Econometrics **May 4, 2023**

Topic 14: Regularization Methods in Thresholded Parameter Space

by Sai Zhang

Key points: The connections and differences of all regularization methods and some interesting phase transition phenomena.

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.*

14.1 Model Setup

Now, consider a generalized linear model (GLM) linking a p-dimensional predictor $\mathbf{x}$ to a scalar response Y. With canonical link, the conditional distribution of Y given $\mathbf{x}$ has density

$$f(y;\theta,\phi) = \exp\left[y\theta - b(\theta) + c(y,\phi)\right]$$

where $\theta = \mathbf{x}'\boldsymbol{\beta}$ with $\boldsymbol{\beta}$ a $p-$dimensional regression coefficient vector, $b(\dot{})$ and $c(\cdot,\cdot)$ are know functions and ϕ is dispersion parameter. Again, $\boldsymbol{\beta} = \left(\beta_{0,1},\cdots,\beta_{0,p}\right)'$ is sparse with many zero components, and $\log p = O(n^a)$ for some $0 < a < 1$.

The penalized negative log-likelihood is

$$Q_n(\boldsymbol{\beta}) = -n^{-1}\left[\mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \mathbf{1}'\mathbf{b}(\mathbf{X}\boldsymbol{\beta})\right] + \|p_\lambda(\boldsymbol{\beta})\|_1$$

where

- $\mathbf{y} = \left(y_1,\cdots,y_n\right)'$, $\mathbf{X} = (\mathbf{x}_1,\cdots,\mathbf{x}_n)'$, each column of $\mathbf{X}$ is rescaled to have L_2-norm $\sqrt{n}$
- $\mathbf{b}(\boldsymbol{\theta}) = (b(\theta_1),\cdots,b(\theta_n))'$ with $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_n)'$
- $\|p_\lambda(\boldsymbol{\beta})\|_1 = \sum_{j=1}^{p} p_\lambda(|\beta_j|)$