Screening Rules for Lasso with Non-Convex Sparse Regularizers

A. Rakotomamonjy

Joint work with G. Gasso and J. Salmon

ICML 2019

This work benefited from the support of the project OATMIL ANR-17-CE23-0012 of the French National Research Agency (ANR), the Normandie Projet GRR-DAISI, European funding FEDER DAISI

Objective of the paper

Lasso and screening

- learning sparsity-induced linear models from high-dimensional data $\mathbf{X} \in \mathbb{R}^{n \times d}$, $\mathbf{y} \in \mathbb{R}^n$

$$\min_{\mathbf{w} \in \mathbb{R}^d} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \sum_{j=1}^{d} \lambda |w_j|$$

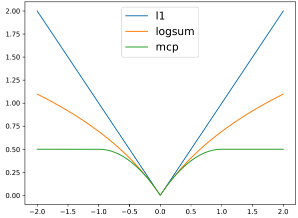
- Screening rule : identify vanishing variables in $\mathbf{w}^\star$. Example with $\hat{\mathbf{w}}, \hat{\mathbf{s}}$ intermediate primal-dual solutions :

$$|\mathbf{x}_j^\top \hat{\mathbf{s}}| + r(\hat{\mathbf{w}}, \hat{\mathbf{s}}) \|\mathbf{x}_j\| < 1 \implies w_j^\star = 0$$

by exploiting sparsity, convexity and duality.

Extension to non-convex regularizers

- non-convex regularizers lead to statistically better models but
- how to do screening when the regularizer is non-convex?



Non-convex Lasso

The problem

$$\min_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2 + \sum_{j=1}^{d} r_\lambda(|w_j|)$$

with the regularizer $r_\lambda(\cdot)$ being smooth and concave on $[0,\infty[$.

The proposed screening strategy

- Solve by majorization-minimization

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}\in\mathbb{R}^d} \quad \tfrac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2 + \tfrac{1}{2\alpha}\|\mathbf{w}-\mathbf{w}^k\|_2^2 \quad + \sum_{j=1}^{d} \lambda_j |w_j| \ ,$$

  with $\lambda_j = r'_\lambda(|w_j|)$
- Screen at two levels
  - within each weighted Lasso
  - propagate screened variables information between 2 successive Lasso.

Screening weighted Lasso

- Optimization problem and screening condition

$$\min_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{w}\|_2^2+\frac{1}{2\alpha}\|\mathbf{w}-\mathbf{w}^k\|_2^2+\sum_{j=1}^{d}\lambda_j|w_j| \qquad |\mathbf{x}_j^\top \mathbf{s}^\star - v_j^\star| - \lambda_j < 0 \implies w_j^\star = 0$$

  with $\mathbf{s}$ and $\mathbf{v}$ being dual variables and $\mathbf{s}^\star = \mathbf{y} - \mathbf{X}\mathbf{w}^\star$ and $\mathbf{w}^\star - \mathbf{w}'^\star = \alpha \mathbf{v}^\star$.
- Our screening test

$$\underbrace{|\mathbf{x}_j^\top \hat{\mathbf{s}} - \hat{v}_j| + \sqrt{2G_\Lambda}\left(\|\mathbf{x}_j\| + \frac{1}{\alpha}\right)}_{T_j^{(\lambda_j)}(\hat{\mathbf{w}},\hat{\mathbf{s}},\hat{\mathbf{v}})} < \lambda_j$$

  given a primal-dual intermediate solution $(\hat{\mathbf{w}}, \hat{\mathbf{s}}, \hat{\mathbf{v}})$, with duality gap G_Λ.

Screened variables propagation

Setting

- After iteration k, we have a weigthed Lasso with weights $\{\lambda_j\}$ and approximate solutions $\hat{\mathbf{w}}$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{v}}$. Screened variables are those

$$T_j^{(\lambda_j)}(\hat{\mathbf{w}}, \hat{\mathbf{s}}, \hat{\mathbf{v}}) < \lambda_j$$

- Before iteration $k+1$
  - change of weights $\{\lambda_j^\nu\}_{j=1,\dots,d}$
  - new primal-dual triplet $(\hat{\mathbf{w}}^\nu$,$\hat{\mathbf{s}}^\nu, \hat{\mathbf{v}}^\nu)$,

Screening propagation test

$$T_j^{(\lambda_j)}(\hat{\mathbf{w}}, \hat{\mathbf{s}}, \hat{\mathbf{v}}) + \|\mathbf{x}_j\|(a + \sqrt{2b}) + c + \frac{1}{\alpha}\sqrt{2b} < \lambda_j^\nu$$

with that $\|\hat{\mathbf{s}}^\nu - \hat{\mathbf{s}}\|_2 \leq a$, $|G_\Lambda(\hat{\mathbf{w}}, \hat{\mathbf{s}}, \hat{\mathbf{v}}) - G_{\Lambda^\nu}(\hat{\mathbf{w}}^\nu, \hat{\mathbf{s}}^\nu, \hat{\mathbf{v}}^\nu)| \leq b$ and $|\hat{v}_j^\nu - \hat{v}_j| \leq c$.

Summary

- First approach for screening with non-convex regularizers
- Convexification and propagation

At poster #190 Pacific Ballroom

- More technical details
- Experimental results on computational gain and on propagation strategy