~~-~~ Block coordinate Descent:

$$\min_{\theta_j \in \mathbb{R}^{p_j}} \frac{1}{2} \| y - \sum_{j=1}^{J} Z_j^T \theta_j \|_2^2 + \lambda \sum_{j=1}^{J} \|\theta_j\|_2$$

$$= \min_{\theta_j \in \mathbb{R}^{p_j}} \frac{1}{2} \| y - \sum_{k \neq j} Z_k^T \theta_k - Z_j^T \theta_j \| + \lambda \sum_{j=1}^{J} \|\theta_j\|_2$$

let $f(\overset{y}{\theta_1} \ldots \theta_J) = \frac{1}{2} \| y - \sum_{j=1}^{J} Z_j^T \theta_j \|_2^2 + \lambda \sum_{j=1}^{J} \|\theta$

~~let~~ $r_j = y - \sum_{k \neq j} Z_k^T \theta_k$

$$\min_{\theta_j \in \mathbb{R}^{p_j}} f(\theta_1 \ldots \theta_J) = \min_{\theta_j \in \mathbb{R}^{p_j}} \sum_{j=1}^{J} f(r_j, \theta_j). = \cancel{\sum_{j=1}}$$

$\Rightarrow$ <u>therefore</u> f is separatable

$\Rightarrow$ we can apply Block Coordinate Descent

Decomposable Function:

$$\min_{\theta_j \in R^{p_j}} \frac{1}{2}\|r_j - Z_j^T\theta_j\|_2^2 + \lambda\|\theta_j\|_2$$

$$r_j := y - \sum_{k \neq j} Z_k^T \theta_k$$

convex and differentiable function $g = \frac{1}{2}\|r_j - Z_j^T\theta_j\|_2^2$

non-differentiable : $h = \lambda\|\theta_j\|_2$

c. proximal Gradient Descent - Gradient step:

$$\nabla g(\theta_{j,k}) = -Z_j^T(r_j - Z_j^T\theta_j)$$

$$a_{j,k} = \theta_{j,k} - t_k \nabla g(\theta_{j,k}) = \theta_{j,k} - t_k(-Z_j^T(r_j - Z_j^T\theta_j$$

t_k : step size

d. proximal function we need to solve:

$$\text{prox}(\theta_{j,k}) = \arg\min_z \frac{1}{2t_k}\|\theta_{j,k} - z\|_2^2 + \lambda\|z$$

$$\text{prox}_h(\theta_{j,k}) = \arg\min \frac{1}{2}\|\theta_{j,k} - z\|_2^2 + \lambda t_k$$

problem 2a

① $\min_{\beta \in \mathbb{R}^p} \sum_{b=1}^{3} \frac{1}{2} \| y_b - X_b \beta_b \|_2^2 + \lambda \|\beta\|_1$

$\Leftrightarrow$ ② $\min_{\beta \in \mathbb{R}^p} \sum_{b=1}^{3} \left(\frac{1}{2} \beta_b^T X_b^T X_b \beta_b - y_b^T X_b \beta_b \right) + \lambda \|\beta\|_1 + K$

① $\Leftrightarrow$ ② $\Rightarrow \frac{1}{2} \beta_b^T X_b^T X_b \beta_b - y_b^T X_b \beta_b + K$

$= \frac{1}{2} \| y_b - X_b \beta_b \|_2^2$

$\Leftrightarrow \quad X_b \beta_b \left(\frac{1}{2} \beta_b^T X_b^T - y_b^T \right) + K$

$= \frac{1}{2} \| y_b - X_b \beta_b \|_2^2$

problem 2b:

$$\alpha_\rho (\beta_b, \theta, u_b) = \sum_{b=1}^{3} \Big(\frac{1}{2} \beta_b^T X_b^T X_b \beta_b - y_b^T X_b \beta_b$$

$$+ \lambda \|\theta\|_1 + u^T (\beta_b - \theta) + \frac{\rho}{2} \| \beta_b - \theta$$

$$\mathcal{L}_\rho(\beta_b, \theta, u_b) = \sum_{b=1}^{3} \left(\frac{1}{2} \beta_b^T X_b^T X_b \beta_b - y_b^T X_b \beta_b \right)$$

$$+ \lambda \|\theta\|_1 + \frac{\rho}{2} \| \beta_b - \theta + u_b \|_2^2$$

c) For β_b, $b = 1, 2, 3$. Since we have quadratic objective shows.

$$\beta_b^{t+1} = (X_b^T X_b + \rho I)^{-1} (X_b^T y_b + \rho(\theta^t - u_b^t))$$

$$\frac{\partial \mathcal{L}_\rho(\beta_b, \theta, u_b)}{\partial \beta_b} = \frac{1}{2}\left(X_b^T X_b + (X_b^T X_b)^T \right) \beta_b - y_b^T X_b + \rho(\beta_b - \theta + u_b)$$

$$\frac{\partial \mathcal{L}_\rho(\beta_b, \theta, u_b)}{\partial \beta_b} = 0 \Leftrightarrow$$

$$\frac{1}{2}(X_b^T X_b + (X_b^T X_b)^T) \beta_b - y_b^T X_b + \rho \beta_b - \rho(\theta - u_b) = 0$$

$$\Leftrightarrow \beta_b (X_b^T X_b + \rho I) = y_b^T X_b + \rho(\theta - u_b)$$

$$\Leftrightarrow \beta_b = (X_b^T X_b + \rho I)^{-1} (y_b^T X_b + \rho(\theta - u_b))$$