

## Topic 13: Non-convex Learning + Lasso

by Sai Zhang

**Key points:** Combining the best of the two, we can use **Lasso plus Concave** method, with Lasso screening and concave component selecting variables, achieving a coordinated intrinsic two-scale learning.

**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.

We are facing a tradeoff:

- **Convex** methods: have appealing prediction power and oracle inequalities, but challenging to provide tight false sign rate control
- **Concave** methods: have good variable selection properties, but challenging to establish global properties and risk properties

Here, we take advantage of the linearity of Lasso (convex *and* concave) and try to combine it with concave regularization to get the best of both.

## 13.1 Model Setup

Again, consider a linear regression model  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , where

- response vector ( $n \times 1$ ):  $\mathbf{y} = (y_1, \dots, y_n)'$
- design matrix ( $n \times p$ ):  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$

here, we consider a scenario where

- $\boldsymbol{\beta}_0 = (\beta_{0,1}, \dots, \beta_{0,p})'$  is *sparse* (with many 0 components)
- ultra-high dimensions:  $\log p = O(n^a)$ , for some  $0 < a < 1$

and consider the penalized least squares

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ (2n)^{-1} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_0 \|\boldsymbol{\beta}\|_1 + \right\} \quad (13.1)$$