Subgradient Method

Problem: $f: \mathbb{R}^n \to \mathbb{R}$, convex

$$\min_{x \in Q \subseteq \mathbb{R}^n} f(x)$$

Fix $R \geq \|x_0 - x^*\|$. The method, $k \geq 0$:

$$x_{k+1} = \pi_Q\left(x_k - \frac{\gamma_k}{\|f'(x_k)\|} f'(x_k)\right)$$

$\{\gamma_k\}$ - controlling parameters

Measure of optimality:

$$\Delta_k = \frac{\langle f'(x_k), x_k - x^* \rangle}{\|f'(x_k)\|}$$

Define $\Delta_k^* = \min\limits_{0 \leq i \leq k-1} \Delta_i$.

Theorem For any $\{\gamma_k\}_{k \geq 0}$:

$$\frac{1}{2}\sum_{i=1}^{k-1} \gamma_i^2 + \frac{R^2}{2} \geq \sum_{i=0}^{k-1} \gamma_i \Delta_i \geq \left(\sum_{i=0}^{k-1} \gamma_i\right) \Delta_k^*$$

We get:

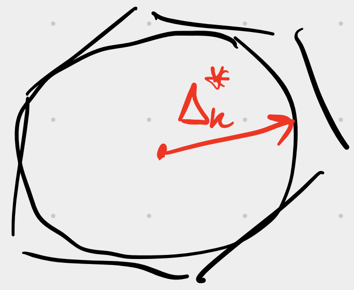
$$\Delta_k^* \leq \frac{\sum_{i=1}^{k-1} \gamma_i^2 + R^2}{2\sum_{i=0}^{k-1} \gamma_i} = \varphi(\gamma) \to \min_{\gamma}$$

- $\varphi(\cdot)$ convex
- $\varphi(\cdot)$ is symmetric γ

$$\gamma_1^* = \ldots = \gamma_K^* = \gamma = \frac{R}{\sqrt{K}}.$$

<u>Corollary</u>

$$\Delta_k^* \le \frac{R}{\sqrt{K}}.$$



Functional Growth

convexity

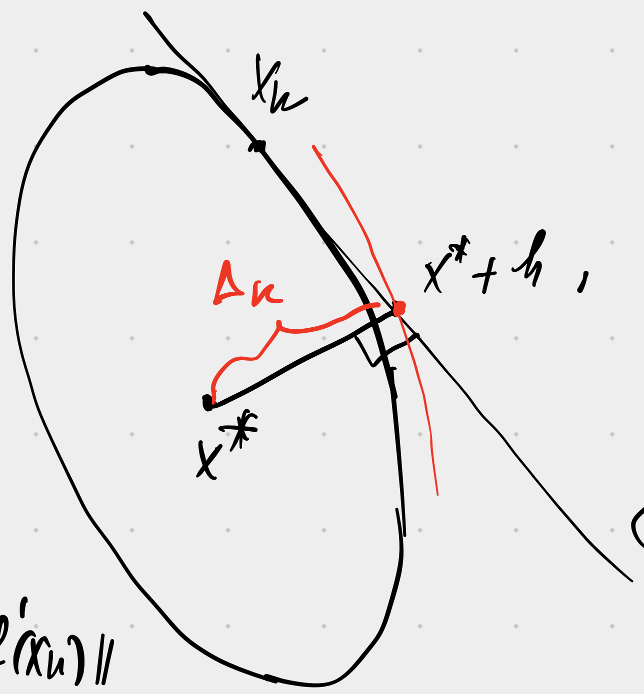
$$\Delta_k = \frac{\langle f'(x_k), x_k - x^* \rangle}{\|f'(x_k)\|} \geq \frac{f(x_k) - f^*}{\|f'(x_k)\|}$$

$$\|f'(x)\| \le M \quad \forall x \in Q \quad \Rightarrow \quad \Delta_k \geq \frac{f(x_k) - f^*}{M}.$$

it's enough:

$$x_{k+1} = \pi_Q\left(x_k - \frac{\gamma_k}{M} f'(x_k)\right)$$

$$\Rightarrow \quad f(x_k) - f^* \le \frac{MR}{\sqrt{K}}.$$



$$h = \frac{\Delta_k}{\|f'(x_k)\|} f'(x_k)$$

$$\Delta_k \|f'(x_k)\| = \langle f'(x_k), x_k - x^* \rangle = \langle f'(x_k), x_k - x^* - h \rangle + \langle f'(x_k), h \rangle \geq$$

$$\geq f(x_k) - f(x^* + h) + \langle f'(x_k), h \rangle = f(x_k) - f^* - (f(x^* + h) - f^*) + \Delta_k \|f'(x_k)\|$$

$$\Rightarrow \quad f(x_k) - f^* \le f(x^* + h) - f^* \qquad h = \Delta_k \frac{f'(x_k)}{\|f'(x_k)\|}$$

$\|h\| = \Delta_k$

Functional Growth at x^*:

$$\omega_f(r) = \max \{ f(x) - f(x^*) : \|x - x^*\| \le r \}$$

$\omega_f(\cdot) \nearrow$ increasing in $r \ge 0$.

Theorem: $x_k^* = \operatorname{argmin} \{ f(y) : y \in \{x_0, \ldots, x_k\} \}$.

$$f(x_k^*) - f^* \le \omega_f(\Delta_k^*), \quad \Delta_k^* \le \frac{R}{\sqrt{k}}.$$

Proof
$$f(x_k) - f^* \le f\left(x^* + \Delta_k \cdot \frac{f'(x_k)}{\|f'(x_k)\|}\right) - f^*$$
$$\le \omega_f\left(\left\| \Delta_k \frac{f'(x_k)}{\|f'(x_k)\|} \right\| \right) = \omega_f(\Delta_k). \quad \square$$

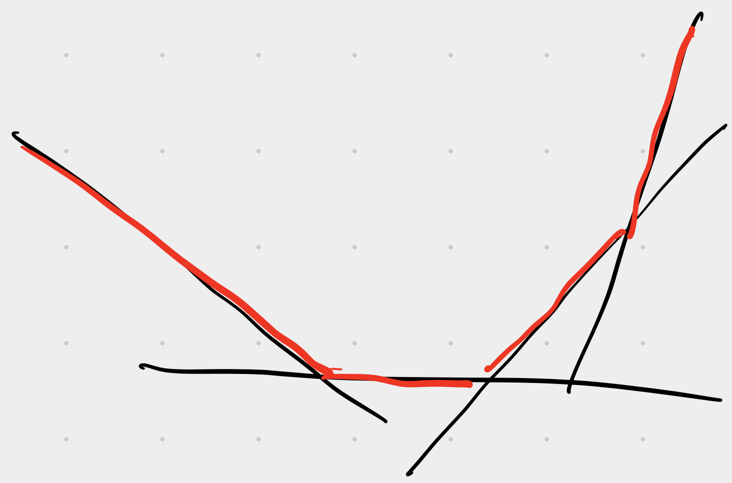
Example

f is Lipschitz: $f(x+h) - f(x) \le M \cdot \|h\|, \ \forall x, h$

$$\omega_f(r) = \max \{ f(x^* + h) - f^* : \|h\| \le r \} \le Mr.$$

$$\Rightarrow \quad f(x_k^*) - f^* \le \frac{MR}{\sqrt{k}}.$$

Consider



$$f(x) = \max_{1 \le i \le m} [\langle a_i, x \rangle - b_i]$$

$$M = \max_{1 \le i \le m} \|a_i\|$$

$\forall x, y: \quad f(y) - f(x) \le M\|y - x\|,$

$$f(y)-f(x) = \max_{1\le i\le m}\left[\langle a_i, y\rangle - b_i \quad - \max_{1\le j\le m}[\langle a_j, x\rangle - b_j]\right]$$

$$\le \max_{1\le i\le m}\left[\langle a_i, y\rangle - b_i - \langle a_i, x\rangle + b_i\right]$$

$$= \max_{1\le i\le m}\langle a_i, y-x\rangle \le \max_{1\le i\le n}\|a_i\|\cdot\|y-x\|.$$

$f'(x) = a_i$, maximum is achieved at i^{th} component

Example

$$f(x) = \max_{1\le i\le m} f_i(x), \quad f_i(x) - \text{smooth, convex.}$$

$$f_i(x) = \frac{1}{2}\langle A_i x, x\rangle - \langle b_i, x\rangle - c_i$$

By Lipschitzness of $f_i'(x)$:

$$f_i(x) \le f_i(x^*) + \langle f_i'(x^*), x - x^*\rangle + \frac{L_i}{2}\|x-x^*\|^2$$

$$w_f(r) = \max_{\|h\|\le r}\max_{1\le i\le m}\left[f_i(x^*+h) - f_i(x^*)\right] \le$$

$$\le \underline{\max_{1\le i\le m}\|f_i'(x^*)\|\cdot r} + \max_{1\le i\le m}\frac{L_i}{2}\cdot r^2.$$

$$\Delta_n^* \le \frac{R}{\sqrt{K}}$$
$$\shortparallel$$
$$r$$

We obtain:

$$f(x_n^*) - f^* \le \frac{\max_{1\le i\le m}\overbrace{\|f_i'(x^*)\|}\cdot R}{\sqrt{K}} + \frac{\max_{1\le i\le m} L_i}{2}\cdot\frac{R^2}{K}.$$

As a particular case: $\min_{x \in \mathbb{R}^n} f_1(x)$, $f_1'(x^*) = 0$

$$\Rightarrow f(x_k^*) - f^* \leq \frac{LR^2}{2K} .$$

$$x_{k+1} = x_k - \lambda_k f'(x_k)$$

① $\lambda_k = \frac{1}{L}$

② $\lambda_k = \frac{\gamma}{\|f'(x_k)\|}$, $\gamma = \frac{R}{\sqrt{K}}$.

Lower Complexity Bound

Problem Class

$$\min_{\|x\| \le R} f(x) \quad , \quad f \text{ is convex, Lipschitz: } |f(y) - f(x)| \le M \cdot \|y - x\|.$$

Subgradient method:

$$f(x_k^*) - f^* \le \frac{MR}{\sqrt{k}} = \varepsilon$$

$$k = \left[\frac{MR}{\varepsilon}\right]^2.$$

Theorem Let $M, R > 0$ be fixed. Then, for any first-order algorithm running for $K \ge 1$ iterations:

$\exists$ a function $f : \mathbb{R}^n \to \mathbb{R}$, $n \ge K$ such that:

1. f is convex, it is Lipschitz with const. $M > 0$.
2. $$f(x_K) - f^* \ge \frac{MR}{2\sqrt{K}}.$$ $\Rightarrow$ Subgradient method is optimal.

Remark $\mathcal{A} = (A_1, A_2, \ldots)$ - algorithm

$x_1 \in Q$ - initial point

$$x_{k+1} = A_k(\mathcal{O}_f(x_1), \ldots, \mathcal{O}_f(x_k)), \quad \mathcal{O}_f(x) = \{f(x), f'(x)\}$$
for any $f'(x) \in \partial f(x)$.

Proof. Let $\delta > 0$ be a fixed small parameter.

Fix $\xi_1, \dots, \xi_k \in \{-1, 1\}$

Fix a permutation

$$t \mapsto \sigma(t) \in \{1, 2, \dots, n\}$$

Consider the family of functions:

$$f_k(x) := M \cdot \max_{1 \le i \le k} \Big[\xi_i \underbrace{\langle e_{\sigma(i)}, x \rangle}_{x^{(\sigma(i))}} - (i-1)\delta \Big]$$

where

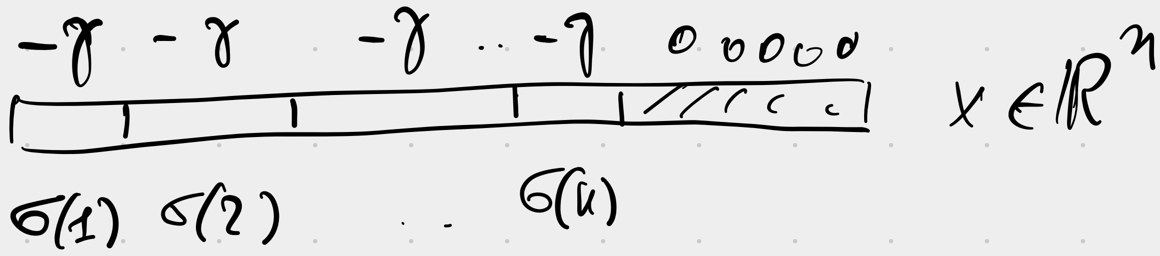
$$e_i \in \mathbb{R}^n \text{ basis vector}$$

1. $f_k(x)$ is Lipschitz with const $M > 0$.

2. $$f_k^* = \min_{\|x\| \le R} f_k(x) = M \cdot \min_{\|x\| \le R} \max_{1 \le i \le k} \Big[\xi_i \langle e_{\sigma(i)}, x \rangle - (i-1)\delta \Big]$$

$$\le M \cdot \min_{\|x\| \le R} \max_{1 \le i \le k} \xi_i \langle e_{\sigma(i)}, x \rangle =$$

$$= M \cdot \min_{\|x\| \le R} \max_{1 \le i \le k} \langle e_{\sigma(i)}, x \rangle = -\frac{MR}{\sqrt{k}}.$$



$$\|(\underbrace{-\gamma, \dots, -\gamma}_{k}, 0, \dots, 0)\|^2 = k \cdot \gamma^2 \overset{?}{=} R^2 \qquad \gamma := -\frac{R}{\sqrt{k}}.$$

2. Resisting strategy.

First Step. x_1 - initial point

Choose $\sigma(1) \in \underset{1 \le i \le n}{\operatorname{argmax}} |\langle e_i, x_1 \rangle|$

We choose $\xi_1 \in \{-1, 1\}$:

$$\xi_1 \langle e_{\sigma(1)}, x_1 \rangle = |\langle e_{\sigma(1)}, x_1 \rangle|$$

$$\xi_1 = \operatorname{sign}(\langle e_{\sigma(1)}, x_1 \rangle)$$

At step $1 \le k \le K-1$

we have built $f_k(x)$.

The algorithm: $x_{k+1} = \mathcal{A}(\mathcal{O}_{f_k}(x_1), \dots, \mathcal{O}_{f_k}(x_k))$.

Let us choose:

$$\sigma(k+1) \in \underset{\substack{1 \le i \le n \\ i \notin \{\sigma(1), \dots, \sigma(k)\}}}{\operatorname{argmax}} |\langle e_i, x_{k+1} \rangle|$$

Specify $\xi_{k+1} \in \{-1, 1\}$:

$$\xi_{k+1} \langle e_{\sigma(k+1)}, x_{k+1} \rangle = |\langle e_{\sigma(k+1)}, x_{k+1} \rangle|$$

3. let $s < k$,

$f_s(\cdot)$ and $f_n(\cdot)$ – informationally indistinguishable

It's enough to prove: $\mathcal{O}_{f_s}(x_s) = \mathcal{O}_{f_n}(x_s)$

Note:

$$f_n(x) = \max\{ f_s(x), \underbrace{M \cdot \max_{s<i\le k} [\xi_i \langle e_{\sigma(i)}, x\rangle - (i-1)\delta]}\}$$

By the definition of ξ_s, $\sigma(s)$:

$$\xi_s \langle e_{\sigma(s)}, x_s\rangle \geq \xi_i \langle e_{\sigma(i)}, x_s\rangle$$
$$\forall\ i > s.$$

Therefore:

$$f_s(x_s) \geq M \cdot [\xi_i \langle e_{\sigma(i)}, x_s\rangle - (i-1)\delta] + M\delta$$

By Lipschitzness:

$$f_s(x) \geq M [\xi_i \langle e_{\sigma(i)}, x\rangle - (i-1)\delta], \quad \|x - x_s\| \leq \delta$$

We have shown:

$$f_s(x) \equiv f_k(x), \quad \|x - x_s\| \leq \delta.$$

$$k > s$$

...