ЧАСТЬ I

БЕЗУСЛОВНАЯ МИНИМИЗАЦИЯ

Глава 1

ОСНОВЫ ТЕОРИИ И МЕТОДОВ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ

Мы начинаем изучение проблем оптимизации с классической задачи безусловной минимизации гладкой функции:

$$\min f(x), \quad x \in \mathbf{R}^n.$$

Этой задаче будет уделено большое внимание не только из-за ее важности, но и потому, что в силу ее простоты для нее наиболее четко видна схема математического исследования общих оптимизационных задач и идейные основы методов оптимизации.

§ 1. Сведения из математического анализа

1. Дифференцирование скалярных функций. Скалярная функция $f(x)$ n-мерного аргумента x (кратко это записывается $f\colon \mathbf{R}^n \to \mathbf{R}^1$) называется *дифференцируемой в точке* x, если найдется вектор $a \in \mathbf{R}^n$ такой, что для всех $y \in \mathbf{R}^n$

$$f(x+y) = f(x) + (a, y) + o(y). \tag{1}$$

Вектор a в (1) называется *производной* или *градиентом* функции $f(x)$ в точке x и обозначается $f'(x)$ или $\nabla f(x)$. Итак, градиент определяется равенством

$$f(x+y) = f(x) + (\nabla f(x), y) + o(y). \tag{2}$$

Иначе можно сказать, что функция дифференцируема в точке x, если она *допускает линейную аппроксимацию первого порядка* в этой точке, т. е. найдется линейная функция $\bar{f}(y) = f(x) + (\nabla f(x), y)$ такая, что $|f(x+y) - \bar{f}(y)| = o(y)$. Ясно, что градиент определяется однозначно, при этом $\nabla f(x)$ — вектор с компонентами $(\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_n)$. Вычислять градиент можно, во-первых, непосредственно из определения, во-вторых, с помощью его координатной записи и, в-третьих, с помощью правила дифференцирования сложной функции (см. ниже (12)).

Пусть, например, $f(x)$ — квадратичная функция

$$f(x) = (Ax, x)/2 - (b, x),$$

где A — симметричная $n \times n$-матрица, $b \in \mathbf{R}^n$. Тогда $f(x+y) = (A(x+y), x+y)/2 - (b, (x+y)) = (Ax, x)/2 - (b, x) + (Ax - b, y) + (Ay, y)/2 = f(x) + (Ax - b, y) + (Ay, y)/2$. Но $|(Ay, y)| \leqslant \|A\| \|y\|^2$, поэтому $(Ay, y)/2 = o(y)$. Итак, $f(x)$ дифференцируема в любой точке x и

$$\nabla f(x) = Ax - b. \tag{3}$$

Функция $f(x)$ называется *дифференцируемой на множестве* $Q \subset \mathbf{R}^n$, если она дифференцируема во всех точках Q. Если $f(x)$ дифференцируема на всем пространстве $\mathbf{R}^n$, то говорят просто, что она *дифференцируема*.

Пусть $f(x)$ дифференцируема на отрезке $[x, x+y]$ (т. е. для точек вида $x+\tau y$, $0 \leqslant \tau \leqslant 1$). Рассмотрим функцию одного переменного $\varphi(\tau) = f(x+\tau y)$ и вычислим ее производную для $0 \leqslant \tau \leqslant 1$:

$$\frac{\varphi(\tau+\Delta\tau)-\varphi(\tau)}{\Delta\tau} = \frac{f(x+(\tau+\Delta\tau)\,y)-f(x+\tau y)}{\Delta\tau} = \\ = \frac{(\nabla f(x+\tau y), \Delta\tau y) + o(\Delta\tau y)}{\Delta\tau},$$

$$\varphi'(\tau) = \lim_{\Delta\tau\to 0} \frac{\varphi(\tau+\Delta\tau)-\varphi(\tau)}{\Delta\tau} = (\nabla f(x+\tau y), y).$$

Таким образом, $\varphi(\tau)$ дифференцируема на $[0, 1]$ и

$$\varphi'(\tau) = (\nabla f(x+\tau y), y). \tag{4}$$

Величина

$$f'(x; y) = \lim_{\varepsilon\to+0} \frac{f(x+\varepsilon y)-f(x)}{\varepsilon} \tag{5}$$

называется *производной по направлению* (или *вариацией*) функции $f(x)$ в точке x по направлению y. Производная по направлению может существовать и для негладких функций. Например, для $f(x) = \|x\|$ имеем $f'(0; y) = \|y\|$. Если $f(x)$ имеет в точке x производную по всем направлениям, линейную по y: $f'(x; y) = (a, y)$, то говорят, что $f(x)$ *дифференцируема по Гато* в точке x. Такая функция имеет частные производные, причем $f'(x; e_i) = \partial f(x)/\partial x_i$ (e_i — координатные орты), $a = (\partial f/\partial x_1, \ldots$ $\ldots, \partial f/\partial x_n)$. Из формулы (4) следует, что если $f(x)$ дифференцируема в точке x, то она дифференцируема и по Гато, причем

$$f'(x; y) = \varphi'(0) = (\nabla f(x), y). \tag{6}$$

Обратное, вообще говоря, неверно. Например, функция f: $\mathbf{R}^n \to \mathbf{R}^1$, $n \geqslant 2$, вида

$$f(x) = \begin{cases} 1, & \text{если } \|x-a\| = \|a\|, \quad x \neq 0, \\ 0 & \text{в остальных точках,} \end{cases} \tag{7}$$

где $a \in \mathbf{R}^n$, $a \neq 0$, дифференцируема в точке 0 по любому направлению и $f'(0; y) = 0$ для всех y, т. е. она дифференцируема по Гато в нуле, однако она не дифференцируема (и даже не непрерывна) в этой точке. Отметим еще, что иногда (чтобы подчеркнуть отличие от дифференцируемости по Гато) употребляют термин «*дифференцируемость по Фреше*» вместо «дифференцируемость».

Если функция $f(x)$ дифференцируема на $[x, x+y]$, то, пользуясь (4) и *формулой Ньютона — Лейбница* $\varphi(1)=\varphi(0)+\int_0^1 \varphi'(\tau)\,d\tau$, получаем запись остаточного члена в (2) в интегральной форме:

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

$$= f(x)+(\nabla f(x),\ y)+\int_0^1 (\nabla f(y+\tau y)-\nabla f(x),\ y)\,d\tau. \tag{8}$$

Другой полезный результат — *теорема о среднем* — следует из *формулы конечных приращений* $\varphi(1)=\varphi(0)+\varphi'(\theta)$, $0\leqslant \theta\leqslant 1$, и (4):

$$f(x+y)=f(x)+(\nabla f(x+\theta y),\ y), \tag{9}$$

где $0\leqslant\theta\leqslant 1$ — некоторое число.

Упражнения.

1. Докажите, что: а) $\nabla\|x\| = x/\|x\|$ при $x\neq 0$; при $x=0$ функция $\|x\|$ недифференцируема; б) $\nabla\|x_+\|^2 = 2x_+$.

2. Докажите, что из непрерывности по x производной Гато следует дифференцируемость.

2. Дифференцирование векторных функций. До сих пор речь шла о дифференцируемости скалярных функций. Совершенно аналогично определяется дифференцируемость векторных функций. Функция g: $\mathbf{R}^n\to\mathbf{R}^m$ называется *дифференцируемой в точке* x, если найдется матрица A размерности $m\times n$ такая, что для всех $y\in\mathbf{R}^n$

$$g(x+y)=g(x)+Ay+o(y). \tag{10}$$

Матрица A называется *производной* или *матрицей Якоби* отображения $g(x)$, и для нее применяется то же обозначение $g'(x)$ или $\nabla g(x)$, что и в скалярном случае. Итак,

$$g(x+y)=g(x)+g'(x)\,y+o(y), \tag{11}$$

т. е. дифференцируемая в точке x функция допускает в этой точке линейную аппроксимацию первого порядка. Очевидно, что для дифференцируемой векторной функции $g(x)=(g_1(x),\ \ldots,\ g_m(x))$ элементы матрицы Якоби определяются формулой $g'(x)_{ij}=\partial g_i(x)/\partial x_j$.

Пусть g: $\mathbf{R}^n\to\mathbf{R}^m$ — дифференцируемая в точке x функция, а h: $\mathbf{R}^m\to\mathbf{R}^s$ дифференцируема в точке $g(x)$. Тогда справедливо *правило дифференцирования сложных функций*

$$[h(g(x))]'=h'(g(x))\,g'(x), \tag{12}$$

где в правой части стоит произведение матриц h' и g'.