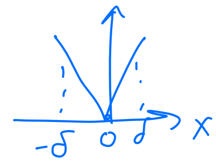
$$\min_{x\in\mathbb{R}^d} f(x)$$

f - ~~L-гладкая~~

$f(x) = |x|$



$$|f'(0+\delta) - f'(0-\delta)| =$$

$$= 2 \le \underbrace{L \cdot 2\delta}_{\cancel{\text{гладкость}}}$$

$$\delta \to 0 \quad L \to \infty$$

<u>Предположение</u> (вместо гладкости)

$f: \mathbb{R}^d \to \mathbb{R}$ является M-липшицевой, если
$$\forall x, y \in \mathbb{R}^d \hookrightarrow |f(x) - f(y)| \le M\|x-y\|_2$$

f - выпуклая и M-липшицева

<u>Субградиент</u>

$f: \mathbb{R}^d \to \mathbb{R}$ выпукл. Вектор $g \in \mathbb{R}^d$ назыв. субградиентом ф. f в т. $x \in \mathbb{R}^d$, если

$$\forall y \in \mathbb{R}^d \hookrightarrow f(y) \ge f(x) + \langle \underset{\nabla f(x)}{g}; y - x \rangle$$ (выпуклость)

<u>Субдифференциал</u> $\partial f(x)$ - мн-во всех субградиентов ф. f в т. x

Условие оптимальности

x^* – минимум выпуклой ф. f $\Leftrightarrow$ $0 \in \partial f(x^*)$

Док-во: $\Leftarrow$ $0 \in \partial f(x^*)$

опред. субградиента вып. ф. f:

$$f(x) \geq f(x^*) + \langle \underbrace{g}_{0 \,\in\, \partial f(x^*)} ; x - x^* \rangle = f(x^*) \qquad \forall x \in \mathbb{R}^d$$

опред. глоб. минимума

$$\Rightarrow \quad f(x) \geq f(x^*) \quad \forall x \in \mathbb{R}^d,$$ тогда

$$\underbrace{f(x) \geq f(x^*) + \underbrace{\langle 0; x - x^* \rangle}_{0} \quad \forall x \in \mathbb{R}^d}_{\text{определение субград в т. } x^*} \Rightarrow 0 \in \partial f(x^*) \ \blacksquare$$

Лемма

$f: \mathbb{R}^d \to \mathbb{R}$ выпуклая, тогда

f M-липшицева

$\Leftrightarrow$

$\forall x \in \mathbb{R}^d$ и $\forall g \in \partial f(x) \hookrightarrow \|g\|_2 \leq M$

Док-во:

$\Rightarrow$ f - выпукла, M-липшицева,

$\measuredangle$ $g \in \partial f(x)$

определение субградиента вып. ф. f

$$f(y) - f(x) \geq \langle g; y - x \rangle \qquad \forall y \in \mathbb{R}^d$$

M- липшицевость

$$\langle g; y-x\rangle \le f(y)-f(x) \le |f(y)-f(x)| \le M\|x-y\|_2$$

$y = x+g$

$$\|g\|_2^2 \le M\|g\|_2 \Rightarrow \boxed{\|g\|_2 \le M}$$

$\Leftarrow$ f - выпукла,

$\forall x \in \mathbb{R}^d,\ \forall g \in \partial f(x) \hookrightarrow \|g\|_2 \le M$

$\triangle$ $g \in \partial f(x)$

выпуклость и опред субград.

$$f(y)-f(x) \ge \langle g; y-x\rangle \quad |\cdot(-1)$$

$$\langle g; x-y\rangle \ge f(x)-f(y)$$

КБШ

$$f(x)-f(y) \le \langle g; x-y\rangle \le \|g\|_2\cdot\|x-y\|_2$$

$\|g\|_2 \le M$

$$f(x)-f(y) \le M\cdot\|x-y\|_2$$

аналогично

$$f(y)-f(x) \le M\|x-y\|_2$$

$$\Rightarrow |f(x)-f(y)| \le M\|x-y\|_2$$

$\blacksquare$

Субградиентный метод

$$\boxed{x^{k+1} = x^k - \gamma g^k \qquad g^k \in \partial f(x^k)}$$

Док-во:

$$\|x^{k+1}-x^*\|_2^2 = \|x^k-\gamma g^k - x^*\|_2^2$$
$$= \|x^k - x^*\|_2^2 - 2\gamma\langle g^k; x^k - x^*\rangle + \gamma^2\|g^k\|_2^2$$

M – липшицевость $\Rightarrow \|g\|_2 \le M$

$$\le \|x^k-x^*\|_2^2 - 2\gamma\langle g^k; x^k-x^*\rangle + \gamma^2 M^2$$

выпуклость и опред. субград

$$\le \|x^k-x^*\|_2^2 - 2\gamma\left(f(x^k)-f(x^*)\right) + \gamma^2 M^2$$

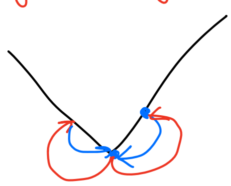
$$2\gamma\left(f(x^k)-f(x^*)\right) \le \|x^k-x^*\|_2^2 - \|x^{k+1}-x^*\|_2^2 + \gamma^2 M^2$$

$$\sum r^k - r^{k+1} = \ldots + \underline{r^i} - \cancel{r^{i+1}} + \cancel{r^{i+1}} - \underline{r^{i+2}}$$

$$\frac{1}{K}\sum_{k=0}^{K-1}$$

$$2\gamma\cdot\frac{1}{K}\sum_{k=0}^{K-1}\left(f(x^k)-f(x^*)\right) \le \frac{\|x^0-x^*\|_2^2 - \cancel{\|x^K-x^*\|_2^2}}{K} + \gamma^2 M^2$$

для гладких задач $f(x^{k+1}) \le f(x^k)$ – так нельзя без гладкости



нер-во Йенсена

$$2\gamma\cdot\Big(f\Big(\underbrace{\frac{1}{K}\sum_{k=0}^{K-1}x^k}_{\text{по средней точке}}\Big) - f(x^*)\Big) \le \frac{\|x^0-x^*\|_2^2}{K} + \gamma^2 M^2$$

$$\left.f\Big(\frac{1}{K}\sum_{k=0}^{K-1}x^k\Big) - f(x^*) \le \frac{\|x^0-x^*\|_2^2}{2\gamma K} + \gamma\frac{M^2}{2}\right)\min_{\gamma}$$

$$\gamma = \frac{\|x^0 - x^*\|_2}{\sqrt{K} \cdot M}$$

Сходимость

$$\boxed{f\left(\frac{1}{K}\sum_{k=0}^{K-1} x^k\right) - f(x^*) \le \frac{\|x^0 - x^*\|_2 \cdot M}{\sqrt{K}}}$$

в худшем случае град. спуск сход. $\frac{1}{K}$

на этом результат оптимален

Проблема: $\gamma = \frac{\|x^0 - x^*\|_2}{M\sqrt{K}}$ зависимость M, $\|x^0 - x^*\|_2$, K

Решение:

$$\gamma^k = \frac{\|x^0 - x^*\|_2}{\sqrt{k+1}\, M}$$

AdaGradNorm ($M \to \|g\|$)

$$\|x^0 - x^*\|_2 \sim D \leftarrow \text{диаметр}$$

(в общ. случае подбирать)

$$\gamma^k = \frac{\|x^0 - x^*\|_2}{\sqrt{(k+1)\cdot M^2}} \longrightarrow \frac{D}{\sqrt{\underbrace{\sum_{t=0}^{k} \underbrace{\|g^t\|_2^2}_{M^2}}_{(k+1)M^2}}}$$

AdaGrad (индив. подбор шага)

$$\gamma_{k,i} = \frac{D_i}{\sqrt{\sum_{t=0}^{k} (g_i^t)^2}} \qquad D_i \to D \text{ (подбирать)}$$

i — номер коорд.

RMSProp

подбирать

$$\gamma^{k,i} = \frac{\gamma}{h_i^k}$$

в AdaGrad смотрим градиенты именно тот же вес

$$(h_i^k)^2 = \beta_2 (h_i^{k-1})^2 + (1-\beta_2)(g_i^k)^2 \qquad \beta_2 \in (0,1)$$

можно просто 1

нужно подбирать

$\beta_2 = 0{,}99$

$= 0{,}995$

Adam = RMSProp + тяжелый шарик

DOG (избавимся от $\|x^0 - x^*\|_2$)

Если метод сходится: $x^k \to x^*$

$$\|x^0 - x^*\|_2 \to \|x^0 - x^k\|_2 \to \underbrace{d_k = \max_{t \in [0,k]} \{\|x^0 - x^t\|_2\}}_{\text{оценка эффективного } D \text{ мн-ва оптим.}}$$

по коорд:

$$|x_i^0 - x_i^*| \to d_{k,i} = \max_t \{|x_i^0 - x_i^t|\}$$

DoG = AdaGrad + $d_{k,i}$