Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima

Po-Ling Loh and Martin Wainwright

Department of Statistics, UC Berkeley

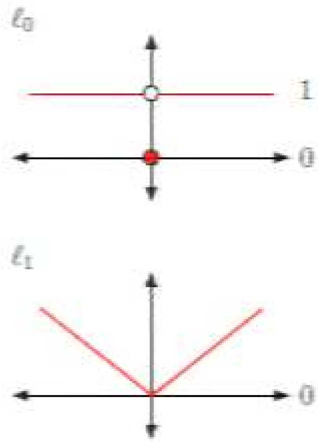
- High-dimensional statistical estimation problems:

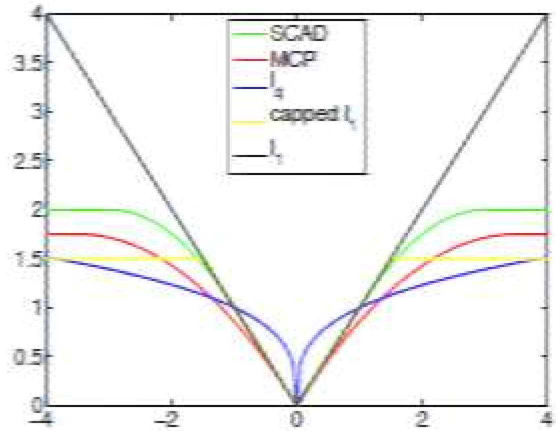
$$\widehat{\beta} \in \arg\min_{\beta} \left\{ \underbrace{\frac{1}{n}\sum_{i=1}^{n} (y_i - x_i^T \beta)^2}_{M\text{-estimator}} + \underbrace{\lambda \|\beta\|_1}_{\text{regularizer}} \right\}$$

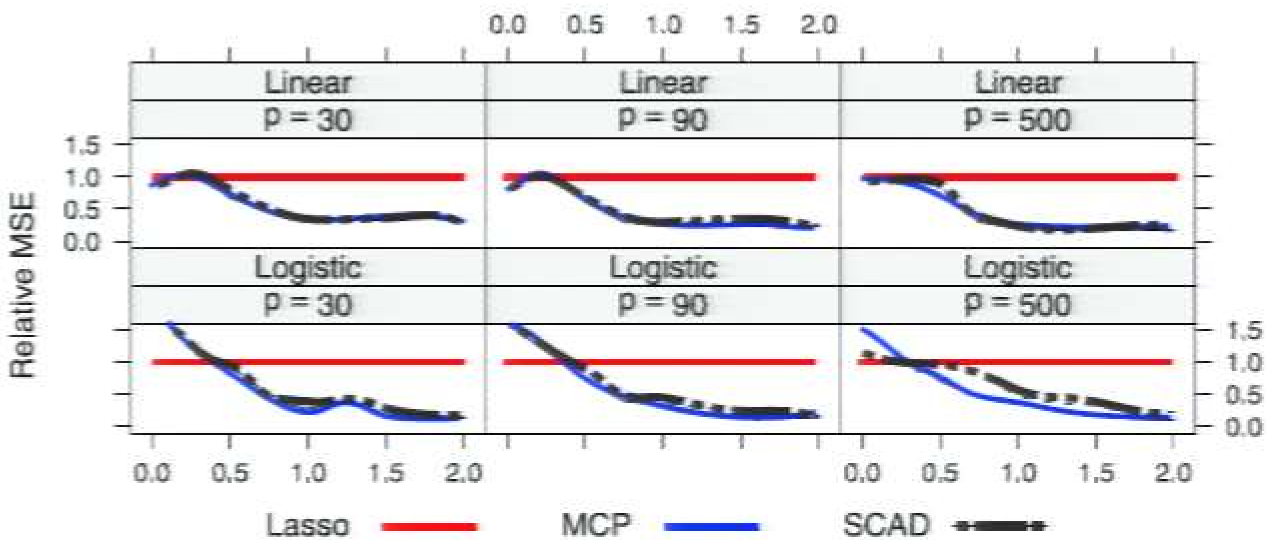
- ℓ_1 is **convex relaxation** of ℓ_0, encourages sparsity

Benefits of nonconvex regularizers

- **BUT** ℓ_1 penalizes large coefficients severely, causing **solution bias**







Breheny & Huang, '11

How to find optima??

- Optimization algorithms only guaranteed to find **local optima**
- Statistical consistency only established for **global optima**

GAP IN THEORY

- Previous theoretical results:
  - Sufficiently good initialization points
  - Sufficiently sparse local optima

Our contributions

Result 1:

- Sufficient conditions on loss/regularizer to guarantee *all local optima* are close

$$\underbrace{\|\widetilde{\beta} - \beta^*\|_2}_{\text{optimization error}} = \mathcal{O}(\underbrace{\|\widehat{\beta} - \beta^*\|_2}_{\text{statistical error}})$$

Result 2:

- Composite gradient descent locates local optima in log-linear time

$$\|\beta^t - \widehat{\beta}\|_2 \leq C\left(\kappa^t\|\beta^0 - \widehat{\beta}\|_2 + \mathcal{O}(\|\widehat{\beta} - \beta^*\|_2)\right)$$

Come to our poster!! **Sat38**