

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

---

Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«Пензенский государственный университет  
архитектуры и строительства»  
(ПГУАС)

В.В. Кузина, А.Н. Кошев

## **ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА**

Рекомендовано Редсоветом университета  
в качестве учебного пособия  
по направлению подготовки 09.03.02  
«Информационные системы и технологии»

Пенза 2016

УДК 519.6 (075.8)  
ББК 22.193я73  
К89

Рецензенты: доктор технических наук, профессор  
И.А. Прошин (ПензГТУ);  
доктор технических наук, профессор  
А.М. Данилов (ПГУАС)

**Кузина В.В.**

К89      Вычислительная математика: учеб. пособие по направлению  
подготовки 09.03.02 «Информационные системы и технологии»/  
В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2016. – 172 с.

Учебное пособие посвящено теоретическим основам вычислительной математики.

Учебное пособие подготовлено на кафедре «Информационно-вычислительные системы» и предназначено для студентов, обучающихся по программе подготовки академического бакалавриата по направлению 09.03.02 «Информационные системы и технологии» очной формы обучения.

© Пензенский государственный университет  
архитектуры и строительства, 2016

© Кузина В.В., Кошев А.Н., 2016

## ПРЕДИСЛОВИЕ

Предлагаемое учебное пособие предназначено для студентов бакалавриата, обучающихся по направлению 09.03.02 «Информационные системы и технологии», при изучении дисциплины «Вычислительная математика».

Пособие разработано на кафедре «Информационно-вычислительные системы» в соответствии с рабочей программой дисциплины и требованиями ФГОС ВО.

Цель учебного пособия по дисциплине «Вычислительная математика» – формирование у студентов базы для развития профессиональных компетенций в области вычислительной математики для их дальнейшего применения в профессиональной деятельности.

Задачи учебного пособия:

- формирование умения применять методы вычислительной математики для решения задач интегрирования, дифференцирования, интерполирования и аппроксимации функций, решения линейных алгебраических и нелинейных уравнений и их систем, дифференциальных уравнений;
- выработка умения реализовывать численные методы в различных интегрированных математических средах (*MathCad*, *Matlab* и др.) и прикладных программах.

Процесс изучения дисциплины направлен на формирование следующих компетенций:

- владение широкой общей подготовкой (базовыми знаниями) для решения практических задач в области информационных систем и технологий;
- способность использовать основные законы естественнонаучных дисциплин в профессиональной деятельности, применять методы математического анализа и моделирования, теоретического и экспериментального исследования.

Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенции):

*знать:*

- математический аппарат современных численных методов;
- основные положения и методы численного дифференцирования и интегрирования, интерполирования и аппроксимации функций, численного решения алгебраических и дифференциальных уравнений и их систем; о приложениях теории в информатике, программировании и вычислительной технике;
- современные методы обработки результатов измерений (аппроксимация, визуализация и оценка погрешности);
- современные математические пакеты программ для решения задач вычислительной математики;

*уметь:*

– осуществлять цепочку приближенных арифметических вычислений с заданной точностью, реализовывать численные методы решения задач на ПЭВМ;

– решать типовые задачи;

– использовать встроенные функции математических пакетов для решения задач вычислительной математики;

– программировать вычислительные алгоритмы и решать типовые задачи на компьютере;

*владеть:*

– базовыми знаниями и навыками методов вычислительной математики;

– навыками решения проблемных задач, требующих применения логико-математического аппарата;

– навыками работы в интегрированных математических средах и с прикладными математическими пакетами программ;

– навыками решения проблемных задач, используя вычислительный эксперимент;

*иметь представление:*

– о математическом аппарате современных численных методов;

– об основных положениях и методах современной вычислительной математики, о приложениях теории в информатике, программировании и вычислительной технике;

– о применении методов математического анализа и моделирования в профессиональной деятельности;

– о математическом моделировании и вычислительном эксперименте.

Знания, умения и приобретенные компетенции будут использованы при изучении следующих дисциплин (модулей) и разделов ООП: численные методы и методы оптимизации; моделирование процессов и систем; математическая статистика и прогнозирование.

## ВВЕДЕНИЕ

Методы вычислительной математики находят все большее применение в различных областях науки и в самых разнообразных сферах профессиональной деятельности человека, что способствует созданию и применению инструмента математического моделирования для решения прикладных задач. Современные численные методы в совокупности с возможностью их алгоритмизации при использовании персональных компьютеров и превращаются в рабочий инструмент для решения научных, технических, экономических и других задач.

Разработка технологических процессов, проектирование объектов, оптимальное управление объектами или процессами предполагают предварительное *математическое моделирование* этих явлений, то есть описание их математическими терминами в виде математических уравнений, формул и дополнительных условий, которым должно удовлетворять решение.

*Модель* (от лат. *modus, modulus*) – мера, способ, образ и т.д. *Моделирование* – замещение одного объекта другим с целью получения информации о важнейших свойствах объекта-оригинала с помощью объекта-модели. Объект изучают с помощью вычислительных методов.

Простая математическая модель – это совокупность алгебраических формул, по которым явно вычисляются искомые величины. Однако чаще всего поведение параметров описывается дифференциальными уравнениями в частных производных. Использование современных быстродействующих ЭВМ позволяет решать сложные задачи на основе вычислительных методов.

Все математические задачи классифицированы, т.е. объединены в некоторые группы. Для каждой группы задач существует набор стандартных методов, которые изучает специальный раздел математики, названный «Вычислительная математика» или «Методы вычислений».

Современная математическая энциклопедия определяет термин «*вычислительная математика*» как раздел математики, включающий круг вопросов, связанных с использованием ЭВМ. При этом утверждается, что содержание термина нельзя считать установившимся, так как эта область математики интенсивно развивается в связи с быстро растущими применениями ЭВМ в новых направлениях. Часто под вычислительной математикой понимается теория численных методов и алгоритмов решения типовых математических задач.

Все методы решения уравнений можно разделить на два класса: точные и приближенные. В *точных методах* решение получают в виде формул за конечное число операций, однако их можно использовать только для решения уравнений специального вида. В общем случае задачу можно решить только приближенно. *Приближенные методы* позволяют получить

решение в виде бесконечной последовательности, сходящейся к точному решению.

Решение сложной математической задачи на ЭВМ включает в себя необходимые этапы выбора метода решения, создания алгоритма, разработки программы и ее тестирования. После этого можно использовать разработанный пакет программ для решения нужной задачи.

В настоящее время существует большое количество специализированных и универсальных математических пакетов. Однако даже для того чтобы воспользоваться уже готовой программой, нужно иметь представление о существующих методах решения, их преимуществах, недостатках и особенностях использования.

В учебном пособии рассматриваются методы вычислительной математики – численные методы решения задач. Под *численными (вычислительными) методами* будем понимать методы вычислений, а также алгоритмы решения и исследования конкретных математических задач.

Решение любой реальной задачи можно разделить на следующие *этапы моделирования*.

1. Общая постановка задачи.
2. Построение физической (экономической, социальной и т.д.) модели процесса или явления, выбор физического приближения.
3. Математическая постановка задачи, исследование влияния наиболее значимых физических факторов на процесс и построение математической модели, т.е. математическое описание физического процесса с помощью математических уравнений и соотношений, связывающих эти факторы.
4. Выбор известных или разработка новых математических методов решения и исследования математических уравнений и соотношений, составляющих математическую модель.
5. Выбор или разработка численных методов, вычислительных алгоритмов и программ для решения математических уравнений и соотношений.
6. Численные решения (получение результата).
7. Анализ полученных численных результатов, сравнение их с результатами модельного эксперимента.

В случае несовпадения результатов численного решения и модельного эксперимента необходимо проверить и уточнить математическую модель, т.е. пройти все этапы – пункты с 1 по 7: правильно ли поставлена задача, достаточно ли исходных данных, не противоречат ли они друг другу, существует ли решение поставленной задачи и единственно ли оно и т.д.

Математическое моделирование можно рассматривать не только как приложение к определенной науке, но и как самостоятельный метод познания, имеющий свою специфику и методологию.

Математическое и компьютерное моделирование, как основа имитационного моделирования, моделей визуализации и пр., помогает развитию

абстрактных представлений об изучаемом процессе или явлении и позволяет глубже и правильнее реализовывать принципы фундаментальных исследований сложных научных и научно-технических проблем.

Создание компьютерных моделей – сложный и высокотехнологичный процесс. В большинстве случаев компьютерная модель строится на основе известных естественнонаучных законов и развитых математических теорий. Математическое описание в компьютерном моделировании играет зачастую ключевую роль. Таким образом, прослеживается диалектическая связь между математическим и компьютерным моделированием.

Симбиоз математического и компьютерного моделирования можно назвать *информационным моделированием*, включающим в себя классическое определение информатизации как средства переработки, передачи и хранения информации. Информационное моделирование можно представить в виде схемы (рис.1). Каждая из составляющих такой структурной схемы имеет свои индивидуальные принципы разработки и свойства. Но только собранные все вместе блоки информационной модели становятся мощным инструментом исследования объекта.

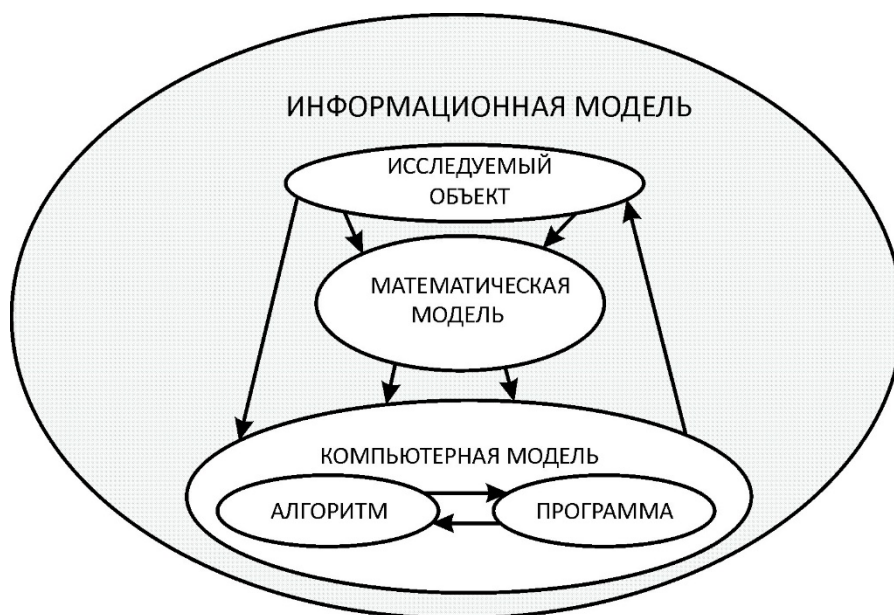


Рис. 1. Общая схема информационного моделирования

Структура каждого блока сама по себе достаточно сложна и требует отдельного рассмотрения. Остановимся на наиболее общих характеристиках составляющих информационной модели.

*Объект* – это некоторая совокупность естественнонаучных, технических, экономических, социальных и пр. знаний об изучаемом феномене, необходимая для построения *математической модели*, то есть некоторой абстракции реального явления, записанной в математических терминах и символах. Математическая модель при этом должна достаточно полно учитывать влияние основных факторов, определяющих объект, и связей

между этими факторами. Математическая модель лежит в основе *компьютерной модели*, которая определяется разработанным по математической модели *алгоритмом* и его компьютерной реализацией – *программой*.

*Алгоритм* – это формальное описание последовательности стандартных операций, которые необходимо выполнить над исходными данными для решения задачи из конкретного класса. Среди главных свойств алгоритма можно выделить следующие. Свойство *определенности*, т.е. понятности для исполнителя, благодаря чему выполнение алгоритма носит «механический» характер. Свойство *результативности*, что дает возможность получить решение задачи за конечное число шагов. И, наконец, свойство *независимости от исходных данных*, что позволяет использовать алгоритм не для одной конкретной задачи, а для класса определенных задач, для решения которых разработан алгоритм.

*Программа* – это компьютерная реализация алгоритма на языке программирования, то есть перевод алгоритма на язык символов, понятный для ЭВМ. Современные компьютерные и программные средства дают возможность осуществлять не только программирование математического решения задачи, но и статическую и динамическую визуализацию решения, что позволяет глубже проникать в причины и следствия рассматриваемого явления и, следовательно, более эффективно управлять этим явлением или процессом.

При этом адекватность математической модели реальному объекту относительна и имеет границы применимости. То есть триада «модель – алгоритм – программа» должна быть дополнена этапами «анализ результатов – уточнение модели». Если результат вычислительного (компьютерного) эксперимента верно отражает поведение объекта, то можно проанализировать его поведение в разных, даже экстремальных условиях, выбирая наиболее выгодные и удобные параметры.

Использование ЭВМ выдвигает дополнительные требования к алгоритму нахождения как точного, так и приближенного решения: он должен быть устойчивым, реализуемым и экономичным. *Устойчивость* означает, что малые погрешности, внесенные в процесс решения, не приводят к большим ошибкам в конечном результате. Погрешности возникают из-за неточного задания исходных данных (неустраняемые погрешности), из-за округления чисел, которое всегда имеет место при компьютерных расчетах, а также связаны с точностью используемого метода. *Реализуемость* алгоритма означает, что решение может быть получено за допустимое время. При этом надо иметь в виду, что время приближенного решения зависит от заданной точности решения. *Экономичным* называется алгоритм, который позволяет получить решение с заданной точностью за минимальное количество операций и, следовательно, за минимальное расчетное время.



Методом исследования математической модели является *вычислительный эксперимент*, то есть исследование естественнонаучных, технологических или технических проблем средствами вычислительной математики. Вычислительный эксперимент важен в условиях, когда проведение натурального (физического) эксперимента опасно, дорого или невозможно. Вычислительный эксперимент не отвергает традиционных математических методов исследования, он активно использует их, реализуя в виде комплексов компьютерных программ.

Компьютерный эксперимент позволяет получать многочисленные следствия, которые требуют больших расчетов. Он дает возможность получить ответ в наглядной форме, позволяет быстро анализировать варианты, проверять предположения и гипотезы.

Численные методы и вычислительные алгоритмы являются важнейшими элементами при проведении вычислительных экспериментов.

Численные методы решения задач используются в случаях, когда не удается найти точное решение возникшей математической задачи. Чаще всего это происходит из-за того, что искомое решение обычно не выражается в привычных для нас элементарных или других функциях. Вычислительные методы позволяют свести решение таких задач к арифметическим и некоторым логическим действиям, т.е. к действиям, которые выполняет компьютер.

Численные методы являются инструментом математического моделирования и используются как на стадии построения модели, так и на стадии решения задачи на компьютере.

Именно от вычислительных методов зачастую зависит не только результат решения задачи, но и сама возможность ее решения. Это связано с тем, что вычислительные методы чаще всего имеют итерационный характер и требуют определенных условий сходимости, при несоблюдении которых результат не будет достигнут.

# 1. ВВЕДЕНИЕ В ВЫЧИСЛИТЕЛЬНУЮ МАТЕМАТИКУ

## 1.1. Общая характеристика численных методов

Методы решения задач принято разделять на аналитические, численные и графические.

*Аналитические методы* представляют собой различные символьные преобразования, приводящие к формулам, выражающим неизвестные величины через известные.

*Графические методы* основаны на графических построениях и являются наиболее наглядными.

*Численные методы* основаны на построении конечной последовательности арифметических действий над числами.

Общим свойством всех численных методов является сведение непрерывной математической задачи к ее конечномерному аналогу. Это достигается путем дискретизации исходной задачи, т.е. переходом от функций непрерывного аргумента к функциям дискретного аргумента.

Решение дискретной задачи принимается за некоторое приближение решения исходной задачи.

Пример 1.

Простейшая задача численного интегрирования.

$$\mathfrak{I} = \int_0^1 f(x) dx = \sum_{i=0}^{N-1} \int_{\frac{i}{N}}^{\frac{i+1}{N}} f(x) dx \approx \sum_{i=0}^{N-1} \int_{\frac{i}{N}}^{\frac{i+1}{N}} f\left(\frac{i}{N}\right) dx = \sum_{i=0}^{N-1} f\left(\frac{i}{N}\right) \cdot \frac{1}{N} = \mathfrak{I}_N.$$

В общем случае  $\mathfrak{I}_N = \sum_{k=0}^{N-1} C_k f(x_k)$ ,  $C_k > 0$ ,  $\sum_{k=0}^{N-1} C_k = 1$  – общая формула

для построения метода численного интегрирования.

Очевидно, что при дискретизации задачи возникает погрешность, обусловленная тремя основными причинами:

1) неточным заданием входных данных (например, неточными начальными условиями, неточными значениями коэффициентов моделирующих уравнений и т.п.) – так называемая *неустраняемая погрешность*;

2) заменой непрерывной задачи дискретным аналогом – *погрешность метода*;

3) округлениями при решении задачи на компьютере в силу конечности числа знаков в представлении десятичных чисел, что обусловлено конечностью разрядной сетки компьютера, – *погрешность округления*.

Для контроля погрешности решения задачи возникает необходимость определять *точность* алгоритма, т.е. возможность решать конкретную

вычислительную задачу с требуемой точностью  $\varepsilon$  за конечное число шагов (итераций) –  $Q(\varepsilon)$ . Так, для примера 1 должны выполняться условия:

$$|\mathfrak{S} - \mathfrak{S}_n| < \varepsilon, \quad Q(\varepsilon) = n.$$

Если алгоритм позволяет решать задачу за допустимое число действий, то он называется *реализуемым*.

Алгоритм, позволяющий решать задачу за наименьшее компьютерное время по сравнению с эквивалентными по точности другими алгоритмами, называется *экономичным*.

Вычислительный алгоритм должен быть составлен таким образом, чтобы избежать аварийного останова компьютера в процессе вычисления.

## 1.2. Приближенные числа и действия над ними

*Приближенным* числом  $Q$  называют число, незначительно отличающееся от точного  $Q_0$  и заменяющее последнее в вычислениях. При этом возникает погрешность  $\Delta Q = Q_0 - Q$ .

*Абсолютной погрешностью*  $\Delta$  приближенного числа  $Q$  называют величину  $\Delta = |Q_0 - Q|$ . Под *предельной* абсолютной погрешностью понимают всякое число  $\Delta_Q$ , удовлетворяющее условию  $\Delta < \Delta_Q$ , т.е.  $-\Delta_Q \leq Q_0 - Q \leq \Delta_Q$  или  $Q_0 = Q \pm \Delta_Q$ .

Однако абсолютная погрешность недостаточна для характеристики точности измерения или вычисления. Так, при двух измерениях длины  $l_1 = (100,8 \pm 0,1)$  м и  $l_2 = (5,2 \pm 0,1)$  м результат первого измерения точнее. Поэтому часто используется *относительная* погрешность  $\delta$  приближенного числа  $Q$ , которая определяется равенством

$$\delta = \frac{\Delta}{|Q|} \quad (Q \neq 0).$$

В приведенном примере

$$\delta_{l_1} = \frac{\Delta l_1}{l_1} = \frac{0,1}{100,8} = 0,00099; \quad \delta_{l_2} = \frac{\Delta l_2}{l_2} = \frac{0,1}{5,2} = 0,019.$$

Очевидно, погрешность  $\delta$  так же, как и значения  $\Delta$  и  $Q_0$ , не всегда можно определить. Поэтому на практике в основном используют предельную относительную погрешность  $\delta_Q$ , определяемую из условия  $\delta < \delta_Q$ .

Если  $Q \neq 0$ , полагают, что для  $\Delta_Q$  и  $\delta_Q$  верны соотношения

$$\delta_Q = \frac{\Delta_Q}{|Q|}, \quad \Delta_Q = |Q| \delta_Q.$$

Запись приближенных чисел должна подчиняться определенным правилам, связанным с понятием о верных цифрах.

$k$ -тая цифра приближенного числа  $Q$  называется *верной*, если абсолютная погрешность  $\Delta$  этого числа не превосходит половины единицы  $k$ -го разряда, т.е. когда  $\Delta = |Q_0 - Q| \leq 0,5 \cdot 10^{m-k+1}$ . В противном случае цифру  $k$ -го разряда называют *сомнительной*.

*Значащими* цифрами приближенного числа  $Q$  называют все его верные цифры, начиная от первой слева, отличной от нуля, до первой сомнительной цифры включительно. Все остальные цифры называются *незначащими*. Записывая конечные результаты приближенных вычислений, незначащие цифры числа отбрасывают.

**Пример 2.**

Значащими цифрами в числах  $Q = 0,00\underline{2016}$  и  $Q = 0,00\underline{201600}$  являются подчеркнутые цифры. При этом если  $\Delta_Q = 0,00000001$ , то верными являются следующие подчеркнутые цифры:  $Q = 0,00\underline{201600}$ .

Приближенное положительное число

$$Q = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \dots + \alpha_n \cdot 10^{m-n+1},$$

где  $m$  – число разрядов  $Q$ ;  $n$  – число верных знаков в  $Q$ ;  $\alpha_k$  – числа от 0 до 9, причем  $\alpha_1 \neq 0$ .

На практике при выполнении вычислений возникает необходимость округлять числа. Число  $Q$  округляют к числу  $Q_1$  до  $n$  значащих цифр так, чтобы погрешность округления не превосходила половины единицы десятичного разряда, который не отбрасывается, т.е. чтобы выполнялось условие  $|Q_1 - Q| \leq 0,5 \cdot 10^{m-n+1}$ . Таким образом, если при округлении числа  $Q$  до  $n$  значащих цифр отбрасывают меньше, чем  $0,5 \cdot 10^{m-n+1}$ , то в числе сохраняются без изменения все первые  $n$  цифр числа  $Q$ . Если отбрасывают не меньше, чем  $0,5 \cdot 10^{m-n+1}$ , то в числе  $Q$  последнюю цифру берут на единицу больше, чем в числе  $Q_1$ .

Это правило называют *правилом дополнения при округлении*: если первая отбрасываемая цифра  $\geq 5$ , то последняя сохраняемая цифра увеличивается на 1.

**Пример 3.**

Рассмотрим число  $\pi = 3,14159\dots$ . Числа 3,1416; 3,142; 3,14 – значения  $\pi$ , округленные соответственно до пяти, четырех, трех значащих цифр с абсолютными погрешностями, меньшими соответственно  $0,5 \cdot 10^{-4}$ ,  $0,5 \cdot 10^{-3}$  и  $0,5 \cdot 10^{-2}$ .

Погрешность, возникающую при округлении числа, следует прибавить к первоначальной абсолютной погрешности числа, затем полученную абсолютную погрешность округлить в сторону увеличения.

Если положительное приближенное число  $Q$  имеет  $n$  верных значащих цифр, то его относительная погрешность  $\delta$  удовлетворяет соотношению

$$\delta \leq \frac{1}{\infty_1} \left( \frac{1}{10} \right)^{n-1},$$

где  $\infty_1$  – первая значащая цифра приближенного числа  $Q$ .

Абсолютные и относительные погрешности записывают не более чем с двумя-тремя значащими цифрами. В приближенном числе  $Q$  не следует сохранять те разряды, которые подвергаются округлению в его абсолютной погрешности  $\Delta Q$ .

Относительная погрешность обычно выражается в процентах. Ориентировочно можно считать, что наличие в приближенном числе  $Q$  только одного верного знака соответствует  $\delta$  порядка 10%, двух верных знаков – порядка 1%, трех верных знаков – порядка 0,1% и т.д.

Рассмотрим  $n$  приближенных чисел  $Q_1, Q_2, \dots, Q_n$  с предельными абсолютными погрешностями  $\Delta Q_1, \Delta Q_2, \dots, \Delta Q_n$  и предельными относительными погрешностями  $\delta_1, \delta_2, \dots, \delta_n$ .

*Погрешность суммы.*

Если  $S = Q_1 + Q_2 + \dots + Q_n$ , то  $\Delta S = \Delta Q_1 + \Delta Q_2 + \dots + \Delta Q_n$  – абсолютная погрешность суммы равна сумме абсолютных погрешностей слагаемых.

Относительная погрешность суммы  $\delta_s = \frac{\Delta S}{S}$ , ( $\Delta Q_n \geq Q$ ), причем  $\gamma \leq \delta_s \leq r$ , где  $\gamma = \min[\delta_{Q_1}, \delta_{Q_2}, \dots, \delta_{Q_n}]$ ,  $r = \max[\delta_{Q_1}, \delta_{Q_2}, \dots, \delta_{Q_n}]$ .

Если среди слагаемых имеется число, абсолютная погрешность которого значительно превосходит абсолютные погрешности остальных слагаемых, то абсолютная погрешность суммы считается равной этой наибольшей погрешности.

**Пример 4.**

Найти сумму приближенных чисел 0,3; 5,2; 0,541; 2,1663; 11,80561, считая в них все знаки верными, т.е. что абсолютная погрешность каждого слагаемого не превосходит половины единицы младшего оставленного разряда. Оценить относительную погрешность суммы.

**Решение.**

Наибольшую абсолютную погрешность  $\Delta=0,05$  имеют два числа: 0,3 и 5,2. Поэтому можно считать, что абсолютная погрешность суммы равна  $2\Delta=0,10$ . Так как количество слагаемых невелико, в расчетах сохраняем только один запасной знак, округляя слагаемые до 0,01 по правилу дополнения:

$$\begin{array}{r}
0,3 \\
5,2 \\
+ 0,54 \\
2,17 \\
11,81 \\
S = 20,02
\end{array}$$

В окончательном результате один знак отбрасываем:  $S = 20,0$ .

Относительная погрешность суммы

$$\delta_s = \frac{\Delta S}{S} = \frac{0,10}{20,0} = 0,005 = 0,5\%.$$

*Погрешность разности.*

Абсолютная погрешность разности равна сумме абсолютных погрешностей уменьшаемого и вычитаемого.

$$\text{Если } t = Q_1 - Q_2, \text{ то } \Delta t = \Delta Q_1 + \Delta Q_2, \quad \delta_t = \frac{\Delta Q_1 + \Delta Q_2}{|t|}.$$

При умножении и делении приближенных чисел складываются их относительные (а не абсолютные) погрешности. Абсолютные погрешности произведения и частного вычисляются по их относительной погрешности.

*Погрешность произведения.*

Если  $\rho = Q_1 Q_2 \dots Q_n$ , то  $\Delta \rho = \delta_\rho |\rho|$ ,  $\delta_\rho = \delta_{Q_1} + \delta_{Q_2} + \dots + \delta_{Q_n}$ . Если  $q = kQ$ , где  $k$  – точное число, то  $\delta_q = \delta_a$ ,  $\Delta q = |k| \Delta a$ .

Для упрощения вычислений рекомендуется перед умножением уравнивать (путем округления) число значащих цифр сомножителей. Суммарная погрешность результата от этого почти не изменится. В произведении обычно берут столько значащих цифр, сколько их было в каждом из сомножителей (после уравнивания).

*Погрешность частного.*

$$\text{Если } r = \frac{a_1}{a_2} \quad (a_1, a_2 > 0), \text{ то } \delta_r = \delta_{a_1} + \delta_{a_2}.$$

*Погрешность степени.*

Относительная погрешность  $m$ -й степени числа в  $m$  раз больше относительной погрешности самого числа.

$$\text{Если } u = a^k, \text{ то } \delta_u = k \delta_a.$$

*Погрешность корня.*

Относительная погрешность корня  $m$ -й степени в  $m$  раз меньше относительной погрешности подкоренного числа.

$$\text{Если } v = \sqrt[m]{a}, \text{ то } \delta_v = \frac{1}{m} \delta_a.$$

Пример 5.

Рассчитать приближённое значение выражения

$$y(x_1, x_2, x_3) = \frac{\sqrt{x_2 - x_1 + x_1 \cdot x_3}}{x_3^2 - x_1 \cdot x_2},$$

где  $x_1 = 0,74$ ;  $x_2 = 0,853$ ;  $x_3 = 0,9224$ , все указанные знаки в числах верные.

Решение.

Расчеты проведем с использованием *MS Excel*. В ячейки **A4–C4** введем значения  $x_i$ ,  $i = 1, 2, 3$  (рис. 2).

Наибольшую абсолютную погрешность  $\Delta = 0,01$  имеет число  $x_1$ , следовательно, в результате вычисления  $y$  верными будут только 2 знака после запятой. Для того чтобы погрешность в результате не перешла и на указанные знаки, расчёт выполняется с 1-2 запасными знаками после запятой.

Исходя из указанного замечания, расчёт точных чисел будем выполнять с 1 запасным знаком, то есть с 3-мя знаками после запятой, а погрешности ввиду малости их значений – с 2-мя запасными знаками, то есть с 4-мя знаками после запятой.

Учитывая, что все указанные знаки в числах верные, значения абсолютных погрешностей принимаем равными:  $\Delta x_1 = 0,01$ ;  $\Delta x_2 = 0,001$ ;  $\Delta x_3 = 0,0001$ , что соответствует ячейкам **E4–G4**.

Тогда относительные погрешности  $\delta x_i = \frac{\Delta x_i}{x_i}$  указанных величин составят:  $\delta x_1 = 0,0135$ ;  $\delta x_2 = 0,0012$ ;  $\delta x_3 = 0,0001$  (ячейки **I4–K4**).

Заданное выражение разбивается на части согласно рангам действий, и в соответствии с ними выполняется расчёт промежуточных значений и их погрешностей.

На рис. 2, а представлены формулы, 2, б – результаты вычислений.

*Расчёт числителя*

Подкоренное выражение:  $x_2 - x_1 = 0,853 - 0,74 = 0,113$  (ячейка **A9**), абсолютная погрешность разности определяется как сумма абсолютных погрешностей уменьшаемого и вычитаемого (ячейка **B9**):

$$\Delta(x_2 - x_1) = \Delta x_2 + \Delta x_1 = 0,001 + 0,01 = 0,011.$$

Относительную погрешность разности (ячейка **C9**) определим по формуле

$$\delta(x_2 - x_1) = \frac{\Delta(x_2 - x_1)}{\Delta x_2 + \Delta x_1} = \frac{0,011}{0,113} = 0,0973.$$

a)

Погрешности_формулы											
A	B	C	D	E	F	G	H	I	J	K	L
$y(x_1, x_2, x_3) = \frac{\sqrt{x_2 - x_1 + x_1 \cdot x_3}}{x_3^2 - x_1 \cdot x_2}$											
Исходные данные			Погрешности исходных данных								
x1	x2	x3		$\Delta_1$	$\Delta_2$	$\Delta_3$		$\delta_1$	$\delta_2$	$\delta_3$	
0,74	0,853	0,9224		0,01	0,001	0,0001		=E4/A4	=F4/B4	=G4/C4	
Расчет погрешностей числителя											
Подкоренное выражение			Корень			Произведение			Числитель		
$x_2 - x_1$	$\Delta$	$\delta$		$\delta$	$\Delta$	$x_1 \cdot x_3$	$\delta$	$\Delta$		$\Delta$	$\delta$
=B4-A4	=F4+E4	=B9/A9	=КОРЕНЬ(A9)	=C9/2	=E9*D9	=A4*C4	=I4+K4	=H9*G9	=D9+G9	=F9+I9	=K9/J9
Расчет погрешностей знаменателя											
Степенное выражение			Произведение			Знаменатель					
$x_3^2$	$\delta$	$\Delta$	$x_1 \cdot x_2$	$\delta$	$\Delta$		$\Delta$	$\delta$			
=C4^2	=2*K4	=B14*A14	=A4*B4	=I4+J4	=E14*D14	=A14-D14	=C14+F14	=H14/G14			
Расчет погрешностей функции y											
y	$\delta$	$\Delta$									
=J9/G14	=L9+I14	=B18*A18									

б)

Пример погрешности											
A	B	C	D	E	F	G	H	I	J	K	L
$y = \frac{\sqrt{x_2 - x_1 + x_1 \cdot x_3}}{x_3^2 - x_1 \cdot x_2}$											
Исходные данные			Погрешности исходных данных								
$x_1$	$x_2$	$x_3$		$\Delta_1$	$\Delta_2$	$\Delta_3$		$\delta_1$	$\delta_2$	$\delta_3$	
0,74	0,853	0,9224		0,01	0,001	0,0001		0,0135	0,0012	0,0001	
Расчет погрешностей числителя											
Подкоренное выражение			Корень			Произведение			Числитель		
$x_2 - x_1$	$\Delta$	$\delta$		$\delta$	$\Delta$	$x_1 \cdot x_3$	$\delta$	$\Delta$		$\Delta$	$\delta$
0,113	0,011	0,0973	0,3362	0,0487	0,0164	0,6826	0,0136	0,0093	1,0187	0,0257	0,0252
Расчет погрешностей знаменателя											
Степенное выражение			Произведение			Знаменатель					
$x_3^2$	$\delta$	$\Delta$	$x_1 \cdot x_2$	$\delta$	$\Delta$		$\Delta$	$\delta$			
0,8508	0,0002	0,0002	0,6312	0,0147	0,0093	0,2196	0,0095	0,0431			
Расчет погрешностей функции y											
y	$\delta$	$\Delta$									
4,6390	0,0682	0,3166									

Рис. 2. Пример приближенного вычисления функции в MS Excel



Вызвав из вкладки *Математические* стандартную функцию *КОРЕНЬ( )*, вычислим в ячейке **D9** квадратный корень, определим его относительную (ячейка **E9**) и абсолютную (ячейка **F9**) погрешности по формулам:

$$\delta(\sqrt{x_2 - x_1}) = \frac{\delta(x_2 - x_1)}{2} = \frac{0,0973}{2} = 0,0487;$$

$$\Delta(\sqrt{x_2 - x_1}) = \delta(\sqrt{x_2 - x_1}) \cdot \sqrt{x_2 - x_1} = 0,0487 \cdot 0,3362 = 0,0164.$$

Вычислив в ячейке **G9** произведение  $x_1 \cdot x_3 = 0,6826$ , определим его относительную (ячейка **H9**) и абсолютную (ячейка **I9**) погрешности по формулам:

$$\delta(x_1 \cdot x_3) = \delta x_1 + \delta x_3 = \mathbf{0,0135} + \mathbf{0,0001} = \mathbf{0,0136};$$

$$\Delta(x_1 \cdot x_3) = \delta(x_1 \cdot x_3) \cdot x_1 \cdot x_3 = 0,0136 \cdot 0,6826 = 0,0093.$$

Числитель представляет собой сумму уже вычисленных квадратного корня и произведения, получим его в ячейке **J9** и определим его абсолютную погрешность, как сумму абсолютных погрешностей слагаемых (ячейка **K9**), и относительную погрешность, поделив абсолютную погрешность числителя на его значение (ячейка **L9**):

$$\Delta_{\text{числ}} = 0,0064 + 0,0093 = 0,0257; \quad \delta_{\text{числ}} = \frac{0,0257}{1,0187} = 0,0252..$$

#### *Расчёт знаменателя*

Степенное выражение  $x_3^2$  вычисляем в ячейке **A14**, его относительную погрешность – в ячейке **B14** по формуле

$$\delta(x_3^2) = 2 \cdot \delta(x_3) = 2 \cdot 0,0011 = 0,0022,$$

а абсолютную погрешность  $\Delta(x_3^2) = 0,0022 \cdot 0,8508 = 0,0002$  – в ячейке **C14**.

В ячейках **D14–F14** вычисляем произведение  $x_1 \cdot x_2$  и его погрешности, а в ячейках **G14–I14** вычисляем знаменатель, как разность степени и произведения  $x_3^2 - x_1 \cdot x_2$ , а также погрешности знаменателя.

#### *Расчёт исходной функции*

Вычисление значения искомой функции  $y$  производится в ячейке **A18** как частное от деления значения числителя на значение знаменателя, его относительную погрешность – в ячейке **B14** по формуле

$$\delta(y) = \delta_{\text{числ}} + \delta_{\text{знам}} = 0,0252 + 0,0431 = 0,0683;$$

а абсолютную погрешность  $\Delta(y) = 0,0683 \cdot 4,6390 = 0,3166$  – в ячейке **C18**.

Округляем полученные значения функции  $y$  и ее погрешностей до указанного ранее точного числа знаков после запятой, то есть до двух, и записываем ответ в виде

$$y = y \pm \Delta(y) = 4,64 \pm 0,32.$$

### 1.3. Устойчивость вычислительного алгоритма

Если в процессе вычислений погрешности округления неограниченно возрастают при применении некоторого численного метода или алгоритма, то такой метод (алгоритм) называется *неустойчивым*.

Если погрешности не накапливаются, то метод (алгоритм) называется *устойчивым*.

Пример 6.

1. Вычислить  $y_i$ ,  $i = 1, 2, \dots, N$  по формуле  $y_{i+1} = y_i + d$ . Предположим, что при определении  $y_i$  вносится погрешность (округления)  $\delta_i$ , то есть вместо  $y_i$  получаем  $\tilde{y}_i = y_i + \delta_i$ . Тогда вместо  $y_{i+1}$  получим  $\tilde{y}_{i+1} = y_i + \delta_i + d$  или  $\tilde{y}_{i+1} = y_{i+1} + \delta_i$  – накопления погрешности не происходит; следовательно, алгоритм устойчив к ошибкам округления.

2. Вычислить  $y_{i+1} = d \cdot y_i$  ( $y_0$  и  $d$  считаются заданными). Допустим, что при определении  $y_i$  вносится погрешность  $\delta_i$ ,  $\tilde{y}_i = y_i + \delta_i$ . Тогда  $\tilde{y}_{i+1} = y_i \cdot d + \delta_i \cdot d$  или  $\tilde{y}_{i+1} = y_{i+1} + \delta_i \cdot d$ .

Таким образом, при  $d > 1$  погрешность возрастает, а значит, алгоритм не является устойчивым.

Если погрешность вычислений при применении алгоритма накапливается по экспоненциальному закону  $\exp(\delta_i)$ , то алгоритм принято называть *неустойчивым*.

Если погрешность накапливается по степенному закону  $\delta^n$ , то алгоритм называется *условно-устойчивым*.

### 1.4. Корректность постановки задачи

Важным вопросом при постановке вычислительной задачи является вопрос зависимости решения от входных данных. При этом возможны два случая:

- 1) задача *корректна*, когда реализуются следующие положения:
  - a) задача разрешима для любых допустимых входных данных;
  - b) имеется единственное решение задачи;
  - c) решение задачи непрерывно зависит от входных данных, а именно: малому изменению входных данных соответствует малое изменение решения, то есть задача устойчива по отношению к начальным условиям;
- 2) задача *некорректна*, когда хотя бы одно из положений *a*, *b*, *c* не выполняется.

Пример 7.

Оценить, насколько корректным является решение задачи численного интегрирования:

$$\mathfrak{I} = \int_0^1 f(x) dx, \quad x \in [0, 1].$$

Пусть  $\tilde{\mathfrak{I}} = \int_0^1 \tilde{f}(x) dx$ , где  $\tilde{f}(x) = f(x) + \delta(f(x))$ , т.е. подынтегральная

функция задана с некоторой погрешностью. Оценим ошибку вычисления:

$$\delta(\mathfrak{I}) = |\mathfrak{I} - \tilde{\mathfrak{I}}| = \left| \int_0^1 f(x) dx - \int_0^1 \tilde{f}(x) dx \right| = \left| \int_0^1 (f(x) - \tilde{f}(x)) dx \right| = \int_0^1 \delta(f(x)) dx.$$

Следовательно,  $|\delta(\mathfrak{I})| \leq \max_{0 \leq x \leq 1} |\delta(f(x))|$  – ошибка не превышает максимума ошибки задания подынтегральной функции  $f(x)$ . Таким образом, результат вычисления интеграла непрерывно зависит от ошибки задания исходного данного –  $f(x)$ .

Оценим ошибку численного решения задачи численного интегрирования по квадратурной формуле

$$\mathfrak{I}_N = \sum_{k=0}^{N-1} C_k f(x_k), \quad C_k > 0, \quad \sum_{k=0}^{N-1} C_k = 1.$$

В этом случае

$$\tilde{\mathfrak{I}}_N = \sum_{k=0}^{N-1} C_k \tilde{f}(x_k);$$

$$\delta(\mathfrak{I}_N) = |\mathfrak{I}_N - \tilde{\mathfrak{I}}_N| = \sum_{k=0}^{N-1} C_k |f(x_k) - \tilde{f}(x_k)| = \sum_{k=0}^{N-1} C_k \delta(f(x_k)).$$

Следовательно,

$$\delta(\mathfrak{I}_N) \leq \sum_{k=0}^{N-1} C_k \cdot \max_{0 \leq x \leq 1} \delta(f(x)) = \max_{0 \leq x \leq 1} \delta(f(x)).$$

Таким образом, задача вычисления интеграла по квадратурной формуле вполне корректна.

Пример 8.

Доказать, что задача дифференцирования не является корректной.

Покажем, что нарушается одно из перечисленных ранее условий корректности –  $a$ ,  $b$  или  $c$ . Для этого рассмотрим некоторую функцию  $u(x)$ , операция дифференцирования которой не является корректной при внесении погрешности определенного вида, а именно:

$$\delta(u(x)) = \frac{1}{N} \cdot \sin(N^2 x).$$

Тогда функция  $\tilde{u}(x)$  с внесенной погрешностью будет иметь вид

$$\tilde{u}(x) = u(x) + \frac{1}{N} \cdot \sin(N^2 x).$$

Для оценки погрешности определим расстояние между двумя функциями  $f(x)$  и  $g(x)$  в метрике  $C$ , т.е. в пространстве непрерывно дифференцируемых функций, которое задается следующим образом:

$$\rho(f(x), g(x)) = \|f(x) - g(x)\|_C = \max_x |f(x) - g(x)|, \quad \alpha \leq x \leq \beta.$$

Следовательно,

$$\|u(x) - \tilde{u}(x)\|_C = \|\delta(u(x))\|_C = \left\| \frac{1}{N} \cdot \sin(N^2 x) \right\|_C = \frac{1}{N} \leq \varepsilon \text{ при } N > \frac{1}{\varepsilon}.$$

Здесь  $\varepsilon$  – любое, наперед заданное малое число.

Для погрешности производной имеем

$$\delta(u') = u' - \tilde{u}' = \left( \frac{1}{N} \cdot \sin(N^2 x) \right)' = N \cdot \cos(N^2 x)$$

$$\|\delta(u')\|_C = \|\tilde{u}' - u'\|_C = \|N \cdot \cos(N^2 x)\|_C = N.$$

Значит, при  $N > \frac{1}{\varepsilon}$  получаем:  $\|\delta(u')\| \geq \frac{1}{\varepsilon}$ .

Таким образом, малому изменению функции соответствует как угодно большое изменение её производной. Следовательно, задача дифференцирования является некорректной. Отсюда следует, что и численное дифференцирование является некорректной задачей. Значит, при вычислении производной функции, заданной с некоторой ошибкой, необходимо согласовывать точность вычисления производной с погрешностью функции и шагом дифференцирования.

**Пример 9.**

Для приближенного вычисления квадратного корня  $x = \sqrt{a}$ ,  $a > 0$  используется итерационная схема:

$$x^{(n)} = \frac{1}{2} \left( x^{(n-1)} + \frac{a}{x^{(n-1)}} \right). \quad (1)$$

Доказать сходимость вычислительной схемы, то есть что  $x^{(n)} \rightarrow \sqrt{a}$  при  $n \rightarrow \infty$ .

**Решение.**

Выберем  $x^{(0)} = 1$ .

Оценим погрешность вычислений  $\varepsilon^{(n)}$ , учитывая, что

$$\frac{x^{(n)}}{\sqrt{a}} = 1 + \varepsilon^{(n)}. \quad (2)$$

Подставив в равенство (2) вместо  $x^{(n)}$  его значение из формулы (1), получим:

$$\frac{x^{(n)}}{\sqrt{a}} = \frac{1}{2} \left( \frac{x^{(n-1)}}{\sqrt{a}} + \frac{1}{\frac{x^{(n-1)}}{\sqrt{a}}} \right) = 1 + \varepsilon^{(n)}. \quad (3)$$

Запишем выражение (3), используя равенство (2):

$$\frac{1}{2} \left( 1 + \varepsilon^{(n-1)} + \frac{1}{1 + \varepsilon^{(n-1)}} \right) = 1 + \varepsilon^{(n)}. \quad (4)$$

Отсюда  $\varepsilon_n = \frac{1}{2} \cdot \frac{\varepsilon^{(n-1)}}{\varepsilon^{(n-1)} + 1} \cdot \varepsilon^{(n-1)}$ .

Учитывая начальные значения, имеем

$$1 + \varepsilon^{(0)} = \frac{1}{\sqrt{a}} > 0 \Rightarrow \varepsilon^{(n)} > 0 \Rightarrow \frac{\varepsilon^{(n-1)}}{1 + \varepsilon^{(n-1)}} < 1.$$

Из соотношения (4) следует, что  $\varepsilon^{(n)} < \frac{1}{2} \varepsilon^{(n-1)}$ . Отсюда видим, что погрешность  $\varepsilon^{(n)}$  убывает быстрее геометрической прогрессии со знаменателем  $\frac{1}{2}$ , то есть  $x_n \rightarrow \sqrt{a}$  при  $n \rightarrow \infty$  (что и требовалось доказать).

### Контрольные вопросы и задания

1. В каких случаях производится дискретизация непрерывной математической задачи?
2. Что представляет собой процесс моделирования?
3. Из каких этапов состоит математическое моделирование?
4. Чем отличаются численные методы решения задач от аналитических методов?
5. Какие виды погрешностей возможны в процессе вычислений и чем обусловлено их появление?
6. Чему равны абсолютные и относительные погрешности суммы, разности, произведения и частного двух приближенных чисел?
7. Всегда ли совпадает количество значащих и верных цифр приближенного числа? От чего это зависит?
8. Какие условия должны выполняться, чтобы задача была корректной?

## 2. СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Система линейных алгебраических уравнений (СЛАУ) имеет следующий вид:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2; \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m. \end{cases} \quad (5)$$

В матричной форме

$$A\mathbf{x} = \mathbf{b}, \quad (6)$$

где  $A$  – матрица коэффициентов размера  $m \times n$ ;

$\mathbf{x}$  – столбец неизвестных;

$\mathbf{b}$  – столбец свободных членов.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{pmatrix}.$$

*Решением системы* называется вектор  $\mathbf{x}$ , обращающий все уравнения системы в тождества. Система называется *совместной*, если она имеет хотя бы одно решение, и *несовместной*, если она не имеет решений. Совместная система называется *определенной*, если она имеет единственное решение, и *неопределенной*, если она имеет более одного решения.

Система называется *системой линейных однородных уравнений*, если в ней все свободные члены равны нулю:

$$\sum_{j=1}^n a_{ij}x_j = 0, \quad i = \overline{1, n}.$$

Система линейных однородных уравнений всегда совместна (имеет множество решений, одним из которых является *нуль-вектор*  $\mathbf{x} = (0, 0, \dots, 0)$ ).

### 2.1. Задачи, приводящие к СЛАУ

Рассмотрим СЛАУ в общем виде (6), когда число строк  $m$  и число столбцов  $n$  матрицы  $A$  необязательно равны между собой.

В зависимости от соотношения величин  $m$  и  $n$  приняты следующие определения:

– если  $m = n$ , то система называется *определенной*;

- если  $m < n$ , то система называется *недоопределенной*;
- если  $m > n$ , то система называется *переопределенной*;
- если  $\mathbf{b} = 0$ , то система называется *однородной*;

Заметим, что если  $m = n$  и  $\mathbf{b} = 0$ , то система имеет множество решений только тогда, когда определитель матрицы коэффициентов равен 0. В противном случае система имеет единственное нулевое решение.

Когда  $m = n$  и  $\mathbf{b} \neq 0$  (система неоднородная), то СЛАУ имеет единственное решение, если определитель матрицы  $A$  не равен нулю.

Перечислим некоторые задачи, приводящие к системам линейных уравнений.

СЛАУ возникают при:

- 1) решению различных задач естествознания (например, при изучении задач о равновесии сил в механике или изучении химических равновесий);
- 2) решении различных задач интерполяции и при приближении некоторых дискретных (таблично заданных) функций непрерывными аналогами;
- 3) минимизации функций нескольких переменных;
- 4) конечноразностных аппроксимациях систем обыкновенных дифференциальных уравнений и уравнений в частных производных;
- 5) отыскании собственных чисел и собственных векторов матриц.

## 2.2. Некоторые характерные типы матриц СЛАУ

Матрицей размера  $m \times n$  называется прямоугольная таблица чисел, содержащая  $m$  строк и  $n$  столбцов:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

или, в компактной форме:

$$A = (a_{ij}),$$

где  $a_{ij}$  – элементы матрицы. Индексы  $i, j$  означают, что элемент расположен на пересечении  $i$ -й строки и  $j$ -го столбца.

Матрица, все элементы которой равны нулю, называется нулевой:  $O = (0_{ij})$ .

Если  $m=n$ , то матрица называется квадратной, а число  $n$  – порядком матрицы. Матрица размера  $m \times 1$  называется вектор-столбцом; матрица размера  $1 \times n$  называется вектор-строкой.

Диагональ матрицы, на которой находятся элементы с одинаковыми индексами, называется главной диагональю, другая диагональ – *побочной*.

Элементы главной диагонали называются диагональными, все остальные элементы — внедиагональными.

Матрица называется *диагональной*, если все ее внедиагональные элементы равны нулю. Частным случаем квадратной диагональной матрицы является *единичная* матрица, все ее диагональные элементы равны единице:

$$E_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

*Верхней (правой) треугольной* называется матрица, все элементы которой, расположенные ниже главной диагонали, равны нулю

$$a_{ij} = 0 \text{ при } i > j.$$

Если в матрице равны нулю все элементы, расположенные выше главной диагонали

$$a_{ij} = 0 \text{ при } i < j,$$

то матрица называется *нижней (левой) треугольной*.

Матрица называется *вырожденной* (сингулярной, или особенной), если ее *определитель* (детерминант) равен нулю  $\Delta = 0$ .

Определитель — это числовая характеристика квадратной матрицы. Порядок определителя равен порядку матрицы.

*Алгебраическое дополнение*  $\Delta_{ij} = (-1)^{i+j} M_{ij}$ , где  $M_{ij}$  — *минор*, т.е. определитель матрицы  $(n-1)$ -го порядка, полученной из матрицы  $A$  удалением  $i$ -й строки и  $j$ -го столбца.

Определитель  $n$ -го порядка можно вычислить разложением по элементам  $i$ -й строки или  $j$ -го столбца:

$$\Delta_{ij} = \sum_{s=1}^n a_{is} \Delta_{is} = \sum_{s=1}^n a_{si} \Delta_{si} \quad (i = \overline{1, n}),$$

где  $\Delta_{si}$  — алгебраическое дополнение элемента  $a_{si}$ .

Замена строк столбцами называется *транспонированием* матрицы:

$$(a_{ij})^T = (a_{ji}).$$

Матрица  $A_n^{-1}$  является *обратной* к матрице  $A_n$ , если  $A_n^{-1} \times A_n = E_n$ .

*Правило определения обратной матрицы:*

- 1) элементы  $a_{ij}$  матрицы  $A_n$  заменяются их алгебраическими дополнениями  $\Delta_{ij}$ ;
- 2) матрица алгебраических дополнений транспонируется, в результате чего получается *присоединенная* или *взаимная* матрица к матрице  $A_n$ , обозначаемая  $\text{Adj}A$ ;



3) вычисляется определитель  $\Delta$  матрицы  $A_n$  и присоединенная матрица  $\text{Adj}A$  умножается на величину, обратную этому определителю.

$$A^{-1} = \frac{1}{\Delta} \begin{bmatrix} \Delta_{11} & \Delta_{21} & \dots & \Delta_{n1} \\ \Delta_{12} & \Delta_{22} & \dots & \Delta_{n2} \\ \dots & \dots & \dots & \dots \\ \Delta_{1n} & \Delta_{2n} & \dots & \Delta_{nn} \end{bmatrix}.$$

Приведем некоторые характерные типы матриц СЛАУ.

1. *Разреженная* (редкая) матрица – матрица, у которой преобладают нулевые элементы.

2. *Плотная* матрица – матрица, у которой незначительное количество нулевых элементов.

3. *Симметричная* матрица (как правило, возникает при решении дифференциальных уравнений разностными методами) – матрица, элементы которой симметричны относительно главной диагонали.

4. *Ленточная* матрица возникает при решении систем дифференциальных уравнений и характеризуется тем, что её ненулевыми элементами являются элементы, стоящие на главной диагонали и около неё.

5. Матрица с *диагональным преобладанием элементов* – матрица, у которой модуль каждого диагонального элемента не меньше суммы модулей других элементов строки или столбца, в которой(м) находится этот диагональный элемент.

6. *Ступенчатая* матрица – матрица, у которой либо все элементы равны нулю (нулевая матрица), либо выполнены следующие условия:

– если некоторая строка матрицы, отличная от первой, не является нулевой, то в начале этой строки стоит больше нулей, чем в начале предыдущей строки;

– если некоторая строка матрицы является нулевой, то и все ее последующие строки – нулевые.

7. Матрицы *высокого порядка* – матрицы со значительным числом строк и столбцов ( $n > 100$ ).

### 2.3. Некоторые свойства матриц и матричных операций

Суммой двух матриц  $A$  и  $B$  одинаковых размеров  $m \times n$  называется матрица  $C$  того же размера, элементы которой равны суммам соответствующих элементов матриц  $A$  и  $B$ :

$$c_{ij} = a_{ij} + b_{ij}, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Произведением матрицы  $A$  на число  $\alpha$  называется матрица  $C$ , элементы которой равны произведениям числа  $\alpha$  на элементы матрицы  $A$ :

$$c_{ij} = \alpha a_{ij}, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Пусть даны матрица  $A$  размера  $m \times p$  и матрица  $B$  размера  $p \times n$ . Произведением матриц  $A$  и  $B$  называется матрица  $C$  размера  $m \times n$ , элементы  $c_{ij}$  которой равны суммам произведений соответствующих элементов  $i$ -й строки матрицы  $A$  и  $j$ -го столбца матрицы  $B$ :

$$c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Степень матрицы (для квадратных матриц) определяется рекуррентной формулой

$$A^r = A \times A^{r-1}.$$

Пусть даны матрицы  $A_n$  и  $B_m$ . Матрица  $C_{n+m} = A_n \oplus B_m$  называется прямой суммой матриц  $A_n$  и  $B_m$ , если  $c_{ij} = a_{ij}$  при  $1 \leq i \leq n, 1 \leq j \leq n$ ;  $c_{kl} = b_{kl}$  при  $n \leq k \leq n+m, n \leq l \leq n+m$ ;  $c_{sr} = 0$  в остальных случаях.

При этом  $C_{n+m}$  имеет блочную структуру:

$$C_{n+m} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

Здесь блоки представляют собой матрицы меньших размеров и при выполнении операций рассматриваются как элементы исходных матриц.

Приведем некоторые свойства матричных операций.

1. Коммутативность сложения:

$$A + B = B + A.$$

2. Некоммутативность умножения (в общем случае):

$$A \times B \neq B \times A.$$

Если  $A \times B = B \times A$ , то матрицы  $A$  и  $B$  называются коммутативными. Если  $A \times B = B \times A = E$ , где  $E$  – единичная матрица, то матрицы  $A$  и  $B$  – взаимобратные.

3. Ассоциативность сложения и умножения:

$$A + (B + C) = (A + B) + C; \quad A \times (B \times C) = (A \times B) \times C.$$

4. Дистрибутивность умножения относительно сложения:

$$A \times (B + C) = A \times B + A \times C.$$

5. Обратная матрица произведения квадратных матриц равна произведению обратных матриц сомножителей, взятому в обратном порядке:

$$(A \times B)^{-1} = B^{-1} \times A^{-1}.$$

6. Транспонированная матрица произведения равна произведению транспонированных матриц сомножителей, взятому в обратном порядке:

$$(A \times B)^T = B^T \times A^T.$$

## 2.4. Собственные значения и собственные векторы матрицы

Пусть  $A_n$  – некоторая квадратная матрица размера  $n$ . Рассмотрим уравнение  $|A_n - \lambda E_n| = 0$ , называемое характеристическим, где  $\lambda$  – некоторая неизвестная величина (скаляр),  $E_n$  – единичная матрица,  $|A_n - \lambda E_n|$  – определитель матрицы  $A_n - \lambda E_n$ .

Решения этого уравнения  $\lambda_1, \lambda_2, \dots, \lambda_n$  называются *собственными значениями* (*характеристическими числами*) матрицы  $A_n$ .

Пример 10.

Дана матрица  $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ . Найти собственные числа матрицы  $A$ .

Решение:

$$\begin{aligned} & \begin{vmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{vmatrix} = 0. \\ & (1-\lambda) \cdot (4-\lambda) - 6 = 0; \\ & \lambda^2 - 5\lambda - 2 = 0; \\ & \lambda_{1,2} = \frac{5 \pm \sqrt{33}}{2}; \quad \lambda_1 \approx -0,372; \quad \lambda_2 \approx 5,372. \end{aligned}$$

Пусть  $h_i = \begin{pmatrix} h_{i1} \\ h_{i2} \end{pmatrix}$  – вектор-столбец, который получается из решения уравнения  $(A_n - \lambda_i E)h_i = 0$ , где  $\lambda_i$  – собственное значение матрицы  $A_n$ . Ненулевой вектор  $h_i$  называется *собственным вектором* матрицы  $A_n$ .

Существует теорема, из которой следует, что если все собственные значения матрицы  $A_n$  попарно различны, то соответствующие им собственные векторы этой матрицы линейно независимы. Каждому собственному значению соответствует один или несколько линейно независимых собственных векторов;  $k$ -кратному корню характеристического уравнения соответствует не более  $k$  линейно независимых собственных векторов.

Пример 11.

Найти собственные векторы матрицы  $A = \begin{pmatrix} -1 & -2 & 12 \\ 0 & 4 & 3 \\ 0 & 5 & 6 \end{pmatrix}$ .

Решение:

Из уравнения

$$\begin{vmatrix} -1-\lambda & -2 & 12 \\ 0 & 4-\lambda & 3 \\ 0 & 5 & 6-\lambda \end{vmatrix} = 0$$

определим собственные значения матрицы  $A$ :  $\lambda_1 = -1$ ,  $\lambda_2 = 1$ ,  $\lambda_3 = 9$ .

Найдем собственный вектор  $h_1$ , соответствующий собственному значению  $\lambda_1 = -1$ :

$$\begin{pmatrix} 0 & -2 & 12 \\ 0 & 5 & 3 \\ 0 & 5 & 7 \end{pmatrix} \times \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \end{pmatrix} = 0;$$

$$\begin{cases} 0 \cdot h_{11} - 2h_{12} + 12h_{13} = 0; \\ 0 \cdot h_{11} + 5h_{12} + 3h_{13} = 0; \\ 0 \cdot h_{11} + 5h_{12} + 7h_{13} = 0. \end{cases}$$

Решением этой системы является, например, вектор  $h_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ .

Найдем собственный вектор  $h_2$ , соответствующий значению  $\lambda_2 = 1$ :

$$\begin{pmatrix} -2 & -2 & 12 \\ 0 & 3 & 3 \\ 0 & 5 & 5 \end{pmatrix} \times \begin{pmatrix} h_{21} \\ h_{22} \\ h_{23} \end{pmatrix} = 0;$$

$$\begin{cases} -2h_{21} - 2h_{22} + 12h_{23} = 0; \\ 0 \cdot h_{21} + 3h_{22} + 3h_{23} = 0; \\ 0 \cdot h_{21} + 5h_{22} + 5h_{23} = 0. \end{cases} h_2 = \begin{pmatrix} 7 \\ -1 \\ 1 \end{pmatrix}.$$

Найдем собственный вектор  $h_3$ , соответствующий собственному значению  $\lambda_3 = 9$ :

$$\begin{pmatrix} -10 & -2 & 12 \\ 0 & -5 & 3 \\ 0 & 5 & -3 \end{pmatrix} \times \begin{pmatrix} h_{31} \\ h_{32} \\ h_{33} \end{pmatrix} = 0;$$

$$\begin{cases} -10h_{31} - 2h_{32} + 12h_{33} = 0; \\ 0 \cdot h_{31} - 5h_{32} + 3h_{33} = 0; \\ 0 \cdot h_{31} + 5h_{32} - 3h_{33} = 0. \end{cases} h_3 = \begin{pmatrix} 1,08 \\ 0,6 \\ 1 \end{pmatrix}.$$

## 2.5. Норма матрицы и её свойства

Рассмотрим  $n$ -мерные векторы  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$  и  $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$  из пространства  $R^n$  (множество всех  $n$ -мерных векторов с вещественными координатами).

Скалярное произведение двух векторов есть число, равное

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i.$$

Нормой (или длиной) вектора  $\|\mathbf{x}\|$  называется скалярное произведение

$$\|\mathbf{x}\| = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2} = \sqrt{(\mathbf{x}, \mathbf{x}^T)} \quad (7)$$

Вектор, норма которого равна единице, называется *единичным*.

Векторы  $\mathbf{x}$  и  $\mathbf{y}$  называются *ортогональными*, если их скалярное произведение  $(\mathbf{x}, \mathbf{y}) = 0$ .

Норма вектора удовлетворяет следующим свойствам:

1.  $\|\mathbf{x}\| \geq 0$ , причем  $\|\mathbf{0}\| = 0$ .
2.  $\|c\mathbf{x}\| = |c| \cdot \|\mathbf{x}\|$ ,  $c = \text{const}$ .
3.  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
4.  $\|\mathbf{x}^T \cdot \mathbf{y}\| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ .

Рассмотрим пространство квадратных матриц  $A = A_{n \times n}$ . В этом случае можно вычислить *норму матрицы*  $A$  по формуле

$$\|A\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}^T\|}{\|\mathbf{x}\|}. \quad (8)$$

Для нормы матрицы выполняются все перечисленные выше свойства нормы вектора и добавляются следующие:

5.  $\|A\mathbf{x}^T\| \leq \|A\| \cdot \|\mathbf{x}\|$
6.  $\|AB\| \leq \|A\| \cdot \|B\|$ .

Норма на заданном пространстве не единственна. Можно определить норму матрицы  $A$  другими способами:

1) как максимальное значение из сумм модулей элементов матрицы по строкам

$$\|A\|_1 = \max_i \sum_{j=1}^n |a_{ij}|; \quad (9)$$

2) как максимальное значение из сумм модулей элементов матрицы по столбцам

$$\|A\|_\infty = \max_j \sum_{i=1}^n |a_{ij}|. \quad (10)$$

3) как евклидову норму

$$\|A\|_2 = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}. \quad (11)$$

Пример 12.

Вычислить норму матрицы

$$A = \begin{pmatrix} 6 & 1 & -1 & 3 \\ 1 & 2 & 0 & 5 \\ 6 & 1 & 1 & 0 \\ 5 & 4 & -1 & 2 \end{pmatrix}.$$

Решение.

Воспользовавшись формулами (9) и (10), получим:

$$\begin{aligned} \|A\|_1 &= \max_i \sum_{j=1}^n |a_{ij}| = \max(6+1+1+3, 1+2+0+5, 6+1+1+0, 5+4+1+2) = \\ &= \max(11, 8, 8, 12) = 12; \end{aligned}$$

$$\begin{aligned} \|A\|_\infty &= \max_j \sum_{i=1}^n |a_{ij}| = \max(6+1+6+5, 1+2+1+4, 1+0+1+1, 3+5+0+2) = \\ &= \max(18, 8, 3, 10) = 18. \end{aligned}$$

Можно показать, что свойствам нормы удовлетворяет величина  $\mu_1$  – максимальное сингулярное число матрицы  $A \times A^T$ . То есть за норму матрицы  $A$  можно принять величину  $\|A\|_2 = \mu_1$ . Тогда норма  $\|A^{-1}\|_2 = \mu_n^{-1}$ , где  $\mu_n$  – минимальное сингулярное число матрицы  $A \times A^T$ .

Напомним, что *сингулярным числом* положительно определенной матрицы  $A$  называется арифметическое значение квадратного корня соответствующего собственного числа матрицы  $A \times A^T$ .

Матрицу  $A = [a_{ij}]_{n \times n}$  называют *положительно определенной*, если для всякого вектора  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$  при  $|\mathbf{x}| \neq 0$  будет выполняться следующее неравенство:  $\sum_{i,j=1}^n a_{ij} x_i x_j > 0$ .

Скалярное произведение  $(\mathbf{x}, A\mathbf{x}^T) = \sum_{i,j=1}^n a_{ij} x_i x_j$  называется *квадратичной*

*формой*. Квадратичная форма считается *положительно определенной*, если её значение неотрицательно для любых значений компонентов вектора  $\mathbf{x}$ .

Матрица  $A$ , соответствующая положительно определенной квадратичной форме, называется *положительно определенной*. Для того чтобы произвольная матрица  $A$  была положительно определенной, необходимо и достаточно, чтобы все главные миноры матрицы  $A$  были положительными.

Все собственные значения положительно определенной симметричной матрицы положительны.

Пример 13.

Проверим, является ли матрица  $A = \begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}$  положительно определенной. Для этого найдем соответствующую ей квадратичную форму:  
 $(\mathbf{x}, A\mathbf{x}^T) = 4x_1x_1 + 3x_1x_2 + x_2x_1 + 2x_2x_2 = 2(2x_1^2 + 2x_1x_2 + x_2^2) = 2[(x_1 + x_2)^2 + x_1^2]$ .

Это скалярное произведение положительно при любых  $x_1$  и  $x_2$ , одновременно не равных нулю.

## 2.6. Матричные преобразования

При исследовании свойств матриц, решении систем алгебраических и дифференциальных уравнений и в прочих задачах, где используется матричное исчисление, применяются различные виды преобразований матриц.

### 2.6.1. Эквивалентное преобразование

Рассматривается квадратная матрица  $A_n$ . Пусть  $P_n$  и  $Q_n$  – неособенные матрицы того же порядка.

Матрица  $\tilde{A} = P_n A_n Q_n$  называется эквивалентной матрице  $A_n$ , то есть  $A_n$  и  $\tilde{A}_n$  называются эквивалентными друг другу ( $A_n \cong \tilde{A}_n$ ).

Представим  $A_n = \begin{pmatrix} \bar{a}_1 \\ \bar{a}_2 \\ \dots \\ \bar{a}_n \end{pmatrix}$  в виде вектор-столбца вектор-строк ( $\bar{a}_i$  –  $i$ -я

вектор-строка матрицы  $A_n$ ) и рассмотрим произведение  $P_n A_n$ :

$$P_n A_n = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{pmatrix} \times \begin{pmatrix} \bar{a}_1 \\ \bar{a}_2 \\ \dots \\ \bar{a}_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n p_{1j} \bar{a}_j \\ \sum_{j=1}^n p_{2j} \bar{a}_j \\ \dots \\ \sum_{j=1}^n p_{nj} \bar{a}_j \end{pmatrix},$$

где  $\sum_{j=1}^n p_{ij} \bar{a}_j$  – сумма строк, умноженных на некоторые коэффициенты  $p_{ij}$ .

Таким образом, результирующая матрица есть не что иное, как эквивалентное преобразование строк исходной матрицы  $A_n$ .

Аналогично, рассматривая матрицу  $A_n$  как вектор-строку, состоящую из вектор-столбцов  $A_n = (\bar{a}_1^T \ \bar{a}_2^T \ \dots \ \bar{a}_n^T)$ , нетрудно получить, что  $A_n Q_n$  есть эквивалентное преобразование столбцов матрицы  $A_n$ . Таким образом, нетрудно показать, что эквивалентное преобразование матрицы сводится к последовательности элементарных преобразований следующих типов:

- 1) перестановка произвольных двух строк (столбцов);
- 2) умножение строки (столбца) на число;
- 3) сложение строк (столбцов);
- 4) прибавление к строке (столбцу) другой строки (столбца), умноженной на число.

С помощью элементарных преобразований произвольную матрицу можно привести к нормальной (или канонической) форме, которая для неособенной матрицы является единичной, а в общем случае имеет вид

$A_n = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ , где  $E_r$  – единичная матрица порядка  $r$ , а  $r = \text{rang}(A_n)$  – ранг матрицы  $A_n$ .

Напомним, что *рангом матрицы* называется наивысший из порядков миноров этой матрицы, определитель которых отличен от нуля.

Ранг матрицы равен количеству линейно независимых вектор-строк или вектор-столбцов матрицы.

На практике для нахождения ранга матрицы используют следующее утверждение: ранг матрицы равен количеству ненулевых строк после приведения матрицы к ступенчатому (треугольному) виду.

**Пример 14.**

$$P \cdot A = \begin{pmatrix} ??? \\ ??? \\ ??? \end{pmatrix} \begin{pmatrix} 123 \\ 456 \\ 031 \end{pmatrix} = \begin{pmatrix} 456 \\ 123 \\ 031 \end{pmatrix}.$$

На какую матрицу требуется умножить данную матрицу  $A$ , чтобы поменять местами первую и вторую строки?

Ответ:  $P = \begin{pmatrix} 010 \\ 100 \\ 001 \end{pmatrix}$ .

Частным случаем эквивалентного преобразования является *преобразование Гаусса – Жордана*. Такое преобразование сводится к преобразованию строк расширенной матрицы системы линейных алгебраических уравнений с целью получения единичной матрицы для первых  $n$  столбцов.



Пример 15.

$$\begin{cases} 2x_1 + x_2 + 4x_3 = 16; \\ 3x_1 + 2x_2 + x_3 = 10; \\ x_1 + 3x_2 + 3x_3 = 16. \end{cases}$$

Выполним элементарные преобразования:

- 1) разделим первую строку на число 2;
- 2) вычтем из второй строки первую, умноженную на число 3, и запишем результат на место второй строки; из третьей строки вычтем первую и результат запишем на место третьей;

3) и т.д., чтобы получить единичные диагональные элементы и нулевые внедиагональные для квадратной матрицы  $A$  коэффициентов системы.

$$\begin{aligned} \begin{pmatrix} 2 & 1 & 4 & 16 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{pmatrix} &\xrightarrow{(1)} \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & \frac{1}{2} & -5 & -14 \\ 0 & \frac{5}{2} & 1 & 8 \end{pmatrix} \xrightarrow{(3)} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 7 & 22 \\ 0 & 1 & -10 & 28 \\ 0 & 0 & 26 & 78 \end{pmatrix} \xrightarrow{(4)} \begin{pmatrix} 1 & 0 & 7 & 22 \\ 0 & 1 & -10 & 28 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{(5)} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{pmatrix}. \end{aligned}$$

В результате матрица  $A$  преобразовалась в единичную, а столбец свободных членов  $b$  – в столбец, элементы которого равны значениям искомых величин  $x_1, x_2, x_3$ , т.е.  $[A, b]$  преобразуется в  $[E, x]$ , откуда сразу получаем:  $x_1 = 1, x_2 = 2, x_3 = 3$ .

### 2.6.2. Преобразование подобия

Если  $P = Q^{-1}$ , то  $\tilde{A} = Q^{-1}AQ$ , а преобразование называется *преобразованием подобия*.

Матрицы  $A$  и  $\tilde{A}$  называются *подобными*.

Одним из свойств преобразования подобия является свойство сохранения значения определителя преобразуемой матрицы.

Действительно:

$$\det(\tilde{A}) = \det(Q^{-1} \cdot A \cdot Q) = \det(Q) \cdot \det(A) \cdot \det(Q^{-1}) = \det(A),$$

так как

$$\det(A \cdot B) = \det(A) \cdot \det(B) \text{ и } \det(A^{-1}) = \frac{1}{\det(A)}.$$

### Преобразование при помощи модальной матрицы

Пусть  $\lambda_i$  – собственные числа матрицы  $A$ ,  $h_i$  – собственные векторы матрицы  $A$ . Матрица,  $n$  столбцов которой являются собственными векторами заданной матрицы  $A$ , называется *модальной матрицей*, которую будем обозначать символом  $H$ .

Преобразование подобия  $\tilde{A} = H^{-1}AH$  приводит матрицу  $A$  к диагональной форме.

**Пример 16.**

Привести к диагональной форме матрицу

$$A = \begin{pmatrix} 3 & 2 & 2 \\ -4 & -5 & -9 \\ 4 & 6 & 10 \end{pmatrix}.$$

**Решение.**

$|A - \lambda \cdot E| = 0$  – характеристическое уравнение.

$$|A - \lambda \cdot E| = \begin{vmatrix} 3 - \lambda & 2 & 2 \\ -4 & -5 - \lambda & -9 \\ 4 & 6 & 10 - \lambda \end{vmatrix}.$$

Алгебраические дополнения элементов первой строки:

$$\Delta_{11}(\lambda) = \begin{vmatrix} -5 - \lambda & -9 \\ 6 & 10 - \lambda \end{vmatrix} = \lambda^2 - 5\lambda + 4;$$

$$\Delta_{12}(\lambda) = (-1)^{1+2} \cdot \begin{vmatrix} -4 & -9 \\ 4 & 10 - \lambda \end{vmatrix} = -4\lambda + 4;$$

$$\Delta_{13}(\lambda) = \begin{vmatrix} -4 & -5 - \lambda \\ 4 & 6 \end{vmatrix} = 4\lambda - 4.$$

Характеристический многочлен:

$$\Delta(\lambda) = (3 - \lambda)(\lambda^2 - 5\lambda + 4) - 2(4\lambda - 4) + 2 \cdot (4\lambda - 4) = -\lambda^3 + 8\lambda^2 - 19\lambda + 12,$$

решения которого:  $\lambda_1 = 1$ ,  $\lambda_2 = 3$ ,  $\lambda_3 = 4$ .

Найдем собственные векторы, соответствующие найденным собственным числам. Для этого решим последовательно три уравнения:  $(A - \lambda_i \cdot E)\mathbf{h} = \mathbf{0}$ ,  $i = 1, 2, 3$ .

**Замечание** (о решении однородных систем линейных алгебраических уравнений).

Рассмотрим однородную систему линейных алгебраических уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0; \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0; \\ \dots \dots \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = 0. \end{cases}$$

Применим к определителю матрицы коэффициентов системы разложение Лапласа по строке с номером  $i$ .

$$\sum_{j=1}^n a_{ij} \det_{ij} = 0, \quad i = 1, \dots, n, \quad \det_{ij} - \text{алгебраическое дополнение к элементу } a_{ij}.$$

Сравним это выражение с записью  $i$ -той строки системы:  $\sum_{j=1}^n a_{ij} x_j = 0$ .

Нетрудно видеть, что  $x_j = K \cdot \det_{ij}$ , где  $K$  – произвольная постоянная.

Таким образом, решение однородной системы линейных алгебраических уравнений можно выразить через соответствующие алгебраические дополнения к элементам некоторого столбца матрицы коэффициентов.

Представим каждое решение характеристического уравнения  $(A - \lambda_i \cdot E)\mathbf{h} = \mathbf{0}$ ,  $i = 1, 2, 3$ , в виде  $h_i = k_i \Delta_{ij}$ , где  $k_i \neq 0$  – произвольное число;  $\Delta_{ij}$  – алгебраические дополнения элементов какой-либо строки матрицы  $A$ . То есть собственные векторы запишем в виде:

$$h_i = k_i \begin{pmatrix} \Delta_{j1}(\lambda_i) \\ \Delta_{j2}(\lambda_i) \\ \Delta_{j3}(\lambda_i) \end{pmatrix}.$$

Следовательно, при  $\lambda_1 = 1$  получим:

$$h_1 = k_1 \begin{pmatrix} 3 \\ 2 \\ -5 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -5 \\ 2 \end{pmatrix}, \quad (\text{выбрали } k_1 = 2).$$

Аналогично:

$$\lambda_2 = 3, \quad k_2 = -\frac{1}{2}, \quad h_2 = k_2 \begin{pmatrix} -2 \\ -8 \\ 8 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ -4 \end{pmatrix};$$

$$\lambda_3 = 4, \quad k_3 = \frac{1}{12}, \quad h_3 = k_3 \begin{pmatrix} 0 \\ -12 \\ 12 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Следовательно, модальная матрица имеет вид:

$$H = \begin{pmatrix} 3 & 1 & 0 \\ -5 & 4 & -1 \\ 2 & -4 & 1 \end{pmatrix}.$$

Найдем в *MS Excel* (рис. 3) обратную к ней матрицу

$$H^{-1} = \begin{pmatrix} 0 & -0,333 & -0,333 \\ 1 & 1 & 1 \\ 4 & 4,667 & 5,667 \end{pmatrix}.$$

Для этого выделим диапазон ячеек (**I2:K4**), и, кликнув (**CL**) на кнопке **f(x)** (функция) на панели инструментов, выбираем из вкладки **Математические** функцию **Мобр**.

Переместив указатель в строку формул в конец формулы **=Мобр(E2:G4)** и нажав одновременно три клавиши **Shift+Ctrl+Enter**, получаем в выделенном диапазоне значения обратной матрицы.

Преобразование подобия  $\tilde{A} = H^{-1}AH$  позволяет получить диагональную матрицу:

$$\tilde{A} = \begin{pmatrix} 0 & -0,333 & -0,333 \\ 1 & 1 & 1 \\ 4 & 4,667 & 5,667 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 & 2 \\ -4 & -5 & -9 \\ 4 & 6 & 10 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 & 0 \\ -5 & 4 & -1 \\ 2 & -4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Для этого выделяем ячейки **A7:C9** и в ячейку **A7** вводим формулу **=Мумнож(I2:K4; A2:C4)**, выбрав функцию **Мумнож** из вкладки **Математические**. Аналогично умножаем полученное произведение на модальную матрицу.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	
1	A				H				H <sup>-1</sup>			
2		3,000	2,000	2,000		3,000	1,000	0,000		0,000	-0,333	-0,333
3		-4,000	-5,000	-9,000		-5,000	4,000	-1,000		1,000	1,000	1,000
4		4,000	6,000	10,000		2,000	-4,000	1,000		4,000	4,667	5,667
5												
6	H <sup>-1</sup> ·A				H <sup>-1</sup> ·A·H							
7		0,000	-0,333	-0,333		1,000	0,000	0,000				
8		3,000	3,000	3,000		0,000	3,000	0,000				
9		16,000	18,667	22,667		0,000	0,000	4,000				

Рис. 3. Пример преобразования при помощи модальной матрицы

### Ортогональное преобразование

Если в преобразовании подобия матрица  $Q$  – ортогональная,  $QQ^T = Q^TQ = E$ , то преобразование называется *ортогональным*.

Как было показано ранее, одним из важных свойств матрицы ортогонального преобразования является свойство сохранять скалярное произведение и длину векторов, то есть  $(Qx, Qy) = (x, y)$ ,  $\|Qx\| = \|x\|$ .

В связи с важностью понятия ортогональной матрицы и ортогонального преобразования необходимо уметь преобразовывать матрицу так, чтобы в

результате получить ортогональную. Этот вопрос непосредственно связан с ортогональным преобразованием векторов.

### Ортогонализация Грама – Шмидта

Рассмотрим следующую задачу: преобразовать два данных неколлинеарных вектора  $a$  и  $b$  во взаимно перпендикулярные векторы  $V_1$  и  $V_2$ .

Решение.

Рассмотрим вектор  $V_2 = b - p$ , где  $p = \text{пр}_a b$  – проекция вектора  $b$  на вектор  $a$ . Найденный таким образом вектор  $V_2$  перпендикулярен вектору  $a$ . Так

как  $(a, b) = |a| \cdot |b| \cdot \cos \varphi = |a| \cdot |\text{пр}_a b| \Rightarrow |p| = \frac{(a, b)}{|a|}$ , то  $V_2 = b - \frac{(a, b)}{|a|^2} \cdot a$ , следова-

тельно, можно принять  $V_1 = a$ ,  $V_2 = b - \frac{V_1^T b}{V_1^T V_1} V_1$ .

Полученные таким образом векторы  $V_1, V_2$  ортогональны друг другу, что легко проверяется умножением.

Если взять три вектора  $a, b$  и  $c$ , то их ортогонализацию можно произвести совершенно аналогично. В результате получим

$$V_1 = a, \quad V_2 = b - \frac{V_1^T b}{V_1^T V_1} V_1, \quad V_3 = c - \frac{V_1^T c}{V_1^T V_1} V_1 - \frac{V_2^T c}{V_2^T V_2} V_2.$$

Ортогональность векторов  $V_1, V_2, V_3$  проверяется непосредственно умножением.

Произвольный набор векторов  $a_1, a_2, \dots, a_n$  можно преобразовать в набор взаимно ортогональных векторов при помощи процесса Грама – Шмидта, который в общем виде представляется формулами:

$$V_1 = a_1, \quad V_i = a_i - \frac{V_1^T a_i}{|V_1|^2} V_1 - \frac{V_2^T a_i}{|V_2|^2} V_2 - \dots - \frac{V_{i-1}^T a_i}{|V_{i-1}|^2} V_{i-1}.$$

Векторы  $q_i = \frac{V_i}{|V_i|}$  образуют ортонормированную систему векторов (они взаимно ортогональны и единичной длины).

Пример 17.

Даны векторы  $a_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ ,  $a_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ ,  $a_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ . Требуется построить систе-

му ортогональных векторов  $V_1, V_2, V_3$  и ортонормированную систему векторов  $q_1, q_2, q_3$ .

Определим  $V_1, V_2, V_3$ , пользуясь полученными выше формулами:

$$V_1 = a_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad V_2 = a_2 - \frac{1}{2}V_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix},$$

$$V_3 = a_3 - \frac{1}{2}V_1 - \frac{1}{3}V_2 = \begin{pmatrix} -2/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

Определим ортонормированную систему. Для этого подсчитаем длины векторов  $V_1, V_2, V_3$ , равные соответственно  $\sqrt{2}, \sqrt{3/2}, \sqrt{4/3}$ .

$$\text{Следовательно, } q_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad q_2 = \frac{1}{\sqrt{3/2}} \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix}, \quad q_3 = \frac{1}{\sqrt{4/3}} \begin{pmatrix} -2/3 \\ 2/3 \\ 2/3 \end{pmatrix}.$$

### 2.6.3. Разложение матрицы на произведение ортогональной и треугольной

В общем случае матрица  $A_n$  с линейно независимыми столбцами (то есть невырожденная) может быть представлена в виде произведения  $A_n = Q_n R_n$ , где  $Q_n$  – ортогональная матрица, а  $R_n$  – верхняя треугольная матрица. Такое представление основано на преобразовании вектор-столбцов  $a_i$  матрицы  $A_n$  по методу Грама – Шмидта.

**Пример 18.**

Преобразуем матрицу  $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ , вектор-столбцами которой являются

векторы  $a_1, a_2, a_3$ , рассмотренные в примере 17.

Тогда из формул для  $V_1, V_2, V_3$  нетрудно получить:

$$\begin{cases} a_1 = V_1 = \sqrt{2} \cdot q_1; \\ a_2 = V_2 + (1/2)V_1 = \sqrt{1/2} q_1 + \sqrt{3/2} q_2; \\ a_3 = V_3 + (1/2)V_1 + (1/3)V_2 = \sqrt{1/2} q_1 + \sqrt{1/6} q_2 + \sqrt{4/3} q_3. \end{cases}$$

Эту систему можно записать в матричной форме:

$$(a_1, a_2, a_3) = (q_1, q_2, q_3) \times \begin{pmatrix} \sqrt{2} & \sqrt{1/2} & \sqrt{1/2} \\ 0 & \sqrt{3/2} & \sqrt{1/6} \\ 0 & 0 & \sqrt{4/3} \end{pmatrix}.$$

Отметим, что верхняя треугольная матрица является транспонированной матрицей коэффициентов записанной выше системы уравнений. Таким образом, мы представили матрицу  $A$  как произведение ортогональной и верхней треугольной матриц.

С л е д с т в и е .

Предположим, что для системы линейных алгебраических уравнений  $Ax = B$  матрица  $A$  представлена в виде произведения ортогональной и треугольной матриц  $A = QR$ . Такое представление дает возможность простого решения системы  $Ax = B$ .

Действительно,  $A^T = (QR)^T = R^T Q^T$ .

Рассмотрим произведение  $A^T \cdot Ax = A^T \cdot B$ .

После подстановки получим  $(R^T Q^T \cdot QR)x = R^T Q^T \cdot B$ .

Поскольку  $Q$  – ортогональная матрица, то  $Q^T Q = E$ . Отсюда  $R^T R x = R^T Q^T B$ . Умножив обе части системы на произведение  $(R^T R)^{-1}$ , получим:

$$(R^T R)^{-1} \cdot (R^T R)x = (R^T R)^{-1} \cdot R^T Q^T B;$$

$$x = (RR^T)^{-1} R^T Q^T B;$$

$$x = R^{-1} Q^T B;$$

$$Rx = Q^T B.$$

Так как  $R$  – треугольная матрица, то система уравнений  $Rx = Q^T B$  легко решается.

Таким образом, если известно разложение матрицы  $A$  коэффициентов некоторой системы линейных алгебраических уравнений  $Ax = B$  на произведение ортогональной и треугольной  $A = QR$ , то для решения этой системы достаточно решить систему  $Rx = Q^T B$  с треугольной матрицей коэффициентов  $R$ .

П р и м е р 19.

Решить систему уравнений:

$$\begin{cases} 3x = 1; \\ 4x + 5y = 0. \end{cases}$$

Решение.

Применяя ортогонализацию Грама – Шмидта к столбцам матрицы коэффициентов  $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$ , получим представление этой матрицы в виде произведения матриц

$$Q = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}, \quad R = \begin{pmatrix} 5 & 4 \\ 0 & 3 \end{pmatrix}.$$

Легко проверить, что  $Q \times R = A$ . Вычислим  $Q^T B$ , где  $B$  – столбец свободных членов системы  $B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ :

$$Q^T B = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix}.$$

Следовательно, система, эквивалентная заданной, имеет вид:

$$\begin{cases} 5x + 4y = 3/5; \\ 3y = -4/5. \end{cases}$$

Находим отсюда  $y = -4/15$ ,  $x = 1/3$ . Эти значения являются одновременно и решением исходной системы уравнений.

## 2.7. Методы решения СЛАУ

Все методы решения можно разделить на три группы:

1. Прямые (точные) методы, позволяющие получить точное решение после конечного числа арифметических и логических операций. Прямые методы применяются в том случае, если порядок системы невелик (до  $10^3$ ), или же если матрица коэффициентов имеет специальный вид (например, является ленточной – когда ненулевые элементы расположены лишь вблизи главной диагонали).

К прямым методам относятся:

- метод Гаусса и его модификации (в том числе метод прогонки);
- метод обратной матрицы;
- метод Крамера;
- метод  $LU$ -разложения матрицы коэффициентов.

2. Итерационные методы, позволяющие после конечного числа операций получить лишь приближенное решение (для получения точного решения требуется выполнить бесконечное число операций). Итерационные методы основаны на получении постепенно «улучшаемых» приближенных решений. Итерационные методы обычно применяют для решения систем высокого порядка ( $n \sim 10^3 \dots 10^8$ ).

К итерационным методам относятся:

- метод простых итераций;
- метод Зейделя;
- методы нижней и верхней релаксации.

3. Вероятностные методы, которые позволяют получить грубую оценку решения для систем высокого порядка ( $n \gg 10^8$ ).



Выбор метода решения зависит от многих факторов: особенностей уравнений, числа неизвестных, использования программных средств и т.д.

### 2.7.1. Прямые методы решения СЛАУ

Наиболее популярным способом решения СЛАУ является **метод Гаусса**, суть которого состоит в последовательном исключении неизвестных. Вычисления по методу Гаусса выполняют в 2 этапа: на 1-м этапе, называемом прямым ходом метода, исходную систему преобразуют к треугольному виду; на 2-м этапе, который называют обратным ходом метода, решают треугольную систему, эквивалентную исходной.

Эквивалентные преобразования не меняют решения системы и сводятся к перестановке уравнений в системе, умножению уравнения на константу, сложению уравнений, замене любого уравнения линейной комбинацией его с другими уравнениями, выражению отдельных переменных через другие и т.п.

Практически все преобразования удобно применять не к исходной системе, а к расширенной матрице – матрице коэффициентов, дополненной столбцом свободных членов.

**Метод обратной матрицы** сводится к решению матричного уравнения (6)  $A \cdot \mathbf{x} = \mathbf{b}$ . Если матрица коэффициентов  $A$  невырождена (т.е. ее определитель отличен от нуля), то решение уравнения существует и единственно. Это матричное уравнение решается умножением обеих его частей слева на обратную матрицу:  $A^{-1}A \cdot \mathbf{x} = A^{-1} \cdot \mathbf{b}$ , в результате чего получаем:  $\mathbf{x} = A^{-1} \cdot \mathbf{b}$ .

**Метод Крамера** основан на применении формул Крамера  $x_k = \frac{\Delta_k}{\Delta}$ ,  $k = 1, \dots, n$ , где  $\Delta$  – определитель матрицы коэффициентов,  $\Delta_k$  – определители матриц, полученных при замене  $k$ -го столбца матрицы коэффициентов столбцом свободных членов.

Следует заметить, что ввиду громоздкости структуры расчетов методы Крамера и обратной матрицы нецелесообразно применять в качестве методов вычислительной математики из-за значительных затрат компьютерного времени на их реализацию, а также возможного накопления погрешностей округления при вычислениях.

Основными требованиями к численному методу решения являются его экономичность и точность, т.е. минимальное число арифметических действий для отыскания приближенного решения с заданной точностью  $\varepsilon > 0$ .

**Метод решения систем с помощью LU-разложения** матриц основан на разложении матрицы коэффициентов  $A$  в произведение треугольных: нижней треугольной  $L$  и верхней треугольной  $U$ , то есть  $A = LU$ . После

разложения задача по решению системы  $A \cdot \mathbf{x} = \mathbf{b}$  сводится к решению двух треугольных систем уравнений. Действительно,  $L \cdot U \cdot \mathbf{x} = \mathbf{b}$ . Обозначив  $U \cdot \mathbf{x} = \mathbf{y}$ , получим:  $L \cdot \mathbf{y} = \mathbf{b}$ . Решая эту систему, определяем  $\mathbf{y}$  (прямой ход). Полученные значения подставляем в систему  $U \cdot \mathbf{x} = \mathbf{y}$ , откуда находим требуемые значения  $\mathbf{x}$  (обратный ход).

Приведем компактную схему определения  $LU$ -разложения матрицы  $A$ .

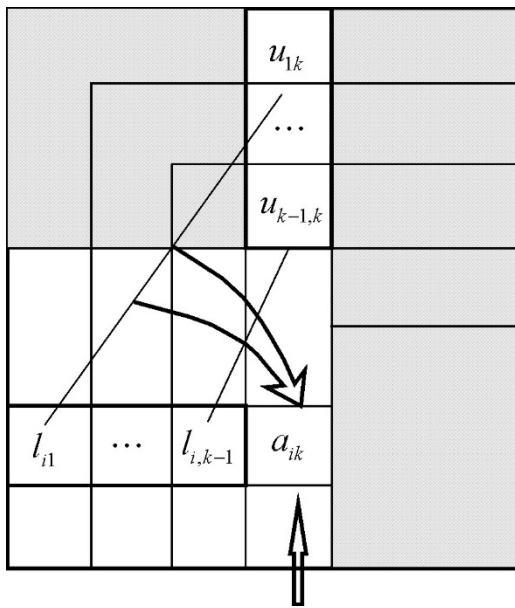
Если  $a_{11} \neq 0$ ,  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ , ...,  $|A| \neq 0$ , то есть все главные миноры не

равны 0, то элементы матриц  $L$  и  $U$  можно определить по следующей схеме (схеме Краута, рис. 4):

$$u_{ii} = 1; l_{i1} = a_{i1}; u_{1j} = \frac{a_{1j}}{l_{11}};$$

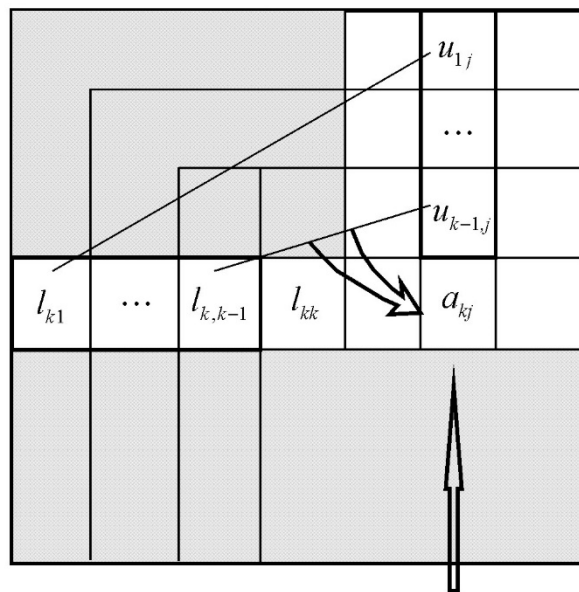
$$l_{ij} = a_{ij} - \sum_{s=1}^{j-1} l_{is} \cdot u_{sj} \quad \text{при } i \geq j; \quad (12)$$

$$u_{ij} = \frac{1}{l_{ii}} \left( a_{ij} - \sum_{s=1}^{i-1} l_{is} \cdot u_{sj} \right) \quad \text{при } j > i.$$



$$l_{ik} = a_{ik} - \sum_{s=1}^{k-1} l_{is} u_{sk}$$

( $k$ -й столбец нижней треугольной матрицы)



$$u_{kj} = \frac{1}{l_{kk}} \left( a_{kj} - \sum_{s=1}^{k-1} l_{ks} u_{sj} \right)$$

( $k$ -я строка верхней треугольной матрицы)

Рис. 4. Схема Краута для  $LU$ -разложения матрицы

Результат представления матрицы  $A$  в виде произведения двух треугольных матриц (операции *факторизации* или *декомпозиции*) в методе  $LU$ -разложения удобно хранить в одной матрице следующей структуры:

$$\begin{pmatrix} l_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ l_{21} & l_{22} & u_{23} & \dots & u_{2n} \\ l_{31} & l_{32} & l_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{pmatrix}.$$

**Пример 20.**

Решить прямыми методами с использованием средств *MS Excel* систему уравнений

$$\begin{cases} 2x_1 + x_2 + 4x_3 = 1; \\ 3x_1 + 2x_2 + x_3 = -1; \\ x_1 + 3x_2 + 3x_3 = 1. \end{cases}$$

**Решение.**

*Метод Крамера*

1. Запускаем *MS Excel*. При решении СЛАУ методом Крамера используем функцию *Монред* из вкладки *Математические* для нахождения определителей четырех матриц, которые заносим последовательно в следующие ячейки:

$$\text{в ячейки } A1:C3 \begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix}; \quad \text{в ячейки } A5:C7 \begin{pmatrix} 1 & 1 & 4 \\ -1 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix};$$

$$\text{в ячейки } A9:C11 \begin{pmatrix} 2 & 1 & 4 \\ 3 & -1 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \text{в ячейки } A13:C15 \begin{pmatrix} 2 & 1 & 1 \\ 3 & 2 & -1 \\ 1 & 3 & 1 \end{pmatrix}.$$

2. В ячейку **D4** вводим формулу =*Монред*(**A1:C3**), для этого на панели инструментов **CL** (кликнув) на кнопке **f(x)** (функция), выбираем из вкладки *Математические* функцию *Монред*.

3. В ячейку **D8** вводим формулу =*Монред*(**A5:C7**).

4. В ячейку **D12** вводим формулу =*Монред*(**A9:C11**).

5. В ячейку **D16** вводим формулу =*Монред*(**A13:C15**).

6. Введя в ячейки **A17**, **B17**, **C17** соответственно формулы: =**D8/D4**, =**D12/D4**, =**D16/D4**, получим значения вектора неизвестных  $x$  (рис. 5).

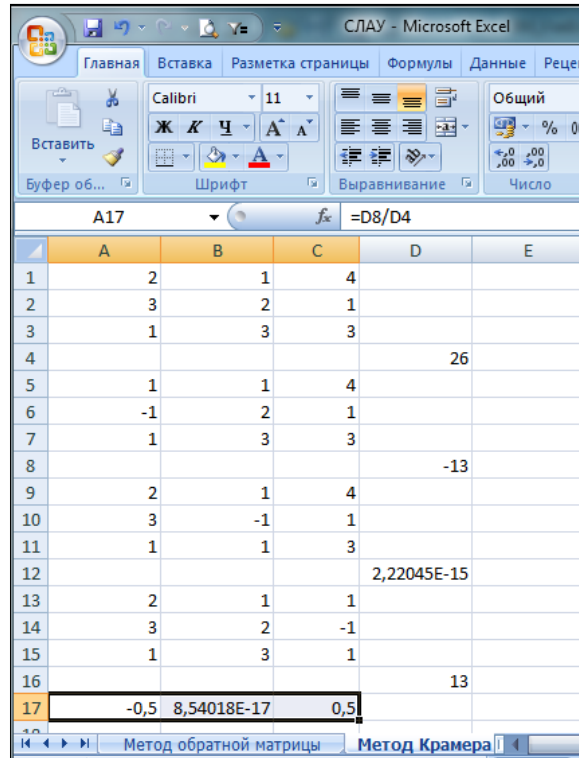


Рис. 5. Решение СЛАУ средствами *MS Excel* методом Крамера

После округления получаем решение  $\mathbf{x} = \begin{pmatrix} -0,5 \\ 0 \\ 0,5 \end{pmatrix}$ .

#### *Метод обратной матрицы*

Для решения СЛАУ методом обратной матрицы вычисляем обратную матрицу для матрицы коэффициентов и умножаем ее на столбец свободных членов:  $\mathbf{x} = A^{-1} \cdot \mathbf{b}$ .

1. Запускаем *MS Excel*. Заполняем ячейки **A2:C4** значениями коэффициентов системы:

$$\begin{pmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{pmatrix}.$$

2. Заносим в ячейки **D2:D4** значения свободных членов:  $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ .

3. Выделяем методом протаскивания диапазон **A6:C8**.

В ячейку **A6** вводим формулу **=Мобр(A2:C4)**. Для этого на панели инструментов **CL** (кликнув) на кнопке **f(x)** (функция), выбираем из вкладки **Математические** функцию **Мобр**.

Переместив указатель в строку формул в конец формулы  $=\text{Мобр}(A2:C4)$  и нажав одновременно три клавиши  $\text{Shift}+\text{Ctrl}+\text{Enter}$ , получаем в выделенном диапазоне значения обратной матрицы.

4. Выделяем ячейки  $E6:E8$  и в ячейку  $E6$  вводим формулу  $=\text{Мумнож}(A6:C8; D1:D3)$ , выбрав функцию  $\text{Мумнож}$  из вкладки  $\text{Математические}$ .

Переместив указатель в строку формул в конец формулы и нажав одновременно три клавиши  $\text{Shift}+\text{Ctrl}+\text{Enter}$ , получаем в выделенном диапазоне значения вектора неизвестных  $x$  (рис. 6).

Не снимая выделения с ячеек, в которых получено решение, вызываем нажатием правой кнопки мыши (RCL) контекстное меню, нажимаем строку «Формат ячеек» и открывшемся окне диалога задаем для числового формата количество десятичных знаков (в данном примере – 2).

Чтобы проверить правильность полученного решения  $x = \begin{pmatrix} -0,5 \\ 0 \\ 0,5 \end{pmatrix}$ ,

выделим диапазон  $G6:G8$  и введем формулу  $=\text{Мумнож}(A2:C4; E6:E8)$ . Нажав одновременно три клавиши  $\text{Shift}+\text{Ctrl}+\text{Enter}$ , получаем в выделенном диапазоне значения вектора свободных членов.

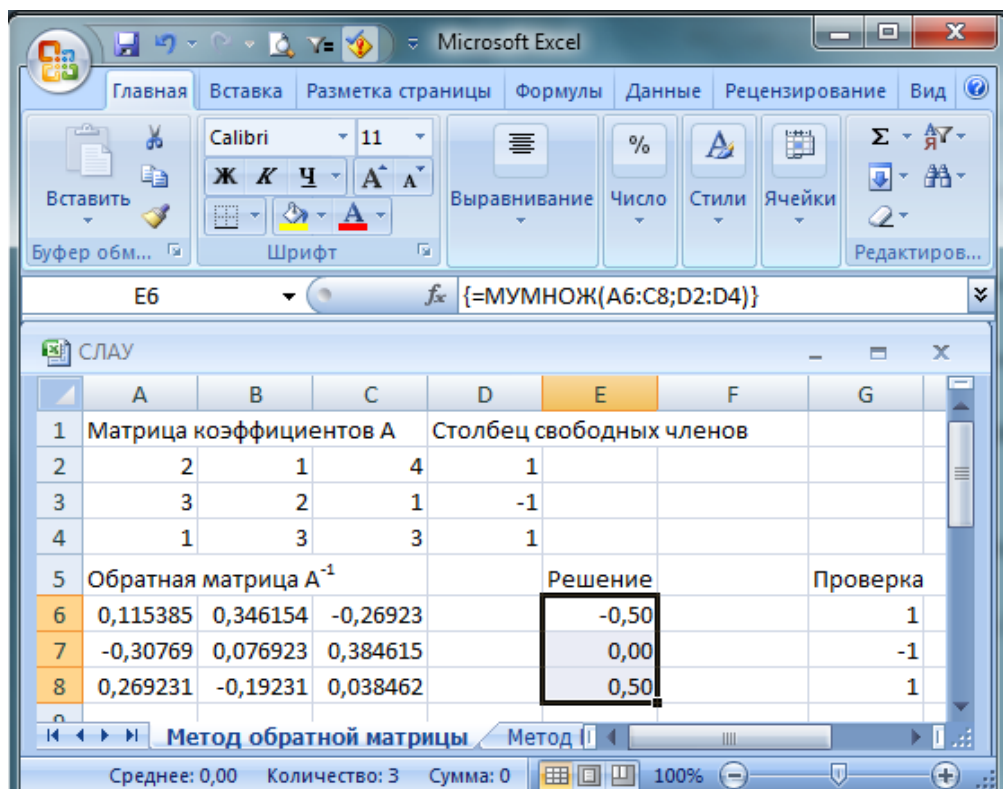


Рис. 6. Решение СЛАУ средствами  $MS\ Excel$  методом обратной матрицы

### Метод Гаусса

1. Запускаем *MS Excel*. Заполняем ячейки **A2:D4** (рис. 7) значениями элементов расширенной матрицы (матрицы коэффициентов системы и вектора свободных членов):

$$A_1 = \begin{pmatrix} 2 & 1 & 4 & 1 \\ 3 & 2 & 1 & -1 \\ 1 & 3 & 3 & 1 \end{pmatrix}.$$

2. Выполним *прямой ход* метода Гаусса решения СЛАУ – приведем матрицу коэффициентов к верхнетреугольному виду с единицами на главной диагонали путем исключения элементов, расположенных ниже главной диагонали.

Для этого, поделив первую строку на первый *ведущий элемент* (в.э.)  $a_{11} = 2 \neq 0$ , получим первую *опорную строку* (о.с.): **A6:D6**.

Чтобы получить нулевые элементы в первом столбце, умножаем первую о.с. на  $a_{21} = 3$  и результат вычитаем из второй строки (**A7:D7**). Аналогично для третьей строки: умножаем первую о.с. на  $a_{31} = 1$  и результат вычитаем из третьей строки (**A8:D8**).

Продолжаем со второй строки: поделив ее на второй в.э.  $\tilde{a}_{22} \neq 0$ , получим вторую о.с.: **A10:D10**. Умножаем вторую о.с. на  $a_{32} = 3$  и результат вычитаем из третьей строки (**A11:D11**).

Продолжаем с третьей строки: поделив ее на третий в.э.  $\tilde{a}_{33} \neq 0$ , получим третью о.с.: **A13:D13**.

Таким образом, в результате прямого хода получили матрицу, которая содержит опорные строки:

$$\hat{A}_1 = \begin{pmatrix} 1 & 0,5 & 2 & 0,5 \\ 0 & 1 & -10 & -5 \\ 0 & 0 & 1 & 0,5 \end{pmatrix}.$$

3. Выполним *обратный ход* метода Гаусса.

По матрице  $\hat{A}_1$  составим систему уравнений

$$\begin{pmatrix} 1 & 0,5 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,5 \\ -5 \\ 0,5 \end{pmatrix}$$

или

$$\begin{cases} x_1 + 0,5x_2 + 2x_3 = 0,5; \\ x_2 - 10x_3 = -5; \\ x_3 = 0,5. \end{cases}$$

Разрешая эту систему, начиная с последнего уравнения, получаем:  $x_3 = 0,5$ ;  $x_2 = 0$ ;  $x_1 = -0,5$  (**G13**, **G10**, **G6** соответственно).

	A	B	C	D	E	F	G	H	
1	Матрица коэффициентов A			Столбец свободных членов					
2	2	1	4	1					
3	3	2	1	-1					
4	1	3	3	1					
5	Прямой ход				Обратный ход				
6	1	0,5	2	0,5		x1=	-0,5		
7	0	0,5	-5	-2,5					
8	0	2,5	1	0,5					
9									
10	0	1	-10	-5		x2=	0		
11	0	0	26	13					
12									
13	0	0	1	0,5		x3=	0,5		
14									

Рис. 7. Решение СЛАУ средствами *MS Excel* методом обратной матрицы

Замечания. 1. Изложенный метод называют методом Гаусса единственного деления. Он имеет ограничение: ведущие элементы должны быть отличны от нуля и не должны быть малыми по модулю, поскольку тогда погрешности вычислений будут большими.

Поэтому в методе Гаусса – Жордана (с выбором главного элемента) в качестве ведущих элементов выбираются максимальные по модулю  $a_{ii}$  путем перебора этих элементов по столбцу, соответствующему этому в.э., или по всем столбцам.

2. Если система имеет разреженную матрицу коэффициентов, то удобно использовать метод прогонки.

#### Метод *LU*-разложения (рис. 8)

1. Заполняем в *MS Excel* ячейки **A2:C4** значениями элементов матрицы коэффициентов *A*, ячейки **D2:D4** – значениями столбца свободных членов.

2. Прямой ход

Выполним операцию факторизации (представим *A* в виде произведения двух треугольных матриц *L* и *U*):

$$\begin{pmatrix} \boxed{2} & 1 & 4 \\ 3 & \boxed{2} & 1 \\ 1 & 3 & 3 \end{pmatrix} \xrightarrow{k=1} \begin{pmatrix} 2 & 0,5 & 2 \\ 3 & \boxed{2} & 1 \\ 1 & 3 & \boxed{3} \end{pmatrix} \xrightarrow{k=2} \begin{pmatrix} 2 & 0,5 & 2 \\ 3 & \boxed{0,5} & -10 \\ 1 & 2,5 & \boxed{3} \end{pmatrix} \xrightarrow{k=3} \begin{pmatrix} 2 & \boxed{0,5} & 2 \\ 3 & 0,5 & \boxed{-10} \\ 1 & 2,5 & 26 \end{pmatrix};$$

при  $k=1$   $l_{11} = a_{11} = 2; l_{21} = a_{21} = 3; l_{31} = a_{31} = 1;$

$$u_{12} = \frac{1}{l_{11}} a_{12} = 0,5; u_{13} = \frac{1}{l_{11}} a_{13} = \frac{4}{2} = 2;$$

	A	B	C	D	E
1	Матрица коэффициентов A				Столбец свободных членов b
2	2	1	4		1
3	3	2	1		-1
4	1	3	3		1
5	Прямой ход:				
6	Матрицы LU				
7	=A2	=B2/\$A\$2	=C2/\$A\$2		
8	=A3	=B3-A8*B7	=1/B8*(C3-A8*C7)		
9	=A4	=B4-A9*B8	=C4-A9*C7-B9*C8		
10	Матрица L			b	y
11	=A7	0	0	=E2	=D11/A11
12	=A8	=B8	0	=E3	=(D12-A12*E11)/B12
13	=A9	=B9	=C9	=E4	=(D13-A13*E11-B13*E12)/C13
14	Обратный ход:				
15	Матрица U			y	x
16	1	=B7	=C7	=E11	=D16-B16*E17-C16*E18
17	0	1	=C8	=E12	=D17-C17*E18
18	0	0	1	=E13	=D18
19					

E16      fx      =D16-B16\*E17-C16\*E18

	A	B	C	D	E	F	G
1	Матрица коэффициентов A				Столбец свободных членов b		
2	2	1	4		1		
3	3	2	1		-1		
4	1	3	3		1		
5	Прямой ход:						
6	Матрицы LU						
7	2	0,5	2				
8	3	0,5	-10				
9	1	2,5	26				
10	Матрица L			b	y		
11	2	0	0	1	0,5		
12	3	0,5	0	-1	-5		
13	1	2,5	26	1	0,5		
14	Обратный ход:						
15	Матрица U			y	x		
16	1	0,5	2	0,5	-0,5		
17	0	1	-10	-5	0		
18	0	0	1	0,5	0,5		
19							

Рис. 8. Решение СЛАНУ средствами MS Excel методом LU-разложения



при  $k=2$   $l_{22} = a_{22} - l_{21}u_{12} = 2 - 3 \cdot 0,5 = 0,5$ ;  $l_{32} = a_{32} - l_{31}u_{12} = 3 - 1 \cdot 0,5 = 2,5$ ;  
 $u_{23} = \frac{1}{l_{22}}(a_{23} - l_{21}u_{13}) = \frac{1}{0,5}(1 - 3 \cdot 2) = -10$ ;  
при  $k=3$   $l_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23} = 3 - 1 \cdot 2 - 2,5 \cdot (-10) = 26$ .

В результате получены две треугольные матрицы:

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 0,5 & 0 \\ 1 & 2,5 & 26 \end{pmatrix}; \quad U = \begin{pmatrix} 1 & 0,5 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix}.$$

Решим систему  $L \cdot \mathbf{y} = \mathbf{b}$ .

$$\begin{pmatrix} 2 & 0 & 0 \\ 3 & 0,5 & 0 \\ 1 & 2,5 & 26 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \text{ или } \begin{cases} 2y_1 = 1; \\ 3y_1 + 0,5y_2 = -1; \\ y_1 + 2,5y_2 + 26y_3 = 1. \end{cases}$$

Отсюда  $y_1 = 0,5$ ;  $y_2 = -5$ ;  $y_3 = 0,5$ .

3. Обратный ход

Решим систему  $U \cdot \mathbf{x} = \mathbf{y}$ .

$$\begin{pmatrix} 1 & 0,5 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,5 \\ -5 \\ 0,5 \end{pmatrix} \text{ или } \begin{cases} x_1 + 0,5x_2 + 2x_3 = 0,5; \\ x_2 - 10x_3 = -5; \\ x_3 = 0,5. \end{cases}$$

Отсюда  $x_3 = 0,5$ ;  $y_2 = 0$ ;  $y_3 = -0,5$ .

## 2.8. Обусловленность СЛАУ

Рассмотрим СЛАУ в матричной форме  $A\mathbf{x} = \mathbf{b}$ , для которой  $A = A_{n \times n}$ .

Понятие обусловленности СЛАУ введено для оценки возможности наиболее точного решения задачи в зависимости от свойств матрицы коэффициентов системы  $A$ . Обусловленность СЛАУ характеризуется так называемым числом обусловленности, обозначаемым  $\text{cond}(A)$  (от английского слова *condition* – условие).

Число обусловленности СЛАУ  $\text{cond}(A)$  вычисляется по формуле

$$\text{cond}(A) = \frac{\max_i \mu_i}{\min_i \mu_i}, \quad (13)$$

где  $\mu_i$  – сингулярные числа матрицы  $A \times A^T$ .

После определения числа обусловленности СЛАУ оно анализируется: чем ближе  $\text{cond}(A)$  к 1, тем лучше обусловлена система линейных алгебраических уравнений.

Если  $\text{cond}(A)$  – большое число, то ошибки округлений, возникающие при решении системы, либо ошибки в числовых выражениях коэффициентов или в правой части системы способны исказить решение СЛАУ до неприемлемой степени.

Пусть дана система линейных алгебраических уравнений

$$A \cdot \mathbf{x} = \mathbf{b}.$$

Тогда

$$\mathbf{x} = A^{-1} \cdot \mathbf{b}.$$

Докажем, что при внесении ошибки  $\delta \mathbf{b}$  в правую часть СЛАУ ошибка  $\delta \mathbf{x}$  в решении системы подчиняется следующему неравенству:

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \text{cond}(A) \cdot \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}. \quad (14)$$

**Доказательство.**

Если в правую часть СЛАУ внести ошибку  $\delta \mathbf{b}$ , решение системы будет найдено с ошибкой  $\delta \mathbf{x}$ , то есть

$$A(\mathbf{x} + \delta \mathbf{x}) = \mathbf{b} + \delta \mathbf{b}.$$

Оценим величину  $\delta \mathbf{x}$ :

$$\begin{aligned} A\mathbf{x} + A\delta \mathbf{x} &= \mathbf{b} + \delta \mathbf{b}, \\ A\delta \mathbf{x} &= \delta \mathbf{b}, \\ \delta \mathbf{x} &= A^{-1}\delta \mathbf{b}. \end{aligned}$$

По свойству (4) нормы имеем:

$$\|\delta \mathbf{x}\| \leq \|A^{-1}\| \cdot \|\delta \mathbf{b}\|. \quad (15)$$

По тому же свойству из равенства  $A \cdot \mathbf{x} = \mathbf{b}$  получим:

$$\|\mathbf{b}\| \leq \|A\| \cdot \|\mathbf{x}\|. \quad (16)$$

Из уравнений (15) и (16) следует:

$$\begin{aligned} \|\delta \mathbf{x}\| \cdot \|\mathbf{b}\| &\leq \|A^{-1}\| \cdot \|A\| \cdot \|\delta \mathbf{b}\| \cdot \|\mathbf{x}\|; \\ \frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} &\leq \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}. \end{aligned}$$

Так как  $\|A\| = \mu_1$ ,  $\|A^{-1}\| = \mu_n^{-1}$  и  $\text{cond}(A) = \frac{\mu_1}{\mu_n}$ , то  $\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \text{cond}(A) \cdot \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}$ .

**Пример 21.**

Исследовать обусловленность СЛАУ  $\begin{cases} x_1 + 0,99x_2 = 1,99; \\ 0,99x_1 + 0,98x_2 = 1,97. \end{cases}$

Внеся небольшую ошибку в правую часть системы, оценить погрешность решения.

Решение.

Матрица коэффициентов  $A = \begin{pmatrix} 1 & 0,99 \\ 0,99 & 0,98 \end{pmatrix}$ ; вектор свободных членов

$$\mathbf{b} = \begin{pmatrix} 1,99 \\ 1,97 \end{pmatrix}, \text{ точное решение системы } \mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Используя функцию  $\text{cond1}(A)$  системы *Mathcad*, вычислим число обусловленности СЛАУ:  $\text{cond}(A) = 39600 \gg 1$ .

Внесем небольшую ошибку в правую часть системы  $\delta \mathbf{b} = \begin{pmatrix} -0,000097 \\ 0,000106 \end{pmatrix}$ .

Система примет вид

$$\begin{cases} x_1 + 0,99x_2 = 1,989903; \\ 0,99x_1 + 0,98x_2 = 1,970106. \end{cases}$$

Найдя ее новое решение  $\mathbf{x} = \begin{pmatrix} 3 \\ -1,0203 \end{pmatrix}$ , можем оценить абсолютную

погрешность решения  $\Delta \mathbf{x} = \begin{pmatrix} 2 \\ 2,0203 \end{pmatrix}$ .

Этот пример наглядно демонстрирует, как при большом значении  $\text{cond}(A)$  небольшие ошибки в правой части системы способны исказить решение СЛАУ до неприемлемой степени.

### Контрольные вопросы и задания

1. Что такое матрица? Приведите примеры треугольной, ленточной, диагональной, единичной матриц.
2. Какие функции *MS Excel* используются при решении СЛАУ прямыми методами?
3. Какая система уравнений называется однородной, неособенной, невырожденной, недоопределенной?
4. Как вычислить скалярное и векторное произведения двух векторов?
5. Как можно определить норму матрицы? Чему равна норма единичного вектора?
6. Что такое собственные значения матрицы, сингулярные числа матрицы?
7. Как вычисляется квадратичная форма матрицы? Какая квадратичная форма матрицы считается положительно определенной?
8. Почему один из прямых методов решения СЛАУ называется методом *LU*-разложения?
9. На что влияет число обусловленности СЛАУ?
10. От чего зависит точность решения СЛАУ?

### 3. ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

#### 3.1. Общие сведения

Идея всех итерационных методов решения СЛАУ состоит в том, чтобы, задав некоторое начальное приближение к решению  $\mathbf{x}^{(0)} \in R^n$ , последовательно уточнять его, определяя последующие приближения  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)}, \mathbf{x}^{(k+1)}$  ... При этом каждое  $\mathbf{x}^{(k+1)}$  выражается через  $\mathbf{x}^{(k)}, \mathbf{x}^{(k-1)}, \mathbf{x}^{(k-2)}$  ... с помощью формул, соответствующих методу решения.

Если при вычислении последующего приближения используется только одна предыдущая итерация, т.е. на каждом шаге последующий слой приближений определяется через предыдущий  $\mathbf{x}^{(k+1)} = f(\mathbf{x}^{(k)})$ , то метод называется *одношаговым* или *двухслойным*.

Если  $\mathbf{x}^{(k+1)}$  вычисляется по двум предыдущим итерациям  $\mathbf{x}^{(k+1)} = f(\mathbf{x}^{(k)}, \mathbf{x}^{(k-1)})$ , то имеем двухшаговый, или трехслойный, метод.

*Каноническая* формула одношаговых итерационных методов записывается следующим образом:

$$B \frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau^{(k+1)}} + A\mathbf{x}^{(k)} = \mathbf{b}, \quad (17)$$

где  $B$  – некоторая матрица размера  $n \times n$ , для которой существует обратная;

$A$  – матрица коэффициентов СЛАУ ( $A = A_{n \times n}$ );

$k$  – номер итерации;

$\tau^{(k+1)}$  – итерационный параметр, причем  $\tau^{(k+1)} > 0$ .

Если  $B = E$  – единичная матрица, то метод называется *явным*. Явная схема имеет вид:

$$\frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau^{(k+1)}} + A\mathbf{x}^{(k)} = \mathbf{b}$$

или

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \tau^{(k+1)} (A\mathbf{x}^{(k)} - \mathbf{b}).$$

В общем случае  $B \neq E$ , и для нахождения  $\mathbf{x}^{(k+1)}$  используется следующая система уравнений:

$$B\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} - \tau^{(k)} (A\mathbf{x}^{(k)} - \mathbf{b}).$$

Такой метод определения  $\mathbf{x}^{(k+1)}$  называется *неявной схемой*.

Естественно, решение системы уравнений для нахождения  $\mathbf{x}^{(k+1)}$  должно быть достаточно простым.

Точность решения СЛАУ – это мера отклонения вектора  $\mathbf{x}^{(k)}$  от вектора  $\mathbf{x}^*$ , где  $\mathbf{x}^*$  – точное решение.

Пусть  $\mathbf{z}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^*$ , тогда, подставляя выражение для  $\mathbf{z}^{(k)}$  в формулу (17), отображающую канонический вид итерационного метода, получим:

$$B \frac{\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}}{\tau^{(k)}} + A\mathbf{z}^{(k)} = 0.$$

Итерационный метод сходится, если норма вектора  $\mathbf{z}^{(k)}$  стремится к 0 при  $k \rightarrow \infty$ , т.е.  $\|\mathbf{z}^{(k)}\| \rightarrow 0, k \rightarrow \infty$ .

Обозначим  $m$  – номер итерации, для которой относительное отличие решения, полученного численным методом, от точного решения задачи не больше, чем некоторое заранее заданное малое число  $\xi$ :

$$\|\mathbf{x}^{(m)} - \mathbf{x}^*\| \leq \xi \|\mathbf{x}^{(0)} - \mathbf{x}^*\|.$$

Обозначим  $g_0$  – общее число арифметических операций для одной итерации решения.

Тогда  $Q(\xi) = mg_0$  определяет количество арифметических операций, необходимых для совершения  $m$  итераций.

Задача состоит в том, чтобы выбрать итерационный метод решения, т.е. матрицу  $B$  и последовательность  $\tau^{(k)}$  таким образом, чтобы минимизировать число  $Q(\xi)$ .

### 3.2. Метод простой итерации

Рассмотрим итерационный процесс решения СЛАУ (17), когда  $B = E$ ,  $\tau^{(k+1)} = \tau$ .

В развернутой форме получим явную двухслойную (одношаговую) схему:

$$x_i^{(k+1)} = x_i^{(k)} - \tau \left( \sum_{j=1}^n a_{ij} x_j^{(k)} - b_i \right).$$

Существуют и другие варианты метода простой итерации. Наиболее распространенная форма записи метода выглядит следующим образом:

$$x_i^{(k+1)} = -\frac{1}{a_{ii}} \left( \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k)} - b_i \right). \quad (18)$$

Преобразуем эту форму записи к общему виду.

Запишем сумму в матричной форме:

$$\sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k+1)} = \sum_{j=1}^n a_{ij} x_j^{(k)} - a_{ii} x_i^{(k)} = (A\mathbf{x}^{(k)})_i - (D\mathbf{x}^{(k)})_i;$$

здесь  $D = (a_{ij} \delta_{ij})$  – диагональная матрица с диагональными элементами  $a_{ii}$ ;

$\delta_{ij}$  – символ Кронекера,

$$\delta_{ij} = \begin{cases} 1 & \text{при } i = j, \\ 0 & \text{при } i \neq j. \end{cases}$$

Тогда из формулы (18) получим:

$$\begin{aligned} a_{ii} x_i^{(k+1)} &= b_i - (A\mathbf{x}^{(k)})_i + (D\mathbf{x}^{(k)})_i; \\ (D\mathbf{x}^{(k+1)})_i &= b_i - (A\mathbf{x}^{(k)})_i + (D\mathbf{x}^{(k)})_i. \end{aligned}$$

Следовательно, общую схему можно записать в виде

$$D \frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau} + A\mathbf{x}^{(k)} = \mathbf{b}, \quad \tau = 1.$$

Метод простой итерации в такой форме записи представляет собой формально неявную схему, но поскольку  $D$  является диагональной матрицей, то  $\mathbf{x}^{(k+1)}$  можно явно выразить через  $\mathbf{x}^{(k)}$ .

### 3.3. Метод Зейделя

Рассмотрим систему линейных алгебраических уравнений в развернутом виде:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2; \\ \dots \dots \dots \dots \dots \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n. \end{cases}$$

Представим эту систему в следующей форме:

$$\left\{ \begin{array}{l} x_1 = -\frac{1}{a_{11}} \left( \sum_{j=2}^n a_{1j} x_j - b_1 \right); \\ \dots \\ x_i = -\frac{1}{a_{ii}} \left( \sum_{\substack{j=1 \\ i \neq j}}^n a_{ij} x_j - b_i \right); \\ \dots \\ x_n = -\frac{1}{a_{nn}} \left( \sum_{j=1}^{n-1} a_{nj} x_j - b_n \right). \end{array} \right.$$

Построим итерационную схему по методу Зейделя:

$$\left\{ \begin{array}{l} x_1^{(k+1)} = -\frac{1}{a_{11}} \left( \sum_{j=2}^n a_{1j} x_j^{(k)} - b_1 \right); \\ x_2^{(k+1)} = -\frac{1}{a_{22}} \left( a_{21} x_1^{(k+1)} + \sum_{j=3}^n a_{2j} x_j^{(k)} - b_2 \right); \\ \dots \\ x_i^{(k+1)} = -\frac{1}{a_{ii}} \left( \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^n a_{ij} x_j^{(k)} - b_i \right); \\ \dots \\ x_n^{(k+1)} = -\frac{1}{a_{nn}} \left( \sum_{j=1}^{n-1} a_{nj} x_j^{(k+1)} - b_n \right). \end{array} \right.$$

В общем виде  $i$ -е уравнение для схемы метода Зейделя можно записать в виде

$$\sum_{j=1}^i a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^n a_{ij} x_j^{(k)} = b_i. \quad (19)$$

Каноническая форма записи для метода Зейделя выражается следующим соотношением:

$$(A^- + D) \frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau} + A \mathbf{x}^{(k)} = \mathbf{b}, \quad \tau = 1, \\ A = A^- + D + A^+, \quad (20)$$

где  $D$ ,  $A^-$ ,  $A^+$  – соответственно диагональная, поддиагональная и наддиагональная матрицы.

### 3.4. Метод релаксации (модификация метода Зейделя для ускорения сходимости)

Рассмотрим метод, представленный следующей канонической формой:

$$(D + \omega A^-) \frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\omega} + A\mathbf{x}^{(k)} = \mathbf{b}, \quad \tau = \omega.$$

Перейдем от канонической формы к вычислительной схеме:

$$\begin{aligned} \left(\frac{1}{\omega} D + A^-\right) \mathbf{x}^{(k+1)} - \left(\frac{1}{\omega} D + A^-\right) \mathbf{x}^{(k)} + A\mathbf{x}^{(k)} &= \mathbf{b}; \\ \frac{1}{\omega} D\mathbf{x}^{(k+1)} + A^-\mathbf{x}^{(k+1)} &= \frac{1}{\omega} D\mathbf{x}^{(k)} + A^-\mathbf{x}^{(k)} - A\mathbf{x}^{(k)} + \mathbf{b}; \\ \frac{1}{\omega} D\mathbf{x}^{(k+1)} &= \frac{1}{\omega} D\mathbf{x}^{(k)} + (\mathbf{b} - A^-\mathbf{x}^{(k+1)} - A^+\mathbf{x}^{(k)} - D\mathbf{x}^{(k)}); \\ D\mathbf{x}^{(k+1)} &= D\mathbf{x}^{(k)} + \omega (\mathbf{b} - A^-\mathbf{x}^{(k+1)} - (A^+ + D)\mathbf{x}^{(k)}). \end{aligned}$$

Выпишем  $i$ -е уравнение:

$$a_{ii}x_i^{(k+1)} = a_{ii}x_i^{(k)} + \omega \left( b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^n a_{ij}x_j^{(k)} \right),$$

или

$$x_i^{(k+1)} = x_i^{(k)} + \frac{\omega}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^n a_{ij}x_j^{(k)} \right), \quad 0 < \omega < 2. \quad (21)$$

Если  $0 < \omega < 1$ , то имеем *метод нижней релаксации*; если  $1 < \omega < 2$  – *метод верхней релаксации*.

### 3.4. Вопросы сходимости стационарных итерационных методов

Метод называется *стационарным*, когда компоненты матрицы  $B$  и итерационный параметр  $\tau$  не зависят от номера итерации  $k$ .

Справедлива следующая *теорема сходимости итерационных методов решения СЛАУ*.

**Для того чтобы стационарный итерационный процесс**  
 $B \frac{x^{(k+1)} - x^{(k)}}{\tau} + Ax^{(k)} = b$  **решения системы  $Ax = b$  был сходящимся:**

$$\|z^{(k)}\| \rightarrow 0, \quad k \rightarrow \infty, \quad z^{(k)} = x^{(k)} - x^*,$$



необходимо, чтобы матрица  $A$  была симметрической положительной матрицей ( $A = A^T$ ,  $A > 0$ ) и выполнялось условие

$$B - \frac{\tau}{2}A > 0,$$

то есть чтобы матрица  $B - \frac{\tau}{2}A$  была положительно определенной.

Рассмотрим схему решения вопросов сходимости на примере метода простой итерации, причем будем считать, что  $A$  – симметрическая положительная матрица:

$$E \frac{x^{(k+1)} - x^{(k)}}{\tau} + Ax^{(k)} = b, \quad A = A^T, \quad A \geq 0.$$

Можно доказать, что  $\frac{A}{\|A\|} \leq E$ , т.е. матрица  $E - \frac{A}{\|A\|}$  положительно определена и, следовательно, справедливо неравенство  $E - \frac{\tau}{2}A \geq \frac{A}{\|A\|} - \frac{\tau}{2}A$ .

Таким образом, если мы найдем условия, при которых правая часть неравенства будет положительной, то и левая часть будет положительной. То есть необходимо, чтобы

$$\frac{A}{\|A\|} - \frac{\tau}{2}A = \left( \frac{1}{\|A\|} - \frac{\tau}{2} \right) A > 0.$$

Так как  $A$  – симметрическая положительная матрица, то необходимо, чтобы коэффициент  $\frac{1}{\|A\|} - \frac{\tau}{2}$  был больше нуля или  $\tau < \frac{2}{\|A\|}$ . Это и будет условием сходимости метода простой итерации.

Если в канонической форме записи итерационного метода принять  $\tau = 1$  и  $B = D$ , то можно доказать другую теорему сходимости метода простой итерации (метода Зейделя и метода релаксации), которая не требует симметричности и положительности матрицы коэффициентов, но требует выполнения условия преобладания диагональных элементов.

Для формулировки теоремы приведем систему  $Ax = b$  к виду

$$x = A'x + \varphi,$$

где  $\varphi$  – столбец свободных членов.

Тогда итерационная схема будет выглядеть следующим образом:

$$x^{(k+1)} = A'x^{(k)} + \varphi.$$

Имеет место следующая теорема.

Для СЛАУ вида  $\mathbf{x} = A'\mathbf{x} + \varphi$  метод простой итерации сходится к решению со скоростью геометрической прогрессии со знаменателем  $g$ , если  $\|A'\| = g < 1$ .

**Доказательство.**

Пусть  $\mathbf{x}^*$  – точное решение, тогда  $\mathbf{x}^* = A'\mathbf{x}^* + \varphi$ , и можно записать равенство  $\mathbf{x}^* - \mathbf{x}^{(k+1)} = A'(\mathbf{x}^* - \mathbf{x}^{(k)})$  или  $\mathbf{z}^{(k+1)} = A'\mathbf{z}^{(k)}$ . По одному из свойств нормы получаем:  $\|\mathbf{z}^{(k+1)}\| \leq \|A'\| \cdot \|\mathbf{z}^{(k)}\|$  или  $\|\mathbf{z}^{(k+1)}\| \leq g\|\mathbf{z}^{(k)}\|$ . Так как  $g < 1$  и каждое последующее приближение по норме меньше предыдущего в  $\frac{1}{g}$  раз, то теорема доказана.

Справедливо также утверждение о том, что любую систему вида  $A\mathbf{x} = \mathbf{f}$  можно привести к эквивалентной системе вида  $\mathbf{x} = A'\mathbf{x} + \varphi$ , для которой выполняются условия последней теоремы сходимости.

Докажем это утверждение.

Рассмотрим систему  $A\mathbf{x} = \mathbf{f}$ , где  $A$  – неособенная матрица, т.е.  $|A| \neq 0$ , и существует обратная матрица  $A^{-1}$ .

Рассмотрим матрицу  $(A^{-1} - \xi)$ , где  $\xi = \xi_{n \times n} = (\xi_{ij})$  – матрица, у которой элементы  $\xi_{ij}$  – малые числа. Умножим обе части исходной системы на матрицу  $(A^{-1} - \xi)$ :

$$\begin{aligned} (A^{-1} - \xi)A\mathbf{x} &= (A^{-1} - \xi) \cdot \mathbf{f}; \\ A^{-1}A\mathbf{x} - \xi A\mathbf{x} &= (A^{-1} - \xi) \cdot \mathbf{f}; \\ \mathbf{x} &= \xi A\mathbf{x} + (A^{-1} - \xi) \cdot \mathbf{f}. \end{aligned}$$

Получим систему  $\mathbf{x} = A'\mathbf{x} + \varphi$ , где  $A' = \xi A$ ,  $\varphi = (A^{-1} - \xi) \cdot \mathbf{f}$ . Следовательно, в силу того что элементы  $\xi_{ij}$  – малые числа, элементы матрицы  $\xi A$  также малые числа, и их всегда можно сделать такими, чтобы выполнялось условие сходимости  $\|A'\| = g < 1$ .

**Пример 22.**

Решить СЛАУ итерационным методом средствами *MS Excel*

$$\begin{cases} 31x_1 + 10x_2 + 5x_3 = 31; \\ 10x_1 + 5x_2 + 31x_3 = 5; \\ 5x_1 + 31x_2 + 10x_3 = -26. \end{cases}$$

Решение.

Приведем систему к сходящемуся виду. Методы простой итерации, Зейделя и релаксации являются сходящимися, если выполняется условие

$$|a_{ii}| > \sum_{j=1}^n |a_{ij}|: |a_{11}| > |a_{12}| + |a_{13}|, |a_{22}| > |a_{21}| + |a_{23}|, |a_{33}| > |a_{31}| + |a_{32}|.$$

Поменяв местами строки, получим матрицу с диагональным преобладанием элементов<sup>1</sup>:

$$\begin{cases} 31x_1 + 10x_2 + 5x_3 = 31; \\ 5x_1 + 31x_2 + 10x_3 = -26; \\ 10x_1 + 5x_2 + 31x_3 = 5. \end{cases}$$

Полученную матрицу коэффициентов вводим в ячейки **B1-D3**, столбец свободных членов – в ячейки **G1-G3** (рис. 9, 10).

Итерационный метод решения в *MS Excel* реализован в надстройке «Поиск решения»<sup>2</sup>. В этом случае каждое уравнение должно быть записано в одной из ячеек как формула для вычисления правой части (рис. 10).

Для вычисления значений правых частей используются произвольно выбранные начальные значения вектора неизвестных  $x$ .

Эти значения заданы в ячейках **B7-B9** и не изменяются, а также в ячейках **C7-C9**, содержимое которых используется в формулах для вычисления значений правых частей и уточняются в ходе итерационной процедуры для получения их требуемых значений (рис. 9).

Вызываем процедуру «Поиск решения» и задаем параметры.

---

<sup>1</sup> Приведение системы к сходящемуся виду достигается выполнением эквивалентных преобразований: умножением строки на константу, последующим сложением строк и т.п.

<sup>2</sup> Надстройку «Поиск решения» необходимо активизировать, если она еще не подключена. Для этого необходимо открыть закладку «Файл» и, нажав на кнопку «Параметры», выбрать категории «Надстройки», затем перейти в «Надстройки Excel», выбрать «Поиск решения», поставить на ней флажок. После чего она переместится в список активных надстроек, и в меню «Данные» появится соответствующая кнопка.

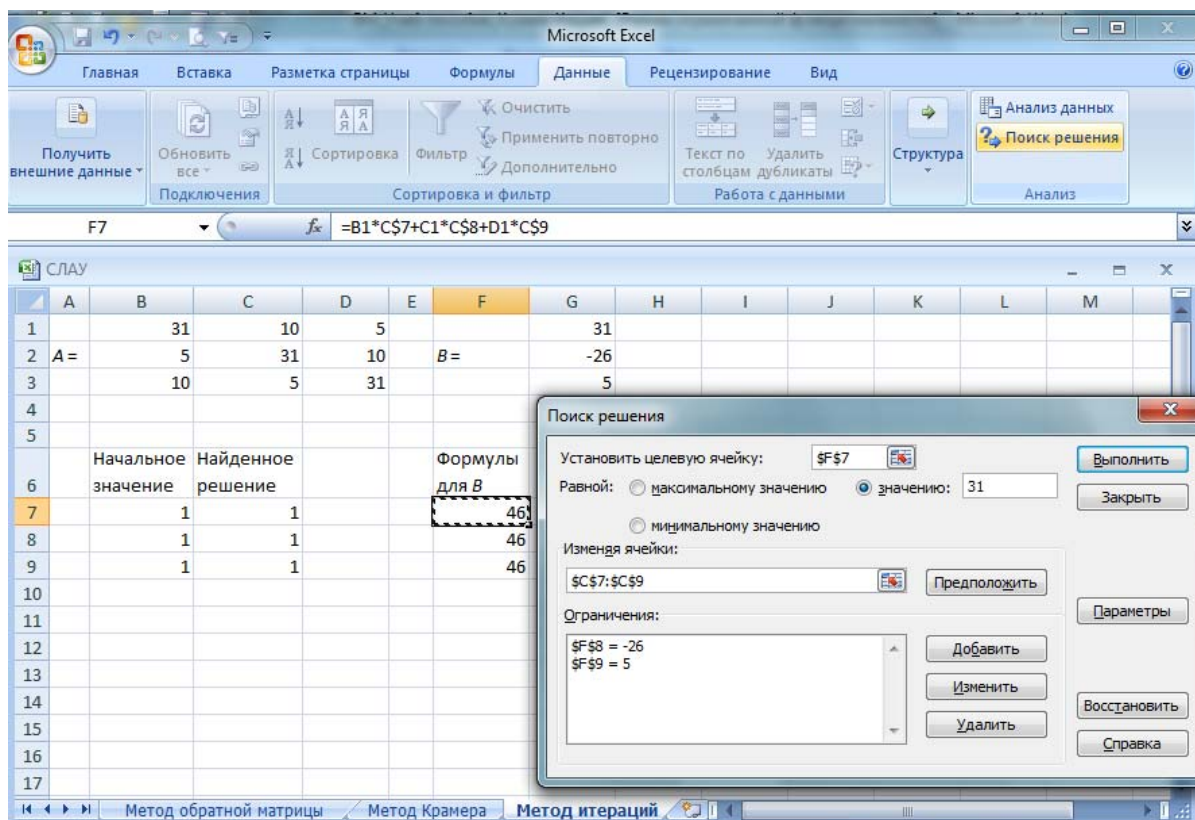


Рис. 9. Подготовка к решению СЛАУ итерационным методом в *MS Excel*

Одна из таких ячеек (например, **F7**) используется как целевая ячейка. На панели процедуры поиска решения требуем установить в ней истинное значение правой части (для нашего примера должно быть 31) за счет подбора значений искомого вектора. Две другие ячейки с формулами (**F8** и **F9**) используются как ограничения (требуем установить в них истинные значения правых частей: -26 и 5 соответственно).

В результате получено следующее решение:  $x_1 = 1,34683619$ ;  $x_2 = -1,02093964$ ;  $x_3 = -0,10850531$ . Решение СЛАУ представлено на рис. 10.

Пример решения этой системы уравнений в среде MathCAD различными итерационными методами (простых итераций, Зейделя, верхней и нижней релаксаций) и с помощью встроенной функции приведен авторами в лабораторном практикуме (см. пример 3, [21]).

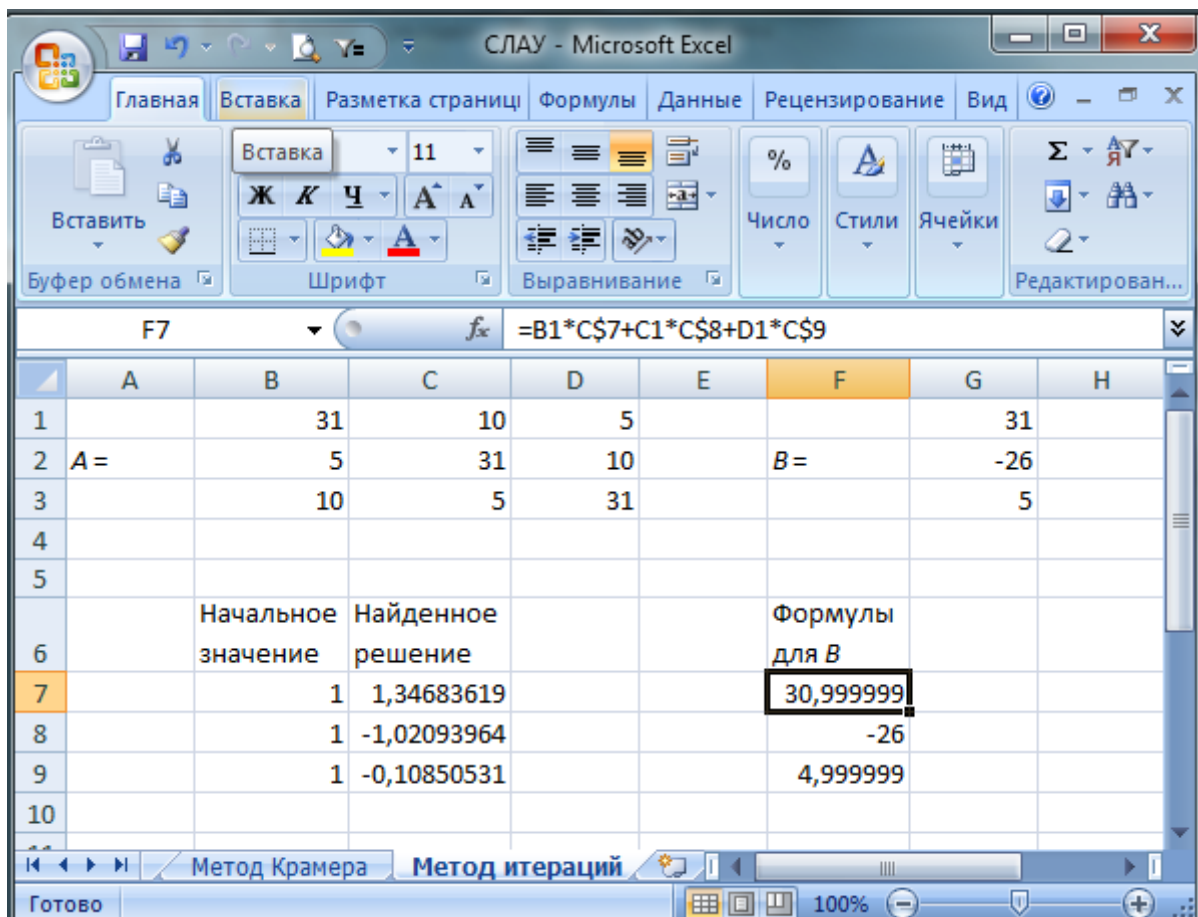
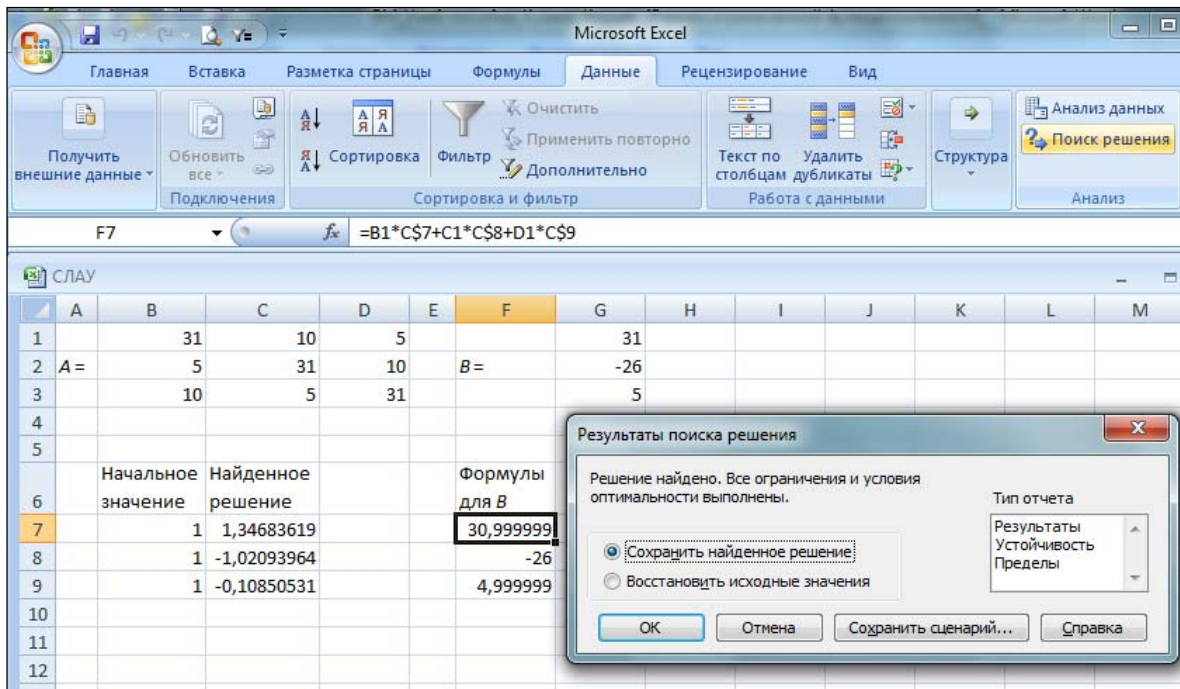


Рис. 10. Решение СЛАУ итерационным методом средствами MS Excel

## Контрольные вопросы и задания

1. Что такое система линейных алгебраических уравнений (СЛАУ) и какие основные группы методов их решения существуют?
2. Назовите прямые и итерационные методы решения СЛАУ и кратко охарактеризуйте их.
3. Как можно проверить правильность решения системы линейных алгебраических уравнений?
4. Всегда ли можно получить решение системы линейных алгебраических уравнений?
5. Что такое «начальное приближение» решения СЛАУ и как его выбирать? Сколько решений СЛАУ может существовать?
6. Какой из методов решения СЛАУ имеет лучшую сходимость при одинаковых начальных условиях: метод простых итераций или метод Зейделя?
7. Каковы достаточные условия сходимости метода Зейделя?
8. Что такое обусловленность системы?
9. Какой метод получается, если в методе релаксации положить  $\omega=1$ ?
10. Для чего СЛАУ в итерационных методах приводится к диагональному виду?
11. Что является условием окончания итерационного процесса при решении СЛАУ?
12. Как выполняются операции над матрицами и векторами в программах *MS Excel* и *MathCAD*?

## 4. РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

### 4.1. Постановка задачи

*Нелинейным уравнением* относительно неизвестной  $x$  называется уравнение, в которое эта неизвестная входит либо в степени не равной единице, либо в виде функции. Нелинейное уравнение в общем виде может быть записано как

$$f(x) = 0,$$

где  $f(x)$  – некоторая непрерывная функция, заданная на интервале  $[a, b]$ .

Всякое число  $\xi \in [a, b]$ , обращающее приведенное уравнение в тождество  $f(\xi) \equiv 0$ , называется *корнем* уравнения или его *решением*. Нелинейное уравнение может иметь несколько корней.

К сожалению, для нелинейных уравнений не существует универсальных методов решения. В большинстве случаев корни нелинейного уравнения можно найти лишь приближенно. При нахождении приближенных значений корней с заданной точностью  $\varepsilon$  необходимо: 1) отделить корни, т.е. определить принадлежащие интервалу  $[a, b]$  отрезки, содержащие только один корень – интервалы изоляции корней; 2) вычислить значения корней с необходимой точностью.

Существуют различные методы нахождения приближенного решения, т.е. способы построения последовательности итераций, однако все они имеют общие этапы:

- 1) задается начальное приближение  $x^{(0)} \in [a, b]$  и точность решения  $\varepsilon$ ;
- 2) вычисляется следующее приближение  $x^{(k+1)} = \varphi(x^{(k)})$ ;
- 3) осуществляется проверка: достигнута ли заданная точность решения?

Используются различные критерии остановки итерационного процесса. Например,  $|x^{(k)} - x^*| < \varepsilon$ , где  $x^*$  – точное решение;  $\varepsilon$  – заданная точность решения. Однако это условие не всегда возможно проверить, так как  $x^*$  неизвестно. Чаще всего используют следующее:  $|\varphi(x^{(k+1)}) - \varphi(x^{(k)})| < \varepsilon$ . Тогда принимается  $x^* = x^{(k+1)}$ .

### 4.2. Правила выделения корня

Отделение корней может производиться графическим или аналитическим способом.

**Графический способ** состоит в построении графика функции  $y = f(x)$ . Абсциссы точек пересечения кривой графика с осью  $Ox$  и будут корнями уравнения. В некоторых случаях уравнение  $f(x) = 0$  удобно представить в

виде  $f_1(x) = f_2(x)$ , а потом построить графики функций  $y = f_1(x)$  и  $y = f_2(x)$ . Абсциссы точек пересечения этих графиков будут корнями уравнения. Таким способом нельзя установить точное значение корней, однако можно указать отрезки, на которых они находятся – *интервалы изоляции корней*. Каждому корню соответствует свой интервал изоляции.

**При аналитическом способе** отделения действительных корней удобно воспользоваться следующей *теоремой*.

**Если функция  $f(x)$  непрерывна на интервале  $[a, b]$  и  $f(a) \cdot f(b) < 0$ , то внутри интервала  $[a, b]$  существует корень уравнения  $f(x) = 0$ .**

**Доказательство.**

По теореме о непрерывной функции,  $f(x)$  принимает на интервале  $[a, b]$  все свои значения, расположенные между  $\min_{x \in [a, b]} f(x)$  и  $\max_{x \in [a, b]} f(x)$ ; так как функция  $f(x)$  на концах интервала  $[a, b]$  имеет разные знаки, внутри интервала существует точка  $\xi$  такая, что  $f(\xi) = 0$ , что и требовалось доказать.

Корень будет единственным, если внутри  $[a, b]$  существует и сохраняет знак первая производная функции  $f'(x)$ .

Исходя из приведенной теоремы, для определения корней необходимо область определения функции, то есть интервал  $[a, b]$ , разбить последовательностью точек  $\alpha_i$  ( $i = 1, \dots, n$ ) на подынтервалы и исследовать знаки функции  $f(\alpha_i)$ . Если в каких-то двух точках  $\alpha_{i_1}, \alpha_{i_2}$  функция  $f(x)$  имеет разные знаки, т.е. произведение  $f(\alpha_{i_1}) \cdot f(\alpha_{i_2}) < 0$ , то между значениями  $[\alpha_{i_1}, \alpha_{i_2}]$  существует корень уравнения.

**Пример 23.**

Определить графически корень нелинейного уравнения  $\lg x + x^2 = 0$  и уточнить его средствами *MS Excel*.

**Решение.**

Для решения уравнения в *MS Excel* необходимо протабулировать функцию  $f(x) = \lg x + x^2$ , построить ее график и выявить точку пересечения (рис. 11).



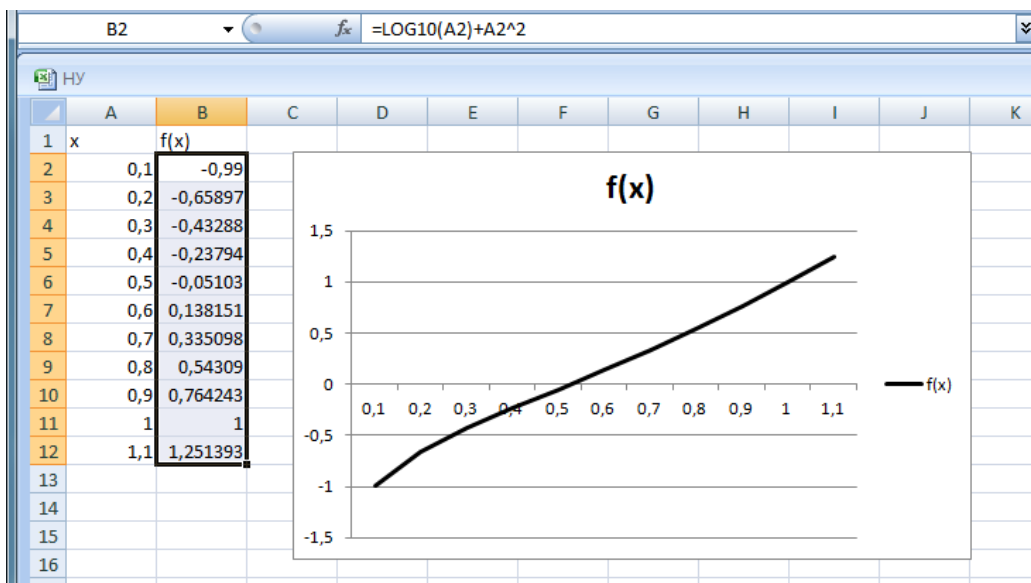


Рис. 11. Отделение корня уравнения

Из графика видно, что решение (точка пересечения графика функции с осью абсцисс) находится на отрезке  $[0,5; 0,6]$ . Таким образом, задача локализации точки решения выполнена.

Для уточнения значений  $x$  и  $f(x)$  выполним подготовительные действия: в ячейке **A2** запишем приближенное значение корня  $x = 0,5$ , в ячейке **B2** – формулу  $f(x)$  (рис. 12). Воспользуемся процедурой «Поиск решения», установив целевую ячейку **B2** равной нулю.

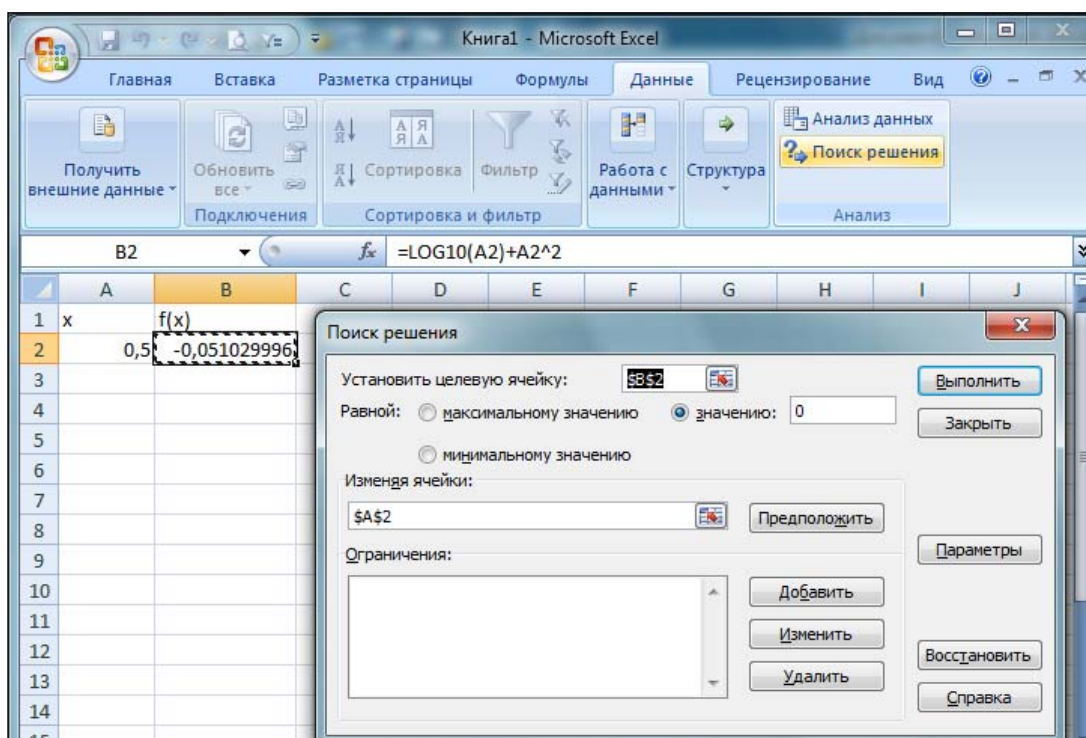


Рис. 12. Подготовка к запуску процедуры уточнения решения

На рис. 13 показан результат уточнения корня:  $f(x) \approx 0$  при  $x = 0,52724$ .

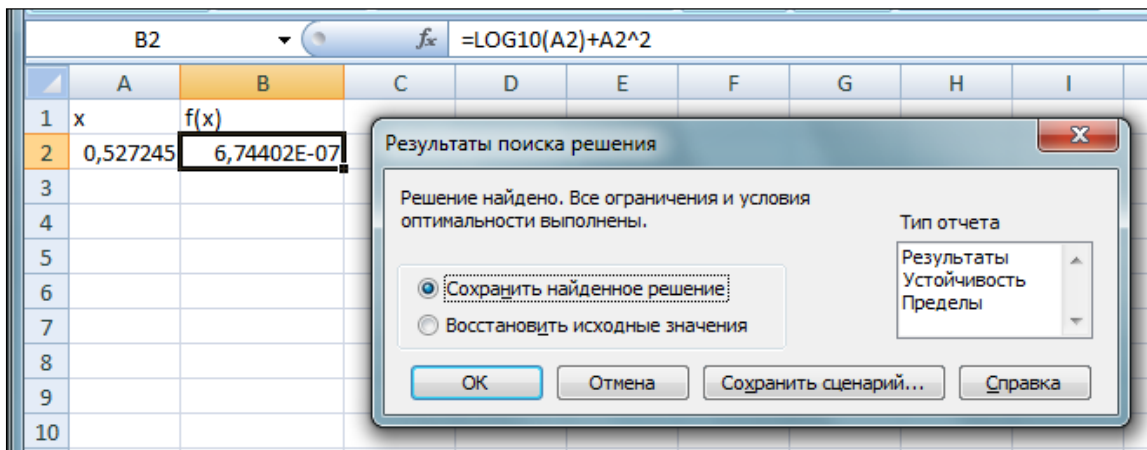


Рис. 13. Подготовка к запуску процедуры уточнения решения

В лабораторном практикуме [21] приводится решение этого примера различными итерационными методами (половинного деления, комбинированным методом хорд и касательных) и в системе *MathCAD* (см. пример 5, [21]).

### 4.3. Итерационные методы решения нелинейных уравнений

Предположим, что корень  $x^*$  уравнения  $f(x) = 0$  существует и  $x^* \in [a, b]$ , то есть  $f(x^*) = 0$  и  $a \leq x^* \leq b$ .

Будем рассматривать методы решения, основанные на итерационной схеме вида  $x^{(n)} = \varphi(x^{(n-1)})$ , сходящейся к решению уравнения, т.е.  $\{x^{(n)}\} \rightarrow x^*$  при  $n \rightarrow \infty$ , где  $x^*$  – корень уравнения. Процесс вычислений заканчивается, когда  $|\varphi(x^{(n-1)}) - \varphi(x^{(n)})| < \varepsilon$ , где  $\varepsilon$  – точность вычислений.

#### 4.3.1. Условия сходимости итерационных методов

Пусть уравнение  $f(x) = 0$  преобразовано к виду  $x = \varphi(x)$ . Это можно сделать различными способами. Однако не для каждой функции  $\varphi(x)$  итерационный процесс  $x^{(n)} = \varphi(x^{(n-1)})$  будет сходящимся.

Пример 24.

Рассмотрим уравнение  $x^2 - k = 0$  и три различных способа построения итерационной схемы:

$$\begin{aligned}
 1) \quad x &= \frac{1}{2} \cdot \left( x + \frac{k}{x} \right); \\
 2) \quad x &= \frac{k}{x}
 \end{aligned}
 \tag{22}$$

$$3) x = 2 \cdot x - \frac{k}{x}.$$

Предположим, что  $k = 2$ ,  $x^{(0)} = 1$ . Покажем численно, что сходящейся является только первая схема.

*1-й способ.*

$$x^{(1)} = \frac{1}{2}(1 + 2) = 1,5;$$

$$x^{(2)} = \frac{1}{2}\left(1,5 + \frac{2}{1,5}\right) = 1,417;$$

$$x^{(3)} = 1,414 \text{ и т.д.}$$

Видим итерационное приближение к корню уравнения:

$$x = \sqrt{k} = \sqrt{2} \quad (x^* = \sqrt{2} \approx 1,414).$$

*2-й способ.*

$$x^{(1)} = \frac{2}{1} = 2;$$

$$x^{(2)} = \frac{2}{2} = 1;$$

$$x^{(3)} = 2 \text{ и т.д.}$$

Видим, что сходимости к решению не наблюдается.

*3-й способ.*

$$x^{(1)} = 2 \cdot 1 - \frac{2}{1} = 0;$$

$$x^{(2)} = 2 \cdot 0 - \frac{2}{0} \text{ — третья схема также не сходится.}$$

Необходимым условием сходимости итерационного метода является условие  $|x^{(n)} - x^*| \rightarrow 0$  при  $n \rightarrow \infty$ , то есть  $\{x^{(n)} - x^*\}$  должна быть последовательностью, сходящейся к нулю.

В качестве условия сходимости последовательности выберем следующее:  $|x^{(n)} - x^*| < \theta \cdot |x^{(n-1)} - x^*|$  при  $\theta < 1$ , то есть потребуем убывания членов последовательности со скоростью геометрической прогрессии со знаменателем  $1/\theta$ . Тогда для итерационного процесса  $x^{(n)} = \varphi(x^{(n-1)})$  получим условие сходимости:

$$|\varphi(x^{(n-1)}) - \varphi(x^*)| < \theta \cdot |x^{(n-1)} - x^*|.$$

Более жестким условием, накладываемым на функцию  $\varphi$ , является *условие Литшица*:

$$\forall x_1, x_2 \in [a, b] \Rightarrow |\varphi(x_1) - \varphi(x_2)| < \theta |x_1 - x_2|, \quad \theta < 1. \quad (23)$$

Функция  $\varphi$ , удовлетворяющая условию Липшица, называется *сжимающим отображением* (сжатой функцией) с коэффициентом сжатия  $\theta$ . Таким образом, для того чтобы итерационный процесс сходился, достаточно, чтобы функция  $\varphi(x)$  была сжимающим отображением.

Предположим, что функция  $\varphi(x)$  – непрерывно дифференцируема. Рассмотрим неравенство, вытекающее из условия Липшица:

$$\frac{|\varphi(x_1) - \varphi(x_2)|}{|x_1 - x_2|} < \theta. \quad (24)$$

Из *теоремы о среднем* следует, что для любых точек  $x_1, x_2 \in [a, b]$  существует такое значение  $x_3 \in [a, b]$ , что  $\frac{\varphi(x_1) - \varphi(x_2)}{x_1 - x_2} = \varphi'(x_3)$ , и, следовательно, в качестве  $\theta$  всегда можно взять  $\max_{x \in [a, b]} |\varphi'(x)|$ .

Таким образом, итерационный процесс  $x^{(n)} = \varphi(x^{(n-1)})$  будет сходиться при  $n \rightarrow \infty$ , если производная функции  $\varphi(x)$  по модулю меньше некоторой константы  $\theta < 1$ , т.е.  $|\varphi'(x)| < \theta$  для любого  $x \in [a, b]$ .

**Пример 25.**

а) Для задачи  $x^2 - k = 0$  и итерационного процесса  $x = \frac{1}{2} \cdot \left( x + \frac{k}{x} \right)$ , то есть когда  $\varphi = \frac{1}{2} \cdot \left( x + \frac{k}{x} \right)$ , можно записать цепочку равенств и неравенство:

$$\begin{aligned} |\varphi(x_1) - \varphi(x_2)| &= \left| \frac{1}{2} \left( x_1 + \frac{k}{x_1} \right) - \frac{1}{2} \left( x_2 + \frac{k}{x_2} \right) \right| = \frac{1}{2} \cdot \left| (x_1 - x_2) + k \cdot \left( \frac{1}{x_1} - \frac{1}{x_2} \right) \right| = \\ &= \frac{1}{2} \cdot \left| (x_1 - x_2) + \frac{k \cdot (x_2 - x_1)}{x_1 \cdot x_2} \right| = \frac{1}{2} \cdot \left| (x_1 - x_2) \cdot \left( 1 - \frac{k}{x_1 \cdot x_2} \right) \right| \leq \frac{1}{2} \cdot |x_1 - x_2| \cdot \left| 1 - \frac{k}{x_1 \cdot x_2} \right|. \end{aligned}$$

Значит, в качестве  $\theta$  можно взять максимальное значение величины  $\frac{1}{2} \cdot \left| 1 - \frac{k}{x_1 \cdot x_2} \right|$  при  $x_1, x_2 \in [a, b]$ . Таким образом, если  $k = 2$ , то итерационный процесс (22) на интервале  $[a, b] = [1, 2]$  сходится, поскольку

$$\max_{[1, 2]} \left( \frac{1}{2} \cdot \left| 1 - \frac{2}{x_1 \cdot x_2} \right| \right) = \frac{1}{2} \cdot \left| 1 - \frac{2}{1 \cdot 1} \right| = \frac{1}{2}, \quad \theta = \frac{1}{2}.$$

б) Для этой же задачи  $x^2 - k = 0$  оценим сходимость итерационного процесса  $x = \frac{k}{x}$  при  $k = 2$ . Тогда  $|\varphi'(x)| = \left| -\frac{2}{x^2} \right|$ ,  $\max_{x \in [1, 2]} |\varphi'(x)| = 2 > 1$ .

Условие Липшица не выполняется. Это говорит о том, что сходимость данного итерационного процесса не гарантирована.

#### 4.3.2. Принцип выбора функции для итерационного процесса

При выборе функции  $\varphi$ , для которой уравнения  $f(x) = 0$  и  $\varphi(x) = x$  эквивалентны, то есть имеют общий корень  $x^*$ , и итерационный процесс  $x^{(n+1)} = \varphi(x^{(n)})$  сходится к этому корню, естественно потребовать, чтобы функция  $\varphi$  была сжимающим отображением и коэффициент сжатия  $\theta$  был минимальным.

Пусть  $r(x)$  – произвольная ограниченная функция, не имеющая корней на интервале поиска корня  $[a, b]$ . Функцию  $\varphi(x)$  можно представить в виде

$$\varphi(x) = x + r(x) \cdot f(x).$$

Чтобы итерационный процесс сходиллся с наивысшей скоростью, необходимо, чтобы  $|\varphi'(x)| \rightarrow \min$ . Вычислим  $\varphi'(x)$ :

$$\varphi'(x) = 1 + r'(x) \cdot f(x) + r(x) \cdot f'(x).$$

Вблизи корня величина  $f(x)$  мала ( $f(x) \approx 0$ ) в сравнении с остальными членами правой части. Примем ее равной нулю, тогда  $\varphi'(x) = 1 + r(x) \cdot f'(x)$ .

Рассмотрим предельный случай, когда  $|\varphi'(x)|$  близка к нулю, и тогда

$r(x) = -\frac{1}{f'(x)}$ . Следовательно,  $\varphi(x) = x - \frac{f(x)}{f'(x)}$ . В этом случае итерационный процесс будет выглядеть следующим образом:

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}. \quad (25)$$

Такой итерационный процесс называется *методом Ньютона*.

Этот процесс имеет высокую скорость сходимости в окрестности корня уравнения. Действительно, оценим  $\varphi'(x)$ :

$$\varphi'(x) = 1 - \frac{(f'(x))^2 - f(x) \cdot f''(x)}{(f'(x))^2} = \frac{f(x) \cdot f''(x)}{(f'(x))^2}. \quad (26)$$

Так как функция  $f(x)$  при  $x \rightarrow x^*$  стремится к 0, то и  $\varphi'(x) \rightarrow 0$  при  $x \rightarrow x^*$ ; и чем ближе мы подходим к корню, тем быстрее сходится процесс.

## 4.4. Итерационные схемы численных методов решения нелинейных уравнений

Пусть требуется найти корень  $x^*$  уравнения  $f(x) = 0$  на интервале  $[a, b]$ , для которого выполняется условие  $f(a) \cdot f(b) < 0$ .

Во всех представленных ниже методах будем искать последовательность интервалов  $[a_k, b_k]$  такую, чтобы  $[a_{k+1}, b_{k+1}] \subset [a_k, b_k] \subset [a, b]$ , а также чтобы  $|a_k - b_k| \rightarrow 0$  при  $k \rightarrow \infty$  и  $x^* \in [a_k, b_k]$ .

### 4.4.1. Метод простой итерации

В методе простой итерации предварительно уравнение  $f(x) = 0$  преобразуется к виду  $x = \varphi(x)$ . Это преобразование может быть выполнено различными способами, но для сходимости, как мы помним, необходимо выполнение условия  $|\varphi'(x)| < \theta < 1$ , причем при  $0 < \varphi'(x) < 1$  реализуется односторонняя (монотонная) сходимость (рис. 14, а), при  $-1 < \varphi'(x) < 0$  имеет место двухсторонняя (немонотонная) сходимость (рис. 14, б). При этом задача сводится к нахождению абсциссы точки пересечения прямой  $y = x$  и кривой  $y = f(x)$ .

Для решения задачи задаются начальное приближение  $x^{(0)} \in [a, b]$  и точность решения  $\varepsilon$ . Вычисляется следующее приближение:  $x^{(k+1)} = \varphi(x^{(k)})$ . Процесс вычислений заканчивается, когда  $|\varphi(x^{(k+1)}) - \varphi(x^{(k)})| < \varepsilon$ . Тогда  $x^* = x^{(k+1)}$ .

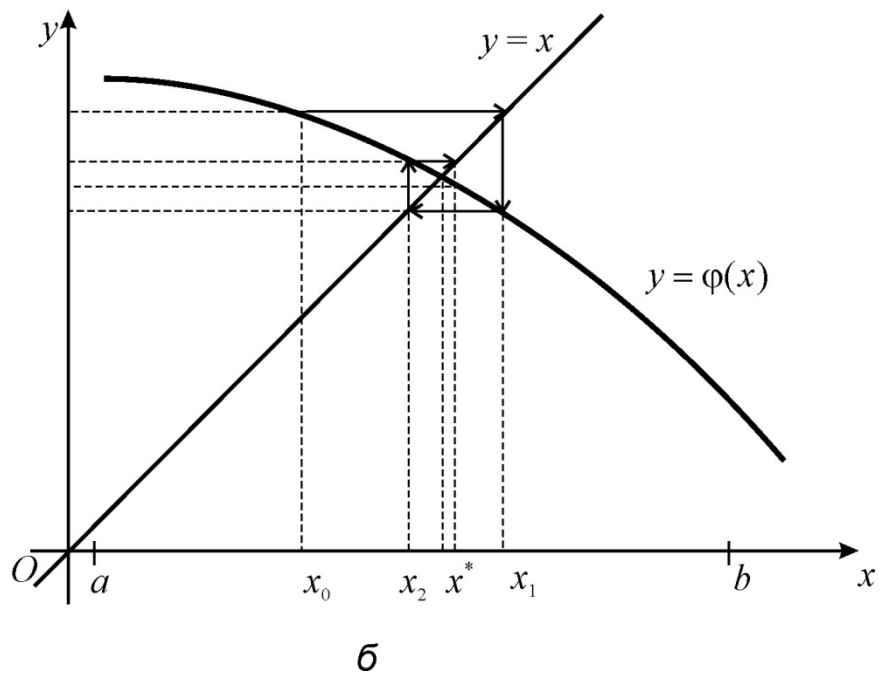
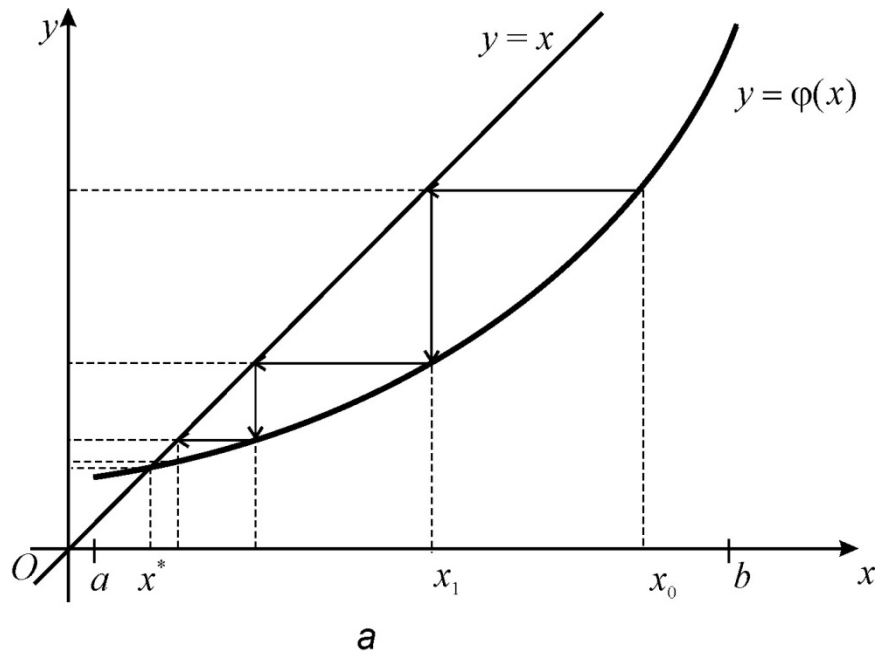


Рис. 14. Иллюстрация метода простой итерации:  
 а – односторонняя (монотонная) сходимость;  
 б – двухсторонняя (немонотонная) сходимость

#### 4.4.2. Метод половинного деления

Метод половинного деления (метод бисекций, метод дихотомии) реализуется следующим образом (рис. 15). Предположим, что уже выполнена  $(k-1)$ -я итерация, в результате чего выделен интервал  $[a_k, b_k]$ , внутри которого содержится корень  $x^*$ , т.е.  $f(a_k) \cdot f(b_k) < 0$ . Тогда на  $k$ -й итерации

делим отрезок пополам точкой  $z_k = \frac{a_k + b_k}{2}$  и вычисляем  $f(z_k)$ . Далее выбираем ту половину отрезка, на концах которой функция имеет противоположные знаки:

- 1) если  $f(z_k) \cdot f(a_k) < 0$ , то  $b_{k+1} = z_k$ ,  $a_{k+1} = a_k$ ;
- 2) если  $f(z_k) \cdot f(a_k) > 0$ , то  $b_{k+1} = b_k$ ,  $a_{k+1} = z_k$ ;
- 3) если  $f(z_k) = 0$ , то  $x^* = z_k$ .

Процесс заканчивается, когда  $|a_m - b_m| < 2\varepsilon$ , где  $\varepsilon$  – малая величина, которая определяет точность решения, а в качестве корня берется координата середины отрезка  $x^* = \frac{a_m + b_m}{2}$ .

Докажем сходимость такого метода.

Для  $k$ -го шага выполняется равенство  $|a_k - b_k| = \frac{1}{2^k} |a - b|$ , так как длина интервала на каждом шаге уменьшается вдвое. Следовательно,  $|a_k - b_k| \rightarrow 0$ ,  $k \rightarrow \infty$ . Кроме того,  $x^* \in [a_k, b_k]$  в силу выбора векторов  $a_k, b_k$ . Значит,

$$\left. \begin{array}{l} a_k \rightarrow x^* \\ b_k \rightarrow x^* \end{array} \right\} \text{при } k \rightarrow \infty.$$

Следовательно, процесс сходится.

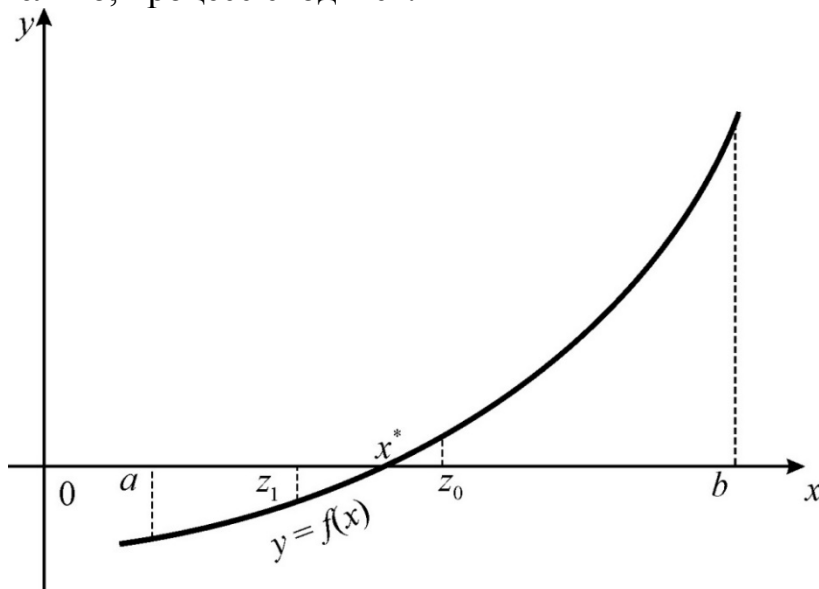


Рис. 15. Иллюстрация метода половинного деления

Метод бисекций сходится для любых непрерывных функций  $f(x)$ . Для достижения точности  $\varepsilon$  необходимо совершить  $N$  итераций,  $N \approx \log_2 \frac{b-a}{2\varepsilon}$ .



Пример 26.

Для функции  $f(x) = (28x)^3 - 56x$  определить графически все корни нелинейного уравнения  $f(x) = 0$  на интервале  $(-0,06; 0,06)$  и уточнить один из них методом половинного деления с использованием *MS Excel*.

Решение.

Построив график функции (рис. 16), выделим интервалы, в которых находятся корни уравнения:  $(-0,05; -0,04)$ ;  $(-0,01; 0,01)$ ;  $(0,04; 0,05)$ .

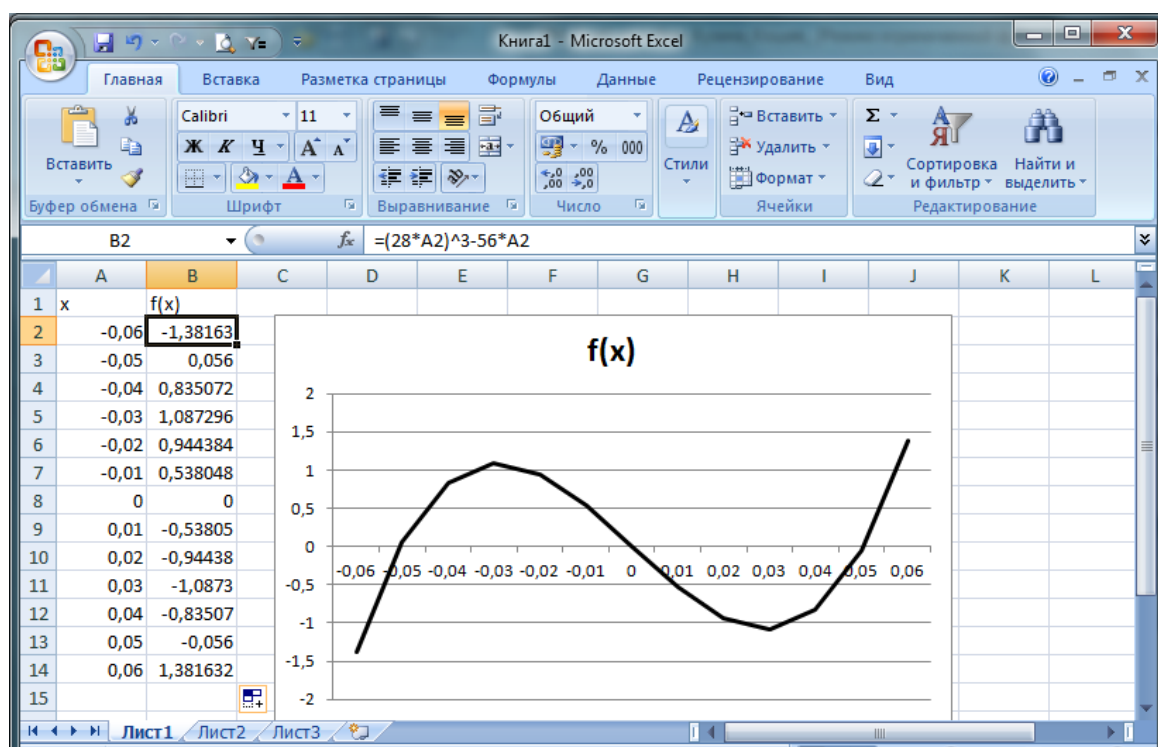


Рис. 16. Отделение корней уравнения

Уточним корень из интервала  $(0,04; 0,05)$  методом половинного деления (рис. 17).

Исходные значения границ интервала запишем в ячейки **A2** и **B2**, в ячейке **C2** получим середину заданного интервала, в ячейках **D2** и **E2** – значения функции на концах интервала, в ячейке **F2** будем определять длину интервала. Необходимую точность укажем в ячейке **H2** (используем абсолютную адресацию). В ячейку **G2** запишем формулу нахождения корня по правилу: если длина текущего отрезка меньше или равна требуемой точности, то в качестве корня уравнения принимаем значение середины этого отрезка.

В следующей строке запишем значения, полученные на первой итерации. Причем в ячейки **A3** и **B3** запишем формулы для получения нового интервала, а в ячейки **C3** – **G3** скопируем формулы из ячеек **C2**–**G2** соответственно.

Копирование формул в очередную строку продолжаем до тех пор, пока не достигнем требуемой точности решения, т.е. пока в столбце **G** не появится искомое значение корня.

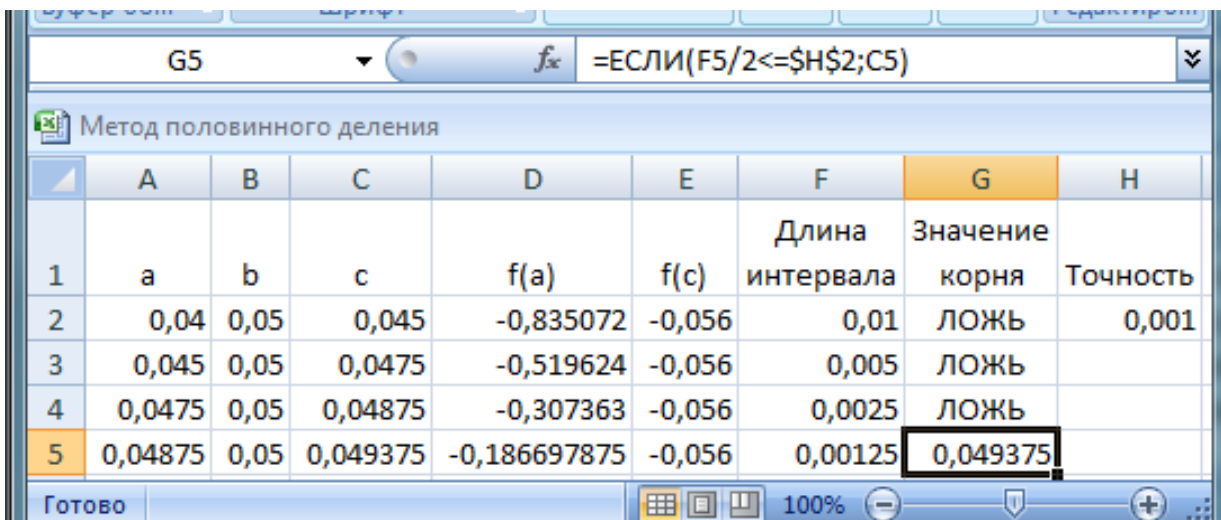
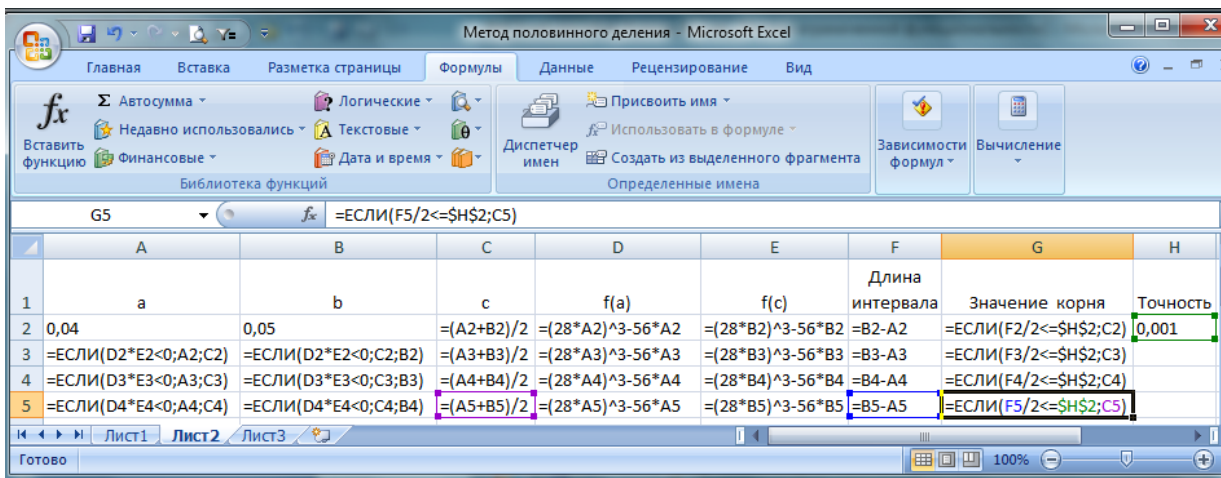


Рис. 17. Уточнение корня уравнения методом половинного деления

#### 4.4.3. Метод хорд (метод секущих)

В методе хорд (рис. 18), в отличие от метода половинного деления, интервал  $[a, b]$  делится не пополам, а в отношении  $|f(a)| : |f(b)|$ , что во многих случаях обеспечивает более быстрое нахождение корня. Метод основан на замене нелинейной функции  $f(x)$  на участке  $[a, b]$  хордой, проходящей через точки с координатами  $(a, f(a))$  и  $(b, f(b))$ , а в качестве приближенного значения корня берется точка пересечения хорды с осью абсцисс.

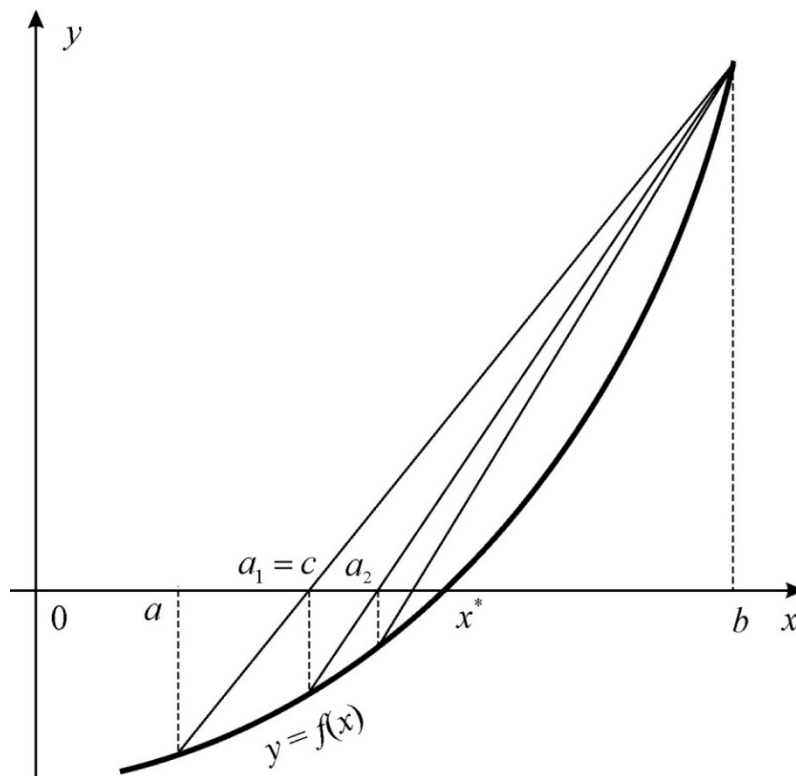


Рис. 18. Иллюстрация метода хорд

Запишем уравнения секущей, проходящей через точки  $(a, f(a))$  и  $(b, f(b))$ :

$$\frac{y - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a},$$

тогда в точке пересечения  $(c, 0)$  этой секущей с осью  $Ox$  получим:

$$\frac{-f(a)}{c - a} = \frac{f(b) - f(a)}{b - a} \Rightarrow c = a + \left( -\frac{f(a) \cdot (b - a)}{f(b) - f(a)} \right).$$

Следовательно, итерационная схема метода хорд представима в виде:

$$a_0 = a; \\ a_k = a_{k-1} + \Delta a_k, \quad \Delta a_k = \frac{-f(a_{k-1}) \cdot (b - a_{k-1})}{f(b) - f(a_{k-1})}, \quad k = 1, 2, \dots \quad (27)$$

Если про вторую производную функции  $f(x)$  ничего не известно, то метод хорд позволяет построить аналогичную итерационную схему, но после определения точки  $c$  интервал, на котором происходит дальнейшее определение корня, выбирается из условия противоположности знаков функции  $f(x)$  на его концах.

#### 4.4.4. Метод Ньютона (метод касательных)

Метод Ньютона (метод касательных) аналогичен методу хорд, но в качестве прямой берется касательная к кривой в точке начального приближения (рис. 19).

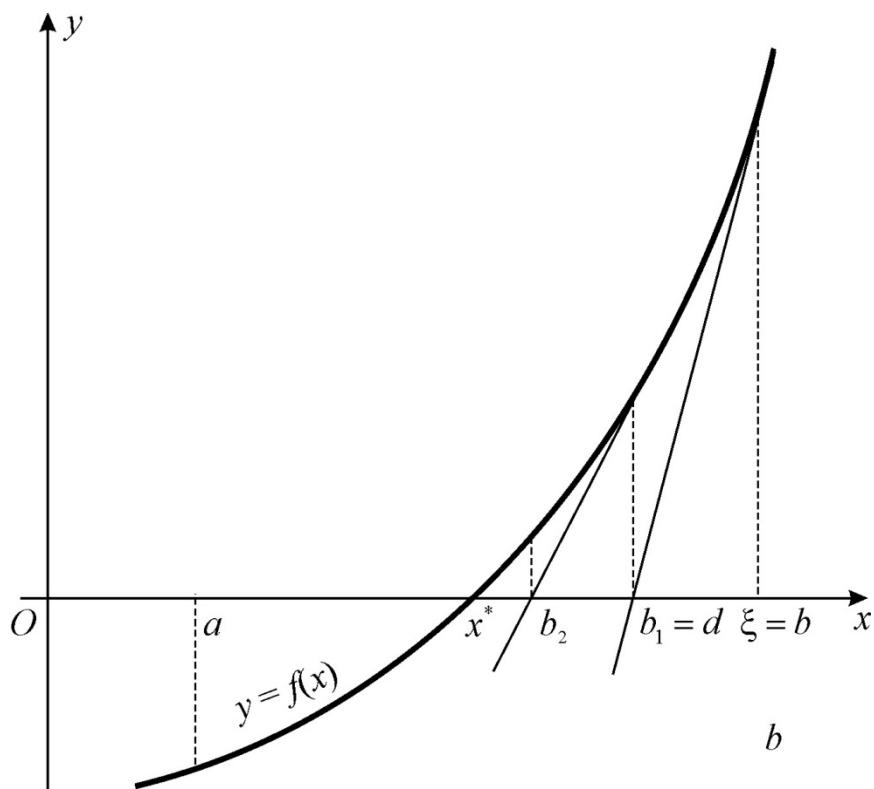


Рис. 19. Иллюстрация метода касательных

Метод касательных применим только в случае, если функция  $f(x)$  не имеет на интервале  $[a, b]$  точек перегиба, то есть  $f''(x)$  сохраняет знак. Точку  $\xi$  ( $a$  или  $b$ ), из которой начинается схема метода касательных, выбираем из условия  $f''(\xi) \cdot f(\xi) > 0$ .

Запишем уравнение касательной, выходящей из точки  $(b, f(b))$ :

$$\frac{y - f(b)}{x - b} = k = f'(b),$$

где  $k$  – тангенс угла наклона касательной к оси  $Ox$ . Пусть  $(0, d)$  – точка пересечения касательной и оси  $Ox$ , тогда  $d = b - \frac{f(b)}{f'(b)}$  и итерационная схема

метода касательных записывается в виде:

$$b_0 = b, \quad \Delta b_n = \frac{f(b_{n-1})}{f'(b_{n-1})}, \quad b_n = b_{n-1} - \Delta b_n. \quad (28)$$

Пример 27.

Для функции  $f(x) = (28x)^3 - 56x$  из примера 26 получим решение методом Ньютона с точностью  $\varepsilon = 0,001$ , используя *MS Excel*.

Решение.

Зададим заголовки столбцов. В следующую строку занесем значения: номер итерации 0; начальное приближение 0,04; формулы для  $f(x)$  и ее производной соответственно:  $= (28*B2)^3 - 56*B2$  и  $= 3*28^3*(B2)^2 - 56$ . В ячейку **E2** введем формулу  $= \text{ЕСЛИ}(\text{ABS}(C2) \leq 0,001; \text{ИСТИНА})$ .

Далее, в ячейку **A3** введем формулу  $= A2 + 1$ , чтобы увеличить номер итерации на 1, а в ячейку **B3** введем формулу для вычисления следующей итерации по методу Ньютона  $= B2 - C2/D2$  в соответствии с формулой (28). Значения ячеек **C3 – E3** получим, протянув вниз формулы из ячеек **C2 – E2** соответственно. После этого выделим диапазон ячеек **A3:E4** и протянем формулы вниз. Фрагмент рабочего листа, приведенный на рис. 20, показывает, что заданная точность достигнута на 4-й итерации.

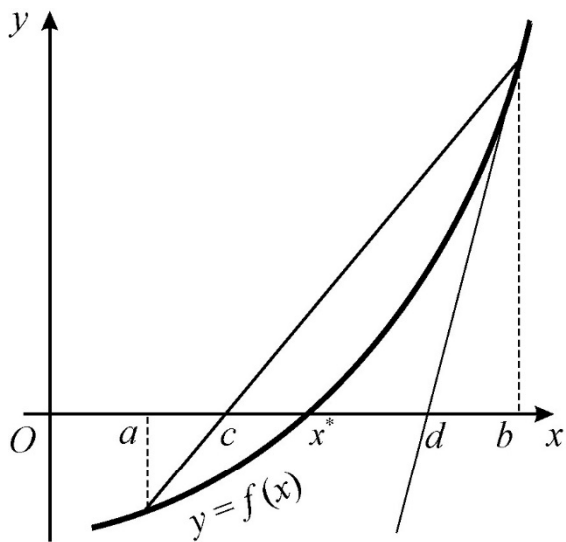
	A	B	C	D	E
	Номер итерации	x	f(x)	f'(x)	Достигнута точность?
1	0	0,04	-0,83507	49,3696	ЛОЖЬ
2	1	0,056915	0,85991	157,3262	ЛОЖЬ
3	2	0,051449	0,108391	118,3203	ЛОЖЬ
4	3	0,050533	0,002827	112,1678	ЛОЖЬ
5	4	0,050508	2,11E-06	112,0001	ИСТИНА

Рис. 20. Решение нелинейного уравнения методом Ньютона в *MS Excel*

#### 4.4.5. Комбинированный метод хорд и касательных

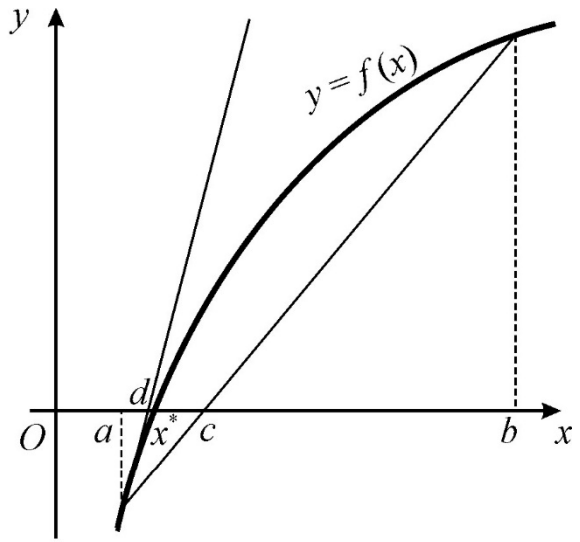
Комбинированный метод хорд и касательных (рис. 21) является одним из наиболее используемых методов поиска корней алгебраических уравнений. Идея метода состоит в одновременном применении метода хорд и метода касательных, которые дают приближение положительного корня с разных сторон. Комбинированный метод удобно использовать, если на отрезке  $[a, b]$ , содержащем только один корень, вторая производная  $f''(x)$  сохраняет знак. Постоянство знака  $f''(x)$  означает, что кривая либо выпуклая ( $f''(x) < 0$ ), либо вогнутая ( $f''(x) > 0$ ).

Рассмотрим 4 различных случая поведения функции на отрезке  $[a, b]$ , в зависимости от знака производных.



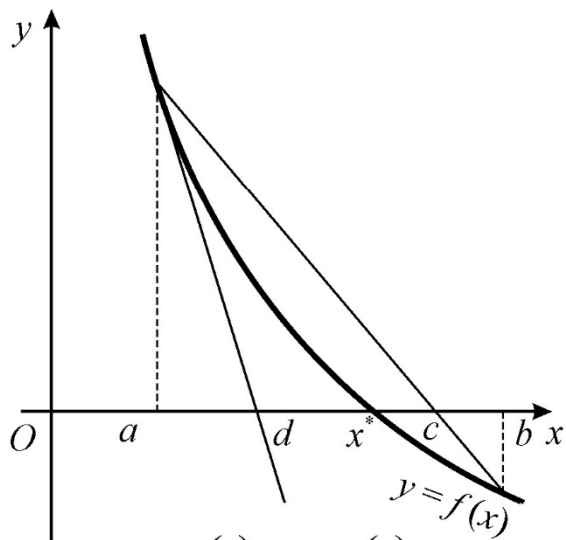
$$f'(x) > 0; f''(x) > 0$$

*a*



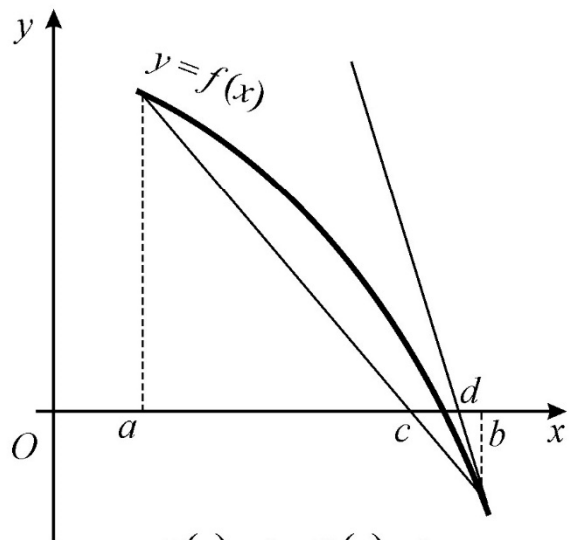
$$f'(x) > 0; f''(x) < 0$$

*б*



$$f'(x) < 0; f''(x) > 0$$

*в*



$$f'(x) < 0; f''(x) < 0$$

*г*

Рис. 21. Иллюстрация комбинированного метода хорд и касательных

В 1-м случае (рис. 21, *a*), когда на отрезке  $[a, b]$  функция  $f(x)$  монотонно возрастает ( $f'(x) > 0$ ), а кривая  $y = f(x)$  вогнута ( $f''(x) > 0$ ), имеем: касательная пересекает ось  $Ox$  со стороны выпуклости (справа), а хорда – со стороны вогнутости графика функции  $y = f(x)$ , слева.

Абсциссы точек пересечения вычисляются по формулам:

$$a_{k+1} = a_k + \Delta a_k; \quad b_{k+1} = b_k - \Delta b_k \quad (k = 0, 1, 2, \dots), \quad (29)$$

где  $\Delta a_k = -\frac{(b_k - a_k)f(a_k)}{f(b_k) - f(a_k)}, \quad \Delta b_k = -\frac{f(b_k)}{f'(b_k)}.$

Процесс вычисления заканчивается на  $m$ -м приближении, когда выполняется неравенство  $|a_m - b_m| < \varepsilon$ , где  $\varepsilon$  – точность вычислений, в качестве корня определяется величина  $x^* = \frac{a_m + b_m}{2}$ .

Аналогично использование формул для случая 4 (рис. 21, з), когда функция  $f(x)$  монотонно убывает ( $f'(x) < 0$ ), а кривая  $y = f(x)$  вогнута ( $f''(x) > 0$ ).

Во 2-м случае (рис. 21, б), когда на отрезке  $[a, b]$  функция  $f(x)$  монотонно возрастает ( $f'(x) > 0$ ), а кривая  $y = f(x)$  выпукла ( $f''(x) < 0$ ), касательная пересекает ось  $Ox$  слева, а хорда – справа.

Абсциссы точек пересечения вычисляются по формулам:

$$a_{k+1} = a_k + \Delta a_k; \quad b_{k+1} = b_k + \Delta b_k \quad (k = 0, 1, 2, \dots), \quad (30)$$

где  $\Delta a_k = -\frac{f(a_k)}{f'(a_k)}, \quad \Delta b_k = -\frac{(b_k - a_k)f(b_k)}{f(b_k) - f(a_k)}.$

Аналогично использование формул для случая 3 (рис. 21, в), когда функция  $f(x)$  монотонно убывает ( $f'(x) < 0$ ), а кривая  $y = f(x)$  выпукла ( $f''(x) > 0$ ).

### Контрольные вопросы и задания

1. Для чего и каким образом выполняют отделение корней нелинейного уравнения?
2. Что такое интервал изоляции корня нелинейного уравнения?
3. Какая теорема используется при исследовании корней нелинейного уравнения?
4. Какая функция называется сжимающим отображением?
5. Что является критерием остановки итерационного процесса?
6. Какие два метода используются в комбинированном методе при решении нелинейного уравнения и для чего?
7. Всегда ли при решении нелинейного уравнения метод Ньютона дает более быструю сходимость, чем метод дихотомии? Приведите примеры.

## 5. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

### 5.1. Общие сведения

В общем виде система нелинейных уравнений (СНУ) может быть записана следующим образом:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0; \\ f_2(x_1, x_2, \dots, x_n) = 0; \\ \dots \\ f_n(x_1, x_2, \dots, x_n) = 0. \end{cases} \quad (31)$$

Если обозначить  $X = (x_1 \dots x_n)$ , а  $F = \begin{pmatrix} f_1 \\ \dots \\ f_n \end{pmatrix}$ , то в векторной форме

получим уравнение  $F(X) = 0$ . Для дальнейших рассмотрений нам понадобится матрица частных производных вектора-функции  $F(X)$  – *матрица Якоби*

$$\left( \frac{\partial F}{\partial X} \right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}. \quad (32)$$

Норма матрицы Якоби в точке  $X^{(0)}$  определяется по формуле

$$\left\| \frac{\partial F}{\partial X}(X^{(0)}) \right\| = \max_{1 \leq i \leq n} \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(X^{(0)}) \right|.$$

Предположим, что векторное уравнение  $F(X) = 0$  преобразовано к виду

$$X = \Phi(X), \text{ где } \Phi(X) \text{ – вектор-функция } \Phi = \begin{pmatrix} \Phi_1 \\ \dots \\ \Phi_n \end{pmatrix}, \text{ таким образом, что оба}$$

уравнения имеют один и тот же корень. В этом случае можно организовать поиск корня по итерационной схеме  $X^{(n)} = \Phi(X^{(n-1)})$ .

Аналогично случаю одного нелинейного уравнения можно показать, что этот итерационный процесс будет сходящимся, если для любых двух точек  $X'$  и  $X''$  из области поиска корня  $G$  выполняется неравенство  $\|\Phi(X') - \Phi(X'')\| < \|X' - X''\| \cdot \theta$ , где  $0 < \theta < 1$  – аналог условия Липшица.



Вектор-функция  $\Phi$ , обладающий такими свойствами, называется сжимающим (сжатым) отображением.

Имеет место следующая теорема.

Для того чтобы итерационный процесс  $X^{(n)} = \Phi(X^{(n-1)})$ ,  $n = 1, 2, \dots$  был сходящимся в области поиска решения  $G$ , то есть  $\{X^{(n)}\} \rightarrow X^* = (x_1^*, x_2^*, \dots, x_n^*)$ , где  $X^*$  – точное решение системы  $X^0$ ,  $X^* \in G$ , достаточно, чтобы выполнялось неравенство  $\max_{X \in G} \|\Phi'(X)\| < 1$  или

$$\max_{X \in G} \max_i \sum_{j=1}^n \left| \frac{\partial \varphi_i}{\partial x_j} \right| < 1.$$

Доказательство.

Докажем, что в этом случае отображение будет сжимающим, что обеспечит сходимость итерационной схемы.

Введем в рассмотрение функцию  $V(s) = \varphi_i(X' + s(X'' - X'))$ , где  $s$  – некоторый параметр,  $0 \leq s \leq 1$ . Тогда  $\varphi_i(X'') - \varphi_i(X') = V(1) - V(0) = V'(s')$ , где  $s'$  – некоторое значение между 0 и 1. Это равенство записано в соответствии с одной из теорем математического анализа (следствие теоремы Лагранжа).

Производная  $V'(s)$  вычисляется по формуле

$$V'(s) = \sum_{k=1}^n \frac{\partial \varphi_i}{\partial x_k} \cdot \frac{dx_k}{ds},$$

а  $\frac{dx_k}{ds} = x'_k - x''_k$ , так как  $x_k(s) = x'_k + s(x''_k - x'_k)$ .

Следовательно,  $|\varphi_i(X') - \varphi_i(X'')| \leq \sum_{k=1}^n \left| \frac{\partial \varphi_i}{\partial x_k} \right| \cdot |x'_k - x''_k|$ .

Переходя к нормам векторов-функций, получим:

$$\|\Phi(X') - \Phi(X'')\| < \max_X \|\Phi'(X)\| \cdot \|X' - X''\|.$$

Значит, можно принять  $\theta = \max \|\Phi'(X)\|$ , и вектор-функция  $\Phi$  будет сжимающим отображением, когда  $\|\Phi'(X)\| < 1$ , что и требовалось доказать.

Пример 28.

Определить графически и уточнить средствами *MS Excel* корень системы нелинейных уравнений

$$\begin{cases} x_1^2 - x_2 = 0; \\ \sin x_1 - x_2 = 0. \end{cases}$$

Решение.

Выполним решение системы в *MS Excel* аналогично решению одного уравнения (пример 23). Для этого на этапе локализации корня необходимо

протабулировать функции, построить их графики и выявить точку пересечения (рис. 22). Получим зависимости  $x_2$  от  $x_1$ : для 1-го уравнения  $x_2 = x_1^2$ ; для 2-го уравнения  $x_2 = \sin x_1$ .

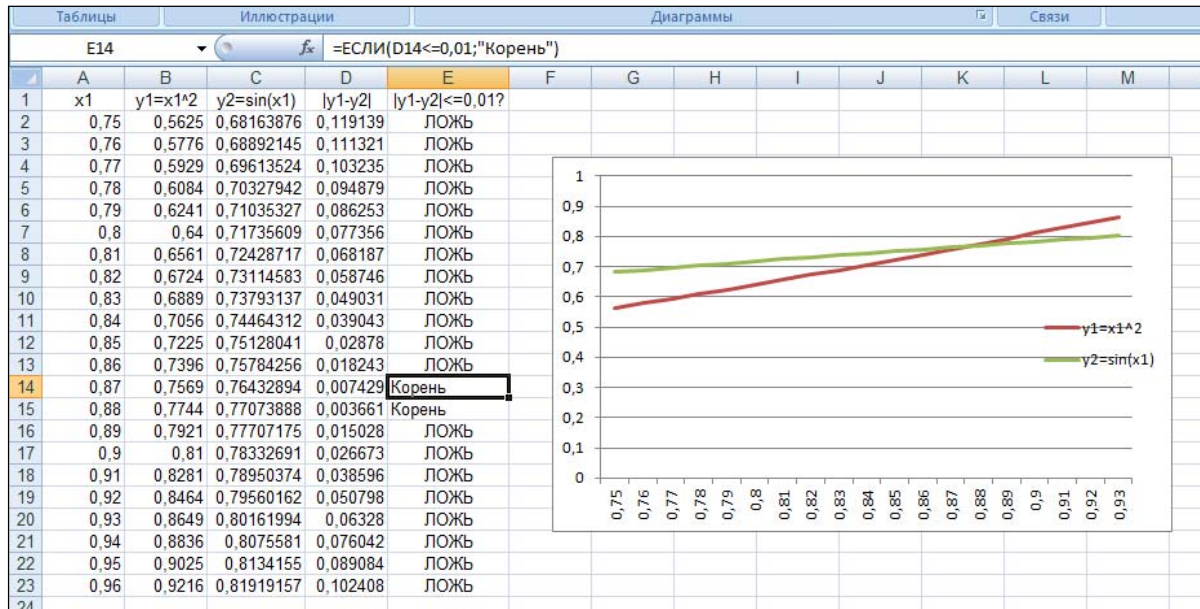


Рис. 22. Графическое решение системы нелинейных уравнений в MS Excel

Из графика видно, что  $x_1$  находится между значениями 0,87 и 0,88 (можно принять  $x_1 = 0,875$ ), а  $x_2 \approx 0,76$ . Для уточнения применяем процедуру «Поиск решения» (рис. 23).

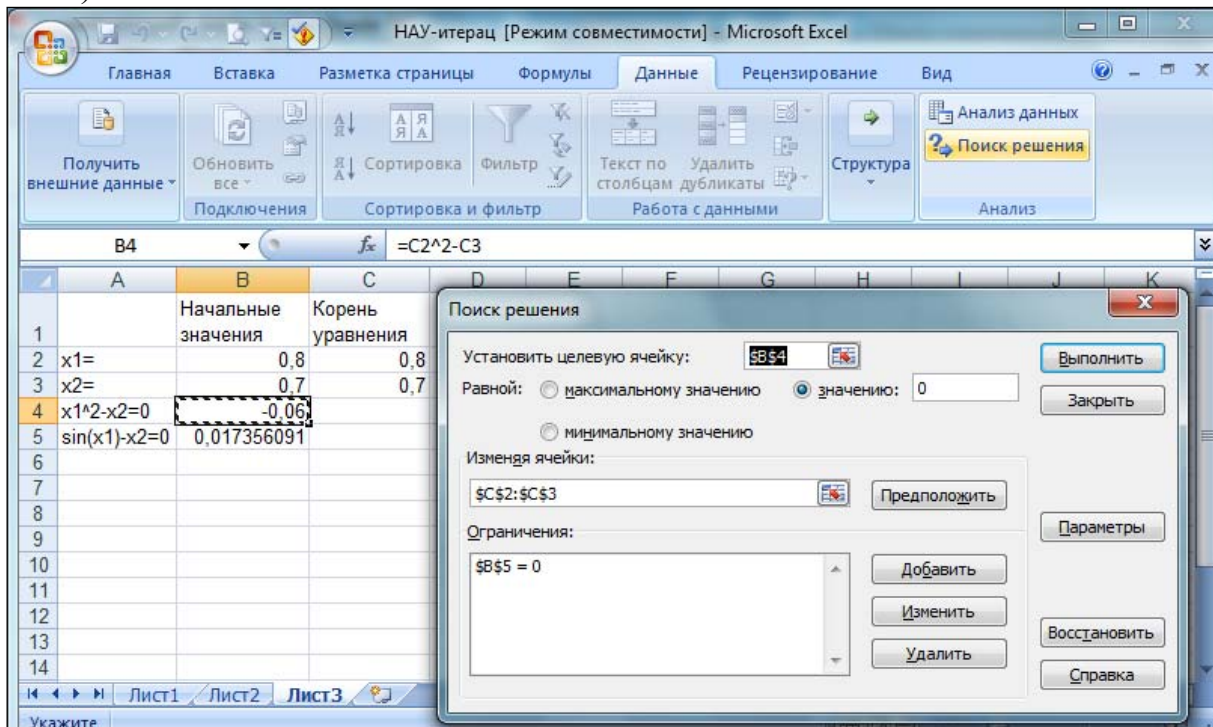
Для выполнения этапа уточнения значений  $x_1$  и  $x_2$ , которые соответствуют равенству в исходных уравнениях, выполним подготовительные действия. В ячейках **B2** и **B3**, а также в ячейках **C2** и **C3** запишем приближенные значения  $x_1$  и  $x_2$  (рис. 23, а). Причем, значения в ячейках **B2** и **B3** – начальные приближения, они останутся неизменными. А значения в ячейках **C2** и **C3** будут изменены процедурой «Поиск решения». Соответственно в формулах для функций системы, содержащихся в ячейках **B4** и **B5**, используются именно ячейки **C2** и **C3**.

После этого вызывается процедура «Поиск решения» и заполняется задание на поиск: в качестве оптимизируемой выбирается ячейка **B4**, а для обеспечения значения 0 в ячейке **B5** задается соответствующее ограничение (рис. 23, а). В результате получено решение (рис. 23, б, в).

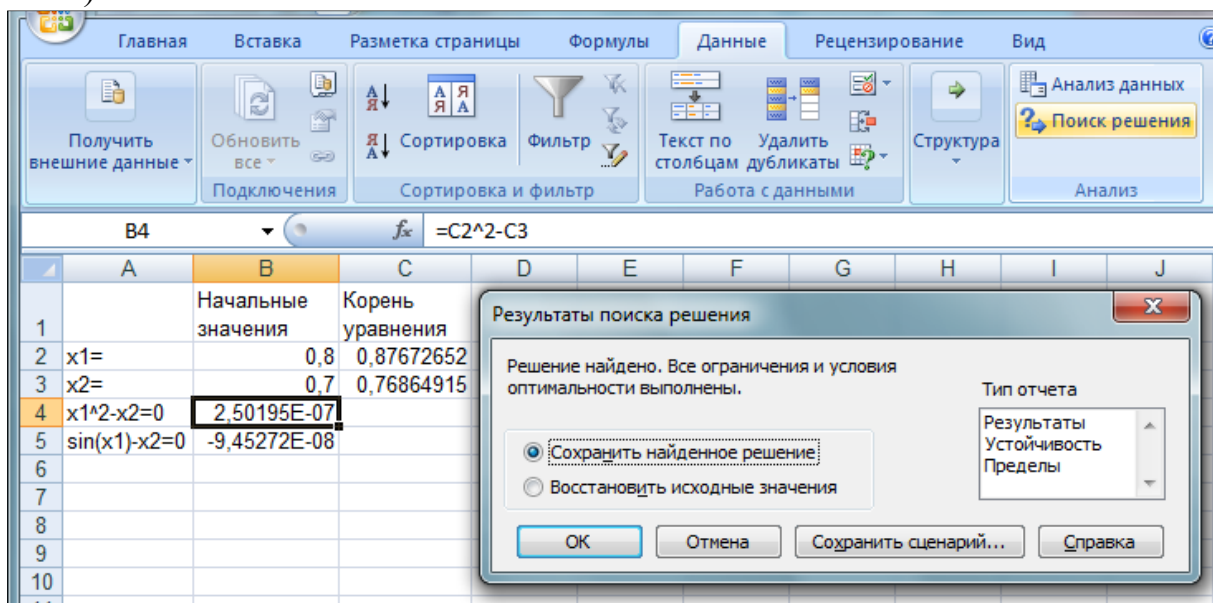
Значения в ячейках **C2** и **C3** изменены процедурой «Поиск решения» таким образом, что в ячейках **B4** и **B5**, содержащих формулы для функций системы нелинейных уравнений, получены значения, очень близкие к нулю.

В лабораторном практикуме решение данной системы различными методами приводится с использованием *MathCAD* (см. пример 11 [21]).

a)



б)



в)

	Начальные значения	Корень уравнения
1		
2	x1= 0,8	0,87672652
3	x2= 0,7	0,76864915
4	x1 <sup>2</sup> -x2=0	2,50195E-07
5	sin(x1)-x2=0	-9,45272E-08
6		

Рис. 23. Уточнение решения системы нелинейных уравнений в MS Excel

## 5.2. Метод Ньютона

Итерационную схему метода Ньютона для решения СНУ можно получить так же, как и схему для решения одного нелинейного уравнения. Построим итерационную схему Ньютона исходя из разложения функций нескольких переменных  $f_i(X)$ ,  $X = (x_1, \dots, x_n)$ , имеющих непрерывные производные любого порядка, в ряды Тейлора до 1-го порядка малости:

$$f_1(X + \Delta X) \approx f_1(X) + \sum_{k=1}^n \frac{\partial f_1}{\partial x_k}(X) \Delta x_k;$$

...

$$f_n(X + \Delta X) \approx f_n(X) + \sum_{k=1}^n \frac{\partial f_n}{\partial x_k}(X) \Delta x_k,$$

или в векторной форме:

$$F(X + \Delta X) \approx F(X) + F'(X)(\Delta X)^T, \quad \Delta X = \{\Delta x_1, \Delta x_2, \dots, \Delta x_n\}, \quad (33)$$

где  $F'(X)$  – матрица Якоби в точке  $X$ .

Значение  $\Delta X$  будем определять из условия равенства нулю значения функции  $F(X + \Delta X) = 0$ , к чему мы и стремимся.

$$F(X) + F'(X)(\Delta X)^T = 0 \Rightarrow F'(X)(\Delta X)^T = -F(X) \Rightarrow (\Delta X)^T = -(F'(X))^{-1} F(X)$$

Тогда, считая  $\Delta x_k = x_k^{(n)} - x_k^{(n-1)}$ ,  $X = X^{(n)}$ , получаем итерационную схему в виде

$$X^{(n)} = X^{(n-1)} - (F'(X^{(n-1)}))^{-1} \cdot F(X^{(n-1)}). \quad (34)$$

Эта итерационная схема является наиболее быстроходящейся, если начальное приближенное решение  $X^{(0)}$  выбрано достаточно близким к решению.

Существует следующая *теорема сходимости для метода Ньютона*.

**Пусть точка  $X^{(0)}$  принадлежит некоторой области поиска корня  $G$ , и в этой точке выполняются следующие условия:**

**1) каждая  $f_i(X)$  имеет непрерывные вторые производные в области  $G$  по всем  $x_k$ , и выполняется неравенство**

$$\left| \frac{\partial^2 f_i(X)}{\partial x_k^2} \right| \leq L, \quad L = \text{const};$$

**2)  $f'_i(X^{(0)}) \neq 0$ ;**

**3)  $|f_i(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})| < \delta$ ;**

**4)  $\|(F'(X^{(0)}))^{-1}\| < M$ ,  $M = \text{const}$ ;**

**5) выполняется неравенство  $h = M^2 L \delta n^2 \leq \frac{1}{2}$ .**

Тогда итерационная схема Ньютона от начального значения  $X^{(0)}$  сойдется к решению системы нелинейных уравнений  $X^*$  ( $F(X^*) = 0$ ).

**З а м е ч а н и е .** При решении СНУ очень сложно отделить корни; поэтому обычно выбирают начальное приближение ближе к предполагаемому решению. Для получения всех возможных решений, если таковых несколько, на практике чаще всего решают задачу, начиная решение из различных начальных приближений, распределенных по всей области допустимых значений  $G$ .

В случае системы двух уравнений начальное приближение удобно находить графически, определив координаты точки пересечения кривых, описываемых уравнениями  $f_1(x_1, x_2) = 0$ ;  $f_2(x_1, x_2) = 0$ .

### 5.3. Метод Ньютона – Рафсона

Итерационная схема этого метода имеет вид:

$$X^{(k+1)} = X^{(k)} - \lambda_k (F'(X^{(k)}))^{-1} F(X^{(k)}), \quad \lambda_k > 0. \quad (35)$$

Здесь  $\{\lambda_k\}$  – или убывающая последовательность, или  $\lambda_k = \text{const}$ , или  $\lambda_k$  выбирается из условия минимизации нормы вектора-функции  $\|F(X^{(k+1)}(\lambda_k), \lambda_k)\|$  по переменной  $\lambda_k$ . Как правило,  $\lambda_k < 1$ .

### 5.4. Метод регуляризации Левенберга – Марквардта

Метод имеет следующую итерационную схему:

$$X^{(k+1)} = X^{(k)} - [F'(X^{(k)}) + \lambda_k E]^{-1} F(X^{(k)}), \quad (36)$$

где  $\lambda_k = \text{const}$ ;  $E$  – единичная матрица,  $F'(X)$  – матрица, плохо обусловленная в области поиска решения.

Слагаемое  $\lambda_k E$  добавляет величины  $\lambda_k$  на диагональ матрицы  $F$ , что приводит к улучшению обусловленности матрицы Якоби  $F'(X)$ . Значение  $\lambda_k$  выбирается из условия минимизации нормы вектора-функции  $\|F(X^{(k+1)}(\lambda_k), \lambda_k)\|$  по переменной  $\lambda_k$ .

### 5.5. Применение градиентных методов для решения СНУ

Можно решать СНУ с помощью сведения ее к задаче минимума суммы квадратов левой части системы (31):

$$S(x_1, x_2, \dots, x_n) = \sum_{i=1}^n [f_i(x)]^2. \quad (37)$$

Очевидно, что каждое решение системы (31) обращает в нуль функцию  $S(x_1, x_2, \dots, x_n)$ . То есть если мы найдем вектор  $X^*(x_1^*, x_2^*, \dots, x_n^*)$  такой, что  $S(X) \approx 0$ , то этот вектор можно принять за решение СЧУ.

Если функции  $f_i(x_1, x_2, \dots, x_n)$  действительны и дифференцируемы в их области поиска решения, то в некоторых случаях минимизацию функции  $S(x_1, x_2, \dots, x_n)$  эффективно проводить градиентными методами.

Градиентом функции  $\Phi(x_1, x_2, \dots, x_n)$  называют вектор  $\nabla\Phi(X)$ , приложенный в точке  $X$  и имеющий направление нормали  $n$  к поверхности уровня функции в данной точке и длину, равную  $\|\nabla\Phi(X)\|$ , где

$$\nabla\Phi(X) = \left( \frac{\partial\Phi}{\partial x_1}(X), \frac{\partial\Phi}{\partial x_2}(X), \dots, \frac{\partial\Phi}{\partial x_n}(X) \right)^T.$$

То есть градиент представляет собой вектор частных производных, показывающий направление наискорейшего возрастания функции.

Поверхностью уровня функции  $\Phi(X)$  называется множество точек, в которых функция принимает постоянное значение, т.е.  $\Phi(X) = \text{const}$ . Если  $n = 2$ , поверхность уровня изображается линией уровня на плоскости.

Вектор  $-\nabla\Phi(X)$ , противоположный вектору градиента, называется антиградиентом функции.

Точка, в которой градиент функции  $\nabla f(x^{(k)}) = 0$ , называется стационарной.

Идея градиентных методов (наискорейшего градиентного спуска, сопряженных градиентов) заключается в том, чтобы исходя из начальной точки, двигаться в направлении антиградиента к стационарной точке, которая и будет решением СЧУ.

Итерационная схема метода градиентного спуска:

$$X^{(k+1)} = X^{(k)} - \lambda_k \frac{\nabla\Phi(X^{(k)})}{\|\nabla\Phi(X^{(k)})\|}, \quad (38)$$

где параметр  $\lambda_k$  может быть выбран одним из следующих способов:

- 1)  $\lambda_k = \text{const}$ ;
- 2)  $\lambda_k \rightarrow 0$  при  $k \rightarrow \infty$ ;
- 3) из условия минимума функции  $\Phi(\lambda_k)$  по переменной  $\lambda_k$ :

$$\left\| \Phi(X^{(k+1)}(\lambda_k), \lambda_k) \right\|^2 \rightarrow \min_{\lambda_k}.$$

## Контрольные вопросы и задания

1. Каким образом выполняется процедура отделения корней при решении одного нелинейного уравнения, системы двух нелинейных уравнений в *MS Excel*?
2. Что такое система нелинейных уравнений, и какие этапы решения ее можно выделить?
3. Какая процедура используется в *MS Excel* для уточнения корней при решении СНУ?
4. Какова цель этапа локализации (отделения) различных решений системы нелинейных уравнений и зачем этот этап необходимо выполнять?
5. Сколько решений систем нелинейных уравнений может существовать?
6. Как зависит сходимость методов Ньютона от выбора точки начального приближения?
7. Что такое норма матрицы Якоби системы нелинейных уравнений?
8. Объясните действия, выполняемые процедурой «Поиск решения», имеющейся в программе *MS Excel*.
9. Объясните действия, выполняемые процедурой *Find( )* программы *MathCAD*.

## 6. ИНТЕРПОЛИРОВАНИЕ ФУНКЦИЙ

### 6.1. Постановка задачи

Пусть  $x_i \in [a, b]$ ,  $i = 0, 1, 2, \dots, n$  – множество значений аргумента, а  $y_i \in R$  – множество значений некоторой функции  $\varphi(x)$ , аналитическое выражение которой в общем случае неизвестно. Требуется по заданным точкам  $(x_i, y_i)$  определить функцию  $y = f(x)$  такую, чтобы:

1)  $f(x_i) = y_i$ ;

2) отличие значений этой функции  $f(x)$  от значений точной функции  $\varphi(x)$  во всех других точках интервала  $[a, b]$  было минимальным, то есть

$$\max_{x \in [a, b]} |f(x) - \varphi(x)| \rightarrow \min.$$

Точки  $(x_i, y_i)$  называются *узлами интерполяции*. Задача по отысканию функции  $f$  называется задачей интерполяции, а сама функция  $f$  – интерполянт. При этом различаются понятия *интерполяции* (интерполирования) – нахождения значения таблично заданной функции внутри интервала интерполирования  $[a, b]$ , и *экстраполяции* (экстраполирования), которое сводится к построению функции в точках за пределами заданного интервала.

Задачи интерполяции возникают при рассмотрении следующих вопросов:

1) при обработке экспериментальных данных для выяснения общих закономерностей поведения изучаемой системы;

2) при проведении многократных вычислений в достаточно узком интервале изменения аргумента с целью упрощения сложных вычислительных формул;

3) при решении задач дифференциального исчисления, таких, как вычисление производных, вычисление определенных интегралов, решение дифференциальных уравнений.

Интерполяция применяется также в задаче *обратного интерполирования*: для таблично заданной функции  $y_i = f(x_i)$  найти  $x$  как функцию от  $y$ . Например, обратное интерполирование можно использовать для нахождения корней алгебраического уравнения.

Общий подход к решению задачи интерполяции заключается в следующем.

*Интерполянта* определяется в виде линейной комбинации линейно независимых функций:

$$f(x) = \sum_{k=0}^n C_k \cdot \Phi_k(x).$$



Здесь  $\Phi_k(x)$  – система известных линейно независимых функций, вид которых устанавливается исследователем. Постоянные  $C_k$  неизвестны, в их определении и заключается задача интерполяции.

Так как для системы точек  $(x_i, y_i)$  должно выполняться условие  $y_i = f(x_i)$ , то можно записать систему, состоящую из  $(n + 1)$ -го уравнения с  $(n + 1)$ -м неизвестным  $c_k$ :

$$\begin{cases} y_0 = \sum c_k \cdot \Phi_k(x_0); \\ y_1 = \sum c_k \cdot \Phi_k(x_1); \\ \vdots \\ y_n = \sum c_k \cdot \Phi_k(x_n). \end{cases}$$

Данная система линейных уравнений будет иметь решение, когда определитель коэффициентов системы не равен нулю:

$$\begin{vmatrix} \Phi_0(x_0) & \Phi_1(x_0) & \dots & \Phi_n(x_0) \\ \Phi_0(x_1) & \Phi_1(x_1) & \dots & \Phi_n(x_1) \\ \dots & \dots & \dots & \dots \\ \Phi_0(x_n) & \Phi_1(x_n) & \dots & \Phi_n(x_n) \end{vmatrix} \neq 0.$$

Для выбора функций  $\Phi_k$  при построении интерполянта чаще всего используются следующие системы функций:

- 1)  $1, x, x^2, \dots, x^n$  – система степенных функций;
- 2)  $\{\cos kx, \sin kx\}, k = 0, \dots, n$  – система тригонометрических функций;
- 3)  $\frac{\alpha_0 + \alpha_1 \cdot x + \alpha_2 \cdot x^2 + \dots + \alpha_s \cdot x^s}{\beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \dots + \beta_p \cdot x^p}$  – система дробно-рациональных

функций.

В первом случае мы имеем дело с полиномиальной интерполяцией, во втором – с тригонометрической интерполяцией и, наконец, в третьем – получаем дробно-рациональную интерполяцию. Кроме перечисленных способов часто используется так называемая сплайн-интерполяция, принципы которой мы рассмотрим далее.

## 6.2. Полиномиальная интерполяция

В этом случае приближение таблично заданной функции осуществляется с помощью системы степенных функций, т.е. интерполянта  $f(x)$  определяется в виде:

$$f(x) = c_0 + c_1 \cdot x + c_2 \cdot x^2 + \dots + c_n \cdot x^n = P_n(x),$$

где  $P_n(x)$  – полином степени  $n$ .

Задача состоит в поиске коэффициентов  $c_0, c_1, \dots, c_n$ , удовлетворяющих системе уравнений  $y_i = P_n(x_i)$ ,  $i = 0, \dots, n$ .

Имеет место следующая теорема (*теорема Вейерштрасса*).

**Для любого как угодно малого числа  $\varepsilon$  существует такой полином степени  $n = n(\varepsilon)$ , что  $\max_{x \in [a, b]} |P(x) - \varphi(x)| < \varepsilon$ , где  $\varphi(x)$  – точное выражение искомой функции.**

Теорема Вейерштрасса представляет собой теоретический результат, который не дает практического способа построения интерполяционного полинома, однако свидетельствует о том, что построение такого полинома возможно.

Рассмотрим интерполяционный полином  $P_n(x) = \sum_{k=0}^n c_k \cdot x^k$  такой, что

$P_n(x_i) = y_i$ , то есть

$$\begin{cases} c_0 + c_1 \cdot x_0 + c_2 \cdot x_0^2 + \dots + c_n \cdot x_0^n = y_0; \\ c_0 + c_1 \cdot x_1 + c_2 \cdot x_1^2 + \dots + c_n \cdot x_1^n = y_1; \\ \dots \\ c_0 + c_1 \cdot x_n + c_2 \cdot x_n^2 + \dots + c_n \cdot x_n^n = y_n. \end{cases}$$

Здесь неизвестными являются  $c_0, c_1, \dots, c_n$ . Определитель матрицы коэффициентов

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix}$$

называется *определителем Вандермонда*.

Существует строгое математическое доказательство, что такой определитель при любых значениях  $x_i$  (таких, что  $x_i \neq x_j$  для всех  $i \neq j$ ) не обращается в нуль (доказать самостоятельно). Следовательно, система линейных алгебраических уравнений для определения коэффициентов  $c_i$  всегда имеет единственное решение. Таким образом, интерполяционный полином для интерполяции табличной функции, значения которой известны в  $(n + 1)$  точках, всегда существует, единственен и имеет степень, равную  $n$ .

### 6.2.1. Интерполяционный полином Лагранжа

В качестве базисных рассмотрим полиномы степени  $n$ , удовлетворяющие следующему условию:

$$l_k^n(x) = \begin{cases} 1, & \text{если } x = x_j, \quad j = k; \\ 0, & \text{если } x = x_j, \quad j \neq k. \end{cases}$$

Такому условию удовлетворяют полиномы вида:

$$l_k^n(x) = \frac{(x - x_0) \cdot (x - x_1) \cdot \dots \cdot (x - x_{k-1}) \cdot (x - x_{k+1}) \cdot \dots \cdot (x - x_n)}{(x_k - x_0) \cdot (x_k - x_1) \cdot \dots \cdot (x_k - x_{k-1}) \cdot (x_k - x_{k+1}) \cdot \dots \cdot (x_k - x_n)} = \quad (39)$$

$$= d_0 + d_1 \cdot x + d_2 \cdot x^2 + \dots + d_n \cdot x^n.$$

Тогда в качестве интерполяционного полинома можно взять сумму вида  $L_n(x) = \sum_{k=0}^n y_k \cdot l_k^n(x)$ . Имеем  $L_n(x_i) = \sum_{k=0}^n y_k \cdot l_k^n(x_i) = y_i$ . Таким образом, значения функции  $L_n(x)$  в узлах интерполяции  $x_i$  совпадают с заданными значениями  $y_i$ .

Полином

$$L_n(x) = \sum_{k=0}^n y_k \cdot \prod_{\substack{i=0 \\ i \neq k}}^n \frac{(x - x_i)}{(x_k - x_i)} \quad (40)$$

называется *интерполяционным полиномом Лагранжа*.

### Оценка погрешности интерполяции полиномом Лагранжа

Если известно, что интерполируемая функция  $\varphi(x)$  имеет непрерывные частные производные до  $(n+1)$ -го порядка включительно, то разность между точным значением функции и её полиномиальным приближением  $\varphi(x) - L_n(x)$  в любой точке  $x$  из интервала интерполирования  $x \in [a, b]$

можно оценить по формуле  $\varphi(x) - L_n(x) = \frac{\varphi^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^n (x - x_j)$ , где  $\xi \in [a, b]$ .

Из этой формулы вытекает, что

$$|R_n(x)| = |\varphi(x) - L_n(x)| \leq \max_{x \in [a, b]} \left| \frac{\varphi^{(n+1)}(x)}{(n+1)!} \prod_{j=0}^n |x - x_j| \right|. \quad (41)$$

### 6.2.2. Интерполяционный полином Ньютона

Обобщением понятия производной является понятие *разделенной разности*. Разделенная разность нулевого порядка совпадает со значением

самой функции  $y(x_i)$ . Разделенной разностью первого порядка называется величина  $y(x_i, x_j) = \frac{y(x_i) - y(x_j)}{x_i - x_j}$ .

Аналогично, величина  $y(x_i, x_j, x_k) = \frac{y(x_i, x_j) - y(x_j, x_k)}{x_i - x_k}$  называется разделенной разностью второго порядка, а  $y(x_0, x_1, \dots, x_k) = \frac{y(x_0, x_1, \dots, x_{k-1}) - y(x_1, x_2, \dots, x_k)}{x_0 - x_k}$  – разделенной разностью порядка  $k$ .

Таблицу разделенных разностей функции  $y(x_i)$ ,  $i = 1, \dots, n$  можно представить в виде следующей схемы:

$$\begin{array}{ccccccc}
 & & & & & & y(x_1) \\
 & & & & & & \\
 & & & & & & y(x_1, x_2) \\
 & & & & & & \\
 y(x_2) & & & & & & y(x_1, x_2, x_3) \\
 & & & & & & \\
 & & & & & & y(x_2, x_3) \quad \vdots \quad y(x_1, x_2, \dots, x_n) \\
 & & & & & & \\
 & & & & & & \vdots \\
 & & & & & & \\
 & & & & & & y(x_{n-1}, x_n) \\
 & & & & & & \\
 & & & & & & y(x_n)
 \end{array}$$

Предположим, что функция  $y$  представляет собой полином  $y = P_n(x)$  порядка  $n$ , тогда разделенные разности также являются полиномами:

$$P(x, x_0) = \frac{P(x) - P(x_0)}{x - x_0} \quad \text{– полином порядка } (n - 1),$$

$$P(x, x_0, x_1) = \frac{P(x, x_0) - P(x_0, x_1)}{x - x_1} \quad \text{– полином порядка } (n - 2),$$

$$P(x, x_0, \dots, x_k) = \frac{P(x, x_0, \dots, x_{k-1}) - P(x_0, \dots, x_k)}{x - x_k} \quad \text{– полином порядка } (n - k - 1).$$

Используя приведенные разности, можно записать полином  $P_n(x)$  в следующей форме:

$$\begin{aligned}
 P(x) &= P(x_0) + P(x, x_0)(x - x_0) = \\
 &= P(x_0) + (x - x_0)P(x_0, x_1) + (x - x_0)(x - x_1)P(x, x_0, x_1) = \dots = \\
 &= P(x_0) + (x - x_0) \cdot P(x_0, x_1) + (x - x_0)(x - x_1) \cdot P(x_0, x_1, x_2) + \dots + \\
 &\quad + (x - x_0)(x - x_1) \dots (x - x_{n-1}) \cdot P(x_0, x_1, \dots, x_n) = \\
 &= P(x_0) + \sum_{k=1}^n (x - x_0) \cdot (x - x_1) \cdot \dots \cdot (x - x_{k-1}) \cdot P(x_0, x_1, \dots, x_k).
 \end{aligned}$$

Если такой полином  $P_n(x)$  взять в качестве интерполяционного полинома для функции  $y$ , то, так как в узлах интерполяции  $P_n(x_i) = y_i$ , совпадут и все разделенные разности, вычисленные в этих точках для функций  $P_n(x)$  и  $y$ . Тогда интерполяционный полином можно записать в виде:

$$P_n(x) = y_0 + \sum_{k=1}^n (x - x_0) \cdot (x - x_1) \cdot \dots \cdot (x - x_{k-1}) \cdot y(x_0, x_1, \dots, x_k). \quad (42)$$

В силу способа построения  $P_n(x_i) = y_i$ , а  $P_n(x)$  представляет собой интерполяционный полином, называемый *интерполяционным полиномом Ньютона*.

Если узлы интерполяции на интервале  $[a, b]$  расставлены равномерно, т.е. шаг интерполяции  $h = x_i - x_{i-1}$  для всех  $i = 0, 1, \dots, n$ , то интерполяционный полином Ньютона можно записать в следующем виде:

$$P_n(x) = y_0 + \frac{\Delta y_0}{h \cdot 1!} \cdot (x - x_0) + \frac{\Delta^2 y_0}{h^2 \cdot 2!} \cdot (x - x_0) \cdot (x - x_1) + \dots + \frac{\Delta^n y_0}{h^n \cdot n!} \cdot (x - x_0) \cdot \dots \cdot (x - x_{n-1}). \quad (43)$$

Здесь  $\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i$  – *неразделенные разности*, которые вычисляются аналогично разделенным разностям, но без деления на разность значений аргумента.

Таким образом, для построения интерполяционного полинома Ньютона необходимо предварительно найти  $(n - 1)$  неразделенных разностей.

Отметим, что для вычисления значений полинома лучше всего использовать схему Горнера, которая заключается в преобразовании полинома

$$P(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_n \cdot x^n$$

к виду

$$P(x) = a_0 + x \cdot (a_1 + x \cdot (a_2 + \dots + x \cdot (a_{n-1} + x \cdot a_n))),$$

что позволяет сократить объем вычислений.

**Пример 29.**

По данным табл. 1 вычислить приближенное значение функции  $f(x)$  в точке  $x = 21,85$  средствами *MS Excel* с использованием интерполяционного полинома Ньютона.

Т а б л и ц а 1

$i$	0	1	2	3
$x_i$	-0,7	5,5	11,7	17,9
$y_i$	25,25	20,03	44,93	10,61

Решение.

Интерполяционный полином Ньютона (43) будет иметь вид:

$$P_4(x) = y_0 + \frac{\Delta y_0}{h \cdot 1!} \cdot (x - x_0) + \frac{\Delta^2 y_0}{h^2 \cdot 2!} \cdot (x - x_0)(x - x_1) + \frac{\Delta^3 y_0}{h^3 \cdot 3!} \cdot (x - x_0)(x - x_1)(x - x_2). \quad (44)$$

Исходные данные из табл. 1 введем в ячейки **A1:E3** (рис. 24). В ячейку **G2** введем шаг, в **I2** – точку  $x$ , в которой требуется вычислить приближенное значение функции  $f(x)$ .

	A	B	C	D	E	F	G	H	I
1	i	0	1	2	3		h		x заданное
2	x	-0,7	5,5	11,7	17,9		=C2-B2		21,85
3	y	25,25	20,03	44,93	10,61				
4	$\Delta y$	=C3-B3	=D3-C3	=E3-D3					
5	$\Delta^2 y$	=C4-B4	=D4-C4						
6	$\Delta^3 y$	=C5-B5							
7	Промежуточные коэффициенты	=(\$I\$2-B2)/\$G\$2	=(\$I\$2-C2)*(\$I\$2-B2)/2/\$G\$2	=(\$I\$2-D2)*(\$I\$2-C2)*(\$I\$2-B2)/6/\$G\$2					
8	Кoeffициенты полинома	=B7*B4	=C7*B5	=D7*B6		=СУММ(B8:D8)			
9	Значение функции					=E8+B3			

	A	B	C	D	E	F	G	H	I
1	i	0	1	2	3		h		x заданное
2	x	-0,7	5,5	11,7	17,9		6,2		21,85
3	y	25,25	20,03	44,93	10,61				
4	$\Delta y$	-5,22	24,9	-34,32					
5	$\Delta^2 y$	30,12	-59,22						
6	$\Delta^3 y$	-89,34							
7	Промежуточные коэффициенты	3,637097	4,795688	2,617002					
8	Кoeffициенты полинома	-18,9856	144,4461	-233,803	-108,342				
9	Значение функции				-83,0925				

Рис. 24. Вычисление значения таблично заданной функции в MS Excel

В ячейку **B4** занесем формулу для вычисления конечной разности первого порядка и скопируем ее во все ячейки, где будут вычисляться конечные разности второго и третьего порядков.

В 7-ю строку введем формулы для вычисления промежуточных коэффициентов в соответствии с формулой (44), затем просуммируем их в ячейке **E8** и вычислим в ячейке **E9** значение функции в заданной точке. На рис. 24, а показан лист *Excel* с формулами (для этого задать: **Формулы – Зависимости формул – Показать формулы**) и числовыми значениями – рис. 24, б.

Построение полинома Ньютона и графика с использованием *MathCAD* для этого примера показано в лабораторном практикуме (см. пример 15 [21]):

$$\begin{aligned}
 P_4(x) &= 25,25 + \frac{-5,22}{6,2 \cdot 1!} \cdot (x + 0,7) + \frac{30,12}{6,2^2 \cdot 2!} \cdot (x + 0,7)(x - 5,5) + \\
 &+ \frac{-89,34}{6,2^3 \cdot 3!} \cdot (x + 0,7)(x - 5,5)(x - 11,7) = \\
 &= -0,062x^3 + 1,423x^2 - 5,99x + 20,338.
 \end{aligned}$$

Вычисленное в *MathCAD* значение функции  $f(21,85) = -83,151$  несколько отличается от полученного в *Excel*, что говорит о погрешности вычислений.

### 6.3. Принципы двумерной интерполяции

Предположим, что функция  $f(x, y)$  зависит от двух переменных и известны ее значения при некоторых значениях аргументов  $\{x_i, y_j\}$ . Предположим, что нам необходимо вычислить значение функции в межузловой точке  $f(x_r, y_s)$ , причем  $r \neq i, s \neq j$  для любых  $i$  и  $j$ . В этом случае для всех значений  $y_j$  строятся интерполяционные полиномы  $P(x, y_j)$ , т.е.  $P(x, y_1), P(x, y_2), \dots, P(x, y_m)$ . Далее для последовательности точек  $P(x_r, y_1), P(x_r, y_2), \dots, P(x_r, y_m)$  строится интерполяционный полином  $PP(x_r, y)$ , и затем вычисляется требуемое значение  $f(x_r, y_s) = PP(x_r, y_s)$ .

#### Пример 30

Вычислить значение функции двух переменных  $z = f(x, y)$  в точке  $z = (0,25; 2,5)$ , если функция задана табл. 2.

Таблица 2

$y_i/x_i$	0,1	0,2	0,3
1	2	3	4
2	3	4	5
3	4	5	6

Решение.

Построим интерполяционные полиномы для всех значений  $y_i$  по формуле (44), используя значения неразделенных разностей из табл. 3.

$h_x = 0,1$ ; принимаем  $x_0 = 0,2$ .

Таблица 3

$x_i$	0,1	0,2	0,3	$x_i$	0,1	0,2	0,3	$x_i$	0,1	0,2	0,3
$z(x_i, y_1)$	2	3	4	$z(x_i, y_2)$	3	4	5	$z(x_i, y_3)$	4	5	6
$\Delta y_1$	1	1		$\Delta y_2$	1	1		$\Delta y_3$	1	1	
$\Delta^2 y_1$	0			$\Delta^2 y_2$	0			$\Delta^2 y_3$	0		

$$P(x, y_1) = 2 + \frac{1}{0,1 \cdot 1!} \cdot (x - 0,1) = 10x + 1; \quad P(0,25; 3) = 3,5;$$

$$P(x, y_2) = 3 + \frac{1}{0,1 \cdot 1!} \cdot (x - 0,1) = 10x + 2; \quad P(0,25; 4) = 4,5;$$

$$P(x, y_3) = 4 + \frac{1}{0,1 \cdot 1!} \cdot (x - 0,1) = 10x + 3; \quad P(0,25; 5) = 5,5.$$

Для последовательности точек  $P(0,25, y_1)$ ;  $P(0,25, y_2)$ ;  $P(0,25, y_3)$  строим интерполяционный полином  $P(0,25, y_1)$ , используя значения неразделенных разностей из табл. 4.

$h_y = 1$ ; принимаем  $y_0 = 4$ .

Таблица 4

$y_i$	1	2	3
$P(x)$	3,5	4,5	5,5
$\Delta P(x)$	1	1	
$\Delta^2 P(x)$	0		

$$P(0,25, y) = 3,5 + \frac{1}{1 \cdot 1!} \cdot (y - 1) = 2,5 + y$$

Вычисляем требуемое значение  $P(0,25; 2,5) = 5$ .

#### 6.4. Кубическая сплайн-интерполяция

В некоторых задачах требуется интерполировать не только саму функцию, но также ее первую и вторую производные. Для решения таких задач используется сплайновая интерполяция. Сплайн – это функция, которая на каждом межузловом интервале совпадает с некоторым полиномом, своим для каждого интервала. Полиномы смежных интервалов склеиваются так, чтобы функция и ее производная были непрерывны.

В случае кубической сплайн-интерполяции на каждом из интервалов  $[x_{i-1}, x_i]$ , где  $i = 1, 2, \dots, n$ , строится интерполирующий кубический полином



(полином третьего порядка) таким образом, чтобы в точках  $x_i$  выполнялись следующие условия:

- 1)  $P(x_i) = y_i$ ;
- 2)  $P'(x_i - 0) = P'(x_i + 0)$ ;
- 3)  $P''(x_i - 0) = P''(x_i + 0)$ ,

где  $P'(x_i \pm 0)$  и  $P''(x_i \pm 0)$  – соответственно первые и вторые производные функции  $P(x_i)$  слева и справа.

Кубический полином, определенный на интервале  $[x_{i-1}, x_i]$ , имеет вид

$$P(x) = a_i + b_i \cdot (x - x_{i-1}) + c_i \cdot (x - x_{i-1})^2 + d_i \cdot (x - x_{i-1})^3, \quad (45)$$

где  $x \in [x_{i-1}, x_i]$ . Для всего интервала будет соответственно  $n$  кубических полиномов, отличающихся коэффициентами  $a_i, b_i, c_i, d_i$ .

Требование выполнения условий 1, 2, 3 приводит к следующим линейным уравнениям, составленным для каждого из  $n$  интервалов:

$$\begin{aligned} P(x_{i-1}) &= a_i; \quad i = 1, 2, \dots, n; \\ P(x_i) &= a_i + b_i \cdot h_i + c_i \cdot h_i^2 + d_i \cdot h_i^3 = y_i; \quad h_i = x_i - x_{i-1}; \end{aligned} \quad (46)$$

$$\begin{cases} b_i + 2 \cdot c_i \cdot (x - x_{i-1}) + 3 \cdot d_i \cdot (x - x_{i-1})^2 = P'(x); \\ 2 \cdot c_i + 6 \cdot d_i \cdot (x - x_{i-1}) = P''(x). \end{cases}$$

На интервале  $[x_i, x_{i+1}]$

$$P(x) = a_{i+1} + b_{i+1} \cdot (x - x_i) + c_{i+1} \cdot (x - x_i)^2 + d_{i+1} \cdot (x - x_i)^3.$$

Следовательно,

$$\begin{cases} b_{i+1} + 2 \cdot c_{i+1} \cdot (x - x_i) + 3 \cdot d_{i+1} \cdot (x - x_i)^2 = P'(x); \\ 2 \cdot c_{i+1} + 6 \cdot d_{i+1} \cdot (x - x_i) = P''(x). \end{cases} \quad (47)$$

Из полученных уравнений найти все коэффициенты нельзя, т.к. система является недоопределенной. Дополним систему условиями гладкости функции (непрерывности первой производной) и гладкости первой производной (непрерывности второй производной) в узлах интерполяции.

Приравнявая производные для  $P(x)$  первого и второго порядка в конце  $i$ -го и в начале  $(i + 1)$ -го интервалов, получим набор уравнений:

$$\begin{cases} b_i + 2 \cdot c_i \cdot h_i + 3 \cdot d_i \cdot h_i^2 = b_{i+1} + 2 \cdot c_{i+1} \cdot h_{i+1} + 3 \cdot d_{i+1} \cdot h_{i+1}^2; \\ 2 \cdot c_i + 6 \cdot d_i \cdot h_i = 2 \cdot c_{i+1} + 6 \cdot d_{i+1} \cdot h_{i+1}; \end{cases} \quad i = 1, 2, \dots, n - 1.$$

Эти соотношения дают  $(2n - 2)$  уравнения, всего получим  $(4n - 2)$  уравнения.

4) В точках  $x_0$  и  $x_n$  считается, что  $f''(x_0) = f''(x_n) = 0$ . Это дает еще два уравнения:

$$\begin{cases} c_1 = 0; \\ c_n + 3 \cdot d_n \cdot h_n = 0. \end{cases} \quad (48)$$

Таким образом, имеем  $4n$  линейных уравнений с  $4n$  неизвестными:  $a_i, b_i, c_i, d_i, i = 1, \dots, n$ . Решив эту систему, найдем совокупность из  $n$  кубических полиномов, значения которых приближенно равны значениям интерполируемой функции в соответствующих точках.

Заметим, что решать полученную систему линейных алгебраических уравнений удобнее всего методом прогонки, описание которого приводится в разд. 10.

Замечание. Чаще всего узлы сплайновой интерполяции располагают равномерно, т.е.  $h_i = \text{const}$ , хотя это необязательно.

Пример 31.

Построить кубический сплайн для функции, заданной таблично (табл. 5):

Таблица 5

$i$	0	1	2
$x_i$	1	2	3
$y_i$	1	5	8

Определить значения функции в точках  $x = 1,25$  и  $x = 2,5$ .

Решение.

Весь интервал делим на подынтервалы:  $[1, 3] = [1, 2] \cup [2, 3]$ , на каждом из которых строим полином 3-го порядка.

На интервале  $[1,2]$ , в соответствии с формулами (45, 47), получаем:

$$y = a_1 + b_1(x-1) + c_1(x-1)^2 + d_1(x-1)^3,$$

$$y' = b_1 + 2c_1(x-1) + 3d_1(x-1)^2,$$

$$y'' = 2c_1 + 6d_1(x-1).$$

На интервале  $[2,3]$  имеем:

$$y = a_2 + b_2(x-2) + c_2(x-2)^2 + d_2(x-2)^3,$$

$$y' = b_2 + 2c_2(x-2) + 3d_2(x-2)^2,$$

$$y'' = 2c_2 + 6d_2(x-2).$$

Составим систему уравнений. Для каждого интервала по формулам (46) получим уравнения, соответствующие прохождению полинома через

начальную и конечную точки. Здесь узлы интерполирования располагаются равномерно,  $h = 1$ . Для интервала  $[1,2]$  получаем:

$$1 = a_1;$$

$$5 = a_1 + b_1h + c_1h^2 + d_1h^3 = a_1 + b_1 + c_1 + d_1,$$

для интервала  $[2,3]$ :

$$5 = a_2;$$

$$8 = a_2 + b_2h + c_2h^2 + d_2h^3 = a_2 + b_2 + c_2 + d_2.$$

Приравниваем в точке  $x = 2$ , являющейся концом 1-го и началом 2-го интервалов, соответствующие производные первого и второго порядков:

$$b_1 + 2c_1 + 3d_1 = b_2;$$

$$2c_1 + 6d_1h = 2c_2.$$

По формулам (48) получаем еще 2 уравнения:

$$c_1 = 0; \quad c_2 + 3d_2 = 0.$$

Полученную систему из 8-ми уравнений с 8-ю неизвестными можно решить любым известным методом. Приведем решение в системе *MathCAD*:

$$a1 := 0 \quad b1 := 0 \quad c1 := 0 \quad d1 := 0 \quad a2 := 0 \quad b2 := 0 \quad c2 := 0 \quad d2 := 0$$

Given

$$a1 = 1$$

$$a1 + b1 + c1 + d1 = 5$$

$$a2 = 5$$

$$a2 + b2 + c2 + d2 = 8$$

$$b1 + 2c1 + 3d1 = b2$$

$$2c1 + 6d1 = 2c2$$

$$c1 = 0$$

$$2c2 + 6d2 = 0$$

$$\text{Find}(a1, b1, c1, d1, a2, b2, c2, d2) =$$

$$\begin{pmatrix} 1 \\ 4.25 \\ 0 \\ -0.25 \\ 5 \\ 3.5 \\ -0.75 \\ 0.25 \end{pmatrix}$$

Вычисленные коэффициенты подставляем в соответствующие полиномы и получаем полиномы для каждого подынтервала:

$$y = 1 + 4,25(x - 1) - 0,25(x - 1)^3 = -0,25x^3 + 0,75x^2 + 3,5x - 3$$

$$y = 5 + 3,5(x - 2) - 0,75(x - 2)^2 + 0,25(x - 2)^3 = 0,25x^3 - 2,25x^2 + 9,5x - 7.$$

Определим значение функции в точке  $x = 1,25$ :

$$y(1,25) = -0,25 \cdot 1,25^3 + 0,75 \cdot 1,25^2 + 3,5 \cdot 1,25 - 3 = 2,059$$

и в точке  $x = 2,5$ :

$$y(2,5) = 0,25 \cdot 2,5^3 - 2,25 \cdot 2,5^2 + 9,5 \cdot 2,5 - 7 = 6,594.$$

Графики кубических сплайнов в системе *MathCAD* показаны на рис. 25.

$$y_1(x) := 1 + 4.25(x - 1) - 0.25(x - 1)^3$$

$$y_2(x) := 5 + 3.5(x - 2) - 0.75(x - 2)^2 + 0.25(x - 2)^3$$

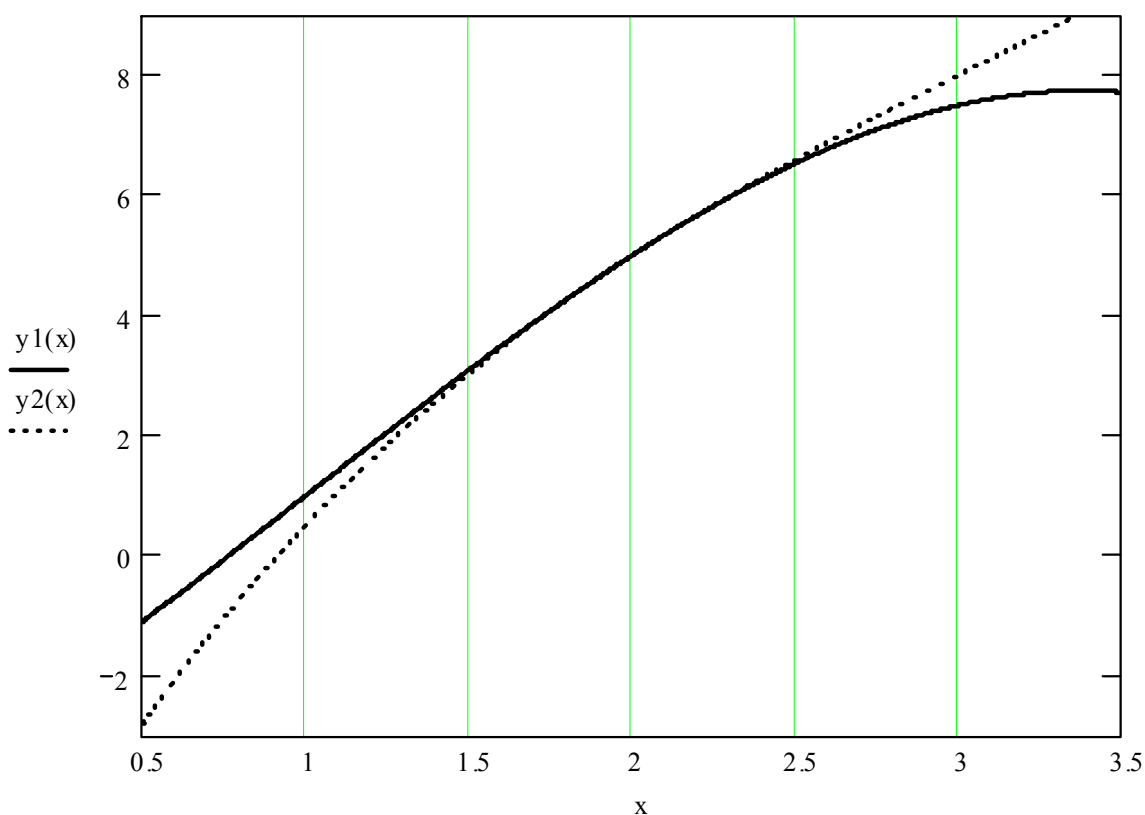


Рис. 25. Графики кубических сплайнов в системе MathCAD

### Контрольные вопросы и задания

1. Какие точки называются узлами интерполирования?
2. Зависит ли выбор метода интерполирования от количества точек исходной табличной функции, от их взаимного расположения?
3. Что такое интерполянта?
4. В чем заключается задача обратного интерполирования?
5. В каком методе полиномиальной интерполяции используются конечные разности?
6. Какие виды интерполирования, кроме полиномиального, Вам известны?
7. Что такое кубическая сплайн-интерполяция?
8. Можно ли по экспериментальным данным получить аналитическую зависимость?
9. В каких случаях интерполяционные полиномы Ньютона и Лагранжа совпадают? Приведите пример.

## 7. АППРОКСИМАЦИЯ ФУНКЦИЙ

### 7.1 Метод наименьших квадратов

В отличие от полиномиальной интерполяции в данном случае решается задача о приближении (*аппроксимации*) функции  $f(x)$ , заданной таблично, такой функцией  $\varphi(x)$ , чтобы среднее квадратичное отклонение  $\varphi(x)$  от  $f(x)$  в заданной области было наименьшим. Функция  $\varphi(x)$  при этом называется *аппроксимирующей функцией*, а метод построения такой функции – методом наименьших квадратов. Основной задачей аппроксимации является построение приближенной (аппроксимирующей) непрерывной функции, наиболее близко проходящей около заданных точек.

При среднеквадратичном приближении мерой отклонения  $\varphi(x)$  от заданной функции  $f(x)$  на множестве точек  $(x_j, y_j)$ ,  $j = 0, \dots, n$  является величина  $S$ , равная сумме квадратов разностей между аппроксимирующей функцией и функцией, заданной таблично:

$$S = \sum_{j=0}^n (\varphi(x_j) - y_j)^2.$$

Практически важным случаем такой аппроксимации является *случай приближения табличной функции многочленом*

$$\varphi(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m, \quad m \neq n. \quad (49)$$

При этом коэффициенты  $a_j$  нужно подобрать так, чтобы достичь наименьшего среднеквадратичного отклонения многочлена от известных значений заданной функции:

$$\min_{a_i, i=0, \dots, m} \left( S = \sum_{j=0}^n (a_0 + a_1x_j + \dots + a_mx_j^m - y_j)^2 \right).$$

Такой метод аппроксимации приводит к системе линейных алгебраических уравнений, которую всегда нетрудно решить любым известным способом.

Поскольку параметры  $a_0, a_1, \dots, a_m$  выступают в роли независимых переменных функции  $S$ , то ее минимум находится путем приравнивания нулю частных производных  $S$  по этим переменным:

$$\frac{\partial S}{\partial a_0} = 0; \quad \frac{\partial S}{\partial a_1} = 0; \quad \dots; \quad \frac{\partial S}{\partial a_m} = 0.$$

Определив частные производные и положив их равными нулю, получим:

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial a_0} = 2 \sum_{j=0}^n (y_j - a_0 - a_1 x_j - a_2 x_j^2 - \dots - a_m x_j^m) \cdot (-1) = 0; \\ \frac{\partial S}{\partial a_1} = 2 \sum_{j=0}^n (y_j - a_0 - a_1 x_j - a_2 x_j^2 - \dots - a_m x_j^m) \cdot (-x_j) = 0; \\ \frac{\partial S}{\partial a_2} = 2 \sum_{j=0}^n (y_j - a_0 - a_1 x_j - a_2 x_j^2 - \dots - a_m x_j^m) \cdot (-x_j^2) = 0; \\ \dots \\ \frac{\partial S}{\partial a_m} = 2 \sum_{j=0}^n (y_j - a_0 - a_1 x_j - a_2 x_j^2 - \dots - a_m x_j^m) \cdot (-x_j^m) = 0. \end{array} \right.$$

Приведя коэффициенты при неизвестных  $a_0, a_1, \dots, a_m$  в соответствующие суммы и перенеся свободные члены вправо от знака равенства, получим следующую систему линейных алгебраических уравнений:

$$\left\{ \begin{array}{l} (n+1)a_0 + a_1 \sum_{j=0}^n x_j + a_2 \sum_{j=0}^n x_j^2 + \dots + a_m \sum_{j=0}^n x_j^m = \sum_{j=0}^n y_j; \\ a_0 \sum_{j=0}^n x_j + a_1 \sum_{j=0}^n x_j^2 + a_2 \sum_{j=0}^n x_j^3 + \dots + a_m \sum_{j=0}^n x_j^{m+1} = \sum_{j=0}^n y_j x_j; \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_0 \sum_{j=0}^n x_j^m + a_1 \sum_{j=0}^n x_j^{m+1} + a_2 \sum_{j=0}^n x_j^{m+2} + \dots + a_m \sum_{j=0}^n x_j^{2m} = \sum_{j=0}^n y_j x_j^m. \end{array} \right. \quad (50)$$

Решив эту систему, найдем коэффициенты  $a_0, a_1, \dots, a_m$  многочлена (49), которые являются искомыми параметрами эмпирической<sup>3</sup> функции. Систему (50) можно записать в более компактном виде:

$$\left\{ \begin{array}{l} b_{00}a_0 + b_{01}a_1 + \dots + b_{0m}a_m = c_0; \\ b_{10}a_0 + b_{11}a_1 + \dots + b_{1m}a_m = c_1; \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ b_{m0}a_0 + b_{m1}a_1 + \dots + b_{mm}a_m = c_m. \end{array} \right.$$

Здесь  $b_{kl} = \sum_{j=0}^n x_j^{k+l}$ ,  $c_k = \sum_{j=0}^n x_j^k y_j$ ,  $k, l = 0, \dots, m$ .

Система линейных уравнений (50) называется *нормальной* системой. Для её решения удобно применять метод Зейделя, который сходится для любой степени аппроксимирующего полинома.

<sup>3</sup> Эмпирическая (выборочная) функция распределения в математической статистике – это приближение теоретической функции распределения, построенное с помощью выборки из него (ru.wikipedia.org).

При  $m = 0$  получим выражение  $\varphi(x) = P_0(x) = a_0$ , которое рассчитаем из уравнения

$$(n + 1) \cdot a_0 = \sum_{j=0}^n y_j$$

или

$$a_0 = \frac{\sum_{j=0}^n y_j}{n + 1}$$

(среднее значение функции  $y$  на отрезке аппроксимации).

При  $m = 1$  полином примет вид:  $\varphi(x) = P_1(x) = a_0 + a_1x$ . Коэффициенты  $a_0$  и  $a_1$  находятся из системы уравнений

$$\begin{cases} (n + 1)a_0 + \sum_{j=0}^n x_j a_1 = \sum_{j=0}^n y_j; \\ \sum_{j=0}^n x_j a_0 + \sum_{j=0}^n x_j^2 a_1 = \sum_{j=0}^n y_j x_j. \end{cases}$$

Это линейная аппроксимация.

При  $m = 2$  имеем следующий полином:  $\varphi(x) = P_2(x) = a_0 + a_1x + a_2x^2$ , а коэффициенты  $a_0$ ,  $a_1$  и  $a_2$  находятся при решении системы трех уравнений с тремя неизвестными:

$$\begin{cases} (n + 1)a_0 + \sum_{j=0}^n x_j a_1 + \sum_{j=0}^n x_j^2 a_2 = \sum_{j=0}^n y_j; \\ \sum_{j=0}^n x_j a_0 + \sum_{j=0}^n x_j^2 a_1 + \sum_{j=0}^n x_j^3 a_2 = \sum_{j=0}^n y_j x_j; \\ \sum_{j=0}^n x_j^2 a_0 + \sum_{j=0}^n x_j^3 a_1 + \sum_{j=0}^n x_j^4 a_2 = \sum_{j=0}^n y_j x_j^2. \end{cases}$$

При  $m = 3$  искомым полином имеет следующий вид:  $\varphi(x) = P_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ , а коэффициенты  $a_0$ ,  $a_1$ ,  $a_2$  и  $a_3$  находятся из системы четырех уравнений с четырьмя неизвестными:

$$\begin{cases} (n+1)a_0 + \sum_{j=0}^n x_j a_1 + \sum_{j=0}^n x_j^2 a_2 + \sum_{j=0}^n x_j^3 a_3 = \sum_{j=0}^n y_j; \\ \sum_{j=0}^n x_j a_0 + \sum_{j=0}^n x_j^2 a_1 + \sum_{j=0}^n x_j^3 a_2 + \sum_{j=0}^n x_j^4 a_3 = \sum_{j=0}^n y_j x_j; \\ \sum_{j=0}^n x_j^2 a_0 + \sum_{j=0}^n x_j^3 a_1 + \sum_{j=0}^n x_j^4 a_2 + \sum_{j=0}^n x_j^5 a_3 = \sum_{j=0}^n y_j x_j^2; \\ \sum_{j=0}^n x_j^3 a_0 + \sum_{j=0}^n x_j^4 a_1 + \sum_{j=0}^n x_j^5 a_2 + \sum_{j=0}^n x_j^6 a_3 = \sum_{j=0}^n y_j x_j^3. \end{cases} \quad (51)$$

Пример 32.

Используя метод наименьших квадратов, построить аппроксимирующий полином 3-й степени для функции, заданной в табличном виде (табл. 6):

Таблица 6

$x$	3	3,2	3,4	3,7	3,9	4
$y$	-14	-10	-8	-12	-16	-18

Решение.

Здесь  $n = 5$ ,  $m = 3$ .

Найдем минимум суммы квадратов отклонений:

$$\min_{a_0, \dots, a_3} \sum_{j=0}^5 (\varphi(x_j, a_0, a_1, a_2, a_3) - y_j)^2.$$

Полином третьей степени выглядит следующим образом:

$$\varphi(x) = P_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$

Коэффициенты  $a_0, a_1, a_2, a_3$  вычислим, решив СЛАУ (51).

Решение в среде табличного процессора *MS Excel* приведено на рис. 26. Представлены графики табличной функции (штриховая линия) и аппроксимирующей функции (сплошная линия).

Решение данного примера в *MathCAD* приводится в лабораторном практикуме (см. пример 24 [21]).

Для аппроксимации таблично заданных функций по методу наименьших квадратов можно применять и функции, отличные от полиномиальных и содержащие некоторые неизвестные коэффициенты, определенные, как и в случае полиномиальной аппроксимации, из условия минимума суммы квадратов отклонений значений аппроксимирующей функции от заданных значений.



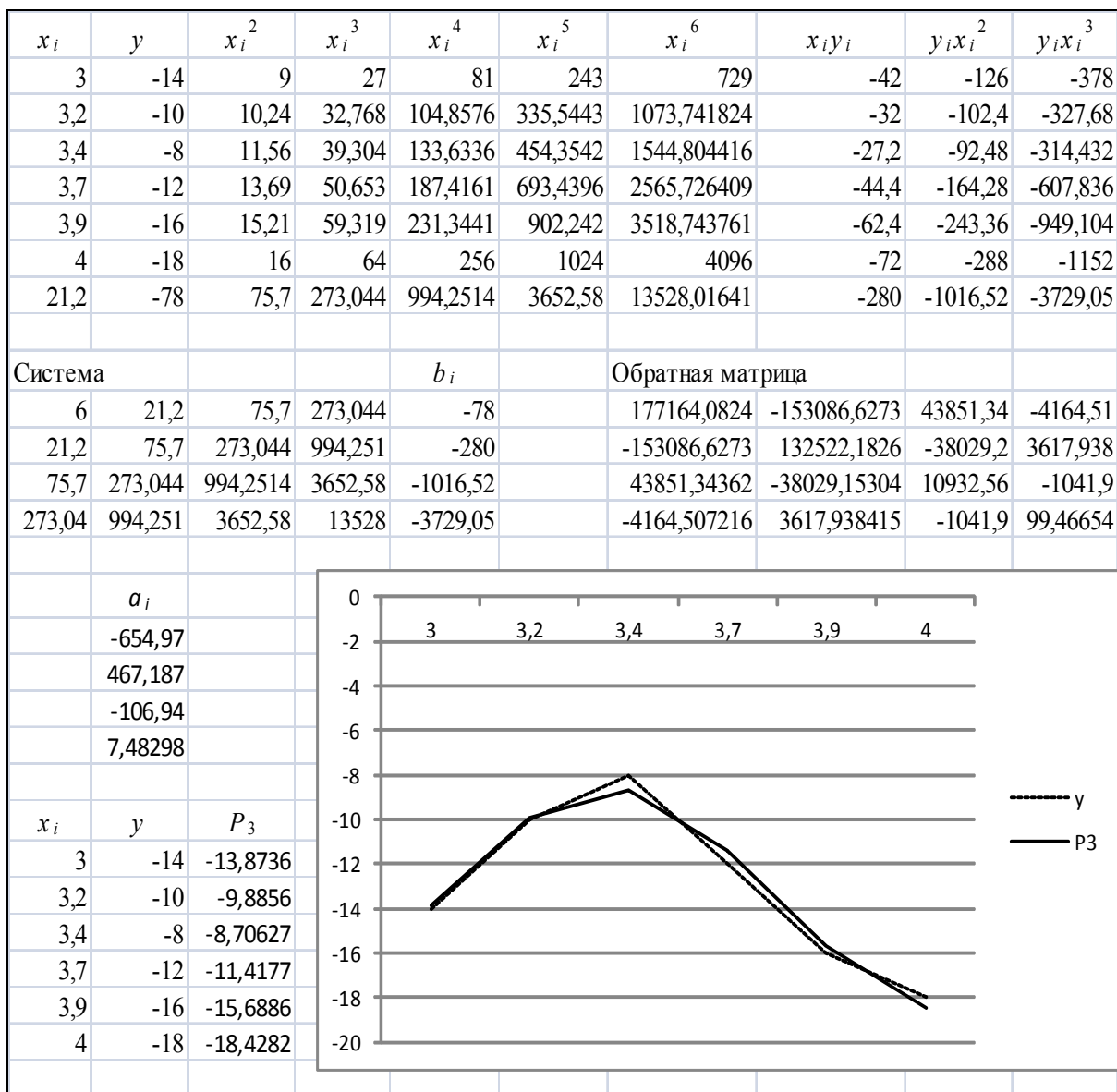


Рис. 26. Пример аппроксимации методом наименьших квадратов в MS Excel

### Пример 33.

Результаты эксперимента характеризуются таблицей (табл. 7). Подобрать аппроксимирующую функцию и решить задачу аппроксимации.

Таблица 7

$j$	0	1	2	3	4
$x_j$	5	10	20	30	40
$y_j$	1	5	20	200	2000

Решение.

Аппроксимирующая функция  $y(x)$  должна удовлетворять условию:

$$S = \sum_{j=0}^n (\varphi(x_j) - y_j)^2 \rightarrow \min.$$

Можно перебрать различные варианты элементарных функций или их комбинации, например:  $y(x) = \sin x$ ,  $y(x) = x^2$ ,  $y(x) = \ln x$ ,  $y(x) = e^x$ ,  $y(x) = a \cdot \sin(x + b) + c \cdot \cos(x + d) + e \cdot x^2$  и т.д.

Сложно найти общий подход к вопросу выбора аппроксимирующих функций. Можно построить график табличной функции и определить вид зависимости.

Заданная табличная функция напоминает экспоненту, поэтому в качестве аппроксимирующей выберем функцию  $y = a \cdot e^{bx}$ . Определим наилучшие значения параметров  $a$  и  $b$ . Для этого вычислим  $\min S$ .

Составим сумму квадратов отклонений  $S(a, b)$ :

$$S(a, b) = \sum_{j=0}^4 (ae^{bx_j} - y_j)^2 =$$

$$= (ae^{5b} - 1)^2 + (ae^{10b} - 5)^2 + (ae^{20b} - 20)^2 + (ae^{30b} - 200)^2 + (ae^{40b} - 2000)^2.$$

Продифференцируем полученное выражение по переменным  $a$  и  $b$ :

$$\frac{\partial S(a, b)}{\partial a} = \sum_{j=0}^4 2(ae^{bx_j} - y_j) \cdot e^{bx_j};$$

$$\frac{\partial S(a, b)}{\partial b} = \sum_{j=0}^4 2(ae^{bx_j} - y_j) \cdot ae^{bx_j} \cdot x_j.$$

Приравняв нулю эти частные производные, получим систему двух уравнений с двумя неизвестными:

$$\frac{\partial S(a, b)}{\partial a} = \sum_{j=0}^4 2(ae^{bx_j} - y_j) \cdot e^{bx_j} = 0;$$

$$\frac{\partial S(a, b)}{\partial b} = \sum_{j=0}^4 2(ae^{bx_j} - y_j) \cdot ae^{bx_j} \cdot x_j = 0.$$

На рис. 27 приводится решение полученной системы нелинейных уравнений в *MathCAD*.

$$x := \begin{pmatrix} 5 \\ 10 \\ 20 \\ 30 \\ 40 \end{pmatrix} \quad y := \begin{pmatrix} 1 \\ 5 \\ 20 \\ 200 \\ 2000 \end{pmatrix}$$

$$\begin{aligned}
 & a := 0.2 \quad b := 0.5 \\
 & \text{Given} \\
 & \sum_{i=0}^4 2 \cdot \left( a \cdot e^{b \cdot x_i} - y_i \right) \cdot e^{b \cdot x_i} = 0 \\
 & \sum_{i=0}^4 2 \cdot \left( a \cdot e^{b \cdot x_i} - y_i \right) \cdot a \cdot e^{b \cdot x_i} \cdot x_i = 0 \\
 & u := \text{Find}(a, b) \quad u = \begin{pmatrix} 0.2 \\ 0.23 \end{pmatrix}
 \end{aligned}$$

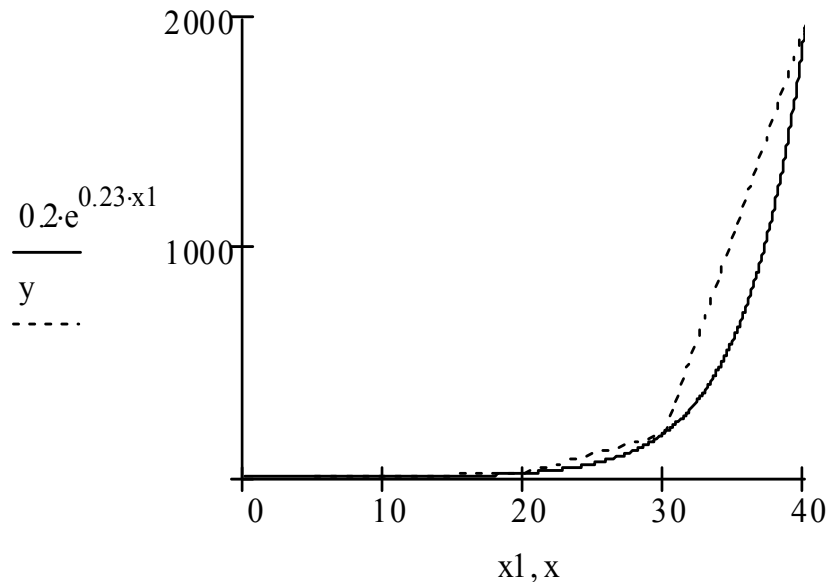


Рис. 27. Аппроксимация экспоненциальной функцией в системе *MathCAD*

Решение системы –  $a = 0,2$  и  $b = 0,23$  – обеспечивает наилучшее приближение функции  $y = 0,2 \cdot e^{0,23 \cdot x}$  к заданной табличной функции.

## 7.2. Оценка параметров эмпирической формулы, построенной по экспериментальным данным на примере линейной модели

Пусть  $x_i$  ( $i = 1, \dots, n$ ) – значения аргумента таблично заданной функции;  $y_{i,j}$  – значения табличной функции в  $j$ -м наблюдении ( $j = 1, \dots, P_i$ );  $P_i$  – число наблюдений величины  $y_i$ .

Пусть  $\bar{x}$  – среднее для  $x_i$ ,  $\bar{y}_i = \frac{\sum_{j=1}^{P_i} y_{i,j}}{P_i}$  – математическое ожидание ве-

личины  $y_{i,j}$ , а  $\bar{\bar{y}} = \frac{\sum_{i=1}^n \bar{y}_i}{n}$  – среднее значение математических ожиданий  $\bar{y}_i$ .

Будем считать, что таблично заданная функция аппроксимируется линейной моделью  $y = a + b \cdot x$ . Тогда оценки коэффициентов линейной модели можно произвести по формулам:

$$b = \frac{\sum x_i \bar{y}_i - \bar{x} \frac{\sum y_i P_i}{\sum P_i}}{\overline{x_i^2} - \bar{x}^2} = \frac{\sum \frac{\sum P_i}{N} x_i \bar{y}_i - \bar{x} \sum y_i P_i}{\sum P_i \cdot (\overline{x_i^2} - \bar{x}^2)}; \quad (52)$$

$$a = \bar{\bar{y}} - b\bar{x}.$$

Для оценки адекватности модели реальной табличной функции вычисляют следующие величины:  $S_r^2$ , называемую взвешенной суммой квадратов, и  $S_l^2$  – меру рассеивания, вызванного случайной ошибкой.

$$S_r^2 = \frac{1}{n-2} \cdot \sum_{i=1}^n P_i \cdot (\bar{y}_i - \hat{y}_i)^2,$$

$$S_l^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{P_i} (y_{i,j} - \bar{y}_i)^2}{\sum_{i=1}^n P_i - n}, \quad (53)$$

где  $\hat{y}_i = a + b \cdot (x_i - \bar{x})$ .

Затем по специальным таблицам вычисляется величина  $F_{1-\alpha}$ , которая называется *критерием Фишера*, причем  $\alpha$  есть уровень значимости – мера допустимой ошибки, которая определяется требуемой точностью аппроксимации табличной функции.

Если  $\frac{S_r^2}{S_l^2} \geq F_{1-\alpha}$ , то выбор линейной модели неадекватен реальной зависимости, в противном случае модель хорошо описывает процесс.

#### Пример 34.

Оценить адекватность линейной модели функции, заданной в табл. 8.

Т а б л и ц а 8

$i$	1		2			3
Аргумент $x_i$	10	10	20	20	30	30
Функция $y_{i,j}$	12,6	12,8	13,8	13,6	16,7	17,5

Рассчитать величины  $a$ ,  $b$ , а также  $S_r^2$ ,  $S_l^2$  и оценить их отношение, считая, что  $F_{1-\alpha} = 5,41$ .

**Решение.**

Аппроксимируем исходную функцию линейной моделью  $y = a + b \cdot x$ , для определения коэффициентов которой воспользуемся формулами (55). Определим математическое ожидание  $\bar{y}_i$  для каждой величины  $y_{i,j}$  и среднее значение математических ожиданий  $\bar{\bar{y}}$ .

$$\bar{y}_1 = \frac{12,6 + 12,8}{2} = 12,7.$$

Аналогично  $\bar{y}_2 = 13,7$ ;  $\bar{y}_3 = 17,1$ .

$$\bar{\bar{y}} = \frac{12,7 + 13,7 + 17,1}{3} = 14,5; \quad \bar{x} = \frac{10 + 20 + 30}{3} = 20;$$

$$b = \frac{2 \cdot (10 \cdot 12,7 + 20 \cdot 13,7 + 30 \cdot 17,1) - 20 \cdot 2 \cdot (12,7 + 13,7 + 17,1)}{6 \cdot \left( \frac{10^2 + 20^2 + 30^2}{3} - 20^2 \right)} = 0,22;$$

$$a = 14,5 - 0,22 \cdot 20 = 10,1;$$

$$y = 0,22 \cdot x + 10,1.$$

Оценим адекватность полученной линейной модели по формулам (53).

$$\hat{y}_1 = 10,1 + 0,22 \cdot (10 - 20) = 7,9;$$

$$\hat{y}_2 = 10,1 + 0,22 \cdot (20 - 20) = 10,1;$$

$$\hat{y}_3 = 10,1 + 0,22 \cdot (30 - 20) = 12,3.$$

$$S_r^2 = \frac{1}{3-2} \cdot 2 \left[ (12,7 - 7,9)^2 + (13,7 - 10,1)^2 + (17,1 - 12,3)^2 \right] = 118,08;$$

$$S_l^2 = \frac{1}{6-3} \cdot \left[ (12,6 - 12,7)^2 + (12,8 - 12,7)^2 + (13,8 - 13,7)^2 + (13,6 - 13,7)^2 + (16,7 - 17,1)^2 + (17,5 - 17,1)^2 \right] = 0,12.$$

Получили  $S_r^2 / S_l^2 = 984 \gg F_{1-\alpha}$ ; следовательно, линейная модель неадекватна реальной зависимости.

### 7.3. Среднеквадратичная аппроксимация в метрическом пространстве

*Метрическим пространством*  $R = (X, r)$  называется множество, в котором между любой парой элементов определено обладающее определенными свойствами расстояние, называемое *метрикой*.

Здесь  $X$  – множество (пространство) элементов (точек),  $r$  – расстояние, т. е. неотрицательная действительная функция  $r(x, y)$ , определенная для любых  $x$  и  $y$  из  $X$  и подчиненная следующим трем аксиомам (аксиомам метрики):

- 1)  $r(x, y) = 0$  тогда и только тогда, когда  $x = y$ ;
- 2)  $r(x, y) = r(y, x)$  (аксиома симметрии);
- 3)  $r(x, z) \leq r(x, y) + r(y, z)$  (аксиома треугольника).

В данном разделе рассматриваются функции из метрического пространства  $L_2[a, b]$ , то есть функции, интегрируемые на интервале  $[a, b]$  вместе со своим квадратом. Скалярное произведение и норма в таком пространстве определены следующим образом:

$$(f, g) = \int_a^b f(x) \cdot g(x) dx; \quad \|f(x)\| = \sqrt{(f, f)}.$$

Общая задача аппроксимации в этом случае заключается в приближении заданной функции  $f(x)$  функциями  $y$  из  $L_2[a, b]$  таким образом, чтобы  $\|f - y\|$  была минимальной. Другими словами, необходимо, чтобы выполнялось условие  $\|f - y\| \leq \varepsilon$ , где  $\varepsilon$  – заданная точность аппроксимации.

Наиболее распространена аппроксимация заданной функции комбинацией из системы ортогональных или ортонормированных функций, то есть функций  $\varphi_k$ ,  $k = 1, 2, \dots$ , для которых выполняется условие:  $(\varphi_k, \varphi_m) = \delta_{k, m}$ , где  $\delta_{k, m}$  – символ Кронекера ( $\delta_{k, m} = 1$ , если  $k = m$ ;  $\delta_{k, m} = 0$ , если  $k \neq m$ ), если система функций ортонормированная, и  $(\varphi_k, \varphi_m) = 0$  при  $k \neq m$ , если система ортогональная.

Тогда в качестве аппроксимирующей функции берется сумма  $y = \sum_{k=0}^n c_k \cdot \varphi_k(x)$ , и задача аппроксимации сводится к отысканию коэффициентов  $c_k$ .

Итак, имеем задачу:  $\min_{\{c_k\}} \left\| f(x) - \sum_{k=0}^n c_k \cdot \varphi_k(x) \right\|$ . Решим эту задачу.

$$\begin{aligned} \|f - y\|^2 &= \int_a^b (f - y)(f - y) dx = \|f\|^2 - 2 \cdot (f, y) + \|y\|^2 = \\ &= \|f\|^2 - 2 \sum_{k=0}^n c_k \cdot f_k + \sum_{k=0}^n c_k^2 + \sum_{k=0}^n f_k^2 - \sum_{k=0}^n f_k^2; \\ (f, y) &= \sum_{k=0}^n c_k \cdot (f, \varphi_k) = \sum_{k=0}^n c_k \cdot f_k, \text{ где } f_k = (f, \varphi_k). \end{aligned}$$

Заметим, что  $\|y\|^2 = \sum_{k=0}^n c_k^2$ .

Тогда  $\|f - y\|^2 = \|f\|^2 + \sum_{k=0}^n (c_k - f_k)^2 - \sum_{k=0}^n f_k^2$ .

Из вида последнего выражения можно заключить, что минимум  $\|f - y\|^2$  будет достигаться тогда, когда  $\sum_{k=0}^n (c_k - f_k)^2 = 0$ , то есть когда  $c_k = (f, \varphi_k)$ .

Значит, наилучшим приближением для функции  $f(x)$  будет  $y = \sum_{k=0}^n f_k \cdot \varphi_k(x)$ ,

при этом  $\|f - y\|^2 = \|f\|^2 - \sum_{k=0}^n f_k^2$ .

Из решения следует, что наилучшее среднеквадратичное приближение существует и единственно. Задача по его отысканию сводится к задаче вычисления интегралов для определения коэффициентов

$$c_k = \int_a^b f(x) \cdot \varphi_k(x) dx. \quad (54)$$

В качестве аппроксимирующей системы функций часто используется ортогональная система многочленов Лежандра, заданная на интервале  $[-1, 1]$ :

$$P_n(x) = \frac{1}{2^n \cdot (n!)} \cdot \frac{d^n}{dx^n} \cdot (x^2 - 1)^n; \quad n = 0, 1, 2, \dots \quad x \in [-1, 1]. \quad (55)$$

В частности, имеем:

$$P_0(x) = 1;$$

$$P_1(x) = x;$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1);$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x);$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \text{ и т.д.}$$

Пусть на интервале  $[-1, 1]$  задана функция  $f(x)$ . Приближим эту функцию линейной комбинацией из полиномов Лежандра

$$Q(x) = c_0 P_0(x) + c_1 P_1(x) + \dots + c_m P_m(x), \quad (56)$$

причем коэффициенты  $c_0, c_1, \dots, c_m$  подберем так, чтобы величина отклонения

$$S_m = \int_{-1}^1 \left[ f(x) - \sum_{k=0}^m c_k P_k(x) \right]^2 dx$$

была минимальной.

В соответствии с полученной формулой (53) для коэффициентов  $c_k$  вычислим их значения:

$$c_k = \frac{1}{\|P_k(x)\|^2} \int_{-1}^1 f(x) P_k(x) dx = \frac{2k+1}{2} \int_{-1}^1 f(x) P_k(x) dx. \quad (57)$$

Заметим, что коэффициент  $\frac{1}{\|P_k(x)\|^2}$  позволяет ортогонализировать систему полиномов Лежандра, привести ее к ортонормированному виду.

**Пример 35.**

Функцию  $f(x) = 3^x$  на интервале  $[-1, 1]$  квадратично аппроксимировать полиномом Лежандра 3-й степени.

**Решение.**

Построим полином по формуле (56), коэффициенты вычисляем согласно формуле (57):

$$c_0 = \frac{1}{2} \int_{-1}^1 3^x dx = 1,2137;$$

$$c_1 = \frac{3}{2} \int_{-1}^1 x \cdot 3^x dx = 1,2371;$$

$$c_2 = \frac{5}{2} \int_{-1}^1 (3x^2 - 1) \cdot \frac{3^x}{2} dx = 0,4384;$$

$$c_3 = \frac{7}{2} \int_{-1}^1 \frac{1}{2} (5x^3 - 3x) \cdot 3^x dx = 0,09345.$$

Получаем:

$$\begin{aligned} Q(x) &= 1,2137 + 1,2371x + \frac{0,4384}{2}(3x^2 - 1) + \frac{0,9345}{2}(5x^3 - 3x) = \\ &= 0,2336x^3 + 0,6576x^2 + 1,0968x + 0,9945. \end{aligned}$$

### Контрольные вопросы и задания

1. Что такое аппроксимация функции и для чего она применяется?
2. Какие виды аппроксимирующих функций используют для приближения таблично заданных функций?
3. Почему метод аппроксимации называется методом наименьших квадратов?
4. Что такое символ Кронекера?
5. Что такое метрическое пространство?
6. Как определяются конечные разности различных порядков?
7. Что такое разделенная разность?
8. Подберите вид аппроксимирующей функции для аппроксимации результатов эксперимента, приведенных в табл. 9.

Таблица 9

x	1	2	3	4	5	6
y	1,1	7,9	27,5	64,5	130,2	217,6



## 8. МЕТОДЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ

### 8.1. Общие сведения

Определенный интеграл от непрерывной на отрезке  $[a, b]$  функции  $f(x)$ , для которой известна первообразная  $F(x)$ , может быть вычислен по формуле Ньютона – Лейбница

$$\int_a^b f(x)dx = F(b) - F(a),$$

где  $F'(x) = f(x)$ .

Если не удается найти первообразную подынтегральной функции или она имеет сложный и неудобный вид, применяют методы численного интегрирования. В случае, когда подынтегральная функция  $f(x)$  задается таблично, аналитические методы просто неприменимы. Задачи численного интегрирования встречаются в вычислительной математике, например, при решении дифференциальных уравнений, определении площадей, объемов, а также всевозможных интегральных характеристик задач естествознания.

*Общая задача численного интегрирования* заключается в нахождении значения определенного интеграла

$$F(a,b) = \int_a^b f(x)dx \quad (58)$$

численными методами.

Такого рода задачи возникают, если не удается найти первообразную подынтегральной функции или она имеет сложный и неудобный вид. В случае, когда подынтегральная функция  $f(x)$  задается таблично, аналитические методы просто неприменимы. Задачи численного интегрирования встречаются в вычислительной математике, например, при решении дифференциальных уравнений, определении площадей, объемов, а также всевозможных интегральных характеристик задач естествознания.

Наша цель – получить формулы для вычисления определенных интегралов и сравнить эти формулы по точности.

Методы численного интегрирования основаны на замене подынтегральной функции  $f(x)$  более простой интерполирующей или аппроксимирующей функцией (прямой, параболой, полиномом), интеграл от которой можно вычислить аналитически. В результате получают формулы интегрирования, называемые *квadrатурными*.

Зададим на отрезке  $[a, b]$  множество  $\omega$  точек  $x_i, i = 1, \dots, N$ , называемое сеткой:  $\omega = \{x_i; a = x_0 < x_1 < \dots < x_N = b\}$ .

Как известно, формула для вычисления интегралов численными методами (см. разд. 1, пример 1) может быть представлена в виде взвешенной суммы ординат подынтегральной функции в отдельных точках:

$$I_N(f) = \sum_{i=0}^N c_i f(x_i).$$

Это общая квадратурная формула, в которой  $x_i$  – узлы разбиения интервала интегрирования,  $c_i$  – некоторые весовые коэффициенты (веса), которые в общем случае могут зависеть от значений  $x_i$  и не зависят от вида функции  $f(x)$ .

Задача численного интегрирования состоит в определении значений  $x_i$  и  $c_i$  таких, чтобы величина  $D_N(f) = \left| \sum_{i=0}^N c_i f(x_i) - \int_a^b f(x) dx \right|$  была минимальной для функции  $f(x)$  из заданного класса функций.

При построении квадратурной формулы удобно рассматривать интеграл от  $a$  до  $b$  как сумму интегралов, заданных на некоторых интервалах

$$[\alpha_i, \beta_i], \text{ то есть } \int_a^b f(x) dx = \sum \int_{\alpha_i}^{\beta_i} f(x) dx.$$

Каждый интеграл вида  $\int_{\alpha}^{\beta} f(x) dx$  сводится к вычислению интеграла единичной длины вида  $\int_0^1 \bar{f}(s) ds$  с помощью следующей замены переменной интегрирования:

$$x = \alpha + (\beta - \alpha) \cdot s, \quad s \in [0,1].$$

При этом получаем:

$$f(x) = f(\alpha + (\beta - \alpha) \cdot s) = \bar{f}(s); \quad d(\alpha + (\beta - \alpha) \cdot s) = (\beta - \alpha) ds;$$

$$\int_{\alpha}^{\beta} f(x) dx = \int_{\alpha}^{\beta} f(\alpha + (\beta - \alpha) \cdot s) d(\alpha + (\beta - \alpha) \cdot s) = (\beta - \alpha) \int_0^1 \bar{f}(s) ds.$$

Таким образом, достаточно получить формулу для вычисления интеграла  $\int_0^1 \bar{f}(s) ds$ , которую затем нетрудно использовать при определении

$$F(a,b) = \int_a^b f(x) dx.$$

Предположим, что на интервале  $[a, b]$  задана равномерная сетка, т.е.  $x_i - x_{i-1} = h$ , ( $h = \text{const}$ ), тогда

$$F(a,b) = \sum_{i=1}^N \int_{x_{i-1}}^{x_i} f(x) dx = \sum_{i=1}^N h \int_0^1 f(x_{i-1} + h \cdot s) ds = \sum_{i=1}^N h \int_0^1 \bar{f}(s) ds = I(\bar{f}).$$

## 8.2. Построение квадратурных формул

Будем искать квадратурную формулу в виде

$$I(\bar{f}) = \int_0^1 \bar{f}(s) ds = \sum_{k=0}^m P_k \bar{f}(s_k). \quad (58)$$

В выражении (58)  $0 \leq s_0 < s_1 < \dots < s_m \leq 1$ . Разбиение интервала  $[0, 1]$  точками  $s_k$  называется *шаблоном квадратурной формулы*.

Чтобы построить квадратурную формулу, необходимо определить весовые коэффициенты (веса)  $P_k$  и узлы шаблона  $s_k$  таким образом, чтобы ошибка численного интегрирования была минимальной.

Допустим, что задан некоторый шаблон  $s_k$ . Найдем  $P_k$ . Потребуем, чтобы формула интегрирования была точной для любого полинома  $P_r(s)$  степени  $r \leq m$  (где  $(m+1)$  – количество точек шаблона). Для этого необходимо и достаточно, чтобы квадратурная формула была точной для всех степенных функций  $s^\sigma$  ( $\sigma \leq m$ ), так как

$$P_r(s) = a_0 + a_1 s + a_2 s^2 + \dots + a_r s^r.$$

Заметим, что

$$I(s^\sigma) = \int_0^1 s^\sigma ds = \frac{s^{\sigma+1}}{\sigma+1} \Big|_0^1 = \frac{1}{\sigma+1}.$$

Из формулы (58) следует, что формула интегрирования будет точной при значениях весов  $P_k$ , удовлетворяющих следующей системе уравнений:

$$\left\{ \begin{array}{l} P_0 + P_1 + \dots + P_m = 1; \\ P_0 s_0 + P_1 s_1 + \dots + P_m s_m = \frac{1}{2}; \\ P_0 s_0^2 + P_1 s_1^2 + \dots + P_m s_m^2 = \frac{1}{3}; \\ \dots \dots \dots \\ P_0 s_0^m + P_1 s_1^m + \dots + P_m s_m^m = \frac{1}{m+1}. \end{array} \right.$$

Эта система является системой линейных алгебраических уравнений с определителем матрицы коэффициентов, называемым *определителем Вандермонда*.

Как известно, определитель Вандермонда никогда не равен нулю, и полученная система линейных уравнений всегда имеет единственное решение.

### 8.2.1. Формула прямоугольников

Если при вычислении интеграла  $\int_a^b f(x)dx$  функцию  $f(x)$  заменить на отрезке  $[a,b]$  горизонтальной прямой (рис. 28), то получим квадратурную формулу прямоугольников.

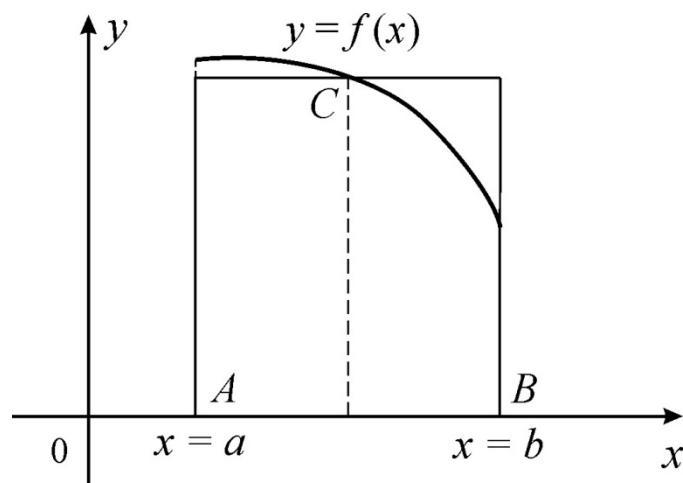


Рис. 28. Иллюстрация метода прямоугольников

Шаблон содержит один узел.

$$m=0, s_0 = \frac{1}{2}, P_0 = 1; \quad I_0(\bar{f}) = \int_0^1 \bar{f}(s)ds = P_0 \bar{f}(s_0) = P_0 \bar{f}\left(\frac{1}{2}\right) = \bar{f}\left(\frac{1}{2}\right);$$

$$I_N(f) = \sum_{i=1}^N h \int_0^1 f(x_{i-1} + hs)ds = \sum_{i=1}^N hf\left(x_{i-1} + \frac{h}{2}\right) = h \sum_{i=1}^N f\left(x_{i-1} + \frac{h}{2}\right);$$

$$I_N(f) = h \sum_{i=1}^N f\left(x_{i-1} + \frac{h}{2}\right). \quad (59)$$

### 8.2.2. Формула трапеций

Если при вычислении интеграла  $\int_a^b f(x)dx$  функцию  $f(x)$  заменить интерполяционным полиномом первой степени (линейная интерполяция), который в узлах  $x_0 = a$  и  $x_1 = b$  принимает соответственно значения  $f(a)$  и  $f(b)$  (рис. 29), то получим квадратурную формулу трапеций.

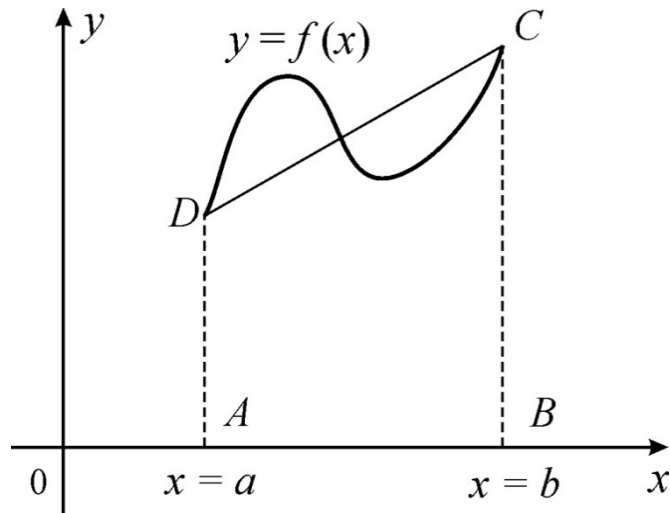


Рис. 29. Иллюстрация метода трапеций

Шаблон содержит два узла.

$$m = 1, s_0 = 0, s_1 = 1, P_0 = P_1 = \frac{1}{2}; \quad I_1(\bar{f}) = \frac{1}{2} \bar{f}(0) + \frac{1}{2} \bar{f}(1);$$

$$I_N(f) = \frac{1}{2} h \sum_{i=1}^N (f(x_{i-1}) + f(x_{i-1} + h)). \quad (60)$$

Таким образом,

$$I_N(f) = h \sum_{i=1}^N c_i f(x_{i-1}),$$

где  $c_0 = c_N = \frac{1}{2}$ , а все остальные  $c_i = 1$  ( $i = 1, \dots, N - 1$ ), то есть

$$I_N(f) = \frac{h}{2} \left( f(x_0) + 2 \cdot \sum_{i=1}^{N-1} f(x_i) + f(x_N) \right).$$

### 8.2.3. Формула Симпсона (формула парабол)

Если при вычислении  $\int_a^b f(x) dx$  функцию  $f(x)$  на  $[a, b]$  заменить интерполяционным полиномом второй степени, построенным по значениям функции  $f(x)$  в трех узлах  $x_0 = a$ ,  $x_1 = \frac{a+b}{2}$ ,  $x_2 = b$ , то получим формулу Симпсона (формулу парабол) (рис. 30).

Шаблон содержит три узла.

$$m = 2, s_0 = 0, s_1 = \frac{1}{2}, s_2 = 1.$$

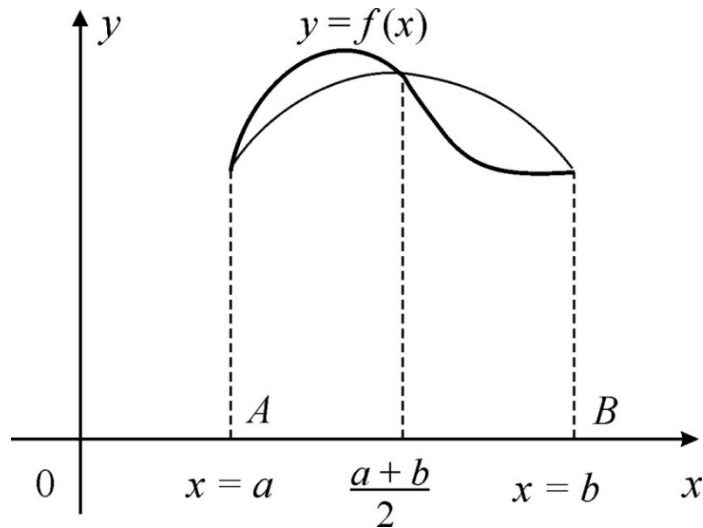


Рис. 30. Иллюстрация метода парабол

Для определения коэффициентов  $P_i$  имеем систему:

$$\begin{cases} P_0 + P_1 + P_2 = 1; \\ P_1 \cdot \frac{1}{2} + P_2 \cdot 1 = \frac{1}{2}; \\ P_1 \cdot \frac{1}{4} + P_2 \cdot 1 = \frac{1}{3}. \end{cases}$$

Следовательно,  $P_0 = \frac{1}{6}$ ,  $P_1 = \frac{4}{6}$ ,  $P_2 = \frac{1}{6}$ , и, значит,

$$I_2(\bar{f}) = \frac{1}{6} \cdot \bar{f}(0) + \frac{4}{6} \cdot \bar{f}\left(\frac{1}{2}\right) + \frac{1}{6} \cdot \bar{f}(1).$$

Отсюда следует квадратурная формула Симпсона:

$$I_N(f) = \frac{1}{6} h \sum_{i=1}^N \left( f(x_{i-1}) + 4f\left(x_{i-1} + \frac{1}{2}h\right) + f(x_{i-1} + h) \right). \quad (61)$$

**З а м е ч а н и е .** В формуле Симпсона число узлов обязательно нечетное, т.е. число отрезков четно.

#### 8.2.4. Формулы Ньютона – Котеса

Данный метод основан на замене подынтегральной функции параболой  $m$ -го порядка (при  $m = 1$  имеем формулу трапеций, при  $m = 2$  – формулу Симпсона).

Потребуем, чтобы формула интегрирования была точной для полинома Лагранжа  $P_m(s)$  (см. разд. 6):

$$P_m(s) = \sum_{k=0}^m l_k^m(s) \cdot f(s_k),$$

где  $l_k^m(s)$  – интерполяционные коэффициенты Лагранжа.

Найдем  $I(P_m)$ :

$$I(P_m) = \int_0^1 P_m(s) ds = \int_0^1 \sum_{k=0}^m l_k^m(s) f(s_k) ds = \sum_{k=0}^m f(s_k) \int_0^1 l_k^m(s) ds.$$

Квадратурная формула  $\int_0^1 f(s) ds = \sum_{k=0}^m P_k f(s_k)$  является точной формулой для интегрирования полинома степени  $m$ , если

$$P_k = \int_0^1 l_k^m(s) ds.$$

*Формула Котеса для четырехточечного шаблона.* Разобьем интервал  $[0,1]$  точками  $s_k = \frac{k}{3}$ ,  $k = 0, 1, 2, 3$  (число интервалов кратно трем), то есть  $s_0 = 0$ ,  $s_1 = \frac{1}{3}$ ,  $s_2 = \frac{2}{3}$ ,  $s_3 = 1$ ;  $m = 3$ .

Тогда

$$P_0 = \int_0^1 \frac{\left(s - \frac{1}{3}\right)\left(s - \frac{2}{3}\right)(s - 1)}{\left(0 - \frac{1}{3}\right)\left(0 - \frac{2}{3}\right)(0 - 1)} ds = \frac{1}{8};$$

$$P_1 = \int_0^1 \frac{(s - 0)\left(s - \frac{2}{3}\right)(s - 1)}{\left(\frac{1}{3} - 0\right)\left(\frac{1}{3} - \frac{2}{3}\right)\left(\frac{1}{3} - 1\right)} ds = \frac{3}{8};$$

Аналогично  $P_2 = \frac{3}{8}$ ;  $P_3 = \frac{1}{8}$ .

Следовательно,

$$I_3(\bar{f}) = \frac{1}{8} \left( \bar{f}(0) + 3\bar{f}\left(\frac{1}{3}\right) + 3\bar{f}\left(\frac{2}{3}\right) + \bar{f}(1) \right); \quad (62)$$

$$I_N(f) = \sum_{i=0}^{N-1} \frac{1}{8} h \left( f(x_i) + 3f\left(x_i + \frac{1}{3}h\right) + 3f\left(x_i + \frac{2}{3}h\right) + f(x_i + 1 \cdot h) \right).$$

Метод Ньютона – Котеса является обобщением рассмотренных нами методов численного интегрирования. Весовые коэффициенты не зависят от функции, зависят от количества точек шаблона. Выпишем их значения:

$m = 1$ ,  $P_0 = P_1 = \frac{1}{2}$  – метод трапеций;

$m = 2$ ,  $P_0 = \frac{1}{6}$ ,  $P_1 = \frac{4}{6}$ ,  $P_2 = \frac{1}{6}$  – метод Симпсона.

Метод Ньютона – Котеса соответственно для четырех-, пяти- и шеститочечного шаблона:

$$m = 3, P_0 = P_3 = \frac{1}{8}, P_1 = P_2 = \frac{3}{8};$$

$$m = 4, P_0 = P_4 = \frac{7}{90}, P_1 = P_3 = \frac{32}{90}, P_2 = \frac{2}{15};$$

$$m = 5, P_0 = P_5 = \frac{19}{288}, P_1 = P_4 = \frac{25}{96}, P_2 = P_3 = \frac{25}{144}.$$

Наиболее употребляемой среди квадратурных формул Ньютона – Котеса является формула Симпсона, поскольку она точна для полиномов не только второй, но и третьей степени.

### 8.2.5. Квадратурная формула Чебышева

Во всех предыдущих методах интервал интегрирования делился на равные участки, а весовые коэффициенты (в общем случае) были различны. Для некоторых задач хорошей точности метода численного интегрирования можно добиться исходя из оптимального выбора узлов интегрирования. Предположим, что все коэффициенты  $P_i$  квадратурной формулы равны

одной и той же константе  $P_i = \frac{1}{m}$ , то есть

$$I(f) = \frac{1}{m} \sum_{i=1}^m f(s_i).$$

Как и при построении предыдущих квадратурных формул, потребуем точности формулы интегрирования для одночленов  $1, s, s^2, s^3, \dots, s^N$ .

Получим систему уравнений для определения  $s_i$ :

$$s_1^k + s_2^k + \dots + s_m^k = \frac{m}{k+1}, \quad k = 1, 2, \dots, m.$$

Решая эту систему нелинейных алгебраических уравнений, можно получить значения неизвестных  $s_i$  узлов разбиения интервала интегрирования.

Например, при  $m = 3$  формула Чебышева имеет вид:

$$\int_0^1 \bar{f}(x) dx \approx J_3[\bar{f}] = \frac{1}{3} \left[ \bar{f}\left(\frac{1}{2} - \frac{1}{4}\sqrt{2}\right) + \bar{f}\left(\frac{1}{2}\right) + \bar{f}\left(\frac{1}{2} + \frac{1}{4}\sqrt{2}\right) \right]. \quad (63)$$

Запишем формулу для всего интервала интегрирования  $[a, b]$ .

Пусть  $\{x_i\}$  – разбиение интервала  $[a, b]$ ,  $x = a = (b - a)s$ ,  $x_i - x_{i-1} = h$ .



Тогда

$$\begin{aligned}
 F(a,b) &= \int_a^b f(x)dx = \sum_{i=1}^N \int_{x_{i-1}}^{x_i} f(x)dx = \sum_{i=1}^N h \int_0^1 \bar{f}(s)ds = \sum_{i=1}^N h \sum_{k=0}^m P_k \bar{f}(s_k) = \\
 &= \sum_{i=1}^N h \sum_{k=0}^m P_k f(x_{i-1} + hs_k) = h \sum_{i=1}^N \sum_{k=0}^m P_k f(x_{i-1} + hs_k).
 \end{aligned}$$

Тогда для трехточечного шаблона формула Чебышева имеет вид:

$$\begin{aligned}
 I_N(f) &= \int_a^b f(x)dx = \\
 &= \frac{h}{3} \sum_{i=1}^N \left[ f\left(x_i + h\left(\frac{1}{2} - \frac{1}{4}\sqrt{2}\right)\right) + f\left(x_i + \frac{h}{2}\right) + f\left(x_i + h\left(\frac{1}{2} + \frac{1}{4}\sqrt{2}\right)\right) \right].
 \end{aligned} \tag{64}$$

### 8.2.6. Квадратурная формула Гаусса

Для получения квадратурной формулы отыскиваются оптимальные значения узлов интегрирования  $s_i$  и коэффициентов  $c_i$  квадратурной формулы  $I(f) = \sum_{i=0}^m c_i f(s_i)$  одновременно.

Потребуем точности формулы для полиномов степени, меньшей либо равной  $(2m + 1)$ .

Учитывая, что  $I(s^m) = \int_0^1 s^m dx = \frac{s^{m+1}}{m+1} \Big|_0^1 = \frac{1}{m+1}$ , получим систему  $(2m + 2)$

алгебраических уравнений с  $(2m + 2)$  неизвестными:

$$\begin{cases}
 c_0 + c_1 + \dots + c_m = 1; \\
 c_0 s_0 + c_1 s_1 + \dots + c_m s_m = \frac{1}{2}; \\
 c_0 s_0^2 + c_1 s_1^2 + \dots + c_m s_m^2 = \frac{1}{3}; \\
 \dots\dots\dots \\
 c_0 s_0^{2m+1} + c_1 s_1^{2m+1} + \dots + c_m s_m^{2m+1} = \frac{1}{2(m+1)}.
 \end{cases}$$

Например, при  $m = 2$  получим следующую формулу:

$$\int_0^1 \bar{f}(s)ds \approx J_3[\bar{f}] = \frac{5}{18} \bar{f}(s_0) + \frac{8}{18} \bar{f}(s_1) + \frac{5}{18} \bar{f}(s_2), \tag{65}$$

где  $s_0 = \frac{1 - \sqrt{0,6}}{2}$ ,  $s_1 = \frac{1}{2}$ ,  $s_2 = \frac{1 + \sqrt{0,6}}{2}$ .

Аналогично формуле Чебышева можно записать:

$$I_N(f) = \int_a^b f(x) dx = \frac{h}{18} \sum_{i=1}^N \left[ f\left(x_i + h\left(\frac{1-\sqrt{0,6}}{2}\right)\right) + f\left(x_i + \frac{h}{2}\right) + f\left(x_i + h\left(\frac{1+\sqrt{0,6}}{2}\right)\right) \right]. \quad (66)$$

### 8.3. Оценки погрешности квадратурных формул

Существуют следующие оценки погрешности квадратурных формул

$$D_N(f) = \left| \sum_{i=0}^N c_i f(x_i) - \int_a^b f(x) dx \right|, \text{ полученные на основе разложения интегрируемой функции в ряд Тейлора, и оценки остаточного члена разложения.}$$

Для формулы прямоугольников

$$D_N(f) = -\frac{h^2}{24} f''(\xi)(b-a); \quad D_N(f) \leq -\frac{h^2}{24} (b-a) \max_{\xi} f''(\xi); \quad (67)$$

здесь и далее  $\xi$  – некоторая точка  $\xi \in [a, b]$ ;  $h$  – шаг интегрирования.

Для формулы трапеций

$$D_N(f) = \frac{h^2}{12} f''(\xi)(b-a). \quad (68)$$

Для формулы парабол (Симпсона)

$$D_N(f) = \frac{b-a}{180} h^4 f''(\xi). \quad (69)$$

Очевидно, что точное значение интеграла в методе прямоугольников можно получить только при  $f(x)$  – кусочно-постоянной, а в методе трапеций – при  $f(x)$  линейной или кусочно-линейной функции.

**Пример 36.**

Вычислить  $I = \int_0^1 \frac{dx}{1+x^2}$  по формулам Ньютона – Лейбница, прямоугольников, трапеций и Симпсона с использованием средств *MS Excel*.

**Решение.**

Для заданной функции формула Ньютона – Лейбница имеет вид:

$$\int_0^1 \frac{dx}{1+x^2} = \text{arctg}x \Big|_0^1 = \text{arctg}1 - \text{arctg}0.$$

Выполним вычисления в *MS Excel*.

1) Протабулируем функцию в диапазоне изменения значения аргумента 0 – 1 (рис. 31). В ячейке **H2** зададим шаг.

- 2) В ячейку **D2** введем формулу вычисления частных сумм  $=1/(1+(B2+(\$H\$2/2))^2)$ , скопируем ее буксировкой до ячейки **D12**.
- 3) В ячейку **D13** введем формулу метода прямоугольников (59):  $=\$H\$2*СУММ(D2:D12)$ .
- 4) В ячейку **D14** введем формулу метода трапеций (60):  $=\$H\$2/2*(C2+2*СУММ(C3:C11)+C12)$ .
- 5) В ячейку **F2** введем формулу вычисления частных сумм  $=C2+4*D2+C3$ , скопируем ее буксировкой до ячейки **F11**.
- 6) В ячейку **F13** введем формулу метода Симпсона (61):  $=\$H\$2/6*СУММ(F2:F11)$ .
- 7) В ячейку **F14** введем формулу Ньютона – Лейбница  $=ATAN(B12)-ATAN(B2)$ .

Сравнивая результаты вычислений, отметим, что наиболее точным оказался результат, полученный по формуле Симпсона.

	A	B	C	D	E	F	G	H
1	i	x	f(x)					h
2	0	0	1	0,997506		5,980124		0,1
3	1	0,1	0,990099	0,977995		5,863618		
4	2	0,2	0,961538	0,941176		5,643676		
5	3	0,3	0,917431	0,890869		5,342975		
6	4	0,4	0,862069	0,831601		4,988472		
7	5	0,5	0,8	0,767754		4,606311		
8	6	0,6	0,735294	0,702988		4,218386		
9	7	0,7	0,671141	0,64		3,840897		
10	8	0,8	0,609756	0,580552		3,484448		
11	9	0,9	0,552486	0,525624		3,154983		
12	10	1	0,5	0,475624				
13				0,833169	метод прямоугольников	0,785398	метод Симпсона	
14				0,784981	метод трапеций	0,785398	формула Ньютона- Лейбница	

Рис. 31. Пример численного интегрирования функции в MS Excel

## Контрольные вопросы и задания

1. Какие формулы называются квадратурными?
2. Что такое шаблон квадратурной формулы?
3. Как получается определитель Вандермонда?
4. Сколько узлов содержит шаблон квадратурной формулы трапеций?
5. В какой квадратурной формуле число узлов должно быть нечетным?
6. В каком методе численного интегрирования исходят из оптимального выбора узлов интегрирования?

7. Вычислите интеграл  $I = \int_0^{\frac{\pi}{2}} \sin x dx$  по формулам Котеса (62), Чебышева (64) и Гаусса (66) в среде *MS Excel* и *MathCAD*.

## 9. РАЗНОСТНЫЕ УРАВНЕНИЯ

### 9.1. Сеточные функции

Как мы уже знаем, в приближенных методах функции непрерывного аргумента заменяются функциями дискретного аргумента – сеточными функциями. *Сеточную функцию* можно рассматривать как функцию целочисленного аргумента:

$$y(i) = y_i, \quad i = 0, \pm 1, \pm 2, \dots$$

Для  $y(i)$  можно ввести операции, являющиеся дискретным (разностным) аналогом операций дифференцирования и интегрирования.

Аналогом первой производной являются разности первого порядка:

$$\Delta y_i = y_{i+1} - y_i \quad \text{– правая разность,}$$

$$\nabla y_i = y_i - y_{i-1} \quad \text{– левая разность,}$$

$$\delta y_i = \frac{1}{2}(\Delta y_i + \nabla y_i) = \frac{1}{2}(y_{i+1} - y_{i-1}) \quad \text{– центральная разность,}$$

при этом ( $\Delta y_i = \nabla y_{i+1}$ ).

Разности второго порядка:

$$\Delta^2 y_i = \Delta(\Delta y_i) = y_{i+2} - 2y_{i+1} + y_i;$$

$$\nabla \Delta y_i = y_{i+1} - 2y_i + y_{i-1} \Rightarrow \nabla \Delta y_i = \Delta^2 y_{i-1}.$$

Аналогично определяется разность  $m$ -го порядка  $\Delta^m y_i = \Delta(\Delta^{m-1} y_i)$ , содержащая значения  $y_i, y_{i+1}, \dots, y_{i+m}$ . Очевидно, что

$$\sum_{j=k}^i \Delta y_j = y_{i+1} - y_k; \quad \sum_{j=k}^i \nabla y_j = y_i - y_{k-1}.$$

#### 9.1.1. Разностные аналоги формул дифференцирования произведения и интегрирования по частям

Пусть  $y_i, v_i$  – произвольные функции целочисленного аргумента. Тогда

$$\Delta(y_i v_i) = y_i \Delta v_i + v_{i+1} \Delta y_i = y_{i+1} \Delta v_i + v_i \Delta y_i; \quad (70)$$

$$\nabla(y_i v_i) = y_{i-1} \nabla v_i + v_i \nabla y_i = y_i \nabla v_i + v_{i-1} \nabla y_i. \quad (71)$$

Формулы (70) и (71) представляют собой аналоги формулы дифференцирования произведения  $(y(x)v(x))' = yv' + yv'$ .

Аналогом формулы интегрирования по частям является формула суммирования по частям:

$$\sum_{i=0}^{N-1} y_i \Delta v_i = - \sum_{i=1}^N v_i \nabla y_i + y_N v_N - y_0 v_0. \quad (72)$$

Действительно, поскольку  $\Delta y_i = \nabla y_{i+1}$ , из формулы (70) следует, что

$$y_i \Delta v_i = \Delta(y_i v_i) - v_{i+1} \Delta y_i = \Delta(y_i v_i) - v_{i+1} \nabla y_{i+1},$$

отсюда получаем:

$$\begin{aligned} \sum_{i=0}^{N-1} y_i \Delta v_i + \sum_{i=1}^N v_i \nabla y_i &= \sum_{i=0}^{N-1} \Delta(y_i v_i) - \sum_{i=0}^{N-1} v_{i+1} \nabla y_{i+1} + \sum_{i=1}^N v_i \nabla y_i = \\ &= y_N v_N - y_0 v_0 - \sum_{i=1}^N v_i \nabla y_i + \sum_{i=1}^N v_i \nabla y_i = y_N v_N - y_0 v_0. \end{aligned}$$

Что и требовалось доказать.

Формулу (72) можно использовать для вычисления сумм.

**Пример 37.**

Вычислить сумму  $S_N = \sum_{i=1}^N i 2^i$ .

Положим  $v_i = i$ ,  $\nabla y_i = 2^i$ , так что  $y_i = y_{i-1} + 2^i = y_0 + \sum_{j=1}^i 2^j = y_0 + 2^{i+1} - 2$ .

Выберем  $y_0 = 2 - 2^{N+1}$ ; тогда  $y_N = 0$ . Так как  $v_0 = 0$ ,  $\Delta v_i = 1$ , то из формулы (72) следует, что

$$S_N = \sum_{i=1}^N v_i \nabla y_i = - \sum_{i=0}^{N-1} y_i \Delta v_i = - \sum_{i=0}^{N-1} y_i = -N(y_0 - 2) - \sum_{i=0}^{N-1} 2^{i+1} = N 2^{N+1} - (2^{N+1} - 2),$$

так что  $S_N = (N-1)2^{N+1} + 2$ .

## 9.2. Разностные уравнения

Линейное уравнение относительно сеточной функции  $y_i = y(i)$  ( $i = 0, \pm 1, \pm 2, \dots$ )

$$a_0(i)y(i) + a_1(i)y(i+1) + \dots + a_m(i)y(i+m) = f(i), \quad (73)$$

где  $a_k(i)$  ( $k = 0, 1, 2, \dots, m$ ),  $f(i)$  – заданные сеточные функции,  
 $a_0(i) \neq 0$ ,  $a_m(i) \neq 0$ ,

называется *линейным разностным уравнением  $m$ -го порядка*.

Если использовать формулы конечных разностей, то можно уравнение (73) переписать в виде:

$$\bar{a}_0(i)y_i + \bar{a}_1(i)\Delta y_i + \dots + \bar{a}_m(i)\Delta^m y_i = f(i). \quad (74)$$

Если коэффициенты  $a_0, a_1, \dots, a_m$  не зависят от  $i$ , то уравнение (73) называется *разностным уравнением с постоянными коэффициентами*.

### 9.2.1. Уравнения первого порядка

При  $m = 1$  из уравнения (73) получаем разностное уравнение первого порядка:

$$a_0(i)y_i + a_1(i)y_{i+1} = f(i), \quad a_0(i) \neq 0, \quad a_1(i) \neq 0 \quad (75)$$

или

$$\bar{a}_0(i)y_i + a_1(i)\Delta y_i = f(i), \quad \bar{a}_0 = a_0 + a_1.$$

Пример 38.

Разностными уравнениями первого порядка являются формулы для вычисления арифметической и геометрической прогрессий:  
 $y_{i+1} = y_i + d$ ,  $y_{i+1} = qy_i$ .

Уравнение (74) можно переписать в виде  $y_{i+1} = q_i y_i + \varphi_i$ , где  $q_i = -\frac{a_0(i)}{a_1(i)}$ ,  $\varphi_i = -\frac{f(i)}{a_1(i)}$ . Отсюда видно, что решение  $y(i)$  определено однозначно при  $i > i_0$ , если задано значение  $y(i_0)$ . Значения  $y_1, y_2, \dots, y_i, \dots$  можно вычислить по формуле

$$y_{i+1} = \left( \prod_{k=0}^i q_k \right) y_0 + \sum_{k=0}^{i-1} \left( \prod_{s=k+1}^i q_s \right) \varphi_k + \varphi_i. \quad (76)$$

Для уравнения с постоянным коэффициентом формула примет вид:

$$y_{i+1} = q^{i+1} y_0 + \sum_{k=0}^i q^{i-k} \varphi_k. \quad (77)$$

### 9.2.2. Уравнения второго порядка с постоянными коэффициентами

Рассмотрим разностное уравнение второго порядка

$$by_{i+1} - cy_i + ay_{i-1} = f_i, \quad i = 1, 2, \dots, \quad a \neq 0, \quad b \neq 0, \quad (78)$$

коэффициенты которого не зависят от  $i$ .

Если  $f_i = 0$ , то уравнение

$$by_{i+1} - cy_i + ay_{i-1} = 0, \quad i = 1, 2, \dots \quad (79)$$

называется *однородным*. Его решение может быть найдено в явном виде.

Пусть  $\bar{y}_i, i = 0, 1, \dots$  – решение однородного уравнения (79),  $y_i^*, i = 0, 1, \dots$  – какое-либо решение неоднородного уравнения (78). Тогда их сумма  $y_i = \bar{y}_i + y_i^*$  также является решением неоднородного уравнения. Действительно,

$$\begin{aligned} & b(\bar{y}_{i+1} + y_{i+1}^*) - c(\bar{y}_{i+1} + y_{i+1}^*) + a(\bar{y}_{i+1} + y_{i+1}^*) = \\ & = [b\bar{y}_{i+1} - c\bar{y}_{i+1} + a\bar{y}_{i+1}] + [by_{i+1}^* - cy_{i+1}^* + ay_{i+1}^*] = f_i. \end{aligned}$$

Это свойство – следствие линейности уравнения (75); оно сохраняет силу для разностного уравнения (73) любого порядка.

Очевидно, что если  $\bar{y}_i$  является решением однородного уравнения (79), то и  $c\bar{y}_i$ , где  $c$  – произвольная постоянная, также удовлетворяет данному уравнению.

Пусть  $y_i^{(1)}$  и  $y_i^{(2)}$  – два решения уравнения (79). Они являются линейно независимыми, если равенство  $c_1 y_i^{(1)} + c_2 y_i^{(2)} = 0, i = 0, 1, 2, \dots$ , что возможно только при  $c_1 = c_2 = 0$ , или если определитель системы

$$\begin{cases} c_1 y_i^{(1)} + c_2 y_i^{(2)} = 0, \\ c_1 y_{i+m}^{(1)} + c_2 y_{i+m}^{(2)} = 0, \quad m = \pm 1, \pm 2, \dots \end{cases}$$

отличен от нуля для всех  $i, m$ . В частности,

$$\Delta_{i,i+1} = \begin{vmatrix} y_i^{(1)} & y_i^{(2)} \\ y_{i+1}^{(1)} & y_{i+1}^{(2)} \end{vmatrix} \neq 0.$$

Принимаем без доказательства два утверждения.

1. Если решения  $y_i^{(1)}$  и  $y_i^{(2)}$  уравнения (79) линейно независимы, то общее решение однородного разностного уравнения (79) имеет вид

$$y_i^0 = c_1 y_i^{(1)} + c_2 y_i^{(2)},$$

где  $c_1, c_2$  – произвольные постоянные.

2. Общее решение неоднородного уравнения (78) можно представить в виде

$$y_i = y_i^0 + y_i^* = c_1 y_i^{(1)} + c_2 y_i^{(2)} + y_i^*, \quad (80)$$

где  $y_i^*$  – частное решение уравнения (78). Для определения  $c_1$  и  $c_2$ , как и в случае дифференциальных уравнений, надо задать дополнительные условия – начальные или краевые.

### 9.2.3. Нахождение общего решения однородного разностного уравнения

Общее решение уравнения (80) можно найти явно. Для этого линейно независимые решения будем искать в виде  $y_i = q^i$ , где  $q \neq 0$  – неизвестное пока число. После подстановки  $y_k = q^k$  в уравнение (79) получим квадратное уравнение  $bq^2 - cq + a = 0$ , имеющее корни

$$q_1 = \frac{c + \sqrt{c^2 - 4ab}}{2b}, \quad q_2 = \frac{c - \sqrt{c^2 - 4ab}}{2b}. \quad (81)$$

В зависимости от значений дискриминанта  $D = c^2 - 4ab$  возможны три случая.

1)  $D > 0$ , корни действительны и различны. Им соответствуют два решения  $y_k^{(1)} = q_1^k$  и  $y_k^{(2)} = q_2^k$ , которые линейно независимы, так как отличен от нуля определитель

$$\Delta_{k,k+1} = \begin{vmatrix} q_1^k & q_1^{k+1} \\ q_2^k & q_2^{k+1} \end{vmatrix} = q_1^k q_2^k (q_2 - q_1) \neq 0.$$

Общее решение уравнения (79) имеет вид

$$y_k = c_1 q_1^k + c_2 q_2^k. \quad (82)$$



2)  $D < 0$ , корни комплексно-сопряженные:

$$q_1 = \frac{c + j\sqrt{|D|}}{2b}, \quad q_2 = \frac{c - j\sqrt{|D|}}{2b},$$

где  $j$  – мнимая единица. Эти корни удобно представить в виде

$$q_1 = \rho e^{j\varphi}, \quad q_2 = \rho e^{-j\varphi}, \quad \rho = \sqrt{\frac{a}{b}}, \quad \varphi = \operatorname{arctg} \frac{\sqrt{|D|}}{c}.$$

Линейно независимыми решениями являются не только функции

$$q_1^k = \rho^k e^{jk\varphi} = \rho^k (\cos k\varphi + j \sin k\varphi),$$

$$q_2^k = \rho^k e^{-jk\varphi} = \rho^k (\cos k\varphi - j \sin k\varphi),$$

но и функции  $y_k^{(1)} = \rho^k \cos k\varphi$ ,  $y_k^{(2)} = \rho^k \sin k\varphi$ , линейная независимость которых следует из линейной независимости функций  $\sin k\varphi$  и  $\cos k\varphi$ . Общее решение уравнения (79) имеет вид

$$y_k = \rho^k (c_1 \cos k\varphi + c_2 \sin k\varphi). \quad (83)$$

3)  $D = 0$ , корни действительны и равны:

$$q_1 = q_2 = \frac{c}{2b} = q_0.$$

Линейно независимыми являются решения

$$y_k^{(1)} = q_0^k, \quad y_k^{(2)} = kq_0^k. \quad (84)$$

Общее решение уравнения (75) имеет вид

$$y_k = c_1 q_0^k + c_2 k q_0^k. \quad (85)$$

**Пример 39.**

Найти общее решение уравнения

$$y_{k+1} - 2py_k + y_{k-1} = 0, \quad a = b = 1, \quad c = 2p > 0.$$

**Решение.**

Возможны три случая:

1)  $p < 1$ .

Положим  $p = \cos \alpha$ ; тогда  $D = 4(\cos^2 \alpha - 1) = -4\sin^2 \alpha < 0$ . Линейно независимые решения имеют вид

$$y_k^{(1)} = \cos k\alpha, \quad y_k^{(2)} = \sin k\alpha;$$

2)  $p > 1$ .

Положим  $p = \operatorname{ch} \alpha$ . Напомним, что  $\operatorname{ch} \alpha = \frac{e^\alpha + e^{-\alpha}}{2}$ ,  $\operatorname{sh} \alpha = \frac{e^\alpha - e^{-\alpha}}{2}$ . Тогда получим для  $q$  квадратное уравнение  $q^2 - 2\operatorname{ch}(\alpha q) + 1 = 0$ ; его дискриминант  $D = 4(\operatorname{ch}^2 \alpha - 1) = 4\operatorname{sh}^2 \alpha$ , а корни имеют вид  $q_{1,2} = \operatorname{ch} \alpha \pm \operatorname{sh} \alpha = e^{\pm \alpha}$ . Линейно независимыми решениями являются функции

$$y_k^{(1)} = \operatorname{ch}(k\alpha), \quad y_k^{(2)} = \operatorname{sh}(k\alpha);$$

3)  $p = 1$ .

В этом случае  $q^2 - 2q + 1 = 0$ ,  $q_{1,2} = 1$ , линейно независимые решения имеют вид

$$y_k^{(1)} = 1, y_k^{(2)} = k.$$

Во всех трех случаях общее решение уравнения будет иметь вид

$$y_k = c_1 y_k^{(1)} + c_2 y_k^{(2)}.$$

**Пример 40.**

Найти общее решение уравнения

$$y_{k+1} - y_k - 6y_{k-1} = 2^{k+1}.$$

**Решение.**

Найдем сначала общее решение однородного уравнения.

$D = 1 + 24 = 25 > 0$ ; корни квадратного уравнения  $q^2 - q - 6 = 0$  равны  $q_1 = 3$ ,  $q_2 = -2$ , так что  $y_k^{(1)} = 3^k$ ,  $y_k^{(2)} = (-2)^k$ .

Частное решение находим в виде  $y_k^* = c \cdot 2^k$ , где  $c = \text{const}$ . Подставив значение  $y_k^*$  в исходное уравнение, получим:

$$c(2^{k+1} - 2^k - 6 \cdot 2^{k-1}) = c \cdot 2^{k-1}(-4) = 2^{k+1}, c = -1.$$

Общее решение неоднородного уравнения представляет собой сумму  $y_i = \bar{y}_i + y_i^*$  общего решения однородного уравнения и частного решения неоднородного. Следовательно, общее решение будет иметь вид

$$y_k = c_1 \cdot 3^k + c_2 (-2)^k - 2^k.$$

### Контрольные вопросы и задания

1. Какой вид имеет линейное разностное уравнение  $m$ -го порядка?
2. От чего зависит порядок разностного уравнения?
3. Приведите примеры разностных уравнений первого порядка.
4. Какая функция называется сеточной? Приведите пример.
5. Что собой представляет сетка для построения дискретного аналога дифференциального уравнения?
6. Что является аналогом первой производной?
7. Какая формула является разностным аналогом формулы дифференцирования по частям, формулы интегрирования по частям?
8. Какие разностные уравнения называются однородными?
9. Как выглядит разностное уравнение первого порядка с постоянными коэффициентами?
10. Как вычисляются разности первого порядка: правая, левая, центральная разность?
11. Если общее решение неоднородного уравнения представлено в виде суммы  $y_i = y_i^0 + y_i^*$ , то чем являются каждое из этих слагаемых?
12. Найдите общее решение уравнения  $y_{k+1} + y_k - 6y_{k-1} = 2^{k+1}$ .

## 10. ЗАДАЧА КОШИ И КРАЕВАЯ ЗАДАЧА ДЛЯ РАЗНОСТНЫХ УРАВНЕНИЙ ВТОРОГО ПОРЯДКА

### 10.1. Разностное уравнение второго порядка с переменными коэффициентами

Рассмотрим разностное уравнение второго порядка с переменными коэффициентами

$$b_i y_{i+1} - c_i y_i + a_i y_{i-1} = f_i, \quad i = 0, 1, \dots, \quad a \neq 0, \quad b \neq 0, \quad (86)$$

или

$$y_{i+1} = \frac{c_i y_i - a_i y_{i-1} + f_i}{b_i}, \quad b \neq 0. \quad (87)$$

Выразим  $y_{i+1}$  и  $y_{i-1}$  через  $y_i$  и разности первого и второго порядков. Тогда уравнение (86) будет иметь вид

$$\Delta \nabla y_i + (b_i - a_i) \Delta y_i - (c_i - a_i - b_i) y_i = f_i, \quad a_i \neq 0, \quad b_i \neq 0.$$

Ранее было отмечено, что для решения разностного уравнения первого порядка требуется начальное условие, например  $y_0 = c_0$ . Для решения разностного уравнения второго порядка необходимы два начальных условия. Если оба условия заданы в двух соседних точках, то имеем *задачу Коши* для разностного уравнения второго порядка. Если же два условия заданы в двух разных (но не соседних) точках, то такая задача называется *краевой*.

Введем обозначение оператора

$$Ly_i = b_i y_{i+1} - c_i y_i + a_i y_{i-1}$$

и сформулируем эти задачи.

**Задача Коши:** найти решение уравнения

$$Ly_i = f_i, \quad i = 1, 2, \dots \quad (88)$$

при дополнительных условиях

$$y_0 = \mu_1, \quad y_1 = \mu_2. \quad (89)$$

Условие (89) запишем в виде  $\Delta y_0 = y_1 - y_0 = \mu_2 - \mu_1 = \bar{\mu}_1$  и будем говорить, что в случае задачи Коши заданы в одной точке  $i = 0$  величины

$$y_0 = \mu_1, \quad \Delta y_0 = \bar{\mu}_1. \quad (90)$$

**Краевая задача:** найти решение уравнения

$$Ly_i = f_i, \quad i = 1, 2, \dots, N-1 \quad (91)$$

при дополнительных условиях

$$y_0 = \mu_1, \quad y_N = \mu_2, \quad N \geq 2. \quad (92)$$

В граничных узлах  $i = 0$  и  $i = N$  можно задать не только значения функций, но и их комбинации с разностями, т.е. выражения  $\alpha_1 \Delta y_0 + \beta_1 y_0$  при  $i = 0$  и  $\alpha_2 \Delta y_N + \beta_2 y_N$  при  $i = N$ . Такие условия можно записать в виде

$$y_0 = \chi_1 y_1 + \mu_1, \quad y_N = \chi_2 y_{N-1} + \mu_2. \quad (93)$$

Если  $\chi_1 = \chi_2 = 0$ , то имеем условия первого рода; при  $\chi_1 = \chi_2 = 1$  – условия второго рода:

$$\Delta y_0 = -\mu_1, \quad \nabla y_N = \mu_2. \quad (94)$$

Если  $\chi_{1,2} \neq 0$  и  $\chi_{1,2} \neq 1$ , то формулы (92) называют условиями третьего рода:

$$-\chi_1 \Delta y_0 + (1 - \chi_1) y_0 = \mu_1; \quad \chi_2 \nabla y_N + (1 - \chi_2) y_N = \mu_2. \quad (95)$$

Возможны краевые задачи с комбинацией этих краевых условий – смешанные задачи: при  $i = 0$  – условия одного типа, при  $i = N$  – условия другого типа.

Решение задачи Коши находится непосредственно из уравнения (88) по рекуррентной формуле (87) с учетом начальных данных  $y_0 = \mu_1$ ,  $y_1 = \mu_2$ . Краевые задачи решаются методом исключения.

### 10.1.1. Решение разностных краевых задач методом прогонки

Краевая задача

$$a_i y_{i-1} - c_i y_i + b_i y_{i+1} = -f_i, \quad a_i \neq 0, \quad b_i \neq 0, \quad i = 1, 2, \dots, N-1, \quad (96)$$

$$y_0 = \chi_1 y_1 + \mu_1, \quad y_N = \chi_2 y_{N-1} + \mu_2$$

представляет собой систему линейных алгебраических уравнений с трехдиагональной матрицей размером  $(N+1) \times (N+1)$ :

$$A = \begin{bmatrix} 1 & -\chi_1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ a_1 & -c_1 & b_1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_i & -c_i & b_i & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & a_{N-1} & -c_{N-1} & b_{N-1} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & -\chi_2 & 1 \end{bmatrix}.$$

Вместо формул (96) можно записать

$$Ay = f, \quad y = (y_0, y_1, \dots, y_N), \quad f = (\mu_1, -f_1, \dots, -f_{N-1}, \mu_2). \quad (97)$$

Для решения краевой задачи (96) используем метод исключения, называемый *методом прогонки*.

Предположим, что

$$y_i = \alpha_{i+1} y_{i+1} + \beta_{i+1} \quad (98)$$

с не определенными пока коэффициентами  $\alpha_{i+1}$  и  $\beta_{i+1}$ . Подставив  $y_{i-1} = \alpha_i y_i + \beta_i$  в формулу (96), получим:

$$(a_i \alpha_i - c_i) y_i + b_i y_{i+1} = -(f_i + a_i \beta_i).$$

Сравнивая это тождество с выражением (98), находим

$$\alpha_{i+1} = \frac{b_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \dots, N-1, \quad (99)$$

$$\beta_{i+1} = \frac{a_i \beta_i + f_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \dots, N-1. \quad (100)$$

Схема решения состоит из следующих действий.

1. Используем краевое условие при  $i = 0$  для определения  $\alpha_1, \beta_1$ . Из формул (98) и (96) для  $i = 0$  находим

$$\alpha_1 = \chi_1, \quad \beta_1 = \mu_1. \quad (101)$$

2. Зная  $\alpha_i$  и  $\beta_i$  и переходя от  $i$  к  $(i+1)$  в формулах (99) и (100), определяем  $\alpha_i$  и  $\beta_i$  для всех  $i = 2, 3, \dots, N$ .

3. Находим  $y_N$  из краевого условия  $y_N = \chi_2 y_{N-1} + \mu_2$  и условия (98) при  $i = N-1$ :  $y_{N-1} = \alpha_N y_N + \beta_N$ . Отсюда:

$$y_N = \frac{\mu_2 + \chi_2 \beta_N}{1 - \alpha_N \chi_2}. \quad (102)$$

4. По формуле (98) вычислим  $y_{N-1}$  через  $y_N$ :

$$y_{N-1} = \alpha_N y_N + \beta_N.$$

Таким образом, краевая задача для уравнения второго порядка сведена к трем задачам Коши для уравнений первого порядка (стрелки указывают направление счета:  $i \rightarrow i+1, i-1 \leftarrow i$ ):

$$\begin{aligned} \alpha_{i+1} &= \frac{b_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \dots, N-1, \\ \bar{\beta}_{i+1} &= \frac{a_i \beta_i + f_i}{c_i - a_i \alpha_i}, \quad i = 1, 2, \dots, N-1, \quad \beta_1 = \mu_1; \\ \bar{y}_i &= \alpha_{i+1} y_{i+1} + \beta_{i+1}, \quad i = N-1, N-2, \dots, 2, 1, 0. \end{aligned} \quad (103)$$

### 10.1.2. Устойчивость метода прогонки

Метод прогонки можно применять, если знаменатели дробей (103) не обращаются в 0, т.е. когда

$$\begin{cases} c_i - a_i \alpha_i \neq 0; \\ 1 - \alpha_N \chi_2 \neq 0. \end{cases} \quad (104)$$

Примем без доказательства следующую теорему и утверждения 1,2, которые могут быть полезны при использовании метода прогонки для практических расчетов.

Т е о р е м а .

Достаточными условиями выполнения системы (103) являются неравенства

$$\begin{cases} |c_i| \geq |a_i| + |b_i|, & i = 1, 2, \dots, N-1, \\ |\chi_1| \leq 1, |\chi_2| \leq 1, & |\chi_1| + |\chi_2| < 2. \end{cases} \quad (105)$$

**Утверждение 1.**

При  $|\alpha_i| \leq 1$  погрешность  $\delta y_i$  не возрастает.

**Утверждение 2.**

Если  $\alpha_{i+1}$  и  $\beta_{i+1}$  вычисляются с погрешностью  $\varepsilon_0$ , то  $\max_{1 \leq i \leq N} |\delta y_i| \leq \varepsilon_0 N^2$ ,

где  $N$  – число узлов.

### Контрольные вопросы и задания

1. Сколько начальных условий необходимо для решения разностного уравнения первого, второго порядка?
2. Чем отличается задача Коши для разностного уравнения второго порядка от краевой задачи?
3. Что такое начальные условия, краевые условия?
4. Как задаются условия в краевой задаче?
5. Какую систему уравнений можно решить методом прогонки?
6. В чем заключается метод прогонки?
7. Какие дискретные аналоги используются для замены первой и второй производных функции, если при решении дифференциального уравнения невозможно получить его аналитическое решение?
8. К каким трем задачам Коши для уравнений первого порядка сводится краевая задача для уравнения второго порядка?
9. Решите методом прогонки краевую задачу:

$$Ly = f, \quad y = (y_0, y_1, y_2, y_3); \quad N = 3;$$

$$f = (0; 0,5; 1; 0,5; 1);$$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -2 & 0,5 & 0 \\ 0 & 0,5 & -2,5 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

# 11. ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

## 11.1. Формулы численного дифференцирования

В тех случаях, когда не представляется возможным произвести вычисление производной некоторой заданной функции аналитическим путем, для дифференцирования применяют численные методы. Такие задачи возникают при исследовании функций, при поисках экстремумов (минимумов и максимумов) функций, при решении алгебраических и дифференциальных уравнений и их систем и в других случаях, когда возникает необходимость в вычислении производных. Функция, подлежащая численному дифференцированию, может быть таблично заданной, или иметь очень сложный вид для вычисления производной в аналитическом виде, или её вычисление возможно только для конкретных значений аргумента.

Формулы для численного дифференцирования получаются посредством дифференцирования интерполяционных полиномов и зависят от числа узлов, окружающих точку, в которой вычисляется производная. Приведем формулы для численного дифференцирования в случае использования трех и пяти узлов; производная при этом определяется в центральном узле:

1) при трех узлах:  $x - h$ ;  $x$ ;  $x + h$ , где  $h$  – шаг дифференцирования:

$$y'(x) = \frac{y(x+h) - y(x-h)}{2h};$$
$$y''(x) = \frac{y(x-h) - 2y(x) + y(x+h)}{h^2};$$
(106)

2) при пяти узлах:  $x - 2h$ ;  $x - h$ ;  $x$ ;  $x + h$ ;  $x + 2h$ :

$$y'(x) = \frac{y(x-2h) - 8y(x-h) + 8y(x+h) - y(x+2h)}{12h};$$
$$y''(x) = \frac{-y(x-2h) + 16y(x-h) - 30y(x) + 16y(x+h) - y(x+2h)}{12h^2}.$$
(107)

Вычислить производную в произвольной точке можно с помощью непосредственного дифференцирования интерполяционного полинома. Например, продифференцируем полином Ньютона:

$$P_n(x) = y_0 + \frac{\Delta y_0}{h \cdot 1!} \cdot (x - x_0) + \frac{\Delta^2 y_0}{h^2 \cdot 2!} \cdot (x - x_0) \cdot (x - x_1) + \dots + R_n(x);$$
$$P'_n(x) = \frac{\Delta y_0}{h \cdot 1!} + \frac{\Delta^2 y_0}{h^2 \cdot 2!} \cdot [(x - x_0) + (x - x_1)] + \dots + R'_n(x).$$
(108)

При этом нетрудно проверить, что остаточный член, характеризующий ошибку вычисления производной, подчиняется неравенству

$$|R'_n(x_i)| \leq \frac{(n-i)!i!}{(n+1)!} M_{n+1} h^n, \quad (109)$$

где  $h$  – шаг интерполяции;  $M_{n+1} = \max_x |y^{(n+1)}(x)|$ .

Поскольку  $y^{(n+1)}(x)$  во многих случаях трудно оценить, то при  $h$  малом считают:

$$y^{(n+1)}(x) \approx \frac{\Delta^{n+1} y_0}{h^{n+1}}$$

и, следовательно,

$$R'_n(x_0) \approx \frac{(n-i)!i!}{(n+1)!} \cdot \frac{\Delta^{n+1} y_0}{h}.$$

Аналогично может быть найдено значение  $R''_n(x_0)$ .

**З а м е ч а н и е .** При нахождении производных  $y'(x)$ ,  $y''(x)$ , ... в фиксированной точке  $x$  в качестве  $x_0$  выбирают ближайшее табличное значение аргумента.

При численном дифференцировании таблично заданной функции  $y = f(x)$  возникают погрешности двух типов:

– *погрешности усечения*, которые вызываются заменой функции  $f(x)$  интерполяционным многочленом  $P_n(x)$ ;

– *погрешности округления*, которые вызываются неточным заданием исходных значений  $y_i$ .

Общая погрешность вычисления производной складывается из погрешности усечения и погрешности округления. Поскольку при уменьшении шага  $h$  погрешность усечения уменьшается, а погрешность округления увеличивается, то существует оптимальный шаг для каждой формулы дифференцирования. Погрешности округления быстро возрастают с ростом порядка производной.

**Пример 41.**

Пусть функция  $y = x^2$  задана с некоторой погрешностью в виде табл. 10.

Т а б л и ц а 1 0

$x$	0,5	0,7	0,9
$y$	0,25	0,5	0,8

Определить значение первой производной в точках  $x = 0,75$  и  $x = 0,7$ .



Решение.

Дополним исходную таблицу значениями конечных разностей (табл. 11).

Таблица 11

$x$	$y$	$\Delta y$	$\Delta^2 y$
0,5	0,25	0,25	0,05
0,7	0,5	0,3	
0,9	0,8		

Вычислим по формуле (108)

$$y'(0,75) = \frac{1}{0,2} \left( 0,25 + \frac{0,05 \cdot [(0,75 - 0,5) + (0,75 - 0,7)]}{0,2 \cdot 2} \right) + R'_n(0,75) = 1,438 + R'_n(0,75).$$

Поскольку  $M_3 = \max_x |y'''(x)| = 0$ , следовательно,  $R'_n(0,75) = 0$ , то  $y'(0,75) \approx 1,438$ . На рис. 32 приведено решение в *MS Excel*.

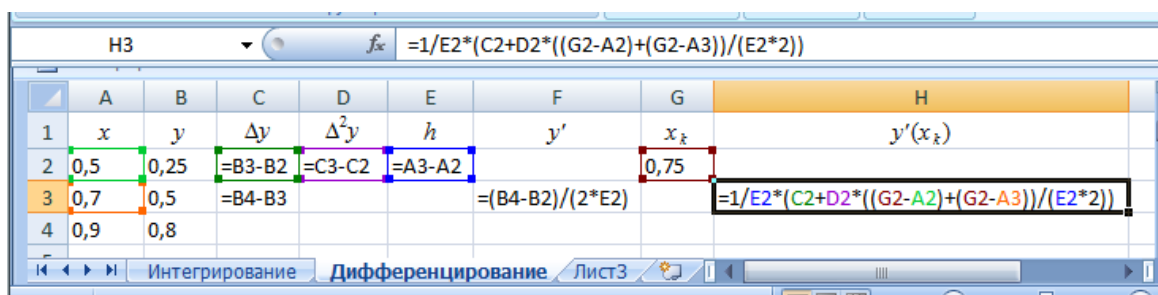
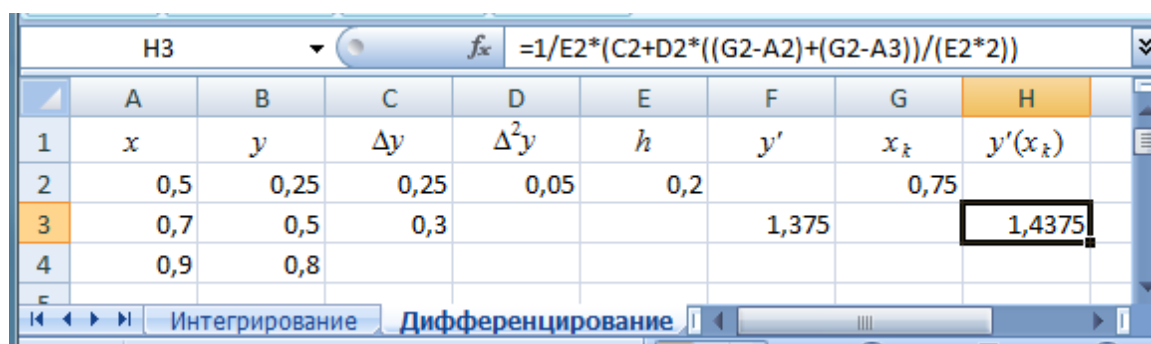


Рис. 32. Пример численного дифференцирования функции в *MS Excel*

Заметим, что для заданной функции  $y = x^2$  точное значение производной  $y'(0,75) = 1,5$ .

Определим по формуле (106)

$$y'(0,7) = \frac{y(0,9) - y(0,5)}{2 \cdot 0,2} = \frac{0,8 - 0,25}{0,4} = 1,375.$$

Точное значение производной  $y'(0,7) = 1,4$ .

## 11.2. Численные методы решения дифференциальных уравнений

Рассмотрим задачу Коши для дифференциального уравнения относительно неизвестной функции  $u(t)$ :

$$\frac{du}{dt} = F(t, u), \quad u(0) = u_0. \quad (110)$$

Здесь  $F(t, u)$  – некоторая функция;  $u_0$  – заданное начальное условие.

Пусть  $F(t, u)$  определена и непрерывна в области  $D = \{t, u : 0 \leq t \leq T, |u - u_0| < U\}$  и удовлетворяет в области  $D$  условию Липшица по переменной  $u$ , то есть для  $\forall u_1, u_2 \in D$  имеем  $|F(t, u_1) - F(t, u_2)| \leq k \cdot |u_1 - u_2|$ ,  $k < 1$ .

Из теории дифференциальных уравнений известно, что такая задача имеет единственное решение, если значение  $T$  достаточно мало.

### 11.2.1. Аппроксимация и сходимость численных методов решения дифференциальных уравнений на примере метода Эйлера

На интервале  $[0, T]$  зададим множество точек  $\bar{\omega}_\tau = \{t_n = \tau \cdot n, n = 0, 1, \dots\}$ , называемое сеткой. При этом  $\tau$  называется *параметром сетки* или параметром численного метода.

Функцию  $u(t)$  в уравнении (109) заменим функцией  $u_n$  дискретного аргумента  $t_n$ ,  $n = 0, 1, \dots$ , а производную  $\frac{du}{dt}$  – ее конечной разностью.

Получим разностный аналог уравнения (110):

$$\frac{u_{n+1} - u_n}{\tau} = F(t_n, u_n), \quad n = 0, 1, \dots \quad (111)$$

или

$$u_{n+1} = u_n + \tau \cdot F(t_n, u_n), \quad u(0) = u_0. \quad (112)$$

Схема (112) позволяет последовательно вычислить все значения  $u_n$ , т.е. найти приближенное решение задачи (110). Такой метод решения дифференциального уравнения носит название *метода Эйлера*. Наша цель – выяснить, как точно дискретная функция  $u_n$  приближает решение дифференциального уравнения (110), то есть какова точность аппроксимации непрерывного дифференциального уравнения разностным уравнением.

Оценим точность аппроксимации метода Эйлера. Положим, что  $u_n = u(t_n) + \delta u_n$ . Подставим значение  $u_n$  в разностное уравнение (111):

$$\frac{u(t_{n+1}) + \delta u_{n+1} - u(t_n) - \delta u_n}{\tau} = F(t_n, u(t_n) + \delta u_n)$$

или

$$\frac{\delta u_{n+1} - \delta u_n}{\tau} = F(t_n, u(t_n) + \delta u_n) - \frac{u(t_{n+1}) - u(t_n)}{\tau}. \quad (113)$$

Оценим правую часть этого уравнения. Для этого представим  $u(t_{n+1})$  в виде ряда Тейлора до членов второго порядка по переменной  $\tau$ :

$$u(t_{n+1}) = u(t_n) + \tau \cdot \frac{du}{dt}(t_n) + o(\tau^2);$$

$$u(t_{n+1}) = u(t_n) + \tau \cdot F(t_n, u(t_n)) + o(\tau^2).$$

Здесь  $o(\tau^2)$  – бесконечно малая величина порядка малости  $\tau^2$ .

После деления на  $\tau$ , учитывая, что  $\frac{o(\tau^2)}{\tau} = o(\tau)$ , получим:

$$\frac{u(t_{n+1}) - u(t_n)}{\tau} = F(t_n, u(t_n)) + o(\tau). \quad (114)$$

Анализируя уравнение (114), видим, что уравнение (111) отличается от него величиной  $o(\tau)$ , т.е. если точные значения решения задачи (110)  $u(t_n)$  подставить в уравнение (111), то получим ошибку в виде величины  $o(\tau)$ . Следовательно, точное решение удовлетворяет разностному уравнению с точностью до  $o(\tau)$ .

Оценим  $\delta u_n$ . Подставив правую часть уравнения (114) в уравнение (113), получим:

$$\frac{\delta u_{n+1} - \delta u_n}{\tau} = F(t_n, u(t_n) + \delta u_n) - F(t_n, u(t_n)) + o(\tau). \quad (115)$$

Из условия Липшица

$$|F(t_n, u(t_n) + \delta u_n) - F(t_n, u(t_n))| \leq k \cdot |\delta u_n|$$

следует, что

$$|\delta u_{n+1}| < (1 + k \cdot \tau) \cdot |\delta u_n| + o(\tau^2). \quad (116)$$

Оценим накопление ошибки  $\delta u_N$  за  $N$  шагов ( $\tau \cdot N = T$ ):

$$|\delta u_N| \leq (1 + k \cdot \tau) \cdot |\delta u_{N-1}| + o(\tau^2);$$

$$|\delta u_N| \leq (1 + k \cdot \tau) \cdot [(1 + k \cdot \tau) \cdot |\delta u_{N-2}| + o(\tau^2)] + o(\tau^2) \leq \dots \leq$$

$$\leq o(\tau^2) + (1 + k \cdot \tau) \cdot o(\tau^2) + (1 + k \cdot \tau)^2 \cdot o(\tau^2) + \dots +$$

$$+ (1 + k \cdot \tau)^{N-1} \cdot o(\tau^2) + (1 + k \cdot \tau)^N \cdot |\delta u_0|.$$

Заметим, что в правой части этого неравенства все слагаемые до последнего члена представляют собой геометрическую прогрессию со знаменателем  $(1 + k \cdot \tau)$  и, значит,

$$|\delta u_N| \leq \frac{(1 + k \cdot \tau)^N - 1}{(1 + k \cdot \tau) - 1} \cdot o(\tau^2) + (1 + k \cdot \tau)^N \cdot |\delta u_0|, \quad N = \frac{T}{\tau}.$$

Рассмотрим  $(1 + k \cdot \tau)^N = (1 + k \cdot \tau)^{\frac{T}{\tau}} = \left[ (1 + k \cdot \tau)^{\frac{1}{\tau \cdot k}} \right]^{k \cdot T} \approx e^{k \cdot T}$ . Следовательно-

$$\text{но, } |\delta u_N| \leq \frac{e^{k \cdot T} - 1}{k \cdot \tau} \cdot o(\tau^2) + e^{k \cdot T} \cdot |\delta u_0| \quad \text{или} \quad |\delta u_N| \leq e^{k \cdot T} \cdot \frac{o(\tau)}{k} + e^{k \cdot T} \cdot |\delta u_0| - \frac{o(\tau)}{k}.$$

Так как  $\delta u(T) = \delta u_N$ ,  $\delta u_0 = u_0 - u(0) = 0$ , получаем:  $|\delta u(T)| \leq e^{k \cdot T} \cdot o(\tau)$ , то есть  $\delta u(T) \rightarrow 0$  при  $\tau \rightarrow 0$ .

Таким образом, при правильном выборе параметра  $\tau$  таблица значений  $t_n, u_n$ , полученная из решения дифференциального уравнения по методу Эйлера, как угодно точно аппроксимирует решение исходной задачи (110). Это означает, что установлена сходимость приближенного решения к точному при  $\tau \rightarrow 0$ . Однако приведенное доказательство не дает оценки значения  $\tau$ , при котором решение получается с заданной точностью.

При практической реализации численных методов решения дифференциальных уравнений обычно поступают следующим образом.

Решают задачу при нескольких значениях  $\tau$  и сравнивают результаты. Когда решение мало меняется при изменении  $\tau$ , считают, что большей точности достичь нельзя.

Порядок точности метода Эйлера минимальный:  $u_n - u(t_n) = o(\tau)$ . Это вызвано грубым способом аппроксимации дифференциального уравнения его разностным аналогом.

Качество аппроксимации – это точность, с которой решение исходной задачи удовлетворяет разностному уравнению. В случае метода Эйлера аппроксимация имеет порядок  $o(\tau)$ .

### 11.2.2. Оценка порядка аппроксимации дифференциального уравнения другими разностными уравнениями

Рассмотрим разностную схему вида

$$\frac{u_{n+1} - u_n}{\tau} = \frac{1}{2} (F(t_n, u_n) + F(t_{n+1}, u_{n+1})). \quad (117)$$

Оценим точность аппроксимации уравнения (110) схемой (117). Подставив величины  $u(t_{n+1})$  и  $u(t_n)$  в уравнение (114), получим:

$$\frac{u(t_{n+1}) - u(t_n)}{\tau} = \frac{1}{2} (F(t_n, u(t_n)) + F(t_{n+1}, u(t_{n+1}))).$$

Используем разложение функций  $u(t_{n+1})$  и  $F(t_{n+1}, u(t_{n+1}))$  в точке  $t_n$  в ряд Тейлора:

$$u(t_{n+1}) = u(t_n) + \tau \cdot \left( \frac{du}{dt} \right)_{t_n} + \frac{\tau^2}{2} \cdot \left( \frac{d^2u}{dt^2} \right)_{t_n} + o(\tau^3);$$

$$F(t_{n+1}, u(t_{n+1})) = F(t_n, u(t_n)) + \tau \cdot \left( \frac{dF}{dt} \right)_{t_n} + o(\tau^2).$$

Получим:

$$\begin{aligned} & \frac{1}{\tau} \left[ \tau \cdot \left( \left( \frac{du}{dt} \right)_{t_n} + \frac{\tau^2}{2} \cdot \left( \frac{d^2u}{dt^2} \right)_{t_n} \right) + o(\tau^3) \right] = \\ & = \frac{1}{2} F(t_n, u(t_n)) + \frac{1}{2} F(t_n, u(t_n)) + \frac{\tau}{2} \cdot \left( \frac{dF}{dt} \right)_{t_n} + o(\tau^2) \Rightarrow \\ & \Rightarrow \left( \frac{du}{dt} \right)_{t_n} = F(t_n, u(t_n)) + o(\tau^2), \end{aligned}$$

так как  $\frac{d^2u}{dt^2} = \frac{dF}{dt}$ . Следовательно, точное решение удовлетворяет разностному уравнению с точностью до  $o(\tau^2)$ .

Метод решения уравнения (110), использующий схему (117), не позволяет явно выразить  $u_{n+1}$  через  $u_n$ . Следовательно, для решения системы уравнений (117) необходимо применять итерационный процесс, эффективность которого основывается на хорошем начальном приближении к решению.

На практике такой же порядок точности, как и в методе (117), можно получить двумя итерациями следующего вида:

- 1) по формуле Эйлера вычислить  $\tilde{u}_{n+1} = u_n + \tau \cdot F(t_n, u_n)$ ;
- 2) подставить  $\tilde{u}_{n+1}$  в правую часть схемы (117):

$$u_{n+1} = u_n + \frac{\tau}{2} (F(t_n, u_n) + F(t_{n+1}, \tilde{u}_{n+1})).$$

Следовательно,

$$\frac{u_{n+1} - u_n}{\tau} = \frac{1}{2} (F(t_n, u_n) + F(t_{n+1}, u_n + \tau \cdot F(t_n, u_n))).$$

Такая разностная схема дает явное выражение для  $u_{n+1}$  через  $u_n$ .

**Примечания:**

1. Разность  $(u_{n+1} - \tilde{u}_{n+1})$  позволяет судить о величине шага  $\tau$  и служит основанием для автоматического выбора шага.

2. Обобщение такого способа построения разностных уравнений позволяет получить разностные методы более высоких порядков точности.

Если дополнительно ввести в рассмотрение несколько различных промежуточных значений  $\tilde{u}$ , вычисляемых последовательно, получим методы типа *Рунге – Кутты*.

### 11.3. Схемы Рунге – Кутта

1) Схема второго порядка.

На первом этапе находим промежуточное значение

$$\bar{u}_n = u_n + \alpha\tau \cdot F(t_n, u_n).$$

На втором этапе определяем значение  $u_{n+1}$  по формуле

$$u_{n+1} = u_n + \tau(1 - \sigma) \cdot F(t_n, u_n) + \sigma\tau \cdot F(t_n + \alpha\tau, \bar{u}_n).$$

Такая схема имеет порядок аппроксимации  $o(\tau^2)$  при  $\sigma \cdot \alpha = \frac{1}{2}$ .

В частном случае, при  $\sigma = \frac{1}{2}$ ,  $\alpha = 1$ , получаем схему, называемую

«предиктор – корректор» (счет – пересчет):

$$\frac{u_{n+1} - u_n}{\tau} = \frac{1}{2} [F(t_n, u_n) + F(t_{n+1}, u_n + \tau \cdot F(t_n, u_n))].$$

Заметим, что такая схема уже была рассмотрена (см. подразд. 11.2.2): сначала схема Эйлера с шагом  $\tau$  (предиктор), затем схема с полусуммой (корректор).

Идея метода «предиктор – корректор» часто используется при написании разностных схем для уравнений математической физики с частными производными.

2) Схема четвертого порядка.

Она имеет вид

$$\frac{u_{n+1} - u_n}{\tau} = \frac{1}{6} [k_1(u_n) + 2k_2(u_n) + 2k_3(u_n) + k_4(u_n)], \quad n = 0, 1, \dots; \quad u(0) = u_0,$$

где  $k_1 = F(t_n, u_n)$ ;  $k_2 = F\left(t_n + \frac{\tau}{2}, u_n + \frac{\tau k_1}{2}\right)$ ;

$$k_3 = F\left(t_n + \frac{\tau}{2}, u_n + \frac{\tau k_2}{2}\right); \quad k_4 = F(t_n + \tau, u_n + \tau k_3).$$

При определении  $u_{n+1}$  по заданному  $u_n$  необходимо четыре раза вычислить функцию  $F(t, u)$ . Вычисления производятся последовательно. Для невязки (точности аппроксимации) получим выражение

$$\psi_n = \frac{1}{6} [k_1(u(t_n)) + 2k_2(u(t_n)) + 2k_3(u(t_n)) + k_4(u(t_n))] - \frac{u(t_{n+1}) - u(t_n)}{\tau}.$$

С помощью разложения в ряд Тейлора в окрестности точки  $t_n$  можно получить оценку  $\psi_n \approx o(\tau^4)$ , то есть схема имеет четвертый порядок аппроксимации.

Предположим, что уже вычислены  $n$  значений  $u_0, u_1, \dots, u_n$ . Тогда можно построить интерполяционный полином  $P_n(t)$  по табличной функции

$(u_i, F(t_i, u_i))$ , где  $i = n, n-1, \dots, n-k+1$ , и при вычислении  $u_{n+1}$  использовать этот полином  $P_n(t)$  вместо функции  $F(t)$  на интервале  $[t_n, t_{n+1}]$  (экстраполяция):

$$u_{n+1} = u_n + \int_{t_n}^{t_{n+1}} P_n(t) dt.$$

Получим принцип построения методов типа *Адамса*. Неудобство метода заключается в том, что для его применения необходимо иметь  $n$  точек  $u_1, u_2, \dots, u_n$ , построенных другим методом.

Если  $F(t, u)$  имеет особенность в некоторой точке, необходимо в окрестности этой точки аппроксимировать  $F(t, u)$  гладкой функцией и использовать эту аппроксимацию для прохождения особой точки.

На ошибку аппроксимации накладывается ошибка округления. Неточность, возникающая при отдельном шаге из-за округлений, не должна превосходить неточность аппроксимации. В этом случае вместе с уменьшением шага интегрирования  $\tau$  необходимо пропорционально увеличивать точность вычислений.

#### 11.4. Разностные методы решения ОДУ второго порядка

Для получения разностного уравнения на примере линейного дифференциального уравнения второго порядка необходимо:

- заменить область непрерывного изменения аргументов дискретным множеством точек (сеткой);
- аппроксимировать на сетке дифференциальное уравнение разностным уравнением.

Рассмотрим линейное дифференциальное уравнение

$$u'' + p(x)u' - q(x)u = f(x), \quad (118)$$

где функции  $p(x)$ ,  $q(x)$ ,  $f(x)$  непрерывны.

Требуется найти решение этого уравнения, удовлетворяющее краевым условиям:

$$u(0) = \gamma_0, \quad u(1) = \gamma_1, \quad \text{где } \gamma_0, \gamma_1 \text{ – заданные числа;}$$

$$q(x) \geq 0, \quad x \in [0, 1].$$

Это частный случай краевой задачи.

Дифференциальное уравнение (118) во внутренних точках  $x = x_i$  ( $i = 1, 2, \dots, N-1$ ) интервала  $[0, 1]$  приближенно можно заменить системой линейных уравнений

$$\frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + p_i \frac{u_{i+1} - u_{i-1}}{2h} - q_i u_i = f_i, \quad (119)$$

где  $i = 1, 2, \dots, N-1$ ,  $u_0 = \gamma_0$ ,  $u_N = \gamma_1$ ,  $p_i = p(x_i)$ ,  $q_i = q(x_i)$ ,  $f_i = f(x_i)$ .

Если принять

$$b_i = \frac{1}{h^2} - \frac{p_i}{2h}, \quad c_i = \frac{2}{h^2} + q_i, \quad a_i = \frac{1}{h^2} + \frac{p_i}{2h},$$

$$f_i = f_i, \quad \chi_1 = \chi_2 = 0, \quad \mu_1 = \gamma_0, \quad \mu_2 = \gamma_1,$$

то получим краевую задачу для разностного уравнения 2-го порядка с переменными коэффициентами, аналогичную задаче (96):

$$a_i u_{i-1} - c_i u_i + b_i u_{i+1} = f_i,$$

$$u_0 = \chi_1 u_1 + \mu_1, \quad u_N = \chi_2 u_{N-1} + \mu_2.$$

Эту задачу удобно решать методом прогонки (см. подразд. 10.1.1).

Отметим, что порядок аппроксимации краевой задачи разностной схемой равен  $o(h^2)$  и разностная схема устойчива.

#### Пример 42

Решить уравнение  $y'' - 2xy' - 2y = -4x$  при краевых условиях:  $y(0) - y'(0) = 0$ ;  $y(1) = 1 + e = 3,718$ .

Решение.

Решение получено в ячейках *MS Excel* (рис. 33).

Выбрав шаг  $h = 0,1$ , заменим исходное уравнение и краевые условия системой конечно-разностных уравнений согласно формуле (119).

$$\begin{cases} \frac{y_{i+2} - 2y_{i+1} + y_i}{0,01} - 2x_i \frac{y_{i+1} - y_i}{0,1} - 2y_i = -4x_i, & i = 1, 2, \dots, 8; \\ y_0 - \frac{y_1 - y_0}{0,1} = 0; \\ y_{10} = 3,718. \end{cases}$$

После приведения подобных членов получим

$$y_{i+2} + (-2 - 0,2x_i)y_{i+1} + (0,98 + 0,2x_i)y_i = -0,01 \cdot 4x_i.$$

Заносим в ячейки **A2–G2** исходные данные:  $a_0 = 1$ ;  $a_1 = -1$ ;  $A = 0$  – коэффициенты начального условия;  $b_0 = 1$ ;  $b_1 = 0$ ;  $B = 3,718$  – коэффициенты конечного условия и шаг  $h = 0,1$ .

Обозначения:

$$y_{i+2} + m_i y_{i+1} + k_i y_i = h^2 f_i;$$

$$a_0 y_0 + a_1 \frac{y_1 - y_0}{h} = A; \quad b_0 y_n + b_1 \frac{y_n - y_{n-1}}{h} = B.$$

Заполняем расчетную таблицу.

В ячейки **B5–B15** заносим  $x_i$  (от 0 до 1 с шагом 0,1), в ячейки **A5–A15** – соответствующие значения индексов  $i$ .



		fx = $\frac{(\$B\$2*\$H6-\$C\$2*\$G\$2)}{(\$B\$2-\$A\$2*\$G\$2)}$								
	A	B	C	D	E	F	G	H	I	
1	a0	a1	A	b0	b1	B	h			
2	1	-1	0	1	0	3,718	0,1			
3						Прямой ход		Обратный ход		
4	i	x <sub>i</sub>	m <sub>i</sub>	k <sub>i</sub>	f <sub>i</sub>	c <sub>i</sub>	d <sub>i</sub>	y <sub>i</sub>	y(x <sub>i</sub> )	
5	0	0,0	-2,00	0,98	0,0	-0,9016	0,0000	1,118	1,000	
6	1	0,1	-2,02	1,00	-0,4	-0,8942	-0,0040	1,230	1,110	
7	2	0,2	-2,04	1,02	-0,8	-0,8866	-0,0116	1,364	1,241	
8	3	0,3	-2,06	1,04	-1,2	-0,8788	-0,0227	1,521	1,394	
9	4	0,4	-2,08	1,06	-1,6	-0,8707	-0,0372	1,704	1,574	
10	5	0,5	-2,10	1,08	-2,0	-0,8623	-0,0550	1,917	1,784	
11	6	0,6	-2,12	1,10	-2,4	-0,8536	-0,0761	2,164	2,033	
12	7	0,7	-2,14	1,12	-2,8	-0,8447	-0,1008	2,455	2,332	
13	8	0,8	-2,16	1,14	-3,2	-0,8354	-0,1291	2,800	2,696	
14	9	0,9						3,214	3,148	
15	10	1,0						3,718	3,718	

		fx = $\frac{(\$B\$2*\$H6-\$C\$2*\$G\$2)}{(\$B\$2-\$A\$2*\$G\$2)}$								
	A	B	C	D	E	F	G	H	I	
1	a0	a1	A	b0	b1	B	h			
2	1	-1	0	1	0	3,718	0,1			
3						Прямой ход		Обратный ход		
4	i	x	m <sub>i</sub>	k <sub>i</sub>	f <sub>i</sub>	c <sub>i</sub>	d <sub>i</sub>	y <sub>i</sub>	y(x <sub>i</sub> )	
5	0	0	=-2,02*\$B5	=0,98+0,2*\$B5	=-4*\$B5	=(\$B\$2-\$A\$2*\$G\$2)/(C5*\$B\$2-\$A\$2*\$G\$2)+D5*\$B\$2	=D5*\$G\$2*(G\$2/(\$B\$2-\$A\$2*\$G\$2))+E5*\$G\$2	=(\$H\$2-\$C\$2*\$G\$2)/(\$B\$2-\$A\$2*\$G\$2)	=B5+EXP(B5^2)	
6	1	0,1	=-2,02*\$B6	=0,98+0,2*\$B6	=-4*\$B6	=1/(C6-D6*\$F6)	=E6*\$G\$2^2-D6*\$F6*\$G6	=F5*(G5-H7)	=B6+EXP(B6^2)	
7	2	0,2	=-2,02*\$B7	=0,98+0,2*\$B7	=-4*\$B7	=1/(C7-D7*\$F6)	=E7*\$G\$2^2-D7*\$F6*\$G6	=F6*(G6-H8)	=B7+EXP(B7^2)	
8	3	0,3	=-2,02*\$B8	=0,98+0,2*\$B8	=-4*\$B8	=1/(C8-D8*\$F7)	=E8*\$G\$2^2-D8*\$F7*\$G7	=F7*(G7-H9)	=B8+EXP(B8^2)	
9	4	0,4	=-2,02*\$B9	=0,98+0,2*\$B9	=-4*\$B9	=1/(C9-D9*\$F8)	=E9*\$G\$2^2-D9*\$F8*\$G8	=F8*(G8-H10)	=B9+EXP(B9^2)	
10	5	0,5	=-2,02*\$B10	=0,98+0,2*\$B10	=-4*\$B10	=1/(C10-D10*\$F9)	=E10*\$G\$2^2-D10*\$F9*\$G9	=F9*(G9-H11)	=B10+EXP(B10^2)	
11	6	0,6	=-2,02*\$B11	=0,98+0,2*\$B11	=-4*\$B11	=1/(C11-D11*\$F10)	=E11*\$G\$2^2-D11*\$F10*\$G10	=F10*(G10-H12)	=B11+EXP(B11^2)	
12	7	0,7	=-2,02*\$B12	=0,98+0,2*\$B12	=-4*\$B12	=1/(C12-D12*\$F11)	=E12*\$G\$2^2-D12*\$F11*\$G11	=F11*(G11-H13)	=B12+EXP(B12^2)	
13	8	0,8	=-2,02*\$B13	=0,98+0,2*\$B13	=-4*\$B13	=1/(C13-D13*\$F12)	=E13*\$G\$2^2-D13*\$F12*\$G12	=F12*(G12-H14)	=B13+EXP(B13^2)	
14	9	0,9						=F13*(G13-H15)	=B14+EXP(B14^2)	
15	10	1						=F13*(G13-H15)	=B15+EXP(B15^2)	

Рис. 33. Пример решения дифуравнения методом прогонки

### Прямой ход

В принятых обозначениях  $m_i = -2 - 0,2x_i$ ;  $k_i = 0,98 + 0,2x_i$ ;  $f_i = 4x_i$ .  
 Заполним ячейки **C5–C13** значениями  $m_i$ , ячейки **D5–D13** значениями  $k_i$ ,  
 ячейки **E5–E13** значениями  $f_i$ .

В ячейки **F5** и **G5** введем соответственно формулы:

$$c_0 = \frac{a_1 - a_0 h}{m_0(a_1 - a_0 h) + k_0 a_1}; \quad d_0 = \frac{k_0 A h}{a_1 - a_0 h} + f_0 h^2.$$

Продолжим заполнение столбцов значениями, вычисленными по рекуррентным формулам:

$$c_i = \frac{1}{m_i - k_i c_{i-1}}; \quad d_i = f_i h^2 - k_i c_{i-1} d_{i-1}.$$

### Обратный ход

В ячейку **H15** вводится формула  $y_n = \frac{\beta_1 c_{n-2} d_{n-2} + B h}{\beta_1 (1 + c_{n-2}) + \beta_0 h}$ .

Ячейки **H14–H6** заполняем, применяя рекуррентные формулы

$$y_{n-1} = c_{n-2}(d_{n-2} - y_n).$$

Ячейка **H5** вычисляется по формуле  $y_0 = \frac{a_1 y_1 - Ah}{a_1 - a_0 h}$ .

Для сравнения с точным значением последний столбец вычисляется по формуле  $y = x + e^{x^2}$ .

### Контрольные вопросы и задания

1. В каком узле определяется производная при численном дифференцировании в случае использования трех и пяти узлов?

2. Как численно определить значение производной в произвольной точке?

3. Какие погрешности возникают при численном дифференцировании таблично заданной функции?

4. Получите общую формулу приближенного вычисления второй производной функции. Примените её для вычисления  $f''(x)$  в точке  $x = 4$ , если  $f(x) = 2x^2$ . Сравните с точным значением.

5. В каких случаях результаты численного дифференцирования по формулам при 3-х и 5 узлах, окружающих точку дифференцирования, совпадают? Покажите, что в этом случае совпадают формулы дифференцирования.

6. Что является решением дифференциального уравнения?

7. Какие методы можно применять для решения системы алгебраических уравнений, полученных при аппроксимации дифференциального уравнения второго порядка разностной схемой?

## 12. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

### 12.1. Уравнения математической физики

Математические постановки задач с дифференциальными уравнениями в частных производных содержат дифференциальные уравнения и дополнительные условия, позволяющие выделить искомые частные решения среди целого семейства решений.

Необходимость решения таких уравнений возникает, когда исследуется объект или процесс, свойства которого описываются функциями нескольких переменных (например, времени и расстояния).

Дифференциальные уравнения в частных производных называются *уравнениями математической физики*.

В общем виде нелинейное уравнение в частных производных может быть записано в виде

$$F(x, t, u, u'_x, u'_t, u''_{xx}, u''_{tt}, u''_{tx}, \dots, u^{(n)}_{x\dots x}, u^{(n)}_{t\dots t}, u^{(n)}_{t\dots x}) = 0,$$

где  $t, x$  – независимые переменные;  $u$  – искомая функция;  $u'_x = \frac{\partial u(x,t)}{\partial x}$ ,

$$u'_t = \frac{\partial u(x,t)}{\partial t}, \quad u''_{xx} = \frac{\partial^2 u(x,t)}{\partial x^2}, \quad u''_{tt} = \frac{\partial^2 u(x,t)}{\partial t^2}, \quad u''_{tx} = \frac{\partial^2 u(x,t)}{\partial t \partial x}, \quad \dots,$$

$$u^{(n)}_{x\dots x} = \frac{\partial^n u(x,t)}{\partial x^n}, \quad u^{(n)}_{t\dots t} = \frac{\partial^n u(x,t)}{\partial t^n}, \quad u^{(n)}_{t\dots x} = \frac{\partial^n u(x,t)}{\partial t^k \partial x^l}, \quad n = k + l, \quad - \text{частные}$$

производные первого, второго и  $n$ -го порядков соответственно по переменным  $t$  и  $x$ .

*Решением уравнения* называется функция  $u = u(t, x)$ , имеющая частные производные до требуемого порядка и обращающая это уравнение в тождество. График решения представляет собой поверхность в пространстве  $Otxi$  (интегральная поверхность).

*Порядком уравнения* называется наивысший порядок частных производных, входящих в уравнение.

Уравнение называется *линейным*, если функция  $F$  линейна относительно искомой функции  $u$  и ее производных.

Если коэффициенты уравнения не зависят от  $x$  и  $t$ , то уравнение называется *уравнением с постоянными коэффициентами*, в противном случае – *уравнением с переменными коэффициентами*.

Уравнение называется *однородным*, если правая часть равна нулю для всех  $x$  и  $t$ , в противном случае – *неоднородным*.

Рассмотрим вопросы сходимости, аппроксимации и устойчивости на примере уравнения в частных производных первого порядка

$$\frac{\partial U}{\partial t} + a \frac{\partial U}{\partial x} = F(x, t),$$

$$-\infty < x < \infty, \quad t \geq 0, \quad U(x, 0) = \Phi(x).$$

Известно аналитическое решение этого уравнения:

$$U(x, t) = \Phi(x - at) + \int_0^t F(x - at + at', t') dt'.$$

Построим численный метод решения задачи сведением ее к системе конечно-разностных уравнений. Для этого введем сетку

$$\omega_x = \{x_k = kh, k = 0, \pm 1, \dots; h - \text{шаг по координате } x\},$$

$$\omega_t = \{t_n = n\tau, n = 0, 1, \dots; \tau - \text{шаг по координате } t\}.$$

Вместо функций  $U(x, t)$ ,  $F(x, t)$ ,  $\Phi(x)$  будем рассматривать сеточные функции, соответственно  $u_{k,n}$ ,  $f_{k,n}$ ,  $\varphi_k$ . Заменив производные в исходном уравнении их разностными аналогами  $F(x_k, t_k) = f_{k,n}$ ,  $\varphi_k = \Phi(x_k)$ , получим:

$$\frac{u_{k,n+1} - u_{k,n}}{\tau} + a \frac{u_{k+1,n} - u_{k,n}}{h} = f_{k,n}, \quad u_{k,0} = \varphi_k. \quad (120)$$

Заметим, что такой способ аппроксимации непрерывного уравнения конечно-разностным аналогом не является единственным.

Вычислительный алгоритм заключается в последовательном нахождении решения на каждом *слое* по переменной  $t$  (слоем называются узлы, лежащие на одной прямой  $t = t_k$  при фиксированном  $k = 0, 1, \dots, m$ ).

$$u_{k,n+1} = \left(1 + a \frac{\tau}{h}\right) u_{k,n} - a \frac{\tau}{h} u_{k+1,n} + \tau f_{k,n}.$$

### 12.1.1. Аппроксимация

При решении задачи разностным методом надо знать, с какой точностью решение разностной задачи приближает решение исходной задачи

$$u_{k,n} = U(x_k, t_n) + \delta u_{k,n} = U_{k,n} + \delta u_{k,n}.$$

Подставив выражения для  $u_k^n$  в уравнение (120), получим:

$$\frac{\delta u_{k,n+1} - \delta u_{k,n}}{\tau} + a \frac{\delta u_{k+1,n} - \delta u_{k,n}}{h} = f_{k,n} - \left( \frac{u_{k,n+1} - u_{k,n}}{\tau} + a \frac{u_{k+1,n} - u_{k,n}}{h} \right). \quad (121)$$

Разложим в ряд Тейлора функцию  $U(x, t)$  в точке  $(x_k, t_n)$  по переменным  $x$  и  $t$  до величин второго порядка малости:

$$u_{k,n+1} = u_{k,n} + \tau \left( \frac{\partial U}{\partial t} \right)_{k,n} + o(\tau^2);$$

$$u_{k+1,n} = u_{k,n} + h \left( \frac{\partial U}{\partial x} \right)_{k,n} + o(h^2).$$

Отсюда следует, что

$$\begin{aligned} & \frac{u_{k,n+1} - u_{k,n}}{\tau} + a \frac{u_{k+1,n} - u_{k,n}}{h} + f_{k,n} = \\ & = \left( \frac{\partial U}{\partial t} + a \frac{\partial U}{\partial x} \right)_{k,n} + F(x_k, t_n) + o(\tau, h). \end{aligned} \quad (122)$$

Так как  $F(x_k, t_n) = \left( \frac{\partial U}{\partial t} + a \frac{\partial U}{\partial x} \right)_{k,n} = f_{k,n}$ , получаем точность аппроксимации исходного дифференциального уравнения конечно-разностным аналогом, равную  $o(\tau, h)$ .

### 12.1.2. Сходимость

Подставив выражение (122) в равенство (121), после преобразований получим:

$$\delta u_{k,n+1} = \left( 1 + a \frac{\tau}{h} \right) \delta u_{k,n} - a \frac{\tau}{h} \delta u_{k+1,n} + \tau o(\tau, h). \quad (123)$$

Выберем  $\tau$  и  $h$ ,  $a \leq \delta$ ,  $0 \leq -a \frac{\tau}{h} \leq 1$ . Тогда

$$\begin{aligned} |\delta u_{k,n+1}| & \leq \left( 1 + a \frac{\tau}{h} \right) |\delta u_{k,n}| + \left( -a \frac{\tau}{h} \right) |\delta u_{k+1,n}| + \tau o(\tau, h) \leq \\ & \leq \max_k (|\delta u_{k,n}|) + \tau o(\tau, h). \end{aligned}$$

Обозначим  $\|\delta u_n\| = \max_k |\delta u_{k,n}|$ . Отсюда  $\|\delta u_{n+1}\| \leq \|\delta u_n\| + \tau o(\tau, h)$ .

Следовательно, за  $N$  шагов по  $\tau$  получим:

$$\|\delta u_N\| \leq \|\delta u_0\| + N \tau o(\tau, h).$$

Значит, если  $t = N\tau = \text{const}$  ( $\tau, h \rightarrow 0$ ,  $N \rightarrow \infty$ ), так как  $\delta u_0 = \Phi_k - \phi_k = 0$ , то  $\|\delta u_N\| \approx o(\tau, h)$ . Следовательно, ошибка аппроксимации стремится к 0 при  $\tau, h \rightarrow 0$ , т.е. разностная схема является сходящейся.

### 12.1.3. Устойчивость

Устойчивость разностной схемы рассмотрим на модели ошибки  $\delta u_{k,n} = \varepsilon(-1)^k$ .

Подставив ее в выражение (123), получим:

$$\delta u_{k,n+1} = \left( 1 + a \frac{\tau}{h} \right) \varepsilon(-1)^k - a \frac{\tau}{h} \varepsilon(-1)^{k+1} = \left( 1 + 2a \frac{\tau}{h} \right) \varepsilon(-1)^k,$$

т.е. ошибка накапливается с множителем  $\left(1 + 2a \frac{\tau}{h}\right)$ . За  $N$  слоев она будет иметь значение

$$\delta u_{k,n+N} = \left(1 + 2a \frac{\tau}{h}\right)^N \varepsilon (-1)^k.$$

Следовательно, при  $\left(1 + 2a \frac{\tau}{h}\right) < 1$  ошибка уменьшается с ростом  $N$ , в противном случае она экспоненциально возрастает.

### Контрольные вопросы и задания

1. Какие уравнения называют уравнениями математической физики?
2. Что такое порядок дифференциального уравнения в частных производных?
3. Какое дифференциальное уравнение в частных производных называется линейным, однородным?
4. Какой порядок имеет точность аппроксимации дифференциального уравнения конечно-разностным аналогом?
5. Единственным ли способом можно аппроксимировать непрерывное дифференциальное уравнение в частных производных конечно-разностным аналогом?
6. Что называют слоем по переменной  $t$ ?
7. В каких случаях ошибка аппроксимации дифференциального уравнения конечно-разностным аналогом возрастает, а в каких – уменьшается?

## 13. УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ ВТОРОГО ПОРЯДКА

### 13.1. Классификация уравнений с частными производными

Рассмотрим принципы классификации линейных дифференциальных уравнений с двумя независимыми переменными на примере линейного дифференциального уравнения второго порядка. Запишем его в канонической форме:

$$A(x,t)\frac{\partial^2 u(x,t)}{\partial x^2} + 2B(x,t)\frac{\partial^2 u(x,t)}{\partial t \partial x} + C(x,t)\frac{\partial^2 u(x,t)}{\partial t^2} + D(x,t)\frac{\partial u(x,t)}{\partial x} + E(x,t)\frac{\partial u(x,t)}{\partial t} + F(x,t)u(x,t) = G(x,t),$$

где  $u(x,t)$  – искомая функция;  $A(x,t), B(x,t), C(x,t), D(x,t), E(x,t), F(x,t)$  – коэффициенты;  $G(x,t)$  – свободный член (правая часть).

Предполагается, что коэффициенты и правая часть являются заданными дважды дифференцируемыми функциями, причем  $|A(x,t)| + |B(x,t)| + |C(x,t)| \neq 0$ .

Иногда используют упрощенную запись этого дифференциального уравнения:

$$Au''_{xx} + 2Bu''_{tx} + Cu'' + Du'_x + Eu' + Fu = G.$$

В зависимости от знака выражения  $AC - B^2$  (дискриминанта), уравнение в данной области относится к одному из следующих типов:

- 1)  $AC - B^2 > 0$  – эллиптический тип;
- 2)  $AC - B^2 = 0$  – параболический тип;
- 3)  $AC - B^2 < 0$  – гиперболический тип;
- 4)  $AC - B^2$  не сохраняет постоянного знака – смешанный тип.

Если искомая функция зависит от времени, уравнение называется *нестационарным*, если не зависит от времени, то *стационарным*.

Во многих уравнениях присутствует сумма частных производных второго порядка, называемая *оператором Лапласа*:

$$\Delta u = \frac{\partial^2 u(x,t)}{\partial x^2} + \frac{\partial^2 u(x,t)}{\partial y^2} + \frac{\partial^2 u(x,t)}{\partial z^2} \quad \begin{array}{l} \text{– для трехмерной функции} \\ u = u(x, y, z), \text{ где } x, y, z \text{ – прост-} \\ \text{ранственные переменные;} \end{array}$$

$$\Delta u = \frac{\partial^2 u(x,t)}{\partial x^2} + \frac{\partial^2 u(x,t)}{\partial y^2} \quad \text{– для двухмерной функции } u = u(x, y).$$

Уравнения параболического типа описывают, например, процессы теплопроводности и диффузии, гиперболического – колебательные системы и волновые движения, эллиптического – течение жидкости в стационарных

потоках, стационарное распределение напряженности электрического и магнитного полей.

Тип уравнения определяет формулировку задачи и численные методы его решения.

В математической физике различают три основных типа задач:

1. *Задача Коши* (с начальными условиями).
2. *Краевые (граничные) задачи* с краевыми (граничными) условиями.
3. *Смешанные задачи* (начально-краевые) с начальными и краевыми условиями.

### 13.2. Уравнение параболического типа (уравнение теплопроводности)

Рассмотрим уравнение теплопроводности (Фурье) вида

$$\frac{\partial U}{\partial t} = a \frac{\partial^2 U}{\partial x^2},$$

$$U(0, x) = U_0(x), \quad -\infty < x < \infty, \quad 0 \leq t \leq T.$$

Конечно-разностную аппроксимацию этой задачи представим следующей разностной схемой:

$$\frac{u_{k,n+1} - u_{k,n}}{\tau} = a \frac{u_{k+1,n} - 2u_{k,n} + u_{k-1,n}}{h^2}, \quad (124)$$

$$u_{k,0} = U_0(x_k), \quad k = 0, \pm 1, \dots, \quad n = 0, 1, \dots, N.$$

Для определения области устойчивости разностной схемы используем так называемый спектральный признак, который соответствует проверке устойчивости на частных решениях вида  $u_{k,n} = \lambda_n e^{ik\varphi}$ ,  $\lambda_n e^{ik\varphi} = \lambda_n (\cos k\varphi + i \cdot \sin k\varphi)$ . Подставив его в выражение (124), получим:

$$\frac{\lambda^{n+1} - \lambda^n}{\tau} = -\mu \lambda^n, \quad \mu = 2a \frac{1 - \cos \varphi}{h^2}.$$

Следовательно,  $\lambda_{n+1} = (1 - \tau\mu)\lambda_n = (1 - \tau\mu)^{n+1}$ .

Для устойчивости решения необходимо, чтобы

$$|1 - \tau\mu| \leq 1, \quad \text{т.е.} \quad \left| 1 - 2a\tau \frac{1 - \cos \varphi}{h^2} \right| \leq 1,$$

$$-1 \leq 1 - \frac{4a\tau}{h^2} \leq 1,$$

откуда следует, что

$$\frac{4a\tau}{h^2} \leq 2 \quad \text{и} \quad \frac{\tau}{h^2} \leq \frac{1}{2a},$$

что и является условием устойчивости.



Разностная схема (124) дает явное выражение для последовательного вычисления  $u_{k,n+1}$  через  $u_{k,n}$ ,  $u_{k-1,n}$ ,  $u_{k+1,n}$  (явная схема), но является условно устойчивой.

Рассмотрим другую схему

$$\frac{u_{k,n+1} - u_{k,n}}{\tau} = a \frac{u_{k+1,n+1} - 2u_{k,n+1} + uu_{k-1,n+1}}{h^2}, \quad (125)$$

$$u_{k,0} = u_0(x_k), \quad k = 0, \pm 1, \dots, \quad n = 0, 1, \dots, N.$$

Величины в правой части схемы (125) берутся с  $(n+1)$ -го слоя. Разностная схема (125) не дает явного выражения для расчета  $u_{k,n+1}$ ; поэтому для нахождения неизвестной функции необходимо решать систему большого числа линейных уравнений. Разностная схема (125) называется *неявной*.

Точки итерационных слоев для данной схемы (расчетная ячейка) можно изобразить в виде схемы (рис. 34).

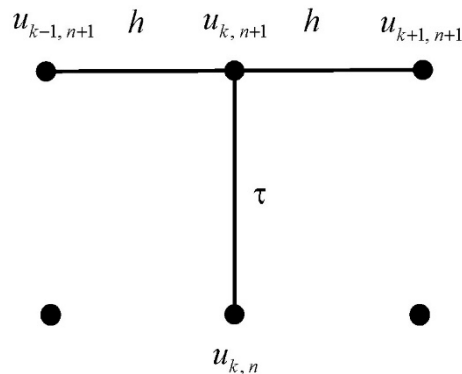


Рис. 34. Расчетная ячейка

Для исследования устойчивости этой схемы на функциях вида  $u_{k,n} = \lambda_n e^{ikp}$  нужно повторить все рассуждения, проведенные для явной схемы. В результате получим доказательство абсолютной устойчивости схемы (125).

### 13.2.1. Решение разностного уравнения для неявной схемы

Схему (125) запишем в виде

$$\frac{u_{k,n+1} - u_{k,n}}{\tau} - a \frac{u_{k+1,n+1} - 2u_{k,n+1} + u_{k-1,n+1}}{h^2} = 0, \quad (126)$$

$$u_{k,0} = u_0(x_k), \quad k = 0, \pm 1, \dots, K-1, \quad n = 0, 1, \dots, N.$$

Рассмотрим задачу (126) с заданными условиями на границах  $x = x_0$ ,  $x = x_K$ :

$$u_{0,n+1} = \alpha_{n+1}, \quad u_{K,n+1} = \beta_{n+1}.$$

Введем обозначение  $r = a \frac{\tau}{h^2}$  и, опустив индекс  $(n+1)$ , запишем схему (126) в виде

$$\begin{cases} u_0 = \alpha, \\ -ru_{k-1} + (1+2r)u_k - ru_{k+1} = u_{k,n}, \quad k=1, 2, \dots, K-1, \\ u_K = \beta. \end{cases} \quad (127)$$

Специфика системы (127) заключается в том, что каждое  $k$ -е уравнение содержит только три неизвестных:  $u_{k-1}$ ,  $u_k$ ,  $u_{k+1}$ .

### 13.2.2. Метод прогонки

Решаем систему (127) методом *прогонки*: поскольку значение  $u_0$  задано, то из уравнения при  $k=1$  находим соотношение между  $u_1$  и  $u_2$ . С помощью этого соотношения, используя следующее уравнение, исключаем  $u_1$  и получаем соотношение между  $u_2$  и  $u_3$  и т.д.

Предположим, что соотношение между  $u_k$  и  $u_{k+1}$  известно:

$$u_{k-1} = L_k u_k + M_k. \quad (128)$$

Тогда из  $k$ -го уравнения получим

$$u_k = \frac{ru_{k+1} + u_{k,n} + rM_k}{1 + 2r - rL_k},$$

а соотношение между  $L_{k+1}$ ,  $M_{k+1}$  и  $L_k$ ,  $M_k$  будет иметь вид

$$L_{k+1} = \frac{r}{1 + 2r - rL_k}, \quad M_{k+1} = \frac{u_{k,n} + rM_k}{1 + 2r - rL_k}. \quad (129)$$

Следовательно, для любого  $k$  можно рассчитать  $L_{k+1}$ ,  $M_{k+1}$  (т. к.  $u_{k,0}$  известно) через  $L_k$ ,  $M_k$ , и, поскольку  $u_0 = \alpha$ , то  $L_1 = 0$ ,  $M_1 = \alpha$ . Формулы (128) позволяют вычислить все  $L_k$ ,  $M_k$ . Далее, так как  $u_K = \beta$ , по формуле (127) находим последовательно все значения  $u_k$ ,  $k=1, 2, \dots, k-1$ .

Основное достоинство метода прогонки заключается в его экономичности. Можно доказать, что при решении системы (127) методом прогонки точность результата совпадает с точностью расчета и точностью задания исходных данных.

Можно доказать также, что разностная схема (127) устойчива при любых соотношениях шагов  $\tau$  и  $h$ .

### 13.2.3. Итерационные методы

Систему (127) перепишем в виде

$$\begin{cases} u_0 = \alpha, \\ u_k = \frac{r(u_{k-1} + u_{k+1}) + u_{k,n}}{1 + 2r}, \quad k = 1, 2, \dots, K - 1, \\ u_K = \beta. \end{cases} \quad (130)$$

Подставив в правую часть уравнения  $m$ -е приближение  $u_k^{(m)}$ , получим в левой части  $u_k^{(m+1)}$ .

Докажем сходимость метода к точному решению.

Пусть  $u_{k,m} = U_k + \delta_{k,m}$ , где  $U_k$  – точное решение системы (130).

Подставив это выражение в систему (130), получим:

$$\begin{cases} \delta_0^{(m+1)} = 0, \\ \delta_k^{(m+1)} = \frac{r}{1 + 2r} (\delta_{k-1}^{(m)} + \delta_{k+1}^{(m)}), \quad k = 1, 2, \dots, K - 1, \\ \delta_K^{(m+1)} = \beta. \end{cases}$$

Оценим  $|\delta_k^{(m+1)}|$ :

$$\max |\delta_k^{(m+1)}| \leq \frac{2r}{1 + 2r} \max |\delta_k^{(m)}|.$$

Следовательно, от итерации к итерации величина  $|\delta_k|$  уменьшается со скоростью геометрической прогрессии со знаменателем  $\frac{2r}{1 + 2r}$ . Значит,  $\delta_k^{(m)} \rightarrow 0$ , и процесс сходится. Скорость сходимости определяется величиной  $\frac{2r}{1 + 2r}$ . Можно показать, что число итераций для получения

достаточно точного решения приблизительно равно  $\sqrt{r}$ , значит, при достаточно небольших значениях  $r$  метод итерации является эффективным, но он медленно сходится при больших  $r$ .

Метод итераций можно применять и в том случае, когда задача нелинейна, т.е.  $a = a(u)$  и, следовательно,  $r = r(u_k)$ .

## 13.3. Уравнения эллиптического типа (уравнения Лапласа и Пуассона)

Исследования стационарных процессов различной физической природы (колебания, теплопроводность и др.), а также задачи, описывающие стационарные состояния систем, полей, часто приводят к уравнениям эллиптического типа.

Двухмерное уравнение эллиптического типа имеет вид

$$\frac{\partial^2 U}{dx^2} + \frac{\partial^2 U}{dy^2} = f(x, y). \quad (131)$$

Эллиптическое уравнение, когда  $f \equiv 0$ , носит название *уравнения Лапласа*. Неоднородное уравнение Лапласа  $f \neq 0$  называется *уравнением Пуассона*.

Рассмотрим контур  $\Gamma$ , ограничивающий расчетную область  $G$ , и функцию  $u_\Gamma = g(x_\Gamma, y_\Gamma)$ , заданную на границе области  $G$  и являющуюся граничным условием (рис. 35).

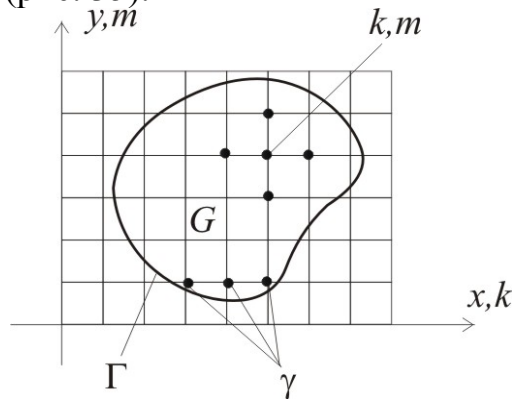


Рис. 35. Расчетная область

### 13.3.1. Построение разностной схемы

- 1) Покрываем область  $G$  расчетной сеткой  $h_x = h_y = h$ .
- 2) Производную заменяем разностными аналогами

$$\frac{u_{k+1,m} - 2u_{k,m} + u_{k-1,m}}{h^2} + \frac{u_{k,m+1} - 2u_{k,m} + u_{k,m-1}}{h^2} = f_{k,m}. \quad (132)$$

где  $f_{k,m} = f(x_k, y_m)$  – сеточная функция.

Это разностное уравнение имеет смысл для любой внутренней точки, т.е. для любой точки  $(k, m)$ , для которой точки  $(k-1, m)$ ,  $(k+1, m)$ ,  $(k, m-1)$ ,  $(k, m+1)$  расположены внутри области  $G$ . Остальные расчетные точки  $(k, m) \in G$  объявим граничными. Их совокупность обозначим символом  $\gamma$ , и значения  $u_{k,m}$  на  $\gamma$  получим простым переносом значения  $g$  из ближайшей точки границы  $\Gamma$ :  $u|_\gamma = g|_\Gamma$ .

### 13.3.2. Аппроксимация

Аппроксимация следует из того, что разностная схема (132) переходит в уравнение (131) при  $h \rightarrow 0$  и  $u|_\gamma \rightarrow g|_\Gamma$ , так как  $\gamma \rightarrow \Gamma$ .

### 13.3.3. Устойчивость

Используя ограниченность области  $G$ , можно доказать, что схема абсолютно устойчива.

#### 13.3.4. Простейший итерационный способ решения

Представим каждое уравнение (132) в виде, разрешенном относительно значения  $u_{k,m}$  в центральной точке расчетной ячейки:

$$u_{k,m} = \frac{1}{4}(u_{k-1,m} + u_{k+1,m} + u_{k,m-1} + u_{k,m+1} - h^2 f_{k,m}). \quad (133)$$

Получили основную расчетную формулу метода Зейделя для решения системы линейных алгебраических уравнений, аппроксимирующих эллиптические уравнения, – итерационного процесса решения задачи (132).

$$u_{k,m}^{(s+1)} = \frac{1}{4}(u_{k-1,m}^{(s+1)} + u_{k+1,m}^{(s)} + u_{k,m-1}^{(s+1)} + u_{k,m+1}^{(s)} - h^2 f_{k,m}), \quad (134)$$

где  $s$  – номер итерации.

Докажем сходимость.

Пусть  $u_{k,m}^{(s)} = u_{k,m} + \delta_{k,m}^{(s)}$ , где  $u_{k,m}$  – точное решение,  $u_{k,m}^{(s)}$  – приближенное решение за  $s$  итераций.

Тогда для ошибки  $\delta_{k,m}^s$  имеем

$$\delta_{k,m}^{(s+1)} = \frac{1}{4}(\delta_{k+1,m}^{(s)} + \delta_{k-1,m}^{(s)} + \delta_{k,m+1}^{(s)} + \delta_{k,m-1}^{(s)}), \quad \delta_{k,m}^{(s+1)}|_{\gamma} = 0.$$

Обозначим  $\max_{k,m} |\delta_{k,m}^{(0)}| = \bar{\delta}$ , оценим  $|\delta_{k,m}^{(1)}|$ . Так как значение  $\delta_{k,m}^{(1)}$  является средним арифметическим четырех значений  $\delta_{k\pm 1, m\pm 1}^{(0)}$ , то  $|\delta_{k,m}^{(1)}| \leq \bar{\delta}$ . Это неравенство справедливо для всех точек на всех итерациях.

Для точек, соседних с граничными, хотя бы одна из  $\delta_{k\pm 1, m\pm 1}^{(0)}$  равна нулю; следовательно, в этой точке  $|\delta_{k,m}^{(1)}| \leq \frac{0 + 3\bar{\delta}}{4} = \frac{3}{4}\bar{\delta}$ . Это соотношение справедливо для всего приграничного слоя точек.

На второй итерации влияние граничных точек распространяется на следующий внутренний слой точек:

$$|\delta_{k,m}^{(2)}| \leq \frac{\frac{3}{4}\bar{\delta} + \bar{\delta}}{4} = \frac{15}{16}\bar{\delta} \text{ и т.д.}$$

Следовательно, на  $s$ -й итерации  $|\delta_{k,m}^{(s)}| \leq \left(1 - \frac{1}{4^s}\right) \bar{\delta}$ . На некоторой  $n$ -й итерации ( $n \approx \frac{1}{h}$ ) получим, что ошибка уменьшилась не менее чем в  $(1 - 4^{-s})$  раз и т.д., т.е.  $\delta \rightarrow 0$  при  $s \rightarrow \infty$ .

### 13.4. Уравнения гиперболического типа

Рассмотрим простейшее уравнение гиперболического типа – *уравнение свободных колебаний однородной ограниченной струны*:

$$\frac{\partial^2 U}{\partial t^2} = a^2 \frac{\partial^2 U}{\partial x^2}; \quad (135)$$

$$U(x, 0) = f(x), \quad \frac{\partial U}{\partial t}(x, 0) = F(x), \quad 0 \leq x \leq l; \quad (136)$$

$$U(0, t) = \varphi(t), \quad U(l, t) = \psi(t), \quad 0 \leq t < \infty. \quad (137)$$

Решим методом сеток.

Конечно-разностную аппроксимацию этой задачи представим следующей разностной схемой:

$$\frac{u_{k,n+1} - 2u_{k,n} + u_{k,n-1}}{\tau^2} = a^2 \frac{u_{k+1,n} - 2u_{k,n} + u_{k-1,n}}{h^2}. \quad (138)$$

При  $\tau = \frac{h}{a}$  уравнение (138) упрощается и принимает вид

$$u_{k,n+1} + u_{k,n-1} = u_{k+1,n} + u_{k-1,n},$$

откуда

$$u_{k,n+1} = u_{k+1,n} + u_{k-1,n} - u_{k,n-1}. \quad (139)$$

Из уравнения (139) видно, что для получения значений  $U(x, t)$  на  $(n+1)$ -м слое используются значения  $U(x, t)$  в двух предыдущих слоях:  $n$ -м и  $(n-1)$ -м. Для начала вычислений по формуле (138) также необходимо знать значения  $U(x, t)$  на двух слоях, в то время как начальные условия (136) задают значения  $U(x, t)$  лишь на 0-м слое. Используя начальные условия, определим значения  $U(x, t)$  на фиктивном  $(-1)$ -м слое. Для этого заменим производную во втором начальном условии конечно-разностным отношением, получим:

$$\frac{u_{k,-1} - u_{k,0}}{-\tau} = \frac{\partial U}{\partial t}(x_k, 0) = F(x_k).$$

Отсюда

$$u_{k,-1} = u_{k,0} - \tau F(x_k). \quad (140)$$

Теперь, зная значения  $U(x, t)$  на фиктивном  $(-1)$ -м слое, начинаем вычисления по формуле (140). Для получения значений  $u_{0,n}$  и  $u_{K,n}$  используем краевые условия (137).

Пример 43

Решить уравнение  $\frac{\partial^2 U}{\partial t^2} = \frac{\partial^2 U}{\partial x^2}$  при граничных и начальных условиях:

$$U(x, 0) = x(\pi - x), \quad \frac{\partial U}{\partial t}(x, 0) = 0, \quad 0 \leq x \leq \pi,$$

$$U(0, t) = U(\pi, t) = 0, \quad 0 \leq t < \infty.$$

Решение.

Здесь  $a = 1 \Rightarrow \tau = h$ . Пусть  $\tau = h = \frac{\pi}{18}$ .

Для расчета воспользуемся формулой (139), для этого определим  $u_{k,-1}$  по формуле (139). Поскольку  $F(x) = 0$ , получим:

$$u_{k,-1} = u_{k,0} - \tau F(x_0) = u_{k,0}.$$

$$u_{10} = \frac{\pi}{18} \left( \pi - \frac{\pi}{18} \right) = 0,518; \quad u_{20} = \frac{2\pi}{18} \left( \pi - \frac{2\pi}{18} \right) = 0,975 \text{ и т.д. (табл. 12);}$$

$$u_{11} = u_{20} + u_{00} - u_{1,-1} = u_{20} + u_{00} - u_{10} = 0,975 + 0 - 0,518 = 0,457;$$

$$u_{21} = u_{30} + u_{10} - u_{20} = 1,371 + 0,518 - 0,975 = 0,914;$$

$$u_{31} = u_{40} + u_{20} - u_{30} = 1,706 + 0,975 - 1,371 = 1,310 \text{ и т.д.}$$

График решения  $u = U(x, t)$  симметричен относительно плоскости  $x = \pi/2$ ; поэтому в табл. 12 приведены данные лишь для  $0 \leq x \leq \pi/2$ .

Таблица 12

$x_k \setminus t_n$	$t_0 = 0$	$t_1 = h$	$t_2 = 2h$	$t_3 = 3h$	$t_4 = 4h$	$t_5 = 5h$
$x_0 = 0$	0	0	0	0	0	0
$x_1 = h$	0,518	0,457	0,457	0,335	0,274	0,214
$x_2 = 2h$	0,975	0,914	0,792	0,731	0,549	0,427
$x_3 = 3h$	1,371	1,310	1,188	1,006	0,884	0,641
$x_4 = 4h$	1,706	1,645	1,524	1,341	1,098	0,914
$x_5 = 5h$	1,980	1,920	1,798	1,616	1,371	1,038
$x_6 = 6h$	2,193	2,133	2,012	1,828	1,556	1,28
$x_7 = 7h$	2,346	2,285	2,163	1,952	1,737	1,374
$x_8 = 8h$	2,437	2,376	2,225	2,072	1,770	1,524
$x_9 = 9h$	2,467	2,407	2,285	2,043	1,859	1,497

## Контрольные вопросы и задания

1. Из каких этапов состоит метод прогонки?
2. Какой порядок имеет уравнение теплопроводности?
3. Что такое оператор Лапласа?
4. В чем отличие явной схемы аппроксимации от неявной?
5. Является ли уравнение Пуассона дифференциальным уравнением с частными производными?
6. Является ли уравнение Пуассона частным случаем уравнения Лапласа?
7. К уравнениям какого типа относится уравнение свободных колебаний однородной ограниченной струны?



## ЗАКЛЮЧЕНИЕ

Настоящее учебное пособие предназначено для усвоения лекционного курса по дисциплине «Вычислительная математика», читаемого студентам бакалавриата по направлению 09.03.02 «Информационные системы и технологии» Пензенского государственного университета архитектуры и строительства, а также тем, вынесенных на самостоятельное изучение.

Рассмотрены важные разделы, изучение которых помогает студенту в усвоении других естественнонаучных и общеобразовательных дисциплин – высшей математики, физики и т.д. Это интерполяция, аппроксимация, решение систем линейных алгебраических уравнений, решение нелинейных уравнений и их систем, решение дифференциальных уравнений и их систем, вычисление интегралов, решение разностных уравнений. Все разделы логически связаны между собой, иллюстрированы примерами. Предложенные в данном учебном пособии контрольные вопросы и задания призваны закрепить теоретические знания и способствовать развитию практических навыков, полученных в результате изучения курса, чтобы оно могло служить руководством для решения задач, при выполнении курсового проекта, подготовке к коллоквиумам, тестам и зачету.

Учебное пособие является частью учебно-методического комплекса по дисциплине. Содержание материала соответствует требованиям государственных стандартов ВО. Авторы надеются, что пособие поможет студентам не только познакомиться с разделами учебного курса, но и получить навыки работы с формальными моделями и системами.

В результате изучения данной дисциплины у студентов и слушателей курса формируются знания основных алгоритмов типовых численных методов решения математических задач, а также умения применять математические методы при решении профессиональных задач повышенной сложности, в том числе в интегрированной математической среде MathCad и *MS Excel*.

Необходимыми предпосылками для успешного усвоения дисциплины являются знания, умения, навыки, компетенции, сформированные у студентов при изучении курсов: математики, дискретной математики, математической логики и теории алгоритмов, дифференциальных уравнений, информатики, вероятности и статистики.

Учебное пособие может быть полезно студентам, обучающимся по направлениям подготовки ВО, реализуемым в вузе, а также аспирантам и инженерам, желающим овладеть основами методов вычислительной математики при решении прикладных задач.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бахвалов, Н.С. Численные методы [Электронный ресурс] / Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельков. – Электрон. текстовые данные. – М.: БИНОМ. Лаборатория знаний, 2012. – 635 с. – Режим доступа: <http://www.iprbookshop.ru/6502>. – ЭБС «IPRbooks», по паролю.
2. Бахвалов, Н.С. Численные методы в задачах и упражнениях [Электронный ресурс]: учеб. пособие / Н.С. Бахвалов, А.В. Лапин, Е.В. Чижонков. – Электрон. текстовые данные. – М.: БИНОМ. Лаборатория знаний, 2013. – 240 с. – Режим доступа: <http://www.iprbookshop.ru/12282>. – ЭБС «IPRbooks», по паролю.
3. Бедарев, И.А. Методы вычислений [Текст]: учеб. пособие / И.А. Бедарев, Ю.В. Кратова, Н. Н. Федорова ; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – 2-е изд., перераб. и доп. – Новосибирск : НГАСУ (Сибстрин), 2009. – 116 с.
4. Васильков, Ю.В. Компьютерные технологии вычислений в математическом моделировании [Текст]: учеб. пособие / Ю.В. Васильков, Н.Н. Василькова – М.: Финансы и статистика, 1999. – 256 с.
5. Вержбицкий, В.М. Основы численных методов [Текст]: учебник для вузов / В.М. Вержбицкий. – М.: Высшая школа, 2002. – 840 с.
6. Вычислительная математика [Текст]: методические указания для самостоятельной работы по направлению подготовки 09.03.02 «Информационные системы и технологии» / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2016. – 30 с.
7. Вычислительная математика [Текст]: методические указания по выполнению курсового проекта для направления подготовки 09.03.02 «Информационные системы и технологии» / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2016. – 32 с.
8. Данко, П.Е. Высшая математика в упражнениях и задачах: учеб. пособие для вузов. В 2-х ч. Ч. 1. / П.Е. Данко, А.Г. Попов, Т.Я. Кожевникова, С.П. Данко. – 7-е изд., испр. – М.: Оникс: Мир и образование, 2008. – 448 с.
9. Демидович, Б.П. Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения [Текст]: учеб. пособие / Б.П. Демидович, И.А. Марон, Э.З. Шувалова. – 3-е изд., перераб. – М.: Наука. Гл. ред. физ.-мат. лит., 1967. – 368 с.
10. Дьяконов, В.П. MATLAB. Полный самоучитель [Электронный ресурс] / Дьяконов В.П. – Электрон. текстовые данные. – М.: ДМК Пресс, 2014. – 768 с. – Режим доступа: <http://www.iprbookshop.ru/7911>. – ЭБС «IPRbooks», по паролю.
11. Задачи и упражнения по математическому анализу [Текст]: учеб. пособие для вузов / под ред. Б.П. Демидовича. – 7-е изд., стереотип. – М.: Наука. Гл. ред. физ.-мат. лит., 1970. – 472 с.

12. Киреев, В.И. Численные методы в примерах и задачах [Текст]: учеб. пособие / В.И. Киреев, А.В. Пантелеев. – М.: Высшая школа, 2004. – 480 с.
13. Кондаков Н.С. Основы численных методов [Электронный ресурс]: практикум / Кондаков Н.С.— Электрон. текстовые данные. – М.: Московский гуманитарный университет, 2014. – 92 с. – Режим доступа: <http://www.iprbookshop.ru/39690>. – ЭБС «IPRbooks», по паролю.
14. Копченова, Н.В. Вычислительная математика в примерах и задачах [Текст]: учеб. пособие / Н.В. Копченова, И.А. Марон.– 2-е изд., стер. – СПб.: Лань, 2008. – 368 с.
15. Контрольно-измерительные материалы по курсу «Вычислительная математика» [Текст]: учебно-методическое пособие / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2013. – 98 с.
16. Кошев, А.Н. Введение в численные методы [Текст]: учеб. пособие / А.Н. Кошев, С.В. Бакушев, И.Г. Гвоздева. – Пенза: ПГАСА, 2000. – 53 с.
17. Кошев, А.Н. Дискретная математика [Текст]: учеб. пособие: в 2-х ч. Ч. I. Элементы дискретной математики. Ч. II. Элементы теории графов / А.Н. Кошев, В.В. Кузина. – Пенза: ПГАСА, 2002. – 156 с.
18. Кошев, А.Н. Вычислительные методы [Текст]: учеб. пособие / А.Н. Кошев, В.В. Кузина. – Пенза: ПГУАС, 2012. – 204 с.
19. Кошев, А.Н. Численные методы и методы оптимизации [Текст]: учеб. пособие: в 2-х ч. / А.Н. Кошев, В.В. Кузина. – Пенза: ПГУАС, 2004. – 136 с.
20. Кошев, А.Н. Численные методы решения задач оптимизации [Текст]: учеб. пособие / А.Н. Кошев, В.В. Кузина. – Пенза: ПГУАС, 2012. – 132 с.
21. Кузина, В.В. Вычислительная математика [Текст]: лабораторный практикум для направления подготовки 09.03.02 «Информационные системы и технологии» / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2016. – 124 с.
22. Кузина, В.В. Вычислительная математика [Текст]: учебно-методическое пособие по подготовке к зачету по направлению подготовки 09.03.02 «Информационные системы и технологии» / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2016. – 64 с.
23. Кузина, В.В. Математическое моделирование в задачах мониторинга состояния водной среды [Текст]: моногр. / В.В. Кузина, А.Н. Кошев. – Пенза: ПГУАС, 2014. – 144 с.
24. Курсовые работы по направлению 230200 «Информационные системы» [Текст]: методические указания для студентов специальности «Информационные системы и технологии» по выполнению курсовых работ / А.Н. Кошев, В.В. Кузина. – Пенза: ПГУАС, 2006.– 28 с.
25. Методы математической физики [Электронный ресурс]: учеб. пособие/ Ю.В. Гриняев [и др.]. – Электрон. текстовые данные. – Томск: Томский государственный университет систем управления и

радиоэлектроники, Эль Контент, 2012. – 148 с. – Режим доступа: <http://www.iprbookshop.ru/13862>. – ЭБС «IPRbooks», по паролю.

26. Мышкис, А.Д. Элементы теории математических моделей [Текст] / А.Д. Мышкис. – 2-е изд., испр. – М.: Едиториал УРСС, 2004. – 192 с.

27. Охорзин, В.А. Прикладная математика в системе MATHCAD [Текст]: учеб. пособие / В.А. Охорзин. – 2-е изд., испр. и доп. – СПб.: Лань, 2008. – 352 с.

28. Пантелеев, А.В. Методы оптимизации в примерах и задачах [Текст]: учеб. пособие / А.В. Пантелеев, Т.А. Летова. – М.: Высшая школа, 2002. – 544 с.

29. Пантина, И.В. Вычислительная математика [Электронный ресурс]: учебник/ Пантина И.В., Синчуков А.В. – Электрон. текстовые данные. – М.: Московский финансово-промышленный университет «Синергия», 2012. – 176 с. – Режим доступа: <http://www.iprbookshop.ru/17012>. – ЭБС «IPRbooks», по паролю.

30. Плис, А.И. Mathcad. Математический практикум для инженеров и экономистов [Текст]: учеб. пособие / А.И. Плис, Н.А. Сливина. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2003. – 656 с.

31. Плохотников, К.Э. Вычислительные методы. Теория и практика в среде MATLAB [Электронный ресурс]: курс лекций. Учебное пособие для вузов/ Плохотников К.Э. – Электрон. текстовые данные. – М.: Горячая линия – Телеком, 2013. – 496 с. – Режим доступа: <http://www.iprbookshop.ru/37120>. – ЭБС «IPRbooks», по паролю.

32. Пшеничный, Б.Н. Численные методы в экстремальных задачах [Текст] / Б.Н. Пшеничный, Ю.М. Данилин. – М.: Наука. Гл. ред. физ.-мат. лит., 1975. – 320 с.

33. Самарский, А.А. Введение в численные методы [Текст]: учеб. пособ. для вузов / А.А. Самарский. – 2-е изд., перераб. и доп.– М.: Наука. Гл. ред. физ.-мат. лит., 1987. – 287 с.

34. Самарский, А.А. Задачи и упражнения по численным методам [Текст]: учеб. пособие / А.А. Самарский, П.Н. Вабищевич, Е.А. Самарская. – М.: Едиториал УРСС, 2000. – 208 с.

35. Седов, Е.С. Основы работы в системе компьютерной алгебры Mathematica [Электронный ресурс]/ Седов Е.С. – Электрон. текстовые данные. – М.: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. – 401 с. – Режим доступа: <http://www.iprbookshop.ru/16717>. – ЭБС «IPRbooks», по паролю.

36. Супрун, А.Н. Вычислительная математика для инженеров-экологов [Текст]: методическое пособие / А.Н. Супрун, В.В. Найденко. – М.: Изд-во Ассоциации строительных вузов, 1996. – 391 с.

37. Тыртышников, Е.Е. Методы численного анализа [Текст]: учеб. пособие для студ. вузов / Е.Е. Тыртышников – М.: ИЦ «Академия», 2007. – 320 с.

Перечень ресурсов информационно-телекоммуникационной сети  
«Интернет», необходимых для освоения дисциплины

1. [http://enc-dic.com/enc\\_math/Vchislitel'naja-matematika-6033.html](http://enc-dic.com/enc_math/Vchislitel'naja-matematika-6033.html)  
Электронный ресурс. Дата обращения 25.05.2016 г.
2. <http://www.intuit.ru/>
3. <http://www.exponenta.ru/>
4. [www.mathnet.ru](http://www.mathnet.ru) – общероссийский математический портал;
5. [http://e.lanbook.com/books/?p\\_f\\_1\\_temp\\_id=18&p\\_f\\_1\\_65=917&p\\_f\\_1\\_63=&p\\_f\\_1\\_67=](http://e.lanbook.com/books/?p_f_1_temp_id=18&p_f_1_65=917&p_f_1_63=&p_f_1_67=) – электронно-библиотечная система, издательство «Лань»;
6. [www.elibrary.ru](http://www.elibrary.ru) – научная электронная библиотека;
7. <http://lib.mexmat.ru/> – электронная библиотека механико-математического факультета МГУ;
8. [http://www.newlibrary.ru/genre/nauka/matematika/kompyutery\\_i\\_matematika/](http://www.newlibrary.ru/genre/nauka/matematika/kompyutery_i_matematika/) – электронная библиотека по математике;
9. [http://www.edu.ru/modules.php?op=modload&name=Web\\_Links&file=index&l\\_op=viewlink&cid=2851](http://www.edu.ru/modules.php?op=modload&name=Web_Links&file=index&l_op=viewlink&cid=2851) – федеральный портал российского профессионального образования: численные методы;
10. [https://mipt.ru/education/chair/computational\\_mathematics/study/materials/compmath/](https://mipt.ru/education/chair/computational_mathematics/study/materials/compmath/) – кафедра вычислительной математики МФТИ: вычислительная математика (3 курс).

## СПИСОК ИСПОЛЬЗУЕМЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ

$\ A\ $	– норма матрицы $A$
$A^T$	– матрица, транспонированная к матрице $A$
$(a, b)$	– скалярное произведение векторов $a$ и $b$
$\text{cond}(A)$	– число обусловленности матрицы $A$
$F_{1-\alpha}$	– критерий Фишера
$\min_X f(x),$ $\max_X f(x)$	– минимальное и максимальное значения функции $f(x)$ на множестве $X$
$R^n$	– $n$ -мерное евклидово арифметическое пространство
$ x $	– длина (модуль), евклидова норма вектора $x$
$x \in \Omega$	– $x$ принадлежит множеству $\Omega$ , т.е. является его элементом
$\Delta$	– абсолютная погрешность
$\delta$	– относительная погрешность
$\delta_{ij}$	– символ Кронекера
$\left( \frac{\partial F}{\partial X} \right)$	– матрица Якоби
$\Delta z$	– приращение функции $z$
$\mu$	– сингулярное число
$\Omega \subset R^n$	– множество $\Omega$ принадлежит $R^n$ , т.е. является его подмножеством
ЛР	– лабораторная работа
МНК	– метод наименьших квадратов
ОДУ	– обыкновенное дифференциальное уравнение
СЛАУ	– система линейных алгебраических уравнений
СНУ	– система нелинейных уравнений
CL	– кликнуть левой кнопкой мыши
RCL	– кликнуть правой кнопкой мыши

## ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ .....	3
ВВЕДЕНИЕ .....	5
1. ВВЕДЕНИЕ В ВЫЧИСЛИТЕЛЬНУЮ МАТЕМАТИКУ .....	10
1.1. Общая характеристика численных методов .....	10
1.2. Приближенные числа и действия над ними .....	11
1.3. Устойчивость вычислительного алгоритма .....	18
1.4. Корректность постановки задачи .....	18
Контрольные вопросы и задания .....	21
2. СИСТЕМЫ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ .....	22
2.1. Задачи, приводящие к СЛАУ .....	22
2.2. Некоторые характерные типы матриц СЛАУ .....	23
2.3. Некоторые свойства матриц и матричных операций .....	25
2.4. Собственные значения и собственные векторы матрицы .....	27
2.5. Норма матрицы и её свойства .....	28
2.6. Матричные преобразования .....	31
2.6.1. Эквивалентное преобразование .....	31
2.6.2. Преобразование подобия .....	33
Ортогонализация Грама – Шмидта .....	37
2.6.3. Разложение матрицы на произведение ортогональной и треугольной .....	38
2.7. Методы решения СЛАУ .....	40
2.7.1. Прямые методы решения СЛАУ .....	41
2.8. Обусловленность СЛАУ .....	49
Контрольные вопросы и задания .....	51
3. ИТЕРАЦИОННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ .....	52
3.1. Общие сведения .....	52
3.2. Метод простой итерации .....	53
3.3. Метод Зейделя .....	54
3.4. Метод релаксации (модификация метода Зейделя для ускорения сходимости) .....	56
3.4. Вопросы сходимости стационарных итерационных методов .....	56
Контрольные вопросы и задания .....	62
4. РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ .....	63
4.1. Постановка задачи .....	63
4.2. Правила выделения корня .....	63
4.3. Итерационные методы решения нелинейных уравнений .....	66
4.3.1. Условия сходимости итерационных методов .....	66
4.3.2. Принцип выбора функции для итерационного процесса .....	69

4.4. Итерационные схемы численных методов решения нелинейных уравнений .....	70
4.4.1. Метод простой итерации .....	70
4.4.2. Метод половинного деления .....	71
4.4.3. Метод хорд (метод секущих) .....	74
4.4.4. Метод Ньютона (метод касательных) .....	76
4.4.5. Комбинированный метод хорд и касательных .....	77
Контрольные вопросы и задания .....	79
<b>5. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ .....</b>	<b>80</b>
5.1. Общие сведения .....	80
5.2. Метод Ньютона .....	84
5.3. Метод Ньютона – Рафсона .....	85
5.4. Метод регуляризации Левенберга – Марквардта .....	85
5.5. Применение градиентных методов для решения СЧУ .....	85
Контрольные вопросы и задания .....	87
<b>6. ИНТЕРПОЛИРОВАНИЕ ФУНКЦИЙ .....</b>	<b>88</b>
6.1. Постановка задачи .....	88
6.2. Полиномиальная интерполяция .....	89
6.2.1. Интерполяционный полином Лагранжа .....	91
6.2.2. Интерполяционный полином Ньютона .....	91
6.3. Принципы двумерной интерполяции .....	95
6.4. Кубическая сплайн-интерполяция .....	96
Контрольные вопросы и задания .....	100
<b>7. АППРОКСИМАЦИЯ ФУНКЦИЙ .....</b>	<b>101</b>
7.1 Метод наименьших квадратов .....	101
7.2. Оценка параметров эмпирической формулы, построенной по экспериментальным данным на примере линейной модели .....	107
7.3. Среднеквадратичная аппроксимация в метрическом пространстве .....	109
Контрольные вопросы и задания .....	112
<b>8. МЕТОДЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ .....</b>	<b>113</b>
8.1. Общие сведения .....	113
8.2. Построение квадратурных формул .....	115
8.2.1. Формула прямоугольников .....	116
8.2.2. Формула трапеций .....	116
8.2.3. Формула Симпсона (формула парабол) .....	117
8.2.4. Формулы Ньютона – Котеса .....	118
8.2.5. Квадратурная формула Чебышева .....	120
8.2.6. Квадратурная формула Гаусса .....	121
8.3. Оценки погрешности квадратурных формул .....	122



Контрольные вопросы и задания .....	124
9. РАЗНОСТНЫЕ УРАВНЕНИЯ .....	125
9.1. Сеточные функции .....	125
9.1.1. Разностные аналоги формул дифференцирования произведения и интегрирования по частям .....	125
9.2. Разностные уравнения .....	126
9.2.1. Уравнения первого порядка .....	126
9.2.2. Уравнения второго порядка с постоянными коэффициентами.....	127
9.2.3. Нахождение общего решения однородного разностного уравнения .....	128
Контрольные вопросы и задания .....	130
10. ЗАДАЧА КОШИ И КРАЕВАЯ ЗАДАЧА ДЛЯ РАЗНОСТНЫХ УРАВНЕНИЙ ВТОРОГО ПОРЯДКА .....	131
10.1. Разностное уравнение второго порядка с переменными коэффициентами .....	131
10.1.1. Решение разностных краевых задач методом прогонки .....	132
10.1.2. Устойчивость метода прогонки .....	133
Контрольные вопросы и задания .....	134
11. ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ. ....	135
11.1. Формулы численного дифференцирования.....	135
11.2. Численные методы решения дифференциальных уравнений ....	138
11.2.1. Аппроксимация и сходимость численных методов решения дифференциальных уравнений на примере метода Эйлера ....	138
11.2.2. Оценка порядка аппроксимации дифференциального уравнения другими разностными уравнениями.....	140
11.3. Схемы Рунге – Кутта.....	142
11.4. Разностные методы решения ОДУ второго порядка.....	143
Контрольные вопросы и задания .....	146
12. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ .....	147
12.1. Уравнения математической физики .....	147
12.1.1. Аппроксимация.....	148
12.1.2. Сходимость.....	149
12.1.3. Устойчивость .....	149
Контрольные вопросы и задания .....	150
13. УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ ВТОРОГО ПОРЯДКА .....	151
13.1. Классификация уравнений с частными производными .....	151
13.2. Уравнение параболического типа (уравнение теплопроводности).....	152

13.2.1. Решение разностного уравнения для неявной схемы.....	153
13.2.2. Метод прогонки .....	154
13.2.3. Итерационные методы .....	155
13.3. Уравнения эллиптического типа (уравнения Лапласа и Пуассона).....	155
13.3.1. Построение разностной схемы.....	156
13.3.2. Аппроксимация.....	156
13.3.3. Устойчивость .....	157
13.3.4. Простейший итерационный способ решения.....	157
13.4. Уравнения гиперболического типа .....	158
Контрольные вопросы и задания.....	160
ЗАКЛЮЧЕНИЕ.....	161
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	162
СПИСОК ИСПОЛЬЗУЕМЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ.....	166



Учебное издание

Кузина Валентина Владимировна  
Кошев Александр Николаевич

**ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА**  
Учебное пособие  
по направлению подготовки 09.03.02  
«Информационные системы и технологии»

В авторской редакции  
Верстка Н.В. Кучина, В.В. Кузина

---

Подписано в печать 18.10.16.                      Формат 60x84/16.  
Бумага офисная «Снегурочка». Печать на ризографе.  
Усл.печ.л. 10,0.      Уч.-изд.л. 10,75.                      Тираж 80 экз.  
Заказ № 640.

---



Издательство ПГУАС.  
440028, г. Пенза, ул. Германа Титова, 28.