ЛЕКЦИЯ № 13

1. Градиентные методы (продолжение)

2. Метод Ньютона (метод второго порядка)

Определение Дифференцируемая функция f называется сильно выпуклой (с константой $l > 0$), если для любых x и y из R^n справедливо

$$f(x+y) \geq f(x) + \langle f'(x), y \rangle + l\|y\|^2/2. \tag{37}$$

Лемма 13 Если функция f является сильно выпуклой (с константой $l > 0$), то она имеет глобальный минимум на R^n.

Доказательство. Перепишем (37), используя неравенство Коши - Буняковского

$$f(x+y) \geq f(x) - \|f'(x)\|\|y\| + l\|y\|^2/2.$$

Вынесем величину $l\|y\|/2$ за скобки

$$f(x+y) \geq f(x) + l\|y\|/2(\|y\| - 2\|f'(x)\|/l).$$

$\Longrightarrow \forall y : \|y\| > r = 2\|f'(x)\|/l$ имеем неравенство $f(x+y) > f(x)$, из которого следует, что минимум $\exists$ и достигается на шаре $B(x,r)$. ■

Лемма 14 Если функция $\boldsymbol{f}$ является сильно выпуклой (с константой $\boldsymbol{l > 0}$) и $\boldsymbol{x^*}$ — ее глобальный минимум, то для любого $\boldsymbol{x \in R^n}$ выполняется неравенство $\boldsymbol{\|f'(x)\|^2 \geq 2l(f(x) - f(x^*))}$.

Доказательство. Подставим в неравенство (37) вместо $\boldsymbol{y}$ вектор $\boldsymbol{x^* - x}$. Получим

$$\boldsymbol{f(x) - f(x^*) + \langle f'(x), x^* - x\rangle +}$$
$$\boldsymbol{+l\|x^* - x\|^2/2 \leq 0.}$$

Градиентные методы

По определению скалярного произведения имеем

$$\langle f'(x)/\sqrt{2l} + \sqrt{l/2}(x^* - x), f'(x)/\sqrt{2l} + \\ + \sqrt{l/2}(x^* - x)\rangle = \\ \|f'(x)/\sqrt{2l} + \sqrt{l/2}(x^* - x)\|^2 \geq 0.$$

Распишем скалярное произведение и получим

$$\|f'(x)\|^2/2l + \langle f'(x), x^* - x\rangle + l\|x^* - x)\|^2/2 \geq \\ \geq 0 \geq f(x) - f(x^*) + \langle f'(x), x^* - x\rangle + \\ + l\|x^* - x\|^2/2.$$

После приведения подобных членов получим требуемое неравенство:

$$\|f'(x)\|^2 \geq 2l(f(x) - f(x^*)).\blacksquare$$

Теорема 20 (Вторая теорема сходимости) Пусть функция f дифференцируема в R^n, является сильно выпуклой, выполняется условие Липшица для градиента $f'(x) : \|f'(x) - f'(y)\| \leq L\,\|x - y\|$ и длина шага α удовлетворяет условию $0 < \alpha < 2/L$.

Тогда $x^k \to x^*$ при $k \to \infty$ и $\|x^k - x^*\| \leq Cq^k,\ 0 \leq q < 1$.

Доказательство. У функции f $\exists$ глобальный минимум x^* (лемма 13). При доказательстве теоремы 19 получили:

$$f(x^{k+1}) \leq f(x^k) - \alpha(1 - L\alpha/2)\|f'(x^k)\|^2$$

отсюда, учитывая неравенство

$\|f'(x^k)\|^2 \geq 2l(f(x^k) - f(x^*))$ (лемма 14)

получим

$$f(x^{k+1}) \leq f(x^k) - l\alpha(2 - L\alpha)(f(x^k) - f(x^*))$$

или,

вычитая $f(x^*)$ из обеих частей, имеем:

$$f(x^{k+1}) - f(x^*) \leq (1 - l\alpha(2 - L\alpha))\times$$
$$\times(f(x^k) - f(x^*)) \qquad (38)$$

Положим $q_1 = 1 - l\alpha(2 - L\alpha)$. Из (38) $\Longrightarrow$

$$f(x^{k+1}) - f(x^*) \leq q_1^{k+1}(f(x^0) - f(x^*)) \qquad (39)$$

Функция $f \neq const$ ((от противного) Пусть $f = const$.

Тогда $f'(x) = 0$ и из неравенства $f(x+y) \geq f(x) + \langle f'(x), y \rangle + l\|y\|^2/2$ имеем $l\|y\|^2/2 \leq 0 \Rightarrow$ противоречие с условием $l > 0$)

$\Longrightarrow$ найдётся начальная точка x^0 такая, что

$$f(x^0) > f(x^*).$$

Тогда при $k = 0$ имеем (неравенство (38))

$$0 \leq f(x^1) - f(x^*) \leq q_1(f(x^0) - f(x^*)).$$

Следовательно $q_1 \geq 0$. Так как $q_1 < 1$, то из (39) следует, что

$$f(x^k) \to f(x^*).$$

Подставим в неравенство

$$f(x+y) \geq f(x) + \langle f'(x), y \rangle + l\|y\|^2/2$$

вместо y вектор $(x^k - x^*)$ и x^* вместо x.

Учитывая, что $\boldsymbol{x^*}$ – глобальный минимум, и, следовательно, $\boldsymbol{f'(x^*) = 0,}$ получим

$$(f(x^k) - f(x^*)) \geq l\|x^k - x^*\|^2/2.$$

Из этого неравенства и (39) имеем:

$$\|x^k - x^*\|^2 \leq 2q_1^k\,(f(x^0) - f(x^*))/l.$$

Перепишем это неравенство в следующем виде

$\boldsymbol{\|x^k - x^*\| \leq Cq^k,}$ где

$$C = \sqrt{2(f(x^0) - f(x^*))/l}, q = \sqrt{q_1}.$$

Таким образом метод имеет линейную оценку скорости сходимости. Из этого неравенства также получим, что $\boldsymbol{x}^{\boldsymbol{k}} \longrightarrow \boldsymbol{x}^*$ при $\boldsymbol{k} \longrightarrow \infty$. $\blacksquare$

Метод Ньютона

Пусть $\boldsymbol{f}$ — дважды непрерывно дифференцируемая функция и есть алгоритм вычисления её вторых производных.

Идея метода: заменить функцию $\boldsymbol{f}$ в окрестности текущего приближения $\boldsymbol{x^k}$ её квадратичной аппроксимацией: $\boldsymbol{q(x) = f(x^k) + f'(x^k)(x - x^k) + \frac{1}{2}(x - x^k)^\top f''(x^k)(x - x^k)}$.

Выбрать в качестве нового приближения $\boldsymbol{x^{k+1}}$ точку минимума функции $\boldsymbol{q(x)}$ (если она $\exists$).

Метод Ньютона

Предположим, что матрица $f''(x^k)$ — положительно определённая $\Longrightarrow$ функция $q(x)$ — сильно выпукла $\Longrightarrow$ у неё единственный минимум $\Longrightarrow$ его можно найти как решение системы уравнений $q'(x^{k+1}) = 0$, которая по определению $q(x)$ эквивалентна следующей системе линейных уравнений

$$f'(x^k) = -f''(x^k)(x^{k+1} - x^k).$$

Метод Ньютона

$\Longrightarrow$ получаем необходимую итерационную формулу

$$\boldsymbol{x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k).}$$

Заметим, что здесь как направление, так и длина шага заданы.

Лемма 15 Пусть f — дважды непрерывно дифференцируемая функция. Если f — сильно выпуклая функция с константой l, то выполняется следующее неравенство:

$$\|[f''(x)]^{-1}\| \leq l^{-1}.$$

Доказательство. Воспользуемся следующей формулой для конечных приращений функции f:

$$f(x+y) - f(x) = \langle f'(x), y \rangle +$$

$$+\frac{\langle f''(x+\tau_2 y)y, y\rangle}{2},$$

где $0 \le \tau_2 \le 1.$ Из неравенства (37) (определения сильной выпуклости) $\Longrightarrow$

$$\langle f''(x+\tau_2 y)y, y\rangle/2 = f(x+y) - f(x) -$$
$$-\langle f'(x), y\rangle \ge l\|y\|^2/2.$$

Заменяя y на ty, получим:

$$\langle f''(x+\tau_2 ty)ty, ty\rangle \ge l\|ty\|^2.$$

Следовательно,

$$t^2\langle f''(x+\tau_2 ty)y, y\rangle \geq t^2 l\|y\|^2.$$

Поделив на t^2 и устремляя t к нулю, будем иметь

$$\langle f''(x)y, y\rangle \geq l\|y\|^2.$$

Пусть $y = (f''(x))^{-1}z \Longrightarrow$

$$\langle z, (f''(x))^{-1}z\rangle \geq l\|(f''(x))^{-1}z\|^2$$

$\Longrightarrow$ (используя неравенство Коши-Буняковского)

Метод Ньютона

$$\|z\|\|(f''(x))^{-1}z\| \geq l\|(f''(x))^{-1}z\|^2 \Longrightarrow$$

$$l\|(f''(x))^{-1}z\| \leq \|z\| \; \forall z.$$

Это означает, что $\|[f''(x)]^{-1}\| \leq l^{-1}$. ■

Пусть последовательность $\{x^k\}$ получена с помощью метода Ньютона и точка x^* — глобальный минимум функции f.

Теорема 21 Пусть дважды непрерывно дифференцируемая функция f сильно выпукла (с константой $l > 0$), вторая производная удовлетворяет условию Липшица: для любых $x, y \in R^n$

$$\|f''(x) - f''(y)\| \le L\,\|x - y\|,$$

и $q = L\|f'(x^0)\|/2l^2 < 1$. Тогда $x^k \to x^*$ при $k \to \infty$ и метод Ньютона имеет квадратичную скорость сходимости

$$\|x^k - x^*\| \le (2l/L)q^{2^k}.$$

Доказательство. Воспользуемся следующей формулой конечных приращений:

$$\boldsymbol{g(x+y) = g(x)+\langle g'(x), y\rangle+\int\limits_0^1 \langle g'(x+\tau y)-}$$

$$\boldsymbol{-g'(x), y\rangle d\tau.}$$

Подставим вместо $\boldsymbol{g}$ производную функции $\boldsymbol{f}$:

Метод Ньютона

$$f'(x+y) = f'(x) + \langle f''(x), y\rangle + \int\limits_0^1 \langle f''(x+\tau y) -$$

$$- f''(x), y\rangle d\tau.$$

В этой формуле использованы следующие соглашения:

$$\langle f''(x), y\rangle = \Big(\langle (f'(x))'_{x_1}, y\rangle, \ldots,$$

$$\langle (f'(x))'_{x_n}, y\rangle\Big),$$

Метод Ньютона

$\int\limits_0^1 \langle f''(x+\tau y) - f''(x), y\rangle d\tau = \Big(\int\limits_0^1 \langle (f'(x+\tau y) - f'(x))'_{x_1}, y\rangle d\tau, \ldots, \int\limits_0^1 \langle (f'(x+\tau y) - f'(x))'_{x_n}, y\rangle d\tau\Big)$. Перепишем формулу $f'(x+y) - f'(x) - \langle f''(x), y\rangle = \int\limits_0^1 \langle f''(x+\tau y) - f''(x), y\rangle d\tau$. Вектор слева обозначим через F, вектор справа через $\boldsymbol{Z}$, вектор под интегралом через $\boldsymbol{Z(\tau) = (Z_1(\tau), \ldots, Z_n(\tau))}$. Понятно, что $\boldsymbol{Z = \int\limits_0^1 Z(\tau) d\tau}$. Очевидно, что $\boldsymbol{\|F\| = \|Z\|}$. Оценим норму $\boldsymbol{Z}$.

$$\boldsymbol{\|Z\|^2 = \sum_j Z_j^2 = \sum_j Z_j \int\limits_0^1 Z_j(\tau) d\tau =}$$

Метод Ньютона

$$= \int\limits_0^1 (\sum_j Z_j Z_j(\tau))d\tau \leq$$

$$\leq \int\limits_0^1 \|Z\|\|Z(\tau)\|d\tau = \|Z\| \int\limits_0^1 \|Z(\tau)\|d\tau \leq$$

$$\leq \|Z\| \int\limits_0^1 \|f''(x+\tau y) - f''(x)\|d\tau \leq \|Z\|L\|y\|^2/2$$

Метод Ньютона

$$\Longrightarrow$$

$$\|F\| = \|Z\| \le L\|y\|^2/2$$

$$\Longrightarrow$$

$$\|f'(x+y) - f'(x) - \langle f''(x), y\rangle\| \le L\|y\|^2/2.$$

Подставим $x = x^k$ и $y = -[f''(x^k)]^{-1}f'(x^k)$

$$\|f'(x^{k+1})\| \le (L/2)\|[f''(x^k)]^{-1}\|^2\|f'(x^k)\|^2.$$

Метод Ньютона

Применяя лемму 15, получим

$$\|f'(x^{k+1})\| \leq (L/2l^2)\|f'(x^k)\|^2.$$

Итерируем это неравенство по k

$$\|f'(x^{k+1})\| \leq (2l^2/L)\,(\underbrace{L\|f'(x^0)\|/2l^2}_{q})^{2^{k+1}}.$$

f — сильно выпуклая функция $\Longrightarrow$

$$f(x+y) \geq f(x) + \langle f'(x), y\rangle + l\|y\|^2/2.$$

y заменим на $y - x$

$$\Longrightarrow$$

$$f(y) \geq f(x) + \langle f'(x), y - x \rangle + l\|y - x\|^2/2.$$

меняем x на y, а y на $x \Longrightarrow$

$$f(x) \geq f(y) + \langle f'(y), x - y \rangle + l\|x - y\|^2/2.$$

Сложим оба неравенства:

$$\langle f'(x) - f'(y), x - y \rangle \geq l\|x - y\|^2.$$

Подставим $y = x^*, x = x^{k+1}$, и учитывая равенство $f'(x^*) = 0$, получим

$$l\|x^{k+1} - x^*\|^2 \leq \langle f'(x^{k+1}), x^* - x^{k+1}\rangle \leq$$
$$\leq \|f'(x^{k+1})\| \; \|x^* - x^{k+1}\| \leq$$
$$\leq (2l^2/L)q^{2^{k+1}}\|x^* - x^{k+1}\| \Longrightarrow$$
$$\|x^{k+1} - x^*\| \leq (2l/L)q^{2^{k+1}} \blacksquare$$