Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Г.И. Беликова, Е.А. Бровкина, Б.Г. Вагер, Л.В. Витковская,

Ю.Л. Матвеев

Численные методы

Учебное пособие для студентов 2-го курса

Санкт-Петербург

2019

УДК 51
ББК 22.193я73

Г.И. Беликова, Е.А. Бровкина, Б.Г. Вагер, Л.В. Витковская, Ю.Л. Матвеев
Численные методы. Учебное пособие. – СПб.: РГГМУ, 2019. –174 с.

Книга содержит курс лекций, которые читаются студентам 2-го курса метеорологического факультета РГГМУ. Основное внимание в учебнике уделено методам решения систем линейных алгебраических уравнений, дифференциальных уравнений с начальными и краевыми условиями, уравнений математической физики. Достаточно подробно изложена теория приближения функций. Большое внимание уделено итерационным процессам. Представлены способы нахождения экстремальных значений функций нескольких переменных.

ISBN 978-5-86813-485-2

Оглавление

Предисловие

Математика, являясь самой древней из всех наук, вместе с тем остаётся вечно молодой.
Академик М.В. Келдыш

«Численные методы» – это совокупность методов и алгоритмов построения приближенных решений задач линейной алгебры, математического анализа, дифференциальных уравнений и других разделов математики, а также методов оценки точности полученных результатов.

Новый этап в развитии «Численных методов» наступил с повсеместным использованием компьютеров. Программное обеспечение любых математических стандартных пакетов разработано на основе этих методов. Для эффективного использования таких стандартных пакетов необходимо освоить соответствующую теорию.

Основой «Численных методов» является высшая алгебра, математический анализ, вариационное исчисление, дифференциальные уравнения, уравнения математической физики и функциональный анализ. Поэтому «Численные методы» обычно начинают читать с четвёртого семестра.

Наше учебное пособие является введением в «Численные методы». В него включены только те методы, которые часто используются для решения задач метеорологии, гидрологии, океанологии и будут понятны студентам второго курса. Наиболее важные и простые для понимания методы написаны более подробно. С другими, такими же важными, но математически не простыми методами, знакомим кратко.

В некоторых главах учебника представлены различные методы решения близких по постановке задач, потому что метод, эффективный для задач с малым объёмом данных, может оказаться неэффективным для аналогичной задачи с большим объёмом данных.

Математика формировалась как наука в течение тысячелетий и постепенно стала неотъемлемой частью общечеловеческой культуры. Она тесно связана с историей и языкознанием. Поэтому книга содержит этимологический и толковый словарь математических терминов и понятий. Представлен справочник с краткими сведениями о математиках, имена которых встречаются в этом учебнике.

«Численные методы», как и другие математические науки, развивают строгость мышления и воображение, а это полезно любому студенту.

Надеемся, что это учебное пособие будет с интересом и пользой изучено студентами нашего университета.

Глава 1

БАЗОВЫЕ ПОНЯТИЯ

Значение математической строгости не следует преувеличивать и доводить до абсурда; здравый смысл в математике не менее уместен.

В. Успенский. «Апология математики».

1.1. Погрешности приближённых вычислений

В численных методах уделяется большое внимание оценке погрешности результата решённой задачи и способам уменьшения погрешностей, возникающих в процессе решения.

1.1.1. Классификация погрешностей.

Погрешности классифицируют по источникам их возникновения.

1. Погрешность в исходных данных.
2. Погрешность выбранной математической модели.
3. Погрешность выбранного численного метода.
4. Погрешность компьютерных вычислений (погрешности округления).

Погрешности 1-го класса неустранимы. Их источники – физические измерения. Множество реальных данных содержат погрешность, которую часто называют ***шумом***. Этот шум отрицательно влияет на любой численный метод.

Погрешности 2-го класса неизбежны потому, что математические модели в метеорологии, океанологии, гидрологии, экологии и других науках – это всегда приближённые модели. Для уменьшения такой погрешности следует попытаться повысить точность модели.

Погрешности 3-го класса связаны с использованием тех или иных формул и алгоритмов, выведенных в рамках «Численных методов». В настоящее время хорошо разработаны методы, в которых для получения точного результата теоретически необходимо сделать бесконечно много действий, а в реальных вычислениях проводится только конечное число операций. Такая ситуация возникает при использовании сходящихся функциональных и числовых рядов, а также в итерационных процессах.

Минимизация погрешностей наиболее приоритетна в процессе разработки любого численного метода. Если во время решения погрешность не увеличивается, тогда говорят, что метод (алгоритм) имеет ***вычислительную устойчивость***.

Компьютерное округление и деление на число, близкое по модулю к нулю, увеличивают погрешности 4-го класса. Накопление ошибок округления может значительно повлиять на результат решения задачи. Поэтому во всех алгоритмах стандартных программ есть специальные операции для максимального уменьшения погрешностей такой природы.

1.1.2. Оценка погрешности.

Для оценки погрешности используются два понятия: ***абсолютная погрешность*** и ***относительная погрешность***.

Пусть $\tilde{x}$ – приближение к некоторому точному значению x.

Абсолютная погрешность определяется следующей формулой

$$\varepsilon_x = |x - \tilde{x}|.$$

Относительная погрешность определяется по формуле

$$r_x = \left|\frac{x - \tilde{x}}{x}\right|, \quad x \neq 0.$$

Если $|x| < 1$, лучше использовать относительную погрешность.

Определение 1. Величина $\widetilde{x}$ называется приближением к числу x с k значащими цифрами, если k – наибольшее натуральное число, для которого справедливо неравенство

$$\left|\frac{x - \tilde{x}}{x}\right| < \frac{10^{-k}}{2}.$$

Например, для относительной погрешности $\varepsilon_{\text{о}}$ компьютерного округления числа справедливо равенство

$\varepsilon_{\text{о}} = 0{,}5 \cdot 10^{-t}$, где t– количество битов, занимаемое числом. Это число очень мало, но оно может существенно вырасти после проведения на компьютере миллиона арифметических действий.

1.1.3. Распространение погрешностей

Разберёмся, как изменяются погрешности в процессе вычислений.

АБСОЛЮТНАЯ ПОГРЕШНОСТЬ СУММЫ

Пусть x– сумма точных значений чисел $\{x_i\}_{i=1}^n$, $\tilde{x}$ – сумма их приближений: $\tilde{x} = \{\widetilde{x_i}\}_{i=1}^n$, а $\{\varepsilon_i\}_{i=1}^n$– соответствующие абсолютные погрешности. Сделаем следующие оценки

$$\varepsilon_x = |x - \widetilde{x}\ | = \left|\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \widetilde{x_i}\right| \leq \sum_{i=1}^{n} |x_i - \widetilde{x_i}| = \sum_{i=1}^{n} \varepsilon_i \implies$$

$$\varepsilon_x \leq \sum_{i=1}^{n} \varepsilon_i. \qquad (1.1)$$

Из неравенства (1.1) следует, что абсолютная погрешность суммы не превосходит суммы абсолютных погрешностей слагаемых.

АБСОЛЮТНАЯ ПОГРЕШНОСТЬ ПРОИЗВЕДЕНИЯ

Предположим, что числа x и y – точные значения, $\tilde{x}$ и $\tilde{y}$ их приближения, а ε_x, ε_y – абсолютные погрешности, тогда

$$xy = (\tilde{x} + \varepsilon_x)(\tilde{y} + \varepsilon_y) = \tilde{x}\tilde{y} + \tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x + \varepsilon_x\varepsilon_y \Longrightarrow$$

$$\varepsilon_{xy} = |xy - \tilde{x}\tilde{y}| = |\tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x + \varepsilon_x\varepsilon_y|. \qquad (1.2)$$

Обратимся к правой части равенства (1.2). Если для приближений будут верными неравенства $|\tilde{x}| > 1$ и $|\tilde{y}| > 1$, то возможно увеличение погрешности за счёт суммы $\tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x$.

ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТЬ ПРОИЗВЕДЕНИЯ И ОТНОШЕНИЯ

Составим относительную погрешность произведения и проведём элементарные преобразования:

$$r_{xy} = \left|\frac{xy - \tilde{x}\tilde{y}}{xy}\right| = \left|\frac{\tilde{x}\varepsilon_y + \tilde{y}\varepsilon_x + \varepsilon_x\varepsilon_y}{xy}\right| = \left|\frac{\tilde{x}\varepsilon_y}{xy} + \frac{\tilde{y}\varepsilon_x}{xy} + \frac{\varepsilon_x\varepsilon_y}{xy}\right|.$$

Предположим, что $\tilde{x}$ и $\tilde{y}$ – хорошие приближения для чисел x, y, тогда

$$\frac{\tilde{x}}{x} \approx 1, \frac{\tilde{y}}{y} \approx 1, \left|\frac{\varepsilon_x}{x} \cdot \frac{\varepsilon_y}{y}\right| \approx 0 \Rightarrow r_{xy} \approx \left|\frac{\tilde{x}\varepsilon_y}{xy} + \frac{\tilde{y}\varepsilon_x}{xy}\right| \leq \left|\frac{\varepsilon_x}{x}\right| + \left|\frac{\varepsilon_y}{y}\right|. \quad (1.3)$$

Из неравенства (1.3) следует, что относительная погрешность произведения xy не превосходит суммы относительных погрешностей сомножителей:

$$r_x = \left|\frac{\varepsilon_x}{x}\right|, r_y = \left|\frac{\varepsilon_y}{y}\right| \Longrightarrow r_{xy} \leq r_x + r_y.$$

Можно доказать, что аналогичное утверждение справедливо для отношения двух величин:

$$r_{\frac{x}{y}} \leq r_x + r_y.$$

Замечание 1. В процессе вычислений обычно не происходит быстрого роста вычислительных погрешностей, так как одни приближения чисел округляются с избытком, а другие с недостатком. Погрешности частично компенсируют друг друга. При любых расчётах используют правило: надо удерживать столько значащих цифр, чтобы погрешность округления была существенно меньше всех остальных погрешностей.

1.2. Итерационные методы

Итерационные методы – это фундаментальное понятие «Численных методов». Слово *итерация* произошло от латинского iteratio – повторение. Анг-

лийское слово *iteration – повторение, итерация.* В математике так называют процесс повторного применения совокупности математических операций для последовательного приближения к точному решению задачи. Поэтому итерационные методы ещё называют ***методами последовательных приближений***.

ИТЕРАЦИОННЫЙ ПРОЦЕСС

Поставлена некоторая задача *M*. Надо найти решение x этой задачи. Пусть известно начальное приближение x_0 к этому решению. Допустим, что разработан алгоритм *A*, при помощи которого строятся следующие приближения:

$$A(x_0) = x_1,\ A(x_1) = x_2, \cdots, A(x_{k-1}) = x_k. \quad (1.4)$$

Предположим, что удалось доказать сходимость последовательности таких приближений $\{x_k\}_{k=1}^{\infty}$. Это означает, что существует предел

$$\lim_{k\to\infty} x_k = x^*, \quad (1.5)$$

который называется ***неподвижной точкой***. Неподвижная точка x^* является точным решением задачи *M*.

Переход от приближения x_{k-1} к приближению x_k с помощью равенства $A(x_{k-1}) = x_k$ называется ***шагом итерации,*** а последовательность (1.4) приближений называется ***итерационной последовательностью***.

Точное решение может быть получено только теоретически при бесконечном числе шагов итерации. Реально мы можем лишь оценить степень близости точного и приближенного решений задачи. Поэтому перед началом итерационного процесса всегда задаётся допустимое значение погрешности расчёта ε.

Например, в итерационном процессе выполняется столько шагов итерации, сколько необходимо для выполнения неравенства

$$\left|\frac{x_k - x_{k-1}}{x_k}\right| \le \varepsilon.$$

СХОДИМОСТЬ ИТЕРАЦИОННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Определение 2. Сходимость последовательности $\{x_k\}_{k=1}^{\infty}$ к некоторой неподвижной точке x^* называется ***линейной***, если существует такое число $C \in (0;1)$ и такое число $N \in \mathbb{N}$, что для любого $k > N$ будет выполняться неравенство

$$|x^* - x_{k+1}| < C|x^* - x_k|. \quad (1.6)$$

Линейную сходимость часто называют ***сходимостью со скоростью геометрической прогрессии***.

Определение 3. Последовательность $\{x_k\}_{k=1}^{\infty}$ ***сходится*** к неподвижной точке x^* ***с порядком p***, если существуют такие числа $C > 0$, $p > 1, N \in \mathbb{N}$, при которых будет верно неравенство

$$|x^* - x_{k+1}| < C|x^* - x_k|^p \qquad \text{при } \forall k > N. \qquad (1.7)$$

Если порядок $p > 1$, такая сходимость считается ***быстрой*** и называется ***сверхлинейной***.

Замечание 2. Пусть при исследовании последовательности $\{x_k\}_{k=1}^{\infty}$ на сходимость неравенства (1.6) или (1.7) выполняются при условии, что каждому значению k, соответствует своё число C_k, и выполняется предельное равенство

$$\lim_{k\to\infty} C_k = C \in (0;1).$$

В таком случае говорят, что есть ***асимптотическая сходимость***.

Определение 4. Говорят, что последовательности $\{x_i\}_{i=1}^{\infty}$, $\{y_i\}_{i=1}^{\infty}$ имеют одинаковый порядок сходимости, если существуют такие положительные константы $C > 0$ и $n > 0$, для которых при любом $i > n$ справедливо одно из двух неравенств

$$|x_i| \le C \cdot |y_i| \text{ или } |y_i| \le C \cdot |x_i|.$$

Замечание 3. Чем выше порядок сходимости последовательности, тем выше скорость сходимости. Следовательно, потребуется меньше шагов итерации для получения необходимой точности решения. Уменьшение числа шагов итерации приводит к сокращению числа арифметических действий, а значит к уменьшению вычислительной погрешности. Порядок и скорость сходимости являются очень важными характеристиками любого итерационного процесса.

Определение 5. Если итерационный процесс сходится при любом начальном условии, его называют ***глобально сходящимся процессом***.

Определение 6. Если итерационный процесс сходится только для того начального условия, которое находится в малой окрестности точного решения, процесс называют ***локально сходящимся***.

Обычно локальные процессы сходятся быстрее глобальных. Некоторые смешанные итерационные алгоритмы начинают работать с глобальных сходящихся процессов, а ближе к точному решению переходят к локальным.

1.3. Устойчивость и корректность

Для любого численного метода и соответствующего алгоритма вводится следующее требование. Маленькая погрешность $|\delta b|$ в начальных условиях должна приводить к малой погрешности $|\delta x|$ результата решения.

Численный метод и соответствующий алгоритм решения, обладающие таким свойством, называются ***устойчивыми относительно входных данных.*** В противном случае и метод, и алгоритм – ***неустойчивые***. Для устойчивого численного метода справедливо неравенство

$$|\delta x| \le C|\delta b|, \text{ где } C - const.$$

Если постоянная C очень велика, устойчивость будет только формальной. Тогда в процессе решения задачи могут возникнуть неустранимые большие погрешности. В этом случае говорят о ***слабой устойчивости.***

Определение 7. Задача называется ***корректно поставленной***, если для любых входных данных заданного множества решение задачи ***существует***, ***единственно и устойчиво относительно входных данных***.

Замечание 4. Существуют методы решения многих некорректно поставленных задач. Эти методы основаны на решении близкой к исходной, но корректно поставленной задаче.

Глава 2

РЕШЕНИЕ УРАВНЕНИЙ

…наука вся состоит из уравнений, начиная с азов арифметики и до теории относительности.
Ричард Браун. «Математика за 30 секунд»

Задача о нахождении корней уравнения $f(x) = 0$ имеет большую и интересную историю. С давних времён творцы математики уделяли особое внимание вычислению корней алгебраических уравнений с действительными коэффициентами. Пытались выразить корни через точные формулы. Искали связь числа корней с порядком уравнения.

В настоящее время для решения уравнений созданы различные эффективные итерационные методы, реализованные в программном обеспечении современных компьютеров.

2.1. Корни алгебраических уравнений

В высшей алгебре всесторонне изучены уравнения такого класса. Даны ответы на вопросы о количестве корней уравнения и о возможности точного нахождения этих корней.

Уравнение называется алгебраическим уравнением n-й степени, если оно представимо в виде

$$x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_{n-1} x + p_n = 0,$$

где коэффициенты $\{p_i\}_{i=1}^{n}$ – заданные действительные числа.

Решить данное уравнение – это значит найти такие значения величины x, при которых уравнение превращается в тождество. Найденные значения x называются корнями уравнения.

В XVI произошло величайшее открытие в алгебре. Итальянские математики: Тарталья, Кардано и Феррари вывели формулы для нахождения корней уравнений 3-й и 4-й степеней.

Почти ещё три века лучшие математические умы пытались вывести формулы для точного нахождения корней уравнений 5-й степени. Только в начале XIX в. норвежский математик Нильс Абель доказал, что нельзя точно решить в общем случае алгебраическое уравнение выше 4-й степени с помощью четырёх арифметических действий и извлечения корня (в радикалах).

ОСНОВНАЯ ТЕОРЕМА АЛГЕБРЫ

Теорема Гаусса. Алгебраическое уравнение n-й степени имеет хотя бы один корень, в общем случае он может быть комплексным.

Следствие из теоремы. Алгебраическое уравнение n-й степени имеет ровно n корней, среди которых могут быть и кратные, и комплексные корни.

Из выше изложенного следует, что задача о нахождении корней алгебраического уравнения выше 4-й степени может быть решена только приближённо. Численные методы являются основным инструментом вычисления этих корней.

2.2. Отделение корней

Корень r уравнения $f(x) = 0$ считается отделённым в области $(a; b)$, если уравнение имеет только один корень $r \in (a; b)$. Корни уравнения считаются отделёнными, если для каждого корня r_i найдена своя область единственности $(a_i; b_i)$, для которой $r_i \in (a_i; b_i)$.

Задача нахождения корней начинается с их отделения, но универсальных способов для этого не существует. Поэтому следует более детально изучить функцию $f(x)$, стоящую в левой части уравнения, построить её график и вспомнить следующие две теоремы из математического анализа.

Теорема 1. Если функция $f(x)$ определена и непрерывна на $[a; b]$ и верно равенство $f(a)f(b) < 0$, тогда в интервале $(a; b)$ есть хотя бы один корень уравнения $f(x) = 0$.

Теорема 2. Пусть функция $f(x)$ определена, непрерывна, монотонна на $[a; b]$ и верно равенство $f(a)f(b) < 0$, тогда в интервале $(a; b)$ есть только один корень уравнения $f(x) = 0$.

Будем предполагать, что все корни уравнения удалось отделить. Поэтому перейдём к рассмотрению различных, достаточно простых, итерационных методов нахождения отделённого корня.

Во многих итерационных методах решения уравнения обычно задают две оценки точности расчёта. Первая – необходимая точность ε найденного корня r. Вторая оценка связана непосредственно с уравнением $f(x) = 0$. Если в уравнение подставить приближённое значение корня $\tilde{r}$, то будет равенство $f(\tilde{r}) = \delta$. Число δ называется ***невязкой.*** Она является второй характеристикой точности найденного корня.

Разберём четыре простейших классических итерационных метода.

2.3. Метод деления пополам

У этого метода есть другие названия: метод ***половинного деления***, ***бисекции***, ***дихотомии***. Его придумал в начале XIX в. чешский математик Бернард Больцано.

Ставится задача о нахождении корня непрерывной функции $y = f(x)$ на отрезке $[a_0; b_0]$, если известно, что $f(a_0)f(b_0) < 0$ и внутри интервала есть только один корень $x = r$. Эта задача эквивалентна решению уравнения

$$f(x) = 0.$$

Метод деления пополам – итерационный метод. Поэтому задаётся нужная точность ε приближения $\tilde{r}$ к корню r и невязка δ для оценки близости к нулю соответствующего значения $f(\tilde{r})$.

АЛГОРИТМ ИТЕРАЦИОННОГО ПРОЦЕССА

Для упрощения алгоритма на каждом шаге итерации будет происходить переименование границ того отрезка, в котором лежит точный корень.

1. Берётся середина отрезка $[a_0; b_0]$: $c_0 = 0{,}5(a_0 + b_0)$. В точке c_0 вычисляется $f(c_0)$. Сравниваются знаки трёх значений функции $f(x)$. Возможны три случая:

$$\begin{cases} f(a_0)f(c_0) < 0 \Longrightarrow \text{ корень } r \in (a_0; c_0) \Longrightarrow [a_0; c_0] = [a_1; b_1], \\ f(c_0)f(b_0) < 0 \Longrightarrow \text{ корень } r \in (c_0; b_0) \Longrightarrow [c_0; b_0] = [a_1; b_1], \\ f(c_0) = 0 \Longrightarrow \text{ корень } r = c_0. \text{ Корень найден.} \end{cases}$$

2. Если $f(c_0) \neq 0$, в середине нового отрезка $c_1 = 0{,}5(a_1 + b_1)$ вычисляется $f(c_1)$. Рассматриваются три случая:

$$\begin{cases} f(a_1)f(c_1) < 0 \Longrightarrow \text{ корень } r \in (a_1; c_1) \Longrightarrow [a_1; c_1] = [a_2; b_2], \\ f(c_1)f(b_1) < 0 \Longrightarrow \text{ корень } r \in (c_1; b_1) \Longrightarrow [c_1; b_1] = [a_2; b_2], \\ f(c_1) = 0 \Longrightarrow \text{ корень } r = c_1. \text{ Корень найден.} \end{cases}$$

.....

K. Вычисляется средняя точка очередного отрезка $c_{k-1} = 0{,}5(a_{k-1} + b_{k-1})$ и значение $f(c_{k-1})$. Рассматриваются три случая:

$$\begin{cases} f(a_{k-1})f(c_{k-1}) < 0 \Longrightarrow \text{ корень } r \in (a_{k-1}; c_{k-1}) \Longrightarrow [a_{k-1}; c_{k-1}] = [a_k; b_k], \\ f(c_{k-1})f(b_{k-1}) < 0 \Longrightarrow \text{ корень } r \in (c_{k-1}; b_{k-1}) \Longrightarrow [c_{k-1}; b_{k-1}] = [a_k; b_k], \\ f(c_{k-1}) = 0 \Longrightarrow \text{ корень } r = c_{k-1}. \text{ Корень найден.} \end{cases}$$

Через K шагов итерации приходим к последовательности неравенств

$$a_0 \leq a_1 \leq \cdots \leq a_{k-1} \leq a_k \leq r \leq b_k \leq b_{k-1} \leq \cdots \leq b_1 \leq b_0.$$

Итерационный процесс закончится, когда длина отрезка $[a_k; b_k]$ станет меньше заданной погрешности ε:

$$|b_k - a_k| < \varepsilon.$$

СХОДИМОСТЬ МЕТОДА

Доказано, что для каждого шага итерации справедливо неравенство

$$|r - c_{k-1}| \leq \frac{b_0 - a_0}{2^k}. \qquad (2.1)$$

Из этого неравенства следует очевидная цепочка предельных равенств:

$$\lim_{k\to\infty} \frac{b_0 - a_0}{2^k} = 0 \Longrightarrow \lim_{k\to\infty} |r - c_{k-1}| = 0 \Longrightarrow \lim_{k\to\infty} |c_k| = r.$$

Последнее предельное равенство говорит о сходимости метода бисекции. В качестве корня $\tilde{r}$ берётся середина отрезка $[a_k; b_k]$:

$$\tilde{r} = 0{,}5(a_k + b_k).$$

Метод устойчив, глобально сходится, но сходится линейно. Такая сходимость считается медленной.

ОЦЕНКА ТОЧНОСТИ

Достоинством метода является априорное знание оценки погрешности (2.1) найденного корня c_k. Но следует учесть поведение самой функции в окрестности этого корня. Если в его окрестности значение модуля первой производной большое, тогда сама функция в точке c_k может существенно отличаться от нуля. В этом случае добавляется условие

$$|f(c_k)| < \delta.$$

Процесс заканчивается, когда выполняются сразу два неравенства

$$\begin{cases} |b_k - a_k| < \varepsilon, \\ |f(c_k)| < \delta. \end{cases}$$

2.4. Метод хорд

Будем искать приближённое значение корня уравнения $f(x) = 0$, если известно, что корень находится внутри отрезка $[a; b]$, $f(a)f(b) < 0$, функция $f(x)$ монотонна в этой области, вторая производная имеет там постоянный знак. Следовательно, в области поиска корня график функции выпуклый вверх или выпуклый вниз.

Для описания алгоритма метода удобно переименовать левую границу отрезка $[a; b]$. Пусть $a = a_0$.

АЛГОРИТМ ИТЕРАЦИОННОГО ПРОЦЕССА

1. Процесс начинается с построения уравнения прямой, проходящей через две заданные точки плоскости $A\big(a_0; f(a_0)\big)$ и $B\big(b; f(b)\big)$:

$$\frac{x - a_0}{b - a_0} = \frac{y - f(a_0)}{f(b) - f(a_0)}. \qquad (2.2)$$

Запишем это уравнение в виде

$$y = f(a_0) + \frac{f(b) - f(a_0)}{b - a_0}\,(x - a_0) \qquad (2.3)$$

и назовём его уравнением хорды AB (рис. 2.1). Выразим из уравнения (2.3) величину x. Пусть x равно a_1 – точке пересечения хорды AB с осью абсцисс:

$$x = a_1 = a_0 - \frac{b - a_0}{f(b) - f(a_0)} f(a_0). \qquad (2.4)$$

2. Вновь строится хорда, соединяющая точки $A_1\big(a_1; f(a_1)\big)$ и $B\big(b; f(b)\big)$ и её уравнение

$$\frac{x - a_1}{b - a_1} = \frac{y - f(a_1)}{f(b) - f(a_1)}.$$

Находится точка пересечения этой хорды с осью OX:

$$x = a_2 = a_1 - \frac{b - a_1}{f(b) - f(a_1)} f(a_1).$$

……...

В результате K повторений ***рекуррентной формулы***

$$a_k = a_{k-1} - \frac{b - a_{k-1}}{f(b) - f(a_{k-1})} f(a_{k-1}) \qquad (2.5)$$

построена последовательность $\{a_k\}_{k=1}^{K}$ точек пересечения семейства хорд $A_k B$ с осью абсцисс (рис. 2.1).

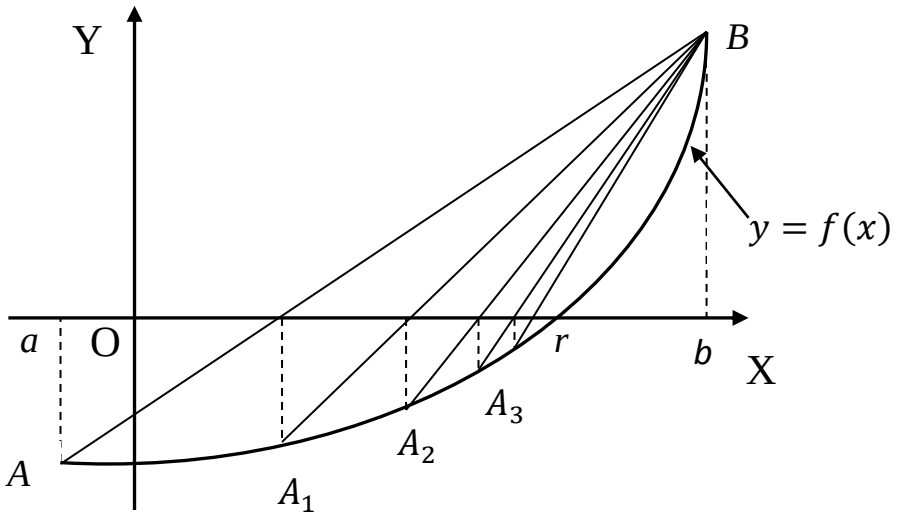

Рис. 2.1

СХОДИМОСТЬ И ОЦЕНКА ТОЧНОСТИ

Доказано, что метод хорд сходится, т.е. справедливо предельное равенство

$$\lim_{k \to \infty} |a_k| = r.$$

На каждом шаге итерации по заданной точности и невязке оценивается близость к точному значению корня уравнения

$$\begin{cases} \left|\dfrac{c_k - c_{k-1}}{c_k}\right| < \varepsilon, \\ |f(c_k)| < \delta. \end{cases} \qquad (2.6)$$

Если оба неравенства выполняются, процесс закончен. В качестве корня берут приближение последней итерации: $\tilde{r} = c_k$. Если хотя бы одно из неравенств не выполняется, итерационный процесс продолжается.

Метод устойчив, глобально сходится, но только линейно. Сходимость становится быстрой в случае близости функции $f(x)$ к линейной функции.

2.5. Метод секущих

С помощью этого метода строится приближённое значение корня уравнения $f(x) = 0$, если известно, что корень находится внутри отрезка $[a; b]$, $f(a)f(b) < 0$, функция $f(x)$ монотонна, дифференцируема в этой области, вторая производная имеет постоянный знак на $[a; b]$.

АЛГОРИТМ ИТЕРАЦИОННОГО ПРОЦЕССА

1. Для начала итерационного процесса необходимо задать два начальных приближения a_0 и a_1. Через две точки плоскости $\mathrm{A}_0\big(a_0; f(a_0)\big)$; $\mathrm{A}_1\big(a_1; f(a_1)\big)$ проводится секущая, которая пересекает ось OX в точке $\mathrm{A}_2(a_2; 0)$. Секущая $\mathrm{A}_0\mathrm{A}_1$ лежит на прямой, поэтому для неё справедливо равенство

$$\frac{a_2 - a_1}{0 - f(a_1)} = \frac{a_1 - a_0}{f(a_1) - f(a_0)}. \qquad (2.7)$$

Из равенства (2.7) выражаем величину a_2

$$a_2 = a_1 - f(a_1)\frac{a_1 - a_0}{f(a_1) - f(a_0)}.$$

2. Вновь проводим следующую секущую через две точки $\mathrm{A}_1\big(a_1; f(a_1)\big)$ и $\mathrm{A}_2\big(a_2; f(a_2)\big)$. Пусть секущая $\mathrm{A}_1\mathrm{A}_2$ пересекает ось OX в точке $\mathrm{A}_3(a_3; 0)$. Для этой секущей тоже верны равенства

$$\frac{a_3 - a_2}{0 - f(a_2)} = \frac{a_2 - a_1}{f(a_2) - f(a_1)} \Rightarrow a_3 = a_2 - f(a_2)\frac{a_2 - a_1}{f(a_2) - f(a_1)}.$$

……...

В результате K повторений ***двухточечной рекуррентной формулы***

$$a_{k+1} = a_k - f(a_k)\frac{a_k - a_{k-1}}{f(a_k) - f(a_{k-1})} \qquad (2.8)$$

построена последовательность $\{a_k\}_{k=2}^{\text{К}}$ точек пересечения семейства секущих с осью абсцисс (рис. 2.2).

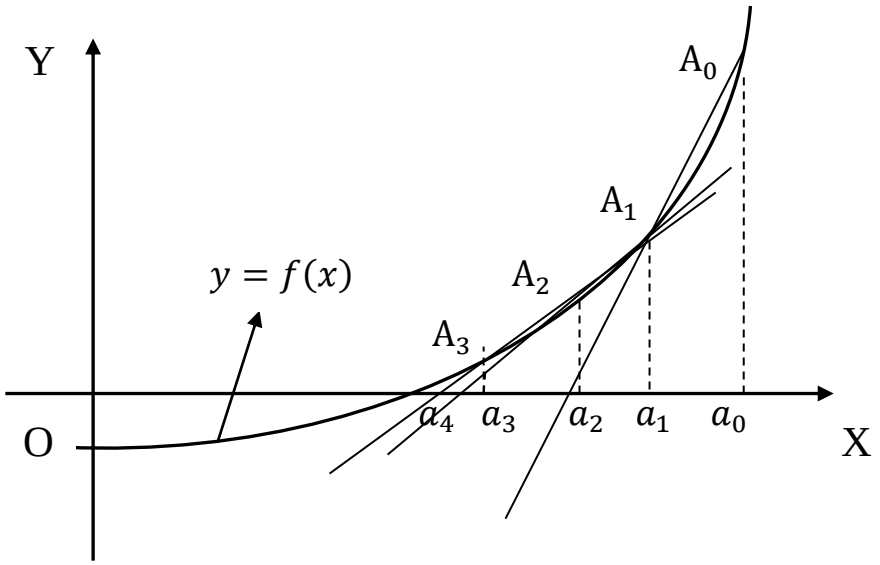

Рис. 2.2

СХОДИМОСТЬ И ОЦЕНКА ТОЧНОСТИ

Метод секущих является локально сходящимся методом. Доказано, что справедливо предельное равенство

$$\lim_{k\to\infty} |a_k| = r.$$

Метод секущих и метод хорд определяются однотипными формулами, но способы построения приближений имеют отличия. Это сказывается на скорости их сходимости. У метода хорд сходимость линейная, т.е. порядок сходимости $p = 1$. Метод секущих сходится быстрее, потому что его порядок сходимости $p \approx 1{,}62$. Чем больше скорость сходимости, тем выше точность. Поэтому можно сказать, что метод секущих более точный.

На каждом шаге итерации по заданной точности или невязке оценивается близость к точному корню с помощью неравенства $|a_{k+1} - a_k| < \varepsilon$ или неравенства $|f(a_{k+1})| < \delta$.

Для сохранения устойчивости метода на каждом шаге проверяется модуль значения знаменателя в рекуррентной формуле (2.8). Итерационный процесс заканчивается в том случае, когда $|f(a_{k+1}) - f(a_k)| \approx 0$.

В качестве приближённого корня берётся результат последнего шага итерации

$$\tilde{r} = a_{k+1}.$$

2.6. Метод Ньютона

В зарубежной литературе его называют методом Ньютона–Рафсона. Этот метод геометрически похож на метод секущих. Последовательности, которые сходятся к точному корню, строятся с помощью касательных. Поэтому метод

часто называют ***методом касательных***. С помощью него находится приближённое значение корня уравнения $f(x) = 0$, если функция $f(x)$ удовлетворяет условию следующей теоремы.

Теорема 3. Пусть функция $f(x)$ задана на $[a; b]$, непрерывны и монотонны её производные $f'(x)$, $f''(x)$. Предположим, что выполняется неравенство

$$f(a)f(b) < 0.$$

Если начальное приближение x_0 к корню так выбрано на $(a; b)$, что

$$f(x_0)f''(x_0) > 0,$$

тогда итерационный процесс, построенный по методу Ньютона, сходится.

АЛГОРИТМ ИТЕРАЦИОННОГО ПРОЦЕССА

В этом методе начальная точка приближения x_0 должна принадлежать малой окрестности точного корня уравнения: $x_0 \in (r - \beta; r + \beta)$.

1. Для построения итерационного процесса можно использовать уравнение касательной, проведённой к графику функции $f(x)$ в точке x_0

$$y - f(x_0) = f'(x_0)(x - x_0\ ).$$

Выразим x из этого уравнения. Пусть величина x равна x_1– точке пересечения касательной с осью абсцисс:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \qquad (2.9)$$

2. Вновь строится касательная к графику функции $f(x)$ в точке x_1. Точка пересечения новой касательной с осью OX выражается через предыдущую точку с помощью равенства (2.9):

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

……

После K повторений рекуррентной формулы

$$x_k = x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})} \qquad (2.10)$$

построена последовательность $\{x_k\}_{k=1}^{K}$ точек пересечения семейства касательных с осью абсцисс (рис. 2.3).

СХОДИМОСТЬ И ОЦЕНКА ТОЧНОСТИ

Метод Ньютона является локально сходящимся методом. Из теоремы 3 следует справедливость предельного равенства

$$\lim_{k\to\infty} |x_k| = r.$$

Последовательность приближений сходится очень быстро, так как порядок сходимости метода $p = 2$. Большая скорость сходимости при небольшом количестве действий даёт ***вычислительную устойчивость*** и хорошую точность. Поэтому итерационный процесс заканчивается при выполнении простого неравенства

$$|x_{k+1} - x_k| < \varepsilon \Longrightarrow \tilde{r} = x_{k+1}.$$

Эффективность метода Ньютона зависит ещё от степени громоздкости производной $f'(x)$. Если для её вычисления требуется сделать большой объём арифметических действий, то лучше использовать метод секущих.

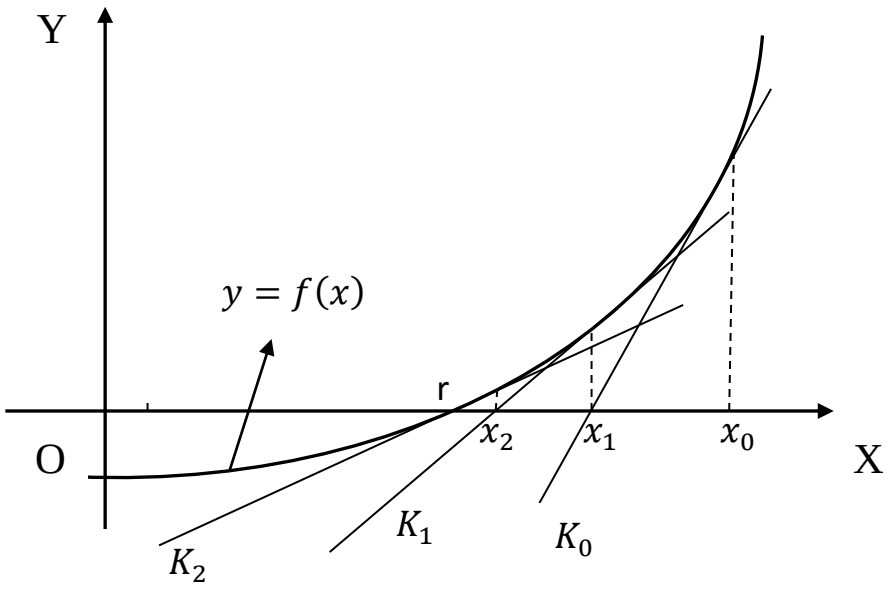

Рис. 2.3

Глава 3

РЕШЕНИЕ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ СИСТЕМ

Творчество математика в такой же степени есть создание прекрасного, как творчество живописца или поэта.
Г. Харди (1877– 1947) – английский математик

Почти 75% всех расчётных математических задач приводят к решению линейных алгебраических систем. Современные компьютерные пакеты программ содержат целый спектр методов решения таких систем.

Методы делятся на две группы: прямые и итерационные. Прямыми методами система решается за конечное число действий. Если они выполняются точно, то и ответ будет точным. Поэтому прямые методы ещё называют ***точными.***

Итерационные методы приводят к точному решению только в результате бесконечного сходящегося процесса. В каждом итерационном методе строится своё рекуррентное соотношение, на основе которого вычисляется следующее приближение.

Выбор метода решения системы зависит от структуры и размерности матрицы исходной системы. Линейную алгебраическую систему уравнений удобно сразу записывать в матричной форме

$$\boldsymbol{AX} = \boldsymbol{B}, \qquad (3.1)$$

где

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ a_{31} & a_{32} \cdots & a_{3n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \cdots & a_{nn} \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \cdots \\ b_n \end{pmatrix}.$$

Для дальнейшего освоения численных методов решения линейных систем необходимо познакомиться с матрицами особой структуры и понятием нормы векторов и матриц.

БАЗОВЫЕ ПОНЯТИЯ

Определение 1. Квадратная матрица $\boldsymbol{A}$ называется ***невырожденной***, или ***неособенной***, если её определитель не равен нулю.

Определение 2. Квадратная матрица $\boldsymbol{A}$ называется ***симметричной***, если

$$a_{ij} = a_{ji} \text{ для } i = 1,2,\cdots,n;\ j = 1,2,\cdots,n.$$

Определение 3. Симметричная матрица $\boldsymbol{A}$ называется ***положительно определённой***, если для любых ненулевых векторов $\boldsymbol{X}$ верно соотношение

$$(\boldsymbol{AX}, \boldsymbol{X}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j > 0.$$

Определение 4. Квадратная матрица $\boldsymbol{A}$ называется ***неопределённой*** матрицей, если знак скалярного произведения $(\boldsymbol{AX}, \boldsymbol{X})$ зависит от вектора $\boldsymbol{X}$.

Определение 5. Квадратная матрица называется ***верхней треугольной матрицей***, если все элементы матрицы ниже главной диагонали равны нулю. Если все элементы матрицы выше главной диагонали равны нулю, такая матрица называется ***нижней треугольной***.

Определение 6. Квадратная матрица называется ***диагональной***, если все её элементы, кроме элементов главной диагонали, равны нулю.

Определение 7. Квадратная матрица $\boldsymbol{A}$ называется ***ортогональной***, если транспонированная матрица $\boldsymbol{A}^{\mathsf{T}}$ равна обратной матрице: $\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{A}^{-1}$.

Из определения вытекает, что определитель ортогональной матрицы всегда не равен нулю.

Определение 8. Система линейных уравнений называется ***нормальной системой***, если матрица этой системы – положительно определённая матрица.

3.1. Нормы векторов и матриц

Норма произвольного вектора $\boldsymbol{X}$ или матрицы $\boldsymbol{A}$ – это неотрицательное число, которое обозначается соответственно в виде $\|\boldsymbol{X}\|$ или $\|\boldsymbol{A}\|$.

По определению норма обладает следующими свойствами:

1. Если $\boldsymbol{X} \neq \mathbf{0}$, то $\|\boldsymbol{X}\| > 0$, если $\boldsymbol{X} = \mathbf{0}$, то $\|\boldsymbol{X}\| = 0$;

   если $\boldsymbol{A} \neq \mathbf{0}$, то $\|\boldsymbol{A}\| > 0$, если $\boldsymbol{A} = \mathbf{0}$, то $\|\boldsymbol{A}\| = 0$.

2. Свойство однородности:

   $\|\mathrm{c} \cdot \mathbf{X}\| = \mathrm{c} \cdot \|\mathbf{X}\|$, где c – любое действительное число;

   $\|\mathrm{c} \cdot \boldsymbol{A}\| = \mathrm{c} \cdot \|\boldsymbol{A}\|$, где c – любое действительное число.

3. $\|\mathbf{X} + \mathbf{Y}\| \leq \|\mathbf{X}\| + \|\mathbf{Y}\|$ для векторов одинаковой размерности;

   $\|\boldsymbol{A} + \mathbf{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$, где $\boldsymbol{A}$ и $\boldsymbol{B}$ – матрицы одного порядка.

ОСНОВНЫЕ ВЕКТОРНЫЕ НОРМЫ

$\|\mathbf{X}\|_1 = (\sum_{n=1}^{N}|x_n|)$ – ***окаэдрическая норма***;

$\|\mathbf{X}\|_p = \left(\sum_{n=1}^{N}\left|x_n^p\right|\right)^{\frac{1}{p}}$ – ***p-норма Гёльдера***, $p \geq 1$;

$\|\mathbf{X}\|_E = (\sum_{n=1}^{N} x_n^2)^{\frac{1}{2}}$ – ***евклидова норма***, или ***сферическая норма***;

$\|\mathbf{X}\|_\infty = \max_n|x_n|$ – ***кубическая норма***.

ОСНОВНЫЕ МАТРИЧНЫЕ НОРМЫ

$$\|\boldsymbol{A}\| = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|,$$

$$\|\boldsymbol{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|,$$

$$\|\boldsymbol{A}\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|,$$

$$\|\boldsymbol{A}\|_E = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a^2{}_{ij}\right)^{\frac{1}{2}}.$$

Определение 9. Матричная норма называется ***согласованной с векторной нормой***, если

$$\|\boldsymbol{AX}\| \le \|\boldsymbol{A}\| \cdot \|\mathbf{X}\|.$$

Например, норма $\|\boldsymbol{A}\|_1$ согласована с нормой $\|\mathbf{X}\|_1$, норма $\|\boldsymbol{A}\|_\infty$ согласована с нормой $\|\boldsymbol{X}\|_\infty$, норма $\|\boldsymbol{A}\|_E$ согласована с нормой $\|\mathbf{X}\|_E$.

Далее мы будем использовать только согласованные нормы, поэтому очевидно, что если $\boldsymbol{AX} = \boldsymbol{f}$, то $\|\boldsymbol{f}\| \le \|\boldsymbol{A}\| \cdot \|\mathbf{X}\|$.

Определение 10. ***Числом обусловленности*** (стандартным числом обусловленности) называется величина

$$cond(\boldsymbol{A}) = \|\boldsymbol{A}\| \cdot \|\boldsymbol{A}^{-1}\| \quad (conditioned - \textit{обусловленный}).$$

Очевидно, что всегда $cond(\boldsymbol{A}) \ge 1$:

$$cond(\boldsymbol{A}) = \|\boldsymbol{A}\| \cdot \|\boldsymbol{A}^{-1}\| \ge \|\boldsymbol{A}^{-1}\boldsymbol{A}\| = \|\boldsymbol{E}\| = 1.$$

Число обусловленности используется для оценки погрешности найденного решения системы.

3.2 Источники погрешностей

Правая часть системы обычно связана с параметрами исходной задачи, поэтому она изначально содержит некоторое ***возмущение*** $\delta\boldsymbol{B}$, с учётом которого систему можно записать в виде

$$\boldsymbol{A}(\boldsymbol{X} + \delta\boldsymbol{X}) = \boldsymbol{B} + \delta\boldsymbol{B}.$$

Рассмотрим, как возмущение правой части системы влияет на погрешность $\delta\boldsymbol{X}$ (возмущение) решения системы. Введём с помощью норм две относительные погрешности

$$\frac{\|\delta\boldsymbol{X}\|}{\|\boldsymbol{X}\|} \text{ и } \frac{\|\delta\boldsymbol{B}\|}{\|\boldsymbol{B}\|}. \qquad (3.2)$$

Найдём связь между нормой погрешности решения и нормой возмущения правой части системы:

$$\boldsymbol{A}(\boldsymbol{X}+\delta\boldsymbol{X})=\boldsymbol{B}+\ \delta\boldsymbol{B};\ \boldsymbol{AX}=\boldsymbol{B}\Longrightarrow\boldsymbol{A}\cdot\delta\boldsymbol{X}=\ \delta\boldsymbol{B}\Longrightarrow$$

$$\delta\boldsymbol{X}=\boldsymbol{A}^{-1}\cdot\delta\boldsymbol{B}\Rightarrow\|\delta\boldsymbol{X}\|\leq\|\boldsymbol{A}^{-1}\|\cdot\|\delta\boldsymbol{B}\|.$$

Теперь можно оценить отношение погрешностей (3.2):

$$\frac{\|\delta\mathbf{X}\|\cdot\|\mathbf{B}\|}{\|\mathbf{X}\|\cdot\|\delta\mathbf{B}\|}\leq\frac{\|\delta\mathbf{X}\|\cdot\|\mathbf{A}\|\cdot\|\mathbf{X}\|}{\|\mathbf{X}\|\cdot\|\delta\mathbf{B}\|}\leq\frac{\|\mathbf{A}^{-1}\|\cdot\|\delta\mathbf{B}\|\cdot\|\mathbf{A}\|}{\|\delta\mathbf{B}\|}\leq\|\mathbf{A}\|\cdot\|\mathbf{A}^{-1}\|\Rightarrow$$

$$\frac{\|\delta\mathbf{X}\|}{\|\mathbf{X}\|}\leq\ cond(\mathbf{A})\cdot\frac{\|\delta\mathbf{B}\|}{\|\mathbf{B}\|}.$$

Из последнего неравенства следует, что число обусловленности можно рассматривать как коэффициент усиления исходного возмущения $\delta\boldsymbol{B}$. Если величина $cond(\boldsymbol{A})$ достаточно близка к единице, то говорят, что исходная система ***хорошо обусловлена*** и возмущение $\delta\boldsymbol{X}$ тоже относительно мало.

Предположим теперь, что матрица системы задана с некоторым возмущением

$$(\boldsymbol{A}+\delta\boldsymbol{A})(\boldsymbol{X}+\delta\boldsymbol{X})=\boldsymbol{B}.$$

Доказано, что и в этом случае число обусловленности $cond(\boldsymbol{A})$ служит коэффициентом роста относительной погрешности решения:

$$\frac{\|\delta\boldsymbol{X}\|}{\|\boldsymbol{X}\|}\leq\ cond(\boldsymbol{A})\cdot\frac{\|\delta\boldsymbol{A}\|}{\|\boldsymbol{A}\|}.$$

Замечание 1. Во всех стандартных программах для решения линейных алгебраических систем оценивается число обусловленности $cond(\boldsymbol{A})$.

Замечание 2. При решении систем следует учитывать и погрешности численного метода.

3.3 Прямые методы решения систем

Для решения линейных алгебраических систем наиболее часто используют пять прямых методов: Крамера, Гаусса, квадратного корня, трёхдиагональной прогонки и матричный метод.

Метод Крамера и матричный метод представлены в начальном курсе высшей алгебры. Их лучше использовать для решения систем 2-го и 3-го порядка. С увеличением порядка системы они становятся неустойчивыми методами относительно ошибок округления, что приводит к быстрому росту погрешностей. В этих методах используется определитель системы, который стоит в знаменателе формул решений. Если модуль определителя намного меньше единицы $(\Delta\ll 1)$, то формулы будут давать большие погрешности. Мы не

будем рассматривать эти два метода, поскольку в рамках численных методов они почти не используются.

3.3.1. Метод Гаусса

Наиболее популярным методом решения линейных алгебраических систем является метод Гаусса (метод исключения). Его реализация состоит из двух частей.

Первая часть – ***прямой ход***. Строится расширенная матрица системы

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & b_n \end{pmatrix}. \qquad (3.3)$$

С помощью элементарных преобразований расширенная матрица (3.3) приводится к трапециевидному виду (3.4). Матрица системы становится верхней треугольной матрицей. У этой матрицы ниже главной диагонали расположены только нули, а на главной диагонали – единицы:

$$\begin{pmatrix} 1 & \breve{a}_{12} & \breve{a}_{13} & \cdots & \breve{a}_{1n} & \breve{b}_1 \\ 0 & 1 & \breve{a}_{23} & \cdots & \breve{a}_{2n} & \breve{b}_2 \\ 0 & 0 & 1 & \cdots & \breve{a}_{3n} & \breve{b}_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \breve{b}_n \end{pmatrix}. \qquad (3.4)$$

Элементарные преобразования расширенной матрицы системы приводят к другой эквивалентной расширенной матрице. Из курса линейной алгебры известно, что системы с эквивалентными расширенными матрицами равносильны, т. е. их решения совпадают.

Вторая часть – ***обратный ход***. С помощью элементарных преобразований матрица (3.4) приводится к матрице (3.5) . В результате чего свободный столбец матрицы (3.5) становится решением исходной системы (3.1).

Если определитель исходной системы не равен нулю, то расширенную матрицу такой системы всегда можно преобразовать к виду

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & x_1 \\ 0 & 1 & 0 & \cdots & 0 & x_2 \\ 0 & 0 & 1 & \cdots & 0 & x_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & x_n \end{pmatrix}. \qquad (3.5)$$

Недостатком метода исключения является накопление вычислительной погрешности. Для снижения скорости роста погрешностей разработана мо-

дификация метода Гаусса. Её наиболее простой вариант – это ***метод Гаусса с выбором главного элемента по столбцу***.

АЛГОРИТМ С ВЫБОРОМ ГЛАВНОГО ЭЛЕМЕНТА ПО СТОЛБЦУ

Прямой ход

1. В первом столбце матрицы системы выбирается самый большой по модулю элемент. Он называется ***главным элементом***. Вся строка с главным элементом делится на этот элемент и меняется местами с первой строкой. Новая первая строка называется ***ведущей строкой*** и начинается с единицы.

2. С помощью 1-й ведущей строки и элементарных преобразований строк расширенной матрицы все элементы первого столбца, стоящие ниже 1-й строки превращаются в нули.

3. Во 2-ом столбце преобразованной матрицы системы, начиная со 2-ой строки, находится самый большой по модулю элемент. Вся строка с найденным новым главным элементом делится на него и меняется местами со 2-ой строкой расширенной матрицы. Новая 2-ая строка теперь будет являться ведущей строкой.

4. С помощью новой ведущей строки и элементарных преобразований строк расширенной матрицы все элементы 2-го столбца, стоящие ниже второй строки превращаются в нули.

……

К. В (k–1)-ом столбце преобразованной матрицы системы, начиная с (k–1)-ой строки, находится самый большой по модулю элемент. Вся строка с новым главным элементом делится на него и меняется местами с $(k-1)$-ой строкой расширенной матрицы. Эта строка теперь будет ведущей строкой.

Прямой ход метода Гаусса заканчивается, когда $k = n$. Главная диагональ матрицы системы теперь будет состоять только из единиц, а ниже главной диагонали – только нули.

Обратный ход

1. В качестве ведущей строки выбирается последняя строка. С помощью единицы, стоящей в последнем столбце последней строки все элементы этого столбца выше последней строки преобразуются в нули.

2. В качестве ведущей строки выбирается $(n-1)$ строка. С помощью единицы, стоящей в $(n-1)$ столбце и элементарных преобразований с расширенной матрицей все элементы в $(n-1)$ столбце и выше $(n-1)$ строки преобразуются в нули.

…….

N–1. В качестве ведущей строки выбирается вторая строка. С помощью единицы, стоящей во 2-м столбце и 2-й строке, элемент из первой строки и 2-го столбца превращается в ноль. На этом заканчивается обратный ход.

Матрица системы становится единичной, а свободный столбец – решением исходной системы (3.1).

Описанный выше алгоритм обладает ***вычислительной устойчивостью***.

Замечание 3. Аналогичным образом можно построить алгоритм метода Гаусса с выбором главного элемента по строке, но соответствующий алгоритм будет более громоздким. Можно построить алгоритм решения системы с выбором главного элемента по всей матрице системы. В этом случае резко увеличится число операций деления, поэтому увеличится вычислительная погрешность.

3.3.2. Метод квадратного корня

Решение многих физических задач связано с решением линейных систем $\boldsymbol{AX} = \boldsymbol{B}$ с симметричной положительно определённой матрицей $\boldsymbol{A}$.

Такие системы лучше решать методом квадратного корня. Он использует приблизительно вдвое меньше арифметических операций и объёма памяти, чем метод Гаусса. Но описание метода квадратного корня не входит в рамки нашего учебника. Оно подробно представлено в работе [27].

3.3.3. Метод трёхдиагональной прогонки

Среди линейных алгебраических систем особое место занимают системы с ***трёхдиагональной матрицей***. Они встречаются в теории аппроксимации, в методах решения обыкновенных дифференциальных уравнений и уравнений в частных производных.

Замечание 4. Термин *прогонка* характерен для Русской математической школы и был введён в середине XX века.

Запишем систему в матричном виде

$$\boldsymbol{AX} = \boldsymbol{f}, \text{ где} \qquad (3.6)$$

$$\boldsymbol{A} = \begin{pmatrix} C_1 & B_1 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ A_2 & C_2 & B_2 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & A_3 & C_3 & B_3 & 0 & 0 & 0 & 0 & \cdots & & 0 \\ 0 & 0 & A_4 & C_4 & B_4 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ & & & & \cdots & & & & & & \\ 0 & 0 & 0 & \cdots & 0 & A_{n-2} & C_{n-2} & B_{n-2} & 0 & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & A_{n-1} & C_{n-1} & B_{n-1} & & \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & A_n & C_n & \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \cdots \\ x_{n-2} \\ x_{n-1} \\ x_n \end{pmatrix}, \boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ \cdots \\ f_{n-2} \\ f_{n-1} \\ f_n \end{pmatrix}.$$

Главная диагональ системы (3.6) состоит из элементов $\{C_i\}_{i=1}^{n}$. Диагональ ниже главной – из элементов $\{A_i\}_{i=2}^{n}$. Диагональ выше главной – из элементов $\{B_i\}_{i=1}^{n-1}$. Остальные элементы матрицы A равны нулю.

Данную систему уравнений можно записать в более простом виде, если ввести ещё четыре вспомогательных нулевых элемента $A_1 = 0,\ B_n = 0,\ x_0 =$

0, $x_{n+1} = 0$. Теперь мы можем записать систему в виде одного обоб-щённого уравнения

$$A_i x_{i-1} + C_i x_i + B_i x_{i+1} = f_i,\ i = 1, 2, 3, \cdots, n. \qquad (3.7)$$

Будем предполагать, что определитель системы не равен нулю, поэтому решение существует и единственно. Система линейна, следовательно, любая пара соседних неизвестных тоже связана линейной зависимостью. Запишем её в виде

$$x_i = \alpha_i x_{i+1} + \beta_i \Longrightarrow x_{i-1} = \alpha_{i-1} x_i + \beta_{i-1.} \qquad (3.8)$$

Для каждой пары неизвестных будет своя единственная пара коэффициентов $\{\alpha_i, \beta_i\}_{i=1}^{n}$. Построим формулы нахождения всех α_i, β_i. Для этого используем линейную зависимость (3.8). Заменим в обобщённом уравнении (3.7) x_{i-1} на $(\alpha_{i-1} x_i + \beta_{i-1})$:

$$A_i(\alpha_{i-1} x_i + \beta_{i-1}) + C_i x_i + B_i x_{i+1} = f_i.$$

Последнее уравнение представим в виде

$$(A_i \alpha_{i-1} + C_i) x_i = -B_i x_{i+1} + f_i - A_i \beta_{i-1} \Longrightarrow$$

$$x_i = \frac{-B_i}{(A_i \alpha_{i-1} + C_i)} x_{i+1} + \frac{f_i - A_i \beta_{i-1}}{(A_i \alpha_{i-1} + C_i)}. \qquad (3.9)$$

Будем предполагать, что в этом равенстве знаменатель

$$(A_i \alpha_{i-1} + C_i) \neq 0.$$

Сравним уравнение (3.9) с исходной линейной зависимостью между той же парой неизвестных

$$x_i = \alpha_i x_{i+1} + \beta_i.$$

В силу единственности решения системы приходим к очевидным равенствам

$$\alpha_i = \frac{-B_i}{(A_i \alpha_{i-1} + C_i)}, \qquad \beta_i = \frac{f_i - A_i \beta_{i-1}}{(A_i \alpha_{i-1} + C_i)}. \qquad (3.10)$$

В итоге построены рекуррентные формулы (3.10), в которых каждая пара α_i, β_i выражена через пару предыдущих коэффициентов α_{i-1}, β_{i-1}. Если найдём значения первых коэффициентов α_1, β_1, то с помощью формул (3.10) сможем вычислить все остальные коэффициенты.

Для нахождения α_1, β_1 обратимся к первому уравнению системы

$$C_1 x_1 + B_1 x_2 = f_1 \Longrightarrow x_1 = \frac{-B_1}{C_1} x_2 + \frac{f_1}{C_1} \Longrightarrow$$

$$\alpha_1 = \frac{-B_1}{C_1}, \quad \beta_1 = \frac{f_1}{C_1}.$$

Остальные коэффициенты α_i, β_i вычисляются по рекуррентным формулам (3.10).

Процесс построения всех коэффициентов $\{\alpha_i, \beta_i\}_{i=1}^{n}$ называется ***прямым ходом прогонки***, а сами коэффициенты называются ***прогоночными коэффициентами.*** После нахождения всех прогоночных коэффициентов очень просто вычисляется решение исходной системы.

Процесс начинается с вычисления x_n. Обратимся к линейной зависимости

$$x_n = \alpha_n \cdot x_{n+1} + \beta_n;\ \alpha_n = \frac{-B_n}{(A_n \cdot \alpha_{n-1} + C_n)}\ ,\ \ B_n = 0 \Longrightarrow \alpha_n = 0 \Longrightarrow x_n = \beta_n.$$

Далее вычисляются $x_{n-1}, x_{n-2}, \cdots, x_1$:

$$x_{n-1} = \alpha_{n-1} \cdot x_n + \beta_{n-1},\ x_{n-2} = \alpha_{n-2} \cdot x_{n-1} + \beta_{n-2}, \cdots, x_1 = \alpha_1 x_2 + \beta_1.$$

Эта часть алгоритма нахождения $\{x_i\}_{i=1}^{n}$ называется ***обратным ходом прогонки.***

Метод прогонки можно успешно применять, если в процессе вычислений не возникнет деление на ноль. Для метода трёхдиагональной прогонки доказана следующая теорема.

Теорема 1. Если выполняется условие ***доминирования*** главной диагонали:

$$\begin{cases} |C_i| \geq |A_i| + |B_i|,\ \{A_i \neq 0\}_{i=2}^{n},\ \{B_i \neq 0\}_{i=1}^{n-1}, \\ |C_1| \geq |B_1|,\ \ |C_n| \geq |A_n|, \end{cases} \tag{3.11}$$

тогда для всех прогоночных коэффициентов $\{\alpha_i\}_{i=1}^{n}$ справедливы два соотношения

$$|\alpha_i| < 1 \ \ \text{и} \ \ (A_i \alpha_{i-1} + C_i) \neq 0\ , i = 1, 2, \cdots, n.$$

В качестве оценки метода прогонки используют два определения.

Определение 11. Метод прогонки называется ***корректным***, если знаменатели всех прогоночных коэффициентов не равны нулю:

$$(A_i \alpha_{i-1} + C_i) \neq 0, \qquad i = 1, 2, \cdots, n.$$

Определение 12. Метод прогонки называется ***устойчивым***, если для всех прогоночных коэффициентов $\{\alpha_i\}_{i=1}^{n}$ справедливо неравенство $|\alpha_i| < 1$.

Доминирование главной диагонали (3.11) – это достаточное условие корректности и устойчивости метода прогонки.

Замечание 5. Доказано, что в случае доминирования главной диагонали метод прогонки устойчив относительно вычислительных округлений даже при

большой размерности исходной системы уравнений. Метод устойчив для хорошо обусловленных систем и в том случае, когда диагональное преобладание (доминирование) нарушается.

3.4. Итерационные методы решения линейных систем

Применение популярного метода Гаусса для систем больших размерностей приводит к накоплению ошибок округления. Поэтому решение может иметь значительную погрешность. В таком случае более эффективны итерационные методы. Они легко реализуются на компьютере. Важным достоинством итерационных методов является возможность заранее задавать необходимую точность искомого решения.

Любой итерационный метод работает в следующем порядке.

1. Задаётся начальное приближение $\boldsymbol{X}^{(0)}$.

2. Создаётся алгоритм, с помощью которого на каждом шаге итерации строится приближение $\boldsymbol{X}^{(k)}$ по предыдущему $\boldsymbol{X}^{(k-1)}$ приближению.

Используются только те итерационные процессы, для которых верно предельное равенство

$$\lim_{k\to+\infty} \boldsymbol{X}^{(k)} = \boldsymbol{X}, \qquad (3.12)$$

где вектор $\boldsymbol{X}$ – точное решение исходной системы (неподвижная точка).
Если равенство (3.12) выполняется, говорят, что ***итерационный процесс сходится.***

Для решения систем следует заранее задать необходимую точность ε. В конце каждого шага итерации оценивается близость двух последовательных приближёний решения системы:

$$\left|\frac{x_i^{(k)} - x_i^{(k-1)}}{x_i^{(k)}}\right| < \varepsilon \text{ для } i = 1,2,\cdots,N. \qquad (3.13)$$

Если это неравенство выполняется для всех $\left\{x_i^{(k)}\right\}_{i=1}^{N}$, итерационный процесс заканчивается. В качестве решения системы можно взять любое из приближений

$$\boldsymbol{X}^{(k)} = \left(x_1^{(k)} x_2^{(k)} x_3^{(k)} \cdots x_n^{(k)}\right) \text{ или } \boldsymbol{X}^{(k-1)} = \left(x_1^{(k-1)} x_2^{(k-1)} x_3^{(k-1)} \cdots x_n^{(k-1)}\right).$$

Чем выше должна быть точность решения, тем больше надо сделать шагов итераций.

В дальнейшем будем использовать следующие определения.

Определение 13. Система линейных алгебраических уравнений называется ***приведённой системой***, если она имеет вид

$$\boldsymbol{X} = \boldsymbol{C}\boldsymbol{X} + \boldsymbol{D}, \qquad (3.14)$$

где $\boldsymbol{X}$ – вектор из неизвестных, $\boldsymbol{D}$ – вектор из числовых элементов, $\boldsymbol{C}$ – матрица системы.

Определение 14. Матрица $\boldsymbol{A}$ называется матрицей со ***строгим диагональным преобладанием***, если для неё выполняется условие

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{N} |a_{ij}|, \qquad i = 1,2,\cdots,N. \qquad (3.15)$$

3.4.1. Метод простых итераций

На первом этапе работы исходную систему $\boldsymbol{AX} = \boldsymbol{B}$ преобразуют в приведённую систему (3.14) . Далее осуществляется итерационный процесс по следующему алгоритму.

1. Задаётся начальное приближение решения $\boldsymbol{X}^{(0)}$.
2. Приведённая система (3.14) записывается в виде рекуррентной формулы

$$\boldsymbol{X}^{(k+1)} = \boldsymbol{C}\boldsymbol{X}^{(k)} + \boldsymbol{D}. \qquad (3.16).$$

После каждого шага итерации оценивается приближение с помощью неравенства (3.13). На каждом последующем шаге итерации происходит пересчёт предыдущего приближения с помощью формулы (3.16).

ДОСТАТОЧНЫЕ УСЛОВИЯ СХОДИМОСТИ

1. Достаточным условием сходимости метода простых итераций является выполнение неравенства $\|\boldsymbol{C}\| < 1$ при любой матричной норме. В этом случае сходимость итерационного процесса обеспечена для любого начального приближения.

2. Итерационный метод сходится и достаточно быстро, если исходная матрица $\boldsymbol{A}$ системы имеет строгое диагональное преобладание (3.15).

Пример. Решим заданную систему методом простых итераций.

$$\begin{cases} 4x_1 + 0.24x_2 - 0.08x_3 = 8, \\ 0.09x_1 + 3x_2 - 0.15x_3 = 9, \\ 0.04x_1 - 0.08x_2 + 4x_3 = 20. \end{cases}$$

Решение. Преобразуем систему к приведённому рекуррентному виду:

$$\begin{cases} x_1 = 2 - 0.06x_2 + 0.02x_3, \\ x_2 = 3 - 0.03x_1 + 0.05x_3, \\ x_3 = 5 - 0{,}01x_1 + 0{,}02x_2 . \end{cases} \Rightarrow \begin{cases} x_1^{(i+1)} = -0.06x_2^{(i)} + 0.02x_3^{(i)} + 2, \\ x_2^{(i+1)} = -0.03x_1^{(i)} + 0.05x_3^{(i)} + 3, \\ x_3^{(i+1)} = -0{,}01x_1^{(i)} + 0{,}02x_2^{(i)} + 5 . \end{cases}$$

Зададим начальное приближение. Пусть $x_1^{(0)} = 2, x_2^{(0)} = 3,\ x_3^{(0)} = 5$. Далее подставляем эти значения в правую часть последней системы и вычисляем первое приближение: $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}$. На этом заканчивается первая итерация. Аналогичным образом вычисляется следующее приближение $x_1^{(2)}, x_2^{(2)}, x_3^{(2)}$.

Таблица 3.1

i –номер итерации	$x_1^{(i)}$	$x_2^{(i)}$	$x_3^{(i)}$
i=0	2	3	5
i=1	1,92	3,19	5,04
i=2	1,9094	3,1944	5,0446
i=3	1,909228	3,194948	5,044794
i=4	1,909109	3,194963	5,044807

Из таблицы 3.1. видно, что если $\varepsilon = 0{,}001$, то для такой точности достаточно сделать четыре шага итерации.

Для использования метода простых итераций необходимо, прежде всего, чтобы все диагональные элементы исходной матрицы системы не равнялись нулю. Для невырожденной матрицы этого всегда можно добиться с помощью перестановок строк исходной матрицы.

Кроме метода простых итераций созданы другие итерационные методы решения линейных алгебраических систем. Остановимся на некоторых из них.

3.4.2 Метод Якоби.

Этот метод является модификацией метода простых итераций. В нём матрицу исходной системы представляют в виде суммы трёх матриц:

$$\boldsymbol{A = L + D + R},$$

$\boldsymbol{L}$ – диагональная матрица, $\boldsymbol{D}$ – верхняя треугольная матрица с нулевой главной диагональю, $\boldsymbol{R}$ – нижняя треугольная матрица с нулевой главной диагональю:

$$\boldsymbol{L}=\begin{pmatrix} a_{11} & \cdots 0 \cdots & 0 \\ 0 \cdots & a_{kk} & 0 \\ 0 \cdots & \cdots 0 \cdots & \cdots a_{nn} \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 0 & \cdots a_{1k} \cdots & a_{1n} \\ 0 & \cdots 0 & a_{kn} \\ 0 \cdots & \cdots\ 0 & 0 \end{pmatrix},$$

$$\boldsymbol{R} = \begin{pmatrix} 0 & \cdots 0 \cdots & 0 \\ a_{21} \vdots & 0 \cdots \vdots & 0 \vdots \\ a_{n1} & \cdots a_{nk} & 0 \end{pmatrix}.$$

Исходная система теперь может быть представлена в виде

$$\boldsymbol{LX} + (\boldsymbol{D} + \boldsymbol{R})\boldsymbol{X} = \boldsymbol{B}.$$

Будем предполагать, что на главной диагонали матрицы $\boldsymbol{L}$ нет нулей. Тогда определитель матрицы $\boldsymbol{L}$ не равен нулю, и у неё есть обратная матрица $\boldsymbol{L}^{-1}$. Поэтому исходную систему можно записать в виде

$$\boldsymbol{X} = -\boldsymbol{L}^{-1}(\boldsymbol{D} + \boldsymbol{R})\boldsymbol{X} + \boldsymbol{L}^{-1}\boldsymbol{B}.$$

Далее превращаем последнее равенство в рекуррентную формулу

$$\boldsymbol{X}^{(k+1)} = -\boldsymbol{L}^{-1}(\boldsymbol{D} + \boldsymbol{R})\boldsymbol{X}^{(k)} + \boldsymbol{L}^{-1}\boldsymbol{B}.$$

Использование этого уравнения для итерационного процесса приводит к решению с заданной точностью. Достаточные условия сходимости в методе Якоби такие же, как в методе простых итераций.

3.4.3. Метод Гаусса–Зейделя

Этот метод является видоизменением метода простых итераций. В предыдущих итерационных методах на каждом шаге итерации строится вектор приближения

$\boldsymbol{X}^{(k)} = \left(x_1^{(k)} x_2^{(k)} x_3^{(k)} \cdots x_n^{(k)}\right)$ полностью на основе вектора

$$\boldsymbol{X}^{(k-1)} = \left(x_1^{(k-1)} x_2^{(k-1)} x_3^{(k-1)} \cdots x_n^{(k-1)}\right),$$

полученного на предыдущем шаге итерации.

В методе Гаусса–Зейделя любой $x_i^{(k)}$ элемент вектора $\boldsymbol{X}^{(k)}$ вычисляется с помощью уже найденных на этом же шаге итерации значений

$\left(x_1^{(k)} x_2^{(k)} x_3^{(k)} \cdots x_{i-1}^{(k)}\right)$ и значений $\left(x_{i+1}^{(k-1)} x_{i+2}^{(k-1)} x_{i+3}^{(k-1)} \cdots x_n^{(k-1)}\right)$ из

предыдущего шага итерации.

Доказано, что если матрица исходной системы имеет диагональное преобладание (3.11), то метод Гаусса–Зейделя сходится быстрее метода Якоби и метода простых итераций. Если у исходной матрицы системы нет диагонального преобладания, но сама матрица симметрична, положительно определена, тогда этот метод тоже сходится, но медленнее метода Якоби.

Бывают такие системы, для которых метод Гаусса–Зейделя сходится, а метод Якоби нет. Бывает и наоборот.

Глава 4

ПРОБЛЕМА СОБСТВЕННЫХ ЗНАЧЕНИЙ

Много шума из ничего – ноль,
но как число, оно заслуживает
особо пристального рассмотрения.
Ричард Браун. «Математика за 30 секунд»

Решение различных задач в гидрометеорологии, экономике, технике и физике часто связано с алгебраической задачей нахождения собственных значений и собственных векторов квадратной матрицы. Эта задача является одной из самых сложных задач линейной алгебры. Поэтому её часто называют ***проблемой собственных значений (чисел)***.

Остановимся кратко только на основных моментах постановки задачи, а затем перейдём к соответствующим численным методам решения.

4.1. Постановка задачи

Собственным вектором квадратной матрицы

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ a_{31} & a_{32} \cdots & a_{3n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \cdots & a_{nn} \end{pmatrix} \qquad (4.1)$$

называется ненулевой вектор

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad (4.2)$$

который удовлетворяет матричному уравнению

$$A\boldsymbol{X} = \lambda \boldsymbol{X}. \qquad (4.3)$$

Число λ в уравнении (4.3) называется ***собственным значением матрицы*** или ***собственным числом матрицы*** $\boldsymbol{A}$.

Замечание 1. Часто говорят, что собственный вектор $\boldsymbol{X}$ принадлежит собственному значению λ.

Будем предполагать, что все элементы рассматриваемых матриц – это действительные числа.

Ставится задача о нахождении таких собственных чисел и собственных векторов, для которых верно равенство (4.3).

Для решения задачи преобразуем уравнение (4.3)

$$\boldsymbol{AX} - \lambda \boldsymbol{X} = 0 \Rightarrow \boldsymbol{AX} - \lambda \boldsymbol{EX} = 0 \Rightarrow (\boldsymbol{A} - \lambda \boldsymbol{E})\boldsymbol{X} = 0. \qquad (4.4)$$

Последнее уравнение – это матричная форма записи линейной однородной системы

$$\begin{cases} (a_{11}-\lambda)x_1 + \quad a_{12}x_2 + \cdots \quad + a_{1n}x_n = 0, \\ a_{21}x_1 + \quad (a_{22}-\lambda)x_2 + \cdots \quad + a_{2n}x_n = 0, \\ a_{31}x_1 + \quad a_{32}x_2 + \cdots\cdots\cdots \quad + a_{3n}x_n = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + \quad a_{n2}x_2 + \cdots \quad + (a_{nn}-\lambda)x_n = 0. \end{cases} \qquad (4.5)$$

Из основного курса высшей алгебры известно, что однородная линейная алгебраическая система имеет ненулевое решение

$$\boldsymbol{X} = (x_1\ x_2\ \cdots x_n)^{\mathsf{T}},$$

если её определитель равен нулю:

$$\Delta = \begin{vmatrix} (a_{11}-\lambda) & a_{12}\cdots & a_{1n} \\ a_{21} & (a_{22}-\lambda)\cdots & a_{2n} \\ a_{31} & a_{32}\cdots\cdots\cdots & a_{3n} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2}\cdots & (a_{nn}-\lambda) \end{vmatrix} = 0. \qquad (4.6)$$

Равенство (4.6) можно представить в виде алгебраического уравнения n-й степени относительно собственного числа λ:

$$(-1)^n(\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + c_{n-1}\lambda + c_n) = 0. \qquad (4.7)$$

Уравнение (4.7) называется ***характеристическим (вековым) уравнением матрицы.*** Корни этого уравнения – собственные числа матрицы $\boldsymbol{A}$. В общем случае собственные числа могут быть комплексными числами.

На основе следствия из теоремы Гаусса (глава 3) можно сказать, что количество собственных чисел с учётом их кратности всегда равно размерности матрицы. Множество всех собственных значений матрицы называется её ***спектром***.

Для дальнейшего знакомства с методами нахождения собственных чисел и векторов матрицы напомним базовые определения и свойства векторов в n-мерном векторном пространстве.

ВЕКТОРЫ

Определение 1. Ненулевой вектор $\boldsymbol{X}$ называется ***линейной комбинацией векторов*** $\{\boldsymbol{X}_i\}_{i=1}^k$, если существуют такие числа $\{\alpha_i \neq 0\}_{i=1}^k$, при которых верно равенство

$$\boldsymbol{X} = \alpha_1\boldsymbol{X}_1 + \alpha_2\boldsymbol{X}_2 + \cdots + \alpha_k\boldsymbol{X}_k.$$

Определение 2. Система векторов $\{\boldsymbol{X}_i\}_{i=1}^{k}$ называется линейно зависимой, если хотя бы один из них можно представить в виде линейной комбинации остальных векторов.

Определение 3. Система векторов называется ***линейно независимой***, если ни один из векторов этой системы не является линейной комбинацией других векторов этой же системы.

Можно сказать, что система векторов линейно независима, если из равенства

$$\alpha_1 \boldsymbol{X}_1 + \alpha_2 \boldsymbol{X}_2 + \cdots + \alpha_k \boldsymbol{X}_k = 0$$

следует, что $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

Определение 4. Любая система линейно независимых векторов $\{\boldsymbol{X}_i\}_{i=1}^{n}$ в n-мерном векторном пространстве называется ***базисом***.

Определение 5. ***Скалярным произведением*** двух векторов называется сумма произведений их компонент (координат), стоящих на одинаковых местах:

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \ (\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} x_i \, y_i.$$

Определение 6. Два вектора называются ***ортогональными***, если их скалярное произведение равно нулю.

Определение 7. Система векторов $\{\boldsymbol{X}_i\}_{i=1}^{n}$ называется ***ортогональным базисом*** в n-мерном векторном пространстве, если все его векторы попарно ортогональны:

$$(\boldsymbol{X}_l, \boldsymbol{X}_m) = \begin{cases} 0, & \text{если } l \neq m, \\ a \neq 0, & \text{если } l = m. \end{cases}$$

Определение 8. Вектор $\boldsymbol{X}$ называется ***нормированным,*** если скалярное произведение $(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{1}$. У таких векторов евклидова норма равна единице:

$$\|\mathbf{X}\|_E = \left(\sum_{n=1}^{N} x_n^2 \right)^{\frac{1}{2}} = 1.$$

Определение 9. Ортогональный базис, все векторы которого имеют единичную норму $\|\mathbf{X}\|_E$, называется ***ортонормированным базисом***.

Свойство 1. Максимальное количество линейно независимых векторов в n-мерном векторном пространстве равно размерности этого пространства.

Свойство 2. Любая система ненулевых ортогональных векторов является линейно независимой системой.

Для собственных векторов и собственных чисел матрицы (4.1) доказаны следующие свойства.

СВОЙСТВА СОБСТВЕННЫХ ВЕКТОРОВ

1. У каждого собственного числа λ существует хотя бы один собственный вектор $\boldsymbol{X}_\lambda$.

2. Если кратность собственного числа λ больше единицы $(k > 1)$, то у этого числа λ есть k линейно независимых собственных векторов

$$\boldsymbol{X}_{\lambda 1}, \boldsymbol{X}_{\lambda 2}, \cdots, \boldsymbol{X}_{\lambda k}.$$

3. Если все собственные числа $\{\lambda_i\}_{i=1}^n$ матрицы (4.1) различны, то все собственные векторы $\{\boldsymbol{X}_{\lambda i}\}_{i=1}^n$ линейно независимы и образуют базис в n-мерном векторном пространстве.

СВОЙСТВА СОБСТВЕННЫХ ЧИСЕЛ

1. Каждому собственному числу соответствует бесчисленное множество собственных векторов, которые отличаются только скалярным множителем, значит эти векторы коллинеарные.

2. Все собственные числа диагональных и верхних треугольных матриц являются диагональными элементами таких матриц. Эти собственные значения – всегда действительные числа.

3. Если λ – собственное число, а $\boldsymbol{X}_\lambda$– собственный вектор матрицы $\boldsymbol{A}$, определитель которой не равен нулю (***обратимая матрица***), тогда число $\frac{1}{\lambda}$ является собственным числом обратной матрицы $\boldsymbol{A}^{-1}$ с тем же собственным вектором $\boldsymbol{X}_\lambda$

4. Пусть λ – собственное число, а $\boldsymbol{X}_\lambda$– собственный вектор матрицы $\boldsymbol{A}$. Тогда при любом постоянном числе σ разность $(\lambda - \sigma)$ будет собственным числом матрицы $\boldsymbol{B} = (\boldsymbol{A} - \sigma \boldsymbol{E})$ с тем же собственным вектором $\boldsymbol{X}_\lambda$.

Это свойство называется ***сдвигом собственных чисел.***

ВАРИАНТЫ ПОСТАНОВКИ ЗАДАЧИ

Возможны два варианта постановки задачи о собственных числах и векторах.

1. Задача о нахождении всех собственных чисел и векторов. Такая постановка называется ***полной проблемой собственных чисел.*** Задача легко (вручную) решается для матриц второго и третьего порядков. Решение строится с помощью прямых методов. В этих методах сначала находятся собственные числа как корни характеристического уравнения (4.7). Далее решается однородная система (4.5). Решением этой системы являются соответствующие собственные векторы. Если порядок матрицы больше трёх, используют другие методы.

2. Задача о нахождении одного или нескольких собственных чисел. Эта постановка называется ***частичной проблемой собственных чисел.*** Решение строится итерационными методами без использования характеристического уравнения.

4.2 Частичная проблема

Для решения задачи используют степенной метод, метод скалярных произведений и другие методы. Остановимся на одном из них.

СТЕПЕННОЙ МЕТОД

Ставится задача о вычислении наибольшего по модулю собственного числа. Предполагается, что модуль такого числа строго больше остальных модулей собственных чисел (***мажорирующее собственное число)***.

В основе степенного метода лежит следующий достаточно простой и красивый итерационный процесс.

1. Предположим, что матрица $\boldsymbol{A}$ n-го порядка имеет n различных собственных значений $\{\lambda_i\}_{i=1}^{n}$. Пусть нумерация собственных чисел и векторов соответствует расположению λ_i по убыванию их модулей:

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$

Возьмём произвольный вектор $\widetilde{\boldsymbol{X}}_0 = \left(\tilde{x}_1^{(0)}\ \tilde{x}_2^{(0)}\ \cdots\ \tilde{x}_n^{(0)}\right)^{\intercal}$ и построим вектор

$$\boldsymbol{Y}_1 = \boldsymbol{A}\widetilde{\boldsymbol{X}}_0, \qquad \boldsymbol{Y}_1 = \left(y_1^{(1)}\ y_2^{(1)}\ \cdots\ y_n^{(1)}\right)^{\intercal}.$$

Нормируем вектор $\boldsymbol{Y}_1$ с помощью числа c_1:

$$c_1 = \max_{1\leq i\leq n}\left|y_i^{(1)}\right|, \quad \widetilde{\boldsymbol{X}}_1 = \frac{1}{c_1}\,\boldsymbol{Y}_1.$$

2. Вновь строим вектор $\boldsymbol{Y}_2$ и проводим такую же нормировку:

$$\boldsymbol{Y}_2 = \boldsymbol{A}\widetilde{\boldsymbol{X}}_1,\ \boldsymbol{Y}_2 = \left(y_1^{(2)}\ y_2^{(2)}\ \cdots\ y_n^{(2)}\right)^{\intercal},\ c_2 = \max_{1\leq i\leq n}\left|y_i^{(2)}\right|,\ \widetilde{\boldsymbol{X}}_2 = \frac{1}{c_2}\,\boldsymbol{Y}_2.$$

……

K. Аналогичным образом строим векторы $\boldsymbol{Y}_k$ и $\widetilde{\boldsymbol{X}}_k$:

$$\boldsymbol{Y}_k = \boldsymbol{A}\widetilde{\boldsymbol{X}}_{k-1},\ \boldsymbol{Y}_k = \left(y_1^{(k)}\ y_2^{(k)}\ \cdots\ y_n^{(k)}\right)^{\intercal},\ c_k = \max_{1\leq i\leq n}\left|y_i^{(k)}\right|,\ \widetilde{\boldsymbol{X}}_k = \frac{1}{c_k}\,\boldsymbol{Y}_k.$$

Докажем, что построенный выше процесс сходится к точному решению задачи, т.е. справедливы сразу два предельных соотношения:

$$\lim_{k\to\infty}\widetilde{\boldsymbol{X}}_k = \boldsymbol{X}_1 \quad \text{и} \quad \lim_{k\to\infty} c_k = \lambda_1, \qquad (4.8)$$

где λ_1 – наибольшее по модулю число матрицы, а $\boldsymbol{X}_1$ – соответствующий собственный вектор матрицы $\boldsymbol{A}$.

СХОДИМОСТЬ СТЕПЕННОГО МЕТОДА

Рассмотрим более подробно шаги описанного выше процесса.

1. Вектор $\widetilde{\boldsymbol{X}}_0$ можно разложить по базису из собственных векторов $\{\boldsymbol{X}_i\}_{i=1}^n$ исходной матрицы:

$$\widetilde{\boldsymbol{X}}_0 = b_1\boldsymbol{X}_1 + b_2\boldsymbol{X}_2 + \cdots + b_n\boldsymbol{X}_n. \qquad (4.9)$$

Предположим, что в равенстве (4.9) все собственные векторы нормализованы, т.е. в каждом собственном векторе максимальное значение его компонент равно единице.

Замечание 2. Всегда можно подобрать такой вектор $\widetilde{\boldsymbol{X}}_0$, для которого $b_1 \neq 0$.

Проследим за цепочкой простейших преобразований

$$\boldsymbol{Y}_1 = \boldsymbol{A}\widetilde{\boldsymbol{X}}_0 \Longrightarrow \boldsymbol{Y}_1 = b_1\boldsymbol{A}\boldsymbol{X}_1 + b_2\boldsymbol{A}\boldsymbol{X}_2 + \cdots + b_n\boldsymbol{A}\boldsymbol{X}_n \Longrightarrow$$

$$\boldsymbol{Y}_1 = b_1\lambda_1\boldsymbol{X}_1 + b_2\lambda_2\boldsymbol{X}_2 + \cdots + b_n\lambda_n\boldsymbol{X}_n \Longrightarrow$$

$$\widetilde{\boldsymbol{X}}_1 = \frac{1}{c_1}\boldsymbol{Y}_1 = \frac{\lambda_1}{c_1}\left(b_1\boldsymbol{X}_1 + b_2\frac{\lambda_2}{\lambda_1}\boldsymbol{X}_2 + \cdots + b_n\frac{\lambda_n}{\lambda_1}\boldsymbol{X}_n\right). \qquad (4.10)$$

2. Умножаем равенство (4.10) на матрицу $\boldsymbol{A}$:

$$\boldsymbol{Y}_2 = \boldsymbol{A}\widetilde{\boldsymbol{X}}_1 \Longrightarrow \frac{\lambda_1}{c_1}\left(b_1\boldsymbol{A}\boldsymbol{X}_1 + b_2\frac{\lambda_2}{\lambda_1}\boldsymbol{A}\boldsymbol{X}_2 + \cdots + b_n\frac{\lambda_n}{\lambda_1}\boldsymbol{A}\boldsymbol{X}_n\right) \Longrightarrow$$

$$\boldsymbol{Y}_2 = \frac{\lambda_1}{c_1}\left(b_1\lambda_1\boldsymbol{X}_1 + b_2\frac{\lambda_2}{\lambda_1}\lambda_2\boldsymbol{X}_2 + \cdots + b_n\frac{\lambda_n}{\lambda_1}\lambda_n\boldsymbol{X}_n\right) \Longrightarrow$$

$$\widetilde{\boldsymbol{X}}_2 = \frac{1}{c_2}\boldsymbol{Y}_2 = \frac{{\lambda_1}^2}{c_1c_2}\left(b_1\boldsymbol{X}_1 + b_2\left(\frac{\lambda_2}{\lambda_1}\right)^2\boldsymbol{X}_2 + \cdots + b_n\left(\frac{\lambda_n}{\lambda_1}\right)^2\boldsymbol{X}_n\right).$$

…….....

K. Очевидно, что вектор $\widetilde{\boldsymbol{X}}_k$ можно сразу представить в виде

$$\widetilde{\boldsymbol{X}}_k = \frac{1}{c_k}\boldsymbol{Y}_k = \frac{{\lambda_1}^k}{c_1c_2\cdots c_k}\left(b_1\boldsymbol{X}_1 + b_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\boldsymbol{X}_2 + \cdots + b_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\boldsymbol{X}_n\right). \quad (4.11)$$

По условию поставленной задачи λ_1 – это максимальное по модулю собственное число матрицы, поэтому

$$\left|\frac{\lambda_j}{\lambda_1}\right| < 1 \text{ для } \forall\, j = 2,\cdots,k \Longrightarrow \lim_{k\to\infty} b_j\left(\frac{\lambda_j}{\lambda_1}\right)^k\boldsymbol{X}_j = \boldsymbol{0} \text{ для } \forall\, j = 2,\cdots,n.$$

Из последнего предельного равенства следует, что

$$\lim_{k\to\infty}\widetilde{\boldsymbol{X}}_k = \lim_{k\to\infty}\frac{b_1{\lambda_1}^k}{c_1c_2\cdots c_k}\boldsymbol{X}_1 = \boldsymbol{V}. \qquad (4.12)$$

Поскольку все векторы $\widetilde{\boldsymbol{X}}_k$ и собственный вектор $\boldsymbol{X}_1$ нормализованы, то предельный вектор $\boldsymbol{V}$ тоже будет нормализованным.

Из определения равенства n-мерных векторов известно, что векторы равны, если равны их компоненты, стоящие на одинаковых местах. Поэтому предел скалярного множителя при векторе $\boldsymbol{X}_1$ существует и равен единице:

$$\lim_{k\to\infty} \frac{b_1 {\lambda_1}^k}{c_1 c_2 \cdots c_k} = 1. \qquad (4.13)$$

Предельное равенство (4.12) теперь можно записать в виде

$$\lim_{k\to\infty} \widetilde{\boldsymbol{X}}_k = \boldsymbol{X}_1. \qquad (4.14)$$

Из предельного равенства (4.14) следует, что последовательность из $\{\widetilde{\boldsymbol{X}}_k\}_{k=1}^{\infty}$ нормализованных векторов сходится к собственному вектору $\boldsymbol{X}_1$.

Для нахождения собственного числа λ_1 используем равенство (4.13) в виде:

$$\lim_{k\to\infty} \frac{b_1 {\lambda_1}^{k-1}}{c_1 c_2 \cdots c_{k-1}} = 1. \qquad (4.15)$$

Разделим равенство (4.13) на (4.15) и после сокращений общих множителей приходим простейшему предельному равенству

$$\lim_{k\to\infty} \frac{\lambda_1}{c_k} = \frac{1}{1} = 1 \Longrightarrow \lambda_1 = \lim_{k\to\infty} c_k.$$

В итоге мы полностью доказали сходимость степенного метода и справедливость предельных соотношений (4.8).

Приведём простейший пример использования степенного метода.

Пример. Найдём наибольшее по модулю собственное число и соответствующий собственный вектор матрицы

$$\boldsymbol{A} = \begin{pmatrix} 0 & 11 & -5 \\ -2 & 17 & -7 \\ -4 & 26 & -10 \end{pmatrix}.$$

Решение.

1. Пусть $\widetilde{\boldsymbol{X}}_0 = (1\ 1\ 1)^{\mathrm{T}}$. Тогда первый шаг итерации имеет вид

$$\begin{pmatrix} 0 & 11 & -5 \\ -2 & 17 & -7 \\ -4 & 26 & -10 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \\ 12 \end{pmatrix} = 12 \begin{pmatrix} \frac{1}{2} \\ \frac{2}{3} \\ 1 \end{pmatrix} = c_1 \widetilde{\boldsymbol{X}}_1 \approx 12 \begin{pmatrix} 0{,}5 \\ 0{,}67 \\ 1 \end{pmatrix}.$$

2. Второй шаг итерации

$$\begin{pmatrix} 0 & 11 & -5 \\ -2 & 17 & -7 \\ -4 & 26 & -10 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{2}{3} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{7}{3} \\ \frac{10}{3} \\ \frac{16}{3} \end{pmatrix} = \frac{16}{3} \begin{pmatrix} \frac{7}{16} \\ \frac{5}{8} \\ 1 \end{pmatrix} = c_2 \, \widetilde{\boldsymbol{X}}_2 \approx 5{,}33 \begin{pmatrix} 0{,}44 \\ 0{,}63 \\ 1 \end{pmatrix}.$$

За восемь шагов итерации приходим к последовательности чисел c_i и векторов $\widetilde{\boldsymbol{X}}_i$, представленных в таблице 4.1.

Таблица 4.1

N	1	2	3	4
c_i	12	5.33	4.50	4.22
$\widetilde{\boldsymbol{X}}_i$	$(0.50 \;\; 0.67 \;\; 1)^{\intercal}$	$(0.44 \;\; 0.63 \;\; 1)^{\intercal}$	$(0.42 \;\; 0.61 \;\; 1)^{\intercal}$	$(0.41 \;\; 0.60 \;\; 1)^{\intercal}$

Таблица 4.1 (окончание)

N	5	6	7	8
c_i	4,11	4,05	4.02	4.01
$\widetilde{\boldsymbol{X}}_i$	$(0.40 \;\; 0.60 \;\; 1)^{\intercal}$	$(0.40 \;\; 0.60 \;\; 1)^{\intercal}$	$(0.40 \;\; 0.60 \;\; 1)^{\intercal}$	$(0.40 \;\; 0.60 \;\; 1)^{\intercal}$

Из таблицы следует, что с точностью до десятых долей собственное число $\lambda = 4{,}0$. Собственный вектор $\boldsymbol{X}_\lambda = (0.4 \;\; 0.6 \;\; 1)^{\intercal}$.

СКОРОСТЬ СХОДИМОСТИ МЕТОДА

У степенного метода линейный порядок сходимости. Из равенства (4.11) следует, что скорость сходимости итерационного процесса пропорциональна величине $\left(\frac{\lambda_j}{\lambda_1}\right)^k$. Чем больше по модулю собственное число λ_1, тем больше скорость сходимости. Для увеличения скорости сходимости используют различные дополнительные методы. Одним из них является достаточно общий метод Эйткена.

МОДИФИКАЦИИ СТЕПЕННОГО МЕТОДА

Перечислим некоторые модификации степенного метода.

1. Метод модифицирован для случая, когда максимальное по модулю собственное число имеет кратность $k > 1$.

2. При совместном использовании этого метода и 4-го свойства собственных чисел (сдвиг собственных чисел**)** находится наименьшее собственное число и соответствующий собственный вектор.

3. В комбинации с другими методами степенной метод можно применять для нахождения всех собственных чисел и векторов матрицы.

4. Степенной метод вместе с различными приёмами уточнения часто используют, когда необходимо построить с большой точностью вектор, собственное число которого приближённо известно.

Различные варианты степенного метода реализованы в стандартных программах математических пакетов.

4.2. Полная проблема

Создано много методов решения этой проблемы для матриц различной структуры. Рассмотрим только один классический метод Якоби на примере симметричных действительных матриц, которые наиболее часто встречаются в прикладных задачах.

В методе используются следующие свойства симметричных матриц.

СВОЙСТВА СИММЕТРИЧНЫХ МАТРИЦ

1. Все собственные числа симметричной матрицы – действительные числа.
2. Если собственные числа симметричной матрицы положительны, тогда матрица положительно определена. Верно обратное утверждение. Если симметричная матрица положительно определена, то её собственные числа всегда положительны.
3. Собственные векторы, соответствующие различным собственным числам симметричной матрицы, ортогональны.
4. Если у симметричной матрицы все собственные числа различны, то все собственные векторы образуют ортогональный базис в n-мерном векторном пространстве.
5. Если $\boldsymbol{A}$ симметричная матрица, то существует такая ортогональная матрица $\boldsymbol{R}$ – ***матрица преобразования,*** которая преобразует исходную матрицу $\boldsymbol{A}$ в диагональную матрицу $\boldsymbol{D}$:

$$\boldsymbol{R}^{-1}\boldsymbol{A}\,\boldsymbol{R} = \boldsymbol{R}^{\mathsf{T}}\,\boldsymbol{A}\,\boldsymbol{R} = \boldsymbol{D}. \qquad (4.16)$$

На главной диагонали такой матрицы $\boldsymbol{D}$ будут стоять все собственные числа исходной матрицы $\boldsymbol{A}$, столбцы матрицы преобразования $\boldsymbol{R}$ станут соответствующими собственными векторами матрицы $\boldsymbol{A}$.

Матрицы $\boldsymbol{A}$ и $\boldsymbol{D}$ в равенстве (4.16) называются ***подобными***. Равенство (4.16) называется ***преобразованием подобия*** или ***вращением матрицы***.

МЕТОД ЯКОБИ

Якоби предложил этот метод в 1846 году. Со временем он превратился в классический ***метод вращений,*** в котором с помощью итерационного процесса строятся с одинаковой степенью точности все собственные числа и соответствующие векторы симметричной матрицы. Метод обладает высокой скоростью сходимости. Его можно использовать для матриц размерности $n \leq 20$. В основе метода лежит 5-е свойство симметричных матриц.

Нахождение собственных значений и собственных векторов матрицы $\boldsymbol{A}$ равносильно построению диагональной матрицы $\boldsymbol{D}$ и ортогональной матрицы вращения $\boldsymbol{R}$, для которых будет выполняться равенство

$$\boldsymbol{D} = \boldsymbol{R}^{\mathsf{T}}\,\boldsymbol{A}\,\boldsymbol{R}.$$

АЛГОРИТМ МЕТОДА ВРАЩЕНИЙ

Алгоритм осуществляет последовательность подобных преобразований матрицы $\boldsymbol{A}$, в результате чего она становится диагональной без изменения спектра. Каждое вращение строится так, чтобы в последующей подобной матрице обнулились два симметричных недиагональных элемента, на месте которых в предыдущей матрице стояли максимальные по модулю элементы.

Основной частью каждого k-го шага этого алгоритма является построение специальной ортогональной матрицы подобия $\boldsymbol{R}_k$ – ***матрицы простых вращений***. Для построения алгоритма удобнее исходную матрицу переименовать: $\boldsymbol{A} = \boldsymbol{A_0}$. Первый шаг итерации начинается с построения первой матрицы простых вращений $\boldsymbol{R}_1$.

ПОСТРОЕНИЕ МАТРИЦЫ ПРОСТЫХ ВРАЩЕНИЙ

1. В исходной матрице $\boldsymbol{A_0}$ находится максимальный по модулю элемент $a_{ij}^{(0)}$. Из-за симметричности матрицы $a_{ij}^{(0)} = a_{ji}^{(0)}$.

2. Выбирается такой угол вращения φ_1, чтобы в подобной матрице $\boldsymbol{A}_1$ равнялись нулю элементы $a_{ij}^{(1)} = a_{ji}^{(1)} = 0$. Матрицы $\boldsymbol{A}_1$ и $\boldsymbol{A_0}$ связывает равенство

$$\boldsymbol{A}_1 = \boldsymbol{R}_1^{\intercal}\, \boldsymbol{A_0}\, \boldsymbol{R}_1. \qquad (4.17)$$

3. Матрица вращений $\boldsymbol{R}_1$ строится с помощью единичной матрицы $\boldsymbol{E}$. Нулевой недиагональный элемент в i-й строке и j-м столбце в $\boldsymbol{E}$ заменятся на $(-\sin\varphi_1)$. Нулевой недиагональный элемент в j-й строке и i-м столбце заменяется на $\sin\varphi_1$. Две диагональные единицы i-й строки и j-го столбца в матрице $\boldsymbol{E}$ заменяются на $\cos\varphi_1$.

Угол φ_1 называется ***углом вращения***. Будем использовать только такие углы вращения, для которых верно неравенство $|\varphi_1| \le \frac{\pi}{4}$.

Матрица простых вращений имеет следующий вид

$$\boldsymbol{R}_1 = \begin{pmatrix} 1 \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 \cdots & \cos\varphi_1 & \cdots & -\sin\varphi_1 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 \cdots & \sin\varphi_1 & \cdots & \cos\varphi_1 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}. \qquad (4.18)$$

4. Значения $\cos\varphi_1$ и $\sin\varphi_1$ находятся с помощью определения равенства матриц. Результат матричного умножения $\boldsymbol{R}_1^{\intercal}\, \boldsymbol{A_0}\, \boldsymbol{R}_1$ – это матрица подобия $\boldsymbol{A_1}$, у которой элемент $a_{ij}^{(1)}$, стоящий в i строке и в j столбце будет равен

$$a_{ij}^{(1)} = (\cos^2\varphi_1 - \sin^2\varphi_1)\ a_{ij}^{(0)} + \cos\varphi_1 \sin\varphi_1 \left(a_{jj}^{(0)} - a_{ii}^{(0)}\right).$$

По условию алгоритма (п. 2) $a_{ij}^{(1)} = a_{ji}^{(1)} = 0$. Из определения равенства матриц вытекает простое тригонометрическое уравнение

$$(\cos^2\varphi_1 - \sin^2\varphi_1)\ a_{ij}^{(0)} + \cos\varphi_1 \sin\varphi_1 \left(a_{jj}^{(0)} - a_{ii}^{(0)}\right) = 0. \quad (4.19)$$

Решение уравнения (4.19) поможет найти значения $\cos\varphi_1$ и $\sin\varphi_1$ в матрице вращений $\boldsymbol{R}_1$. Упростим уравнение с помощью школьных формул:

$$a_{ij}^{(0)}\cos 2\varphi_1 + \frac{\left(a_{jj}^{(0)} - a_{ii}^{(0)}\right)}{2}\sin 2\varphi_1 = 0 \Longrightarrow \tan 2\varphi_1 = \frac{2\ a_{ij}^{(0)}}{\left(a_{ii}^{(0)} - a_{jj}^{(0)}\right)}. \quad (4.20)$$

Выразим $\cos 2\varphi_1$ через $\tan 2\varphi_1$:

$$\cos 2\varphi_1 = \frac{1}{\sqrt{1 + \tan^2 2\varphi_1}}. \quad (4.21)$$

Используем формулы понижения степени:

$$\cos\varphi_1 = \sqrt{\frac{1 + \cos 2\varphi_1}{2}}, \quad \sin\varphi_1 = \pm\sqrt{\frac{1 - \cos 2\varphi_1}{2}}. \quad (4.22)$$

Из п. 3 известно, что $|\varphi_1| \leq \frac{\pi}{4}$, поэтому знак числа $\sin\varphi_1$ будет совпадать со знаком $\tan 2\varphi_1$ и со знаком произведения $a_{ij}^{(0)}\left(a_{ii}^{(0)} - a_{jj}^{(0)}\right)$.

Построение первой матрицы вращения $\boldsymbol{R}_1$ закончено. Затем элементарно строится транспонированная матрица $\boldsymbol{R}_1^{\mathrm{T}}$.

Первый шаг итерации заканчивается нахождением подобной матрицы $\boldsymbol{A_1}$. по формуле (4.17).

Второй шаг алгоритма аналогичен первому. В матрице $\boldsymbol{A_1}$ выбирается максимальный по модулю элемент $a_{ij}^{(1)}$ и запоминается номер строки i и номер столбца j, в котором находится этот элемент. С помощью формул (4.20) – (4.22) вычисляются $\cos\varphi_2$ и $\sin\varphi_2$. Следующая матрица простых вращений $\boldsymbol{R}_2$ строится по такому же принципу как матрица $\boldsymbol{R}_1$. Второй шаг итерации заканчивается построением матрицы $\boldsymbol{A_2}$:

$$\boldsymbol{A_2} = \boldsymbol{R}_2^{\mathrm{T}}\,\boldsymbol{A}_1\,\boldsymbol{R}_2.$$

……

На k-м шаге в матрице $\boldsymbol{A}_{k-1}$ выбирается максимальный по модулю элемент $a_{ij}^{(k-1)}$ и запоминаются его индексы. По формулам (4.20) – (4.22) находятся косинус и синус угла вращения φ_k. Далее строится матрица простых вращений $\boldsymbol{R}_k$ и подобная матрица $\boldsymbol{A}_k$:

$$\boldsymbol{A}_k = \boldsymbol{R}_k^{\mathrm{T}}\, \boldsymbol{A}_{k-1}\, \boldsymbol{R}_k.$$

В результате работы такого алгоритма строятся две последовательности: последовательность подобных матриц

$$\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_2, \ \cdots, \boldsymbol{A}_{k-1}, \ \boldsymbol{A}_k \qquad (4.23)$$

и последовательность матриц простых вращений

$$\boldsymbol{E}, \boldsymbol{T}_1, \boldsymbol{T}_2, \ \cdots, \boldsymbol{T}_{k-1}, \ \boldsymbol{T}_k. \qquad (4.24)$$

Доказано, что последовательность (4.23) сходится к диагональной матрице $\boldsymbol{D}$, на главной диагонали которой стоят собственные числа исходной матрицы $\boldsymbol{A}$:

$$\lim_{k\to\infty} \boldsymbol{A}_k = \boldsymbol{D}.$$

Доказано, что последовательность (4.24) сходится к матрице вращений $\boldsymbol{T}$. Её столбцы – это соответствующие собственные векторы матрицы $\boldsymbol{A}$:

$$\lim_{k\to\infty} \boldsymbol{T}_k = \boldsymbol{T}.$$

Итерационный процесс, лежащий в основе описанного выше алгоритма, сходится. Поэтому всегда наступит такой шаг итерации, начиная с которого модули всех недиагональных элементов матрицы $\boldsymbol{A}_k$ будут меньше заданного малого числа ε. Диагональные элементы матрицы $\boldsymbol{A}_k$ – это приближённые значения собственных чисел матрицы $\boldsymbol{A}$.

Соответствующая матрица вращений $\boldsymbol{T}_k$ является последовательным произведением всех предыдущих матриц простого вращения:

$$\boldsymbol{T}_k = \boldsymbol{T}_1\, \boldsymbol{T}_2 \cdot \cdots \cdot \boldsymbol{T}_{k-2}\, \boldsymbol{T}_{k-1}.$$

Каждый столбец матрицы $\boldsymbol{T}_k$ приближённо равен соответствующему собственному вектору исходной матрицы $\boldsymbol{A}$.

Рассмотрим очень простой пример использования метода вращений.

Пример. Найдём все собственные числа и векторы матрицы

$$\boldsymbol{A} = \begin{pmatrix} 1 & -3 & -1 \\ -3 & 1 & 1 \\ -1 & 1 & 5 \end{pmatrix}.$$

Решение.

1. Пусть $\boldsymbol{A}_0 = \boldsymbol{A}$. Находим максимальный по модулю недиагональный элемент матрицы $\boldsymbol{A}_0$:

$$|a_{12}| = |a_{21}| = 3 \Longrightarrow i = 1, j = 2.$$

Строим матрицу простых вращений

$$\boldsymbol{R}_1 = \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 & 0 \\ \sin\varphi_1 & \cos\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\tan 2\varphi_1 = \frac{2a_{12}}{a_{11} - a_{22}} = \frac{-6}{1-1} = (-\infty) \Longrightarrow 2\varphi = -\frac{\pi}{2} \Longrightarrow \varphi = -\frac{\pi}{4} \Longrightarrow$$

$$\cos\varphi_1 = \frac{\sqrt{2}}{2}, \sin\varphi_1 = -\frac{\sqrt{2}}{2}.$$

С учётом найденных тригонометрических величин угла φ_1 матрица $\boldsymbol{R}_1$ записывается в виде

$$\boldsymbol{R}_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Вычисляем матрицу $\boldsymbol{A}_1$:

$$\boldsymbol{A}_1 = \boldsymbol{R}_1^{\mathrm{T}} \boldsymbol{A}_0 \boldsymbol{R}_1 =$$

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -3 & -1 \\ -3 & 1 & 1 \\ -1 & 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \Longrightarrow$$

$$\boldsymbol{A}_1 = \begin{pmatrix} 4 & 0 & -\sqrt{2} \\ 0 & -2 & 0 \\ -\sqrt{2} & 0 & 5 \end{pmatrix}.$$

2. На втором шаге находим максимальный по модулю недиагональный элемент матрицы $\boldsymbol{A}_1$:

$$|a_{13}| = |a_{31}| = \sqrt{2} \Longrightarrow i = 1, j = 3.$$

Записываем следующую матрицу простых вращений

$$\boldsymbol{R}_2 = \begin{pmatrix} \cos\varphi_2 & 0 & -\sin\varphi_2 \\ 0 & 1 & 0 \\ \sin\varphi_2 & 0 & \cos\varphi_2 \end{pmatrix}.$$

Вычисляем тригонометрические величины по формулам (4.20) – (4.22):

$$\tan 2\varphi_2 = \frac{2a_{13}}{a_{11} - a_{33}} = \frac{-4\sqrt{2}}{4-5} = 2\sqrt{2}, \ \cos\varphi_2 = \sqrt{\frac{2}{3}}, \ \sin\varphi_2 = \sqrt{\frac{1}{3}}.$$

Подставляем найденные значения в матрицу $\boldsymbol{R}_2$:

$$\boldsymbol{R}_2 = \begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & -\sqrt{\frac{1}{3}} \\ 0 & 1 & 0 \\ \sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix}.$$

С помощью матрицы $\boldsymbol{R}_2$ делаем следующее вращение

$$\boldsymbol{A}_2 = \boldsymbol{R}_2^{\mathrm{T}} \boldsymbol{A}_1 \boldsymbol{R}_2 =$$

$$\begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & \sqrt{\frac{1}{3}} \\ 0 & 1 & 0 \\ -\sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix} \cdot \begin{pmatrix} 4 & 0 & -\sqrt{2} \\ 0 & -2 & 0 \\ -\sqrt{2} & 0 & 5 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & -\sqrt{\frac{1}{3}} \\ 0 & 1 & 0 \\ \sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix} \Longrightarrow$$

$$\boldsymbol{A}_2 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

За два шага итерации построена диагональная матрица $\boldsymbol{A}_2$. На её главной диагонали расположены точные значения собственных чисел исходной матрицы $\boldsymbol{A}$. Понадобилось только два простых вращения. Поэтому матрица вращений $\boldsymbol{R}$, преобразовавшая матрицу $\boldsymbol{A}$ в $\boldsymbol{A}_2$, строится как произведение двух матриц простого вращения:

$$\boldsymbol{R} = \boldsymbol{R}_1 \boldsymbol{R}_2 \Longrightarrow$$

$$\boldsymbol{R} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & -\sqrt{\frac{1}{3}} \\ 0 & 1 & 0 \\ \sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{6}} \\ -\sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{6}} \\ \sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix}.$$

Первый столбец матрицы $\boldsymbol{R}$ равен собственному вектору $\boldsymbol{X}_{\lambda=3}$, второй столбец – вектор $\boldsymbol{X}_{\lambda=-2}$, третий столбец – вектор $\boldsymbol{X}_{\lambda=6}$. Полученные точные собственные векторы можно записать в более простом виде:

$$\boldsymbol{X}_{\lambda=3} = \frac{1}{\sqrt{3}}(1 \quad -1 \quad 1)^{\intercal},\ \boldsymbol{X}_{\lambda=-2} = \frac{1}{\sqrt{2}}(1 \quad 1 \quad 0)^{\intercal},\ \boldsymbol{X}_{\lambda=6} = \frac{1}{\sqrt{6}}(-1 \quad 1 \quad 1)^{\intercal}.$$

СКОРОСТЬ СХОДИМОСТИ МЕТОДА

На каждом шаге итерации вместо построенных нулевых элементов могут возникнуть ненулевые. Этот факт мало влияет на скорость сходимости, так как после каждого шага с большой скоростью уменьшается сумма квадратов всех недиагональных элементов.

Если в результате k вращений все недиагональные элементы матрицы $\boldsymbol{A}_k$ отличаются от нуля на величину порядка $O(\varepsilon)$, то её диагональные элементы отличаются от собственных чисел на величину порядка $O(\varepsilon^2)$.

Метод вращений легко реализуется на компьютерах в виде различных стандартных программ с малым количеством арифметических действий. Поэтому такие программы дают хорошую вычислительную точность.

МОДИФИКАЦИЯ МЕТОДА ВРАЩЕНИЙ

1. Метод Хаусхольдера основан на вращениях, но за один шаг итерации в этом методе обнуляются более двух элементов преобразованной матрицы и при последующих вращениях эти нули сохраняются.

QR-МЕТОД (АЛГОРИТМ)

Этот итерационный метод разработали независимо друг от друга российский математик В.Н. Кублановский (1961 г.) и английский математик Дж. Фрэнсис (1962). В настоящее время QR метод считается наиболее надёжным. Он используют для решения полной проблемы собственных чисел. В основе его лежит следующее свойство любой квадратной матрицы.

QR разложение. Любую квадратную матрицу $\boldsymbol{A}$ можно представить в виде произведения двух матриц $\boldsymbol{A} = \boldsymbol{Q} \cdot \boldsymbol{R}$, где $\boldsymbol{Q}$– ортогональная матрица, $\boldsymbol{R}$ – верхняя треугольная матрица.

Глава 5

ПРИБЛИЖЕНИЕ ФУНКЦИЙ

Самое непостижимое в этом мире –
это то, что он постижим.
Альберт Эйнштейн

Приближение функций – это достаточно обширный раздел «Численных методов». Главная задача аппроксимации функций – приближение известной табличной функции непрерывной функцией с заданными свойствами. Методы аппроксимации функций используют и в тех случаях, когда необходимо заменить непрерывную функцию со сложным аналитическим видом на более простую непрерывную функцию.

Построенная непрерывная функция, совпадающая со значениями исходной табличной функции, называется ***интерполяционной функцией***, или ***интерполянтом***. Если построенная функция не совпадает со значениями табличной, но соответствует заданному критерию близости, такую функцию называют ***аппроксимирующей функцией***, или ***аппроксимантом***.

5.1. Интерполирование

Основная задача интерполирования – это построение непрерывной функции $F(x)$, значения которой будут совпадать со значениями табличной $f(x)$:

$$F(x_i) = f(x_i) = y_i, \text{где } i = 0,1,\cdots,N.$$

Функция $F(x)$ – интерполяционная функция.

Поставленная задача имеет бесчисленное множество решений. Для единственности решения надо дополнительно знать свойства исходной табличной функции. Если точного знания нет, можно сделать некоторое предположение, которое называется ***интерполяционной гипотезой***. С помощью дополнительного знания или предположения выбирается класс функций. Затем строится интерполяционная функция, принадлежащая выбранному классу.

Интерполяционные функции чаще всего используют для восстановления значения табличной функции в заданной внутренней точке x^*, которая не совпадает с ***узлами интерполяции***:

$$x^* \in (x_0; x_N),\ x^* \neq x_i, \qquad i = 1,2,\cdots,N-1.$$

Процесс восстановления называется ***интерполяцией функции в точке***. Если точка x^* находится рядом с точками x_0, x_N , но вне отрезка $[x_0;\ x_N]$, нахождение значения функции в такой точке называется ***экстраполяцией функции в точке.*** Погрешность экстраполяции резко растёт с удалением от граничных точек.

Если расстояние между узлами интерполяции – постоянная величина h, все узлы интерполяции можно записать в виде формулы

$$x_i = x_0 + ih, \text{ где } \ i = 0,1,\cdots,N.$$

Число h в этом случае называется ***шагом интерполяции,*** а набор $\{x_i\}_{i=0}^{N}$ называется ***равномерной сеткой***.

Если исходная функция задана на равномерной сетке, то для нахождения значения функции в точке x^* лучше строить интерполянт с учётом только тех узлов сетки, число которых справа и слева от точки x^* одинаково. В таком случае будет более высокая точность интерполяции в точке x^*.

Часто в качестве интерполяционной функции выбираются обобщённые полиномы, алгебраические многочлены или кубические сплайны.

5.1.1. Обобщённые полиномы

Определение 1. Две функции $\varphi_1(x)$ и $\varphi_2(x)$ называются ***линейно-независимыми на*** $[\boldsymbol{a}; \boldsymbol{b}]$, если при любом числе $\gamma \in \mathbb{R}$ верно неравенство

$$\varphi_1(x) \neq \gamma\varphi_2(x) \text{ или } \varphi_2(x) \neq \gamma\varphi_1(x).$$

Определение 2. Бесконечная система функций $\{\varphi_n(x)\}_{n=0}^{\infty}$ называется линейно независимой на отрезке $[a, b]$, если равенство

$$\sum_{n=0}^{\infty} c_n \varphi_n(x) = 0$$

верно только в случае равенства нулю всех числовых коэффициентов $\{c_n\}_{n=0}^{\infty}$. Функции, входящие в эту систему, называются ***базисными,*** или ***координатными функциями***.

Определение 3. Если конечная система функций $\{\varphi_n(x)\}_{n=0}^{M}$ является линейно независимой, тогда выражение

$$\sum_{n=0}^{M} c_n \varphi_n(x) \qquad (5.1)$$

называется ***обобщённым полиномом***.

Теорема 1. Если в табличной функции $\{y_i, x_i\}_{i=0}^{M}$ все узлы интерполяции различны и для заданного набора линейно независимых функций $\{\varphi_n(x)\}_{n=0}^{M}$ определитель $\Delta \neq 0$:

$$\Delta = \begin{vmatrix} \varphi_0(x_0) & \varphi_1(x_0)\cdots & \varphi_M(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1)\cdots & \varphi_M(x_1) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \varphi_0(x_M) & \varphi_1(x_M)\cdots & \varphi_M(x_M) \end{vmatrix} \neq 0, \qquad (5.2)$$

тогда существует единственный обобщённый интерполяционный многочлен, для которого выполняются равенства

$$y_i = \sum_{n=0}^{M} c_n \varphi_n(x_i), \quad 0 \le i \le M. \qquad (5.3)$$

Система функций $\{\varphi_n(x)\}_{n=0}^{M}$, удовлетворяющая условию (5.2), называется ***чебышевской системой функций.***

5.1.2. Алгебраические многочлены

Алгебраические полиномы с действительными коэффициентами хорошо изучены и просто вычисляются. С ними легко совершать арифметические действия, элементарно дифференцировать и интегрировать. Поэтому такие многочлены удобно использовать в качестве интерполяционных функций.

Предположим, что задана табличная функция $\{x_i, y_i\}_{i=0}^{N}$. Построим для этой функции алгебраический интерполяционный полином

$$P_M(x) = \sum_{k=0}^{M} a_k x^k.$$

Полином $P_M(x)$ содержит $(M+1)$ неизвестных. Он должен удовлетворять равенствам:

$$\begin{cases} P_M(x_0) = y_0, \\ P_M(x_1) = y_1, \\ \cdots \\ P_M(x_N) = y_N. \end{cases} \Longrightarrow \begin{cases} a_0 + a_1 x_0 + a_2 x_0^2 + \cdots a_M x_0^M = y_0, \\ a_0 + a_1 x_1 + a_2 x_1^2 + \cdots a_M x_1^M = y_1, \\ \cdots \\ a_0 + a_1 x_N + a_2 x_N^2 + \cdots a_M x_N^M = y_N. \end{cases}$$

Используем эту систему для нахождения коэффициентов $\{a_k\}_{k=0}^{M}$.

Система имеет единственное решение, если её определитель не равен нулю. Следовательно, количество неизвестных должно равняться числу уравнений: $M+1 = N+1$. Обратимся к определителю системы

$$\Delta_V = \begin{vmatrix} 1 & x_0 & x_0^2 \cdots & x_0^N \\ 1 & x_1 & x_1^2 \cdots & x_1^N \\ & & \cdots & \\ 1 & x_N & x_N^2 \cdots & x_N^N \end{vmatrix}.$$

Этот определитель называется ***определителем Вандермонда***. Доказано, что если все узлы $\{x_i\}_{i=0}^{N}$ различны, то определитель Δ_V всегда не равен нулю.

Замечание 1. Определитель Вандермонда является частным случаем определителя (5.2).

Вывод. Для заданной табличной функции с различными ***узлами*** $\{x_i\}_{i=0}^N$ существует единственный интерполяционный алгебраический многочлен $P_N(x)$ степени N.

Есть несколько видов полиномов, представляющих интерполяционный алгебраический многочлен. Рассмотрим два из них: полином Лагранжа $PL_N(x)$ и полином Ньютона $PN_N(x)$. Они построены разными способами, но по сути являются различными формами записи одного и того же многочлена

$$P_N(x) = a_0 + a_1 x + \cdots + a_i x^i + \cdots a_N x^N. \qquad (5.4)$$

Если ставится задача вычисления значения табличной функции в точке $x^* \in (x_0, x_N)$ с помощью полинома $PL_N(x)$ или $PN_N(x)$, то в таком случае не надо приводить эти полиномы к виду (5.4).

5.1.3. Полиномы Лагранжа

Полином Лагранжа строится в виде

$$PL_N(x) = \sum_{i=0}^{N} y_i \, \varphi_{N,i}(x), \qquad (5.5)$$

где функции $\varphi_i(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_N)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_N)}$

называются ***базисными полиномами Лагранжа.*** Они выступают в роли ***символа Кронекера***:

$$\varphi_i(x_k) = \begin{cases} 0, & i \neq k, \\ 1, & i = k. \end{cases}$$

Полином Лагранжа лучше использовать, когда интерполируются несколько разных табличных функций, значения которых заданы на одной сетке $\{x_i\}_{i=0}^N$.

Интерполяционный полином Лагранжа $PL_N(x)$ имеет простой вид, поэтому он часто используется для теоретического анализа. Однако для практических вычислений представление (5.5) используется для небольших значений N, так как базисные полиномы Лагранжа сильно ***осциллируют*** при увеличении степени многочлена.

Если исходная функция $f(x)$ представлена равномерной таблицей с шагом h и является $(n+1)$ раз непрерывно дифференцируемой, тогда погрешность интерполяции полиномом Лагранжа имеет порядок $O(h^{n+1})$. С увеличением количества узлов в таблице погрешность интерполяционного многочлена не всегда уменьшается.

5.1.4. Полиномы Ньютона

Интерполяционный многочлен Ньютона записывается в виде

$$PN_N(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) +$$

$$+a_3(x - x_0)(x - x_1)(x - x_2) + \cdots a_N(x - x_0)(x - x_1)(x-x_2)\cdots(x-x_{N-1}).$$

Полином $PN_N(x)$ строится с помощью рекуррентной формулы

$$PN_N(x) = PN_{N-1}(x) + a_N(x - x_0)(x - x_1)(x - x_2)\cdots(x - x_{N-1}), \qquad (5.6)$$

где $k \le N$.

Поэтому такие полиномы хорошо использовать в тех случаях, когда постепенно увеличивается количество узлов интерполяции. Коэффициенты $\{a_k\}_{k=0}^{N}$ в формуле (5.6) строятся на основе ***разделённых разностей (разностных отношений)***:

$$\Delta_0(x_k) = y_k,$$

$$\Delta_1(x_{k-1}, x_k) = \frac{\Delta_0(x_k) - \Delta_0(x_{k-1})}{x_k - x_{k-1}},$$

$$\Delta_2(x_{k-2}, x_{k-1}, x_k) = \frac{\Delta_1(x_{k-1}, x_k) - \Delta_1(x_{k-2}, x_{k-1})}{x_k - x_{k-2}},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\Delta_j\left(x_{k-j}, x_{k-j+1}, \cdots x_k\right) = \frac{\Delta_{j-1}(x_{k-j+1}, \cdots, x_k) - \Delta_{j-1}\left(x_{k-j}, \cdots, x_{k-1}\right)}{x_k - x_{k-j}}.$$

Посмотрим на табличном примере порядок вычисления разделённых разностей.

Таблица 5.1

x_k	$\Delta_0(x_k)$	$\Delta_1(\cdots\cdots)$	$\Delta_2(\cdots\cdots)$	$\Delta_3(\cdots\cdots)$	$\Delta_4(\cdots\cdots)$
x_0	y_0	----------	----------	----------	----------
x_1	y_1	$\Delta_1(x_0, x_1)$	----------	----------	----------
x_2	y_2	$\Delta_1(x_1, x_2)$	$\Delta_2(x_0, x_1, x_2)$	----------	----------
x_3	y_3	$\Delta_1(x_2, x_3)$	$\Delta_2(x_1, x_2, x_3)$	$\Delta_3(x_0, x_1, x_2, x_3)$	----------
x_4	y_4	$\Delta_1(x_3, x_4)$	$\Delta_2(x_2, x_3, x_4)$	$\Delta_3(x_1, x_2, x_3, x_4)$	$\Delta_4(x_0, x_1, x_2, x_3, x_4)$

Все коэффициенты $\{a_k\}_{k=0}^{N}$ строятся по формуле

$$a_k = \Delta_k(x_0, x_1, \cdots, x_k).$$

Если исходная функция $f(x)$ представлена равномерной таблицей с шагом h и является $(n+1)$ раз непрерывно дифференцируемой, тогда погрешность интерполяции полиномом Ньютона имеет порядок $O(h^{n+1})$.

5.1.5. Кубические интерполяционные сплайны

Интерполяционные полиномы, построенные по таблицам большого объёма, имеют большую степень. Полиномы большой степени сильно осциллируют, поэтому их использование может привести к ошибочным результатам.

Иной подход к построению интерполяционной функции связан с ***кусочно-полиномиальными*** функциями. К этому виду функций относятся кубические сплайны, которые были наиболее популярны во второй половине XX-го века.

К настоящему времени они хорошо изучены. Созданы более точные численные алгоритмы и программы их построения. Кубические сплайны стали классическими интерполяционными функциями со своими достоинствами и недостатками.

ПОСТРОЕНИЕ КУБИЧЕСКОГО СПЛАЙНА

Допустим, что некоторый физический процесс теоретически может быть описан дважды непрерывно дифференцируемой функцией $f(x)$, но фактически представлен в виде таблицы наблюдений $\{x_i, y_i\}_{i=0}^{N}$. Тогда в качестве интерполяционной функции лучше использовать кубический сплайн.

Если объём таблицы равен $(N+1)$, тогда кубический сплайн состоит из N звеньев. Каждое звено – это кубический полином, который включает четыре коэффициента и один узел x_{i-1} из исходной таблицы:

$$z_i = a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3, \quad x_{i-1} \le x < x_i.$$

Кубический интерполяционный сплайн аналитически представляется как сумма звеньев:

$$S_3(x) = \sum_{i=1}^{N} (a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3), \quad (5.7)$$

где $x_{i-1} \le x < x_i$.

Для построения $S_3(x)$ найдём $4N$ неизвестных величин $\{a_i; b_i; c_i; d_i\}_{i=1}^{N}$. Они являются решением системы, которая строится благодаря следующим свойствам кубического сплайна.

1. $S_3(x)$ – интерполяционный сплайн, поэтому его значения совпадают с табличными значениями:

$$S_3(x_{i-1}) = y_{i-1}, \qquad i = 1, \cdots, N+1 \Longrightarrow$$

$$a_i + b_i(x_{i-1} - x_{i-1}) + c_i(x_{i-1} - x_{i-1})^2 + d_i(x_{i-1} - x_{i-1})^3 = y_{i-1} \Rightarrow$$

$$a_i = y_{i-1} \quad \text{для } i = 1, \cdots, N. \quad (5.8)$$

Будем предполагать, для простоты, что шаг таблицы постоянный:

$$h = x_i - x_{i-1}, \quad i = 1, 2, \cdots, N.$$

Последнее звено сплайна в последнем узле x_N должно равняться y_N:

$$a_N + b_N h + c_N h^2 + d_N h^3 = y_N. \quad (5.9)$$

Первое свойство дало $(N+1)$ уравнение.

2. $S_3(x)$ – непрерывная функция, поэтому значения соседних звеньев должны совпадать во всех внутренних узлах таблицы:

$$a_i + b_i(x_i - x_{i-1}) + c_i(x_i - x_{i-1})^2 + d_i(x_i - x_{i-1})^3 =$$

$$= a_{i+1} + b_{i+1}(x_i - x_i) + c_{i+1}(x_i - x_i)^2 + d_{i+1}(x_i - x_i)^3 \Rightarrow$$

$$\Rightarrow a_i + b_i h + c_i h^2 + d_i h^3 = a_{i+1}, \quad i = 1, 2 \cdots, N-1. \qquad (5.10)$$

Второе свойство дало $(N-1)$ уравнение.

3. $S_3(x)$ имеет непрерывную первую производную, поэтому значения первых производных соседних звеньев сплайна должны совпадать во всех внутренних узлах таблицы. Дифференцируем два соседних звена сплайна и приравняем значения производных в общем узле:

$$b_i + 2c_i(x_i - x_{i-1}) + 3d_i(x_i - x_{i-1})^2 =$$

$$b_{i+1} + 2c_{i+1}(x_i - x_i) + 3d_{i+1}(x_i - x_i)^2 \Rightarrow$$

$$b_i + 2c_i h + 3d_i h^2 = b_{i+1}, \quad i = 1, \cdots, N-1. \qquad (5.11)$$

Третье свойство дало $(N-1)$ уравнение.

4. $S_3(x)$ имеет непрерывную вторую производную, поэтому значения вторых производных соседних звеньев сплайна должны совпадать во всех внутренних узлах таблицы. Берём вторую производную от двух соседних звеньев и приравняем их значения в общем узле:

$$2c_i + 6d_i(x_i - x_{i-1}) = 2c_{i+1} + 6d_{i+1}(x_i - x_i) \Rightarrow 2c_i + 6d_i h = 2c_{i+1} \Rightarrow$$

$$\Rightarrow \quad c_i + 3d_i h = c_{i+1}\,,\ i = 1, \cdots, N-1. \qquad (5.12)$$

Четвёртое свойство дало $(N-1)$ уравнение.

Замечание 2. Сплайн $S_3(x)$ состоит из кубических полиномов, поэтому он дважды непрерывно дифференцируем для $\forall x \in [x_0; x_N]$.

Построена система из $(4N-2)$-х уравнений, в которую входят $4N$ неизвестных. Для замкнутости системы необходимо ещё два уравнения. Их строят из дополнительных граничных условий. Они могут быть различными. Приведём три примера.

1. Естественные граничные условия:

$$S_3''(x_0) = S_3''(x_N) = 0.$$

2. Заданы значения вторых производных на границах:

$$S_3''(x_0) = m_0, \qquad S_3''(x_N) = m_N.$$

3. Вторая производная постоянна для первого и последнего звеньев:

$$S_3''(x) = m_0, \quad x_0 \le x < x_1; \qquad S_3''(x) = m_1, \quad x_{N-1} \le x < x_N.$$

Для каждого из этих трёх случаев доказано, что существует единственный сплайн.

Система уравнений $(5.8) - (5.12)$ имеет следующий вид

$$\begin{cases} a_i = y_{i-1}, \\ a_i + b_i h + c_i h^2 + d_i h^3 = a_{i+1}, \\ b_i + 2c_i h + 3d_i h^2 = b_{i+1}, \\ c_i + 3d_i h = c_{i+1}, \\ a_N = y_N - (b_N h + c_N h^2 + d_N h^3). \end{cases} \qquad (5.13)$$

На основе этой системы построим новую систему, в которую войдут только неизвестные $\{c_i\}_{i=1}^{N}$. Для этого проведём с каждой группой уравнений простые преобразования.

1. Из группы уравнений (5.12) выразим d_i через c_i и c_{i+1}:

$$d_i = \frac{c_{i+1} - c_i}{3h}. \qquad (5.14)$$

2. Подставим полученное выражение в группу уравнений системы (5.10) и заменим в этой группе a_i на y_{i-1}, a_{i+1} на y_i:

$$y_{i-1} + b_i h + c_i h^2 + \frac{c_{i+1} - c_i}{3h} h^3 = y_i.$$

3. Из полученного равенства выразим b_i через c_i и c_{i+1}:

$$b_i = -\frac{2}{3} h c_i - \frac{1}{3} h c_{i+1} + \frac{y_i - y_{i-1}}{h}. \qquad (5.15)$$

С помощью формулы (5.15) выразим b_{i+1} через c_{i+1} и c_{i+2}:

$$b_{i+1} = -\frac{2}{3} h c_{i+1} - \frac{1}{3} h c_{i+2} + \frac{y_{i+1} - y_i}{h}. \qquad (5.16)$$

4. В группу уравнений (5.11) подставляем выражения (5.14) – (5.16), полученные для d_i, b_i, b_{i+1}:

$$-\frac{2}{3} h c_i - \frac{1}{3} h c_{i+1} + \frac{y_i - y_{i-1}}{h} + 2c_i h + 3\frac{c_{i+1} - c_i}{3h} h^2 = -\frac{2}{3} h c_{i+1} - \frac{1}{3} h c_{i+2} + \frac{y_{i+1} - y_i}{h}.$$

5. Приведём подобные члены и запишем эти уравнения в каноническом виде:

$$\left(-\tfrac{2}{3} + 2 - 1\right) h c_i + \left(-\tfrac{1}{3} + 1 + \tfrac{2}{3}\right) h c_{i+1} + \tfrac{1}{3} h c_{i+2} = \frac{y_{i-1} - 2y_i + y_{i+1}}{h} \Rightarrow$$

$$\Rightarrow c_i + 4c_{i+1} + c_{i+2} = 3\frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}, \qquad i = 1, \cdots, N-2. \qquad (5.17)$$

Для замкнутости системы (5.17) используем естественные граничные условия:

$$S_3''(x_0) = 2c_1 + 6d_1(x_0 - x_0) = 0 \Rightarrow c_1 = 0,$$

$$S_3''(x_N) = 2c_N + 6d_N h = 0.$$

Предположим, что в последнем равенстве $c_N = 0$. По формуле (5.14)

$$d_N = \frac{c_{N+1} - c_N}{3h}.$$

Коэффициент c_{N+1} не используется в построении кубического сплайна, поэтому его можно приравнять нулю. В итоге можно полагать, что если

$$S_3''(x_N) = 0, \text{то } c_N = 0 \Longrightarrow d_N = 0.$$

Построена линейная система для нахождения коэффициентов $\{c_i\}_{i=2}^{N-1}$:

$$\boldsymbol{AC} = \boldsymbol{f}, \quad \text{где} \tag{5.18}$$

$$\boldsymbol{A} = \begin{pmatrix} 410000000\cdots\cdots\cdots\cdots\cdots 0 \\ 141000000\cdots\cdots\cdots\cdots\cdots 0 \\ 014100000\cdots\cdots\cdots\cdots\cdots 0 \\ 001410000\cdots\cdots\cdots\cdots\cdots 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 000\cdots\cdots\cdots\cdots\cdots\cdots 1410 \\ 0000\cdots\cdots\cdots\cdots\cdots\cdots 141 \\ 000000\cdots\cdots\cdots\cdots\cdots 14 \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} c_2 \\ c_3 \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ c_{N-2} \\ c_{N-1} \end{pmatrix},$$

$$\boldsymbol{f} = {}^{3}/_{h^2} \begin{pmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ y_{N-3} - 2y_{N-2} + y_{N-1} \\ y_{N-2} - 2y_{N-1} + y_N \end{pmatrix}.$$

Матрица $\boldsymbol{A}$ трёхдиагональна с доминантной главной диагональю. Система (5.18) решается методом прогонки (глава 3). После нахождения всех коэффициентов $\{c_i\}_{i=1}^{N}$ вычисляются $\{a_i; b_i; d_i\}_{i=1}^{N}$.

1. Коэффициенты $\{d_i\}_{i=1}^{N-1}$ находим по формуле (5.14). Напомним, что при естественных граничных условиях коэффициент $d_N = 0$.

2. Коэффициенты $\{b_i\}_{i=1}^{N-1}$ вычисляются по формуле (5.15). Последний коэффициент b_N вычисляется проще остальных:

$$b_N = -\frac{2}{3}hc_N - \frac{1}{3}hc_{N+1} + \frac{y_N - y_{N-1}}{h} \Longrightarrow \quad b_N = \frac{y_N - y_{N-1}}{h}.$$

3. Коэффициенты $\{a_i\}_{i=1}^{N}$ известны из равенства (5.8):

$$a_i = y_{i-1}.$$

Найдено всё семейство коэффициентов $\{a_i; b_i; c_i; d_i\}_{i=1}^{N}$, которое определяет единственный кубический интерполяционный сплайн $S_3(x)$ для заданной табличной функции $\{x_i, y_i\}_{i=0}^{N}$.

Построение кубического сплайна завершено.

СВОЙСТВА КУБИЧЕСКОГО СПЛАЙНА

1. Среди всех дважды непрерывно дифференцируемых на $[a; b]$ функций, интерполирующих табличную функцию $\{x_i, y_i\}_{i=0}^{N}$, кубический сплайн $S_3(x)$ меньше всего осциллирует.

2. Свойство минимизации (минимум осреднённой кривизны).

Предположим, что $f(x)$ – дважды непрерывно дифференцируемая на $[a; b]$ функция. Но она представлена в табличном виде $\{x_i, y_i\}_{i=0}^{N}$. Пусть для этой функции построен кубический интерполяционный сплайн $S_3(x)$ с граничными условиями

$$S_3'(a) = f'(a), \qquad S_3'(b) = f'(b),$$

тогда справедливо неравенство

$$\int_a^b \left(S_3''(x)\right)^2 dx \le \int_a^b \left(f''(x)\right)^2 dx. \qquad (5.19)$$

Смысл неравенства (5.19) очень прост. Вторая производная определяет кривизну графика функции. Поэтому интерполяционный кубический сплайн обеспечивает минимум осреднённой кривизны.

ОЦЕНКА ПОГРЕШНОСТИ СПЛАЙНА

Если таблица исходной функции задана с постоянным достаточно малым шагом h $(|h| \ll 1)$, тогда порядок точности интерполирования составляет $O(h^4)$. При неравномерном шаге таблицы точность обычно выше. Точность восстановления первой производной на порядок хуже точности приближения исходной функции с помощью того же сплайна. Точность восстановления второй производной на два порядка хуже точности приближения исходной функции тем же сплайном..

Если табличная функция не является дважды непрерывно дифференцируемой, тогда кубический сплайн не гарантирует хорошей точности.

Погрешность интерполяции функции в точке $x^* \in (x_0; x_N)$ увеличивается с приближением точки x^* к граничным точкам интервала x_0, x_N. Поэтому кубические сплайны не эффективны для экстраполирования функций.

ПРИМЕНЕНИЕ СПЛАЙНА

Кубические интерполяционные сплайны используют, прежде всего, для приближённого интегрирования, дифференцирования, для интерполяции значений функции вдали от границ интервала $(x_0; x_N)$. Сплайны успешно применяются для приближённого решения уравнений математической физики эллиптического типа. Если дифференцируемая функция представлена таблицей большого объёма, для такой функции удобно строить кубический интерполяционный сплайн.

5.1.6. Особенности интерполяционных функций.

Интерполяционные функции используют для нахождения значений табличной функции $f(x)$ в точке x^*, если она находится между узлами таблицы:

$$x_{i-1} < x^* < x_i.$$

Более точный результат можно получить, если интерполяционную функцию строить только по 4 – 6 узлам, окружающим значение аргумента x^*, в котором надо восстановить отсутствующее значение табличной функции.

Интерполяционные функции можно использовать для интегрирования табличной функции $f(x)$ по области $[x_0, x_N]$ и дифференцирования внутри этой области.

Предположим, что для заданной табличной функции мы уже построили некоторую интерполяционную функцию, а после этого добавили ещё один узел и соответствующее значение табличной функции. В таком случае многие интерполяционные функции придётся строить заново. Эта особенность является существенным недостатком интерполяционных функций.

5.2. Аппроксимация

Рассмотрим табличную функцию, в которой значения функции $\{y_i\}_{i=1}^N$ получены из эксперимента или наблюдений. Эти данные всегда содержат погрешность, поэтому нет смысла использовать интерполяционные функции $F(x)$:

$$F(x_i) = y_i\ , 1 < i < N. \qquad (5.20)$$

В таких случаях используются другие методы приближения. От функции $F(x)$ не требуются выполнения условия (5.20). $F(x)$ теперь будем называть аппроксимирующей функцией, или аппроксимантом.

Методы получения аппроксимирующей функции называются коротко ***аппроксимацией.*** Наиболее распространён метод наименьших квадратов и аппроксимация с помощью равномерного приближения ортогональными, почти ортогональными или ортогональными с весом функциями.

5.2.1. Метод наименьших квадратов

ПОСТАНОВКА ЗАДАЧИ

Задана некоторая табличная функция $f(x)$: $\{x_i, y_i\}_{i=0}^{N}$. Ставится задача о построении такой непрерывной функции $F(x)$, чтобы сумма квадратов отклонений от этой функции до заданных значений табличной функции была минимальной:

$$\sum_{i=1}^{N}(F(x_i) - y_i)^2 = min,$$

где величина $(F(x_i) - y_i)$ – ***отклонение***.

В такой постановке задача имеет бесконечное множество решений. Но, если заранее определить класс функций, к которому должна принадлежать $F(x)$, тогда можно найти единственное решение. Рассмотрим простейший случай.

ЛИНЕЙНАЯ АППРОКСИМАЦИЯ

Пусть задана табличная функция $\{x_i, y_i\}_{i=1}^{N}$, у которой все абсциссы разные. Найдём такую линейную функцию $F(x) = kx + b$, чтобы

$$\sum_{i=1}^{N}(kx_i + b - y_i)^2 = \Phi = min. \qquad (5.21)$$

В равенстве (5.21) неизвестны только два параметра k и b. Величину Φ можно рассматривать как функцию, зависящую от двух переменных:

$$\Phi = \Phi(k, b).$$

Это степенная функция, поэтому она непрерывно дифференцируема. Для неё справедливо необходимое условие экстремума:

$$\begin{cases} \dfrac{\partial \Phi(k,b)}{\partial k} = 0, \\ \dfrac{\partial \Phi(k,b)}{\partial b} = 0. \end{cases} \Rightarrow \begin{cases} \displaystyle\sum_{i=1}^{N} 2\,(kx_i + b - y_i)x_i = 0, \\ \displaystyle\sum_{i=1}^{N} 2\,(kx_i + b - y_i) = 0. \end{cases}$$

Приведём полученную систему к каноническому виду:

$$\begin{cases} k\displaystyle\sum_{i=1}^{N} {x_i}^2 + b\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i, \\ k\displaystyle\sum_{i=1}^{N} x_i + bN = \sum_{i=1}^{N} y_i\,. \end{cases}$$

Построенная система является нормальной системой. Доказано, что определитель этой системы Δ всегда не равен нулю. Следовательно, решение $\{k_0, b_0\}$ системы существует, единственно, его можно найти по формулам Крамера:

$$k = k_0 = \frac{\Delta_1}{\Delta}, \quad b = b_0 = \frac{\Delta_2}{\Delta},$$

$$\text{где } \Delta = \begin{vmatrix} \sum_{i=1}^{N} x_i{}^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & N \end{vmatrix},$$

$$\Delta_1 = \begin{vmatrix} \sum_{i=1}^{N} x_i y_i, & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} y_i, & N \end{vmatrix}, \qquad \Delta_2 = \begin{vmatrix} \sum_{i=1}^{N} x_i{}^2 & \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} y_i \end{vmatrix}.$$

В итоге построена аппроксимирующая функция

$$F(x) = k_0 x + b_0.$$

Замечание 3. Очевидно, что $\Phi(k, b)$ – квадратичная функция. Её старшие коэффициенты положительны. Геометрический образ функции $\Phi(k, b)$ – параболоид, поверхность которого бесконечно простирается вверх. В его вершине $M(k_0, b_0)$ функция $\Phi(k, b)$ достигает наименьшего значения. Поэтому достаточно выполнения только необходимого условия существования экстремальной точки.

ПОКАЗАТЕЛЬНАЯ АППРОКСИМАЦИЯ

Пусть задана табличная функция $\{x_i, y_i\}_{i=1}^{N}$ с различными абсциссами и положительными ординатами. Задача состоит в построении показательной функции

$$F(x) = e^{kx+b}$$

по методу наименьших квадратов.

Прологарифмируем значения исходной функции: $\ln y_i = z_i$ и введём функцию

$$Z(x) = \ln F(x) = \ln e^{kx+b} \Longrightarrow Z(x) = kx + b.$$

Теперь наша задача превратилась в задачу линейной аппроксимации:

$$\Phi(k, b) = \sum_{i=1}^{N} (kx_i + b - z_i)^2 = min.$$

Решаем систему

$$\begin{cases} k\sum_{i=1}^{N} x_i{}^2 + b\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i z_i, \\ k\sum_{i=1}^{N} x_i + bN = \sum_{i=1}^{N} z_i\,. \end{cases}$$

Вычисленные по формулам Крамера коэффициенты k_0 и b_0 приводят к искомой аппроксимирующей показательной функции

$$F(x) = e^{k_0 x + b_0}.$$

ЛОГАРИФМИЧЕСКАЯ АППРОКСИМАЦИЯ

Задана табличная функция $\{x_i, y_i\}_{i=1}^{N}$ с различными положительными абсциссами. Аппроксимируем её логарифмической функцией

$$F(x) = \ln(\,kx + b)\ \,,(\,kx + b) > 0$$

по методу наименьших квадратов.
Преобразуем значения табличной функции: $ln\, y_i = t_i$. Введём функцию

$$T(x) = e^{ln(\,kx+b)} \Rightarrow T(x) = kx + b.$$

Теперь наша задача превратилась в задачу линейной аппроксимации:

$$\Phi(k,b) = \sum_{i=1}^{N} (kx_i + b - t_i)^2 = min.$$

Решаем систему

$$\begin{cases} k\sum_{i=1}^{N} x_i{}^2 + b\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i t_i, \\ k\sum_{i=1}^{N} x_i + bN = \sum_{i=1}^{N} t_i\,. \end{cases}$$

Вычисляем по формулам Крамера коэффициенты k_0 и b_0 и приходим к аппроксимационной логарифмической функции

$$F(x) = ln(\,k_0 x + b_0).$$

Рассмотренные выше три вида аппроксимации используют только для монотонных функций. Очевидно, что для немонотонных функций следует использовать только немонотонные аппроксимирующие функции.

АППРОКСИМАЦИЯ КВАДРАТНЫМ ПОЛИНОМОМ

Задана немонотонная табличная функция $\{x_i, y_i\}_{i=1}^{N}$ с различными абсциссами. Аппроксимируем эту таблицу квадратичной функцией

$$F(x) = ax^2 + bx + c.$$

Неизвестные коэффициенты находим по методу наименьших квадратов:

$$\sum_{i=1}^{N} (ax_i{}^2 + bx_i + c - y_i)^2 = \Phi = min.$$

Параметры $\{a, b, c\}$. неизвестны. Величину Φ можно рассматривать как степенную функцию трёх переменных: $\Phi = \Phi(a, b, \mathrm{c})$. Для неё справедливо необходимое условие экстремума:

$$\begin{cases} \dfrac{\partial \Phi}{\partial a} = 0, \\ \dfrac{\partial \Phi}{\partial b} = 0, \\ \dfrac{\partial \Phi}{\partial c} = 0, \end{cases} \Rightarrow \begin{cases} \displaystyle\sum_{i=1}^{N} 2\,(ax_i{}^2 + bx_i + c - y_i) \cdot x_i{}^2 = 0, \\ \displaystyle\sum_{i=1}^{N} 2\,(ax_i{}^2 + bx_i + c - y_i) \cdot x_i = 0, \\ \displaystyle\sum_{i=1}^{N} 2\,(ax_i{}^2 + bx_i + c - y_i) \cdot 1 = 0. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} a\displaystyle\sum_{i=1}^{N} x_i{}^4 + b\sum_{i=1}^{N} x_i{}^3 + c\sum_{i=1}^{N} x_i{}^2 = \sum_{i=1}^{N} y_i x_i{}^2, \\ a\displaystyle\sum_{i=1}^{N} x_i{}^3 + b\sum_{i=1}^{N} x_i{}^2 + c\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i x_i, \\ a\displaystyle\sum_{i=1}^{N} x_i{}^2 + b\sum_{i=1}^{N} x_i + c \cdot N = \sum_{i=1}^{N} y_i\,. \end{cases} \qquad (5.22)$$

Система (5.22) является нормальной, поэтому у неё существует единственное решение, его можно вычислить по формулам Крамера. Нахождение коэффициентов $\{a_0,\ b_0,\ c_0\}$ завершает построение аппроксиманта

$$F(x) = a_0 x^2 + b_0 x + c_0.$$

ОЦЕНКА ТОЧНОСТИ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Предположим, что для заданной табличной функции $\{x_i, y_i\}_{i=1}^{N}$ уже построена аппроксимирующая функция $F(x)$. Для определения качества аппроксима-

ции построенной функции можно использовать различные формулы оценки погрешности. Ниже представлены три основных оценки.

1. ***Максимальная погрешность***

$$E_\infty(F) = \max_{1\le i\le N} |F(x_i) - y_i|.$$

2. ***Средняя погрешность***

$$E_1(F) = \frac{1}{N}\sum_{i=1}^{N} |F(x_i) - y_i|.$$

3. ***Среднеквадратичная погрешность***

$$E_2(F) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (F(x_i) - y_i)^2}.$$

Чем ближе график аппроксимирующей функции расположен к соответствующим точкам y_i табличной функции, тем меньше погрешность**.** Аппроксимация функций по методу наименьших квадратов даёт малую среднеквадратическую погрешность, но при этом возможны большие погрешности в некоторых узлах сетки.

5.2.2. Аппроксимация ортогональными функциями

Определение 4. Функция $\varphi_n(x)$ называется ***интегрируемой с квадратом в области*** $[\boldsymbol{a}; \boldsymbol{b}]$, если интеграл от квадрата этой функции по области $[a; b]$ равен положительному числу I_n:

$$I_n = \int_a^b \varphi_n^2(x)\, dx \neq 0. \qquad (5.23)$$

ОРТОГОНАЛЬНЫЕ ФУНКЦИИ

Бесконечная (или конечная) система интегрируемых с квадратом функций $\{\varphi_k(x)\}_{k=1}^{+\infty}$ называется ***ортогональной с весом в области*** $[\boldsymbol{a}; \boldsymbol{b}]$**,** если для этих функций выполняется равенство

$$\int_a^b \rho(x)\varphi_n(x)\varphi_m(x)dx = \begin{cases} 0, \text{если } n \neq m, \\ \alpha_n > 0, \text{если } n = m. \end{cases} \qquad (5.24)$$

Функция $\rho(x)$ в подынтегральном выражении (5.24) называется ***весовой функцией (весом)***. При отсутствии весовой функции ($\rho(x) = 1$) система интегрируемых с квадратом функций называется ***ортогональной в области*** $[\boldsymbol{a}; \boldsymbol{b}]$, если для этих функций выполняется равенство

$$\int_a^b \varphi_n(x)\varphi_m(x)dx = \begin{cases} 0, \text{если } n \neq m, \\ \alpha_n > 0, \text{если } n = m. \end{cases} \qquad (5.25)$$

Система ортогональных функций является линейно независимой системой, а система линейно независимых функций в общем случае может быть не ортогональной. Но любую линейно независимую систему функций можно преобразовать в ортогональную систему.

Определённый интеграл от произведения двух функций часто называется **ска*лярным произведением функций*** и записывается в виде $\big(\varphi_n(x), \varphi_m(x)\big)$:

$$\big(\varphi_n(x), \varphi_m(x)\big) = \int_a^b \varphi_n(x)\varphi_m(x)dx.$$

РАВНОМЕРНАЯ СХОДИМОСТЬ

Рассмотрим некоторый, сходящийся в области $x \in [a, b]$, функциональный ряд

$$\sum_{n=1}^{\infty} \mathrm{c}_n\varphi_n(x) = \sum_{n=1}^{M} \mathrm{c}_n\varphi_n(x) + R_M, \qquad (5.26)$$

где $\sum_{n=1}^{M} \mathrm{c}_n\varphi_n(x) = S_M(x)$ – частичная сумма ряда, c_n – числовые коэффициенты, R_M – остаток ряда.

Для любых сходящихся рядов выполняется предельное соотношение

$$\lim_{M\to+\infty} R_M = 0.$$

Замечание 4. Выражение

$$\sum_{n=1}^{M} \mathrm{c}_n\varphi_n(x)$$

обычно называется ***линейной комбинацией*** функций $\varphi_n(x)$.

Определение 5. Функциональный ряд (5.26) называется ***равномерно сходящимся*** на $[a, b]$, если для $\forall\ \varepsilon > 0$ найдётся такое натуральное число N, что при $\forall$ натуральном числе $M > N$ и $\forall\ x \in [a, b]$ будет верно неравенство

$$|S(x) - S_M(x)| < \varepsilon,$$

где $S(x)$ – сумма сходящегося ряда (5.26).

Уточним смысл этого определения. Ряд равномерно сходится в заданной области, если существует такой номер N, начиная с которого модуль разности между суммой ряда и частичной суммой $S_M(x)$ становится меньше любого заданного числа $\varepsilon > 0$ для любого x из заданной области $[a; b]$.

В рамках математического анализа в теории функциональных рядов доказана следующая теорема.

Теорема 2. Любая непрерывная на $[a;b]$ функция $f(x)$ может быть единственным образом разложена в равномерно сходящийся ряд по заданной ортогональной на $[a,b]$ системе функций $\{\varphi_n(x)\}_{n=1}^{+\infty}$:

$$f(x)=\sum_{n=1}^{\infty}c_n\varphi_n(x)=\sum_{n=1}^{M}c_n\varphi_n(x)+R_M. \qquad (5.27)$$

Ряды, построенные по ортогональной системе функций, называются ***рядами Фурье.***

СВОЙСТВА РАВНОМЕРНО СХОДЯЩИХСЯ РЯДОВ

Доказано, что для равномерно сходящихся рядов справедливы следующие свойства.

1. Если ряд равномерно сходится на $[a;b]$, то его сумма $S(x)$ – непрерывная функция на этом отрезке.

2. Если ряд равномерно сходится на $[a;b]$, то этот ряд можно почленно интегрировать на этом отрезке:

$$\int_a^b\sum_{n=1}^{\infty}c_n\varphi_n(x)dx=\sum_{n=1}^{\infty}c_n\int_a^b\varphi_n(x)\,dx.$$

3. Если ряд равномерно сходится на $[a;b]$, то этот ряд можно почленно дифференцировать на этом отрезке:

$$\left(\sum_{n=1}^{\infty}c_n\varphi_n(x)\right)'=\sum_{n=1}^{\infty}c_n\varphi_n'(x).$$

4. Если ряд равномерно сходится, то для любого $\varepsilon>0$ $\exists$ такое натуральное число N, что при $\forall$ натуральном $M>N$, $\forall$ натуральном $k>0$ и для всех $x\in[a,b]$ будет верно неравенство

$$\left|\sum_{n=1}^{M}c_n\varphi_n(x)-\sum_{n=1}^{M+k}c_n\varphi_n(x)\right|<\varepsilon. \qquad (5.28)$$

Из 4-го свойства следует, что любую непрерывную на $[a;b]$ функцию можно сколь угодно точно аппроксимировать частичными суммами равномерно сходящегося ряда.

КОЭФФИЦИЕНТЫ ФУРЬЕ

Ниже представлен очень простой и красивый вывод формулы для вычисления коэффициентов $\{c_n\}_{n=1}^{+\infty}$ в ортогональном разложении функции $f(x)$. Вывод основан на условии ортогональности и 2-м свойстве равномерно сходящихся рядов.

$$f(x) = \sum_{n=1}^{\infty} c_n \varphi_n(x) \Longrightarrow f(x)\varphi_k(x) = \sum_{n=1}^{\infty} c_n \varphi_n(x)\varphi_k(x) \Longrightarrow$$

$$\int_a^b f(x)\varphi_k(x)dx = \int_a^b \sum_{n=1}^{\infty} c_n \varphi_n(x)\varphi_k(x)dx = \sum_{n=1}^{\infty} c_n \int_a^b \varphi_n(x)\varphi_k(x)dx.$$

Благодаря ортогональности системы функций $\{\varphi_n(x)\}_{n=1}^{+\infty}$ ряд

$$\sum_{n=1}^{\infty} c_n \int_a^b \varphi_n(x)\varphi_k(x)dx$$

имеет только одно ненулевое слагаемое, в котором $n = k$:

$$c_k \int_a^b \varphi_k(x)\varphi_k(x)dx = c_k \int_a^b \varphi_k^2(x)dx.$$

В результате этих преобразований приходим к равенству

$$\int_a^b f(x)\varphi_k(x)dx = c_k \int_a^b \varphi_k^2(x)dx,$$

из которого следует формула для нахождения любого коэффициента в ортогональном разложении функции $f(x)$:

$$c_k = \frac{\int_a^b f(x)\varphi_k(x)dx}{\int_a^b \varphi_k^2(x)dx}, \text{для } \forall\, k \in \mathbb{N}. \qquad (5.29)$$

В случае ортогональности с весом формула (5.29) преобразуется к виду

$$c_k = \frac{\int_a^b \rho(x) f(x)\varphi_k(x)dx}{\int_a^b \rho(x)\varphi_k^2(x)dx}, \text{для } \forall\, k \in \mathbb{N}. \qquad (5.30)$$

Коэффициенты c_k, построенные по формуле (5.29)или (5.30), называются ***коэффициентами Фурье***.

5.2.3. Тригонометрические ряды Фурье

Связь между математикой и искусством удивительна, но не случайна. Гармония – закон природы, по которому выстроено мироздание.
Шумихин С. и Шумихина А.
«Число π. История длиною в 4000 лет».

Рассмотрим бесконечную систему функций $\left\{\cos\frac{n\pi x}{l}, \sin\frac{n\pi x}{l}\right\}_{n=0}^{+\infty}$ на отрезке $[-l; +l]$. Легко доказать, что эта система функций удовлетворяет условию ортогональности (5.25), а значит является ортогональной системой на $[-l; +l]$.

Любую интегрируемую функцию на отрезке $[-l; +l]$ можно представить в виде равномерно сходящегося тригонометрического ряда Фурье:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right). \qquad (5.31)$$

Коэффициенты в этом разложении вычисляются по формуле (5.29):

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx, \quad b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi x}{l}\,dx, (n = 0,1,2,3,\cdots).$$

Каждое слагаемое $\left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\right)$ в ряде (5.31) называется ***гармоникой***. В ряде прикладных задач удобнее использовать гармоники в преобразованном виде

$$a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l} = \sqrt{a_n^2 + b_n^2}\left(\frac{a_n}{\sqrt{a_n^2 + b_n^2}}\cos\frac{n\pi x}{l} + \frac{b_n}{\sqrt{a_n^2 + b_n^2}}\sin\frac{n\pi x}{l}\right).$$

Выравжение $\sqrt{a_n^2 + b_n^2} = A_n$ называется ***амплитудой.*** Коэффициенты

$$\frac{a_n}{\sqrt{a_n^2 + b_n^2}}, \qquad \frac{b_n}{\sqrt{a_n^2 + b_n^2}}$$

можно рассматривать соответственно как синус и косинус (или наоборот) угла φ_n, который называется ***фазой***. Для завершения преобразования используем формулу

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \text{ или } \cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \Rightarrow$$

$$a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l} = A_n \sin\left(\frac{n\pi x}{l} + \varphi_n\right) = A_n \cos\left(\frac{n\pi x}{l} - \varphi_n\right).$$

В результате ряд (5.31) преобразуется к одному из двух видов

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}A_n\sin\left(\frac{n\pi x}{l}+\varphi_n\right)=\frac{a_0}{2}+\sum_{n=1}^{\infty}A_n\cos\left(\frac{n\pi x}{l}-\varphi_n\right).$$

Разложение функции в тригонометрический ряд Фурье с последующим анализом каждой гармоники называется в прикладной математике ***гармоническим анализом***.

РАЗЛОЖЕНИЕ ЧЁТНОЙ ФУНКЦИИ

Чётная функция $f(x)$ раскладывается в ряд Фурье на отрезке $[-l,+l]$ только по косинусам:

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\frac{n\pi x}{l},\quad a_n=\frac{2}{l}\int_0^l f(x)\cos\frac{n\pi x}{l}dx,(n=0,1,2,3,\cdots),$$

$$b_n=\frac{1}{l}\int_{-l}^{l}f(x)\,sin\frac{n\pi x}{l}dx=0,(n=1,2,3,\cdots).$$

Замечание 5. Подынтегральная функция в последнем равенстве является нечётной. Интеграл от нечётной функции по симметричной области всегда равен нулю.

РАЗЛОЖЕНИЕ НЕЧЁТНОЙ ФУНКЦИИ

Нечётная функция $f(x)$ раскладывается в ряд Фурье на отрезке $[-l,+l]$ только по синусам:

$$f(x)=\sum_{n=1}^{\infty}b_n\,sin\frac{n\pi x}{l},\quad b_n=\frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}dx,(n=1,2,3,\cdots),$$

$$a_n=\frac{1}{l}\int_{-l}^{l}f(x)\,cos\frac{n\pi x}{l}dx=0,(n=0,1,2,3,\cdots).$$

РАЗЛОЖЕНИЕ ФУНКЦИИ НА ОТРЕЗКЕ $[0,l]$

Системы функций $\left\{\cos\frac{n\pi x}{l}\right\}_{n=0}^{+\infty}$ и $\left\{\sin\frac{n\pi x}{l}\right\}_{n=1}^{+\infty}$ ортогональны на $[0,l]$. Поэтому интегрируемую на этом отрезке функцию $f(x)$ можно разложить в тригонометрический ряд Фурье только по синусам или только по косинусам:

$$f(x)=\sum_{n=1}^{\infty}b_n\sin\frac{n\pi x}{l},\ b_n=\frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}dx,(n=1,2,3,\cdots);$$

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\frac{n\pi x}{l},\quad a_n=\frac{2}{l}\int_0^l f(x)\cos\frac{n\pi x}{l}dx,(n=0,2,3,\cdots).$$

РАВЕНСТВО ПАРСЕВАЛЯ–СТЕКЛОВА

Если функция $f(x)$ представлена в виде ряда Фурье на отрезке $[-l; l]$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

тогда справедливо равенство Парсеваля–Стеклова

$$\int_{-l}^{l} f^2(x)\,dx = l\left(\frac{a_0^2}{2} + \sum_{n=1}^{\infty} (a_n^2 + b_n^2)\right). \qquad (5.32)$$

Для чётных и нечётных функций равенство (5.32) упрощается. Если $f(x)$ – чётная функция, то

$$\int_{0}^{l} f^2(x)\,dx = \frac{l}{2}\left(\frac{a_0^2}{2} + \sum_{n=1}^{\infty} a_n^2\right).$$

Если $f(x)$ – нечётная функция, то

$$\int_{0}^{l} f^2(x)\,dx = \frac{l}{2}\sum_{n=1}^{\infty} b_n^2.$$

Доказательство равенства (5.32) представлено в приложении 2.

5.2.4. Тригонометрические полиномы Фурье

Рассмотрим интегрируемую функцию $f(x)$ на отрезке $[0; l]$. Ставится задача о нахождении такого тригонометрического полинома (многочлена)

$$\Phi_M(x) = \frac{a_0}{2} + \sum_{n=1}^{M} a_n \cos\frac{n\pi x}{l} \quad \text{или} \quad \Phi_M(x) = \sum_{n=1}^{M} b_n \sin\frac{n\pi x}{l},$$

чтобы была обеспечена ***близость в среднем*** δ_M к функции $f(x)$:

$$\delta_M = \int_{a}^{b} \big(f(x) - \Phi_M(x)\big)^2\,dx. \qquad (5.33)$$

Индекс M в обозначении полинома $\Phi_M(x)$ равен числу слагаемых в этом полиноме и называется ***порядком***, или ***степенью полинома***.

Аналогично ставится задача о построении тригонометрического полинома

$$\Phi_M(x) = \frac{a_0}{2} + \sum_{n=1}^{M} a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}$$

для интегрируемой на $[-l; l]$ функции $f(x)$.

Доказана следующая теорема.

Теорема 3. Среди всех тригонометрических многочленов заданного порядка M величина δ_M принимает наименьшее значение, если многочлен $\Phi_M(x)$ является частичной суммой тригонометрического ряда Фурье функции $f(x)$. Для таких полиномов Фурье всегда справедливо предельное равенство

$$\lim_{M\to+\infty} \delta_M = 0. \qquad (5.34)$$

Поэтому говорят, что тригонометрические ряды Фурье ***сходятся в среднем***.

АППРОКСИМАЦИЯ ТАБЛИЧНОЙ ФУНКЦИИ

Особое место в теории аппроксимации занимает приближение таблично заданных функций тригонометрическими полиномами Фурье, которые всегда используют в обработке временных рядов. Такая аппроксимация помогает находить периодические процессы в рядах наблюдений.

Пусть функция $f(x)$ задана таблично: $\{x_i, y_i\}_{i=0}^{N}$ в равноотстоящих узлах $x_i \in [a; b]$. Значения аргумента в этих узлах можно записать в виде

$$x_i = \left(a + \frac{i}{N}(b-a)\right). \qquad (5.35)$$

Аппроксимирующую функцию будем строить в виде частичной суммы $\Phi_M(x)$ тригонометрического ряда Фурье. Для этого, прежде всего, перейдём от отрезка $[a, b]$ к отрезку $[-\pi, \pi]$ с помощью линейного преобразования

$$t = \frac{2\pi(x-a)}{(b-a)} - \pi. \qquad (5.36)$$

Узлы x_i преобразуются в узлы $t_i \in [-\pi; \pi]$:

$$t_i = \frac{2\pi}{(b-a)}\left((a + \frac{i}{N}(b-a)) - a\right) - \pi = \frac{2\pi i}{N} - \pi \Rightarrow t_i = \frac{\pi(2i-N)}{N}.$$

Полином Фурье приводится к виду

$$\Phi_M(t) = \frac{a_0}{2} + \sum_{n=1}^{M} (a_n \cos nt + b_n \sin nt), \qquad (5.37)$$

$$a_0 = \frac{1}{N}\sum_{i=0}^{N} y_i\,,\ \ a_n = \frac{2}{N}\sum_{i=0}^{N} y_i \cos(nt_i),\ \ b_n = \frac{2}{N}\sum_{i=0}^{N} y_i \sin(nt_i), n = 1, \cdots, M.$$

После построения полинома $\Phi_M(t)$ возвращаемся к исходному аргументу x с помощью равенства (5.35).

Построение многочлена Фурье для дискретной функции основано на следующей теореме.

Теорема 4. Если $f(t)$– периодическая функция с периодом 2π, и порядок полинома $\Phi_M(t)$ связан с числом узлов таблицы следующим неравенством

$$N \geq 2M + 1 \Longrightarrow M \leq \frac{N-1}{2},$$

тогда полином (5.37) минимизирует сумму квадратов отклонений

$$\sum_{i=1}^{N} \left(f(t_i) - \Phi_M(t_i)\right)^2.$$

ОЦЕНКА ПОГРЕШНОСТИ АППРОКСИМАЦИИ

В основе оценки погрешности аппроксимации полиномами $\Phi_M(x)$ лежит среднеквадратичная погрешность

$$\sigma_M = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(f(x_i) - \Phi_M(x_i)\right)^2} \qquad (5.38)$$

или оценка погрешности с учётом равенства Парсеваля:

$$min\,\delta_M = \sum_{i=1}^{N}(y_i)^2 - N\left(\frac{a_0^2}{2} + \sum_{n=1}^{M}(a_n^2 + b_n^2)\right). \qquad (5.39)$$

С увеличением числа слагаемых в $\Phi_M(x)$ уменьшаются правые части равенств (5.38), (5.39), но при этом возможно увеличение осцилляции аппроксиманта $\Phi_M(x)$.

АЛГОРИТМ ПОСТРОЕНИЯ

Допустим, что известна погрешность δ исходных табличных данных, тогда выбор оптимального порядка полинома Фурье можно сделать с помощью следующего алгоритма.

Пусть $K = 1$. Вычисляется первое слагаемое полинома Фурье: $\{\Phi_1(x_i)\}_{i=0}^{N}$.

1. Число K увеличивается на единицу, вычисляется следующее слагаемое полинома и сам полином $\{\Phi_K(x_i)\}_{i=0}^{N}$.

2. Сравнивается σ_K (5.38) с исходной погрешностью δ. Возможны три случая.

– Если $\sigma_1 \gg \delta$, тогда происходит возвращение к пункту 1.

– Если при некотором значении K $\sigma_K \leq \delta$, то построен аппроксимирующий полином $\Phi_K(x)$ оптимального порядка.
– Если при очередном увеличении числа K $\sigma_K \ll \delta$, тогда последнее слагаемое в полиноме $\Phi_K(x)$ надо убрать как недостоверное. Оставшийся после этого аппроксимант будет иметь оптимальный порядок.

СВЯЗЬ ОБЪЁМА ТАБЛИЦЫ С ПОРЯДКОМ ПОЛИНОМА ФУРЬЕ

Интерес представляет связь между объёмом N исходной табличной функции и оптимальным количеством K слагаемых в многочлене $\Phi_K(x)$. Рассмотрим три ситуации.

1. Если число N достаточно большое, а оптимальный порядок K намного меньше: $K \ll N$, тогда можно считать, что полином построен удачно.
2. Если при большом числе N оптимальное K тоже большое: $K \sim N$, надо искать другие базисные функции и строить новый аппроксимант.
3. Если при маленьком объёме исходной табличной функции поиск оптимального порядка полинома $\Phi_K(x)$ весьма затруднителен, следует тоже подбирать другие базисные функции.

ИНТЕРПОЛЯЦИОННЫЕ ПОЛИНОМЫ ФУРЬЕ

Рассмотрим отдельно случай, когда исходная табличная функция имеет чётное количество узлов $N = 2M$. Пусть, например, полином Фурье содержит M гармоник. В каждую гармонику входит пара a_k, b_k. В этом случае полином определяется N коэффициентами. Построим систему из N линейных относительно $\{a_k, b_k\}_{i=1}^{M}$ алгебраических уравнений

$$\begin{cases} y_1 = \Phi_M(x_1), \\ y_2 = \Phi_M(x_2), \\ \dots\dots\dots\dots\dots, \\ y_N = \Phi_M(x_N). \end{cases} \qquad (5.40)$$

Если количество гармоник в полиноме Фурье (5.37) равно половине объёма аппроксимируемой таблицы:

$$M = \frac{N}{2},$$

тогда у системы (5.40) существует единственное решение. Многочлен Фурье становится интерполяционным многочленом $\Phi_M(x)$.

5.2.5. Многочлены Чебышева

Эти полиномы определены на отрезке $[-1; 1]$. В тригонометрической форме они записываются очень кратко и просто в виде

$$T_n(x) = \cos(n \arccos x), \text{где } n = 0,1,2,3, \dots. \qquad (5.41)$$

В математическом мире полиномы Чебышева обозначаются всегда символом $T_n(x)$, так как на латинском языке фамилия Чебышев – Tschebicheff.

СВОЙСТВА ПОЛИНОМОВ $T_n(x)$

Полиномы Чебышева имеют много замечательных свойств. Отметим некоторые из них.

1. Любой полином Чебышева $T_n(x)$ $(n \neq 0)$ имеет n различных действительных корней $\{x_k\}_{k=1}^{n}$, которые лежат в области $[-1; 1]$:

$$T_n(x) = 0 \Rightarrow n \arccos x = \pi k - \frac{\pi}{2} \Longrightarrow \arccos x = \frac{\pi(2k-1)}{2n},$$

$$0 \leq \frac{\pi(2k-1)}{2n} \leq \pi \Longrightarrow 0 \leq 2k-1 \leq 2n \Rightarrow 1 \leq 2k \leq 2n+1 \Longrightarrow$$

$$\frac{1}{2} \leq k \leq \frac{2n+1}{2} \Rightarrow 1 \leq k \leq n \Longrightarrow x_k = \cos\frac{\pi(2k-1)}{2n} \quad \text{для } k = 1, \ldots\ , n.$$

Полученные корни называются абсциссами (узлами) Чебышева. На оси абсцисс эти узлы расположены неравномерно. Они сгущаются к концам отрезка $[-1; 1]$.

2. Количество экстремальных значений у полинома $T_n(x)$ на отрезке $[-1; 1]$ равно $(n+1)$. Модули всех экстремальных значений равны единице. Легко доказать, что экстремальные значения достигаются в точках

$$x_э = \cos\frac{\pi k}{n} \quad \text{для } k = 0, 1, \ldots\ , n.$$

3. Многочлены $T_n(x)$, начиная с $T_3(x)$, можно построить с помощью рекуррентной формулы

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x).$$

Далее представлен очень простой вывод этого знаменитого соотношения.

Очевидно, что $T_0(x) = 1$ и $T_1(x) = x$. Введём вспомогательное обозначение. Пусть $\alpha = \arccos x$. Используем известные формулы:

$$\cos(n+1)\alpha = \cos n\alpha \cos\alpha - \sin n\alpha \sin\alpha,$$

$$\cos(n-1)\,\alpha = \cos n\alpha \cos\alpha + \sin n\alpha \sin\alpha.$$

Сложение этих равенств приводит к формуле

$$\cos(n+1)\,\alpha + \cos(n-1)\alpha = 2\cos\alpha \cos n\alpha \Rightarrow$$

$$\Rightarrow \cos(n+1)\,\alpha = 2\cos\alpha \cos n\alpha - \cos(n-1)\alpha.$$

Полученная формула верна для любого натурального числа n, поэтому верна аналогичная формула

$$\cos n\,\alpha = 2\cos\alpha\cos(n-1)\alpha - \cos(n-2)\alpha.$$

Переходя в последнем равенстве к обозначениям полиномов Чебышева, получим рекуррентную формулу

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x).$$

С помощью этой формулы многочлены Чебышева можно записать в алгебраическом виде. Ниже представлены первые восемь полиномов $T_n(x)$.

$T_0(x) = 1,$ $T_1(x) = x,$

$T_2(x) = 2x^2 - 1,$ $T_3(x) = 4x^3 - 3x,$

$T_4(x) = 8x^4 - 8x^2 + 1,$ $T_5(x) = 16x^5 - 20x^3 + 5x,$

$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1,$ $T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x.$

4. Полиномы с чётными номерами содержат только чётные степени x, а полиномы с нечётными номерами – только нечётные степени x. Благодаря этому полиномы Чебышева обладают следующим свойством.

5. Свойство симметрии. Если номер полинома Чебышева – чётное число, то полином – чётная функция:

$$k = 2n \Rightarrow T_k(x) = T_k(-x).$$

Если номер полинома – нечётен, то полином Чебышева – нечётная функция:

$$k = 2n + 1 \Rightarrow T_k(x) = -T_k(-x).$$

6. Старший коэффициент многочлена $T_n(x)$ всегда равен 2^{n-1}.

7. Многочлен Чебышева одинаково уклоняется от оси абсцисс в положительном и отрицательном направлениях.

8. Чебышев доказал, что среди всех алгебраических многочленов, заданных на $[-1; 1]$ с единичным старшим коэффициентом, многочлен

$$\hat{T}_n(x) = \frac{1}{2^{n-1}}T_n(x)$$

является единственным полиномом, наименее уклоняющимся от нуля. Это свойство представляет большой интерес в теории аппроксимации. Приближение функции многочленами $\hat{T}_n(x)$ минимизирует максимальное значение погрешности аппроксимации. Это означает, что для $\hat{T}_n(x)$ выполняется ***условие минимакса***.

9. Многочлены $T_n(x)$ ортогональны с весом $\rho(x) = \frac{1}{\sqrt{1-x^2}}$ на $[-1; 1]$:

$$\int_{-1}^{+1} \frac{1}{\sqrt{1-x^2}} T_n(x)T_m(x)dx = \begin{cases} 0, & \text{если } n \neq m, \\ \frac{\pi}{2}, & \text{если } n = m \neq 0, \\ \pi, & \text{если } n = m = 0. \end{cases}$$

10. Для многих функций разложение в равномерно сходящийся ряд по многочленам $T_n(x)$ сходится намного быстрее разложений в другие сходящийся ряды

$$f(x) = \sum_{n=0}^{\infty} c_n T_n(x) = \sum_{n=1}^{M} c_n T_n(t) + R, \qquad (5.42)$$

где коэффициенты c_n вычисляются по формуле

$$c_n = \frac{2}{\pi} \int_{-1}^{1} \frac{f(x)T_n(x)}{\sqrt{1-x^2}} dx. \qquad (5.43)$$

АППРОКСИМАЦИЯ ПОЛИНОМАМИ ЧЕБЫШЕВА

Предположим, что надо аппроксимировать непрерывную функцию $f(x)$ многочленами $T_n(x)$, но $f(x)$ задана на отрезке $[a; b]$, не совпадающим с отрезком $[-1; 1]$. В этом случае предварительно делаем простое линейное преобразование с помощью уравнения прямой, проходящей через две заданные точки:

$$\frac{x - x_1}{x_2 - x_1} = \frac{t - t_1}{t_2 - t_1} \Rightarrow \frac{x - a}{b - a} = \frac{t + 1}{1 + 1} \Rightarrow x = a + \frac{(b-a)}{2}(t + 1) \Rightarrow$$

$$f(x) = f\left(a + \frac{(b-a)}{2}(t + 1)\right) = f(t), \qquad t \in [-1; 1].$$

Аппроксимация функции $f(t)$ многочленами Чебышева представляется в виде суммы

$$f(t) = \sum_{n=1}^{M} c_n T_n(t), \qquad (5.44)$$

где коэффициенты c_n вычисляются по формуле (5.43).

ОЦЕНКА ПОГРЕШНОСТИ

Ряд по многочленам $T_n(x)$ сходится достаточно быстро, поэтому первое в остатке R (5.42) слагаемое ряда можно рассматривать как модуль погрешности чебышевского приближения.

Приближение полиномами $T_n(x)$ уменьшает максимальные погрешности, но при этом возможна большая среднеквадратичная погрешность.

Знакомство с многочленами Чебышева завершим цитатой из книги [7]: «*Многочлены Чебышева – не экспонаты музея математических редкостей, а гибкий и универсальный инструмент познания, который очень активно применяется в современных исследованиях».*

5.2.6. Рациональная аппроксимация

В рамках этой книги мы только очень кратко опишем современный и очень мощный инструмент аппроксимации – рациональные функции.

Определение 6. Рациональной (дробно-рациональной) функцией называется функция

$$R_{N,M}(x) = \frac{P_N(x)}{Q_M(x)}, \qquad \text{где } x \in [a; b], \qquad (5.45)$$

$$P_N(x) = \sum_{k=0}^{N} p_k\, x^k, \qquad Q_M(x) = 1 + \sum_{k=1}^{M} q_k\, x^k.$$

Ставится задача аппроксимации непрерывной функции $f(x)$ функцией (5.45). Построение такой аппроксимации основано на условии ***минимакса***. Поэтому такие рациональные функции строятся с помощью полиномов Чебышева.

Существует несколько методов построения функций (5.45). Наиболее известным среди них является метод Паде ($Pade$). Познакомиться с этим методом можно в работе [10].

Интересно отметить, что требование минимакса для (5.45) выполняется с очень высокой степенью точности для небольших $\{1; 2; 3\}$ значений N и M.

Эксперименты показали, что во многих случаях погрешность аппроксимации станет наименьшей, если степени числителя и знаменателя будут одинаковыми или степень числителя на единицу больше степени знаменателя.

Следует отметить, что все встроенные функции в программном обеспечении, куда входят $\sin x, \cos x, e^x, \ln x, \tan x,$ и др., вычисляются в компьютерах с очень высокой точностью благодаря именно рациональной аппроксимации.

Глава 6

ДИФФЕРЕНЦИАЛЬНЫЕ УРАНЕНИЯ

Всё хорошее в конце концов заканчивается,
но только не в математике.
Ричард Браун. «Математика за 30 секунд»

6.1. Базовые понятия

Дифференциальным уравнением называется уравнение, слагаемые которого содержат производные некоторой функции $y(x)$. Наличие хотя бы одной производной обязательно, а сама функция $y(x)$ и её аргумент могут не быть в уравнении. В общем виде дифференциальное уравнение можно записать в виде

$$F\left(x, y(x), y'(x), y''(x), \cdots, y^{(n)}(x)\right) = 0. \qquad (6.1)$$

Если при подстановке в исходное дифференциальное уравнение функции $y(x)$ и её производных $y'(x), y''(x), \cdots, y^{(n)}(x)$ уравнение становится тождеством, тогда функция $y(x)$ – решение дифференциального уравнения.

Предположим, что в уравнении (6.1) удалось выразить старшую производную в явном виде:

$$y^{(n)}(x) = G\left(x, y(x), y'(x), y''(x), \cdots, y^{(n-1)}(x)\right). \qquad (6.2)$$

В этом случае говорят, что уравнение записано в ***нормальной форме***.

Замечание 1. Дифференциальные уравнения, в которых искомая функция зависит только от одного аргумента, часто называют ***обыкновенными дифференциальными уравнениями.***

Определение 1. ***Порядком дифференциального уравнения*** называется порядок старшей производной, входящей в это уравнение.

Определение 2. Дифференциальное уравнение называется ***линейным***, если оно линейно относительно неизвестной функции и её производных:

$$y^{(n)}(x) + p_1 y^{(n-1)}(x) + \cdots + p_{n-1} y'(x) + p_n y(x) = f(x).$$

Если это условие не выполняется, уравнение называется ***нелинейным.***

Определение 3. Линейное дифференциальное уравнение называется ***однородным,*** если все слагаемые уравнения имеют первый порядок степени относительно искомой функции и её производных:

$$y^{(n)}(x) + p_1 y^{(n-1)}(x) + \cdots + p_{n-1} y'(x) + p_n y(x) = 0.$$

Если это условие не выполняется, уравнение называется ***неоднородным.***

Определение 4. ***Общим решением дифференциального уравнения*** называется решение, полученное в явном виде:

$$y(x) = g(x, c_1, c_2, \cdots, c_n). \qquad (6.3)$$

Определение 5. ***Общим интегралом дифференциального уравнения*** называется решение, полученное в неявном виде:

$$Q(x, y, c_1, c_2, \cdots, c_n) = 0. \qquad (6.4)$$

Параметры $c_1, c_2, \cdots, c_n$ в равенствах (6.3) и (6.4) являются произвольными постоянными. Их количество равно порядку уравнения.

Определение 6. ***Частным решением*** дифференциального уравнения называется решение, полученное из общего решения или общего интеграла в виде

$$y(x) = g(x) \text{ или } Q(x, y) = 0.$$

Численные значения параметров $c_1, c_2, \cdots, c_n$ для частного решения находятся из дополнительных условий, которым удовлетворяет искомая функция в поставленной задаче. Количество дополнительных условий равно порядку дифференциального уравнения. В качестве дополнительных условий обычно рассматривают начальные или краевые (граничные) условия.

Замечание 2. Для некоторых классов простейших дифференциальных уравнений удаётся найти точное решение с помощью техники интегрирования. Поэтому любой процесс построения решения дифференциального уравнения по традиции часто называют ***интегрированием***, а график решения называют ***интегральной кривой***.

6.2. Задача Коши

Предположим, что мы нашли общее решение уравнения первого порядка, и оно равно

$$y(x) = g(x, c_1).$$

Для нахождения параметра c_1 надо знать значения x_0 и y_0 , при которых решение уравнения удовлетворяет равенству

$$y(x_0) = y_0. \qquad (6.5)$$

Условие (6.5) называется ***начальным условием*** для дифференциального уравнения первого порядка.

Рассмотрим вид общего решения дифференциального уравнения второго порядка

$$y(x) = g(x, c_1, c_2).$$

Для нахождения параметров c_1, c_2 нам надо знать такие значения x_0, y_0, при которых решение уравнения удовлетворяет равенству $y(x_0) = y_0$ и знать значение v_0 производной функции $y(x)$ в точке x_0:

$$y'(x_0) = v_0. \qquad (6.6)$$

Условия (6.5) и (6.6) являются ***начальными условиями*** для уравнения второго порядка.

Обратимся к общему решению дифференциального уравнения третьего порядка

$$y(x) = g(x, c_1, c_2, c_3).$$

Для нахождения параметров c_1 , c_2, c_3 надо знать x_0, y_0, при которых решение уравнения удовлетворяет равенству (6.5), знать первую производную (6.6) и значение w_o второй производной функции $y(x)$ в точке x_0:

$$y''(x_0) = w_o. \qquad (6.7)$$

Условия (6.5) – (6.7) являются начальными условиями для уравнения третьего порядка. Количество начальных условий должно равняться порядку уравнения.

Определение 7. Задача нахождения частного решения уравнения с заданными начальными условиями называется ***задачей Коши.***

Задача Коши может быть поставлена для систем. Предположим, что система состоит из m дифференциальных уравнений первого порядка:

$$\begin{cases} y_1' = f_1(x, y_1, y_2, y_3, \cdots, y_m), \\ y_2' = f_2(x, y_1, y_2, y_3, \cdots, y_m), \\ y_3' = f_3(x, y_1, y_2, y_3, \cdots, y_m), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m' = f_m(x, y_1, y_2, y_3, \cdots, y_m). \end{cases}$$

Начальные условия для такой системы имеют следующий вид

$$y_1(x_0) = y_{0,1},\ y_2(x_0) = y_{0,2},\ y_3(x_0) = y_{0,3}, \cdots, y_m(x_0) = y_{0,m},$$

где x_0 и $\{y_{0,k}\}_{k=1}^{m}$ – заданные числа.

В этой системе искомым решением являются функции $\{y_k(x)\}_{k=1}^{m}$.

6.2.1. Существование и единственность

До решения задачи Коши, прежде всего, возникает вопрос о существовании и единственности решения поставленной задачи. Ответ даёт следующая теорема.

Теорема 1. (Теорема Пикара). Существует единственное решение задачи Коши для уравнения

$$y^{(n)}(x) = G\left(x, y(x), y'(x), y''(x), \cdots, y^{(n-1)}(x)\right), \qquad (6.8)$$

если в начальной точке $\left(x_0\ , y_0\ , y_0', y_0'', \cdots, y_0^{(n-1)}\right)$ и некоторой её окрестности функция G непрерывна относительно всех своих аргументов и непрерывны первые частные производные функции G по $y, y', y'', \cdots, y^{(n-1)}$.

Эта теорема обеспечивает достаточные условия существования и единственности решения.

Есть другой подход к вопросу существования единственного решения уравнения. Обратимся к задаче Коши для дифференциального уравнения первого порядка.

$$y' = f(x, y) \text{ с начальным условием } y(x_0) = y_0. \quad (6.9)$$

Определение 8. Предположим, что функция $f(x, y)$ задана и непрерывна в области

$$D = \{x \in [a; b], y \in [c; d]\}.$$

Говорят, что функция $f(x, y)$ удовлетворяет в области D ***условию Липшица*** по переменной y, если существует такая постоянная $L > 0$, при которой выполняется неравенство

$$|f(x, y_2) - f(x, y_1)| < L|y_2 - y_1| \text{ для } \forall\, y_2 \in D \text{ и } \forall\, y_1 \in D. \quad (6.10)$$

Величина L в неравенстве (6.10) называется ***постоянной Липшица*** для функции $f(x, y)$.

Теорема 2. Если правая часть уравнения $y' = f(x, y)$ с начальным условием $y(x_0) = y_0$ удовлетворяет условию Липшица в области

$$D = \{x \in [x_0; b], y \in [y_0; d]\},$$

тогда у задачи Коши решение в этой области существует и единственно.

Если задача Коши (6,9) имеет единственное решение $y(x)$, то через точку $M(x_0, y_0)$ с заданными координатами проходит только одна интегральная кривая.

В теории дифференциальных уравнений есть способы построения общих решений некоторых классов дифференциальных уравнений. Но, в основном, дифференциальные уравнения и системы таких уравнений удаётся решить только приближённо с помощью различных численных методов, которые строят только частные решения по заданным начальным или краевым условиям. Рассмотрим простейшие из таких методов.

6.2.2. Метод Эйлера для уравнения первого порядка

Метод Эйлера важно изучить для понимания более точных и сложных методов построения решений дифференциальных уравнений и систем.

Поставлена задача Коши для уравнения первого порядка

$$y' = f(x, y),\ y(x_0) = y_0. \quad (6.11)$$

Предположим, что правая часть уравнения (6.11) удовлетворяет условию Липшица в заданной области интегрирования.

Будем строить решение задачи Коши на $[a, b]$. Для этого разобьём отрезок на n частей:

$$a = x_0 < x_1 < x_2 < \cdots < x_i < x_{i+1} < \cdots < x_n = b.$$

Пусть расстояние между соседними узлами – постоянная величина h (***шаг интегрирования***). Решение строим в виде табличной функции $y_i = y(x_i)$.

Допустим, что искомая функция дважды непрерывно дифференцируема в заданной области D, поэтому в каждом узле сетки x_{i+1} функцию можно представить в виде многочлена Тейлора:

$$y(x_{i+1}) = y(x_i + h) = y_{i+1} = y_i + y_i' h + \frac{y_i''(c)}{2!} h^2,\ c \in (x_i, x_{i+1}) \Longrightarrow$$

$$y_{i+1} = y_i + y_i' h + \mathrm{O}(h^2). \qquad (6.12)$$

Значение производной y_i' равно значению правой части $f(x_i, y_i)$ уравнения (6.11). Если в равенстве (6.12) отбросить величину $\mathrm{O}(h^2)$, то получим рекуррентную ***формулу Эйлера***

$$y_{i+1} = y_i + f(x_i, y_i)h. \qquad (6.13)$$

Численный метод построения решения задачи Коши (6.11) по формуле (6.13) называется методом Эйлера.

ОЦЕНКА ТОЧНОСТИ МЕТОДА

Локальный порядок точности формулы Эйлера равен $\mathrm{O}(h^2)$. Очевидно, что чем меньше шаг интегрирования, тем выше точность расчёта по этой формуле. Теоретически при стремлении шага интегрирования к нулю приближёние по методу Эйлера сходится к точному решению. Но из-за низкого порядка точности сходимость решения по методу Эйлера очень медленная. Для повышения точности надо проводить много операций, а это приводит к накоплению вычислительных погрешностей.

ГЕОМЕТРИЧЕСКИЙ СМЫСЛ МЕТОДА ЭЙЛЕРА

Сравним формулу Эйлера (6.13) с уравнением касательной к графику дифференцируемой функции $y(x)$ в точке $M(x_i, y_i)$:

$$y(x) = y_i + y'(x_i)(x - x_i).$$

Из сравнения следует, что график решения, найденного по методу Эйлера, является кусочно-линейной кривой. Каждое звено этого графика (рис. 6.1) лежит на касательной, проведённой к интегральной кривой в точке $M(x_i, y_i)$.

6.2.3. Метод Рунге–Кутта для уравнения первого порядка

В начале XX в. математики сначала Рунге, а затем Кутта предложили простой численный метод решения дифференциального уравнения первого порядка. Позже было разработано несколько вариантов (схем) этого метода. Наиболее популярным стал метод Рунге–Кутта четвёртого порядка. С помощью него находится решение задачи Коши

$$y' = f(x, y),\ \ y(x_0) = y_0.$$

Для понимания смысла метода Рунге–Кутта 4-го порядка используем понятие дифференциала функции. Напомним, что дифференциал функции – это линейная часть приращения функции относительно приращения Δx в некоторой точке x_i. Дифференциал вычисляется по формуле

$$dy_i = y'(x_i)\Delta x.$$

Для простоты будем строить решение на равномерной сетке $\{x_i\}_{i=1}^{n}$ с шагом $h = \Delta x$.

Процесс построения решения в каждом узле x_i начинается с вычисления четырёх дифференциалов. Дифференциал k_1 вычисляется в точке $M_1(x_i, y_i)$:

$$k_1 = h\, f(x_i, y_i);$$

дифференциал k_2 – в точке $M_2\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}\right)$:

$$k_2 = h\, f\left(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}\right);$$

дифференциал k_3 – в точке $M_3\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}\right)$:

$$k_3 = h\, f\left(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}\right);$$

дифференциал k_4 – в точке $M_4(x_i + h, y_i + k_3)$:

$$k_4 = h\, f(x_i + h, y_i + k_3).$$

Далее строится приращение $\tilde{\Delta} y_i$ как среднее взвешенное $k_1,\ k_2,\ k_3,\ k_4$:

$$\tilde{\Delta} y_i = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

В качестве значения функции в точке x_{i+1} берётся

$$y_{i+1} = y_i + \tilde{\Delta} y_i.$$

В итоге приходим к формуле Рунге – Кутта 4-го порядка:

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \qquad (6.14)$$

Замечание 3. Точный вывод этой формулы представлен в книге [6].

6.2.4. Метод Рунге – Кутта для системы

Для наглядности обратимся к технике решения системы двух уравнений в нормальной форме:

$$\begin{cases} y_1' = f_1(x, y_1, y_2,), \\ y_2' = f_2(x, y_1, y_2,) \end{cases}$$

с начальными условиями $y_1(x_0) = y_{0,1},\ y_2(x_0) = y_{0,2}$.

1. В каждом узле x_i, начиная с x_0, вычисляется группа дифференциалов в следующем порядке

$$k_1 = h\, f_1\big(x_i, y_{i,1}, y_{i,2}\big),\ m_1 = h\, f_2\big(x_i, y_{i,1}, y_{i,2}\big);$$

$$k_2 = h\, f_1\left(x_i + \frac{h}{2}, y_{i,1} + \frac{k_1}{2}, y_{i,2} + \frac{m_1}{2}\right),\ m_2 = h\, f_2\left(x_i + \frac{h}{2}, y_{i,1} + \frac{k_1}{2}, y_{i,2} + \frac{m_1}{2}\right);$$

$$k_3 = h\, f_1\left(x_i + \frac{h}{2}, y_{i,1} + \frac{k_2}{2}, y_{i,2} + \frac{m_2}{2}\right),\ m_3 = h\, f_2\left(x_i + \frac{h}{2}, y_{i,1} + \frac{k_2}{2}, y_{i,2} + \frac{m_2}{2}\right);$$

$$k_4 = h\, f_1\big(x_i + h, y_{i,1} + k_3, y_{i,2} + m_3\big),\ m_4 = h\, f_2\big(x_i + h, y_{i,1} + k_3, y_{i,2} + m_3\big).$$

2. Вычисляются приращения $\tilde{\Delta} y_{i,1}$ и $\tilde{\Delta} y_{i,2}$:

$$\tilde{\Delta} y_{i,1} = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4);$$

$$\tilde{\Delta} y_{i,2} = \frac{1}{6}(m_1 + 2m_2 + 2m_3 + m_4).$$

3. По формуле Рунге – Кутта строятся значения функций y_1 и y_2 в точке $x_{i+1} = x_i + h$:

$$y_{(i+1),1} = y_{i,1} + \tilde{\Delta} y_{i,1} = y_{i,1} + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4),$$

$$y_{(i+1),2} = y_{i,2} + \tilde{\Delta} y_{i,2} = y_{i,2} + \frac{1}{6}(m_1 + 2m_2 + 2m_3 + m_4).$$

Доказано, что локальный порядок точности этого метода равен O (h^5)

Рис. 6.1

6.2.5. Алгоритм решения задачи Коши.

Построение численного решения задачи Коши в узлах $\{x_i\}_{i=1}^{n}$ области $[a, b]$ только с помощью формулы Эйлера или формулы Рунге–Кутта может привести к решению, которое не будет иметь ничего общего с решением поставленной задачи. Этот печальный факт связан с накоплением вычислительной погрешности при переходе от узла x_i к x_{i+1} (рис. 6.2).

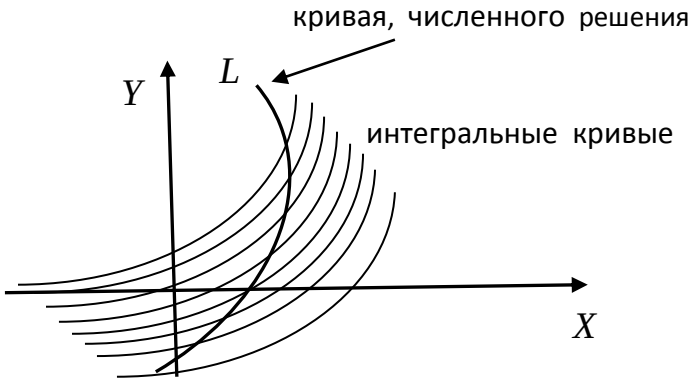

Рис. 6.2

Как правило, оптимальный шаг интегрирования h нельзя выбрать априори. Если мы хотим построить хорошее приближение к точному решению задачи Коши, необходимо на протяжении всего процесса вычисления по формуле Эйлера или Рунге–Кутта использовать такой алгоритм вычисления, в котором не будет накапливаться вычислительная погрешность. Значения приближенного решения должны лежать внутри достаточно узкого канала между теоретическими интегральными кривыми исходного уравнения (рис. 6.3).

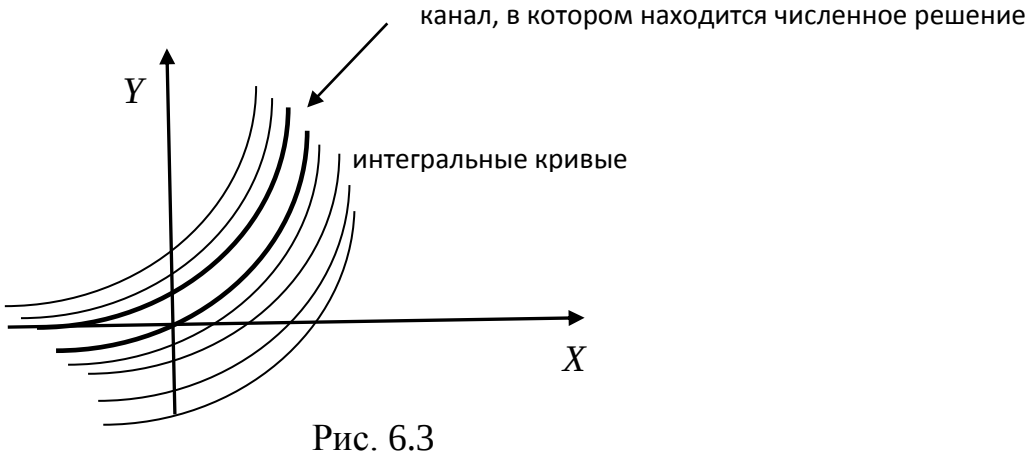

Рис. 6.3

Один из самых простых и достаточно эффективных алгоритмов решения задачи Коши – это алгоритм с ***автоматическим выбором шага***. Этот алгоритм и его модификации реализованы и успешно используются в стандартных программах различных математических пакетов. Опишем кратко один из простейших вариантов этого алгоритма.

АЛГОРИТМ С АВТОМАТИЧЕСКИМ ВЫБОРОМ ШАГА ДЛЯ ОДНОГО УРАВНЕНИЯ

1. Входные данные для алгоритма: правая часть дифференциального уравнения $f(x,y)$, начальные условия $y(x_0)=y_0$, необходимая точность решения ε и начальный шаг интегрирования $h = h_{нач}$.

2. По формуле Рунге–Кутта (6.14) вычисляется y_1 в точке $x_1 = x_0 + h$.

3. Исходный шаг интегрирования делится пополам: $h = 0{,}5\,h_{нач}$. С этим половинным шагом дважды используется формула (6.14) для вычисления $\bar{y}_1$ в той же точке $x_1 = x_0 + h$.

4. Сравниваются два значения искомой функции y_1 и $\bar{y}_1$ по формуле относительной погрешности:

$$\left|\frac{y_1 - \bar{y}_1}{y_1}\right| = \delta. \qquad (6.15)$$

Рассматриваются три случая:

– Если $\delta < \varepsilon$, тогда по формуле (6.14) вычисляется значение решения y_2 в точке $x_2 = x_1 + h_{нач}$.

– Если $\delta \ll \varepsilon$ (намного меньше), тогда начальный шаг $h_{нач}$ увеличивается вдвое $h = 2h_{нач}$ и далее строится решение y_2 в точке $x_2 = x_1 + h$.

– Если $\delta > \varepsilon$, тогда шаг $h_{нач}$ уменьшается вдвое и в качестве исходного шага берётся $h = 0{,}5h_{нач}$ и с этим шагом находится решение y_1 в точке $x_1 = x_0 + h$, а затем строится решение $\bar{y}_1$ (п. 3). Далее происходит оценка погрешности (6.15) и выбор оптимального шага интегрирования (п. 4).

Шаги 2 – 4 алгоритма повторяются в каждой последующей точке интегрирования $x_{i+1} = x_i + h$. Благодаря постоянной корректировке шага h, все вычисленные значения решения задачи Коши находятся в достаточно узком канале между близкими теоретическими интегральными кривыми исходного уравнения (рис.6.3). Ширина этого канала зависит от заданной точности ε.

АЛГОРИТМ С АВТОМАТИЧЕСКИМ ВЫБОРОМ ШАГА ДЛЯ СИСТЕМЫ УРАВНЕНИЙ

Структура алгоритма сохраняется в случае решения систем, содержащих m неизвестных функций. Остановимся подробнее на 4-м шаге алгоритма. С помощью формулы (6.15) в каждой точке x_i вычисляется относительная погрешность δ_k для всех функций:

$$\left\{\left|\frac{y_{i,k} - \bar{y}_{i,k}}{y_{i,k}}\right| = \delta_k\right\}_{k=1}^{m}; \quad i = 1,2,\cdots,n.$$

Интегрирование переходит в точку $x_{i+1} = x_i + h$ только при условии $\delta_k < \varepsilon$ или $\delta_k \ll \varepsilon$ для каждой пары

$$y_{i,k}\ ,\ \bar{y}_{i,k},\quad k = 1,2,\cdots,m,\ \ i = 1,2,\cdots,n.$$

6.2.6. Задачи Коши для уравнений n-го порядка

Любое дифференциальное уравнение n-го порядка можно преобразовать в систему уравнений первого порядка. Для этого вводят новые переменные:

$$y'(x) = g_1(x),\ y''(x) = g_2(x), \cdots, y^{(n-1)}(x) = g_{n-1}(x).$$

Дифференциальное уравнение

$$y^{(n)}(x) = F\left(x, y(x), y'(x), y''(x), \cdots, y^{(n-1)}(x)\right) \qquad (6.16)$$

превращается в систему дифференциальных уравнений первого порядка

$$\begin{cases} g'_{n-1}(x) = F(x, y, g_1, g_2, \cdots, g_{n-1}), \\ g'_{n-2}(x) = g_{n-1}(x), \\ g'_{n-3}(x) = g_{n-2}(x), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ g'_1(x) = g_2(x), \\ y'(x) = g_1(x). \end{cases} \qquad (6.17)$$

Начальные условия для дифференциального уравнения (6.16)

$$y(x_0) = y_0\ ,\ \ y'(x_0) = y'_0, \cdots,\ \ y^{(n-1)}(x_0) = y_0^{(n-1)}$$

становятся начальными условиями для системы (6.17)

$$y(x_0) = y_0\ ,\ \ g_1(x_0) = y'_0, \cdots,\ \ g_{n-1}(x_0) = y_0^{(n-1)}.$$

Задачу Коши для системы (6.17) можно решить методом Рунге–Кутта. В процессе решения уравнения (6.16) с помощью системы (6.17) находятся не только значения функции $y(x)$, но и одновременно с этим вычисляются значения всех производных этой функции до $(n-1)$-го порядка включительно.

6.3. Краевая задача

Задача Коши и задача с краевыми условиями в нескольких заданных точках существенно отличаются по методам их решения. Для краевых задач порядок дифференциального уравнения может быть равен любому натуральному числу $m \geq 2$. Количество краевых условий должно равняться m. Краевые условия каким-либо образом распределены между границами области интегрирования.

Разберём простейшие методы решения краевых задач на примере дифференциальных уравнений второго порядка

$$y''_{xx} = F(x, y, y'_x),\ a \leq x \leq b. \qquad (6.18)$$

Доказана следующая теорема.

Теорема 3. Дифференциальное уравнение (6.18) второго порядка с заданными краевыми условиями в граничных точках $x = a,\ x = b$ имеет единственное решение, если выполняются следующие условия:

1. $F(x, y, y'_x)$, $F(x, y, y'_x)'_y$, $F(x, y, y'_x)'_{y'_x}$ – непрерывные функции по трём переменным;

2. $F(x, y, y'_x)'_x > 0$;

3. существует такое число $M > 0$, что $\left| F(x, y, y'_x)'_y \right| < M$, следовательно, частная производная по переменной y ограничена.

6.3.1. Краевые условия

Будем рассматривать линейную краевую задачу для уравнения второго порядка

$$y''(x) + p(x)y'(x) + q(x)y(x) = f(x). \qquad (6.19)$$

Пусть аргумент x изменяется на отрезке $[a; b]$. Точки a, b – граничные точки, в которых задаются краевые условия. Определим три рода краевых условий для искомой функции $y(x)$.

1. Заданы значения функции:

$$y(a) = \gamma_1, y(b) = \gamma_2. \qquad (6.20)$$

2. Заданы значения производной:

$$y'(a) = \gamma_1, . y'(b) = \gamma_2. \qquad (6.21)$$

3. Заданы значения ***линейной комбинации*** искомой функции и её первой производной:

$$\alpha_1 y(a) + \beta_1 y'(a) = \gamma_1, \qquad \alpha_2 y(b) + \beta_2 y'(b) = \gamma_2,$$

где числа α_1, β_1, α_2, β_2 известны. Такие краевые условия часто называют ***смешанными***. Частным случаем смешанных условий являются краевые условия двух видов:

$$y(a) = \gamma_1, y'(b) = \gamma_2 \text{ или } y'(a) = \gamma_1, y(b) = \gamma_2. \quad (6.22)$$

Если в любых краевых условиях $\gamma_1 = \gamma_2 = 0$, условия называют ***однородными***.

6.3.2. Разностная аппроксимация производных

Предположим, что функция $y(x)$ задана на отрезке $[a; b]$ и дважды непрерывно дифференцируема в этой области. Область $[a; b]$ разбита на n равных частей с достаточно малым $(h \ll 1)$ шагом $h = (b - a)/n$. Пусть $\{x_i\}_{i=0}^{n}$ – ***узлы разбиения отрезка***.

Выразим значение функции $y(x)$ в узле x_{i+1} через значение функции и её производных в предыдущем узле x_i. Для этого используем ***многочлен Тейлора:***

$$y(x_{i+1}) = y_{i+1} = y_i + \frac{y'_i}{1!}h + \frac{y''(c)}{2!}h^2, \qquad \exists\, c \in (x_i; x_{i+1}) \Longrightarrow$$

$$y_{i+1} = y_i + \frac{y'_i}{1!}h + \mathrm{O}(h^2) \Longrightarrow y'_i = \frac{y_{i+1} - y_i}{h} + \mathrm{O}(h). \qquad (6.23)$$

Равенство (6.23) приводит к формуле

$$y_i' = \frac{y_{i+1} - y_i}{h}. \qquad (6.24)$$

Формула (6.24) называется ***правосторонней разностной аппроксимацией первой производной*** в любом внутреннем узле разбиения x_i и в левой граничной точке a области $[a; b]$:

$$y'(a) = y_0' = \frac{y_1 - y_0}{h}.$$

Порядок точности формулы (6.24) равен $O(h)$. Запишем равенство (6.23) в виде

$$y_{i-1} = y_i - \frac{y_i'}{1!}h + O(h^2) \Longrightarrow y_i' = \frac{y_i - y_{i-1}}{h} + O(h).$$

Последнее равенство приводит к формуле

$$y_i' = \frac{y_i - y_{i-1}}{h}. \qquad (6.25)$$

Её называют ***левосторонней разностной аппроксимацией первой производной.*** Порядок точности равен $O(h)$. Такую аппроксимацию можно использовать для приближения первой производной в правой граничной точке области интегрирования:

$$y'(b) = y_n' = \frac{y_n - y_{n-1}}{h}.$$

Пусть функция $y(x)$ трижды непрерывно дифференцируема на $[a; b]$ с тем же разбиением. Выразим y_{i+1} и y_{i-1} через значения функции и её производных в точке x_i:

$$y_{i+1} = y_i + \frac{y_i'}{1!}h + \frac{y_i''}{2!}h^2 + \frac{y'''(c)}{3!}h^3, \quad c \in (x_i, x_{i+1}),$$

$$y_{i-1} = y_i - \frac{y_i'}{1!}h + \frac{y_i''}{2!}h^2 - \frac{y'''(c)}{3!}h^3, \quad c \in (x_{i-1}, x_i).$$

От первого равенства отнимем второе:

$$y_{i+1} - y_{i-1} = 2y_i'h + O(h^3).$$

Выразим из последнего равенство значение первой производной:

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2h} + O(h^2). \qquad (6.26)$$

Равенство (6.26) приводит к формуле

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2h}. \qquad (6.27)$$

Эта формула называется ***центрально-разностной аппроксимацией первой производной.*** Её можно использовать в любом внутреннем узле x_i отрезка $[a; b]$. Порядок точности формулы (6.27) равен $O(h^2)$.

Предположим, что функция $y(x)$ имеет непрерывную производную 4-го порядка. Запишем для неё два многочлена Тейлора, используя соседние узлы $x_{i-1},\ x_i,\ x_{i+1}$:

$$y_{i+1} = y_i + \frac{y_i'}{1!}h + \frac{y_i''}{2!}h^2 + \frac{y_i'''}{3!}h^3 + \frac{y_i^{(4)}(c)}{4!}h^4, \qquad c \in (x_i, x_{i+1}),$$

$$y_{i-1} = y_i - \frac{y_i'}{1!}h + \frac{y_i''}{2!}h^2 - \frac{y_i'''}{3!}h^3 + \frac{y_i^{(4)}(c)}{4!}h^4, \quad c \in (x_{i-1}, x_i).$$

Сложим эти многочлены:

$$y_{i+1} + y_{i-1} = 2y_i + 2\frac{y_i''}{2!} \cdot h^2 + O(h^4).$$

Выразим из последнего равенства значение второй производной:

$$y_i'' = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} + O(h^2). \qquad (6.28)$$

Равенство (6.28) приводит к формуле

$$y_i'' = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}. \qquad (6.29)$$

Она называется ***центрально-разностной аппроксимацией второй производной***. Эту формулу можно использовать в любом внутреннем узле x_i отрезка $[a; b]$. Порядок точности этой формулы $O(h^2)$.

6.3.3. Сеточный метод решения

Применение сеточного метода (***разностного метода***, ***конечно-разностного метода***) для решения краевых задач сводится к простой процедуре замены производных в дифференциальном уравнении на их разностную аппроксимацию. После такой замены уравнение преобразуется в систему линейных алгебраических уравнений. Простота структуры полученной системы наиболее привлекательна в сеточном методе.

КРАЕВЫЕ УСЛОВИЯ ПЕРВОГО РОДА

В качестве примера разберём решение линейного неоднородного дифференциального уравнения второго порядка

$$y''(x) + p(x)y'(x) + q(x)y(x) = f(x) \qquad (6.30)$$

с заданными непрерывными функциями $p(x),\ q(x),\ f(x)$ и краевыми условиями первого рода

$$x \in [a;b], \qquad y(a) = \gamma_1, \qquad y(b) = \gamma_2.$$

Построим на $[a;b]$ сетку с постоянным шагом интегрирования h и узлами разбиения $x_i,\ i = 1, \cdots, n-1$.

Уравнение (6.30) в любом внутреннем узле x_i можно аппроксимировать разностным уравнением. Для этого заменим $y''(x_i)$ и $y'(x_i)$ их центральными разностями (6.27) и (6.28). Пусть $y(x_i) = y_i$.

Разностная аппроксимация уравнения (6.30) имеет вид

$$\frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} + p(x_i)\frac{y_{i+1} - y_{i-1}}{2h} + q(x_i)y_i = f(x_i).$$

Для упрощения этого уравнения введём новые обозначения и приведём подобные члены. Пусть $p(x_i) = p_i,\ q(x_i) = q_i,\ f(x_i) = f_i$:

$$\left(\frac{1}{h^2} - \frac{p_i}{2h}\right)y_{i-1} + \left(q_i - \frac{2}{h^2}\right)y_i + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right)y_{i+1} = f_i. \qquad (6.31)$$

Это уравнение является ***разностной аппроксимацией дифференциального уравнения*** (6.30)

Запишем уравнение (6.31) в каждом внутреннем узле разбиения $\{x_i\}_{i=1}^{n-1}$:

$$\begin{cases} \left(\frac{1}{h^2} - \frac{p_1}{2h}\right)y_0 + \left(q_1 - \frac{2}{h^2}\right)y_1 + \left(\frac{1}{h^2} + \frac{p_1}{2h}\right)y_2 = f_1, \\ \left(\frac{1}{h^2} - \frac{p_2}{2h}\right)y_1 + \left(q_2 - \frac{2}{h^2}\right)y_2 + \left(\frac{1}{h^2} + \frac{p_2}{2h}\right)y_3 = f_2, \\ \left(\frac{1}{h^2} - \frac{p_3}{2h}\right)y_2 + \left(q_3 - \frac{2}{h^2}\right)y_3 + \left(\frac{1}{h^2} + \frac{p_3}{2h}\right)y_4 = f_3, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left(\frac{1}{h^2} - \frac{p_i}{2h}\right)y_{i-1} + \left(q_i - \frac{2}{h^2}\right)y_i + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right)y_{i+1} = f_i, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \left(\frac{1}{h^2} - \frac{p_{n-1}}{2h}\right)y_{n-2} + \left(q_{n-1} - \frac{2}{h^2}\right)y_{n-1} + \left(\frac{1}{h^2} + \frac{p_{n-1}}{2h}\right)y_n = f_{n-1}. \end{cases} \qquad (6.32)$$

В первом и последнем уравнениях значения y_0 и y_n известны из краевых условий первого рода: $y_0 = y(a) = \gamma_1,\ y_n = y(b) = \gamma_2$, поэтому первое слагаемое в 1-ом уравнении и последнее слагаемое в последнем уравнении можно перенести в правую часть системы:

$$\begin{cases} \left(q_1-\dfrac{2}{h^2}\right)y_1+\left(\dfrac{1}{h^2}+\dfrac{p_1}{2h}\right)y_2=-\left(\dfrac{1}{h^2}-\dfrac{p_1}{2h}\right)\gamma_1+f_1, \\ \left(\dfrac{1}{h^2}-\dfrac{p_2}{2h}\right)y_1+\left(q_2-\dfrac{2}{h^2}\right)y_2+\left(\dfrac{1}{h^2}+\dfrac{p_2}{2h}\right)y_3=\ f_2, \\ \left(\dfrac{1}{h^2}-\dfrac{p_3}{2h}\right)y_2+\left(q_3-\dfrac{2}{h^2}\right)y_3+\left(\dfrac{1}{h^2}+\dfrac{p_3}{2h}\right)y_4=\ f_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \left(\dfrac{1}{h^2}-\dfrac{p_i}{2h}\right)y_{i-1}+\left(q_i-\dfrac{2}{h^2}\right)y_i+\left(\dfrac{1}{h^2}+\dfrac{p_i}{2h}\right)y_{i+1}=\ f_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \left(\dfrac{1}{h^2}-\dfrac{p_{n-1}}{2h}\right)y_{n-2}+\left(q_{n-1}-\dfrac{2}{h^2}\right)y_{n-1}=-\left(\dfrac{1}{h^2}+\dfrac{p_{n-1}}{2h}\right)\gamma_n+f_{n-1}. \end{cases} \qquad (6.33)$$

Система (6.33) замкнутая, так как содержит одинаковое количество уравнений и неизвестных. Матрица системы трёхдиагональная. Поэтому значения $\{y_i\}_{i=1}^{n-1}$ находятся методом прогонки (глава 3).

КРАЕВЫЕ УСЛОВИЯ ВТОРОГО РОДА

Краевые условия второго рода для уравнения (6.30) имеют вид

$$y'(a)=\gamma_1, \qquad y'(b)=\gamma_2.$$

Аппроксимируем первую производную $y'(a)$ правосторонней разностью:

$$y'(a)=y_0'=\frac{y_1-y_0}{h}=\gamma_1 \Longrightarrow -y_0+y_1=h\gamma_1 \qquad (6.34)$$

и первую производную $y'(b)$ левосторонней разностью:

$$y'(b)=y_n'=\frac{y_n-y_{n-1}}{h}=\gamma_2 \Longrightarrow -y_{n-1}+y_n=h\gamma_2. \qquad (6.35)$$

Последнее равенство в (6.34) станет первым уравнением в системе для нахождения значений $\{y_i\}_{i=0}^{n}$, а последнее равенство в (6.35) будет последним уравнением в новой системе.

Система состоит из $(n+1)$-го уравнения и содержит столько же неизвестных. Её уравнения, кроме первого и последнего, строятся аналогично уравнениям в системе (6.33). Матрица системы трёхдиагональна, поэтому система легко решается методом прогонки.

Ниже представлена такая система (6.36).

$$\begin{cases} -y_0 + y_1 = h \cdot \gamma_1 \\ \left(q_1 - \frac{2}{h^2}\right) y_1 + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right) y_2 = -\left(\frac{1}{h^2} - \frac{p_1}{2h}\right) \gamma_1 + f_1, \\ \left(\frac{1}{h^2} - \frac{p_2}{2h}\right) y_1 + \left(q_2 - \frac{2}{h^2}\right) y_2 + \left(\frac{1}{h^2} + \frac{p_2}{2h}\right) y_3 = f_2, \\ \left(\frac{1}{h^2} - \frac{p_3}{2h}\right) y_2 + \left(q_3 - \frac{2}{h^2}\right) y_3 + \left(\frac{1}{h^2} + \frac{p_3}{2h}\right) y_4 = f_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_i}{2h}\right) y_{i-1} + \left(q_i - \frac{2}{h^2}\right) y_i + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right) y_{i+1} = f_i, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_{n-1}}{2h}\right) y_{n-2} + \left(q_{n-1} - \frac{2}{h^2}\right) y_{n-1} + \left(\frac{1}{h^2} + \frac{p_{n-1}}{2h}\right) y_n = f_{n-1}. \\ -y_{n-1} + y_n = h \cdot \gamma_2. \end{cases} \quad (6.36)$$

СМЕШАННЫЕ КРАЕВЫЕ УСЛОВИЯ

Смешанные краевые условия, или краевые условия третьего рода (6.22) приводят к одной из представленных ниже систем с трёхдиагональной матрицей.

Решение исходного дифференциального уравнения (6.30) с краевыми условиями

$$y'(a) = \gamma_1, y(b) = \gamma_2$$

сводится к решению линейной алгебраической системы с n неизвестными и n уравнениями:

$$\begin{cases} -y_0 + y_1 = h \cdot \gamma_1, \\ \left(\frac{1}{h^2} - \frac{p_1}{2h}\right) y_0 + \left(q_1 - \frac{2}{h^2}\right) y_1 + \left(\frac{1}{h^2} + \frac{p_1}{2h}\right) y_2 = f_1, \\ \left(\frac{1}{h^2} - \frac{p_2}{2h}\right) y_1 + \left(q_2 - \frac{2}{h^2}\right) y_2 + \left(\frac{1}{h^2} + \frac{p_2}{2h}\right) y_3 = f_2, \\ \left(\frac{1}{h^2} - \frac{p_3}{2h}\right) y_2 + \left(q_3 - \frac{2}{h^2}\right) y_3 + \left(\frac{1}{h^2} + \frac{p_3}{2h}\right) y_4 = f_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_i}{2h}\right) y_{i-1} + \left(q_i - \frac{2}{h^2}\right) y_i + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right) y_{i+1} = f_i, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_{n-1}}{2h}\right) y_{n-2} + \left(q_{n-1} - \frac{2}{h^2}\right) y_{n-1} = -\left(\frac{1}{h^2} + \frac{p_{n-1}}{2h}\right) \gamma_2 + f_{n-1}. \end{cases} \quad (6.37)$$

Решение исходного уравнения (6.30) с краевыми условиями

$$y(a) = \gamma_1, y'(b) = \gamma_2$$

сводится к решению линейной алгебраической системы, которая тоже состоит из n уравнений, содержит n неизвестных:

$$\begin{cases} \left(q_1 - \frac{2}{h^2}\right)y_1 + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right)y_2 = -\left(\frac{1}{h^2} - \frac{p_1}{2h}\right)\gamma_1 + f_1, \\ \left(\frac{1}{h^2} - \frac{p_2}{2h}\right)y_1 + \left(q_2 - \frac{2}{h^2}\right)y_2 + \left(\frac{1}{h^2} + \frac{p_2}{2h}\right)y_3 = f_2, \\ \left(\frac{1}{h^2} - \frac{p_3}{2h}\right)y_2 + \left(q_3 - \frac{2}{h^2}\right)y_3 + \left(\frac{1}{h^2} + \frac{p_3}{2h}\right)y_4 = f_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_i}{2h}\right)y_{i-1} + \left(q_i - \frac{2}{h^2}\right)y_i + \left(\frac{1}{h^2} + \frac{p_i}{2h}\right)y_{i+1} = f_i, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \left(\frac{1}{h^2} - \frac{p_{n-1}}{2h}\right)y_{n-2} + \left(q_{n-1} - \frac{2}{h^2}\right)y_{n-1} + \left(\frac{1}{h^2} + \frac{p_{n-1}}{2h}\right)y_n = f_{n-1}. \\ -y_{n-1} + y_n = h \cdot \gamma_2. \end{cases} \quad (6.38)$$

Система (6.38) решается методом прогонки.

ТОЧНОСТЬ РАЗНОСТНОГО МЕТОДА

Доказано, что если в уравнении (6.30) функции $p(x)$, $q(x)$, $f(x)$ дважды непрерывно дифференцируемы, то разностное решение равномерно сходится к точному решению с погрешностью $O(h^2)$ при $h \to 0$.

В случае краевых условий 2-го и 3-го рода на границах отрезка $[a; b]$ можно использовать аппроксимации (6.24) и (6.25). Они имеют порядок точности $\mathrm{O}(h)$. Но точная аппроксимация только в одной или в двух точках понижает точность всего решения. Поэтому в граничных точках лучше построить более точное приближение первой производной с помощью ***метода фиктивных областей***. Опишем кратко идею этого простого метода.

Введём два фиктивных узла

$$x_{-1} = a - h = x_0 - h; \;\; x_{n+1} = b + h = x_n + h.$$

Выразим значения неизвестной функции $y(x)$ в двух фиктивных узлах через её значения в ближайших реальных узлах на $[a; b]$. Для этого используем центрально-разностную аппроксимацию первой производной, порядок точности которой $\mathrm{O}(h^2)$:

$$y'(a) = y'_0 = \gamma_1 = \frac{y_1 - y_{-1}}{2h} \Longrightarrow y_{-1} = y_1 - \gamma_1 2h, \quad (6.39)$$

$$y'(b) = y'_n = \gamma_2 = \frac{y_{n+1} - y_{n-1}}{2h} \Longrightarrow y_{n+1} = y_{n-1} + \gamma_2 2h. \quad (6.40)$$

Запишем разностную аппроксимацию (6.31) исходного дифференциального уравнения (6.30) в точке x_0 и в точке x_n:

$$\left(\frac{1}{h^2}-\frac{p_0}{2h}\right)y_{-1}+\left(q_0-\frac{2}{h^2}\right)y_0+\left(\frac{1}{h^2}+\frac{p_0}{2h}\right)y_1=f_0\,,$$

$$\left(\frac{1}{h^2}-\frac{p_n}{2h}\right)y_{n-1}+\left(q_n-\frac{2}{h^2}\right)y_n+\left(\frac{1}{h^2}+\frac{p_n}{2h}\right)y_{n+1}=f_n\,.$$

Сделаем в этих уравнениях подстановки (6.39) и (6.40):

$$\left(\frac{1}{h^2}-\frac{p_0}{2h}\right)(y_1-\gamma_1 2h)+\left(q_0-\frac{2}{h^2}\right)y_0+\left(\frac{1}{h^2}+\frac{p_0}{2h}\right)y_1=f_0\,,$$

$$\left(\frac{1}{h^2}-\frac{p_n}{2h}\right)y_{n-1}+\left(q_n-\frac{2}{h^2}\right)y_n+\left(\frac{1}{h^2}+\frac{p_n}{2h}\right)(y_{n-1}+\gamma_2 2h)=f_n\,.$$

Приведём подобные члены:

$$\left(q_0-\frac{2}{h^2}\right)y_0+\frac{2}{h^2}y_1=f_0+\gamma_1\left(\frac{2}{h}-p_0\right), \qquad (6.41)$$

$$\frac{2}{h^2}y_{n-1}+\left(q_n-\frac{2}{h^2}\right)y_n=f_n-\gamma_2\left(\frac{2}{h}+p_n\right). \qquad (6.42)$$

Построенные уравнения (6.41), (6.42) имеют порядок точности $O(h^2)$.

Для повышения точности решения уравнения (6.30) с граничными условиями 2-го рода надо первое уравнение системы (6.36) заменить на уравнение (6.41), а последнее уравнение заменить на уравнение (6.42). В случае граничных условий 3-го рода таким же образом меняется либо первое уравнение системы (6.37), либо последнее уравнение системы (6.38).

Метод фиктивных областей успешно используется в различных задачах для аппроксимации производных в граничных точках заданной области.

6.3.4. Методы аналитического приближения

Перейдём к методам аналитической аппроксимации решений краевой задачи для линейных дифференциальных уравнений второго порядка. Для этого напомним некоторые понятия из теории функциональных рядов.

Определение 9. Функции $\varphi_1(x)$ и $\varphi_2(x)$ называются ***линейно независимыми*** на $[a;b]$, если при любом числе $\gamma\in\mathbb{R}\setminus\{0\}$ верно неравенство

$$\varphi_1(x)\neq\gamma\varphi_2(x), \text{ или } \varphi_2(x)\neq\gamma\varphi_1(x).$$

Определение 10. Бесконечная система функций $\{\varphi_n(x)\}_{n=0}^{\infty}$ называется ***линейно независимой*** на отрезке $[a,b]$, если равенство

$$\sum_{n=0}^{\infty}c_n\,\varphi_n(x)=0$$

верно только в случае равенства нулю числовых коэффициентов $\{c_n\}_{n=0}^{\infty}$.

Функции, входящие в линейно независимую систему, называются ***базисными,*** или ***координатными*** функциями.

Теорема 4. Любую непрерывную на $[a; b]$ функцию $y(x)$ можно разложить на этом отрезке в равномерно сходящийся ряд по бесконечному набору линейно независимых функций:

$$y(x) = \sum_{n=0}^{\infty} c_n \varphi_n(x) = \sum_{n=0}^{M} c_n \varphi_n(x) + R_M.$$

Такую систему функций $\{\varphi_n(x)\}_{n=0}^{\infty}$ называют ***полной системой***. Для неё справедлива следующая теорема.

Теорема 5. Если для некоторой функции $\eta(x)$ и полной системы функций $\{\varphi_n(x)\}_{n=0}^{\infty}$ выполняется равенство

$$\int_a^b \eta(x)\varphi_n(x)dx = 0$$

для любой функции $\varphi_n(x)$ из полной системы, тогда функция $\eta(x) \equiv 0$ на всём отрезке $[a; b]$.

6.3.5. Метод Бубнова–Галёркина

Решение линейного дифференциального уравнения второго порядка

$$y''(x) + p(x)y'(x) + q(x)y(x) = f(x) \qquad (6.43)$$

с линейными краевыми условиями 1-го рода будем строить в виде

$$y(x) = y_M = \varphi_0(x) + \sum_{n=1}^{M} c_n \varphi_n(x) = \varphi_0 + \sum_{n=1}^{M} c_n \varphi_n. \qquad (6.44)$$

Запишем аналитическое представление 1-й и 2-й производных функции $y(x)$:

$$y'_M = \varphi'_0(x) + \sum_{n=1}^{M} c_n \varphi'_n(x) = \varphi'_0 + \sum_{n=1}^{M} c_n \varphi'_n, \qquad (6.45)$$

$$y''_M = \varphi''_0(x) + \sum_{n=1}^{M} c_n \varphi''_n(x) = \varphi''_0 + \sum_{n=1}^{M} c_n \varphi''_n. \qquad (6.46)$$

ПОСТРОЕНИЕ БАЗИСНЫХФУНКЦИЙ

Предположим, что $x \in [a; b]$, $y(a) = \gamma_1$, $y(b) = \gamma_2$. Функцию $\varphi_0(x)$ строим так, чтобы она была максимально проста и на границах области совпадала с заданными краевыми условиями. Очевидно, что в этом случае лучше всего взять линейную функцию. Используем уравнение прямой, проходящей через две заданные точки

$$\frac{x-a}{b-a}=\frac{\varphi_0-\gamma_1}{\gamma_2-\gamma_1}\Rightarrow\varphi_0=\gamma_1+\frac{(\gamma_2-\gamma_1)(x-a)}{b-a}\Rightarrow$$

$$\varphi'_0=\frac{\gamma_2-\gamma_1}{b-a}=c_0\Rightarrow\varphi''_0=0.$$

Остальные базисные функции строятся так, чтобы они были линейно независимы и равнялись нулю в граничных точках. В качестве примера можно взять функции

$$\varphi_1(x)=\varphi_1=(x-a)(x-b),\qquad \varphi_2(x)=\varphi_2=x(x-a)(x-b),$$

$$\varphi_3(x)=\varphi_2=x^2(x-a)(x-b),\qquad \varphi_4(x)=\varphi_4=x^3(x-a)(x-b),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\varphi_M(x)=\varphi_M=x^{M-1}(x-a)(x-b).$$

АЛГОРИТМ ПОСТРОЕНИЯ РЕШЕНИЯ

Количество базисных функций в аналитическом представлении решения (6.44) заранее неизвестно, оно зависит от заданной точности расчёта ε.

Процесс нахождения неизвестных коэффициентов $\{c_n\}_{n=1}^{M}$ состоит из 4-х шагов.

1. Равенства (6.44) – (6.46) подставляются в исходное уравнение (6.43):

$$\sum_{n=1}^{M}c_n\,\varphi''_n+p\left(c_0+\sum_{n=1}^{M}c_n\,\varphi'_n\right)+q\left(\varphi_0+\sum_{n=1}^{M}c_n\,\varphi_n\right)-f=\eta,\qquad(6.47)$$

где $p=p(x),\ q=q(x),\ f=f(x)$.

Параметр η– ***невязк*а**. Чем больше слагаемых в частичных суммах в (6.47), тем меньше значение невязки. Она равна нулю только при подстановке точного решения в исходное дифференциальное уравнение.

В уравнении (6.47) приведём подобные члены относительно неизвестных коэффициентов $\{c_n\}_{n=1}^{M}$:

$$c_1\left(\varphi_1''+p\varphi_1'+\varphi_1\right)+c_2\left(\varphi_2''+p\varphi_2'+\varphi_2\right)+\cdots$$

$$+c_M\left(\varphi_M''+p\varphi_M'+\varphi_M\right)+pc_0+q\varphi_0-f=\eta.\qquad(6.48)$$

2. С помощью уравнения (6.48) строим систему. Для этого умножим уравнение сначала на базисную функцию φ_1, затем на φ_2 и т.д. Последнее уравнение системы получается умножением (6.48) на φ_M.

Для компактной записи системы введём новые обозначения. Пусть

$$\varphi_n''+p\varphi_n'+\varphi_n=L_n,\ \text{где}\ n=1,2,\cdots,M.$$

Построенная система имеет следующий вид

$$\begin{cases} c_1L_1\varphi_1 + c_2L_2\varphi_1 + \cdots + c_ML_M\varphi_1 + (q\varphi_0 + pc_0 - f)\varphi_1 = \eta\varphi_1, \\ c_1L_1\varphi_2 + c_2L_2\varphi_2 + \cdots + c_ML_M\varphi_2 + (q\varphi_0 + pc_0 - f)\varphi_2 = \eta\varphi_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_1L_1\varphi_M + c_2L_2\varphi_M + \cdots + c_ML_M\varphi_M + (q\varphi_0 + pc_0 - f)\varphi_M = \eta\varphi_M. \end{cases} \quad (6.49)$$

Базисные функции $\{\varphi_n\}_{n=1}^{M}$ линейно независимы. Поэтому уравнения построенной системы (6.49) тоже линейно независимы.

3. Для превращения системы (6.49) в линейную алгебраическую систему проинтегрируем по области $[a; b]$ каждое уравнение системы:

$$\begin{cases} c_1\int\limits_a^b L_1\varphi_1\,dx + c_2\int\limits_a^b L_2\varphi_1\,dx + ..+c_M\int\limits_a^b L_M\varphi_1\,dx + \int\limits_a^b \big(q\varphi_0 + pc_0 - f\big)\varphi_1\,dx = \int\limits_a^b \eta\varphi_1\,dx, \\ c_1\int\limits_a^b L_1\varphi_2\,dx + c_2\int\limits_a^b L_2\varphi_2\,dx+..+c_M\int\limits_a^b L_M\varphi_2\,dx + \int\limits_a^b \big(q\varphi_0 + pc_0 - f\big)\varphi_2\,dx = \int\limits_a^b \eta\varphi_2\,dx, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_1\int\limits_a^b L_1\varphi_M\,dx + c_2\int\limits_a^b L_2\varphi_M\,dx+..+c_M\int\limits_a^b L_M\varphi_M\,dx + \int\limits_a^b \big(q\varphi_0 + pc_0 - f\big)\varphi_M\,dx = \int\limits_a^b \eta\varphi_M\,dx. \end{cases}$$

Значение невязки η достаточно мало. На основе теоремы 5 можно считать, что невязка ортогональна всем базисным функциям $\{\varphi_n\}_{n=1}^{M}$ на отрезке $[a; b]$:

$$\int\limits_a^b \eta\varphi_1\,dx = \int\limits_a^b \eta\varphi_2\,dx = \cdots = \int\limits_a^b \eta\varphi_M\,dx = 0 \Longrightarrow$$

$$\begin{cases} c_1\int\limits_a^b L_1\varphi_1\,dx + c_2\int\limits_a^b L_2\varphi_1\,dx + ..+c_M\int\limits_a^b L_M\varphi_1\,dx = \int\limits_a^b (f - q\varphi_0 - pc_0)\varphi_1\,dx, \\ c_1\int\limits_a^b L_1\varphi_2\,dx + c_2\int\limits_a^b L_2\varphi_2\,dx+..+c_M\int\limits_a^b L_M\varphi_2\,dx = \int\limits_a^b (f - q\varphi_0 - pc_0)\varphi_2\,dx, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \\ c_1\int\limits_a^b L_1\varphi_M\,dx + c_2\int\limits_a^b L_2\varphi_M\,dx+..+c_M\int\limits_a^b L_M\varphi_M\,dx = \int\limits_a^b (f - q\varphi_0 - pc_0)\varphi_M\,dx. \end{cases} \quad (6.50)$$

4. Определитель построенной системы не равен нулю, так как все её уравнения линейно независимы. Следовательно, существует единственное решение. Наши базисные функции – многочлены. Входящие в исходное уравнение (6.43) заданные функции $p(x)$, $q(x)$, $f(x)$ обычно аналитически очень просты. Поэтому все интегралы системы можно вычислить точно (в квадратурах).

После нахождения коэффициентов $\{c_n(x)\}_{n=1}^{M}$ приближённое аналитическое решение дифференциального уравнения (6.43) построено полностью:

$$y(x) = \varphi_0(x) + \sum_{n=1}^{M} c_n \varphi_n(x).$$

ОЦЕНКА ТОЧНОСТИ МЕТОДА БУБНОВА–ГАЛЁРКИНА

Количество базисных функций в представлении решения (6.44) заранее неизвестно Оно зависит от заданной точности расчёта ε. Поэтому сначала ограничиваются числом $M = 2$. По описанному выше алгоритму строится первое приближение

$$y_M(x) = \varphi_0(x) + \sum_{n=1}^{M} c_n \varphi_n(x). \qquad (6.51)$$

Далее добавляется ещё одна базисная функция $(M = 3)$ и по такому же алгоритму строится следующее приближение

$$y_{M+1}(x) = \varphi_0(x) + \sum_{n=1}^{M+1} \tilde{c}_n \varphi_n(x). \qquad (6.52)$$

В области $[a; b]$ выбираются точки $\{x_i\}_{i=1}^{K}$, в которых сравниваются значения полученных приближений:

$$\left|\frac{y_{M+1}(x_i) - y_M(x_i)}{y_M(x_i)}\right| < \varepsilon. \qquad (6.53)$$

Если неравенство (6.53) выполняется во всех точках $\{x_i\}_{i=1}^{K}$, тогда в качестве решения исходной краевой задачи выбирается приближение (6.51). Если неравенство не выполняется, хотя бы в одной точке x_i, тогда весь алгоритм повторяется. Перед повтором алгоритма параметр M увеличивают на единицу.

Процесс заканчивается, когда на $(L + 1)$ -м повторении алгоритма во всех точках $\{x_i\}_{i=1}^{K}$ выполняется неравенство

$$\left|\frac{y_{M+1+L}(x_i) - y_{M+L}(x_i)}{y_{M+L}(x_i)}\right| < \varepsilon.$$

Окончательное приближение к точному решению будет иметь вид

$$y_{M+L}(x) = \varphi_0(x) + \sum_{n=1}^{M+L} c_n \varphi_n(x).$$

ОСОБЕННОСТИ МЕТОДА БУБНОВА–ГАЛЁРКИНА

1. Решение задачи очень чувствительно к тому, насколько удачно выбрана система базисных функций для заданной краевой задачи.

2. При добавлении базисных функций в следующем приближении появляются новые коэффициенты c_n и меняются старые. Если в качестве базисных функций выбрана ортогональная система, тогда найденные на предыдущем приближении коэффициенты c_n не будут меняться с увеличением слагаемых. Поэтому ортогональные базисные функции обычно удобнее неортогональных.

6.3.6. Метод Ритца

С помощью этого метода можно найти приближённое решение уравнения в аналитическом виде. Основой метода является раздел высшей математики «Вариационное исчисление». Исторические сведения о нём можно прочесть в приложении 1.Здесь ограничимся только основными понятиями и свойствами.

Базовым понятием этого раздела математики является функционал.

Определение 11(а). ***Функционалом*** называется взаимно однозначное соответствие между функциональным множеством и числовым множеством.

Можно дать другое определение.

Определение 11(б). ***Функционал*** – это оператор, отображающий пространство функций в числовое множество.

Примером функционала является определённый интеграл, подынтегральное выражение которого зависит от некоторой функции $y(x)$ с заданными свойствами или от функции и её производных.

Определение 12. Функция, при которой функционал достигает своего экстремального значения в заданной области, называется ***экстремалью***.

«Вариационное исчисление» посвящёно исследованию функционалов на экстремум и построению общих методов нахождения экстремалей.

Численные методы, которые строят приближённое решение вариационных задач в явном аналитическом виде, называют ***прямыми методами***. Развитие таких методов дало возможность находить экстремали с любой степенью точности. Первым из таких методов стал метод Ритца (1908).

УРАВНЕНИЕ ЭЙЛЕРА

Дифференциальные уравнения и функционалы тесно связаны между собой. В качестве примера приведём фундаментальную теорему для функционала, который часто встречается в различных приложениях.

Теорема 6. Функционал

$$I(y) = \int_a^b F(x, y, y')dx \qquad (6.54)$$

с дважды непрерывно дифференцируемой по всем своим переменным подынтегральной функцией $F(x, y, y')$ имеет экстремаль $y(x)$ с заданными краевыми условиями 1-го рода только в том случае, когда функция $y(x)$ – решение уравнения

$$F''_{y'y'} \cdot y'' + F''_{y'y} \cdot y' + F''_{y'x} - F'_y = 0 \qquad (6.55)$$

с теми же краевыми условиями.

Дифференциальное уравнение (6.55) называется ***уравнением Эйлера.*** Его можно записать в более компактной форме

$$(F(x, y, y'))'_{y'} - F'_y = 0. \qquad (6.56)$$

С помощью этой теоремы можно построить решение уравнения (6.56). Для этого достаточно найти экстремаль соответствующего функционала. Если экстремаль не находится точно, её можно построить приближённо методом Ритца.

ПОСТРОЕНИЕ УРАВНЕНИЯ ПО ФУНКЦИОНАЛУ

В качестве примера функционала (6.54) используем функционал

$$I(y) = \int_a^b (\, p(x)(y')^2 + q(x)y^2 - 2f(x)y)dx \qquad (6.57)$$

с заданными краевыми условиями $y(a) = \gamma_1,\ y(b) = \gamma_2$.

Построим соответствующее уравнение Эйлера. Для этого возьмём частные производные:

$$F''_{y'y'} = 2p(x), \qquad F''_{y'y} = 0, \qquad F''_{y'x} = 2p'(x)y', \qquad F'_y = 2q(x)y - 2f(x)$$

и подставим их в уравнение Эйлера (6.55):

$$2p(x)y'' + 2p'(x)y' - 2q(x)y + 2f(x) = 0 \Longrightarrow$$

$$-\Big(p(x)y'(x)\Big)'_x + q(x)y(x) - f(x) = 0 \Longrightarrow$$

$$-\big(p(x)y'(x)\big)'_x + q(x)y(x) = f(x). \qquad (6.58)$$

Уравнение (6.58) – уравнение Эйлера для функционала (6.57).

Из теоремы 6 следует, что решение уравнения (6.58) с заданными краевыми условиями совпадает с экстремалью функционала (6.57) с теми же краевыми условиями.

Прежде чем приступить к описанию метода Ритца, покажем, как любое линейное дифференциальное уравнение второго порядка с переменными коэффициентами можно преобразовать в уравнение Эйлера.

ПЕРЕХОД К УРАВНЕНИЮ ЭЙЛЕРА

Предположим, что дифференциальное уравнение задано в виде

$$y''(x) + g(x)y'(x) + z(x)y(x) = f(x). \qquad (6.59)$$

Функции $g(x)$, $z(x)$ и $f(x)$ непрерывны, а $g(x)$ – дифференцируема на отрезке $[a; b]$. Умножим уравнение (6.59) на положительную функцию

$$p(x) = e^{\int_a^x g(t)dt} \Longrightarrow p'(x) = p(x)g(x) \Longrightarrow$$

$$p(x)y''(x) + p(x)g(x)y'(x) + p(x)z(x)y(x) = p(x)f(x).$$

Сумма двух первых слагаемых этого уравнении является производной от произведения $y'(x)p(x)$:

$$p(x)y''(x) + p(x)g(x)y'(x) = \Big(p(x)y'(x)\Big)'_x \Longrightarrow$$

$$\Big(p(x)y'(x)\Big)'_x + p(x)z(x)y(x) = p(x)f(x).$$

Умножим последнее уравнение на (–1) и введём новые обозначения

$$u(x) = -p(x)z(x), \qquad w(x) = -p(x)f(x).$$

Теперь преобразованное уравнение (6.59) стало уравнением Эйлера

$$-\big(p(x)y'(x)\big)'_x + u(x)y(x) = w(x). \qquad (6.60)$$

РЕШЕНИЕ УРАВНЕНИЯ МЕТОДОМ РИТЦА

1. Решение уравнения (6.58) строится в виде обобщённого многочлена

$$y_M(x) = \varphi_0(x) + \sum_{n=1}^{M} c_n\,\varphi_n(x) \qquad (6.61)$$

по системе линейно независимых на $[a; b]$ функций $\{\varphi_n(x)\}_{n=0}^{M}$. Базисные функции в этой системе можно строить также как в методе Бубнова–Галёркина:

$$\varphi_0(x) = \gamma_1 + \frac{(\gamma_2 - \gamma_1)(x-a)}{b-a},$$

$$\varphi_1(x) = (x-a)(x-b), \qquad \varphi_2(x) = x(x-a)(x-b),$$

$$\varphi_3(x) = x^2(x-a)(x-b), \quad \varphi_4(x) = x^3(x-a)(x-b),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\varphi_M(x) = x^{M-1}(x-a)(x-b).$$

2. Для нахождения неизвестных коэффициентов $\{c_n\}_{n=1}^{M}$ заменяем в функционале (6.57) неизвестную функцию на её приближение (6.61):

$$I(y) = \int_a^b (p(x)(y')^2 + q(x)y^2 - 2f(x)y)dx \Longrightarrow$$

$$I(y_M) = \int_a^b \left(p\left(\varphi'_0 + \sum_{n=1}^{M} c_n \varphi'_n\right)^2 + q\left(\varphi_0 + \sum_{n=1}^{M} c_n \varphi_n\right)^2 - 2f\left(\varphi_0 + \sum_{n=1}^{M} c_n \varphi_n\right)\right) dx, \qquad (6.62)$$

где $p = p(x)$, $q = q(x)$, $f = f(x)$, $\varphi_0 = \varphi_0(x)$, $\varphi_n = \varphi_n(x)$.

3. Правая часть равенства (6.62) зависит от значений $\{c_n\}_{n=1}^{M}$. Поэтому функционал $I(y_M)$ становится функцией, которая тоже зависит от этих коэффициентов:

$$I(y_M) = I(c_1, c_2, \cdots, c_{M-1}, c_M). \qquad (6.63)$$

Функция (6.63) дифференцируема по каждой переменной c_n, следовательно, для неё справедливо необходимое условие существования экстремума. В случае функции нескольких переменных это условие представляется системой

$$\begin{cases} \dfrac{\partial I}{\partial c_1} = 0, \\ \dfrac{\partial I}{\partial c_2} = 0, \\ \ldots\ldots\ldots \\ \dfrac{\partial I}{\partial c_k} = 0, \\ \ldots\ldots\ldots \\ \dfrac{\partial I}{\partial c_M} = 0. \end{cases} \qquad (6.64)$$

В каждом уравнении этой системы приведём подобные члены относительно $\{c_n\}_{n=1}^{M}$. После этого любое уравнение системы (6.64), например, уравнение с номером k имеет следующий вид

$$\int_a^b \left(p\left(\varphi'_0 + \sum_{n=1}^{M} c_n \varphi'_n \right) \varphi'_k + q\left(\varphi_0 + \sum_{n=1}^{M} c_n \varphi_n \right) \varphi_k - f\varphi_k \right) dx = 0 \Longrightarrow$$

$$c_1 \int_a^b \left(p\varphi'_1 \varphi'_k + q\varphi_1 \varphi_k \right) dx + c_2 \int_a^b \left(p\varphi'_2 \varphi'_k + q\varphi_2 \varphi_k \right) dx + \cdots +$$

$$+ c_M \int_a^b \left(p\varphi'_M \varphi'_k + q\varphi_M \varphi_k \right) dx = \int_a^b \left(f\varphi_k - p\varphi'_0 - q\varphi_0 \right) dx \qquad (6.65)$$

Из уравнения (6.65) видно, что система (6.64) является линейной системой относительно коэффициентов $\{c_n\}_{n=1}^{M}$. Легко проверить, что матрица этой системы будет симметричной.

В рамках вариационного исчисления доказано, что если функции

$$p(x) > 0 \quad \text{и} \quad q(x) \geq 0$$

и имеют место краевые условия первого рода, тогда существует единственное решение системы (6.64).

После решения системы становится известным приближённое решение уравнения (6.58). Построенное приближенное решение (6.61) сходится к точному решению $y(x)$:

$$\lim_{M \to +\infty} \left(\varphi_0(x) + \sum_{n=1}^{M} c_n \varphi_n(x) \right) = y(x).$$

ОЦЕНКА ТОЧНОСТИ МЕТОДА РИТЦА

Системы с симметричными матрицами лучше решать методом квадратного корня. Этот метод содержит вдвое меньше арифметических действий и использует вдвое меньше оперативной памяти, чем метод Гаусса, поэтому он даёт более точные результаты за меньшее время расчёта.

Алгоритм нахождения оптимального количества базисных функций в методе Ритца аналогичен соответствующему алгоритму в методе Бубнова–Галёркина.

6.3.7. Метод конечных элементов

Во второй половине XX века был разработан метод конечных элементов (МКЭ). Он повысил точность метода Ритца. В рамках МКЭ строятся специальные базисные функции – ***финитные функции***, которые не только имеют простейший аналитический вид, но и приводят к системе (6.64) с трёхдиагональной матрицей.

ФИНИТНЫЕ ФУНКЦИИ

Опишем процесс построения финитных функций на простейшем примере. Предположим, что задача решается с нулевыми (однородными) краевыми условиями 1-го рода на отрезке $[0;1]$. Разобьём отрезок на M равных частей с шагом $h=\frac{1}{M}$. Каждый узел разбиения вычисляется по формуле

$$x_n = nh, \text{ где } n = 0,1,2,\cdots,M.$$

Определим систему финитных функций $\{\varphi_n(x)\}_{n=1}^{M}$ на отрезке $[0;1]$

$$\varphi_n(x)=\begin{cases} 0, \text{ если } 0 \le x \le x_{n-1}, \\ \frac{x-x_{n-1}}{h}, \text{ если } x_{n-1} < x \le x_n, \\ \frac{x_{n+1}-x}{h}, \text{ если } x_n < x \le x_{n+1}, \\ 0, \text{ если } x_{n+1} < x \le 1. \end{cases} \quad (n=1,2,\cdots,M-1) \qquad (6.66)$$

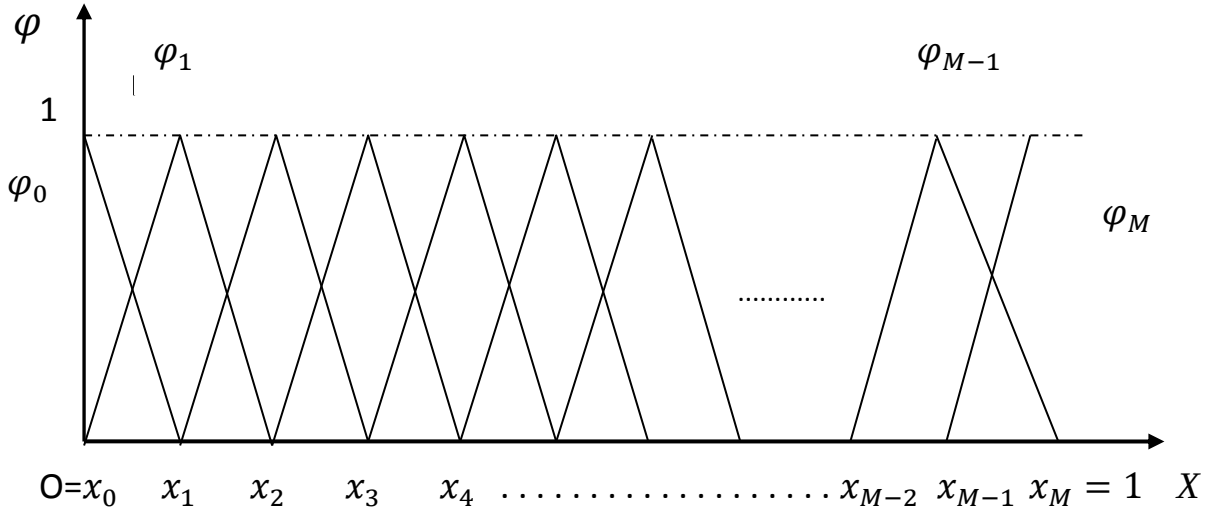

Рис. 6.4.

Каждая $\varphi_n(x)$ – непрерывная, кусочно-линейная функция и не равна нулю только в области $[x_{n-1};x_{n+1}]$. В МКЭ эти области называют ***конечными элементами.*** На рис. 6.4. представлены эскизы графиков таких функций.

С помощью интегрирования простейших степенных функций и свойств интегралов просто доказать, что система финитных функций $\{\varphi_n(x)\}_{n=1}^{M}$ ***почти ортогональна*:**

$$(\varphi_n,\varphi_j)=\int_0^1 \varphi_n(x),\varphi_j(x)dx,$$

$$(\varphi_n, \varphi_j) = \begin{cases} 0, & n \le j-2 \\ h/6, & n = j-1, \\ 2h/3, & n = j, \\ h/6, & n = j+1, \\ 0, & n \ge j+2. \end{cases} \qquad (6.67)$$

Из соотношения (6.67) видно, что любая финитная функция $\varphi_n(x)$ системы ортогональна всем функциям той же системы кроме двух соседних функций $\varphi_{n-1}(x)$ и $\varphi_{n+1}(x)$.

Производные финитных функций имеют простейший вид:

$$\frac{d\varphi_n(x)}{dx} = \begin{cases} 0, \ 0 \le x \le x_{n-1}, \\ \frac{1}{h}, \quad x_{n-1} \le x \le x_n, \\ \frac{-1}{h}, \quad x_n \le x \le x_{n+1}, \\ 0, \ x_{n+1} \le x \le 1. \end{cases} \qquad (6.68)$$

Вернёмся к решению уравнения Эйлера (6.58) с нулевыми (однородными) краевыми условиями 1-го рода на отрезке $[0; 1]$. Решение будем искать в виде обобщённого многочлена

$$y(x) = \sum_{n=0}^{M} c_n \varphi_n(x), \qquad (6.68)$$

где $\varphi_n(x)$ – финитные функции (6.66), а неизвестные коэффициенты c_n вычисляются методом Ритца.

Интересно отметить, что в каждом узле x_n сетки финитная функция $\varphi_n(x_n) = 1$, а в остальных узлах она равна нулю. Из равенства (6.69), следует, что

$$y(x_n) = y_n = c_n.$$

Коэффициенты $c_0 = c_M = 0$ из-за нулевых краевых условий. Остальные коэффициенты $\{c_n\}_{n=1}^{M-1}$ находятся из линейной системы (6.64). Благодаря почти ортогональности системы финитных функций и структуре каждого уравнения системы, матрица системы (6.64) становится трёхдиагональной.

ОЦЕНКА ТОЧНОСТИ МКЭ

Чрезвычайная простота аналитического вида финитных функций (6.66) и их производных (6.68) существенно упрощает каждое уравнение системы. Поэтому МКЭ имеет минимальные вычислительные погрешности при решении дифференциального уравнения (6.58).

Оптимальный выбор количества финитных функций строится по такому же принципу, как в методах Бубнова–Галёркина и Ритца.

Глава 7

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Математика, как и искусство, –
это особый способ познания.
В. Успенский. «Апология математики»

Уравнением в частных производных называется уравнение, слагаемые которого содержат частные производные неизвестной функции нескольких переменных. Сама неизвестная функция может не входить в уравнение.

К уравнениям в частных производных приводят многие задачи газодинамики, теплопроводности, электромагнитных полей, процессов переноса в газах и др. Поэтому такие уравнения часто называют уравнениями математической физики.

7.1. Начальные понятия

Уравнение в частных производных в общем случае имеет бесчисленное множество решений. Для выделения из этого множества единственного решения необходимо ввести дополнительные условия.

Обычно в физических задачах в качестве независимых переменных выступают время и пространственные координаты. Для переменной по времени вводятся ***начальные условия.*** Если область, в которой строится решение, имеет границу, то задаются ***краевые условия***.

Определение 1. Порядком уравнения в частных производных называется порядок старшей частной производной, входящей в уравнение.

Для единственности решения поставленной задачи необходимо, чтобы количество начальных условий равнялось порядку уравнения.

Определение 2. Уравнение называется линейным, если оно линейно относительно неизвестной функции и её частных производных.

Определение 3. Линейное уравнение называется однородным, если все слагаемые уравнения имеют первый порядок степени относительно искомой функции и её частных производных.

Из курса математической физики известно, что линейные уравнения второго порядка делятся на три типа.

УРАВНЕНИЯ ПАРАБОЛИЧЕСКОГО ТИПА

Примеры:

– однородное одномерное уравнение теплопроводности

$$\frac{\partial u}{\partial t} = k^2 \frac{\partial^2 u}{\partial x^2};$$

– неоднородное одномерное уравнение теплопроводности

$$\frac{\partial u}{\partial t} = k^2 \frac{\partial^2 u}{\partial x^2} + f(x,t);$$

– неоднородное двумерное уравнение теплопроводности

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f(x, y, t).$$

УРАВНЕНИЯ ГИПЕРБОЛИЧЕСКОГО ТИПА

Примеры:

– однородное одномерное волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} = k^2 \frac{\partial^2 u}{\partial x^2};$$

– неоднородное одномерное волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} = k^2 \frac{\partial^2 u}{\partial x^2} + f(x, t);$$

– неоднородные двумерное и трехмерное волновые уравнения

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f(x, y, t),$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + f(x, y, z, t).$$

УРАВНЕНИЯ ЭЛЛИПТИЧЕСКОГО ТИПА

Примеры:

– уравнения Лапласа

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

– уравнения Пуассона

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y, t), \qquad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = f(x, y, z, t).$$

В дальнейшем будем разбирать методы решения только линейных уравнений второго порядка.

КРАЕВЫЕ И НАЧАЛЬНЫЕ УСЛОВИЯ

Рассмотрим линейные краевые и начальные условия для одномерных уравнений параболического или гиперболического типа. Пусть искомая функция $u(x, t)$ находится в области $D: \{x \epsilon [a; b], \ t \epsilon [0; \mathrm{T}]\}$.

Краевые условия первого рода – это аналитически заданные законы изменения функции $u(x, t)$ в граничных точках $x = a$, $x = b$ на всём временном интервале:

$$u(a, t) = \gamma_1(t), \qquad u(b, t) = \gamma_2(t). \qquad (7.1)$$

Краевые условия второго рода – это заданные законы изменения частной производной $\frac{\partial u}{\partial x}$ в граничных точках в течение всего временного интервала:

$$\left.\frac{\partial u}{\partial x}\right|_{x=a} = \gamma_1(t), \quad \left.\frac{\partial u}{\partial x}\right|_{x=b} = \gamma_2(t). \qquad (7.2)$$

Краевые условия третьего рода (смешанные краевые условия) – это краевые условия одного из двух видов:

$$u(a,t) = \gamma_1(t), \qquad \left.\frac{\partial u}{\partial x}\right|_{x=b} = \gamma_2(t)$$

или (7.3)

$$\left.\frac{\partial u}{\partial x}\right|_{x=a} = \gamma_1(t), \qquad u(b,t) = \gamma_2(t).$$

Линейные краевые условия называются ***однородными условиями***, если

$$\gamma_1(t) = \gamma_2(t) = 0.$$

Во многих физических задачах в качестве одной из независимых переменных выступает время. Поэтому начальные условия – это известный закон изменения искомой функции в начальный момент времени $(t = 0)$. Для уравнения первого порядка необходимо и достаточно одного начального условия

$$u(x,0) = \varphi(x).$$

Для уравнения второго порядка надо знать начальное поведение функции и её производной по времени в области $[a;b]$:

$$u(x,0) = \varphi(x), \qquad \left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_1(x),$$

где $\varphi(x)$, $\varphi_1(x)$ – заданные непрерывные функции.

Для краевых и начальных условий обязательны ***условия согласования***. Это означает, что если в начальный момент времени функция $u(x,0) = \varphi(x)$, то в точках плоскости (рис.7.1) с координатами $(a;0)$ и $(b;0)$ должны выполняться соответствующие равенства

$$\gamma_1(0) = \varphi(a), \qquad \gamma_2(0) = \varphi(b). \qquad (7.4)$$

7.2. Сеточный метод

Идея метода сеток принадлежит Эйлеру. Но в те времена реализация этой идеи была невозможна в полной мере из-за отсутствия вычислительной техники. В настоящее время сеточный метод стал одним из наиболее эффективных методов приближённого решения уравнений в частных производных.

В методе сеток непрерывную область, в которой решается задача, прежде всего, заменяют дискретным множеством узлов. Производные, входящие в уравнение, заменяются их разностными аппроксимациями.

Все начальные и граничные условия, а также заданные в уравнении функции преобразуют к дискретному виду. Дифференциальное уравнение превращается в разностное уравнение, а затем в систему линейных уравнений. Результат решения уравнения в частных производных имеет табличный вид.

Перейдём к более детальному описанию метода сеток.

АППРОКСИМАЦИЯ ЧАСТНЫХ ПРОИЗВОДНЫХ

Рассмотрим дважды дифференцируемую функцию двух переменных $u(x,t)$ в прямоугольной замкнутой области D (рис 7.1).

$$D: \{x\epsilon\ [a;b]\ , t\epsilon[0;t_m]\ \}\ .$$

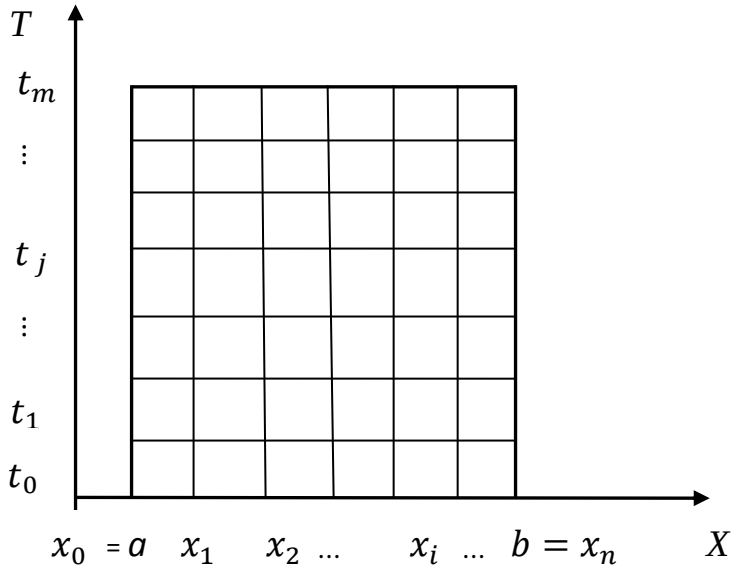

Рис. 7.1.

Покроем эту область ***прямоугольной сеткой*** (рис 7.1). Для простоты используем сетку с постоянными шагами разбиения по каждой переменной:

$$h = h_x = x_{i+1} - x_i,\ i = 0,1,\cdots,n-1;\ \ \tau = \tau_t = t_{j+1} - t_j,\ j = 0,1,\cdots,m-1.$$

У всех узлов построенной сетки есть двойная индексация (i,j). Узлы сетки, попавшие на границу области, называются граничными узлами. Остальные узлы – внутренние. Каждую горизонтальную линию сетки будем называть ***временным слоем.***

Аппроксимируем частные производные в любом внутреннем узле сетки. Для этого вспомним, что при дифференцировании функции по одной переменной другая переменная в этот момент становится константой. Для построения интересующих нас формул используем правостороннюю аппроксимацию первой производной и центрально-разностную аппроксимацию второй производной функции $y(x)$ в любом внутреннем узле разбиения области изменения переменной x:

$$y_i' = \frac{y_{i+1} - y_i}{h}; \qquad y_i'' = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}.$$

Порядки точности этих аппроксимаций (глава 6) равны соответственно

$$\mathrm{O}(h) \quad \text{и} \quad \mathrm{O}(h^2).$$

Очевидно, что приближение первой частной производной по x или t в любом внутреннем узле сетки (рис 7.1) можно представить в виде правосторонней аппроксимации:

$$\left.\frac{\partial u}{\partial x}\right|_{(i,j)} = \frac{u_{i+1,j} - u_{i,j}}{h}, \qquad \left.\frac{\partial u}{\partial t}\right|_{(i,j)} = \frac{u_{i,j+1} - u_{i,j}}{\tau}. \qquad (7.5)$$

Приближение вторых частных производных по x или t в любом внутреннем узле сетки можно представить в виде центрально-разностной аппроксимации:

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{(i,j)} = \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2}, \quad \left.\frac{\partial^2 u}{\partial t^2}\right|_{(i,j)} = \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{\tau^2}. \quad (7.6)$$

7.3. Одномерное уравнение теплопроводности

Обратимся к решению одномерного уравнения теплопроводности с заданными начальным и краевыми условиями:

$$\frac{\partial u}{\partial t} = k^2 \frac{\partial^2 u}{\partial x^2} + f(x,t). \qquad (7.7)$$

Неизвестная функция $u = u(x,t)$ – это температура любой точки стержня, абсцисса которой равна $x \in [a;b]$. Задана постоянная k^2 – коэффициент теплопроводности и функция $f(x,t)$. Если в стержне нет источника тепла, то $f(x,t) = 0$. В этом случае уравнение теплопроводности является однородным.

Предположим, что стержень ограничен, однороден и такой тонкий, что его толщиной можно пренебречь. Начальное условие задаётся распределением тепла $\varphi(x)$ в стержне в нулевой момент времени:

$$u(x,0) = \varphi(x). \qquad (7.8)$$

В качестве краевых условий рассмотрим условия первого рода:

$$u(a,t) = \gamma_1(t), \quad (b,t) = \gamma_2(t).$$

Значения функции $u(x,t)$ в граничных узлах и в узлах на нулевом временном слое элементарно вычисляются с помощью начального и краевых условий.

Сеточный метод решения уравнения (7.7) строит решение $u(x,t)$ во всех узлах сетки. Поэтому решение будет представлено в табличном виде:

$$u(x_i, t_j) = u_{i,j}, \text{ где } \quad i = 0,1,\cdots,n; \quad j = 0,1,\cdots,m.$$

Пусть $f(x_i, t_j) = f_{i,j}$. Для непосредственной реализации сеточного метода заменим в уравнении (7.7) производные на их аппроксимации:

$$\frac{u_{i,j+1} - u_{i,j}}{\tau} = k^2 \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + f_{i,j}. \qquad (7.9)$$

Полученное уравнение называется разностным уравнением. Это алгебраическое уравнение, связывающее между собой четыре значения неизвестной функции в четырёх узлах сетки. На основе этого разностного уравнения разработаны две схемы численного решения уравнения. Одна схема называется явной, другая неявной.

ЯВНАЯ СХЕМА

Из уравнения (7.9) выразим явно $u_{i,j+1}$:

$$u_{i,j+1} = u_{i,j} + \frac{k^2\tau}{h^2}\left(u_{i-1,j} - 2u_{i,j} + u_{i+1,j}\right) + \tau f_{i,j}. \qquad (7.10)$$

В этом уравнении значение неизвестной функции в любом внутреннем узле на последующем временном слое зависит от значений этой же функции в трёх различных узлах на предыдущем временном слое. Для наглядности изобразим эту зависимость в виде ***шаблона явной схемы для уравнения теплопроводности*** (рис 7.2).

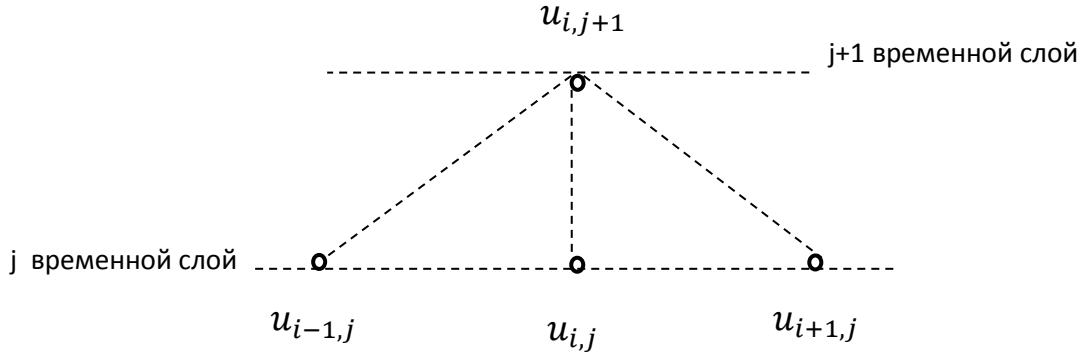

Рис. 7.2

АЛГОРИТМ РЕШЕНИЯ ПО ЯВНОЙ СХЕМЕ

Приведём простейший алгоритм построения значений функции $u(x,t)$ во всех внутренних узлах сетки.

1. Задаём нулевое значение индексу j $(j = 0)$. В этом случае все величины $u_{i-1,0}$, $u_{i,0}$, $u_{i+1,0}$ известны из начальных условий при любых значениях индекса $\{i\}_{i=1}^{n-1}$. По формуле (7.10) вычисляем все значения $u_{i,1}$ на первом временном слое.

2. Задаём следующее значение индексу j $(j = 1)$. Все значения функции $u(x,t)$ уже вычислены во всех внутренних узлах сетки на предыдущем шаге

алгоритма. Величины $u_{0,1}$ и $u_{n,1}$ известны из краевых условий. Поэтому по формуле (7.10) легко вычислить все значения $u_{i,2}$ на 2-м временном слое.

3. Пусть $j = 2$. Значения $u_{i,2}$ уже вычислены для $\{i\}_{i=1}^{n-1}$. Значения $u_{0,2}$ и $u_{n,2}$ известны из краевых условий. Вновь по формуле (7.10) находятся все значения $u_{i,3}$ на 3-м временном слое.

……

M. Описанный процесс заканчивается после вычисления всех значений $u_{i,m}$ на последнем временном слое $(j = m)$.

ОЦЕНКА КАЧЕСТВА ЯВНОЙ СХЕМЫ

Численная схема называется ***устойчивой схемой***, если решение разностного уравнения мало меняется при малом изменении всех исходных данных задачи. Это значит, что решение должно непрерывно зависеть от исходных данных при уменьшении размеров всех ячеек сетки. Доказано, что в этом случае численная схема сходится к точному решению.

При исследовании явной схемы на устойчивость было обнаружено, что она не является устойчивой. Но простота формулы (7.10) и алгоритма сделали явную схему настолько привлекательной, что четырём математикам удалось найти достаточное условие устойчивости явной схемы. Не удивительно, что это условие называют в честь разных математиков: условие Куранта, условие Куранта-Леви, условие Фридрихса.

Доказано, что явная схема будет устойчивой, если шаг по времени τ и шаг по пространству h будут связаны условием

$$\frac{k^2\tau}{h^2} \leq \frac{1}{2}. \qquad (7.11)$$

После вывода условия (7.11) явную схему стали называть ***условно устойчивой***. Если предположить, что параметр $k^2 = 1$, то, согласно неравенству (7.11), для устойчивости схемы шаг по времени должен быть на порядок меньше шага по пространству:

$$\tau \leq \frac{h^2}{2}.$$

Такая зависимость между шагами – существенный недостаток явной схемы. Поэтому была создана другая схема сеточного метода.

НЕЯВНАЯ СХЕМА

Основой ***неявной схемы*** тоже является разностное уравнение (7.9). Но аппроксимация первой производной по времени будет выражена через предыдущий временной слой:

$$\frac{u_{i,j} - u_{i,j-1}}{\tau} = k^2 \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + f_{i,j}. \qquad (7.12)$$

Из уравнения (7.12) выражаем значения функции $u(x,t)$ на временном слое j через её значение на временном слое $j-1$:

$$\frac{k^2}{h^2} u_{i-1,j} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right) u_{i,j} + \frac{k^2}{h^2} u_{i+1,j} = \frac{u_{i,j-1}}{\tau} + f_{i,j}. \quad (7.13)$$

Полученное разностное уравнение называется неявной схемой.

Представим связь значений функции $u(x,t)$ в виде ***шаблона неявной схемы для уравнения теплопроводности*** (рис 7.3).

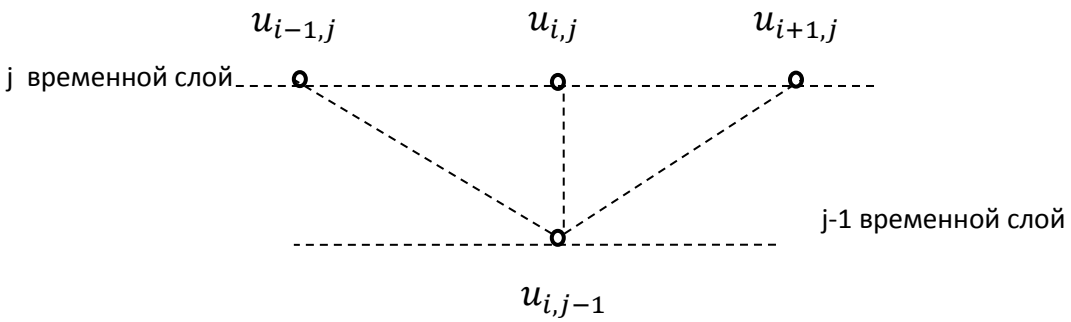

Рис 7.3.

АЛГОРИТМ РЕШЕНИЯ ПО НЕЯВНОЙ СХЕМЕ

Решение начинается с первого временного слоя ($j = 1$). Строится линейная алгебраическая система. Опишем процесс построения этой системы.

1. Для первых трёх узлов сетки $(i = 0, 1, 2)$ схема (7.13) имеет вид

$$\frac{k^2}{h^2} u_{0,1} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right) u_{1,1} + \frac{k^2}{h^2} u_{2,1} = \frac{u_{1,0}}{\tau} + f_{1,1}. \qquad (7.14)$$

В этом уравнении величина $u_{0,1}$ известна из краевых условий. Поэтому в уравнении присутствуют только два неизвестных числа $u_{1,1}$ и $u_{2,1}$ и его лучше представить в виде

$$-\left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right) u_{1,1} + \frac{k^2}{h^2} u_{2,1} = \frac{u_{1,0}}{\tau} + f_{1,1} - \frac{k^2}{h^2} u_{0,1}. \qquad (7.15)$$

Запишем уравнение (7.14) для узлов сетки с номерами $(i = 1, 2, 3)$

$$\frac{k^2}{h^2} u_{1,1} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right) u_{2,1} + \frac{k^2}{h^2} u_{3,1} = \frac{u_{2,0}}{\tau} + f_{2,1}.$$

Очевидно, что следующее уравнение будет иметь вид

$$\frac{k^2}{h^2}u_{2,1} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right)u_{3,1} + \frac{k^2}{h^2}u_{4,1} = \frac{u_{3,0}}{\tau} + f_{3,1}.$$

……

Последнее уравнение строится для узлов с номерами $(i = n-2, n-1, n)$:

$$\frac{k^2}{h^2}u_{n-2,1} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right)u_{n-1,1} + \frac{k^2}{h^2}u_{n,1} = \frac{u_{n-1,0}}{\tau} + f_{n-1,1}. \qquad (7.16)$$

Величина $u_{n,1}$ известна из краевых условий, поэтому перепишем уравнение (7.16) в виде

$$\frac{k^2}{h^2}u_{n-2,1} - \left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right)u_{n-1,1} = \frac{u_{n,0}}{\tau} + f_{n-1,1} - \frac{k^2}{h^2}u_{n,1}. \qquad (7.17)$$

В итоге построена линейная система из $(n-1)$-го уравнения с таким же числом неизвестных $u_{i,1}$. Матрица этой системы трёхдиагональна. Её главная диагональ доминантна:

$$\left|-\left(\frac{2k^2}{h^2} + \frac{1}{\tau}\right)\right| > \left|\frac{k^2}{h^2}\right| + \left|\frac{k^2}{h^2}\right|. \qquad (7.18)$$

Это условие, как известно из 3-ей главы (п. 3.3.3), является достаточным условием для успешного решения системы методом трёхдиагональной прогонки. В случае доминирования главной диагонали метод прогонки устойчив относительно вычислительных округлений даже при большой размерности исходной системы уравнений.

2. После решения построенной системы мы знаем значения $\{u_{i,1}\}_{i=1}^{n}$ искомой функции $u(x,t)$ во всех узлах 1-го временного слоя. Для нахождения значений на 2-м временном слое строится аналогичная система:

$$\begin{cases} -\left(\dfrac{2k^2}{h^2} + \dfrac{1}{\tau}\right)u_{1,2} + \dfrac{k^2}{h^2}u_{2,2} = \dfrac{u_{1,1}}{\tau} + f_{1,2} - \dfrac{k^2}{h^2}u_{0,2}, \\ \dfrac{k^2}{h^2}u_{1,2} - \left(\dfrac{2k^2}{h^2} + \dfrac{1}{\tau}\right)u_{2,2} + \dfrac{k^2}{h^2}u_{3,2} = \dfrac{u_{2,1}}{\tau} + f_{2,2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dfrac{k^2}{h^2}u_{i-1,2} - \left(\dfrac{2k^2}{h^2} + \dfrac{1}{\tau}\right)u_{i,2} + \dfrac{k^2}{h^2}u_{i+1,2} = \dfrac{u_{i,1}}{\tau} + f_{i,2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dfrac{k^2}{h^2}u_{n-2,2} - \left(\dfrac{2k^2}{h^2} + \dfrac{1}{\tau}\right)u_{n-1,2} = \dfrac{u_{n,1}}{\tau} + f_{n-1,2} - \dfrac{k^2}{h^2}u_{n,2}. \end{cases} \qquad (7.19)$$

В правой части системы (7.19) присутствуют значения искомой функции, полученные на предыдущем временном слое с помощью первой системы.

Построение значений функции $u(x,t)$ на каждом временном слое связано с решением линейной системы аналогичной системе (7.19), поэтому количество систем равно числу временных слоёв. Результатом решения уравнения теплопроводности будет таблица значений $u_{i,j}$ размерности $m \times n$.

Решение уравнения (7.7) с краевыми условиями 2-го или 3-го рода по неявной схеме проходит аналогичным образом. Построение первого и последнего уравнений для таких краевых условий подробно описано в 6-й главе (п.6.3.3).

ОЦЕНКА КАЧЕСТВА НЕЯВНОЙ СХЕМЫ

Решение уравнения теплопроводности по неявной схеме сводится к решению линейных систем методом трёхдиагональной прогонки. Этот метод подробно разобран в 3-й главе (п.3.3.3). Здесь мы оценим только качество схемы.

Матрицы всех линейных систем, построенных на каждом временном слое, имеют доминантную главную диагональ (7.19). Этого условия достаточно для корректности и устойчивости метода прогонки. Доказано, что в случае доминирования главной диагонали метод прогонки устойчив относительно вычислительных округлений даже при большой размерности исходной системы уравнений. Этот метод устойчив для хорошо обусловленных систем и в том случае, когда диагональное преобладание нарушается.

В неявной схеме шаг по времени не зависит от шага по пространству.

7.4. Двумерное уравнение теплопроводности

МЕТОД ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ

Используем этот метод для решения уравнения теплопроводности

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f(x,y,t) \qquad (7.20)$$

Метод переменных направлений является сеточным. В его основе лежит неявная схема. Решение $u(x,y,t)$ строится в области

$$D: \{\ x\epsilon\ [a;b]\ , y\epsilon\ [c;d], t\epsilon[0;\mathrm{T}]\ \}.$$

В этой области задано начальное условие:

$$u(x,y,0) = \varphi(x,y). \qquad (7.21)$$

Будем полагать, для простоты, что на границе Г области D краевые условия однородны: $u(x,y,t)|_{\Gamma} = 0,$ а коэффициент теплопроводности равен 1.

Далее строим разностные сетки. Введём постоянные шаги разбиения для каждой переменной: h_x, h_y и шаг разбиения по времени τ. С помощью этих

шагов на отрезке $[a;b]$ строятся точки $\{x_i\}_{i=0}^{n}$, на отрезке $[c;d]$ – точки $\{y_k\}_{k=0}^{l}$, на отрезке [0, T] –точки $\{t_j\}_{j=0}^{m}$.

Каждый узел разбиения области D имеет тройную индексацию. Значение искомой функции $u(x,y,t)$ в любом узле $(i,\ k,\ j)$ запишем кратко в виде $u_{i,k}^{j}$. Значение $f(x_i,y_k,t_j)=f_{i,k}^{j}$.

Каждый временной слой – плоскость, параллельная плоскости XOY. После разбиения области D на любом временном слое появляется сетка.

РАЗНОСТНАЯ АППРОКСИМАЦИЯ ПРОИЗВОДНЫХ

Разностные аппроксимации частных производных в любом внутреннем узле каждой сетки можно записать в виде

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{(i,k,j)}=\frac{u_{i-1,k}^{j}-2u_{i,k}^{j}+u_{i+1,k}^{j}}{h_x^2},\qquad \left.\frac{\partial^2 u}{\partial y^2}\right|_{(i,k,j)}=\frac{u_{i,k-1}^{j}-2u_{i,k}^{j}+u_{i,k+1}^{j}}{h_y^2},$$

$$\left.\frac{\partial u}{\partial t}\right|_{(i,k,j)}=\frac{u_{i,k}^{j}-u_{i,k}^{j-1}}{\tau}.$$

Если подставить эти аппроксимации производных в уравнение (7.20), то в результате получится очень громоздкое разностное уравнение, к которому невозможно непосредственно применить столь удачную неявную схему.

Математики Д. Кронк и Ф. Николсон разработали простой и интересный метод, позволяющий использовать неявную схему. Познакомимся с алгоритмом этого метода.

АЛГОРИТМ МЕТОДА ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ

Прежде всего, с помощью начального условия (7.21) вычисляются значения искомой функции во всех узлах нулевого временного слоя, т.е. в узлах нулевой сетки (рис.7.4).

Шаг по времени делится пополам, Переход от одного временного слоя j к последующему слою $(j+1)$ состоит из двух полушагов $\tau/2$. Номер 1-го слоя на первом полушаге обозначим в виде $j+1/2$. Номер 2-го слоя на втором полушаге будет равен $(j+1)$. В начале алгоритма $j=0$.

1. Первое направление (вдоль оси *OX*).

Разностное уравнение записывается в виде

$$\frac{u_{i,k}^{j+\frac{1}{2}}-u_{i,k}^{j}}{\tau}=\frac{u_{i-1,k}^{j+\frac{1}{2}}-2u_{i,k}^{j+\frac{1}{2}}+u_{i+1,k}^{j+\frac{1}{2}}}{h_x^2}+\frac{u_{i,k-1}^{j}-2u_{i,k}^{j}+u_{i,k+1}^{j}}{h_y^2}+f_{i,k}^{j+\frac{1}{2}}.$$

Это уравнение преобразуется в неявную схему:

$$\frac{1}{h_x^2}u_{i-1,k}^{j+\frac{1}{2}} - \left(\frac{2}{h_x^2}+\frac{1}{\tau}\right)u_{i,k}^{j+\frac{1}{2}} + \frac{1}{h_x^2}u_{i+1,k}^{j+\frac{1}{2}} = \left(\frac{2}{h_y^2}-\frac{1}{\tau}\right)u_{i,k}^{j} - \frac{u_{i,k-1}^{j}+u_{i,k+1}^{j}}{h_y^2} - f_{i,k}^{j+\frac{1}{2}}. \qquad (7.22)$$

Зафиксируем индекс j. Для $j=0$ все числовые величины в правой части этого уравнения известны из начальных условий. Зафиксируем индекс $k=1$. С помощью уравнения (7.22), изменяя только индекс i $(i=1,\cdots,n-1)$ построим систему линейных уравнений. Благодаря однородности краевых условий в первом и последнем уравнениях этой системы

$$u_{0,k}^{j+\frac{1}{2}} = 0 \quad \text{и} \quad u_{n,k}^{j+\frac{1}{2}} = 0.$$

Очевидно, что система будет иметь трёхдиагональную матрицу с доминирующей главной диагональю. После решения системы методом прогонки получим значения искомой функции во всех узлах 1-й сетки с индексом $k=1$. Далее для каждого фиксированного $k=2,\cdots,l$ строится и решается аналогичная система. На этом заканчивается первый шаг, в конце которого станут известными значения функции во всех узлах 1-й сетки.

2. Второе направление (вдоль оси OY).

Разностное уравнение записывается в виде

$$\frac{u_{i,k}^{j+1}-u_{i,k}^{j+\frac{1}{2}}}{\tau} = \frac{u_{i-1,k}^{j+\frac{1}{2}}-2u_{i,k}^{j+\frac{1}{2}}+u_{i+1,k}^{j+\frac{1}{2}}}{h_x^2} + \frac{u_{i,k-1}^{j+1}-2u_{i,k}^{j+1}+u_{i,k+1}^{j+1}}{h_y^2} + f_{i,k}^{j+1}.$$

Преобразуем это уравнение в неявную схему:

$$\frac{1}{h_y^2}u_{i,k-1}^{j+1} - \left(\frac{2}{h_y^2}+\frac{1}{\tau}\right)u_{i,k}^{j+1} + \frac{1}{h_y^2}u_{i,k+1}^{j+1} = \left(\frac{2}{h_x^2}-\frac{1}{\tau}\right)u_{i,k}^{j+\frac{1}{2}} - \frac{u_{i-1,k}^{j+\frac{1}{2}}+u_{i+1,k}^{j+\frac{1}{2}}}{h_y^2} - f_{i,k}^{j+1}. \qquad (7.23)$$

Все числовые величины в правой части уравнения известны из 1-го слоя. Зафиксируем теперь индекс $i=1$. С помощью уравнения (7.23), изменяя только индекс k $(k=1,\cdots,l-1)$ построим систему уравнений. Краевые условия однородны, поэтому в первом и последнем уравнениях системы

$$u_{i,0}^{j+1} = 0 \quad \text{и} \quad u_{i,l}^{j+1} = 0.$$

У системы тоже будет трёхдиагональная матрица с доминирующей главной диагональю. После решения каждой системы методом прогонки получим значения искомой функции во всех узлах 2-й сетки с индексом $i=1$. Далее для каждого фиксированного $i=2,\cdots,n$ строится и решается аналогичная система. В итоге становятся известными значения искомой функции во всех узлах 2-й сетки (рис. 7.4).

3. Зафиксируем индекс j. Пусть $j = 1$. Теперь мы можем приступить к построению значений функции $u(x, y, t)$ на следующем временном слое 2τ. Для этого достаточно повторить алгоритм, начиная с 1-го пункта.

Решение двумерного уравнения завершается после нахождения значений искомой функции на последнем временном слое $(j = m)$.

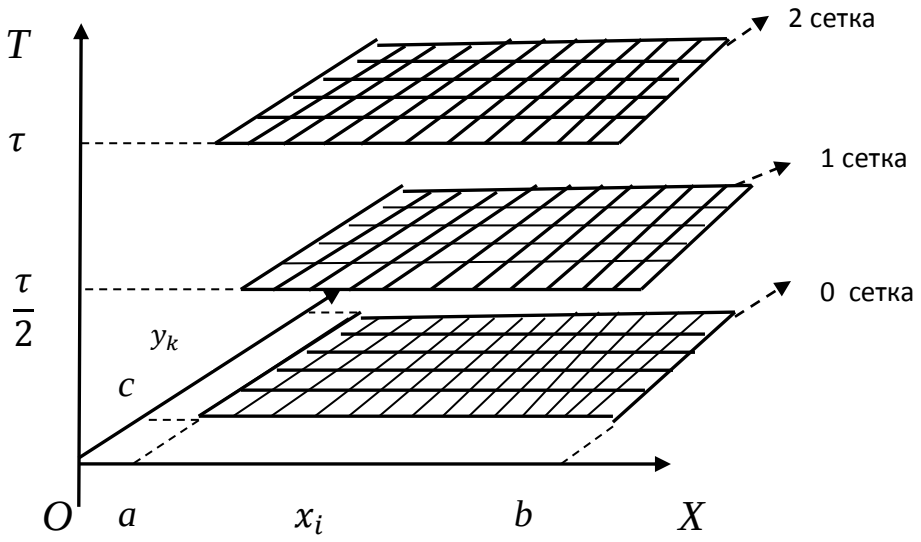

Рис.7.4.

ОЦЕНКА КАЧЕСТВА МЕТОДА ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ

Процесс решения двумерного уравнения теплопроводности методом переменных направлений приводит к решению большого числа $(n + l - 2)m$ линейных систем. Но поскольку в основе этого метода лежит неявная схема, метод переменных направлений – устойчивый метод. Его можно успешно использовать и для трёхмерного уравнения теплопроводности

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + f(x, y, z, t).$$

Для этого достаточно шаг по времени τ разбить на три равных шага $\tau/3$ **.** К описанному выше алгоритму, добавится ещё одно направление.

7.5. Одномерное волновое уравнение

Построим разностную схему для решения сеточным методом простейшего волнового уравнения

$$\frac{\partial^2 u}{\partial t^2} = k^2 \frac{\partial^2 u}{\partial x^2}$$

в области $D: \{ x\epsilon\ [a; b]\ , t\epsilon[0; T]\ \}$.

Заданы начальные условия

$$u(x,0)=\varphi_1(x), \qquad \left.\frac{\partial u}{\partial t}\right|_{t=0}=\varphi_2(x)$$

и краевые условия первого рода

$$u(a,t)=\gamma_1(t),\ \ u(b,t)=\gamma_2(t).$$

Введём сетку с постоянными шагами разбиения по каждой переменной:

$$h=\frac{b-a}{n}, \quad h=h_x=x_{i+1}-x_i, \qquad i=0,1,\cdots,n-1;$$

$$\tau=\frac{T}{m}, \qquad \tau=\tau_t=t_{j+1}-t_j, \qquad j=0,1,\cdots,m-1.$$

У всех узлов построенной сетки есть двойная индексация (i,j).

Заменим в волновом уравнении вторые частные производные на их разностные аппроксимации в любом внутреннем узле сетки

$$\frac{u_{i,j-1}-2u_{i,j}+u_{i,j+1}}{\tau^2}=k^2\,\frac{u_{i-1,j}-2u_{i,j}+u_{i+1,j}}{h^2}.$$

Из этого уравнения выразим явно величину $u_{i,j+1}$

$$u_{i,j+1}=\left(\frac{k\tau}{h}\right)^2\left(u_{i-1,j}-2u_{i,j}+u_{i+1,j}\right)+2u_{i,j}-u_{i,j-1}. \qquad (7.24)$$

Уравнение (7.24) является явной схемой для решения волнового уравнения. Безразмерный параметр $\frac{k\tau}{h}=r$ в схеме называется ***параметром Куранта***. Шаблон явной схемы для волнового уравнения представлен на рисунке 7.5.

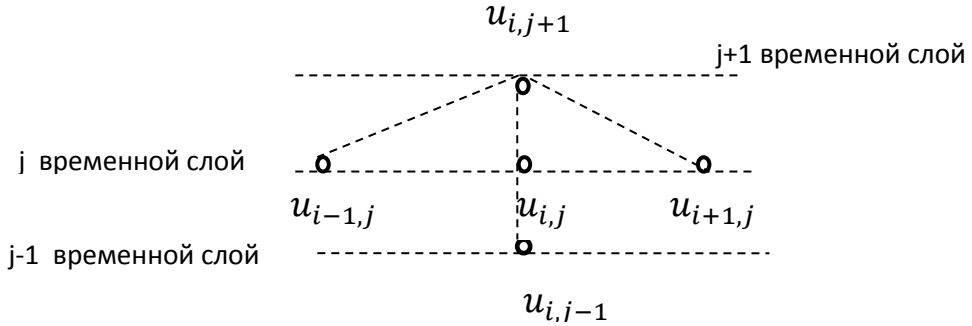

Рис. 7.5

Из схемы (7.24) и шаблона видно, что для вычисления величины $u_{i,j+1}$ надо знать значения функции на двух предыдущих временных слоях. Поэтому построение значений функции на 1-м временном слое связано с фиктивным временным слоем с номером $j=-1$. Для нахождения значений $u_{i,-1}$ на этом слое используем второе начальное условие

$$\left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_2(x).$$

В каждом узле сетки на нулевом временном слое заменяем эту частную производную левосторонней разностной аппроксимацией:

$$\frac{u_{i,-1} - u_{i,0}}{-\tau} = \varphi_2(x_i) \Longrightarrow u_{i,-1} = u_{i,0} - \tau\varphi_2(x_i) \Longrightarrow u_{i,-1} = \varphi_1(x_i) - \tau\varphi_2(x_i).$$

С помощью последнего равенства вычисляем значения $u_{i,-1}$ на фиктивном слое для всех значений индекса $i = 0,2,\cdots,n$. Теперь по явной схеме (7.24) легко находятся значения $u_{i,1}$ во всех внутренних узлах первого временного слоя. Далее по той же явной схеме (7.24) вычисляются значения искомой функции во всех узлах последующих временных слоёв.

ОЦЕНКА КАЧЕСТВА ЯВНОЙ СХЕМЫ

Из критерия фон Неймана следует, что явная схема решения волнового уравнения будет устойчивой, если параметр Куранта $r \leq 1$. Поэтому такая явная схема условно устойчива.

7.6. Уравнения Лапласа и Пуассона

Построение математической модели различных стационарных процессов часто приводит к уравнению Лапласа

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \Delta u = 0. \qquad (7.25)$$

Напомним, что дважды непрерывно дифференцируемая функция $u(x,y)$ называется ***гармонической***, если она удовлетворяет уравнению Лапласа в некоторой области D с непрерывной границей $\Gamma(D)$.

Для единственности решения уравнения (7.25) должно быть дополнительно задано краевое условие одного из трёх видов:

1. $u(x,y) = \gamma(x,y)$ на $\Gamma(D)$;
2. $\frac{\partial u}{\partial n} = \gamma(x,y)$ на $\Gamma(D)$, где $\frac{\partial u}{\partial n}$ – производная по нормали к границе области;
3. $\alpha_1 u(x,y) + \alpha_2 \frac{\partial u}{\partial n} = \gamma(x,y)$ на $\Gamma(D)$.

Решение уравнения Лапласа с первым краевым условием называется задачей Дирихле, со вторым – задачей Неймана, с третьим – смешанной задачей.

Перейдём к решению задачи Дирихле сеточным методом. Для простоты будем решать задачу в прямоугольнике, стороны которого параллельны координатным осям (рис. 7.6). Пусть шаги разбиения по каждой переменной будут

постоянными и одинаковыми: $h = h_x = h_y$. Очевидно, что каждый узел сетки имеет двойную индексацию (i, j).

Во всех граничных узлах сетки значения искомой функции известны из краевого условия:

$$u_{1,j} = \gamma(x_1, y_j) = \gamma_{1,j},\ u_{n,j} = \gamma(x_n, y_j) = \gamma_{n,j}\ ,\ \text{где}\ j = 1, \cdots, m. \quad (7.26)$$

$$u_{i,1} = \gamma(x_i, y_1) = \gamma_{i,1},\ u_{i,m} = \gamma(x_i, y_m) = \gamma_{i,m}\ ,\ \text{где}\ i = 1, \cdots, n. \quad (7.27)$$

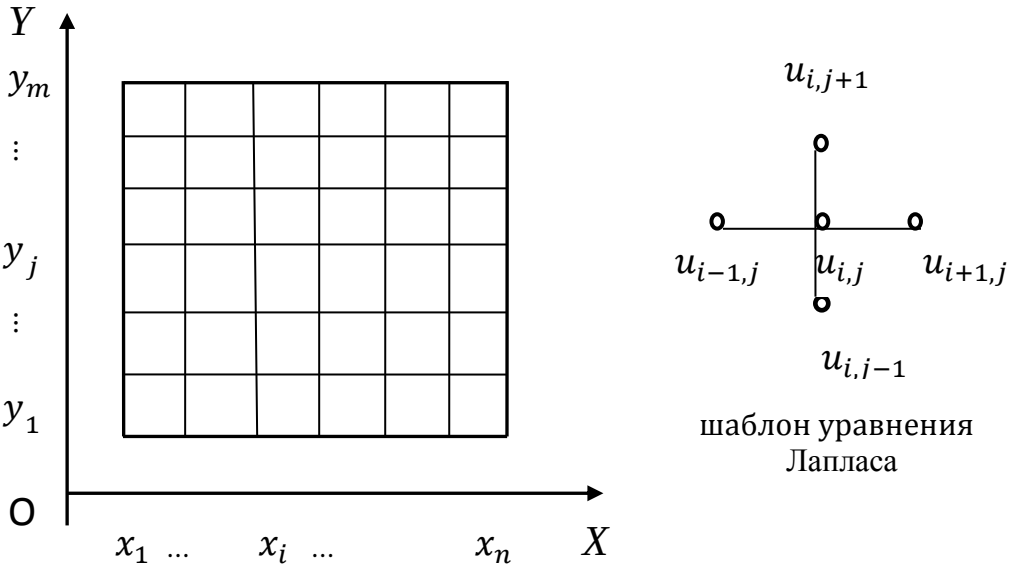

Рис. 7.6

Для построения разностной схемы уравнения Лапласа используем центральную аппроксимацию вторых частных производных

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} = 0.$$

Выразим из этого уравнения величину $u_{i,j}$, которая находится во внутреннем узле сетки (шаблон на рис. 7.6)

$$u_{i,j} = \frac{u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}}{4}\ . \quad (7.28)$$

Если запишем уравнение (7.28) для каждого внутреннего узла сетки, то получим линейную систему уравнений. Эта система будет неоднородной, поскольку величины $u_{1,j}$, $u_{n,j}$, $u_{i,1}$, $u_{i,m}$ известны из краевого условия $(7.26–7.27)$. Доказано, что определитель такой системы не равен нулю, значит, существует единственное решение.

Количество уравнений $Q = (m-2)(n-2)$ в системе будет равно числу всех внутренних узлов сетки. Чем меньше шаг сетки, тем выше точность решения. Но с уменьшением шага существенно увеличивается количество узлов, а значит и число уравнений в системе, что приведёт к росту погрешности.

Поэтому для реализации схемы (7.28) обычно не строят систему, а используют метод итераций.

АЛГОРИТМ МЕТОДА ИТЕРАЦИЙ

Обозначим номер каждого шага итерации верхним индексом $k = 1, \cdots, L$ и приведём уравнение (7.28) к виду

$$u_{i,j}^{k+1} = \frac{u_{i-1,j}^{k} + u_{i+1,j}^{k} + u_{i,j-1}^{k} + u_{i,j+1}^{k}}{4}. \qquad (7.29)$$

1. Во всех внутренних узлах сетки задаётся начальное приближение значений $u_{i,j}^{0}$. Для этого можно использовать краевое условие. Вычисляется число R_0, равное среднему арифметическому $(2n + 2m - 4)$ значений функции $\gamma(x, y)$ в граничных узлах:

$$R_0 = \frac{1}{(2n + 2m - 4)} \left(\sum_{j=1}^{m} (\gamma_{1,j} + \gamma_{n,j}) + \sum_{i=1}^{n} (\gamma_{i,1} + \gamma_{i,m}) \right).$$

Возьмём за начальное приближение число R_0:

$$R_0 = u_{i,j}^{0}, \qquad j = 2, \cdots, m-1;\ i = 2, \cdots, n-1.$$

2. На каждом шаге итерации с помощью уравнения (7.29) вычисляются новые значения $u_{i,j}^{k+1}$ во всех внутренних узлах сетки.

3. После каждого шага итерации сравниваются значения функции $u(x, y)$, построенные за два последовательных шага итерации:

$$\left| \frac{u_{i,j}^{k+1} - u_{i,j}^{k}}{u_{i,j}^{k}} \right| < \varepsilon, \qquad (7.30)$$

где ε – заданная относительная погрешность.

Если неравенство (7.30) выполняется во всех внутренних узлах сетки, можно считать, что построено решение задачи Дирихле. Если неравенство (7.30) не выполняется хотя бы в одном внутреннем узле, тогда надо сделать ещё шаг итерации, используя формулу (7.29).

Доказано, что такой итерационный процесс сходится, следовательно, на некотором L-м шаге итерации обязательно будет выполняться неравенство (7.30) для всех узлов сетки.

Аналогичный алгоритм решения задачи Дирихле можно использовать для уравнения Пуассона:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y)$$

и трёхмерных уравнений Лапласа и Пуассона

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=0, \qquad \frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=f(x,y,z).$$

ОЦЕНКА КАЧЕСТВА СХЕМЫ

Доказано, что задача Дирихле является корректно поставленной задачей. Это означает, что решение всегда существует, единственно и устойчиво относительно граничных условий. Построенное решение сходится к точному решению при стремлении шага разбиения к нулю.

7.7. Замечания о сеточном методе

В заключении отметим, что при использовании сеточного метода необходимы не только теоретические знания, но и большой опыт его применения. Ниже приведены некоторые заключения, полученные опытным путём.

1. Реально может оказаться, что схема первого порядка точности на грубых сетках может дать более точный результат, чем схема второго порядка на той же сетке.

2. Обычно априорные оценки точности далеки от оптимальных и для конкретных задач могут быть сильно завышены.

3. Оценки точности схем часто являются асимптотическими при стремлении шага сетки к нулю.

5. Порядок точности результата решения задачи может быть выше порядка аппроксимации производных.

6. Для случая многих переменных порядок аппроксимации производных по разным переменным может быть различным. Очевидно, что порядок точности решения по разным переменным тоже может быть различным.

Глава 8

ПОИСК ЭКСТРЕМУМОВ ФУНКЦИЙ

Будет ли конец для всего, что нас окружает?
Если верить математикам, то нет.
Ричард Браун. «Математика за 30 секунд»

Определение 1. Функция $f(x)$ называется унимодальной на $[a;b]$, если существует только одно такое число x_0, что $f(x)$ убывает (возрастает) на $[a;x_0]$ и возрастает (убывает) на $[x_0;b]$.

Далее представлены методы только для унимодальных на $[a;b]$ функций.

8.1. Метод золотого сечения

Метод золотого сечения используется для нахождения минимума непрерывной функции одного аргумента. Для понимания идеи этого метода и алгоритма его реализации очень кратко напомним историю появления выражения *золотое сечение*.

Среди различных средних величин уникальными свойствами обладает одно среднее, которое специальным образом делит отрезок длины a на две части. Отношение целого отрезка a к его большей части x должно равняться отношению большей части x к меньшей части $(a-x)$:

$$a:x = x:(a-x). \qquad (8.1)$$

Эта пропорция приводит к уравнению:

$$x^2 + ax - a^2 = 0.$$

Полученное уравнение имеет один положительный корень:

$$x = \frac{a(\sqrt{5}-1)}{2}. \qquad (8.2)$$

Если рассматривать отрезок единичной длины $(a=1)$,, то уравнение имеет вид

$$x^2 + x - 1 = 0.$$

Положительный корень этого уравнения равен

$$x = \frac{(\sqrt{5}-1)}{2} = 0.6180339\ldots\ . \qquad (8.3)$$

Такая пропорция была известна ещё пифагорейцам. В эпоху Возрождения Леонардо да Винчи назвал её золотым сечением. Оно используется в архитектуре, скульптуре, живописи, музыке и математике. Более подробно о золотом сечении можно прочесть в приложении 1.

Предположим, что унимодальная функция $f(x)$ имеет минимум в неизвестной внутренней точке отрезка $[0;1]$. Для его нахождения строится алгоритм, в основе которого лежит оптимальное сужение области поиска минимума с помощью золотого сечения.

1. Отрезок $[0;1]$ делится на три части с помощью числа r $(0.5 < r < 1)$. Число r находится из пропорции золотого сечения (8.1)

$$\frac{r}{1} = \frac{1-r}{r} \Rightarrow r = \frac{\sqrt{5}-1}{2}.$$

2. Строим две внутренние точки (рис. 8.1): $x_1 = 1 - r;\ x_2 = r.$

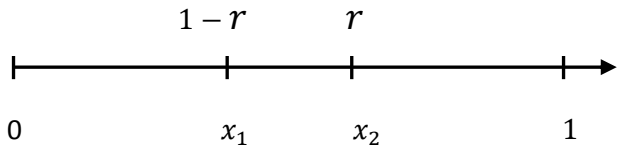

Рис. 8.1

3. Вычисляем значения

$$f(0), \quad f(x_1), \quad f(x_2), \quad f(1)$$

и выбираем среди них наименьшее. Пусть наименьшее значение не находится в точках $x = 0$ и $x = 1$.

Если $f(x_1) < f(x_2)$, переходим к поискам минимума на $[0; x_2]$ (рис. 8.2).

Если $f(x_1) > f(x_2)$, переходим к поискам минимума на $[x_1; 1]$ (рис. 8.3).

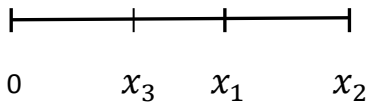

Рис. 8.2

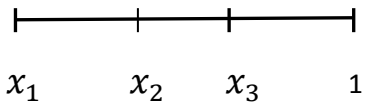

Рис. 8.3

4. Предположим, что $f(x_1) < f(x_2)$. В этом случае точка x_1 находится внутри $[0; x_2]$. Для того, чтобы вновь разделить этот отрезок на три части, достаточно построить только одну точку x_3 согласно пропорции золотого сечения.

Далее вычисляется $f(x_3)$ и сравниваются между собой значения функции

$$f(0), \quad f(x_3), \quad f(x_1), \quad f(x_2).$$

Если $f(x_3)$ – наименьшее значение, то выбирается отрезок $[0; x_1]$. В противном случае выбирается отрезок $[x_3; x_2]$. В одном из этих отрезков вновь вычисляется точка x_4 с помощью золотого сечения. Затем опять сравниваются значения функции в соответствующих точках и т. д.

Доказано, что метод золотого сечения – это сходящийся итерационный процесс. Поэтому на каком-то шаге итерации длина отрезка $|\delta x|$, в котором находится минимум, станет меньше заданной погрешности ε:

$$|\delta x| < \varepsilon.$$

В качестве наименьшего значения функции можно взять $f(x_i)$ или $f(x_{i+1})$, где x_i и x_{i+1} – граничные точки отрезка δx. Можно использовать их среднее арифметическое:

$$f_{min} = \frac{f(x_i) + f(x_{i+1})}{2}.$$

8.2. Поиск минимума функций двух переменных

Рассмотрим дифференцируемую, унимодальную функцию $z = F(x, y)$. Графический образ такой функции – поверхность в трёхмерном пространстве с декартовой системой координат. Минимальное значение функции – это значение координаты z низшей точки поверхности.

Эффективными методами поиска минимума являются итерационные методы спуска. Они сводятся к построению таких траекторий на поверхности графика функции, вдоль которых на каждом шаге итерации функция убывает, можно сказать, спускается к низшей точке поверхности. Поэтому такие методы ещё называют методами спуска. Они отличаются друг от друга способом спуска.

8.2.1. Метод наискорейшего спуска

Второе название метода – ***градиентный метод***. Прежде чем перейти к его описанию, напомним, что градиентом функции $z = F(x, y)$ называется вектор

$$grad\, F(x_0, y_0) = F'_{x_0} \cdot \bar{\imath} + F'_{y_0} \cdot \bar{\jmath}.$$

Доказано, что градиент ортогонален к поверхности графика функции в точке $N(x_0, y_0)$ и он показывает направление наискорейшего возрастания функции $F(x, y)$ при достаточно малом движении из точки $N(x_0, y_0)$. Ясно, что антиградиент равен

$$-grad\, F(x_0, y_0).$$

Он показывает направление наискорейшего убывания функции при малом движении из точки $N(x_0, y_0)$. Поэтому градиентный метод ещё называют методом наискорейшего спуска.

Обратимся к эскизу графика функции (рис. 8.4). Проведём семейство плоскостей $F(x, y) = const$. Линии пересечения этих плоскостей с поверхностью $z = F(x, y)$ спроектируем на плоскость XOY (рис. 8.5). Такие проекции называются ***линиями уровня***.

Замечание 1. Точность вычисления градиента сильно влияет на скорость сходимости метода наискорейшего спуска.

Замечание 2. Очевидно, что в точке минимума или максимума градиент дифференцируемой функции равен нулю.

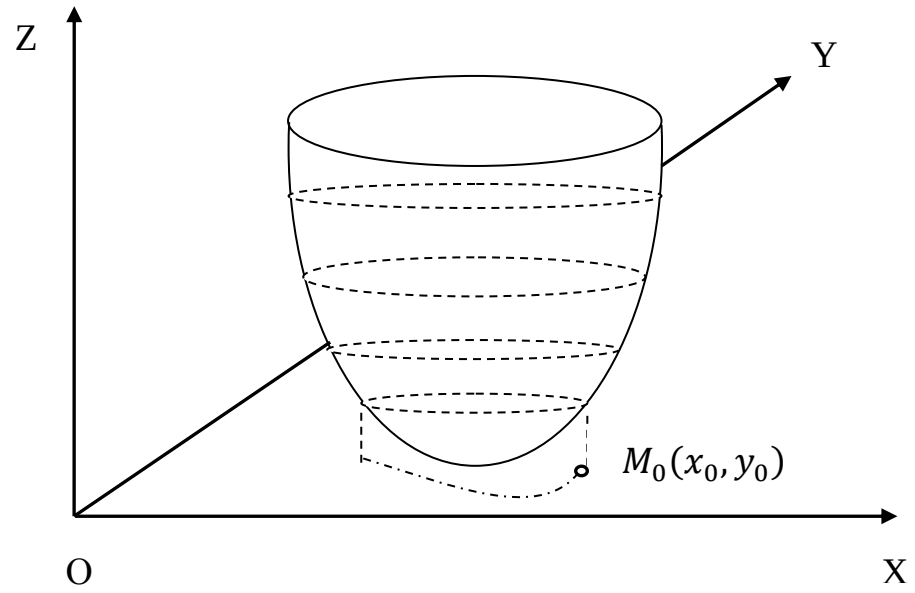

Рис. 8.4

АЛГОРИТМ МЕТОДА

1. На некоторой линии уровня задаём начальную точку $M_0(x_0, y_0)$ и допустимые величины погрешности: ε_x, ε_y, δ. Чем ближе начальная точка к окрестности нахождения точки минимума, тем быстрее сходится итерационный процесс поиска.

2. Для нахождения новой точки, более близкой к точке минимума, строим радиус-вектор $\boldsymbol{r_0} = \overline{0M_0}$ и вектор $\overline{M_0M_1} = -t_1 gradF(x_0, y_0)$, где t_1 – длина шага вдоль направления $-gradF(x_0, y_0)$. Будем искать такую новую точку $M_1(x_1, y_1)$ на другой линии уровня, чтобы быть ближе к точке минимума (рис. 8.5). Для этого вводим радиус-вектор $\boldsymbol{r_1} = \overline{0M_1}$, удовлетворяющий равенству

$$\boldsymbol{r}_1 = \boldsymbol{r_0} - t_1 gradF(x_0, y_0). \qquad (8.4)$$

3. Строим координаты точки $M_1(x_1, y_1)$. Переходим от равенства векторов (8.4) к равенствам соответствующих координат:

$$x_1 = x_0 - t_1 F'_x; \;\; y_1 = y_0 - t_1 F'_y. \qquad (8.5)$$

4. Точка $M_1(x_1, y_1)$ лежит на линии уровня, следовательно, она принадлежит области определения функции $z = F(x, y)$:

$$z(M_1) = F(x_1, y_1) = F\big(x_0 - t_1 F'_x(x_0, y_0);\;\; y_0 - t_1 F'_y(x_0, y_0)\big) = F(t_1).$$

5. Спуск на каждом шаге должен быть наискорейшим. Поэтому величину t_1 находим из условия равенства нулю первой производной: $F'(t_1) = 0$.

Предположим, что найдено решение t_1 этого уравнения (глава 2). С помощью равенств (8.5) находим координаты первой точки приближения.

6. Сравниваем координаты двух точек M_0 и M_1:

$$\begin{cases} \left|\dfrac{x_1 - x_0}{x_0}\right| < \varepsilon_x, \\ \left|\dfrac{y_1 - y_0}{y_0}\right| < \varepsilon_y \end{cases} \qquad (8.6)$$

или значения функции в этих точках:

$$\left|\frac{F(x_1,y_1)-F(x_0,y_0)}{F(x_0,y_0)}\right| < \delta. \qquad (8.7)$$

Если выполняются условия (8.6) или (8.7), то можно считать, что в точке M_1 достигается минимум. Если условия не выполняются, то происходит 2-й шаг итерации по описанному выше алгоритму.

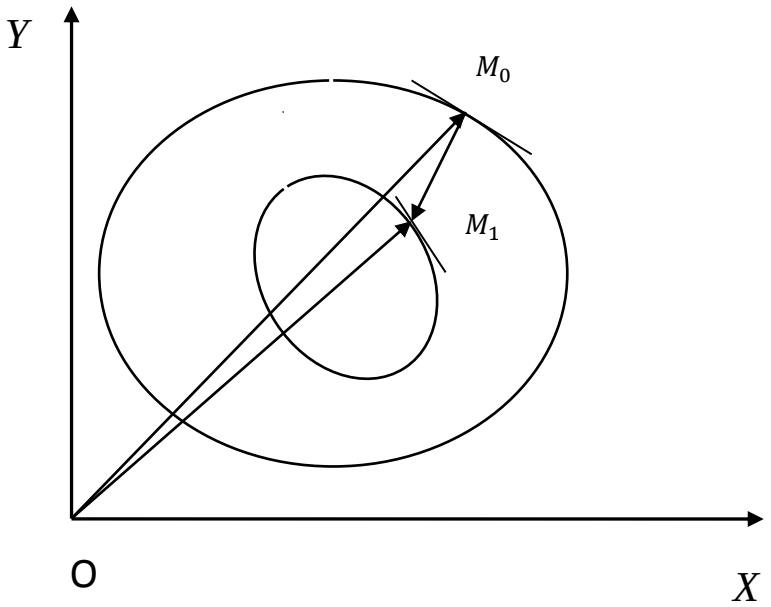

Рис. 8.5

На (n+1)-м шаге итерации координаты точки M_{n+1} выражаются через координаты предыдущей точки M_n и градиент функции в этой точке:

$$x_{n+1} = x_n + t_{n+1}F_x'(x_n, y_n);\ \ y_{n+1} = y_n + t_{n+1}F_y'(x_n, y_n). \qquad (8.8)$$

Выражения (8.8) подставляются в исходную функцию:

$$F(x_{n+1}, y_{n+1}) = F\big(x_n + t_{n+1}F_x'(x_n, y_n), y_n + t_{n+1}F_y'(x_n, y_n\big) = F(t_{n+1}). \qquad (8.9)$$

Функция теперь зависит только от неизвестного параметра t_{n+1}, значение которого является решением уравнения

$$F'(t_{n+1}) = 0. \qquad (8.10)$$

После нахождения t_{n+1} вычисляем координаты (8.8) точки M_{n+1}. Далее сравниваются результаты двух последних шагов итерации:

$$\begin{cases} \left|\frac{x_{n+1}-x_n}{x}\right| < \varepsilon_x, \\ \left|\frac{y_{n+1}-y_n}{y_n}\right| < \varepsilon_y \end{cases} \quad \text{или} \quad \left|\frac{F(x_{n+1},y_{n+1})-F(x_n,y_n)}{F(x_n,y_n)}\right| < \delta. \qquad (8.11)$$

Доказано, что градиентный метод сходится, поэтому всегда существует такой (n+1)-й шаг итерации, на котором будут выполняться неравенства (8.11). Точка $M_{n+1}(x_{n+1}, y_{n+1})$ – точка минимума исходной функции.

Замечание 3. Траектория спуска состоит из звеньев. Легко доказать, что соседние звенья этой траектории ортогональны друг другу. Для этого достаточно с помощью равенства (8.9), записать уравнение (8.10) в виде скалярного произведения:

$$F'(t_{n+1}) = F'_x(M_{n+1}) \cdot F'_y(M_n) = \big(gradF(M_{n+1}), gradF(M_n)\big) = 0.$$

Скалярное произведение соседних градиентов равно нулю, значит, они ортогональны.

8.2.2. Метод спуска по координатам

АЛГОРИТМ МЕТОДА

1. Выберем нулевое приближение – начальную точку $M_0(x_0, y_0)$.
2. Зафиксируем вторую переменную y в исходной функции $z = F(x, y)$:

$$z(x) = F(x, y_0).$$

Ищем то значение аргумента, при котором функция $z(x)$ имеет минимум:

$$z'(x) = 0. \qquad (8.12)$$

3. Предположим, что решение уравнения (8.12) равно x_1.Возвращаемся к исходной функции $z = F(x, y)$ и фиксируем первую переменную:

$$z(y) = F(x_1, y).$$

Вновь находится то значение аргумента, при котором функция $z(y)$ имеет минимум:

$$z'(y) = 0. \qquad (8.13)$$

Пусть решение уравнения (8.13) равно y_1.

4. В качестве первой точки приближения выбираем точку $M_1(x_1, y_1)$. Далее происходит сравнение (8.6) координат двух точек M_0 и M_1 или сравнение значений исходной функции в этих точках (8.7).

Если выполняются условия (8.6) или (8.7), то можно считать, что в точке M_1 достигается минимум. Если эти условия не выполняются, то происходит следующий шаг итерации (рис. 8.6).

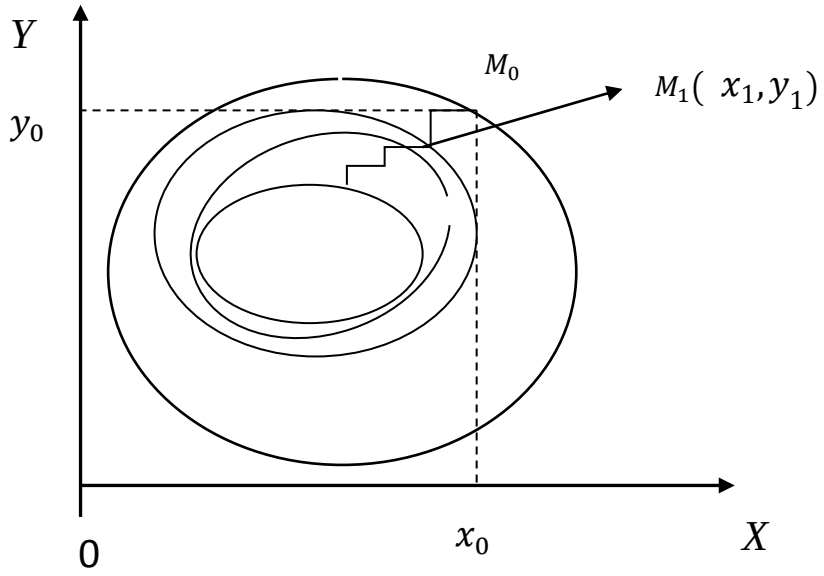

Рис.8.6

Каждая последующая точка приближения $M_{n+1}(x_{n+1}, y_{n+1})$ строится с помощью точки $M_n(x_n, y_n)$, построенной на предыдущем шаге итерации.

На (n+1)-м шаге итерации переходим к функции

$z(x) = F(x, y_n)$, находим x_{n+1} – корень уравнения $\mathrm{z}'(x) = 0$.

Вновь используем функцию

$z(y) = F(x_{n+1}, y)$ и вычисляем y_{n+1} – корень уравнения $\mathrm{z}'(y) = 0$.

Доказано, что метод спуска по координатам сходится. Поэтому на некотором (n+1)-шаге будут выполняться неравенства (8.11). Точка $M_{n+1}(x_{n+1}, y_{n+1})$ будет точкой минимума.

ОЦЕНКА ПОГРЕШНОСТИ МЕТОДОВ

Скорость сходимости описанных выше методов зависит от вида графика исходной функции. Метод, приспособленный к одному типу графиков функций, может оказаться плохим для других.

Дифференцируемая функция в достаточно малой окрестности экстремальной точки изменяется очень мало. Поэтому при оценке погрешностей в такой окрестности могут быть верными неравенства

$$\begin{cases} \left|\dfrac{F(x_{n+1}, y_{n+1}) - F(x_n, y_n)}{F(x_n, y_n)}\right| < \left|\dfrac{x_{n+1} - x_n}{x_n}\right|, \\ \left|\dfrac{F(x_{n+1}, y_{n+1}) - F(x_n, y_n)}{F(x_n, y_n)}\right| < \left|\dfrac{y_{n+1} - y_n}{y_n}\right|. \end{cases}$$

Глава 9

ИНТЕГРИРОВАНИЕ И ДИФФЕРЕНЦИРОВАНИЕ

Математик изучает свою науку вовсе не потому, что она полезна. Он изучает её потому, что она прекрасна.

Лузин Н.Н. (1883 – 1950) – знаменитый русский математик

9.1. Численное интегрирование

Ставится вопрос о вычислении определённого интеграла

$$I = \int_a^b f(x)dx. \qquad (9.1)$$

Если интеграл (9.1) можно вычислить с помощью формулы Ньютона – Лейбница

$$\int_a^b f(x)dx = F(x)\Big|_a^b = F(b) - F(a), \qquad (9.2)$$

тогда говорят, что функция $f(x)$ интегрируемая, в противном случае $f(x)$ – неинтегрируемая функция. Количество таких функций велико.

Численные методы интегрирования используют для неинтегрируемых функций и в том случае, когда функция $f(x)$ задана очень громоздким аналитическим выражением.

Суть любого метода численного интегрирования заложена в определении определённого интеграла.

Определение 1. Определённым интегралом называется предел интегральных сумм:

$$\int_a^b f(x)dx = \lim_{\substack{\Delta x_i \to 0 \\ n \to \infty}} \sum_{i=1}^{n} f(x_i)\Delta x_i, \qquad (9.3)$$

где x_i – узлы разбиения области интегрирования, n – число интервалов разбиения $\Delta x_i = x_{i+1} - x_i$.

Задача численного интегрирования состоит в том, чтобы вычислить интеграл, используя при этом значения функции $f(x)$ в конечном числе точек области интегрирования.

9.1.1. Простейшие квадратурные формулы

Обратимся к трём простейшим классическим формулам приближённого интегрирования, две из которых имеют элементарную геометрическую интерпретацию и соответствующие названия.

ФОРМУЛА ПРЯМОУГОЛЬНИКОВ

Область интегрирования $[a; b]$ разбивается на n частей:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

На каждом отрезке $[x_i; x_{i+1}]$ аппроксимируем функцию $f(x)$ её значением в средней точке $x_{i+\,0{,}5} = 0{,}5(x_i + x_{i+1})$. Такое приближение функции называется ***кусочно-постоянной аппроксимацией***.

Запишем приближённое значение интеграла на каждом отрезке $[x_i; x_{i+1}]$

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx f\left(x_{i+\,1/2}\right)(x_{i+1} - x_i) \qquad (9.4)$$

Правую часть равенства (9.4) можно рассматривать как площадь (рис. 9.1) прямоугольника.

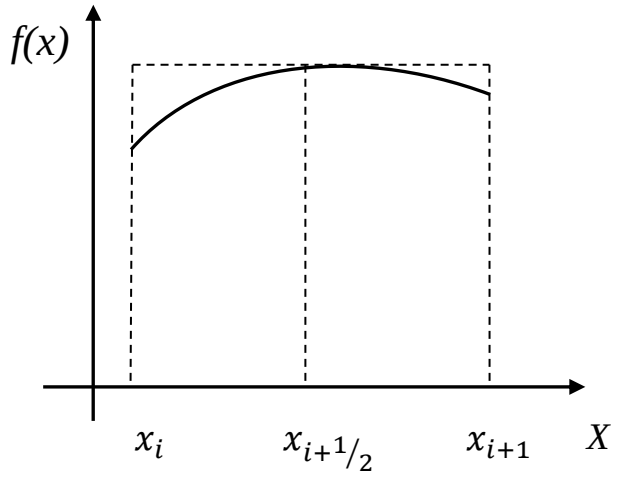

Рис. 9.1

Определённый интеграл обладает свойством аддитивности по области интегрирования, поэтому

$$\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx = \int_a^b f(x)dx = \sum_{i=0}^{n-1} f\left(x_{i+\,1/2}\right)(x_{i+1} - x_i)\,. \qquad (9.5)$$

Последнее равенство в (9.5) называется ***формулой прямоугольников***, или ***формулой средних.***

ФОРМУЛА ТРАПЕЦИЙ

Область интегрирования $[a; b]$ разбивается на n частей:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

В этом способе шаг разбиения обычно берётся постоянным: $h = (b - a) : n$.

На каждом отрезке $[x_i; x_{i+1}]$ аппроксимируем функцию $f(x)$ линейной функцией, проходящей через две граничные точки x_i , x_{i+1} каждого отрезка. Такое приближение функции называется ***кусочно-линейной аппроксимацией***. Геометрическая интерпретация определённого интеграла известна. Интеграл

$$\int_{x_i}^{x_{i+1}} f(x)dx = I_1 \qquad (9.6)$$

можно рассматривать как площадь фигуры, расположенной между отрезком $[x_i; x_{i+1}]$ и соответствующей частью графика функции $f(x)$. После кусочно-линейной аппроксимации площадь фигуры приближённо равна площади трапеции (рис.9.2):

$$I_i \approx S_i = \frac{1}{2}\big(f(x_i) + f(x_{i+1})\big)h \implies \int_{x_i}^{x_{i+1}} f(x)dx \approx \frac{h}{2}\big(f(x_i) + f(x_{i+1})\big).$$

Суммируем площади всех полученных на $[a; b]$ трапеций:

$$\sum_{i=1}^{n} S_i = h\sum_{i=1}^{n-1} f(x_i) + \frac{h}{2}\big(f(x_0) + f(x_n)\big) \implies$$

$$\sum_{i=1}^{n} S_i \approx \sum_{i=1}^{n} I_i = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx = \int_a^b f(x)dx \implies$$

$$\int_a^b f(x)dx = h\sum_{i=1}^{n-1} f(x_i) + \frac{h}{2}\big(f(x_0) + f(x_n)\big). \qquad (9.7)$$

Формула (9.7) называется ***формулой трапеций.***

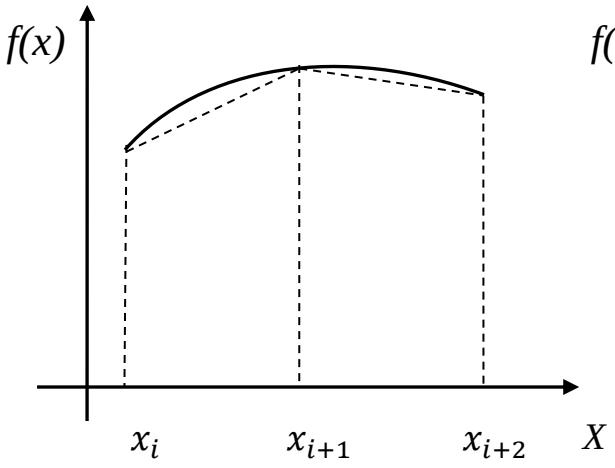

Рис. 9.2

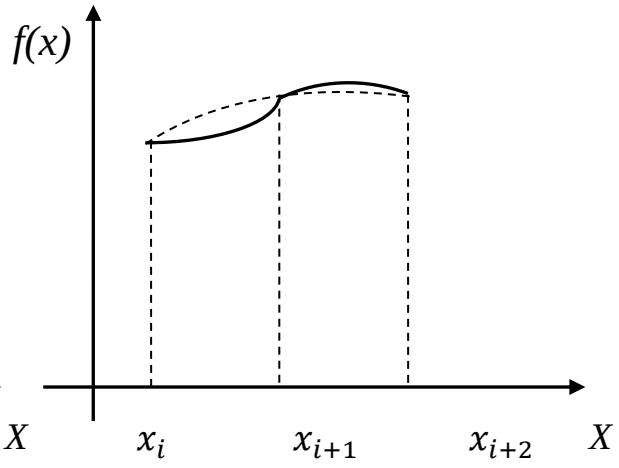

Рис. 9.3

ФОРМУЛА СИМПСОНА

Рассмотрим в общих чертах основную идею построения этой формулы. Разобьём область интегрирования на чётное число интервалов $n = 2k$. Шаг разбиения $h = (b-a)/(2k)$. Для каждых двух смежных областей разбиения (рис.9.3) построим ***квадратичный интерполяционный многочлен***

$$p_j(x) = a_j x^2 + b_j x + c_j, \quad x \in [x_i; x_{i+2}], \quad j = 1,2,\cdots,k. \qquad (9.8)$$

Значения $p_j(x)$ совпадают в трёх узлах x_i, x_{i+1}, x_{i+2} со значениями подынтегральной функции $f(x)$, т.е. выполняются три равенства

$$p_j(x_i) = f(x_i) = f_i, \;\; p_j(x_{i+1}) = f(x_{i+1}) = f_{i+1}, \;\; p_j(x_{i+2}) = f(x_{i+2}) = f_{i+2}.$$

Эти равенства приводят к линейной алгебраической системе (9.9), благодаря которой коэффициенты a_j, b_j, c_j выражаются через три узла разбиения и значения функции $f(x)$ в этих узлах.

$$\begin{cases} a_j x_i^2 + b_j x_i + c_j = f_i, \\ a_j x_{i+1}^2 + b_j x_{i+1} + c_j = f_{i+1}, \\ a_j x_{i+2}^2 + b_j x_{i+2} + c_j = f_{i+2}. \end{cases} \qquad (9.9)$$

Таким же образом каждая тройка коэффициентов a_j, b_j, c_j выражается через соответствующие узлы разбиения и значения подынтегральной функции.

Построенные многочлены $p_j(x)$ элементарно интегрируются:

$$\int_{x_i}^{x_{i+2}} f(x)dx \approx \int_{x_i}^{x_{i+2}} \left(a_j x^2 + b_j x + c_j\right)dx = \left(\frac{1}{3}a_j x^3 + \frac{1}{2}b_j x^2 + c_j x\right)\Big|_{x_i}^{x_{i+2}}.$$

Ряд несложных, но громоздких преобразований приводят к известной ***формуле Симпсона***:

$$\int_a^b f(x)dx = \frac{h}{3}(f(a) + f(b)) + \frac{2h}{3}\sum_{j=1}^{k-1} f\left(x_{2j}\right) + \frac{4h}{3}\sum_{j=1}^{k} f\left(x_{2j-1}\right). \qquad (9.10)$$

ОЦЕНКА ТОЧНОСТИ ФОРМУЛ

Квадратурные формулы, в отличие от определения интеграла, выражаются через конечные суммы (частичные суммы). Поэтому погрешностью любой из таких формул является модуль остаточного члена

$R = \left|\int_a^b f(x)dx - \sum_{i=0}^{n} A_i f_i\right|$, где $\sum_{i=0}^{n} A_i f_i$ –любая квадратурная формула.

Погрешность квадратурных формул связана, прежде всего, с шагом h разбиения области интегрирования и оценивается с его помощью. Из определения интеграла следует, что чем меньше шаг h, тем больше точность, поэтому должно выполняться неравенство $0 < h < 1$.

Для формулы трапеций доказано, что

$$R = \frac{(a-b)f''(c)}{12}h^2 \Longrightarrow R = \mathrm{O}(h^2), \qquad c \in (a;b).$$

Последнее равенство означает, что формула трапеций имеет 2-й ***порядок точности.***

Для формулы Симпсона доказано, что

$$R = \frac{(a-b)f^{(4)}(c)}{180}h^4 \Longrightarrow R = \mathrm{O}(h^4), \quad c \in (a;b).$$

У формулы Симпсона 4-й порядок точности.

Если интеграл вычислен с помощью разных формул, то сравнение таких результатов следует делать при одинаковом шаге h.

ВЫЧИСЛЕНИЕ ИНТЕГРАЛОВ С ЗАДАННОЙ ТОЧНОСТЬЮ

Оптимальный шаг разбиения $h_{\text{орт}}$ области интегрирования обычно неизвестен. Его можно найти, если заранее задать нужную точность ε вычисления интеграла и построить итерационный процесс. Проще всего это делается с помощью формулы Симпсона.

Задаётся начальный шаг h_1. С этим шагом вычисляется интеграл I_{h_1}. Шаг h_1 делится на два: $h_2 = 0{,}5h_1$ и снова вычисляется следующее значение интеграла I_{h_2}. Сравниваются полученные значения:

$$\left|\frac{I_{h_2} - I_{h_1}}{I_{h_1}}\right| < \varepsilon. \qquad (9.11)$$

Если неравенство (9.11) выполняется, то интеграл вычислен с заданной точностью. Если не выполняется, шаг h_2 делится на два: $h_3 = 0{,}5h_2$ и вычисляется интеграл I_{h_3} и т. д. В конце каждого итерации проверяется справедливость неравенства

$$\left|\frac{I_{h_{n+1}} - I_{h_n}}{I_{h_n}}\right| < \varepsilon. \qquad (9.12)$$

Доказано, что такой процесс сходится, поэтому на некотором $(n+1)$-м шаге итерации неравенство (9.12) станет верным. Численное значение интеграла будет равно I_{h_n} с точностью ε, а $h_{\text{орт}} = h_n$.

Итерационный процесс с использованием формулы Симпсона сходится медленно. В настоящее время разработаны и реализованы способы, которые увеличивают скорость сходимости таких итераций.

9.1.2. Интегрирование и кубические сплайны

Разберём один из вариантов другого способа интегрирования. Если подынтегральная функция $f(x)$ достаточно ***гладкая***, её аппроксимируют на отрезке $[a;b]$ другой такой же гладкой, но легко интегрируемой функцией и берут от неё интеграл. Порядок точности численного интегрирования обычно выше порядка точности построенного приближения для функции $f(x)$.

В качестве примера разберём случай, когда надо проинтегрировать дважды непрерывно дифференцируемую функцию $f(x) \in C^2_{[a;b]}$. Функцию можно до интегрирования аппроксимировать кубическим интерполяционным сплайном, а затем без каких-либо проблем взять интеграл от построенного сплайна.

Если функция $f(x)$ задана в виде таблицы, тогда сплайн строится на основе этой таблицы. Если $f(x)$ – аналитическая функция, то узлы разбиения можно задавать различными способами с учётом свойств $f(x)$, а затем вычислить значения $f(x_i)$ в выбранных узлах.

Предположим, что приближение к функции $f(x)$ уже построено в виде кубического интерполяционного сплайна:

$$f(x) \approx S_3(x) = \sum_{i=1}^{n} (a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3),$$

где $a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$, $x_{i-1} \le x < x_i$.

В построенном сплайне $S_3(x)$ все коэффициенты $\{a_i, b_i, c_i, d_i\}_{i=1}^{n}$ известны.

Проинтегрируем построенный сплайн:

$$\int_a^b S_3(x)dx = \int_a^b \sum_{i=1}^{N} (a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3)dx \Longrightarrow$$

$$\int_a^b S_3(x)dx = \sum_{i=1}^{N} \int_{i-1}^{i} (a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3)\, dx =$$

$$= \sum_{i=1}^{N} \left(\int_{x_{i-1}}^{x_i} a_i\, dx + \int_{x_{i-1}}^{x_i} b_i(x - x_{i-1})\, dx + \int_{x_{i-1}}^{x_i} c_i(x - x_{i-1})^2\, dx + \int_{x_{i-1}}^{x_i} d_i(x - x_{i-1})^3\, dx \right) =$$

$$= \sum_{i=1}^{N} \left(a_i(x_i - x_{i-1}) + \frac{b_i}{2}(x_i - x_{i-1})^2 + \frac{c_i}{3}(x_i - x_{i-1})^3 + \frac{d_i}{4}(x_i - x_{i-1})^4 \right).$$

Последняя сумма является приближённым значением исходного интеграла. Если шаг разбиения постоянный, тогда интеграл сводится к более простому виду

$$\int_a^b f(x)dx = \sum_{i=1}^{N}\left(a_i h + \frac{b_i}{2}h^2 + \frac{c_i}{3}h^3 + \frac{d_i}{4}h^4\right). \qquad (9.13)$$

ОЦЕНКА ПОГРЕШНОСТИ

Из п. 5.1.5. известно, что если таблица исходной функции задана с постоянным достаточно малым шагом h ($|h| \ll 1$), тогда порядок точности построенного интерполяционного сплайна $S_3(x)$ равен $O(h^4)$. Порядок точности формулы (9.13) составляет $O(h^5)$.

9.2. Численное дифференцирование

Численное дифференцирование – это процесс нахождения численного значения производной некоторой таблично заданной функции. Существует два основных метода с разнообразными их модификациями.

9.2.1. Дифференцирование и интерполяционные функции

Один из простейших способов численного дифференцирования состоит в том, что по табличной функции строится, прежде всего, интерполяционный многочлен (глава 5). После этого происходит непосредственное вычисление производной необходимого порядка для заданного значения аргумента. В качестве примера используем кубический интерполяционный сплайн $S_3(x)$.

Предположим, что есть дифференцируемая функция $f(x) \in C^2_{[a;b]}$, но она задана таблично $\{f_i, x_i\}_{i=0}^n$. Вычислим производные $f'(x^*)$ и $f''(x^*)$ для двух случаев: значение аргумента x^* не совпадает с узловыми значениями $\{x_i\}_{i=0}^n$ и совпадает с узлами таблицы.

ПОРЯДОК ДИФФЕРЕНЦИРОВАНИЯ

1. Строим по заданной таблице кубический сплайн

$$S_3(x) = \sum_{i=1}^{n}(a_i + b_i(x - x_{i-1}) + c_i(x - x_{i-1})^2 + d_i(x - x_{i-1})^3).$$

2. Определяем номера ближайших узлов таблицы, между которыми находится число x^*: $x_{j-1} < x^* < x_j$.

3. Выбираем j-е звено сплайна:

$$Z_j = a_j + b_j(x - x_{j-1}) + c_j(x - x_{j-1})^2 + d_j(x - x_{j-1})^3.$$

4. Дважды дифференцируем полином Z_j:

$$Z'_j = b_j + 2c_j(x - x_{j-1}) + 3d_j(x - x_{j-1})^2, \qquad Z''_j = 2c_j + 6d_j(x - x_{j-1}).$$

5. Вычисляем значения производных:

$$f'(x^*) \approx Z'_j(x^*) = b_j + 2c_j(x^* - x_{j-1}) + 3d_j(x^* - x_{j-1})^2,$$

$$f''(x^*) \approx Z''_j(x^*) = 2c_j + 6d_j(x^* - x_{j-1}).$$

В результате за 5 простейших шагов удалось интерполировать значения производных в точке x^* от таблично заданной функции.

Если точка x^* совпадает с узлом x_{j-1}, тогда $f'(x^*) = b_j,\ f''(x^*) = 2c_j$.

Если точка x^* совпадает с узлом x_j, тогда $f'(x^*) = b_{j+1},\ f''(x^*) = 2c_{j+1}$.

9.2.2. Дифференцирование и многочлен Тейлора

С помощью разностной аппроксимации можно приближённо вычислить производные табличной функции, используя только её значения в заданных узлах. Для производных первого и второго порядков такие формулы были выведены в 6-ой главе (п. 6.3.3.) с помощью многочлена Тейлора и найден порядок их точности. Аналогичным образом можно вывести разностные формулы для производных более высокого порядка. Далее представлены четыре формулы центрально-разностной аппроксимации. Порядок их точности равен $O(h^2)$.

$$f'_i = \frac{f_i - f_{i-1}}{2h}, \qquad f''_i = \frac{f_{i-1} - 2f_i + f_{i+1}}{h^2},$$

$$f_i^{(3)} = \frac{f_{i+2} - 2f_{i+1} + 2f_{i-1} - f_{i-2}}{2h^3}, f_i^{(4)} = \frac{f_{i+2} - 4f_{i+1} + 6f_0 - 4f_{i-1} + f_{i-2}}{h^4}.$$

С помощью многочлена Тейлора выведены формулы более высокого порядка точности $O(h^4)$, но они очень громоздки.

9.2.3. Погрешность численного дифференцирования

Численное дифференцирование в общем случае является некорректной задачей. Во всех разностных формулах дифференцирования оценки их погрешности обратно пропорциональны квадрату шага h. Поэтому для табличной функции с уменьшением шага погрешность дифференцирования с некоторого критического значения h начнёт возрастать. Этот момент называют ***дилеммой длины шага***. Чем выше порядок производной, тем меньше порядок вычислительной погрешности.

Заключение

ЧИСЛЕННЫЕ МЕТОДЫ В ГИДРОМЕТЕОРОЛОГИИ

Мы успели сорок тысяч разных
книжек прочитать.
И узнали что к чему и что почём
И очень точно.
Б. Ш. Окуджава

Впервые проблему прогноза погоды, как задачу математики, поставили в 1904 г. Это была задача решения уравнений гидромеханики с начальными условиями.

В начале 20-х годов XX в. Ричардсон предложил метод численного прогноза погоды. Рассчитанное по его методу приземное давление в 50 раз превосходило реальное давление. Эксперимент проводился в окрестности немецкого города Нюрнберг. В настоящее время понятны причины такой неудачи. Вычисления делались с помощью логарифмической линейки, а условие Куранта–Леви, гарантирующее устойчивость сеточного метода, нашли только в 1928 г. Сейчас можно сказать, что конечно-разностный метод Ричардсона был просто неустойчив.

Благодаря созданию электронно-вычислительных машин в начале 50-х годов появился первый численный прогноз погоды.

В институте перспективных исследований в Принстоне группа учёных создала ряд математических моделей различных атмосферных процессов. В эту группу входил знаменитый Джон фон Нейман. Он считал, что прогноз погоды может стать верным только с помощью компьютеров.

В начале 60-х годов появилась первая девяти-уровенная модель. Она была основана на неупрощённых уравнениях в частных производных.

Дальнейший прогресс в развитии вычислительной техники привёл к созданию более точных математических моделей и к улучшению технологии прогноза погоды. Выяснилось, что основные климатические характеристики, полученные с помощью различных моделей, а затем осреднённые по всему набору этих моделей, гораздо ближе к реальным характеристикам, чем полученные с помощью отдельных моделей.

В России разработка численного решения математических моделей поведения климата началась только в 70-х годах XX в. по инициативе президента Академии наук СССР Гурия Ивановича Марчука.

В настоящее время «Численные методы» – хорошо разработанный и теоретически обоснованный раздел современной высшей математики, благодаря которому создано надёжное программное обеспечение для решения множества прикладных задач. Следует отметить, что все встроенные функции в программном обеспечении, куда входят $\sin x$, $\cos x$, e^x, $\ln x$, $\tan x$, и др., вычисляются в компьютерах с очень высокой точностью благодаря именно рациональной аппроксимации.

Приведём несколько простейших примеров, связанных с работой нашего университета.

1. В первичной обработке наблюдений повсеместно используется теория аппроксимации: метод наименьших квадратов и гармонический анализ.

2. Метод конечных элементов давно стал классическим. Он используется в гидрологи и океанологии для решения уравнений мелкой воды с учётом естественной границы (берегов) данного океана.

3. Задача на собственные числа также используется в метеорологии для исследования спектра собственных колебаний в системе океан-атмосфера.

Литература

1. Александрова Н.В. История математических терминов, понятий, обозначений: Словарь-справочник. Изд 3-е, исп. – М.: Изд. ЛКИ, 2008 –248 с.

2. Барахнин В.Б., Шапеев В. П. Введение в численный анализ: Учебное пособие. – СПб.: Издательство «Лань», 2005. – 112 с.

3. Вагер Б.Г. Численные методы решения дифференциальных уравнений: Учебное пособие для вузов. – РГГМУ – СПб., 2004.–110 с.

4. Вагер Б.Г. , Серков Н.К. Сплайны при решении прикладных задач метеорологии и гидрологии. Монография. Ленинград. Гидрометеоиздат, 1987. –160 с.

5. Гордин В.А. Математика, компьютер, прогноз погоды и другие сценарии математической физики. – М.: ФИЗМАТЛИТ, 2010. – 736 с.

6. Гордон Е.П., Лыкосов В. Н. и др. – Новосибирск: Наука, 2013. –199 с. Вычислительно-информационные технологии мониторинга и моделирования климатических изменений и их последствий.

7. Вержбицкий В. М. Численные методы (линейная алгебра и нелинейные уравнения): Учебное пособие для вузов. – Выс. шк., 2000. – 266 с. ил.

8. Данилов Ю.А. Многочлены Чебышева. – М.: Едиториал УРСС, 2003.– 160 с.

9. Демидович Б. П., Марон И. А., Шувалова Э. З. Численные методы анали-за. – М.: Физматгиз, 1963. – 600 с.

10. Джон Г. Мэтьюз, Куртис Д. Финк Численные методы. Использование MATLAB. 3-е издание: Пер. с англ. – М.: Издательский дом «Вильямс», 2001. – 720 с.

11. Дж. Холл, Дж. Уатт Современные численные методы решения обыкновенных дифференциальных уравнений. – М.: «Мир», 1979. – 312с.

12. Ефимов А.В., Золотарёв Ю. Г. Математический анализ (специальные разделы). – М.: «Высшая школа», 1980. – 296 с.

13. Зализняк В.Е. Численные методы. Основы научных вычислений: учебник и практикум для бакалавриата.– М.: Издательство Юрайт, 2015. – 356 с.

14. Земляков А.Н. Введение в алгебру и анализ: культурно-исторический дискурс. Элективный курс: Учебное пособие.–М.: БИНОМ. Лаборатория знаний, 2007.– 320 с.

15. Зельдович Я.Б., Мышкис А.Д. Элементы прикладной математики. – М.: «Наука», 1972. – 592 с.

16. Калиткин Н. Н. Численные методы. Главная редакция физико-математической литературы изд-ва «Наука», М., 1978. – 512 с.

17. Краснов М. Л., Макаренко Г. И., Киселёв А. И. Вариационное исчисле-ние. – «Наука», М.: 1973. –192 с.

18. Курош А.Г. Курс высшей алгебры.– «Наука», М.: 1965. –432 с.

19. Лыкосов В.Н., Глазунов А. В. и др. Суперкомпьютерное моделирование в физике климатической системы: Учебное пособие. – М.: Издательство Московского университета, 2012. – 408 с., ил.

20. Марчук Г.И. Методы вычислительной математики. – М.: «Наука», 1989. – 608 с.

21. Микаэль Лоне. Большой роман о математике. – Москва: Эксмо. 2018. – 320 с.– (Non – fiction. Best).

22. Михлин С. Г. Курс математической физики. – М.: «Наука», 1968. – 576 с.

23. Мышкис А.Д. Математика, специальные курсы. – М.: «Наука», 1971. – 632 с.

24. Петров Ю. П. История и философия науки. Математика, вычислительная техника, информатика. – СПб.: БХВ- Петербург, 2005. –448 с.: ил.

25. Ричард Браун (редактор). Математика за 30 секунд.– М.: « РИПОЛ класс-сик», 2015.– 160 с.

26. Успенский В.А. Апология математики. – СПб.: Амфора.2009. – 554 с.

27. Фаддеев М. А., Марков К. А. Основные методы вычислительной математики: Учебное пособие. – СПб.: Издательство «Лань», 2008. –160 с. ил. – (Учебники для вузов. Специальная литература).

28. Файншмидт В.Л. Дифференциальное и интегральное исчисление функций нескольких аргументов (Учебник). – СПб.: БХВ-Петербург, 2007. – 208 с.: ил.

29. Шевцов Г. С., Крюкова О. Г., Мызникова Б. И. Численные методы линейной алгебры: Учебное пособие. – СПб.: Издательство «Лань», 2011. – 496 с.

30. Шумихин С. Шумихина А. Число π. История длиною в 4000 лет. – М.: Эксмо, 2011. – (Тайны мироздания).

ПРИЛОЖЕНИЕ 1

Толковый и этимологический словарь математических терминов и понятий

Уточняйте значение слов,
и вы избавите человечество
от половины заблуждений.
Рене Декарт

А

АЛГЕБРАИЧЕСКОЕ УРАВНЕНИЕ

Уравнение называется алгебраическим, если оно представимо в виде

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n = 0,$$

где коэффициенты $\{a_i\}_{i=1}^{n}$ – заданные действительные числа. Решить данное уравнение, значит найти такие значения величины x, при которых это уравнение превращается в тождество. Найденные значения x называются корнями уравнения.

В XVI в. произошло величайшее открытие в алгебре. Итальянские математики Тарталья, Кардано и Феррари вывели формулы для нахождения корней уравнений 3-й и 4-й степеней.

В течение следующих трёх веков лучшие математические умы пытались вывести формулы для точного нахождения корней уравнений 5-ой степени.

В начале XIX в. очень молодой гениальный норвежский математик Нильс Абель вывел эти формулы. Многие известные математики того времени проверяли их и сочли правильными. Ошибку нашёл сам Нильс Абель. В результате размышлений о причине ошибочного результата он доказал (1823) , что в общем случае алгебраическое уравнение выше 4-й степени не решается точно с помощью четырёх арифметических действий и извлечения корней (в радикалах).

АЛГОРИТМ

Последовательность точно описанных операций, которые однозначно выполняются в определённом порядке и преобразуют исходные данные в решение задач заданного класса.

Слово произошло от географического названия – Хорезм. Так называли древнее государство Средней Азии в низовьях реки Амударьи. В конце VIII в. и начале IX в. там жил великий арабский математик Абу Абдуллах Мухаммад ибн Муса аль-Хорезми (аль-Хорезми – из Хорезма).

Этот учёный предложил некоторые методы решения различных задач арифметики. В средневековой Европе знали работы Мухаммада ибн Мусы. Его полное имя писали на латинском языке (.....аль-Хорезми – Algoritmi).

С античных времён математики искали общие методы и порядок решения различных задач математики. Но понимание алгоритма как самостоятельного математического инструмента встречается только с 1912 г. в работах Э. Бо-

реля – великого французского математика. Современное понятие было закреплено за словом *алгоритм* только в 30-х годах XX века в работах А. Тьюринга, Е. Поста, А.Маркова и др.

На русском языке слово впервые появилось в 1835 г. В Петербурге долго говорили и писали *алгорифм.*

Программа для компьютера – это запись алгоритма решения задачи или класса задач на определённом языке программирования. Со второй половины XX века у каждой стандартной компьютерной программы есть свой алгоритм, разработанный с максимальной степенью общности и максимальным быстродействием с учётом современных достижений в математике.

АНАЛИЗ

Греческое слово *αναλυσις – решение, разрешение*. Сначала слово *анализ* означало переход от данной единицы к низшей. В «Началах» Евклида встречаются слова *анализ* и *синтез.* В новую математику термин настойчиво вводил Ф. Виет.

АППРОКСИМАЦИЯ

Приближённое выражение математических объектов через другие, более простые**.** Буквальное значение термина – *приближение*. Термин происходит от латинского *approximo – приближаюсь*. Основы теории аппроксимации заложили Вейерштрасс и Чебышёв в середине XIX в.

Б

БАЗИС

Набор из n линейно независимых элементов в n-мерном векторном или функциональном пространстве.

В

ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

Первая задача этой ветви математического анализа известна с глубокой древности. Она называется задачей Дидо́. Согласно легенде Дидо́ была царицей одного из государств Древней Греции. Спасаясь от преследования царя соседнего государства, она попала в Северную Африку. Там попросила у местных жителей участок земли, который можно покрыть шкурой вола. Аборигены удивились её скромности и сразу дали согласие. Царица резала шкуру на тонкие полоски и связывала их друг с другом. В результате получилась очень длинная и тонкая нить, которой она охватила достаточно большой участок земли. Там Дидо развернула строительство, которое потом превратилось в знаменитый город древности Карфаген.

Начиная с XVII века у математиков стали появляться задачи о нахождении функций, обеспечивающих максимум или минимум некоторого свойства исследуемого объекта. Существенно позднее, после изобретения общего метода решения таких задач, метод стал называться «Вариационным исчислением», а найденные функции стали называть экстремалями.

В 1687 г. аналогичную задачу стал изучать Ньютон. Ему надо было выбрать такую форму корабля, при которой сопротивление воды стало бы наименьшим. Швейцарский математик Иоганн Бернулли (1667 – 1748) первый попытался обобщить такого рода задачи. Он предложил математикам за год решить задачу о брахистохроне: «*Среди всех линий, соединяющих две заданные точки, найти кривую, двигаясь по которой под действием только силы тяжести и без трения, материальная точка пройдёт этот путь за минимальное время*». Лейбниц решил эту, по его мнению, прекрасную задачу. Кроме Лейбница её решили ещё три человека: Якоб Бернулли, Лопиталь и Ньютон.

Группе математиков, включая И. Бернулли, его старшего брата Я. Бернулли и Лейбница, удалось ещё решить некоторые задачи такого типа. Но честь создания совершенно нового метода решения таких задач принадлежит только Эйлеру. Ему было всего 25 лет, когда он получил первые блестящие результаты в этом направлении.

Второй этап развития «Вариационного исчисления» связан с Ж. Лагранжом (1736 – 1813) и А. Лежандром (1752 – 1833). Французский математик Лагранж в 19 лет придумал новый метод решения – метод вариации и сразу написал об этом мировой математической звезде Эйлеру. Ответ был быстрым. Метод привёл в восторг Эйлера. Переписка между ними продолжалась много лет.

Следующий этап развития связан с немецкими математиками К. Якоби (1804 – 1851) и К. Вейерштрассом (1815 – 1897). В XX веке этот раздел математики продолжали успешно развивать. Задачами вариационного исчисления занимались и получили прекрасные результаты математики России: Н. Крылов, Н. Боголюбов, В. Кротов, Н. Гернет, Б. Давыдов Л. Понтрягин, Л. Канторович, Л. Михлин и др.

В 1908 г. немецкий математик и физик Ритц разработал эффективный метод нахождения приближённых решений вариационных задач. Теоретическое обоснование метода Ритца было сделано только в 1929 г. русскими математиками Н. Крыловым и Н. Боголюбовым.

До настоящего времени расширяется диапазон применения вариационного исчисления. Оно используется в геометрии, физике, в квантовой механике, теории оптимального управления, в ракетной и космической технике.

ВЕЛИЧИНА

Величиной называют параметр, функцию, коэффициент или другие математические объекты, если они равны каким-то числам.

ВЫРАЖЕНИЕ

Формула или некоторая её часть.

Г

ГЛАДКАЯ ФУНКЦИЯ

Функция, имеющая непрерывную хотя бы первую производную. Чем выше порядок непрерывной производной, тем выше порядок гладкости функции.

Д

ДЕДУКЦИЯ

Латинское слово *deduction – выведение*. Способ рассуждения, при котором новое положение выводится чисто логическим путём из некоторых исходных утверждений.

ДИССИПАЦИЯ

Мера скорости рассеяния.

ДЕТЕРМИНАНТ

Слово происходит от латинского *determine – ограничивать, определять*. Впервые встречается у Гаусса (1801). Но у него это слово означало «дискриминант квадратичной формы». В современном значении это слово стал употреблять Коши.

Идея введения детерминанта принадлежит Лейбницу. Он пришёл к детерминантам при решении линейных систем уравнений (1700). В 1750 г. детерминанты были изобретены вновь женевским математиком Габриэлем Крамером (1750), а затем Лагранжем (1770). Крамер ввёл знакомые сегодня термины: строка и столбец детерминанта.

Метод Крамера очень быстро вошёл в школьную математическую программу тех времён. Французский математик Вандермонд опубликовал первое исследование, посвящённое свойствам детерминанта. Якоби (1826 – 1841) обозначил его греческой буквой Δ и стал использовать слово *element* для описания определителя.

Правило вычисления определителя 3-го порядка методом треугольников придумал страсбургский профессор Саррюс (1846). Теорема о разложении определителя по элементам любой строки или столбца встречалась в частных случаях у Лапласа, Вандермонда и Безу, но только Коши смог её полностью сформулировать и доказать.

Первое полное изложение теории определителей принадлежит Коши. Он посвятил этой теории 14 мемуаров и превратил её в самостоятельный раздел математики. Ему принадлежит термин *детерминант n-го порядка*.

Следующим этапом развития теории определителей стали 30 работ Якоби с завершающей статьёй «О построении и свойствах детерминантов». Он ввёл и первый использовал функциональные определители.

Английский математик Сильвестр впервые назвал якобианом известный функциональный определитель, чтобы воздать должное трудам Якоби. Теория бесконечных детерминантов была заложена работами Пуанкаре.

ДИСКРЕТНОСТЬ

Термин образован от латинского слова *discretus – сложенный из отдельных частей, разорванный*.

Е

e

Впервые это число появилось в 1618 г. в работе «Описание удивительной таблицы логарифмов» шотландского астронома, изобретателя логарифмов Джона Непера. Он использовал в качестве основания логарифма число e^{-1}, которое называли числом Непера. В 1668 г. немецкий математик Меркатор (Кауфман) первый стал использовать термин *натуральный* для логарифма по основанию e. В 1683 году Я. Бернулли доказал, что

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e.$$

Это был первый случай в истории математики, когда число определялось как предел. Обозначение e введено Эйлером. Этот символ сразу стал общепринятым. Скорее всего, Эйлер выбрал такое обозначение, потому что e – первая буква латинского слова *exponentia– показательный.*

Иррациональность и трансцендентность числа e была доказана только в XIX в. французскими математиками Ж. Лиувиллем и Ш. Эрмитом.

З

ЗОЛОТОЕ СЕЧЕНИЕ

Среди различных средних величин уникальными свойствами обладает одно, делящее данный отрезок длины a на две части b и $a-b$ в геометрической пропорции. Это означает, что отношение целого отрезка a к его большей части b равно отношению большей части b к его меньшей части $a-b$:

$$a:b = b:(a-b).$$

Пусть $a:b = \Phi$, тогда эту пропорцию можно преобразовать к виду

$$\Phi^2 - \Phi - 1 = 0.$$

Полученное уравнение имеет два положительных корня:

$$\Phi = \frac{\sqrt{5}+1}{2} = 1{,}618\cdots, \qquad \varphi = \frac{\sqrt{5}-1}{2} = 0{,}618\cdots.$$

Эти два корня – иррациональные, взаимно обратные числа. Их сумма равна единице. Константы Φ и φ являются фундаментальными константами.

Золотая пропорция была известна ещё пифагорейцам. В эпоху Возрождения Леонардо да Винчи назвал её золотым сечением. Его широко используют в архитектуре, скульптуре, живописи, музыке и математике.

Древние греческие вазы обычно имеют очень красивую форму. Анализ их размеров привёл к интересной закономерности. Большинство таких ваз вписываются в прямоугольник с отношением сторон, равным золотому сечению.

Золотое сечение есть в геометрических фигурах с осью симметрии пятого порядка – это пятиконечная звезда и пятиугольник. Формы различных цветов, морских звёзд, морских ежей и формы многих других живых объектов имеют золотое сечение. Эту интересную закономерность описал Лука Пачоли в книге «О божественной пропорции», иллюстрированную его другом Леонардо да Винчи.

В 1855 г. немецкий учёный Цейзинг в работе «Эстетические исследования» показал, что золотое сечение близко к отношению высоты человека к расстоянию между его пупком и подошвами ног. Сразу после рождения человека это отношение равно двум. Постепенно оно уменьшается и примерно к 21 году становится равным 1,625. Эта величина очень близка числу Ф.

Опыты немецкого психолога Фехнера доказали, что среди различных отношений человек, как правило, выбирает золотое сечение. В 1876 г. он показал множеству людей различные прямоугольники и просил выбрать из них тот, который больше всего нравится своей формой. Большинство выбрало прямоугольник с отношением сторон, равным отношению золотого сечения.

Удивительна связь между золотым сечением и законами расположения листьев растений, геометрией живых организмов, пропорцией тела человека. Эта связь не случайна. Однако причины такой связи пока неизвестны.

И

ИЗОЛИРОВАННЫЙ

Термин происходит от французского слова *isole – отдельный*.

ИЗОМОРФИЗМ

Взаимно однозначное соответствие между определёнными алгебраическими системами. В таких алгебраических системах сохраняются операции и их свойства. Изоморфные системы одинаковы в математическом смысле. Можно сказать, что изоморфизм – существенное внутреннее сходство, тождественность в главном.

ИНВАРИАНТ

Французское *invariant – неизменяющийся*. В математике инвариантом называют выражение, которое остаётся неизменным при некоторых преобразованиях переменных. Инвариантом является первый дифференциал функции одного аргумента. Форма дифференциала не меняется при переходе к новой независимой переменной.

ИНДЕКС

Латинское *index – показатель, указатель, титул, надпись*. С развитием книгопечатания индексы стали писать ниже строки. Первым это сделал Лейбниц.

ИНТЕГРАЛ

Впервые (1690) это слово использовал Якоб Бернулли. Возможно, он образовал этот термин от латинского *integer – целый* или от другого слова *integro – восстанавливать*. После долгих обсуждений Лейбниц и Иоганн Бернулли согласились с этим названием в 1696 г. Знак интеграла придумал Лейбниц. Он

просто стал использовать с 1675 г. первую букву слова *summa*, которая в те времена писалась как современный знак интеграла.

ИНТЕРВАЛ

Термин происходит от латинского *intervallum – промежуток, расстояние*.

ИНТЕРПОЛЯЦИЯ

Восстановление значений табличной функции в заданной области с помощью известных значений этой функции внутри той же области. Термин происходит от латинского *interpolare – подделывать, подновлять*. Это слово первоначально означало подделку рукописи. Слово в его современном смысле впервые употребил английский математик Д. Валлис (1656) при составлении астрономических и математических таблиц.

ИНТЕРПОЛЯЦИОННЫЕ ПОЛИНОМЫ (МНОГОЧЛЕНЫ)

Многочлены, численные значения которых совпадают со значениями табличной функции в заданных узлах. С помощью таких многочленов восстанавливаются значения табличной функции между узлами.

Интерполирование с помощью полиномов впервые ввёл Джеймс Грегори в 1670 г. Интерполяционная формула Ньютона была опубликована в 1711г., Лагранжа в 1792 г. Дальнейшее развитие теории интерполирования связано с Гауссом, Коши, Энке, Леверье и Чебышевым. Независимо друг от друга Рунге (1904) и Борель (1903) показали, что рост степени многочлена не увеличивает точность приближения.

С 1918 г. началась строгая разработка теории интерполирования.

ИНТЕРПРЕТАЦИЯ

Латинское *interpretatio – толкование, объяснение*.

ИТЕРАЦИЯ

Термин появился только в XIX в. Происходит от латинского *iteratio – повторение*. Английское слово *iteration – повторение, итерация.* В математике этим словом называют процесс повторного применения совокупности математических операций при решении различных задач. Метод итерации ещё называют методом последовательных приближений. Первые попытки изучения методов итераций сделал Л. Эйлер (1778).

ИТЕРИРОВАНИЕ

Переход от одного шага итерации к следующему шагу.

ИНТЕРПОЛИРОВАНИЕ

Термин используют в тех случаях, когда надо восстановить значение функции в точке, которая расположена между крайними (граничными) узлами таблично заданной функции.

К

КАНОН

Слово образовано от греческого слова *κανον – предписание, правило*.

КАНОНИЧЕСКИЙ

От греческого слова *κανονικοζ – составленный по правилам, нормальный*.

КВАДРАТУРА

От латинского *quadratura – придание квадратной формы.* Смысл этого слова связан с историей греческой математики. Вычисление площади плоской фигуры греки сводили к построению квадрата той же площади. С точки зрения геометрии определенный интеграл – это тоже площадь. Ещё в начале XIX в. математики считали, что определённый интеграл следует интерпретировать только как площадь. Поэтому процесс вычисления определённого интеграла называли и сейчас по традиции называют квадратурой.

КВАНДРАНТ

Один из четырёх углов на плоскости, образованных двумя перпендикулярными осями координат. Четверть круга. Сектор с центральным углом в 90^0 .

КОРРЕКТНОСТЬ

От латинского *correctement – правильно.*

КОЭФФИЦИЕНТ

Числовой множитель при буквенном выражении, известный множитель при неизвестном выражении, постоянный множитель при переменной величине. Термин составлен из латинских *co – вместе* и *efficiens – производящий*. Буквальное значение слова *coefficiens – содействующий*. Систематически стал использовать этот термин в математике английский математик Д. Валлис (XVII в).

КРАТНОЕ

Число, равное данному числу, умноженному на другое целое число.

Л

ЛЕММА

Вспомогательное утверждение, которое используется для доказательства тех или иных теорем.

ЛОГАРИФМ

Это изобретение Д. Непера (1594) – шотландского математика, астронома и друга Кеплера. Свою работу о логарифмах Непер опубликовал только через 20 лет. В математическом мире логарифмы встретили с полным восторгом. Через сто лет после публикации этой работы Лаплас писал: «*Изобретение логарифмов превратило труд нескольких месяцев в вычисления нескольких дней, удваивая этим жизнь астрономам*».

Непер придумал слово *логарифм*, опираясь на греческие слова $\lambda\acute{o}\gamma o\varsigma$ и $\acute{a}\rho\iota\theta\mu o\varsigma$ – числа отношений. Дело в том, что логарифмы возникли при сравнении членов двух числовых последовательностей. Немецкий математик Меркатор (1668) назвал логарифмы с основанием e натуральными. Выражение *logarithmi naturali* использовали Галлей (1695) и Коши (1823).

Термин *логарифм* получил современный смысл благодаря английскому математику Д. Валлису.

ЛОКАЛЬНЫЙ

Термин происходит от латинского *localis – местный*.

М

МАКСИМУМ, МИНИМУМ

Это русское изображение латинских слов *maximum, minimum – наибольшее, наименьшее*. Отдельные задачи на нахождение экстремумов были решены древнегреческими математиками. До XVII в. для решения каждой такой задачи составлялся индивидуальный метод. Первый общий алгоритм изобрёл П. Ферма (1629).

МАТРИЦА

Термин происходит от латинского слова *matrix*. Впервые этот термин стали использовать в середине XIX в. Квадратные матрицы 2×2, 3×3 есть в работах Эйлера (1771). Он называл их просто квадратами. Гаусс использовал матрицы для линейных преобразований (1801) и называл их таблицами.

МЕТОД

От греческого слова $\mu\varepsilon\theta o\delta o\varsigma$ – *дорога, ведущая за чем-либо*. Платон и Аристотель стали использовать это слово как название совокупности математических действий, необходимых для получения правильного результата.

– **конечных разностей** (разностей)

Этот метод возник и развивался теоретически на протяжении XVII в. Его использовали для составления тригонометрических, логарифмических, навигационных и других таблиц, необходимых в географии, небесной механике, теоретической и практической астрономии. Развитием этого метода занимались Кеплер, Гюйгенс, Ньютон и др.

– наименьших квадратов

Впервые метод был опубликован Лежандром в 1805 г. Гаусс прочёл эту публикацию и заявил, что он пользовался этим методом ещё в 1795 г. Спор о приоритете не разгорелся только потому, что математический мир признал, что оба учёных открыли метод независимо друг от друга.

Гаусс обосновал метод тремя различными способами (1821–1823). Лаплас после статьи Лежандра связал метод с теорией вероятностей. Гаусс тоже использовал метод наименьших квадратов при решении некоторых задач теории вероятностей.

МЕТРИКА

Неотрицательная функция $\rho(x, y)$ двух точек множества, которая удовлетворяет трём условиям:

1. $\rho(x, y) = 0 \Leftrightarrow x = y$;

2. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$;

3. $\rho(x, y) = \rho(y, x)$.

МЕТРИЧЕСКОЕ ПРОСТРАНСТВО

Пространство, в котором (на котором) можно измерить расстояние между любыми элементами этого пространства.

МОДУЛЬ

От латинского *modulus – мера*. Используют выражения: *модуль функции, модуль вектора, модуль комплексного числа.*

Н

НЕСОВМЕСТНОСТЬ

Система называется несовместной, если у неё нет решений, удовлетворяющих всем уравнениям системы**.**

НОЛЬ

Особое место в позиционной системе занимает число 0. У нуля не было геометрического образа, поэтому он очень долго и тяжело входил в мир чисел. Ноль признали числом только в XVII в. после работ Рене Декарта.

Число ноль возникло в Индии. Только там увлекались огромными числами. Например, у индусов было 68000 божеств, у брамов сто миллионов божеств. В легенде о мудром Арджуме пытались вычислить количество всех мельчайших частиц во вселенной. Именно индийский ум создал понятие небытия, которое играет важную роль в буддийской философии. Всё это привело к введению знака, который обозначал отсутствие цифры. Индусы называли такой знак словом *sunya*, что значит *пустой*. Арабы перевели это слово по смыслу и получили слово *al-sifr*. Леонардо Фибоначчи нашёл соответствующее слово в латинском языке, поэтому называл в своей книге современный ноль словом *zephirum*. Отсюда произошло французское и английское слово *zero* для обозначения нуля. Сторонники индийской нумерации в Италии сохранили арабское звучание *cifra*, которое с XIII в. перешло во все европейские языки. На французском – *chiffre*, на немецком – *ziffer*, на английском – *cipher* .

В латинских переводах арабских трактатов XII в. знак 0 назывался кружком *ciculus* или *nulla figura – никакой знак*. В XV в. итальянцы заменили арабское слово *cifra* словом *nulla.* Его заимствовали немцы и стали писать *null*. Из Германии нуль перешёл в Россию. Введение нового термина освободило арабское слово *cifra* от значения 0. Его стали постепенно использовать для обозначения числового знака.

Слово *цифр*а очень долго использовалось для обозначения нуля. Ещё в 1783 г. Эйлер использовал вместо слова *ноль* термин *цифра*. Также поступал Гаусс. В Англии такое же использование слова можно встретить в произведениях Шекспира и Теккерея. Английское слово *cipher* до сих пор сохраняет значение *ноль*. В России в конце XVIII в. число 0 называли цифрой. В это же время появилось в русском языке слово нуль, а позже ноль. Европейцы не вдавались в историю происхождения изображения чисел. Поэтому привычные для нас числа называли и сегодня часто называют арабскими**.**

НОРМА

Неотрицательное число $\|x\|$, которое сопоставляется каждому элементу некоторого множества и удовлетворяет следующим условиям:

1. $\|x\| = 0 \Longrightarrow x = 0;$

2. $\|\beta x\| = |\beta| \cdot \|x\|$, где $\beta \in \mathbb{R}$;

3. $\|x + y\| \leq \|x\| + \|y\|$.

– **вектора**

В качестве нормы можно взять квадратный корень из скалярного произведения вектора на себя:

$$\|\bar{a}\| = \sqrt{\bar{a}^2}.$$

– **функции**

В пространстве функций, интегрируемых с квадратом: $f(x) \in L_2[a, b]$, норма вычисляется по формуле

$$\|f(x)\| = \sqrt{\int_a^b f^2(x)dx}.$$

Всё множество интегрируемых с квадратом функций называется гильбертовым пространством.

– **матрицы**

В пространстве матриц порядка $(n \times m)$ часто используют нормы:

1. $\|A\| = \max_j \sum_i |a_{ij}|$ – наибольшая из сумм модулей элементов матрицы по столбцам;

2. $\|A\| = \max_i \sum_j |a_{ij}|$ – наибольшая из сумм модулей элементов матрицы по строкам;

3. $\|A\|_E = \left(\sum_{ij} |a_{ij}|^2\right)^{\frac{1}{2}}$ – евклидова норма.

О

ОКТАНТ

Одна из восьми частей, на которые разбивается трёхмерное пространство координатными плоскостями декартовой системы координат.

ОПЕРАТОР

Знак, символ, обозначение некоторого математического действия или преобразования. Термин произошёл от латинского *operator – работник*. Например, символ $\sum -$ оператор суммирования.

ОПРЕДЕЛИТЕЛЬ – см. детерминант**.**

ОРТ

Термин ввёл английский математик Оливер Хевисайд (1892) как сокращение слова *orientation – ориентация*.

ОСЦИЛЛЯЦИЯ

Термин произошёл от английского слова *oscillation – качание, колебание, вибрация.*

П

ПАРАМЕТР

От греческого слова $\pi\alpha\rho\alpha\mu\varepsilon\tau\rho\varepsilon\omega$ – *измеряю что-нибудь, сравнивая с чем-то другим*. Лейбниц называл параметром любую произвольную постоянную величину, входящую в уравнение.

ПЕРЕМЕННАЯ

Термин ввёл в конце XVII в. Лейбниц при обсуждении понятия функции.

π

Значение отношения длины окружности к её диаметру. Иногда это число называют *константой Архимеда*. Его вычисление проводились ещё с IV в. до н. э. В библии было написано, что это отношение равно трём. В разных странах это число находили с различной степенью точности. В IV в. н. э. китайский математик Лю Хуэй нашёл приближение $\pi \approx 3{,}14159$. Эта точность удерживалась до XV в. Тогда в обсерватории Улугбека в Самарканде получили приближение числа π с точностью до 16-и знаков после запятой. В 1673 г. Лейбниц открыл числовой ряд

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots.$$

Первый раз привычное для нас обозначение появилось в 1706 г. в книге Уильяма Джонса «Новое введение в математику». Знаменитую константу назвали, скорее всего, таким именем по первой букве греческого слова $\pi\varepsilon\rho\iota\mu\varepsilon\tau\rho o\varsigma$ – *периметр*. Позже Эйлер стал постоянно пользоваться этим обозначением. Окончательно знаменитое число стали называть числом π с 1736 года только благодаря Эйлеру. Французский математик А. Лежандр (1752 – 1833) доказал иррациональность этого знаменитого числа. В 1882 г. немецкий математик Ф. Линдеман доказал трансцендентность этого числа. Более того, математики нашли связь между удивительными константами π и e. Ниже представлены простейшие из них

$$\pi^{\mathrm{e}} = \mathrm{e}^{\pi};\ \sqrt{\pi^{\pi}} + \sqrt{\mathrm{e}^{\mathrm{e}}} = \pi^2;\ \pi\mathrm{e}^{\,2} = \mathrm{e}^{\pi};\ \ \frac{\pi^2 + \mathrm{e}^{\,2}}{\pi^3 - \mathrm{e}^{\,3}} = \frac{\pi}{2}.$$

Сегодня нет проблем вычислить число π с любой степенью точности. Интересно отметить, что в Америке и Франции есть праздник числа π.

ПОЛНАЯ СИСТЕМА ФУНКЦИЙ

Если любую функцию некоторого метрического пространства можно сколь угодно точно аппроксимировать линейной комбинацией из заданной системы функций $\{\varphi_i\}_{i=1}^{\infty}$ этого же метрического пространства, такая система функций называется полной.

ПРЕДЕЛ

В начале XIX в. математики давали только словесное описание предела. Современное обозначение предела придумали Гамильтон, Вейерштрасс и Риман. В русском издании лекций Коши (1831) слово *limit* перевели как *предел* и обозначение *lim* заменяли на *пр.*

ПРИРАЩЕНИЕ

Разность двух значений переменной величины.

ПОСТУЛАТ

Термин образован от латинского слова *postulare – требовать*. В греческой геометрии постулаты играли роль аксиом. Примером этому является знаменитый пятый постулат геометрии Евклида. Он гласит, что через одну точку на плоскости можно провести только одну прямую, не пересекающуюся с данной прямой.

ПРОБЛЕМА

От греческого слова $\pi\rho o\beta\lambda\eta\mu\alpha$ *– то, что поставлено вперёд, задание*.

ПРОГРАММА

От греческого слова $\pi\rho o\gamma\rho\alpha\mu\mu\alpha$ *– объявление, наказ.*

Р

РАВЕНСТВО

До появления специального знака слово *равенство* писали в математике на разных языках. В 1557 г. английский математик и врач Рекорд предложил знак =. Он писал: «*Ничего нет более равного, чем две параллельные прямые»*. Этот знак равенства широко распространился только к концу XVIII века. К сожалению, изобретатель знака, который наиболее часто встречается в математике, умер в Лондонской долговой тюрьме.

РАДИКАЛ

Радикалом в математике называют знак $\sqrt[n]{\ }$ извлечения корня произвольной степени $n \geq 2$ Декарт ввёл знак квадратного корня с чертой над подкоренным выражением $\sqrt[2]{\ }$, а Ньютон распространил обозначение Декарта на корни любой степени, используя его в книге «Всеобщая арифметика».

Древние греки вместо слов *извлечь корень* говорили: «*найти сторону по данной площади квадрата*», а квадратный корень называли стороной. На латинский язык слова *сторона, бок*, *корень* переводятся одним и тем же словом *radix*. Термины *корень* и *радикал* стали математическими благодаря первым переводчикам «Начал» Евклида. Переводили сначала с греческого языка на арабский язык, а затем на латынь.

РЕКУРРЕНТНОСТЬ

От латинского *recuro – бегу назад, возвращаюсь*. Очень удобное свойство некоторых последовательностей, при котором любой их член может быть вычислен через предыдущие члены этой же последовательности. Формула, с помощью которой вычисляется любой член такой последовательности, называется рекуррентной формулой.

С

СЕКУЩАЯ

Прямая, пересекающая заданную линию по меньшей мере в двух точках.

СИГНУМ

От латинского слова *signum – знак.*

СИМВОЛ

Русское написание греческого слова $\sigma\upsilon\mu\beta o\lambda o\varsigma$ – *условный знак, примета*. Древние греки обозначали символом устный опознавательный знак (пароль) для членов определённого общества.

СКАЛЯР

От латинского *scale – шкала, лестница.* Французский математик Виет первым стал называть величины *скалярами*. В таком смысле слово скаляр впервые вошло в математику. Современное понятие *скалярная величина*, в отличие от векторной, придумал Гамильтон.

СОБСТВЕННОЕ ЗНАЧЕНИЕ (собственное число**)**

От английского *proper value.* Известный математик Гамильтон называл такие величины собственными числами. Использовали название *characteristic value – характеристическое значение.* Современное название становится общепринятым только с начала XX века.

СООТНОШЕНИЕ

Равенство или неравенство, выражающее зависимость между величинами.

СПЛАЙН

От английского *splain – гибкая линейка*. Термин происходит от названия тонкой деревянной или металлической полоски, которую использовали чертёжники для проведения гладкой линии через заданные точки. Впервые такие гибкие линейки стали использовать в авиастроении при конструировании обтекаемых профилей.

Термин *сплайн* ввёл в математику Шёнберг в середине XX века. Сначала сплайнами называли кусочно-полиномиальные функции заданной гладкости. Простейший вид таких функций использовал ещё Л. Эйлер при решении дифференциальных уравнений (метод ломаных).

Т

ТАБЛИЦА

От латинского слова *tabula – доска, стол*.

ТЕОРИЯ

От греческого слова $\theta\varepsilon\omega\rho\acute{\iota}\alpha$ – *наблюдение, исследование, опыт.*

ТЕРМИН

Латинское слово *terminus – межа, граница, конец*. Наименование некоторого понятия.

ТИП

Термин образован от греческого $\tau\upsilon\pi o\varsigma$ – *образ, отображение*.

ТОЖДЕСТВО

Равенство выражений с одной или несколькими переменными, левая и правая части которого равны при любых значениях переменных. Знак тождества ≡ впервые использовал немецкий математик Риман (1857).

ТОЧКА

Лобачевский говорил, что это слово происходит от прикосновения отточенного пера, поэтому *точка* означает остриё гусиного пера, которым писали во времена Лобачевского, следовательно, слово образовано от глагола *точить*.

ТРАНСЦЕНДЕНТНЫЙ

От латинского *transcendens – выходящий за пределы.* Буквальный смысл – недоступный познанию.

У

УСЛОВИЕ ЛИПШИЦА

Условие $|f(x_1) - f(x_2)| \leq L|x_1 - x_2|^n$ ввёл немецкий математик Рудольф Липшиц в 1864 г. Он изобрёл знак модуля и первый раз использовал его именно в этом неравенстве.

Ф

ФАКТОРИАЛ

От английского слова *factor – множитель*. Термин ввели в 1800 г., но обозначение были разные. Только в 1916 г. Совет Лондонского математического общества рекомендовал всем математикам принять знак $n!$ и предложили читать его так: «*n-восхищение».*

ФУНКЦИЯ

Термин впервые ввёл в математику в XVII веке Г.В.Лейбниц.

Э

ЭКСТРЕМУМ

От латинского *extremum – крайний, последний*. Этот термин предложил немецкий математик Дюбуа Раймон (1879) для обозначения минимума и максимума в тех случаях, когда не обязательно их различие.

ПРИЛОЖЕНИЕ 2

Равенство Парсеваля–Стеклова

Предположим, что функция представлена в виде ряда Фурье

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l} \right), \qquad (1)$$

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx, (n = 0,1,2,3,\cdots), \qquad (2)$$

$$b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi x}{l}\,dx, (n = 1,2,3,\cdots) \qquad (3)$$

Умножим первое равенство на $f(x)$ и разделим на l:

$$\frac{1}{l}f^2(x) = \frac{1}{l}\cdot\frac{f(x)a_0}{2} + \frac{1}{l}\sum_{n=1}^{\infty}\left(a_n f(x)\cos\frac{n\pi x}{l} + b_n f(x)\sin\frac{n\pi x}{l}\right).$$

Проинтегрируем это равенство почленно:

$$\frac{1}{l}\int_{-l}^{l} f^2(x)\,dx = \frac{1}{l}\int_{-l}^{l}\frac{f(x)a_0}{2}\,dx +$$

$$+\sum_{n=1}^{\infty}\left(a_n\frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx + b_n\frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi x}{l}\,dx\right).$$

Используя равенства (2) и (3) приходим к равенству Парсеваля – Стеклова

$$\int_{-l}^{l} f^2(x)\,dx = l\left(\frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\right).$$

Это равенство справедливо и для таблично заданных функций $f(x)$.

ПРИЛОЖЕНИЕ 3

Именной справочник

Математик изучает свою науку вовсе не потому, что она полезна. Он изучает её потому, что она прекрасна.
Лузин Н.Н. (1883 – 1950) знаменитый русский математик.

Абель Нильс Генрих (1802 – 1829) – норвежский математик.
По традиции новые результаты в науке называют в честь того, кто их открыл. Поэтому в различных разделах высшей математики встречаются теоремы Абеля, абелевы интегралы, уравнения Абеля, абелевы группы, преобразования Абеля. Он сделал огромный вклад в математику, занимаясь ею всего семь лет.

С тринадцати лет Нильс вместе со своими братьями учился в школе, где математику вёл один из лучших учителей Норвегии. Математика давала ему столько радости, что он занимался ею всё свободное время. Только иногда он ходил в театр или играл в шахматы.

В 1820 г. умер отец Нильса. После этого жизнь Абеля резко изменилась. Он постоянно подрабатывал, что бы помочь своей семье и самому не умереть от голода. В 1821 г. Абель окончил школу и поступил в университет. Там оценили его выдающиеся математические способности, желание учиться, отличное поведение, поэтому ему разрешили бесплатно жить в общежитии. В те времена студентам не платили стипендию. Несколько профессоров сами платили Абелю стипендию.

В 1822 г. Нильс успешно сдал экзамены за первый курс и получил звание кандидата философии. К этому времени Абель прочитал все книги по математике, которые можно было достать, и начал заниматься собственным творчеством.

В 1823 г. в журнале «Естественнонаучный журнал» напечатали первую статью Абеля. Она связана с функциональными уравнениями. В этом же году он выполняет большую научную работу, в которой описывает общий метод проверки дифференциального уравнения на интегрируемость. За эту работу Абелю назначили государственную стипендию. После университета он уехал за границу для продолжения образования. Тогда же стал заниматься теорией алгебраических уравнений.

До Абеля математики уже научились решать точно (с помощью формул) алгебраические уравнения до четвёртой степени включительно. Надо было сделать следующий шаг – найти формулы для точного решения уравнений пятой степени. Над этой проблемой в течение трёх веков безуспешно трудились лучшие математические умы.

Абель взялся за эту задачу. Через несколько недель он вывел формулы. Но затем обнаружил, что они не универсальны. Отрицательный результат привёл Абеля к мысли: «*Может быть, алгебраическое уравнение выше четвёртой*

степени в общем случае не решается с помощью четырёх арифметических действий и корней (в радикалах)»?

К концу 1823 г. Абель доказал, что уравнение вида

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 = 0, \qquad где\ n \geq 5$$

не может быть разрешено в радикалах.

Больцано Бернард (1781–1848) – чешский богослов, философ, логик и математик.

Больцано оказал большое влияние на Коши. Первый аналитически доказал, что если правая часть уравнения *f(x)=0* принимает значения противоположных знаков, то между этими значениями лежит хотя бы один действительный корень уравнения. Первый дал определение предела через (ε, δ) и определение бесконечно малой величины на основе понятия предела. Ввёл в математику следующие понятия: односторонняя непрерывность, окрестность точки, предельная точка последовательности. Нашёл критерии сходимости.

В 1851 г. посмертно опубликовали его работу «Парадоксы бесконечности». Благодаря этой работе математический мир по праву стал считать Больцано основателем теории множеств. Он дал определения конечным и бесконечным множествам, ввёл понятие взаимно однозначного соответствия.

В течение 15-и лет его математические рукописи были запрещены для публикации, так как императора Франца-Иосифа разгневали вольнодумные проповеди Больцано. В 1930 г. нашли и опубликовали рукопись Больцано «Учение о функциях», из которой видно, что он на несколько десятилетий обогнал своих современников.

Вандермонд Александр (1735 – 1796) – французский математик. Опубликовал первое исследование, посвященное определителям (1772).

Вейерштрасс Карл Теодор Вильгельм (1815 – 1897) – немецкий математик, почётный член Петербургской Академии Наук.

Родился в Вестфалии. С 14 лет работал в католической прогимназии, где преподавал математику, физику, ботанику, географию, историю, чистописание и немецкий язык. Свободное от работы время посвящал математике. В 1854 г. становится доктором наук за работы в области математического анализа. С 1856 г. Вейерштрасс – профессор Берлинского университета. В число его многочисленных учеников входила Софья Васильевна Ковалевская – знаменитый русский математик.

Гаусс Карл Фридрих (1777 – 1855) – немецкий математик.

Гаусс родился в семье мастера-водопроводчика, в маленьком немецком герцогстве Брауншвейг. Блестящие математические способности он проявлял с раннего детства. Гаусс учился в Геттингенском университете и первое выдающееся открытие опубликовал в студенческие годы. Он доказал, что циркулем и

линейкой можно построить правильный 17-и угольник. Это открытие осталось для Гаусса на всю жизнь самым любимым результатом его творчества.

Всемирную славу Гауссу принесла работа по определению орбиты планеты по малому числу наблюдений. Он разработал новый способ вычисления орбиты, а затем предсказал положение карликовой планеты (астероида), которую потом назвали Церерой. Эта планета появилась на том месте, которое предсказал Гаусс. Вскоре он вновь достаточно точно рассчитал положение другой малой планеты Паллады. В 1804 г. Гаусс изложил разработанный им метод вычисления орбит в знаменитой монографии «Теория движения небесных тел». За великие заслуги в астрономии его именем назвали один из крупнейших кратеров Луны.

Многие результаты Гаусса были незаконченными или описаны только в письмах к своим коллегам. Учёные Германии собирали и исследовали научное наследие Гаусса вплоть до начала второй мировой войны. Результатом такой работы стала публикация 12-и томного собрания трудов Гаусса.

Дирихле Петер Густав Лежен (1805–1859) – немецкий математик.
Друг и ученик Гаусса. Какое-то время жил в Париже и работал там учителем. Тесно связан с немецкой и французской математическими школами. Первый дал чёткое определение функции. Ввёл понятия условной и абсолютной сходимости рядов. Он первый вывел условие сходимости рядов Фурье. Внёс большой вклад в теорию чисел и интегральное исчисление. Был блестящим преподавателем. Знаменитый математик Минковский так писал о высоком педагогическом даре Дирихле: «*Он обладал искусством соединять минимум слепых формул с максимумом зрячей мысли*».

Зейдель Людвиг (1821– 1896) – немецкий астроном и математик.

Кронекер Леопольд (1823– 1891) – немецкий математик.
Представитель Берлинской математической школы. Ему принадлежит знаменитая фраза: «*Целые числа сотворил господь бог, а всё прочее – дело людских рук*».

Лагранж Жозеф Луи (1736 – 1813) – французский математик.
Император Наполеон Бонапарт любил беседовать с Лагранжем на философские, государственные и математические темы. Наполеон считал Лагранжа самым скромным математиком XVIII века и величественной пирамидой математических наук. Он сделал Лагранжа сенатором, графом и командиром ордена почётного легиона. Король Франции Людовик XVI и королева Мария осыпали Лагранжа почестями. Некоторое время он даже жил в Лувре. После казни короля и королевы революционная "общественность" не тронула Лагранжа.

Лагранж родился в итальянском городе Турине. Дед его был французом, женился на итальянке. Лагранжа крестили под итальянским именем Joseph

Louis. В итальянской энциклопедии он – итальянский учёный. В других странах его считают французским математиком.

Лагранж был одиннадцатым и последним ребёнком в семье. Его отец ко времени рождения Лагранжа потерял своё состояние. Позже Лагранж писал: «*Если бы я унаследовал состояние, то мне, вероятно, не пришлось бы связать свою судьбу с математикой»*.

В детстве Лагранж увлекался латынью и случайно изучил научную книгу «О преимуществах аналитического метода», написанную на латыни. Автор книги Э. Галлей – друг Ньютона. Его именем названа комета Галлея. Книга посвящена применению математического анализа в оптике. С тех пор и на всю жизнь Лагранж заболел математическим анализом. К 17-и годам он изучил все известные к тому времени труды по математике и приступил к самостоятельным исследованиям.

В августе 1755г. юный Лагранж послал письмо знаменитому Эйлеру, в котором изложил новые идеи в вариационном исчислении. Эйлер высоко оценил содержание этого письма. В 19 лет Лагранж становится профессором Королевской артиллерийской школы. Ещё через год его избрали в Берлинскую Академию наук. В родном Турине он организовал свою Академию наук. По просьбе Эйлера Лагранж переехал жить в Берлин и занял место руководителя Берлинской Академии, вместо уехавшего в Санкт-Петербург Эйлера.

Лагранж настолько плодотворно использовал математику в механике, что пошёл дальше Ньютона. Сейчас классическая механика наполовину может быть названа «лагранжевой». Во время берлинского периода Лагранжу пять раз присуждали премии Парижской Академии за работы в области небесной механики. Король Пруссии Фридрих II был покровителем Лагранжа и проводил многие часы в беседах с ним. Через полгода после смерти Фридриха II Лагранж переехал в Париж. В 1790 г. его включили в комиссию по стандартизации системы мер и весов. Благодаря Лагранжу была введена десятичная метрическая система.

В 1794 – 1795 годах во Франции открылись две знаменитые школы: Политехническая школа (Ecole Polytechique) для подготовки офицеров и инженеров, а так же Нормальная школа (Ecole Normale) для подготовки учителей. В этих школах Лагранж читал лекции по математическому анализу и элементарной математике.

Важно отметить, что в работе Лагранжа «Теория аналитических функций» (1797) впервые появились новые термины и обозначения, которые с тех пор успешно используют во всём математическим мире. Это термины: *производная*, *первообразная* и обозначения: y' и $f'(x)$.

Ландау Эдмунд (1877– 1938) – немецкий математик. Внёс большой вклад в «Аналитическую теорию чисел». Ввёл понятия *О-большое* и *о-маленькое.*

Липшиц Рудольф Отто Сигизмунд (1832– 1903) – немецкий математик.

Впервые ввёл условие $|f(x_1) - f(x_2)| \leq L \cdot |x_1 - x_2|^n$,

которое получило широкое распространение в математике. Интересно, что Липшиц изобрёл обозначение модуля.

Нейман Джон фон (1903 – 1957) – венгерский математик. Он сыграл выдающуюся роль в разработке теории вычислительных машин.

Джон фон Нейман родился в Будапеште. Там окончил университет. В 1930 г. переехал в США. Стал профессором Принстонского университета и сотрудником Института перспективных исследований в Принстоне. Его коллеги вспоминали, что улыбка сразу вспыхивала на его лице, если в задаче появлялись логические или математические парадоксы. Он мог часами говорить с друзьями о различных научных вопросах. После науки он больше всего интересовался историей. Хорошо знал латынь, греческий, английский, немецкий и испанский языки. В 1946 г. совместно с другими учёными опубликовал статью об основных принципах построения вычислительных машин.

Ньютон Исаак (1643 – 1727) – английский математик. Имя Исаака Ньютона известно миллионам людей, его называют величайшим учёным в истории человечества. Он открыл многие законы механики и всемирного тяготения. Ньютон является одним из создателей дифференциального и интегрального исчисления. В астрономии сделал открытия в небесной механике и построил зеркальный телескоп.

Ньютон писал: «*Если я видел дальше других, то потому, что стоял на плечах гигантов»*. В своей научной деятельности Ньютон придерживался принципа Галилея – искать не физическое, а математическое описание.

В жизни Ньютона нет ярких внешних событий. Учёба, научная работа, работа в Монетном дворе. Ньютон жил во время зарождения английской опытной науки, которая сыграла большую роль в истории мировой культуры и в научной жизни самого Ньютона.

В школе Ньютон учился средне. В свободное время он делал механические игрушки. Построил мельницу, которую приводила в движение мышь. Сделал водяные и солнечные часы. Запускал воздушных змей с фонариками. Ньютон много читал, рисовал, писал стихи. В 1661 году он поступил в колледж Святой Троицы Кембриджского университета. Там встретился с выдающимся учителем – Исааком Барроу. С этого момента и до 1696 г. жизнь Ньютона была связана с Кембриджем. В 1665 г. из-за эпидемии чумы он уехал на два года в деревню на юго-восток Англии. В этот период Ньютон занимался научной работой по механике, математике и оптике. Понял, что открытый им закон всемирного тяготения даёт ключ ко всей механике. Он установил, что белый свет включает в себя все цвета радуги от красного до фиолетового. Ньютон разработал общий метод решения многих задач математического анализа.

Результаты своей научной работы Ньютон не торопился опубликовывать. Классическое произведение Ньютона «Математические начала натуральной философии» издали только в 1687 г. После выхода этой книги имя Ньютона стало широко известно. Но книга была очень трудна для чтения, поэтому по-

явились популярные изложения этого сочинения. Постепенно, почти за сто лет, математики довели содержание его книги до полной ясности. Известный математик Клейн (1849 – 1925) так определил значение этого научного труда: «*Эта книга открыла перед человечеством новый мир – Вселенную, управляемую единым сводом законов физики с их точным математическим выражением. Работа Ньютона доказала всему миру, что природа основана на математических принципах и что законы природы описываются математическим языком».* В 27 лет Ньютон стал профессором Кембриджского университета, а затем членом Лондонского королевского общества.

Анализом бесконечно малых занимались многие учёные, начиная с Архимеда и включая некоторых учёных XVII в. Но Ньютон обобщил и систематизировал труды своих предшественников. В результате он построил интегральное и дифференциальное исчисление. Ньютон пришёл к понятию производной, когда определял в механике скорость движения. Функцию от времени он называл флюэнтой – текущей величиной (от латинского слова fluere – течь), а производную – флюксией. Ньютон решал обратные задачи, в которых по флюксиям находил флюэнты, т.е. по производным находил первообразные.

Ньютон считал, что задача науки состоит в раскрытии блистательных замыслов творца. Он всю жизнь изучал и интерпретировал религиозные произведения, а в конце жизни посвятил себя богословию (теологии). Ньютон верил, что Бог сотворил мир. Он писал: «*Мне хотелось найти такие начала, которые были бы совместимы с верой людей в Бога; мой труд оказался не напрасным».*

В 1696 г. начался лондонский период жизни Ньютона – время его славы и признания. Он становится членом парламента и директором Монетного двора.

В Кембридже на статуе Ньютона высечено: «Разумом он превосходил род человеческий». В доме, где он родился, есть запись: «Природа и её законы были покрыты мраком. Бог сказал: да будет Ньютон – и стал свет».

Сам о себе Ньютон сказал так: «*Не знаю, чем я могу казаться миру, но сам себе я кажусь только мальчиком, который играет на берегу и отыскивает цветные камешки и красивую раковину, а в это время великий океан истины расстилается передо мной неисследованным».*

Парсеваль Марк-Антуан (1755 – 1836) – французский математик.
Основные труды по дифференциальным уравнениям и теории функций действительного переменного. В 1799 г. Парсеваль нашёл соотношение между функцией и коэффициентами в разложении этой функции в тригонометрический ряд Фурье:

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,dx = \frac{a_0^2}{2} + \sum_{n=1}^{\infty}\left(a_n^2 + b_n^2\right).$$

Позже аналогичные обобщённые равенства для различных функциональных пространств доказали другие математики, среди которых были русские математики Ляпунов А.М. (1845 – 1918) и Стеклов В.А. (1864–1926).

Пикар Эмиль (1856 – 1941) – французский математик и философ.
Он говорил шутя: «*Математика флиртует с философией, и это идёт на пользу самой философии*». Основные его достижения – в теории многочленов, интегральных функций и теории линейных дифференциальных уравнений, которую он строил аналогично теории алгебраических уравнений.

Пуанкаре Анри (1854 – 1912**) –** французский математик.
Он глубоко разбирался в огромном количестве математических дисциплин и внёс существенный вклад в каждую из них. В качестве примера можно привести его исследование расходящихся рядов. На основе результатов этого исследования Пуанкаре разработал теорию асимптотических разложений. Интересны его работы о сущности вероятности, космогонии и теории относительности.

Сильвестр Джемс Джозеф (1814 – 1897) – англо-американский математик.
Был не только математиком, поэтом, остряком, но и наряду с Лейбницем, непревзойдённым за всю историю математики создателем новых терминов. Он с юмором называл себя математическим Адамом. Используя латинское слово *discriminans,* что означает *разделяющий*, придумал термин *дискриминант*.

Сильвестр преподавал в военной академии. Дважды работал в Америке и там вошёл в группу математиков, которые заложили основы математических исследований в американских университетах.

Симпсон Томас (1716 – 1761) – английский математик. Определил понятие *средней ошибки*. Он первый стал рассматривать и изучать распределение ошибок. Одним из первых стал употреблять современное обозначение $\sin\alpha$.

Фурье Жан Батист Жозеф (1768 – 1830) – французский математик и физик. Почётный член Петербургской Академии Наук.
Фурье родился в семье портного. В 8 лет остался сиротой. Окончил военную школу и работал там учителем. Его научные интересы были связаны с алгеброй, математической физикой и дифференциальными уравнениями.

В 1798 г. Наполеон возглавил военный поход в Египет. В его свиту входили 165 учёных, среди них был Фурье. Во время пребывания в этой стране он изучал египетские древности и занимался дипломатической работой. После возвращения во Францию Фурье руководил осушением болот и восстановлением дорог.

Он первый представил распределение температуры в виде суммы гармоник. Вывел формулы для вычисления коэффициентов в разложении функции в ряд по тригонометрическим функциям. Для математиков XIX в. этот факт казался очень странным. Даже великий Эйлер не мог принять идеи Фурье, хотя сам умел раскладывать некоторые функции в ряд по синусам. Работа Фурье «Аналитическая теория тепла» (1822) послужила основой создания теории тригонометрических рядов. Вопрос о сходимости тригонометрического ряда изучался достаточно долго. В настоящее время благодаря работам немецкого

математика Гильберта (1862 – 1943) теория тригонометрических рядов Фурье стала частью общей теории разложения функций в ортогональные ряды.

Чебышев Пафнутий Львович (1821 –1894) – русский математик.
В 1852 г. Чебышев уехал в командировку во Францию. Там он изучил конструкцию парового двигателя, который в те времена был основой передовой французской промышленности. В паровых двигателях тогда использовали шарнирные механизмы. Эти механизмы двигали поршень. Движение поршня должно минимально отличаться от прямолинейного движения. Изучение поведения таких шарнирных механизмов привело Чебышева к созданию знаменитых полиномов, наименее уклоняющихся от нуля. Многие математические открытия Чебышева связаны с прикладными работами. Он считал, что большая часть вопросов практики сводится к задачам о нахождении наибольших и наименьших величин. Так, например, он разработал проекцию карты Европейской части России, для которой искажение масштаба было минимальным.

Заслуги П.Л. Чебышева в области математики признал весь научный мир ещё при его жизни. Чебышева избрали иностранным членом Парижской Академии наук, Итальянской королевской Академии, Лондонского королевского общества, Шведской Академии наук. П.Л. Чебышев оставил после себя много учеников, среди которых были такие известные во всём мире математики как А. М. Ляпунов и А. А. Марков.

Якоби Карл Густав (1804 – 1851) – немецкий математик.
После учёбы в Берлинском университете он стал молодым профессором в Кенигсберге и возглавил там кружок талантливых учеников, куда входили впоследствии известные математики Людвиг Гессе и Феликс Клейн.

Якоби отличался пылким и страстным характером. Клейн вспоминал, что его влияние на учеников было исключительным. Все подчинялись его образу мышления. Появлялся огромный интерес к очередной, поставленной им задаче. Якоби внёс большой вклад в развитие «Вариационного исчисления».

Знак Δ для обозначения определителя придумал Якоби. Слово *element* в определителях впервые использовал Якоби. Его старший брат известен как русский физик Борис Якоби. Он жил в Петербурге. Один из первых в России стал изучать электричество и изобрёл гальванопластику.

Предметный указатель

УЧЕБНОЕ ИЗДАНИЕ

Беликова Галина Иосифовна, Бровкина Екатерина Анатольевна,
Вагер Борис Георгиевич, Витковская Лариса Валерьевна,
Матвеев Юрий Леонидович

ЧИСЛЕННЫЕ МЕТОДЫ

Учебное пособие

Публикуется в авторской редакции

Подписано в печать 29.11.19. Формат 60 x 90/8. Гарнитура Times New Roman.
Печать цифровая. Усл. печ. 21,75. Тираж 150 экз. Заказ № 826.
РГГМУ, 19207, Санкт-Петербург, Воронежская, 79.
