

Методы оптимизации

Михайлов Максим

8 апреля 2021 г.

Оглавление

Лекция 1	10 февраля	3
1	Теория погрешности	3
2	Задачи оптимизации. Вводное.	7
3	Одномерная минимизация функций. Прямые методы.	8
3.1	Метод дихотомии	8
Лекция 2	17 февраля	10
3.2	Метод золотого сечения	10
3.3	Метод Фибоначчи	11
3.4	Метод парабол	12
3.5	Комбинированный метод Брента	13
Лекция 3	24 февраля	14
3.6	Метод равномерного перебора	14
4	Методы оптимизации, использующие производную	14
4.1	Методы средней точки	14
4.2	Метод хорд (<i>метод секущей</i>)	15
4.3	Метод Ньютона (<i>метод касательной</i>)	16
Лекция 4	3 марта	17
4.4	Модификации метода Ньютона	19
5	Метод минимизации многомодальных функций (<i>метод ломаных</i>)	19
Лекция 5	10 марта	21
6	Минимизация функций многих переменных	21
6.1	Постановка задачи	21
6.2	Свойства выпуклых множеств и выпуклых функций	23
6.3	Необходимое и достаточное условие безусловного экстремума	24

Лекция 1

10 февраля

Этот курс — о минимизации (*максимизации*) функционалов. Кроме конкретных методов оптимизации, планируется рассмотреть форматы хранения матриц, о методах работы с ними и рассмотреть 1-2 (*может быть 3*) СЛАУ с использованием различных форматов.

Т.к. значения, получаемые компьютерами — не точные, нам требуется теория погрешности.

1 Теория погрешности

Все погрешности разделяются на два класса:

1. Неустраняемая — обусловлена неточностью исходных данных. Например, неточное знание физических констант или других параметров задачи. Тем не менее, необходимо знать эту погрешность, чтобы ставить рамки погрешности для решения.
2. Устранимая — погрешность процесса решения задачи. Эту погрешность можно уменьшить выбором метода решения задачи.

(a) Погрешность модели

(b) Остаточная погрешность (*погрешность аппроксимации*)

Например, аппроксимация ряда первыми n его членами или аппроксимация по теореме Вейерштрасса квадратичной функцией.

(c) Погрешность округления

(d) Накапливаемая погрешность

2c и **2d** часто объединяют в вычислительную погрешность.

Определение. Пусть X^* — точное решение, а X — найденное (*приближенное*) решение. Тогда $X^* - X$ называется **погрешностью**, а её модуль $\Delta X = |X^* - X|$ — **абсолютная погрешность**.

Разумеется, ΔX представляет сугубо теоретический интерес, т.к. X^* неизвестна и ΔX нельзя вычислить.

Определение. В качестве требования к решению часто предоставляется **предельная абсолютная погрешность** $\Delta_X \geq |X^* - X|$.

Определение. Также существует **относительная погрешность** $\delta X = \left| \frac{X^* - X}{|X|} \right|$

Относительная погрешность позволяет выражать погрешность относительно значений самой величины. Например, при измерении длины парты погрешность 1 см не очень хорошо, а при измерении расстояния между городами — приемлемо.

Определение. **Предельная относительная погрешность** $\delta_X \geq \left| \frac{X^* - X}{|X|} \right|$

Определение. **Значащие цифры** некоторого числа — все цифры в его изображении, отличные от нуля, а также нули, если они содержатся между значащими цифрами или расположены в конце числа и указывают на сохранение разряда точности.

Определение. Если значащая цифра приближенного значения a , находящаяся в разряде, в котором выполняется условие $\Delta \leq 0.5 \cdot 10^k$, т.е. абсолютное значение погрешности не превосходит половину единицы этого разряда (k — номер этого разряда), то такая цифра называется **верной в узком смысле**.

Цифра называется **верной в широком смысле**, если в определении выше используется 1 вместо 0.5.

Пример. $a = 3.635$, $\Delta a = 0.003$

- $k = 0 \quad \frac{1}{2} \cdot 10^0 = \frac{1}{2} \geq \Delta a$
- $k = -1 \quad \frac{1}{2} \cdot 10^{-1} = 0.05 \geq \Delta a$
- $k = -2 \quad \frac{1}{2} \cdot 10^{-2} = 0.005 \geq \Delta a$
- $k = -3 \quad \frac{1}{2} \cdot 10^{-3} = 0.0005 < \Delta a$

Таким образом, цифра 5 является сомнительной, остальные — верные.

Пример. Рассмотрим следующие способы записи одного и того же выражения:

$$\left(\frac{\sqrt{2} - 1}{\sqrt{2} + 1} \right)^3 = (\sqrt{2} - 1)^6 = (3 - 2\sqrt{2})^3 = 99 - 70\sqrt{2}$$

Посчитаем все выражения с различными приближениями $\sqrt{2}$:

- $\frac{7}{5} = 1.4$
- $\frac{17}{12} = 1.41666$
- $\frac{707}{500} = 1.414$
- $\sqrt{2} = 1.4142135624$

$\sqrt{2}$	$\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)^3$	$(\sqrt{2}-1)^6$	$(3-2\sqrt{2})^3$	$99-70\sqrt{2}$
$\frac{7}{5}$	$\frac{1}{216} \approx 0.0046$	$\frac{64}{15625} \approx 0.0051$	$\frac{1}{125} = 0.008$	1
$\frac{17}{12}$	$\frac{125}{24389} \approx 0.00513$	$\frac{15625}{2985354} \approx 0.0052$	$\frac{1}{216} \approx 0.0046$	$-\frac{1}{6} = -0.6(6)$
$\frac{707}{500}$	$\frac{8869743}{1758416743} \approx 0.005044$	$\frac{78672340886049}{15625 \cdot 10^{12}} \approx 0.00504$	$\frac{636056}{125000000} \approx 0.00509$	0.02

$$\Delta_{(X \pm Y)} = \Delta_X + \Delta_Y$$

$$\Delta_{(X \cdot Y)} \approx |Y| \Delta_X + |X| \Delta_Y$$

$$\Delta_{(X/Y)} \approx \left| \frac{1}{Y} \right| \Delta_X + \left| \frac{X}{Y^2} \right| \Delta_Y$$

$$|\Delta u| = |f(x_1 + \Delta x_1, \dots, x_n + \Delta x_n) - f(x_1 \dots x_n)|$$

$$|\Delta u| \approx |df(x_1 \dots x_n)| = \left| \sum_{i=1}^n \frac{\partial u}{\partial x_i} \Delta x_i \right| \leq \sum_{i=1}^n \left| \frac{\partial u}{\partial x_i} \right| |\Delta x_i|$$

$$\Delta_u = \sum_{i=1}^n \left| \frac{\partial u}{\partial x_i} \right| \Delta x_i$$

$$|\delta u| = \sum_{i=1}^n \left| \frac{\partial \ln u}{\partial x_i} \right| |\Delta x_i|$$

$$\delta_u = \sum_{i=1}^n \left| \frac{\partial \ln u}{\partial x_i} \right| |\Delta x_i|$$

$$\delta_{(X \pm Y)} = \left| \frac{X}{X \pm Y} \right| \delta_X + \left| \frac{Y}{X \pm Y} \right| \delta_Y$$

$$\delta_{(X \cdot Y)} = \delta_X + \delta_Y$$

$$\delta_{(X/Y)} = \delta_X + \delta_Y$$

Вернемся к прошлому примеру и посчитаем относительную погрешность.

$$\triangleleft x = \frac{7}{5}$$

$$\delta_{f_1} = 3 \left| \frac{1}{x-1} - \frac{1}{x+1} \right| \cdot |\delta x| = 6.25|\delta x|$$

$$\delta_{f_2} = 6 \left| \frac{1}{x-1} \right| \cdot |\delta x| = 15|\delta x|$$

$$\delta_{f_3} = 6 \left| \frac{1}{3-2x} \right| \cdot |\delta x| = 30|\delta x|$$

$$\delta_{f_4} = \left| \frac{90}{99-70x} \right| \cdot |\delta x| = 70|\delta x|$$

Таким образом, наибольшую погрешность даёт f_4 , наименьшую — f_1 .

Пример.

$$y^2 - 140y + 1 = 0$$

$$y = 70 - \sqrt{4899}$$

$$\sqrt{4899} \approx 69.99$$

$$y \approx 70 - 69.99 = 0.01$$

Посчитаем другим методом — избавимся от вычитания похожих чисел.

$$y = \frac{1}{70 + \sqrt{4899}}$$

$$y = \frac{1}{139.99} \approx \frac{1}{140} = 0.00714285 \approx 0.007143$$

Можно заметить, что результат весьма точнее.

Пример. Рассмотрим задачу вычисления суммы $S = \sum_{j=1}^{10^6} \frac{1}{j^2}$.

Если суммировать по формуле $S_n = S_{n-1} + \frac{1}{n^2}$, то из-за того, что сначала суммируются большие числа, а потом малые, погрешность велика: $\Delta = 10^6 \cdot 2^{-1} \approx 2 \cdot 10^{-4}$

Если же суммировать с конца, то $\Delta = \mathcal{O}\left(\frac{1}{n}\right) \approx 6 \cdot 10^{-8}$

Рекомендации для увеличения точности вычислений:

1. Если складывать или вычитать последовательность чисел, то лучше начинать с малых членов.
2. Желательно избегать вычитания двух почти равных чисел, по возможности преобразую формулу.
3. Необходимо сводить к минимуму число математических операций. Это также способствует ускорению работы алгоритма.

4. Если ЯП и компьютер позволяют использовать числа разных типов, то числа с большим числом разрядов всегда повышают точность вычислений (*в ущерб памяти*).

Дробные числа нужно сравнивать с помощью ε , т.е. $|a - b| \leq \varepsilon$

2 Задачи оптимизации. Вводное.

Здесь и далее целевая функция — функция, которую мы минимизируем.

Обозначение. Пусть целевая функция — $f(x)$. Это обозначается как $f(x) \xrightarrow{x \in U} \min$.

$f(x) \rightarrow \max \Rightarrow -f(x) \rightarrow \min$. Таким образом, мы без потери общности рассматриваем задачу минимизации.

Определение. Если $\exists x^* \in U \quad f(x^*) \leq f(x) \quad \forall x \in U$, то такой x^* называется **точкой (глобального) минимума**

Обозначение. Множество всех точек минимума обозначается $U^* = \{x_i^* \mid i = 1 \dots k\}$

Мы рассматриваем класс функций таких, что $U^* \neq \emptyset$

Определение. Функция $f(x)$ называется **унимодальной** на $[a, b]$, если она:

1. Непрерывна на $[a, b]$
2. $\exists \alpha, \beta : a \leq \alpha \leq \beta \leq b$, такие что:
 - (а) Если $a < \alpha$, то на $[a, \alpha]$ $f(x)$ строго монотонно убывает.
 - (б) Если $\beta < b$, то на $[\beta, b]$ $f(x)$ строго монотонно возрастает.
 - (с) $\forall x \in [\alpha, \beta] \quad f(x) = f_* = \min_{[a, b]} f(x)$

Свойства.

1. Если функция унимодальна на $[a, b]$, то она унимодальна и на $[c, d] \subset [a, b]$
2. Если f унимодальна на $[a, b]$, $a \leq x_1 < x_2 \leq b$, тогда:
 - (а) Если $f(x_1) \leq f(x_2)$, то $x^* \in [a, x_2]$
 - (б) Если $f(x_1) > f(x_2)$, то $x^* \in [x_1, b]$

Определение. $f(x)$, заданная на $[a, b]$, называется **выпуклой** на этом отрезке, если

$$\forall x', x'' \in [a, b], \alpha \in [0, 1] \quad f(\alpha x' + (1 - \alpha)x'') \leq \alpha f(x') + (1 - \alpha)f(x'')$$

Свойства.

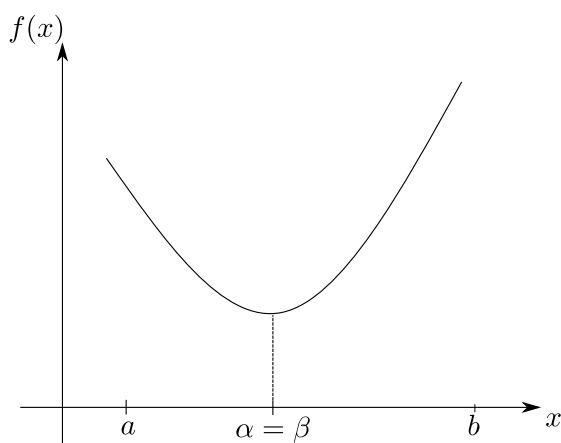


Рис. 1.1:
Вырожденные α и β ,
униmodalная
функция

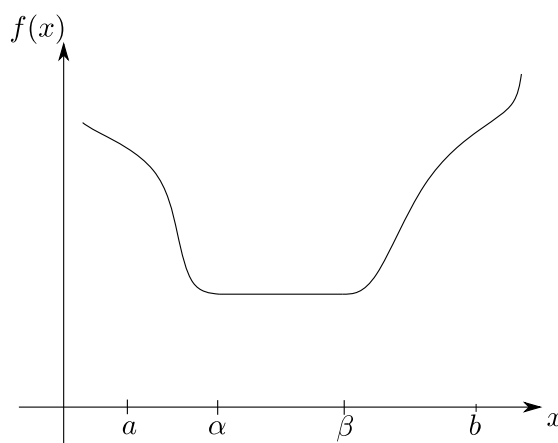


Рис. 1.2:
Униmodalная
функция

1. Если $f(x)$ выпукло на $[a, b]$, то $\forall [x', x''] \subset [a, b]$, то её график расположен ниже хорды между x' и x''
2. Всякая выпуклая функция на отрезке является униmodalной на нём.

Определение. Стационарные точки — точки x , для которых $f'(x) = 0$.

Мы будем рассматривать одномерные задачи оптимизации, т.к. многомерные задачи часто сводятся к одномерным.

3 Одномерная минимизация функций. Прямые методы.

Прямые методы — методы, не использующие производные целевой функции.

3.1 Метод дихотомии

Этот метод — тернарный поиск.

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

$$\tau = \frac{b - x_1}{b - a} = \frac{x_2 - a}{b - a} \rightarrow \frac{1}{2}$$

$$x^* \in [a_i, b_i] \quad \forall i$$

Шаг 1: Находим x_1 и x_2 , вычисляем $f(x_1)$ и $f(x_2)$

Шаг 2: Сравниваем $f(x_1)$ и $f(x_2)$.

- Если $f(x_1) \leq f(x_2)$, переходим к отрезку $[a, x_2]$, т.е. $b = x_2$
- Иначе переходим к $[x_1, b]$, т.е. $a = x_1$

Шаг 3: $\varepsilon_n = \frac{b-a}{2}$, где n — номер итерации.

- Если $\varepsilon_n > \varepsilon$, переходим к новой итерации.
- Если $\varepsilon_n \leq \varepsilon$, завершаем поиск и переходим к шагу 4.

Шаг 4: $X^* \approx \overline{X} = \frac{a+b}{2}$

Примечание. δ выбирается на интервале $(0, 2\varepsilon)$. Чем меньше δ , тем больше относительное уменьшение длины отрезка на каждой итерации. При чрезмерно малом δ сравнение $f(x_1)$ и $f(x_2)$ будет затруднительно, т.к. они близки.

Мы можем оценить число необходимых итераций:

$$n \geq \log_2 \frac{b-a-\delta}{2\varepsilon-\delta}$$

Лекция 2

17 февраля

3.2 Метод золотого сечения

Рассмотрим отрезок $[0, 1]$. Пусть $x_2 = \tau$, тогда симметрично расположенная $x_1 = 1 - \tau$. Пусть дальше был выбран отрезок $[0, \tau]$, тогда пусть $x'_2 = 1 - \tau$. Чтобы новые точки делили отрезок в таком же соотношении, необходимо, чтобы $\frac{1}{\tau} = \frac{\tau}{1-\tau} \Rightarrow \tau^2 = 1 - \tau \Rightarrow \tau = \frac{\sqrt{5}-1}{2} \approx 0.61803$. Таким образом, $x_1 = 1 - \tau = \frac{3-\sqrt{5}}{2}$, $x_2 = \tau = \frac{\sqrt{5}-1}{2}$

В общем случае для отрезка $[a, b]$:

$$x_1 = a + \frac{3 - \sqrt{5}}{2}(b - a), x_2 = a + \frac{\sqrt{5} - 1}{2}(b - a) \quad (1)$$

Вычислим погрешность:

$$\Delta_n = \tau^n(b - a) \quad \varepsilon_n = \frac{\Delta_n}{2} = \frac{1}{2} \left(\frac{\sqrt{5} - 1}{2} \right)^n (b - a)$$

Для заданного ε условия окончания $\varepsilon_n \leq \varepsilon$.

Результат метода:

$$x^* = \frac{a_{(n)} + b_{(n)}}{2}$$

Оценка числа шагов для достижения искомой точности:

$$n \geq \ln \left(\frac{\frac{2\varepsilon}{b-a}}{\ln \tau} \right) \approx 2 \cdot 1 \cdot \ln \left(\frac{b-a}{2\varepsilon} \right)$$

Шаг 1: Находим x_1 и x_2 по формуле (1), вычисляем $f(x_1)$ и $f(x_2)$. $\varepsilon_n = \frac{b-a}{2}$, $\tau = \frac{\sqrt{5}-1}{2}$.

Шаг 2: – Если $\varepsilon_n > \varepsilon$, переходим к шагу 3.

– Если $\varepsilon_n \leq \varepsilon$, переходим к шагу 4.

Шаг 3: Сравниваем $f(x_1)$ и $f(x_2)$.

– Если $f(x_1) \leq f(x_2)$, то $b = x_2, x_2 = x_1, x_1 = b - \tau(b - a)$. Мы запоминаем $f(x_2)$ для следующего шага, т.к. оно равно $f(x_1)$ на этом шаге.

– Иначе $a = x_1, x_1 = x_2, f(x_1) = f(x_2)$. Мы запоминаем $f(x_1)$ для следующего шага, т.к. оно равно $f(x_2)$ на этом шаге.

Шаг 4: $X^* \approx \bar{X} = \frac{a(n)+b(n)}{2}$

3.3 Метод Фибоначчи

Мы знаем, что $F_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}$, а также при $n \rightarrow +\infty$ $F_n \approx \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n}{\sqrt{5}}$

Рассмотрим нулевую итерацию:

$$x_1 = a + \frac{F_n}{F_{n+2}}(b - a) \quad x_2 = a + \frac{F_{n+1}}{F_{n+2}}(b - a)$$

Рассмотрим k -тую итерацию:

$$x_1 = a_{(k)} + \frac{F_{n-k+1}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+1}}{F_{n+2}}(b_0 - a_0)$$

$$x_2 = a_{(k)} + \frac{F_{n-k+2}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+2}}{F_{n+2}}(b_0 - a_0)$$

Пусть $k = n$, тогда:

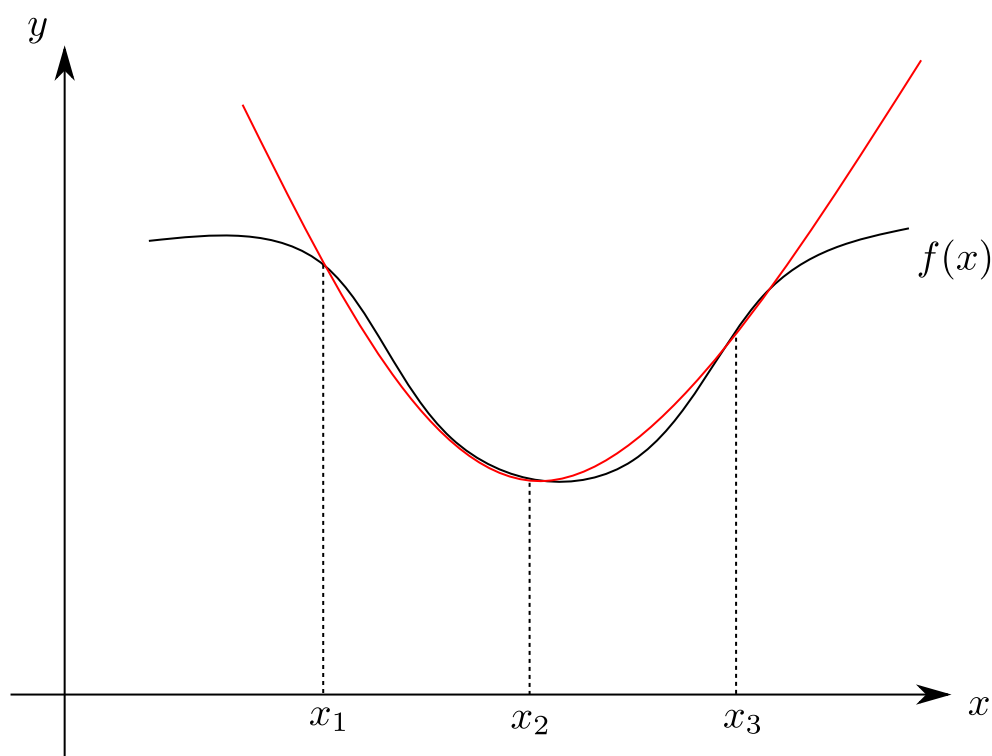
$$x_1 = a_n + \frac{F_1}{F_{n+2}}(b_0 - a_0) \quad x_2 = a_n + \frac{F_2}{F_{n+2}}(b_0 - a_0)$$

Условие на погрешность:

$$\frac{b_n - a_n}{2} = \frac{b_0 - a_0}{F_{n+2}} < \varepsilon$$

Какое брать n ? Такое, что $\frac{b_0 - a_0}{\varepsilon} < F_{n+2}$

Есть проблема, при большом n $\frac{F_n}{F_{n+2}}$ есть бесконечная десятичная дробь, вследствие чего образуется погрешность.

Рис. 2.1: Функция $f(x)$ и её приближение параболой.

3.4 Метод парабол

Пусть $\exists x_1, x_2, x_3 \in [a, b]$, такие что $\begin{cases} x_1 < x_2 < x_3 \\ f(x_1) \geq f(x_2) \leq f(x_3) \end{cases}$

Тогда приближающая парабола имеет вид $q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$.

Мы имеем условия на коэффициенты этой параболы: $\begin{cases} q(x_1) = f(x_1) = f_1 \\ q(x_2) = f(x_2) = f_2 \\ q(x_3) = f(x_3) = f_3 \end{cases}$

Коэффициенты можно найти следующим образом:

$$a_0 = f_1 \quad a_1 = \frac{f_2 - f_1}{x_2 - x_1} \quad a_2 = \frac{1}{x_3 - x_2} \left(\frac{f_3 - f_1}{x_3 - x_1} - \frac{f_2 - f_1}{x_2 - x_1} \right)$$

Тогда результат итерации есть $\bar{x} = \frac{1}{2} \left(x_1 + x_2 - \frac{a_1}{a_2} \right)$, на следующей лекции будет рассказан переход к следующей итерации.

Точки x_1, x_2, x_3 для новой итерации выбираются следующим образом:

1. (а) Если $x_1 < \bar{x} < x_2 < x_3$ и $f(\bar{x}) \geq f(x_2)$, то $x^* \in [\bar{x}, x_3]$, $x_1 = \bar{x}$, точки x_2 и x_3 не меняются.

- (b) Если $x_1 < \bar{x} < x_2 < x_3$ и $f(\bar{x}) < f(x_2)$, то $x^* \in [x_1, x_2]$, $x_3 = x_2$, $x_2 = \bar{x}$, точка x_1 не меняется.
2. (a) Если $x_1 < x_2 < \bar{x} < x_3$ и $f(\bar{x}) \leq f(x_2)$, то $x^* \in [x_2, x_3]$, $x_1 = x_2$, $x_2 = \bar{x}$, точка x_3 не меняется.
- (b) Если $x_1 < x_2 < \bar{x} < x_3$ и $f(\bar{x}) > f(x_2)$, то $x^* \in [x_1, \bar{x}]$, $x_3 = \bar{x}$, точки x_1 и x_2 не меняются.

Примечание. Метод парабол имеет квадратичную сходимость.

Примечание. Метод парабол требует гладкость функции, что неверно для предыдущих методов.

3.5 Комбинированный метод Брента

Для собственного изучения.

Лекция 3

24 февраля

3.6 Метод равномерного перебора

Шаг 1: Если $f(x_0) > f(x_0 + \delta)$, то $k = 1, x_1 = x_0 + \delta, h = \delta$

иначе $x_1 = x_0, h = -\delta$

Шаг 2: $h = 2h, x_{k+1} = x_k + h$

Шаг 3: Если $f(x_k) > f(x_{k+1})$, то $k = k + 1$ и переходим к шагу 2. Иначе прекращаем поиск и искомое лежит в $[x_{k-1}, x_{k+1}]$

4 Методы оптимизации, использующие производную

В рамках этой главы $f(x)$ — дифференцируемая или дважды дифференцируемая выпуклая функция.

Есть три классических метода, использующих производную:

- Средней точки
- Метод хорд
- Метод Ньютона

$f'(x) = 0$ — необходимое и достаточное условие глобального минимума. Таким образом, условие остановки вычислений — $f'(x) \approx 0$, т.е. $|f'(x)| \leq \varepsilon$

4.1 Методы средней точки

Средняя точка $\bar{x} = \frac{a+b}{2}$.

Общая идея алгоритма:

- Если $f'(x) > 0$, то $\bar{x} \in$ участку монотонного возрастания $f(x)$ и $x^* < \bar{x}$, т.е. минимум лежит на $[a, \bar{x}]$
- Если $f'(x) < 0$, то аналогично можем вывести, что минимум лежит на $[\bar{x}, b]$
- Если $f'(x) = 0$, то мы нашли решение.

Перепишем это в виде алгоритма:

Шаг 1: $\bar{x} = \frac{a+b}{2}$, вычислим $f'(\bar{x})$

Шаг 2: Если $|f'(x)| \leq \varepsilon$, то $x^* = \bar{x}$ и завершаем вычисление.

Шаг 3: Сравниваем $f'(x)$ с нулём:

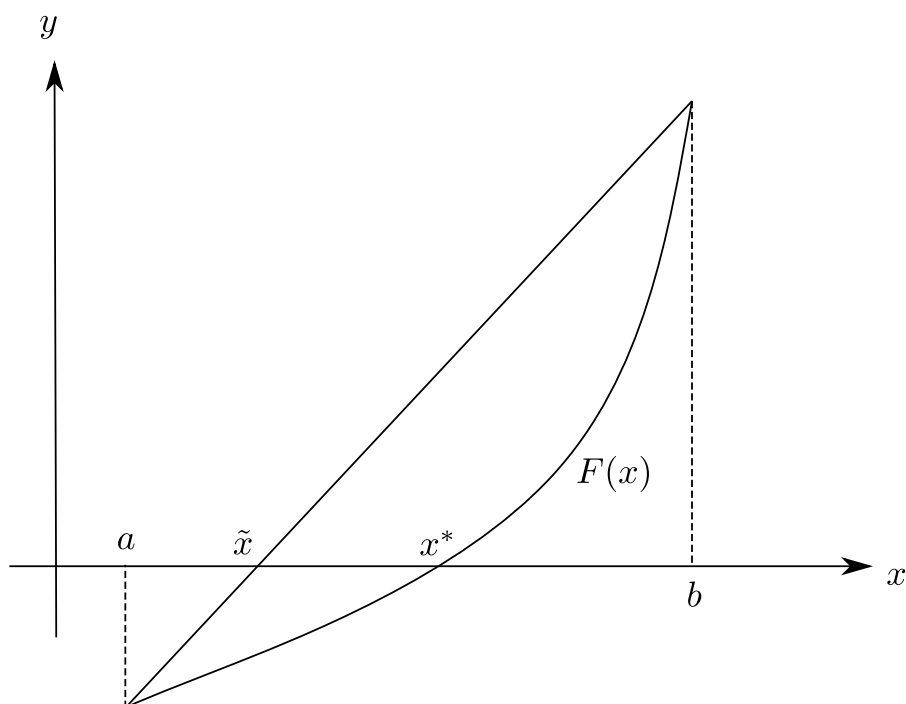
- Если $f'(x) > 0$, то $x^* \in [a, \bar{x}]$ и $b = \bar{x}$
- Иначе $x^* \in [\bar{x}, b]$ и $a = \bar{x}$

Длина отрезка после n итераций есть $\Delta_n = \frac{b-a}{2^n}$

4.2 Метод хорд (метод секущей)

Если $\exists f'(x)$ на $[a, b]$, $f'(a) \cdot f'(b) < 0$ и $f'(x)$ непрерывна на $[a, b]$, то $\exists x \in (a, b) : f'(x) = 0$.

$F(x) = f'(x)$. Пусть \tilde{x} — точка пересечения хорды $F(x)$ с осью Ox на $[a, b]$



Можем тривиально вывести \tilde{x} из уравнения прямой по двум точками:

$$\tilde{x} = a - \frac{f'(a)}{f'(a) - f'(b)}(a - b) \quad (2)$$

Шаг 1: Считаем \tilde{x} по (2)

Шаг 2: Если $|f'(\tilde{x})| \leq \varepsilon$, то $x^* = \tilde{x}$ и мы заканчиваем вычисление.

Иначе шаг 3.

Шаг 3: Переходим к новому отрезку:

- Если $f'(\tilde{x}) > 0$, то $x^* \in [a, \tilde{x}]$, $b = \tilde{x}$, $f'(b) = f'(\tilde{x})$, переходим к шагу 1
- иначе $x^* \in [\tilde{x}, b]$, $a = \tilde{x}$, $f'(a) = f'(\tilde{x})$, переходим к шагу 1

Примечание. Если $f'(a) \cdot f'(b) \geq 0$, то $x^* = a$ или $x^* = b$.

4.3 Метод Ньютона (*метод касательной*)

Если f выпуклая на $[a, b]$ и дважды непрерывно дифференцируемая, то уравнение $f'(x) = 0$ решается методом Ньютона.

Пусть $x_0 \in [a, b]$ — начальное приближение x^* . $F(x) = f'(x)$ линеаризуема в окрестности x_0 , т.е.

$$F(x) \approx F(x_0) + F'(x_0)(x - x_0)$$

Пусть x_1 — следующее приближение к x^* . Это будет пересечение касательной с Ox . Найдём эту точку.

$$\begin{aligned} F(x_0) + F'(x_0)(x_1 - x_0) &= 0 \\ x_1 &= x_0 - \frac{F(x_0)}{F'(x_0)} \end{aligned}$$

Таким образом, мы можем получить $\{x_k\}_{k=1}^n$ — итерационную последовательность.

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

Условие остановки такое же, как в предыдущих методах: $|f'(x_k)| \leq \varepsilon$

Лекция 4

3 марта

Пусть x_k — текущая оценка решения x^*

Рассмотрим ряд Тейлора:

$$f(x_k + p) = f(x_k) + pf'(x_k) + \frac{1}{2}p^2 f''(x_k) + \dots$$

$$\begin{aligned} f(x^*) &= \min_x f(x) \\ &= \min_p f(x_k + p) \\ &= \min_p \left(f(x_k) + pf'(x_k) + \frac{1}{2}p^2 f''(x_k) + \dots \right) \\ &\approx \min_p \left(f(x_k) + pf'(x_k) + \frac{1}{2}p^2 f''(x_k) \right) \end{aligned}$$

Приравняем производную выражения под \min к нулю:

$$f'(x_k) + pf''(x_k) = 0$$

$$p = -\frac{f'(x_k)}{f''(x_k)}$$

Тогда $x^* \approx x_k + p$ и $x_{k+1} = x_k + p = x_k - \frac{f'(x_k)}{f''(x_k)}$

Главное преимущество метода Ньютона — квадратичная скорость сходимости, т.е. если x_k достаточно близка к x^* и $f''(x^*) > 0$, то $|x_{k+1} - x^*| \leq \beta |x_k - x^*|^2$

Метод Ньютона может потерпеть неудачу в следующих случаях:

1. $f(x)$ плохо аппроксимируется первыми тремя членами в ряде Тейлора. Тогда x_{k+1} может быть хуже (как аппроксимация) x_k .
2. $f''(x_k) = 0$, тогда p не определен.
3. Кроме f нужно вычислять f' и f'' , что затруднительно в реальных задачах.

Мы можем аппроксимировать производную по определению:

$$f'(x_k) \approx \frac{f(x_k + h) - f(x_k)}{h}$$

Эта формула называется правой разностной схемой, у нее есть улучшение, называемое центральной разностной схемой:

$$f'(x_k) \approx \frac{f(x_k + h) - f(x_k - h)}{2h}$$

Если $f(x)$ — квадратичная функция, то метод Ньютона сходится за один шаг при любом выборе x_0 .

Достаточное условие монотонной сходимости метода Ньютона

Пусть $x^* \in [a, b]$ и $f(x)$ трижды непрерывно дифференцируемая и выпуклая на $[a, b]$ функция. Тогда $\{x_k\}$ будет сходиться к пределу x^* монотонно, если $0 < \frac{x^* - x_{k+1}}{x^* - x_k} < 1$

$$f'(x^*) = 0 = f'(x_k) + f''(x_k)(x^* - x_k) + \frac{f'''(x)}{2}(x^* - x_k)^2$$

$$\frac{x^* - x_{k+1}}{x^* - x_k} = \frac{x^* - x_k + \frac{f'(x_k)}{f''(x_k)}}{x^* - x_k} = 1 - \frac{2}{2 + \frac{f'''(x)(x^* - x_k)^2}{f'(x_k)}}$$

Последовательность итераций $\{x_k\}$ монотонна, если $\frac{f'''(x)}{f'(x_k)} > 0$, таким образом условие монотонной сходимости метода Ньютона — постоянство на $x \in [x^*, x_0]$ знака $f'''(x)$ и его совпадение с $f'(x_0)$.

Пример. $f(x) = x \cdot \arctg(x) - \frac{1}{2}$???

$$f'(x) = \arctg x \quad f''(x) = \frac{1}{1+x^2} > 0 \quad f'''(x) = -\frac{2x}{(1+x^2)^2}$$

$f'(x) \cdot f'''(x) < 0$, таким образом не будет монотонной сходимости.

Пусть $x_0 = 1$.

k	x_k	$f'(x_k)$	$f''(x_k)$
0	1	0.785	$\frac{1}{2}$
1	-0.57	-0.518	a
2	0.117	0.	
4	$9 \cdot 10^{-8}$		

4.4 Модификации метода Ньютона

Метод Ньютона-Рафсона

$$x_{k+1} = x_k - \tau_k \frac{f'(x_k)}{f''(x_k)}, 0 < \tau_k \leq 1$$

τ_k — константы. Если $\tau = 1$, то метод Ньютона-Рафсона вырождается в метод Ньютона.

Для нахождения τ_k зададим $\varphi(\tau)$:

$$\varphi(\tau) = f\left(x_k - \tau \frac{f'(x_k)}{f''(x_k)}\right) \rightarrow \min$$

Тогда

$$\tau_k = \frac{(f'(x_k))^2}{(f'(x_k))^2 + (f'(\tilde{x}))^2}, \text{ где } \tilde{x} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Метод Марквардта

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

, где $\mu_k > 0$

μ_0 выбирают на порядок выше значения $f''(x_0)$, $\mu_{k+1} = \begin{cases} \frac{\mu_k}{2} & , \text{ если } f(x_{k+1}) < f(x_k) \\ \mu_{k+1} = 2\mu_k & , \text{ если } f(x_{k+1}) \geq f(x_k) \end{cases}$

5 Метод минимизации многомодальных функций (метод ломаных)

Определение. $f(x)$, $x \in [a, b]$ удовлетворяет условию Липшица, если $\forall x_1, x_2 \in [a, b] |f(x_1) - f(x_2)| \leq L|x_1 - x_2|$

Шаг 1 Возьмём $x_1^* = \frac{1}{2L}(f(a) - f(b) + L(a + b))$ и $p_1^* = \frac{1}{2}(f(a) + f(b) + L(a - b))$. Добавим в рассматриваемое множество $x'_1 = x_1^* - \Delta_1$ и $x''_1 = x_1^* + \Delta_1$, где $\Delta_1 = \frac{1}{2L}(f(x_1^*) - p_1)$

Шаг 2 Из пар (x'_1, p_1) и (x''_1, p_1) выберем пару с минимальной $p : (x_2^*, p_2^*)$ и исключим из рассматриваемого множества.

Шаг n В результате мы получим множество из n пар (x, p) . Исключаем пару с минимальной p и вместо неё

Пример. $f(x) = \frac{\sin x}{x}$, $[a, b] = [10, 15]$, $\varepsilon = 0.01$

Проверим условие Липшица:

$$|f'(x)| = \left| \frac{x \cos x - \sin x}{x^2} \right| < \frac{x |\cos x| + \sin |x|}{x^2} < \frac{x + 1}{x^2} \leq 0.11$$

n	x_n^*	p_n^*	$2L\Delta_n$	x_n'	x_n''	p_n
1	12.056	-0.281	0.240	10.963	13.149	-0.161
2	10.963	-0.161	0.070	10.646	11.280	-0.126
3	13.149	-0.161	0.203	12.227	14.701	-0.096
4	10.646	-0.126	0.038	10.474	10.818	-0.107
5	11.280	-0.126	0.041	11.094	11.466	-0.106
6	10.474	-0.107	0.024	10.364	10.584	-0.095
7	10.818	-0.107	0.160	10.745	10.891	-0.099
8	11.094	-0.106	0.016	11.020	11.168	-0.098
9	11.466	-0.106	0.028	11.338	11.594	-0.092
10	10.891	-0.099	$0.008 < \varepsilon$			

Лекция 5

10 марта

6 Минимизация функций многих переменных

6.1 Постановка задачи

Необходимо найти $x^* = (x_1 \ x_2 \ \dots \ x_n)^T \in U \subset E_n$, где U — множество допустимых значений, а E_n — евклидово пространство размера n , при этом $f(x^*) = \min_{x \in U} f(x)$.

Примечание.

1. Как и в одномерном случае, задача минимизации эквивалентна задаче максимизации и в общем случае называется задачей поиска экстремума.
2. Если U задается ограничениями на вектор x , то такая задача оптимизации называется задачей поиска условного экстремума.
3. Если $U = E_n$, т.е. не имеет ограничений, то такая задача оптимизации называется задачей поиска безусловного экстремума.
4. Решением задачи поиска экстремума называется пара $(x^*, f(x^*))$.

Определение. Если $f(x^*) \leq f(x) \ \forall x \in U$, то x^* называется **глобальным минимумом**.

Определение. Если $\exists \varepsilon > 0 : \|x - x^*\| < \varepsilon \Rightarrow f(x^*) \geq f(x)$, то x^* называется **локальным минимумом**.

Примечание.

$$\|x\| = \sqrt{\sum_i x_i^2}$$

Определение. Поверхностью уровня функции $f(x)$ называется множество точек, в которых функция принимает постоянное значение.

Определение. Градиентом $\nabla f(x)$ непрерывно дифференцируемой функции $f(x)$ в x называется:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

Примечание. Градиент направлен по нормали к поверхности нормали уровня, т.е. перпендикулярно к касательной плоскости, проведенной в точке x в сторону наибольшего возрастания функции.

Определение. Матрица Гессе $\mathbf{H}(x)$ дважды непрерывно дифференцируемой в точке x функции $f(x)$ называется матрица частных производных производных второго порядка, вычисленных в данной точке.

$$\mathbf{H}(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix}$$

1. $\mathbf{H}(x)$ симметрична, имеет размер $n \times n$.
2. Можно определить антиградиент — вектор, равный по модулю градиенту и направленный противоположно. Антиградиент указывает в сторону наибольшего убывания $f(x)$.
3. $\Delta f(x) = f(x + \Delta x) - f(x) = \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \mathbf{H}(x) \Delta x + o(\|\Delta x\|^2)$, где $o(\|\Delta x\|^2)$ есть сумма всех членов разложения, имеющих порядок выше второго. Можем заметить, что $\Delta x^T \mathbf{H}(x) \Delta x$ — квадратичная форма.

Определение. Квадратичная форма $\Delta x^T \mathbf{H}(x) \Delta x$ ¹ называется:

- Положительно определенной, если $\forall \Delta x \neq 0 \quad \Delta x^T \mathbf{H}(x) \Delta x > 0$
- Отрицательно определенной, если $\forall \Delta x \neq 0 \quad \Delta x^T \mathbf{H}(x) \Delta x < 0$
- Положительно полуопределенной, если $\forall \Delta x \quad \Delta x^T \mathbf{H}(x) \Delta x \geq 0$ и имеется $\Delta x \neq 0 : \Delta x^T \mathbf{H}(x) \Delta x = 0$
- Отрицательно полуопределенной, если $\forall \Delta x \quad \Delta x^T \mathbf{H}(x) \Delta x \leq 0$ и имеется $\Delta x \neq 0 : \Delta x^T \mathbf{H}(x) \Delta x = 0$
- Неопределенной, если $\exists \Delta x, \tilde{\Delta x} : \Delta x^T \mathbf{H}(x) \Delta x > 0, \tilde{\Delta x}^T \mathbf{H}(x) \tilde{\Delta x} < 0$
- Тождественно равной нулю, если $\forall \Delta x \quad \Delta x^T \mathbf{H}(x) \Delta x = 0$

¹ и соответствующая ей матрица $\mathbf{H}(x)$

6.2 Свойства выпуклых множеств и выпуклых функций

Определение. Пусть $x, y \in E_n$, множество точек вида $\{z\} \subset E_n : z = \alpha x + (1 - \alpha)y$, т.е. z это отрезок $[x, y]$.

Определение. $U \subset E_n$ выпуклое, если вместе с точками $x, y \in U$ оно содержит весь отрезок z .

Определение. Функция $f(x)$, заданная на выпуклом множестве $U \subset E_n$, называется:

- выпуклой, если:

$$\forall x, y \in U, \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

- строго выпуклой, если:

$$\forall x, y \in U, \alpha \in (0, 1) \quad f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

- сильно выпуклой с константой $l > 0$, если:

$$\forall x, y \in U, \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{l}{2}\alpha(1 - \alpha)\|x - y\|^2$$

Свойства.

1. Функция $f(x)$ выпуклая, если её график целиком лежит не выше отрезка, соединяющего две её произвольные точки.
2. Функция $f(x)$ строго выпуклая, если её график целиком лежит ниже отрезка, соединяющего две её произвольные, но не совпадающие точки.²
3. Если функция сильно выпуклая, то она одновременно строго выпуклая и выпуклая.
4. Если функция строго выпуклая, то она выпуклая.
5. Выпуклость функции можно определить по $H(x)$:
 - Если $H(x) \geq 0 \quad \forall x \in E_n$, то $f(x)$ выпуклая.
 - Если $H(x) > 0 \quad \forall x \in E_n$, то $f(x)$ строго выпуклая.
 - Если $H(x) \geq lE^3 \quad \forall x \in E_n$, то $f(x)$ сильно выпуклая.

Свойства (выпуклых функций).

1. Если $f(x)$ — выпуклая функция на множестве U , то всякая точка локального минимума — глобальный минимум на U .

² Пример будет на следующей лекции

³ единичная матрица

2. Если выпуклая функция достигает своего минимума в двух различных точках, то она достигает минимума во всех точках отрезка, соединяющего эти точки.
3. Если $f(x)$ строго выпуклая функция на множестве U , то она может достигать своего глобального минимума на U не более чем в одной точке.

6.3 Необходимое и достаточное условие безусловного экстремума

Необходимое условие экстремума первого порядка

Пусть $x^* \in E_n$ — точка локального минимума⁴ $f(x)$ на E_n и $f(x)$ дифференцируема в точке x^* . Тогда $\nabla f(x)$ в точке x^* равен нулю: $\nabla f(x^*) = 0$ или $\frac{\partial f(x^*)}{\partial x_i} = 0 \quad \forall i \in 1 \dots n$. Точка x^* называется **стационарной**.

Необходимое условие экстремума второго порядка

Пусть $x^* \in E_n$ — точка локального минимума⁵ $f(x)$ на E_n и $f(x)$ дважды дифференцируема в точке x^* . Тогда $\mathbf{H}(x^*)$ положительно полуопределена или отрицательно полуопределена.

Достаточное условие экстремума

Пусть $f(x)$ в $x^* \in E_n$ дважды дифференцируема, $\nabla f(x^*) = 0$ и $\mathbf{H}(x) > 0$ (или $\mathbf{H}(x) < 0$). Тогда x^* — точка локального минимума⁶ $f(x)$ на E_n .

Проверка выполнений условий экстремума

- Вычисление угловых миноров $\mathbf{H}(x)$
- Вычисление главных миноров $\mathbf{H}(x)$

Есть два способа это сделать:

1. Исследование положительной или отрицательной определенности угловых и главных миноров $\mathbf{H}(x)$.
2. Анализ собственных значений $\mathbf{H}(x)$.

⁴ или максимума

⁵ или максимума

⁶ или максимума