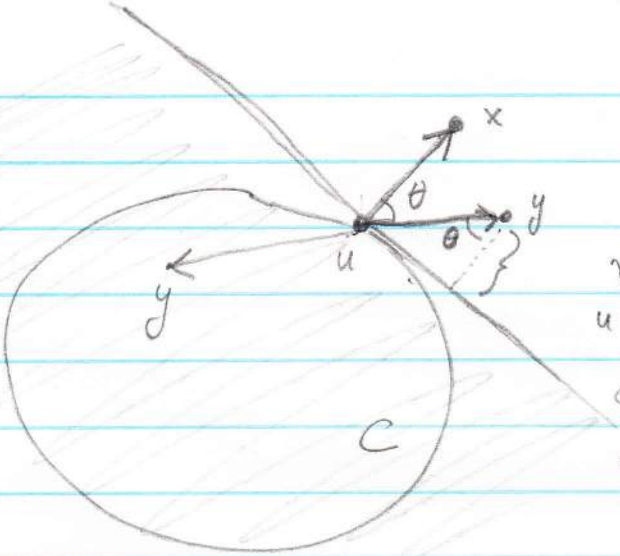


$u = P_C(x)$

$$\min_{u \in C} \tfrac{1}{2}\|u - x\|_2^2 = f(u)$$

$\langle \nabla f(u), y-u \rangle \geq 0$ all $y \in C$

$\langle u-x, y-u \rangle \geq 0$ all $y \in C$

$$C \subseteq H = \{y : \langle u-x, y-u \rangle \geq 0\}$$

Claim: for any y

$$\boxed{\operatorname{dist}(y, C) \geq \frac{(x-u)^T (y-u)}{\|x-u\|_2}}$$

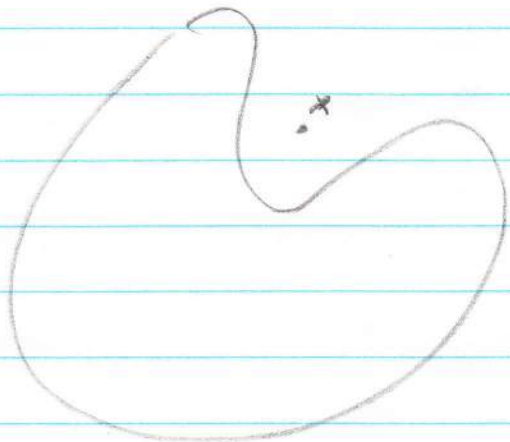
- $y \in H$: RHS is ≤ 0, ✓
- $y \notin H$: $\dfrac{(x-u)^T(y-u)}{\|x-u\|_2} = \cancel{\|x-u\|_2}\underbrace{\|y-u\|_2 \cos\theta}$

$$= \operatorname{dist}(y, H) \quad ✓$$
$$\leq \operatorname{dist}(y, C)$$

$$\underset{\displaystyle f(y)}{\overset{\shortparallel}{\operatorname{dist}(y,C)}} \geq \frac{(x-u)^T(y-x+x-u)}{\|x-u\|_2}$$

$$= \underset{\displaystyle f(x)}{\overset{\shortparallel}{\|x-u\|_2}} + \underbrace{\frac{(x-u)^T}{\|x-u\|_2}}_{g} (y-x)$$

true for all y



$t_k = \frac{1}{k}$

$$f(x_{best}^{(k)}) - f^* \le \underbrace{\frac{R^2 + G^2 \sum_{i=1}^{k} t_i^2}{2 \sum_{i=1}^{k} t_i}}$$

$R = \|x^{(0)} - x^*\|_2$

Fixed: $t_k = t$ Bound is $\frac{R^2 + G^2 k t^2}{2kt}$

$$= \frac{R^2}{2kt} + \frac{G^2 t}{2}$$

Dimishing:

$$\lim_{k \to \infty} f(x_{best}^{(k)}) - f^* \le 0 + 0 = 0$$

Fixed: suppose want bound to be $\le \varepsilon$

make $\frac{R^2}{2kt} \le \frac{\varepsilon}{2}$ and $\frac{G^2 t}{2} \le \frac{\varepsilon}{2}$

so $t = \frac{\varepsilon}{G^2}$ and

$$k = \frac{R^2}{t\varepsilon} = \frac{R^2 G^2}{\varepsilon^2}$$

ie. convergence rate is $O\left(\frac{1}{\varepsilon^2}\right)$

compare to rate $O\left(\frac{1}{\varepsilon}\right)$ for grad desc.

$$\nabla \operatorname{dist}(x, C_i) = \frac{x - P_{C_i}(x)}{\|x - P_{C_i}(x)\|_2}$$

$$f(x) = \max_{i=1\ldots m} f_i(x) = \max_{i=1\ldots m} \operatorname{dist}(x, C_i)$$

$$\partial f(x) = \operatorname{Conv}\left(\bigcup_{\substack{i \text{ active},\\ f_i(x) = f(x)}} \partial f_i(x) \right)$$

$$\partial f(x) = \operatorname{Conv}\left(\bigcup_{\substack{C_i \text{ is farthest}\\ \text{from } x}} \left\{ \frac{x - P_{C_i}(x)}{\|x - P_{C_i}(x)\|_2} \right\} \right)$$

---

repeat:

- find i such that C_i farthest from x
- take $g = \frac{x - P_{C_i}(x)}{\|x - P_{C_i}(x)\|_2}$
- $x^+ = x - t \cdot \frac{x - P_{C_i}(x)}{\|x - P_{C_i}(x)\|_2}$

take Polyak step size

$$t = \frac{f(x) - f^*}{\|g\|_2^2}$$

$$= \frac{\|x - P_{C_i}(x)\|_2 - 0}{1}$$

so

$$x^+ = x - \cancel{\|x - P_{C_i}(x)\|_2} \cdot \frac{(x - P_{C_i}(x))}{\cancel{\|x - P_{C_i}(x)\|_2}}$$

$$= P_{C_i}(x)$$

alternating projections algorithm

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

$$\sum_{i=1}^{n} \left(\underbrace{-y_i x_i^T\beta + \log(1+e^{-x_i^T\beta})}_{f_i(\beta)} \right)$$

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

suppose f_i differentiable

cylic rule, step size t

stochastic: $x_{k+m} = x_k - t \sum_{i=1}^{m} \nabla f_i(x_{k+i-1})$

batch: $x_{k+1} = x_k - t \sum_{i=1}^{m} \nabla f_i(x_k)$

$$x_{k+m} = x_k - t\nabla f(x_k) + E$$