Coordinate Descent Algorithm

For general $\mathbf{X}$, Lasso can be solved via coordinate descent. At each iteration, repeatedly solve a one-dimensional Lasso problem for β_j while holding all other $(p-1)$ coefficients $\hat{\beta}_k$ $(k \neq j)$ at their current values:

$$\min_{\beta_j} \sum_{i=1}^{n} (y_i - \sum_{k \neq j} x_{ik} \hat{\beta}_k - x_{ij} \beta_j)^2 + \lambda \sum_{k \neq j} |\hat{\beta}_k| + \lambda |\beta_j|.$$

$$\Longrightarrow \min_{\beta_j} \sum_{i=1}^{n} (r_i - x_{ij} \beta_j)^2 + \lambda |\beta_j|.$$

Why does this algorithm work? – Tseng (2001)