Introduction à l'optimisation non convexe
MCP

S.Canu
ITII /INSA de Rouen Normandie

25 novembre 2024

les moindres carrés pénalisés

- Données
  - $X \in \mathbb{R}^{n \times p}$ les p variables observées n fois
  - $y \in \mathbb{R}^n$ la variable réponse
- Inconnues
  - $\beta \in \mathbb{R}^p$

le cout pénalisé

$$J_{\text{pen}}(\beta) \;=\; \frac{1}{2}\|y - X\beta\|^2 + \textbf{pen}(\beta) \;,$$

Qu'est-ce qu'une bonne pénalité

- algo efficace
- parcimonie
- non biaisée
- stable

les pénalités déjà étudiées

le cout pénalisé

$$J_{\text{pen}}(\beta) \;=\; \tfrac{1}{2}\|y - X\beta\|^2 + \mathbf{pen}(\beta)\ ,$$

$$\mathbf{pen}(\beta) = \sum_{j=1}^{p} \mathbf{pen}_\lambda(|\beta_j|)$$

Moindrs carrés

$$\mathbf{pen}_\lambda(|\beta_j|) = 0$$

Régression ridge

$$\mathbf{pen}_\lambda(|\beta_j|) = \lambda\beta_j^2$$

Le lasso

$$\mathbf{pen}_\lambda(|\beta_j|) = \lambda|\beta_j|$$

le cas 1d

- Données
  - $X \in \mathbb{R}^{n \times p}$ l'unique $p = 1$ variable observée n fois
  - $y \in \mathbb{R}^n$ la variable réponse
- une seule inconnue
  - $\beta \in \mathbb{R}$

le cout pénalisé

$$J_{\text{pen}}(\beta) \;=\; \frac{1}{2}\|y - X\beta\|^2 + \textbf{pen}(\beta)\ ,$$

la solution

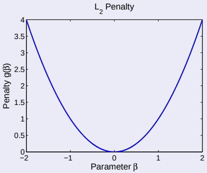
$$\widehat{\beta} = \textbf{shr}(\beta)\widehat{\beta}_{MC}$$

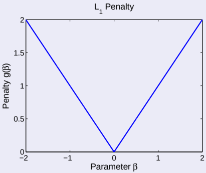
$$J'_{\text{pen}}(\beta) \;=\; X^\top(y - X\beta) + \textbf{pen}'(\beta)\ ,$$

$$J'_{\text{pen}}(\widehat{\beta}) \;= 0 \qquad \Leftrightarrow \qquad \widehat{\beta} = (1 - \frac{\textbf{pen}'(\widehat{\beta})}{X^\top y})\frac{X^\top y}{X^\top X}$$

Proposition d'étude des pénalités en 1d

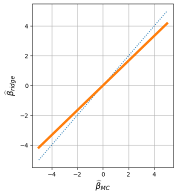
la pénalité comme une fonction : **pen**(β)

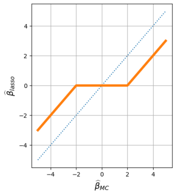




la solution comme une fonction de la solution des moindres carrés

$$\widehat{\beta} = \textbf{shr}(\widehat{\beta}_{MC})$$





le MCP (Zhang, 2010)

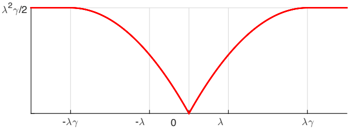
- Données
  - $X \in \mathbb{R}^{n \times p}$ les p variables observées n fois
  - $y \in \mathbb{R}^n$ la variable réponse
  - $\lambda, \gamma \geq 0$ des (hyper) paramètres
- Inconnue
  - $\beta \in \mathbb{R}^p$

le cout de MCP

$$J_{\text{MCP}}(\beta) \;=\; \tfrac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \mathbf{pen}_{\lambda,\gamma}(|\beta_j|) \;,$$

où $\mathbf{pen}_{\lambda,\gamma}$ est la pénalité du MCP

$$\mathbf{pen}_{\lambda,\gamma}(t) = \begin{cases} t - \dfrac{t^2}{2\lambda\gamma} & \text{if } t \leq \gamma\lambda \\ \dfrac{\gamma\lambda}{2} & \text{else.} \end{cases}$$



Algo pour le MCP

- CW
- DC
- Proximal

Algo CW pour le MCP

$$J_{\text{MCP}}(\beta) \;=\; \tfrac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{p} \mathbf{pen}_{\lambda,\gamma}(|\beta_j|) \;,$$

Algo CW pour le MCP

$$\begin{array}{l} \text{Pour } j = 1, p \\ \quad \text{fixer toutes les variables sauf } \beta_j \\ \quad \text{résoudre le MCP en une dimension :} \\ \qquad \hat{\beta}_j^{MCP} \;=\; \arg\min_{\beta\in\mathbb{R}} J^{(j)}_{\text{MCP}}(\beta) \end{array} \tag{1}$$

Le cout MCP en une dimension

$$J^{(j)}_{\text{MCP}}(\beta) = \tfrac{1}{2}\|y - X\beta^{(-j)} - X_j\beta\|^2 + \mathbf{pen}_{\lambda,\gamma}(|\beta|)$$

R. Mazumder and J. Friedman and T. Hastie, "Sparse Net : Coordinage Descent with Non-Convex Penalties" 2009

Algo CW pour le MCP

$$\text{Pour } j = 1, p \;:\; \hat{\beta}_j^{MCP} = \arg\min_{\beta \in \mathbb{R}} J_{\text{MCP}}^{(j)}(\beta)$$

$$\begin{aligned} J_{\text{MCP}}^{(j)}(\beta) &= \tfrac{1}{2}\|y - X\beta^{(-j)} - X_j\beta\|^2 + \textbf{pen}_{\lambda,\gamma}(|\beta|) \\ &= \tfrac{1}{2}\|z - X_j\beta\|^2 + \textbf{pen}_{\lambda,\gamma}(|\beta|) \end{aligned} \qquad (2)$$

avec $z = y - X\beta^{(-j)}$ et $\beta^{(-j)} = (\beta_1, \dots, \beta_{j-1}, 0, \beta_{j+1}, \dots, \beta_p)$

Maths	Info
Pour $j = 1, p$	for $j = 1 : p$
$\beta^{(-j)} = (\beta_1, \dots, \beta_{j-1}, 0, \beta_{j+1}, \dots, \beta_p)$	$bj = beta; bj(j) = 0;$
$z = y - X\beta^{(-j)}$	$z = y - X * bj$
$\hat{\beta}_j^{MCP} = \arg\min_{\beta \in \mathbb{R}} J_{\text{MCP}}^{(j)}(\beta, X_j, z)$	$beta(j) =$
fin de pour	end

Chaque étape de l'algo CW pour le MCP revient à résoudre :

$$\min_{\beta \in \mathbb{R}} J_{\mathrm{MCP}}^{(j)}(\beta) \;=\; \tfrac{1}{2}\|z - X_j\beta\|^2 + \begin{cases} \lambda|\beta| - \dfrac{|\beta|^2}{2\gamma} & \text{if } |\beta| \leq \gamma\lambda \\ \dfrac{\gamma\lambda^2}{2} & \text{else.} \end{cases}$$

$$\partial_\beta J_{\mathrm{MCP}}^{(j)}(\beta) = X_j^\top (X_j\beta - z) \; + \; \begin{cases} \lambda\alpha & \text{if } \beta = 0 \\ \lambda \mathrm{sign}(\beta) - \dfrac{\beta}{\gamma} & \text{if } |\beta| \leq \gamma\lambda \\ 0 & \text{else.} \end{cases}$$

pour $\alpha \in [-1, 1]$

$$\partial_\beta J_{\mathrm{MCP}}^{(j)}(\beta) = 0 \Leftrightarrow \begin{cases} X_j^\top (X_j\beta - z) \; + \; \lambda\alpha = 0 & \text{if } \beta = 0 \\ X_j^\top (X_j\beta - z) \; + \; \lambda \mathrm{sign}(\beta) - \dfrac{\beta}{\gamma} = 0 & \text{if } |\beta| \leq \gamma\lambda \\ X_j^\top (X_j\beta - z) = 0 & \text{else.} \end{cases}$$

$$\partial_\beta J^{(j)}_{\mathrm{MCP}}(\beta) = 0 \Leftrightarrow \begin{cases} X_j^\top (X_j\beta - z) \ + \ \lambda\alpha = 0 & \text{if } \beta = 0 \\ X_j^\top (X_j\beta - z) \ + \ \lambda \mathrm{sign}(\beta) - \dfrac{\beta}{\gamma} = 0 & \text{if } |\beta| \leq \gamma\lambda \\ X_j^\top (X_j\beta - z) = 0 & \text{else.} \end{cases}$$

puisque $X_j^\top X_j = 1$

$$\begin{cases} X_j^\top z \ + \ \lambda\alpha = 0 & \text{if } \beta = 0 \\ \beta - \dfrac{\beta}{\gamma} = X_j^\top z \ - \ \lambda \mathrm{sign}(\beta) & \text{if } |\beta| \leq \gamma\lambda \\ \beta = X_j^\top z & \text{else.} \end{cases}$$

soit

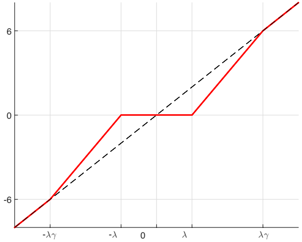
$$\begin{cases} \text{if } \dfrac{X_j^\top z}{\lambda} \in [-1, 1] & \beta = 0 \\ \beta = (X_j^\top z \ - \ \lambda \mathrm{sign}(X_j^\top z)) \dfrac{\gamma}{\gamma - 1} & \text{if } |X_j^\top z| \leq \gamma\lambda \\ \beta = X_j^\top z & \text{else.} \end{cases}$$

Codage

$$\begin{cases} \text{if } \frac{X_j^\top z}{\lambda} \in [-1, 1] & \beta = 0 \\ \beta = \left(X_j^\top z \ - \ \lambda \text{sign}(X_j^\top z)\right) \dfrac{\gamma}{\gamma - 1} & \text{if } |X_j^\top z| \leq \gamma\lambda \\ \beta = X_j^\top z & \text{else.} \end{cases}$$

$g = X_j^\top z$

- si $|g| < \lambda$ then $\beta = 0$
- elseif $|g| > \lambda\gamma$ then $\beta = g$
- elseif $\beta = (g - \text{sign}(g))\frac{\gamma}{\gamma-1}$



$$\beta = \text{sign}(g) \max\left(0, \min(abs(g), (abs(g) - \lambda)\frac{\gamma}{\gamma - 1})\right)$$

MCP comme une DC

$$
\begin{aligned}
J_{\text{MCP}}(\beta) &= \tfrac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{p} \begin{cases} \lambda|\beta_j| - \dfrac{\beta_j^2}{2\gamma} & \text{if } |\beta_j| \le \gamma\lambda \\ \dfrac{\gamma\lambda^2}{2} & \text{else.} \end{cases} \\
&= \tfrac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} |\beta_j| - \sum_{j=1}^{p} \begin{cases} \dfrac{\beta_j^2}{2\gamma} & \text{if } |\beta_j| \le \gamma\lambda \\ \lambda|\beta_j| - \dfrac{\gamma\lambda^2}{2} & \text{else.} \end{cases} \\
&= \qquad f(\beta) \qquad - \qquad g(\beta)
\end{aligned}
$$

$f(\beta)$ c'est le cout du lasso et

$$
g(\beta) = \sum_{j=1}^{p} \begin{cases} \dfrac{\beta_j^2}{2\gamma} & \text{if } |\beta_j| \le \gamma\lambda \\ \lambda|\beta_j| - \dfrac{\gamma\lambda^2}{2} & \text{else.} \end{cases}
$$

DCA pour le MCP

$$\begin{aligned}
\beta^{(k+1)} &= \arg\min_\beta f(\beta) - \beta^\top \nabla_\beta g(\beta^{(k)}) \\
&= \tfrac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} |\beta_j| - \sum_{j=1}^{p} |\beta_j| sign(\beta_j) \nabla_\beta g(\beta_j^{(k)}) \\
&= \tfrac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \quad \text{avec} \quad w_j = 1 - s\nabla_\beta g(\beta_j^{(k)})/\lambda
\end{aligned}$$

$$g(\beta) = \sum_{j=1}^{p} \begin{cases} \dfrac{\beta_j^2}{2\gamma} & \text{if } |\beta_j| \leq \gamma\lambda \\ \lambda|\beta_j| - \dfrac{\gamma\lambda^2}{2} & \text{else.} \end{cases} \Rightarrow s\nabla g(\beta) = \sum_{j=1}^{p} \begin{cases} \dfrac{|\beta_j|}{\gamma} & \text{if } |\beta_j \\ \lambda & \text{else.} \end{cases}$$

$$w_j = 1 - \partial_\beta g(\beta_j^{(k)})\lambda = \begin{cases} 1 - \dfrac{|\beta_j|}{\gamma\lambda} & \text{if } |\beta_j| \leq \gamma\lambda \\ 0 & \text{else.} \end{cases}$$