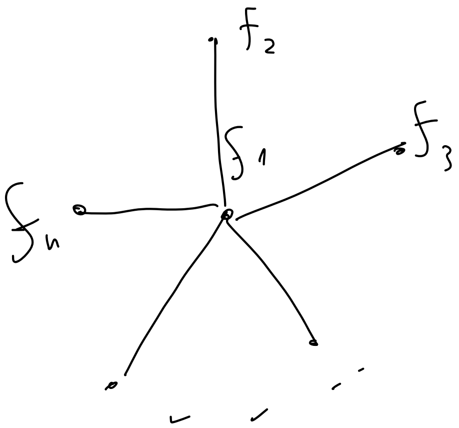
$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\frac{1}{N} \sum_{j=1}^{N} l(x, z_j)}_{f_i(x)}$$



похожесть $f_i \approx f_j$

1) $\|\nabla f_i(x) - \nabla f_j(x)\| < \delta \quad \forall x$

2) $|f_i(x) - f_j(x)| < \delta \quad \forall x$

- $\|\nabla^2 f_i(x) - \nabla^2 f(x)\| < \delta$

почему хорошее?

— l — L-гладкая, выпуклая, дважды дифф. по x

δ — ? в общем $\delta \sim L$

— данные на всех устройствах имеют одну природу

Th. (Матричное нер-во Хёффдинга) $\{X_i\}_{i=1}^{N}$ — квадр. матрицы (размера d), симметричные ($X_i^T = X_i$)

$$\mathbb{E}[X_i] = 0 \qquad X_i^2 \preceq A^2 \qquad \underset{const}{A^T = A}$$

Тогда с вер. $1-p$

$$\left\| \sum_{i=1}^{N} X_i \right\| \leq \sqrt{8N \|A^2\| \ln \frac{d}{p}}$$

В нашем случае $\nabla^2 L(x, z_i) \preceq L I \quad \forall x, z_i$

$$X_j = \frac{1}{N}\left(\nabla^2 l(x, z_j) - \nabla^2 f(x)\right)$$

$\mathbb{E}[\cdot] = \nabla^2 f(x)$

$$\mathbb{E} X_j = 0$$

$$\sum X_i + \nabla^2 f(x) = \nabla^2 f_i(x)$$

$$X_j^2 \preceq \frac{4L^2}{N^2} I$$

Тогда по матричн. н. Хефдинга

$$\left\| \sum X_i \right\| \le \sqrt{8N \cdot \frac{4L^2}{N^2} \ln \frac{d}{p}} = \frac{\sqrt{32}\, L}{\sqrt{N}} \sqrt{\ln \frac{d}{p}}$$

$$\| \nabla^2 f_i(x) - \nabla^2 f(x) \|$$

$$\| \nabla^2 f_i(x) - \nabla^2 f(x) \| < \delta \sim O\left(\frac{L}{\sqrt{N}}\right)$$

$\delta \to 0$ при $N \to \infty$

кол-во лок. данных

Зеркальный спуск:

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \left(\gamma \langle \nabla f(x^k); x \rangle + V(x, x^k)\right)$$

$$V(x, y) = \varphi(x) - \varphi(y) - \langle \nabla \varphi(y); x - y \rangle$$

Предположение для сход. метода

$$\mu_\varphi \nabla^2\varphi(x) \preceq \nabla^2 f(x) \preceq L_\varphi \nabla^2 \varphi(x)$$

$$\Updownarrow$$

$$\mu_\varphi V(x,y) \le f(x) - f(y) - \langle \nabla f(y); x-y\rangle \le L_\varphi V(x,y)$$

у нас везде $V = \frac{1}{2}\|\ \|^2$

Сходимость ЗС в усл. пред.

$$\langle \gamma\nabla f(x^k) + \nabla\varphi(x^{k+1}) - \nabla\varphi(x^k);\ x^{k+1} - x^*\rangle = 0$$

$$\gamma\langle \nabla f(x^k); x^{k+1}-x^k\rangle + V(x^{k+1}; x^k) =$$

$$= V(x^*, x^k) - V(x^*; x^{k+1}) - \gamma\langle \nabla f(x^k); x^k - x^*\rangle$$

$$\gamma = \frac{1}{L_\varphi}$$

$$\underline{\langle \nabla f(x^k); x^{k+1}-x^k\rangle} + L_\varphi V(x^{k+1}; x^k) =$$

$$= L_\varphi V(x^*; x^k) - L_\varphi V(x^*; x^{k+1}) - \langle \nabla f(x^k); x^k - x^*\rangle$$

$+$

L_φ гладкость (отн.)

$$f(x^{k+1}) - f(x^k) \le \underline{\langle \nabla f(x^k); x^{k+1}-x^k\rangle} + L_\varphi V(x^{k+1}; x^k)$$

$$f(x^{k+1}) - f(x^k) \le L_\varphi V(x^*; x_k) - L_\varphi V(x^*; x_{k+1}) - \underline{\langle \nabla f(x^k); x^k - x^* \rangle}$$

$+$

μ_φ - сильно вып (отн.)

$$\mu_\varphi V(x^*; x^k) \le f(x^*) - f(x^k) - \underline{\langle \nabla f(x^k); x^* - x^k \rangle}$$

$$0 \le \underline{f(x^{k+1}) - f(x^*)} \le (L_\varphi - \mu_\varphi) V(x^*; x^k) - L_\varphi V(x^*; x^{k+1})$$

≥ 0

$$V(x^*; x^{k+1}) \le \left(1 - \frac{\mu_\varphi}{L_\varphi}\right) V(x^*; x^k)$$

ЗС $\gamma = \frac{1}{L_\varphi}$

$$O\left(\frac{L_\varphi}{\mu_\varphi} \log\frac{1}{\varepsilon}\right)$$

ЗС + похожесть (δ) + f - μ-сильно сильно вып

Δ $\varphi(x) = f_1(x) + \frac{\delta}{2}\|x\|^2$ $\quad$ f_1 - лок. функция на 1 устр-ве

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \left(\gamma \langle \nabla f(x^k); x \rangle + V(x, x^k)\right)$$

1 раз полный град и все

все ост. (подсчет argmin)

отдельно на 1 устр-ве

кол-во ит. ЗС с φ = кол-во коммуник.

$$\mu_\varphi \nabla^2 \varphi(x) \preceq \nabla^2 f(x) \preceq L_\varphi \nabla^2 \varphi(x)$$

? ? $\quad \nabla^2\varphi(x) = \nabla^2 f_1(x) + \delta I$

L_φ:

$$\|\nabla^2 f_1(x) - \nabla^2 f(x)\| \le \delta \Rightarrow \nabla^2 f(x) - \nabla^2 f_1(x) \preceq \delta I$$

$$\Downarrow$$

$$\nabla^2 f(x) \preceq \underbrace{\delta I + \nabla^2 f_1(x)}_{\nabla^2 \varphi(x)}$$

$$L_\varphi = 1$$

μ_φ: из сильн. вып. f

$$\mu I \preceq \nabla^2 f(x) \Rightarrow \frac{2\delta}{\mu} \cdot \Rightarrow 2\delta I \preceq \frac{2\delta}{\mu} \nabla^2 f(x)$$

$$\Downarrow$$

$$\delta I \preceq \frac{2\delta}{\mu} \nabla^2 f(x) - \delta I$$

из усл.:

$$\nabla^2 f_1(x) - \nabla^2 f(x) \preceq \delta I$$

$$\nabla^2 f_1(x) - \nabla^2 f(x) \preceq \frac{2\delta}{\mu} \nabla^2 f(x) - \delta I$$

$$\nabla^2 f_1(x) + \delta I \preceq \left(\frac{2\delta}{\mu} + 1\right) \nabla^2 f(x)$$

$$\mu_\varphi = \left(\frac{2\delta}{\mu}+1\right)^{-1}$$

Тогда ЗС $\quad O\left(\frac{L_\varphi}{\mu_\varphi}\log\frac{1}{\varepsilon}\right)$

$$\| $$

$$O\left(\left(\frac{\delta}{\mu}+1\right)\log\frac{1}{\varepsilon}\right)$$ итер ЗС

$\|$

коммуник.

для GD число комм. - $O\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$

здесь $O\left(\left(1+\frac{\delta}{\mu}\right)\log\frac{1}{\varepsilon}\right)$ $\quad \delta \sim \frac{L}{\sqrt{N}}$

число комм $\downarrow$ делать N больше!

2014 год O. Shamir

нижние оценки 2015 год O. Shamir

$$O\left(\left(1+\sqrt{\frac{\delta}{\mu}}\right)\log\frac{1}{\varepsilon}\right)$$

2022 год D. Kovalev, A. Beznosikov, E. Borodich

первый опт метод

$$O\left(\left(1+\sqrt{\frac{\delta}{\mu}}\right)\log\frac{1}{\varepsilon}\right)$$