Градиентный спуск:

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

$$x \in (E, \|\cdot\|_*)$$

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

NB $\|\cdot\|_p$ – норм $\|\cdot\|_q$ – сопряж. $\frac{1}{p} + \frac{1}{q} = 1$

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

- Модификация (А. Немировский, Д. Юдин)

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

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

Шаг градиентного спуска в „зеркальном" пр-ве

Опр. (сильная выпуклость по произв. норме)

$d: X \to \mathbb{R}$ является сильно выпуклой с $\mu > 0$, если $\forall x, y \in X \Leftrightarrow d(x) \geq d(y) + \langle \nabla d(y); x - y \rangle + \frac{\mu}{2}\|x-y\|_*^2$

произв. норма

Опр. (Дивергенция Брэгмана)

1-сильно выпуклую по $\|\cdot\|$ функцию d, зад. на X. Дивергенция Брэгмана, порожд. ф. d есть функция двух аргументов $V(x,y): X \times X \to \mathbb{R}$

$\forall\, x, y \in X \qquad V(x,y) = d(x) - d(y) - \langle \nabla d(y); x - y \rangle$

Примеры:

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

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

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

$$V(x,y) = \sum_{i=1}^{d} x_i \log \frac{x_i}{y_i}$$ ← KL-дивергенция (измерение расст. между распред.)

- $d(X) = \operatorname{tr}(X \log X)$ X - матрица

$$V(X,Y) = \operatorname{tr}(X \log X - X \log Y - X + Y)$$
(квантовая дивергенция фон Неймана)

Свойства:

- ассиметричные (см. KL-дивергенцию)
- сильная выпуклость (по определению

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

- неотрицательность
- невыпукла по 2 аргументу

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

f — выпуклая

$\overline{X}$ — выпуклое

Метод зеркального спуска

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

Примеры

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

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

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

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

(градиентный спуск)

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

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

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

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

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

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

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

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

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

$$\gamma \nabla f(x^k) + \nabla d(x) - \nabla d(x^k) = 0$$

$$x^{k+1}$$

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

$\varphi \leftarrow$ ([illegible], [illegible])

- на симплексе

$$d(x) = \sum_{i=1}^{d} x_i \log x_i \qquad \mathcal{X} = \triangle$$

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

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

Формальная запись:

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

$$\text{s.t.} \quad -x_i \leq 0$$

$$\sum_{i=1}^{d} x_i - 1 = 0$$

Лагранжиан:

$$L(x, \lambda, \nu) = \gamma \langle \nabla f(x^k); x \rangle + V(x, x^k) + \sum_{i=1}^{d} \lambda_i (-x_i) + \nu \left(\sum_{i=1}^{d} x_i - 1 \right)$$

по компонентам

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

оптимизируем по x_i для получения двойств.

$$\left(\left(a + \log \frac{x_i}{b} \right) x_i \right)' = 0$$

$$a + \log \frac{x_i^*}{b} + 1 = 0$$

$$x_i^* = b \exp(-a - 1)$$

$$\inf_x L(x, \lambda, \nu) = \sum_{i=1}^{d} \left(a_i + \log \frac{x_i}{b_i} \right) x_i - \nu$$

$$= \sum_{i=1}^{d} \left(\cancel{a_i} - \cancel{a_i} - 1 \right) b_i \exp(-a_i - 1) - \nu$$

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

Двойств. задача

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

$$\boxed{\lambda_i^* = 0}$$

KKT: $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

$$\nabla_x \left(\sum_{i=1}^{d} \left(\gamma [\nabla f(x^k)]_i + \log \frac{x_i^*}{x_i^k} - \underbrace{\lambda_i^*}_{0} + \nu^* \right) x_i^* - \nu^* \right) = 0$$

$$\nabla_{x_i} \left[\left(\gamma [\nabla f(x^k)]_i + \log \frac{x_i^*}{x_i^k} - \cancel{\lambda_i^*} + \nu^* \right) x_i^* \right] = 0$$

$$\gamma [\nabla f(x^k)]_i + \nu^* + \log \frac{x_i^*}{x_i^k} + 1 = 0$$

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

?

$\sum x_i = 1$ $\qquad \exp(1+\nu^*)$ – нормировка

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