Лекция 9

4. Численные методы одномерной оптимизации

4.1. Постановка задачи

Пусть X — некоторое подмножество $\mathbb{R}$, $f(x)$ — числовая функция, определённая на X, $f : X \to \mathbb{R}$. Будем рассматривать задачу минимизации функции $f(x)$ на множестве X. Напомним некоторые определения из математического анализа.

Точка $x_* \in X$ называется *точкой минимума* функции $f(x)$ на множестве X, если $f(x_*) \leq f(x)$ для всех $x \in X$; величина $f(x_*)$ называется *наименьшим* или *минимальным значением* $f(x)$ на X и обозначают $\min\limits_{x \in X} f(x) = f(x_*)$. Множество всех точек минимума $f(x)$ на X обозначается через X_*.

В зависимости от свойств множества X и функции $f(x)$ множество X_* может содержать одну, несколько или бесконечно много точек, а также возможны случаи, когда X_* пусто.

Пример 1.

а) если $f(x) = \ln x$, $X = (0, 1]$, то $X_* = \varnothing$;

б) если $f(x) = x^2$, $X = [-1, 1]$, то $X_* = \{0\}$;

в) если $f(x) = \sin^2(\pi x)$, $X = \mathbb{R}$, то $X_* = \mathbb{Z}$.

Функция $f(x)$ называется *ограниченной снизу* на множестве X, если существует такое число M, что $f(x) \geq M$ для всех $x \in X$. Легко видеть, что функция $f(x)$ не ограничена снизу тогда и только тогда, когда существует последовательность $\{x_n\} \subset X$, для которой $\lim\limits_{n\to\infty} f(x_n) = -\infty$.

Пусть функция $f(x)$ ограничена снизу на множестве X. Тогда число f_* называется *нижней гранью* $f(x)$ на X, если:

1. $f(x) \geq f_*$ при всех $x \in X$;
2. для всякого $\varepsilon > 0$ найдётся точка $x_\varepsilon \in X$, для которой $f(x_\varepsilon) < f_* + \varepsilon$.

Если функция $f(x)$ не ограничена снизу на множестве X, то в качестве нижней грани $f(x)$ на X принимается $f_* = -\infty$. Нижнюю грань $f(x)$ на X обозначают через $\inf\limits_{x\in X} f(x) = f_*$.

Если $X_* \neq \varnothing$, то, очевидно, что

$$\inf_{x\in X} f(x) = \min_{x\in X} f(x) = f(x_*).$$

В этом случае говорят, что функция $f(x)$ на X достигает нижней грани. Подчеркнём, что $\inf\limits_{x\in X} f(x) = f_*$ существует всегда, а $\min\limits_{x\in X} f(x) = f_*$ может не существовать.

Пример 2. Пусть

$$f(x) = \frac{1}{1 + x^2}, \quad x \in \mathbb{R}.$$

В этом случае $\inf\limits_{x\in X} f(x) = 0$, а $\min\limits_{x\in X} f(x)$ не существует.

Последовательность $\{x_n\} \subset X$ называется *минимизирующей* для функции $f(x)$ на множестве X, если

$$\lim_{n\to\infty} f(x_n) = \inf_{x\in X} f(x) = f_*.$$

Из определения и существования нижней грани следует, что минимизирующая последовательность всегда существует.

Последовательность $\{x_n\}$ *сходится* к множеству X, если

$$\lim_{n\to\infty} d(x_n, X) = 0, \quad d(x_n, X) = \inf_{x\in X} |x_n - x|.$$

Заметим, что если $X_* \neq \varnothing$, то всегда существует минимизирующая последовательность, сходящаяся к X_*; например, можно взять стационарную последовательность $x_n = x_*$ $(n = 1,\ 2,\ \dots)$, где x_* — какая-либо точка из X_*. Следующий пример показывает, что не всякая минимизирующая последовательность сходится к X_*, $X_* \neq \varnothing$.

Пример 3. Пусть

$$f(x) = \frac{x^2}{1 + x^4}, \quad X = \mathbb{R}.$$

Очевидно, здесь $f_* = 0$ и $X_* = \{0\}$. Последовательность $x_n = n$ $(n = 1, 2, \dots)$ является минимизирующей, так как $\lim\limits_{n\to\infty} f(n) = 0$, но $d(x_n, X_*) = n$ не стремится к нулю.

Условимся задачу на минимум записывать в виде:

$$f(x) \to \min, \quad x \in X. \tag{1.1}$$

Получить точное решение задачи (1.1) удаётся лишь в редких случаях. Поэтому на практике строят минимизирующую последовательность $\{x_n\}$, сходящуюся ко множеству X_*, и в качестве приближения для f_* и точки $x_* \in X_*$ берут соответственно величину $f(x_n)$ и точку x_n при достаточно большом n.

Как показывает пример 3, не всякая минимизирующая последовательность может быть использована для получения приближённого решения задачи (1.1). Мы будем рассматривать лишь такие задачи, у которых любая минимизирующая последовательность сходится к X_*. Один такой класс задач даётся следующей теоремой, называемой *теоремой Вейерштрасса.*

Теорема 1.1. *Пусть $X \subset \mathbb{R}$ — замкнутое и ограниченное множество, функция $f(x)$ непрерывна на X. Тогда $f(x)$ ограничена снизу на X, множество X_* точек минимума $f(x)$ на X непусто, замкнуто и любая минимизирующая последовательность $\{x_n\}$ сходится к X_*.*

4.2. Унимодальные функции

Функция $f(x)$ называется *унимодальной* на отрезке $X = [a, b]$, если она непрерывна на $[a, b]$ и существуют числа α и β, $a \le \alpha \le \beta \le b$, такие, что:

1) если $a < \alpha$, то $f(x)$ строго убывает на $[a, \alpha]$;
2) если $\beta < b$, то $f(x)$ строго возрастает на $[\beta, b]$;
3) при $x \in [\alpha, \beta]$ $f(x) = f_* = \min\limits_{x\in[a,b]} f(x)$, так что $X_* = [\alpha, \beta]$.

Отметим, что возможно вырождение в точку одного или двух отрезков из $[a, \alpha]$, $[\alpha, \beta]$ и $[\beta, b]$. В частности, если $\alpha = \beta$, то $f(x)$ называется *строго унимодальной* на отрезке $[a, b]$.

Из определения вытекают следующие свойства унимодальных функций.

1. Любая из точек локального минимума унимодальной функции является и точкой её глобального минимума на отрезке $[a, b]$.

2. Функция, унимодальная на отрезке $[a, b]$, является унимодальной и на любом меньшем отрезке $[c, d] \subset [a, b]$.

3. Пусть $f(x)$ — унимодальная функция на отрезке $[a, b]$ и $a \le x_1 < x_2 \le b$. Тогда:

$$\begin{aligned} f(x_1) \le f(x_2) \quad &\Rightarrow \quad x_* \in [a, x_2]; \\ f(x_1) > f(x_2) \quad &\Rightarrow \quad x_* \in [x_1, b], \end{aligned} \tag{2.1}$$

где x^* — одна из точек минимума $f(x)$ на отрезке $[a, b]$.

4.3. Метод деления отрезка пополам

Пусть функция $f(x)$ — унимодальная на отрезке $[a, b]$. Поиск минимума $f(x)$ на $[a, b]$ начинается с выбора двух точек x_1 и x_2:

$$x_1 = \frac{b + a - \delta}{2}, \quad x_2 = \frac{b + a + \delta}{2}, \tag{3.1}$$

где $\delta > 0$ — малое число. В качестве нового отрезка берут либо $[a, x_2]$, либо $[x_1, b]$.

При этом отношение длин нового и исходного отрезков

$$\tau = \frac{b - x_1}{b - a} = \frac{x_2 - a}{b - a}$$

близко к $1/2$, этим и объясняется название метода. В конце вычислений в качестве приближённого значения x_* берут середину последнего из найденного отрезка.

Опишем алгоритм метода деления отрезка пополам.

1. Определить x_1 и x_2 по формулам (3.1). Вычислить $f(x_1)$ и $f(x_2)$.

2. Сравнить $f(x_1)$ и $f(x_2)$. Если $f(x_1) \le f(x_2)$, то перейти к отрезку $[a, x_2]$, положив $b = x_2$, иначе — к отрезку $[x_1, b]$, положив $a = x_1$.

3. Найти достигнутую точность $\varepsilon_n = \dfrac{b - a}{2}$. Если $\varepsilon_n > \varepsilon$, то возвращаемся к **1**. если $\varepsilon_n \le \varepsilon$, то завершить поиск x_*, перейдя к **4**.

4. Положить $x_* \approx \overline{x} = \dfrac{a + b}{2}$, $f_* \approx f(\overline{x})$.

После n итераций длина отрезка поиска станет

$$\Delta_n = \frac{b - a}{2^n} + \left(1 - \frac{1}{2^n}\right)\delta. \tag{3.2}$$

При этом будет достигнута точность определения точки минимума

$$\varepsilon_n = \frac{\Delta_n}{2} = \frac{b - a}{2^{n+1}} + \left(1 - \frac{1}{2^n}\right)\frac{\delta}{2}. \tag{3.3}$$

Исходя из условия $\varepsilon_n \le \varepsilon$, получим число итераций, необходимое для определения x_* с точностью ε:

$$n \ge \log_2 \frac{b - a - \delta}{2\varepsilon - \delta}. \tag{3.4}$$

Величина δ выбирается пользователем и должна находиться в интервале $(0, 2\varepsilon)$.

4.4. Метод золотого сечения

Золотым сечением отрезка называют деление отрезка на две части так, чтобы отношение длины большей части отрезка к длине всего отрезка было равно равно отношению длины меньшей части к большей. Пусть отрезок имеет длину l и точка деления делит его на две части l_1, l_2: $l_1 > l_2$, $l_1 + l_2 = l$. По условию

$$\frac{l_1}{l} = \frac{l_2}{l_1}.$$

Следовательно,

$$l_1^2 = l_2 l = l_2(l_1 + l_2).$$

Разделив обе части этого равенства на l_1^2, получим квадратное уравнение для определения $\tau = l_2/l_1$:

$$\tau^2 + \tau - 1 = 0.$$

Отбрасывая отрицательный корень, находим

$$\tau = \frac{\sqrt{5} - 1}{2} \approx 0{,}618033988.$$

Золотое сечение отрезка $[a, b]$ производят две симметрично расположенные точки (рис. 4.1):

$$x' = a + (1 - \tau)(b - a), \quad x'' = a + \tau(b - a). \tag{4.1}$$

Рис. 1. Золотое сечение отрезка

Легко проверяется, что

$$\frac{x' - a}{x'' - a} = \frac{x'' - x'}{x' - a}, \quad \frac{b - x''}{b - x'} = \frac{x'' - x'}{b - x''}.$$

Таким образом, точка x', в свою очередь, производит золотое сечение отрезка $[a, x'']$, а точка x'' — золотое сечение отрезка $[x', b]$.

Рассмотрим метод золотого сечения. Пусть на отрезке $[a, b]$ задана унимодальная функция F. Необходимо найти минимум F на $[a, b]$.

Удобно обозначить $[a_0, b_0] = [a, b]$. Пусть точки x_0' и x_0'' осуществляют золотое сечение отрезка $[a_0, b_0]$ (по формулам (4.1)). Вычислим значения $F(x_0')$, $F(x_0'')$. Если $F(x_0') < F(x_0'')$, то, ввиду унимодальности функции F, её минимум x_* находится на отрезке $[a_0, x_0'']$. Положим $a_1 = a_0$ и $b_1 = x_0''$. Если же $F(x_0') \geq F(x_0'')$, то $x_* \in [x_0', b_0]$ и мы полагаем $[a_1, b_1] = [x_0', b_0]$.

После k шагов получим интервал $[a_k, b_k]$. Точки x_k' и x_k'' осуществляют золотое сечение отрезка $[a_k, b_k]$. Сравнивая значения $F(x_k')$ и $F(x_k'')$, положим $[a_{k+1}, b_{k+1}]$ равным отрезку $[a_k, x_k'']$, если $F(x_k') < F(x_k'')$, и равным $[x_k', b_k]$ в противном случае. В качестве очередного приближения к x_* можно взять любую точку $x_{k+1} \in [a_{k+1}, b_{k+1}]$.

При использовании метода золотого сечения после сужения интервала $[a_k, b_k]$ значение функции в одной из внутренних точек нового интервала $[a_{k+1}, b_{k+1}]$ оказывается известным: в точке $x''_{k+1} = x'_k$ при интервале $[a_{k+1}, b_{k+1}] = [a_k, x''_k]$ или в точке $x'_{k+1} = x''_k$ при интервале $[a_{k+1}, b_{k+1}] = [x'_k, b_k]$.

Так как из формул (4.1) следует

$$x''_k - a_k = b_k - x'_k = \tau\,(b_k - a_k),$$

то получаем

$$b_{k+1} - a_{k+1} = \tau\,(b_k - a_k), \quad k = 0, 1, \ldots. \tag{4.2}$$

Из соотношения (4.2) легко получить равенство

$$b_k - a_k = \tau^k\,(b - a), \quad k = 0, 1, \ldots. \tag{4.3}$$

Отсюда видно, что $b_k - a_k \to 0$ при $k \to \infty$. Процесс останавливают при достаточно малом отрезке $[a_k, b_k]$. Приведём оценку погрешности после N шагов

$$|x_N - x_*| \le \tau^N\,(b - a). \tag{4.4}$$

Лекция 10

4.5. Выпуклые функции

Функция $f(x)$, заданная на отрезке $[a, b]$, называется *выпуклой* на этом отрезке, если для всех x', $x'' \in [a, b]$ и произвольного числа $\alpha \in [0, 1]$ выполняется неравенство

$$f[\alpha x' + (1 - \alpha)x''] \leq \alpha f(x') + (1 - \alpha)f(x''). \tag{5.1}$$

Перечислим основные свойства выпуклых функций.

1. Если функция $f(x)$ выпукла на $[a, b]$, то на любом отрезке $[x', x''] \subset \subset [a, b]$ её график расположен не выше хорды, проведённой через точки графика с абсциссами x' и x''.

2. Дифференцируемая на отрезке $[a, b]$ функция $f(x)$ выпукла на этом отрезке тогда и только тогда, когда производная $f'(x)$ является возрастающей функцией на отрезке $[a, b]$.

3. Дважды дифференцируемая на отрезке $[a, b]$ функция $f(x)$ выпукла на этом отрезке тогда и только тогда, когда для всякого $x \in [a, b]$ выполняется неравенство $f''(x) \geq 0$.

4. Условие выпуклости для дифференцируемой на отрезке $[a, b]$ функции $f(x)$ означает, что на этом отрезке любая касательная к графику $f(x)$ лежит не выше этого графика.

5. Если $f(x)$ — выпуклая дифференцируемая на отрезке $[a, b]$ функция и в точке $x_* \in [a, b]$ выполняется равенство

$$f'(x_*) = 0, \tag{5.2}$$

то x_* является точкой глобального минимума $f(x)$ на $[a, b]$.

Таким образом, равенство (5.2) для выпуклой дифференцируемой функции является не только необходимым условием глобального минимума, но и его достаточным условием.

6. Всякая выпуклая непрерывная на отрезке $[a, b]$ функция является и унимодальной на этом отрезке.

4.6. Метод средней точки

Если определение значений производной $f'(x)$ не представляет затруднений, то в процедуре исключения отрезков метода деления отрезка пополам вычисление двух значений $f(x)$ вблизи середины очередного отрезка можно заменить вычислением одного значения $f'(x)$ в его средней точке $\overline{x} = (a + b)/2$.

В самом деле, если $f'(\overline{x}) > 0$, то точка $\overline{x}$ лежит на участке строгого возрастания $f(x)$, поэтому $x_* < \overline{x}$, и точку минимума следует искать на отрезке $[a, \overline{x}]$. При $f'(\overline{x}) < 0$ точка $\overline{x}$ лежит на участке строгого убывания $f(x)$, поэтому $x_* > \overline{x}$, и точку минимума следует искать на отрезке $[\overline{x}, b]$. Равенство $f'(\overline{x}) = 0$ означает, что точка минимума найдена точно: $x_* = \overline{x}$.

Такое исключение отрезков требует на каждой итерации только одного вычисления $f'(x)$ и уменьшает отрезок поиска ровно в два раза.

Опишем алгоритм метода средней точки.

1. Положить $\overline{x} = \dfrac{a + b}{2}$. Вычислить $f'(\overline{x})$.

2. Проверка на окончание поиска: если $|f'(\overline{x})| \leq \varepsilon$, то положить $x_* = \overline{x}$, $f_* \approx f(\overline{x})$ и завершить поиск, иначе — перейти к **3**.

3. Сравнить $f'(\overline{x})$ с нулём. Если $f'(\overline{x}) > 0$, то продолжить поиск на отрезке $[a, \overline{x}]$, положив $b = \overline{x}$, иначе — перейти к отрезку $[\overline{x}, b]$, положив $a = \overline{x}$. Перейти к **1**.

4.7. Метод хорд

Как уже отмечалось, равенство $f'(x) = 0$ является необходимым и достаточным условием глобального минимума выпуклой дифференцируемой функции $f(x)$. Поэтому если на концах отрезка $[a, b]$ производная $f'(x)$ имеет разные знаки, т. е. $f'(a)f'(b) < 0$, то на интервале (a, b) $f'(x)$ обращается в нуль.

Рассмотрим метод хорд для поиска корня уравнения $F(x) = 0$ на отрезке $[a, b]$, $F(a)F(b) < 0$. Построим прямую, проходящую через точки $(a, F(a))$ и $(b, F(b))$, т. е. построим хорду. Хорда

$$y = F(a) + \frac{F(b) - F(a)}{b - a}\,(x - a).$$

пересекает ось абсцисс в точке

$$\widetilde{x} = a - \frac{F(a)}{F(b) - F(a)}\,(b - a).$$

Далее вычисляется $F(\overline{x})$. В зависимости от знака $F(\overline{x})$ выбирается интервал $[a, \overline{x}]$ или $[\overline{x}, b]$.

Опишем алгоритм метода хорд.

1. Найти $\widetilde{x}$ по формуле:

$$\widetilde{x} = a - \frac{f'(a)}{f'(b) - f'(a)}\,(b - a). \tag{7.1}$$

Вычислить $f'(\widetilde{x})$. Перейти к **2**.

2. Проверка на окончание поиска: если $|f'(\widetilde{x})| \leq \varepsilon$, то положить $x_* = \widetilde{x}$, $f_* \approx f(\widetilde{x})$ и завершить поиск, иначе — перейти к **3**.

3. Сравнить $f'(\widetilde{x})$ с нулём. Если $f'(\widetilde{x}) > 0$, то продолжить поиск на отрезке $[a, \widetilde{x}]$, положив $b = \widetilde{x}$, иначе — перейти к отрезку $[\widetilde{x}, b]$, положив $a = \widetilde{x}$. Перейти к **1**.

4.8. Метод Ньютона

Будем считать, что $f(x)$ является дважды непрерывно дифференцируемой выпуклой функцией. Пусть x_k — приближённое значение точки минимума, полученное на k-й итерации. Аппроксимируем $f(x)$ рядом Тейлора:

$$f(x) \approx f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}\,f''(x_k)(x - x_k)^2. \tag{8.1}$$

Как известно, парабола $y = A + Bx + Cx^2$, $C > 0$, достигает минимума в точке $x = -B/2C$. Используем этот факт, находим следующее приближение к точки минимума:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f'(x_k)}. \tag{8.2}$$

4.9. Метод парабол

Пусть на плоскости (x, y) даны три точки

$$(x_1, y_1),\ (x_2, y_2),\ (x_3, y_3),$$

где x_1, x_2, x_3 различны. Соответствующий полином Лагранжа:

$$L(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)}\, y_1 + \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)}\, y_2 + \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)}\, y_3. \tag{9.1}$$

Обозначим

$$\Delta = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2).$$

Запишем $L(x)$ в виде:

$$L(x) = \frac{1}{\Delta}((x-x_2)(x-x_3)(x_3-x_2)\, y_1 - (x-x_1)(x-x_3)(x_3-x_1)\, y_2 + (x-x_1)(x-x_2)(x_2-x_1)\, y_3).$$

Пусть $L(x) = Ax^2 + Bx + C$. Тогда

$$\begin{aligned} A &= \frac{1}{\Delta}\left\{(x_3 - x_2)y_1 - (x_3 - x_1)y_2 + (x_2 - x_1)y_3\right\} \\ B &= \frac{1}{\Delta}\left\{-(x_3^2 - x_2^2)y_1 + (x_3^2 - x_1^2)y_2 - (x_2^2 - x_1^2)y_3\right\} \\ C &= \frac{1}{\Delta}\left\{x_2x_3(x_3 - x_2)y_1 - x_1x_3(x_3 - x_1)y_2 + x_1x_2(x_2 - x_1)y_3\right\}. \end{aligned} \tag{9.2}$$

Предполагая, что $A > 0$, найдём

$$x_{min} = -\frac{B}{2A}, \quad y_{min} = -\frac{B^2}{4A} + C. \tag{9.3}$$

Опишем метод парабол. Рассмотрим унимодальную на отрезке $[a, b]$ функцию $f(x)$. Выберем три точки x_1, x_2 и x_3 отрезка $[a, b]$, для которых выполняются неравенства

$$x_1 < x_2 < x_3, \quad f(x_1) \geq f(x_2) \leq f(x_3). \tag{9.4}$$

Построим интерполяционный трёхчлен Лагранжа, проходящий через точки $(x_1, f(x_1))$, $(x_2, f(x_2))$, $(x_3, f(x_3))$ графика функции $f(x)$. Будем считать, хотя бы одно из неравенств (9.4) для $f(x)$ является строгим. Если $f(x_1) = f(x_2) = f(x_3)$, то из унимодальности $f(x)$ следует, что она достигает минимума в каждой точке отрезка $[x_1, x_3]$. Тогда из (9.4) следует, что ветви искомой параболы направлены вверх, а точка минимума трёхчлена принадлежит отрезку $[x_1, x_3]$.

Определяя коэффициенты трёхчлена по формулам (9.2), найдём точку минимума $\overline{x}$ по формуле (9.3). Выберем новые точки x_1, x_2 и x_3, удовлетворяющие (9.4). Далее описанная процедура повторяется с новыми точками.