Econometrics **May 12, 2023**

Topic 15: Sparse Orthogonal Factor Regression

by Sai Zhang

Key points: Sparcity and dimensionality reduction for Multivariate Linear Regression models.

Disclaimer: *The note is built on Prof. Jinchi Lv's lectures of the course at USC, DSO 607, High-Dimensional Statistics and Big Data Problems.*

15.1 Motivation

Consider a Mutlivariate Linear Regression (MLR) model

$$\underset{n\times q}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \cdot \underset{p\times q}{\mathbf{C}} + \underset{n\times q}{\mathbf{E}}$$

How to apply regularization methods to this model? There are several approaches to consider

- **Shrinkage**: ridge regression to overcome multicollinearity
- **sparsity**: variable selection in multivariate setting
- **Reduced-rank**
  - **Dimension reduction** via reducing rank of **C**
  - $\min \|\mathbf{Y} - \mathbf{XC}\|_F^2$ s.t. $\text{rank}(\mathbf{C}) \leq r$
- **Combinations**
- **Low-rank** plus **sparse decomposition**: robust PCA, latent variable graphical models, covariance estimation
- **Regularized matrix** or **tensor regression**

Or, we can introduce a very attractive sparsity structure to achieve simultaneous dimension reduction and variable selection. This structure should be characterized by

- Having a few **distinct** channels/pathways relating responses and predictors
- Each of such associations may involve only **a smaller subset**, but not all of the responses and predictors

that is

$$\begin{aligned}
\mathbf{Y} &= \mathbf{XC} + \mathbf{E} \\
&= \mathbf{X} \cdot \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1q} \\ c_{21} & c_{22} & \cdots & c_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pq} \end{pmatrix} + \mathbf{E} \\
&= \mathbf{X} \cdot \begin{pmatrix} 0 & u_{12} & \cdots & u_{1r} \\ u_{21} & 0 & \cdots & c_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pr} \end{pmatrix} \cdot \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & \cdots & v_{q1} \\ v_{12} & v_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ v_{1r} & v_{2r} & \cdots & v_{qr} \end{pmatrix} + \mathbf{E}
\end{aligned}$$

This way, we can have

- **Sparsity**: selection of both **latent** and **original** variables
- **Low-rank SVD**: different subsets of responses allowed to be associated with different subsets of predictors

Consider an example:

Example 15.1.1: Dimension Reduction and Variable Selection via Sparse SVD

Consider the case where $p = 1000, q = 100$, then C, as a $p \times q$ matrix, contains 100000 coefficients. Meanwhile, for a rank-3 SVD model:

$$\mathbf{C} = d_1 \mathbf{u}_1 \mathbf{v}_1' + d_2 \mathbf{u}_2 \mathbf{v}_2' + d_3 \mathbf{u}_3 \mathbf{v}_3'$$

where $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are all $p \times 1$, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are all $q \times 1$, d_1, d_2, d_3 are all scalars. Hence, there are only $3 \times (1000 + 100 + 1) = 3303$ paramaters to estimate. If futher assume sparcity, the dimension would be even lower.