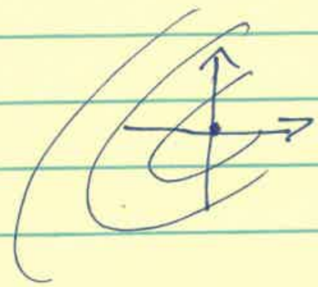

$$f = \underset{\substack{\text{convex}\\ \text{diff.}}}{g} + \underset{\text{convex}}{h}$$

separable: $h(x) = \sum_i h_i(x_i)$

if x is cw-min

then $f(y) - f(x) = g(y) - g(x) + h(y) - h(x)$

$$\geq \nabla g(x)^T (y - x) + \sum \left(h_i(y_i) - h_i(x_i) \right)$$

$$= \sum \Big(\underbrace{\nabla_i g(x) \cdot (y_i - x_i) + h_i(y_i) - h_i(x_i)}_{\geq 0} \Big)$$

why? Fix an i. $f(x + \delta e_i) = g(x + \delta e_i) + \underbrace{h(x + \delta e_i)}_{= \sum_{j \neq i} h_j(x_j) + h_i(x_i + \delta)}$

$0 \in \nabla_i g(x) + \partial h_i(x_i)$

$\iff -\nabla_i g(x) \in \partial h_i(x_i)$

$\iff h_i(y_i) \geq h_i(x_i) - \nabla_i g(x)(y_i - x_i)$

$\underline{h_i(y_i) - h_i(x_i) + \nabla_i g(x)(y_i - x_i) \geq 0}$ ✓

$$X^T X \beta = X^T y. \qquad f(\beta) = \frac{1}{2}\|y - X\beta\|_2^2 .$$

$$0 = \nabla_i f(\beta)$$

$$0 = X_i^T(y - X\beta)$$

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

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

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

$\iff$ Gauss-Seidel for $X^T X \beta = X^T y$.

$$\beta^+ = \beta + t \cdot X^T(y - X\beta)$$

$$t = \underset{s \geq 0.}{\operatorname{argmin}} \; f(\beta + s \cdot X^T(y - X\beta))$$

$$= \frac{r^T r}{r^T A r} \qquad r = y - X\beta \qquad A = X^T X.$$

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

$$= \frac{X_i^T(y - X_{-i}\beta_{-i})}{X_i^T X_i} \checkmark$$

$$
\text{KKT} \left\{ \begin{array}{l} X_1^T (y - X\beta) = \lambda s_1 \\ X_2^T (y - X\beta) = \lambda s_2 \quad \leftarrow \\ \quad \vdots \\ X_p^T (y - X\beta) = \lambda s_p \end{array} \right.
$$

active set optimization.