

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ХАРЬКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ»

В. П. Северин

МЕТОДЫ ОДНОМЕРНОГО ПОИСКА

Учебно-методическое пособие по курсу
«Методы оптимизации» для студентов
направлений 6.040302 «Информатика»,
6.040303 «Системный анализ»

Утверждено
редакционно-издательским
советом НТУ «ХПИ»,
протокол №2 от 07.12.2011

Харьков
НТУ «ХПИ»
2012

УДК 519.85(075)

ББК 22.18я73

С 28

Рецензенты: *И. Ф. Домнин*, д-р техн. наук, проф., Институт ионосферы НАН Украины;

Н. В. Ткачук, д-р техн. наук, проф., Национальный технический университет «Харьковский политехнический институт»

Розглянуті теорія і методи одновимірного пошуку. Викладена теорія необхідних і достатніх умов екстремуму функції однієї змінної. Приведено обґрунтування побудови, теоретичний опис і алгоритми більш ніж десяти основних методів одновимірного пошуку. Представлені алгоритми дозволяють полегшити вивчення методів і прискорити виконання лабораторних робіт. Дані завдання для лабораторних робіт. Призначено для студентів технічних спеціальностей.

Северин В. П.

С 28 Методы одномерного поиска : учебно-метод. пособ. по курсу «Методы оптимизации» / В. П. Северин. – Х. : НТУ «ХПИ», 2012. – 112 с. – На русск. яз.

ISBN 978-966-5930973-3

Рассмотрены теория и методы одномерного поиска. Изложена теория необходимых и достаточных условий экстремума функции одной переменной. Приведены обоснование построения, теоретическое описание и алгоритмы более десяти основных методов одномерного поиска. Представленные алгоритмы позволяют облегчить изучение методов и ускорить выполнение лабораторных работ. Даны задания для лабораторных работ. Предназначено для студентов технических специальностей.

Ил. 26. Табл. 4. Библиогр.: 22 назв.

УДК 519.85(075)

ББК 22.18я73

ISBN 978-966-5930973-3

© В.П. Северин, 2012 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. ВВЕДЕНИЕ В МЕТОДЫ ОПТИМИЗАЦИИ	7
1.1. Предмет методов оптимизации	7
1.2. Понятия и определения теории оптимизации	9
1.3. Примеры задач оптимизации	11
1.4. Классификация методов оптимизации	15
Лабораторная работа	19
Контрольные вопросы	22
2. ОСНОВЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ	24
2.1. Экстремум функции одной переменной	24
2.2. Условия экстремума функции одной переменной	26
2.3. Унимодальные функции и их свойства	29
2.4. Интервал неопределенности минимума унимодальной функции	32
2.5. Поиск интервала неопределенности методом Свенна	35
Лабораторная работа	38
Контрольные вопросы	41
3. МЕТОДЫ УМЕНЬШЕНИЯ ИНТЕРВАЛА НЕОПРЕДЕЛЕННОСТИ	43
3.1. Метод равномерного поиска	43
3.2. Метод дихотомии	45
3.3. Метод деления интервала пополам	49
3.4. Числа Фибоначчи и их свойства.....	53
3.5. Метод Фибоначчи	57
3.6. Золотое сечение и его свойства	62
3.7. Метод золотого сечения	67
3.8. Метод адаптации шага	70
3.9. Сравнение методов уменьшения интервала неопределенности	76
Лабораторная работа	78
Контрольные вопросы	81

4. МЕТОДЫ АППРОКСИМАЦИИ И ИНТЕРПОЛЯЦИИ	84
4.1. Метод квадратичной интерполяции с тремя точками	84
4.2. Метод кубической интерполяции с четырьмя точками	88
4.3. Метод квадратичной интерполяции с двумя точками	92
4.4. Метод секущих	95
4.5. Метод кубической интерполяции с двумя точками	96
4.6. Метод Ньютона-Рафсона	99
4.7. Сравнение методов аппроксимации и интерполяции	101
Лабораторная работа	104
Контрольные вопросы	107
СПИСОК ЛИТЕРАТУРЫ	110

ВВЕДЕНИЕ

Математические методы оптимизации, их алгоритмы и компьютерные программы являются эффективным элементом наукоемких технологий, разработка которых в настоящее время особо актуальна. Публикуемые книги по методам оптимизации пользуются большим спросом как на прилавках книжных магазинов, так и в библиотеках.

Курс лекций по дисциплине «Методы оптимизации», который в течение ряда лет читался автором на факультете информатики и управления НТУ «ХПИ» для студентов специальностей «Информатика», «Социальная информатика», «Системный анализ и управление», рассчитан на два семестра. В первом семестре изучаются разделы «Методы одномерного поиска» и «Методы безусловной минимизации», во втором – «Методы условной оптимизации» и «Специальные методы оптимизации». В предлагаемом первом модуле дисциплины «Методы оптимизации» дано введение в методы оптимизации, рассмотрены теория и методы одномерного поиска.

Первый раздел посвящен определению и рассмотрению тех основных понятий, которые присущи всему курсу «Методы оптимизации». Представлен предмет методов оптимизации, введены основные понятия теории, приведены примеры задач оптимизации и дана классификация методов. Различается оптимизация в широком и узком смысле, обосновывается необходимость изучения многих методов оптимизации. Вводятся понятия переменных параметров, допустимой области, целевой функции, задачи оптимизации и ее решения, метода оптимизации. Рассматриваются примеры задач оптимизации, показывается полезность их решения и важность применения современных эффективных методов. Представляются различные принципы классификации методов оптимизации.

В большинстве методов многомерной оптимизации используются методы одномерного поиска. Во втором разделе рассматриваются во-

просы, связанные с теорией и практикой одномерного поиска. Излагается теория необходимых и достаточных условий экстремума функции одной переменной. Рассматриваются свойства унимодальной функции, вводится понятие интервала неопределенности точки минимума унимодальной функции, приводятся основные принципы поиска интервала неопределенности и его уменьшения. Обосновывается метод Свена для поиска интервала неопределенности.

Третий раздел посвящен в основном методам уменьшения интервала неопределенности. Обосновываются методы равномерного поиска, дихотомии, деления интервала пополам. Рассматриваются свойства чисел Фибоначчи, и конструируется метод Фибоначчи, изучаются свойства золотого сечения, и обосновывается метод золотого сечения. Представляется метод адаптации шага, который не требует нахождения интервала неопределенности. Выполняется сравнение методов уменьшения интервала неопределенности по их эффективности.

В четвертом разделе рассматриваются методы нулевого, первого и второго порядков, основанные на аппроксимации унимодальной целевой функции полиномом с последующей интерполяцией точки минимума функции. Обосновываются методы квадратичной и кубической интерполяции, метод секущих, одномерный метод Ньютона-Рафсона. Дается сравнение скорости асимптотической сходимости методов аппроксимации и интерполяции.

Приведены реализуемые алгоритмы всех рассмотренных методов одномерного поиска. Представленные алгоритмы позволяют облегчить понимание методов и ускорить выполнение лабораторных работ. Даны задания для лабораторных работ.

Особенностью пособия является более полное изложение методов одномерного поиска по сравнению с большинством учебно-методических изданий. Включен оригинальный материал, например метод адаптации шага, который не изложен в другой учебно-методической литературе.

Для усвоения материала достаточно владения стандартными курсами математического анализа и линейной алгебры.

1. ВВЕДЕНИЕ В МЕТОДЫ ОПТИМИЗАЦИИ

Данный раздел посвящен определению и рассмотрению тех основных понятий, которые будут изучаться во всем курсе «Методы оптимизации». Различается оптимизация в широком и узком смысле, показывается происхождение термина «оптимизация», определяется предмет методов оптимизации, обосновывается необходимость изучения многих методов оптимизации. Вводятся понятия переменных параметров оптимизации, пространства параметров и допустимой области, критерия качества и целевой функции, задачи оптимизации и решения задачи оптимизации, метода оптимизации. Рассматриваются примеры задач оптимизации, показывается полезность их решения и важность применения современных эффективных методов. Рассматриваются различные принципы классификации методов оптимизации и дается классификация методов по применяемым средствам, по математической формулировке задач оптимизации, по порядку производных. Приводится описание лабораторной работы для начала практического изучения методов оптимизации.

1.1. Предмет методов оптимизации

Слово «*оптимальный*» в переводе с греческого *optimus* означает «наилучший», поэтому *оптимизация в широком смысле* означает поиск наилучшего варианта из множества различных альтернатив.

Оптимизация различных объектов и их систем является основой инженерной деятельности, поскольку цель труда любого инженера – это создание новых, более эффективных систем при использовании меньших возможных ресурсов, в том числе финансовых. В настоящее время знание методов оптимизации студентами высшей школы так же необходимо, как и знание основ математического анализа, физики, информатики и других дисциплин.

Для поиска наилучшего варианта некоторой системы необходимо задать для нее критерий качества и определить, с помощью каких параметров на него можно воздействовать. Выбранный критерий качества представляет функцию переменных параметров, которую необходимо минимизировать или максимизировать. *Оптимизация в узком смысле* заключается в минимизации или максимизации функции при наличии или отсутствии ограничений.

Оптимизация может осуществляться различными средствами – как с помощью простой арифметики, так и с помощью сложных аналитических и численных процедур. Существует множество методов оптимизации, самыми мощными из которых являются численные методы, которые наиболее полно используют возможности современной вычислительной техники.

Термин «оптимизация» впервые ввёл в математике Лейбниц. Методы оптимизации начали развиваться после Второй мировой войны, а большинство классических методов оптимизации создано в 50–60-е годы прошлого века. Наравне с понятием методов оптимизации применяют его синоним – *математическое программирование*. С другой стороны, математическое программирование является частью исследования операций – подраздела вычислительной математики, изучающего свойства различных операций.

Операцией в математике называется отображение декартового произведения некоторых множеств в другое множество:

$$M_1 \times M_2 \times \dots M_n \rightarrow M.$$

Распространенным примером операции является функция нескольких переменных $f(x)$, поэтому исследование функций – это часть исследования операций.

Запросы человеческой деятельности ставят множество задач оптимизации различных типов, и даже для разных задач одинакового типа не существует единого универсального метода решения. Это приводит к настоящей необходимости изучения многих методов оптимизации и анализа их достоинств и недостатков с целью наиболее

эффективного применения методов.

Человеческий гений изобрел множество эффективных численных методов оптимизации, которые имеют большую как познавательную, так и практическую ценность. Численные методы для решения различных типовых задач оптимизации и составляют предмет учебной дисциплины «Методы оптимизации».

1.2. Понятия и определения теории оптимизации

Для формального представления задач оптимизации и методов их решения необходимо ввести основные понятия и определения теории оптимизации.

1. *Параметры оптимизации* $x_1, x_2, \dots, x_n$ – это те независимые переменные, которые можно изменять при решении задачи оптимизации, n – число переменных. Эти переменные также называются *переменными* или *варьируемыми параметрами*. Упорядоченный набор переменных образует *вектор переменных параметров* $x = (x_1, x_2, \dots, x_n)$. Будем полагать, что каждый параметр принадлежит множеству вещественных чисел $R = (-\infty, +\infty)$: $x_i \in R, i = \overline{1, n}$.

2. *Пространство параметров* – это множество векторов переменных параметров x , образующих n -мерное пространство R^n . При этом полагают, что пространство параметров представляет декартово произведение n множеств вещественных чисел $R^n = R \times R \times \dots \times R$. Для фиксированного вектора x его можно представить точкой пространства параметров. Размерность пространства параметров n определяется размерностью его вектора x , для чего используется обозначение $n = \dim x$. Здесь $\dim$ – это сокращение от английского слова dimension – размерность.

3. *Допустимая область* – это множество векторов параметров, удовлетворяющих ограничениям оптимизационной задачи, и обозначаемое символом D . Ограничения могут задаваться в виде ра-

венств и неравенств:

$$D = \{ x \mid g_i(x) \leq 0, i = \overline{1, p}; h_j(x) = 0, j = \overline{1, r} \}.$$

При наличии ограничений говорят о задаче условной оптимизации, а при их отсутствии – о задаче безусловной оптимизации.

4. *Критерий качества* – это функционал, отображающий качество оптимизируемой системы. Функционал – это любой оператор, результатом которого является число. Критерий качества обычно представляется функцией вектора переменных параметров $f(x)$, принимающей вещественные значения. Такая функция выполняет математическую операцию отображения пространства R^n в R , т. е. $f: R^n \rightarrow R$.

5. *Целевая функция* – это критерий качества с указанием типа его оптимальности (минимальности или максимальности): $f(x) \rightarrow \min$ или $f(x) \rightarrow \max$. При разработке современного программного обеспечения методов оптимизации для определенности полагают, что целевую функцию необходимо минимизировать. Если необходима максимизация целевой функции $g(x)$ ($g(x) \rightarrow \max$), то переходят к вспомогательной функции $f(x) = -g(x)$, минимизация которой приведет к максимизации исходной функции:

$$f(x) \rightarrow \min \Rightarrow g(x) \rightarrow \max.$$

Если $g(x) \neq 0$, то можно также положить $f(x) = 1/g(x)$.

Таким образом, любую задачу максимизации целевой функции можно свести к соответствующей задаче минимизации.

6. *Задача оптимизации* – это задача, которая заключается в минимизации целевой функции $f(x)$ при отсутствии или наличии дополнительных условий. Если дополнительные условия отсутствуют, то имеет место задача безусловной оптимизации, которая заключается в минимизации целевой функции $f(x)$ и кратко записывается в виде $\min f(x)$. Если имеются дополнительные условия в виде обязательной

принадлежности вектора переменных x допустимой области D , то ставится задача условной оптимизации, которая записывается в виде $\min_{x \in D} f(x)$. При известной целевой функции $f(x)$ и допустимой области D говорят, что задана модель задачи оптимизации.

7. *Решение задачи оптимизации* – это значение вектора переменных параметров x^* , которое удовлетворяет условию задачи оптимизации. Решение задачи безусловной оптимизации определяют равенствами:

$$f(x^*) = \min f(x) \quad \text{или} \quad x^* = \arg \min f(x).$$

Оптимальное значение целевой функции обозначается через $f^* = f(x^*)$.

8. *Метод оптимизации* – это метод для выполнения итерационного процесса решения задачи оптимизации, который позволяет исключить полный перебор вариантов и получить решение задачи x^* более рациональным способом. Любой метод оптимизации формирует последовательность точек $\{x^{(k)}\}$ в пространстве параметров R^n , которая должна сходиться к решению задачи оптимизации: $x^{(k)} \xrightarrow{k \rightarrow \infty} x^*$. Эффективность метода оптимизации определяется как полученной точностью решения задачи оптимизации, так и величиной затраченных на ее решение вычислительных ресурсов.

1.3. Примеры задач оптимизации

Существует огромное число задач оптимизации, которые различаются по степени известности и полезности. Рассмотрим наиболее известные и показательные задачи оптимизации.

1. Задача Дидоны.

По древнему преданию легендарная царица Дидона основала город Карфаген в 825 году до нашей эры на морском побережье север-

ной Африки при ироничном условии вождей местных племен уступить ей столько земли, сколько она сможет охватить бычьей шкурой. Дидона разрежала шкуру на тонкие бечевки, связала их и с учетом морского берега охватила прямоугольный участок, имеющий максимальную площадь и достаточный для основания города. Так была решена одна из древнейших задач оптимизации.

Обозначим через x_1 и x_2 длины сторон прямоугольного участка, площадь которого $S(x) = x_1 x_2$ необходимо максимизировать при ограниченной длине трех сторон $l(x) = 2x_1 + x_2$, $l(x) = C$, так как четвертая сторона прямоугольника была ограничена морем

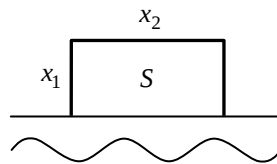


Рис. 1.1. Задача Дидоны

(рис. 1.1). Таким образом, необходимо максимизировать целевую функцию $f(x) = S(x)$ при ограничении-равенстве $2x_1 + x_2 = C$. Заметим, что решение этой задачи позволило минимизировать количество строительного материала для построения оборонительной стены вокруг города при наибольшей возможной его площади.

Задача Дидоны является задачей нелинейного программирования, поскольку в ней целевая функция нелинейная относительно переменных.

2. Задача оптимизации плана производства.

Необходимо составить план производства n видов продукции с использованием m типов ресурсов так, чтобы получить максимальную прибыль предприятия. Введем обозначения: a_{ij} – норма расхода i -го ресурса для производства единицы j -й продукции, $i = \overline{1, m}$, $j = \overline{1, n}$; b_i – запас i -го ресурса, $i = \overline{1, m}$; c_j – норма прибыли от продажи единицы j -й продукции, $j = \overline{1, n}$; x_j – планируемое количество j -го вида продукции, $j = \overline{1, n}$. Общая стоимость производимой продукции

определяется целевой функцией

$$f(x) = \sum_{j=1}^n c_j x_j \rightarrow \max ,$$

что должно быть достигнуто выбором n значений $x_j \geq 0$, которые являются переменными параметрами. Но при этом должны быть учтены ограничения наличия ресурсов:

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = \overline{1, m}.$$

Здесь целевая функция и функции ограничений являются линейными, поэтому такая задача является задачей линейного программирования.

3. Транспортная задача.

Необходимо составить план перевозок некоторого товара с m складов в n магазинов так, чтобы затраты на эти перевозки были минимальными. Введем обозначения: a_i – количество единиц товара на i -м складе, $i = \overline{1, m}$; b_j – количество единиц товара, заказанное j -м магазином, $j = \overline{1, n}$; c_{ij} – стоимость перевозки единицы товара с i -го склада в j -й магазин; x_{ij} – планируемое количество товара, перевозимое с i -го склада в j -й магазин. Общая стоимость перевозок определяется целевой функцией

$$f(x) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min ,$$

минимум которой должен быть достигнут выбором mn значений переменных параметров $x_{ij} \geq 0$. При этом должны быть учтены ограничения наличия товара на складах и потребностей магазинов:

$$\sum_{j=1}^n x_{ij} \leq a_i, \quad i = \overline{1, m}; \quad \sum_{i=1}^m x_{ij} = b_j, \quad j = \overline{1, n}.$$

Целевая функция и функции ограничений являются линейными, поэтому транспортная задача также является задачей линейного программирования.

4. Задача минимизации стоимости ракеты.

При создании трехступенчатой ракеты для реализации космических программ США в 60-х годах прошлого века разработана нелинейная целевая функция ее стоимости в миллионах долларов $f(x) \rightarrow \min$, зависящая от 25 переменных параметров, включающих значения масс элементов ракеты и топлива, тяговые параметры, времена работы ступеней, а также другие конструктивные и технологические параметры. На переменные параметры наложены 26 ограничений: 9 равенств и 17 неравенств, 10 ограничений представлены нелинейными функциями. Таким образом, эта задача условной оптимизации является задачей нелинейного программирования. Американскими учеными Б. С. Рашом, Дж. Браконом и Г. П. МакКормиком получено решение задачи методом последовательной безусловной минимизации со значением целевой функции $f^* = 2330$. Однако, последующие исследования показали, что целевая функция многоэкстремальная, и на самом деле был найден ее локальный минимум. Глобальный минимум с $f^* = 1359$ был вычислен методом Ψ -преобразования В. К. Чичинадзе, специально предназначенным для решения многоэкстремальных задач оптимизации. Таким образом, применение более эффективного метода оптимизации позволило бы сэкономить 971 миллион долларов, то есть 42 % от первоначально рассчитанной стоимости ракеты.

5. Задача коммивояжера.

Имеется граф со взвешенными ребрами. Вершины соответствуют городам, а веса ребер – стоимости перемещения между городами. Требуется найти оптимальный путь коммивояжера, который представляет замкнутый цикл, включающий все вершины и имеющий минимальный

вес. Эта задача является задачей дискретной оптимизации, поскольку переменными являются вершины графа – дискретные значения.

6. Задача оптимизации системы управления.

Модель системы автоматического управления представляется системой дифференциальных уравнений:

$$\frac{dX}{dt} = F(x, X, u), \quad y = CX,$$

где x – вектор переменных параметров системы, X – вектор ее состояния, u – внешнее воздействие, y – управляемая величина. При ступенчатом изменении $u = 1(t)$, где

$$1(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0, \end{cases}$$

система переходит в новое состояние. Характер этого перехода определяется функцией управляемой величины $y = y(x, t)$, по которой определяют прямые показатели качества системы управления: перерегулирование $\sigma(x)$, показатель колебаний $\zeta(x)$ и время регулирования $t_c(x)$. На первые два показателя накладываются ограничения $\delta(x) \leq \delta_m$, $\lambda(x) \leq \lambda_m$, а третий показатель минимизируется $t_c(x) \rightarrow \min$. Данная задача является задачей нелинейного программирования.

1.4. Классификация методов оптимизации

Для успешного применения методов оптимизации необходимо четко представлять их назначение и возможности. Существует множество методов оптимизации, которые можно классифицировать по различным признакам. Рассмотрим основные признаки классификации.

I. По применяемым средствам.

1. Аналитические методы. Используют классические методы

дифференциального исчисления. Для решения задач безусловной оптимизации применяется необходимое условие экстремума – равенство нулю производных целевой функции:

$$\frac{\partial f}{\partial x_i} = 0, \quad i = \overline{1, n}.$$

Для решения задач условной оптимизации используется метод множителей Лагранжа. Аналитические методы дают возможность получить точное решение задачи оптимизации. Общий их недостаток в том, что они могут применяться только для простых задач с дифференцируемыми целевыми функциями, заданными аналитически, при небольшом числе переменных.

2. Методы перебора вариантов. Анализируются различные варианты решения задачи оптимизации и путем их сравнения выбирается наилучший вариант. Недостаток этих методов состоит в их большой трудоемкости и малой точности.

3. Графические методы. Основаны на построении графиков функций одной или нескольких переменных. Экстремум функции определяется непосредственно путем анализа ее графика. Они просты и наглядны, сразу показывают, существует ли решение. Применимы только для малого числа переменных.

4. Численные методы. Самый широкий класс методов, наиболее полно использующий возможности современной вычислительной техники. Это итерационные методы, которые заключаются в последовательном переходе от одного решения $x^{(k)}$ к лучшему решению $x^{(k+1)}$. В отличие от аналитических методов, с помощью которых получают точное решение, численные методы дают приближенное решение.

5. Экспериментальные методы. В некоторых задачах построение модели задачи оптимизации затруднительно, поэтому в них оптимальное решение x^* получают методом оптимального эксперимента – как и в численных методах, переход к лучшему решению проводится итерационно, однако значения самой целевой функции $f(x)$ определяют

ся экспериментально. Это самые дорогостоящие методы.

Для практических приложений наиболее эффективны численные методы оптимизации.

II. По математической формулировке задачи оптимизации.

1. Методы одномерного поиска. Заключаются в нахождении $\min f(x)$, $\dim x = 1$.

2. Методы безусловной минимизации. Заключаются в нахождении $\min f(x)$, $\dim x \geq 2$.

3. Методы условной оптимизации. Необходимо найти $\min_{x \in D} f(x)$,
где

$$D = \{x \mid g_i(x) \leq 0, i = \overline{1, p}; \quad h_j(x) = 0, j = \overline{1, r}\}.$$

Если хотя бы одна из функций задачи оптимизации нелинейная, то методы называются методами нелинейного программирования.

4. Методы выпуклой оптимизации. Формулировка задачи аналогична пункту 3, однако $f(x)$ и D являются выпуклыми. В этом случае решение x^* существует и единственно.

5. Методы линейного программирования. Формулировка задачи аналогична пункту 3, только все функции линейны.

6. Методы минимизация суммы квадратов. Ищется минимум функции

$$S(x) = \sum_{k=1}^m [f_k(x)]^2,$$

представляющей сумму квадратов нелинейных функций $f_k(x)$, $k = \overline{1, m}$. Если $f_k(x)$ – линейные функции, то применяется метод наименьших квадратов, который сводит решение задачи оптимизации к решению системы линейных алгебраических уравнений.

7. Методы дискретного программирования. Задачи оптимизации, решаемые этими методами, характеризуется тем, что переменные па-

параметры x_i , $i = \overline{1, n}$ могут принимать только дискретные значения. Если x_i принимают только целые значения, то соответствующая задача является задачей целочисленного программирования. Она решается соответствующими методами целочисленного программирования. Данный класс методов широко использует дискретную математику и различные комбинаторные схемы.

III. По порядку производных.

Численные методы оптимизации в свою очередь классифицируют по максимальному порядку применяемых в этих методах производных целевой функции.

1. Методы нулевого порядка (методы прямого поиска). В этих методах для оптимизации целевой функции используются только значения самой функции, а значения ее производных не используются. Преимущество методов нулевого порядка заключается в сравнительно простой модификации для решения широкого круга различных задач оптимизации.

2. Методы первого порядка. Эти методы применяются для оптимизации дифференцируемых целевых функций и основаны на использовании первых производных функции. Первые производные задаются градиентом целевой функции, представляющим вектор-столбец ее частных производных по переменным параметрам

$$\text{grad } f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T,$$

где T – знак транспонирования. Поэтому данные методы часто называются градиентными методами. Для градиента также широко используется обозначение $\nabla f(x) = \text{grad } f(x)$.

3. Методы второго порядка. Эти методы применяются для минимизации дважды дифференцируемых целевых функций и основаны на использовании матрицы вторых частных производных целевой функции – матрицы Гессе

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Матрица Гессе или гессиан также обозначается $\nabla^2 f(x) = H(x)$.

Лабораторная работа

Тема лабораторной работы.

Разработка общих подпрограмм для методов одномерного поиска.

Цель лабораторной работы.

Разработать общие подпрограммы, которые позволяют отображать на экране компьютера процесс минимизации функции одной переменной в виде таблицы и графика.

Порядок выполнения лабораторной работы.

1. Получить у преподавателя вариант целевой функции $f(x)$ одной переменной $x \in R$ с начальной точкой поиска минимума $x_0 \in R$.

2. Аналитически найти точку $x^* \in R$ минимума заданной функции $f(x)$ и вычислить минимальное значение функции $f^* = f(x^*)$.

3. Написать подпрограмму вычисления значений функции $f(x)$ с входным параметром x и выходным параметром f .

4. Протестировать подпрограмму вычисления функции $f(x)$ при вычислении значений $f_0 = f(x_0)$ и $f^* = f(x^*)$, сравнить f^* с f_0 .

5. Написать подпрограмму итерационного процесса приближения к точке минимума x^* функции $f(x)$ из начальной точки x_0 в соответствии с требованиями:

1) итерации должны выполняться по рекуррентным формулам:

$$h_k = x^* - x_k, \quad x_{k+1} = x^* + \alpha h_k, \quad f_k = f(x_k), \quad k = 0, 1, 2, \dots,$$

где $\alpha \in (0; 1)$ – заданное число;

2) итерации продолжаются до тех пор, пока выполняются неравенства:

$$|h_k| > \varepsilon, \quad k \leq N,$$

где ε – допустимая погрешность, N – предельное число точек;

3) входные параметры подпрограммы: $f(x)$, x_0 , x^* , α , ε , N ;

4) выходные параметры подпрограммы: X – массив всех точек поиска x_k , F – массив соответствующих значений функции $f_k = f(x_k)$;

5) на каждой итерации выводить на экран компьютера строку, содержащую номер точки поиска k , шаг поиска h_k , значение функции f_k и значение переменной x_k ;

6) после окончания процесса оптимизации на экране должна отображаться таблица, имеющая соответствующий заголовок и представляющая процесс минимизации функции;

7) под таблицей необходимо отобразить количество вычислений целевой функции, конечный шаг, конечное значение функции и соответствующее ему значение независимой переменной.

6. Написать подпрограмму отображения итерационного процесса на двумерном графике функции $f(x)$ в соответствии с требованиями:

1) входные параметры подпрограммы: $f(x)$, X , F ;

2) определить границы интервала оси абсцисс:

$$a = \min\{X_k\}, \quad b = \max\{X_k\};$$

3) отобразить на экране график функции $f(x)$ на интервале $[a, b]$;

4) на построенном графике функции отобразить все последовательные точки поиска (x_k, f_k) , $k = 0, 1, 2, \dots$, соединенные отрезками прямой линии;

5) на том же графике отобразить начальную и конечную точки поиска маркерами, различными по форме и цвету.

7. Написать общую программу для выполнения итерационного процесса и его отображения на экране в соответствии с требованиями:

1) задать в программе значения x_0 , x^* , α , ε , N ;

2) путем вызова подпрограммы итерационного процесса с использованием подпрограммы вычисления функции $f(x)$ и заданных значений x_0 , x^* , α , ε , N реализовать итерационный процесс с построением таблицы процесса на экране и получением массивов значений переменной X и функции F ;

3) с помощью подпрограммы графического отображения итерационного процесса для минимизации функции $f(x)$ по массивам X и F отобразить график процесса на экране компьютера.

8. Запустить на счет общую программу итерационного процесса и выполнить анализ полученных результатов:

1) на основании табличного и графического представления итерационного процесса сделать вывод о его сходимости к точке минимума функции;

2) установить, является ли последовательность значений функции монотонно убывающей последовательностью;

3) сделать вывод о правильности работы общей программы и всех используемых ею подпрограмм.

9. Ответить на вопросы преподавателя по теме лабораторной работы.

10. Оформить и сдать отчет о проведенной лабораторной работе.

С о д е р ж а н и е о т ч е т а :

1) титульный лист установленного образца с указанием министерства и учебного заведения, названия учебной дисциплины, темы рабо-

ты, номера варианта, фамилии и инициалов исполнителя и принимающего, города, года;

2) введение с кратким описанием предмета методов оптимизации и целью лабораторной работы;

3) описание целевой функции, значение начальной точки поиска, аналитическое решение задачи минимизации целевой функции, точка минимума и минимальное значение целевой функции;

4) теоретическое описание итерационного процесса минимизации целевой функции;

5) распечатки используемых компьютерных программ;

6) табличное представление итерационного процесса оптимизации и его итоги – количество вычислений целевой функции, конечный шаг, конечное значение функции и соответствующее ему значение независимой переменной;

7) двумерный график итерационного процесса со всеми точками поиска;

8) анализ полученных результатов;

9) выводы о проделанной лабораторной работе.

Контрольные вопросы

1. Что означает слово «оптимальный»?

2. Почему недопустимы выражения «самый оптимальный», «оптимальнее», «более оптимальный»?

3. Что понимают под оптимизацией в широком смысле?

4. Что понимают под оптимизацией в узком смысле?

5. Какие методы оптимизации наиболее полно используют возможности современной вычислительной техники?

6. Кто впервые ввёл в математике термин «оптимизация»?

7. Приведите синонимы понятия методов оптимизации.

8. Почему необходимо изучать множество методов оптимизации, а не один метод?

9. Что составляет предмет учебной дисциплины «Методы оптими-

зации»?

10. Дайте определения параметров оптимизации и вектора переменных параметров.

11. Что такое пространство параметров и какова его размерность?

12. Дайте определение допустимой области.

13. Что представляет критерий качества?

14. Дайте определение целевой функции.

15. В чем заключается задача оптимизации?

16. Дайте определение решения задачи оптимизации.

17. Что представляет собой метод оптимизации?

18. Приведите примеры задач оптимизации.

19. В чем заключается полезность решения задач оптимизации?

20. Какие основные признаки классификации методов оптимизации?

21. Приведите классификацию методов оптимизации по применяемым средствам.

22. Приведите классификацию методов оптимизации по математической формулировке оптимизационной задачи.

23. Приведите классификацию методов оптимизации по порядку используемых производных целевой функции.

2. ОСНОВЫ ОДНОМЕРНОЙ ОПТИМИЗАЦИИ

Одномерный поиск используется в большинстве методов многомерной оптимизации и во многом определяет их практическую эффективность. В данном разделе рассматриваются вопросы, связанные с основными положениями теории одномерного поиска. Даются определения минимума, максимума и экстремума функции одной переменной, приводятся свойства экстремума. Формулируются и доказываются необходимые и достаточные условия экстремума функции одной переменной. Вводится понятие унимодальной функции, рассматриваются ее свойства, даются понятия выпуклой и вогнутой функций. Вводится понятие интервала неопределенности точки минимума унимодальной функции, приводятся основные принципы поиска интервала неопределенности и его уменьшения. Обосновывается метод Свенна для поиска интервала неопределенности и дается алгоритм метода. Приводится описание лабораторной работы по нахождению интервала неопределенности точки минимума функции одной переменной.

2.1. Экстремум функции одной переменной

Методы одномерного поиска предназначены для вычисления экстремума целевой функции $f(x)$ одной переменной $x \in R$.

Если в некоторой точке x^* функция $f(x)$ принимает значение $f^* = f(x^*)$, наименьшее по сравнению с ее значениями в некоторой окрестности этой точки $C_\varepsilon(x^*) = (x^* - \varepsilon, x^* + \varepsilon)$, где $\varepsilon > 0$, то точка x^* называется *точкой минимума* функции $f(x)$, а число f^* называют просто *минимумом* функции (рис. 2.1).

Если же для всех $x \in C_\varepsilon(x^*)$ и $x \neq x^*$ будет $f(x) > f(x^*)$, то в

точке x^* функция $f(x)$ имеет максимум f^* . Минимум и максимум имеют общее название *экстремум*.

Для исследования функции $f(x)$ на экстремум используют понятие производной этой функции

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

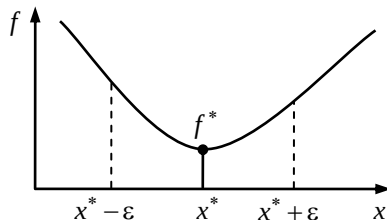


Рис. 2.1. Минимум функции

Основанное на определении производной полезное свойство экстремума дается теоремой Ферма.

Теорема Ферма. Если функция $f(x)$ в некоторой внутренней точке x^* интервала X принимает в этом интервале наименьшее или наибольшее значение и дифференцируема в точке x^* , то $f'(x^*) = 0$.

Доказательство. Пусть для определенности функция $f(x)$ в точке x^* принимает наименьшее в интервале X значение. Возьмем такое приращение Δx , что $x^* + \Delta x \in X$. Составим отношение

$$d(\Delta x) = \frac{f(x^* + \Delta x) - f(x^*)}{\Delta x}.$$

Его числитель всегда положителен. Поэтому, если $\Delta x > 0$, то $d(\Delta x) > 0$, а значит и $\lim_{\Delta x \rightarrow +0} d(\Delta x) \geq 0$, то есть $f'(x^* + 0) \geq 0$. Если же

$\Delta x < 0$, то $d(\Delta x) < 0$, а значит $\lim_{\Delta x \rightarrow -0} d(\Delta x) \leq 0$, то есть $f'(x^* - 0) \leq 0$.

Поскольку функция $f(x)$ дифференцируема в точке x^* , из неравенств $f'(x^* + 0) \geq 0$ и $f'(x^* - 0) \leq 0$ следует, что $f'(x^* + 0) = f'(x^* - 0) = 0$, то

есть $f'(x^*) = 0$, что и требовалось доказать.

Геометрический смысл теоремы Ферма состоит в том, что касательная к графику функции, проведенная в точке экстремума, параллельна оси абсцисс, что показано на рис. 2.2.

Точки x , в которых выполняется равенство $f'(x) = 0$, называются *стационарными* или *критическими точками*.

Функция, имеющая несколько минимумов или максимумов, называется *многоэкстремальной*. Наименьший из всех минимумов многоэкстремальной функции называется *глобальным минимумом*, а остальные минимумы называются *локальными*

(рис. 2.3). При этом глобальный минимум можно определить путем нахождения всех локальных минимумов и выбора наименьшего из них. Аналогично определяются *глобальный* и *локальные максимумы*. Рассматриваемые в данном пособии методы одномерного поиска предназначены для вычисления локального минимума целевой функции.

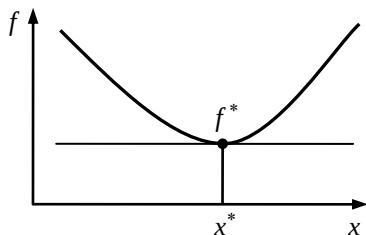


Рис. 2.2. Смысл теоремы Ферма

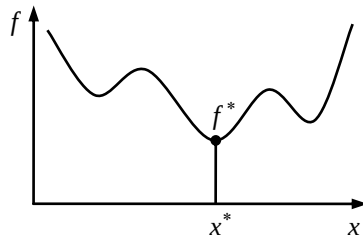


Рис. 2.3. Глобальный минимум

2.2. Условия экстремума функции одной переменной

Необходимый признак существования экстремума функции одной переменной $f(x)$ в точке x^* представляется следующей теоремой.

Теорема 1. Если функция $f(x)$ имеет в точке x^* экстремум,

то ее производная в этой точке либо равна нулю, либо не существует.

Доказательство. Предположим, что функция $f(x)$ имеет в точке x^* экстремум. Тогда существует такой интервал, содержащий внутри себя точку x^* , что $f^* = f(x^*)$ есть или наибольшее, или наименьшее значение $f(x)$ в этом интервале. При этом возможны два случая.

1. Функция $f(x)$ дифференцируема в точке x^* . Тогда в силу теоремы Ферма $f'(x^*) = 0$ (рис. 2.2).

2. Функция $f(x)$ не дифференцируема в точке x^* . В этом случае $f'(x^*)$ не существует (рис. 2.4).

Достаточным этот признак не является, так как равенство нулю производной или ее отсутствие в данной точке еще не означает, что эта точка является экстремальной.

Необходимым и достаточным признаком существования экстремума функции $f(x)$ в точке x^* является из-

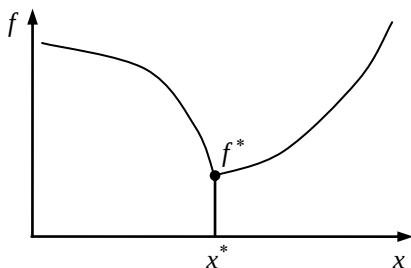


Рис. 2.4. Функция не дифференцируема

менение знака производной $f'(x)$ при переходе через точку x^* (рис. 2.5). Если при возрастании переменной x производная $f'(x)$ в точке x^* меняет знак с минуса на плюс, то это точка минимума функции (рис. 2.5, а). Если же знак $f'(x)$ в точке x^* меняется с плюса на минус, то это точка максимума (рис. 2.5, б). Поэтому имеет место следующая теорема.

Теорема 2. Для того, чтобы дифференцируемая функция $f(x)$ имела в точке x^* экстремум, необходимо и достаточно, чтобы

ее производная $f'(x)$ при переходе через эту точку меняла знак.

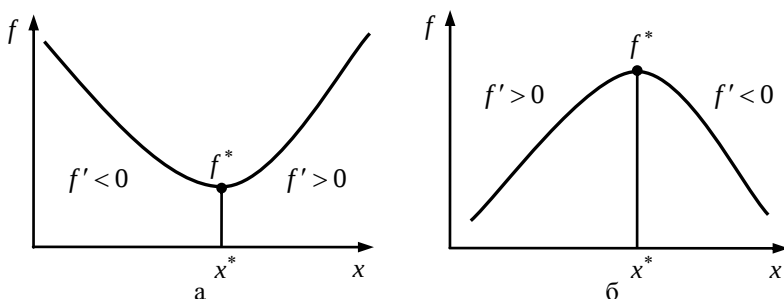


Рис. 2.5. Необходимый и достаточный признак экстремума

Если же $f'(x^*) = 0$ или $f'(x^*)$ не существует, но при переходе через точку x^* знак $f'(x)$ не меняется, то экстремум в точке x^* не существует. Примером является точка перегиба графика функции с горизонтальной касательной (рис. 2.6).

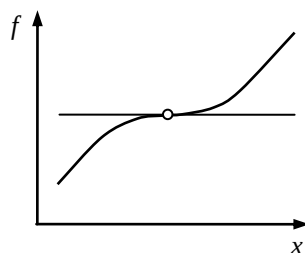


Рис. 2.6. Точка перегиба

Исследовать функцию на экстремум можно и с помощью второй производной.

Теорема 3. Пусть в точке x^* функция $f(x)$ дважды дифференцируема и $f'(x^*) = 0$, $f''(x^*) \neq 0$. Тогда функция $f(x)$ в точке x^* имеет экстремум: минимум, если $f''(x^*) > 0$, и максимум, если $f''(x^*) < 0$.

Доказательство. Пусть $f''(x^*) > 0$. Это значит, что

$$\lim_{x \rightarrow x^*} \frac{f'(x) - f'(x^*)}{x - x^*} > 0.$$

А так как $f'(x^*) = 0$, то

$$\lim_{x \rightarrow x^*} \frac{f'(x)}{x - x^*} > 0.$$

Отсюда следует, что при x достаточно близких к x^* также будет

$$\frac{f'(x)}{x - x^*} > 0.$$

Пусть $x < x^*$, тогда $x - x^* < 0$, а значит и $f'(x) < 0$. Если же $x > x^*$, то $x - x^* > 0$, а значит и $f'(x) > 0$. Итак, при переходе через точку x^* производная $f'(x)$ меняет знак с минуса на плюс, а, значит, функция $f(x)$ имеет в этой точке минимум. Аналогично доказывается существование максимума.

Доказанная теорема 3 является более «узкой», чем предыдущая теорема 2, так как в точке экстремума даже $f'(x)$ может не существовать, а тогда тем более не существует $f''(x)$. В подобных случаях можно применять только теорему 2, которая применима и в том случае, когда $f'(x^*) = 0$ и $f''(x^*) = 0$.

Все приведенные условия существования экстремумов справедливы как для локальных экстремумов, так и для глобальных. Поэтому они не позволяют отличить глобальный экстремум от локального экстремума.

2.3. Унимодальные функции и их свойства

Существуют функции, для которых условия оптимальности дают возможность определить глобальный экстремум. Рассмотрим ряд определений.

Функция $f(x)$ называется *строго убывающей* в интервале $X \subseteq R$, если для любых $u, v \in E$, $u < v$ выполняется неравенство

$f(u) > f(v)$ (рис. 2.7, а). Для дифференцируемой убывающей функции $f'(x) < 0$.

Функция $f(x)$ называется *строго возрастающей* в интервале $X \subseteq R$, если для любых $u, v \in E$, $u < v$ выполняется неравенство $f(u) < f(v)$ (рис. 2.7, б). Для дифференцируемой возрастающей функции $f'(x) > 0$.

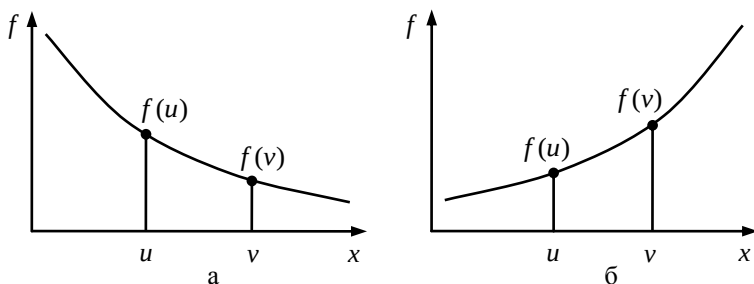


Рис. 2.7. Убывающая и возрастающая функции

Функция $f(x)$ называется *унимодальной*, если существует единственная точка ее минимума x^* и слева от x^* эта функция строго убывает, а справа от x^* строго возрастает (рис. 2.8, а, б). Это означает, что для любых $u, v \in R$, $u < v < x^*$, выполняется неравенство $f(u) > f(v)$, а для любых $u, v \in R$, таких что $x^* < u < v$, выполняется неравенство $f(u) < f(v)$. Для дифференцируемой унимодальной функции $f'(x) < 0$ для любого $x < x^*$ и $f'(x) > 0$ при любом $x > x^*$ (рис. 2.5, а).

Унимодальная функция не обязательно непрерывна (рис. 2.8, б).

Функция $f(x)$ называется *выпуклой* в интервале $X \subseteq R$, если для всех $u, v \in X$ и $\lambda \in [0; 1]$ выполняется неравенство

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda)f(v). \quad (2.1)$$

Если это неравенство является строгим для всех $u, v \in R$, $u \neq v$ и $\lambda \in (0; 1)$, то функция $f(x)$ называется *строго выпуклой*.

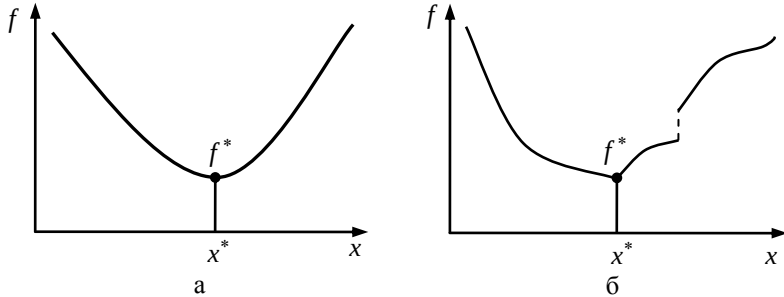


Рис. 2.8. Унимодальные функции

Возьмем фиксированные u , v , $\lambda \in (0; 1)$, и определим $w = \lambda u + (1-\lambda)v$, $y = f(w)$, $z = \lambda f(u) + (1-\lambda)f(v)$. Точка (w, y) лежит на графике функции, а точка (w, z) – на графике секущей (рис. 2.9, а). Неравенство (2.1) примет вид $y \leq z$. Таким образом, геометрически выпуклость функции $f(x)$ означает, что любая точка произвольной хорды графика $f(x)$ располагается не ниже соответствующей точки самого графика. Это также означает, что любая касательная к графику функции $f(x)$ расположена не выше графика самой функции. Если функция $f(x)$ дважды дифференцируемая и строго выпуклая, то $f''(x) > 0$.

Функция $f(x)$ называется (строго) *вогнутой*, если $-f(x)$ (строго) выпукла (рис. 2.9, б). Для дифференцируемой строго вогнутой функции $f''(x) < 0$.

Если строго выпуклая функция имеет минимум, то этот минимум

единственный. Такая функция является унимодальной. Однако определению унимодальной функции могут удовлетворять и функции, не являющиеся непрерывными и выпуклыми.

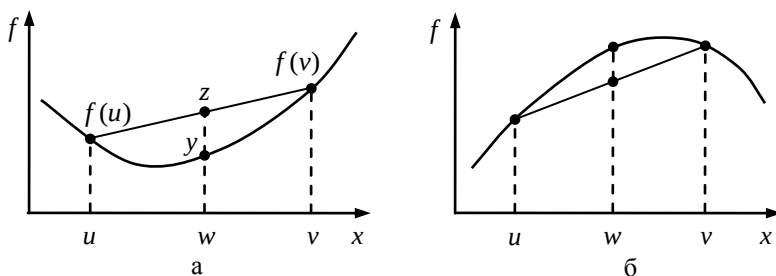


Рис. 2.9. Выпуклая и вогнутая функции

Если функция обладает свойством унимодальности, то локальный минимум является и глобальным минимумом. Если же функция не является унимодальной, то возможно наличие нескольких локальных минимумов.

Почти все методы одномерного поиска, рассматриваемые в этом пособии, предназначены для численной минимизации унимодальных функций.

2.4. Интервал неопределенности минимума унимодальной функции

Интервал $[a, b] \subset R$, содержащий неизвестную точку минимума x^* функции $f(x)$, называется *интервалом неопределенности*.

Таким образом, если известно, что $x^* \in [a, b]$, но само значение x^* не известно заранее, то $[a, b]$ – интервал неопределенности.

Одномерный поиск минимума целевой функции обычно состоит из двух основных этапов:

- 1) поиск интервала неопределенности;
- 2) уменьшение интервала неопределенности.

Поиск интервала неопределенности заключается в том, что из некоторой начальной точки последовательно осуществляются шаги до тех пор, пока значения функции в пробных точках не перестанут уменьшаться. Если найдены такие три различные точки унимодальной функции $f(x)$, что значение функции во внутренней точке не превосходит ее значений в крайних точках, то крайние точки являются границами интервала неопределенности. Пусть, например, при $a < c < b$ выполняются неравенства $f(a) > f(c)$, $f(c) \leq f(b)$. Тогда $[a, b]$ – интервал неопределенности, что показано на рис. 2.10.

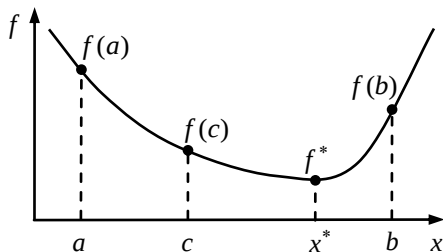


Рис. 2.10. Интервал неопределенности

Методы уменьшения интервала неопределенности основаны на следующей теореме.

Теорема. Пусть $[a, b]$ – интервал неопределенности унимодальной функции $f(x)$, минимум которой f^* достигается в точке x^* . Пусть u, v – такие внутренние точки интервала неопределенности, что $a < u < v < b$, $f_u = f(u)$ и $f_v = f(v)$ – значения функции в этих точках. Тогда, если $f_u < f_v$, то $x^* \in [a, v]$, если же $f_u \geq f_v$, то $x^* \in [u, b]$.

Доказательство. Рассмотрим случай, когда $f_u < f_v$. Предположим, что утверждение теоремы неверно, то есть $x^* \in (v, b]$. Поскольку x^* – точка минимума унимодальной функции $f(x)$, то по определению левее точки x^* функция монотонно убывает. Это означа-

ет, что для $u < v < x^*$ должно выполняться неравенство $f_u > f_v$, что противоречит условию. Аналогично доказывается утверждение для случая $f_u \geq f_v$.

Эту теорему называют *правилом исключения интервалов*. Действительно, из нее следует, что если выполняется неравенство $f_u < f_v$, то полуоткрытый интервал $(v, b]$ заведомо не содержит точку минимума x^* и его следует исключить, приняв в качестве нового интервала неопределенности интервал $[a, v]$ меньшей длины, чем исходный интервал $[a, b]$ (рис. 2.11, а). Если же выполняется противоположное неравенство $f_u \geq f_v$, то заведомо не содержит точку минимума x^* полуоткрытый интервал $[a, u)$, и его следует исключить, приняв в качестве нового интервала неопределенности меньший интервал $[u, b]$ (рис. 2.11, б).

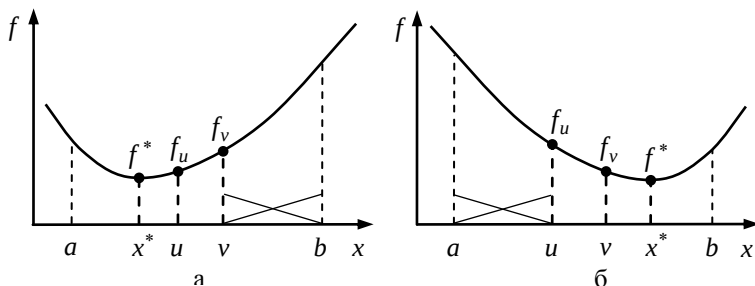


Рис. 2.11. Правило исключения интервалов

Таким образом, по правилу исключения интервалов можно построить процедуру поиска точки минимума целевой функции путем вычисления значений функции во внутренних точках текущего интервала неопределенности и последующего исключения частей этого интервала на основании сравнения значений функции во внутренних точках. Поиск завершается, когда интервал неопределенности умень-

шится до достаточно малых размеров. На этом принципе основывается большинство методов одномерного поиска из следующих разделов.

2.5. Поиск интервала неопределенности методом Свенна

Для поиска интервала неопределенности $[a, b]$ унимодальной функции $f(x)$ Свенном был предложен метод удвоения шага. Для начала работы метода задается начальная точка x_0 и начальный шаг $h > 0$. В точке x_0 вычисляется значение функции $f_0 = f(x_0)$ и выполняется переход к следующей точке $x_1 = x_0 + h$ с вычислением значения функции $f_1 = f(x_1)$. Если $f_1 < f_0$, то выполняются итерации в положительном направлении относительно точки x_0 по рекуррентной формуле удвоения шага

$$x_{k+1} = x_k + 2^k h, \quad k = \overline{1, m} \quad (2.2)$$

с вычислением значений функции $f_k = f(x_k)$ до тех пор, пока не будет получена точка с не меньшим значением функции, чем значение функции в предыдущей

точке: $f_{m-1} > f_m$, $f_m \leq f_{m+1}$. Соответствующий процесс вычислений для значения $m = 2$ представлен на рис. 2.12.

Очевидно, что в этом случае будут определены концы

интервала неопределенности $a = x_{m-1}$, $b = x_{m+1}$ со значениями функции $f_a = f_{m-1}$, $f_b = f_{m+1}$ и внутренняя точка интервала $c = x_m$ со значением функции

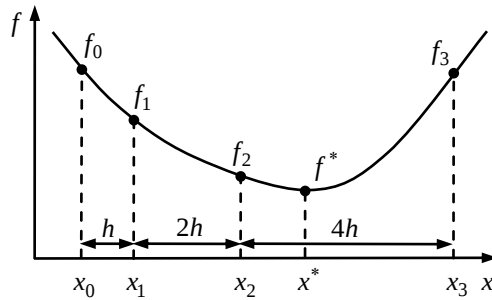


Рис. 2.12. Поиск в положительном направлении

$f_c = f_m$. Таким образом, аналогично рис. 2.10 имеем интервал неопределенности $[a, b]$.

Если после начального шага $f_1 \geq f_0$, то направление поиска меняется на противоположное и определяются $x_2 = x_0 - h$, $f_2 = f(x_2)$. Если $f_2 \geq f_0$, то интервал неопределенности найден и $a = x_2$, $b = x_1$, $f_a = f_2$, $f_b = f_1$, $c = x_0$, $f_c = f_0$ (рис. 2.13).

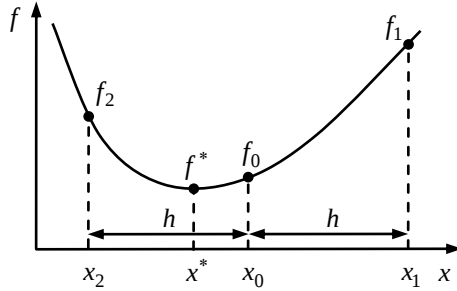


Рис. 2.13. Изменение направления поиска

Если же $f_2 < f_0$, то итерации выполняются в отрицательном направлении по формуле удвоения шага

$$x_{k+1} = x_k - 2^{k-1}h, \quad k = \overline{2, n} \quad (2.3)$$

с вычислением значений функции $f_k = f(x_k)$ до тех пор, пока значения функции не перестанут уменьшаться: $f_{n-1} > f_n$, а $f_n \leq f_{n+1}$, что представлено на рис. 2.14. Таким

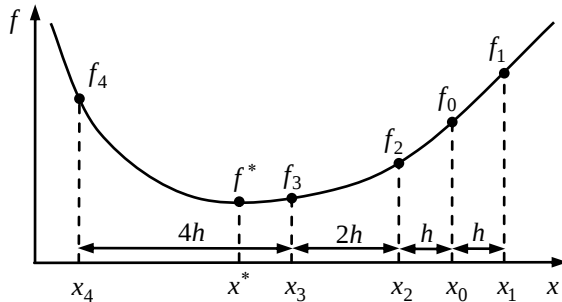


Рис. 2.14. Поиск в отрицательном направлении

образом, будет определен интервал неопределенности со значениями

концов $a = x_{n+1}$, $b = x_{n-1}$, $f_a = f_{n+1}$, $f_b = f_{n-1}$ и внутренняя точка интервала $c = x_n$ со значением функции $f_c = f_n$.

В результате работы метода Свенна устанавливаются значения границ интервала неопределенности. В любом случае интервал неопределенности $[a, b]$ определяется тремя последними точками поиска метода Свенна x_{n-1} , x_n , x_{n+1} . При этом становятся известными значения функции на концах интервала неопределенности $f_a = f(a)$ и $f_b = f(b)$, а также внутренняя точка интервала c и значение функции в ней $f_c = f(c)$ (рис. 2.10).

В методе Свенна используются только значения целевой функции, поэтому это метод нулевого порядка. Метод Свенна с использованием формул (2.2) и (2.3) можно реализовать по следующему алгоритму.

Алгоритм метода Свенна.

Входные параметры: c – начальная точка поиска; h – начальный шаг.

Выходные параметры: a и b – границы интервала неопределенности; f_a и f_b – значения функции на концах интервала; c и f_c – внутренняя точка интервала и значение в ней функции.

1. Вычислить $f_c = f(c)$, $b = c + h$, $f_b = f(b)$.
2. Если $f_b \geq f_c$, то перейти к шагу 6.
3. Положить $a = c$, $f_a = f_c$, $c = b$, $f_c = f_b$, $h = 2h$.
4. Вычислить $b = c + h$, $f_b = f(b)$.
5. Если $f_b < f_c$, то перейти к шагу 3, иначе перейти к шагу 11.
6. Вычислить $a = c - h$, $f_a = f(a)$.
7. Если $f_a \geq f_c$, то перейти к шагу 11.
8. Положить $b = c$, $f_b = f_c$, $c = a$, $f_c = f_a$, $h = 2h$.
9. Вычислить $a = c - h$, $f_a = f(a)$.

10. Если $f_a < f_c$, то перейти к шагу 8.

11. Остановиться.

На шагах 1–5 выполняется поиск интервала в положительном направлении, а на шагах 6–10 – в отрицательном направлении.

Лабораторная работа

Тема лабораторной работы.

Разработка подпрограммы поиска интервала неопределенности.

Цель лабораторной работы.

Разработать подпрограмму метода Свенна для поиска интервала неопределенности унимодальной целевой функции.

Порядок выполнения лабораторной работы.

1. Получить у преподавателя вариант унимодальной целевой функции $f(x)$ одной переменной $x \in R$ с начальной точкой поиска минимума $x_0 \in R$.

2. Аналитически найти точку $x^* \in R$ минимума заданной функции $f(x)$ и вычислить минимальное значение функции $f^* = f(x^*)$.

3. Написать подпрограмму вычисления значений функции $f(x)$ с входным параметром x и выходным параметром f .

4. Протестировать подпрограмму вычисления функции $f(x)$ при вычислении значений $f_0 = f(x_0)$ и $f^* = f(x^*)$, сравнить f^* с f_0 .

5. На основании алгоритма метода Свенна написать подпрограмму для поиска интервала неопределенности точки минимума x^* функции $f(x)$ из начальной точки x_0 в соответствии с требованиями:

1) за основу подпрограммы для поиска интервала неопределенности взять подпрограмму итерационного процесса из предыдущей лабораторной работы и модифицировать эту подпрограмму согласно алгоритму метода Свенна;

2) входные параметры подпрограммы: $f(x)$ – целевая функция,

x_0 – начальная точка поиска, h – начальный шаг, N – предельное число точек;

3) выходные параметры подпрограммы: X – массив всех точек поиска x_k ; F – массив соответствующих значений функции $f_k = f(x_k)$; a и b – границы интервала неопределенности; f_a и f_b – значения функции на концах интервала; c и f_c – внутренняя точка интервала и значение в ней функции;

4) на каждой итерации выводить на экран компьютера строку, содержащую номер точки поиска k , шаг поиска h_k , значение функции f_k и значение переменной x_k ;

5) итерации продолжаются до тех пор, пока выполняется неравенство $k \leq N$;

6) после окончания процесса оптимизации на экране должна отображаться таблица, имеющая соответствующий заголовок и представляющая процесс минимизации функции;

7) под таблицей необходимо отобразить количество вычислений целевой функции, конечный шаг, конечное значение функции и соответствующее ему значение независимой переменной, а также значения выходных параметров a , b , f_a , f_b , c , f_c .

6. Написать общую программу для выполнения вычислительного процесса поиска интервала неопределенности и отображения на экране этого процесса в соответствии с требованиями:

1) за основу общей программы для поиска интервала неопределенности взять общую программу итерационного процесса из предыдущей лабораторной работы и модифицировать эту программу;

2) задать в программе значения x_0 , x^* , h , ε , N ;

3) путем вызова подпрограммы метода Свенна с использованием подпрограммы вычисления функции $f(x)$ и значений x_0 , h , N реализовать итерационный процесс поиска интервала неопределенности с построением таблицы процесса на экране и получением массивов зна-

чений переменной X и функции F , границ интервала неопределенности a и b , значений функции на концах интервала f_a и f_b , внутренней точки интервала c и значения в ней функции f_c ;

4) с помощью подпрограммы графического отображения итерационного процесса, разработанной на предыдущей лабораторной работе, по массивам X и F отобразить график процесса поиска интервала неопределенности на экране компьютера.

7. Запустить на счет общую программу поиска интервала неопределенности из нескольких начальных точек

$$x_{01} < x^*, \quad x_{02} = x^*, \quad x_{03} > x^*$$

при нескольких значениях шага

$$h_{01} < |x_0 - x^*|, \quad h_{02} = |x_0 - x^*|, \quad h_{03} > |x_0 - x^*|.$$

8. Выполнить анализ полученных результатов:

1) на основании табличной и графической информации о работе метода поиска интервала неопределенности определить конечные точки метода Свенна (a, f_a) , (b, f_b) , (c, f_c) и найденный этим методом интервал неопределенности $[a, b]$;

2) на основании табличного и графического представления процесса поиска интервала неопределенности сделать вывод о верном нахождении границ интервала;

3) сделать вывод о правильности работы общей программы и всех используемых ею подпрограмм.

9. Ответить на вопросы преподавателя по теме лабораторной работы.

10. Оформить и сдать отчет о проведенной лабораторной работе.

С о д е р ж а н и е о т ч е т а :

1) титульный лист установленного образца с указанием организации, названия учебной дисциплины, темы работы, номера варианта, исполнителя и принимающего, города, года;

2) введение с кратким общим описанием предназначения методов

одномерного поиска и целью лабораторной работы;

3) целевая функция, начальная точка поиска, аналитическое решение задачи минимизации функции, точка минимума и минимальное значение функции;

4) теоретическое описание метода Свенна;

5) алгоритм метода Свенна;

6) распечатки используемых компьютерных программ;

7) табличное представление процесса поиска интервала неопределенности и его итоги – количество вычислений целевой функции, конечный шаг, границы интервала неопределенности a и b , значения функции на концах интервала f_a и f_b , внутренняя точка интервала c и значение в ней функции f_c ;

8) двумерный график процесса поиска интервала неопределенности со всеми точками;

9) анализ полученных результатов;

10) выводы о проделанной лабораторной работе.

Контрольные вопросы

1. Для решения какой задачи предназначены методы одномерного поиска?

2. Дайте определения минимума и максимума функции одной переменной.

3. Сформулируйте и докажите теорему Ферма.

4. Дайте определение стационарной точки.

5. Дайте определения глобального и локального минимумов.

6. Сформулируйте необходимый признак существования экстремума функции одной переменной.

7. Сформулируйте необходимый и достаточный признак существования экстремума функции одной переменной.

8. Как исследовать функцию одной переменной на экстремум с помощью второй производной этой функции?

9. Какие функции называются строго убывающими и строго возрастающими?
10. Дайте определение унимодальной функции.
11. Какие функции называются выпуклыми и строго выпуклыми?
12. Какие функции называются вогнутыми и строго вогнутыми?
13. В чем заключается особенность выпуклой функции, имеющей минимум?
14. Дайте определение интервала неопределенности.
15. Из каких этапов состоит большинство методов одномерного поиска?
16. Какое наименьшее количество точек на графике функции надо знать, чтобы определить интервал неопределенности?
17. Какое наименьшее количество точек внутри интервала неопределенности необходимо определить, чтобы уменьшить этот интервал?
18. Сформулируйте правило исключения интервалов.
19. Для чего предназначен метод Свенна?
20. По какому принципу изменяется величина шага в методе Свенна?
21. Как заканчиваются вычисления в методе Свенна?
22. Какого порядка метод Свенна?

3. МЕТОДЫ УМЕНЬШЕНИЯ ИНТЕРВАЛА НЕОПРЕДЕЛЕННОСТИ

Данный раздел в основном посвящен методам уменьшения интервала неопределенности точки минимума унимодальной функции. Обосновываются методы равномерного поиска, дихотомии, деления интервала пополам. Рассматриваются свойства чисел Фибоначчи, на основании которых конструируется метод Фибоначчи. Изучаются свойства золотого сечения, по которым обосновывается метод золотого сечения. Для рассмотренных методов дана оценка их эффективности. Представляется метод адаптации шага, который не требует нахождения интервала неопределенности. Выполняется сравнение методов уменьшения интервала неопределенности по коэффициенту уменьшения длины интервала при заданном количестве вычислений функции. Приводятся алгоритмы для всех рассмотренных методов. Дается описание лабораторной работы по изучению методов уменьшения интервала неопределенности точки минимума функции одной переменной.

3.1. Метод равномерного поиска

Для отыскания минимума функции одной переменной $f(x)$ существуют пассивные и последовательные стратегии поиска. При пассивной стратегии все n точек $x_1, x_2, \dots, x_n$, в которых будут вычислены значения функции $f(x)$, определены заранее до вычисления функции в этих точках.

Метод равномерного поиска соответствует пассивной стратегии с равномерным заполнением n точками поиска интервала неопределенности $[a, b]$, которые делят этот интервал на $n+1$ часть одинаковой длины. Для этого находится длина интервала неопределенности

$L = b - a$, определяется расстояние между точками $\delta = L/(n+1)$, вычисляются координаты точек $x_k = a + k\delta$ для $k = \overline{1, n}$ и соответствующие им значения функции $f_k = f(x_k)$ (рис. 3.1).

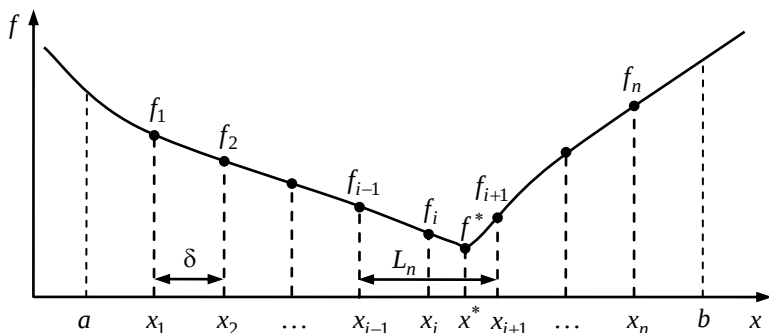


Рис. 3.1. Метод равномерного поиска

Далее определяется точка x_i с наименьшим значением функции $f_i = \min_k f_k$. Поскольку в двух соседних точках относительно точки x_i значения функции не меньше, чем в этой точке, то в результате выполнения равномерного поиска будет получен конечный интервал неопределенности $[x_{i-1}, x_{i+1}]$ длины $L_n = 2\delta$ (рис. 3.1). Точка минимума x^* будет находиться в этом интервале. При этом погрешность вычисления точки x^* относительно лучшей точки x_i не превысит δ .

Эффективность метода равномерного поиска можно оценить с помощью отношения длин конечного и начального интервалов неопределенности:

$$r_n = \frac{L_n}{L} = \frac{2\delta}{L} = \frac{2L}{L(n+1)}, \quad r_n = \frac{2}{n+1}. \quad (3.1)$$

Эта величина показывает, во сколько раз уменьшится длина исходного

интервала при n вычислениях целевой функции.

Выясним, сколько раз необходимо вычислить значения функции, чтобы интервал неопределенности уменьшить в сто раз:

$$r_n = \frac{1}{100}, \quad \frac{2}{n+1} = \frac{1}{100}, \quad n+1 = 200, \quad n = 199.$$

Если же интервал необходимо уменьшить в 1000 раз, то $r_n = 0,001$ и $n = 1999$.

Метод равномерного поиска можно реализовать по следующему алгоритму.

Алгоритм метода равномерного поиска.

Входные параметры: a и b – границы интервала неопределенности, n – количество точек деления интервала.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $L = b - a$, $h = L/(n+1)$, $x = a + h$, $f_x = f(x)$.
2. Положить $k = 2$, $v = x$.
3. Положить $v = v + h$ и вычислить $f_v = f(v)$.
4. Если $f_v < f_x$, то положить $x = v$, $f_x = f_v$.
5. Если $k < n$, то положить $k = k + 1$ и перейти к шагу 3.
6. Остановиться.

Метод равномерного поиска позволяет вычислить минимум даже многоэкстремальной целевой функции в заданном интервале с гарантированной точностью. Однако для унимодальных функций существуют намного более эффективные последовательные методы поиска.

3.2. Метод дихотомии

В отличие от пассивного метода минимизации функции одной переменной $f(x)$ в последовательных методах точки поиска выбирают последовательно с учетом результатов предыдущих вычислений. Все

далее рассматриваемые методы одномерного поиска являются последовательными.

Рассмотрим деление интервала неопределенности $[a, b]$ длины $L = b - a$ двумя внутренними точками u и v , такими что $u < v$, со значениями функции $f_u = f(u)$, $f_v = f(v)$. По правилу исключения интервалов получим новый интервал неопределенности: $[a, v]$, если $f_u < f_v$ (рис. 2.11, а); $[u, b]$, если $f_u \geq f_v$ (рис. 2.11, б). Обозначим длины новых интервалов $l_1 = v - a$ и $l_2 = b - u$.

Поскольку заранее неизвестно, $f_u < f_v$ или $f_u \geq f_v$, то при двух вычислениях функции $n = 2$ длина L_2 нового интервала неопределенности будет удовлетворять условию $L_2 \leq \max\{l_1, l_2\}$. Величина L_2 не будет зависеть от результата сравнения значений функции во внутренних точках u и v , если $l_1 = l_2$. Таким образом, точки u и v целесообразно располагать симметрично относительно центра $c = (a + b)/2$ исходного интервала $[a, b]$.

Пусть δ – расстояние точек u и v от центра интервала c . Тогда эти точки $u = c - \delta$, $v = c + \delta$, длина нового интервала $L_2 = L/2 + \delta$

(рис. 3.2). При этом чем меньше значение δ , тем меньше длина нового интервала неопределенности. Поэтому δ берут достаточно малым, так что можно полагать $L_2 = L/2$, то есть интервал неопределенности уменьшается в два раза. В зависимости от результата сравнения f_u и f_v в качестве следующего интервала неопределенности берут $[a, v]$ или $[u, b]$, и к новому интервалу

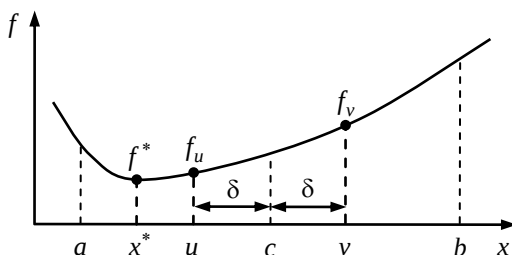


Рис. 3.2. Метод дихотомии

применяют аналогичные действия. В этом и заключается *метод дихотомии*. Термин *дихотомия* происходит от греческого слова *dichotomia* – разделение надвое. Итерации продолжаются до тех пор, пока интервал неопределенности не уменьшится до заданного значения. Погрешность вычисления точки минимума x^* не превысит половины длины конечного интервала неопределенности.

На каждой итерации метода дихотомии с номером k вычисляется два значения функции, поэтому количество вычислений функций $n = 2k$ – четное число. После первой итерации длина нового интервала неопределенности $L_2 = L/2 + \delta$. После второй итерации получим

$$L_4 = \frac{L_2}{2} + \delta = \frac{L}{4} + \frac{3\delta}{2}.$$

После третьей итерации

$$L_6 = \frac{L_4}{2} + \delta = \frac{L_2}{4} + \frac{3\delta}{2} = \frac{L}{8} + \frac{7\delta}{4}.$$

После итерации с номером k

$$L_{2k} = \frac{L}{2^k} + \frac{2^k - 1}{2^{k-1}} \delta.$$

В этом равенстве второе слагаемое правой части не превосходит 2δ и стремится к этому значению при $k \rightarrow \infty$. Поскольку $k = n/2$, последнее равенство перепишем в виде

$$L_n = \frac{L}{2^{n/2}} + \frac{2^{n/2} - 1}{2^{n/2-1}} \delta.$$

При пренебрежимо малом значении δ в результате n вычислений функции получим интервал длины

$$L_n = \frac{L}{2^{n/2}}.$$

Эффективность метода дихотомии определяется аналогично методу равномерного поиска путем вычисления отношения длин конечного и начального интервалов неопределенности:

$$r_n = \frac{L_n}{L}, \quad r_n = \frac{1}{2^{n/2}}. \quad (3.2)$$

Пусть, например, интервал неопределенности необходимо уменьшить не менее чем в сто раз:

$$r_n \leq \frac{1}{100}, \quad \frac{1}{2^{n/2}} \leq \frac{1}{100}, \quad 2^{n/2} \geq 100, \quad n \geq 2\log_2 100 = 4\log_2 10 = 13,3.$$

Это означает, что можно положить $n=14$, то есть в 7 итерациях функцию достаточно вычислить 14 раз.

Если же интервал неопределенности необходимо уменьшить не менее чем в 1000 раз, то $n \geq 2\log_2 1000 = 19,9$ и можно положить $n=20$.

Итак, последовательный метод дихотомии гораздо эффективнее пассивного метода равномерного поиска.

Недостатком метода дихотомии является необходимость вычисления двух значений функции на каждой итерации и зависимость надежности метода от значения параметра δ . При слишком малом значении δ и ошибках вычисления значений функции интервал неопределенности может быть утерян. При слишком большом значении δ эффективность метода дихотомии существенно снижается.

Алгоритм метода дихотомии.

Входные параметры: a и b – концы интервала неопределенности, δ – расстояние между внутренними точками, ε – допустимая погрешность.

Выходные параметры: u и f_u – лучшая точка и значение функции в ней.

1. Вычислить $c = (a+b)/2$.

2. Вычислить $u = c - \delta$, $f_u = f(u)$, $v = c + \delta$, $f_v = f(v)$.
3. Если $f_u < f_v$, то положить $b = v$, иначе положить $a = u$.
4. Если $v - a > \varepsilon$, то перейти к шагу 1.
5. Остановиться.

В этом алгоритме между его входными параметрами должно выполняться соотношения $a < b$, $\delta < \varepsilon$.

3.3. Метод деления интервала пополам

Метод деления интервала пополам применяется для уменьшения интервала неопределенности точки минимума x^* целевой функции $f(x)$ и является усовершенствованием метода дихотомии с целью исключения недостатков этого метода.

На первой итерации разделим интервал неопределенности $[a, b]$ длины $L = b - a$ пополам точкой $c = (a + b)/2$ и вычислим в ней значение целевой функции $f_c = f(c)$. В результате получим два меньших интервала: $[a, c]$ и $[c, b]$ (рис. 3.3).

Разделим левый интервал $[a, c]$ пополам точкой $u = (a + c)/2$ и вычислим в ней значение целевой функции $f_u = f(u)$. Если $f_u < f_c$, то по правилу исключения интервалов исключим правый интервал $(c, b]$ и получим новый интервал неопределенности $[a, c]$ с известной средней точкой (u, f_u) , что представлено на рис. 3.3.

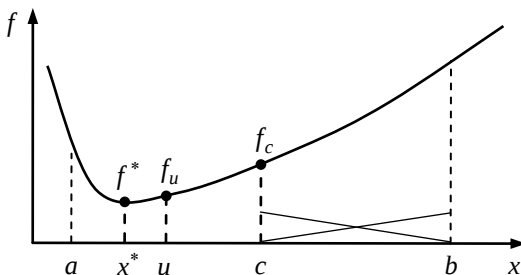


Рис. 3.3. Исключение правого интервала

Если же $f_u \geq f_c$, то разделим правый интервал $[c, b]$ пополам точкой $v = (c + b)/2$ и вычислим в ней значение целевой функции $f_v = f(v)$.

Если $f_v < f_c$, то по правилу исключения интервалов исключим левый интервал $[a, c]$ и получим новый интервал неопределенности $[c, b]$ с известной средней точкой (v, f_v) (рис. 3.4).

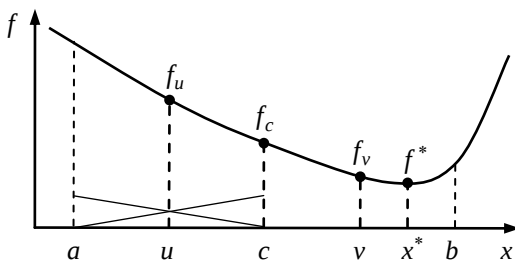


Рис. 3.4. Исключение левого интервала

Если же $f_v \geq f_c$, то исключаются крайние интервалы $[a, u]$ и $(v, b]$, а интервал $[u, v]$ становится новым интервалом неопределенности с известной средней точкой (c, f_c) (рис. 3.5).

В любом случае после двух или трех вычислений целевой функции будет найден новый интервал неопределенности, вдвое меньший исходного интервала, с известной средней точкой.

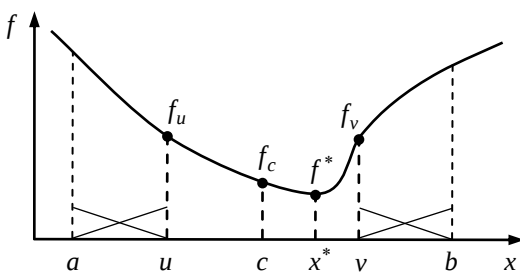


Рис. 3.5. Исключение крайних интервалов

В последующих итерациях, которые выполняются аналогично первой, но без вычисления средней точки, интервал неопределенности будет делиться пополам уже после одного или двух вычислений целевой функции. Итерации продолжаются до тех пор, пока длина интервала

неопределенности не станет меньше заданного значения. Погрешность вычисления точки минимума x^* не превысит половины длины конечного интервала неопределенности.

Если на каждой итерации интервалом неопределенности становится левая половина предыдущего интервала неопределенности, то результате n вычислений функции получим конечный интервал неопределенности длины

$$L_n = \frac{L}{2^{n-1}}.$$

Если же на каждой итерации интервалом неопределенности становится не левая половина предыдущего интервала неопределенности, то результате n вычислений функции получим конечный интервал неопределенности длины

$$L_n = \frac{L}{2^{(n-1)/2}}.$$

В общем случае

$$\frac{L}{2^{n-1}} \leq L_n \leq \frac{L}{2^{(n-1)/2}}.$$

Эффективность метода деления интервала пополам определяется аналогично эффективности метода дихотомии:

$$r_n = \frac{L_n}{L}, \quad \frac{1}{2^{n-1}} \leq r_n \leq \frac{1}{2^{(n-1)/2}}. \quad (3.3)$$

Пусть, например, интервал неопределенности необходимо уменьшить не менее чем в сто раз и $r_n \leq 0,01$:

$$\frac{1}{2^{n-1}} \leq \frac{1}{100}, \quad 2^{n-1} \geq 100, \quad n-1=7, \quad n=8;$$

$$\frac{1}{2^{(n-1)/2}} \leq \frac{1}{100}, \quad 2^{n/2} \geq 100, \quad \frac{n-1}{2}=7, \quad n=15.$$

Здесь использовано неравенство $2^7 = 128 > 100$. Таким образом, $8 \leq n \leq 15$.

Если же интервал неопределенности необходимо уменьшить не менее чем в 1000 раз, то, используя неравенство $2^{10} = 1024 > 1000$, получим $11 \leq n \leq 21$.

Итак, метод деления интервала пополам в наихудшем случае может потратить только на одно вычисление функции больше, чем метод дихотомии, а в наилучшем случае он использует почти вдвое меньшее количество вычислений. К тому же метод деления интервала пополам не использует малого параметра, а в нем применяются только простые арифметические операции сложения чисел и деления на два. Поэтому данный метод надежнее и во многих случаях эффективнее метода дихотомии.

Алгоритм метода деления интервала пополам.

Входные параметры: a, b – концы интервала неопределенности, e – допустимая погрешность.

Выходные параметры: c и f_c – лучшая точка и значение функции в ней.

1. Вычислить $c = (a+b)/2$, $f_c = f(c)$.
2. Вычислить $u = (a+c)/2$, $f_u = f(u)$.
3. Если $f_u < f_c$, то положить $b = c$, $c = u$, $f_c = f_u$ и перейти к шагу 7.
4. Вычислить $v = (c+b)/2$, $f_v = f(v)$.
5. Если $f_v < f_c$, то положить $a = u$, $c = v$, $f_c = f_v$ и перейти к шагу 7.
6. Положить $a = u$, $b = v$.
7. Если $c - a > e$, то перейти к шагу 2.
8. Остановиться.

В этом алгоритме на первых двух шагах вычисляются средние точки интервала и левой половины интервала, а на шаге 4 – средняя точка правой половины интервала. Точки сравниваются на шагах 3 и 5.

3.4. Числа Фибоначчи и их свойства

Леонардо Пизано Фибоначчи – известный математик эпохи Средневековья, который в 1202 году написал книгу «Liber abacci», содержащую почти все арифметические и алгебраические сведения того времени. В частности, по этой книге европейцы познакомились с индусскими или «арабскими» цифрами. Там же содержалась знаменитая «задача о размножении кроликов».

«Пусть в огороженном месте имеется пара кроликов (самка и самец) в первый день января. Эта пара кроликов производит новую пару кроликов в первый день февраля и затем в первый день каждого следующего месяца. Каждая новорожденная пара кроликов становится зрелой уже через месяц и затем через месяц дает жизнь новой паре кроликов. Сколько пар кроликов будет в огороженном месте через год, то есть через 12 месяцев с начала размножения?»

Решение этой задачи приводит к *числам Фибоначчи*:

$$F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, F_6 = 13, F_7 = 21, \dots,$$

где индекс числа Фибоначчи соответствует номеру месяца, а значение числа – количеству пар кроликов в этом месяце. Такой последовательности чисел, начиная с F_2 , отвечает закономерность: *каждый член последовательности равен сумме двух предыдущих*. Это правило может быть записано в виде первой в истории математики *рекуррентной формулы*

$$F_n = F_{n-1} + F_{n-2}. \quad (3.4)$$

Числа Фибоначчи F_n для номера n от 0 до 51 представлены в табл. 3.1.

Таким образом, по этой таблице получим решение приведенной «задачи о размножении кроликов»: через год в огороженном месте окажется $F_{13} = 377$ пар кроликов.

Числа Фибоначчи кроме описания количественных характеристик

процесса размножения кроликов представляют собой удобный математический инструмент для описания различных явлений природы: процессов развития растений и животных, форм живых организмов и неживой материи.

Таблица 3.1 – Числа Фибоначчи

n	F_n	n	F_n	n	F_n	n	F_n
0	1	13	377	26	196418	39	102334155
1	1	14	610	27	317811	40	165580141
2	2	15	987	28	514229	41	267914296
3	3	16	1597	29	832040	42	433494437
4	5	17	2584	30	1346269	43	701408733
5	8	18	4181	31	2178309	44	1134903170
6	13	19	6765	32	3524578	45	1836311903
7	21	20	10946	33	5702887	46	2971215073
8	34	21	17711	34	9227465	47	4807526976
9	55	22	28657	35	14930352	48	7778742049
10	89	23	46368	36	24157817	49	12586269025
11	144	24	75025	37	39088169	50	20365011074
12	233	25	121393	38	63245986	51	32951280099

Числа Фибоначчи обладают многими интересными математическими свойствами. В частности,

$$F_1^2 - F_0 F_2 = 1^2 - 1 \cdot 2 = -1, \quad F_2^2 - F_1 F_3 = 2^2 - 1 \cdot 3 = 1,$$

$$F_3^2 - F_2 F_4 = 3^2 - 2 \cdot 5 = -1, \quad F_4^2 - F_3 F_5 = 5^2 - 3 \cdot 8 = 1, \quad \dots$$

Методом математической индукции докажем справедливость общей математической формулы

$$F_n^2 - F_{n-1} F_{n+1} = (-1)^n. \quad (3.5)$$

Пусть эта формула верна при некотором $n = k$, то есть

$$F_k^2 - F_{k-1}F_{k+1} = (-1)^k. \quad (3.6)$$

Докажем справедливость формулы (3.5) и при $n = k + 1$, то есть, что

$$F_{k+1}^2 - F_kF_{k+2} = (-1)^{k+1}. \quad (3.7)$$

На основании равенства (3.4) имеем

$$F_{k+1}^2 - F_kF_{k+2} = F_{k+1}(F_k + F_{k-1}) - F_k(F_{k+1} + F_k).$$

Раскрывая скобки в правой части, получим

$$F_{k+1}^2 - F_kF_{k+2} = F_{k+1}F_k + F_{k+1}F_{k-1} - F_kF_{k+1} - F_k^2,$$

и с учетом равенства (3.6)

$$F_{k+1}^2 - F_kF_{k+2} = -(F_k^2 - F_{k+1}F_{k-1}) = -(-1)^k$$

приходим к равенству (3.7).

Докажем теперь справедливость формулы

$$F_{n-1}F_{n-2} - F_nF_{n-3} = (-1)^n. \quad (3.8)$$

На основании равенства (3.4) имеем

$$F_{n-1}F_{n-2} - F_nF_{n-3} = F_{n-1}(F_n - F_{n-1}) - F_n(F_{n-1} - F_{n-2}).$$

Раскрывая скобки, получим

$$F_{n-1}F_{n-2} - F_nF_{n-3} = F_{n-1}F_n - F_{n-1}^2 - F_nF_{n-1} + F_nF_{n-2}.$$

С учетом равенства (3.5)

$$F_{n-1}F_{n-2} - F_nF_{n-3} = -(F_{n-1}^2 - F_nF_{n-2}) = -(-1)^{n-1},$$

откуда и следует формула (3.8).

Получим выражения для чисел Фибоначчи в явном виде. Для этого будем искать решение рекуррентного уравнения (3.4) среди геометрических прогрессий с n -м членом x^n

$$x^n = x^{n-1} + x^{n-2}.$$

Сокращая на x^{n-2} , получим уравнение

$$x^2 - x - 1 = 0, \quad (3.9)$$

корни которого

$$x_1 = \frac{1+\sqrt{5}}{2}, \quad x_2 = \frac{1-\sqrt{5}}{2}. \quad (3.10)$$

Итак, последовательности $\{x_1^n\}$ и $\{x_2^n\}$ удовлетворяют рекуррентному уравнению (3.4). Ему же удовлетворяет и любая линейная комбинация этих последовательностей $\{c_1 x_1^n + c_2 x_2^n\}$. Выберем коэффициенты c_1 и c_2 так, чтобы

$$F_n = c_1 x_1^n + c_2 x_2^n. \quad (3.11)$$

Для этого достаточно потребовать выполнения двух условий: $F_0 = 1$, $F_1 = 1$. Имеем:

$$\begin{cases} c_1 + c_2 = 1, \\ c_1 x_1 + c_2 x_2 = 1. \end{cases}$$

Решим эту систему уравнений по правилу Крамера:

$$\Delta = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} = x_2 - x_1 = \frac{1-\sqrt{5}}{2} - \frac{1+\sqrt{5}}{2} = \frac{-2\sqrt{5}}{2} = -\sqrt{5},$$

$$\Delta_1 = \begin{vmatrix} 1 & 1 \\ 1 & x_2 \end{vmatrix} = x_2 - 1 = \frac{1-\sqrt{5}}{2} - 1 = \frac{-1-\sqrt{5}}{2} = -\frac{1+\sqrt{5}}{2} = -x_1,$$

$$\Delta_2 = \begin{vmatrix} 1 & 1 \\ x_1 & 1 \end{vmatrix} = 1 - x_1 = 1 - \frac{1+\sqrt{5}}{2} = \frac{1-\sqrt{5}}{2} = x_2,$$

$$c_1 = \frac{\Delta_1}{\Delta} = \frac{-x_1}{-\sqrt{5}} = \frac{x_1}{\sqrt{5}}, \quad c_2 = \frac{\Delta_2}{\Delta} = \frac{x_2}{-\sqrt{5}} = -\frac{x_2}{\sqrt{5}}.$$

Подставляя найденные значения c_1 и c_2 в формулу (3.11), получим

$$F_n = \frac{x_1}{\sqrt{5}} x_1^n - \frac{x_2}{\sqrt{5}} x_2^n = \frac{x_1^{n+1} - x_2^{n+1}}{\sqrt{5}}.$$

С учетом значений корней (3.10) окончательно имеем

$$F_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}{\sqrt{5}}. \quad (3.12)$$

Это равенство называется формулой Бине. Таким образом, все числа Фибоначчи, по определению входящие в множество целых чисел, могут быть выражены в явном виде через разность степеней дробей, включающих иррациональное число $\sqrt{5}$.

3.5. Метод Фибоначчи

Метод Фибоначчи основан на числах Фибоначчи и применяется для уменьшения интервала неопределенности точки минимума x^* целевой функции $f(x)$.

При рассмотрении метода дихотомии для интервала неопределенности $[a, b]$ длины $L = b - a$ двумя внутренними точками, показано, что точки деления целесообразно располагать симметрично относительно центра интервала на малом расстоянии между ними. Но в методе дихотомии на каждой итерации значения целевой функции необходимо вычислять в двух точках.

В методе Фибоначчи также две первые точки деления x_1 и x_2 расположим симметрично относительно центра интервала, вычислим в

них значения функции и применим правило исключения интервалов. На следующей итерации в оставшемся интервале неопределенности с одной известной внутренней точкой найдем вторую внутреннюю точку x_3 , симметричную оставшейся внутренней точке, вычислим в точке x_3 значение функции и опять применим правило исключения интервалов. Таким образом, сохраняя принцип симметрии внутренних точек, на каждой последующей итерации будем вычислять только одно значение функции.

Пусть на уменьшение интервала неопределенности отводится n вычислений целевой функции. Чтобы обеспечить наибольшее уменьшение интервала, последнюю точку поиска x_n , как и в методе дихотомии, расположим симметрично относительно оставшейся внутренней точки x_{n-1} на малом расстоянии δ от нее. Обозначим конечную длину интервала неопределенности, полученную при n вычислениях целевой функции, через L_n . Окончание процесса деления интервала неопределенности показано на рис. 3.6.

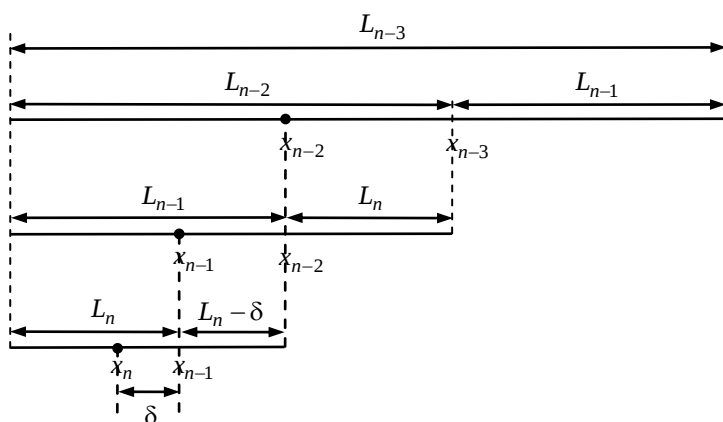


Рис. 3.6. Окончание деления интервала методом Фибоначчи

Найдем длины интервалов неопределенности:

$$\begin{aligned} L_{n-1} &= L_n + (L_n - \delta) = 2L_n - \delta, & L_{n-2} &= L_{n-1} + L_n = 3L_n - \delta, \\ L_{n-3} &= L_{n-2} + L_{n-1} = 5L_n - 2\delta, & L_{n-4} &= L_{n-3} + L_{n-2} = 8L_n - 3\delta, \quad \dots \end{aligned}$$

Перепишем эти равенства с использованием чисел Фибоначчи:

$$\begin{aligned} L_{n-1} &= F_2 L_n - F_0 \delta, & L_{n-2} &= F_3 L_n - F_1 \delta, \\ L_{n-3} &= F_4 L_n - F_2 \delta, & L_{n-4} &= F_5 L_n - F_3 \delta, \quad \dots \end{aligned}$$

Отсюда выведем общую формулу для длин интервалов:

$$L_{n-j} = F_{j+1} L_n - F_j \delta, \quad j = \overline{1, n-1}. \quad (3.13)$$

Полагая $j = n-1$, найдем исходную длину интервала $L_1 = L = b - a$

$$L = F_n L_n - F_{n-2} \delta.$$

Отсюда получим, что при n вычислениях функции в методе Фибоначчи конечная длина интервала неопределенности станет равной

$$L_n = \frac{L}{F_n} + \frac{F_{n-2}}{F_n} \delta. \quad (3.14)$$

Для начала поиска необходимо найти положение первой точки деления исходного интервала, которая должна лежать на расстоянии L_2 от конца интервала. Полагая в формуле (3.13) $j = n-2$, получим

$$L_2 = F_{n-1} L_n - F_{n-3} \delta.$$

Подставим в эту формулу выражение для L_n из формулы (3.14):

$$L_2 = F_{n-1} \left(\frac{L}{F_n} + \frac{F_{n-2}}{F_n} \delta \right) - F_{n-3} \delta = \frac{F_{n-1}}{F_n} L + \frac{F_{n-1} F_{n-2} - F_n F_{n-3}}{F_n} \delta.$$

С учетом равенства (3.8) окончательно получим

$$L_2 = \frac{1}{F_n}[(b-a)F_{n-1} + (-1)^n \delta]. \quad (3.15)$$

По свойству симметрии для двух первых точек деления исходного интервала неопределенности $[a, b]$ можно положить:

$$x_1 = b - L_2, \quad x_2 = a + L_2.$$

Отсюда следует, что

$$L_2 = b - x_1, \quad x_2 = a + b - x_1.$$

Если на некоторой итерации известна внутренняя точка x_k интервала $[a_k, b_k]$, то симметричная ей точка находится по формуле

$$x_{k+1} = a + b - x_k. \quad (3.16)$$

Путем сравнения x_k и x_{k+1} , $f_k = f(x_k)$ и $f_{k+1} = f(x_{k+1})$ по правилу исключения интервалов уменьшается интервал неопределенности.

При пренебрежимо малом значении δ в результате n вычислений функции по формуле (3.14) получим интервал длины

$$L_n = \frac{L}{F_n}.$$

Эффективность метода Фибоначчи определим отношением длин конечного и начального интервалов неопределенности:

$$r_n = \frac{L_n}{L}, \quad r_n = \frac{1}{F_n}. \quad (3.17)$$

Если интервал неопределенности необходимо уменьшить не менее чем в сто раз, то по числам Фибоначчи из табл. 3.1 получим:

$$r_n \leq \frac{1}{100}, \quad \frac{1}{F_n} \leq \frac{1}{100}, \quad F_n \geq 100, \quad F_{11} = 144 > 100, \quad n = 11.$$

Если же интервал неопределенности необходимо уменьшить не менее

чем в 1000 раз, то $F_n \geq 1000$, $F_{16} = 1597 > 1000$, и можно положить $n = 16$.

Итак, метод Фибоначчи эффективнее метода дихотомии. Недостатком метода Фибоначчи являются необходимость задания количества вычислений n для значений целевой функции и зависимость надежности метода от значения параметра δ , которое должно удовлетворять условию $\delta < L/F_{n+1}$. Но при слишком малом значении δ и ошибках вычисления значений функции интервал неопределенности может быть утерян.

Формулы (3.15) и (3.16) вместе с правилом исключения интервалов являются основой алгоритма метода Фибоначчи.

А л г о р и т м м е т о д а Ф и б о н а ч ч и .

Входные параметры: a и b – концы интервала неопределенности, n – число вычислений функции, δ – расстояние между внутренними точками на последней итерации.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Положить $F_0 = 1$, $F_1 = 1$, $F = F_0 + F_1$, $s = 1$, $i = 3$.
2. Положить $F_0 = F_1$, $F_1 = F$, $F = F_0 + F_1$, $s = -s$.
3. Если $i < n$, то положить $i = i + 1$ и перейти к шагу 2.
4. Вычислить $x = b - [(b - a) \cdot F_1 + \delta \cdot s] / F$, $f_x = f(x)$.
5. Положить $k = 1$.
6. Вычислить $y = a + b - x$, $f_y = f(y)$.
7. Если $f_y > f_x$, то перейти к шагу 10.
8. Если $y < x$, то положить $b = x$, $x = y$, $f_x = f_y$ и перейти к шагу 12.
9. Положить $a = x$, $x = y$, $f_x = f_y$ и перейти к шагу 12.
10. Если $y < x$, то положить $a = y$ и перейти к шагу 12.
11. Положить $b = y$.

12. Если $k < n$, то положить $k = k + 1$ и перейти к шагу 6.
13. Если $f_y < f_x$, то положить $x = y$, $f_x = f_y$.
14. Остановиться.

3.6. Золотое сечение и его свойства

Самое известное математическое сочинение античной науки «Начала», написанное Евклидом в 3 веке до нашей эры, содержит геометрическую задачу «о делении отрезка в крайнем и среднем отношении». Смысл этой задачи поясняет рис. 3.7: требуется разделить отрезок AB точкой C в таком отношении, чтобы длина большей части отрезка CB так относилась к длине меньшей части AC , как длина всего отрезка AB к длине его большей части CB

$$\frac{AB}{CB} = \frac{CB}{AC}. \quad (3.18)$$

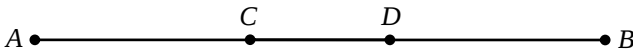


Рис. 3.7. Деление отрезка в крайнем и среднем отношении

Обозначим отношение (3.18) через x . Тогда, учитывая, что $AB = AC + CB$, отношение (3.18) представим в виде

$$x = \frac{AC + CB}{CB} = 1 + \frac{AC}{CB} = 1 + \frac{1}{\frac{CB}{AC}},$$

откуда вытекает алгебраическое уравнение

$$x^2 = x + 1. \quad (3.19)$$

Это уравнение совпадает с уравнением (3.9), имеющем корни (3.10).

Из геометрического смысла отношения (3.18) вытекает, что решением задачи является положительный корень

$$\tau = \frac{1 + \sqrt{5}}{2} = 1,6180339887498949 \dots \quad (3.20)$$

Леонардо да Винчи назвал это иррациональное число *золотым сечением*. Уравнение (3.19) называется *уравнением золотого сечения*. На отрезке AB существует еще одна точка D (рис. 3.7), которая делит его золотым сечением, так как

$$\frac{AB}{AD} = \frac{AD}{DB} = \frac{1 + \sqrt{5}}{2}.$$

Золотое сечение широко встречается в геометрии. Из «Начал» Евклида известен способ геометрического построения золотого сечения с использованием линейки и циркуля, показанный на рис. 3.8.

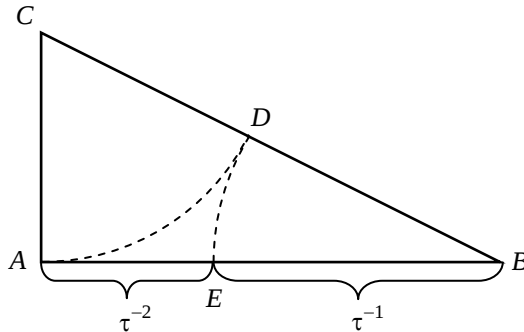


Рис. 3.8. Геометрическое построение золотого сечения

Построим прямоугольный треугольник ABC со сторонами $AB=1$ и $AC=1/2$. Тогда по теореме Пифагора

$$CB = \sqrt{1 + \left(\frac{1}{2}\right)^2} = \frac{\sqrt{5}}{2}.$$

Проведя дугу окружности AD с центром в точке C до ее пересечения с отрезком CB в точке D , получим отрезок

$$BD = CB - CD = \frac{\sqrt{5} - 1}{2} = \tau^{-1}.$$

Проведя дугу DB с центром в точке B до ее пересечения с отрезком AB в точке E , получим деление отрезка AB золотым сечением, поскольку

$$\frac{AB}{EB} = \frac{EB}{AE} = \tau, \quad AB = 1 = EB + AE = \tau^{-1} + \tau^{-2}.$$

Знаменитый астроном Иоганн Кеплер выразил свое восхищение золотым сечением: «В геометрии существует два сокровища – теорема Пифагора и деление отрезка в крайнем и среднем отношении. Первое можно сравнить с ценностью золота, второе можно назвать драгоценным камнем».

Золотое сечение обладает интересными алгебраическими свойствами. Из уравнения золотого сечения (3.19)

$$\tau^2 = \tau + 1. \quad (3.21)$$

Это тождество может быть представлено в виде

$$\tau = 1 + \frac{1}{\tau} \quad (3.22)$$

откуда

$$\tau - 1 = \frac{1}{\tau}. \quad (3.23)$$

Чтобы получить обратное к τ число, достаточно вычесть из τ единицу.

Если в правую часть равенства (3.22) вместо τ подставить его значение, задаваемое этим же равенством, то получим представление τ в виде многоэтажной дроби

$$\tau = 1 + \frac{1}{1 + \frac{1}{\tau}}.$$

Если продолжить такую подстановку в правой части бесконечное число раз, то в результате получим цепную дробь

$$\tau = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}}.$$

Вернемся к тождеству (3.21), которое представим в форме

$$\tau = \sqrt{1 + \tau}. \quad (3.24)$$

Если в правой части этого тождества вместо τ подставить его значение, задаваемое этим же тождеством, то получим представление τ в виде

$$\tau = \sqrt{1 + \sqrt{1 + \tau}}.$$

Подставляя в правую часть этого тождества выражение (3.24) вместо τ и повторяя эту операцию бесконечное число раз, получим представление золотого сечения в радикалах

$$\tau = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \dots}}}}.$$

Умножив все члены тождества (3.21) на τ , а затем разделим их на τ , получим два новых тождества:

$$\tau^3 = \tau^2 + \tau, \quad \tau = 1 + \tau^{-1}.$$

В результате приходим к тождеству, связывающему степени золотого сечения,

$$\tau^n = \tau^{n-1} + \tau^{n-2},$$

где $n = 0, \pm 1, \pm 2, \pm 3, \dots$. Любая целая степень золотого сечения равна сумме двух предыдущих.

Рассмотрим последовательность степеней золотого сечения

$$\dots, \tau^{-n}, \tau^{-(n-1)}, \dots, \tau^{-2}, \tau^{-1}, \tau^0 = 1, \tau^1, \tau^2, \dots, \tau^{n-1}, \tau^n, \dots,$$

называемую золотой прогрессией. Это единственная последовательность, которая одновременно является геометрической и арифметической прогрессией. Целые части положительных степеней золотого сечения τ^n для n от 1 до 52 представлены в табл. 3.2.

Таблица 3.2 – Целые части степеней золотого сечения

n	$[\tau^n]$	n	$[\tau^n]$	n	$[\tau^n]$	n	$[\tau^n]$
1	1	14	842	27	439204	40	228826127
2	2	15	1364	28	710646	41	370248451
3	4	16	2206	29	1149851	42	599074578
4	6	17	3571	30	1860497	43	969323029
5	11	18	5777	31	3010349	44	1568397607
6	17	19	9349	32	4870846	45	2537720636
7	29	20	15126	33	7881196	46	4106118243
8	46	21	24476	34	12752042	47	6643838879
9	76	22	39602	35	20633239	48	10749957122
10	122	23	64079	36	33385282	49	17393796001
11	199	24	103681	37	54018521	50	28143753123
12	321	25	167761	38	87403803	51	45537549124
13	521	26	271442	39	141422324	52	73681302247

Учитывая (3.20), имеем:

$$\tau = \frac{1+\sqrt{5}}{2}, \quad \tau^{-1} = \frac{-1+\sqrt{5}}{2}.$$

Формулу Бине (3.12) представим в виде

$$F_n = \frac{\tau^{n+1} - (-\tau)^{-(n+1)}}{\sqrt{5}}. \quad (3.25)$$

Таким образом, формула Бине устанавливает связь между числами Фибоначчи и золотым сечением.

3.7. Метод золотого сечения

Недостатками рассмотренного метода Фибоначчи являются необходимость задания количества вычислений n для значений целевой функции $f(x)$ и зависимость надежности метода от значения малого параметра δ . На каждой итерации метода Фибоначчи внутренняя точка текущего интервала неопределенности делит его в отношении двух последовательных чисел Фибоначчи. Так на начальном этапе в соответствии с равенством (3.15) при пренебрежимо малом значении δ интервал неопределенности $[a, b]$ длины $L = b - a$ делится в отношении

$$\frac{L_2}{L} = \frac{F_{n-1}}{F_n}.$$

По формуле Бине (3.25) при $n \rightarrow \infty$ число Фибоначчи F_n эквивалентно $\tau^{n+1}/\sqrt{5}$, поэтому

$$\lim_{n \rightarrow \infty} \frac{F_{n-1}}{F_n} = \frac{\tau^n \sqrt{5}}{\sqrt{5} \tau^{n+1}} = \frac{1}{\tau}.$$

Следовательно, в методе Фибоначчи при больших n интервал неопределенности делится по золотому сечению. Обозначим $t = 1/\tau$. В соответствии с равенством (3.22) $t = \tau - 1$ и по формуле (3.20)

$$t = \frac{\sqrt{5}-1}{2} = 0,6180339887498949\dots \quad (3.26)$$

Как и метод Фибоначчи, метод золотого сечения использует принцип симметрии для двух внутренних точек деления интервала неопределенности. Две первые точки определяются по золотому сечению:

$$x_1 = b - tL, \quad x_2 = a + tL,$$

где $x_1 < x_2$. В этих точках вычисляются значения функции $f_1 = f(x_1)$, $f_2 = f(x_2)$ и применяется правило исключения интервалов, после чего интервал неопределенности уменьшается в τ раз и имеет длину $L_2 = tL$. Оставшаяся внутренняя точка делит новый интервал неопределенности также по золотому сечению. Следующая точка поиска определяется симметрично оставшейся точке по формуле (3.16).

На каждой итерации метода золотого сечения вычисляется значение функции только в одной внутренней точке, а вторая точка со своим значением функции переходит из предыдущей итерации. После этого применяется правило исключения интервалов, и интервал неопределенности уменьшается в τ раз. Итерации продолжаются до тех пор, пока длина интервала неопределенности не станет меньше допустимой погрешности ε , то есть пока не выполнится условие $L_n < \varepsilon$.

Оценим эффективность метода золотого сечения. В результате n вычислений функции получим интервал длины

$$L_n = t^{n-1}L = \frac{L}{\tau^{n-1}}.$$

Отношение длин конечного и начального интервалов неопределенности составит

$$r_n = \frac{1}{\tau^{n-1}}. \quad (3.27)$$

Если интервал неопределенности необходимо уменьшить не менее чем в сто раз, то на основании степеней золотого сечения τ^n из табл. 3.2 получим:

$$r_n \leq \frac{1}{100}, \quad \frac{1}{\tau^{n-1}} \leq \frac{1}{100}, \quad \tau^{n-1} \geq 100, \quad [\tau^{10}] = 122 > 100, \quad n = 11.$$

Если же интервал неопределенности необходимо уменьшить не менее чем в 1000 раз, то $\tau^{n-1} \geq 1000$, $[\tau^{15}] = 1364 > 1000$, и можно положить $n = 16$. Эти результаты показывают, что эффективность метода золотого сечения практически такая же, как и метода Фибоначчи. Для более точного сравнения эффективности этих методов найдем предел отношения для них длин конечных интервалов неопределенности L_n :

$$\lim_{n \rightarrow \infty} \frac{L/\tau^{n-1}}{L/F_n} = \lim_{n \rightarrow \infty} \frac{F_n}{\tau^{n-1}} = \lim_{n \rightarrow \infty} \frac{\tau^{n+1}/\sqrt{5}}{\tau^{n-1}} = \frac{\tau^2}{\sqrt{5}} = 1,1708\dots$$

При больших n длина интервала неопределенности после n вычислений значений целевой функции в методе золотого сечения лишь примерно на 17 % больше, чем в методе Фибоначчи. Однако метод золотого сечения обладает тем существенным преимуществом, что в нем вычисления можно прекратить в любой момент, так как точки, в которых проводятся вычисления, не зависят от их общего числа n .

Алгоритм метода золотого сечения.

Входные параметры: a и b – концы интервала неопределенности, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $t = 0,5 \cdot (\sqrt{5} - 1)$, $h = t \cdot (b - a)$, $x = b - h$, $f_x = f(x)$, $y = a + h$, $f_y = f(y)$.
2. Если $f_x > f_y$, то перейти к шагу 5.
3. Положить $b = y$, $y = x$, $f_y = f_x$.
4. Вычислить $d = b - a$, $x = b - t \cdot d$, $f_x = f(x)$ и перейти к шагу 7.
5. Положить $a = x$, $x = y$, $f_x = f_y$.

6. Вычислить $d = b - a$, $y = a + t \cdot d$, $f_y = f(y)$.

7. Если $d > \varepsilon$, то перейти к шагу 2.

8. Если $f_y < f_x$, то положить $x = y$, $f_x = f_y$.

9. Остановиться.

Очевидно, что алгоритм метода золотого сечения проще алгоритма метода Фибоначчи. Параметр метода t золотого сечения следует задавать как можно точнее по формуле (3.26), иначе в результате погрешностей вычислений интервал, содержащий минимум, может быть утерян. Признаком подобной ситуации в приведенном алгоритме является выполнение неравенства $x \geq y$. При этом рекомендуется перейти к шагу 1 алгоритма.

3.8. Метод адаптации шага

Рассмотренный ранее метод деления интервала пополам требует нахождения интервала неопределенности, который ищется методом Свенна. Дальнейшим развитием сочетания этих двух методов является метод адаптации шага, который разработан харьковским математиком В. Ф. Коропом. Метод заключается в умножении текущего шага поиска на некоторый коэффициент, который вычисляется на основании информации предшествующего поиска.

Для начала работы метода адаптации шага задаются начальная точка x_0 , начальный шаг $h_0 > 0$ и начальное значение множителя шага $r_0 = 0$. В точке x_0 вычисляется значение целевой функции $f_0 = f(x_0)$.

Итерация с номером $k+1$ метода адаптации шага выполняются по результатам итерации с номером $k = 0, 1, 2, \dots$. Вначале итерации вычисляется следующая пробная точка поиска:

$$y_{k+1} = x_k + h_k, \quad g_{k+1} = f(y_{k+1}). \quad (3.28)$$

Затем определяется лучшая точка поиска путем сравнения нового зна-

чения целевой функции g_{k+1} с лучшим значением предыдущих итераций f_k :

$$x_{k+1} = \begin{cases} y_{k+1}, & g_{k+1} < f_k; \\ x_k, & g_{k+1} \geq f_k; \end{cases} \quad f_{k+1} = \begin{cases} g_{k+1}, & g_{k+1} < f_k; \\ f_k, & g_{k+1} \geq f_k. \end{cases} \quad (3.29)$$

На основании сравнения g_{k+1} с f_k и предыдущего значения множителя шага r_k вычисляется новое значение множителя шага

$$r_{k+1} = \begin{cases} 2, & g_{k+1} < f_k \wedge r_k \geq 0,5; \\ 0,5, & g_{k+1} < f_k \wedge r_k < 0,5; \\ 0,25, & g_{k+1} \geq f_k \wedge r_k = 2; \\ -0,5, & g_{k+1} \geq f_k \wedge r_k \neq 2. \end{cases} \quad (3.30)$$

Умножением на это значение величины предыдущего шага находится значение следующего шага

$$h_{k+1} = r_{k+1} h_k. \quad (3.31)$$

Итерации продолжаются до тех пор, пока не выполнится условие $|h_{k+1}| < \varepsilon$, где ε – допустимая погрешность.

Рассмотрим особенности процесса одномерного поиска для метода адаптации шага по формулам (3.28)–(3.31). Пусть в первой итерации из точки x_0 со значением функции f_0 выполняется начальный шаг с переходом к пробной точке $y_1 = x_0 + h_0$ и вычислением значения функции $g_1 = f(y_1)$. Если $g_1 < f_0$, то полагают $x_1 = y_1$, $f_1 = g_1$, множителю адаптации шага задают значение $r_1 = 0,5$, вычисляют следующий шаг $h_1 = r_1 h_0 = 0,5 h_0$ и переходят к следующей пробной точке $y_2 = x_1 + h_1$ с вычислением значения функции $g_2 = f(x_2)$.

Если значения функции продолжают уменьшаться, то выполняются итерации в положительном направлении по рекуррентной формуле удвоения шага

$$r_k = 2, \quad h_k = r_k h_{k-1}, \quad y_{k+1} = x_k + h_k, \quad k = \overline{3, m} \quad (3.32)$$

с вычислением значений функции $g_k = f(y_k)$ и выполнением равенств $x_k = y_k$, $f_k = g_k$ до тех пор, пока не будет получена точка y_{k+1} с не меньшим значением функции $g_{k+1} = f(y_{k+1})$, чем значение функции в предыдущей точке: $f_{m-1} > f_m$, $f_m \leq g_{m+1}$. Этот процесс вычислений для $m=3$ представлен на рис. 3.9, где $h_2 = r_2 h_1 = h_0$, $h_3 = r_3 h_2 = 2h_0$.

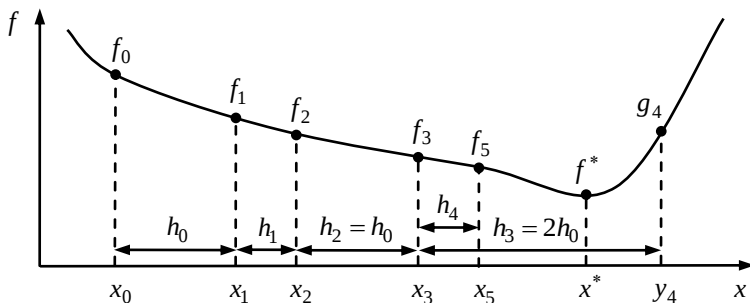


Рис. 3.9. Поиск методом адаптации шага в положительном направлении

В этом случае действия метода адаптации шага похожи на действия метода Свенна за исключением более осторожного шага $h_1 = h_0/2$, что подтверждается сравнением формул (2.2) и (3.32), а также рис. 2.12 и рис. 3.9:

$$x_{k+1} = x_k + 2^{k-2} h_0, \quad k = \overline{1, m-1}.$$

Далее в соответствии с формулами (3.29) и (3.30) полагают $x_{m+1} = x_m$, $f_{m+1} = f_m$, $r_{m+1} = 0,25$, вычисляют значение следующего шага $h_{m+1} = r_{m+1} h_m$ и выполняют следующий шаг: $y_{m+2} = x_{m+1} + h_{m+1}$, $g_{m+2} = f(x_{m+2})$. Если $g_{m+2} < f_{m+1}$, то полагают $x_{m+2} = y_{m+2}$,

$f_{m+2} = g_{m+2}$, $r_{m+2} = 0,5$ и вычисляют $h_{m+2} = r_{m+2}h_{m+1}$ (рис. 3.9).

Если же после начального шага $g_1 \geq f_0$, то в соответствии с формулой (3.29) в итерации с номером $k=1$ полагают $x_1 = x_0$, $f_1 = f_0$, множителю адаптации шага по формуле (3.30) задают значение $r_1 = -0,5$ и вычисляют $h_1 = r_1 h_0 = -0,5h_0$, $y_2 = x_1 + h_1 = x_0 - 0,5h_0$, $g_2 = f(y_2)$, что представлено на рис. 3.10.

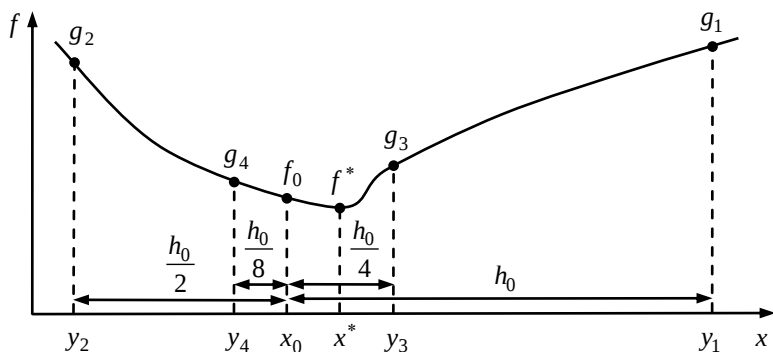


Рис. 3.10. Изменение направления поиска методом адаптации шага

Если и в последующих итерациях с номерами $k=2, 3, \dots$ значения целевой функции не уменьшаются по сравнению со значением f_0 , то есть выполняется условие $g_k = f(y_k) \geq f_{k-1} = f_0$, то полагают

$$r_k = -0,5, \quad h_k = r_k h_{k-1}, \quad y_{k+1} = x_k + h_k = x_0 + h_k.$$

Таким образом, в этом случае на каждой итерации направление поиска меняется на противоположное (рис. 3.10).

Если же на первой итерации $g_1 \geq f_0$, то по формулам (3.29) и (3.30) полагают $x_1 = x_0$, $f_1 = f_0$, $r_1 = -0,5$, $h_1 = r_1 h_0 = -0,5h_0$. На второй итерации вычисляются $y_2 = x_1 + h_1 = x_0 - 0,5h_0$, $g_2 = f(y_2)$ и, если

$g_2 < f_0$, то полагают $x_2 = y_2$, $f_2 = g_2$, множителю адаптации шага по формуле (3.30) задают значение $r_2 = 0,5$ и вычисляют $h_2 = r_2 h_1 = -0,25 h_0$.

Если значения функции продолжают уменьшаться, то выполняются итерации в отрицательном направлении по рекуррентной формуле удвоения шага (3.32) с вычислением значений функции $g_k = f(y_k)$ и выполнением равенств $x_k = y_k$, $f_k = g_k$ до тех пор, пока не будет получена точка y_{k+1} с не меньшим значением функции $g_{k+1} = f(y_{k+1})$, чем значение функции в предыдущей точке: $f_{m-1} > f_m$, $f_m \leq g_{m+1}$. Этот процесс вычислений для $m=4$ представлен на рис. 3.11, где $h_3 = r_3 h_2 = -0,5 h_0$, $h_4 = r_4 h_3 = -h_0$.

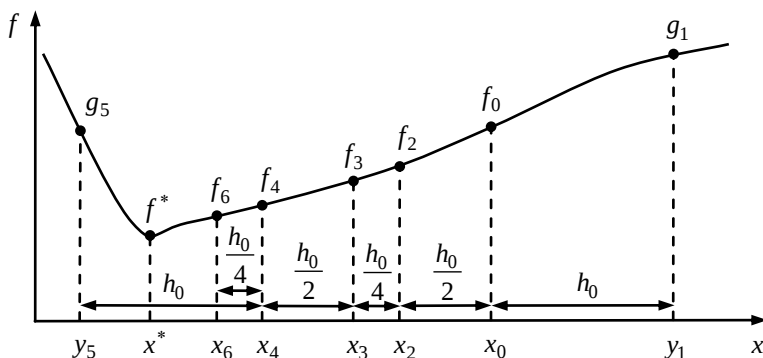


Рис. 3.11. Поиск методом адаптации шага в отрицательном направлении

В этом случае действия метода адаптации шага также похожи на действия метода Свенна за исключением более осторожных трех первых шагов в отрицательном направлении $h_1 = -h_0/2$, $h_2 = -h_0/4$ и $h_3 = -h_0/2$, что подтверждается сравнением формул (2.3) и (3.32), а также рис. 2.14 и рис. 3.11:

$$x_{k+1} = x_k - 2^{k-4} h_0, \quad k = \overline{2, m-1}.$$

Далее в соответствии с формулами (3.29) и (3.30) полагают $x_{m+1} = x_m$, $f_{m+1} = f_m$, $r_{m+1} = 0,25$, вычисляют значение следующего шага $h_{m+1} = r_{m+1} h_m$ и выполняют следующий шаг: $y_{m+2} = x_{m+1} + h_{m+1}$, $g_{m+2} = f(x_{m+2})$. Если $g_{m+2} < f_{m+1}$, то полагают $x_{m+2} = y_{m+2}$, $f_{m+2} = g_{m+2}$, $r_{m+2} = 0,5$ и вычисляют $h_{m+2} = r_{m+2} h_{m+1}$ (рис. 3.11).

В любом случае лучшие точки поиска x_k будут приближаться к оптимальной точке x^* до тех пор, пока не выполнится условие $|h_{k+1}| < \varepsilon$. Таким образом, метод адаптации шага основан как на идее удвоения шага в методе Свенна при удачном направлении поиска, так и на идее деления шага на два или на четыре, что позволяет быстро приближаться к оптимальной точке. Если же задать начальное значение $r_0 \geq 0,5$, то сходство с методом Свенна увеличится.

Алгоритм метода адаптации шага.

Входные параметры: x – начальная точка поиска, h – начальный шаг, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Положить $r = 0$ и вычислить $f_x = f(x)$.
2. Вычислить $y = x + h$, $f_y = f(y)$.
3. Если $f_y \geq f_x$, то перейти к шагу 7.
4. Положить $x = y$, $f_x = f_y$.
5. Если $r \geq 0,5$, то положить $r = 2$, иначе положить $r = 0,5$.
6. Перейти к шагу 8.
7. Если $r = 2$, то положить $r = 0,25$, иначе положить $r = -0,5$.
8. Положить $h = h \cdot r$.
9. Если $|h| > \varepsilon$, то перейти к шагу 2.

10. Остановиться.

Сравнивая алгоритм метода адаптации шага с алгоритмами методов Свенна и деления интервала пополам, которые совместно решают ту же задачу минимизации целевой функции одной переменной, что и данный метод, можно сделать вывод, что алгоритм метода адаптации шага проще объединения двух указанных алгоритмов. К тому же в методе адаптации шага применяются только простые операции сложения, удвоения шага, деления шага на два и на четыре. Поэтому данный метод является надежным и эффективным на практике.

3.9. Сравнение методов уменьшения интервала неопределенности

Мы рассмотрели различные методы одномерного поиска, предназначенные для уменьшения интервала неопределенности точки минимума x^* унимодальной целевой функции одной переменной $f(x)$ без использования производных. Эти методы нулевого порядка, позволяющие находить минимум унимодальной функции произвольного вида, по объему исходной информации для их алгоритмов можно разделить на две группы.

1. Методы, использующие значения границ интервала неопределенности. Это методы равномерного поиска, дихотомии, деления интервала пополам, Фибоначчи и золотого сечения.

2. Метод, не использующий интервал неопределенности. Это один только метод адаптации шага.

Для работы методов из первой группы необходимо найти интервал неопределенности методом Свенна. Эффективность методов из этой группы можно оценить по значениям коэффициента уменьшения длины интервала неопределенности при n вычислениях значений целевой функции

$$r_n = \frac{L_n}{L}.$$

Формулы этого коэффициента для метода равномерного поиска (МРП) (3.1), метода дихотомии (МД) (3.2), метода деления интервала пополам (МДИП) (3.3), метода Фибоначчи (МФ) (3.17) и метода золотого сечения (МЗС) (3.27) для удобства сравнения сведены в табл. 3.3.

Таблица 3.3 – Коэффициенты уменьшения интервала

МРП	МД	МДИП	МФ	МЗС
$r_n = \frac{2}{n+1}$	$r_n = \frac{1}{2^{n/2}}$	$\frac{1}{2^{n-1}} \leq r_n \leq \frac{1}{2^{(n-1)/2}}$	$r_n = \frac{1}{F_n}$	$r_n = \frac{1}{\tau^{n-1}}$

На основании этой таблицы, а также табл. 3.1 и табл. 3.2 для чисел Фибоначчи и степеней золотого сечения составлена табл. 3.4, в которой для тех же методов представлено количество вычислений целевой функции, необходимое для достижения требуемого уменьшения интервала неопределенности, которое задано значением коэффициента уменьшения интервала r_n .

Таблица 3.4 – Количество вычислений функции

r_n	МРП	МД	МДИП	МФ	МЗС
10^{-1}	19	8	[5; 9]	6	6
10^{-2}	199	14	[8; 15]	11	11
10^{-3}	1999	20	[11; 21]	16	16
10^{-4}	19999	28	[15; 29]	20	21
10^{-5}	199999	34	[18; 35]	25	25
10^{-6}	1999999	40	[21; 41]	30	30
10^{-7}	19999999	48	[25; 49]	35	35
10^{-8}	199999999	54	[28; 55]	39	40
10^{-9}	1999999999	60	[31; 61]	44	45
10^{-10}	19999999999	68	[35; 69]	49	49

По этой таблице видно, что теоретически самым эффективным методом уменьшения интервала неопределенности является метод Фибоначчи. Незначительно уступает ему в эффективности метод золотого сечения. Верхняя оценка количества вычислений функции методом деления интервала пополам лишь на единицу превышает те же показатели для метода дихотомии, однако нижняя оценка почти вдвое ниже метода дихотомии и даже меньше, чем у метода Фибоначчи. Метод равномерного поиска для достижения той же точности, что и другие методы, тратит на несколько порядков больше вычислений значений функции. Самым надежным из этих методов уменьшения интервала неопределенности является метод деления интервала пополам, не зависящий от дополнительных параметров. Его надежность сравнима с надежностью метода адаптации шага.

Лабораторная работа

Тема лабораторной работы.

Разработка подпрограммы для метода уменьшения интервала неопределенности.

Цель лабораторной работы.

Разработать подпрограмму метода уменьшения интервала неопределенности унимодальной целевой функции.

Порядок выполнения лабораторной работы.

1. Получить у преподавателя задание на программирование одного из методов уменьшения интервала неопределенности:

- 1) метод равномерного поиска;
- 2) метод дихотомии;
- 3) метод деления интервала пополам;
- 4) метод Фибоначчи;
- 5) метод золотого сечения;
- 6) метод адаптации шага.

2. На основании алгоритма для заданного метода уменьшения интервала неопределенности написать компьютерную подпрограмму в

соответствии с требованиями:

1) входные параметры подпрограммы: $f(x)$ – целевая функция, N – предельное число точек, ε – допустимая погрешность, входные параметры алгоритма заданного метода;

2) выходные параметры подпрограммы: X – массив всех точек поиска x_k ; F – массив соответствующих значений функции $f_k = f(x_k)$; x и f_x – лучшая точка поиска и значение в ней функции;

3) на каждой итерации выводить на экран компьютера строку, содержащую номер точки поиска k , длину интервала неопределенности L_k , значение функции f_k и значение переменной x_k ;

4) итерации продолжаются до тех пор, пока выполняются неравенства:

$$|L_k| > \varepsilon, \quad k \leq N;$$

5) после окончания процесса оптимизации на экране должна отображаться таблица, имеющая соответствующий заголовок и представляющая процесс минимизации функции;

6) под таблицей необходимо отобразить количество вычислений целевой функции, длину конечного интервала неопределенности, минимальное значение функции и соответствующее ему значение независимой переменной.

3. Написать общую программу для выполнения вычислительного процесса минимизации целевой функции одной переменной и отображения на экране этого процесса в соответствии с требованиями:

1) за основу общей программы для минимизации функции взять общую программу поиска интервала неопределенности из предыдущей лабораторной работы и модифицировать эту программу;

2) задать в программе значения x_0 , x^* , h , ε , N ;

3) путем вызова подпрограммы метода Свенна с использованием подпрограммы вычисления функции $f(x)$ и значений x_0 , h , N реализовать итерационный процесс поиска интервала неопределенности с

построением таблицы процесса на экране и получением массивов значений переменной X и функции F , границ интервала неопределенности a и b ;

4) путем вызова разработанной подпрограммы уменьшения интервала неопределенности реализовать итерационный процесс уменьшения интервала с построением таблицы процесса на экране и получением массивов значений переменной X и функции F , лучшей точки поиска x и значения в ней функции f_x ;

5) с помощью подпрограммы графического отображения итерационного процесса по массивам X и F отобразить график процесса минимизации целевой функции одной переменной на экране компьютера.

4. Запустить на счет общую программу минимизации функции одной переменной для функций предыдущих лабораторных работ из нескольких начальных точек

$$x_{01} < x^*, \quad x_{02} = x^*, \quad x_{03} > x^*$$

при нескольких значениях шага

$$h_{01} < |x_0 - x^*|, \quad h_{02} = |x_0 - x^*|, \quad h_{03} > |x_0 - x^*|.$$

5. Выполнить анализ полученных результатов:

1) на основании табличной и графической информации о работе методов одномерного поиска определить конечные точки метода Свенна и найденный этим методом интервал неопределенности;

2) на нескольких начальных итерациях по табличным данным и графикам сопоставить начало работы метода уменьшения интервала с его рабочими формулами;

3) оценить эффективность метода уменьшения интервала и сопоставить ее с теоретической эффективностью.

6. Ответить на вопросы преподавателя по теме лабораторной работы.

7. Оформить и сдать отчет о проведенной лабораторной работе.

Содержание отчета.

1) титульный лист установленного образца с указанием организации, названия учебной дисциплины, темы работы, номера варианта, исполнителя и принимающего, города, года;

2) постановка задачи минимизации целевой функции, включающая задание функции, начальную точку и точку минимума;

3) теоретическое описание метода уменьшения интервала;

4) алгоритм метода;

5) распечатки используемых компьютерных программ;

6) табличное представление процесса минимизации и его итоги – количество вычислений значений целевой функции, минимальное значение функции и соответствующее ему значение точки минимума, достигнутую точность;

7) двумерный график траектории поиска со всеми точками;

8) анализ полученных результатов;

9) выводы о проделанной лабораторной работе.

Контрольные вопросы

1. Чем отличается пассивная стратегия поиска?
2. Какое назначение метода равномерного поиска?
3. Опишите метод равномерного поиска.
4. Как оценивается эффективность метода равномерного поиска?
5. Какие преимущества и недостатки метода равномерного поиска?
6. Чем отличаются последовательные методы одномерного поиска?
7. Какое назначение метода дихотомии?
8. Какие основные принципы метода дихотомии?
9. Опишите метод дихотомии.
10. Как оценивается эффективность метода дихотомии?
11. Какого порядка метод дихотомии?
12. В чем заключаются преимущества и недостатки метода дихо-

томии?

13. Какое назначение метода деления интервала пополам?

14. Опишите метод деления интервала пополам.

15. Как оценивается эффективность метода деления интервала пополам?

16. Какого порядка метод деления интервала пополам?

17. Какие достоинства и недостатки метода деления интервала пополам?

18. Какая задача привела к открытию чисел Фибоначчи?

19. По какой рекуррентной формуле определяются числа Фибоначчи?

20. Запишите 10 первых чисел Фибоначчи.

21. Укажите свойства чисел Фибоначчи.

22. Выведите формулу Бине для чисел Фибоначчи.

23. Какое назначение метода Фибоначчи?

24. Опишите метод Фибоначчи.

25. Выведите общую формулу для длин интервалов неопределенности в методе Фибоначчи.

26. Как начинается поиск минимума в методе Фибоначчи?

27. Как оценивается эффективность метода Фибоначчи?

28. Какого порядка метод Фибоначчи?

29. В чем заключаются преимущества и недостатки метода Фибоначчи?

30. Какая задача привела к открытию золотого сечения?

31. Выведите уравнение золотого сечения.

32. Приведите численное представление золотого сечения.

33. Опишите способ геометрического построения золотого сечения с использованием линейки и циркуля.

34. Укажите алгебраические свойства золотого сечения.

35. Какова особенность золотой прогрессии?

36. Какова связь между числами Фибоначчи и золотым сечением?

37. Какое назначение метода золотого сечения?

38. Опишите метод золотого сечения.

39. В чем заключается сходство и различие методов Фибоначчи и золотого сечения.

40. Как начинается поиск минимума в методе золотого сечения?

41. Как оценивается эффективность метода золотого сечения?

42. Сравните эффективность методов Фибоначчи и золотого сечения.

43. Какого порядка метод золотого сечения?

44. Какие преимущества и недостатки метода золотого сечения?

45. Какое назначение метода адаптации шага?

46. Опишите метод адаптации шага.

47. По какому правилу изменяется шаг поиска в методе адаптации шага?

48. Как оценивается эффективность метода адаптации шага?

49. Какого порядка метод адаптации шага?

50. Какие достоинства и недостатки метода адаптации шага?

51. На какие группы можно разделить методы уменьшения интервала неопределенности?

52. Сравните эффективность методов уменьшения интервала неопределенности.

4. МЕТОДЫ АППРОКСИМАЦИИ И ИНТЕРПОЛЯЦИИ

Не существует единственного самого эффективного метода для одномерного поиска минимума функции произвольного вида. В этом разделе рассматриваются методы нулевого, первого и второго порядков, основанные на аппроксимации унимодальной целевой функции полиномом с последующей интерполяцией точки минимума целевой функции в точке минимума полиномиальной функции. Обосновываются методы квадратичной интерполяции с тремя точками, кубической интерполяции с четырьмя точками, квадратичной интерполяции с двумя точками, секущих, кубической интерполяции с двумя точками, Ньютона-Рафсона. Дается сравнение скорости асимптотической сходимости методов аппроксимации и интерполяции. Приводятся алгоритмы для всех рассмотренных методов. Дается описание лабораторной работы по изучению методов аппроксимации и интерполяции для минимизации унимодальной целевой функции одной переменной.

4.1. Метод квадратичной интерполяции с тремя точками

Рассмотренные ранее методы одномерного поиска уменьшают интервал неопределенности в наперед заданное число раз независимо от вида самой целевой функции. Однако, если функции близки к полиномам, то более эффективны методы полиномиальной аппроксимации и интерполяции.

Можно показать, что любая гладкая функция $f(x)$ в малой окрестности ее точки минимума x^* близка к квадратичной функции. Поэтому такую функцию можно аппроксимировать квадратичным многочленом

$$y = \alpha x^2 + \beta x + \gamma, \quad (4.1)$$

где коэффициенты α , β и γ определяются по трем точкам, полученным с помощью алгоритма метода Свенна – (a, f_a) , (x, f_x) и (b, f_b) , где $a < x < b$. Дифференцируя уравнение параболы (4.1), имеем

$$y' = 2\alpha x + \beta.$$

По необходимому условию существования экстремума $y' = 0$ получим точку минимума параболы

$$z = -\frac{\beta}{2\alpha}. \quad (4.2)$$

Для определения этой точки достаточно определить лишь параметры α и β уравнения (4.1). Подставляя в уравнение параболы (4.1) координаты трех известных точек функции $f(x)$, составим систему уравнений:

$$\begin{cases} \alpha a^2 + \beta a + \gamma = f_a, \\ \alpha x^2 + \beta x + \gamma = f_x, \\ \alpha b^2 + \beta b + \gamma = f_b. \end{cases}$$

Вычитая из второго уравнения первое, а из третьего – второе, имеем:

$$\begin{cases} \alpha(x^2 - a^2) + \beta(x - a) = f_x - f_a, \\ \alpha(b^2 - x^2) + \beta(b - x) = f_b - f_x. \end{cases}$$

Решим полученную систему по правилу Крамера:

$$D = \begin{vmatrix} x^2 - a^2 & x - a \\ b^2 - x^2 & b - x \end{vmatrix} = (x^2 - a^2)(b - x) - (b^2 - x^2)(x - a),$$

$$D_\alpha = \begin{vmatrix} f_x - f_a & x - a \\ f_b - f_x & b - x \end{vmatrix} = (f_x - f_a)(b - x) - (f_b - f_x)(x - a),$$

$$D_{\beta} = \begin{vmatrix} x^2 - a^2 & f_x - f_a \\ b^2 - x^2 & f_b - f_x \end{vmatrix} = (x^2 - a^2)(f_b - f_x) - (b^2 - x^2)(f_x - f_a),$$

$$\alpha = \frac{D_{\alpha}}{D}, \quad \beta = \frac{D_{\beta}}{D}.$$

Для уменьшения количества арифметических операций введем обозначения:

$$p = (x - a)(f_b - f_x), \quad q = (b - x)(f_a - f_x). \quad (4.3)$$

Тогда получим выражения для определителей:

$$D_{\alpha} = -(p + q), \quad D_{\beta} = p(x + a) + q(b + x).$$

Подставив определители в формулу для точки минимума параболы (4.2), окончательно получим:

$$z = \frac{p(x + a) + q(b + x)}{2(p + q)}. \quad (4.4)$$

Таким образом, первоначальный интервал неопределенности $[a, b]$, как и в большинстве методов предыдущего раздела, делится двумя точками x и z на три части. По правилу исключения интервалов на основании сравнения значений целевой функции во внутренних точках интервала неопределенности $f_x = f(x)$ и $f_z = f(z)$ та часть, которая заведомо не содержит минимума, отбрасывается, и интервал $[a, b]$ уменьшается. Для оставшегося интервала с известной одной внутренней точкой опять применяется квадратичная интерполяция по формуле (4.4). Итерации продолжаются до тех пор, пока либо расстояние между точками x и z не станет меньше заданной наперед величины ε , либо длина интервала неопределенности не станет меньше допустимой погрешности.

Вследствие ошибок округления при вычислениях может оказаться, что значение z , вычисленное по формуле (4.4), окажется вне ин-

тервала неопределенности. В таком случае необходимо предусмотреть выход из цикла, несмотря на то, что заданная точность не была достигнута.

Рассмотренный метод нулевого порядка называется методом квадратичной интерполяции с тремя точками. На основании рабочих формул (4.3) и (4.4) составлен алгоритм метода квадратичной интерполяции.

Алгоритм метода квадратичной интерполяции с тремя точками.

Входные параметры: a и b – концы интервала неопределенности, f_a и f_b – значения функции на концах интервала, x и f_x – внутренняя точка интервала и значение функции в ней, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $p = (x - a)(f_b - f_x)$, $q = (b - x)(f_a - f_x)$, $s = p + q$.
2. Если $s \leq 0$, то перейти к шагу 13.
3. Вычислить $y = 0,5 \cdot (p(a + x) + q(b + x)) / s$.
4. Если $y \leq a$ или $y \geq b$, то перейти к шагу 13.
5. Вычислить $f_y = f(y)$ и, если $f_y > f_x$, то перейти к шагу 8.
6. Если $y < x$, то положить $b = x$, $f_b = f_x$, $x = y$, $f_x = f_y$ и перейти к шагу 10.
7. Положить $a = x$, $f_a = f_x$, $x = y$, $f_x = f_y$ и перейти к шагу 10.
8. Если $y < x$, то положить $a = y$, $f_a = f_y$ и перейти к шагу 10.
9. Положить $b = y$, $f_b = f_y$.
10. Если $|x - y| \geq \varepsilon$, то перейти к шагу 1.
11. Если $f_a < f_x$, то положить $x = a$, $f_x = f_a$.
12. Если $f_b < f_x$, то положить $x = b$, $f_x = f_b$.
13. Остановиться.

4.2. Метод кубической интерполяции с четырьмя точками

Метод кубической интерполяции основан на аппроксимации минимизируемой функции $f(x)$ полиномом третьей степени

$$y = \alpha x^3 + \beta x^2 + \gamma x + \delta. \quad (4.5)$$

Запишем необходимое условие минимума этого полинома

$$y' = 3\alpha x^2 + 2\beta x + \gamma = 0.$$

Решая это квадратное уравнение, получим его корни:

$$z_{1,2} = \frac{-\beta \pm \sqrt{\beta^2 - 3\alpha\gamma}}{3\alpha}.$$

Минимуму функции соответствует тот корень, для которого

$$y''(x) = 6\alpha x + 2\beta > 0.$$

Имеем

$$y''(z_{1,2}) = 6\alpha \frac{-\beta \pm \sqrt{\beta^2 - 3\alpha\gamma}}{3\alpha} + 2\beta = \pm 2\sqrt{\beta^2 - 3\alpha\gamma}.$$

Условию минимума удовлетворяет корень

$$z = \frac{\sqrt{\beta^2 - 3\alpha\gamma} - \beta}{3\alpha}. \quad (4.6)$$

Четыре неизвестных параметра полинома (4.5) определим по четырём точкам минимизируемой функции $f(x)$: (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , (x_4, y_4) . Три из этих точек получим в результате поиска интервала неопределённости методом Свенна, а четвертую точку – с помощью квадратичной интерполяции по первым трем точкам. Подставляя в равенство (4.5) координаты четырех точек, составим систему из четырех линейных алгебраических уравнений

$$\alpha x_i^3 + \beta x_i^2 + \gamma x_i + \delta = y_i, \quad i = \overline{1, 4}.$$

Решение этой системы для трех параметров формулы (4.6) получим по правилу Крамера:

$$\alpha = \frac{D_\alpha}{D}, \quad \beta = \frac{D_\beta}{D}, \quad \gamma = \frac{D_\gamma}{D},$$

где

$$D = \begin{vmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ x_3^3 & x_3^2 & x_3 & 1 \\ x_4^3 & x_4^2 & x_4 & 1 \end{vmatrix}, \quad D_\alpha = \begin{vmatrix} y_1 & x_1^2 & x_1 & 1 \\ y_2 & x_2^2 & x_2 & 1 \\ y_3 & x_3^2 & x_3 & 1 \\ y_4 & x_4^2 & x_4 & 1 \end{vmatrix},$$

$$D_\beta = \begin{vmatrix} x_1^3 & y_1 & x_1 & 1 \\ x_2^3 & y_2 & x_2 & 1 \\ x_3^3 & y_3 & x_3 & 1 \\ x_4^3 & y_4 & x_4 & 1 \end{vmatrix}, \quad D_\gamma = \begin{vmatrix} y_1 & x_1^2 & y_1 & 1 \\ y_2 & x_2^2 & y_2 & 1 \\ y_3 & x_3^2 & y_3 & 1 \\ y_4 & x_4^2 & y_4 & 1 \end{vmatrix}.$$

Для рационального вычисления этих определителей введем обозначения:

$$p_1 = y_1(x_3 - x_2)(x_4 - x_2)(x_4 - x_3), \quad p_2 = y_2(x_3 - x_1)(x_4 - x_1)(x_4 - x_3), \quad (4.7)$$

$$p_3 = y_3(x_2 - x_1)(x_4 - x_1)(x_4 - x_2), \quad p_4 = y_4(x_2 - x_1)(x_3 - x_1)(x_3 - x_2), \quad (4.8)$$

$$q_1 = x_1 + x_2, \quad q_2 = x_3 + x_4, \quad r_1 = x_1 \cdot x_2, \quad r_2 = x_3 \cdot x_4. \quad (4.9)$$

Тогда определители примут вид:

$$D = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3),$$

$$D_\alpha = -p_1 + p_2 - p_3 + p_4, \quad (4.10)$$

$$D_\beta = p_1(x_2 + q_2) - p_2(x_1 + q_2) + p_3(q_1 + x_4) - p_4(q_1 + x_3), \quad (4.11)$$

$$D_\gamma = -p_1(q_2x_2 + r_2) + p_2(q_2x_1 + r_2) - p_3(r_1 + q_1x_4) + p_4(r_1 + q_1x_3). \quad (4.12)$$

По формуле (4.6) найдем минимум кубического полинома

$$z = \frac{\sqrt{D_\beta^2 - 3D_\alpha D_\gamma} - D_\beta}{3D_\alpha}. \quad (4.13)$$

Таким образом, интервал неопределенности разделен тремя внутренними точками на четыре части. По правилу исключения интервалов отбрасывается одна из крайних частей и к новому интервалу с двумя внутренними точками опять применяется кубическая интерполяция. Итерации продолжаются до тех пор, пока расстояние между внутренними точками не станет меньше допустимой погрешности.

Рассмотренный метод нулевого порядка называется методом кубической интерполяции с четырьмя точками. На основании формул (4.7)–(4.13) составлен алгоритм метода кубической интерполяции.

Алгоритм метода кубической интерполяции с четырьмя точками.

Входные параметры: a и b – концы интервала неопределенности, f_a и f_b – значения функции на концах интервала, x и f_x – внутренняя точка интервала и значение функции в ней, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $p = (x - a)(f_b - f_x)$, $q = (b - x)(f_a - f_x)$, $s = p + q$.
2. Если $s \leq 0$, то перейти к шагу 32.
3. Вычислить $y = 0,5 \cdot (p(a + x) + q(b + x)) / s$.
4. Если $y \leq a$ или $y \geq b$, то перейти к шагу 32.
5. Вычислить $f_y = f(y)$.
6. Если $y < x$, то положить

$$z = x, \quad f_z = f_x, \quad x = y, \quad f_x = f_y, \quad y = x, \quad f_y = f_x.$$

7. Вычислить $f_{xa} = f_x - f_a$, $f_{yx} = f_y - f_x$, $f_{by} = f_b - f_y$.

8. Вычислить $d_{xa} = x - a$, $d_{yx} = y - x$, $d_{by} = b - y$.

9. Вычислить $s_{xa} = x + a$, $s_{yx} = y + x$, $s_{by} = b + y$.

10. Вычислить

$$p_{xa} = f_{xa} \cdot d_{yx} \cdot d_{by}, \quad p_{yx} = f_{yx} \cdot d_{by} \cdot d_{xa}, \quad p_{by} = f_{by} \cdot d_{xa} \cdot d_{yx}.$$

11. Вычислить

$$D_\alpha = p_{xa} \cdot (s_{yx} - s_{by}) + p_{yx} \cdot (s_{by} - s_{xa}) + p_{by} \cdot (s_{xa} - s_{yx}).$$

12. Если $D_\alpha = 0$, то перейти к шагу 31.

13. Вычислить $a_2 = a^2$, $x_2 = x^2$, $y_2 = y^2$, $b_2 = b^2$.

14. Вычислить

$$q_{xa} = x_2 + x \cdot a + a_2, \quad q_{yx} = y_2 + y \cdot x + x_2, \quad q_{by} = b_2 + b \cdot y + y_2.$$

15. Вычислить

$$D_\beta = p_{xa} \cdot (q_{by} - q_{yx}) + p_{yx} \cdot (q_{xa} - q_{by}) + p_{by} \cdot (q_{yx} - q_{xa}).$$

16. Вычислить

$$D_\gamma = p_{xa} (q_{yx} s_{by} - q_{by} s_{yx}) + p_{yx} (q_{by} s_{xa} - q_{xa} s_{by}) + p_{by} (q_{xa} s_{yx} - q_{yx} s_{xa}).$$

17. Вычислить $D_3 = 3 \cdot D_\alpha$, $D = D_\beta^2 - D_3 \cdot D_\gamma$.

18. Если $D < 0$, то перейти к шагу 31.

19. Вычислить $S_D = \sqrt{D}$.

20. Если $D_\beta > 0$, то положить $z = -(S_D + D_\beta)/D_3$,

иначе положить $z = D_\gamma/(S_D - D_\beta)$.

21. Если $z \leq a$ или $z \geq b$, то перейти к шагу 31.

22. Вычислить $f_z = f(z)$.

23. Если $f_x \geq f_y$, то перейти к шагу 27.

24. Положить $b = y$, $f_b = f_y$.

25. Если $z \geq x$, то положить $y = z$, $f_y = f_z$ и перейти к шагу 30.

26. Положить $y = x$, $f_y = f_x$, $x = z$, $f_x = f_z$ и перейти к шагу 30.

27. Положить $a = x$, $f_a = f_x$.

28. Если $z < y$, то положить $x = z$, $f_x = f_z$ и перейти к шагу 30.

29. Положить $x = y$, $f_x = f_y$, $y = z$, $f_y = f_z$.

30. Если $|x - y| \geq \varepsilon$, то перейти к шагу 7.

31. Если $f_y < f_x$, то положить $x = y$, $f_x = f_y$.

32. Остановиться.

В связи с большим числом операций в формулах кубической интерполяции этот метод подвержен большим ошибкам вычислений и не позволяет находить точку минимума с высокой точностью. Однако, в алгоритмах многомерной оптимизации, где интервал неопределенности необходимо уменьшать в 100–1000 раз, он позволяет существенно сократить число вычислений целевой функции по сравнению с методом квадратичной интерполяции.

4.3. Метод квадратичной интерполяции с двумя точками

Метод квадратичной интерполяции с двумя точками является методом одномерного поиска первого порядка для минимизации унимодальной дифференцируемой целевой функции $f(x)$, который аппроксимирует ее квадратичным полиномом и использует значения ее первой производной $f'(x)$.

Пусть даны две точки x_1 , x_2 , значения функции в них $f_1 = f(x_1)$, $f_2 = f(x_2)$ и значение производной в одной из этих точек $f'_1 = f'(x_1)$ или $f'_2 = f'(x_2)$. Построим квадратичную аппроксимирующую функцию

$$q(x) = \alpha x^2 + \beta x + \gamma, \quad (4.14)$$

параметры которой α , β и γ удовлетворяют системе линейных алгебраических уравнений:

$$\begin{cases} q(x_1) = \alpha x_1^2 + \beta x_1 + \gamma = f_1, \\ q(x_2) = \alpha x_2^2 + \beta x_2 + \gamma = f_2, \\ q'(x_1) = 2\alpha x_1 + \beta = f'_1. \end{cases}$$

Искомый минимум функции (4.14) находится в точке

$$\bar{x} = -\frac{\beta}{2\alpha}.$$

Решая систему линейных алгебраических уравнений, получим:

$$\alpha(x_2^2 - x_1^2) + \beta(x_2 - x_1) = f_2 - f_1,$$

$$\beta = f'_1 - 2\alpha x_1,$$

то есть

$$\alpha(x_2^2 - x_1^2) + (f'_1 - 2\alpha x_1)(x_2 - x_1) = f_2 - f_1,$$

$$\alpha(x_2 - x_1)^2 = f_2 - f_1 - f'_1(x_2 - x_1),$$

откуда

$$\alpha = \frac{f_2 - f_1 - (x_2 - x_1)f'_1}{(x_2 - x_1)^2}.$$

Тогда

$$\bar{x} = -\frac{\beta}{2\alpha} = x_1 - \frac{f'_1}{2\alpha} = x_1 - \frac{1}{2} \frac{f'_1(x_2 - x_1)^2}{f_2 - f_1 - (x_2 - x_1)f'_1}.$$

Окончательно имеем

$$\bar{x} = x_1 - \frac{1}{2} \frac{(x_2 - x_1)^2 f'_1}{f_2 - f_1 - (x_2 - x_1)f'_1}.$$

Мы получили итерационную формулу

$$x_{k+1} = x_{k-1} - \frac{1}{2} \frac{(x_k - x_{k-1})^2 f'_{k-1}}{f_k - f_{k-1} - (x_k - x_{k-1})f'_{k-1}}, \quad k = 2, 3, \dots, \quad (4.15)$$

где $f_k = f(x_k)$, $f_{k-1} = f(x_{k-1})$, $f'_{k-1} = f'(x_{k-1})$.

Используя значение производной $f'_k = f'(x_k)$ итерационную формулу можно представить также в виде

$$x_{k+1} = x_k - \frac{1}{2} \frac{(x_k - x_{k-1})^2 f'_k}{(x_k - x_{k-1}) f'_k - f_k + f_{k-1}}, \quad k = 2, 3, \dots \quad (4.16)$$

Вычисления продолжаются, пока $|x_{k+1} - x_k|$ не станет меньше заданной допустимой погрешности ε вычисления точки минимума x^* , то есть пока не выполнится неравенство $|x - y| < \varepsilon$.

Очевидно, формулы (4.15) и (4.16) проще формул квадратичной интерполяции с тремя точками (4.3) и (4.4).

Метод квадратичной интерполяции с двумя точками, основанный на итерационных формулах (4.15) и (4.16), эффективен при минимизации выпуклых целевых функций.

Алгоритм метода квадратичной интерполяции с двумя точками.

Входные параметры: x и y — начальные точки, ε — допустимая погрешность.

Выходные параметры: x и f_x — лучшая точка и значение функции в ней.

1. Вычислить $f_x = f(x)$, $f'_x = f'(x)$, $f_y = f(y)$.
2. Вычислить $s = x - y$, $p = s f'_x$, $d = 0,5 s p / (p - f_x + f_y)$.
4. Положить $y = x$, $f_y = f_x$, $f'_y = f'_x$.
5. Положить $x = x - d$.
6. Вычислить $f_x = f(x)$, $f'_x = f'(x)$.
7. Если $|x - y| > \varepsilon$, то перейти к шагу 2.
8. Остановиться.

Начальные точки для этого алгоритма могут быть получены методом Свенна.

4.4. Метод секущих

Метод секущих также является методом квадратичной интерполяции первого порядка для минимизации унимодальной дифференцируемой целевой функции $f(x)$ и аппроксимирует ее квадратичным полиномом, используя значения только ее первой производной $f'(x)$.

Пусть даны две точки x_1, x_2 и значения в них производной $f'_1 = f'(x_1), f'_2 = f'(x_2)$. Построим квадратичную аппроксимирующую функцию (4.14), параметры которой удовлетворяют системе линейных алгебраических уравнений:

$$\begin{cases} q'(x_1) = 2\alpha x_1 + \beta = f'_1, \\ q'(x_2) = 2\alpha x_2 + \beta = f'_2. \end{cases}$$

Тогда

$$\alpha = \frac{f'_2 - f'_1}{2(x_2 - x_1)}$$

и

$$\bar{x} = x_1 - \frac{x_2 - x_1}{f'_2 - f'_1} f'_1.$$

Мы получили итерационную формулу

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'_k - f'_{k-1}} f'_{k-1}, \quad k = 2, 3, \dots \quad (4.17)$$

Итерационную формулу можно представить также в виде

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'_k - f'_{k-1}} f'_k, \quad k = 2, 3, \dots \quad (4.18)$$

Формулы (4.17) и (4.18) называются формулами секущих или формулами хорд. Метод одномерного поиска, основанный на итерационных формулах (4.17) или (4.18), также можно назвать методом квад-

ратичной интерполяции с двумя точками. Он эффективен при минимизации выпуклых целевых функций. Метод, основанный на формулах (4.17) или (4.18), часто называется методом секущих или методом хорд.

Алгоритм метода секущих.

Входные параметры: x и y – начальные точки, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $f'_x = f'(x)$, $f'_y = f'(y)$.
2. Вычислить $d = f'_x(x - y) / (f'_x - f'_y)$.
3. Положить $y = x$, $f'_y = f'_x$.
4. Положить $x = x - d$.
5. Вычислить $f_x = f(x)$, $f'_x = f'(x)$.
6. Если $|x - y| > \varepsilon$, то перейти к шагу 2.
7. Остановиться.

Начальные точки для этого алгоритма получим методом Свенна.

4.5. Метод кубической интерполяции с двумя точками

Метод кубической интерполяции с двумя точками является методом первого порядка для минимизации унимодальной дифференцируемой целевой функции $f(x)$. В этом методе функция аппроксимируется кубическим полиномом, для определения точки минимума которого используются как значения самой функции, так и значения ее первой производной $f'(x)$.

Пусть даны две точки x_1 , x_2 , значения функции в них $f_1 = f(x_1)$, $f_2 = f(x_2)$ и значения производной $f'_1 = f'(x_1)$, $f'_2 = f'(x_2)$. Построим кубическую аппроксимирующую функцию в виде полинома третьей степени

$$p(x) = \alpha x^3 + \beta x^2 + \gamma x + \delta. \quad (4.19)$$

Найдем производную этой функции

$$p'(x) = 3\alpha x^2 + 2\beta x + \gamma.$$

Из необходимого условия экстремума $p'(x) = 0$ имеем квадратное уравнение

$$3\alpha x^2 + 2\beta x + \gamma = 0,$$

решая которое получим корни

$$\bar{x}_{1,2} = \frac{-\beta \pm \sqrt{\beta^2 - 3\alpha\gamma}}{3\alpha}.$$

Минимуму функции соответствует тот корень, для которого

$$p''(x) = 6\alpha x + 2\beta > 0.$$

Имеем

$$p''(\bar{x}_{1,2}) = 6\alpha \frac{-\beta \pm \sqrt{\beta^2 - 3\alpha\gamma}}{3\alpha} + 2\beta = \pm 2\sqrt{\beta^2 - 3\alpha\gamma}.$$

Условию минимума удовлетворяет корень

$$\bar{x} = \frac{\sqrt{\beta^2 - 3\alpha\gamma} - \beta}{3\alpha}. \quad (4.20)$$

Однако, эта формула теряет смысл, если в результате аппроксимации окажется $\alpha = 0$. В этом случае

$$p(x) = \beta x^2 + \gamma x + \delta, \quad p'(x) = 2\beta x + \gamma, \quad \bar{x} = \frac{-\gamma}{2\beta}.$$

Для того, чтобы учесть оба случая $\alpha \neq 0$ и $\alpha = 0$, умножим числитель и знаменатель правой части формулы (4.20) на сопряженное для числителя выражение $\sqrt{\beta^2 - 3\alpha\gamma} + \beta$ и после преобразований получим

$$\bar{x} = \frac{-\gamma}{\sqrt{\beta^2 - 3\alpha\gamma} + \beta}. \quad (4.21)$$

Параметры аппроксимирующей функции (4.19) α , β , γ и δ должны удовлетворять системе линейных алгебраических уравнений:

$$\begin{cases} p(x_1) = \alpha x_1^3 + \beta x_1^2 + \gamma x_1 + \delta = f_1, \\ p(x_2) = \alpha x_2^3 + \beta x_2^2 + \gamma x_2 + \delta = f_2, \\ p'(x_1) = 3\alpha x_1^2 + 2\beta x_1 + \gamma = f'_1, \\ p'(x_2) = 3\alpha x_2^2 + 2\beta x_2 + \gamma = f'_2. \end{cases}$$

Решая эту систему, после преобразований формулы (4.21) получим

$$\bar{x} = x_2 - \frac{f'_2 + z - w}{f'_1 + f'_2 + 2z}(x_2 - x_1),$$

где

$$z = f'_1 + f'_2 - 3 \frac{f_2 - f_1}{x_2 - x_1}, \quad w = \sqrt{z^2 - f'_1 f'_2}.$$

По этим формулам построим итерационную формулу

$$x_{k+1} = x_k - \frac{f'_k + z_k - w_k}{f'_k + f'_{k-1} + 2z_k}(x_k - x_{k-1}), \quad (4.22)$$

где

$$z_k = f'_k + f'_{k-1} - 3 \frac{f_k - f_{k-1}}{x_k - x_{k-1}}, \quad w_k = \sqrt{z_k^2 - f'_k f'_{k-1}}. \quad (4.23)$$

Очевидно, эта итерационная формула проще формул кубической интерполяции с четырьмя точками (4.7)–(4.13). Вычисления продолжаютсЯ, пока $|x_{k+1} - x_k|$ не станет меньше заданной допустимой погрешности. Метод кубической интерполяции с двумя точками, основанный на итерационной формуле (4.22) с обозначениями (4.23), эф-

фективен при минимизации выпуклых целевых функций.

Алгоритм метода кубической интерполяции с двумя точками.

Входные параметры: x и y – начальные точки, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Вычислить $f_x = f(x)$, $f'_x = f'(x)$, $f_y = f(y)$, $f'_y = f'(y)$.
2. Вычислить $s = x - y$, $d = f'_x + f'_y$, $z = d - 3(f_x - f_y)/(x - y)$.
3. Вычислить $w = \sqrt{z^2 - f'_x f'_y}$, $r = (f'_x + z - w)/(d + 2z)$.
4. Положить $y = x$, $f_y = f_x$, $f'_y = f'_x$.
5. Положить $x = x - rs$.
6. Вычислить $f_x = f(x)$, $f'_x = f'(x)$.
7. Если $|x - y| > \varepsilon$, то перейти к шагу 1.
8. Остановиться.

4.7. Метод Ньютона-Рафсона

Метод Ньютона-Рафсона для минимизации унимодальной дважды дифференцируемой целевой функции $f(x)$ использует значения ее первой и второй производных $f'(x)$ и $f''(x)$. Это метод второго порядка.

Классический подход при минимизации функции одной переменной $f(x)$ состоит в использовании необходимого условия экстремума, то есть в решении уравнения

$$f'(x) = 0.$$

Если унимодальная функция $f(x)$ дважды дифференцируема и выпукла в интервале неопределенности $[a, b]$, то решить такое уравне-

ние можно методом Ньютона. Пусть $x_0 \in [a, b]$ – начальная точка поиска точки минимума функции x^* . Линеаризуем функцию $f'(x)$ в окрестности начальной точки, приближенно заменив дугу графика этой функции касательной в точке $(x_0, f'(x_0))$

$$f'(x) \approx f'(x_0) + f''(x_0)(x - x_0). \quad (4.24)$$

Выберем в качестве следующего приближения к x^* точку x_1 пересечения касательной с осью абсцисс. Приравнявая нулю правую часть равенства (4.24), имеем

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Мы получили итерационную формулу

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, \quad k = 0, 1, 2, \dots \quad (4.25)$$

Метод одномерного поиска, основанный на этой формуле, называется методом Ньютона-Рафсона или методом касательных. Для квадратичной функции $f(x)$ функция $f'(x)$ линейна. Поэтому равенство (4.24) будет точным, а метод Ньютона-Рафсона будет сходиться за один шаг при любом выборе точки x_0 .

В общем случае сходимость метода Ньютона-Рафсона существенно зависит от выбора начальной точки x_0 . Для надежной работы этого метода необходимо, чтобы функция $f(x)$ в интервале неопределенности $[a, b]$ была выпуклой, то есть, чтобы для любого $x \in [a, b]$ было $f''(x) > 0$. Если к тому же функция $f(x)$ трижды дифференцируема и для всех точек последовательности (4.25) знаки производных $f'(x)$ и $f''(x)$ совпадают, то метод Ньютона-Рафсона обладает квадратичной скоростью сходимости.

Очевидно, итерационная формула (4.25) проще итерационных

формул всех остальных методов интерполяции. Вычисления продолжаютсЯ до тех пор, пока $|x_{k+1} - x_k|$ не станет меньше заданной допустимой погрешности.

Алгоритм метода Ньютона-Рафсона.

Входные параметры: x – начальная точка, ε – допустимая погрешность.

Выходные параметры: x и f_x – лучшая точка и значение функции в ней.

1. Положить $y = x$.
2. Вычислить $f_x = f(x)$, $f'_x = f'(x)$, $f''_x = f''(x)$.
3. Положить $x = x - f'_x / f''_x$.
4. Если $|x - y| > \varepsilon$, то положить $y = x$ и перейти к шагу 2.
5. Остановиться.

4.7. Сравнение методов аппроксимации и интерполяции

Мы рассмотрели различные методы аппроксимации и интерполяции, предназначенные для вычисления точки минимума x^* унимодальной целевой функции одной переменной $f(x)$ как с использованием значений производных функции, так и без использования значений производных. По этому признаку рассмотренные методы можно разделить на три группы.

1. Методы нулевого порядка. Это метод квадратичной интерполяции с тремя точками и метод кубической интерполяции с четырьмя точками, основанные на вычислении значений самой целевой функции и не использующие значений ее производных.

2. Методы первого порядка. Это метод квадратичной интерполяции с двумя точками, метод секущих и метод кубической интерполяции с двумя точками, использующие значения первой производной целевой функции.

3. Метод второго порядка. Это один только метод Ньютона-

Рафсона, использующий значения первой и второй производных целевой функции.

Эти методы одномерного поиска можно сравнить по скорости сходимости. Методы одномерного поиска должны порождать последовательность точек $\{x_k\}$, для которой

$$\lim_{k \rightarrow \infty} x_k = x^*.$$

Тогда по свойству предела

$$\left| x_{k+1} - x^* \right| < \left| x_k - x^* \right|,$$

поэтому выполняется неравенство

$$\frac{\left| x_{k+1} - x^* \right|}{\left| x_k - x^* \right|} < 1,$$

В этом случае говорят о *глобальной сходимости* метода. Если глобальная сходимость метода установлена, то вызывает интерес оценка ее эффективности.

С практической точки зрения эффективность алгоритма зависит от числа итераций, необходимых для получения приближения оптимальной точки x^* с допустимой погрешностью ε . Если допустить, что время вычисления итераций одинаково для всех алгоритмов, то наилучшим среди них будет тот, который требует наименьшего числа итераций.

Поведение последовательности точек $\{x_k\}$ в окрестности оптимальной точки x^* позволяет установить характер *асимптотической сходимости*. Если выполняется неравенство

$$\lim_{k \rightarrow \infty} \frac{\left| x_{k+1} - x^* \right|}{\left| x_k - x^* \right|} = \alpha < 1,$$

то говорят, что имеет место *линейная сходимость* и что α – соответствующий *коэффициент сходимости*. Рассмотренные в разделе 3 методы нулевого порядка для уменьшения интервала неопределенности обладают линейной сходимостью.

Если выполняется неравенство

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = 0,$$

то говорят, что имеет место *сверхлинейная сходимость*. При этом, если существует такое $\gamma > 1$, что

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^\gamma} < +\infty,$$

то говорят, что имеет место *сходимость порядка γ* . В частности, если

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} < +\infty,$$

то говорят, что имеет место *квадратичная сходимость*.

Метод квадратичной интерполяции нулевого порядка с тремя точками имеет сверхлинейную скорость сходимости порядка $\gamma = 1,32\dots$, если $f(x)$ имеет непрерывные производные четвертого порядка и $f'(x^*) = 0$, $f''(x^*) \neq 0$.

Метод секущих первого порядка с двумя точками имеет сверхлинейную скорость сходимости порядка $\gamma = \tau = 1,618\dots$, если $f(x)$ трижды непрерывно дифференцируема и $f'(x^*) = 0$, $f''(x^*) \neq 0$.

Метод кубической интерполяции первого порядка с двумя точками и метод Ньютона-Рафсона имеют квадратичную скорость сходимости, если $f(x)$ дважды непрерывно дифференцируема и $f'(x^*) = 0$,

$$f''(x^*) \neq 0.$$

Следовательно, в малой окрестности оптимальной точки метод кубической интерполяции с двумя точками и метод Ньютона-Рафсона при указанных предположениях относительно целевой функции $f(x)$ сходятся быстрее остальных методов.

Однако в общем случае при решении конкретной задачи оптимизации необходимы вычислительные эксперименты для выбора самого эффективного метода, позволяющего решить задачу оптимизации с допустимой погрешностью при использовании минимальных вычислительных ресурсов.

Лабораторная работа

Тема лабораторной работы.

Разработка подпрограммы для метода аппроксимации и интерполяции.

Цель лабораторной работы.

Разработать подпрограмму метода аппроксимации и интерполяции для минимизации унимодальной целевой функции.

Порядок выполнения лабораторной работы.

1. Получить у преподавателя задание на программирование одного из методов аппроксимации и интерполяции:

- 1) метод квадратичной интерполяции с тремя точками;
- 2) метод кубической интерполяции с четырьмя точками;
- 3) метод квадратичной интерполяции с двумя точками;
- 4) метод секущих;
- 5) метод кубической интерполяции с двумя точками;
- 6) метод Ньютона-Рафсона.

2. На основании алгоритма для заданного метода аппроксимации и интерполяции написать компьютерную подпрограмму в соответствии с требованиями:

- 1) входные параметры подпрограммы: $f(x)$ – целевая функция,

N – предельное число точек, ε – допустимая погрешность, входные параметры алгоритма заданного метода;

2) выходные параметры подпрограммы: X – массив всех точек поиска x_k ; F – массив соответствующих значений функции $f_k = f(x_k)$; x и f_x – лучшая точка поиска и значение в ней функции;

3) на каждой итерации выводить на экран компьютера строку, содержащую номер точки поиска k , шаг перехода в новую точку h_k , значение функции f_k и значение переменной x_k ;

4) итерации продолжаются до тех пор, пока выполняются неравенства:

$$|h_k| > \varepsilon, \quad k \leq N;$$

5) после окончания процесса оптимизации на экране должна отображаться таблица, имеющая соответствующий заголовок и представляющая процесс минимизации функции;

6) под таблицей необходимо отобразить количество вычислений целевой функции, конечный шаг поиска, минимальное значение функции и соответствующее ему значение независимой переменной.

3. Написать общую программу для выполнения вычислительного процесса минимизации целевой функции одной переменной и отображения на экране этого процесса в соответствии с требованиями:

1) за основу общей программы для минимизации функции взять общую программу из предыдущей лабораторной работы и модифицировать эту программу;

2) задать в программе значения x_0 , x^* , h , ε , N ;

3) путем вызова подпрограммы метода Свенна с использованием подпрограммы вычисления функции $f(x)$ и значений x_0 , h , N выполнить итерационный процесс поиска интервала неопределенности с построением таблицы процесса на экране и получением массивов значений переменной X и функции F , границ интервала неопределенности a и b , значений функции на концах интервала f_a и f_b , внут-

ренной точки интервала c и значения в ней функции f_c ;

4) путем вызова разработанной подпрограммы метода аппроксимации и интерполяции реализовать итерационный процесс минимизации функции с построением таблицы процесса на экране и получением массивов значений переменной X и функции F , лучшей точки поиска x и значения в ней функции f_x ;

5) с помощью подпрограммы графического отображения итерационного процесса по массивам X и F отобразить график процесса минимизации целевой функции одной переменной на экране компьютера.

4. Запустить на счет общую программу минимизации функции одной переменной для функций предыдущих лабораторных работ из нескольких начальных точек

$$x_{01} < x^*, \quad x_{02} = x^*, \quad x_{03} > x^*$$

при нескольких значениях шага

$$h_{01} < |x_0 - x^*|, \quad h_{02} = |x_0 - x^*|, \quad h_{03} > |x_0 - x^*|.$$

5. Выполнить анализ полученных результатов:

1) на основании табличной и графической информации о работе методов одномерного поиска определить конечные точки метода Свенна и показать найденный этим методом интервал неопределенности;

2) на нескольких начальных итерациях по табличным данным и графикам сопоставить начало работы метода аппроксимации и интерполяции с его рабочими формулами;

3) оценить эффективность метода аппроксимации и интерполяции и сопоставить ее эффективностью метода уменьшения интервала неопределенности из предыдущей лабораторной работы.

6. Ответить на вопросы преподавателя по теме лабораторной работы.

7. Оформить и сдать отчет о проведенной лабораторной работе.

Содержание отчета.

1) титульный лист установленного образца с указанием организации, названия учебной дисциплины, темы работы, номера варианта, исполнителя и принимающего, города, года;

2) постановка задачи минимизации целевой функции, включающая задание функции, начальную точку и точку минимума;

3) теоретическое описание метода аппроксимации и интерполяции;

4) алгоритм метода;

5) распечатки используемых компьютерных программ;

6) табличное представление процесса минимизации и его итоги – количество вычислений значений целевой функции, минимальное значение функции и соответствующее ему значение точки минимума, достигнутую точность;

7) двумерный график траектории поиска со всеми точками;

8) анализ полученных результатов;

9) выводы о проделанной лабораторной работе.

Контрольные вопросы

1. Какое назначение метода квадратичной интерполяции с тремя точками?

2. Опишите метод квадратичной интерполяции с тремя точками.

3. Как начинается поиск минимума в методе квадратичной интерполяции с тремя точками?

4. Какого порядка метод квадратичной интерполяции с тремя точками?

5. Какие преимущества и недостатки метода квадратичной интерполяции с тремя точками?

6. Какое назначение метода кубической интерполяции с четырьмя точками?

7. Опишите метод кубической интерполяции с четырьмя точками.

8. Как начинается поиск минимума в методе кубической интерпо-

ляции с четырьмя точками?

9. Какого порядка метод кубической интерполяции с четырьмя точками?

10. Какие преимущества и недостатки метода кубической интерполяции с четырьмя точками?

11. Какое назначение метода квадратичной интерполяции с двумя точками?

12. Опишите метод квадратичной интерполяции с двумя точками.

13. Какие преимущества и недостатки метода квадратичной интерполяции с двумя точками?

14. Какое назначение метода секущих?

15. Опишите метод секущих.

16. Какого порядка метод секущих?

17. Какие преимущества и недостатки метода секущих?

18. Какое назначение метода кубической интерполяции с двумя точками?

19. Опишите метод кубической интерполяции с двумя точками.

20. Какие преимущества и недостатки метода кубической интерполяции с двумя точками?

21. Какое назначение метода Ньютона-Рафсона?

22. Выведите итерационную формулу метода Ньютона-Рафсона.

23. Опишите метод Ньютона-Рафсона.

24. Какого порядка метод Ньютона-Рафсона?

25. Какие преимущества и недостатки метода Ньютона-Рафсона?

26. На какие группы можно разделить методы аппроксимации и интерполяции?

27. Дайте определение глобальной сходимости метода одномерного поиска.

28. Дайте определение асимптотической сходимости метода одномерного поиска.

29. Как определяется линейная сходимость метода одномерного поиска?

30. Приведите примеры методов одномерного поиска, обладаю-

щих свойством линейной сходимости?

31. Как определяется сверхлинейная сходимость метода одномерного поиска?

32. Приведите примеры методов одномерного поиска, обладающих свойством сверхлинейной сходимости?

33. Как определяется квадратичная сходимость метода одномерного поиска?

34. Приведите примеры методов одномерного поиска, обладающих свойством квадратичной сходимости?

35. Чем определяется выбор метода одномерного поиска для решения конкретной задачи оптимизации?

СПИСОК ЛИТЕРАТУРЫ

1. 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. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2003. – 440 с.
2. Банди Б. Методы оптимизации. Вводный курс / Б. Банди ; пер. с англ. О.В. Шихеевой. – М. : Радио и связь, 1988. – 128 с.
3. Бейко И.В. Методы и алгоритмы решения задач оптимизации / И.В. Бейко, Б.Н. Бублик, П.Н. Зинько. – К. : Вища школа, 1983. – 512 с.
4. Воробьев Н.Н. Числа Фибоначчи / Н.Н. Воробьев. – 4-е изд., доп. – М. : Наука. Гл. ред. физ.-мат. лит., 1978. – 142 с.
5. Гилл Ф. Практическая оптимизация / Ф. Гилл, У. Мюррей, М. Райт ; пер. с англ. В.Ю. Лебедева. – М. : Мир, 1985. – 512 с.
6. Дэннис Д. Численные методы безусловной оптимизации и решения нелинейных уравнений / Д. Дэннис, Р. Шнабель ; пер. с англ. О.П. Бурдакова – М. : Мир, 1988. – 440 с.
7. Измаилов А.Ф. Численные методы оптимизации : учеб. пособ. / А.Ф. Измаилов, М.В. Солодов. – М. : ФИЗМАТЛИТ, 2005. – 304 с.
8. Карманов В.Г. Математическое программирование / В.Г. Карманов. – 2-е изд., перераб. и доп. – М. : Наука. Гл. ред. физ.-мат. лит., 1980. – 256 с.
9. Короп В.Ф. Принцип равномерного вклада и регулирование шагов покоординатного случайного поиска / В.Ф. Короп // Численные методы нелинейного программирования. – К. : Наук. думка. – 1976. – С. 33–41.
10. Мину М. Математическое программирование. Теория и алгоритмы / М. Мину ; пер. с фр. А.И. Штерна. – М. : Наука. Гл. ред. физ.-мат. лит., 1990. – 488 с.
11. Полак Д. Численные методы оптимизации. Единый подход / Д. Полак ; пер. с англ. Ф.И. Ерешко. – М. : Мир, 1974. – 376 с.
12. Пантелеев А. В. Методы оптимизации в примерах и задачах :

учеб. пособ. / А. В. Пантелеев, Т.А. Летова. – М. : Высш. шк., 2002. – 544 с.

13. Раскин Л.Г. Математическое программирование : учебно-метод. пособ. / Л.Г. Раскин. – Харьков : ПП «ГЕММА», 1986. – 68 с.

14. Реклейтис Г. Оптимизация в технике : в 2 кн. / Г. Реклейтис, А. Рейвиндран, К. Рэгсдел. Кн. 1 ; пер. с англ. В.Я. Алтаева, В.И. Моторина. – М. : Мир, 1986. – 349 с.

15. Струченков В.И. Методы оптимизации. Основы теории, задачи, обучающие компьютерные программы : учеб. пособ. / В.И. Струченков. – М. : Экзамен, 2005. – 256 с.

16. Сухарев А.Г. Курс методов оптимизации : учеб. пособ. / А.Г. Сухарев, А.В. Тимохов, В.В. Федоров. – 2-е изд. – М. : ФИЗМАТЛИТ, 2005. – 368 с.

17. Форсайт Д. Машинные методы математических вычислений / Д. Форсайт, М. Малькольм, К. Моулер ; пер. с англ. Х.Д. Икрамова. – М. : Мир, 1980. – 280 с.

18. Химмельблау Д. Прикладное нелинейное программирование / Д. Химмельблау ; пер. с англ. И.М. Быховской, Б.Т. Вавилова. – М. : Мир, 1975. – 536 с.

19. Fletcher R. Practical methods of optimization / R. Fletcher. – 2-d edition. – UK, Chichester : John Wiley & Sons Ltd, 2000. – 436 p.

20. Luenberger D. G. Linear and nonlinear programming / D.G. Luenberger, Y. Yinu. – 3-d edition. – USA, New York : Springer, 2008. – 546 p.

21. Bazaraa M.S. Nonlinear programming : theory and algorithms / M.S. Bazaraa, H.D. Sherali, C.M. Shetty. – 3-d edition. – USA, Hoboken, New Jersey : John Wiley & Sons Inc., 2006. – 853 p.

22. Sun W. Optimization theory and methods. Nonlinear programming / W. Sun, Y. X. Yuan. – USA, New York: Springer, 2006. – 687 p.

Навчальне видання

СЕВЕРИН Валерій Петрович

МЕТОДИ ОДНОВИМІРНОГО ПОШУКУ

Навчально-методичний посібник з курсу
«Методи оптимізації» для студентів
напрямків 6.040302 «Інформатика»,
6.040303 «Системний аналіз»

Російською мовою

Роботу до видання рекомендував М. І. Безменов

В авторській редакції

План 2012 р., п. 2/19-12

Підп. до друку 14.02.1012 р. Формат 60×84 1/16. Папір друк. №2.
Друк – ризографія. Гарнітура Times New Roman. Ум. друк. арк. 6,5.
Наклад 100 прим. Зам. № 60. Ціна договірна.

Видавничий центр НТУ «ХПІ»

Свідоцтво про державну реєстрацію ДК № 3657 від 24.12.2009 р.

61002, Харків, вул. Фрунзе, 21

Друкарня НТУ «ХПІ», 61002, Харків, вул. Фрунзе, 21