Journal of Multivariate Analysis 150 (2016) 42–54

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A note on fast envelope estimation

R. Dennis Cook [a], Liliana Forzani [b,c], Zhihua Su [d,*]

[a] School of Statistics, University of Minnesota, Minneapolis, MN 55455, United States
[b] Facultad de Ingeniería Química, Universidad Nacional del Litoral, Argentina
[c] Instituto Matemática Aplicada Litoral, CONICET-UNL, Santa Fe, Argentina
[d] Department of Statistics, University of Florida, Gainsville, FL 32611-8545, United States

ARTICLE INFO

Article history:
Received 19 April 2016
Available online 22 May 2016

AMS subject classifications:
62H12

Keywords:
Envelopes
Grassmann manifold
Reducing subspaces

ABSTRACT

We propose a new algorithm for envelope estimation, along with a new $\sqrt{n}$-consistent method for computing starting values. The new algorithm, which does not require optimization over a Grassmannian, is shown by simulation to be much faster and typically more accurate than the best existing algorithm proposed by Cook and Zhang (2016).

1. Introduction

The goal of envelope methods is to increase efficiency in multivariate parameter estimation and prediction by exploiting variation in the data that is effectively immaterial to the goals of the analysis. Envelopes achieve efficiency gains by basing estimation on the variation that is material to those goals, while simultaneously excluding that which is immaterial. It now seems evident that immaterial variation is often present in multivariate analyses and that the estimative improvement afforded by envelopes can be quite substantial when the immaterial variation is large, sometimes equivalent to taking thousands of additional observations.

Algorithms for envelope estimation require optimization of a non-convex objective function over a Grassmannian, which can be quite slow in all but small or modest sized problems, possibly taking hours or even days to complete an analysis of a sizable problem. Local optima are another complication that may increase the difficulty of the computations and the analysis generally. Until recently, envelope methods were available only in Matlab, as these computing issues hindered implementation in R.

In this article we propose new easily computed $\sqrt{n}$-consistent starting values and a novel non-Grassmann algorithm for optimization of the most common envelope objective function. These computing tools are much faster than current algorithms in sizable problems and can be implemented straightforwardly in R. The new starting values have proven quite effective and can be used as fast standalone estimators in exploratory analyses. An R package that implements the algorithm was developed and is available at http://www.stat.ufl.edu/~zhihuasu/Renvlp.

In the remainder of this introduction we review envelopes and describe the computing issues in more detail. We let $\mathbf{P}_{(\cdot)}$ denote a projection with $\mathbf{Q}_{(\cdot)} = \mathbf{I} - \mathbf{P}_{(\cdot)}$, let $\mathbb{R}^{r\times c}$ be the set of all real $r \times c$ matrices, and let $\mathbb{S}^{k\times k}$ be the set of all real and symmetric $k \times k$ matrices. If $\mathbf{M} \in \mathbb{R}^{r\times c}$, then $\text{span}(\mathbf{M}) \subseteq \mathbb{R}^r$ is the subspace spanned by columns of $\mathbf{M}$. vec is the vectorization operator that stacks the columns of a matrix. A subspace $\mathcal{R} \subseteq \mathbb{R}^p$ is said to be a reducing subspace of $\mathbf{M} \in \mathbb{R}^{p\times p}$ if $\mathcal{R}$ decomposes $\mathbf{M}$ as $\mathbf{M} = \mathbf{P}_{\mathcal{R}}\mathbf{M}\mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}}\mathbf{M}\mathbf{Q}_{\mathcal{R}}$. If $\mathcal{R}$ is a reducing subspace of $\mathbf{M}$, we say that $\mathcal{R}$ reduces $\mathbf{M}$.

* Corresponding author.
E-mail addresses: dennis@stat.umn.edu (R.D. Cook), liliana.forzani@gmail.com (L. Forzani), zhihuasu@stat.ufl.edu (Z. Su).

http://dx.doi.org/10.1016/j.jmva.2016.05.006
0047-259X/

1.1. Review of envelopes

Envelopes were originally proposed and developed by Cook et al. [2,3] in the context of multivariate linear regression,

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\varepsilon}_i \in \mathbb{R}^r$ is a normal error vector with mean 0, variance $\boldsymbol{\Sigma} > 0$ and is independent of $\mathbf{X}$, $\boldsymbol{\alpha} \in \mathbb{R}^r$ and $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$ is the regression coefficient matrix in which we are primarily interested. Immaterial variation can occur in $\mathbf{Y}$ or $\mathbf{X}$ or both. Cook et al. [3] operationalized the idea of immaterial variation in the response vector by asking if there are linear combinations of $\mathbf{Y}$ whose distribution is invariant to changes in $\mathbf{X}$. Specifically, let $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ denote the projection onto a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ with the properties (1) the distribution of $\mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X}$ does not depend on the value of the non-stochastic predictor $\mathbf{X}$ and (2) $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ is independent of $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ given $\mathbf{X}$. These conditions imply that the distribution of $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ is not affected by $\mathbf{X}$ marginally or through an association with $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$. Consequently, changes in the predictor affect the distribution of $\mathbf{Y}$ only via $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ and so we refer to $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ informally as the material part of $\mathbf{Y}$ and to $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ as the immaterial part of $\mathbf{Y}$.

Conditions (1) and (2) hold if and only if (a) $\mathcal{B} := \mathrm{span}(\boldsymbol{\beta}) \subseteq \mathcal{E}$ (so $\mathcal{E}$ *envelopes* $\mathcal{B}$) and (b) $\mathcal{E}$ reduces $\boldsymbol{\Sigma}$. The $\boldsymbol{\Sigma}$-envelope of $\mathcal{B}$, denoted $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, is defined formally as the intersection of all reducing subspaces of $\boldsymbol{\Sigma}$ that contain $\mathcal{B}$. Let $u = \dim\{\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})\}$ and let $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0) \in \mathbb{R}^{r\times r}$ be an orthogonal matrix with $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times u}$ and $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$. This leads directly to the envelope version of model (1),

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X}_i + \boldsymbol{\varepsilon}_i, \quad \text{with } \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^\top + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^\top, \quad i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^{u\times p}$ gives the coordinates of $\boldsymbol{\beta}$ relative to basis $\boldsymbol{\Gamma}$, and $\boldsymbol{\Omega} \in \mathbb{S}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{S}^{(r-u)\times(r-u)}$ are positive definite matrices. While $\boldsymbol{\eta}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ depend on the basis $\boldsymbol{\Gamma}$ selected to represent $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ depend only on $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and not on the basis. All parameters in (2) can be estimated by maximizing its likelihood with the envelope dimension u determined by using standard methods like likelihood ratio testing, information criteria, cross-validation or a hold-out sample, as described by Cook et al. [3]. The envelope estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is just the projection of the ordinary least squares estimator $\mathbf{B}$ of $\boldsymbol{\beta}$ onto the estimated envelope, $\widehat{\boldsymbol{\beta}} = \mathbf{P}_{\widehat{\mathcal{E}}}\mathbf{B}$, and $\sqrt{n}\{\mathrm{vec}(\widehat{\boldsymbol{\beta}}) - \mathrm{vec}(\boldsymbol{\beta})\}$ is asymptotically normal with mean 0 and covariance matrix given by Cook et al. [3], where u is assumed to be known. An introductory example of response envelopes is available in Cook and Zhang [5].

Similar reasoning leads to partial envelopes for use when only selected columns of $\boldsymbol{\beta}$ are of interest (Su and Cook [10]), to predictor envelopes allowing for immaterial variation in $\mathbf{X}$ (Cook et al. [1]), to predictor-response envelopes allowing simultaneously for immaterial variation in $\mathbf{X}$ and $\mathbf{Y}$ (Cook and Zhang [6]) and to heteroscedastic envelopes for comparing the means of multivariate populations with unequal covariance matrices (Su and Cook [11]).

Cook and Zhang [5] extended envelopes beyond multivariate linear models by proposing the following estimative construct for vector-valued parameters. Let $\widetilde{\boldsymbol{\theta}}$ denote an estimator of a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^m$ based on a sample of size n and assume, as is often the case, that $\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges in distribution to a normal random vector with mean 0 and covariance matrix $\mathbf{V}(\boldsymbol{\theta}) > 0$ as $n \to \infty$. To accommodate the presence of nuisance parameters, decompose $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\boldsymbol{\psi}^\top, \boldsymbol{\phi}^\top)^\top$, where $\boldsymbol{\phi} \in \mathbb{R}^p$, $p \le m$, is the parameter vector of interest and $\boldsymbol{\psi} \in \mathbb{R}^{m-p}$ is the nuisance parameter vector. The asymptotic covariance matrix of $\widetilde{\boldsymbol{\phi}}$ is represented as $\mathbf{V}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\theta})$, which is the $p \times p$ lower right block of $\mathbf{V}(\boldsymbol{\theta})$. Then Cook and Zhang [5] defined the envelope for improving $\widetilde{\boldsymbol{\phi}}$ as the smallest reducing subspace of $\mathbf{V}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\theta})$ that contains $\mathrm{span}(\boldsymbol{\phi})$, $\mathcal{E}_{\mathbf{V}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\theta})}\{\mathrm{span}(\boldsymbol{\phi})\} \subseteq \mathbb{R}^p$. This definition links the envelope to a particular pre-specified method of estimation through the covariance matrix $\mathbf{V}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\theta})$, while normal-theory maximum likelihood is the only method of estimation allowed by the previous approaches. The goal of an envelope is to improve that pre-specified estimator, perhaps a maximum likelihood, least squares or robust estimator. Second, the matrix to be reduced – here $\mathbf{V}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\theta})$ – is dictated by the method of estimation. Third, the matrix to be reduced can now depend on the parameter being estimated, in addition to perhaps other parameters. Cook and Zhang [5] sketched application details for generalized linear models, weighted least squares, Cox regression and described an extension to matrix-valued parameters.

1.2. Computational issues

The approaches reviewed in the last section all require estimation of an envelope, now represented generically as $\mathcal{E}_{\mathbf{M}}(\mathcal{U})$, the smallest reducing subspace of $\mathbf{M} \in \mathbb{S}^{r\times r}$ that contains $\mathcal{U} \subseteq \mathbb{R}^r$, where $\mathbf{M} > 0$. Let $u = \dim\{\mathcal{E}_{\mathbf{M}}(\mathcal{U})\}$, let $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times u}$ be a semi-orthogonal basis matrix for $\mathcal{E}_{\mathbf{M}}(\mathcal{U})$, let $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ be an orthogonal matrix, let $\widehat{\mathbf{M}}$ be a $\sqrt{n}$-consistent estimator of $\mathbf{M}$, and let $\widehat{\mathbf{U}}$ be a positive semi-definite $\sqrt{n}$-consistent estimator of a basis matrix $\mathbf{U}$ for $\mathcal{U}$. With u specified, the most common objective function used for envelope estimation is

$$L_u(\mathbf{G}) = \ln|\mathbf{G}^\top\widehat{\mathbf{M}}\mathbf{G}| + \ln|\mathbf{G}^\top(\widehat{\mathbf{M}} + \widehat{\mathbf{U}})^{-1}\mathbf{G}|, \tag{3}$$

and the envelope is estimated as $\widehat{\mathcal{E}}_{\mathbf{M}}(\mathcal{U}) = \mathrm{span}\{\arg\min L_u(\mathbf{G})\}$, where the minimum is taken over all semi-orthogonal matrices $\mathbf{G} \in \mathbb{R}^{r\times u}$. Objective function (3) corresponds to maximum likelihood estimation under normality for many envelopes, including those associated with (1). Otherwise it provides a $\sqrt{n}$-consistent estimator of the projection onto $\mathcal{E}_{\mathbf{M}}(\mathcal{U})$ provided $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{U}}$ are $\sqrt{n}$-consistent (Cook and Zhang [7], who also provided additional background on $L_u(\mathbf{G})$).