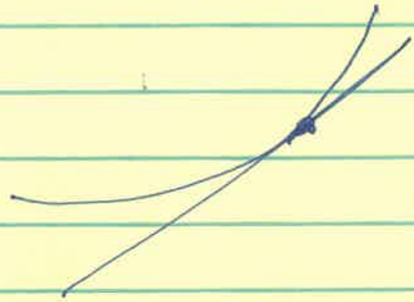


g subgrad of f at x
means

$$f(y) \geq f(x) + g^T(y-x) \quad \text{all } y.$$

∇f Lipschutz constant L

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|x-y\|_2^2 \quad \leftarrow$$

$y = x^+ = x - t\nabla g(x)$ plug in.

take $t \leq \frac{1}{L}$. $\Rightarrow$

$$f(x^+) \leq f(x) - \frac{t}{2}\|\nabla f(x)\|_2^2. \quad \leftarrow$$

Eg. $t_k = 1/k$ square summable
not summable.

---

f is Lipschutz with constant G_2 $\Longrightarrow$ any $g \in \partial f(x)$ satisfies $\|g\|_2 \leq G_2$

Proof.

$$\|x^{(k)} - x^*\|_2^2 = \|x^{(k-1)} - t_k g^{(k-1)} - x^*\|_2^2$$

$$= \|x^{(k-1)} - x^*\|_2^2 + \underbrace{t_k^2\|g^{(k-1)}\|_2^2}_{\leq t_k^2 G_2^2.} - 2t_k\underbrace{(g^{(k-1)})^T(x^{(k-1)} - x^*)}$$

$$f(x^*) \geq f(x^{(k-1)}) + (g^{(k-1)})^T(x^* - x^{(k-1)})$$

$$(g^{(k-1)})^T(x^* - x^{(k-1)}) \leq -f(x^{(k-1)}) + f(x^*)$$

$$\|x^{(k)} - x^*\|_2^2 \le \|x^{(k-1)} - x^*\|_2^2 + t_k^2 G^2 - 2t_k\left(f(x^{(k-1)}) - f(x^*)\right)$$

Iterate over $i = 1, \dots k \dots$

$$\|x^{(k)} - x^*\|_2^2 \le \underbrace{\|x^{(0)} - x^*\|_2^2}_{R^2} + \sum_{i=1}^{k} t_i^2 G^2 - 2\sum_{i=1}^{k} t_i\left(f(x^{(i-1)}) - f(x^*)\right)$$

$$0 \le R^2 + \sum_{i=1}^{k} t_i^2 G^2 - 2\sum t_i\left(f(x^{(i-1)}) - f(x^*)\right)$$

$$2\sum t_i \cdot \left(f(x^{(k)}_{best}) - f(x^*)\right) \le 2\sum t_i\left(f(x^{(i-1)}) - f(x^*)\right) \le R^2 + \sum t_i^2 G^2$$

$$\boxed{f(x^{(k)}_{best}) - f(x^*) \le \frac{R^2 + \sum t_i^2 G^2}{2\sum t_i}} \leftarrow$$

$t_i = t$ all i

RHS: $\dfrac{R^2 + t^2 k G^2}{2tk} \xrightarrow[k \to \infty]{\text{as}} \dfrac{G^2 t}{2}$ ✓

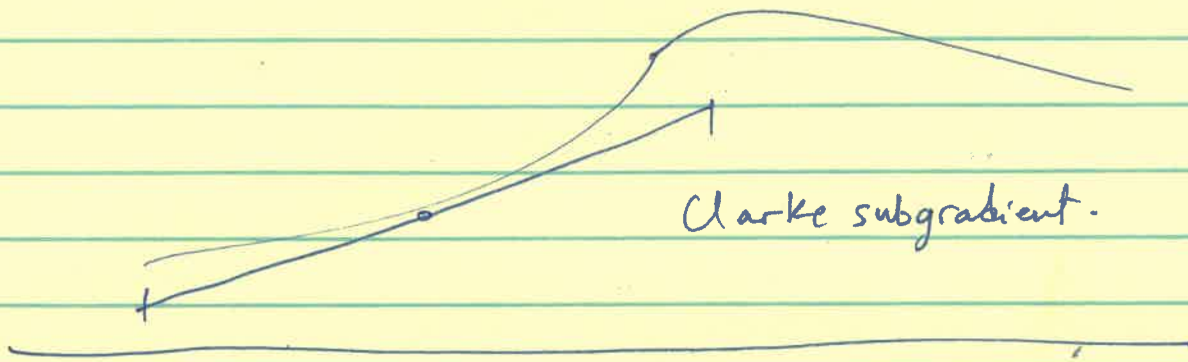
t_i diminishing $\dfrac{R^2 + \sum t_i^2 G^2}{2\sum t_i} \xrightarrow[k \to \infty]{\text{as}} 0$ ✓

$1/\varepsilon$ vs $1/\varepsilon^2$?

$\varepsilon = 10^{-2} = 0.01$

$1/\varepsilon$ 100 iterations

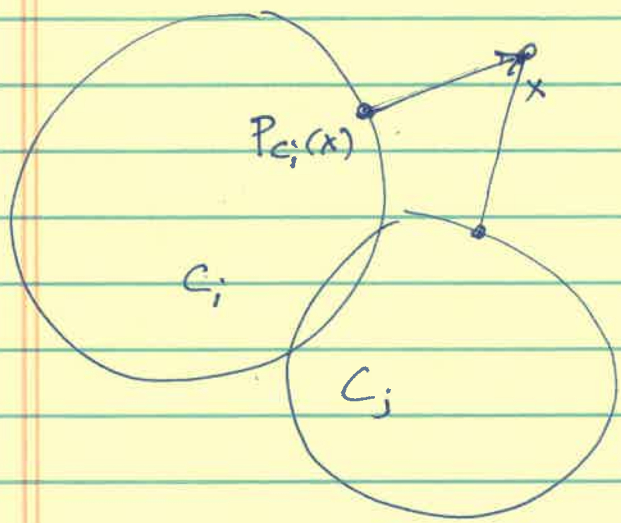
$1/\varepsilon^2$ 10K iterations



Clarke subgradient.

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

$$\nabla f_i(x) = \frac{x - P_{C_i}(x)}{\|x - P_{C_i}(x)\|_2}$$



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

$$x^{(k)} = x^{(k-1)} - t_k \, g^{(k-1)}$$

$$= x^{(k-1)} - \frac{\cancel{f(x^{(k-1)}) - f^*}}{1} \, \frac{x^{(k-1)} - P_{C_i}(x^{(k-1)})}{\cancel{\|x^{(k-1)} - P_{C_i}(x^{(k-1)})\|_2}}$$

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

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