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

$x \in (E, \|\cdot\|)$ $\qquad$ $\nabla f(x) \in (E^*; \|\cdot\|_*)$

$\| \; \|_2$ $\qquad$ $\| \; \|_* = \| \; \|_2$

А. Немировский, Д. Юдин

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

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

град. спуск. в „зеркальном" пр-ве

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

не обяз. 2

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

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 $x, y \in \mathcal{X}$

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

Примеры:

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

$$\frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle y; x - y \rangle$$

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

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

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

1-сильн. вып. в $\| \ \|_1$

- $d(x) = \sum_{i=1}^{d} x_i \log x_i$ на вероятн. симплексе

$$\Delta = \{ x \in \mathbb{R}^d \mid x_i \geq 0 \ \ \sum x_i = 1\}$$

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

KL (дивергенция Кульбака – Лейблера)

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

матрица $d \times d$

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

квантовая дивер. фон Неймана

- $d(X) = -\log\det X$

$$V(X,Y) = tr(XY^{-1} - I) - \log\det(XY^{-1})$$

Св-ва дивер. Брэгмана

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

Свойство дивергенции Брэгмана

Для любых точек $x, y \in \mathcal{X}$ следует что $V(x,y) \geq \frac{1}{2}\|x-y\|^2$.

3)

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

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

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

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

Метод зеркального спуска

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

Примеры:

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

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

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

  $$\gamma \langle \nabla f(x^k); x - x^k \rangle + \cancel{\frac{\gamma^2}{2}} \|\nabla f(x^k)\|_2^2 + \frac{1}{2}\|x - x^k\|_2^2$$

  $$= \frac{1}{2}\|x - x^k + \gamma \nabla f(x^k)\|_2^2$$

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

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

  $\mathcal{X} \neq \mathbb{R}^d$ выпуклое

  $\operatorname{argmin}_{x \in \mathcal{X}} \frac{1}{2}\|x - x^k + \gamma \nabla f(x^k)\|_2^2$

  град. спуск с проекцией

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

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

$$V(x, x^k) = d(x) - d(y) - \langle \nabla d(y); x - y\rangle = d(x) - d(x^k) - \langle \nabla d(x^k); x - x^k\rangle$$

x^{k+1} $\nabla = 0$

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

$$\left[\begin{array}{l} \nabla d(x^{k+1}) = \nabla d(x^k) - \gamma \nabla f(x^k) \\ \varphi(x^{k+1}) = \varphi(x^k) - \gamma \nabla f(x^k) \end{array}\right.$$

Идея: решать argmin аналитически

Док-во сходимости

- f – L-гладкой
- f – выпуклой

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

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

Теорема (свойство L - гладкой функции)

Пусть дана L - гладкая относительно нормы $\|\cdot\|$ функция $f : \mathcal{X} \to \mathbb{R}$. Тогда для любых $x, y \in \mathcal{X}$ выполнено

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

КБШ : раньше $\langle a; b\rangle \le \|a\|_2 \cdot \|b\|_2$

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

$\|a\|_*$: $\nabla f - \nabla f$ $\quad$ $\|b\|$: $x - y$

Док-во:

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

$$V(x, x^k) = d(x) - d(x^k) - \langle \nabla d(x^k); x - x^k\rangle$$

$$\min_{x \in \mathcal{X}} g(x)$$

$$\langle \nabla g(x^*); x - x^*\rangle \ge 0 \quad \forall x \in \mathcal{X}$$

$x^* = x^{k+1}$

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

- по перавен. $x = x^{k+1}$, $y = x^k$ $z = x$

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

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

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

раньше x^k $\quad$ x^k

Домножим

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

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

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

вычитаем

$$\gamma\left(f(x^{(k+1)}) - \cancel{f(x^{(k)})}\right) + \gamma\left(\cancel{f(x^{(k)})} - f(x)\right)$$

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

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

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

$$\le V(x; x^{(k)}) - V(x; x^{(k+1)}) - \frac{1}{2}\underbrace{(1 - \gamma L)}_{\ge 0}\|x^{(k+1)} - x^{(k)}\|^2$$

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

$$\le V(x; x^{(k)}) - V(x; x^{(k+1)})$$

$\frac{1}{K}\sum$

$$\gamma \cdot \frac{1}{K}\sum_{k=0}^{K-1} f(x^{k+1}) - f(x) \le \frac{V(x;x^0) - \cancel{V(x;x^K)}}{K} \Big| \gamma$$

Иенсен

$$f\left(\frac{1}{K}\sum x^k\right) - f(x) \le \frac{V(x;x^0)}{K}$$

$$x = x^*$$

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

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

$\gamma = \frac{1}{L}$ $\qquad \frac{L\,V(x^*;x^0)}{K}$

для град. спуск $\qquad \frac{L_2 \|x^0 - x^*\|_2^2}{2K}$

- L_2 vs L?

$\|\ \|. = \|\ \|_p \qquad p \in [1;2]$

$\|\ \|_* = \|\ \|_q \qquad \frac{1}{p} + \frac{1}{q} = 1 \qquad p = 1 \quad q = \infty$

$p = 2 \quad q = 2$

$p = 1 \qquad q = \infty$

$\|\ \|_1 \qquad \|\ \|_\infty$

$\|\ \|_2 \le \|\ \|_1 \qquad\qquad \|\ \|_\infty \le \|\ \|_2$

$$\begin{array}{ccccc} \| & \cdot & \|_\infty \le L & \| & \|_1 \\ & \wedge & & & \vee \\ \| & \cdot & \|_2 \le L_2 & \| & \|_2 \end{array} \Rightarrow L \le L_2$$

- $V(x,y)$ vs $\frac{1}{2}\|x-y\|_2^2$

$$V(x,y) \ge \frac{1}{2}\|x-y\|_p^2 \ge \frac{1}{2}\|x-y\|_2^2$$

$p \in [1;2]$

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

$$X = \Delta = \{x \in \mathbb{R}^d \mid x_i \ge 0 \quad \sum x_i = 1\}$$

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

$$\min \{\gamma \langle \nabla f(x^k); x \rangle + \sum x_i \log \frac{x_i}{x_i^k}\}$$

s.t. $x_i \ge 0 \rightarrow -x_i \le 0$

$\sum x_i = 1 \rightarrow \sum x_i - 1 = 0$

Лагранжиан

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

по коорд

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

Вернемся к двойственной $\Rightarrow \inf\limits_{x} L$

произв. по x_i

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

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

$$\max_{\substack{\lambda \geq 0 \\ \nu \in \mathbb{R}}} \left[g(\lambda, \nu) = L(x_i^*, \lambda, \nu) \right]$$

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

$$\lambda_i^* = 0$$

Вернемся к град. по x от L

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

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

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

$$x_i^{k+1} = \underbrace{x_i^k \exp\left(-\gamma\,[\nabla f(x^k)]_i\right)}_{+} \cdot \underbrace{\exp(-\nu - 1)}_{\text{нормировка?}}$$

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

Ремарка в случае симплекса

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

а $L_{1,\infty}$ лучше L_2 в худшем случае в d

Выигрыш в $\frac{d}{\log d}$ раз.