Градиентный спуск:

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

$x \in (E, \|\cdot\|_{\cancel{2}})$

что можем сказать про $\nabla f(x)$? $\nabla f(x) \in (E^*, \|\cdot\|_*)$

до этого $\|\cdot\| = \|\cdot\|_2 \Rightarrow (E^*, \|\cdot\|_*) = (E, \|\cdot\|)$

<u>Модификация</u> (А. Немировский, Д. Юдин)

$$\varphi(x^{k+1}) = \varphi(x^k) - \gamma \nabla f(x^k)$$

$\varphi: E \to E^*$ $\quad \varphi^{-1}: E^* \to E$

шаг град. спуска делаем в „двойственном" пр-ве

<u>Опр.</u> (сильная выпуклость отн. произв. нормы)

$d: \bar{X} \to \mathbb{R}$ является сильно выпуклой в норме $\underline{\|\cdot\|}$ с $\mu > 0$, если

$$\forall x, y \in \bar{X} \hookrightarrow d(x) \geq d(y) + \langle \nabla d(y); x - y \rangle + \frac{\mu}{2}\|x-y\|^2_{\cancel{2}}$$

произв. норма

<u>Опр.</u> (дивергенция Брэгмана)

$\lhd$ 1-сильно выпуклая отн. $\|\cdot\|$ на $\bar{X}$ функция d.

Дивергенция Брэгмана, порожденная ф. d, есть ф. от двух аргументов $V(x,y): \mathbb{X}\times\mathbb{X}\to\mathbb{R}$

$\forall x,y\in\mathbb{X}$ $\quad V(x,y)=d(x)-d(y)-\langle\nabla d(y); x-y\rangle$

Пример

- $d(x)=\frac{1}{2}\|x\|_2^2$ на $\mathbb{R}^d$

$$V(x,y)=\frac{1}{2}\|x\|_2^2-\frac{1}{2}\|y\|_2^2-\langle y; x-y\rangle$$
$$=\frac{1}{2}\|x\|_2^2-\frac{1}{2}\|y\|_2^2+\|y\|_2^2-\langle y; x\rangle$$
$$=\frac{1}{2}\left(\|x\|_2^2-2\langle x;y\rangle+\|y\|_2^2\right)$$
$$=\frac{1}{2}\|x-y\|_2^2$$

- $d(x)=\underbrace{\sum_{i=1}^{d}x_i\log x_i}_{\text{энтропия}}$ на $\Delta=\left\{x\in\mathbb{R}^d \,/\, \sum_{i=1}^{d}x_i=1,\ x_i\geq 0\right\}$

$$V(x,y)=\sum_{i=1}^{d}x_i\log\frac{x_i}{y_i} \leftarrow \text{KL-дивергенция}$$

- $d(X)=tr(X\log X)$ $\quad X$ - матрица

$$V(X,Y)=tr(X\log X-X\log Y-X+Y)$$

(квантовая дивергенция фон Неймана)

Свойства

- ассиметричность (KL-дивергенция)
- сильная выпуклость (из определения)

$$V(x,y)\geq\frac{1}{2}\|x-y\|^2$$

($KL(x\|y) \geq \frac{1}{2}\|x-y\|_1^2$ / нер-во Пинскера)

- неотрицательность
- невыпуклая по 2 аргументу
- теорему Пифагора для V:

$\forall x, y, z \in \underline{X} \hookrightarrow$

$$V(z,x) + V(x,y) - V(z,y) = \langle \nabla d(y) - \nabla d(x); z - x \rangle$$

Док-во: $V(x,y) = d(x) - d(y) - \langle \nabla d(y); x - y \rangle$

$$V(z,x) + V(x,y) = \text{по определению}$$

$$= d(z) - \cancel{d(x)} - \langle \nabla d(x); z - x \rangle$$

$$+ \cancel{d(x)} - d(y) - \langle \nabla d(y); x - y \rangle$$

$$= \underbrace{d(z) - d(y) - \langle \nabla d(y); z - y \rangle}_{V(z,y)}$$

$$- \langle d(x); z - x \rangle$$

$$+ \langle \nabla d(y); z - x \rangle$$

$$= V(z,y) + \langle \nabla d(y) - \nabla d(x); z - x \rangle \;\blacksquare$$

Метод зеркального спуска

$$\boxed{x^{k+1} = \underset{x \in \underline{X}}{\operatorname{argmin}} \left\{ \underbrace{\gamma \langle \nabla f(x^k); x \rangle}_{\gamma(f(x^k) + \langle \nabla f(x^k); x - x^k \rangle)} + V(x, x^k) \right\}}$$

Пример

- $d(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathcal{X} = \mathbb{R}^d$

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^d} \Big\{ \underbrace{\langle \gamma \nabla f(x^k); x\rangle + \frac{1}{2}\|x-x^k\|_2^2}_{\nabla = 0} \Big\}$$

$$\gamma \nabla f(x^k) + \underset{\overset{\uparrow}{x^{k+1}}}{x} - x^k = 0$$

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

(градиентный спуск)

- $d(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathcal{X} \neq \mathbb{R}^d$

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} \Big\{ \underbrace{\gamma\langle \nabla f(x^k); x\rangle + \frac{1}{2}\|x-x^k\|_2^2}_{+ \frac{\gamma^2}{2}\|\nabla f(x^k)\|_2^2 - \gamma\langle \nabla f(x^k); x^k\rangle} \Big\}$$

$$= \arg\min_{x\in\mathcal{X}} \Big\{ \frac{1}{2}\Big(\|x-x^k\|_2^2 + 2\gamma\langle \nabla f(x^k); x-x^k\rangle + \gamma^2\|\nabla f(x^k)\|_2^2\Big) \Big\}$$

$$= \arg\min_{x\in\mathcal{X}} \Big\{ \frac{1}{2}\|x \underbrace{- x^k + \gamma \nabla f(x^k)}_{-y^k}\|_2^2 \Big\} \qquad y^k = x^k - \gamma \nabla f(x^k)$$

$$= \arg\min_{x\in\mathcal{X}} \Big\{ \|x - y^k\|_2^2 \Big\} \leftarrow$$ град. спуск с проекцией

- Условие оптимальности на $\mathbb{R}^d$

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

$$= \arg\min \{ \underbrace{\gamma \langle \nabla f(x^k); x \rangle + d(x) - d(x^k) - \langle \nabla d(x^k); x - x^k \rangle}_{\nabla = 0} \}$$

$$\gamma \nabla f(x^k) + \nabla d(\underset{\overset{\uparrow}{x^{k+1}}}{x}) - \nabla d(x^k) = 0$$

$$\underbrace{\nabla d}_{\varphi}(x^{k+1}) = \underbrace{\nabla d}_{\varphi}(x^k) - \gamma \nabla f(x^k)$$

$$x^{k+1} = (\nabla d)^{-1}(\nabla d(x^k) - \gamma \nabla f(x^k))$$

Опр. Непрер. дифферен. ф. $f: \underline{\overline{X}} \to \mathbb{R}$

f – L-гладкая относительно нормы $\|\cdot\|$ на $\underline{\overline{X}}$,

если $\forall x, y \in \underline{\overline{X}} \hookrightarrow \|\nabla f(x) - \nabla f(y)\|_* \le L \|x - y\|$

Теорема $f: \underline{\overline{X}} \to \mathbb{R}$ L-гладкая отн. $\|\cdot\|$, тогда

$$\forall x, y \in \underline{\overline{X}} \hookrightarrow |f(y) - f(x) - \langle \nabla f(x); y - x \rangle| \le \frac{L}{2} \|x - y\|^2$$

Док-во: ф. Ньютона-Лейбница

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + \tau(y - x)); y - x \rangle \, d\tau$$

$$= \langle \nabla f(x); y - x \rangle + \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x); y - x \rangle d\tau$$

$$|f(y) - f(x) - \langle \nabla f(x); y - x \rangle| = \left| \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x); y - x \rangle d\tau \right|$$

$|\Sigma| \le \Sigma |\ |$

$$\le \int_0^1 |\langle \nabla f(x+\tau(y-x)) - \nabla f(x); y - x \rangle| d\tau$$

КБШ $\langle a; b \rangle \le \|a\|_* \cdot \|b\|$

$$\le \int_0^1 \|\nabla f(x+\tau(y-x)) - \nabla f(x)\|_* \cdot \|y - x\| d\tau$$

L- гладкость отн. $\|\circ\|$

$$\le \int_0^1 L\tau \|y - x\|^2 d\tau$$

$$= \frac{L}{2} \|y - x\|^2 \quad \blacksquare$$

Док-во сходимости

Условие оптимальности для

$$x^{k+1} = \underset{x \in \underline{\overline{X}}}{\operatorname{argmin}} \{ \underbrace{\gamma \langle \nabla f(x^k); x \rangle + V(x, x^k)}_{g(x)} \}$$

$$\forall x \in \bar{\underline{X}} \hookrightarrow \langle \nabla g(\underset{\underset{x^{k+1}}{\uparrow}}{x^*}); \underset{\underset{x^{k+1}}{\uparrow}}{x^*} - x \rangle \le 0$$

$$\langle \gamma \nabla f(x^k) + \nabla d(x^{k+1}) - \nabla d(x^k); x^{k+1} - x \rangle \le 0$$

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

Св-во V: $V(z,x) + V(x,y) - V(z,y) = \langle \nabla d(\underset{\underset{x^k}{\shortparallel}}{y}) - \nabla d(\underset{\underset{x^{k+1}}{\shortparallel}}{x}); \underset{\underset{x}{\shortparallel}}{z} - \underset{\underset{x^{k+1}}{\shortparallel}}{x} \rangle$

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

$$= V(x; x^k) - V(x; x^{k+1}) - V(x^{k+1}, x^k)$$

$x^{k+1} \to x^k$ в скаляр произведении, $x \to x^*$

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

Гладкость

$$| f(y) - f(x) - \langle \nabla f(x); y - x \rangle | \le \frac{L}{2} \|x - y\|^2$$

$$\gamma \left(f(x^{k+1}) - f(x^k) - \underline{\langle \nabla f(x^k); x^{k+1} - x^k \rangle} \right) \le \frac{L}{2} \|x^{k+1} - x^k\|^2 \cdot \gamma$$

$$\gamma \langle \nabla f(x^k); x^k - x^* \rangle \le V(x^*, x^k) - V(x^*, x^{k+1}) - V(x^{k+1}; x^k) + \gamma \frac{L}{2} \|x^k - x^{k+1}\|^2 + \gamma f(x^k) - \gamma f(x^{k+1})$$

выпуклость функции f:

$$\gamma \left(\cancel{f(x^k)} - f(x^*)\right) \le V(x^*, x^k) - V(x^*, x^{k+1}) - V(x^{k+1}; x^k) + \gamma \frac{L}{2} \|x^k - x^{k+1}\|^2 + \cancel{\gamma f(x^k)} - \gamma f(x^{k+1}) \quad \Big| \cdot \frac{1}{\gamma}$$

$$f(x^{k+1}) - f(x^*) \le \frac{1}{\gamma} \left(V(x^*, x^k) - V(x^*, x^{k+1}) - V(x^{k+1}; x^k) \right) + \frac{L}{2} \|x^k - x^{k+1}\|^2$$

$$V(x^{k+1}, x^k) \ge \frac{1}{2} \|x^k - x^{k+1}\|^2$$

$$\le \frac{1}{\gamma} \left(V(x^*, x^k) - V(x^*, x^{k+1}) \right) + \frac{1}{\gamma} \underbrace{(\gamma L - 1)}_{\le 0} \underbrace{V(x^{k+1}; x^k)}_{\ge 0}$$

$$\gamma \le \frac{1}{L}$$

$$\leq \frac{1}{\gamma}\left(V(x^*, x^k) - V(x^*, x^{k+1})\right)$$

$$\frac{1}{K}\sum_{k=1}^{K} \qquad \sum V^k - V^{k+1} = \ldots + V^i - \cancel{V^{i+1}} + \cancel{V^{i+1}} - V^{i+2}$$

$$\frac{1}{K}\sum_{k=1}^{K} f(x^k) - f(x^*) \leq \frac{V(x^*, x^0) - \cancel{V(x, x^K)}}{K}$$

$$\leq \frac{V(x^*, x^0)}{K}$$

Йенсен

$$f\left(\frac{1}{K}\sum_{k=1}^{K} x^k\right) - f(x^*) \leq \frac{L\, V(x^*, x^0)}{K}$$

$$\gamma = \frac{1}{L}$$

Сходимость зерк. спуска для L-гладких, выпуклых задач

$$\boxed{f\left(\frac{1}{K}\sum_{k=1}^{K} x^k\right) - f(x^*) \leq \frac{L\, V(x^*, x^0)}{K}}$$

сублинейная сходимость $\frac{1}{K}$, как и у град. спуска

только у град спуске L_2, $\|x^0 - x^*\|_2^2$

⊕ $L \leq L_2$

$$\|\nabla f(x) - \nabla f(y)\|_q \leq L\|x-y\|_p \qquad \frac{1}{p} + \frac{1}{q} = 1$$

$p \in [1; 2]$ $q \in [2; +\infty)$

- $p = 1$ $\quad \|\ \|_2 \le \|\ \|_1 \qquad q = \infty \qquad \|\ \|_\infty \le \|\ \|_2$

$$\begin{array}{ccc} \|\ \|_2 & \le & L_2 \|\ \|_2 \\ \rotatebox{90}{$\le$} & & \rotatebox{-90}{$\le$} \\ \|\ \|_\infty & \le & L \|\ \|_1 \end{array}$$

⊖ $V(x,y) \geq \frac{1}{2}\|x-y\|^2_{p=1} \geq \frac{1}{2}\|x-y\|_2^2 \qquad p \in [1;2]$

Зеркальный спуск на симплексе

$$x^{k+1} = \underset{x \in \overline{X}}{\operatorname{argmin}} \left\{ \gamma \langle \nabla f(x^k); x \rangle + V(x, x^k) \right\}$$

$$V(x,y) = \sum_{i=1}^{d} x_i \log\left(\frac{x_i}{y_i}\right) \qquad \overline{X} = \Delta$$

$$\boxed{x_i^{k+1} = \frac{x_i^k \exp\left(-\gamma [\nabla f(x^k)]_i\right)}{\sum_{j=1}^{d} x_j^k \exp\left(-\gamma [\nabla f(x^k)]_j\right)}}$$

подробнее в семинаре по ККТ

⊕ $V(x,y) \approx \log d \qquad \frac{1}{2}\|x-y\|_2^2 \sim 1$

но на L можно выиграть в d раз