Envelopes and tensor linear regression

Xin (Henry) Zhang; Department of Statistics, Florida State University

supported by FSU-CRC FYAP program in Summer 2015

1. Envelopes in multivariate linear model

Multivariate linear model of $\mathbf{Y}_i \in \mathbb{R}^r$ on $\mathbf{X}_i \in \mathbb{R}^p$:

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

where $\boldsymbol{\epsilon}_i$ is i.i.d. error with mean 0 covariance $\boldsymbol{\Sigma} > 0$, and is independent of $\mathbf{X}_i$.
Goal: efficient estimation of $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$.
Response Envelope model: Suppose there is a subspace $\mathcal{E} \subseteq \mathbb{R}^r$, and let $\mathbf{P}_\mathcal{E}$ and $\mathbf{Q}_\mathcal{E} = \mathbf{I}_r - \mathbf{P}_\mathcal{E}$ denote projections onto $\mathcal{E}$ and $\mathcal{E}^\perp$, such that

$$\mathbf{Q}_\mathcal{E}\mathbf{Y}|\mathbf{X} \sim \mathbf{Q}_\mathcal{E}\mathbf{Y}, \quad \mathbf{Q}_\mathcal{E}\mathbf{Y} \perp \mathbf{P}_\mathcal{E}\mathbf{Y}|\mathbf{X}. \tag{2}$$

- $\mathbf{P}_\mathcal{E}\mathbf{Y}$: material part
- $\mathbf{Q}_\mathcal{E}\mathbf{Y}$: immaterial part

Equivalently:

$$\mathrm{span}(\boldsymbol{\beta}) \subseteq \mathcal{E}, \quad \boldsymbol{\Sigma} = \mathbf{P}_\mathcal{E}\boldsymbol{\Sigma}\mathbf{P}_\mathcal{E} + \mathbf{Q}_\mathcal{E}\boldsymbol{\Sigma}\mathbf{Q}_\mathcal{E}. \tag{3}$$

The envelope is then the smallest such subspace $\mathcal{E}$.
Parameters in the envelope regression:

$$\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\theta}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{4}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times u}$ is a semi-orthogonal basis for the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r\times(r-u)}$ is the orthogonal completion of $\boldsymbol{\Gamma}$, $\boldsymbol{\theta} \in \mathbb{R}^u$, $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u)\times(r-u)}$.

2. An example: Cattle data from Kenward (1987)

- Compare two treatment for the control of the parasite
- 30 cows were randomly assigned to each treatment
- Weights were measured at weeks 2, 4, ..., 18, 19

The model is $\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}X_i + \boldsymbol{\epsilon}_i$, where $\mathbf{Y}_i \in \mathbb{R}^{10}$ is the weight profile of each cow and $X_i \in \{0, 1\}$ indicating two groups.

- Standard estimation: $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_0$.
- Envelope estimation: $\widehat{\boldsymbol{\beta}}_{\mathrm{Env}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\theta}}$ estimated via maximizing the likelihood function.
- Comparing two methods: the bootstrap standard error of each regression coefficient $\widehat{\beta}_{\mathrm{Env},k}$ is **2.6 to 5.9 times smaller than** that of $\widehat{\beta}_{\mathrm{OLS},k}$ for $k = 1, \ldots, 10$.

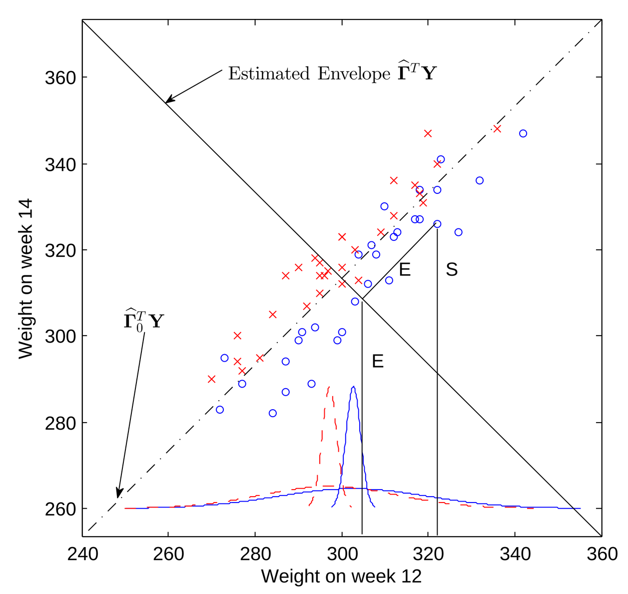

Figure 1: *Visualize the working mechanism of envelope regression: a simpler regression problem of the bivariate response $\mathbf{Y} = (Y_6, Y_7)$ on the binary predictor X of the cattle data.*

3. Envelope models and methods for tensor regression

Motivations:

1. Data in the form of tensor (multidimensional array) are becoming more and more common in both scientific and business applications, especially in brain imaging analysis.
2. Envelope method is a new and fast evolving tool for dimension reduction and improving efficiency in multivariate parameter estimation. Substantial gains are achievable by incorporating envelope method to classical regression problems such as OLS, PLS, RRR, GLM, etc.
3. We propose a parsimonious tensor envelope regression of a tensor-valued response on a scalar- or vector-valued predictor. It models all voxels of the tensor response jointly, while accounting for the inherent structural information among the voxels. Efficiency gain is achieved with improved interpretation.

Some tensor notations:

- Multidimensional array $\mathbf{A} \in \mathbb{R}^{r_1\times\cdots\times r_m}$ is called an m-th order tensor.
- Mode-k matricization turns a tensor $\mathbf{A}$ into a matrix $\mathbf{A}_{(k)} \in \mathbb{R}^{r_k\times(\prod_{j\neq k} r_j)}$.
- Mode-k product of a tensor $\mathbf{A}$ and a matrix $\mathbf{B} \in \mathbb{R}^{d\times r_k}$ is defined as $\mathbf{A}\times_k\mathbf{B} \in \mathbb{R}^{r_1\times\cdots\times r_{k-1}\times d\times r_{k+1}\times\cdots\times r_m}$.
- We write $\mathbf{A} = [\![\mathbf{C}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(m)}]\!]$ for the Tucker decomposition, which is defined as $\mathbf{A} = \mathbf{C}\times_1\mathbf{B}^{(1)}\times_2\cdots\times_m\mathbf{B}^{(m)}$, where $\mathbf{C} \in \mathbb{R}^{d_1\times\cdots\times d_m}$ is the core tensor and $\mathbf{B}^{(k)} \in \mathbb{R}^{r_k\times d_k}$, $k = 1, \ldots, m$, are factor matrices.

Tensor response regression

- $\mathbf{Y}_i \in \mathbb{R}^{r_1\times\cdots\times r_m}$ tensor-valued response on $\mathbf{X}_i \in \mathbb{R}^p$ vector-valued predictor, $i = 1, \ldots, n$ i.i.d. samples.
- $\boldsymbol{\epsilon}_i \in \mathbb{R}^{r_1\times\cdots\times r_m}$ error tensor with mean 0 and covariance $\mathrm{cov}\{\mathrm{vec}(\boldsymbol{\epsilon})\} = \boldsymbol{\Sigma}$ of size $(\prod_{k=1}^m r_k)^{\otimes 2}$.
- We assume a separable Kronecker covariance structure: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_m\otimes\cdots\otimes\boldsymbol{\Sigma}_1$.
- Tensor linear model:

$$\mathbf{Y}_i = \mathbf{B}\times_{(m+1)}\mathbf{X}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n. \tag{5}$$

- Vectorized model: $\mathrm{vec}(\mathbf{Y}_i) = \mathbf{B}_{(m+1)}^T\mathbf{X}_i + \mathrm{vec}(\boldsymbol{\epsilon}_i)$.
- Goal: estimating $\mathbf{B} \in \mathbb{R}^{r_1\times\cdots\times r_m\times p}$. For example, a standard way is fitting individual elements of $\mathbf{Y}$ on $\mathbf{X}$ one-at-a-time.

Tensor envelope: $\mathcal{T}_{\boldsymbol{\Sigma}}(\mathbf{B}) = \mathcal{E}_{\boldsymbol{\Sigma}_m}(\mathbf{B}_{(m)})\otimes\cdots\otimes\mathcal{E}_{\boldsymbol{\Sigma}_1}(\mathbf{B}_{(1)})$ is the intersection of all reducing subspaces $\mathcal{E}$ of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_m\otimes\cdots\otimes\boldsymbol{\Sigma}_1$ that contain $\mathrm{span}(\mathbf{B}_{(m+1)}^T)$ and can be written as $\mathcal{E} = \mathcal{E}_m\otimes\cdots\otimes\mathcal{E}_1$, where $\mathcal{E}_k \subseteq \mathbb{R}^{r_k}$, $k = 1, \ldots, m$.
Tensor envelope parameterization:

- Let $(\boldsymbol{\Gamma}_k, \boldsymbol{\Gamma}_{0k}) \in \mathbb{R}^{r_k\times r_k}$ be an orthogonal matrix such that $\mathrm{span}(\boldsymbol{\Gamma}_k) = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\mathbf{B}_{(k)})$, $\boldsymbol{\Gamma}_k \in \mathbb{R}^{r_k\times u_k}$.
- Regression coefficient tensor

$$\mathbf{B} = [\![\boldsymbol{\Theta}; \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_m, \mathbf{I}_p]\!] \quad \text{for some } \boldsymbol{\Theta} \in \mathbb{R}^{u_1\times\cdots\times u_m\times p}$$

- Covariance matrices

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k\boldsymbol{\Omega}_k\boldsymbol{\Gamma}_k^T + \boldsymbol{\Gamma}_{0k}\boldsymbol{\Omega}_{0k}\boldsymbol{\Gamma}_{0k}^T, \quad k = 1, \ldots, m$$

- Total number of parameters is reduced by

$$p\{\prod_{k=1}^m r_k - \prod_{k=1}^m r_k\}$$

4. Estimation

1. Initialize $\mathbf{B}^{(0)}$ and $\boldsymbol{\Sigma}^{(0)} = \boldsymbol{\Sigma}_m^{(0)}\otimes\cdots\otimes\boldsymbol{\Sigma}^{(m)}$ from standard methods.
2. [Numerical Grassmannian optimization] Estimate envelope basis $\{\boldsymbol{\Gamma}_k\}_{k=1}^m$ based on $\mathbf{B}^{(0)}$ and $\boldsymbol{\Sigma}^{(0)}$. The 1D envelope algorithm (Cook and Zhang 2014) is used to obtain a stable and $\sqrt{n}$-consistent envelope basis estimates.
3. [Analytical solutions] Estimate other parameters $\boldsymbol{\Theta}$, $\{\boldsymbol{\Omega}_k\}_{k=1}^m and \{\boldsymbol{\Omega}_{0k}\}_{k=1}^m$ based on $\{\boldsymbol{\Gamma}_k\}_{k=1}^m$.
4. [Analytical solutions] Obtain $\mathbf{B}$ and $\boldsymbol{\Sigma}$ from the envelope parameterization.

5. Some numerical results

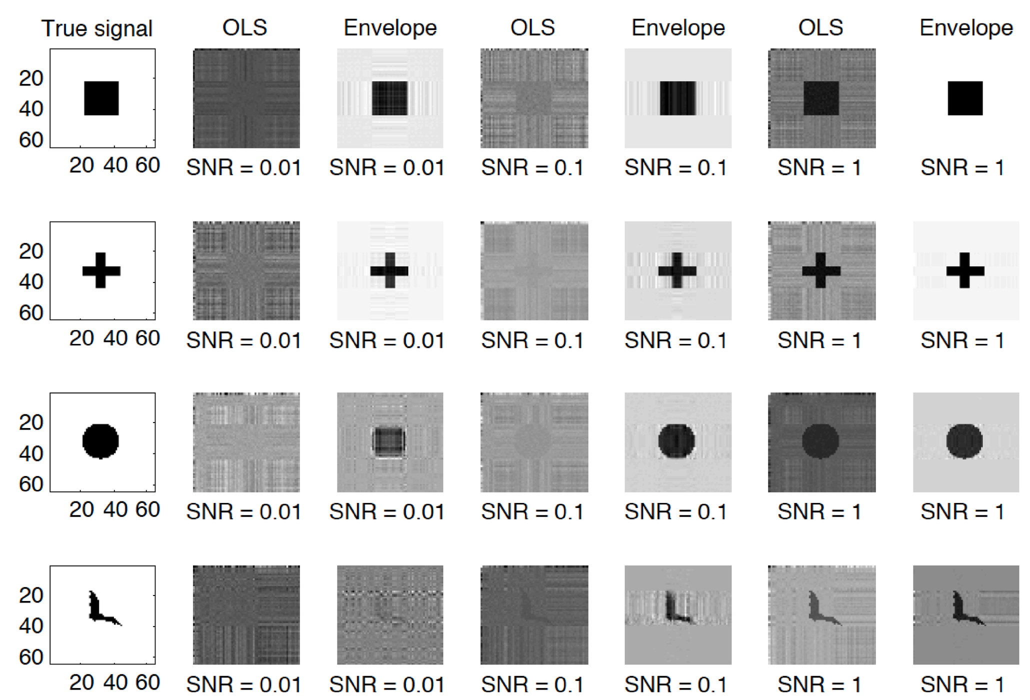

Figure 2: *Comparison with OLS: The true and estimated regression coefficient tensors under various signal shapes and signal-to-noise ratios (SNR).*

5.1 Simulations

To visualize the regression coefficient tensor $\mathbf{B}$ and its estimators, we consider the following matrix-valued (order-2 tensor) response regression model,

$$\mathbf{Y}_i = \mathbf{B}X_i + \sigma\cdot\boldsymbol{\epsilon}_i, \quad, i = 1, \ldots, n,$$

X_i is either 0 or 1; $\boldsymbol{\epsilon}_i$ follows a matrix normal distribution with covariance $\|\boldsymbol{\Sigma}_1\|_F = \|\boldsymbol{\Sigma}_2\|_F = 1$, $\sigma > 0$ controls the signal-to-noise-ratio (SNR) $\mathbf{Y}_i$, $\boldsymbol{\epsilon}_i$ and $\mathbf{B}$ all have the same dimension 64×64
Sample size is small: $n = 20$

5.2 ADHD data analysis

285 combined ADHD subjects and 491 normal controls
comparing two groups after adjusting for age and sex (i.e. number of predictors $p = 3$)
downsized MRI images from $256\times198\times256$ to $30\times36\times30$
$\mathbf{B}$ has the dimension $30\times36\times30\times3 \Rightarrow 97,200$ coefficients

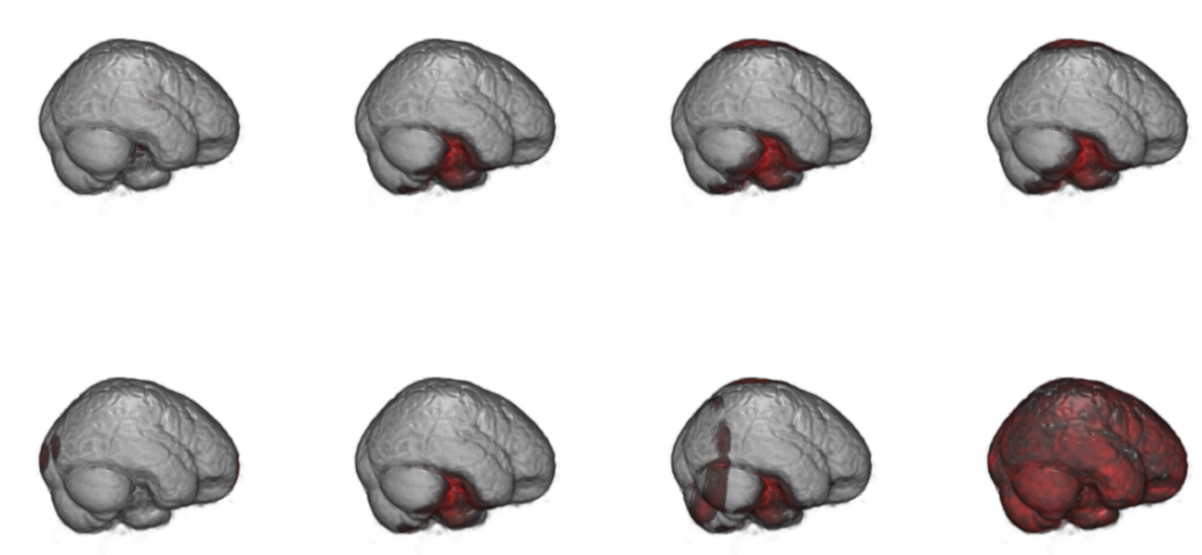

Figure 3: *ADHD Coefficients. Top row: $u_1 = u_2 = 10$ and u_3 varies as $\{1, 2, 10, 20\}$, where $u_1 = u_2 = u_3 = 10$ is selected by BIC if we force the three dimensions to be the same. Bottom row: (u_1, u_2, u_3) varies as $\{(8, 9, 1),\ (9, 10, 2),\ (10, 11, 3),\ (30, 30, 36)(OLS)\}$, where $(u_1, u_2, u_3) = (9, 10, 2)$ is selected by BIC.*

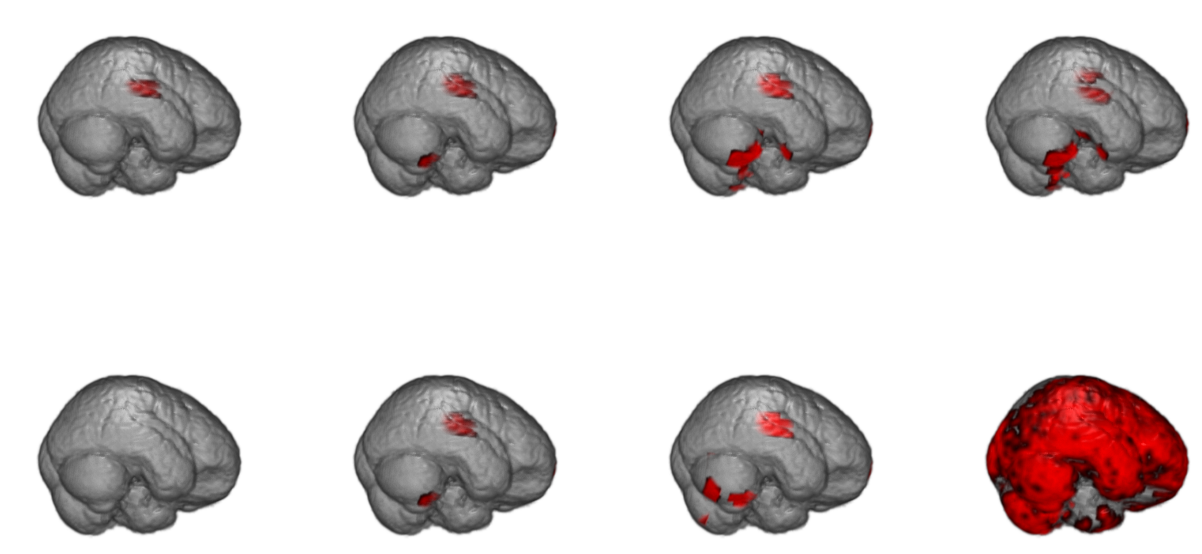

Figure 4: *ADHD P-value maps. Red regions represent $p < 0.05$.*

6. Key References

Cook, R.D. and Zhang, X. (2015), Foundations for envelope models and methods, *J. of Amer. Stat. Assoc.*, **110**, 599–611.
Cook, R.D. and Zhang, X. (2016), Algorithms for envelope estimation, *J. of Comput. Graph. Stat.*, In press.
Li, L. and Zhang, X. (2015), Parsimonious Tensor Response Regression, arXiv preprint arXiv:1501.07815