

## 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

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

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

The penalized negative log-likelihood is

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

where

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

Next, define **robust spark**  $\kappa_c$

#### Definition 14.1.1: Robust Spark $\kappa_c$

The robust spark  $\kappa_c$  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

$$\kappa_c \leq n + 1$$

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

#### Proposition 14.1.2: Order of $\kappa_c$

## 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.