ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ РФ
ФБГОУ ВО «ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

В.Н. ТАРАСОВ, Н.Ф. БАХАРЕВА

ЧИСЛЕННЫЕ МЕТОДЫ
ТЕОРИЯ
АЛГОРИТМЫ
ПРОГРАММЫ

ИЗДАНИЕ ВТОРОЕ
ПЕРЕРАБОТАННОЕ

Рекомендовано ГОУ ВПО МГТУ
им. Н.Э. Баумана в качестве учебного
пособия для студентов высших учебных заведний,
обучающихся по направлению подготовки
«Информатика и вычислительная техника»
Рег.№ рецензии 119 от 16.07.2008 г. МГУП

Самара 2017

ББК 22.19я7
Т19
УДК [519.95+004.421](075.8)

Рецензенты: заведующий кафедрой «Информационные системы и технологии» СГАУ, Заслуженный работник высшей школы РФ, академик международной академии информатизации, д.т.н., профессор ***С.А. Прохоров***; кафедра «Программное обеспечение ЭВМ и информационные технологии» МГТУ им. Н.Э. Баумана, д.т.н., профессор ***В.М. Градов***.

Тарасов В.Н., Бахарева Н.Ф.

Численные методы. Теория, алгоритмы, программы. – Самара: ИНУЛ ПГУТИ, 2017. – 264 с.

Т19

ISBN 5-7410-0451-2

Учебное пособие предназначено для студентов специальностей направления подготовки 09.03.01 – Информатика и вычислительная техника.

ISBN 5-7410-0451-2

Содержание

ПРЕДИСЛОВИЕ

Данное издание учебного пособия «Численные методы. Теория, алгоритмы, программы» включает все основные (классические) разделы дисциплины «Вычислительная математика», предусмотренные государственным образовательным стандартом для студентов направления подготовки 230100 – Информатика и вычислительная техника. Учебное пособие рассчитано на стандартный семестровый курс.

Наряду с теоретическими основами численных методов, пособие содержит также полный комплекс лабораторных работ, включающий схемы алгоритмов методов, коды программ, решения контрольных примеров и варианты заданий. На взгляд авторов, такое представление материала учебного пособия наиболее полно отражает специфику направления подготовки – Информатика и вычислительная техника, т.к. простое использование закрытых математических пакетов типа MathCAD и Matlab с точки зрения построения алгоритмов вычислительных методов мало информативно. Наоборот, умение программировать вычислительные методы поможет лучше понять содержимое математических пакетов программ и их работу.

Все результаты расчетов контрольных примеров лабораторной части учебного пособия проверены в пакете MathCAD.

Читатель, заинтересованный в более глубоком и детальном изучении курса численных методов, должен обратиться к более полным руководствам. Некоторые из них приведены в списке литературы. Так же можно обратиться к Internet ресурсам. Например, современные достижения в этой области можно увидеть на Web – сайтах Института вычислительной математики РАН: www.inm.ras.ru и научного журнала «Вычислительные методы и программирование. Новые вычислительные технологии» - www.num-meth.srcc.msu.ru.

Введение

Математическое моделирование и вычислительный эксперимент

1. Схема вычислительного эксперимента. Эффективное решение крупных естественнонаучных и народнохозяйственных задач сейчас невозможно без применения быстродействующих электронно-вычислительных машин (ЭВМ). В настоящее время выработалась технология исследования сложных проблем, основанная на построении и анализе с помощью ЭВМ математических моделей изучаемого объекта. Такой метод исследования называют *вычислительным экспериментом*.

Пусть, например, требуется исследовать какой-то физический объект, явление, процесс. Тогда схема вычислительного эксперимента выглядит так, как показано на рисунке 1. Формулируются основные законы, управляющие данным объектом исследования (I) и строится соответствующая *математическая модель* (II), представляющая обычно запись этих законов в форме системы уравнений (алгебраических, дифференциальных, интегральных и т. д.).

Рисунок 1 – Этапы построения и анализа с помощью ЭВМ математической модели объекта

При выборе физической и, следовательно, математической модели мы пренебрегаем факторами, не оказывающими существенного влияния на ход изучаемого процесса. Типичные математические модели, соответствующие физическим явлениям, формулируются в виде уравнений математической физики. Большинство реальных процессов описывается нелинейными уравнениями и лишь в первом

приближении (при малых значениях параметров, малых отклонениях от равновесия и др.) эти уравнения можно заменить линейными.

После того как задача сформулирована в математической форме, необходимо найти ее решение. Но что значит решить математическую задачу? Только в исключительных случаях удается найти решение в явном виде, например в виде ряда. Иногда утверждение «задача решена» означает, что доказано существование и единственность решения. Ясно, что этого недостаточно для практических приложений. Необходимо еще изучить качественное поведение решения и найти те или иные количественные характеристики.

Именно на этом этапе требуется привлечение ЭВМ и, как следствие, развитие численных методов (см. III на рис. 1). Под *численным методом* здесь понимается такая интерпретация математической модели («дискретная модель»), которая доступна для реализации на ЭВМ. Например, если математическая модель представляет собой дифференциальное уравнение, то численным методом может быть аппроксимирующее его разностное уравнение совместно с алгоритмом, позволяющим отыскать решение этого разностного уравнения. Результатом реализации численного метода на ЭВМ является число или таблица чисел. Отметим, что в настоящее время помимо собственно численных методов имеются также методы, которые позволяют проводить на ЭВМ аналитические выкладки. Однако аналитические методы для ЭВМ не получили пока достаточно широкого распространения.

Чтобы реализовать численный метод, необходимо составить программу для ЭВМ (см. IV на рис. 1) или воспользоваться готовой программой.

После отладки программы наступает этап проведения вычислений и анализа результатов (V). Полученные результаты изучаются с точки зрения их соответствия исследуемому явлению и, при необходимости, вносятся исправления в численный метод и уточняется математическая модель.

Такова в общих чертах схема вычислительного эксперимента. Его основу составляет триада: *модель — метод* (*алгоритм*) — *программа*. Опыт решения крупных задач показывает, что метод математического моделирования и вычислительный эксперимент соединяют в себе преимущества традиционных теоретических и экспериментальных методов исследования. Можно указать такие крупные области применения вычислительного эксперимента, как энергетика,

аэрокосмическая техника, обработка данных натурного эксперимента, совершенствование технологических процессов.

Пример. Академик Самарский Александр Андреевич совместно с академиком Тихоновым Андреем Николаевичем с 1948 г. Разрабатывал численные методы и вел первые в СССР прямые расчеты мощности взрыва атомной бомбы. В этих расчетах были заложены основы разностных схем решения систем дифференциальных уравнений и параллельных вычислений.

2. Вычислительный алгоритм. Предметом данной книги является изложение вопросов, отражающих этапы III, IV, V вычислительного эксперимента. Таким образом, здесь не обсуждаются исходные задачи и их математическая постановка.

Необходимо подчеркнуть, что процесс исследования исходного объекта методом математического моделирования и вычислительного эксперимента неизбежно носит приближенный характер, потому что на каждом этапе вносятся те или иные погрешности. Так, построение математической модели связано с упрощением исходного явления, недостаточно точным заданием коэффициентов уравнения и других входных данных. По отношению к численному методу, реализующему данную математическую модель, указанные погрешности являются *неустранимыми,* поскольку они неизбежны в рамках данной модели.

При переходе от математической модели к численному методу возникают погрешности, называемые *погрешностями метода.* Они связаны с тем, что всякий численный метод воспроизводит исходную математическую модель приближенно.

Наиболее типичными погрешностями метода являются *погрешность дискретизации* и *погрешность округления.*

Поясним причины возникновения таких погрешностей.

Обычно построение численного метода для заданной математической модели разбивается на два этапа: а) формулирование дискретной задачи, б) разработка вычислительного алгоритма, позволяющего отыскать решение дискретной задачи. Например, если исходная математическая задача сформулирована в виде системы дифференциальных уравнений, то для численного решения необходимо заменить ее системой конечного, может быть, очень большого числа линейных или разностных алгебраических уравнений. В этом случае говорят, что проведена *дискретизация исходной математической задачи.* Простейшим примером дискретизации является построение

разностной схемы, путем замены дифференциальных выражений конечно-разностными отношениями. В общем случае дискретную модель можно рассматривать как конечномерный аналог исходной математической задачи. Ясно, что решение дискретизированной задачи отличается от решения исходной задачи. Разность соответствующих решений и называется *погрешностью дискретизации.* Обычно дискретная модель зависит от некоторого параметра (или множества параметров) дискретизации, при стремлении которого к нулю должна стремиться к нулю и погрешность дискретизации. При этом число алгебраических уравнений, составляющих дискретную модель, неограниченно возрастает. В случае разностных методов таким параметром является шаг сетки.

Как уже отмечалось, дискретная модель представляет собой систему большого числа алгебраических уравнений. Невозможно найти решение такой системы точно и в явном виде. Поэтому приходится использовать тот или иной численный алгоритм решения системы алгебраических уравнений. Входные данные этой системы, а именно коэффициенты и правые части, задаются в ЭВМ не точно, а с округлением.

В процессе работы алгоритма погрешности округления обычно накапливаются, и в результате решение, полученное на ЭВМ, будет отличаться от точного решения дискретизированной задачи. Результирующая погрешность называется *погрешностью округления* (иногда ее называют *вычислительной погрешностью*).

Величина этой погрешности определяется двумя факторами: точностью представления вещественных чисел в ЭВМ и чувствительностью данного алгоритма к погрешностям округления.

Алгоритм называется *устойчивым,* если в процессе его работы вычислительные погрешности возрастают незначительно, и *неустойчивым* — в противоположном случае. При использовании неустойчивых вычислительных алгоритмов накопление погрешностей округления приводит в процессе счета к переполнению арифметического устройства ЭВМ.

Итак, следует различать погрешности модели, метода и вычислительную. Какая же из этих трех погрешностей является преобладающей? Ответ здесь неоднозначен. Видимо, типичной является ситуация, возникающая при решении задач математической физики, когда погрешность модели значительно превышает погрешность метода, а погрешностью округления в случае устойчивых алгоритмов

можно пренебречь по сравнению с погрешностью метода. С другой стороны, при решении, например, систем обыкновенных дифференциальных уравнений возможно применение столь точных методов, что их погрешность будет сравнима с погрешностью округления. В общем случае нужно стремиться, чтобы все указанные погрешности имели один и тот же порядок. Например, нецелесообразно пользоваться разностными схемами, имеющими точность 10^{-6}, если коэффициенты исходных уравнений задаются с точностью 10^{-2}.

3. Требования к вычислительным методам. Одной и той же математической задаче можно поставить в соответствие множество различных дискретных моделей. Однако далеко не все из них пригодны для практической реализации.

Вычислительные алгоритмы, предназначенные для быстродействующих ЭВМ, должны удовлетворять многообразным и зачастую противоречивым требованиям.

Попытаемся здесь сформулировать основные из этих требований в общих чертах.

Можно выделить две группы требований к численным методам. Первая группа связана с адекватностью дискретной модели исходной математической задаче, и вторая группа – с реализуемостью численного метода на ЭВМ.

К первой группе относятся такие требования, как сходимость численного метода, выполнение дискретных аналогов законов сохранения, качественно правильное поведение решения дискретной задачи.

Поясним эти требования. Предположим, что дискретная модель математической задачи представляет собой систему большого, но конечного числа алгебраических уравнений. Обычно, чем точнее мы хотим получить решение, тем больше уравнений приходится брать. Говорят, что численный метод *сходится,* если при неограниченном увеличении числа уравнений решение дискретной задачи стремится к решению исходной задачи.

Поскольку реальная ЭВМ может оперировать лишь с конечным числом уравнений, на практике сходимость, как правило, не достигается. Поэтому важно уметь оценивать погрешность метода в зависимости от числа уравнений, составляющих дискретную модель. По этой же причине стараются строить дискретную модель таким образом, чтобы она правильно отражала качественное поведение реше-

ния исходной задачи даже при сравнительно небольшом числе уравнений.

Например, дискретной моделью задачи математической физики может быть разностная схема. Для ее построения область изменения независимых переменных заменяется дискретным множеством точек – *сеткой,* а входящие в исходное уравнение производные заменяются, на сетке, конечно-разностными отношениями. В результате получаем систему алгебраических уравнений относительно значений искомой функции в точках сетки.

Число уравнений этой системы равно числу точек сетки. Известно, что дифференциальные уравнения математической физики являются следствиями интегральных законов сохранения. Поэтому естественно требовать, чтобы для разностной схемы выполнялись аналоги таких законов сохранения. Разностные схемы, удовлетворяющие этому требованию, называются *консервативными.* Оказалось, что при одном и том же числе точек сетки консервативные разностные схемы более правильно отражают поведение решения исходной задачи, чем неконсервативные схемы.

Сходимость численного метода тесно связана с его корректностью. Предположим, что исходная математическая задача поставлена корректно, т.е. ее решение существует, единственно и непрерывно зависит от входных данных. Тогда дискретная модель этой задачи должна быть построена таким образом, чтобы свойство корректности сохранилось. Таким образом, в понятие *корректности численного метода* включаются свойства однозначной разрешимости соответствующей системы уравнений и ее устойчивости по входным данным. Под *устойчивостью* понимается непрерывная зависимость решения от входных данных, равномерная относительно числа уравнений, составляющих дискретную модель.

Вторая группа требований, предъявляемых к численным методам, связана с возможностью реализации данной дискретной модели на данной ЭВМ, т. е. с возможностью получить на ЭВМ решение соответствующей системы алгебраических уравнений за приемлемое время. Основным препятствием для реализации корректно поставленного алгоритма является ограниченный объем оперативной памяти ЭВМ и ограниченные ресурсы времени счета. Реальные вычислительные алгоритмы должны учитывать эти обстоятельства, т. е. они должны быть экономичными как по числу арифметических действий, так и по требуемому объему памяти.

Численные методы алгебры и анализа

1 Решение систем линейных алгебраических уравнений

Рассмотрим систему линейных алгебраических уравнений:

$$\begin{cases} a_{11}x_1+a_{12}x_2+...+a_{1m}x_m=b_1 \\ a_{21}x_1+a_{22}x_2+...+a_{2m}x_m=b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1+a_{m2}x_2+...+a_{mm}x_m=b_m \end{cases} \quad (1.1)$$

или в матричной форме:

$$\boldsymbol{Ax}=\boldsymbol{b}, \quad (1.2)$$

где: $\boldsymbol{A}=\{a_{ij}\}$ квадратная матрица размерности $(m\times m,)$; $\boldsymbol{x}=(x_1,\ldots,x_m)^T$; T – операция транспонирования; $\boldsymbol{b}=(b_1,\ldots.,b_m)^T$; $det\ \boldsymbol{A}\neq 0$.

Предположим, что определитель матрицы $\boldsymbol{A}$ не равен нулю. Тогда решение $\boldsymbol{x}$ существует и единственно. На практике встречаются системы, имеющие большой порядок. Методы решения системы (1.1) делятся на две группы:

1) прямые (точные методы);

2) итерационные методы (приближенные).

1.1 Точные методы

В точных методах решение $\boldsymbol{x}$ находится за конечное число действий, но из-за погрешности округления и их накопления прямые методы можно назвать точными, только отвлекаясь от погрешностей округления.

1.1.1 Метод Гаусса

Вычисления с помощью метода Гаусса (который называют также методом последовательного исключения неизвестных) состоят из двух основных этапов: прямого хода и обратного хода. Прямой ход метода заключается в последовательном исключении неизвестных из системы для преобразования ее к эквивалентной системе с треугольной матрицей. На этапе обратного хода производят вычисления значений неизвестных. Рассмотрим простейший вариант метода Гаусса, называемый схемой единственного деления.

Прямой ход метода

1-й шаг. Предположим, что $a_{11} \neq 0$. Поделим первое уравнение на этот элемент, который назовем *ведущим* элементом первого шага :

$$x_1 + c_{12}x_2 + ... + c_{1m}x_m = y_1. \qquad (1.3)$$

Остальные уравнения системы (1.1) запишем в виде

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{im}x_m = b_i, \qquad (1.4)$$

где $i = \overline{2,m}$.

Уравнение (1.3) умножаем на a_{i1} и вычитаем из i-го уравнения системы (1.4). Это позволит обратить в нуль коэффициенты при x_1 во всех уравнениях, кроме первого.

Получим эквивалентную систему вида:

$$\begin{cases} x_1 + c_{12}x_2 + ... + c_{1m}x_m = y_1 \\ a_{22}^{(1)}x_2 + ... + a_{2m}^{(1)}x_1 = b_2^{(1)} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{m2}^{(1)}x_2 + ... + a_{mm}^{(1)}x_1 = b_m^{(1)} \end{cases}. \qquad (1.5)$$

$$\begin{aligned} a_{ij}^{(1)} &= a_{ij} - c_{1j}a_{i1} \\ b_i^{(1)} &= b_i - y_1 a_{i1} \end{aligned},$$

где $i, j = \overline{2,m}$. Система (1.5) имеет матрицу вида:

$$\begin{pmatrix} 1 & x & ... & x \\ 0 & x & ... & x \\ ... & ... & ... & ... \\ 0 & x & ... & x \end{pmatrix}.$$

Работаем с укороченной системой, т.к. x_1 входит только в 1-ое уравнение

$$\begin{cases} a_{22}^{(1)}x_2 + ... + a_{2m}^{(1)}x_m = b_2^{(1)} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{m2}^{(1)}x_2 + ... + a_{mm}^{(1)}x_m = b_m^{(1)} \end{cases}.$$

2-й шаг. На этом шаге исключаем неизвестное x_2 из уравнений с номерами $i=3, 4,\ldots,m$. Если ведущий элемент второго шага $a_{22}^{(1)} \neq 0$, то из укороченной системы аналогично исключаем неизвестное x_2 и получаем матрицу коэффициентов такого вида:

$$\begin{pmatrix} 1 & x & x & ... & x \\ 0 & 1 & x & ... & x \\ 0 & 0 & x & ... & x \\ ... & \cdots & ... & \cdots & ... \\ 0 & 0 & x & ... & x \end{pmatrix}.$$

Аналогично повторяем указанные действия для неизвестных x_3, $x_4,\ldots,x_{m-1}$ и приходим к системе:

$$\begin{cases} x_1 + c_{12}x_2 + \ldots + c_{1m}x_m = y_1 \\ \qquad x_2 + \ldots + c_{2m}x_m = y_2 \\ \qquad\qquad \ldots\ldots\ldots\ldots \\ x_{m-1} + c_{m-1,m}x_m = y_{m-1} \\ \qquad\qquad c_{mm}x_m = y_m \end{cases} \qquad (1.6)$$

Эта система с верхней треугольной матрицей:

$$\begin{pmatrix} 1 & x & x & ... & x & x \\ 0 & 1 & x & ... & x & x \\ 0 & 0 & 1 & x & ... & x \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & ... & 1 & x \\ 0 & 0 & 0 & ... & ... & x \end{pmatrix}.$$

Обратный ход метода. Из последнего уравнения системы (1.6) находим x_m, из предпоследнего x_{m-1}, …, из первого уравнения – x_1.

Общая формула: $x_m = y_m / c_{mm}$,

$$x_i = y_i - \sum_{j=i+1}^{m} c_{ij}x_j, \ (i = m-1,\ldots,1).$$

Для реализации метода Гаусса требуется примерно $(1/3)m^3$ арифметических операций, причем большинство из них приходится на прямой ход.

Ограничение метода единственного деления заключается в том, что ведущие элементы на k-ом шаге исключения не равны нулю, т.е. $a_{kk}^{k-1} \neq 0$.

Но если ведущий элемент близок к нулю, то в процессе вычисления может накапливаться погрешность. В этом случае на каждом шаге исключают не x_k, а x_j (при $j \neq k$). Такой подход называется методом выбора главного элемента. Для этого выбирают неизвестные x_j с наибольшим по абсолютной величине коэффициентом либо в строке, либо в столбце, либо во всей матрице. Для его реализации требуется $\frac{m(m^2+3m-1)}{3}$ – арифметических действий.

Пример 1. Используя схему Гаусса, решить систему уравнений с точностью до 0,001.

$$0{,}14x_1 + 0{,}24x_2 - 0{,}84x_3 = 1{,}11$$
$$1{,}07x_1 - 0{,}83x_2 + 0{,}56x_3 = 0{,}48,$$
$$0{,}64x_1 + 0{,}43x_2 - 0{,}38x_3 = -0{,}83.$$

Вычисления производим по схеме единственного деления:

Коэффициенты при неизвестных			Свободные члены
x_1	x_2	x_3.	
0,14 1,07 0,64	0,24 -0,83 0,43	-0,84 0,56 -0,38	1,11 0,48 -0,83
1	1,7143	-6,0000	7,926
	-2,6643 0,6672	6,98 -3,4600	-8,0036 -5,9043
	1	-2,6198	3,0040
		1,7121	-3,9000
		1	-2,2279
		1	-2,2779 -2.9636 -0,6583

Ответ: $x_1 \approx -0{,}658$; $x_2 \approx -2{,}964$; $x_3 \approx -2{,}278$.

1.1.2 Связь метода Гаусса с разложением матрицы на множители. Теорема об *LU* разложении.

Пусть дана система $\boldsymbol{Ax}=\boldsymbol{b}$ (1.1), которая при прямом ходе преобразуется в эквивалентную систему (1.6) и запишем ее в виде

$$\boldsymbol{Cx}=\boldsymbol{y}, \qquad (1.6^*)$$

где $\boldsymbol{C}$ – верхняя треугольная матрица с единицами на главной диагонали, полученная из (1.6) делением последнего уравнения системы на c_{mm}.

Как связаны в системе (1.1) элементы $\boldsymbol{b}$ и элементы $\boldsymbol{y}$ из (1.6*)?

Если внимательно посмотреть на прямой ход метода Гаусса, то можно увидеть, что

$$\begin{aligned} b_1 &= a_{11}y_1 \\ b_2 &= a_{21}y_1 + a_{22}^{(1)}y_2 \end{aligned}.$$

Для произвольного j имеем

$$b_j = d_{j1}y_1 + d_{j2}y_2 + \ldots + d_{jj}y_i, \qquad (1.7)$$

где $j=\overline{1,m}$, d_{ji} – числовые коэффициенты:

$$d_{jj}=a_{jj}^{(j-1)}. \qquad (1.8)$$

Можно записать систему:

$$\boldsymbol{b}=\boldsymbol{Dy},$$

где $\boldsymbol{D}$–нижняя треугольная матрица с элементами $a_{jj}^{(j-1)}$ на главной диагонали ($j=\overline{1,m}$, $a_{11}^{(0)}=a_{11}$).

В связи с тем, что в методе Гаусса угловые коэффициенты не равны нулю $a_{jj}^{(j-1)}\neq 0$, то на главной диагонали матрицы $\boldsymbol{D}$ стоят не нулевые элементы. Следовательно, эта матрица имеет обратную, тогда $\boldsymbol{y}=\boldsymbol{D}^{-1}\boldsymbol{b}$, $\boldsymbol{Cx}=\boldsymbol{D}^{-1}\boldsymbol{b}$.

Тогда

$$\boldsymbol{D}\times\boldsymbol{Cx}=\boldsymbol{b}. \qquad (1.9)$$

В результате использования метода Гаусса, получили разложение матрицы $\boldsymbol{A}$ на произведение двух матриц

$$\boldsymbol{A} = \boldsymbol{D} \times \boldsymbol{C},$$

где $\boldsymbol{D}$ − нижняя треугольная матрица, у которой элементы на главной диагонали не равны нулю, а $\boldsymbol{C}$ – верхняя треугольная матрица с единичной диагональю.

Таким образом, если задана матрица $\boldsymbol{A}$ и вектор $\boldsymbol{b}$, то в методе Гаусса сначала производится разложение этой матрицы $\boldsymbol{A}$ на произведение $\boldsymbol{D}$ и $\boldsymbol{C}$, а затем последовательно решаются две системы:

$$\boldsymbol{Dy}=\boldsymbol{b},\ \boldsymbol{Cx}=\boldsymbol{y}. \qquad (1.10)$$

Из последней системы находят искомый вектор $\boldsymbol{x}$. При этом разложение матрицы $\boldsymbol{A}$ на произведение $\boldsymbol{C}$x$\boldsymbol{D}$ – есть прямой ход метода Гаусса, а решение систем (1.10) обратный ход. Обозначим нижнюю треугольную матрицу через $\boldsymbol{L}$, верхнюю треугольную матрицу − $\boldsymbol{U}$.

Теорема об LU разложении

Введем обозначения: Δ_j − угловой минор порядка j матрицы $\boldsymbol{A}$, т.е.

$$\Delta_1 = a_{11},$$
$$\Delta_2 = det\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\Delta_m = det(\boldsymbol{A}).$$

Теорема. *Пусть все угловые миноры матрицы* $\boldsymbol{A}$ *не равны нулю* ($\Delta_j \neq 0$ *для* $j=\overline{1,m}$). *Тогда матрицу* $\boldsymbol{A}$ *можно представить единственным образом в виде произведения* $\boldsymbol{A}=\boldsymbol{L}*\boldsymbol{U}$.

Идея доказательства. Рассмотрим матрицу $\boldsymbol{A}$ второго порядка и будем искать разложение этой матрицы в виде $\boldsymbol{L}$ и $\boldsymbol{U}$.

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}=\begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix}*\begin{pmatrix} 1 & u_{12} \\ 0 & 1 \end{pmatrix}.$$

Сопоставляя эти два равенства, определяем элементы матриц $\boldsymbol{L}$ и $\boldsymbol{U}$ (перемножим и приравняем неизвестные). Система имеет единственное решение. Методом математической индукции сказанное можно обобщить для матрицы размерности $m \times m$.

Следствие. Метод Гаусса (схему единственного деления) можно применять только в том случае, когда угловые миноры матрицы A не равны нулю.

1.1.3 Метод Гаусса с выбором главного элемента

Может оказаться так, что система (1.1) имеет единственное решение, хотя какой либо из миноров матрицы A равен нулю. Заранее неизвестно, что все угловые миноры матрицы A не равны нулю. В этом случае можно использовать метод Гаусса с выбором главного элемента.

1. Выбор главного элемента по столбцу, когда на k-ом шаге исключения в качестве главного элемента выбирают максимальный по модулю коэффициент при неизвестном x_k в уравнениях с номерами $i=k, k+1,\ldots,m$. Затем уравнение, соответствующее выбранному коэффициенту, меняют местами с k-ым уравнением системы, чтобы ведущий элемент занял место коэффициента a_{kk}^{k-1}.После перестановки исключение неизвестного x_k выполняют как в схеме единственного деления.

ПРИМЕР. Пусть дана система второго порядка

$$\begin{cases} a_{11}x_1+a_{12}x_2=b_1 \\ a_{21}x_1+a_{22}x_2=b_2 \end{cases}$$

Предположим, что $|a_{21}|>|a_{11}|$, тогда переставим уравнения

$$\begin{cases} a_{21}x_1+a_{22}x_2=b_2 \\ a_{11}x_1+a_{12}x_2=b_1 \end{cases}$$

и применяем первый шаг прямого хода метода Гаусса. В этом случае имеет место перенумерация строк.

2. Выбор главного элемента по строке, т.е. производится перенумерация неизвестных системы.

При $|a_{12}|>|a_{11}|$, на первом шаге вместо неизвестного x_1 исключают x_2:

$$\begin{cases} a_{12}x_2+a_{11}x_1=b_1 \\ a_{22}x_2+a_{21}x_1=b_2 \end{cases}.$$

К этой системе применяем первый шаг прямого хода метода Гаусса.

3. Поиск главного элемента по всей матрице заключается в совместном применении методов 1 и 2. Всё это приводит к уменьшению вычислительной погрешности, но может замедлить процесс решения задачи.

Пример 2. Решить систему уравнений методом главного элемента с точностью до 0,0001

$$2{,}74x_1 - 1{,}18x_2 + 3{,}17x_3 = 2{,}18;$$
$$1{,}12x_1 + 0{,}83x_2 - 2{,}16x_3 = -1{,}15;$$
$$0{,}18x_1 + 1{,}27x_2 + 0{,}76x_3 = 3{,}23.$$

Вычисления производим по следующей схеме:

Коэффициенты при неизвестных			Свободные члены
x_1	x_2	x_3	
2,74 1,12 0,18	-1,18 0,83 1,27	**3,17** -2,16 0,76	2,18 -1,15 3,23
2,9870 -0,4768	0,0259 1,5528	—	0,3355 2,7075
—	**1,5569**	—	2,7602
0,0970	1,7728	1,2638	

Ответ:

$$x_2 = 2{,}7602 / 1{,}5569 = 1{,}7728;$$

$$x_1 = (0{,}3355 - 0{,}0259*1{,}7728) / 2{,}9870 = 0{,}0\ 970;$$

$$x_3 = (2{,}18\text{-}2{,}74*0{,}0970 + 1{,}18*1{,}7728) / 3{,}17 = 1{,}2638.$$

1.1.4 Метод Холецкого (метод квадратных корней)

Пусть дана система

$$\boldsymbol{Ax}=\boldsymbol{b}, \qquad (1.11)$$

где $\boldsymbol{A}$ − симметричная положительно определенная матрица.

Тогда решение системы (1.11) проводится в два этапа:

1. Симметричная матрица $\boldsymbol{A}$ представляется как произведение двух матриц

$$\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{L}^T.$$

Рассмотрим метод квадратных корней на примере системы 4-го порядка:

$$\begin{pmatrix} a_{11} & \dots & a_{14} \\ \dots & \dots & \dots \\ a_{41} & \dots & a_{44} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix} * \begin{pmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{34} \\ 0 & 0 & 0 & l_{44} \end{pmatrix}.$$

Перемножаем матрицы в правой части разложения и сравниваем с элементами в левой части:

$$l_{11}=\sqrt{a_{11}},\ l_{21}=\frac{a_{12}}{\sqrt{a_{11}}},\ l_{31}=\frac{a_{13}}{\sqrt{a_{11}}},\ l_{41}=\frac{a_{14}}{\sqrt{a_{11}}},\ l_{22}=\sqrt{a_{22}-l_{21}^2},$$

$$l_{32}=\frac{a_{32}-l_{21}l_{31}}{l_{22}},\ l_{33}=\sqrt{a_{33}-l_{31}^2-l_{32}^2},\ l_{44}=\sqrt{a_{44}-l_{41}^2-l_{42}^2-l_{43}^2}.$$

2. Решаем последовательно две системы

$$\boldsymbol{Ly}=\boldsymbol{b},$$
$$\boldsymbol{L}^T\boldsymbol{x}=\boldsymbol{y}.$$

Замечания

1) Под квадратным корнем может получиться отрицательное число, следовательно в программе необходимо предусмотреть использование правил действия с комплексными числами.

2) Возможно переполнение, если угловые элементы близки к нулю.

1.2 Итерационные методы решений систем алгебраических уравнений

Итерационные методы обычно применяются для решения систем большой размерности и они требуют приведения исходной системы к специальному виду.

Суть итерационных методов заключается в том, что решение $\boldsymbol{x}$ системы (1.1) находится как предел последовательности $\lim\limits_{n\to\infty} x(n)$.

Так как за конечное число итераций предел не может быть достигнут, то задаётся малое число ε − точность, и последовательные приближения вычисляют до тех пор, пока не будет выполнено неравенство

$$\left\| \boldsymbol{x}^n - \boldsymbol{x}^{n-1} \right\| < \varepsilon,$$

где $n=n(\varepsilon)$ – функция ε, $||\boldsymbol{x}||$ − норма вектора.

Определения основных норм в пространстве векторов и матриц.

Для вектора $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$ нормы вычисляются по следующим формулам:

$$\|\boldsymbol{x}\|_1 = \max_{1\le i\le n} |x_i|;$$

$$\|\boldsymbol{x}\|_2 = \sum_{i=1}^{n} |x_i|;$$

$$\|\boldsymbol{x}\|_3 = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Согласованные с ними нормы в пространстве матриц:

$$\|\boldsymbol{A}\|_1 = \max_{1\le i\le n} \left(\sum_{j=1}^{n} |a_{ij}| \right);$$

$$\|\boldsymbol{A}\|_2 = \max_{1\le j\le n} \left(\sum_{i=1}^{n} |a_{ij}| \right);$$

$\|\boldsymbol{A}\|_3 = \sqrt{\sum\limits_{i,j=1}^{n} |a_{ij}|^2}$ − величина, называемая евклидовой нормой матрицы $\boldsymbol{A}$.

Прямые методы рассчитаны для решения систем, порядок которых не больше 100, иначе используются итерационные методы.

1.2.1 Метод Якоби (простых итераций)

Исходную систему (1.11)

$$\boldsymbol{Ax=b}$$

преобразуем к виду:

$$x_i = -\sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j - \sum_{j=i+1}^{m}\frac{a_{ij}}{a_{ii}}x_j + \frac{b_i}{a_{ii}}, \qquad (1.12)$$

где $i=1,2,\ldots,m$; $a_{ii}\neq 0$.

Первая сумма равна нулю, если верхний предел суммирования меньше нижнего.

Так (1.12) при $i=1$ имеет вид

$$x_1 = -\sum_{j=2}^{m}\frac{a_{1j}}{a_{11}}x_j + \frac{b_1}{a_{11}}.$$

По методу Якоби x_i^{n+1} (n+1 приближение x_i) ищем по формуле

$$x_i^{n+1} = -\sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^n - \sum_{j=i+1}^{m}\frac{a_{ij}}{a_{ii}}x_j^n + \frac{b_i}{a_{ii}}. \qquad (1.13)$$

где n – номер итерации (0,1,…); $i=\overline{1,m}$.

Итерационный процесс (1.13) начинается с начальных значений x_i^0, которые в общем случае задаются произвольно, но предпочтительнее за x_i^0 взять свободные члены исходной системы.

Условие окончания счета:

$$\max_i \left| x_i^{n+1} - x_i^n \right| < \varepsilon, \qquad \text{где } i=\overline{1,m}.$$

Пример 3. Решить систему линейных уравнений методом итераций с точностью до 0,001.

$$0{,}68\,x_1 + 0{,}05\,x_2 - 0{,}11\,x_3 + 0{,}08\,x_4 = 2{,}15;$$
$$0{,}11\,x_1 - 0{,}84\,x_2 - 0{,}28\,x_3 - 0{,}06\,x_4 = 0{,}83;$$
$$-0{,}08\,x_1 + 0{,}15\,x_2 + x_3 - 0{,}12\,x_4 = 1{,}16;$$
$$0{,}21\,x_1 - 0{,}13\,x_2 + 0{,}27\,x_3 + x_4 = 0{,}44.$$

Решение.

$x_1 = 0{,}32\,x_1 - 0{,}05x_2 + 0{,}11\,x_3 - 0{,}08x_4 + 2{,}15;$

$x_2 = 0{,}11\,x_1 + 0{,}16\,x_2 - 0{,}28\,x_3 - 0{,}06\,x_4 - 0{,}83;$

$x_3 = 0{,}08\,x_1 - 0{,}15\,x_2 + 0{,}12\,x_4 + 1{,}16;$

$x_4 = -0{,}21\,x_1 + 0{,}13\,x_2 - 0{,}27\,x_3 + 0{,}44.$

Здесь $\|A\|_1 = \max\{0{,}56; 0{,}61; 0{,}35; 0{,}61\} = 0{,}61 < 1$, значит, итерационный процесс сходится.

Вычисления располагаем в таблице:

k	x_1	x_2	x_3	x_4
0	2,15	-0,83	1,16	0,44
1	2,9719	-1,0775	1,5093	-6,4326
2	3,3555	-1,0721	1,5075	-0,7317
3	3,5017	-1,0106	1,5015	-0,8111
4	3,5511	-0,9277	1,4944	-6*8321
5	3,5637	-0,9563	1,4834	-0,8298
6	3,5678	-0,9566	1,4890	-0,8332
7	3,5700	-0,9575	1,4889	-0,8356
8	3,5709	-0,9573	1,4890	-0,8362
9	3,5712	-0,9571	1,4889	-0,8364
10	3.5713	-0.9570	1.4890	-0.8364

Сходимость в тысячных долях имеет место уже на 10-м шаге. Ответ: $x_1 \approx 3{,}571$; $x_2 \approx -0{,}957$; $x_3 \approx 1{,}489$; $x_4 \approx -0{,}836$.

1.2.2 Метод Зейделя

Система (1.11) преобразуется к виду (1.12) и организуется итерационная процедура, где неизвестные x_i на $n+1$ шаге определяются по формулам

$$x_i^{n+1} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{n+1} - \sum_{j=i+1}^{m} \frac{a_{ij}}{a_{ii}} x_j^n + \frac{b_i}{a_{ii}}. \qquad (1.14)$$

Например,

$$x_1^{n+1} = -\sum_{j=2}^{m} \frac{a_{1j}}{a_{11}} x_j^n + \frac{b_1}{a_{11}}, \qquad (1.15)$$

$$x_2^{n+1} = -\sum_{j=3}^{m} \frac{a_{2j}}{a_{22}} x_j^n + \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{n+1}, \qquad (1.16)$$

и так далее.

Итерационные процессы (1.13) и (1.14) сходятся, если норма матрицы $\boldsymbol{A}$ ($\boldsymbol{A}$ − матрица коэффициентов при неизвестных в правой части систем (1.13) и (1.14)) удовлетворяет условию:

$$||\boldsymbol{A}||<1.$$

1.2.3 Матричная запись методов Якоби и Зейделя

Исходную матрицу системы (1.11) представим в виде суммы трёх матриц

$$\boldsymbol{A}=\boldsymbol{A}_1+\boldsymbol{D}+\boldsymbol{A}_2,$$

где $\boldsymbol{D}$ − диагональная матрица;

$\boldsymbol{D} = diag[a_{11}a_{22}...a_{mm}]$;

$\boldsymbol{A}_1$ − нижняя треугольная матрица;

$\boldsymbol{A}_2$ − верхняя треугольная матрица.

Пример: Дана матрица размерности (3×3):

$$\underbrace{\begin{pmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ a_{31} & a_{32} & 0 \end{pmatrix}}_{\boldsymbol{A}_1} + \underbrace{\begin{pmatrix} 0 & a_{12} & a_{13} \\ 0 & 0 & a_{23} \\ 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{A}_2} + \underbrace{\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}}_{\boldsymbol{D}} = A.$$

Тогда исходную систему (1.11) можно записать в виде

$$\boldsymbol{x}=-\boldsymbol{D}^{-1}\boldsymbol{A}_1\,\boldsymbol{x} - \boldsymbol{D}^{-1}\boldsymbol{A}_2\,\boldsymbol{x}+\boldsymbol{D}^{-1}\,\boldsymbol{b}.$$

Тогда метод Якоби можно записать в виде:

$$\boldsymbol{x}^{n+1}=-\boldsymbol{D}^{-1}\boldsymbol{A}_1\boldsymbol{x}^n-\boldsymbol{D}^{-1}\boldsymbol{A}_2\boldsymbol{x}^n+\boldsymbol{D}^{-1}\boldsymbol{b}$$

или

$$\boldsymbol{D}\boldsymbol{x}^{n+1}+(\boldsymbol{A}_1+\boldsymbol{A}_2)\,\boldsymbol{x}^n=\boldsymbol{b}. \qquad (1.17)$$

В матричной форме метод Зейделя будет выглядеть:

$$\boldsymbol{x}^{n+1}=-\boldsymbol{D}^{-1}\boldsymbol{A}_1\boldsymbol{x}^{n+1}-\boldsymbol{D}^{-1}\boldsymbol{A}_2\boldsymbol{x}^n+\boldsymbol{D}^{-1}\boldsymbol{b}$$

или

$$(\boldsymbol{D}+\boldsymbol{A}_1)\boldsymbol{x}^{n+1}+\boldsymbol{A}_2\,\boldsymbol{x}^n=\boldsymbol{b}. \qquad (1.18)$$

Преобразуем формулы (1.17) и (1.18):

$$\boldsymbol{D}(\boldsymbol{x}^{n+1}-\boldsymbol{x}^n)+\boldsymbol{A}\boldsymbol{x}^n=\boldsymbol{b}, \qquad (1.19)$$

$$(\boldsymbol{D}+\boldsymbol{A}_1)(\boldsymbol{x}^{n+1}-\boldsymbol{x}^n)+\boldsymbol{A}\boldsymbol{x}^n=\boldsymbol{b}. \qquad (1.20)$$

Из (1.19) и (1.20) видно, что если итерационный метод сходится, то он сходится к точному решению. Иногда при решении задач большой размерности, в итерационные методы вводятся числовые параметры, которые могут зависеть от номера итерации.

Пример для метода Якоби.

$$\boldsymbol{D}\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^n}{t_{n+1}}+\boldsymbol{A}\boldsymbol{x}^n=\boldsymbol{b},$$

где t – числовой параметр.

Возникают вопросы:

1) При каких значениях t сходимость будет наиболее быстрой?

2) При каких значениях t метод сходится?

На примере двух методов просматривается вывод о том, что одни и те же методы можно записывать несколькими способами. Поэтому вводят каноническую (стандартную) форму записи:

$$\boldsymbol{D}_{n+1}\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^n}{\boldsymbol{t}_{n+1}}+\boldsymbol{A}\boldsymbol{x}^n=\boldsymbol{b}. \qquad (1.21)$$

Формула (1.21) получена путем объединения (1.19) и (1.20).

Матрица $\boldsymbol{D}_{n+1}$ здесь задает тот или иной метод. Если существует обратная матрица к этой матрице, то из последней системы мы можем найти все неизвестные.

1. Метод (1.21) – явный, если матрица $\boldsymbol{D}_n$ совпадает с единичной матрицей и неявный – в противном случае.

2. Метод (1.21) – стационарный, если матрица $\boldsymbol{D}_{n+1}=\boldsymbol{D}$, и параметр t не зависит от номера итерации и нестационарный – в противном случае.

1.2.4 Метод Ричардсона

Явный метод с переменным параметром t:

$$\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^{n}}{t_{n+1}}+\boldsymbol{A}\boldsymbol{x}^{n}=\boldsymbol{b}, \qquad (1.21а)$$

называется методом Ричардсона.

1.2.5 Метод верхней релаксации (обобщённый метод Зейделя)

$$(\boldsymbol{D}+\omega\boldsymbol{A}_1)\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^{n}}{\omega}+\boldsymbol{A}\boldsymbol{x}^{n}=\boldsymbol{b}, \qquad (1.21б)$$

где ω – числовой параметр.

Если матрица $\boldsymbol{A}$ − симметричная и положительно определена, то последний метод сходится при ($0 < \omega < 2$). Последнюю формулу запишем в следующем виде:

$$(\boldsymbol{E}+\omega\boldsymbol{D}^{-1}\boldsymbol{A}_1)\boldsymbol{x}^{n+1}=((1-\omega)\boldsymbol{E}-\omega\boldsymbol{D}^{-1}\boldsymbol{A}_2)\boldsymbol{x}^{\mathrm{n}}+\omega\boldsymbol{D}^{-1}\boldsymbol{b}, \qquad (1.22)$$

где $\boldsymbol{E}$ – единичная матрица.

Тогда для вычисления неизвестных x_i ($i=\overline{1,m}$) можно записать итерационную процедуру в виде:

$$x_{\mathrm{i}}^{\mathrm{n+1}}+\omega\sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{n+1}=(1-\omega)x_i^{n}-\omega\sum_{j=i+1}^{m}\frac{a_{ij}}{a_{ii}}x_j^{n}+\omega\frac{b_i}{a_{ii}}. \qquad (1.23)$$

Например, для x_1 это будет такое выражение:

$$x_1^{n+1}=(1-\omega)x_1^{n}-\omega\sum_{j=2}^{m}\frac{a_{1j}}{a_{11}}x_j^{n}+\omega\frac{b_1}{a_{11}}.$$

1.2.6 Сходимость итерационных методов

Рассмотрим систему

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},$$

где $\boldsymbol{A}$ – невырожденная действительная матрица.

Для решения системы рассмотрим одношаговый стационарный метод

$$\boldsymbol{D}\frac{\boldsymbol{x}^{n+1}-\boldsymbol{x}^{n}}{t}+\boldsymbol{A}\boldsymbol{x}^{n}=\boldsymbol{b}, \qquad (1.24)$$

при n=0,1,2….

Предположим, что задан начальный вектор решения. Тогда метод (1.24) сходится, если норма вектора

$$\left\| \boldsymbol{x}-\boldsymbol{x}^{n} \right\|_{n\to\infty} \to 0.$$

Теорема. *Условие сходимости итерационного метода.*

Пусть $\boldsymbol{A}$ *– симметричная положительно определенная матрица и выполнено условие* $\boldsymbol{D}$ - $0.5t\boldsymbol{A} > 0$ (*где* $t > 0$). *Тогда итерационный метод* (1.24) *сходится.*

Следствие 1. Пусть $\boldsymbol{A}$ – симметричная и положительно определенная матрица с диагональным преобладанием, то есть

$$|a_{jj}| > \sum_{\substack{i=1 \\ i\neq j}}^{m} |a_{ij}|,$$

при j=1,2,…,m. Тогда метод Якоби сходится.

Следствие 2. Пусть $\boldsymbol{A}$ – симметричная и положительно определенная матрица с диагональным преобладанием, тогда метод верхней релаксации сходится при (0< ω<2).

Проверяется, при каком значении ω метод достигает заданной точности быстрее.

В частности, при ω = 1 метод верхней релаксации превращается в метод Зейделя, следовательно, при ω = 1 метод Зейделя сходится.

Теорема. *Итерационный метод* (*1.24*) *сходится при любом начальном векторе* $\boldsymbol{x}^0$ *тогда и только тогда, когда все собственные значения матрицы*

$$\boldsymbol{S}=\boldsymbol{E}-t\boldsymbol{D}^{-1}\boldsymbol{A}$$

по модулю меньше единицы.

2 Плохо обусловленные системы линейных алгебраических уравнений

Дана система линейных алгебраических уравнений

$$\boldsymbol{Ax}=\boldsymbol{b} \qquad (2.1)$$

Если система плохо обусловлена, то это значит, что погрешности коэффициентов матрицы $\boldsymbol{A}$ и правых частей $\boldsymbol{b}$ или же погрешности их округления сильно искажают решение системы.

Для оценки обусловленности системы вводят число обусловленности M_A

$$M_A=\| \boldsymbol{A}^{-1} \| \cdot \| \boldsymbol{A} \|.$$

Чем больше M_A, тем система хуже обусловлена.

Свойства числа обусловленности:

1) $M_E = 1$;

2) $M_A \geq 1$;

3) $M_A \geq |\lambda_{max}| / |\lambda_{min}|$, где λ_{max}, λ_{min} – соответственно максимальное и минимальное собственные числа матрицы $\boldsymbol{A}$;

4) $M_{AB} \leq M_A \times M_B$;

5) Число обусловленности матрицы $\boldsymbol{A}$ не меняется при умножении матрицы на произвольное число $\alpha \neq 0$.

Найдем выражение для полной оценки погрешности решения системы.

Пусть в системе (2.1) возмущены коэффициенты матрицы $\boldsymbol{A}$ и правая часть $\boldsymbol{b}$, т.е.

$$\delta \boldsymbol{A}=\tilde{\boldsymbol{A}}-\boldsymbol{A},\ \delta \boldsymbol{b}=\tilde{\boldsymbol{b}}-\boldsymbol{b},\ \delta \boldsymbol{x}=\tilde{\boldsymbol{x}}-\boldsymbol{x}.$$

Теорема. *Пусть матрица* $\boldsymbol{A}$ *имеет обратную матрицу, и выполняется условие* $\|\delta \boldsymbol{A}\| < \left\|\boldsymbol{A}^{-1}\right\|^{-1}$. *Тогда матрица* $\tilde{A}=\delta A+A$ *имеет обратную и справедлива следующая оценка относительной погрешности:*

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \frac{M_A}{1 - M_A \dfrac{\|\delta \boldsymbol{A}\|}{\|\boldsymbol{A}\|}} \left(\frac{\|\delta \boldsymbol{A}\|}{\|\boldsymbol{A}\|} + \frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} \right).$$

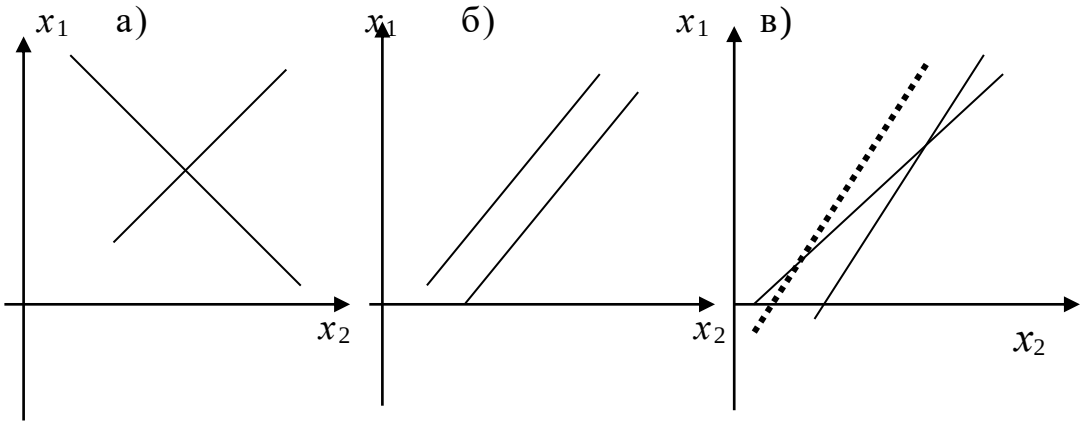


Рисунок 2 – а) система имеет единственное решение;
б) система не имеет решения;
в) система плохо обусловлена.

В случае в) малейшее возмущение системы сильно меняет положение точки пересечения прямых.

В качестве примера рассмотрим систему

$$\begin{cases} 1.03x_1 + 0.991x_2 = 2.51 \\ 0.991x_1 + 0.943x_2 = 2.41. \end{cases}$$

Решение этой системы

$$x_1 \approx 1.981$$
$$x_2 \approx 0.4735.$$

Оценим влияние погрешности правых частей на результат. Рассмотрим “возмущенную” систему с правой частью $\boldsymbol{b}^*$ = (2.505, 2.415) и решим эту систему:

$$x_1^* \approx 2.877$$
$$x_2^* \approx -0.4629.$$

Относительная погрешность правой части

$\delta(\boldsymbol{b}) = 0.005/2.51 \approx 0.28\%$ привела к относительной погрешности решения $\delta(\boldsymbol{x}^*) = 0.9364/1.981 \approx 47.3\%$.

Погрешность возросла примерно в 237 раз. Число обусловленности системы (2.1) приблизительно равно 237.

Подобные системы называются плохо обусловленными. Возникает вопрос: какими методами можно решать такие системы?

2.1 Метод регуляризации для решения плохо обусловленных систем

Рассмотрим систему

$$\boldsymbol{Ax}=\boldsymbol{b}. \qquad (2.1)$$

Для краткости перепишем эту систему в эквивалентный форме

$$(\boldsymbol{Ax}-\boldsymbol{b},\quad \boldsymbol{Ax}-\boldsymbol{b})=0. \qquad (2.2)$$

Для примера рассмотрим систему

$$\begin{cases} 2x_1 - x_2 = 1 \\ x_1 - 2x_2 = 2 \end{cases}.$$

Тогда ее можно представить как

$$(2x_1\text{-}x_2\text{-}1)^2+(x_1\text{-}2x_2\text{-}2)^2=0. \qquad (2.2^*)$$

Решение системы (2.2) совпадает с решением системы (2.2*).

Если коэффициенты $\boldsymbol{A}$ или $\boldsymbol{b}$ известны неточно, то решение также является не точным, поэтому вместо равенства $(\boldsymbol{Ax}-\boldsymbol{b}, \boldsymbol{Ax}-\boldsymbol{b})=0$ можем потребовать приближенного выполнения равенства $(\boldsymbol{Ax}-\boldsymbol{b}, \boldsymbol{Ax}-\boldsymbol{b})\approx 0$ и в этом виде задача становится не определенной и нужно добавить дополнительные условия.

В качестве дополнительного условия вводят требование, чтобы решение как можно меньше отклонялось от заданного $\boldsymbol{x}_0$ т.е. $(\boldsymbol{x}\text{-}\boldsymbol{x}_0, \boldsymbol{x}\text{-}\boldsymbol{x}_0)$ было минимальным. Следовательно, приходим к регуляризованной задаче вида

$$(\boldsymbol{Ax}\text{-}\boldsymbol{b}, \boldsymbol{Ax}\text{-}\boldsymbol{b})+\alpha(\boldsymbol{x}\text{-}\boldsymbol{x}_0, \boldsymbol{x}\text{-}\boldsymbol{x}_0)=min, \qquad (2.3)$$

где $\alpha > 0$.

Используя свойства скалярного произведения, выражение (2.3) перепишем в виде

$$(\boldsymbol{x},\boldsymbol{A}^T\boldsymbol{Ax})\text{-}2(\boldsymbol{x},\boldsymbol{A}^T\boldsymbol{b})+(\boldsymbol{b},\boldsymbol{b})+\alpha[(\boldsymbol{x},\boldsymbol{x})\text{-}2(\boldsymbol{x},\boldsymbol{x}_0)+(\boldsymbol{x}_0,\boldsymbol{x}_0)]=min. \qquad (2.4)$$

Варьируя $\boldsymbol{x}$ в уравнении (2.4), получим уравнение вида

$$(\boldsymbol{A}^T\boldsymbol{A}+\alpha\boldsymbol{E})\boldsymbol{x}=\boldsymbol{A}^T\boldsymbol{b}+\alpha\boldsymbol{x}_0. \qquad (2.5)$$

Система (2.5) – система линейных алгебраических уравнений, эквивалентная системе (2.1). Систему (2.5) решаем с помощью метода Гаусса или с помощью метода квадратных корней. Решая систему (2.5) найдем решение, которое зависит от числа α.

Выбор управляющего параметра α. Если α=0, то система (2.5) перейдет в плохо обусловленную систему (2.1).

Если же α– велико, то система (2.5) переходит в хорошо обусловленную систему и решение этой системы может сильно отличаться от решения системы (2.1).

Оптимальное значение α – это такое число, при котором система (2.5) удовлетворительно обусловлена.

На практике пользуются невязкой вида $\boldsymbol{r}_\alpha=\boldsymbol{A}\boldsymbol{x}_\alpha-\boldsymbol{b}$, и эту невязку сравнивают по норме с известной погрешностью правых частей $\delta\boldsymbol{b}$ и с влиянием погрешности коэффициентов матрицы $\delta\boldsymbol{A}$.

Если α– слишком велико, то $\boldsymbol{r}_\alpha >> \delta\boldsymbol{b}$ или $\delta\boldsymbol{A}$. Если α– мало, то $\boldsymbol{r}_\alpha << \delta\boldsymbol{b}$ или $\delta\boldsymbol{A}$.

Поэтому проводят серию расчетов, при различных α и в качестве оптимального значения выбирают то значение α, когда выполнено следующее условие

$$\|\boldsymbol{r}_\alpha\| \approx \|\delta\boldsymbol{b}\| + \|\delta\boldsymbol{A}\cdot\boldsymbol{x}\| \ .$$

Для выбора вектора $\boldsymbol{x}_0$ нужно знать приближенное решение или же, если приближенное решение трудно определить, то $\boldsymbol{x}_0$ =0.

2.2 Метод вращения (Гивенса)

Метод Гивенса, как и метод Гаусса состоит из прямого и обратного ходов.

Прямой ход метода. Исключаем неизвестное x_1 из всех уравнений, кроме первого. Для исключения x_1 из 2-го уравнения вычисляют числа

$$\alpha_{12}=\frac{a_{11}}{\sqrt{a_{11}^2+a_{21}^2}}, \qquad \beta_{12}=\frac{a_{21}}{\sqrt{a_{11}^2+a_{21}^2}},$$

где α и β такие, что $\alpha_{12}^2+\beta_{12}^2=1,\ -\beta_{12}a_{11}+\alpha_{12}a_{21}=0.$

Первое уравнение системы заменяем линейной комбинацией первого и второго уравнений с коэффициентами α_{12} и β_{12}, а второе уравнение такой же комбинацией с α_{12} и $-\beta_{12}$. В результате получим систему

$$\begin{cases} a_{11}^{(1)}x_1+a_{12}^{(1)}x_2+\ldots+a_{1m}^{(1)}x_m=b_1^{(1)} \\ \qquad a_{22}^{(1)}x_2+\ldots+a_{2m}^{(1)}x_m=b_2^{(1)} \\ a_{31}x_1+a_{32}x_2+\ldots+a_{3m}x_m=b_3. \\ \qquad\qquad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mm}x_m=b_m \end{cases} \qquad (2.6)$$

Здесь

$$a_{1j}^{(1)}=\alpha_{12}a_{1j}+\beta_{12}a_{2j},\quad a_{2j}^{(1)}=\alpha_{12}a_{2j}-\beta_{12}a_{1j},\quad b_1^{(1)}=\alpha_{12}b_1+\beta_{12}b_2,$$
$$b_2^{(1)}=\alpha_{12}b_2-\beta_{12}b_1,$$

где $j=\overline{1,m}$.

Преобразование системы (2.1) к системе (2.6) эквивалентно умножению слева матрицы $\boldsymbol{A}$ и вектора $\boldsymbol{b}$ на матрицу $\boldsymbol{C}_{12}$ вида

$$\boldsymbol{C}_{12}=\begin{pmatrix} \alpha_{12} & \beta_{12} & 0 & \ldots & 0 \\ -\beta_{12} & \alpha_{12} & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Аналогично для исключения x_1 из третьего уравнения вычисляем числа

$$\alpha_{13}=\frac{a_{11}^{(1)}}{\sqrt{(a_{11}^{(1)})^2+(a_{31}^{(1)})^2}}\ \text{ и }\ \beta_{13}=\frac{a_{31}}{\sqrt{(a_{11}^{(1)})^2+(a_{31}^{(1)})^2}},$$

такие, что $\alpha_{13}^2+\beta_{13}^2=1,\ \alpha_{13}a_{31}-\beta_{13}a_{11}^{(1)}=0.$

Затем первое уравнение системы (2.6) заменяем линейной комбинацией первого и третьего уравнений с коэффициентами α_{13},β_{13}, а третье уравнение системы (2.6) заменяем линейной комбинацией тех же уравнений, но с коэффициентами α_{13} и $-\beta_{13}$. Это преобразование эквивалентно умножению слева на матрицу

$$\boldsymbol{C}_{13}=\begin{pmatrix} \alpha_{13} & 0 & \beta_{13} & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ -\beta_{13} & 0 & \alpha_{13} & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Исключая неизвестное x_1 из всех последующих уравнений получим систему

$$\boldsymbol{A}^{(1)}\,\boldsymbol{x}=\boldsymbol{b}^{(1)},$$

где матрица на первом шаге $\boldsymbol{A}^{(1)}=\boldsymbol{C}_{1m}\dots\boldsymbol{C}_{13}\boldsymbol{C}_{12}\boldsymbol{A}$, , а вектор правых частей $\boldsymbol{b}^{(1)}=\boldsymbol{C}_{1m}\dots\boldsymbol{C}_{13}\boldsymbol{C}_{12}\boldsymbol{b}$.

Здесь и далее через $\boldsymbol{C}_{kj}$ обозначена матрица элементарного преобразования, отличающаяся от единичной матрицы $\boldsymbol{E}$ только четырьмя элементами.

Действие матрицы $\boldsymbol{C}_{kj}$ на вектор $\boldsymbol{x}$ эквивалентно повороту вектора $\boldsymbol{x}$ вокруг оси, перпендикулярной плоскости OX_kX_j на угол φ_{kj} такой, что

$$\alpha_{kj}=\cos\varphi_{kj},\ \beta_{kj}=\sin\varphi_{kj}.$$

Операцию умножения на матрицу $\boldsymbol{C}_{kj}$ называют плоским вращением или преобразованием Гивенса.

Первый этап состоит из m-1 шагов, в результате чего получается система

$$\begin{cases} a_{11}^{(m-1)}x_1+a_{12}^{(m-1)}x_2\dots+a_{1m}^{(m-1)}x_m=b_1^{(m-1)} \\ a_{22}^{(1)}x_2+\dots+a_{2m}^{(1)}x_m=b_2^{(1)} \\ \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ a_{m2}^{(1)}x_2+\dots+a_{mm}^{(1)}x_m=b_m^{(1)}. \end{cases} \qquad (2.7)$$

В матричной форме получаем $\boldsymbol{A}^{(1)}\,\boldsymbol{x}=\boldsymbol{b}^{(1)}$.

На втором этапе, состоящем из m-2 шагов, из уравнений системы (2.7) с номерами 3,4,…,m исключают неизвестное x_2. В результате получим систему

$$\begin{cases} a_{11}^{(m-1)}x_1 + a_{12}^{(m-1)}x_2 + a_{13}^{(m-1)}x_3 \ldots + a_{1m}^{(m-1)}x_m = b_1^{(m-1)} \\ a_{22}^{(m-1)}x_2 + a_{23}^{(m-1)}x_3 \ldots + a_{2m}^{(m-1)}x_m = b_2^{(m-1)} \\ a_{33}^{(2)}x_3 + \ldots + a_{3m}^{(2)}x_m = b_m^{(2)} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{m3}^{(2)}x_3 + \ldots + a_{mm}^{(2)}x_m = b_m^{(2)}. \end{cases}$$

В матричной форме получаем $\boldsymbol{A}^{(2)}\boldsymbol{x}=\boldsymbol{b}^{(2)}$, где $\boldsymbol{A}^{(2)}=\boldsymbol{C}_{2m}\ldots\boldsymbol{C}_{24}\boldsymbol{C}_{23}\boldsymbol{A}^{(1)}$, $\boldsymbol{b}^{(2)}=\boldsymbol{C}_{2m}\ldots\boldsymbol{C}_{24}\boldsymbol{C}_{23}\boldsymbol{b}^{(1)}$.

После завершения (m-1)-го шага придем к системе с верхней треугольной матрицей вида

$$\boldsymbol{A}^{(m-1)}\boldsymbol{x}=\boldsymbol{b}^{(m-1)},$$

где $\boldsymbol{A}^{(m-1)}=\boldsymbol{C}_{m-1,m}\boldsymbol{A}^{(m-2)}$, $\boldsymbol{b}^{(m-1)}=\boldsymbol{C}_{m-1,m}\boldsymbol{b}^{(m-2)}$.

Обратный ход метода вращений проводится точно так же, как и для метода Гаусса.

3 Решение нелинейных уравнений и систем нелинейных уравнений

Рассмотрим систему нелинейных уравнений с m неизвестными вида

$$\begin{cases} f_1(x_1,...,x_m)=0 \\ f_2(x_1,...,x_m)=0 \\ \quad\quad . \\ . \quad . \quad . \quad . \\ f_m(x_1,...,x_m)=0 \end{cases} . \qquad (3.1)$$

Задача решения такой системы является более сложной, чем нахождение корней одного нелинейного уравнения, и чем задача решения линейных алгебраических уравнений. В отличие от систем линейных уравнений здесь использование прямых методов исключено и решение находится с использованием итерационных методов, т.е. находится приближенное решение

$$\boldsymbol{x}^* = (x_1^*, \ldots, x_m^*),$$

удовлетворяющее при заданном $\varepsilon > 0$ условию $\left\|\boldsymbol{x}^* - \boldsymbol{x}\right\| < \varepsilon$.

Задача (3.1) совсем может не иметь решения или же число решений может быть произвольным. Введем векторную запись решения задачи:

$$\begin{gathered} \boldsymbol{x}=(x_1,\ldots,x_m)^T, \\ \boldsymbol{f}=(f_1,\ldots,f_m)^T, \\ \boldsymbol{f}(\boldsymbol{x})=0. \end{gathered} \qquad (3.2)$$

Будем считать, что функции f_i непрерывно дифференцируемы в некоторой окрестности точки $\boldsymbol{x}$. Введем матрицу Якоби

$$\boldsymbol{f}'(\boldsymbol{x})=\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_m} \end{pmatrix}.$$

Как и в случае решения одного уравнения начинаем с этапа локализации решения (отделения корней).

Пример. Дана система 2-х уравнений с двумя неизвестными

$$\begin{cases} x_1^3 + x_2^3 = 8x_1x_2 \\ x_1 \ln x_2 = x_2 \ln x_1 \end{cases}.$$

Найдем на плоскости место расположения решения.

Строим графики уравнений этой системы: а) – график 1-го уравнения, б) – график 2-го уравнения, в) – совмещенные графики.

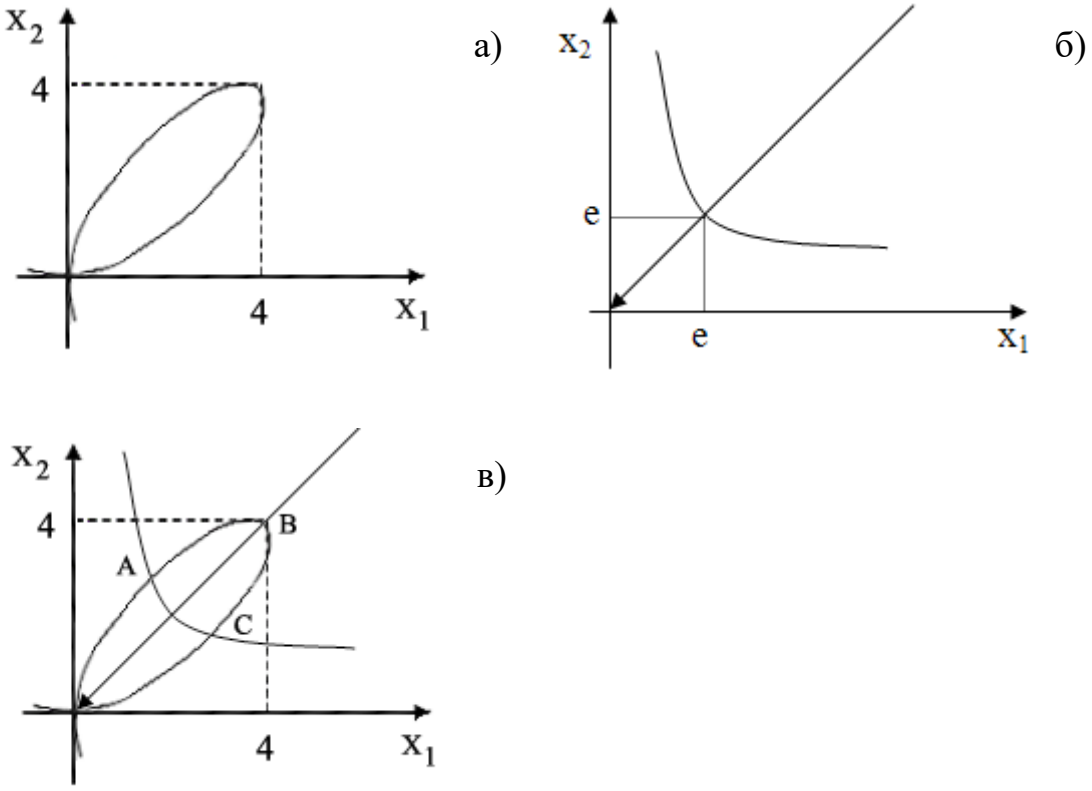


Рисунок 3 – Графики уравнений системы

Определяем границы координат пересечения графиков. Данная система имеет три решения. Координаты точек (*B*,*C*,*A*):

B: $x_1=4$, $x_2=4$

C: $3{,}5 < x_1 < 4$; $1{,}5 < x_2 < 2{,}5$.

Точки *A* и *C* симметричны относительно прямой $x_1=x_2$. Координаты точки *C* определим приближенно: $x_1 \approx 3{,}8$, $x_2 \approx 2$.

Обусловленность и корректность решения системы (3.1). Предположим что система (3.1) имеет решение $\boldsymbol{x}$ и в некоторой

окрестности этого решения матрица Якоби не вырождена. Это означает, что в указанной окрестности нет других решений системы.

В одномерном случае нахождение корня нелинейного уравнения приводит к определению интервала неопределенности $(x^*-\delta,\ x^*+\delta)$. Так как значения функции $f(x)$ чаще всего вычисляются на ЭВМ с использованием приближенных методов нельзя ожидать, что в окрестности корня относительная погрешность окажется малой. Сама погрешность корня ведет себя крайне нерегулярно и в первом приближении может восприниматься как некоторая случайная величина. На рисунке 4а представлена идеальная ситуация, отвечающая исходной математической постановке задачи, а на рисунке 4б – реальная, соответствующая вычислениям значений функции f на ЭВМ.

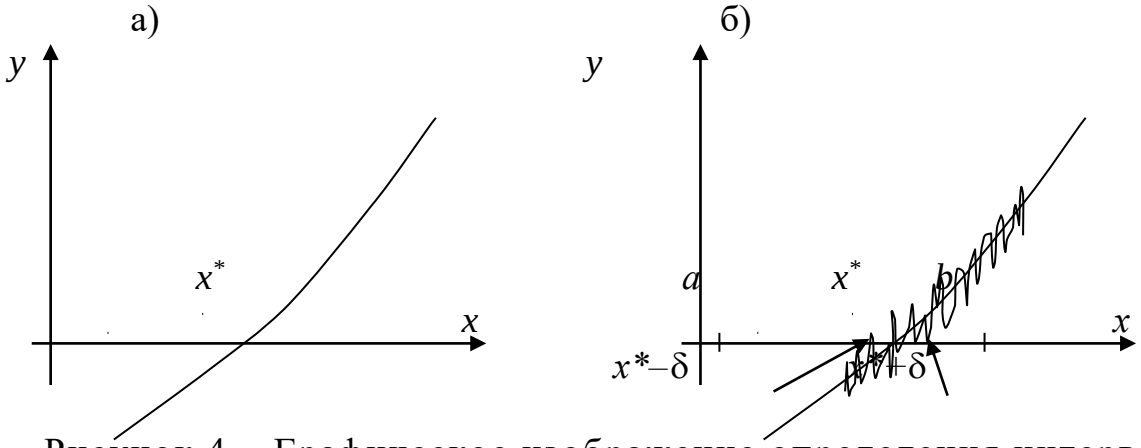


Рисунок 4 – Графическое изображение определения интервала неопределенности

В этом случае мы не можем определить, какая же точка в интервале неопределённости является решением. Радиус интервала неопределенности δ прямо пропорционален погрешности вычисления значения f. Кроме того, δ возрастает (обусловленность задачи ухудшается) с уменьшением $\left|f'(x^*)\right|$. Оценить величину δ довольно сложно, но выполнить это необходимо по следующим причинам:

– не имеет смысла ставить задачу о вычислении корня с точностью $\varepsilon < \delta$;

– после попадания очередного приближения в интервал неопределенности или близко от него, вычисления следует прекратить (этот момент для итерационных методов определяется крайне нерегулярным поведением приближений).

Если случай многомерный, то получаем некоторую область неопределённости D, и можем получить оценку радиуса ε этой области:

$$\bar{\varepsilon} \le \left\| (\boldsymbol{f}'(\boldsymbol{x}))^{-1} \right\| \cdot \Delta(\boldsymbol{f})$$
$$\left\| \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^*) \right\| \le \Delta(\boldsymbol{f}).$$

Роль абсолютного числа обусловленности играет норма матрицы, обратной матрице Якоби $f'(\boldsymbol{x})$. Чем число обусловленности больше, тем хуже эта система обусловлена.

3.1 Метод простых итераций

Систему (3.1) преобразуем к следующему эквивалентному виду:

$$\begin{cases} x_1 = \varphi_1(x_1,\ldots,x_m) \\ x_2 = \varphi_2(x_1,\ldots,x_m) \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ x_m = \varphi_m(x_1,\ldots,x_m) \end{cases}. \tag{3.3}$$

Или в векторной форме

$$\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{x}) \tag{3.4}$$

Пусть задано начальное приближение $\boldsymbol{x}^{(0)} = (x_1^{(0)} \ldots, x_m^{(0)})^T$. Подставляем его в правую часть системы (3.4) и получаем $\boldsymbol{x}^{(1)} = \boldsymbol{\varphi}(\boldsymbol{x}^{(0)})$, продолжая подстановку, находим $\boldsymbol{x}^{(2)}$ и т.д. Получим последовательность точек $\{x^{(0)}, x^{(1)}, \ldots, x^{(\kappa+1)}\}$, которая приближается к искомому решению $\boldsymbol{x}$.

3.1.1 Условия сходимости метода.

Пусть $\boldsymbol{\varphi}'(\boldsymbol{x})$ – матрица Якоби, соответствующая системе (3.4) и в некоторой Δ-окрестности решения $\boldsymbol{x}$ функции $\varphi_i(x)$ (i=1,2,…,m) дифференцируемы и выполнено неравенство вида:

$$\left\| \boldsymbol{\varphi}'(\boldsymbol{x}) \right\| \le q,$$

где ($0 \le q < 1$), q − постоянная.

Тогда независимо от выбора $\boldsymbol{x}^{(0)}$ из Δ-окрестности корня итерационная последовательность $\{\boldsymbol{x}^k\}$ не выходит за пределы данной окрестности, метод сходится со скоростью геометрической прогрессии и справедлива оценка погрешности

$$\left\| \boldsymbol{x}^{(n)} - \boldsymbol{x} \right\| \le q^n \left\| \boldsymbol{x}^{(0)} - \boldsymbol{x} \right\|.$$

3.1.2 Оценка погрешности.

В данной окрестности решения системы, производные функции $\varphi_i(\boldsymbol{x})$ (i=1,…,m) должны быть достаточно малы по абсолютной величине. Таким образом, если неравенство $\left\| \boldsymbol{\varphi}'(\boldsymbol{x}) \right\| \le q$ не выполнено, то исходную систему (3.1) следует преобразовать к виду (3.3).

Пример. Рассмотрим предыдущий пример и приведем систему к удобному для итераций виду

$$x_1 = \sqrt[3]{8x_1x_2 - x_2^3},$$
$$x_2 = x_2 + \frac{x_2}{\ln x_2} - \frac{x_1}{\ln x_1}.$$

Проверяем условие сходимости вблизи точки С. Вычислим матрицу Якоби

$$\boldsymbol{\varphi}'(x_1,x_2) = \begin{pmatrix} \dfrac{8x_2}{3(8x_1x_2 - x_2^3)^{2/3}} & \dfrac{8x_1 - 3x_2^2}{3(8x_1x_2 - x_2^3)^{2/3}} \\ \dfrac{1}{\ln^2 x_1} - \dfrac{1}{\ln x_1} & 1 + \dfrac{1}{\ln x_2} - \dfrac{1}{\ln^2 x_2} \end{pmatrix}.$$

Так как $x_1 \approx 3{,}8$, $x_2 \approx 2$, то при этих значениях вычисляем норму матрицы $\boldsymbol{\varphi}'(\boldsymbol{x})$

$$\| \boldsymbol{\varphi}'(\boldsymbol{x}) \| \approx \| \boldsymbol{\varphi}'(3{,}8\ ;\ 2) \| \approx 0{,}815.$$

Запишем итерационную процедуру

$$x_1^{(k+1)} = \sqrt[3]{8x_1^{(k)}x_2^{(k)} - (x_2^{(k)})^3},$$

$$x_2^{(k+1)} = x_2^{(k)} + \frac{x_2^{(k)}}{\ln x_2^{(k)}} - \frac{x_1^{(k)}}{\ln x_1^{(k)}}.$$

Следовательно, метод простых итераций будет сходиться со скоростью геометрической прогрессии, знаменатель которой $q \approx 0{,}815$. Вычисления поместим в таблице 1.

Таблица 1 Решение системы нелинейных уравнений

k	0	1	…	8	9
$x_1^{(k)}$	3,80000	3,75155	….	3,77440	x_1=3,77418
$x_2^{(k)}$	2,00000	2,03895	…	2,07732	x_2=2,07712

При k=9 критерий окончания счета выполняется при $\varepsilon = 10^{-3}$ и можно положить $x_1 = 3{,}774 \pm 0{,}001$ и $x_2 = 2{,}077 \pm 0{,}001$.

3.2 Метод Ньютона

Суть метода состоит в том, что система нелинейных уравнений сводится к решению систем линейных алгебраических уравнений. Пусть дана система (3.1) и задано начальное приближение $\boldsymbol{x}^{(0)}$. Приближение к решению $\boldsymbol{x}$ строим в виде последовательности $\boldsymbol{x}^{(0)}$, $\boldsymbol{x}^{(1)}$, …, $\boldsymbol{x}^{(n)}$.

В исходной системе (3.1) каждую функцию $f_i(x_1, x_2, ..., x_n)$, где $i = \overline{1, m}$, раскладывают в ряд Тейлора в точке $\boldsymbol{x}^{(n)}$ и заменяют линейной частью её разложения

$$f_i(\boldsymbol{x}) \approx f_i(\boldsymbol{x}^{(n)}) + \sum_{j=1}^{m} \frac{\partial f_i(\boldsymbol{x}^{(n)})}{\partial x_j}(x_j - x_j^{(n)}).$$

В результате получим систему линейных алгебраических уравнений

$$\begin{cases} f_1(\boldsymbol{x}^{(n)}) + \sum\limits_{j=1}^{m} \dfrac{\partial f_1(\boldsymbol{x}^{(n)})}{\partial x_j}(x_j - x_j^{(n)}) = 0 \\ \cdots\cdots\cdots\cdots \\ f_m(\boldsymbol{x}^{(n)}) + \sum\limits_{j=1}^{m} \dfrac{\partial f_m(\boldsymbol{x}^{(n)})}{\partial x_j}(x_j - x_j^{(n)}) = 0 \end{cases}. \qquad (3.5)$$

В матричной форме

$$\boldsymbol{f}(\boldsymbol{x}^{(n)}) + \boldsymbol{f}'(\boldsymbol{x}^{(n)}) * (\boldsymbol{x} - \boldsymbol{x}^{(n)}) = 0, \qquad (3.6)$$

где $\boldsymbol{f}'$ – матрица Якоби.

Предположим, что матрица невырожденная, то есть существует обратная матрица $\left[\boldsymbol{f}'(\boldsymbol{x}^{(n)})\right]^{-1}$.

Тогда система (3.6) имеет единственное решение, которое и принимается за очередное приближение $\boldsymbol{x}^{(n+1)}$. Отсюда выражаем решение $\boldsymbol{x}^{(n+1)}$ по итерационной формуле:

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \left[\boldsymbol{f}'(\boldsymbol{x}^{(n)})\right]^{-1} * \boldsymbol{f}(\boldsymbol{x}^{(n)}). \qquad (3.7)$$

Формула (3.7) и есть итерационная формула метода Ньютона для приближенного решения системы нелинейных уравнений.

Замечание. В таком виде формула (3.7) используется редко в виду того, что на каждой итерации нужно находить обратную матрицу. Поэтому поступают следующим образом: вместо системы (3.6) решают эквиваленту ей систему линейных алгебраических уравнений вида

$$\boldsymbol{f}'(\boldsymbol{x}^{(n)}) * \Delta\boldsymbol{x}^{(n+1)} = -\boldsymbol{f}(\boldsymbol{x}^{(n)}). \qquad (3.8)$$

Это система линейных алгебраических уравнений относительно поправки $\Delta\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n+1)} - \boldsymbol{x}^{(n)}$. Затем полагают

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \Delta\boldsymbol{x}^{(n+1)}. \qquad (3.9)$$

3.2.1 Сходимость метода

Теорема. *Пусть в некоторой окрестности решения* $\boldsymbol{x}$ *системы* (3.1) *функции* f_i (*при* $i=\overline{1,m}$) *дважды непрерывно дифференцируемы и матрица Якоби не вырождена. Тогда найдется такая малая* δ *окрестность вокруг решения* $\boldsymbol{x}$*, что при выборе начального приближения* $\boldsymbol{x}^0$ *из этой окрестности итерационный метод* (3.7) *не выйдет за пределы этой окрестности решения и справедлива оценка вида*

$$\left\| \boldsymbol{x}^{(n+1)} - \boldsymbol{x} \right\| \le \frac{1}{\delta} \left\| \boldsymbol{x}^{(n)} - \boldsymbol{x} \right\|^2,$$

где n – номер итерации.

Метод Ньютона сходится с квадратичной скоростью. На практике используется следующий критерий остановки:

$$\left\| \boldsymbol{x}^{(n)} - \boldsymbol{x}^{(n+1)} \right\| < \varepsilon.$$

4 Решение проблемы собственных значений

Пусть дана квадратная матрица $\boldsymbol{A}$ размерностью ($m*m$) и существует такое число λ, что выполняется равенство

$$\boldsymbol{A}\cdot\boldsymbol{x}=\lambda\cdot\boldsymbol{x},\ \boldsymbol{x}\neq 0,$$

тогда такое число λ называется собственным значением матрицы $\boldsymbol{A}$, а $\boldsymbol{x}$ – соответствующим ему собственным вектором.

Перепишем это равенство в эквивалентной форме

$$(\boldsymbol{A}-\lambda\boldsymbol{E})\boldsymbol{x}=0\ . \qquad (4.1)$$

Система (4.1) – однородная система линейных алгебраических уравнений. Для существования нетривиального решения системы (4.1) должно выполняться условие

$$det(\boldsymbol{A}-\lambda\boldsymbol{E})=0. \qquad (4.2)$$

Определитель в левой части уравнения является многочленом m-ой степени относительно λ, его называют − характеристическим определителем (характеристическим многочленом). Следовательно, уравнение (4.2) имеет m корней или m собственных значений. Среди них могут быть как действительные, так и комплексные корни.

Задача вычисления собственных значений сводится к нахождению корней характеристического многочлена (4.2). Корни могут быть найдены одним из итерационных методов (в частности методом Ньютона).

Если найдено некоторое собственное значение матрицы $\boldsymbol{A}$, то, подставив это число в систему (4.1) и решив эту систему однородных уравнений, находим собственный вектор $\boldsymbol{x}$, соответствующий данному собственному значению.

Собственные вектора будем при нахождении нормировать (вектор $\boldsymbol{x}$ умножаем на $||\boldsymbol{x}||^{-1}$, и таким образом они будут иметь единичную длину), нахождение собственных значений матрицы $\boldsymbol{A}$ и соответствующих им собственных векторов и есть полное решение проблемы собственных значений. А нахождение отдельных собственных значений и соответствующих им векторов − называется решением частной проблемы собственных значений.

Эта проблема имеет самостоятельное значение на практике. Например, в электрических и механических системах собственные значения отвечают собственным частотам колебаний, а собственные вектора характеризуют соответствующие формы колебаний.

Эта задача легко решается для некоторых видов матриц: диагональных, треугольных и трехдиагональных матриц.

К примеру, определитель треугольной или диагональной матрицы равен произведению диагональных элементов, тогда и собственные числа равны диагональным элементам.

Пример. Матрица $\boldsymbol{A}$ – диагональная $\boldsymbol{A}=\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix}$.

Тогда $det(\boldsymbol{A}-\lambda\boldsymbol{E})=(a-\lambda)^3$, а характеристическое уравнение $(a-\lambda)^3=0$ имеет трехкратный корень $\lambda=a$.

Собственными векторами для матрицы $\boldsymbol{A}$ будут единичные векторы

$$\boldsymbol{e}_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \boldsymbol{e}_2=\begin{pmatrix}0\\1\\0\end{pmatrix},\quad \boldsymbol{e}_3=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Пример. Найдем собственные числа матрицы

$$\boldsymbol{A}=\begin{pmatrix} 2 & -9 & 5 \\ 1{,}2 & -5{,}999 & 6 \\ 1 & -1 & -7{,}5 \end{pmatrix}.$$

Составим характеристический многочлен

$$P_3(\lambda)=det(\boldsymbol{A}-\lambda\boldsymbol{E})=det\begin{pmatrix} 2-\lambda & -9 & 5 \\ 1{,}2 & -5{,}3999-\lambda & 6 \\ 1 & -1 & -7{,}5-\lambda \end{pmatrix}=$$
$$=-\lambda^3-10{,}8999\lambda^2-26{,}49945\lambda-21{,}002.$$

Используя метод Ньютона, определим один из корней уравнения $P_3(\lambda)=0$, а именно $\lambda\approx$-7,87279.

Разделив многочлен $P_3(\lambda)$ на $(\lambda-\lambda_1)$ получим многочлен второй степени: $P_2(\lambda)=\lambda^2+3{,}02711\lambda+2{,}66765$. Решив квадратное уравнение, находим оставшиеся два корня:

$\lambda_{2,3}\approx -1{,}51356\pm 0{,}613841*i$ (комплексные сопряженные корни).

Существуют прямые методы нахождения собственных значений и итерационные методы. Прямые методы неудобны для нахождения собственных значений для матриц высокого порядка. В таких случаях с учетом возможностей компьютера более удобны итерационные методы.

4.1 Прямые методы нахождения собственных значений

4.1.1 Метод Леверрье

Метод разделяется на две стадии:

- раскрытие характеристического уравнения,
- нахождение корней многочлена.

Пусть $det(\boldsymbol{A}-\lambda\boldsymbol{E})$ – есть характеристический многочлен матрицы $\boldsymbol{A}=\{a_{ij}\}$ $(i,j=1,2,\ldots,m)$, т.е. $det(\boldsymbol{A}-\lambda\boldsymbol{E})=\lambda^m+p_1\lambda^{m-1}+\cdots+p_m$, и $\lambda_1,\lambda_2,\ldots,\lambda_m$ – есть полная совокупность корней этого многочлена (полный спектр собственных значений).

Рассмотрим суммы вида

$$S_k=\lambda_1^k+\lambda_2^k+\cdots+\lambda_m^k \quad (k=1,2,\ldots,m), \text{ т.е.}$$

$$\begin{aligned} S_1&=\lambda_1+\lambda_2+\ldots+\lambda_m=SpA\\ S_2&=\lambda_1^2+\lambda_2^2+\cdots+\lambda_m^2=SpA^2\\ &\cdots\cdots\cdots\\ S_m&=\lambda_1^m+\lambda_2^m+\cdots+\lambda_m^m=SpA^m \end{aligned} \quad , \tag{4.3}$$

где $SpA=\sum_{i=1}^{m}a_{ii}$ – след матрицы.

В этом случае при $k\le m$ справедливы формулы Ньютона для всех $(1\le k\le m)$

$$S_k+p_1S_{k-1}+\cdots+p_{k-1}S_1=-kp_k, \tag{4.4}$$

Откуда получаем

при k=1 $p_1 = -S_1$,

при k=2 $p_2 = -1/2 \cdot (S_2 + p_1 \cdot S_1)$, (4.5)

. . . . . .

при $k=m$ $p_m = -1/n \cdot (S_m + p_1 \cdot S_{m-1} + p_2 \cdot S_{m-2} + \ldots + p_{m-1} \cdot S_1)$.

Следовательно, коэффициенты характеристического многочлена p_i можно определить, если известны суммы $S_1, S_2, \ldots, S_m$. Тогда схема алгоритма раскрытия характеристического определителя методом Леверрье будет следующей:

1) вычисляют степень матрицы: $\boldsymbol{A}^k = \boldsymbol{A}^{k-1} \cdot \boldsymbol{A}$ для $k=1,\ldots,m$;

2) определяют S_k – суммы элементов, стоящих на главной диагонали матриц $\boldsymbol{A}^k$;

3) по формулам (4.5) находят коэффициенты характеристического уравнения $p_i (i=1,2,\ldots,m)$.

4.1.2 Усовершенствованный метод Фадеева

Алгоритм метода:

1) вычисляют элементы матриц $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_m$:

$$\boldsymbol{A}_1 = \boldsymbol{A}; \quad SpA_1 = q_1; \quad \boldsymbol{B}_1 = \boldsymbol{A}_1 - q_1 \cdot \boldsymbol{E};$$
$$\boldsymbol{A}_2 = \boldsymbol{A} * \boldsymbol{B}_1; \quad \frac{SpA_2}{2} = q_2; \quad \boldsymbol{B}_2 = \boldsymbol{A}_2 - q_2 \cdot \boldsymbol{E};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\boldsymbol{A}_m = \boldsymbol{A} * \boldsymbol{B}_{m-1}; \quad \frac{SpA_m}{m} = q_m; \quad \boldsymbol{B}_m = \boldsymbol{A}_m - q_m \cdot \boldsymbol{E},$$

(в конце подсчета $\boldsymbol{B}_m$ – нулевая матрица для контроля);

2) определяют коэффициенты характеристического уравнения p_i: $q_1 = -p_1$, $q_2 = -p_2, \ldots, q_m = -p_m$.

Существуют и другие методы раскрытия характеристического определителя: метод Крылова, Данилевского и др.

4.1.3 Метод Данилевского

Две матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ называются подобными, если одна получается из другой путем преобразования с помощью некоторой не вырожденной матрицы $\boldsymbol{S}$:

$$\boldsymbol{B} = \boldsymbol{S}^{-1} \cdot \boldsymbol{AS},$$

если это равенство справедливо, то матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ подобны, а само преобразование называется преобразованием подобия (переход к новому базису в пространстве m - мерных векторов).

Пусть $\boldsymbol{y}$ − результат применения матрицы $\boldsymbol{A}$ к вектору $\boldsymbol{x}$

$$\boldsymbol{y}=\boldsymbol{A}\cdot\boldsymbol{x}.$$

Сделаем замену переменных:

$$\boldsymbol{x}=\boldsymbol{S}\cdot\boldsymbol{x}'\ ,\ \boldsymbol{y}=\boldsymbol{S}\cdot\boldsymbol{y}'.$$

Тогда равенство $\boldsymbol{y}=\boldsymbol{A}\cdot\boldsymbol{x}$ преобразуется к виду

$$\boldsymbol{y}'=\boldsymbol{S}^{-1}\cdot\boldsymbol{A}\cdot\boldsymbol{S}\cdot\boldsymbol{x}'.$$

В этом случае матрица $\boldsymbol{B}$ и матрица $\boldsymbol{A}$ имеют одни и те же собственные числа. Это можно легко увидеть раскрыв определитель

$$det(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}-\lambda\boldsymbol{E})=det(\boldsymbol{S}^{-1}(\boldsymbol{A}-\lambda\boldsymbol{E})\boldsymbol{S})=$$
$$=det(\boldsymbol{S}^{-1})\cdot det(\boldsymbol{A}-\lambda\boldsymbol{E})\cdot det(\boldsymbol{S})=det(\boldsymbol{A}-\lambda\boldsymbol{E}).$$

Следовательно, матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ − подобные, имеют одни и те же собственные значения. Но собственные векторы $\boldsymbol{x}$ и $\boldsymbol{x}'$ не совпадают, они связаны между собой простым соотношением

$$\boldsymbol{x}\ =\ \boldsymbol{S}\cdot\boldsymbol{x}'.$$

Такую матрицу $\boldsymbol{A}$ с помощью преобразования подобия или же последовательности таких преобразований можно привести к матрице Фробениуса вида:

$$\boldsymbol{F}=\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1,m-1} & f_{1m} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Детерминант матрицы $\boldsymbol{F}$ – $det(\boldsymbol{F})$ можно разложить по элементам первой строки:

$$det(\boldsymbol{F}-\lambda\boldsymbol{E})=(-1)^m(\lambda^m-p_1\lambda^{m-1}-\cdots-p_m).$$

Тогда коэффициенты характеристического многочлена матрицы $\boldsymbol{A}$ будут

$$p_1 = f_{11}\,,\; p_2 = f_{12},\ldots,\; p_m = f_{1m}.$$

Второй случай. Матрицу $\boldsymbol{A}$ преобразованием подобия можно привести к матрице $\boldsymbol{B}$ верхнего треугольного вида

$$\boldsymbol{B}=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ 0 & b_{22} & \cdots & b_{2m} \\ . & . & . & . \\ 0 & 0 & \cdots & b_{mm} \end{pmatrix}.$$

Тогда собственными числами будут диагональные элементы матрицы $\boldsymbol{B}$:

$$det(\boldsymbol{B}-\lambda\boldsymbol{E})=(b_{11}-\lambda)(b_{22}-\lambda)\cdots(b_{mm}-\lambda).$$

Третий случай. Матрицу $\boldsymbol{A}$ с помощью преобразования подобия можно привести к Жордановой форме $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{\Lambda}$

$$\boldsymbol{\Lambda}=\begin{pmatrix} \lambda_1 & S_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & S_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_m \end{pmatrix},$$

где λ_i – собственные числа матрицы $\boldsymbol{A}$; S_i – константы (0 или 1); если S_i=1, то $\lambda_i=\lambda_{i+1}$.

К четвёртому случаю относятся матрицы, которые с помощью преобразования подобия можно привести к диагональному виду (матрица простой структуры):

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{D}=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \cdots & \ldots & \cdots & \ldots\\ 0 & 0 & \cdots & \lambda_n\end{pmatrix},$$

у которой, как известно, собственными числами являются диагональные элементы.

4.1.4 Метод итераций определения первого собственного числа матрицы.

Пусть дано характеристическое уравнение:

$$det(\boldsymbol{A}-\lambda\cdot\boldsymbol{E}) = 0,$$

где $\lambda_1, \lambda_2,\ldots, \lambda_n$ – собственные значения матрицы $\boldsymbol{A}$.

Предположим, что $|\lambda_1|>|\lambda_2|\geq|\lambda_3|\geq\ldots\geq|\lambda_m|$, т.е. λ_1 – наибольшее по модулю собственное число.

Тогда для нахождения приближенного значения λ_1 используется следующая схема:

1) выбирают произвольно начальный вектор $\boldsymbol{y}^{(0)}$;

2) строят последовательность итераций вида:

$$\begin{aligned}
&\boldsymbol{y}^{(1)}=\boldsymbol{A}\boldsymbol{y}^{(0)},\\
&\boldsymbol{y}^{(2)}=\boldsymbol{A}\cdot\boldsymbol{A}\boldsymbol{y}^{(0)}=\boldsymbol{A}^2\boldsymbol{y}^{(0)},\\
&\ldots\ldots\ldots\\
&\boldsymbol{y}^{(m)}=\boldsymbol{A}\cdot\boldsymbol{A}^{m-1}\boldsymbol{y}^{(0)}=\boldsymbol{A}^m\boldsymbol{y}^{(0)},\\
&\boldsymbol{y}^{(m+1)}=\boldsymbol{A}\cdot\boldsymbol{A}^m\boldsymbol{y}^{(0)}=\boldsymbol{A}^{m+1}\boldsymbol{y}^{(0)}.
\end{aligned}$$

3) выбирают $\boldsymbol{y}^{(m)}=\boldsymbol{A}^m\boldsymbol{y}^{(0)}$ и $\boldsymbol{y}^{(m+1)}=\boldsymbol{A}^{m+1}\boldsymbol{y}^{(0)}$, тогда

$$\lambda_1=\lim_{n\to\infty}\frac{y_i^{(m+1)}}{y_i^{(m)}} \text{ или } \lambda_1\approx\frac{y_i^{(m+1)}}{y_i^{(m)}},$$

где $y_i^{(m)}$ и $y_i^{(m+1)}$ – соответствующие координаты векторов $\boldsymbol{y}^{(m)}$ и $\boldsymbol{y}^{(m+1)}$.

Возникает вопрос выбора начального вектора $\boldsymbol{y}^{(0)}$. При неудачном выборе можем не получить значения нужного корня, или же предела может не существовать. Этот факт при вычислении можно заметить по прыгающим значениям этого отношения, следовательно, нужно изменить $\boldsymbol{y}^{(0)}$. В качестве первого собственного вектора можно взять вектор $\boldsymbol{y}^{(n+1)}$ и пронормировать его.

Пример. Найти наибольшее по модулю собственное значение и соответствующий ему собственный вектор матрицы $\boldsymbol{A}$

$$\boldsymbol{A}=\begin{pmatrix}3 & 1 & 0\\ 1 & 2 & 2\\ 0 & 1 & 1\end{pmatrix}.$$

1) Выбираем начальный вектор $\boldsymbol{y}^{(0)}=\begin{pmatrix}1\\ 1\\ 1\end{pmatrix}$.

2) Вычисляем последовательно векторы $\boldsymbol{y}^{(1)}$, $\boldsymbol{y}^{(2)}$, ..., $\boldsymbol{y}^{(10)}$. Вычисления помещаем в таблицу 2.

Таблица 2 – Вычисление векторов $\boldsymbol{y}^{(n+1)}$

$\boldsymbol{y}^{(0)}$	$\boldsymbol{A}\cdot\boldsymbol{y}^{(0)}$	$A^2\cdot\boldsymbol{y}^{(0)}$	$\boldsymbol{A}^3\cdot\boldsymbol{y}^{(0)}$	.....	$\boldsymbol{A}^9\cdot\boldsymbol{y}^{(0)}$	$\boldsymbol{A}^{10}\cdot\boldsymbol{y}^{(0)}$
1	4	17	69		243569	941370
1	5	18	67		210663	812585
1	2	7	25		73845	284508

3) Вычисляем отношения координат векторов $y_i^{(10)}$ и $y_i^{(9)}$

$$\lambda_1^{(1)}=\frac{y_1^{(10)}}{y_1^{(9)}}=3{,}865;\ \lambda_1^{(2)}=\frac{y_2^{(10)}}{y_2^{(9)}}=3{,}857;\ \lambda_1^{(3)}=\frac{y_3^{(10)}}{y_3^{(9)}}=3{,}853.$$

4) Вычисляем λ_1 как среднее арифметическое $\lambda_1^{(1)}$, $\lambda_1^{(2)}$, $\lambda_1^{(3)}$.

$$\lambda_1=\frac{\lambda_1^{(1)}+\lambda_1^{(2)}+\lambda_1^{(3)}}{3}=3{,}858.$$

5) Определим соответствующий числу λ_1собственный вектор:

$$\boldsymbol{y}^{(10)}=\boldsymbol{A}^{(10)}\cdot\boldsymbol{y}^{(0)}=\begin{pmatrix}941370\\ 812585\\ 284508\end{pmatrix}.$$

6) Нормируем вектор $\boldsymbol{y}^{(10)}$, разделив на его длину

$$\left\|\boldsymbol{y}^{(10)}\right\|_3=\sqrt{941370^2+812585^2+284508^2}=1{,}28\cdot10^6$$

получим вектор

$$\boldsymbol{x}_1 = \begin{pmatrix} 0{,}74 \\ 0{,}64 \\ 0{,}22 \end{pmatrix}.$$

Далее можем определить второе собственное число

$$\lambda_2 \approx \frac{y_i^{(n+1)} - \lambda_1 y_i^{(n)}}{y_i^{(n+1)} - \lambda_1 y_i^{(n-1)}},$$

где $i=1,2,\ldots,n$.

При вычислении собственных чисел подобным образом, будет накапливаться ошибка. Данная методика позволяет приближенно оценить собственные значения матрицы.

5 Задача приближения функции

Постановка задачи.

Пусть на отрезке $[a,b]$ функция $\boldsymbol{y}=f(\boldsymbol{x})$ задана таблицей своих значений $y_0 = f(x_0),...,y_n = f(x_n)$.

Допустим, что вид функции $f(x)$ неизвестен. На практике часто встречается задача вычисления значений функции $y=f(x)$ в точках x, отличных от $x_0,...,x_n$. Кроме того, в некоторых случаях, не смотря на то, что аналитическое выражение $y=f(x)$ известно, оно может быть слишком громоздким и неудобным для математических преобразований (например, специальные функции). Кроме этого значения y_i могут содержать ошибки эксперимента.

Определение . Точки $x_0, ..., x_n$ называются узлами интерполяции.

Требуется найти аналитическое выражение функции $F(x)$, совпадающей в узлах интерполяции со значениями данной функции, т.е.

$$F(x_0) = y_0, F(x_1) = y_1, ..., F(x_n) = y_n.$$

Определение. Процесс вычисления значений функции $F(x)$ в точках отличных от узлов интерполирования называется интерполированием функции $f(x)$. Если $x \in [x_0, x_n]$, то задача вычисления приближенного значения функции в т. x называется интерполированием, иначе – экстраполированием.

Геометрически задача интерполирования функции одной переменной означает построение кривой, проходящей через заданные точки $(x_0, y_0),(x_1, y_1),...,(x_n, y_n)$ (рисунок 5). То есть задача в такой постановке может иметь бесконечное число решений.

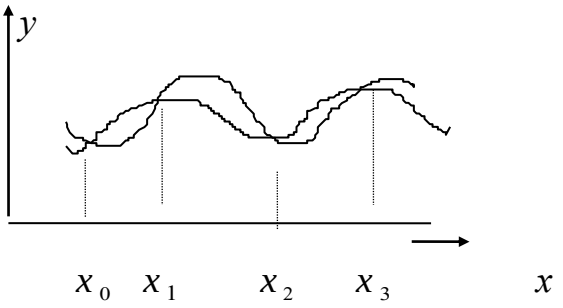


Рисунок 5 – Геометрическая иллюстрация задачи интерполирования функции

Задача становится однозначной, если в качестве $F(x)$ выбрать многочлен степени не выше n, такой что:

$$F_n(x_0)=y_0,\ F_n(x_1)=y_1,\ \ldots,\ F_n(x_n)=y_n.$$

Определение. Многочлен $F_n(x)$, отвечающий вышеназванным условиям, называется интерполяционным многочленом.

Знание свойств функции f позволяет осознанно выбирать класс G аппроксимирующих функций. Широко используется класс функций вида

$$\Phi_m(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_m\varphi_m(x), \qquad (5.1)$$

являющихся линейными комбинациями некоторых базисных функций $\varphi_0(x), \ldots, \varphi_m(x)$.

Будем искать приближающую функцию в виде многочлена степени m, с коэффициентами $c_0, \ldots, c_m$, которые находятся в зависимости от вида приближения. Функцию $\Phi_m(x)$ называют обобщенным многочленом по системе функций $\varphi_0(x), \varphi_1(x), \ldots, \varphi_m(x)$, а число m – его степенью. Назовем обобщенный многочлен $\Phi_m(x)$ интерполяционным, если он удовлетворяет условию

$$\Phi_m(x_i)=y_i,\ (i=0,1,\ldots,n). \qquad (5.2)$$

Покажем, что условие (5.2) позволяет найти приближающую функцию единственным образом

$$\begin{cases} c_0\varphi_0(x_0)+c_1\varphi_1(x_0)+\ldots+c_m\varphi_m(x_0)=y_0 \\ c_0\varphi_0(x_1)+c_1\varphi_1(x_1)+\ldots+c_m\varphi_m(x_1)=y_1 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ c_0\varphi_0(x_n)+c_1\varphi_1(x_n)+\ldots+c_m\varphi_m(x_n)=y_n, \end{cases} \tag{5.3}$$

Система (5.3) есть система линейных алгебраических уравнений относительно коэффициентов $c_0, c_1, \ldots, c_m$.

Эта система n линейных уравнений имеет единственное решение, если выполняется условие $m=n$ и определитель квадратной матрицы $\boldsymbol{P}$

$$detP = \begin{vmatrix} \varphi_0(x_0),\varphi_1(x_0),\ldots, & \varphi_n(x_0) \\ \varphi_0(x_1), \varphi_1(x_1) ,\ldots, & \varphi_n(x_1) \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \\ \varphi_0(x_n),\varphi_1(x_n),\ldots, & \varphi_n(x_n) \end{vmatrix} \neq 0.$$

Определение. Система функций $\varphi_0(x),\ldots,\varphi_n(x)$ называется Чебышевской системой функций на $[a,b]$,если определитель матрицы отличен от нуля $detP\neq 0$ при любом расположении узлов $x_i\in[a,b]$, $i=0,1,\ldots n$, когда среди этих узлов нет совпадающих.

Если мы имеем такую систему функций, то можно утверждать, что существует единственный для данной системы функций интерполяционный многочлен $\Phi_m(x)$, коэффициенты которого определяются единственным образом из системы (5.3).

Пример. При $m\leq n$ система функций 1, x, $x^2, \ldots, x^m$ линейно независима в точках $x_0, x_1, \ldots, x_n$, если они попарно различны.

5.1 Интерполяционный многочлен Лагранжа

Рассмотрим случай, когда узлы интерполирования не равноотстоят друг от друга на отрезке $[a,b]$.

Тогда шаг $h=x_{i+1}-x_i\neq\text{const}$. Задача имеет единственное решение, если в качестве интерполирующей функции $F(x)$ взять алгебраический многочлен

$$L_n(x)=a_0+a_1\,x+a_2\,x^2+\ldots+a_nx^n,$$

где a_i неизвестные постоянные коэффициенты.

Используя условие (5.2) можем записать

$$L_n(x_0)=y_0,\ L_n(x_1)=y_1,...,\ L_n(x_n)=y_n. \qquad (5.4)$$

Запишем это в виде:

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + ... + a_nx_0^n = y_0 \\ a_0 + a_1x_1 + a_2x_1^2 + ... + a_nx_1^n = y_1 \\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ a_0 + a_1x_n + a_2x_n^2 + ... + a_nx_n^n = y_n. \end{cases} \qquad (5.5)$$

Эта система однозначно разрешима, так как система функций 1, x, $x^2,...,x^n$ линейно независима в точках x_0, $x_1,...,x_n$. Однозначная разрешимость следует из того факта, что определитель этой системы (определитель Вандермонда)

$$\begin{vmatrix} 1 & x_0 & x_0^2 & ... & x_0^n \\ 1 & x_1 & x_1^2 & ... & x_1^n \\ ... & ... & ... & ... & ... \\ 1 & x_n & x_n^2 & ... & x_n^n \end{vmatrix} = \prod_{0\le j<i\le n} (x_i - x_j) \neq 0.$$

Без вывода приведем одну из форм записи интерполяционного многочлена Лагранжа

$$L_n(x)=y_0\cdot\frac{(x-x_1)...(x-x_n)}{(x_0-x_1)...(x_0-x_n)}+ y_1\cdot\frac{(x-x_0)(x-x_2)...(x-x_n)}{(x_1-x_0)(x_1-x_2)...(x_1-x_n)}+...+y_n\cdot\frac{(x-x_0)...(x-x_{n-1})}{(x_n-x_0)...(x_n-x_{n-1})}. \qquad (5.6)$$

Определение. Этот многочлен называется интерполяционным многочленом Лагранжа и сокращенно записывается в виде

$$L_n(x)=\sum_{i=0}^{n} y_i \frac{(x-x_0)(x-x_1)...(x-x_{i-1})(x-x_{i+1})...(x-x_n)}{(x_i-x_0)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_n)}. \qquad (5.7)$$

На практике часто пользуются линейной и квадратичной интерполяцией. В этом случае формула Лагранжа имеет вид

$L_1(x)= y_0\frac{(x-x_0)}{(x_0-x_1)}+y_1\frac{(x-x_0)}{(x_1-x_0)}$ – при линейной интерполяции;

$L_2(x)= y_0\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}+y_1\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}+y_2\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$ – при квадратичной интерполяции.

Рассмотрим теперь случай с равноотстоящими узлами. Тогда интерполяционная формула Лагранжа заметно упрощается. В этом случае шаг $h=x_{i+1}-x_i=\text{const}$. Введем в рассмотрение многочлен вида

$$Q_i(x)=\frac{(x-x_0)(x-x_1)...(x-x_{i-1})...(x-x_n)}{(x_i-x_0)(x_i-x_1)...(x_i-x_{i-1})...(x_i-x_n)}.$$

Введем обозначение $q=\frac{x-x_0}{h}$, отсюда следует, что

$$x-x_0= q\cdot h,$$
$$x-x_1= q\cdot h-h=h\cdot(q-1),$$
$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$
$$x-x_i = q\cdot h-i\cdot h=h\cdot(q-i),$$
$$x-x_n= q\cdot h-n\cdot h=h\cdot(q-n).$$

Тогда многочлен Q_i примет вид

$$Q_i(x)=\frac{q(q-1)\cdot[q-(i-1)]\cdot[q-(i+1)]...(q-n)\cdot h^n}{i\cdot h(i-1)\cdot h...h(-h)...[-(n-i)\cdot h]}.$$

Произведя простейшие преобразования, получим выражение вида:

$$Q_i(q)=\frac{q\cdot(q-1)\ldots(q-n)\cdot(-1)^{n-i}}{(q-i)\cdot i!(n-i)!}=(-1)^{n-i}\cdot\frac{C_n^i}{q-i}\cdot\frac{q(q-1)\ldots(q-n)}{n!},$$

где C_n^i – число сочетаний из n элементов по i $C_n^i=\frac{n!}{i!(n-i)!}$.

Тогда интерполяционный многочлен Лагранжа для равноотстоящих узлов имеет вид:

$$L_n(x)=\frac{q(q-1)\ldots(q-n)}{n!}\cdot\sum_{i=0}^{n}(-1)^{n-i}\cdot\frac{C_n^i}{q-i}\cdot y_i.$$

5.1.1 Оценка погрешности интерполяционного многочлена

Оценить погрешность интерполяционной формулы Лагранжа можно только тогда, когда известно аналитическое выражение интерполируемой функции, а точнее, если известно максимальное значение (n+1)-ой производной функции $f(x)$ на отрезке $[a,b]$. Пусть

$$|R_n(x)| = | f(x) - L_n(x)|,$$

где $R_n(x)$ –погрешность;

$f(x)$ − точное значение функции в точке x;

$L_n(x)$ − приближенное значение, полученное по полиному Лагранжа.

Если обозначить через $M_{n+1}=f^{(n+1)}(\xi)=\max\limits_{x\in[a,b]}\left|f^{(n+1)}(x)\right|$, где $\xi\in[a,b]$, причем $x_0=a$, $x_n=b$, то

$$|R_n(x)| \le \frac{M_{n+1}}{(n+1)!}|(\xi-x_0)(\xi-x_1)\ldots(\xi-x_n)|.$$

5.2 Интерполяционные полиномы Ньютона

5.2.1 Интерполяционный многочлен Ньютона для равноотстоящих узлов

Вычисление значений функции для значений аргумента, лежащих в начале таблицы удобно проводить, пользуясь первой интерполяционной формулой Ньютона. Для этого введем понятие конечной разности.

Определение. Конечной разностью перового порядка называется разность между значениями функции в соседних узлах интерполяции. Тогда конечные разности в точках $x_0, x_1, \ldots, x_{n-1}$

$$\Delta y_0 = y_1 - y_0 = f(x_1) - f(x_0) = \Delta f(x_0),$$

$$\Delta y_1 = y_2 - y_1 = f(x_2) - f(x_1) = \Delta f(x_1),$$

. . . . . .

$$\Delta y_{n-1} = y_n - y_{n-1} = f(x_n) - f(x_{n-1}) = \Delta f(x_{n-1}).$$

Конечная разность второго порядка имеет вид:

$$\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i,$$

. . . . . .

$$\Delta^n y_i = \Delta(\Delta^{n-1} y_i).$$

Рассмотрим некоторые свойства конечных разностей. Вторая конечная разность в точке x_i

$$\Delta^2 y_i = \left[f(x_{i+1} + \Delta x) - f(x_i + \Delta x) \right] - \left[f(x_{i+1}) - f(x_i) \right] =$$

$$= f(x_{i+2}) - 2 \cdot f(x_{i+1}) + f(x_i) = y_{i+2} - 2 \cdot y_{i+1} + y_i$$

Аналогично третья конечная разность

$$\Delta^3 y_i = y_{i+3} - 3 \cdot y_{i+2} + 3 \cdot y_{i+1} - y_i.$$

Общее выражение для конечной разности n-го порядка имеет вид

$$\Delta^n y_i = y_{n+i} - C_n^1 y_{n+i-1} + C_n^2 y_{n+i-2} - \ldots$$

$$\ldots + (-1)^m C_n^m y_{n+i-m} + \ldots (-1)^n y_i,$$

а вообще, конечная разность порядка m от конечной разности порядка n

$$\Delta^m(\Delta^n y)=\Delta^{m+n} y.$$

Конечные разности n-го порядка от многочлена степени n – есть величины постоянные, а конечные разности n+1-го порядка равны нулю.

Для вычисления значений функции в начале таблицы требуется построить интерполяционный многочлен степени n такой, что выполнены условия интерполяции

$$P_n(x_0)=y_0,\ldots,P_n(x_n)=y_n.$$

В силу единственности многочлена степени n, построенного по n+1 значениям функции $f(x)$ многочлен $P_n(x)$, в конечном счете, совпадает с многочленом Лагранжа. Найдем этот многочлен в виде:

$$P_n(x) = a_0 + a_1 \cdot (x - x_0) + a_2(x - x_0)(x - x_1) + \ldots + a_n(x - x_0)\ldots(x - x_{n-1}),$$

где $a_i(i=0,1,\ldots, n)$ – неизвестные коэффициенты. Для нахождения a_0 положим $x = x_0$. Тогда $P(x_0) = a_0$, отсюда $a_0=y_0$.

Для вычисления a_1 рассмотрим первую конечную разность для многочлена $P_n(x)$ в точке x.

$$\Delta P_n(x) = P_n(x + h) - P_n(x) = [a_0 + a_1(x - x_0 + h) + \ldots + a_n(x - x_0 + h) \cdot \ldots$$

$$\cdot\ldots(x - x_{n-1} + h)] - \left[a_0 + a_1(x - x_0) + \ldots + a_n(x - x_0)\ldots(x - x_{n-1})\right].$$

В результате преобразований получим

$$\Delta P_n(x)=h\cdot a_1+2ha_2(x-x_0)+\ldots+n\cdot ha_n(x-x_0)\ldots(x-x_{n-1}).$$

Вычислим первую конечную разность многочлена $P_n(x)$ в точке x_0

$\Delta P_n(x_0)=a_1\cdot h$, но $\Delta P_n(x_0)= f(x_1)- f(x_0)=y_1-y_0=\Delta y_0$,

откуда $a_1 = \frac{\Delta y_0}{h}$.

Чтобы определить коэффициент a_2, составим конечную разность второго порядка $\Delta^2 P_n(x)=\Delta P_n(x+h)-\Delta P_n(x)$. Отсюда после преобразования получим $a_2=\frac{\Delta^2 y_0}{2!h^2}$. Вычисляя конечные разности более высоких порядков и полагая $x=x_0$, придем к общей формуле для определения коэффициентов: $a_i=\frac{\Delta^i y_0}{i!h^i}$ $(i=0,1,2,\ldots,n)$.

Подставим значения a_i в многочлен, в результате получим первую интерполяционную формулу Ньютона:

$$P_n(x)=y_0+\frac{\Delta y_0}{1!h}\cdot(x-x_0)+\ldots+\frac{\Delta^n y_0}{n!h^n}\cdot(x-x_0)\ldots(x-x_{n-1}).$$

Первую интерполяционную формулу можно записать в том виде, в котором ее удобнее использовать для интерполирования в начале таблицы. Для этого введем переменную $q=(x-x_0)/h$, где h– шаг интерполирования. Тогда первая формула примет вид

$$P_n(x)=y_0+q\cdot\Delta y_0+\frac{q(q-1)}{2!}\cdot\Delta^2 y_0+\ldots+\frac{q(q-1)\ldots(q-n+1)}{n!}\cdot\Delta^n y_0.$$

5.2.2 Вторая интерполяционная формула Ньютона

Эта формула используется для интерполирования в конце таблицы. Построим интерполяционный многочлен вида

$$P_n(x)=a_0+a_1(x-x_n)+a_2(x-x_n)(x-x_{n-1})+\ldots+a_n(x-x_n)\ldots(x-x_1).$$

Неизвестные коэффициенты $a_0,a_1,\ldots,a_n$ подберем так, чтобы были выполнены равенства

$$P_n(x_0)=y_0,\ P_n(x_1)=y_1,\ldots,\ P_n(x_n)=y_n.$$

Для этого необходимо и достаточно, чтобы

$$\Delta^i P_n(x_{n-i})=\Delta^i y_{n-i}\ (i=0,1,\ldots,n).$$

В случае, если положить $x=x_n$, то сразу определяется коэффициент a_0

$$P_n(x) = y_n = a_0.$$

Из выражения для первой конечной разности найдем a_1:

$$\Delta P_n(x) = 1 \cdot ha_1 + 2 \cdot ha_2(x - x_{n-1}) + \ldots + n \cdot ha_n(x - x_{n-1})(x - x_{n-2})\ldots(x - x_1).$$

Отсюда, полагая $x=x_{n-1}$, получим $a_1 = \frac{\Delta y_{n-1}}{h}$. Из выражения для второй конечной разности найдем a_2: $a_2 = \frac{\Delta^2 y_{n-2}}{2!h^2}$. Общая формула для коэффициента a_i имеет вид $a_i = \frac{\Delta^i y_{n-i}}{i!h^i}$.

Подставим эти коэффициенты в формулу многочлена и получим вторую интерполяционную формулу Ньютона:

$$P_n(x) = y_n + \frac{\Delta y_{n-1}}{h}(x - x_n) + \ldots + \frac{\Delta^n y_0}{n!h^n}(x - x_n)\ldots(x - x_1).$$

На практике используют формулу Ньютона в другом виде. Положим $q=(x-x_n)/h$. Тогда

$$P_n(x) = y_n + q\Delta y_{n-1} + \frac{q(q+1)}{2!}\Delta^2 y_{n-2} + \ldots + \frac{q(q+1)\ldots(q+n-1)}{n!}\Delta^n y_0.$$

5.3 Интерполирование сплайнами

Многочлен Лагранжа или Ньютона на всем отрезке $[a,b]$ с использованием большого числа узлов интерполирования часто приводит к плохому приближению, что объясняется накоплением погрешностей в ходе вычислений. Кроме того, из-за расходимости процесса интерполирования увеличение числа узлов не обязательно приводит к повышению точности вычислений.

Поэтому построим такой вид приближения, который:

– позволяет получить функцию, совпадающую с табличной функцией в узлах;

– приближающая функция в узлах таблицы имеет непрерывную производную до нужного порядка;

В силу вышесказанного на практике весь отрезок $[a,b]$ разбивается на частичные интервалы и на каждом из них приближающая функция $f(x)$ заменяется многочленом невысокой степени. Такая интерполяция называется кусочно-полиномиальной интерполяцией.

Определение. Сплайн – функцией называют кусочно-полиномиальную функцию, определенную на отрезке $[a,b]$ и имеющую на этом отрезке некоторое число непрерывных производных.

Слово сплайн означает гибкую линейку, которую используют для проведения гладких кривых через определенное число точек на плоскости. Преимущество сплайнов – сходимость и устойчивость процесса вычисления. Рассмотрим частный случай (часто используемый на практике), когда сплайн определяется многочленом третьей степени.

5.3.1 Построение кубического сплайна

Пусть на отрезке $[a,b]$ в узлах сетки заданы значения некоторой функции $f(x)$, т.е. $a = x_0 < x_1 < x_2 ... < x_n = b$, $y_i = f(x_i) (i = 0,1,\ldots, n)$.

Сплайном, соответствующим этим узлам функции $f(x)$ называется функция $S(x)$, которая:

1) на каждом частичном отрезке является многочленом третьей степени;

2) функция $S(x)$ и ее две первые производные $S'(x), S''(x)$ непрерывны на $[a,b]$;

3) $S(x_i) = f(x_i)$.

На каждом частичном отрезке $[x_{i-1},x_i]$ будем искать сплайн $S(x) = S_i(x)$, где $S_i(x)$ многочлен третьей степени

$$S_i(x) = a_i + b_i(x - x_i) + \frac{c_i}{2} \cdot (x - x_i)^2 + \frac{d_i}{6} \cdot (x - x_i)^3. \qquad (5.8)$$

То есть для $x \in [x_{i-1},x_i]$ нужно построить такую функцию $S_i(x)$, где a_i, b_i, c_i, d_i подлежат определению. Для всего отрезка интерполирова-

ния $[a,b]$, таким образом, необходимо определить $4 \cdot n$ неизвестных коэффициента.

$$S'(x) = b_i + c_i(x - x_i) + \frac{d_i}{2} \cdot (x - x_i)^2,$$

$$S''(x) = c_i + d_i(x - x_i),$$

$$S_i(x) = a_i = y_i.$$

Доопределим $a_0 = f(x_0) = y_0$. Требование непрерывности функции $S(x)$ приводит к условия $S_i(x_i)=S_{i+1}(x_i)$, $(i=0,1,\ldots,n\text{-}1)$.

Отсюда из (5.8) получаем следующие уравнения:

$$a_i = a_{i+1} + b_{i+1}(x_i - x_{i+1}) + \frac{c_{i+1}}{2}(x_i - x_{i+1})^2 + \frac{d_{i+1}}{6}(x_i - x_{i+1})^3 \ (\text{i}= 1,2,\ldots,n\text{-}1).$$

Введем шаг интерполирования $h_i = x_i - x_{i-1}$.

Тогда последнее равенство можно переписать в виде

$$h_i \cdot b_i - \frac{h_i^2}{2} \cdot c_i + \frac{h_i^3}{6} \cdot d_i = f_i - f_{i-1} (i= 1,2,\ldots,n).$$

Из непрерывности первой производной следует

$$h_i \cdot c_i - \frac{h_i^2}{2} \cdot d_i = b_i - b_{i-1} \ (i=2,3,\ldots,n),$$

а из непрерывности второй производной

$$h_i d_i = c_i - c_{i-1} \ (i=2,3,\ldots,n).$$

Объединив все три вида уравнений, получим систему из $3n$-2 уравнений относительно $3n$ неизвестных b_i, c_i, d_i. Два недостающих уравнения получим, задав граничные условия для функции $S(x)$. Для этого воспользуемся граничными условиями для сплайн-функции в виде $S''(a) = S''(b) = 0$ (концы гибкой линейки свободны).

Тогда получим систему уравнений

$$\begin{cases} h_i \cdot d_i = c_i - c_{i-1}, c_0 = c_n = 0, (i = 1,2,\ldots,n) \\ h_i \cdot c_i - \dfrac{h_i^2}{2} \cdot d_i = b_i - b_{i-1}, (i = 2,3,\ldots,n) \\ h_i \cdot b_i - \dfrac{h_i^2}{2} \cdot c_i + \dfrac{h_i^3}{6} \cdot d_i = f_i - f_{i-1}, (i = 1,2,\ldots,n). \end{cases} \quad (5.9)$$

Решая систему методом подстановки (исключаем из (5.9) неизвестные b_i, d_i), получим систему:

$$\begin{cases} h_i c_{i-1} + 2(h_i + h_{i+1}) \cdot c_i + h_{i+1} c_{i+1} = 6 \cdot \left(\dfrac{y_{i+1} - y_i}{h_{i+1}} - \dfrac{y_i - y_{i-1}}{h_i} \right), \ (i = 1,2,\ldots,n) \\ c_0 = c_n = 0 \end{cases} \quad (5.10)$$

Система (5.10) имеет трехдиагональную матрицу. Эта система может быть решена методом прогонки или Гаусса. Метод прогонки рассматривается в пункте 6.7.2.

После решения системы коэффициенты сплайна d_i, b_i определим через коэффициенты c_i с помощью явных формул

$$d_i = \frac{c_i - c_{i-1}}{h_i},$$

$$b_i = \frac{h_i}{2} \cdot c_i - \frac{h_i^2}{6} \cdot d_i + \frac{y_i - y_{i-1}}{h_i} \quad (i = 1,2,\ldots,n).$$

Существуют специальные виды записи сплайнов на каждом из промежутков $[x_i, x_{i+1}]$ [9], которые позволяют уменьшить число неизвестных коэффициентов сплайна. Вводятся обозначения

$S'(x_i) = m_i$, $i = 0,1,\ldots,n$,

$h_i = x_{i+1} - x_i$ и $t = (x - x_i)/h_i$.

На отрезке $[x_i, x_{i+1}]$ кубический сплайн записывается в виде

$$S(x) = y_i (1-t)^2 (1+2t) + y_{i+1} t^2 (3-2t) + m_i h_i t (1-t)^2 - m_{i+1} t^2 (1+t) h_i.$$

Кубический сплайн, записанный в таком виде, на каждом из промежутков $[x_i, x_{i+1}]$ непрерывен вместе со своей первой производной на $[a,b]$.

Выберем m_i таким образом, чтобы и вторая производная была непрерывна во всех внутренних узлах. Отсюда получим систему уравнений:

$$\lambda_i m_{i-1} + 2m_i + \mu_i m_{i+1} = 3\left(\mu_i \frac{y_{i+1} - y_i}{h_i} + \lambda_i \frac{y_i - y_{i-1}}{h_{i-1}}\right),$$

где $\mu_i = \frac{h_{i-1}}{h_{i-1} + h_i}$, $\lambda_i = 1 - \mu_i = \frac{h_i}{h_{i-1} + h_i}$, $i = 1,2,...,n-1$.

К этим уравнениям добавим уравнения, полученные из граничных условий

$$2m_0 + m_1 = 3\frac{y_1 - y_0}{h_0}, \quad m_{n-1} + 2m_n = 3\frac{y_n - y_{n-1}}{h_{n-1}}.$$

В результате получаем аналогичную систему с трехдиагональной матрицей. Решаем систему линейных уравнений относительно коэффициентов m_i методом пргонки.

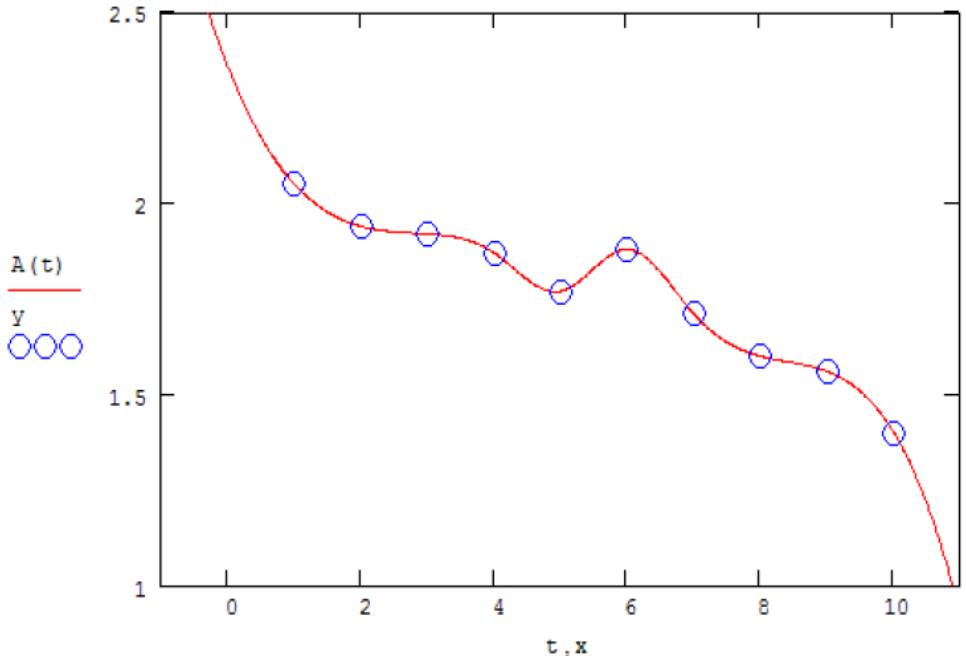
Пример построения кубического сплайна в пакете Mathcad.

Кубическая сплайн-интерполяция

$x := (1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10)^T$

$y := (2.05 \quad 1.94 \quad 1.92 \quad 1.87 \quad 1.77 \quad 1.88 \quad 1.71 \quad 1.60 \quad 1.56 \quad 1.40)^T$

```
s := cspline(x, y)
A(t) := interp(s, x, y, t)
```



A(1) = 2.05	A(3.5) = 1.909	A(6) = 1.88	A(8.5) = 1.583
A(1.5) = 1.977	A(4) = 1.87	A(6.5) = 1.819	A(9) = 1.56
A(2) = 1.94	A(4.5) = 1.8	A(7) = 1.71	A(9.5) = 1.507
A(2.5) = 1.925	A(5) = 1.77	A(7.5) = 1.636	A(10) = 1.4
A(3) = 1.92	A(5.5) = 1.831	A(8) = 1.6	

5.3.2 Сходимость процесса интерполирования кубическими сплайнами

Доказывается, что при неограниченном увеличении числа узлов на одном и том же отрезке $[a,b]$ $S(x) \to f(x)$. Оценка погрешности интерполяции $R(x) = f(x) - S(x)$ зависит от выбора сетки и степени гладкости функции $f(x)$.

При равномерной сетке

$$x_i = a + i \cdot h \, (i = 0,1,\ldots,n)$$

$$|f(x) - S_h(x)| \le \frac{M_4 \cdot h^4}{8},$$

где $M_4 = \max\limits_{[a,b]} |f^{IV}(x)|$.

Другие постановки задачи интерполирования функций.

1. Если функция периодическая, то используется тригонометрическая интерполяция с периодом l, которая строится с помощью тригонометрического многочлена

$$T_n(x) = a_0 + \sum_{k=1}^{n}(a_k \cos\frac{\pi kx}{l} + b_k \sin\frac{\pi kx}{l}),$$

коэффициенты которого находятся из системы уравнений

$$T_n(x_i) = f(x_i) \ (i = 1,2,\ldots, 2n+1).$$

2. Выделяют приближение функций рациональными, дробно – рациональными и другими функциями. В данном пособии эти вопросы не рассматриваются.

5.4 Аппроксимация функций методом наименьших квадратов

К такой задаче приходят при статистической обработке экспериментальных данных с помощью регрессионного анализа. Пусть в результате исследования некоторой величины x значениям $x_1, x_2, x_3, \ldots, x_n$ поставлены в соответствие значения $y_1, y_2, y_3, \ldots, y_n$ некоторой величины y.

Требуется подобрать вид аппроксимирующей зависимости $y=f(x)$, связывающей переменные x и y. Здесь могут иметь место следующие случаи. Во-первых: значения функции $f(x)$ могут быть заданы в достаточно большом количестве узлов; во-вторых: значения таблично заданной функции отягощены погрешностями. Тогда проводить приближения функции с помощью интерполяционного многочлена нецелесообразно, т.к.

- число узлов велико и пришлось бы строить несколько интерполяционных многочленов;
- построив интерполяционные многочлены, мы повторили бы те же самые ошибки, которые присущи таблице.

Будем искать приближающую функцию из следующих соображений:

1) приближающая функция не проходит через узлы таблицы и не повторяет ошибки табличной функции;

2) чтобы сумма квадратов отклонений приближающей функции от таблично заданной в узлах таблицы была минимальной.

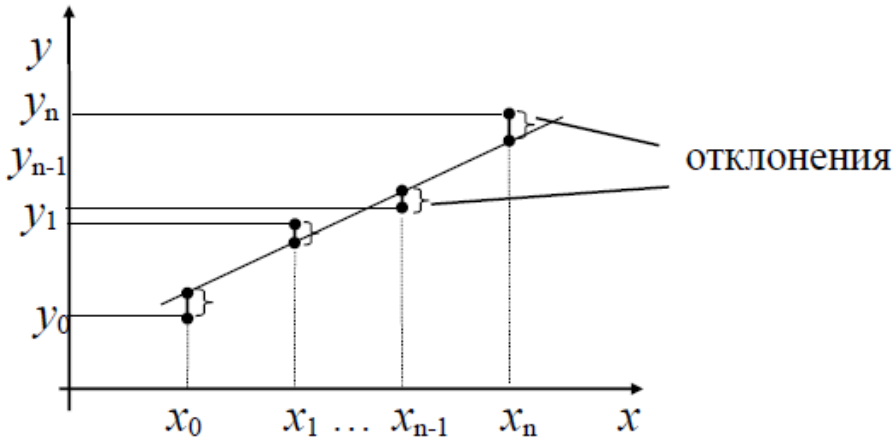


Рисунок 6 – Графическое изображение отклонений приближающей функции от таблично заданной

Рассмотрим линейную задачу наименьших квадратов.

Определение. Уровень погрешности, допускаемый при снятии характеристики измеряемой величины, называется шумом таблицы.

Пусть функция $y=f(x)$ задана таблицей приближенных значений $y_i \approx f(x_i)$, $i=0,1,\ldots,n$, полученных с ошибками $\varepsilon_i = y_i^0 - y_i$, где $y_i^0 = f(x_i)$.

Пусть даны функции $\varphi_0(x), \varphi_1(x), \ldots, \varphi_m(x)$, назовем их базисными функциями.

Будем искать приближающую (аппроксимирующую) функцию в виде линейной комбинации базисных функций

$$y = \Phi_m(x) \equiv c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_m\varphi_m(x). \qquad (5.11)$$

Такая аппроксимация называется линейной, а $\Phi_m(x)$ – обобщенным многочленом.

Будем определять коэффициенты обобщенного многочлена $c_0, \ldots, c_m$ используя критерий метода наименьших квадратов. Согласно этому критерию вычислим сумму квадратов отклонений таблично заданной функции от искомого многочлена в узлах:

$$\delta_m = \sum_{i=0}^{n} (y_i - \Phi_m(x_i))^2 = \sum_{i=0}^{n} (y_i - c_0\varphi_0(x_i) - \ldots - c_m\varphi_m(x_i))^2. \qquad (5.12)$$

Выражение для δ_m можно рассматривать как функцию от неизвестных $c_0, \ldots, c_m$. Нас интересует, при каких значениях $c_0, \ldots, c_m$, значение δ_m будет минимально.

Для этого воспользуемся условием существования экстремума функции, а именно, найдем частные производные от δ_m по всем переменным $c_0, \ldots, c_m$ и приравняем их к нулю. Получим систему вида:

$$\begin{cases} \dfrac{\partial \delta_m}{\partial c_0} = -2\sum\limits_{i=0}^{n}(y_i - c_0\varphi_0(x_i) - \ldots - c_m\varphi_m(x_i))\cdot\varphi_0(x_i) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial \delta_m}{\partial c_m} = -2\sum\limits_{i=0}^{n}(y_i - c_0\varphi_0(x_i) - \ldots - c_m\varphi_m(x_i))\cdot\varphi_m(x_i) = 0 \end{cases} . \quad (5.13)$$

Система (5.13) – система линейных уравнений относительно $c_0, \ldots, c_m$.

Чтобы систему (5.13) записать компактно, ведем определение.

Определение. Скалярным произведением функций f на g на множестве точек $x_0, \ldots, x_n$ называется выражение

$$(f, g) = \sum_{i=0}^{n} f(x_i)\cdot g(x_i)\cdot$$

Тогда систему (5.13) можно записать в виде:

$$\begin{cases} c_0(\varphi_0,\varphi_0) + c_1(\varphi_0,\varphi_1) + \ldots + c_m(\varphi_0,\varphi_m) = (\varphi_0, y) \\ c_0(\varphi_1,\varphi_0) + c_1(\varphi_1,\varphi_1) + \ldots + c_m(\varphi_1,\varphi_m) = (\varphi_1, y) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ c_0(\varphi_m,\varphi_0) + c_1(\varphi_m,\varphi_1) + \ldots + c_m(\varphi_m,\varphi_m) = (\varphi_m, y) \end{cases} . \quad (5.13а)$$

Системы (5.13) или (5.13а) будем называть нормальной системой уравнений.

Решив ее, мы найдем коэффициенты $c_0, \ldots, c_m$ и следовательно, найдем вид аппроксимирующего многочлена. Напомним, что это возможно, если базисные функции линейно независимы, а все узлы различны.

Осталось определить степень многочлена m. Прямому вычислению поддаются только значения среднеквадратичного отклонения δ_m, анализируя которые будем выбирать степень многочлена.

Алгоритм выбора степени многочлена m.

В случае, когда $m=n$ мы получим интерполяционный многочлен, поэтому выберем $m<<n$. Так же необходимо задать числа ε_1 и ε_2, учитывая следующее:

1) $\varepsilon_1 >0$ и $\varepsilon_2>0$ должны быть такими, чтобы δ_m находилось между ними;

2) первоначально m выбирают произвольно, но учитывая условие, что $m<<n$;

3) выбрав m, строят системы (5.13) и (5.13а), решив которые, находят $c_0, \ldots, c_m$;

4) используя найденные коэффициенты, вычисляется δ_m и проверяется, попало ли оно в промежуток между ε_1 и ε_2. Если попало, то степень многочлена выбрана правильно, иначе

а) если $\delta_m > \varepsilon_1$, то степень необходимо уменьшить хотя бы на единицу;

б) если $\delta_m < \varepsilon_2$, то степень необходимо увеличить хотя бы на единицу.

5) затем строим приближающую функцию .

Очень часто для приближения по методу наименьших квадратов используются алгебраические многочлены степени $m \le n$, т.е. $\varphi_k(x) = x^k$. Тогда нормальная система (5.13) принимает следующий вид:

$$\sum_{j=0}^{m}\Big(\sum_{i=0}^{n} x_i^{j+k}\Big)c_j = \sum_{i=0}^{n} y_i\, x_i^k \ , \ (k= 0,1,\ldots,m). \qquad (5.14)$$

Запишем систему (5.14) в развернутом виде в двух наиболее простых случаях $m=1$ и $m=2$.

В случае многочлена первой степени $P_1(x)=c_0+c_1x$, нормальная система имеет вид

$$\begin{cases} (n+1)c_0 + \Big(\sum\limits_{i=0}^{n} x_i\Big)c_1 = \sum\limits_{i=0}^{n} y_i \\ \Big(\sum\limits_{i=0}^{n} x_i\Big)c_0 + \Big(\sum\limits_{i=0}^{n} x_i^2\Big)c_1 = \sum\limits_{i=0}^{n} y_i x_i. \end{cases} \qquad (5.15)$$

Для многочлена второй степени $P_2(x)=c_0+c_1x+c_2x^2$, нормальная система имеет вид

$$\begin{cases} (n+1)c_0 + (\sum_{i=0}^{n} x_i)c_1 + (\sum_{i=0}^{n} x_i^2)c_2 = \sum_{i=0}^{n} y_i \\ (\sum_{i=0}^{n} x_i)c_0 + (\sum_{i=0}^{n} x_i^2)c_1 + (\sum_{i=0}^{n} x_i^3)c_2 = \sum_{i=0}^{n} y_i x_i \\ (\sum_{i=0}^{n} x_i^2)c_0 + (\sum_{i=0}^{n} x_i^3)c_1 + (\sum_{i=0}^{n} x_i^4)c_2 = \sum_{i=0}^{n} y_i x_i^2 \end{cases} . \qquad (5.16)$$

6 Численные методы решения задачи Коши для обыкновенных дифференциальных уравнений и систем дифференциальных уравнений

Будем рассматривать задачу Коши для системы обыкновенных дифференциальных уравнений(ОДУ).Запишем систему в векторной форме

$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{f}(t,\boldsymbol{u}), \qquad (6.1)$$

где: $\boldsymbol{u}$ – искомая вектор-функция; t– независимая переменная; $\boldsymbol{u}(t) = (u_1(t),...,u_m(t))$; $\boldsymbol{f}(t,\boldsymbol{u}) = (f_1,...,f_m)$, m– порядок системы; $u_1(t),...,u_m(t)$ – координаты; $t \geq 0$; $\boldsymbol{u}(0) = \boldsymbol{u}^0$.

Запишем систему (6.1) в развернутом виде

$$\frac{du_i}{dt} = f_i(t,u_1,...,u_m), \qquad (6.2)$$

где: $i=1,\ldots,m$; $u_i(0) = u_i^0$.

В случае i=1 – это будет ОДУ 1-го порядка, а при i=2 – система из двух уравнений первого порядка.

В случае i=1 решение задачи Коши предполагает нахождение интегральной кривой, проходящей через заданную точку и удовлетворяющую заданному начальному условию.

Задача состоит в том, чтобы найти искомую вектор-функцию $\boldsymbol{u}$, удовлетворяющую (6.1) и заданным начальным условиям.

Известны условия, гарантирующие существование и единственность решения (6.1) или (6.2).

Предположим, что функции f_i(i=1, ..., m) непрерывны по всем аргументам в некоторой замкнутой области $D=\{t \leq a, u_i - u_i^0 \leq b\}$, где a,b– известные константы.

Из непрерывности функций следует их ограниченность, т.е. функции f_i сверху ограничены некоторой константой M: $|f_i| < M$ (где $M \geq 0$) всюду в области D и пусть в области D функции f_i удовлетворяют условию Липшица по аргументам $u_1,...,u_m$. Это значит, что

$$|f_i(t,u_1',....,u_m') - f_i(t,u_1'',...,u_m'')| \leq L(|u_1' - u_1''| + ... + |u_m' - u_m''|)$$

для любых двух точек $(t,u_1',....,u_m')$ и $(t,u_1'',...,u_m'')$ из области D. Тогда существует единственное решение задачи (6.1)

$$u_1 = u_1(t),....,u_m = u_m(t) \text{ ,определенное при } t \le T = min\{ a,b/M \} \quad (6.3)$$

и принимающее при $t=0$ заданные начальные значения.

Существует два класса методов для решения задачи (6.1):

1) семейство одношаговых методов(Рунге-Кутта);

2) семейство многошаговых(m-шаговых) методов.

Сначала рассмотрим одношаговые методы. Для простоты возьмем одно уравнение

$$\frac{du}{dt} = f(t,u), \quad (6.4)$$

где: $u(0) = u_0$; $t>0$.

По оси t введем равномерную сетку с шагом $\tau>0$, т.е. рассмотрим систему точек $\omega_\tau = \{t_n = n \cdot \tau, n = 0,1,2,...\}$. Обозначим через $u(t)$ точное решение (6.4), а через $y_n = y(t_n)$ приближенные значения функций u в заданной системе точек.

Приближенное решение является сеточной функцией, т.е. определено только в точках сетки ω_τ.

6.1 Семейство одношаговых методов решения задачи Коши

6.1.1 Метод Эйлера (частный случай метода Рунге-Кутта)

Уравнение (6.4) заменяется разностным уравнением

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n),\ n=0,1,2,\ldots,.\ y_0 = u_0$$

В окончательной форме значения y_{n+1} можно определить по явной формуле

$$y_{n+1} = y_n + \tau \cdot f(t_n, y_n). \quad (6.5)$$

Вследствие систематического накопления ошибок метод используется редко или используется только для оценки вида интегральной кривой.

Определение 1. Метод сходится к точному решению в некоторой точке t , если $|y_n - u(t_n)| \to 0$, при $\tau \to 0$, $t_n = t$.

Метод сходится на интервале $(0,t]$, если он сходится в каждой точке этого интервала.

Определение 2. Метод имеет p-й порядок точности, если существует такое число $p>0$, для которого $|y_n - u(t_n)| = O(\tau^p)$, при $\tau \to 0$, где: τ – шаг интегрирования; O – малая величина порядка τ^p.

Так как $\frac{u_{n+1} - u_n}{\tau} = u'(t_n) + O(\tau)$, то метод Эйлера имеет первый порядок точности.

Порядок точности разностного метода совпадает с порядком аппроксимации исходного дифференциального уравнения.

6.1.2 Методы Рунге-Кутта

Метод Рунге-Кутта второго порядка точности

Отличительная особенность методов Рунге-Кутта от метода (6.5) заключается в том, что значение правой части уравнения вычисляется не только в точках сетки, но и также в середине отрезков(промежуточных точках).

Предположим, что приближенное значение y_n решения задачи в точке $t=t_n$ уже известно. Для нахождения y_{n+1} поступают следующим образом:

1) используют схему Эйлера в таком виде

$$\frac{y_{n+1/2} - y_n}{0{,}5\tau} = f(t_n, y_n) \qquad (6.6)$$

и отсюда вычисляют промежуточное значение $y_{n+1/2}$;

2) воспользуемся разностным уравнением вида

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n + 0{,}5\tau, y_{n+1/2}), \qquad (6.7)$$

откуда найдем значение y_{n+1}. Далее подставим значение $y_{n+1/2} = y_n + 0{,}5\tau\, f_n$ в уравнение (6.7). Тогда получим разностное уравнение

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n + 0{,}5\cdot\tau,\, y_n + 0{,}5\cdot\tau\cdot f_n), \qquad (6.8)$$

где $f_n = f(t_n, y_n)$.

Можно показать, что метод (6.8) имеет второй порядок точности, т.е. $|\, y_n - u(t)\,| = O(\tau^2)$.

Реализация метода (6.8) в виде двух этапов (6.6),(6.7) называется методом предиктор–корректор (прогноза и коррекции) в том смысле, что на первом этапе (6.6) приближенное значение предсказывается с невысокой точностью $O(\tau)$, а на втором этапе (6.7) это предсказанное значение исправляется и результирующая погрешность имеет второй порядок $O(\tau^2)$.

Будем рассматривать явные методы. Задаем числовые коэффициенты a_i, b_{ij}, i=2,...,m; j=1,2,...,m-1 и σ_i, i=1,2,...,m. Последовательно вычисляем функции

$$k_1 = f(t_n, y_n);$$
$$k_2 = f(t_n + a_2\tau,\, y_n + b_{21}\tau\cdot k_1);$$
$$k_3 = f(t_n + a_3\tau,\, y_n + b_{31}\tau\cdot k_1 + b_{32}\tau\cdot k_2);$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$k_n = f(t_n + a_n\tau,\, y_n + b_{m1}\tau\cdot k_1 + \ldots + b_{mm-1}\tau\cdot k_{m-1}).$$

Затем из формулы $\frac{y_{n+1} - y_n}{\tau} = \sum_{i=1}^{m} \sigma_i k_i$ находим новое значение y_{n+1}=$y(t_{n+1})$. Здесь a_i, b_{ij}, σ_i–числовые параметры, которые определяются или выбираются из соображений точности вычислений. Чтобы это уравнение аппроксимировало уравнение (6.4), необходимо потребовать выполнения условия $\sum_{i=1}^{m} \sigma_i = 1$. При m=1 получается метод Эйлера, при m=2 получаем семейство методов

$$y_{n+1} = y_n + \tau(\sigma_1 k_1 + \sigma_2 k_2), \qquad (6.9)$$

где: $k_1 = f(t_n, y_n)$; $k_2 = f(t_n + a_2\tau, y_n + b_{21}\tau k_1)$; $y_0 = u_0$.

Семейство определяет явные методы Рунге-Кутта. Подставив нужные σ_1 и σ_2, получаем окончательную формулу. Точность этих методов совпадает с точностью аппроксимирующего метода и равна $O(\tau^2)$.

Невязкой, или погрешностью аппроксимации метода (6.9) называется величина

$$\delta_n = -\frac{u_{n+1} - u_n}{\tau} + \sigma_1 f(t_n, u_n) + \sigma_2 f(t_n + a_2\tau, u_n + b_{21}\tau\, f(t_n, u_n)),$$

полученная заменой в (6.9) приближенного решения точным решением.

При $\sigma_1+\sigma_2=1$ получим первый порядок точности. Если же потребовать дополнительно $\sigma_2 b_{21} = \sigma_2 a_2 = 0{,}5$, то получим методы второго порядка точности вида

$$\frac{y_{n+1} - y_n}{\tau} = (1-\sigma) f(t_n, y_n) + \sigma f(t_n + a\tau, y_n + a\tau\, f(t_n, y_n)) \text{ при } \sigma \cdot a = 0{,}5.$$

Приведем один из методов Рунге-Кутта третьего порядка точности

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{6}(k_1 + 4k_2 + k_3),$$

где: $k_1 = f(t_n, y_n)$; $k_2 = f(t_n + \frac{\tau}{2}, y_n + \frac{\tau}{2}k_1)$; $k_3 = f(t_n + \tau, y_n - \tau k_1 + 2\tau k_2)$.

Метод Рунге-Кутта 4-го порядка точности

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{6}(k_1 + 2\cdot k_2 + 2\cdot k_3 + k_4),$$

где: $k_1 = f(t_n, y_n)$; $k_2 = f(t_n + \frac{\tau}{2}, y_n + \frac{\tau}{2}\cdot k_1)$;

$k_3 = f(t_n + \frac{\tau}{2}, y_n + \frac{\tau}{2}\cdot k_2)$; $k_4 = f(t_n + \tau, y_n + \tau \cdot k_3)$.

Методы Рунге-Кутта сходятся и порядок их точности совпадает с порядком аппроксимации разностным отношением.

Теорема. *Пусть правая часть уравнения* (6.4) $f(t,u)$ *удовлетворяет условию Липшица по аргументу и с константой L, и пусть* δ_n *– невязка метода Рунге-Кутта. Тогда для погрешности метода при* $n\cdot\tau \le T$ *справедлива оценка*

$$\left| y_n - u(t_n) \right| \le Te^{\alpha T} \max \left| \delta_n \right|,$$

где: $\alpha = \sigma Lm(1+Lb\tau)^{m-1};\ \sigma = \max_i \left| \sigma_i \right|;\ b = \max_{i,j} \left| b_{ij} \right|.$

На практике обычно пользуются правилом Рунге. Для этого сначала проводят вычисления с шагом τ , затем – $\tau/2$. Если y_n^{τ} – решение при шаге τ, а $y_{2n}^{\tau/2}$ – при шаге $\tau/2$, p – порядок точности метода, то справедлива оценка.

$$\left| y_{2n}^{\tau/2} - u(t_n) \right| \le \frac{\left| y_{2n}^{\tau/2} - y_n^{\tau} \right|}{2^p - 1}$$

Тогда за оценку погрешности для метода четвертого порядка точности при шаге $\tau/2$ принимают величину

$$\max_i \left| \frac{y_n^{\tau} - y_{2n}^{\tau/2}}{15} \right|.$$

6.2 Многошаговые разностные методы решения задачи Коши для обыкновенных дифференциальных уравнений

Рассмотрим задачу Коши для обыкновенного дифференциального уравнения

$$\frac{du}{dt} = f(t,u), \qquad (6.10)$$

где $u(0) = u_0$.

Для решения задачи Коши для уравнения (6.10) при $t>0$ введем равномерную сетку с постоянным шагом τ

$$\omega_\tau = \{ t_n = n\cdot\tau,\ n = 0,1,\ldots \}.$$

Введем понятие линейного m шагового разностного метода для решения задачи (6.10). Линейным m-шаговым разностным методом называется система разностных уравнений

$$\frac{a_0 y_n + a_1 y_{n-1} + ... + a_m y_{n-m}}{\tau} = b_0 f_n + b_1 f_{n-1} + ... + b_m f_{n-m}, \qquad (6.11)$$

где: $n=m,m+1...$; a_k, b_k – числовые коэффициенты, не зависящие от n; $k=0,1,...,m$, причем $a_0 \neq 0$.

Систему (6.11) будем рассматривать как рекуррентные соотношения, выражающие новые значения $y_n = y(t_n)$ через ранее найденные значения $y_{n-1}, y_{n-2},..., y_{n-m}$, причем расчет начинают с индекса $n=m$, т.е. с уравнения

$$\frac{a_0 y_m + a_1 y_{m-1} + ... + a_m y_0}{\tau} = b_0 f_m + b_1 f_{m-1} + ... + b_m f_0.$$

Отсюда следует, что для начала расчета по формулам (6.11) надо знать m предыдущих значений функции y, причем $y_0 = u_0$ определяется исходной задачей (6.10). Эти предыдущие m значений могут быть найдены одним из одношаговых методов Рунге-Кутта.

Отличие от одношаговых методов состоит в том, что по формулам (6.11) расчет ведется только в точках сетки.

Определение. Метод (6.11) называется *явным*, если коэффициент $b_0=0$. Тогда значение y_n легко выражается через $y_{n-1}, y_{n-2},..., y_{n-m}$. В противном случае метод называется *неявным*, и для нахождения y_n придется решать нелинейное уравнение вида

$$\frac{a_0}{\tau} y_n - b_0 f(t_n, y_n) = \sum_{k=1}^{m} \left(b_k t_{n-k} - \frac{a_k}{\tau} y_{n-k} \right). \qquad (6.12)$$

Обычно это уравнение решают методом Ньютона при начальном значении $y_n^{(0)} = y_{n-1}$. Коэффициенты уравнения (6.11) определены с точностью до множителя, тогда, чтобы устранить этот произвол, вводят условие $\sum_{k=0}^{m} b_k = 1$, с тем условием, что правая часть (6.11) аппроксимирует правую часть дифференциального уравнения (6.10).

На практике используют частный случай методов (6.11), т.н. методы Адамса, когда производная $u'(t)$ аппроксимируется разностным отношением, включающим две соседние точки t_n и t_{n-1}. Тогда $a_0 = -a_1 = 1$; $a_k=0$, $k=2,...,m$ и

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^{m} b_k \, f_{n-k} \cdot \qquad (6.13)$$

Это и есть методы Адамса. При b_0=0 метод будет явным, в противном случае – неявным.

6.2.1 Задача подбора числовых коэффициентов $a_к$, $b_к$

Выясним, как влияют коэффициенты a_k, b_k на погрешность аппроксимации уравнения (6.11), на устойчивость и сходимость.

Определение. *Невязкой*, или погрешностью аппроксимации методов (6.11) называется функция

$$r_n = -\sum_{k=0}^{m} \frac{a_k}{\tau} u_{n-k} + \sum_{k=0}^{m} b_k f(t_{n-k}, u_{n-k}), \qquad (6.14)$$

которая получается в результате подстановки точного решения $u(t)$ дифференциального уравнения (6.10) в разностное уравнение (6.11) .

Если разложить функции $u_{n-k} = u(t_n - k \cdot \tau)$ в ряд Тейлора в точках $t = t_n$ равномерной сетки, окончательно получим функцию

$$r_n = -(\sum_{k=0}^{m} \frac{a_k}{\tau}) u(t_n) + \sum_{l=1}^{p} (\sum_{k=0}^{m} (-k\tau)^{l-1} (a_k \frac{k}{l} + b_k)) \frac{u^{(l)}(t_n)}{(l-1)!} + O(\tau^p) \cdot \qquad (6.15)$$

Из вида функции r_n следует, что порядок аппроксимации будет равен p, если выполнены условия

$$\begin{gathered} \sum_{k=0}^{m} a_k = 0; \\ \sum_{k=0}^{m} k^{l-1} (k a_k + l b_k) = 0; \\ \sum_{k=0}^{m} b_k = 1, \end{gathered} \qquad (6.16)$$

где l=1,...,p.

Условия (6.16) представляют собой систему линейных уравнений из p+2 уравнений относительно неизвестных $a_0, \ldots, a_m$, $b_0, \ldots, b_m$. Их количество равно 2(m+1). Решив систему (6.16), получаем неизвестные числовые коэффициенты. Для неявных m-шаговых методов

наивысшим достижимым порядком аппроксимации будет $p=2m$, а для явных – $p=2m-1$.

Запишем систему (6.16) для методов Адамса

$$\begin{aligned} l \cdot \sum_{k=1}^{m} k^{l-1} b_k &= 1; \\ b_0 &= 1 - \sum_{k=1}^{m} b_k, \end{aligned} \tag{6.17}$$

где $l=2,\ldots,p$. Отсюда видно, что наивысший порядок аппроксимации для неявного m-шагового метода Адамса $p=m+1$, для явного ($b_0=0$) – $p=m$.

6.2.2 Устойчивость и сходимость многошаговых разностных методов

Наряду с системами уравнений (6.11) будем рассматривать т.н. однородные разностные уравнения вида

$$a_0 V_n + a_1 V_{n-1} + \ldots + a_m V_{n-m} = 0, \tag{6.18}$$

где $n=m,m+1,\ldots$.

Будем искать его решение в виде функции

$$V_n = q^n,$$

где q – число, подлежащее определению. Подставив V_n в (6.18) получаем уравнение для нахождения q

$$a_0 q^m + a_1 q^{m-1} + \ldots + a_{m-1} q + a_m = 0. \tag{6.19}$$

Уравнение (6.19) принято называть характеристическим уравнением для разностных методов (6.11). Говорят, что разностный метод (6.11) удовлетворяет условию корней, если все корни уравнения (6.19) $q_1, \ldots, q_m$ лежат внутри или на границе единичного круга комплексной плоскости, причем на границе нет кратных корней.

Разностный метод (6.11), удовлетворяющий условию корней, называется устойчивым методом.

Теорема. *Пусть разностный метод* (6.11) *удовлетворяет условию корней и выполнено условие* $\left|\frac{\partial f(t,u)}{\partial u}\right| \le L$ *при* $0 \le t \le T$. *Тогда при*

$m \cdot \tau \le t_n = n \cdot \tau \le T$, $n \ge m$ и достаточно малых τ будет выполнена оценка

$$|y_n - u(t_n)| \le M\left(\max_{0 \le j \le m-1} |y_j - u(t_j)| + \max_{0 \le k \le n-m} |r_k|\right), \qquad (6.20)$$

где: r_k– погрешность аппроксимации; $y_j - u(t_j)$– погрешность в задании начальных условий; M –константа , зависящая от L, T и не зависящая от n.

Методы Адамса удовлетворяют условию корней, т.к. для них a_0=-a_1=1, следовательно, q=q_1=1.

6.2.3 Примеры *m*-шаговых разностных методов Адамса

Явные методы. При m=1 порядок точности p=1. Тогда метод описывается формулой

$$\frac{y_n - y_{n-1}}{\tau} = f_{n-1}.$$

В этом случае получаем метод Эйлера. При m=2 порядок точности p=2. Тогда метод описывается формулой

$$\frac{y_n - y_{n-1}}{\tau} = \frac{3}{2} f_{n-1} - \frac{1}{2} f_{n-2}.$$

При m=3 порядок точности p=3. Тогда метод описывается формулой

$$\frac{y_n - y_{n-1}}{\tau} = \frac{1}{12}(23 f_{n-1} - 16 f_{n-2} + 5 f_{n-3}).$$

При m=4 порядок точности p=4. Метод описывается формулой

$$\frac{y_n - y_{n-1}}{\tau} = \frac{1}{24}(55 f_{n-1} - 59 f_{n-2} + 37 f_{n-3} - 9 f_{n-4}).$$

Неявные m-шаговые методы Адамса:

m=1, p=2, $\frac{y_n - y_{n-1}}{\tau} = \frac{1}{2}(f_n + f_{n-1})$ – метод трапеций;

$m=2,\ p=3,\ \dfrac{y_n - y_{n-1}}{\tau} = \dfrac{1}{12}(5f_n + 8f_{n-1} - f_{n-2});$

$m=3,\ p=4,\ \dfrac{y_n - y_{n-1}}{\tau} = \dfrac{1}{24}(9f_n + 19f_{n-1} - 5f_{n-2} + f_{n-3}).$

Неявные методы содержат искомое значение y_n нелинейным образом, поэтому для его нахождения применяют итерационные методы решения нелинейных уравнений.

6.3 Численное интегрирование жестких систем обыкновенных дифференциальных уравнений

Жесткие системы можно сравнить с плохо обусловленными системами алгебраических уравнений.

Рассмотрим систему дифференциальных уравнений(ДУ)

$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{f}(t,\boldsymbol{u}), \tag{6.21}$$

где $\boldsymbol{u}(0) = \boldsymbol{u}_0$. Для решения (6.21) рассмотрим разностные методы вида

$$\frac{1}{\tau} \cdot \sum_{k=0}^{m} a_k \cdot y_{n-k} = \sum_{k=0}^{m} b_k \cdot f(t_{n-k}, y_{n-k}), \tag{6.22}$$

где $n = m, m+1, m+2, \ldots$.

Устойчивость и сходимость методов (6.22) определяется расположением корней характеристического уравнения, т.е. $|q| \le 1$ – корни принадлежат единичному кругу.

Среди методов (6.22) выделим те, которые позволяют получить асимптотически устойчивые решения.

Пример. В качестве частного случая (6.21) рассмотрим уравнение вида

$$\frac{du}{dt} = \lambda \cdot u, \tag{6.23}$$

где: $u(0) = u_0$; $\lambda < 0$; $u(t) = u_0 e^{\lambda t}$ – решение ДУ. При $\lambda < 0$ решение есть монотонно убывающая функция при $t \to \infty$. Для этого решения можно записать при любом шаге $\tau > 0$

$$|u(t+\tau)| \le |u(t)|, \qquad (6.24)$$

что означает устойчивость решения $u(t)$.

Рассмотрим для задачи (6.23) метод Эйлера

$$\frac{y_{n+1}-y_n}{\tau} = \lambda \cdot y_n,$$

где: n=0,1,2,…, $y_{n+1} = q \cdot y_n$, q – промежуточный параметр, равный $1+\tau\lambda$.

Оценка (6.24) для метода Эйлера будет выполнена тогда и только тогда, когда $|q|\le 1$. Шаг τ лежит в интервале $0 < \tau \le 2/|\lambda|$. Метод Эйлера для задачи (6.23) устойчив при выполнении этого условия.

Определение 1. Разностный метод (6.22) называется абсолютно устойчивым, если он устойчив при любом $\tau > 0$.

Определение 2. Разностный метод называется условно устойчивым, если он устойчив при некоторых ограничениях на шаг τ.

Например, метод Эйлера для (6.23) условно устойчив, при $0 < \tau \le 2/|\lambda|$. Примером абсолютно устойчивого метода является неявный метод Эйлера

$$\frac{y_{n+1}-y_n}{\tau} = \lambda \cdot y_{n+1},$$

т.к. в этом случае $|q| = \left|(1-\lambda\tau)^{-1}\right| < 1$, при любых $\tau > 0$.

Замечание. Условная устойчивость является недостатком явных методов в связи с тем, что приходится выбирать мелкий шаг интегрирования.

Пример для задачи (6.23). Если $\lambda = -200$, тогда $\tau \le 0.01$. Если мы рассмотрим интервал (0,1], то необходимо будет 100 шагов. Неявные методы со своей стороны приводят к решению на каждом шаге нелинейного уравнения, но это уже недостаток неявного метода.

6.3.1 Понятие жесткой системы ОДУ

Замечание. Все вышерассмотренные методы легко реализуются на примере одного уравнения и легко переносятся на системы ДУ, но при решении систем возникают дополнительные трудности, связанные с разномасштабностью описанных процессов.

Рассмотрим пример системы двух уравнений:

$$\begin{cases} \dfrac{d\,u_1}{dt} + a_1 \cdot u_1 = 0 \\ \dfrac{du_2}{dt} + a_2 \cdot u_2 = 0 \end{cases},$$

где: $t > 0$; $a_1, a_2 > 0$.

Эта система однородных независимых ДУ имеет решение

$$\begin{cases} u_1(t) = u_1(0) \cdot e^{-a_1 \cdot t} \\ u_2(t) = u_2(0) \cdot e^{-a_2 \cdot t} \end{cases}.$$

Это решение монотонно убывает с ростом t. Пусть коэффициент a_2 на порядок больше a_1, т.е. $a_2 >> a_1$. В этом случае компонента u_2 затухает гораздо быстрее u_1, и тогда, начиная с некоторого момента времени t , решение задачи $u(t)$ почти полностью будет определяться поведением компоненты u_1. Однако при численном решении данной задачи шаг интегрирования τ будет определяться, как правило, компонентой u_2, не существенной с точки зрения поведения решения системы. Рассмотрим метод Эйлера для решения данной системы

$$\begin{cases} \dfrac{u_1^{(n+1)} - u_1^{(n)}}{\tau} + a_1 \cdot u_1^{(n)} = 0 \\ \dfrac{u_2^{(n+1)} - u_2^{(n)}}{\tau} + a_2 \cdot u_2^{(n)} = 0 \end{cases}.$$

Он будет устойчив, если на шаг τ наложены ограничения

$$\begin{matrix} \tau \cdot a_1 \le 2 \\ \tau \cdot a_2 \le 2 \end{matrix}.$$

Учитывая, что $a_2 >> a_1 > 0$, получаем окончательное ограничение на τ

$$\tau \le \frac{2}{|a_2|}.$$

Такие трудности могут возникнуть при решении любых систем ОДУ.

Рассмотрим в качестве примера систему

$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{A}\cdot\boldsymbol{u}, \qquad (6.25)$$

где $\boldsymbol{A}$– квадратная матрица $m*m$.

Если матрица $\boldsymbol{A}$ имеет большой разброс собственных чисел, то возникают проблемы с разномасштабностью описываемых системой процессов.

Допустим, что матрица $\boldsymbol{A}$ постоянна (т.е. не зависит от t). Тогда система (6.21) будет называться жесткой, если:

1) вещественные части собственных чисел $\lambda_k < 0$ для всех k, где k=1,…,m;

2) число $S = \frac{\max\limits_{1\le k\le m} |Re\lambda_k|}{\min\limits_{1\le k\le m} |Re\lambda_k|}$ велико (десятки и сотни), и число S называется числом жесткости системы.

Если же матрица $\boldsymbol{A}$ зависит от t, то и собственные числа зависят от t и λ_k зависят от t.

Решение жесткой системы (6.25) содержит как быстро убывающие, так и медленно убывающие составляющие. Начиная с некоторого t >0 решение системы определяется медленно убывающей составляющей. При использовании явных разностных методов быстро убывающая составляющая отрицательно влияет на устойчивость, поэтому приходится брать шаг интегрирования τ слишком мелким.

6.3.2 Некоторые сведения о других методах решения жестких систем

На практике разностные методы (6.22) для решения жестких систем используются в виде методов Гира (неявный разностный метод) и метода матричной экспоненты (метод Ракитского).

6.3.2.1 Методы Гира

Это частный случай методов (6.22), когда коэффициент $b_0 = 1$, $b_1 = b_2 = ... = b_m = 0$. Запишем числовые коэффициенты, которые определяются из условия p-го порядка точности аппроксимации системы разностными методами

$$a_0 = -\sum_{k=1}^{m} a_k \; ; \; \sum_{k=1}^{m} k \cdot a_k = -1; \; \sum_{k=1}^{m} k^l \cdot a_k = 0, \tag{6.26}$$

где l=2,…,p.

Решив систему линейных уравнений (6.26) с учетом предыдущих условий, получаем все нужные коэффициенты.

Трехшаговый метод Гира (частный случай методов (6.22) с учетом условий (6.26)) имеет вид

$$\frac{11}{6} y_n - 3y_{n-1} + \frac{3}{2} y_{n-2} - \frac{1}{3} y_{n-3} = \tau \cdot f(t_n, y_n). \tag{6.27}$$

Метод имеет третий порядок точности.

При m=4, получаем четырехшаговый метод Гира

$$\frac{25 \cdot y_n - 48 \cdot y_{n-1} + 36 \cdot y_{n-2} - 16 \cdot y_{n-3} - 3 \cdot y_{n-4}}{12 \cdot \tau} = f(t_n, y_n). \tag{6.28}$$

Запишем систему (6.26) в виде

$$\begin{cases} a_1 + 2 \cdot a_2 + ... + m \cdot a_m = -1 \\ a_1 + 2^2 \cdot a_2 + ... + m^2 \cdot a_m = 0 \\ \quad ..... \\ a_1 + 2^m \cdot a_2 + ... + m^m \cdot a_m = 0 \end{cases}. \tag{6.29}$$

Решив (6.29) для каждого случая можем найти коэффициенты a_k, k=1,2,…,m.

Чисто неявные разностные методы обладают хорошими свойствами устойчивости, поэтому используются для решения жестких систем уравнений.

6.3.2.2 Метод Ракитского (матричной экспоненты) решения систем ОДУ

$$\frac{d\boldsymbol{u}}{dt} = \boldsymbol{A} \cdot \boldsymbol{u}, \tag{6.30}$$

где: $\boldsymbol{u} = (u_1,...,u_n)$; $\boldsymbol{u}(0) = \boldsymbol{u}_0$; $\boldsymbol{A}$– матрица размерности $n*n$.

Допустим, что матрица $\boldsymbol{A}$ – постоянная, т.е. ее элементы не зависят от времени. Система (6.30) – однородная, с постоянными коэффициентами. Запишем аналитическое решение (6.30)

$$\boldsymbol{u} = e^{\boldsymbol{A}t} \cdot \boldsymbol{u}_0, \tag{6.31}$$

где $e^{\boldsymbol{A}t}$– матричная экспонента и

$$e^{\boldsymbol{A}t} = E + \boldsymbol{A} \cdot t + \frac{(\boldsymbol{A} \cdot t)^2}{2!} + ... + \frac{(\boldsymbol{A} \cdot t)^n}{n!} + ... \tag{6.32}$$

Проинтегрируем уравнение (6.30) при значениях $t=\tau, 2\tau, 3\tau,....$

Если точно знать матрицу $e^{\boldsymbol{A} \cdot \tau}$, то точное решение в указанных точках можно получить по формуле (6.31), т.е. решение можно записать

$$\boldsymbol{u}|_{t=\tau} = e^{\boldsymbol{A} \cdot \tau} \cdot \boldsymbol{u}_0;$$
$$\boldsymbol{u}|_{t=2\tau} = e^{\boldsymbol{A} \cdot \tau} \cdot \boldsymbol{u}\big|_{x=\tau};$$

Таким образом, задача сводится к тому, чтобы достаточно точно знать матрицу $e^{\boldsymbol{A} \cdot \tau}$. На практике поступают следующим образом: при больших τ нельзя воспользоваться рядом Тейлора в связи с его бесконечностью, т.е. для удовлетворительной точности пришлось бы взять много членов ряда, что трудно. Поэтому поступают так: отрезок $[0, \tau]$ разбивают на k частей, чтобы длина $h=\tau/k$ удовлетворяла условию $||\boldsymbol{A} \cdot h||<0.1$. Тогда запишем по схеме Горнера

$$e^{\boldsymbol{A} \cdot h} = \boldsymbol{E} + \boldsymbol{A} \cdot h(\boldsymbol{E} + \frac{\boldsymbol{A} \cdot h}{2}(\boldsymbol{E} + \frac{\boldsymbol{A} \cdot h}{3}(\boldsymbol{E} + \frac{\boldsymbol{A} \cdot h}{4}))).$$

Каждый столбец матрицы $e^{\boldsymbol{A} \cdot h}$ – w_j вычисляют по формуле

$$w_j = (e^{\boldsymbol{A} \cdot h})^k \cdot w_j^0,$$

где w_j^0– вектор столбец, в i-ой строке которого 1, а в остальных – нули.

Если эта матрица найдена, то решение находится по (6.31).

Для исследования разностных методов при решении жестких систем рассматривают модельное уравнение

$$\frac{du}{dt} = \lambda \cdot u, \qquad (6.33)$$

где λ– произвольное комплексное число.

Для того, чтобы уравнение (6.33) моделировало исходную систему (6.30) его нужно рассматривать при таких значениях λ, которые являются собственными числами матрицы $\boldsymbol{A}$. Многошаговые разностные методы (6.31) имеют вид

$$\sum_{k=0}^{m}(a_k - \mu \cdot b_k) \cdot y_{n-k}, \qquad (6.34)$$

где: $n=m, m+1\ldots$; $\mu=\tau\cdot\lambda$.

Если решение уравнения (6.34) искать в виде $y_n = q^n$, то для нахождения числа q получим характеристическое уравнение вида

$$\sum_{k=0}^{m}(a_k - \mu \cdot b_k) \cdot q^{m-k} = 0.$$

Для устойчивости метода достаточно выполнения условия корней $|q_k| \le 1$. В случае жестких систем используются более узкие определения устойчивости.

Предварительные сведения. Областью устойчивости разностных методов называется множество всех точек комплексной плоскости $\mu=\tau\cdot\lambda$, для которых разностный метод применительно к уравнению (6.33) устойчив.

Определение 1. Разностный метод называется A-устойчивым, если область его устойчивости содержит левую полуплоскость $Re\,\mu<0$.

Замечание. Решение модельного уравнения (6.33) асимптотически устойчиво при значениях $Re\,\mu<0$, поэтому сущность A-устойчивого метода заключается в том, что A-устойчивый разностный метод является абсолютно устойчивым, если устойчиво решение исходного дифференциального уравнения.

Так как класс A-устойчивых методов узок, то пользуются $A(\alpha)$-устойчивым методом.

Определение 2. Разностный метод (6.31) называется $A(\alpha)$-устойчивым, если область его устойчивости содержит угол меньший α, т.е. $|arg(-\mu)|<\alpha$, где $\mu=\tau\cdot\lambda$.

Исходя из этого, определяется, что при $\alpha=90°=\pi/2$ $A(\pi/2)$ устойчивость совпадает с определением A-устойчивого метода.

6.4 Краевые задачи для обыкновенных дифференциальных уравнений

Постановка краевой задачи.

Рассматриваем дифференциальное уравнение порядка n ($n\geq 2$)

$$y^{(n)} = f(x, y, y', ..., y^{(n-1)}). \qquad (6.35)$$

Если сделаем замену переменных вида

$$\begin{cases} y_0' = y_1; \\ y_1' = y_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ y_{n-1}' = f(x, y_1, ..., y_{n-1}); \\ y_k(x_0) = y_0^{(k)}, \end{cases} \qquad (6.36)$$

где: $y_k(x_0) = y_0^{(k)}$; $k=0,1,\ldots,(n-1)$, то задача (6.35) сводится к задаче Коши для нормальной системы ОДУ порядка n.

Типовые примеры краевых задач.

Рассмотрим дифференциальное уравнение

$$F(x, y, y', ..., y^{(n)}) = 0. \qquad (6.37)$$

Для уравнения (6.37) краевая задача формулируется следующим образом: найти решение $y=y(x)$, удовлетворяющее уравнению (6.37), для которой значения ее производных в заданной системе точек $x = x_i$ удовлетворяют n независимым краевым условиям, в общем виде нелинейным. Эти краевые условия связывают значения искомой

функции y и ее производных до $(n\text{-}1)$ порядка на границах заданного отрезка.

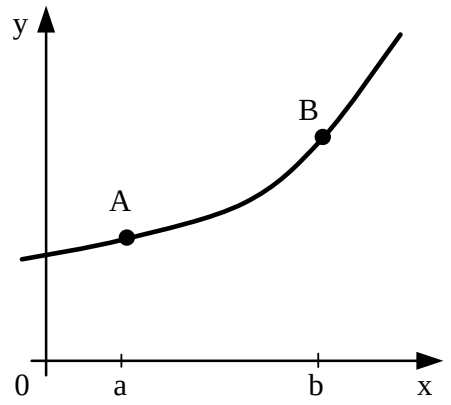


Рисунок 7 – Краевые условия для случая 1

1. Рассмотрим уравнение второго порядка $y'' = f(x, y, y')$. Необходимо найти решение уравнения, удовлетворяющее заданным краевым условиям: $y(a)$=A, $y(b)$=B, т.е. необходимо найти интегральную кривую, проходящую через две заданные точки (рисунок 7).

2. Рассмотрим уравнение $y'' = f(x, y, y')$ с краевыми условиями $y'(a) = A_1$, $y'(b) = B_1$.

Из графика на рисунке 8 видно, что $tg(\alpha)$=A_1, $tg(\beta)$=B_1.

Здесь интегральная кривая пересекает прямые $x=a$ и $x=b$ под заданными углами α и β соответственно.

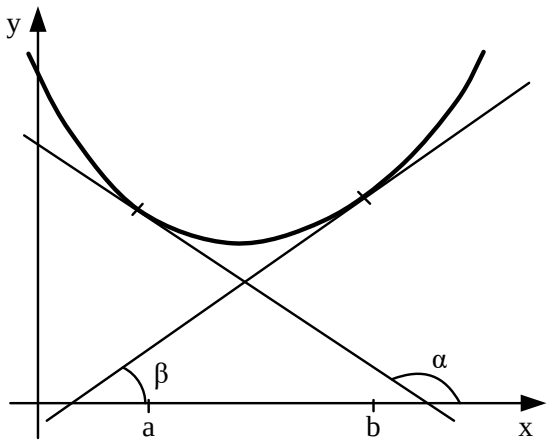


Рисунок 8 – Краевые условия для случая 2

3. Смешанная краевая задача.

Рассмотрим то же самое уравнение $y'' = f(x, y, y')$с краевыми условиями $y(a)$=A_1, $y'(b)$=B_1 .

Геометрическую иллюстрацию этих краевых условий легко представить, используя рисунки 7 и 8.

Замечание. Краевая задача для уравнения (6.37) в общем случае может не иметь решений, иметь единственное решение, иметь несколько решений или бесконечное множество решений.

4. Поражение заданной цели баллистическим снарядом. Дифференциальные уравнения движения снаряда с учетом сопротивления воздуха имеют вид

$$\begin{cases} \ddot{x} = -E\cos\Theta \\ \ddot{y} = -E\sin\Theta - g \end{cases},$$

где: $\ddot{x}$ – вторая производная по времени; $E=E(y,v)$ – известная функция высоты и скорости; $v=\sqrt{\dot{x}^2+\dot{y}^2}$; $g=g(y)$ – ускорение силы тяжести; Θ – угол наклона к горизонту касательной к траектории движения снаряда; $\Theta = arctg\frac{\dot{y}}{\dot{x}}$.

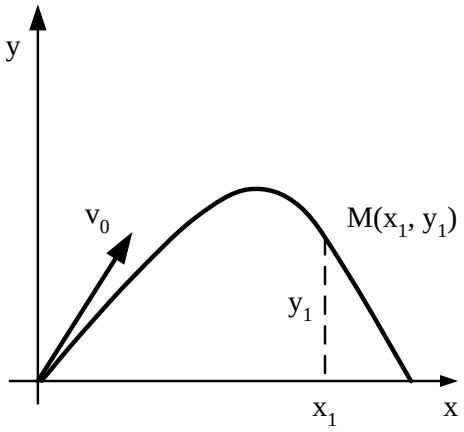


Рисунок 9 – Траектория снаряда

Предполагая, что при $t=t_0$ снаряд выпущен из точки, совпадающей с началом координат с начальной скоростью v_0 под углом Θ_0, а в момент $t=t_1$ он поразит неподвижную мишень в точке $M(x_1,y_1)$ получаем краевые условия

$$\begin{cases} x=0, \quad y=0, \quad \dot{x}=v_0\cos\Theta_0, \quad \dot{y}=v_0\sin\Theta_0, \quad \text{при } t=t_0, \\ x=x_1, \quad y=y_1, \quad \text{при } t=t_1. \end{cases}$$

Здесь неизвестны значения Θ_0 и t_1. Решив данную краевую задачу, можем найти начальный угол $\Theta_0 = arctg\frac{\dot{y}_0}{\dot{x}_0}$, где: $\dot{x}_0=\dot{x}(t_0); \dot{y}_0=\dot{y}(t_0)$; Θ_0 – угол, при котором поражается цель в точке M.

Аналогично ставятся краевые задачи для систем дифференциальных уравнений.

6.5 Решение линейной краевой задачи

Рассмотрим важный частный случай решения краевой задачи, когда дифференциальное уравнение и краевые условия линейны.

Для этого рассмотрим уравнение

$$p_0(x)\cdot y^{(n)}(x)+p_1(x)\cdot y^{(n-1)}(x)+...+p_n(x)\cdot y(x)=f(x), \qquad (6.38)$$

где: $p_i(x)$ и $f(x)$ известные непрерывные функции на отрезке $[a, b]$.

Предположим, что в краевые условия входят две абсциссы $x=a$, $x=b$. Это двухточечные краевые задачи. Краевые условия называются линейными, если они имеют вид

$$R_v(y)=\sum_{k=0}^{n-1}(\alpha_k^{(v)}\cdot y^{(k)}(a)+\beta_k^{(v)}\cdot y^{(k)}(b))=\gamma, \qquad (6.39)$$

где: α, β, γ – заданные константы. Причем они одновременно не равны нулю, т.е.

$$\sum_{k=0}^{n-1}[|\alpha_k^{(v)}|+|\beta_k^{(v)}|]\neq 0, \text{ при } v=1,2,\ldots,n.$$

Например, краевые условия во всех трех рассмотренных ранее задачах линейны, т.к. их можно записать в виде

$$\begin{aligned}\alpha_0\cdot y(a)+\alpha_1\cdot y'(a)&=\gamma_1\\ \beta_0\cdot y(b)+\beta_1\cdot y'(b)&=\gamma_2\end{aligned},$$

причем $\alpha_0=1,\alpha_1=0,\gamma_1=A,\beta_0=1,\beta_1=0,\gamma_2=B$ – для первой задачи.

6.6 Решение двухточечной краевой задачи для линейного уравнения второго порядка сведением к задаче Коши

Запишем линейное уравнение второго порядка в виде

$$y''+p(x)\cdot y'+q(x)\cdot y=f(x), \qquad (6.40)$$

где: p, q, f – известные непрерывные функции на некотором отрезке $[a,b]$.

Требуется найти решение уравнения (6.40), удовлетворяющее заданным краевым условиям

$$\begin{aligned} \alpha_0 \cdot y(a) + \alpha_1 \cdot y'(a) &= A, \\ \beta_0 \cdot y(b) + \beta_1 \cdot y'(b) &= B. \end{aligned} \qquad (6.41)$$

Причем константы α и β одновременно не равны нулю

$$\begin{aligned} |\alpha_0| + |\alpha_1| &\neq 0, \\ |\beta_0| + |\beta_1| &\neq 0. \end{aligned}$$

Решение задачи (6.40), (6.41) будем искать в виде линейной комбинации

$$y = C \cdot u + V,$$

где C – константа, u – общее решение соответствующего однородного уравнения

$$u'' + p(x) \cdot u' + q(x) \cdot u = 0, \qquad (6.42)$$

а V– некоторое частное решение неоднородного уравнения

$$V'' + p(x) \cdot V' + q(x) \cdot V = f(x). \qquad (6.43)$$

Потребуем ,чтобы первое краевое условие было выполнено при любом C,

$$C(\alpha_0 \cdot u(a) + \alpha_1 \cdot u'(a)) + (\alpha_0 \cdot V(a) + \alpha_1 \cdot V'(a)) = A,$$

откуда следует, что $\alpha_0 \cdot u(a) + \alpha_1 \cdot u'(a) = 0,$

$$\alpha_0 \cdot V(a) + \alpha_1 \cdot V'(a) = A.$$

Тогда

$$\begin{aligned} u(a) &= \alpha_1 k \\ u'(a) &= -\alpha_0 k \end{aligned}, \qquad (6.44)$$

где k– некоторая константа, не равная нулю.

Значение функции V и ее производная в точке a могут быть, например,выбраны равными

$$\begin{matrix} V(a) = A/\alpha_0 \\ V'(a) = 0 \end{matrix}, \qquad (6.45)$$

если коэффициент $\alpha_0 \neq 0$ и

$$\begin{matrix} V(a) = 0 \\ V'(a) = A/\alpha_1 \end{matrix}, \qquad (6.46)$$

если коэффициент $\alpha_1 \neq 0$.

Из этих рассуждений следует, что функция u – есть решение задачи Коши для однородного уравнения (6.42) с начальными условиями (6.44), а функция V – есть решение задачи Коши для неоднородного уравнения (6.43) с начальными условиями (6.45) или (6.46) в зависимости от условий. Константу C надо подобрать так, чтобы выполнялись условия (6.41) (вторая строчка) в точке $x=b$

$$C(\beta_0 \cdot u(b) + \beta_1 \cdot u'(b)) + \beta_0 \cdot V(b) + \beta_1 \cdot V'(b) = B.$$

Отсюда следует, что

$$C = \frac{B - (\beta_0 \cdot V(b) + \beta_1 \cdot V'(b)}{\beta_0 \cdot u(b) + \beta_1 \cdot u'(b)},$$

где знаменатель не должен быть равен нулю, т. е.

$$\beta_0 \cdot u(b) + \beta_1 \cdot u'(b) \neq 0. \qquad (6.47)$$

Если условие (6.47) выполнено, то краевая задача (6.35), (6.36) имеет единственное решение. Если же (6.47) не выполняется, то краевая задача (6.35), (6.36) либо не имеет решения, либо имеет бесконечное множество решений.

6.7 Методы численного решения двухточечной краевой задачи для линейного уравнения второго порядка

6.7.1 Метод конечных разностей

Рассмотрим линейное дифференциальное уравнение

$$y'' + p(x) \cdot y' + q(x) \cdot y = f(x) \tag{6.48}$$

с двухточечными краевыми условиями

$$\begin{cases} \alpha_0 y(a) + \alpha_1 y'(a) = A \\ \beta_0 y(b) + \beta_1 y'(b) = B \end{cases}; \tag{6.49}$$

$$(|\alpha_0| + |\alpha_1| \neq 0, \quad |\beta_0| + |\beta_1| \neq 0),$$

где: p, q, f – известные непрерывные функции на некотором отрезке $[a,b]$.

Одним из наиболее простых методов решения этой краевой задачи является сведение ее к системе конечно-разностных уравнений.

Основной отрезок $[a,b]$ делим на n равных частей с шагом $h=(b-a)/n$, то есть рассматриваем равномерную сетку $x_i = x_0 + i \cdot h$, $i=0,1,\ldots,n$.

В каждом внутреннем узле дифференциальное уравнение (6.48) аппроксимируем, используя формулы численного дифференцирования второго порядка точности

$$\begin{aligned} y'_i &= \frac{y_{i+1} - y_{i-1}}{2 \cdot h}; \\ y''_i &= \frac{y_{i+2} - 2 \cdot y_i + y_{i+1}}{h^2}, \end{aligned} \tag{6.50}$$

где $i=1,\ldots, n-1$.

Для граничных точек $x_0 = a$ и $x_n = b$, чтобы не выходить за границы отрезка, используем формулы численного дифференцирования первого порядка точности

$$y'_0 = \frac{y_1 - y_0}{h}, \quad y'_n = \frac{y_{n-1} - y_n}{-h}. \tag{6.51}$$

Используя отношение (6.50) исходное дифференциальное уравнение (6.48) аппроксимируем конечно-разностными уравнениями

$$\frac{y_{i+1}-2\cdot y_i+y_{i-1}}{h^2}+p_i\frac{y_{i+1}-y_{i-1}}{2\cdot h}+q_i\cdot y_i=f_i, \qquad (6.52)$$

где i=1,...,n-1. Учитывая краевые условия, получим еще два уравнения

$$\begin{cases} \alpha_0\cdot y_0+\alpha_1\dfrac{y_1-y_0}{h}=A \\ \beta_0\cdot y_n+\beta_1\dfrac{y_{n-1}-y_n}{-h}=B \end{cases}. \qquad (6.53)$$

Таким образом, получена линейная система n+1 уравнений с n+1 неизвестными $y_0, y_1,...,y_n$, представляющими собой значения искомой функции $y=y(x)$ в точках $x_0, x_1,...,x_n$. Разностная схема (6.52)-(6.53) аппроксимируеткраевую задачу (6.48) - (6.49) с порядком 1 по h за счет краевых условий. Решив эту систему, получим таблицу значений искомой функции y.

6.7.2 Метод прогонки (одна из модификаций метода Гаусса)

При применении метода конечных разностей к краевым задачам для дифференциальных уравнений второго порядка получается система линейных алгебраических уравнений с трехдиагональной матрицей, т.е. каждое уравнение системы содержит три соседних неизвестных. Для решения таких систем разработан специальный метод – «метод прогонки».

Для этого систему (6.52) перепишем в виде

$$y_{i+1}+m_i\cdot y_i+n_i\cdot y_{i-1}=\tilde{f}_i\cdot h^2 \qquad (6.54)$$

для внутренних точек (i=1,…,n-1), где:

$$m_i=\frac{2-q_i\cdot h^2}{1+\frac{p_i\cdot h}{2}};\quad n_i=\frac{1-\frac{p_i\cdot h}{2}}{1+\frac{p_i\cdot h}{2}};\quad \tilde{f}_i=\frac{f_i}{1+\frac{p_i\cdot h}{2}}.$$

На концах отрезка $x_0=a$ и $x_n=b$ производные заменяем разностными отношениями

$$y'_0 = \frac{y_1 - y_0}{h} \quad \text{и} \quad y'_n = \frac{y_{n-1} - y_n}{-h}.$$

Учитывая эту замену, получим еще два уравнения

$$\begin{aligned} &\alpha_0 \cdot y_0 + \alpha_1 \frac{y_1 - y_0}{h} = A; \\ &\beta_0 \cdot y_n + \beta_1 \frac{y_{n-1} - y_n}{-h} = B. \end{aligned} \tag{6.55}$$

Обратим внимание на внешний вид записи системы (6.54), (6.55). В каждом уравнении системы присутствует три ненулевых элемента. В первом и последнем – по два ненулевых коэффициента.

Разрешая уравнение (6.54) относительно y_i, получим

$$y_i = \frac{\tilde{f}_i}{m_i} h^2 - \frac{1}{m_i} y_{i+1} - \frac{n_i}{m_i} y_{i-1}. \tag{6.56}$$

Предположим, что с помощью полной системы (6.54), (6.55) из уравнения (6.56) исключена неизвестная y_{i-1}. Тогда это уравнение примет вид

$$y_i = c_i (d_i - y_{i+1}), \tag{6.57}$$

где: c_i, d_i – некоторые коэффициенты; i=1,2,…,n-1. Отсюда

$$y_{i-1} = c_{i-1}(d_{i-1} - y_i).$$

Подставляя это выражение в уравнение (6.54), получим

$$y_{i+1} + m_i y_i + n_i c_{i-1}(d_{i-1} - y_i) = \tilde{f}_i h^2,$$

а отсюда

$$y_i = \frac{(\tilde{f}_i h^2 - n_i c_{i-1} d_{i-1}) - y_{i+1}}{m_i - n_i c_{i-1}}. \tag{6.58}$$

Сравнивая (6.57) и (6.58), получим для определения c_i и d_i рекуррентные формулы

$$c_i = \frac{1}{m_i - n_i c_{i-1}}; \quad d_i = \tilde{f}_i h^2 - n_i c_{i-1} d_{i-1}; \quad i = 1,\ldots,n\text{-}1. \qquad (6.59)$$

Из первого краевого условия (6.55) и из формулы (6.57) при i=0 находим

$$c_0 = \frac{\alpha_1}{\alpha_0 h - \alpha_1}; d_0 = \frac{Ah}{\alpha_1}. \qquad (6.60)$$

На основании формул (6.59), (6.60) последовательно определяются коэффициенты c_i, d_i (i=1,…,n-1) до c_{n-1} и d_{n-1} включительно (прямой ход).

Обратный ход начинается с определения y_n. Для этого из второго краевого условия (6.55) и из формулы (6.57) при $i=n$-1 найдем

$$y_n = \frac{\beta_0 h + \beta_1 c_{n-1} d_{n-1}}{\beta_0 h + \beta_1 (c_{n-1} + 1)}. \qquad (6.61)$$

Далее по формуле (6.57) последовательно находим $y_{n-1}, y_{n-2}, \ldots, y_0$.

Заметим, что метод прогонки обладает устойчивым вычислительным алгоритмом.

7 Приближенное решение дифференциальных уравнений в частных производных

В реальных физических процессах искомая функция зависит от нескольких переменных, а это приводит к уравнениям в частных производных от искомой функции. Как и для обыкновенных дифференциальных уравнений(ОДУ), в этом случае для выбора одного конкретного решения, удовлетворяющего уравнению в частных производных, кроме начальных условий, необходимо задавать дополнительные условия (т.е. краевые условия). Чаще всего такие задачи на практике не имеют аналитического решения и приходится использовать численные методы их решения, в том числе метод сеток, метод конечных разностей и так далее. Мы будем рассматривать класс линейных уравнений в частных производных второго порядка. В общем виде в случае двух переменных эти уравнения записываются в виде

$$A(x,y)\frac{\partial^2 u}{\partial x^2}+B(x,y)\frac{\partial^2 u}{\partial x\partial y}+C(x,y)\frac{\partial^2 u}{\partial y^2}+a(x,y)\frac{\partial u}{\partial x}+$$
$$+b(x,y)\frac{\partial u}{\partial y}+c(x,y)u=F(x,y), \qquad (7.1)$$

где: A, B, C, a, b, c − заданные непрерывные функции двух переменных, имеющие непрерывные частные производные, u – искомая функция. Для сокращения записи введем обозначения

$$u_{xx}=\frac{\partial u}{\partial x^2};\quad u_{xy}=\frac{\partial^2 u}{\partial x\partial y};\quad u_{yy}=\frac{\partial^2 u}{\partial y^2};\ u_x=\frac{\partial u}{\partial x};\qquad u_y=\frac{\partial u}{\partial y}.$$

Будем рассматривать упрощенную форму записи (7.1) вида

$$A(x,y)u_{xx}+B(x,y)u_{xy}+C(x,y)u_{yy}+a(x,y)u_x+$$
$$+b(x,y)u_y+c(x,y)u=F(x,y) \qquad (7.2)$$

и рассмотрим частный случай (7.2), когда $a=b=c=F\equiv 0$, т.е.

$$A(x,y)u_{xx}+B(x,y)u_{xy}+C(x,y)u_{yy}=0. \qquad (7.3)$$

Путем преобразований уравнение (7.3) может быть приведено к каноническому виду (к одному из трех стандартных канонических форм) эллиптическому типу, гиперболическому типу, параболическому типу. Причем тип уравнения будет определяться коэффициентами A, B, C, a именно – знаком дискриминанта

$$D=B^2-4\cdot A\cdot C.$$

Если $D<0$, то имеем уравнение эллиптического типа в точке с координатами x, y; если $D=0$, то (7.3) − параболического типа; если $D>0$, то (7.3) − гиперболического типа; если D не сохраняет постоянного знака, то (7.3) − смешанного типа.

Замечание. Если A, B, C − константы, тогда каноническое уравнение (7.3) называется полностью эллиптического, параболического, гиперболического типа.

Введем понятие оператора Лапласа для сокращенной записи канонических уравнений вида

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}.$$

Используя это определение, запишем сокращенные канонические уравнения всех трех типов

1. $\Delta u=0$. Это уравнение эллиптического типа, так называемое уравнение Лапласа. В механике это уравнение описывает стационарные тепловые поля, установившееся течение жидкости и т.д.

2. $\Delta u=-f$, где f − заданная непрерывная функция. Это уравнение Пуассона имеет эллиптический тип и описывает процесс теплопередачи с внутренним источником тепла.

3. $a^2\Delta u=\partial u/\partial t$, где a − константа. Не во всех уравнениях в качестве переменных будут выступать стандартные переменные x, y. Может быть также переменная времени. Это уравнение диффузии описывает процесс теплопроводности и является уравнением параболического типа.

4. $\frac{\partial^2 u}{\partial t^2}=a^2\Delta u$, a– константа. Это уравнение гиперболического типа − так называемое волновое уравнение и оно описывает процесс распространения волн.

7.1 Метод сеток для решения смешанной задачи для уравнения параболического типа (уравнения теплопроводности)

Смешанная задача означает, что следует найти искомую функцию, удовлетворяющую заданному уравнению в частных производных, краевым, а так же начальным условиям. Различить эти условия можно в том случае, если одна из независимых переменных − время, а другая − пространственная координата. При этом условия, относящиеся к начальному моменту времени, называются начальными, а условия, относящиеся к фиксированным значениям координат − краевыми.

Рассмотрим смешанную задачу для однородного уравнения теплопроводности

$$\frac{\partial u}{\partial t}=k\frac{\partial^2 u}{\partial x^2},\ k=\text{const}>0. \qquad (7.4)$$

Задано начальное условие

$$u(x,0)=f(x) \qquad (7.5)$$

и заданы краевые условия первого рода

$$u(0,t) = \mu_1(t);$$
$$u(a,t) = \mu_2(t). \qquad (7.6)$$

Требуется найти функцию $u(x,t)$, удовлетворяющую в области D ($0<x<a$, $0<t\le T$) условиям (7.5) и (7.6).

К задаче (7.4) – (7.6) приводит задача о распространении тепла в однородном стержне длины a, на концах которого поддерживается заданный температурный режим.

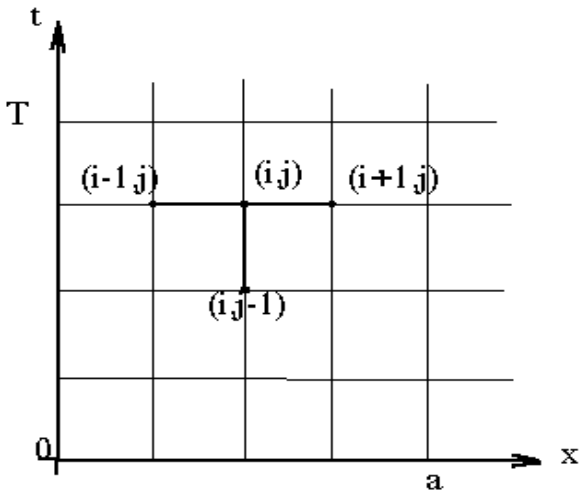


Рисунок 10 – Четырехточечный шаблон неявной схемы

При проведении замены $t \to t/k$ получим $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$, т.е. k=1. Задача решается *методом сеток*: строим в области D равномерную прямоугольную сетку с шагом h по оси x и шагом τ по оси t (см. рисунок 10).

Приближенные значения искомой функции $u(x_i,t_j)$ в точках (x_i,t_j) обозначим через u_{ij}.Тогда

$$x_i = i \cdot h;\ \ h = a/n;\ \ i=0,1,...,n;\ \ t_j = j \cdot \tau;\ \ j=0,1,...,m;\ \ \tau = T/m.$$

Заменим производные в (7.4) разностными отношениями

$$\frac{\partial u}{\partial t} = \frac{u_{i,j} - u_{i,j-1}}{\tau} + O(\tau);$$
$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + O(h^2).$$

В результате получим неявную двухслойную разностную схему с погрешностью $O(\tau+h^2)$

$$\frac{u_{i,j} - u_{i,j-1}}{\tau} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2}$$

Используя подстановку $\lambda = \tau/h^2$, выразим из этой схемы $u_{i,j-1}$

$$u_{i,j-1} = (2\lambda + 1)u_{i,j} - \lambda u_{i+1,j} - \lambda u_{i-1,j}. \qquad (7.7)$$

Получаем разностную схему, которой аппроксимируем уравнение (7.4) во внутренних узлах сетки. Число уравнений меньше числа неизвестных u_{ij} . Из краевых условий получим уравнения

$$u_{0,j} = \mu_1(t_j); \quad u_{n,j} = \mu_2(t_j) ,$$

которые с (7.7) образуют неявную схему. Ее шаблон изображен на рисунке 10.

Получаем систему линейных уравнений с трехдиагональной матрицей. Решив ее любым способом (в частности, методом прогонки), получаем значения функции на определенных временных слоях. Так, на нулевом временном слое используем начальное условие $u_{i,0}=f(x_i)$, т.к. j=0. На каждом следующем слое искомая функция определяется из решения полученной системы. Неявная схема устойчива для любых значений параметра $\lambda = \tau / h^2 > 0$.

Есть и явная схема (рисунок 11), но она устойчива только при $\lambda \le 1/2$, т.е. при $\tau \le h^2/2$. Вычисления по этой схеме придется вести с малым шагом по τ, что приводит к большим затратам машинного времени.

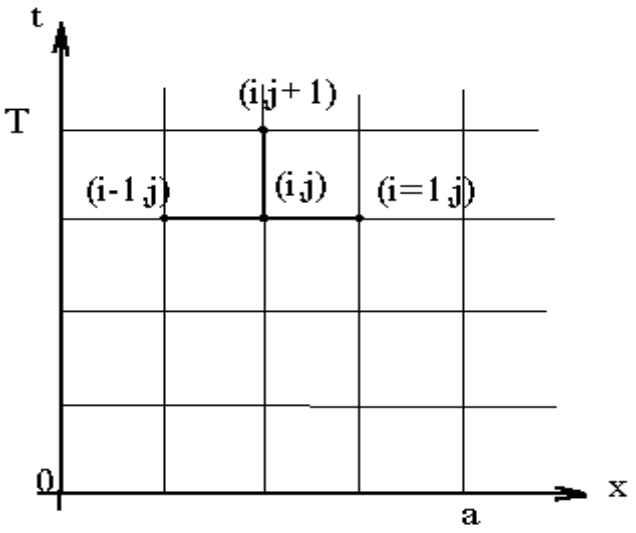


Рисунок 11 – Четырехточечный шаблон явной схемы

7.2 Решение задачи Дирихле для уравнения Лапласа методом сеток

Рассмотрим уравнение Лапласа

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \cdot \qquad (7.8)$$

Будем рассматривать уравнение Лапласа в прямоугольной области

$\Omega = \{(x,y), 0 \le x \le a, 0 \le y \le b\}$ с краевыми условиями

$$u(0,y) = f_1(y); \quad u(a,y) = f_2(y); \quad u(x,0) = f_3(x); \quad u(x,b) = f_4(x),$$

где f_1, f_2, f_3, f_4 – заданные функции. Заметим, что чаще всего область бывает не прямоугольной.

Введем обозначения $u_{ij}=u(x_i,y_j)$. Накладываем на прямоугольную область сетку $x_i = h \cdot i$; $i=0,1,\ldots,n$; $y_j = l \cdot j$; $j=0,1,\ldots,m$. Тогда $x_n = h \cdot n$, $y_m = l \cdot m = b$.

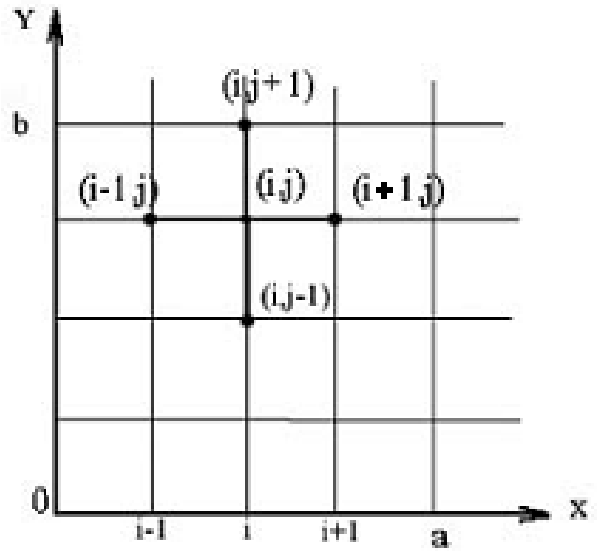


Рисунок 12 – Схема узлов «крест»

Частные производные аппроксимируем по формулам

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{i+1} - 2 \cdot u_{i,j} + u_{i-1,j}}{h^2} + O(h^2);$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{u_{i,j+1} - 2 \cdot u_{i,j} + u_{i,j-1}}{l^2} + O(l^2),$$

и заменим уравнение Лапласа конечно-разностным уравнением

$$\frac{u_{i+1,j} - 2 \cdot u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2 \cdot u_{i,j} + u_{i,j-1}}{l^2} = 0, \qquad (7.9)$$

где: $i=1,\ldots,n-1$, $j=1,\ldots,m-1$ (т.е. для внутренних узлов).

Погрешность замены дифференциального уравнения разностным составляет величину $O(h^2+l^2)$. Уравнения (7.9) и значения $u_{i,j}$ в граничных узлах образуют систему линейных алгебраических уравнений относительно приближенных значений функции $u(x,y)$ в узлах сетки. Выразим $u_{i,j}$ при $h=l$, и заменим систему

$$
\begin{aligned}
&u_{ij} = (u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1})/4;\\
&u_{i0} = f_3(x_i);\\
&u_{im} = f_4(x_i);\\
&u_{0j} = f_1(y_j);\\
&u_{nj} = f_2(y_j).
\end{aligned}
\qquad (7.10)
$$

Систему (7.10) линейных алгебраических уравнений можно решить любым итерационным методом (Зейделя, простых итераций и т.д.). При построении системы использовалась схема типа «крест» (рисунок 12). Строим последовательность итераций по методу Зейделя

$$
u_{i,j}^{(s+1)} = \frac{1}{4}(u_{i-1,j}^{(s+1)} + u_{i+1,j}^{(s)} + u_{i,j+1}^{(s)} + u_{i,j-1}^{(s+1)}),
$$

где s– текущая итерация. Условие окончания итерационного процесса

$$
\max_{i,j} \left| u_{ij}^{(s+1)} - u_{ij}^{(s)} \right| < \varepsilon \quad . \qquad (7.11)
$$

Условие (7.11) ненадежно и на практике используют другой критерий

$$
\max_{i,j} \left| u_{ij}^{(s+1)} - u_{ij}^{(s)} \right| \le \varepsilon(1-\nu),
$$

где $\nu = \dfrac{\max \left| u_{ij}^{(s+1)} - u_{ij}^{(s)} \right|}{\max \left| u_{ij}^{(s)} - u_{ij}^{(s-1)} \right|}.$

Схема «крест»– явная устойчивая схема (малое изменение входных данных ведет к малому изменению выходных данных).

7.3 Решение смешанной задачи для уравнения гиперболического типа методом сеток

Рассмотрим уравнение колебания однородной и ограниченной струны.

Задача состоит в отыскании функции $u(x,t)$ при $t>0$, удовлетворяющей уравнению гиперболического типа

$$\frac{\partial^2 u}{\partial t^2} = c \cdot \frac{\partial^2 u}{\partial x^2}, \qquad (7.12)$$

где: $0<x< a$; $0<t\leq T$, начальным условиям

$$u(x,0) = f(x);$$
$$\frac{\partial u}{\partial t}(x,0) = g(x); \qquad (7.13)$$
$$0 \leq x \leq a$$

и краевым условиям

$$u(0,t) = \mu_1(t);$$
$$u(a,t) = \mu_2(t); \qquad (7.14)$$
$$0 \leq t \leq T.$$

Выполним замену переменных ct на t и получим уравнение

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

и в дальнейшем будем считать c=1.

Для построения разностной схемы решения задачи (7.12)-(7.14) построим в области $D = \{(x,t), 0 \leq x \leq a, 0 \leq t \leq T\}$ сетку $x_i = ih$; i=0,1,…,n; $a = h \cdot n$; $t_j = j\tau$; j=0,1,…,m; $\tau m = T$.

Аппроксимируем (7.12) в каждом внутреннем узле сетки на шаблоне типа «крест» (рисунок 12). Используем для аппроксимации частных производных разностные производные второго порядка точности относительно шага и получаем разностную аппроксимацию уравнения (7.12)

$$\frac{u_{i,j-1} - 2 \cdot u_{i,j} + u_{i,j+1}}{\tau^2} = \frac{u_{i+1,j} - 2 \cdot u_{i,j} + u_{i-1,j}}{h^2}, \qquad (7.15)$$

где u_{ij} – приближенное значение функции $u(x,t)$ в узле (x_i, t_j).

Полагая $\lambda = \tau/h$, перепишем (7.15), выразив $u_{i,j+1}$. Таким образом, получим трехслойную разностную схему

$$u_{i,j+1} = \lambda^2 (u_{i+1,j} - u_{i-1,j}) + 2(1-\lambda^2) u_{i,j} - u_{i,j-1}, \qquad (7.16)$$

где: $i=1,\ldots,n$; $j=1,\ldots,m$. Задаем нулевые граничные условия $\mu_1(t)=0$, $\mu_2(t)=0$. Тогда в (7.16) $u_{0j}=0$, $u_{nj}=0$ для всех j.

Схема (7.16) называется трехслойной, т.к. она связывает значения искомой функции на трех временных слоях $j-1$, j, $j+1$. Схема (7.16) явная, т.е. позволяет в явном виде выразить u_{ij} через значения функции с предыдущих двух слоев.

Численное решение задачи состоит в вычислении приближенных значений u_{ij} решения $u(x,t)$ в узлах (x_i,t_j) при $i=1,\ldots,n$; $j=1,\ldots,m$. Алгоритм решения основан на том, что решение на каждом следующем слое ($j=2,3,..,n$) можно получить пересчетом решений с двух предыдущих слоев ($j=0,1,..,n-1$) по формуле (7.16). При $j=0$ решение известно из начального условия $u_{i0}=f(x_i)$. Для вычисления решения на первом слое ($j=1$) положим

$$\frac{\partial u}{\partial t}(x,0)\approx\frac{u(x,\tau)-u(x,0)}{\tau}, \qquad (7.17)$$

тогда $u_{i1}=u_{i0}-\tau g(x_i)$, $i=1,2,\ldots,n$. Для вычисления решений на следующих слоях можно использовать формулу (7.16). Решение на каждом последующем слое получается пересчетом решений с двух предыдущих слоев по формуле (7.16).

Описанная схема аппроксимирует задачу (7.12) – (7.14) с точностью $O(\tau+h^2)$. Невысокий порядок аппроксимации по τ объясняется использованием грубой аппроксимации для производной по t в формуле (7.17).

Схема будет устойчивой, если выполнено условие Куранта $\tau<h$. Это означает, что малые погрешности, возникающие при вычислении решения на первом слое, не будут неограниченно возрастать при переходе к каждому новому временному слою. Недостаток схемы в том, что сразу после выбора шага h в направлении x, появляется ограничение на величину шага τ по переменной t.