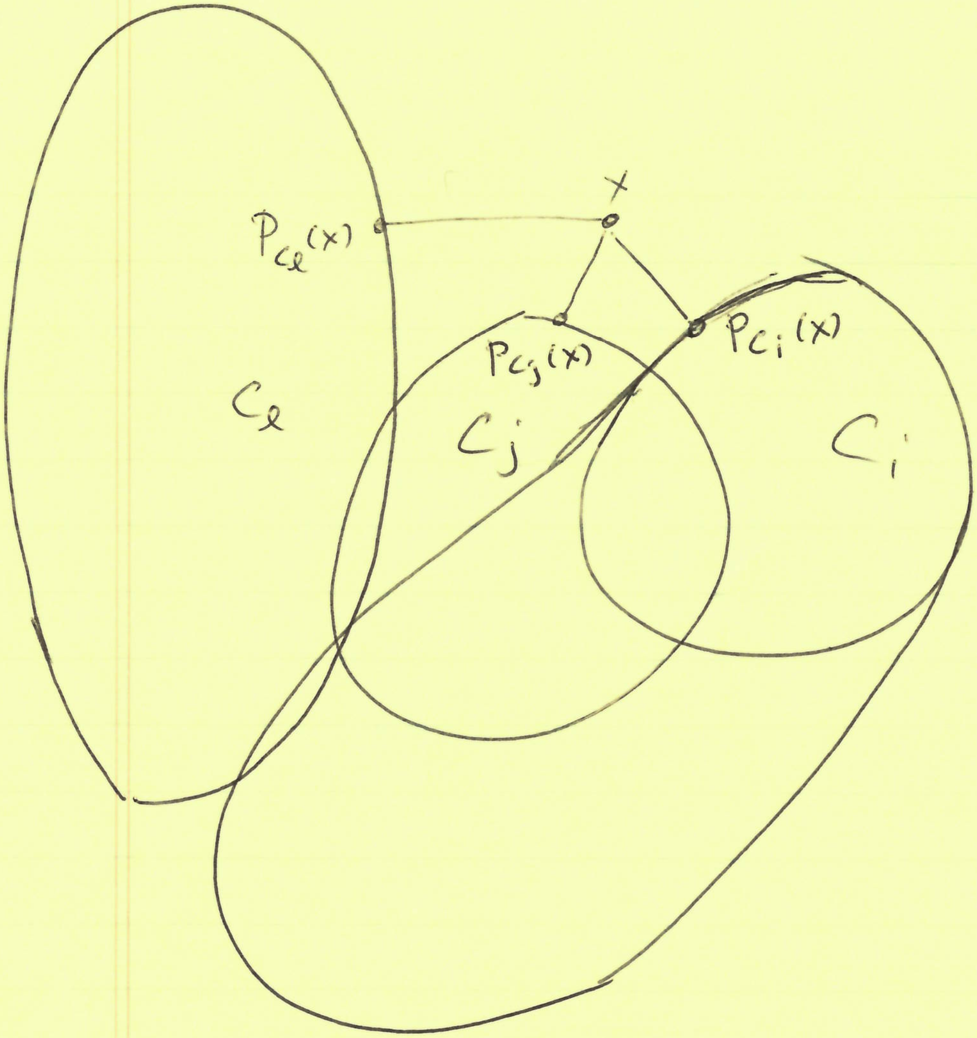


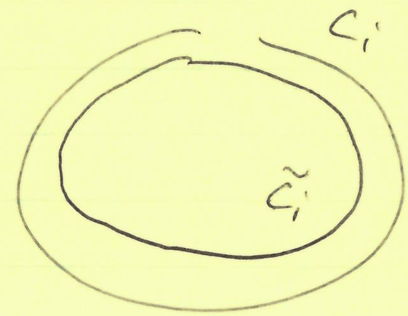
C_i farthest away from $x^{(k-1)}$

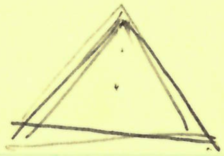
$$f(x^{(k-1)}) = \text{dist}(x^{(k-1)}, C_i)$$
$$= \|x^{(k-1)} - P_{C_i}(x^{(k-1)})\|_2$$

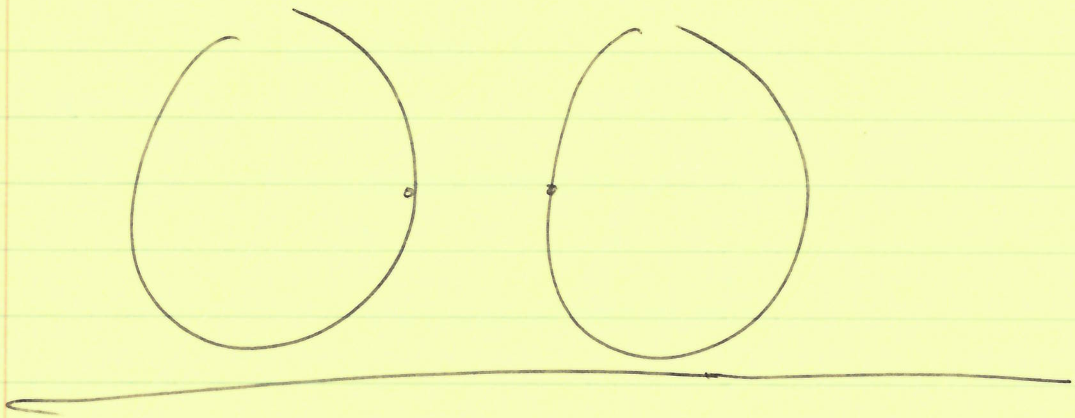
subgradient update: $$x^{(k-1)} - \cancel{f(x^{(k-1)})} \cdot \frac{(x^{(k-1)} - P_{C_i}(x^{(k-1)}))}{\cancel{\|x^{(k-1)} - P_{C_i}(x^{(k-1)})\|_2}}$$

$$= P_{C_i}(x^{(k-1)})$$

$$C_i = \{x : g_i(x) \le t_i\}$$
$$\tilde{C}_i = \{x : g_i(x) \le t_i - \delta\}$$

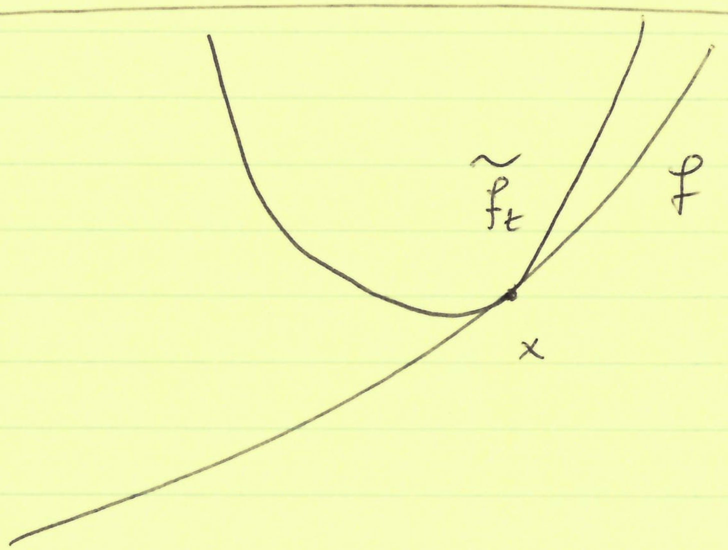






f_i each being G - Lipschitz
$i=1,\dots m$

$\Rightarrow \sum_{i=1}^{m} f_i$ is (mG)- Lipschitz



3

$$\text{prox}(\beta) = \underset{z}{\text{argmin}} \ \frac{1}{2t\lambda} \|\beta - z\|_2^2 + \|z\|_1$$

$$= \underset{z}{\text{argmin}} \ \frac{1}{2} \|\beta - z\|_2^2 + (t\lambda) \cdot \|z\|_1$$

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

$$\beta^+ = S_{\lambda t}(\beta - t \cdot \nabla g(\beta))$$

ISTA

Backtracking for grad descent

check:

$$f(x - t\nabla f(x)) > f(x) - \frac{t}{2} \|\nabla f(x)\|_2^2$$

prox grad

check:

$$g(x - tG_t(x)) > g(x) - t\nabla g(x)^T G_t(x) + \frac{t}{2} \|G_t(x)\|_2^2$$

if TRUE: shrink $t = \beta t$, check again

if FALSE: take $x^+ = x - t \cdot G_t(x)$

$= \text{prox}_t(x - t\nabla g(x))$