$$\min_{\beta_i} \|\underbrace{y - \sum_{j \neq i} X_j \beta_j}_{y - X_{-i}\beta_{-i}} - X_i \beta_i\|_2^2$$

$$0 = X_i^T(y - X_{-i}\beta_{-i} + X_i\beta_i)$$

$$\beta_i^+ = \frac{X_i^T(y - X_{-i}\beta_{-i})}{X_i^T X_i}, \quad i = 1 \ldots p.$$

costs $O(np)$ flops (if done naively)

compare one full cycle (updates $i = 1 \ldots p$) to one iteration of gradient descent.

$$\beta^+ = \beta + t \cdot X^T(y - X\beta) \qquad O(np) \text{ flops.}$$

smarter updates

$$\beta_i^+ = \frac{X_i^T(y - X\beta + X_i\beta_i)}{X_i^T X_i}$$

$$= \frac{X_i^T r}{X_i^T X_i} + \beta_i$$

this requires $O(n)$ flops

- update residual r $\quad O(n)$
- multiply $X_i^T r$ $\quad O(n)$

$$\min_{\beta_i} \|y - \sum_{j\neq i} X_j\beta_j - X_i\beta_i\|_2^2 + \lambda|\beta_i|$$

quadratic in β_i (univariate) $+ \quad \lambda|\beta_i|$

$$\beta_i^{+} = S_{\frac{\lambda}{\|X_i\|_2^2}}\left(\frac{X_i^T(y - X_{-i}\beta_{-i})}{X_i^T X_i}\right)$$

cycle through $i = 1 \dots p$.

$$= S_{\frac{\lambda}{\|X_i\|_2^2}}\left(\frac{X_i^T r}{X_i^T X_i} + \beta_i\right)$$

$$\min_{\beta} \|y - X\beta\|_2^2 + \underbrace{\sum_{i=1}^{p} 1\{|\beta_i| \le s\}}$$

$\iff$ s.t. $|\beta_i| \le s$ all $i = 1..p$

$\iff$ s.t. $\|\beta\|_\infty \le s$