x is coord. wise minimizer

$$
\begin{aligned}
f(y) - f(x) &= g(y) + h(y) - (g(x) + h(x)) \\
&\geq \nabla g(x)^T (y-x) + \sum_{i=1}^{n} h_i(x_i) - h_i(y_i) \\
&= \sum_{i=1}^{n} \Big(\underbrace{\nabla_i g(x)(y_i - x_i) + h_i(y_i) - h_i(x_i)}_{\geq 0.} \Big) \\
&\geq 0.
\end{aligned}
$$

$$\min_{\beta} \quad \| y - \sum_{j=1}^{p} X_j \beta_j \|_2^2$$

min over β_i (all others fixed)

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

$$\Longleftrightarrow \quad \min_{\beta_i} \quad \|X_i\|_2^2 \beta_i^2 - 2\big(y - \sum_{j \neq i} X_j \beta_j\big)^T X_i \beta_i + \text{const}$$

$$\Longleftrightarrow \quad \beta_i = \frac{X_i^T \big(y - \sum_{j \neq i} X_j \beta_j\big)}{\|X_i\|_2^2} = \underbrace{\frac{X_i^T (y - X_{-i} \beta_{-i})}{\|X_i\|_2^2}}$$

cycle $i = 1, \ldots p$

scheme for full cycle costs $O(np)$ flops

Gradient update:

$$\beta^+ = \beta + t \, X^T (y - X\beta) \quad \Big\} \text{ costs } O(np)$$

incremental gradient

$$\min \sum_{i=1}^{n} f_i(\beta)$$ linear regression

$$\sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$

row of matrix X.

coordinate descent

$$\| y - \sum X_j \beta_j \|_2^2$$

col of matrix X

Rewrite i^{th} coord update as

$$\beta_i \leftarrow \frac{X_i^T r}{\|X_i\|_2^2}$$

Single update given r, takes $O(n)$ flops

$r = y - X_{-i}\beta_{-i}$ $\quad$ i^{th} step

$r \leftarrow r + X_i\beta_i - X_{i+1}\beta_{i+1}$ $\quad$ $(i+1)$st step

update to r also costs $O(n)$ flops

therefore done cleverly, full c.d cycle takes $O(np)$!

$$\min_{\beta} \overbrace{\|y - X\beta\|_2^2}^{g(\beta)} + \lambda \sum_{i=1}^{p} \overbrace{|\beta_i|}^{h_i(\beta_i)}$$

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

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

$$\min_{\beta} \; \|y - X\beta\|_2^2 \quad \text{st} \quad \|\beta\|_\infty \leq s$$

$$\iff \min_{\beta} \; \underbrace{\|y - X\beta\|_2^2}_{g(\beta)} + \sum_{i=1}^{p} \underbrace{1\{|\beta_i| \leq s\}}_{h_i(\beta_i)}$$