspace has been the target of much theoretical and methodological inquiry over the past 30-ish years, the first SDR methods being based on the equivalence in Definition 1,

$$Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X} \;\text{ if and only if }\; \mathbf{X} \mid \left(Y, \mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X}\right) \sim \mathbf{X} \mid \mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X}.$$

The monograph by B. Li (2018) is a good place to start for those unfamiliar with SDR.

SDR does not require a model for $Y \mid \mathbf{X}$ and that can be a distinct advantage in the model-selection stage of an analysis. On the other hand, it has little to offer when a model is available. For instance, suppose we have a traditional linear regression model, $Y = \alpha + \boldsymbol{\beta}^T\mathbf{X} + \epsilon$, a logistic regression logit$(p) = \alpha + \boldsymbol{\beta}^T\mathbf{X}$ or a Cox model with hazard function $\lambda(t \mid \mathbf{X}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T\mathbf{X})$. In each of these and many other models, $Y \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\beta}^T\mathbf{X}$ and $\mathcal{S}_{Y|\mathbf{X}} = \text{span}(\boldsymbol{\beta})$. SDR is then of little help because it leads us to do what we would have done anyway—estimate $\boldsymbol{\beta}$ without any assistance from dimension reduction.

2 | RESPONSE ENVELOPES

Envelope methodology can be seen as a descendent of SDR that is designed to sharpen the reductive process and in doing so extend applicability to model-based analyses, while still useful as model-free methodology. To introduce basic ideas, we begin with envelopes for response reduction and later turn to envelopes for predictor reduction. A comprehensive treatment of these and other envelope contexts is available from Cook (2018).

Envelopes were used first by Cook, Li, and Chiaromonte (2007) to facilitate dimension reduction and later developed by them (Cook, Li, & Chiaromonte, 2010) to improve efficiency in analyses based on the standard multivariate (multiresponse) linear model.

$$\mathbf{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i, \; i = 1, \ldots, n, \tag{1}$$

where the stochastic response $\mathbf{Y} \in \mathbb{R}^r$, the nonstochastic predictor vectors $\mathbf{X}_i \in \mathbb{R}^p$, $\boldsymbol{\beta}_0 \in \mathbb{R}^r$, and the error vectors $\boldsymbol{\varepsilon}_i$ are independent copies of $\boldsymbol{\varepsilon} \sim N_r(0, \mathbf{\Sigma})$, where $N_r(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the r-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance–covariance matrix $\mathbf{\Sigma}$. The normality assumption on $\boldsymbol{\varepsilon}$ serves to facilitate this exposition; it is not necessary for the methodology described herein. We assume that $\boldsymbol{\beta}$ is the focus of the analysis and let $\mathbf{B}$ denote its maximum likelihood estimator based on Model 1. Cook et al. (2010) demonstrated that an envelope estimator of the unknown coefficient matrix $\boldsymbol{\beta}$ has the potential to reduce substantially the variability in $\mathbf{B}$ and that these gains will be passed on to other tasks such as prediction.

2.1 | Model for response envelopes

To sketch the development of envelopes for Model 1, we ask if there are linear combinations $\mathbf{G}_0^T\mathbf{Y}$ of $\mathbf{Y}$ so that $\mathbf{G}_0^T\mathbf{Y} \perp\!\!\!\perp \mathbf{X}$, in which case the conditional distribution of $\mathbf{G}_0^T\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is invariant to changes in $\mathbf{x}$ and we call $\mathbf{G}_0^T\mathbf{Y}$ an X-invariant. The matrix $\mathbf{G}_0$ is $r \times (r-u)$ so the statement posits that $r - u$ linear combinations of $\mathbf{Y}$ are invariant to changes in $\mathbf{X}$. For instance, suppose that the elements of $\mathbf{Y} = (Y_1, \ldots, Y_r)^T$ are repeated measurements at times $t_1 < t_2 < \cdots < t_r$ on an individual after receiving one of two treatments indicated by $\mathbf{X} = 0, 1$. If there is a lag between the time that a treatment is administered and manifestation of its effects then the distribution of the first few elements of $\mathbf{Y}$ will be invariant to changes in $\mathbf{X}$. If there are such linear combinations that can be estimated efficiently, then the variation in $\mathbf{B}$ may be reduced substantially. There could also be invariant linear combinations that are unknown a priori, but subsequently uncovered by envelope methodology.

The linear transformation $\mathbf{G}_0^T\mathbf{Y}$ is X-invariant if and only if $\mathbf{A}\mathbf{G}_0^T\mathbf{Y}$ is X-invariant for any nonsingular linear transformation $\mathbf{A}$. Consequently, X-invariance is really a requirement on the subspace span$(\mathbf{G}_0) \subseteq \mathbb{R}^r$ rather than a particular basis $\mathbf{G}_0$. Recognizing this, a formal statement of the requirement that $\mathbf{G}_0^T\mathbf{Y}$ be X-invariant can be phrased in terms of a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ with semiorthogonal basis matrix $\mathbf{G}$ of dimension $r \times u$ and its orthogonal complement $\mathcal{E}^\perp$ with semiorthogonal basis matrix $\mathbf{G}_0$ so $(\mathbf{G}, \mathbf{G}_0)$ is an orthogonal matrix:

$$\text{(i) } \mathbf{Q}_\mathcal{E}\mathbf{Y} \mid \mathbf{X} = \mathbf{x}_1 \sim \mathbf{Q}_\mathcal{E}\mathbf{Y} \mid \mathbf{X} = \mathbf{x}_1 \text{ and (ii) } \mathbf{P}_\mathcal{E}\mathbf{Y} \perp\!\!\!\perp \mathbf{Q}_\mathcal{E}\mathbf{Y} \mid \mathbf{X}.$$

Statement (i) requires that the marginal distribution of $\mathbf{G}_0^T\mathbf{Y} \mid \mathbf{X}$ be independent of the value of $\mathbf{X}$. In terms of Model 1, it holds if and only if span$(\boldsymbol{\beta}) \subseteq \mathcal{E}$. Statement (ii) insures that $\mathbf{G}_0^T\mathbf{Y}$ cannot furnish subordinate information via an association with $\mathbf{G}^T\mathbf{Y}$. It holds if and only if $\mathcal{E}$ is a reducing subspace of $\boldsymbol{\Sigma}$. That is, $\mathcal{E}$ must decompose $\boldsymbol{\Sigma} = \mathbf{P}_\mathcal{E}\boldsymbol{\Sigma}\mathbf{P}_\mathcal{E} + \mathbf{Q}_\mathcal{E}\boldsymbol{\Sigma}\mathbf{Q}_\mathcal{E}$. Together these conditions imply that the impact on $\mathbf{Y}$ of changing $\mathbf{X}$ rests wholly within the X-variant linear combinations $\mathbf{G}^T\mathbf{Y}$. The intersection of all subspaces that satisfy (i) and (ii) is called the $\boldsymbol{\Sigma}$-envelope of span$(\boldsymbol{\beta})$ and denoted as $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. Let $0 \le u \le r$ denote the dimension of the envelope.

Given an envelope dimension u, let the columns of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ be semiorthogonal bases for the envelope and its orthogonal complement, and express $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, where the matrix $\boldsymbol{\eta}$ gives the coordinates of $\boldsymbol{\beta}$ in terms of basis $\boldsymbol{\Gamma}$. We can then write the envelope version of Model 1 as

$$\begin{aligned} \mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i + \varepsilon_i,\ i = 1, \ldots, n, \\ \boldsymbol{\Sigma} &= \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}, \end{aligned} \tag{2}$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are positive definite matrices, $\boldsymbol{\Gamma}^T\mathbf{Y}$ is X-variant and $\boldsymbol{\Gamma}_0^T\mathbf{Y}$ is X-invariant. Cook et al. (2010) derived maximum likelihood estimators for parameters in this model and showed that the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \mathbf{P}_{\hat{\mathcal{E}}}\mathbf{B}$, where $\hat{\mathcal{E}}$ denotes the maximum likelihood estimator of the estimated envelope. They also showed that asymptotically the envelope estimator $\hat{\beta}$ is always better than the standard maximum likelihood estimator $\mathbf{B}$.

The model dimension u is a model-selection parameter. If $u = 0$ then $\boldsymbol{\beta} = 0$ and changing $\mathbf{X}$ has no effect on the distribution of $\mathbf{Y}$. If $u = r$ the envelope Model 2 reduces to Model 1 and the whole of $\mathbf{Y}$ is material to the regression. Information criteria, like Akaike's information criterion or the Bayesian information criterion, likelihood ratio testing or cross validation can be used to select u. Eck and Cook (2017) avoided the selecting u by proposing a model averaging estimator that averages the estimators of $\boldsymbol{\beta}$ over the values of u. Zhang and Mai (2018) proposed a model-free method of selecting the envelope dimension. These methods can be used with many but not all of the envelope types discussed herein.

2.2 | Estimation with response envelopes

Let $\mathbf{S}_{\mathbf{Y}\mid\mathbf{X}}$ denote the sample covariance matrix of the residual vectors from the fit of Model 1 and let $\mathbf{S}_\mathbf{Y}$ denote the marginal sample covariance matrix of the response vectors. Then, after maximizing the likelihood for envelope Model 2 over all other parameters, the estimator of $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is

$$\hat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \arg\min_\mathcal{S}\left[\log\det{}_0\left(\mathbf{P}_\mathcal{S}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{P}_\mathbf{S}\right) + \log\det{}_0(\mathbf{Q}_\mathcal{S}\mathbf{S}_\mathbf{Y}\mathbf{Q}_\mathcal{S})\right],$$

where $\det_0$ denotes the product of the nonzero eigenvalues of the argument matrix and the minimum is taken over all subspaces $\mathcal{S}$ of $\mathbb{R}^r$ with dimension u. This objective function is not convex and can have multiple local optima, which can cause computational problems. Computation is discussed in a later section. Having obtained $\hat{\mathcal{E}}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \mathbf{P}_{\hat{\mathcal{E}}}\mathbf{B}$, as stated previously. The asymptotic normal distribution of $\hat{\boldsymbol{\beta}}$ and the maximum likelihood estimators of the remaining parameters in Model 2 are available from Cook et al. (2010) and from Cook (2018), including methods for inferring about the unknown envelope dimension u.

2.3 | Schematic illustration of response envelopes

Figure 1 gives a schematic illustration of a response envelope for estimation of the coefficient vector in the multivariate regression of a bivariate response $\mathbf{Y} = (Y_1, Y_2)^T$ on a binary predictor $X = 0,1$:

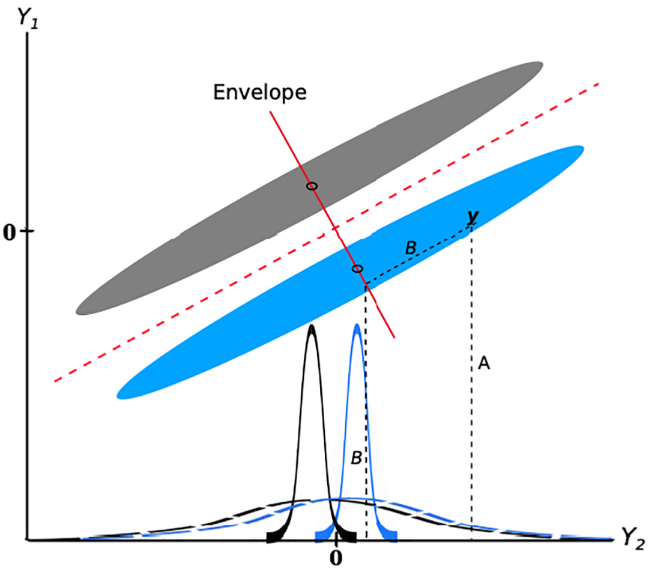

FIGURE 1 Schematic illustration of response envelopes for the regression of a bivariate response $\mathbf{Y} = (Y_1, Y_2)^T$ on a binary predictor. The distributions of $\mathbf{Y} \mid X$ are represented by the gray and blue ellipses. The red solid and dashed lines represent the envelope and its orthogonal complement

$$\mathbf{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta} X + \varepsilon. \tag{3}$$

Similar illustrations were given by Su and Cook (2011), Cook (2018), and elsewhere in the literature. Zhang and Mai (2019) used a version of Figure 1 to illustrate analysis under their envelope discriminant subspace. Model 3 allows the comparison of the means of two bivariate normal populations with the same variance. When $X = 0$, the population is $N_2(\boldsymbol{\beta}_0, \boldsymbol{\Sigma})$ and when $X = 1$ the population is $N_2(\boldsymbol{\beta}_0 + \boldsymbol{\beta}, \boldsymbol{\Sigma})$. For instance, we might imagine that Y_1 and Y_2 stand for systolic and diastolic blood pressure in the presence and absence of a drug. Envelopes have little to offer in this context if the two populations are well separated on both Y_1 and Y_2. However, an envelope may offer substantial variance reduction if the populations are close and difficult to distinguish, as depicted in Figure 1 by the populations represented as gray and blue ellipses.

Suppose for ease of illustration that we wish to infer about the second element of $\boldsymbol{\beta}$, $\beta_2 = E(Y_2 \mid X = 0) - E(Y_2 \mid X = 1)$. Standard maximum likelihood inference under Model 3 is based on projecting the populations on to the Y_2-axis, as represented by the line segment "A" in Figure 1. This results in the gray and blue univariate distributions represented with dashes along the horizontal axis. These distributions overlap substantially, so it may take a very large sample size to infer that the mean difference β_2 is not zero.

Under conditions (i) and (ii), an envelope analysis recognizes that the blue and gray bivariate distributions are identical along the orthogonal complement of the envelope, which is represented by the red dashed line in Figure 1. In other words, the variation in the direction of the orthogonal complement of the envelope is X-invariant, as it contains no information to distinguish the two populations. An envelope analysis first removes that X-invariant variation by projecting the populations onto the envelope, represented by the solid red line in Figure 1, and then projecting onto the horizontal axis for inference on β_2. This operation is represented by the two line segments labeled "B" in Figure 1, and the resulting envelope distributions are represented by the solid blue and gray curves along the horizontal axis. These distributions are much less variable than the original univariate distributions represented by dashes, which reflects the potential for substantial variance reduction by envelopes.

The red line in Figure 1 represents the population envelope. In application the envelope must be estimated, as discussed previously. The estimation uncertainty in the envelope has the effect of causing the red line in Figure 1 to wobble, which translates to increased variation in the envelope distributions represented along the horizontal axis. However, regardless of this estimation uncertainty, the variation in the envelope distributions will never be greater than the original dashed distributions (Cook et al., 2010).

Figure 1 is potentially misleading because the envelope aligns with the second eigenvector of $\boldsymbol{\Sigma}$. This choice was necessary for visualization: the only nontrivial possibility with $r = 2$ responses is to have the envelope align with one of the eigenvectors of $\boldsymbol{\Sigma}$. Being defined as the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains span($\boldsymbol{\beta}$), an envelope need not align with an eigenspace of $\boldsymbol{\Sigma}$ when there are more than two responses.

2.4 | Extensions and adaptations

There are now a number of extensions and applications of this basic envelope methodology for response reduction, each demonstrating the potential for substantial efficiency gains. Partial envelopes (Su & Cook, 2011) were tailored to applications in which a subset of the columns of $\boldsymbol{\beta}$ is of special interest, and they provide for a novel method of prediction (Cook, 2018, Section 3.4). Envelope methodology for sparse regressions was developed by Su, Zhu, Chen, and Yang (2016), and Khare, Pal, and Su (2016) constructed a Bayesian version of envelopes, both based on Model 1 as the starting point. L. Li and Zhang (2017) proposed tensor envelopes for analysis of neuroimaging applications with tensor-valued responses, Ding and Cook (2018) extended response envelopes to regressions with matrix-valued responses and Rekabdarkolaee, Wang, Naji, and Fluentes (2017) reported good efficiency gains in their adaptation of envelopes to spatial data. Ding, Su, Zhu, and Wang (2019) adapted envelopes for use in quantile regression, and Su and Cook (2013) adapted envelopes for estimation of the means $\boldsymbol{\mu}_k$ of several normal population $N_r(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $k = 1,\ldots,K$, with different variance covariance matrices. Park, Su, and Zhu (2017) extended this treatment to group-wise envelopes, which allow for predictors $\mathbf{X}_k$ in each of the K populations and thus permit both distinct regression coefficients and distinct error structures for different populations. Zhang and Mai (2019) used envelopes to improve linear and quadratic discrimination, basing their advances on a new population construction called the envelope discriminant space.

Nearly, all of these methods are based on normality as a starting point for methodological developments that also provide $\sqrt{n}$-consistent estimators under mild technical conditions, like finite fourth moments. See Cook (2018) for a review and additional extensions of envelope methodology.

3 | INNER AND SCALED ENVELOPES

Response envelopes do not always lead to efficiency gains over standard estimators. That happens in the context of Model 2 when the envelope is the full space, so $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \mathbb{R}^r$ and thus all linear combinations of $\mathbf{Y}$ are X-variant, a conclusion that might be of interest in its own right. In such cases, inner or scaled envelopes can still lead to notable reduction in estimative variation.

Recall that, in the context of Model 1, the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is defined as the *smallest* reducing subspace of $\boldsymbol{\Sigma}$ that *contains* span($\boldsymbol{\beta}$). In contrast, the inner envelope $\mathcal{I}\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ was defined by Su and Cook (2012) as the *largest* reducing subspace of $\boldsymbol{\Sigma}$ that is *contained within* span($\boldsymbol{\beta}$). The inner envelope is in effect the largest reducing subspace of $\boldsymbol{\Sigma}$ so that the part of $\boldsymbol{\beta}$ given by the projection $\mathbf{P}_{\mathcal{I}\mathcal{E}}\boldsymbol{\beta}$ onto the inner envelope can be estimated with reduced variability relative to $\mathbf{P}_{\mathcal{I}\mathcal{E}}\mathbf{B}$, while the projection $\mathbf{Q}_{\mathcal{I}\mathcal{E}}\boldsymbol{\beta}$ onto its orthogonal complement is estimated with about the same variability as $\mathbf{Q}_{\mathcal{I}\mathcal{E}}\mathbf{B}$. Since $\boldsymbol{\beta} = \mathbf{P}_{\mathcal{I}\mathcal{E}}\boldsymbol{\beta} + \mathbf{Q}_{\mathcal{I}\mathcal{E}}\boldsymbol{\beta}$, reduced variability is achieved overall. Methodology based on inner envelopes was developed by Su and Cook (2012).

The success of envelope methodology for response reduction can depend on the scaling of the response vector. Consider rescaling the response vector in Model 1 as $\mathbf{Y} \mapsto \boldsymbol{\Lambda}^{-1}\mathbf{Y}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of scaling constants. The reduction of variation gained from envelopes in the $\mathbf{Y}$ scale is not generally the same as that from envelopes in the scale of $\boldsymbol{\Lambda}^{-1}\mathbf{Y}$. This allows for the possibility of gaining greater variance reduction by estimating an appropriate scaling matrix $\boldsymbol{\Lambda}$, a line of inquiry that was introduced and developed by Cook and Su (2013). They showed that scaled envelopes can lead to reduced estimative variability, particularly when the envelope in the original $\mathbf{Y}$ scale produces little or no gain over the standard estimator $\mathbf{B}$.

4 | PREDICTOR ENVELOPES

Cook, Helland, and Su (2013) proposed using envelopes to reduce the random predictor vector $\mathbf{X} \in \mathbb{R}^p$ in Model 1. Here we consider only the univariate linear regression model,

$$Y_i = \alpha + \boldsymbol{\beta}^T(\mathbf{X}_i - \mu_{\mathbf{X}}) + \epsilon_i,\ i = 1,\ldots,n, \tag{4}$$

where $E(\mathbf{X}) = \boldsymbol{\mu}_{\mathbf{X}}$, $\mathrm{var}(\mathbf{X}) = \boldsymbol{\Sigma}_{\mathbf{X}}$, $E(\epsilon) = 0$, $\mathrm{var}(\epsilon) = \sigma^2$, $\mathbf{X} \perp\!\!\!\perp \epsilon$, and coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$ is defined as the transpose of that in Model 1. Let $\mathbf{b}$ denote the least squares estimator of $\boldsymbol{\beta}$ and let $\mathbf{S}_{\mathbf{X}}$ denote the sample covariance matrix of the predictors. Without loss of generality, we cast the following discussion in terms of an orthogonal matrix $(\boldsymbol{\Phi}, \boldsymbol{\Phi}_0)$, where the dimension of $\boldsymbol{\Phi}$ is $p \times q$, $q \le p$. Envelopes are motivated for Model 4 by asking if there are linear combinations $\boldsymbol{\Phi}^T\mathbf{X}$ of the predictor vector that carry all of the information about Y that is available from $\mathbf{X}$. If such linear combinations were known then we could reduce $\mathbf{X} \mapsto \mathbf{\Phi}^T\mathbf{X}$, fit the linear regression of Y on $\mathbf{\Phi}^T\mathbf{X}$ and then use the envelope estimator $\hat{\beta}_{\mathbf{\Phi}} = \mathbf{P}_{\mathbf{\Phi}(\mathbf{S_X})}\mathbf{b}$ to estimate $\boldsymbol{\beta}$. Here, $\mathbf{P}_{\mathbf{\Phi}(\mathbf{S_X})} = \mathbf{\Phi}\left(\mathbf{\Phi}^T\mathbf{S_X}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{S_X}$ denotes the projection onto span($\mathbf{\Phi}$) in the $\mathbf{S_X}$ inner product. Envelopes guide the choice $\mathbf{\Phi}$ by imposing the two conditions (a) $Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{\Phi}^T\mathbf{X}$ and (b) $\mathbf{\Phi}_0^T\mathbf{X} \perp\!\!\!\perp \mathbf{\Phi}^T\mathbf{X}$. Condition (a) is the same as equivalence 3 in Definition 1, the condition that underlies SDR. It ensures that $\mathbf{\Phi}^T\mathbf{X}$ carries all of the information about Y that is available from $\mathbf{X}$. Condition (b) is what sets envelopes apart from SDR. It ensures that Y is unaffected by changes in $\mathbf{\Phi}_0^T\mathbf{X}$ because of an association with $\mathbf{\Phi}^T\mathbf{X}$. Together, these conditions hold if and only if $\left(Y, \mathbf{\Phi}^T\mathbf{X}\right) \perp\!\!\!\perp \mathbf{\Phi}_0^T\mathbf{X}$, so $\mathbf{\Phi}_0^T\mathbf{X}$ is entirely unrelated to the regression. As in SDR, $\mathbf{\Phi}$ is not estimable, but span($\mathbf{\Phi}$) will typically be estimable, and for this reason it is often useful to restate conditions (a) and (b) in terms of projection onto $\mathcal{E} = \text{span}(\mathbf{\Phi})$:

$$\text{(I) } Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{P}_{\mathcal{E}}\mathbf{X} \text{ and (II) } \mathbf{Q}_{\mathcal{E}}\mathbf{X} \perp\!\!\!\perp \mathbf{P}_{\mathcal{E}}\mathbf{X}.$$

In Model 4, condition (I) is equivalent to requiring that span($\boldsymbol{\beta}$) $\subseteq \mathcal{E}$ and condition (II) is equivalent to requiring the decomposition $\mathbf{\Sigma_X} = \mathbf{P}_{\mathcal{E}}\mathbf{\Sigma_X}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\mathbf{\Sigma_X}\mathbf{Q}_{\mathcal{E}}$. The intersection of all subspaces with these properties is called the $\mathbf{\Sigma_X}$-envelope of span ($\boldsymbol{\beta}$) and denoted as $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta})$, since it envelopes the material information in the data while excluding immaterial information. Let q denote the dimension of the predictor envelope $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta})$.

4.1 | Model for predictor envelopes

For a selected dimension q, Model 4 can now be reparameterized using a basis $\mathbf{\Phi}$ for $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta})$ to get its envelope version

$$\begin{aligned} Y_i &= \alpha + \boldsymbol{\eta}^T\mathbf{\Phi}^T(\mathbf{X}_i - \mu_{\mathbf{X}}) + \epsilon_i \\ \mathbf{\Sigma_X} &= \mathbf{\Phi}\mathbf{\Delta}\mathbf{\Phi}^T + \mathbf{\Phi}_0\Delta_0\mathbf{\Phi}_0^T. \end{aligned} \tag{5}$$

Cook et al. (2013) derived maximum likelihood estimators for quantities in this model and showed that the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is, as indicated previously, $\hat{\boldsymbol{\beta}} = \mathbf{P}_{\hat{\mathcal{E}}(\mathbf{S_X})}\mathbf{b}$, where $\hat{\mathcal{E}}$ denotes the maximum likelihood estimator of the estimated envelope. They also showed that asymptotically the envelope estimator $\hat{\boldsymbol{\beta}}$ is always better than the standard maximum likelihood estimator $\mathbf{b}$. In other work, simultaneous envelopes (Cook & Zhang, 2015b) allow for simultaneously reducing the response and random predictor vectors.

4.2 | Predictor envelopes and partial least squares regression

There is a close population connection between envelopes estimated by maximizing the likelihood for Model 4 and SIMPLS, the algorithm for partial least squares regression proposed by de Jong (1993): holding p fixed, SIMPLS provides a $\sqrt{n}$-consistent moment-based estimator of the envelope $\mathcal{E}_{\mathbf{\Sigma_X}}(\boldsymbol{\beta})$ (Cook et al., 2013). Envelope and partial least squares regressions thus estimate the same population quantity, but differ in the method for doing so. Rimal, Almoy, and Saebo (2019) conducted a large simulation study comparing predictions from envelope and partial least squares regressions, concluding that envelopes perform distinctly better. Zhu and Su (2019) developed envelope-based sparse partial least squares for linear and generalized linear models, and showed that their envelope-based method has better performance that current sparse partial least squares (PLS) methods. Building upon the envelope work of Cook et al. (2013), Zhang and Li (2017) extended PLS from regression with a vector predictor to a tensor predictor.

Predictor scaling can be problematic in PLS. Some prefer to scale the predictors to have marginal standard deviations equal to 1, while others do not seem to scale. Since PLS is not invariant or equivariant under rescaling of predictors, the actual scale used can have a substantial impact on subsequent predictions. Cook and Su (2016) formally incorporated predictor scaling into predictor envelopes and developed corresponding estimation methods. They also argued that scaled predictor envelopes avoid the need to select an a priori scaling method and can yield predictions that are better than those from existing PLS methods.

Cook and Forzani (2017, 2019) studied the asymptotic behavior of predictions based on SIMPLS in high-dimensional regressions. They showed that for regressions in which many predictors are marginally correlated with the response, SIMPLS predictions can converge at the root-n rate regardless of the relationship between n and p. However, in sparse regressions where few predictors contribute information about the response, SIMPLS predictions may be inconsistent. The usefulness of SIMPLS then seems to rest primarily with the abundance of informative predictors.

5 | CONNECTIONS WITH OTHER MULTIVARIATE METHODS

We have already discussed the connection between envelope methods, sufficient dimension reduction and partial least squares regression. There are also close connections between envelopes and other long-standing multivariate methods. In this section we describe three such connections. Other connections were described by Cook (2018).

5.1 | Canonical correlations

Cook and Zhang (2015b) demonstrated a population connection between envelopes and canonical correlation analysis. To explain this, consider the canonical correlations between the $p \times 1$ random vector $\mathbf{X}$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ and the $r \times 1$ random vector $\mathbf{Y}$ with covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}$. Let $\boldsymbol{\Sigma}_{\mathbf{X},\mathbf{Y}}$ denote the matrix of covariances between the elements of $\mathbf{X}$ and $\mathbf{Y}$, and let $\rho_{\mathbf{X},\mathbf{Y}} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1/2}\boldsymbol{\Sigma}_{\mathbf{X},\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1/2}$ be the matrix of correlations between the elements of the standardized random vectors $\boldsymbol{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{X}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{Y}$. Let $\boldsymbol{\rho}_{\mathbf{X},\mathbf{Y}} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ be the compact singular value decomposition of $\boldsymbol{\rho}_{\mathbf{X},\mathbf{Y}}$, where $\mathbf{U}$ and $\mathbf{V}$ are semi-orthogonal matrices and $\mathbf{D}$ is a diagonal matrix with positive diagonal elements $d_1 > \ldots > d_q > 0$. Then, the population canonical vectors are the columns of $\boldsymbol{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{U}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{V}$, the canonical variates are $\mathbf{U}^T\boldsymbol{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{X}$ and $\mathbf{V}^T\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{Y}$, and the canonical correlations are $d_1,\ldots,d_q$. Cook and Zhang (2015b) showed that $\mathrm{span}\left(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{U}\right)$ is contained in the envelope used for predictor reduction and that $\mathrm{span}\left(\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{V}\right)$ is contained in the envelope for response reduction. In short, the population canonical vectors are contained in the response and predictor envelopes. They also demonstrated that the sample versions of the canonical vectors provide relatively poor estimators of the corresponding envelopes.

5.2 | Reduced rank regression

In reduced rank regression (Reinsel & Velu, 1998) the $r \times p$ coefficient matrix $\boldsymbol{\beta}$ in Model 1 is allowed to have reduced rank $d < \min(r, p)$, yielding the representation $\boldsymbol{\beta} = \mathbf{H}\mathbf{h}$ where $\mathbf{H}$ is an $r \times d$ matrix of full column rank and $\mathbf{h}$ is a $d \times p$ matrix containing the coordinates of $\boldsymbol{\beta}$ in terms of basis $\mathbf{H}$. The corresponding multivariate model is then

$$\mathbf{Y}_i = \boldsymbol{\beta}_0 + \mathbf{H}\mathbf{h}\mathbf{X}_i + \varepsilon_i, \ i = 1,\ldots,n, \tag{6}$$

As in Model 1, the error vectors $\boldsymbol{\varepsilon}_i$ are independent copies of a $N_r(0, \boldsymbol{\Sigma})$ random vector. Model 6 is like the envelope Model 2 because in both models the coefficient matrix is represented as a basis matrix, $\boldsymbol{\Gamma}$ in Model 2 and $\mathbf{H}$ in Model 6, times a coordinate matrix. But Model 6 is unlike the envelope Model 2 because the error covariance matrix $\boldsymbol{\Sigma} > 0$ is unrestricted in reduced rank regression but restricted in the envelope model. For this reason, we can view the reduced rank regression Model 6 as lying "between" the standard multivariate linear Model 1 and the envelope Model 2. Maximum likelihood estimators of the parameters and an introduction to the literature were given by Cook, Forzani, and Zhang (2015).

To contrast Models 2 and 6, consider a regression in which $\boldsymbol{\beta} \neq 0$ and $d = 1$, so reduced rank regression will yield its greatest dimension reduction and its greatest reduction in estimative variation. It is possible in the setting for an envelope analysis based on Model 2 to have an envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ of dimension r, in which case the envelope model reduces to the usual multivariate Model 1 and there is no estimative gain. This is then a case in which reduced rank regression may be preferable. On the other hand, consider a regression in which $p = 1$. In this case, reduced rank regression offers no reduction in estimative variation, while an envelope analysis could still provide substantial reduction. This type of reasoning leads to the general conclusion that both models have relative strengths and weaknesses.

Cook et al. (2015) develop a reduced rank envelope model that combines the strengths of Models 6 and 2. Under Model 6, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\mathbf{H})$, the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathrm{span}(\mathbf{H})$. Parametrizing Model 6 in terms of a semi-orthogonal basis matrix $\boldsymbol{\Gamma}$ of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathbf{H})$, we arrive at the reduced rank envelope model,

$$\begin{aligned}\mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{h}\mathbf{X}_i + \varepsilon_i, \ i = 1,\ldots,n,\\ \boldsymbol{\Sigma} &= \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Delta}_0\boldsymbol{\Gamma}_0^T.\end{aligned} \tag{7}$$

Cook et al. (2015) developed likelihood-based estimation and inference under this model and demonstrated that it does incorporate the advantages of Models 2 and 6.

5.3 | Supervised singular value decomposition

Low-rank approximation of an $n \times r$ data matrix $\mathbb{Y}$ is often used as a dimension reduction paradigm in the applied sciences. Letting $\hat{\mathbb{Y}}$ denote a rank $u \le \min(n, r)$ approximation to $\mathbb{Y}$, it is know that the Frobenius norm of $\mathbb{Y} - \hat{\mathbb{Y}}$ is minimized by selecting $\hat{\mathbb{Y}}$ to be the rank u compact singular value decomposition of $\mathbb{Y}$. G. Li, Yang, Nobel, and Shen (2015) developed methodology to allow a singular value decomposition to be supervised by a concomitant $n \times p$ data matrix $\mathbb{X}$. An envelope model arises naturally in their development as a stepping stone to the solution.

6 | A GENERAL APPROACH TO ENVELOPES

Response and predictor envelopes and their various methodological descendents are largely for reduction of estimative and predictive variation in linear models. Cook and Zhang (2015a) developed a general approach to envelope estimation that applies well beyond the linear model context. It can be used for dimension reduction in linear, logistic, Cox, and similar regression models where SDR is of little help.

Let $\tilde{\boldsymbol{\theta}}$ denote an estimator of an $r \times 1$ parameter vector $\boldsymbol{\theta}$ with the property that $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges in distribution to a $N_r(0, \boldsymbol{\Delta}(\boldsymbol{\theta}))$ random vector. As indicated by the notation, the asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ can depend on $\boldsymbol{\theta}$. This context is quite general and covers generalized linear models, like logistic and Poisson regression. To allow for the presence of nuisance parameters, partition $\boldsymbol{\theta} = (\boldsymbol{\psi}^T, \boldsymbol{\phi}^T)^T$ where the parameters of interest are held in the $p \times 1$ vector $\boldsymbol{\phi}$. Let $\boldsymbol{\Delta}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ denote the asymptotic covariance matrix of $\tilde{\phi}$. We occasionally use the abbreviated notation $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ to denote $\boldsymbol{\Delta}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$. We next describe how envelopes can be used to reduce variability in the estimator $\tilde{\boldsymbol{\theta}}$.

Let $\mathcal{E}_{\Delta_\phi}(\phi)$ denote the $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$-envelope span($\boldsymbol{\phi}$); that is, the smallest reducing subspace of $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ that contains span($\boldsymbol{\phi}$). This definition of an envelope requires no structure beyond the asymptotic normality of $\tilde{\boldsymbol{\theta}}$ and in consequence it serves as a widely applicable starting point for envelope methodology. It also reproduces past methodology for linear models, and so serves as a generalized method of envelope construction.

To gain intuition into how an envelope can be used to reduce the variation in $\tilde{\phi}$, let $\mathbf{P}_{\mathcal{E}}$ denote the projection onto a known envelope $\mathcal{E}_{\Delta_\phi}(\phi)$ and write

$$\sqrt{n}(\tilde{\phi} - \phi) = \sqrt{n}\mathbf{P}_{\mathcal{E}}(\tilde{\phi} - \phi) + \sqrt{n}\mathbf{Q}_{\mathcal{E}}\tilde{\phi}.$$

The first addend on the right side converges to a normal random vector with mean 0 and variance $\mathbf{P}_{\mathcal{E}}\boldsymbol{\Delta}_{\boldsymbol{\phi}}\mathbf{P}_{\mathcal{E}}$, the second addend converges to a normal random vector with mean 0 and variance $\mathbf{Q}_{\mathcal{E}}\boldsymbol{\Delta}_{\boldsymbol{\phi}}\mathbf{Q}_{\mathcal{E}}$, and the two addends are asymptotically independent. In consequence, we define the envelope estimator of $\boldsymbol{\phi}$ as $\hat{\phi} = \mathbf{P}_{\mathcal{E}}\tilde{\phi}$. Decomposing $\tilde{\phi} = \hat{\phi} + \mathbf{Q}_{\mathcal{E}}\tilde{\phi}$, we see that the term $\mathbf{Q}_{\mathcal{E}}\tilde{\phi}$ is an estimator of 0 that represents extraneous variation which is effectively removed by the envelope estimator.

7 | COMPUTING

The maximum likelihood estimators of response and predictor envelopes—$\mathcal{E}_{\boldsymbol{\Sigma}}$ and $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}$—and their descendants can be written in the general form

$$\hat{\mathcal{E}} = \operatorname{span}\left\{\arg\min_{\mathbf{G}}\left[\operatorname{logdet}(\mathbf{G}^T\hat{\mathbf{M}}\mathbf{G}) + \operatorname{logdet}\left(\mathbf{G}^T(\hat{\mathbf{M}} + \hat{\mathbf{U}})^{-1}\mathbf{G}\right)\right]\right\},$$

where $\hat{\mathbf{M}}$ and $\hat{\mathbf{U}}$ are observed positive definite and semi-definite matrices that depend on the application, and the minimum is over semiorthogonal matrices of a given dimension. For instance, in reference to the general approach discussed in the previous section, we can take $\hat{\mathbf{M}}$ and $\hat{\mathbf{U}}$ to be estimators of $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}\boldsymbol{\phi}^T$. The objective function for this minimization is not convex and, in consequence, finding the solution can be difficult as there may be multiple local optima. The algorithm proposed by Cook, Forzani, and Su (2016) gets around these difficulties by using root-n consistent starting values and transforming the problem so optimization can be carried out by standard algorithms. Cook and Zhang (2016, 2018) developed sequential algorithms that facilitate computation. Comprehensive R and MatLab packages that implement a variety of envelope methods based on these and other algorithms are described and linked at z.umn.edu/envelopes. That page also provides links to a variety of other methods that are not included in the R and MatLab package, including tensor and matrix response envelopes, spatial envelopes, sparse envelopes, Bayesian envelopes, reduced-rank envelopes and envelopes for functional data.

8 | CONCLUSIONS

Envelope constructions have opened a new chapter in multivariate analysis. Envelope methodology has developed rapidly over the past decade and it seems likely that advances will continue apace over the next decade. Several illustrative applications of envelope methodology are available from Cook (2018).

Following Fisher's guidance, envelope methodology allows the relevant (material) information to be separated from the irrelevant (immaterial) information, and in doing so, it can provide substantial reduction in estimative and predictive variation. Inferring that the response envelope is equal to the full space, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \mathbb{R}^r$, implies that the whole of the response vector is material to the regression, a conclusion that might be useful in some applications. Inferring that the response envelope is a proper subset, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) \subset \mathbb{R}^r$, implies that a portion of the response vector is immaterial to the regressions. Similar statements apply to other types of envelopes.

ACKNOWLEDGMENT

The author thanks the referees for helpful comments on a previous version of this article.

CONFLICT OF INTEREST

The author has declared no conflicts of interest for this article.

ORCID

R. Dennis Cook https://orcid.org/0000-0002-3488-4743

RELATED WIREs ARTICLES

Multivariate methods

REFERENCES

Cook, R. D. (1998). *Regression graphics*. New York, NY: Wiley.

Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression. *Statistical Science*, *22*(1), 1–26. Retrieved from https://projecteuclid.org/euclid.ss/1185975631. https://doi.org/10.1214/088342306000000682

Cook, R. D. (2018). *An introduction to envelopes*. Hoboken, NJ: Wiley.

Cook, R. D., & Forzani, L. (2017). Big data and partial least squares prediction. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, *to appear*, *46*, 62–78. https://doi.org/10.1002/cjs.11316

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

Cook, R. D., Forzani, L., & Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis*, *150*, 42–54. https://doi.org/10.1016/j.jmva.2016.05.006

Cook, R. D., Forzani, L., & Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika*, *102*(2), 439–456. https://doi.org/10.1093/biomet/asv001

Cook, R. D., Helland, I. S., & Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society B*, *75*(5), 851–877. https://doi.org/10.1111/rssb.12018

Cook, R. D., Li, B., & Chiaromonte, F. (2007). Dimension reduction in regression without matrix inversion. *Biometrika*, *94*(3), 569–584. https://doi.org/10.1093/biomet/92.4.937

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

Cook, R. D., & Su, Z. (2013). Scaled envelopes: Scale-invariant and efficient estimation in multivariate linear regression. *Biometrika*, *100*(4), 939–954. https://doi.org/10.1093/biomet/ast026

Cook, R. D., & Su, Z. (2016). Scaled predictor envelopes and partial least-squares regression. *Technometrics*, *58*(2), 155–165. https://doi.org/10.1080/00401706.2015.1017611

Cook, R. D., & Zhang, X. (2015a). Foundations for envelope models and methods. *Journal of the American Statistical Association*, *110*(510), 599–611. https://doi.org/10.1080/01621459.2014.983235

Cook, R. D., & Zhang, X. (2015b). Simultaneous envelopes for multivariate linear regression. *Technometrics*, *57*(1), 11–25. https://doi.org/10.1080/00401706.2013.872700

Cook, R. D., & Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, *25*(1), 284–300. https://doi.org/10.1080/10618600.2015.1029577

Cook, R. D., & Zhang, X. (2018). Fast envelope algorithms. *Statistica Sinica*, *28*, 1179–1197. https://doi.org/10.5705/ss.202016.0037

Cox, D. R., & Mayo, D. G. (2010). II Objectivity and conditionality in frequentist inference. In D. G. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* (pp. 276–304). Cambridge, England: Cambridge University Press.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *18*(3), 251–263. https://doi.org/10.1016/0169-7439(93)85002-X

Ding, S., & Cook, R. D. (2018). Matrix-variate regressions and envelope models. *Journal of the Royal Statistical Society B*, *80*, 387–408. https://doi.org/10.1111/rssb.12247

Ding, S., Su, Z., Zhu, G., & Wang, L. (2019). Envelope quantile regression. *Statistica Sinica*. Retrieved from http://www3.stat.sinica.edu.tw/ss_newpaper/SS-2018-0060_na.pdf

Eck, D. J., & Cook, R. D. (2017). Weighted envelope estimation to handle variability in model selection. *Biometrica*, *104*(3), 743–749. https://doi.org/10.1093/biomet/asx035

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *222*(594–604), 309–368. Retrieved from http://rsta.royalsocietypublishing.org/content/222/594-604/309. https://doi.org/10.1098/rsta.1922.0009

Khare, K., Pal, S., & Su, Z. (2016). A Bayesian approach for envelope models. *Annals of Statistics*, *45*(1), 196–222. Retrieved from http://projecteuclid.org/euclid.aos/1487667621. https://doi.org/10.1214/16-AOS1449

Li, B. (2018). *Sufficient dimension reduction: Methods and applications with r*. New York, NY: Chapman and Hall/CRC Press.

Li, G., Yang, D., Nobel, A. B., & Shen, H. (2015). Supervised singular value decomposition and its asymptotic properties. *Journal of Multivariate Analysis*, *146*, 7–17. Retrieved from. https://doi.org/10.1016/j.jmva.2015.02.016

Li, L., & Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, *112*(519), 1131–1146. https://doi.org/10.1080/01621459.2016.1193022

Park, Y., Su, Z., & Zhu, H. (2017). Groupwise envelope models for imaging genetic analysis. *Biometrics*, *73*(4), 1243–1253. https://doi.org/10.1111/biom.12689

Reinsel, G. C., & Velu, R. P. (1998). *Multivariate reduced-rank regression: Theory and applications*. New York, NY: Springer.

Rekabdarkolaee, H. M., Wang, Q., Naji, Z., & Fluentes, M. (2017). New parsimonious multivariate spatial model: Spatial envelope. *Statistica Sinica* Retrieved from http://www3.stat.sinica.edu.tw/ss_newpaper/SS-2017-0455_na.pdf

Rimal, R., Almoy, T., & Saebo, S. (2019, March). *Comparison of multi-response prediction methods*. (arXiv:1903.08426v1)

Su, Z., & Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, *98*(1), 133–146. https://doi.org/10.1093/biomet/asq063

Su, Z., & Cook, R. D. (2012). Inner envelopes: Efficient estimation in multivariate linear regression. *Biometrika*, *99*(3), 687–702. https://doi.org/10.1093/biomet/ass024

Su, Z., & Cook, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica*, *23*(1), 213–230. https://doi.org/10.5705/ss.2010.240

Su, Z., Zhu, G., Chen, X., & Yang, Y. (2016). Sparse envelope model: Estimation and response variable selection in multivariate linear regression. *Biometrika*, *103*(3), 579–593. https://doi.org/10.1093/biomet/asw036

Zhang, X., & Li, L. (2017). Tensor envelope partial least-squares regression. *Technometrics*, *59*(4), 426–436. https://doi.org/10.1080/00401706.2016.1272495

Zhang, X., & Mai, Q. (2018). Model-free envelope dimension selection. *Electronic Journal of Statistics*, *12*(2), 2193–2216. Retrieved from https://projecteuclid.org/euclid.ejs/1531814505 (https://arxiv.org/abs/1709.03945). https://doi.org/10.1214/18-EJS1449

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

Zhu, G., & Su, Z. (2019). Envelope-based sparse partial least squares. *Annals of Statistics*, *47*, To appear. https://doi.org/10.1080/00401706.2016.1272495

How to cite this article: Cook RD. Envelope methods. *WIREs Comput Stat*. 2020;12:e1484. https://doi.org/10.1002/wics.1484