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

$x^k \in (E, \|\cdot\|)$ $\qquad \nabla f(x^k) \overset{?}{\in} (E, \|\cdot\|)$

$\qquad\qquad\qquad\qquad \in (E^*, \|\cdot\|_*)$

до этого: $\|\cdot\| = \|\cdot\|_2$, а двойств. $\|\cdot\|_* = \|\cdot\|_2$

Идея (А. Немировский, Д. Юдин)

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

зеркальный спуск

φ действ. из E в E^*

φ^{-1} действ. из E^* в E

Определение μ-сильной выпуклости

Пусть дана непрерывно дифференцируемая на выпуклом множестве $\mathcal{X}$ функция $d : \mathcal{X} \to \mathbb{R}$. Будем говорить, что она является μ-сильно выпуклой ($\mu > 0$) относительно нормы $\|\cdot\|$ на множестве $\mathcal{X}$, если для любых $x, y \in \mathcal{X}$ выполнено

$$d(x) \geq d(y) + \langle \nabla d(y), x - y \rangle + \frac{\mu}{2}\|x - y\|^2.$$

произв. норма

Определение

Пусть дана дифференцируемая 1-сильно выпуклая относительно нормы $\|\cdot\|$ на множестве $\mathcal{X}$ функция d. Дивергенцией Брэгмана, порожденной функцией d на множестве $\mathcal{X}$, называется функция двух аргументов $V(x, y) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ такая, что для любых $x, y \in \mathcal{X}$

$$V(x, y) = d(x) - d(y) - \langle \nabla d(y), x - y \rangle.$$

← новый измеритель расст.

Пример:

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

$$= \frac{1}{2}\|x-y\|_2^2$$

- $d(x) = \sum_{i=1}^{d} x_i \log x_i$ — сильно-вып. отн $\|\cdot\|_1$ — $\Delta = \{ x \in \mathbb{R}^d \mid x_i \geq 0, \sum x_i = 1 \}$

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

- $d(X) = \mathrm{trace}(X \log X)$ — квантовая дивергенция

  $V(X,Y) = \mathrm{trace}(X \log X - X \log Y - X + Y)$

Свойства:

- асимметричность
- сильная выпуклость: $V(x,y) \geq \frac{1}{2}\|x-y\|_2^2 > 0$
- невыпуклая по второму аргументу

Равенство параллелограмма/теорема Пифагора

Для любых точек $x, y, z \in \mathcal{X}$ следует что

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

Док-во:

$$V(z,x) + V(x,y) \overset{\text{по опред. д. Бр.}}{=} d(z) - \cancel{d(x)} - \langle \nabla d(x); z-x \rangle + \cancel{d(x)} - d(y) - \langle \nabla d(y); x-y \rangle$$

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

$$\min_{x \in X} f(x)$$

выпукла

вып.

Метод зеркального спуска:

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

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

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

$-x^k$

$+\frac{\gamma^2}{2}\|\nabla f(x^k)\|_2^2$

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

$$y^k = x^k - \gamma \nabla f(x^k)$$

град. спуск с проекцией

- $X = \mathbb{R}^d$

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} \underbrace{\left\{ \gamma \langle \nabla f(x^k); x \rangle + d(x) - d(x^k) - \langle \nabla d(x^k); x - x^k \rangle \right\}}_{\text{услов оптим. } \nabla = 0}$$

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

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

Определение L-гладкой функции

Пусть дана непрерывно дифференцируемая на $\mathcal{X}$ функция $f : \mathcal{X} \to \mathbb{R}$. Будем говорить, что данная функция имеет L-Липшицев градиент (говорить, что она является L-гладкой) относительно нормы $\|\cdot\|$ на $\mathcal{X}$, если для любых $x, y \in \mathcal{X}$ выполнено

$$\|\nabla f(x) - \nabla f(y)\|_* \le L\|x - y\|.$$

Св-ва:

разница с ℓ_2

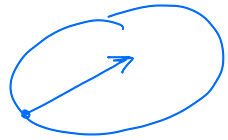
$$|f(y) - f(x) - \langle \nabla f(x); y - x \rangle| \le \frac{L}{2}\|x - y\|^2$$

Док-во сходимости:

f – выпуклая $\quad X$ – выпуклое

f – L-гладкая в $\|\cdot\|$

$$\arg\min_{x \in X} \left\{ \langle \gamma \nabla f(x^k); x \rangle + d(x) - \cancel{d(x^k)} - \langle \nabla d(x^k); x - \cancel{x^k} \rangle \right\}$$



усл. оптим. для вып. зад. на в. мн.-ве

$$\langle \gamma \nabla f(x^k) + \nabla d(\overset{x^*}{x^{k+1}}) - \nabla d(x^k); x^{k+1} - x \rangle \le 0 \quad \forall x \in \mathcal{X}$$

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

т. Пифагора для д. Брэгмана

Равенство параллелограмма/теорема Пифагора

Для любых точек $x, y, z \in \mathcal{X}$ следует что

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

$y = x^k, \quad x = x^{k+1}, \quad z = x, \quad x = x^{k+1}$

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

$+$

L-гладкость:

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

$y = x^{k+1}, \quad x = x^k$

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

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

выпуклость f: $\quad f(x^{(k)}) - f(x) \le \langle \nabla f(x^{(k)}); x^{k} - x \rangle$

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

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

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

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

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

$$\gamma\left(f(x^{(k+1)}) - f(x)\right) \le V(x, x^{k}) - V(x, x^{(k+1)}) \quad \Big| : \gamma$$

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

$$\frac{1}{K}\sum f(x^{(k+1)}) - \underset{x^*}{f(x)} \le \frac{V(\underset{x^*}{x}; x^{0})}{\gamma K}$$

Теорема сходимость зеркального спуска для L-гладких относительно $\|\cdot\|$ и выпуклых функций

Пусть задача оптимизации на выпуклом множестве $\mathcal{X}$ с L-гладкой относительно нормы $\|\cdot\|$, выпуклой целевой функцией f решается с помощью зеркального спуска с шагом $\gamma \le \frac{1}{L}$. Тогда справедлива следующая оценка сходимости

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

аналогично град. спуску для выпуклой задачи

для зерк. спуска $\gamma = \frac{1}{L}$ $\qquad \le \frac{L V(x^*; x^0)}{K}$

для град. спуска $\gamma \le \frac{1}{L_2}$ $\qquad \le \frac{L_2 \|x^* - x^0\|_2^2}{2K}$

L или L_2 ?

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

$$\|\quad\|_2 \le L_2 \|\quad\|_2$$

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

$$\|\quad\|_q \le \|\quad\|_2 \qquad \|\quad\|_2 \le \|\quad\|_p$$

$L \le L_2$ зерк. спуск лучше

но $\frac{1}{2}\|x^0 - x^*\|_p^2 \le V(x^*, x^0)$ $\qquad p \in [1;2]$

$\geq$

$\frac{1}{2}\|x^0 - x^*\|_2^2$ $\qquad$ зерк. спуск. хуже

Зерк. спуск для симплекса

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

$$V(x,y) = \sum_{i=1}^{d} x_i \log \frac{x_i}{y_i} \quad \text{на } \triangle$$

$$\triangle = \{ x \in \mathbb{R}^d \mid x_i \ge 0,\ \sum x_i = 1 \}$$

$\Downarrow$

$$\min_{x \in \mathbb{R}^d} \langle \gamma \nabla f(x^k); x \rangle + \sum x_i \log \frac{x_i}{x_i^k}$$

s.t. $-x_i \le 0$

$\sum x_i - 1 = 0$

Лагранжиан $L(x, \lambda, \nu) =$

$$\langle \gamma \nabla f(x^k), x \rangle + \sum x_i \log \frac{x_i}{x_i^k} + \sum \lambda_i(-x_i) + \nu\left(\sum x_i - 1\right)$$

$$= \sum_{i=1}^{d} \left(\gamma\, x_i [\nabla f(x^k)]_i + x_i \log \frac{x_i}{x_i^k} - \lambda_i x_i + \nu x_i \right) - \nu$$

Двойств.:

$$L'_{x_i} = \gamma [\nabla f(x^k)]_i + \log \frac{x_i}{x_i^k} + 1 - \lambda_i + \nu = 0$$

$$x_i = x_i^k \exp\left(-1 + \lambda_i - \nu - \gamma [\nabla f(x^k)]_i\right)$$

$$\inf_x L(x, \lambda, \nu) = \sum_{i=1}^{d} \left(- x_i^k \exp\left(-1 + \lambda_i - \gamma [\nabla f(x^k)]_i - \nu\right)\right) - \nu$$

$$\max_{\lambda_i \geq 0,\, \nu \in \mathbb{R}} \left[\sum_{i=1}^{d} \left(- x_i^k \exp\left(-1 + \underbrace{\lambda_i}_{\lambda_i^* = 0} - \gamma [\nabla f(x^k)]_i - \nu\right)\right) - \nu \right]$$

$\lambda_i^* = 0$ KKT :

$$x_i^* = x_i^k \exp\left(-1 + \cancel{\lambda_i} - \nu^* - \gamma [\nabla f(x^k)]_i\right)$$

$\|$

x_i^{k+1}

$$x_i^{k+1} = x_i^k \exp\left(-\gamma [\nabla f(x^k)]_i\right) \cdot \exp(-1 - \nu^*)$$

$\sum x_i^{k+1} = 1$ $\exp(-1-\nu^*)$ – нормировка

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

для евклидово: $\frac{1}{2}\|x-y\|_2^2 = 1$) хуже з. с.

дивергенция Брэгмана: $\log d$

с точки зрения L: $\| \ \|_\infty \le L \| \ \|_1$

$$\| \ \|_2 \le L_2 \| \ \|_2$$

$$L \sim \frac{L_2}{d}$$

в лучшем случае з. с. даёт улучшение: $\frac{d}{\log d}$ раз.