Breakpoint model

1 The problem

Consider a response variable Y and an explanatory variable X related by the following equation:

$$Y = \beta X + \gamma(X - \psi)_+ + \varepsilon$$

where $(\beta, \gamma, \psi) \in \mathbb{R}^3$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, and $(x)_+ = x$ if $x > 0$ and 0 otherwise.

Denote by $\Theta = (\beta, \gamma, \psi, \sigma^2)$, we can express the likelihood relative to n iid observations as:

$$\mathcal{L}(\Theta) = \prod_{i=1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(Y_i - \beta X_i - \gamma(X_i - \psi)_+)^2}{2\sigma^2} \right)$$

So maximizing the likelihood with respect to $\Theta_\mu = (\beta, \gamma, \psi)$ is equivalent to minimizing the mean squared error:

$$\ell(\Theta_\mu) = \sum_{i=1} (Y_i - \beta X_i - \gamma(X_i - \psi)_+)^2$$

One difficulty is that this objective function is not differientiable in ψ at $(X_i)_{i=1}^n$.

2 Proximal gradient method

$\ell(\Theta_\mu)$ might not be strictly convex but it is convex. So we can try applying a proximal gradient algorithm. This means updating the estimate by:

$$\begin{aligned}
\Theta_{\mu,k+1} &= \text{prox}_{\alpha_k \ell}(\Theta_{\mu,k}) = \underset{\Theta_\mu \in \mathbb{R}^2}{\arg\min} \left(\ell(\Theta_\mu) + \frac{1}{2\alpha_k} ||\Theta_\mu - \Theta_{\mu,k}||^2 \right) \\
&= \underset{\Theta_\mu \in \mathbb{R}^3}{\arg\min} \left(\sum_{i=1} (Y_i - \beta X_i - \gamma(X_i - \psi)_+)^2 + \frac{(\beta - \beta_k)^2 + (\gamma - \gamma_k)^2 + (\psi - \psi_k)^2}{2\alpha_k} \right)
\end{aligned}$$

where α_k is a pre-defined stricly positive real value.