$t_k = 1/k$

$$t_k = \frac{f(x^{(k-1)}) - f(x^*)}{\|g^{(k-1)}\|_2}$$

$O(1/\sqrt{k})$

if want $f(x^{(k)}_{best}) - f(x^*) \le \varepsilon$
then need $O(1/\varepsilon^2)$ iterations

vs $O(1/\varepsilon)$ iterations

$$f_i(x) = \mathrm{dist}(x, C_i) = \min_{y \in C_i} \|y - x\|_2$$

$$\min_{y \in C} g(x,y) = h(x)$$

$$\max_{i=1,\dots m} f_i(x) = f(x)$$

$$\min_{x \in C} f(x) = \min_{x \in \mathbb{R}^n} f(x) + I_C(x)$$

$$[P_c(x)]_i = \max\{x_i, 0\} \qquad c = \{y : y_i \ge 0 \text{ all } i\}$$

$$C = \{x : Ax = b\} = x_0 + \mathrm{null}(A)$$

$$x_0 +$$

$$P_c(x) = \arg\min_v \| x_0 + v - x \|_2^2 \quad \text{s.t. } v \in \text{null}(A)$$

$$P_c(x) = \underbrace{B(B^T B)^{-1} B^T}_{P_{\text{null}(A)}}(x - x_0) + x_0$$

$$= P_{\text{null}(A)}(x - x_0) + x_0$$

$$= (I - P_{\text{row}(A)})(x - x_0) + x_0$$

$$= (I - A^T(AA^T)^{-1}A)(x - x_0) + x_0 \qquad A^T(AA^T)^{-1}b$$

$$= (I - A^T(AA^T)^{-1}A)(x - A^T(AA^T)^{-1}b) + \cancel{x_0}$$

$$= \boxed{x + A^T(AA^T)^{-1}(b - Ax)} \qquad A \text{ has full row rank}$$

x_0 solved $Ax = b$, $x_0 = A^T(AA^T)^{-1}b$

$$\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{s.t. } y = X\beta$$

$$\beta^{(k)} = P_C\left(\beta^{(k-1)} - t_k g^{(k-1)}\right)$$

$$g^{(k-1)} \in \partial \|\beta^{(k-1)}\|_1$$

$$g_i^{(k-1)} = \text{sign}(\beta_i^{(k-1)}) \quad \text{if } \beta_i^{(k-1)} \neq 0$$

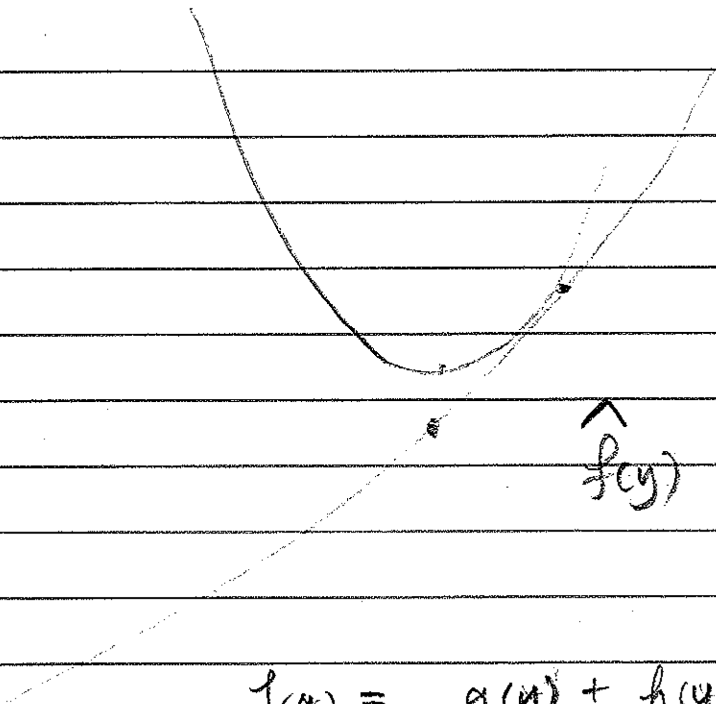
$$\in [-1, 1] \quad \text{else}$$

$$P_c(\beta) = \beta + X^T(XX^T)^{-1}(y - X\beta)$$

$$P_c(\beta^{(k-1)} - t_k g^{(k-1)}) = \beta^{(k-1)} - t_k g^{(k-1)} + X^T(XX^T)^{-1}(y - \cancel{X\beta^{(k-1)}})$$

y ''

$$= \boxed{\beta^{(k-1)} + (I - X^T(XX^T)^{-1}X)t_k g^{(k-1)}} + X t_k g^{(k-1)}$$



$$\hat{f}(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2t}\|y-x\|_2^2$$

$$\begin{aligned} f(y) &= g(y) + h(y) \\ &\approx \hat{g}(y) + h(y) \\ &= \underbrace{g(x) + \nabla g(x)^T (y-x) + \frac{1}{2t}\|y-x\|_2^2 + h(y)} \end{aligned}$$

$$\begin{aligned} x^+ &= \operatorname*{argmin}_y \\ &= \operatorname*{argmin}_y \frac{1}{2t}\|(x - t\nabla g(x)) - y\|_2^2 + h(y) \\ &= \operatorname*{argmin}_y \frac{1}{2t}\|x - y - t\nabla g(x)\|_2^2 + h(y) \end{aligned}$$

$$\operatorname{prox}_{h,t}(x) = \operatorname*{argmin}_z \frac{1}{2t}\|x - z\|_2^2 + h(z)$$

$$\boxed{x^+ = \operatorname{prox}_{h,t}\big(\overset{a}{\boxed{x - t\nabla g(x)}}\big)}$$

$$= x - t\,\boxed{G_t(x)} \longleftarrow \text{generalized grad.}$$

i.e. $G_t(x) = \dfrac{x - \operatorname{prox}(x - t\nabla g(x))}{t}$

$$\min_{\beta} \underbrace{\frac{1}{2}\|y - X\beta\|_2^2}_{} + \lambda\|\beta\|_1$$

$$f(\beta) = g(\beta) + h(\beta)$$

$$\operatorname{prox}_t(x) = \operatorname*{argmin}_z \frac{1}{2}\|z - x\|_2^2 + t\lambda\|z\|_1$$

$$= S_{t\lambda}(x)$$

$$\beta^{(k)} = \operatorname{prox}_t(\beta^{(k-1)} - t_k \nabla g(\beta^{(k-1)}))$$

$$= S_{\lambda t}(\beta^{(k-1)} - t_k \nabla g(\beta^{(k-1)}))$$

$\nabla g(\beta) =$
$-X^T(y - X\beta)$

$$= S_{\lambda t}\left(\beta^{(k-1)} + t_k X^T(y - X\beta^{(k-1)})\right)$$