МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра Вычислительная техника

УТВЕРЖДАЮ
Проректор по учебной работе
________ О.Г. Локтионова
«20» 06 2017 г.

МЕТОДЫ ОДНОМЕРНОЙ МИНИМИЗАЦИИ

Методические указания к лабораторным занятиям
по дисциплине «Методы оптимизации» для студентов
направления подготовки 09.04.01

Курск 2017

УДК 519.8

Составитель Ж.Т. Жусубалиев

Рецензент
Доктор технических наук, профессор *С.А. Филист*

Методы одномерной минимизации: методические указания к лабораторным занятиям по дисциплине «Методы оптимизации»/ Юго-Зап. гос. ун-т; сост.: Ж.Т. Жусубалиев. Курск, 2017. – 21 с.: ил. 4, табл. 2. – Библиогр.: с. 21.

Рассматриваются методы поиска безусловного минимума функции одной переменной и алгоритмы их численной реализации на ЭВМ. Приведены задачи для упражнений.

Методические указания соответствуют требованиям программы, утвержденной учебно-методическим объединением по направлению подготовки «Информатика и вычислительная техника».

Предназначены для студентов специальности направления подготовки 09.04.01 дневной и заочной форм обучения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60×84 $^{1}/_{16}$.
Усл. печ. л. . Уч.-изд. л. . Тираж 30 экз. Заказ . Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

1. Постановка задачи

Требуется найти безусловный минимум функции $f(x)$ одной переменной, т.е. такую точку $x_* \in R$, что $f(x_*) = \min\limits_{x \in R} f(x)$.

Поставленная задача одномерной минимизации может быть решена с помощью необходимых и достаточных условий безусловного экстремума. Однако проблема получения решения уравнения $\frac{df(x)}{dx} = 0$ может оказаться весьма сложной. Более того, в практических задачах функция $f(x)$ может быть не задана в аналитическом виде или часто неизвестно, является ли она дифференцируемой [1]. Поэтому получение численного решения минимума функции $f(x)$ является актуальным.

Для методов одномерной минимизации типично задание априорной информации о положении точки минимума с помощью начального интервала неопределенности $L_0 = [a_0, b_0]$ (рис. 1). Предполагается, что точка минимума x_* принадлежит интервалу L_0, но её точное значение неизвестно. Большинство известных методов одномерной минимизации применяется для класса унимодальных функций.

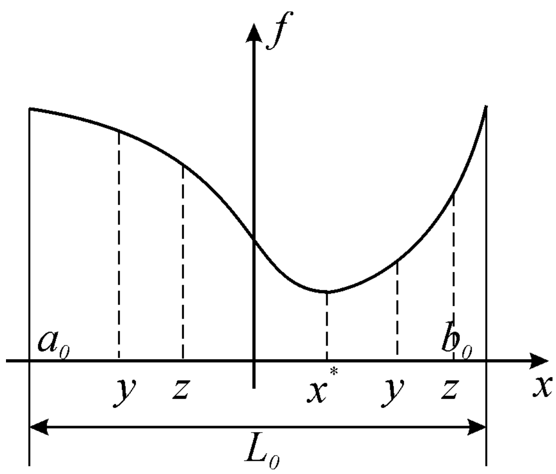

Рис. 1

Функция $f(x)$ *называется унимодальной на интервале* $L_0 = [a_0, b_0]$, *если она достигает глобального минимума на* $[a_0, b_0]$ *в* *единственной точке* x_**, причем слева от* x_* *эта функция строго убывает, а справа от* x_* *строго возрастает. Если* $a_0 \le y < z < x_*$*, то* $f(y) > f(z)$*, а если* $x_* < y < z \le b_0$*, то* $f(y) < f(z)$*, (рис. 1).*

Методы одномерной минимизации широко применяются в методах многомерной минимизации первого и второго порядков для нахождения оптимальной величины шага.

Решение задачи минимизации включает три этапа.

1. Выбор начального интервала неопределенности. Границы a_0, b_0 интервала должны быть такими, чтобы функция $f(x)$ была унимодальной.

2. Уменьшение интервала неопределенности, на котором реализуется конечная последовательность преобразований исходного интервала с тем, чтобы уменьшить его длину до заранее установленной величины.

3. Проверка условия окончания. Поиск заканчивается, когда длина текущего интервала неопределенности $[a_k, b_k]$ оказывается меньше установленной величины.

2. Алгоритм выбора начального интервала неопределенности

Для выбора начального интервала неопределенности можно применить **алгоритм Свенна** [1]. Согласно этому алгоритму $k+1$ пробная точка определяется по рекуррентной формуле

$x_{k+1} = x_k + 2^k \cdot \Delta$, $k = 0, 1, 2, \ldots$,

где x_0 - произвольно выбранная начальная точка; Δ – подбираемая некоторым способом величина шага, знак Δ определяется путем сравнения значений $f(x_0)$, $f(x_0 + |\Delta|)$ и $f(x_0 - |\Delta|)$.

Если $f(x_0 - |\Delta|) \ge f(x_0) \ge f(x_0 + |\Delta|)$, то согласно предположению об унимодальности, точка минимума должна располагаться правее точки x_0 и величина Δ выбирается положительной.

Если изменить знаки неравенств на противоположные, то значение Δ следует выбирать отрицательным.

Если же $f(x_0-|\Delta|)\geq f(x_0)\leq f(x_0+|\Delta|)$, то точка минимума лежит между $x_0-|\Delta|$ и $x_0+|\Delta|$, и поиск интервала неопределенности завершен.

В случае, когда $f(x_0-|\Delta|)\leq f(x_0)\geq f(x_0+|\Delta|)$, функция не является унимодальной.

Пример. Найти начальный интервал неопределенности для поиска минимума функции $f(x)=100(x-0{,}24)^2$.

Решение.

Воспользуемся алгоритмом Свенна. Зададим $x_0=1$ и $\Delta=1$. Вычислим значения функции в точках $x_0-\Delta=0$; $x_0=1$; $x_0+\Delta=2$: $f(0)=5{,}76$, $f(1)=57{,}76$, $f(2)=309{,}76$. Так как $f(0)<f(1)<f(2)$, то $\Delta=-1$, $b_0=1$, $x_1=x_0-\Delta=0$.

Найдем следующую точку $x_2=x_1+2\Delta=1-2=-1$. Так как $f(x_2)=153{,}76>f(x_1)=5{,}76$, то процедура завершается. $a_0=x_2=-1$. Искомый интервал неопределенности имеет вид $[a_0,\, b_0]=[-1,\, 1]$.

3. Методы нулевого порядка

Ряд методов минимизации основан на сравнении значений целевой функции $f(x)$, вычисляемых в точках x_0, x_1, …, x_N. Эти методы часто называют методами *нулевого порядка* или *прямого поиска*, а точки x_0, x_1, …, x_N называют пробными точками. Можно выделить две группы методов прямого поиска, соответствующие двум принципиально различным ситуациям:

- все N пробных точек x_k, $k=\overline{1,N}$, в которых будут вычисленные значения функции, выбирают заранее (до вычисления функции в этих точках).
- пробные точки x_k выбирают последовательно. Для выбора последующей пробной точки используют значения функции, вычисленные в предыдущих точках.

В первом случае поиск минимума функции называют *пассивным*, а во втором – *последовательным* [1, 2].

3.1. Метод равномерного поиска

Метод относится к пассивным стратегиям. Задается начальный интервал неопределенности $L_0 = [a_0, b_0]$ и количество вычислений функции N. Вычисления производятся в N равноотстоящих друг от друга точках (при этом интервал L_0 делится на $N+1$ равных интервалов).

Путем сравнения величин $f(x_i)$, $i = 1, ..., N$ находится точка x_k, в которой функция принимает наименьшее значение. Искомая точка минимума x_* будет лежать в интервале $[x_{k-1}, x_{k+1}]$.

Алгоритм включает следующие шаги:

Шаг 1. Задать начальный интервал неопределенности $L_0 = [a_0, b_0]$, N – количество вычислений функции.

Шаг 2. Вычислить пробные точки в соответствии с формулой $x_i = a_0 + i\dfrac{(b_0 - a_0)}{N+1}$, $i = 1, ..., N$.

Шаг 3. Вычислить значения функции во всех пробных точках: $f(x_i)$, $i = 1, ..., N$.

Шаг 4. Среди точек x_i, $i = 1, ..., N$, найти такую, в которой функция принимает наименьшее значение: $f(x_k) = \min\limits_{1 \le i \le N} f(x_i)$.

Гарантированная оценка погрешности:

$$|\bar{x} - x_*| \le \frac{b_0 - a_0}{N+1}.$$

3.2. Метод деления интервала пополам

Метод относится к последовательным стратегиям. Рассматриваемый метод позволяет исключить половину текущего интервала неопределенности на каждой итерации. Иногда этот метод называют *трехточечным поиском на равных интервалах*, поскольку его реализация основана на выборе трех пробных точек, равномерно распределенных в интервале поиска.

Алгоритм включает следующие шаги:

Шаг 1. Задать начальный интервал неопределенности $[a_0, b_0]$ и требуемую точность $\varepsilon > 0$.

Шаг 2. Вычислить среднюю точку $x_S = \frac{a_0 + b_0}{2}$, $L_0 = b_0 - a_0$. Вычислить значение целевой функции в средней точке $f(x_S)$.

Шаг 3. Вычислить две пробные точки: $x_1 = a_0 + \frac{L_0}{4}$ и $x_2 = b_0 - \frac{L_0}{4}$. Заметим, что точки x_1, x_S и x_2 делят интервал $[a_k, b_k]$ на четыре равные части. Вычислить значения $f(x_1)$ и $f(x_2)$.

Шаг 4. Сравнить значения $f(x_1)$ и $f(x_S)$:

а) если $f(x_1) < f(x_S)$, исключить интервал $[x_S, b_0]$, положив $b_0 = x_S$. Средней точкой нового интервала становится точка x_1 (рис. 2, а). Следовательно, необходимо положить $x_m = x_1$. Перейти к шагу 6.

б) если $f(x_1) \geq f(x_S)$, то перейти к шагу 5.

Шаг 5. Сравнить $f(x_2)$ с $f(x_S)$:

а) если $f(x_2) < f(x_S)$, исключить интервал $[a_0, x_S]$, положив $a_0 = x_S$. Средней точкой нового интервала становится точка x_2 (рис. 2, б). Перейти к шагу 6;

б) если $f(x_2) \geq f(x_s)$, исключить интервалы $[a_0, x_S)$, $(x_2, b_0]$, положив $a_0 = x_1$, $b_0 = x_2$. Средней точкой нового интервала становится точка x_S (рис. 2, в).

Шаг 6. Вычислить $L_0 = b_0 - a_0$ и проверить условие окончания поиска. Если $L_0 \leq \varepsilon$, закончить поиск. В качестве приближенного решения задачи можно взять середину последнего интервала поиска. В противном случае вернуться к шагу 2.

Замечание.

Средняя точка последовательно получаемых интервалов всегда совпадает с одной из пробных точек x_1 и x_2 или x_S, найденных на предыдущей итерации. Следовательно, на каждой итерации требуется не более двух вычислений значения функции.

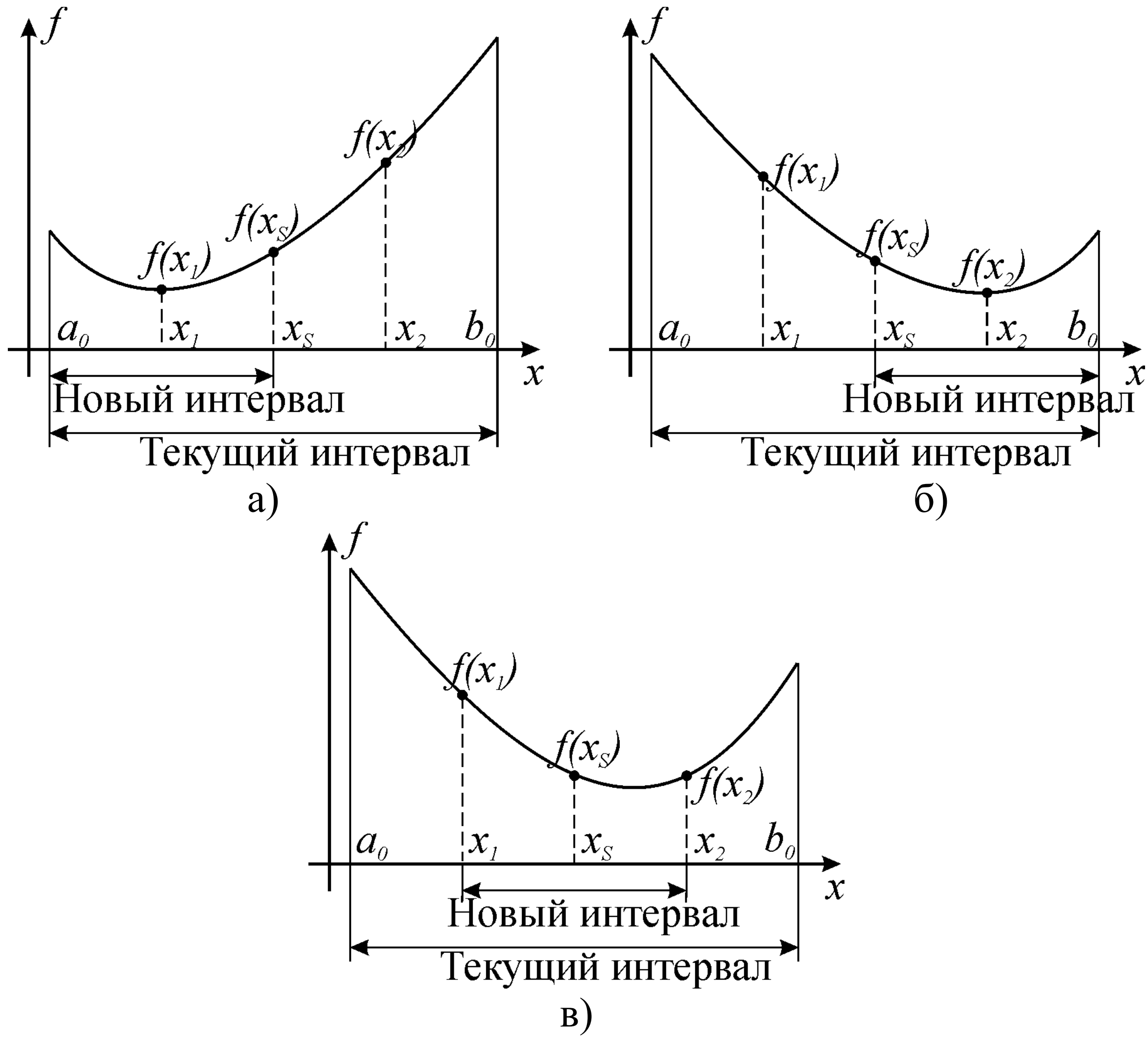

Рис. 2

3.3. Метод дихотомии

Метод относится к последовательным стратегиям. Алгоритм метода опирается на анализ значений функции в двух точках. Для нахождения этих точек текущий интервал неопределенности делится пополам, и в обе стороны от середины откладывается отрезок длиной δ, где δ – малое положительное число. Условия окончания процесса поиска стандартные: поиск заканчивается, когда длина текущего интервала неопределенности оказывается меньше установленной величины.

Алгоритм включает следующие шаги:

Шаг 1. Задать начальный интервал неопределенности $L_0 = [a_0, b_0]$, $\delta > 0$ – малое число, $\varepsilon > 0$ – точность.

Шаг 2. Положить $k = 0$.

Шаг 3. Вычислить $l_k = \frac{a_k + b_k}{2} - \delta$, $f(l_k)$, $r_k = \frac{a_k + b_k}{2} + \delta$, $f(r_k)$.

Шаг 4. Сравнить значения $f(l_k)$ и $f(r_k)$:

а) если $f(l_k) \le f(r_k)$, положить $a_{k+1} = a_k$, $b_{k+1} = r_k$ и перейти к шагу 5;

б) если $f(l_k) \ge f(r_k)$, положить $a_{k+1} = l_k$, $b_{k+1} = b_k$.

Шаг 5. Вычислить $L = b_{k+1} - a_{k+1}$ и проверить условие окончания:

а) если $L \le \varepsilon$, процесс поиска завершается и $x_* \in L = [a_{k+1}, b_{k+1}]$. В качестве приближенного решения можно взять середину последнего интервала: $x_* \cong \frac{a_{k+1} + b_{k+1}}{2}$;

б) если $L > \varepsilon$, то положить $k = k + 1$ и перейти к шагу 3.

3.4. Метод золотого сечения

Как известно, золотым сечением отрезка называется такое его деление на две не равные части, при котором отношение длины всего отрезка к длине его большей части равно отношению большей части к длине меньшей.

Золотое сечение отрезка $[a, b]$ осуществляется каждой из двух точек, симметрично расположенных относительно центра отрезка (см. рис. 3):

$$l = a + \frac{2}{3+\sqrt{5}}(b-a),\ r = a + \frac{2}{1+\sqrt{5}}(b-a);$$

$$\frac{b-a}{b-l} = \frac{b-l}{l-a} = \frac{1+\sqrt{5}}{2},\ \frac{b-a}{r-a} = \frac{r-a}{b-r} = \frac{1+\sqrt{5}}{2}.$$

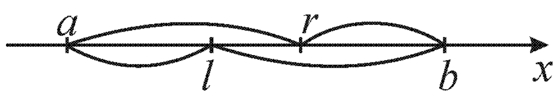

Рис. 3

Чтобы выполнить процедуру сокращения интервала неопределенности необходимо двумя внутренними точками разделить текущий интервал неопределенности $[a_k, b_k]$ на три части. Эти точки выбирают симметрично относительно середины отрезка $[a_k, b_k]$ таким образом, чтобы каждая из них производила золотое сечение этого отрезка.

Пробные точки l_k и r_k находятся по формулам:

$$l_k = a_k + \frac{2}{3+\sqrt{5}}(b_k - a_k),\ r_k = a_k + \frac{2}{1+\sqrt{5}}(b_k - a_k).$$

Важно отметить тот факт, что какой бы из отрезков $[a_k, r_k]$ или $[l_k, b_k]$ не был бы выбран за очередной интервал неопределенности, точка x_{k+1} совпадает с одной из пробных точек l_{k+1} или r_{k+1}, поэтому на каждой итерации достаточно вычислить значение функции лишь в одной недостающей пробной точке.

Алгоритм метода включает следующие шаги.

Шаг 1. Задать начальный интервал неопределенности $L_0 = [a_0, b_0]$, $\varepsilon > 0$ – точность.

Шаг 2. Положить $k = 0$.

Шаг 3. Вычислить $l_k = a_k + \frac{2}{3+\sqrt{5}}(b_k - a_k)$, $r_k = a_k + \frac{2}{1+\sqrt{5}}(b_k - a_k)$.

Шаг 4. Вычислить $f(l_k)$ и $f(r_k)$.

Шаг 5. Сравнить значения $f(l_k)$ и $f(r_k)$:

а) если $f(l_k) \le f(r_k)$, положить $a_{k+1} = a_k$, $b_{k+1} = r_k$ и $l_{k+1} = a_{k+1} + \frac{2}{3+\sqrt{5}}(b_{k+1} - a_{k+1})$, $r_{k+1} = l_k$ (рис. 4, а). Перейти к шагу 6;

б) если $f(l_k) > f(r_k)$, положить $a_{k+1} = l_k$, $b_{k+1} = b_k$ и $l_{k+1} = r_k$, $r_{k+1} = a_{k+1} + \frac{2}{1+\sqrt{5}}(b_{k+1} - a_{k+1})$ (рис. 4, б).

Шаг 6. Вычислить $\Delta = b_{k+1} - a_{k+1}$ и проверить условие окончания:

а) если $\Delta \le \varepsilon$, процесс поиска завершается и $x_* \in [a_{k+1}, b_{k+1}]$. В качестве приближенного решения можно взять середину последнего интервала: $x_* \cong \frac{a_{k+1} + b_{k+1}}{2}$;

б) если $\Delta > \varepsilon$, то положить $k = k+1$ и перейти к шагу 4.

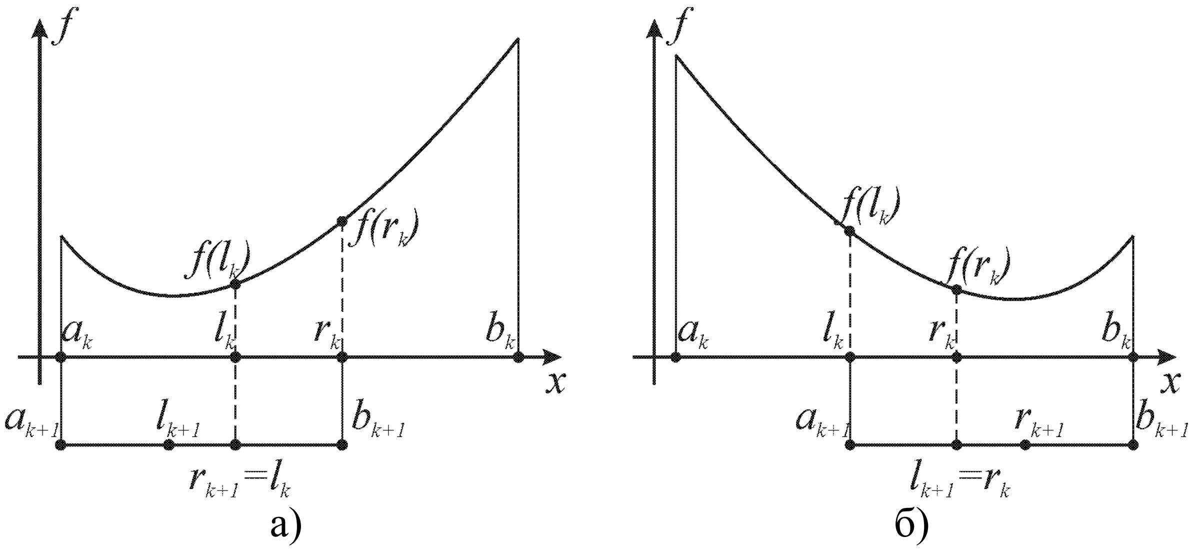

Рис. 4

3.5. Метод Фибоначчи

В методе Фибоначчи реализован алгоритм, обеспечивающий максимальное гарантированное сокращение интервала неопределенности при заданном количестве вычислений функции и претендующий на оптимальность. Этот метод основан на использовании чисел Фибоначчи, задаваемых рекуррентной формулой:

$$F_k = F_{k-1} + F_{k-2},\ k = 2, 3, 4, \ldots,\ F_0 = F_1 = 1.$$

Числа Фибоначчи представляют собой ряд, первые члены которого: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, …

Метод относится к последовательным стратегиям. Задается начальный интервал неопределенности и количество N вычислений функции. Алгоритм уменьшения интервала опирается на анализ значений функции в двух точках. Точки вычисления функции находятся с использованием последовательности из $N+1$ чисел Фибоначчи. Как в методе золотого сечения, на первой итерации требуются два вычисления функции, а на каждой последующей – только по одному. Поиск заканчивается, когда длина текущего интервала неопределенности оказывается меньше установленной величины.

Алгоритм

Шаг 1. Задать начальный интервал неопределенности $L_0 = [a_0, b_0]$, $\varepsilon > 0$ – допустимую длину конечного интервала.

Шаг 2. Найти количество N вычислений функции как наименьшее целое число, при котором удовлетворяется условие $F_N \geq \frac{|L_0|}{\varepsilon}$, и числа Фибоначчи $F_0, F_1, ..., F_N$.

Шаг 3. Положить $k = 0$.

Шаг 4. Вычислить $l_0 = a_0 + \frac{F_{N-2}}{F_N}(b_0 - a_0)$, $r_0 = a_0 + \frac{F_{N-1}}{F_N}(b_0 - a_0)$.

Шаг 5. Вычислить $f(l_k)$, $f(r_k)$.

Шаг 6. Сравнить значения $f(l_k)$ и $f(r_k)$:

а) если $f(l_k) \leq f(r_k)$, положить $a_{k+1} = a_k$, $b_{k+1} = r_k$, $r_{k+1} = l_k$, $l_{k+1} = a_{k+1} + \frac{F_{N-k-2}}{F_{N-k-1}}(b_{k+1} - a_{k+1})$. Перейти к шагу 7;

б) если $f(l_k) > f(r_k)$, положить $a_{k+1} = l_k$, $b_{k+1} = b_k$, $l_{k+1} = r_k$, $r_{k+1} = a_{k+1} + \frac{F_{N-k-2}}{F_{N-k-1}}(b_{k+1} - a_{k+1})$.

Шаг 7. Проверить условие окончания поиска: если $k \neq N-3$, положить $k = k+1$ и перейти к шагу 5, иначе завершить поиск.

В качестве приближенного решения можно взять любую точку последнего интервала, например, его середину: $x_* \cong \frac{a_{N-1} + b_{N-1}}{2}$.

4. Методы полиноминальной аппроксимации

В методах прямого поиска имеется информация только о значениях минимизируемой функции в выбранных точках и предполагается, что она непрерывна и является унимодальной на рассматриваемом отрезке.

Если функцию в некоторой окрестности точки ее минимума можно достаточно точно аппроксимировать многочленом, то для ее минимизации часто используют так называемые *методы полиномиальной аппроксимации*. Их общая особенность состоит в вычислении коэффициентов многочлена по известным значениям функции в отдельных точках и последующем нахождении минимума этого многочлена с использованием необходимых и достаточных условий экстремума.

4.1. Метод квадратичной интерполяции

Рассмотрим метод квадратичной интерполяции, в котором минимизируемая функция приближенно заменяется параболой, проходящей через три точки (x_1, f_1), (x_2, f_2) и (x_3, f_3), где $f_i = f(x_i),\ i = 1,\ 2,\ 3$.

Известно, что через три различные точки, не лежащие на одной прямой можно провести только одну параболу. Аппроксимирующая квадратичная функция записывается в следующем виде:

$$\varphi(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2).$$

Заметим, что значения квадратичной функции $\varphi(x)$ совпадут со значениями минимизируемой функции $f(x)$ в трех указанных точках, то есть $\varphi(x_1) = f_1$, $\varphi(x_2) = f_2$ и $\varphi(x_2) = f_2$. Коэффициенты a_0, a_1 и a_2 квадратичной функции $\varphi(x)$ определяются по известным значениям $f_i = f(x_i),\ i = 1,\ 2,\ 3$ путем решения следующей системы линейных уравнений:

$$f_1 = \varphi(x_1) = a_0;$$

$$f_2 = \varphi(x_2) = a_0 + a_1(x_2 - x_1);$$

$$f_3 = \varphi(x_3) = a_0 + a_1(x_3 - x_1) + a_2(x_3 - x_1)(x_3 - x_2).$$

Данная система легко решается рекурсивным методом. Из первого уравнения находим $a_0 = f_1$. Поскольку $f_2 = f_1 + a_1(x_2 - x_1)$, то $a_1 = \dfrac{f_2 - f_1}{x_2 - x_1}$. Наконец, при $x = x_3$ получаем:

$$f_3 = f_1 + \frac{f_2 - f_1}{x_2 - x_1}(x_3 - x_1) + a_2(x_3 - x_1)(x_3 - x_2).$$

Решая это уравнение относительно a_2, имеем:

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

Если найденные выражения для коэффициентов a_0, a_1 и a_2 подставить в необходимое условие экстремума квадратичной функции $\varphi(x)$: $\varphi'(x) = a_1\left(x - x_2\right) + a_2\left(x - x_1\right) = 0$, то получим ее единственную стационарную точку

$$x_* = \frac{x_1 + x_2}{2} - \frac{a_1}{2a_2}. \qquad (1)$$

На первой итерации метода квадратичной интерполяции при помощи (1) вычисляют точку минимума $x_*^{(1)}$ интерполяционного многочлена и затем $f_*^{(1)} = f\left(x_*^{(1)}\right)$.

Для вычислений на второй итерации из четырех точек $x_*^{(1)}$, x_1, x_2 и x_3 выбирают новую тройку по следующему правилу:

- если $x_*^{(1)} \in [x_2, x_3]$ и $f_*^{(1)} \le f_2$, то $x_1^{(2)} = x_2$, $x_3^{(2)} = x_3$, а $x_2^{(2)} = x_*^{(1)}$.
- если $x_*^{(1)} \in [x_2, x_3]$ и $f_*^{(1)} > f_2$, то $x_1^{(2)} = x_1$, $x_3^{(2)} = x_*^{(1)}$, а $x_2^{(2)} = x_2$.
- если $x_*^{(1)} \in [x_1, x_2]$ и $f_*^{(1)} \le f_2$, то $x_1^{(2)} = x_1$, $x_3^{(2)} = x_2$, а $x_2^{(2)} = x_*^{(1)}$.
- если $x_*^{(1)} \in [x_1, x_2]$ и $f_*^{(1)} > f_2$, то $x_1^{(2)} = x_*^{(1)}$, $x_3^{(2)} = x_3$, а $x_2^{(2)} = x_2$.

Далее, из (1) находят $x_*^{(2)}$, а затем описываемые процедуры повторяют на третьем шаге и так далее, до тех пор, пока длина интервала неопределенности, в котором лежит искомая точка минимума функции $f(x)$, не станет меньше заданного допустимого значения ε.

Описанный метод относится к методам нулевого порядка.

4.2. Метод с использованием кубичной аппроксимации

В соответствии с методом функция $f(x)$, подлежащая минимизации, аппроксимируется полиномом третьего порядка. Логическая схема метода аналогична схеме метода с использованием квадратичной аппроксимации. Однако в данном случае построение аппроксимирующего полинома проводится на основе меньшего числа точек, поскольку в каждой точке можно вычислять значения как функции, так и ее производной.

Выполнение алгоритма начинается в произвольно выбранной точке x_1. Находится точка x_2, такая, что производные $f'(x_1)\cdot f'(x_2)<0$. Таким образом искомая точка x_* находится внутри отрезка $[x_1,\ x_2]$. Аппроксимирующая кубичная функция записывается в следующем виде:

$$\varphi(x)=a_0+a_1(x-x_1)+a_2(x-x_1)(x-x_2)+a_3(x-x_1)^2(x-x_2). \quad (2)$$

Заметим, что значения квадратичной функции $\varphi(x)$ совпадут со значениями минимизируемой функции $f(x)$ в точках x_1, x_2, то есть $\varphi(x_1)=f_1$, $\varphi(x_2)=f_2$ и $\varphi(x_2)=f_2$, где $f_i=f(x_i)$, $i=1,\ 2$. Коэффициенты a_0, a_1, a_2 и a_3 кубичной функции $\varphi(x)$ определяются по известным значениям $f_i=f(x_i)$, $i=1,\ 2$, $f_i'=f'(x_i)$, $i=1,\ 2$ путем решения следующей системы линейных уравнений:

$$f_1=\varphi(x_1)=a_0;$$
$$f_2=\varphi(x_2)=a_0+a_1(x_2-x_1);$$
$$f_1'=\varphi'(x_1)=a_1+a_2(x_1-x_2);$$
$$f_2'=\varphi'(x_2)=a_1+a_2(x_2-x_1)+a_3(x_2-x_1)^2.$$

Данная система легко решается рекурсивным методом:

$$a_0=f_1;$$
$$a_1=\frac{f_1-f_2}{x_1-x_2};$$
$$a_2=\frac{f_1'}{x_1-x_2}-\frac{f_1-f_2}{(x_1-x_2)^2};$$
$$a_3=\frac{f_1'+f_2'}{(x_1-x_2)^2}-2\frac{f_1-f_2}{(x_1-x_2)^3}.$$

После того, как коэффициенты найдены, действуя аналогично методу квадратичной аппроксимации, приходим к решению, определяющему стационарную точку аппроксимирующего кубичного полинома $\varphi(x)$. При этом приравнивание к нулю производной $\varphi(x)$ приводит к квадратному уравнению $\varphi'(x)=0$.

Решение этого уравнения записывается в следующем виде:

$$x_* = \begin{cases} x_2, \text{ если } \mu < 0; \\ x_2 - \mu(x_2 - x_1), \text{ если } 0 \le \mu \le 1; \\ x_1, \text{ если } \mu > 1, \end{cases} \tag{3}$$

где

$$\mu = \frac{f_2' + \omega - z}{f_2' - f_1' + 2\omega}, \; z = \left(\frac{3(f_1 - f_2)}{x_2 - x_1}\right) + f_1' + f_2',$$

$$\omega = \begin{cases} \sqrt{z^2 - f_1' f_2'}, \text{ если } x_1 < x_2, \\ -\sqrt{z^2 - f_1' f_2'}, \text{ если } x_1 > x_2. \end{cases}$$

Формула для вычисления ω обеспечивает надлежащий выбор одного из двух корней квадратного уравнения; для значений μ, заключенных в интервале от 0 до 1, формула (3) гарантирует, что получаемая точка x_* расположена между x_1 и x_2.

Затем снова выбираются две точки для реализации процедуры кубичной аппроксимации – x_* и одна из точек x_1 или x_2, причем значения производной исследуемой функции в этих точках должны быть противоположны по знаку, и процедура кубичной аппроксимации повторяется.

Алгоритм включает следующие шаги.

Шаг 1. Задать начальную точку x_0, величину шага Δ и параметры сходимости ε_1, ε_2 – малые положительные числа.

Шаг 2. Вычислить $f'(x_0)$.

Шаг 3. Проверить знак $f'(x_0)$:

а) если $f'(x_0) < 0$, вычислить $x_{k+1} = x_k + 2^k \Delta$, $k = 0, 1, 2, ...$ вплоть до точки x_M, в которой $f'(x_{M-1}) f'(x_M) \le 0$;

б) если $f'(x_0) > 0$, вычислить $x_{k+1} = x_k - 2^k \Delta$, $k = 0, 1, 2, ...$ вплоть до точки x_M, в которой $f'(x_{M-1}) f'(x_M) \le 0$.

Шаг 4. Положить $x_1 = x_{M-1}$, $x_1 = x_M$. Вычислить значения $f_i = f(x_i), i = 1, 2$, $f_i' = f'(x_i), i = 1, 2$.

Шаг 5. Найти стационарную точку x аппроксимирующего кубического полинома по формуле (3) и значение $f(x)$ в точке минимума $\varphi(x)$.

Шаг 6. Проверить условие убывания:

а) если $f(x) < f(x_1)$, перейти к шагу 7;

б) если $f(x) \ge f(x_1)$, вычислять $x = x + \frac{1}{2}(x - x_1)$ пока не будет выполняться неравенство $f(x) \le f(x_1)$.

Шаг 7. Проверка на окончание поиска: если $|f'(x)| \le \varepsilon_1$ и $\left|\frac{x - x_1}{x}\right| \le \varepsilon_2$, поиск закончить и $x_* \cong x$, иначе положить либо $x_1 = x$, если $f'(x)f'(x_2) < 0$, либо $x_2 = x_1$, $x_1 = x$, если $f'(x)f'(x_1) < 0$ и перейти к шагу 5.

Заметим, что описанный метод относится к методам первого порядка.

Пример. Минимизировать функцию $f(x) = 2x^2 + 16/x$ с использованием кубичной аппроксимации.

Решение.

Шаг 1. Зададим начальную точку $x_0 = 1$, величину шага $\Delta = 1$ и параметры сходимости $\varepsilon_1 = 10^{-2}$, $\varepsilon_2 = 3 \cdot 10^{-2}$.

Шаг 2. Вычислим $f'(x_0) = f'(1) = -12 < 0$.

Шаг 3. Так как $f'(x_0) < 0$, то $x_1 = x_0 + 2^0 \cdot 1 = 1 + 1 = 2$, $f'(x_1) = f'(2) = 4$.

Шаг 4. Так как $f'(1)f'(2) = -48 < 0$, то $x_1 = 1$, $x_2 = 2$

Результаты дальнейших вычислений сведем в таблицу 1.

Таблица 1

№ итерации	x_1	x_2	$f(x_1)$	$f(x_2)$	$f'(x_1)$	$f'(x_2)$	x	$f(x)$
1	1	2	18	16	-12	4	1,5657	15,1219
2	1,5657	2	15,1219	16	-0,264	4	1,5880	15,1191

$f'(1{,}5880) = 0{,}0072 < 10^{-2}$, $\left|\frac{1{,}5880 - 1{,}5657}{1{,}5880}\right| = 0{,}0140 < 3 \cdot 10^{-2}$. Условия окончания поиска выполнены, следовательно, поиск закончен, $x_* \cong 1{,}5880$.

5. Методы минимизации с использованием производных

Если в дополнение к условию непрерывности ввести требование дифференцируемости функции, то эффективность поисковых процедур можно существенно повысить. Необходимое условие существования локального минимума функции f в некоторой точке x – обращение производной функции f в нуль в точке x, т.е. $f'(x)=0$. Классической схемой, ориентированной на нахождение корня нелинейного уравнения, является схема, разработанная Ньютоном и позднее уточненная Рафсоном.

5.1. Метод Ньютона – Рафсона

Предположим, что функция f дважды дифференцируема. Работа алгоритма начинается в точке x_1, которая представляет начальное приближение стационарной точки или корня уравнения $f'(x)=0$. Затем строится линейная аппроксимация функции $f'(x)$ в точке x_1, и точка, в которой аппроксимирующая линейная функция обращается в нуль, принимается в качестве следующего приближения. Если точка x_k принята в качестве текущего приближения к стационарной точке, то линейная функция, аппроксимирующая функцию $f'(x)$ в точке x_k, записывается в виде:

$$\varphi(x,\, x_k)=f'(x_k)+f''(x_k)(x-x_k).$$

Приравнивая правую часть к нулю, получаем следующее:

$$x_{k+1}=x_k-f'(x_k)/f''(x_k). \qquad (5)$$

В зависимости от выбора начальной точки и вида функции алгоритм может как сходиться к истинной стационарной точке, так и расходиться

Пример. Найти минимум функции $f(x)=2x^2+16/x$ методом Ньютона – Рафсона.

Решение.

Зададим значение начальной точки $x_0=1$, величина допустимого отклонения $\varepsilon=10^{-1}$. Результаты вычислений сведем в таблицу 2.

Таблица 2

k	x_k	$f'(x_k)$	$f''(x_k)$	x_{k+1}
0	1	-12	36	1,33
1	1,33	-3,73	17,6	1,54

2	1,54	-0,59	12,76	1,59
3	1,59	0,03		

$|f'(x_3)| = 0{,}03 < \varepsilon$, следовательно, поиск завершен $x_* \cong 1{,}59$.

5.2. Метод средней точки

Для нахождения корня уравнения $f'(x) = 0$ можно воспользоваться эффективным алгоритмом исключения интервалов, на каждой итерации которого рассматривается лишь одна пробная точка. Например, если в точке x выполняется неравенство $f'(x) < 0$, то с учетом предположения об унимодальности естественно утверждать, что точка минимума не может находиться левее точки x. Другими словами, интервал $(-\infty; x]$ подлежит исключению. С другой сторон, если $f'(x) > 0$, то точка минимума не может находиться правее x и интервал $[x; \infty)$ можно исключить. Данная идея лежит в основе метода средней точки, называемого иногда поиском Больцано.

Алгоритм поиска включает следующие шаги.

Шаг 1. Задать интервал $[a, b]$, параметр сходимости ε – положительное малое число, при этом $f'(a) < 0$, $f'(b) > 0$.

Шаг 2. Вычислить $x = \dfrac{a+b}{2}$, $f'(x)$:

а) если $|f'(x)| \le \varepsilon$, закончить поиск и $x_* \cong x$;

б) если $|f'(x)| > \varepsilon$ и $f'(x) < 0$, то $a = x$ и перейти к шагу 3;

в) если $|f'(x)| > \varepsilon$ и $f'(x) > 0$, то $b = x$ и перейти к шагу 3

Метод исключения интервалов основан лишь на исследовании знака производной независимо от значений, которые эта производная принимает.

5.3. Метод секущих

Метод секущих предусматривает рассмотрение как знака производной, так и ее значения и является комбинацией метода Ньютона и общей схемы исключения интервалов. Предположим, что в процессе поиска стационарной точки функции $f(x)$ в интервале $[a, b]$ обнаружены две точки l и r – приближения к искомой точке x_*. В этом случае алгоритм метода секущих позволяет аппроксимировать функцию $f'(x)$ «секущей прямой» (прямой линией, соединяющей две точки $x_{k-1} = l$ и $x_k = r$):

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

и найти такую точку, в которой секущая графика $f'(x)$ пересекает ось абсцисс.

Таким образом, следующее приближение к стационарной точке x_* может быть получено подстановкой в (5) вместо $f''(x)$ выражения $\dfrac{f'(x_{k-1}) - f'(x_k)}{x_{k-1} - x_k}$. В результате получим:

$$\begin{cases} x_{k+1} = x_k - \dfrac{x_k - y_k}{f'(x_k) - f'(y_k)} \cdot f'(x_k); \\ y_k = x_{k-1}. \end{cases}$$

В качестве x_0 выбирается тот конец отрезка $[a, b]$, на котором совпадают знаки $f'(x)$ и $f'''(x)$, а в качестве x_1 – точка пересечения с осью абсцисс хорды, стягивающей дугу графика функции $f'(x)$ на отрезке $[a, b]$:

$$x_1 = \frac{af'(b) - bf'(a)}{f'(b) - f'(a)}.$$

Если $|f'(x)| \le \varepsilon$, поиск завершен, $x_* \cong x$. В отличие от метода средней точки метод секущих в ряде случаев позволяет исключить более половины интервала поиска.

6. Задачи для упражнений

1. Найти минимум функции $f(x) = x^2 - 6x + 14$ на отрезке $[-2; 4]$ методами равномерного поиска, деления интервала пополам, дихотомии.
2. Найти минимум функции $f(x) = x^2 + 6x + 12$ на отрезке $[-4; 1]$ методами равномерного поиска, деления интервала пополам, дихотомии, золотого сечения, Фибоначчи.

3. Найти минимум функции $f(x)$ на отрезке $[a, b]$:

1) $f(x) = e^x + x^2$, $[-1; 0]$;
2) $f(x) = e^x + 1/x$, $[0{,}5; 1]$;
3) $f(x) = e^x - \ln(x)$, $[0{,}3; 1]$;
4) $f(x) = e^x + 1/(1+x)$, $[-0{,}5; 0{,}5]$;
5) $f(x) = tg(x) + 1/x$, $[0{,}5; 1]$;
6) $f(x) = tg(x) + e^{-x} + x$, $[-1; 0]$;
7) $f(x) = x^2 + \sin x$, $[-1; 0]$;
8) $f(x) = e^x - \sin x$, $[0; 1]$;
9) $f(x) = x^4 + 2x^2 + 4x$, $[-1; 0]$.

---

1. Пантелеев, А.В. Методы оптимизации в примерах и задачах: учеб. пособие / А.В. Пантелеев, Т.А. Летова; 2-е изд., исправл.; М.: Высш. шк., 2005. 544 с.: ил.
2. Аттетков, А.В. Методы оптимизации: учебник для вузов / А.В. Аттетков, С.В. Галкин, В.С. Зарубин / Под ред. В.С. Зарубина, А.П. Крищенко; 2-е изд.; М.: Изд-во МГТУ им. Н.Э. Баумана, 2003. 440 с.
3. Реклейтис, Г. Оптимизация в технике / Г. Реклейтис, А. Рейвиндрон, К. Рэгсдел; в 2-х кн.; книга 1; М.: Мир, 1986. 370 с.