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 $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_n)'$
- $\|p_\lambda(\boldsymbol{\beta})\|_1 = \sum_{j=1}^{p} p_\lambda(|\beta_j|)$

Next, define **robust spark** κ_c

Definition 14.1.1: Robust spark κ_c

The robust spark κ_c of the $n \times p$ design matrix $\mathbf{X}$ is defined as the smallest possible positive integer s.t. there exists an $n \times \kappa_c$ submatrix of $\frac{1}{\sqrt{n}}\mathbf{X}$ having a singular value less than a given positive constant c (Zheng et al., 2014), and

$$\kappa_c \le n+1$$

Bounding sparse model size can control collinearity and ensure model identifiability and stability, and as $c \to 0+$, κ_c approaches the spark. Robust spark can be some large number diverging with n:

Proposition 14.1.2: Order of κ_c

Assume $\log p = o(n)$ and that the rows of the $n \times p$ random design matrix $\mathbf{X}$ are i.i.d. as $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ has smallest eigenvalue bounded from below by some positive constant. Then there exist positive constants c and $\tilde{c}$ s.t. with asymptotic probability one, $\kappa_c \geq \frac{\tilde{c}n}{\log p}$

Next, we define a thresholded parameter space

Definition 14.1.3: Thresholded parameter space

$$\mathcal{B}_{\tau,c} = \left\{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\|_0 < \frac{\kappa_c}{2}, \text{ and for each } j, \beta_j = 0 \text{ or } |\beta_j| \geq \tau\right\}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$. τ is some positive threshold on parameter magnitude:

Here, τ is very important:

- τ is key to distinguishing between important covariates and noise covariates for the purpose of variable selection
- τ typically needs to satisfy $\tau\sqrt{n/\log p} \xrightarrow{n\to\infty} \infty$

It turns out that the solution to the regularizaiton problem has the (very natural) hard-thresholding property:

Proposition 14.1.4: Hard-thresholding property

or the L_0-penalty $p_\lambda(t) = \lambda \mathbf{1}_{t\neq 0}$, the global minimizer $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_p)'$ of the regularization problem over $\mathbb{R}^p$ satisfies that each component $\hat{\beta}_j$ is either 0 or has magnitude larger than some positive threshold

This hard-thresholding property is shared by many other penalties such as SICA penalties. This property guarantees sparcity of the model: weak signals are generally difficult to stand out comparing to noise variables due to impact of high dimensionality

References

Zemin Zheng, Yingying Fan, and Jinchi Lv. High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 627–649, 2014.