Robust High-dimensional Statistics I:

Penalized M-Estimation

Dr. Abhik Ghosh

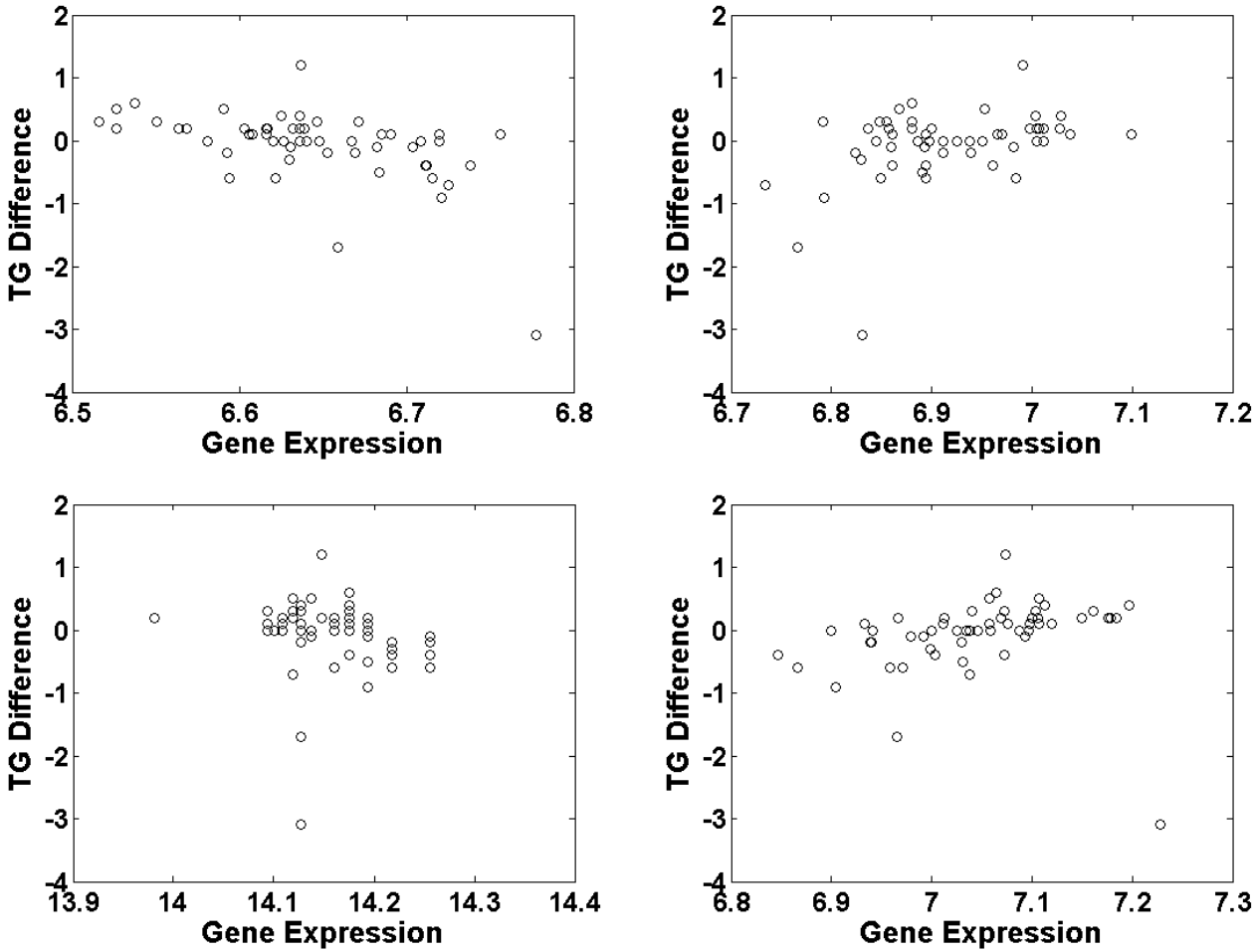
Indian Statistical Institute, Kolkata, India.

2022

Table of contents

Outline

Outliers in Covariates may lead to incorrect results!

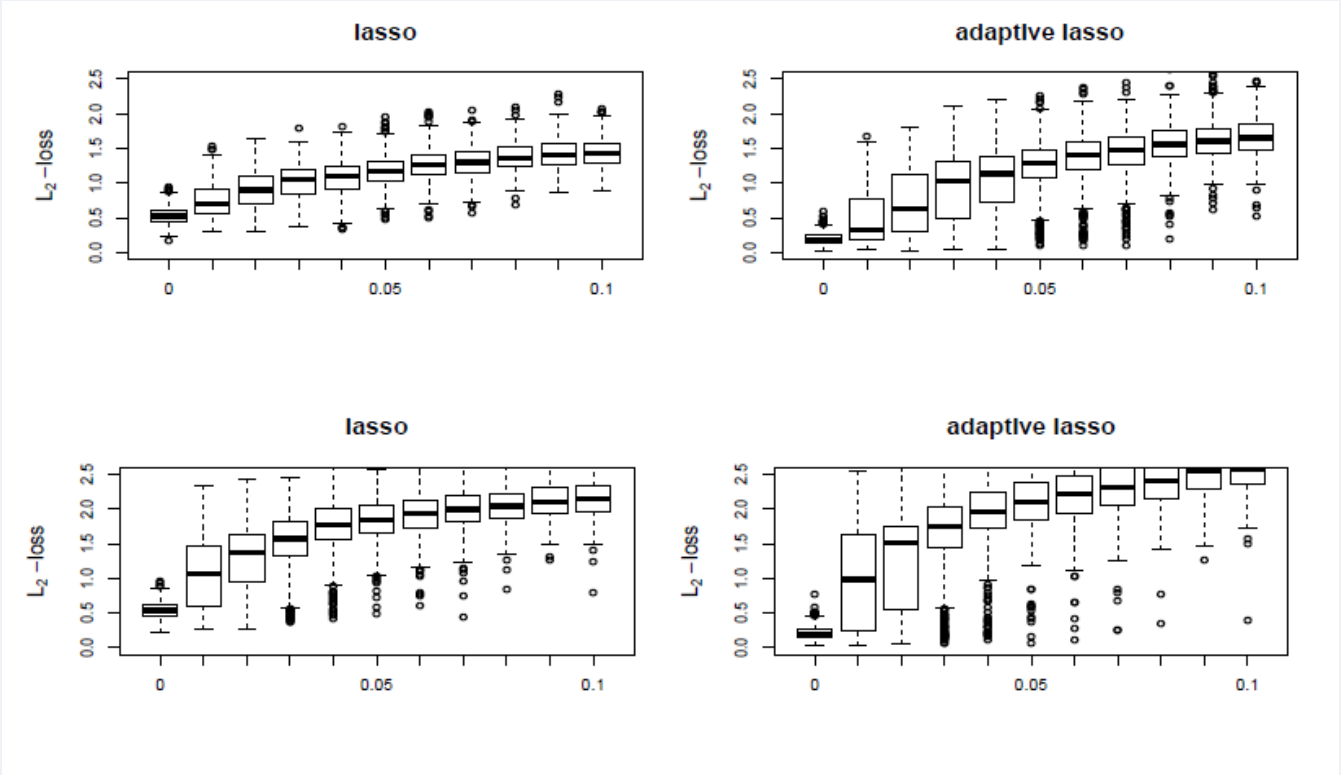


It is difficult to examine for such outlier effects for a vast pool of covariates in high-dimensional data!!

Prediction Accuracy under high-dimensional Poisson regression model!

Simulation results with outlying points – bottom panel is for more distant outliers

(Avella-Medina and Ronchetti, 2018)



Prediction accuracy increases significantly with slight increase in contamination proportion and also their distances!!

Why does A Robust Method Help?

1. High-dimensional data are extremely prone to outliers or noise contamination.
2. It is difficult to separately identify and eliminate outliers from large high-dimensional datasets.
3. **Robust methods can generate stable results even in the presence of outliers or any other type of noises in the data**.
4. **Robust procedures, developed under a parametric model, are much more efficient compared to any non-parametric methods**.

One Approach: **Penalized M-estimators**

Outline

M-Estimators in Low-dimensional Set-ups

Generalized linear model (GLM): Given a covariate value $\mathbf{X} = \boldsymbol{x}$, the response variable Y has density

$$f(y; \boldsymbol{x}^T\beta) = \exp\{y\theta - b(\theta) + c(y)\}, \quad \text{with} \quad E[Y|\boldsymbol{x}] = b'(\theta) = g^{-1}(\boldsymbol{x}^T\beta), \tag{1}$$

M-estimator of β is defined as the solution of $\frac{1}{n}\sum_{i=1}^n \psi(\boldsymbol{z}_i, \beta) = \mathbf{0}_p$, where $\boldsymbol{z}_i = (y_i, \mathbf{x}_i)$.

$$\widehat{\beta}_M = \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^n \rho(\boldsymbol{z}_i, \beta). \tag{2}$$

Example

- ρ is the negative log-likelihood – MLE!
- For linear regression, if we put $r_i = (y_i - \mathbf{x}_i\beta)$ then:
  - $\rho(\boldsymbol{z}_i, \beta) = \frac{r_i^2}{2}$ and $\psi(\boldsymbol{z}_i, \beta) = r_i$ – OLS!
  - A popular robust choice is the **Huber's loss** given by

$$\rho_k(\boldsymbol{z}_i, \beta) = \frac{r_i^2}{2} I(|r_i| \le k) + k(|r_i| - k/2) I(|r_i| > k), \quad \psi_k(\boldsymbol{z}_i, \beta) = r_i I(|r_i| \le k) + k\,\text{sign}(r_i) I(|r_i| > k)$$

Down-weights the contributions of observations with large residuals (outliers)!

Penalized M-estimators

$$\widehat{\boldsymbol{\beta}}_M = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{z}_i, \boldsymbol{\beta}) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \right\}$$

where $p_\lambda(\cdot)$ is a penalty function (lasso, SCAD, MCP, ...).

It is a special case of general penalized estimator!

We must use an appropriately chosen ρ function in order to get robust results!

Selection of the regularization parameter λ

- Cross-Validation is not useful! (why?)
- **Robust Extended BIC** (Avella-Medina and Ronchetti, 2018): Minimize

$$\text{RobEBIC}(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{z}_i, \widehat{\boldsymbol{\beta}}) + \left(\frac{\log n}{n} + \gamma \frac{\log p}{n} \right) |\widehat{S}|,$$

with $\gamma \in [0, 1]$. Suggested choice $\gamma = 0.5$.

Measure of Robustness: Influence Functions

The **Influence Function (IF)** is a classical tool of measuring the asymptotic local robustness of any estimator (Hampel et al., 1986).

Consider a contaminated version of the true joint distribution G given by $G_\epsilon = (1-\epsilon)G + \epsilon \wedge_{\boldsymbol{z}_t}$, where ϵ is the contamination proportion
and $\wedge_{\boldsymbol{z}_t}$ is the degenerate distribution at the contamination point $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t)$.
Then, the IF of any functional $\boldsymbol{T}$ at G is defined as the *limiting (standardized) bias due to infinitesimal contamination*:

$$\mathcal{IF}(\boldsymbol{z}_t, \boldsymbol{T}, G) = \lim_{\epsilon \to 0} \frac{\boldsymbol{T}(G_\epsilon) - \boldsymbol{T}(G)}{\epsilon} = \frac{\partial}{\partial \epsilon} \boldsymbol{T}(G_\epsilon)\Big|_{\epsilon=0}.$$

Boundedness of the IF in contamination points $\boldsymbol{z}_t$ indicates (local bias) robustness of T (why?).

Higher the maximum absolute value of the IF, lower the extent of robustness is!!

IF of Penalized M-estimators

$$\widehat{\beta}_M = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{z}_i, \beta) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \right\}.$$

$$T(G) = \arg\min_{\beta} \left\{ \int \rho(\boldsymbol{z}, \beta) dG(\boldsymbol{z}) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \right\}.$$

Avella-Medina (2017)

$$\mathcal{IF}(\boldsymbol{z_t}, \boldsymbol{T}, \mathbf{G}) = -\left[\boldsymbol{M_{S,S}}(\boldsymbol{\beta_0}) + \mathbf{P_2}(\boldsymbol{\beta_0})\right]^{-1} \left\{\psi_{\mathbf{S}}(\boldsymbol{z_t}, \boldsymbol{\beta_0}) + \mathbf{P_1}(\boldsymbol{\beta_0})\right\}, \tag{3}$$

where $\beta_0 = T(G)$, $S = \{j : \beta_{0j} \neq 0\}$, and $\psi(\boldsymbol{z}, \beta) = \frac{\partial}{\partial \beta} \rho(\boldsymbol{z}, \beta)$, $\boldsymbol{M}(\beta) = E_G \left[\frac{\partial^2}{\partial \beta^2} \rho(\boldsymbol{Z}, \beta) \right]$,

$$P_1(\beta_0) = \left(p'_\lambda(|\beta_{0j}|)\text{sign}(\beta_{0j}) \ : \ j \in S\right), \quad P_2(\beta_0) = \text{Diag}\left\{p''_\lambda(|\beta_{0j}|) \ : \ j \in S\right\}.$$

A penalized M-estimator will be (locally bias) robust if $\psi(\boldsymbol{z}, \beta) = \frac{\partial}{\partial \beta} \rho(\boldsymbol{z}, \beta)$ is bounded in $\boldsymbol{z}$.

Outline

Penalized M-estimators

$$\widehat{\boldsymbol{\beta}}_M = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{z}_i, \boldsymbol{\beta}) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \right\}$$

Theoretical Properties

- ρ is convex – General Decomposable Regularizers (Negahban et al., 2012)
- ρ is non-convex due to involvement of additional nuisance parameter.
- ρ may be itself non-convex, even without involving any nuisance parameter! (common for robust inference)

Non-Convex Penalized M-estimators: General Assumptions

M-estimator based on general data $\mathbb{Z}_n = \{\boldsymbol{z}_i, i = 1, \ldots, n\}$, may be defined as

$$\widehat{\boldsymbol{\beta}}_{\mathbf{M}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}, \mathbb{Z}_{\mathbf{n}}) + \mathbf{P}_{\boldsymbol{\lambda}}(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}, \mathbb{Z}_n) = \frac{1}{n}\sum_{i=1}^{n} \rho(\boldsymbol{z}_i, \boldsymbol{\beta})$; Penalty $P_\lambda(\boldsymbol{\beta}) = \sum_{j=1}^{p} p_\lambda(|\beta_j|)$. Both may be non-convex.

Assumption (P)

- $p_\lambda(s)$ is differentiable and non-decreasing in $s \in [0, \infty)$ with $p'_\lambda(0+) = \lambda L > 0$.
- p_λ is symmetric around zero with $p_\lambda(0) = 0$.
- $p_\lambda(s)/s$ is non-increasing in s and $p_\lambda(s) + \frac{\mu}{2}t^2$ is convex for some $\mu > 0$.

LASSO, SCAD and MCP satisfy Assumption (P) [For $L = 1$, $\mu = 1, (a-1)^{-1}, a^{-1}$, respectively]

Restricted Strong Convexity (Loh and Wainwright, 2015)

$$\langle \nabla L_n(\boldsymbol{\beta}_0 + \boldsymbol{d}) - \nabla L_n(\boldsymbol{\beta}_0), \boldsymbol{d} \rangle \geq \begin{cases} \alpha_1 ||\boldsymbol{d}||_2^2 - \tau_1 \frac{\log p}{n} ||\boldsymbol{d}||_1^2, & \text{for all} \quad ||\boldsymbol{d}||_2 \leq 1, \\ \alpha_2 ||\boldsymbol{d}||_2 - \tau_2 \sqrt{\frac{\log p}{n}} ||\boldsymbol{d}||_1, & \text{for all} \quad ||\boldsymbol{d}||_2 \geq 1, \end{cases} \tag{4}$$

where α_js are strictly positive constants and theτ_js are nonnegative constants.

Non-Convex Penalized M-estimators: General Results

Feasible Solutions

Additional Constraint: $||\beta||_1 \leq R$ for some $R > 0$.

First order necessary Condition for a local M-estimator $\widehat{\beta}$

$$\left\langle \nabla L_n(\widehat{\beta}) + \nabla \rho_\lambda(\widehat{\beta}), \beta - \widehat{\beta} \right\rangle \geq 0, \qquad \text{for all feasible } \beta \in \mathbb{R}^p. \tag{5}$$

Main Results (Loh and Wainwright, 2015)

Suppose that Assumption (P) and Restricted Strong Convexity Condition hold with $3\mu < 4\alpha_1$ and take λ satisfying

$$\frac{4}{L} \max\left\{ ||\nabla L_n(\beta_0)||_\infty, \alpha_2 \sqrt{\frac{\log p}{n}} \right\} \leq \lambda \leq \frac{\alpha_2}{6RL}.$$

Then, if $n \geq 16R^2 \max\{\tau_1^2, \tau_2^2\}/\alpha_2^2$, we have the following results for any local M-estimator $\widehat{\beta}$ satisfying the first order equation (5):

$$||\widehat{\beta} - \beta_0||_2 \leq \frac{6\lambda L\sqrt{s}}{4\alpha_1 - 3\mu}, \qquad ||\widehat{\beta} - \beta_0||_1 \leq \frac{24\lambda L s}{4\alpha_1 - 3\mu},$$

Prediction error: $$\left\langle \nabla L_n(\widehat{\beta}) - L_n(\beta_0), \widehat{\beta} - \beta_0 \right\rangle \leq \lambda^2 L^2 s \left(\frac{9}{4\alpha_1 - 3\mu} + \frac{27\mu}{(4\alpha_1 - 3\mu)^2} \right).$$

Non-Convex Penalized M-estimators: GLMs

Generalized linear model (GLM): Given a covariate value $\mathbf{X} = \boldsymbol{x}$, the response variable Y has density

$$f(y; \boldsymbol{x}^T\beta) = \exp\{y\theta - b(\theta) + c(y)\}, \quad \text{with} \quad E[Y|\boldsymbol{x}] = b'(\theta) = g^{-1}(\boldsymbol{x}^T\beta). \tag{6}$$

In canonical form $\theta = \boldsymbol{x}^T\beta$, i.e., $b' = g^{-1}$.

Negative log-likelihood loss function $L_n(\beta) = -\frac{1}{n}\sum_{i=1}^n \left(y_i \mathbf{x}_i^T\beta - b(\mathbf{x}_i^T\beta)\right)$.

Main Results (Loh and Wainwright, 2015)

Suppose that Assumption (P) hold and the covariates are sub-Gaussian. Choose λ and R such that β_0 is feasible and $c\sqrt{\frac{\log p}{n}} \leq \lambda \leq \frac{c'}{R}$ for some $c, c' > 0$. Then, if $n \geq CR^2 \log p$, any local M-estimator $\widehat{\beta}$ satisfies

$$||\widehat{\beta} - \beta_0||_2 \leq \frac{c_0\lambda\sqrt{s}}{4\alpha_1 - 3\mu}, \qquad ||\widehat{\beta} - \beta_0||_1 \leq \frac{c_0'\lambda s}{4\alpha_1 - 3\mu},$$

with probability at least $1 - c_1\exp(-c_2 \log p)$, for an appropriate choice of α_1 satisfying $2\alpha_1 > \mu$.

Relaxing RSC condition for Linear Regression Models

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

Non-convex Penalized M-estimator

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \rho(y_i, \boldsymbol{x}_i; \boldsymbol{\beta})$; Penalty $P_\lambda(\boldsymbol{\beta}) = \sum_{j=1}^p p_\lambda(|\beta_j|)$. Both may be non-convex.
Additional feasibility constraint: $||\boldsymbol{\beta}||_1 \leq R$ for some $R > 0$.

Local Restricted Strong Convexity (Loh, 2017)

There exists $\alpha > 0$, $\tau \geq 0$ and $r > 0$ such that, for all $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{B}_r(\boldsymbol{\beta}_0)$

$$\langle \nabla L_n(\boldsymbol{\beta}_1) - \nabla L_n(\boldsymbol{\beta}_2), \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \rangle \geq \alpha ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_2^2 - \tau \frac{\log p}{n} ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||_1^2. \tag{7}$$

Main Results (Loh, 2017)

Suppose that Assumption (P) and Local Restricted Strong Convexity Condition hold with $L = 1$, $3\mu < 4\alpha$ and take λ satisfying $\lambda \geq \max\left\{4||\nabla L_n(\boldsymbol{\beta}_0)||_\infty, 8\tau R \frac{\log p}{n}\right\}$.

Then, if $n \geq Cs \log p / r^2$, there exists a local M-estimator $\widehat{\boldsymbol{\beta}}$ such that $||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_2 \leq r$ and additionally

$$||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_2 \leq \frac{24\lambda\sqrt{s}}{4\alpha - 3\mu}, \qquad ||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_1 \leq \frac{96\lambda s}{4\alpha - 3\mu}.$$

Example: Unweighted Regression M-estimators

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

Non-convex Penalized M-estimator: The Unweigheted version (Hampel et al., 1986)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \rho(\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i)$;

Assumptions on ρ

- There exists $\kappa 1, \kappa 2 \geq 0$ such that $|\rho'(u)| \leq \kappa 1$ and $\rho''(u) \geq -\kappa 2$ for all u.
- There exists a small $T > 0$ such that $\alpha_T = \min_{|u| \leq T} \rho''(u) > 0$. Put $\epsilon_T = P(|\epsilon_i| \geq T/2)$.

Sufficient Conndition for Local RSC (Loh, 2017)

Suppose that the above assumptions hold and x_is are sub-Gaussian with parameter τ^2 such that

$$c\tau^2 \left[\sqrt{\epsilon_T} + e^{-\frac{c'T^2}{\tau^2 r^2}}\right] \leq \frac{\alpha_T}{\alpha_T + \kappa 2}\frac{\Lambda_{\min}(\Sigma_x)}{2}.$$

Then, if $n \geq c_0 s \log p$, with probability at least $1 - c\exp(-c' \log p)$, the loss function L_n satisfies the Local SRC Condition with

$$\alpha = \alpha_T \frac{\Lambda_{\min}(\Sigma_x)}{16} \quad \text{and} \quad \tau = \frac{C(\alpha_T + \kappa 2)^2 \tau^2 T^2}{r^2}.$$

Example: Weighted Mallows M-estimators

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

Non-convex Penalized Mallows M-estimator: A Weigheted Version (Hampel et al., 1986)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n w(x_i)\rho(\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i)$, $\quad w(\mathbf{x}_i) = \min\{1, b/||B\mathbf{x}_i||_2\}$.

Assumptions on ρ

- There exists $\kappa 1, \kappa 2 \geq 0$ such that $|\rho'(u)| \leq \kappa 1$ and $\rho''(u) \geq -\kappa 2$ for all u.
- There exists a small $T > 0$ such that $\alpha_T = \min_{|u|\leq T} \rho''(u) > 0$. Put $\epsilon_T = P(|\epsilon_i| \geq T/2)$.

Sufficient Conndition for Local RSC (Loh, 2017)

Suppose that the above assumptions hold and x_is are sub-Exponential with parameter τ such that

$$c\tau^2 b|||B^{-1}|||_2 \left[\sqrt{\epsilon_T} + e^{-\frac{c'T}{\tau r}}\right] \leq \frac{\alpha_T}{\alpha_T + \kappa 2}\frac{\Lambda_{\min}(E[w(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T])}{2}.$$

Then, if $n \geq c_0 s \log p$, with probability at least $1 - c\exp(-c' \log p)$, the loss function L_n satisfies the Local SRC Condition with

$$\alpha = \alpha_T \frac{\Lambda_{\min}(E[w(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T])}{16} \quad \text{and} \quad \tau = \frac{C(\alpha_T + \kappa 2)^2\tau^2 T^2}{r^2}.$$

Example: Weighted Hill-Ryan M-estimators

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

Non-convex Penalized Hill-Ryan M-estimator: A Weigheted Version (Hampel et al., 1986)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n w(x_i)\rho((\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i)w(x_i)), \qquad w(\mathbf{x}_i) = \min\{1, b/||B\mathbf{x}_i||_2\}.$

Assumptions on ρ

- There exists $\kappa 1, \kappa 2 \geq 0$ such that $|\rho'(u)| \leq \kappa 1$ and $\rho''(u) \geq -\kappa 2$ for all u.
- There exists a small $T > 0$ such that $\alpha_T = \min_{|u|\leq T} \rho''(u) > 0$. Put $\epsilon_T = P(|\epsilon_i| \geq T/2)$.

Sufficient Conndition for Local RSC (Loh, 2017)

Suppose that the above assumptions hold and

$$cb^2|||B^{-1}|||_2^2 \left[\sqrt{\epsilon_T} + e^{-\frac{c'T^2}{b^2|||B^{-1}|||_2^2\sigma_x^2 r^2}}\right] \leq \frac{\alpha_T}{\alpha_T + \kappa 2}\frac{\Lambda_{\min}(E[w(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T])}{2}.$$

Then, if $n \geq c_0 s \log p$, with probability at least $1 - c\exp(-c' \log p)$, the loss function L_n satisfies the Local SRC Condition with

$$\alpha = \alpha_T \frac{\Lambda_{\min}(E[w(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T])}{16} \quad \text{and} \quad \tau = \frac{C(\alpha_T + \kappa 2)^2 b^2 |||B^{-1}|||_2^2 T^2}{r^2}.$$

Variable Selection Optimality: Linear Regression

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero. Define $S = \{j : \beta_{0j} \neq 0\}$.

Non-convex Penalized M-estimator: The Unweigheted version

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}$$

Loss $L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \rho(\boldsymbol{x}_i^T\boldsymbol{\beta} - y_i)$;

Assumptions

- On penalty: Assumption (P) with $L = 1$ and $\rho'_\lambda(t) = 0$ for all $t \geq \gamma\lambda$ with $\gamma > 0$.
- On loss: Twice differentiable in ℓ_2-ball of radius r around $\boldsymbol{\beta}_0$ and Local RSC with $3\mu < 4\alpha$.

Main Results (Loh, 2017)

Suppose that $||\boldsymbol{\beta}_0||_1 \leq R/2$, $\frac{160\lambda s}{2\alpha - \mu} < R$ and $\min_{j\in S}|\boldsymbol{\beta}_{0j}| \leq C\sqrt{\frac{\log s}{n}} + \gamma\lambda$.
Then, if $n \geq c_0 \max\{s^2, s\log p\}$, with probability at least $1 - c\exp(-c'\min\{s, \log p\})$, any local M-estimator $\widehat{\boldsymbol{\beta}}$ satisfying $||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_2 \leq r$ also satisfies

$$\{j : \widehat{\beta}_j \neq 0\} \subseteq S, \quad \textbf{and} \quad \widehat{\boldsymbol{\beta}}_{S_0} = \widehat{\boldsymbol{\beta}}^{\mathcal{O}}_{S_0} := \arg\min_{\boldsymbol{\beta}_S : ||\boldsymbol{\beta}_S - \boldsymbol{\beta}_{0S}||_2 \leq r} L_n(\boldsymbol{\beta}_S, \mathbf{0}).$$

Asymptotic Normality Result: Linear Regression

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim f$, independently for each i, having mean zero. Define $S = \{j : \beta_{0j} \neq 0\}$.

Non-convex Penalized M-estimator: The Unweigheted version

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{\mathbf{L_n}(\boldsymbol{\beta}) + \mathbf{P}_\lambda(\boldsymbol{\beta})\}, \qquad \mathbf{L_n}(\boldsymbol{\beta}) = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \rho(\boldsymbol{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} - \mathbf{y_i}).$$

Assumptions

- On penalty: Assumption (P) with $L = 1$ and $\rho'_\lambda(t) = 0$ for all $t \geq \gamma\lambda$ with $\gamma > 0$.
- On loss: Thrice continuously differentiable and Local RSC with $3\mu < 4\alpha$.

Main Results (Loh, 2017)

Suppose that $||\boldsymbol{\beta}_0||_1 \leq R/2$, $\frac{160\lambda s}{2\alpha - \mu} < R$ and $\min_{j \in S} |\beta_{0j}| \leq C\sqrt{\frac{\log s}{n}} + \gamma\lambda$.
Let $\widehat{\boldsymbol{\beta}}$ be any local M-estimator satisfying $||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_2 \leq r$.

- If $s(\log s)^3/n \to 0$, then $||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_2 = O_p(\sqrt{s/n})$.
- If $s^2(\log s)/n \to 0$, then for any bounded sequence $\{\boldsymbol{v}_n\} \subseteq \mathbb{R}^p$, we have

$$\frac{\sqrt{n}}{\sigma(\boldsymbol{v}_n)} \boldsymbol{v}_n^T \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} N(0,1), \qquad \sigma(\boldsymbol{v}_n) = \frac{1}{n} \frac{[\boldsymbol{v}_n^T(\mathbf{X}^T\mathbf{X})\boldsymbol{v}_n]}{\{E[\rho''(\epsilon_i)]E[\rho'(\epsilon_i)^2]\}}.$$

Consistency Under Contamination

- All the results discussed so far are derived assuming that the data has come from a true model!
- Hence, they all provide nice (asymptotic) properties of penalized M-estimators under pure data only!

How do these estimators behave under data contamination?

Consistency under Grossly Error Contamination: Linear Regression

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim (1-\epsilon)f_0 + \epsilon g$, independently for each i, with symmetric f_0.

General M-estimators with Elastic-Net Penalty

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \rho(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) + \sum_{j=1}^{p} \left(\lambda \left|\beta_j\right| + \alpha \left|\beta_j\right|^2 \right) \right\}.$$

Main Result (Zhang et al., 2021)

For any $\pi > 0$, there exists constant C_is independent of (ϵ, n, p, s, M) such that, for $\lambda = C_0 M\sqrt{\frac{\log p}{n}} + C_1\frac{\epsilon}{\sqrt{n}}$, the following results hold with probability at least $1-\pi$:

- Any local M-estimator $\widehat{\boldsymbol{\beta}}$ satisfies $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \le \eta := \frac{\epsilon C_1 + \sqrt{s}\lambda C_2}{1-\epsilon}$, whenever $n \ge C_3 s \log p$.
- For $n \ge C_3 s(\log p)^2$ and $\eta \le C_2 - C_3\epsilon$, then a local M-estimator $\widehat{\boldsymbol{\beta}}$ is indeed a **unique** stationary point (and hence, is also the **global minimizer**).

Assumptions

- The true parameter $\boldsymbol{\beta}_0$ satisfies $\|\boldsymbol{\beta}_0\|_2 \le r/3$ and $|S_0| \le s$, where $r = O(1/\alpha)$.
- Covariate $\mathbf{x}_i$s are IID with zero mean and τ^2-sub-Gaussian, i.e., $E[\exp(\mathbf{x}_i^T\boldsymbol{l})] \le \exp(\tau^2||\boldsymbol{l}||_2^2/2)$.
- Each $\mathbf{x}_i$ has a density and is bounded (in absolute) by $M\tau$ for some constant M (indep. of p).
- $\psi(z) = \rho'(z)$ is bounded, twice differentiable with bounded derivatives and odd in z with $\psi(z) \ge 0$ for all $z \ge 0$.
- Joint covariance matrix of all $\mathbf{x}_i$s is positive definite and $h(z) := \int f_0(u)\psi(z+u)du > 0$ for all $z > 0$, and $h'(0) > 0$.

Consistency under Grossly Error Contamination: Example 1

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim (1-\epsilon)f_0 + \epsilon g$, independently for each i, with symmetric f_0.

Penalized M-estimators with Welsch's exponential squared loss (nonconvex) and Elastic-Net Penalty

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \rho_k(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) + \sum_{j=1}^{p} \left(\lambda \left|\beta_j\right| + \alpha \left|\beta_j\right|^2 \right) \right\}, \qquad \rho_k(z) = \left[1 - e^{-kz^2/2}\right]/k.$$

Assumptions

Covariate $\mathbf{x}_i$s are IID multivariate uniform on $[-\tau, \tau]^p$, $f_0 \equiv N(0, \sigma^2)$, $\|\boldsymbol{\beta}_0\|_2 \leq r/3$ and $|S_0| \leq s$.

Main Result (Zhang et al., 2021)

For any $\pi > 0$, there exists constant C_π such that, for $\lambda = 2C_\pi\tau\sqrt{\frac{\log p}{n}} + \frac{k\tau\epsilon}{2\sqrt{n}}$ and $n \gg s\log p$, the following results hold with probability at least $1-\pi$:

- Any local M-estimator $\widehat{\boldsymbol{\beta}}$ satisfies $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \leq \eta := \frac{4\sqrt{e}\epsilon(1+2\tau)(1+k\sigma^2)^{3/2}}{(1-\epsilon)\tau\sqrt{k}} e^{\frac{32kr^2\tau^2}{3(1+k\sigma^2)}}$.
- Whenever $12\sqrt{3e}\epsilon(1+2\tau)(1+k\sigma^2)^3 e^{\frac{32kr^2\tau^2}{3(1+k\sigma^2)}} \leq (1-\epsilon)\left[\tau^2(1-\epsilon) - \epsilon(1+k\sigma^2)^{3/2}\right]$, then any M-estimator $\widehat{\boldsymbol{\beta}}$ is **unique** and hence the **global minimizer**.

Special Case: LASSO!! ($\eta = \infty$)

Consistency under Grossly Error Contamination: Example 2

Standard **linear regression model** (LRM): $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,
where random error $\epsilon_i \sim (1-\epsilon)f_0 + \epsilon g$, independently for each i, with symmetric f_0.

Penalized M-estimators with Tukey's Bisquare Loss loss (nonconvex) and Elastic-Net Penalty

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \rho_k(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) + \sum_{j=1}^{p}\left(\lambda\left|\beta_j\right| + \alpha\left|\beta_j\right|^2\right) \right\}, \rho_k(z) = \frac{k^2}{6}\left[1 - \left(1 - \frac{z^2}{k^2}\right)^3\right] I(|z| \le k).$$

Assumptions

Covariate $\mathbf{x}_i$s are IID multivariate uniform on $[-\tau, \tau]^p$, $f_0 \equiv N(0, \sigma^2)$, $\|\boldsymbol{\beta}_0\|_2 \le r/3$ and $|S_0| \le s$.

Main Result (Zhang et al., 2021)

For any $\pi > 0$, there exists constant C_π such that, for $\lambda = 2C_\pi\tau\sqrt{\frac{\log p}{n}} + 2k\tau\epsilon$ and $n \gg s\log p$, the following results hold with probability at least $1-\pi$:

- Any local M-estimator $\widehat{\boldsymbol{\beta}}$ satisfies $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \le \eta := \frac{28\sqrt{2\pi}\epsilon(1+2\tau)}{(1-\epsilon)\tau\sigma^3k^2} e^{\frac{k^2+64r^2\tau^2}{\sigma^2}}$.
- Whenever $56\sqrt{6\pi}\epsilon(1+2\tau)e^{\frac{k^2+64r^2\tau^2}{\sigma^2}} \le (1-\epsilon)\sigma^3k^2\left[\tau^2(1-\epsilon)M(k,\sigma) - 4\epsilon\right]$ with $M(k,\sigma) = 2k\int_0^1(1-t)(1+t)(1-5t^2)f_0(kt)dt$, then any M-estimator $\widehat{\boldsymbol{\beta}}$ is the **unique global minimizer**.

Special Case: LASSO!! ($\eta = \infty$)

Outline

LAD-LASSO for Linear Regression

Standard **linear regression model** (LRM):

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

LAD-LASSO Estimator (The first attempt!!)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right| + \lambda ||\boldsymbol{\beta}||_1 \right\}$$

Theoretical Properties: Estimation and Model Selection Consistency

- Low-dimensional Context (Wang et al., 2007)
- High-dimensional Context (Gao and Huang, 2010) - Requires irrepresentable Condition!

Better results may be possible to obtain using general penalty functions! **(Open problem)**

Sparse Least Trimmed Square Estimator: Linear Regression

Standard **linear regression model** (LRM):

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where random error $\epsilon_i \sim f$, independently for each i, having mean zero.

sLTS Estimator (Alfons et al., 2013)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{h} [\mathbf{r}^2(\boldsymbol{\beta})]_{i:n} + h\lambda||\boldsymbol{\beta}||_1 \right\}, \qquad \mathbf{r}_i^2(\boldsymbol{\beta}) = (\mathbf{y}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2.$$

Robustness: Breakdown Point

- **Finite-Sample breakdown point** $\frac{n-h+1}{n}$.
- This result also hold if we define $r_i = \rho(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})$ for any convex, symmetric and positive ρ with $\rho(0) = 0$.

Theoretical Properties, e.g., Consistency, are required to fully justify the method! **(Open problem)**

Sparse LTS for Logistic Regression

Given $\mathbf{x}_i$, $y_i \sim Ber(\pi(\mathbf{x}_i^T\beta))$, independently for $i = 1, \ldots, n$, where $\pi(s) = \frac{e^s}{1+e^s}$.

The enetLTS (Kurnaz et al., 2018)

$$\widehat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{h} [\mathbf{D}(\beta)]_{i:n} + h\lambda \mathbf{P}_\alpha(\beta) \right\},$$

Deviance $D_i(\beta) = -y_i \mathbf{x}_i^T \beta + \log(1 + e^{\mathbf{x}_i^T \beta})$.

Penalty $P_\alpha(\beta) = \sum_{j=1}^{p} \left[\alpha|\beta_j| + (1-\alpha)\frac{\beta_j^2}{2} \right]$.

- Kurnaz et al. (2018) examined its usefulness empirically for logistic regression.
- They also empirically examined the elastic-net penalty for sLTS loss under linear regression.

Theoretical Properties, e.g., Consistency, are required to fully justify the method! **(Open problem)**

Penalized Quasilikelihood for GLM: Definition

Generalized linear model (GLM): Given a covariate value $\mathbf{X} = \boldsymbol{x}$, the response variable Y has density

$$f(y; \boldsymbol{x}^T\beta) = \exp\{y\theta - b(\theta) + c(y)\}, \quad \text{with} \quad E[Y|\boldsymbol{x}] = b'(\theta) = g^{-1}(\boldsymbol{x}^T\beta), \tag{8}$$

Put $\mu_i = E[Y|\boldsymbol{x}_i]$, $\nu(\mu_i) = Var[Y|\boldsymbol{x}_i]$

The Quasilikelihood (Cantoni and Ronchetti, 2001)

$$Q_M(y_i, \boldsymbol{x}_i^T\beta) = \int_{\widetilde{\mu}}^{\mu_i} \psi\left(\frac{y_i - t}{\sqrt{\nu(t)}}\right) w(x_i)dt - \frac{1}{n}\sum_{j=1}^{n}\int_{\widetilde{\mu}_e}^{\mu_j} E\left[\psi\left(\frac{y_i - t}{\sqrt{\nu(t)}}\right)\right] w(x_j)dt,$$

where $\widetilde{\mu}$ and $\widetilde{\mu}_e$ are chosen so as to make the respective integrand zero.
Here, ψ downweights the outliers and w downweights the leverage points!

Penalized Quasilikelihood-based M-estimator (Avella-Medina and Ronchetti, 2018)

$$\widehat{\beta} = \arg\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n} Q_M(y_i, \boldsymbol{x}_i^T\beta) + \sum_{j=1}^{p} p_\lambda(|\beta_j|)\right\}.$$

Penalized Quasilikelihood for GLM: Properties under Pure data

Penalized Quasilikelihood-based M-estimator (Avella-Medina and Ronchetti, 2018)

$$\widehat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} Q_M(y_i, x_i^T \beta) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \right\}, \qquad \Psi_i(\beta) = \frac{\partial}{\partial \beta} Q_M(y_i, x_i^T \beta)$$

Assumptions (in a neighbourbood of β_0)

- $|\psi'(r)|$, $|\psi'(r)r|$ are bounded. Also, first & second order derivatives of Q_M with respect to η are bounded.
- All eigenvalues of $\mathbb{M}(\beta) = E\left[\frac{\partial}{\partial \beta} \Psi_i(\beta)\right]$ are bounded away from 0, ∞; $\mathbb{Q}(\beta) = E\left[\Psi_i(\beta)\Psi_i(\beta)^T\right] < \infty$.
- Penalty satisfies Assumption (P) and $||\mathbf{X}_{S^c}^T \Gamma'(\mathbf{X}\beta)\mathbf{X}_S(\mathbf{X}_S^T \Gamma'(\mathbf{X}\beta)\mathbf{X}_S)^{-1}||_\infty \leq \min\left\{c_3 \frac{p'_\lambda(0+)}{p'_\lambda(d_n)},\ O(n^{\alpha_1})\right\}$, where $\Gamma(\mathbf{X}\beta) = \frac{\partial}{\partial \eta} \sum_{i=1}^{n} Q_M(y_i, \mu_i)$, $0 < c_3 < 1$ and $0 < \alpha_1 < 0.5$.

Main Result under $\log p = O(n^\alpha)$ (Avella-Medina and Ronchetti, 2018)

Suppose that $0 < \alpha < 0.5$, $s^3 = o(n)$, $\lambda\sqrt{ns} \to 0$, $\lambda\kappa(p) = o(1)$ and $d_n \gg \lambda \gg \max\{\sqrt{s/n}, n^{-\alpha}\sqrt{\log n}\}$. Then, there exists a local M-estimator $\widehat{\beta}$ such that

$$\mathbf{P}(\widehat{\beta}_{S^c} = \mathbf{0}) \to \mathbf{1}, \quad \text{and} \quad \sqrt{n}\boldsymbol{v}^T \mathbb{M}_S(\beta_0)\mathbb{Q}_S(\beta_0)^{-1/2}\left(\widehat{\beta}_S - \beta_{0S}\right) \xrightarrow{\mathcal{D}} \mathbf{N}(\mathbf{0}, \mathbf{1}),$$

for any s-dimensional unit vector $\boldsymbol{v}$.

Penalized Quasilikelihood for GLM: Properties under Contaminated data

Penalized Quasilikelihood-based M-estimator (Avella-Medina and Ronchetti, 2018)

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{Q_M}(\mathbf{y_i}, \boldsymbol{x_i^T}\boldsymbol{\beta}) + \sum_{\mathbf{j}=1}^{\mathbf{p}} \mathbf{p}_{\boldsymbol{\lambda}}(|\boldsymbol{\beta_j}|) \right\}, \qquad \boldsymbol{\Psi_i}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{Q_M}(\mathbf{y_i}, \boldsymbol{x_i^T}\boldsymbol{\beta})$$

Assumptions (in a neighbourbood of β_0)

- $|\psi'(r)|$, $|\psi'(r)r|$ are bounded. Also, first & second order derivatives of Q_M with respect to η are bounded.
- All eigenvalues of $\mathbb{M}(\boldsymbol{\beta}) = E\left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\Psi}_i(\boldsymbol{\beta})\right]$ are bounded away from 0, ∞; $\mathbb{Q}(\boldsymbol{\beta}) = E\left[\boldsymbol{\Psi}_i(\boldsymbol{\beta})\boldsymbol{\Psi}_i(\boldsymbol{\beta})^T\right] < \infty$.
- Penalty satisfies Assumption (P) and $||\mathbf{X}_{S^c}^T \Gamma'(\mathbf{X}\boldsymbol{\beta})\mathbf{X}_S(\mathbf{X}_S^T \Gamma'(\mathbf{X}\boldsymbol{\beta})\mathbf{X}_S)^{-1}||_\infty \leq \min\left\{c_3 \frac{p'_\lambda(0+)}{p'_\lambda(d_n)},\ O(n^{\alpha_1})\right\}$, where $\Gamma(\mathbf{X}\boldsymbol{\beta}) = \frac{\partial}{\partial \eta}\sum_{i=1}^n Q_M(y_i, \mu_i)$, $0 < c_3 < 1$ and $0 < \alpha_1 < 0.5$.

Main Result under $\log p = O(n^\alpha)$ (Avella-Medina and Ronchetti, 2018)

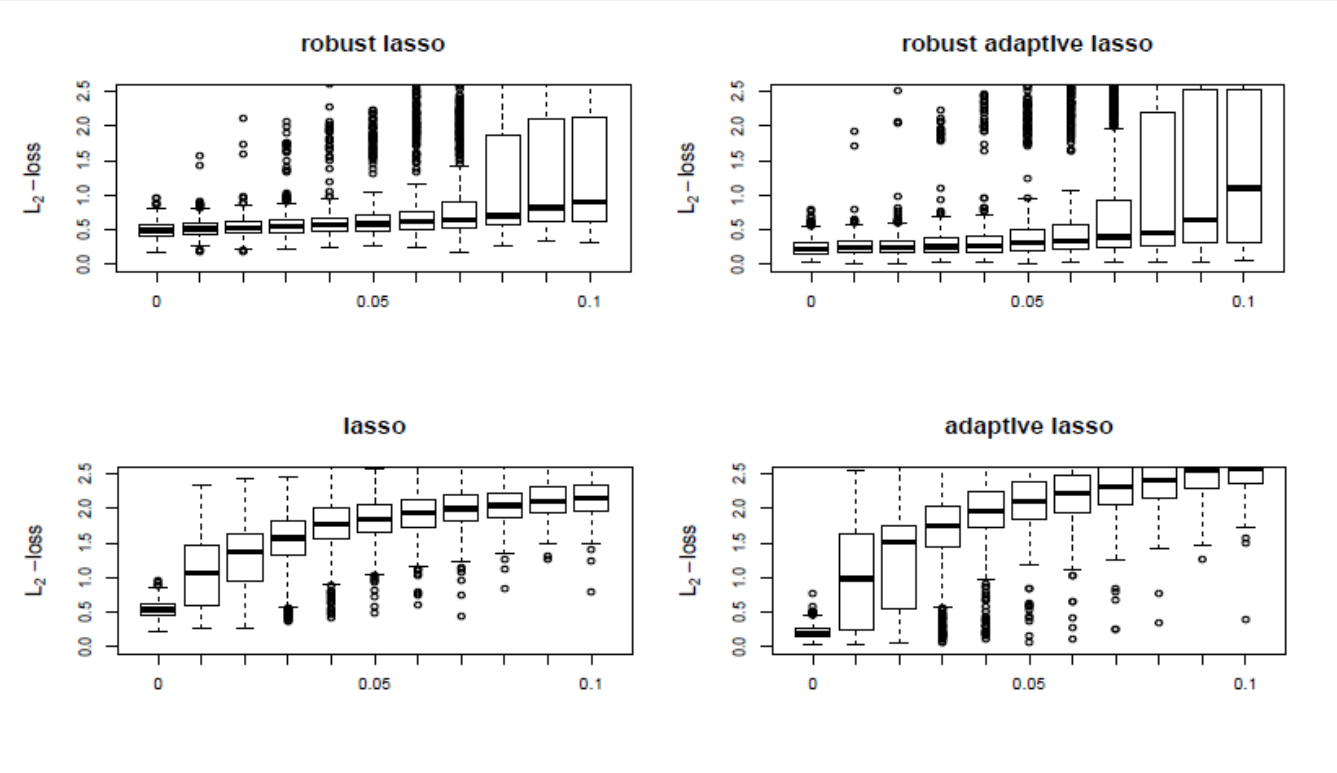
Suppose that $0 < \alpha < 1$, $s = o(n)$, $d_n \gg n^{-\zeta}\log n$, $p'_\lambda(d_n) = o(n^{-\zeta}\log n)$, $\lambda\kappa(p) = o(1)$ and $\lambda \gg n^{-(1-\alpha)/2}\log n$.
Then, for large enough n, there exists $\epsilon > 0$ such that any local M-estimator $\widehat{\boldsymbol{\beta}}$ in the ϵ-gross error neighborhood satisfies the following with probability at least $1 - 2[s/n + (p-s)\exp(-n^\alpha \log n)]$:

$$\widehat{\boldsymbol{\beta}}_{\mathbf{Sc}} = \mathbf{0}, \quad \text{and} \quad ||\widehat{\boldsymbol{\beta}}_{\mathbf{S}} - \boldsymbol{\beta}_{\mathbf{0S}}||_\infty = \mathbf{O}(\mathbf{n}^{-\zeta}\log \mathbf{n} + \epsilon).$$

Illustration: High-dimensional Poisson regression model!

Simulation results: Prediction Accuracy in presence of distant outlying points

(Avella-Medina and Ronchetti, 2018)



Quasi-likelihood based M-estimator can successfully control the outliers, leading to much stable results in most cases, even when the contamination proportion is high!!

Outline

Summary

- We discussed the need for robust methods in high-dimensional data analyses
- We discussed the penalized M-estimators and their robustness via influence function analysis
- Several existing theoretical results for general penalized M-estimators are discussed along with examples.
- Several open problems in the context of different penalized M-estimators are mentioned.

References

- Alfons, A., Croux, C., and Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Ann. Appl. Stat.*, **7**, 226–248.
- Avella-Medina, M. (2017). Influence functions for penalized M-estimators. *Bernoulli*, **23**, 3178–3196.
- Avella-Medina, M., and Ronchetti, E. (2018). Robust and consistent variable selection in high-dimensional generalized linear models. *Biometrika,* **105,** 1, 31–44.
- Cantoni, E., and Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.*, **96(455)**, 1022–1030.
- Gao, X., and Huang, J. (2010). Asymptotic analysis of high-dimensional LAD regression with LASSO. *Stat. Sinica*, 1485–1506.
- Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., and Stahel W.(1986). *Robust Statistics: The Approach Based on Influence Functions*.

  New York, USA: John Wiley & Sons.
- Kurnaz, F. S., Hoffmann, I., and Filzmoser, P. (2018). Robust and sparse estimation methods for high-dimensional linear and logistic regression. *Chemo. Int. Lab. Sys.*, 172, 211–222.
- Loh, P. L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust *M*-estimators. *Ann. of Stat.*, **45(2)**, 866–896.
- Loh, P. L., and Wainwright, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Machine Learn. Res.*, **16(1)**, 559–616.
- Negahban, S. N., Ravikumar, P., Wainwright, M. J., and Yu, B. (2012). A unified framework for high-dimensional analysis of *M*-estimators with decomposable regularizers. *Stat. Sci.*, **27(4)**, 538–557.
- Wang, H., Li, G., and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-lasso. *J. Bus. Econ. Stat.*, **25**, 347–355.
- Zhang, R., Mei, Y., Shi, J., and Xu, H. (2021). Robustness and Tractability for Non-convex M-estimators. *Stat. Sinica*, doi:10.5705/ss.202019.0324.

THANK YOU!

Contact me at:
abhik.ghosh@isical.ac.in