

Introduction à l'optimisation non convexe

MCP

S.Canu
ITII /INSA de Rouen Normandie

23 novembre 2025

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

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

Qu'est-ce qu'une bonne pénalité

- algo efficace
- parcimonie
- non biaisée
- stable

les pénalités déjà étudiées

le cout pénalisé

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

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

Moindrs carrés

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

Régression ridge

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

Le lasso

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

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

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

la solution

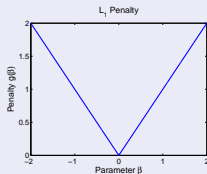
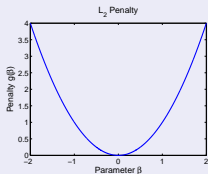
$$\hat{\beta} = \text{shr}(\beta) \hat{\beta}_{MC}$$

$$J'_{\text{pen}}(\beta) = X^T (y - X\beta) + \text{pen}'(\beta),$$

$$J'_{\text{pen}}(\hat{\beta}) = 0 \quad \Leftrightarrow \quad \hat{\beta} = \left(1 - \frac{\text{pen}'(\hat{\beta})}{X^T y}\right) \frac{X^T y}{X^T X}$$

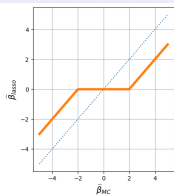
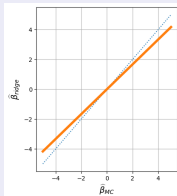
Proposition d'étude des pénalités en 1d

la pénalité comme une fonction : $\text{pen}(\beta)$



la solution comme une fonction de la solution des moindres carrés

$$\hat{\beta} = \text{shr}(\hat{\beta}_{MC})$$



les qualité d'une bonne pénalité

- Provoquer la parcimonie -> Lasso
- Non biaisée -> capped L1
- Continue -> MCP (*minimax concave penalty*)

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

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

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

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



Algo pour le MCP

- CW
- DC
- Proximal

Algo CW pour le MCP

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

Algo CW pour le MCP

Pour $j = 1, p$

fixer toutes les variables sauf β_j

résoudre le MCP en une dimension :

(1)

$$\hat{\beta}_j^{\text{MCP}} = \arg \min_{\beta \in \mathbb{R}} J_{\text{MCP}}^{(j)}(\beta)$$

Le cout MCP en une dimension

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

Algo CW pour le MCP

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

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

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

Maths

Pour $j = 1, p$

$$\beta^{(-j)} = (\beta_1, \dots, \beta_{j-1}, 0, \beta_{j+1}, \dots, \beta_p)$$

$$z = y - X\beta^{(-j)}$$

$$\hat{\beta}_j^{MCP} = \arg \min_{\beta \in \mathbb{R}} J_{MCP}^{(j)}(\beta, X_j, z)$$

fin de pour

Info

for $j = 1 : p$

$$bj = \text{beta}; \quad bj(j) = 0;$$

$$z = y - X * bj$$

$$\text{beta}(j) =$$

end

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

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

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

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

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

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

puisque $X_j^{\top} X_j = 1$

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

soit

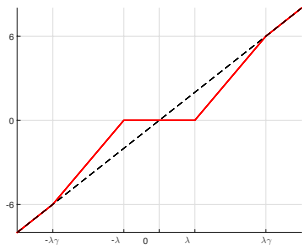
$$\begin{cases} \text{if } \frac{X_j^{\top} z}{\lambda} \in [-1, 1] & \beta = 0 \\ \beta = (X_j^{\top} z - \lambda\text{sign}(X_j^{\top} z)) \frac{\gamma}{\gamma - 1} & \text{if } |X_j^{\top} z| \leq \gamma\lambda \\ \beta = X_j^{\top} z & \text{else.} \end{cases}$$

Codage

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

$$g = X_j^\top z$$

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



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

MCP comme une DC

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

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

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

DCA pour le MCP

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

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

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