

МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО
ОБРАЗОВАНИЯ РСФСР

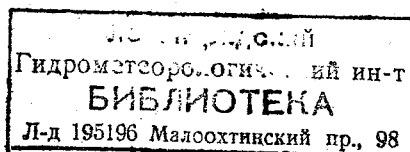
ЛЕНИНГРАДСКИЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ ИНСТИТУТ

А. С. БАЛУЕВА

ПРОСТЕЙШИЕ ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ГИДРОМЕТЕОРОЛОГИИ

(учебное пособие)

Под редакцией проф. И. П. Гинзбурга



ЛЕНИНГРАД

1975

*Одобрено Ученым советом
Ленинградского гидрометеорологического института*

В учебном пособии содержится ряд методов численного анализа, линейной алгебры и метод сеток в применении к задачам математической физики. Материал отвечает программе курса для океанологического и метеорологического факультетов.

Книга предназначена для студентов институтов и факультетов геофизического профиля, а также может быть рекомендована инженерам гидрометслужбы, интересующимся методами вычислительной математики.

© Ленинградский гидрометеорологический
институт (ЛГМИ), 1975 г.

ПРЕДИСЛОВИЕ

Настоящее учебное пособие предназначено для студентов гидрометеорологических институтов и факультетов, а также оно будет полезно для сотрудников гидрометеослужбы. При составлении пособия учитывались все возрастающие требования к научным сотрудникам и инженерам-гидрометеорологам в отношении использования быстродействующих вычислительных машин и применения численных методов.

Книга написана на основании чтения лекций и проведения математического практикума в Ленинградском гидрометеорологическом институте. Она включает ряд методов, используемых при решении задач геофизики.

В первой главе пособия даны простейшие методы численного анализа. Вторая и третья главы посвящены численным методам линейной алгебры и методу сеток. Книга иллюстрирована примерами из различных областей гидрометеорологии.

При подборе и решении примеров значительную помощь оказал В. Н. Веретенников. В расчете примеров приняли также участие студенты-океанологи М. А. Белянцев, А. И. Радошевский, В. А. Паялов, С. В. Котов. Всем указанным лицам автор приносит свою искреннюю благодарность.

Автор считает своим приятным долгом поблагодарить редактора книги профессора Ленинградского государственного университета И. П. Гинзбурга за ряд ценных указаний и советов.

ВВЕДЕНИЕ

Всякая физическая задача, решаемая математическим методом, приводит нас к уравнениям, связывающим те или иные переменные или постоянные величины. Часть этих величин задана, часть подлежит определению. При этом мы должны убедиться, что задача поставлена корректно, т. е. установить существование и единственность решения уравнений, показать, что имеет место устойчивость по начальным и граничным условиям и, наконец, указать конструктивный метод нахождения решения.

Задачи гидрометеорологического профиля, как правило, настолько сложны, что для нахождения решения большинства из них приходится применять приближенные методы. Кроме того, применение быстродействующих вычислительных машин накладывает определенные условия на алгоритм, определяющий решение задачи.

1. Мы не будем касаться постановки задач и проверки корректности постановки. Нас интересуют приближенные методы решения уже поставленных задач, причем будем заниматься численными методами, чаще всего применяемыми в метеорологии, океанологии и гидрологии.

Будем считать, что каждая рассматриваемая задача имеет решение, которое назовем точным, в отличие от описываемого приближенного. Всякий приближенный метод предполагает замену точных уравнений новыми приближенными, в том или ином смысле близкими к данным. Решения этих новых уравнений и являются приближенными. Очевидно, новые уравнения должны быть проще исходных. Получив приближенное решение, следует найти или хотя бы оценить погрешность, т. е. разность между истинным и приближенным решением. Эта погрешность состоит из погрешности метода, погрешности исходных данных и погрешности округления. Погрешность метода может быть определена только для каждого конкретного метода. Следует также иметь в виду, что часто приближенный метод является методом последовательных приближений, когда мы получаем последовательность приближенных решений, пределом кото-

рой является точное решение. В этом случае требуется установить существование предела последовательности и показать, что этот предел является точным решением. При расчете по методу последовательных приближений применяют формулы типа рекуррентных. Очень существенно при выборе метода количество вычислительных операций: чем меньше операций содержит метод, тем меньше затрачивается машинного времени, меньше используется машинная память и тем меньше погрешность счета, получаемая за счет округления. Существует еще одно условие, которое необходимо соблюдать при выборе метода. Это условие — устойчивость счета. Схема является устойчивой, если исходная или случайная погрешность остается при счете органиченной. К сожалению, очень часто проверить устойчивость счета трудно. Приведем пример неустойчивой схемы. Пусть необходимо найти последовательность значений $y_2, y_3, \dots$ по формулам

$$y_{n+1} = -3y_n + 6y_{n-1}$$

и пусть y_0 имеет погрешность $(-\epsilon)$, а y_1 погрешность (ϵ) . Тогда непосредственный расчет показывает, что y_2 имеет погрешность (-9ϵ) , а y_3 погрешность 33ϵ , y_4 погрешность (-153ϵ) и т. д. Хорошо видно, что счет по указанным формулам вести невозможно — он неустойчив.

2. Рассмотрим теперь, как оценить погрешность исходных данных. Все исходные данные задачи, полученные при наблюдениях или эксперименте, являются приближенными числами, так как могут быть получены только с той точностью, которую дает прибор.

Пусть x — приближенное значение числа x^* . Будем называть $|x - x^*|$ абсолютной погрешностью числа x . Величина Δ_x , удовлетворяющая неравенству

$$\Delta_x \geq |x - x^*|, \quad (1)$$

где Δ_x выбрано возможно меньшим, является предельной абсолютной погрешностью числа x . Очевидно, практически мы будем иметь дело только с приближенным значением числа x и его предельной погрешностью, x и Δ_x — это показание прибора и точность прибора. Точное значение числа, а также абсолютная погрешность обычно остаются неизвестными.

Приближенное число обычно записывается в виде десятичной дроби

$$x = a_1 10^n + a_2 10^{n-1} + \dots + a_m 10^{n-m+1}. \quad (2)$$

Все цифры числа, начиная с первой не нулевой, называются значащими цифрами. Количество значащих цифр можно увеличивать и уменьшать, не изменяя числа. Так, число 1,110 записано с четырьмя значащими цифрами; оно же с тремя 1,11 и 1,1100 с пятью.

Цифру a_m числа называют верной, если для предельной абсолютной погрешности Δ_x этого числа (2) имеет место неравенство

$$\Delta_x \leq 10^{n-m+1}. \quad (3)$$

Если условие (3) не выполнено, то цифру называют сомнительной. Очевидно, если a_m верная, то верные также и $a_1, \dots, a_{m-1}$. Например, в числе $x = 1002,31$ верными являются все цифры, если $\Delta_x \leq 0,01$. Если же $\Delta_x \leq 0,1$, то, чтобы в числе были записаны только верные цифры, его следует записать в виде: $1002,3$. Кроме предельной абсолютной погрешности полезно вести предельную относительную погрешность

$$\delta_x = \frac{\Delta_x}{|x|}, \quad (4)$$

которая лучше характеризует точность числа. Связь между предельной относительной погрешностью и числом верных значащих цифр может быть легко установлена с помощью введенных ранее определений. Пусть m значащих цифр числа x верные, т. е. $\Delta_x \leq 10^{n-m+1}$. Тогда предельная относительная погрешность может быть записана в виде

$$\delta_x = \frac{10^{n-m+1}}{|a_1 10^n + \dots + a_m 10^{n-m+1}|} \leq \frac{10^{n-m+1}}{|a_1| 10^n} = \frac{10^{-m+1}}{|a_1|}. \quad (5)$$

Таким образом, получим: если число x имеет m верхних значащих цифр, то за предельную относительную погрешность принимают величину $\frac{10^{-m+1}}{|a_1|}$, и если известна предельная относительная погрешность δ_x , то за число верных цифр принимают наименьшее из чисел m , удовлетворяющих неравенству

$$\delta_x \leq \frac{10^{-m+1}}{|a_1|}.$$

Рассмотрим вопрос о погрешности простейших функций и арифметических операций. Пусть дана непрерывная и дифференцируемая функция своих аргументов $y^* = f(x_1^*, x_2^*, \dots, x_n^*)$ в некоторой области Ω . Так как значения аргументов обычно даны приближенно

$$x_1 = x_1^* + \Delta_1, \quad x_2 = x_2^* + \Delta_2, \quad \dots, \quad x_n = x_n^* + \Delta_n,$$

то значение функции мы получим тоже приближенно

$$y = f(x_1, x_2, \dots, x_n) = y^* + \Delta_y.$$

Требуется определить предельную абсолютную погрешность функции Δ_y . Применим формулу Лагранжа

$$y - y^* = \sum_{k=1}^n \frac{\partial f}{\partial x_k} \Big|_M (x_k - x_k^*);$$

где M — точка области Ω с координатами $(x_1 + \theta(x_1 - x_1^*), x_2 + \theta(x_2 - x_2^*), \dots, x_n + \theta(x_n - x_n^*))$; $0 < \theta < 1$.

Будем предполагать, что $\frac{\partial f}{\partial x_k}$ в области Ω изменяются достаточно медленно и погрешность аргументов достаточно мала. В этом случае можно приближенно принять для предельной абсолютной погрешности формулу

$$\Delta_y \approx \sum_{k=1}^n \left| \frac{\partial f}{\partial x_k} \right| \Delta_k, \quad (6)$$

а для относительной — формулу

$$\delta_y \approx \sum_{k=1}^n \frac{1}{|y|} \left| \frac{\partial f}{\partial x_k} \right| \Delta_k = \sum_{k=1}^n \left| \frac{\partial \ln f}{\partial \ln x_k} \right| \delta_k, \quad (7)$$

где δ_k — предельные относительные погрешности аргументов.

Из формул (6) и (7) можно получить приближенные формулы для погрешности арифметических действий.

Сложение

$$\begin{aligned} y &= x_1 + x_2 + \dots + x_n, \\ \Delta_y &= \Delta_1 + \Delta_2 + \dots + \Delta_n, \\ \delta_y &= \frac{1}{|y|} (|x_1| \delta_1 + |x_2| \delta_2 + \dots + |x_n| \delta_n). \end{aligned} \quad (8)$$

Формулы (8) показывают, что предложенная абсолютная погрешность суммы не превышает сумму предельных абсолютных погрешностей слагаемых.

Вычитание

$$\begin{aligned} y &= x_1 - x_2; \quad x_1 > x_2 > 0; \\ \Delta_y &= \Delta_1 + \Delta_2; \quad \delta_y = \frac{x_1}{y} \delta_1 + \frac{x_2}{y} \delta_2. \end{aligned} \quad (9)$$

Если $x_1 \gg x_2$, то $\delta_y \approx \delta_1$. Если x_2 близко к x_1 и, следовательно, y мало, то δ очень велика. Из-за этого при вычислениях следует избегать вычитания близких друг к другу чисел, изменяя расчетные формулы с помощью тождественных преобразований.

Умножение

$$y = x_1 \cdot x_2 \cdot \dots \cdot x_n,$$

$$\delta_y = \delta_1 + \delta_2 + \dots + \delta_n. \quad (10)$$

Относительная погрешность произведения равна сумме относительных погрешностей сомножителей.

Деление

$$y = \frac{x_1}{x_2}; \quad \delta_y = \delta_1 + \delta_2. \quad (11)$$

Относительная погрешность частного равна сумме относительных погрешностей делимого и делителя.

Мы остановились на разборе погрешностей первых четырех действий арифметики и рекомендуем читателям самим рассмотреть другие функции.

3. Используя приближенные данные при расчете, мы сталкиваемся с необходимостью округления чисел.

Вспомним правило округления. Чтобы округлить число до m значащих цифр, отбрасывают все цифры, стоящие справа от a_m . При этом, если первая отброшенная цифра меньше 5, то a_m оставляют без изменения, в противном случае к a_m прибавляют единицу.

Почти при каждом действии мы получаем погрешность округления ибо, в процессе вычислений сохраняется только конечное число десятичных знаков. Проблема определения погрешности округления особенно важна при машинном расчете, так как в этом случае велик объем работы и трудно оценивать эту погрешность. Существует ряд методов оценки погрешности округления, потому что до того как мы начинаем составлять программу для машины, следует иметь представление об ожидаемой точности.

I. Область ответа. Каждую исходную величину x , входящую в расчет, можно оценить сверху и снизу ($x - \Delta_x$, $x + \Delta_x$), получив таким образом область, в которой лежит эта величина. При выполнении действий над числами вычисляется новая область с учетом возможных округлений. Таким образом, можно на каждой стадии расчета получить область, в которой лежит верный ответ. Но такой подход к оценке погрешности округления удваивает работу и требует большей машинной памяти.

II. Двойная точность. В этом случае тоже считают дважды с обычной и двойной точностью и разряды, совпадающие в двух ответах, считают верными. Эти двойные расчеты делают в нескольких типичных вычислениях и принимают, что та же точность будет

для остальных. Этот метод считается одним из наиболее простых и удобных. В настоящее время разработан еще ряд методов, которых мы здесь не будем касаться.

Примеры.

1. Написать число π с пятью верными знаками. Определить предельные абсолютную и относительную погрешности этого приближения.

Ответ. $\pi \cong 3,1416$, $\Delta_x \leq 10^{-4}$, $\delta_x \cong 0,04\%$.

2. Написать число e с пятью верными значащими цифрами. Найти абсолютную и относительную предельные погрешности.

Ответ. $e \approx 2,7182$, $\Delta_x \leq 10^{-4}$, $\delta_x \cong 0,04\%$.

3. Проверить, будет ли устойчив расчет по формуле

$$y_{n+1} = -4y_n + 5y_{n-1}.$$

Примем при этом, что y_0 имеет погрешность $(-\epsilon)$, а y_1 — погрешность (ϵ) :

Ответ. Счет неустойчив, так как погрешность y_2 (-9ϵ) , а погрешность y_3 (41ϵ) , и при дальнейших расчетах погрешность очень быстро возрастает.

4. Составить формулу для расчета площади кругового кольца с радиусами, соответственно равными $R = 100$ см и $R + a = 101$ см.

Ответ. Формулу, по которой обычно вычисляется площадь, использовать нельзя, потому что величины $(R + a)^2$ и R^2 близки между собой. Следует преобразовать формулу

$$S = \frac{\pi}{2} [(R + a)^2 - R^2] = \frac{\pi}{2} (a^2 + 2aR).$$



Глава I

ПРОСТЕЙШИЕ ЧИСЛЕННЫЕ МЕТОДЫ

§ 1. Численные решения алгебраических и трансцендентных уравнений

1. Нахождение начального приближения

При применении целого ряда методов численного решения уравнений вида

$$P(x) = 0 \quad (1.1)$$

необходимо заранее знать более или менее близкое к действительному значение корня, которое мы будем называть начальным приближением. Укажем ряд приемов нахождения начального приближения. Если $y = P(x)$ — вещественная функция своего аргумента, то, построив графики этой функции, мы можем найти начальное приближение как точку пересечения кривой $y = P(x)$ с осью абсцисс. При этом иногда бывает целесообразно представить функцию $P(x)$ в виде

$$P(x) = P_1(x) - P_2(x),$$

и уравнение (1.1) запишется так:

$$P_1(x) = P_2(x). \quad (1.2)$$

В этом случае за начальное приближение принимают абсциссу точки пересечения кривых $y = P_1(x)$ и $y = P_2(x)$.

Можно также уравнение (1.1) заменить близким к нему в каком-нибудь смысле более простым уравнением $P_3(x) = 0$ и принять корень этого уравнения за начальное приближение к корню уравнения (1.1).

Для нахождения начального приближения корней уравнения (1.1) можно также составить таблицу функции $y = P(x)$ и найти промежутки, в которых лежат корни.

2. Метод хорд

Пусть для уравнения (1.1) известен промежуток (a, b) , в котором лежит один корень. В этом промежутке функция $y = P(x)$ строго монотонная и выполняется условие

$$P(a) \cdot P(b) < 0.$$

Будем предполагать также, что функция $y = P(x)$ в промежутке (a, b) дифференцируема и $P'(x) > 0$ (или $P'(x) < 0$) в (a, b) .

Заменим дугу AB хордой AB (рис. 1). Обозначим через x_1 точку пересечения хорды AB и оси Ox . Если $P(x_1) \neq 0$, то или $P(a)P(x_1) < 0$, или $P(x_1)P(b) < 0$. Выберем тот промежуток, где указанное условие выполняется, и переобозначим его (a_1, b_1) . Соответствующие точки дуги обозначим через A_1 и B_1 и построим новую хорду A_1B_1 . Точку пересечения этой хорды с осью Ox назовем x_2 — вторым приближением. Повторяя процесс k раз, напишем последовательность приближений

$$x_1, x_2, \dots, x_k, \dots$$

Можно показать, что $x_k \rightarrow x^*$, где x^* — искомый корень уравнения $P(x) = 0$.

Построим формулу для нахождения указанных приближений. Уравнение прямой, проходящей через точки $A(a, P(a))$ и $B(b, P(b))$, имеет вид

$$\frac{y - P(a)}{P(b) - P(a)} = \frac{x - a}{b - a}. \quad (1.3)$$

Координату точки пересечения прямой (1.3) и оси Ox получим так:

$$x_1 = a - \frac{P(a)}{P(a, b)}, \quad (1.4)$$

где

$$P(a, b) = \frac{P(b) - P(a)}{b - a}. \quad (1.5)$$

3. Метод Ньютона

Метод Ньютона или метод касательных может быть применен также к алгебраическим и трансцендентным уравнениям.

Рассмотрим уравнение (1.1). Пусть известно начальное приближение x_0 . Будем предполагать, что в окрестности точки $x = x_0$ функция $P(x)$ непрерывна и дифференцируема. Представим эту функцию в виде отрезка ряда Тейлора в окрестности точки $x = x_0$

$$P(x) \cong P(x_0) + P'(x_0)(x - x_0). \quad (1.6)$$

Заменим левую часть уравнения (1.1) приближенно по формуле (1.6)

$$P(x_0) + P'(x_0)(x - x_0) = 0. \quad (1.7)$$

Очевидно, полученное уравнение определяет касательную к кривой $y = P(x)$, построенную в точке $x = x_0$.

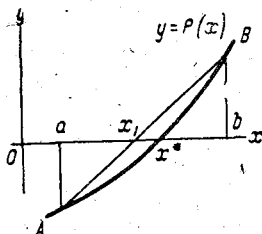


Рис. 1. Метод хорд

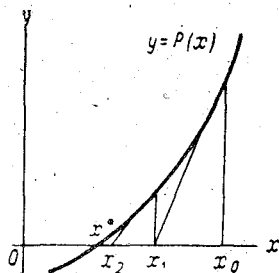


Рис. 2. Метод касательных

Координата x_1 точки пересечения касательной с осью абсцисс определяется формулой

$$x_1 = x_0 - \frac{P(x_0)}{P'(x_0)}. \quad (1.8)$$

Принимаем это значение за первое приближение к корню. Аналогично, повторяя описанный процесс, получим рекуррентное соотношение

$$x_k = x_{k-1} - \frac{P(x_{k-1})}{P'(x_{k-1})}. \quad (1.9)$$

Геометрическая интерпретация метода очень проста. Она представлена на рис. 2.

Можно применять также модифицированный метод Ньютона, когда последовательные приближения вычисляются по формулам:

$$x'_k = x'_{k-1} - \frac{P(x'_{k-1})}{P'(x'_{k-1})}; \quad x_0 = x'_0. \quad (1.10)$$

В этом случае на каждом шаге используется только одно значение производной $P'(x_0)$. Такая модификация метода упрощает вычисления, но сходимость оказывается более медленной.

Укажем без вывода условия сходимости метода Ньютона.

Теорема. Пусть для уравнения (1.1) и начального приближения x_0 выполняются следующие условия:

- 1) $P'(x_0) \neq 0$, $\frac{1}{|P'(x_0)|} \leq B$;
- 2) имеет место неравенство $\left| \frac{P(x_0)}{P'(x_0)} \right| \leq \eta$;
- 3) для любого x из промежутка $|x - x_0| \leq \delta$ справедливо неравенство $|P''(x)| \leq k$;
- 4) числа B, η, k подчинены условию $h = Bk\eta \leq \frac{1}{2}$.

Если

$$\delta \geq \frac{1 - \sqrt{1 - 2h}}{h} \eta,$$

то уравнение (1.1) в промежутке $|x - x_0| \leq \delta$ имеет корень x^* , к которому сходятся последовательные приближения

$$x_k = x_{k-1} - \frac{P(x_{k-1})}{P'(x_{k-1})}.$$

Пример 1. Найти наименьший и наибольший корни уравнения

$$x^3 - x^2 - 3x + 2 = 0.$$

с помощью метода хорд.

Решение. Для нахождения промежутков, в которых лежат корни, построим табл. 1, соответствующую функции $y = x^3 - x^2 - 3x + 2$.

Таблица 1

x	$-\infty$	-2	$-1,5$	0	1	2	∞
y	$-\infty$	-4	$0,75$	2	-1	0	∞

Из табл. 1 следует, что наибольший корень $x = 2$, а наименьший лежит в интервале $(-2, -1,5)$.

Применим теперь метод хорд для нахождения наименьшего корня.

Первое приближение

$$P(-2) = -4 < 0, P(-1,5) = 0,75 > 0, a = -2, b = -1,5,$$

$$x_1 = a - \frac{(b - a) P(a)}{P(b) - P(a)} = -1,590.$$

Второе приближение

$$P(-1,590) = 0,222 > 0, \quad a_1 = -2, \quad b_1 = -1,590,$$

$$x_2 = -1,612.$$

Третье приближение

$$P(-1,612) = 0,048 > 0, \quad a_2 = -2, \quad b_2 = -1,612,$$

$$x_3 = -1,618.$$

С точностью до сотых искомый корень $x^* = -1,62$.

Пример 2. Методом Ньютона найти отрицательный корень уравнения

$$x^4 - 2x - 4 = 0.$$

Решение. Найдем начальное приближение следующим способом: построим на одном чертеже (рис. 3) графики функций

$$y = x^4, \quad y = 2(x + 2).$$

На рис. 3 видно, что отрицательный корень лежит в интервале $(-1,5, -1)$.

Начальное приближение по методу Ньютона примем $x_0 = -1,5$.

Первое приближение

$$P'(x) = 4x^3 - 2,$$

Рис. 3. Начальное приближение

$$x_1 = x_0 - \frac{P(x_0)}{P'(x_0)} = -1,238.$$

Второе приближение

$$x_2 = x_1 - \frac{P(x_1)}{P'(x_1)} \cong -1,152.$$

Действуя таким же способом, найдем

$$x_3 \cong -1,144,$$

$$x_4 \cong -1,143.$$

Окончательный ответ: $x^* = -1,14$.

§ 2. Решение алгебраических уравнений по методу Лобачевского

Метод Лобачевского применяется для отыскания корней уравнений n -й степени с вещественными коэффициентами.

Основная идея метода заключается в преобразовании уравнения в некоторое другое, корнями которого служат высокие степени корней данного уравнения. Например, если исходное уравнение имеет два корня 3 и 4, отношение которых 0,75, то корни преобразованного уравнения будут 3^m и 4^m . Если взять $m = 32$, то отношение корней преобразованного уравнения приближенно равно 0,0001. Таким образом, меньшим корнем уравнения можно пренебречь. Говорят, что корни уравнения отделены, если отношение каждого корня к ближайшему большему есть величина малая сравнительно с единицей. Преобразование уравнений, имеющее целью отделение корней, называется квадрированием. Процесс построения нового уравнения разделен на отдельные шаги, так чтобы каждый раз корни нового уравнения были квадратами корней предыдущего.

Покажем, как проводится один из шагов квадрирования. Пусть данное уравнение имеет вид

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (1.11)$$

и пусть корнями его являются $x_1, x_2, \dots, x_n$. Тогда можно записать

$$f(x) = a_0 (x - x_1) (x - x_2) \dots (x - x_n). \quad (1.12)$$

Построим функцию

$$f(-x) = a_0 (-x - x_1) \dots (-x - x_n) = (-1)^n a_0 (x + x_1) \dots (x + x_n).$$

Вычислим произведение

$$f(x) \cdot f(-x) = a_0^2 (x^2 - x_1^2) (x^2 - x_2^2) \dots (x^2 - x_n^2).$$

Положим $y = -x$. Тогда

$$\begin{aligned} \varphi(y) &= (-1)^n a_0^2 (-y - x_1^2) \dots (-y - x_n^2) = \\ &= a_0^2 (y + x_1^2) \dots (y + x_n^2) \end{aligned} \quad (1.13)$$

или

$$\varphi(y) = a_0^{(1)} y^n + a_1^{(1)} y^{n-1} + \dots + a_n^{(1)} = 0. \quad (1.14)$$

Таким образом, мы получили уравнение, корнями которого являются: $-x_1^2, -x_2^2, \dots, -x_n^2$.

Случай 1. Корни уравнения вещественны и различны

Рассмотрим уравнение n -й степени после k -го квадрирования

$$a_0^{(k)} z^n + a_1^{(k)} z^{n-1} + \dots + a_n^{(k)} = 0. \quad (1.17)$$

Корнями этого уравнения являются: $-x_1^m, -x_2^m, \dots, -x_n^m$, где $m=2^k$ и $x_1, x_2, \dots, x_n$ — корни исходного уравнения (1.11).

Запишем соотношение между корнями и коэффициентами уравнения (1.17).

$$\begin{aligned} \frac{a_1^{(k)}}{a_0^{(k)}} &= x_1^m + x_2^m + \dots + x_n^m, \\ \frac{a_2^{(k)}}{a_0^{(k)}} &= x_1^m x_2^m + x_1^m x_3^m + \dots + x_{n-1}^m x_n^m, \\ &\vdots \\ \frac{a_n^{(k)}}{a_0^{(k)}} &= x_1^m x_2^m \dots x_n^m, \end{aligned} \quad (1.18)$$

Эти соотношения получают, приравнявая коэффициенты при одинаковых степенях x в тождестве

$$a_0^{(k)} z^n + a_1^{(k)} z^{n-1} + \dots + a_n^{(k)} = a_0^{(k)} (z + x_1^m) \dots (z + x_n^m).$$

Будем считать, что корни уравнения (1.11) убывают по модулю, т. е. $|x_1| > |x_2| > \dots > |x_n|$. Тогда естественно преобразовать систему (1.18) к виду

$$\begin{aligned} \frac{a_1^{(k)}}{a_0^{(k)}} &= x_1^m \left[1 + \left(\frac{x_2}{x_1} \right)^m + \dots + \left(\frac{x_n}{x_1} \right)^m \right], \\ \frac{a_2^{(k)}}{a_0^{(k)}} &= x_1^m x_2^m \left[1 + \left(\frac{x_2}{x_1} \right)^m + \dots + \left(\frac{x_{n-1} x_n}{x_1 x_2} \right)^m \right], \\ &\dots \dots \dots \\ \frac{a_n^{(k)}}{a_0^{(k)}} &= x_1^m x_2^m \dots x_n^m. \end{aligned} \quad (1.19)$$

Учитывая сказанное ранее, можно получить приближенные равенства:

$$\begin{aligned} x_1^m &\cong \frac{a_1^{(k)}}{a_0^{(k)}}, \\ x_1^m x_2^m &\cong \frac{a_2^{(k)}}{a_0^{(k)}}, \\ &\dots \dots \dots \\ x_1^m x_2^m \dots x_n^m &\cong \frac{a_n^{(k)}}{a_0^{(k)}}, \end{aligned}$$

из которых последовательно находим:

$$\left. \begin{aligned} x_1^m &\cong \frac{a_1^{(k)}}{a_0^{(k)}}, \\ x_2^m &\cong \frac{a_2^{(k)}}{a_1^{(k)}}, \\ &\dots \dots \dots \\ x_n^m &\cong \frac{a_n^{(k)}}{a_{n-1}^{(k)}}. \end{aligned} \right\} \quad (1.20)$$

Равенства (1.20) позволяют найти абсолютные значения корней уравнения (1.11). Знаки корней определяют подстановкой в уравнения.

Следует отметить, что если корни исходного уравнения вещественны и различны, то квадрирование прекращается, когда при очередном преобразовании удвоенные произведения коэффициентов делаются пренебрежимо малы по сравнению с квадратом соответствующего коэффициента, а значит, окончательные коэффициенты положительны. Указанный факт легко проверить. Пусть после k -го квадрирования получено нужное нам разделение корней. Тогда, с заданной степенью точности,

$$x_i^{(k)} \approx \frac{a_i^{(k)}}{a_{k-1}^{(k)}}.$$

Но и после $(k+1)$ -го квадрирования тем более справедлива формула

$$x_i^{(k+1)} \approx \frac{a_i^{(k+1)}}{a_{k-1}^{(k+1)}},$$

и так как $y_i^{k+1} = -[x_i^{(k)}]^2$, то очевидно $a_i^{(k+1)} \approx [a_i^{(k)}]^2$.

Случай 2. Комплексные сопряженные корни

В этом случае уравнение, записанное в форме (1.12) — через корни, будет содержать квадратичные множители, соответствующие комплексным сопряженным корням. Покажем, как пойдет процесс квадрирования. Пусть уравнение имеет различные вещественные корни и пару комплексных сопряженных корней x_2 и x_3 , $r = |x_2| = |x_3|$. Тогда $|x_1| > r > |x_4| > \dots > |x_n|$. Очевидно,

$$x_2^m + x_3^m = 2r^m \cos m\varphi; \quad x_2^m x_3^m = r^{2m}.$$

После k -го квадрирования, где $m=2^k$, получим

[illegible]

Все формулы (1.21), кроме второй, так же как и в предыдущем случае, можно заменить приближенными

$$x_1^m \approx \frac{a_1^{(k)}}{a_0^{(k)}},$$

$$x_1^m r^{2m} \approx \frac{a_3^{(k)}}{a_0^{(k)}},$$

$$x_1^m r^{2m} x_4^m \approx \frac{a_4^{(k)}}{a_0^{(k)}},$$

• • • • •

$$x_1^m r^{2m} \dots x_n^m \approx \frac{a_n^{(k)}}{a_0^{(k)}},$$

или

$$\left. \begin{aligned} x_1^m &\approx \frac{a_1^{(k)}}{a_0^{(k)}}, \\ r^{2m} &\approx \frac{a_3^{(k)}}{a_1^{(k)}}, \\ x_4^m &\approx \frac{a_4^{(k)}}{a_3^{(k)}}, \\ &\dots \\ x_n^m &\approx \frac{a_n^{(k)}}{a_{n-1}^{(k)}}. \end{aligned} \right\} \quad (1.22)$$

Формулы (1.22) позволяют определить абсолютное значение корней и модуль сопряженных комплексных чисел $r = |x_2| = |x_3|$. Очевидно, при квадрировании все коэффициенты, кроме a_2 , будут вести себя так же, как и в предыдущем случае. Коэффициент $a_2^{(k)}$, содержащий слагаемое $2x_1 r^m \cos m\varphi$, меняет знак при изменении k . Следовательно, квадрирование нужно прекращать, когда все коэффициенты, кроме $a_2^{(k)}$, будут приближенно равны квадратам коэффициентов предыдущего шага. Изменение знака $a_2^{(k)}$ указывает на присутствие комплексных сопряженных корней x_2 и x_3 . Величину аргумента φ можно определить после определения знаков вещественных корней, а также модуля комплексных корней по формуле

$$x_1 + x_2 + \dots + x_n = \frac{a_1}{a_0}$$

или

$$x_1 + 2r \cos \varphi + x_4 + \dots + x_n = \frac{a_1}{a_0}.$$

Случай 3. Вещественные, равные по абсолютной величине корни

Пусть уравнение (1.11) содержит два равных по абсолютной величине корня $|x_2|$ и $|x_3|$. Очевидно, в этом случае имеем

$|x_1| > |x_2| = |x_3| > \dots > |x_n|$, и после квадрирования получим связь между корнями и коэффициентами в виде

$$x_1^m + 2x_2^m + x_4^m + \dots + x_n^m = \frac{a_1^{(k)}}{a_0^{(k)}},$$

$$2x_1^m x_2^m + x_1^m x_4^m + \dots = \frac{a_2^{(k)}}{a_0^{(k)}},$$

$$x_1^m x_2^{2m} + \dots = \frac{a_3^{(k)}}{a_0^{(k)}},$$

.....

$$x_1^m x_2^{2m} x_4^m \dots x_n^m = \frac{a_n^{(k)}}{a_0^{(k)}},$$

или приближенно

$$\left. \begin{aligned} x_1^m &\approx \frac{a_1^{(k)}}{a_0^{(k)}}, \\ 2x_1^m x_2^m &\approx \frac{a_2^{(k)}}{a_0^{(k)}}, \\ x_1^m x_2^{2m} &\approx \frac{a_3^{(k)}}{a_0^{(k)}}, \\ &\dots \\ x_1^m x_2^{2m} \dots x_n^m &\approx \frac{a_n^{(k)}}{a_0^{(k)}}. \end{aligned} \right\} \quad (1.23)$$

После преобразования

$$x_1^m \cong \frac{a_1^{(k)}}{a_0^{(k)}}, \quad x_2^{2m} \cong \frac{a_3^{(k)}}{a_1^{(k)}}, \quad x_4^m \cong \frac{a_n^{(k)}}{a_3^{(k)}}, \quad \dots \quad x_n^m \cong \frac{a_n^{(k)}}{a_{n-1}^{(k)}}.$$

Эти равенства дают возможность найти абсолютное значение корней. При квадрировании (табл. 2) мы замечаем, что когда корни отделены, то все коэффициенты, кроме a_2 , станут приближенно равны квадратам соответствующих коэффициентов преды-

душего шага. Только коэффициент a_2 на соответствующем шаге вычисляется по формуле

$$a_2^{(k+1)} = \frac{1}{2} (a_2^{(k)})^2,$$

что можно легко заметить из формул (1.23). Табл. 2 является примером квадрирования.

Таблица 2

n	$a_0^{(k)}$	$a_1^{(k)}$	$a_2^{(k)}$	$\dots$	$a_n^{(k)}$
0	a_0	a_1	a_2	$\dots$	a_n
	a_0^2	a_1^2 $-2a_0 a_1$	a_2^2 $-2a_1 a_3$ $+2a_0 a_4$	$\dots$	a_n^2
1	$a_0^{(1)}$	$a_1^{(1)}$	$a_2^{(1)}$		$a_n^{(1)}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

Пример. Найти корни уравнения

$$1,23x^5 - 2,52x^4 - 16,1x^3 + 17,3x^2 + 29,4x - 1,34 = 0.$$

Результаты вычислений приведены в табл. 3.

На основании результатов квадрирования можно заключить, что все корни уравнения вещественны и различны. Эти корни определяются из уравнений:

$$7,54 \cdot 10^2 x_1^{32} - 2,346 \cdot 10^{22} = 0,$$

$$2,346 \cdot 10^{42} x_2^{32} - 3,95 \cdot 10^{37} = 0,$$

$$3,95 \cdot 10^{37} x_3^{32} - 8,744 \cdot 10^{46} = 0,$$

$$8,744 \cdot 10^{46} x_4^{32} - 2,148 \cdot 10^{47} = 0,$$

$$2,148 \cdot 10^{47} x_5^{32} - 1,175 \cdot 10^4 = 0.$$

Для определения знаков подставим в уравнение значения $\pm x_1$, $\pm x_2$, $\pm x_3$, $\pm x_4$, $\pm x_5$. Окончательно получим: $x_1 = 4,066$; $x_2 = -2,99$; $x_3 = 1,959$; $x_4 = -1,029$; $x_5 = 0,044$.

Таблица 3

m	$a_0^{(k)}$	$a_1^{(k)}$	$a_2^{(k)}$	$a_3^{(k)}$	$a_4^{(k)}$	$a_5^{(k)}$
0	1,23	-2,52	-16,1	17,3	29,4	-1,34
	1,513	$0,635 \cdot 10$ $3,961 \cdot 10$	$2,592 \cdot 10^2$ $0,872 \cdot 10^2$ $0,723 \cdot 10^2$	$2,993 \cdot 10^2$ $9,467 \cdot 10^2$ $0,068 \cdot 10^2$	$8,644 \cdot 10^2$ $0,464 \cdot 10^2$	1,796
2	$1,51 \cdot 10$	$4,596 \cdot 10$	$4,187 \cdot 10^2$	$1,253 \cdot 10^3$	$9,108 \cdot 10^2$	1,796
	2,289	$2,112 \cdot 10^3$ $-1,267 \cdot 10^3$	$1,753 \cdot 10^5$ $-1,152 \cdot 10^5$ $0,028 \cdot 10^5$	$1,570 \cdot 10^6$ $0,763 \cdot 10^6$ $0,000 \cdot 10^6$	$8,296 \cdot 10^5$ $-0,045 \cdot 10^5$	3,226
4	2,289	$0,845 \cdot 10^3$	$0,629 \cdot 10^5$	$0,807 \cdot 10^6$	$8,251 \cdot 10^5$	3,226
	5,240	$0,714 \cdot 10^6$ $-0,288 \cdot 10^6$	$0,396 \cdot 10^{10}$ $-0,136 \cdot 10^6$ $0 \cdot 10^6$	$0,651 \cdot 10^{12}$ $0,104 \cdot 10^{12}$ $0 \cdot 10^{12}$	$6,808 \cdot 10^{11}$ $0 \cdot 10^{11}$	10,41
8	5,240	$0,426 \cdot 10^6$	$0,260 \cdot 10^{10}$	$0,547 \cdot 10^{12}$	$0,6808 \cdot 10^{12}$	10,41
	$2,746 \cdot 10$	$1,815 \cdot 10^{11}$ $-0,272 \cdot 10^{11}$	$6,76 \cdot 10^{18}$ $0,47 \cdot 10^{18}$ —	$2,992 \cdot 10^{24}$ $0,035 \cdot 10^{24}$ —	$4,635 \cdot 10^{23}$ —	$1,084 \cdot 10^2$
16	$7,54 \cdot 10^2$	$2,38 \cdot 10^{22}$	$3,96 \cdot 10^{37}$	$-8,744 \cdot 10^{46}$	$2,148 \cdot 10^{47}$	$1,084 \cdot 10^2$
	$7,541 \cdot 10^2$	0,035 $2,346 \cdot 10^{22}$	0,01 $3,95 \cdot 10^{37}$	— $8,744 \cdot 10^{46}$	— $2,148 \cdot 10^{47}$	$1,175 \cdot 10^4$
32	$7,541 \cdot 10^2$	$2,346 \cdot 10^{22}$	$3,95 \cdot 10^{37}$	$8,744 \cdot 10^{46}$	$2,148 \cdot 10^{47}$	$1,175 \cdot 10^4$

§ 3. Алгебраическое интерполирование

1. Постановка задачи

Многие численные методы — численное дифференцирование, интегрирование, метод сеток и др. — основаны на замене заданной функции полиномом n -й степени, близким к данной функции в том или другом смысле.

Пусть функция $y = f(x)$ задана в точках $x_0, x_1, x_2, \dots, x_n$ значениями $y_0, y_1, \dots, y_n$. Потребуем, чтобы полином n -й степени $P_n(x)$ и функция $f(x)$ совпадали между собой в этих точках. Такое представление функции называется алгебраическим интерполированием. Точки, в которых вычисляется функция, — узлами интерполирования, а полином — интерполяционным. Запишем интерполяционный полином в виде

$$f(x) = a_0 + a_1x + \dots + a_nx^n.$$

Тогда указанное условие приведет к системе

$$\left. \begin{array}{l} a_0 + a_1 x_0 + \dots + a_n x_0^n = y_0, \\ a_0 + a_1 x_1 + \dots + a_n x_1^n = y_1, \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ a_0 + a_1 x_n + \dots + a_n x_n^n = y_n \end{array} \right\} \quad (1.24)$$

$(n+1)$ уравнения с $(n+1)$ -м неизвестным.

Система имеет единственное решение, так как ее основной определитель

$$\Delta = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix}$$

не равен нулю. Определить Δ носит имя Вандермонда.

Найдя из системы (1.24) коэффициенты $a_0, a_1, \dots, a_n$, определим полином $P_n(x)$ и можем представить функцию так:

$$f(x) = P_n(x) + R_n(x), \quad (1.25)$$

где $R_n(x)$ — погрешность или остаточный член интерполяционного полинома. Коэффициенты $a_0, a_1, \dots, a_n$ непосредственно из системы (1.24) не определяют. Будем отыскивать интерполяционный полином в другой форме.

2. Интерполяционный полином Лагранжа

Рассмотрим одну из основных форм интерполяционного полинома, принадлежащую Лагранжу. Для этого введем вспомогательные полиномы $Q_k^{(n)}(x)$ n -го порядка, такие, что все узлы интерполирования, кроме x_k , будут корнями каждого из полиномов $Q_k^{(n)}(x)$, который определим следующим образом:

$$Q_k^{(n)}(x) = \begin{cases} 1 & \text{при } i = k \\ 0 & \text{при } i \neq k, \end{cases}$$

т. е. в том узле, который не является корнем, полином обращается в единицу. Построим полиномы $Q_k^{(n)}(x)$ по их корням:

$$Q_k^{(n)}(x) = c(x-x_0) \dots (x-x_{k-1})(x-x_{k+1}) \dots (x-x_n).$$

Исходя из второго условия, получим

$$1 = c(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n),$$

$$c = \frac{1}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)}.$$

Тогда

$$Q_k^{(n)}(x) = \frac{(x-x_0) \dots (x-x_{k-1})(x-x_{k+1}) \dots (x-x_n)}{(x_k-x_0) \dots (x_k-x_{k-1})(x_k-x_{k+1}) \dots (x_k-x_n)}.$$

Вводя дополнительно полином $\omega(x)$, корнями которого являются все узлы интерполирования

$$\omega_{n+1}(x) = (x-x_0)(x-x_1) \dots (x-x_n),$$

мы можем представить

$$Q_k^{(n)}(x) = \frac{\omega(x)}{\omega'(x_k)(x-x_k)},$$

так как

$$\begin{aligned} \omega'_{n+1}(x) &= (x-x_1) \dots (x-x_n) + (x-x_0)(x-x_2) \dots (x-x_k) + \dots + \\ &+ (x-x_0) \dots (x-x_{n-1}); \end{aligned}$$

$$\omega'_{n+1}(x_k) = (x_k-x_0) \dots (x_k-x_{k-1})(x_k-x_{k+1}) \dots (x_k-x_n).$$

Используя вспомогательные полиномы, можно построить интерполяционный полином в форме Лагранжа $L_n(x) \equiv P_n(x)$

$$L_n(x) = \sum_{k=0}^n \frac{\omega_{n+1}(x)}{(x-x_k) \omega'_{n+1}(x_k)} y_k$$

или

$$L_n(x) = \sum_{k=0}^n \frac{(x-x_0) \dots (x-x_{k-1})(x-x_{k+1}) \dots (x-x_n)}{(x_k-x_0) \dots (x_k-x_{k-1})(x_k-x_{k+1}) \dots (x_k-x_n)} y_k. \quad (1.26)$$

Рассмотрим частные случаи для $n=0$, $n=1$, $n=2$:

$$L_0(x) = y_0;$$

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

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

Интересен также случай, когда промежутки между узлами интерполирования равны, т. е. $x_1-x_0=x_2-x_1=\dots x_n-x_{n-1}=h$.

Введем новую переменную $t = \frac{x-x_0}{h}$. Формула (1.26) примет вид

$$L_n(x) = L_n(x_0+th) = \sum_{k=0}^n (-1)^{n-k} \frac{t(t-1) \dots (t-n)}{k! (n-k)! (t-k)} y_k. \quad (1.27)$$

Интерполяционный полином Лагранжа используется при численном дифференцировании. Но при непосредственном интерполировании эта формула неудобна.

Конструкция формулы такова, что при добавлении лишнего узла интерполирования приходится пересчитывать все слагаемые формулы. Кроме того, значения слагаемых, соответствующих различным узлам интерполирования, мало отличаются друг от друга, и среди узлов, используемых в расчете, невозможно выделить главные и второстепенные.

Исходя из этих свойств, полином Лагранжа невыгодно применять при нахождении значений функции в точках, лежащих между узлами интерполирования. В этом случае удобна другая

форма интерполяционного полинома, где можно отличить главные узлы от второстепенных и решить, сколько нужно использовать узлов при данной точности вычисления. Для построения такого полинома введем новое понятие.

3. Разделенные и конечные разности

Пусть функция $y = f(x)$ задана в виде таблицы. Построим выражение

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

Обозначим его через $f(x_k, x_{k-1})$ и назовем разделенной разностью первого порядка. Построим теперь разность второго порядка

$$\frac{f(x_{k+1}, x_k) - f(x_k, x_{k-1})}{x_{k+1} - x_{k-1}} = f(x_{k+1}, x_k, x_{k-1}).$$

Аналогично можно построить разделенную разность n -го порядка

$$\frac{f(x_n, \dots, x_1) - f(x_{n-1}, \dots, x_0)}{x_n - x_0} = f(x_n, x_{n-1}, \dots, x_0).$$

Составим таблицу разделенных разностей при $n=4$:

x_0	$f(x_0)$				
x_1	$f(x_1)$	$f(x_1, x_0)$			
x_2	$f(x_2)$	$f(x_2, x_1)$	$f(x_2, x_1, x_0)$		
x_3	$f(x_3)$	$f(x_3, x_2)$	$f(x_3, x_2, x_1)$	$f(x_3, x_2, x_1, x_0)$	
x_4	$f(x_4)$	$f(x_4, x_3)$	$f(x_4, x_3, x_2)$	$f(x_4, x_3, x_2, x_1)$	$f(x_4, x_3, x_2, x_1, x_0)$

Таблица разделенных разностей треугольная. Для функции, заданной таблицей, содержащей $n+1$ значение, можно построить разделенные разности от первого до n -го порядка. Разделенные разности симметричны относительно своих аргументов. Действительно

$$f(x_k, x_{k-1}) = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} = \frac{f(x_{k-1}) - f(x_k)}{x_{k-1} - x_k} = f(x_{k-1}, x_k).$$

Найдем выражение разделенной разности любого порядка непосредственно через значения функции $f(x_0), f(x_1), \dots$. Для этого

рассмотрим сначала разделенные разности первого и второго порядка

$$\begin{aligned} f(x_1, x_0) &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0}; \\ f(x_2, x_1, x_0) &= \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0} = \frac{f(x_2) - f(x_1)}{(x_2 - x_0)(x_2 - x_1)} - \\ &- \frac{f(x_1) - f(x_0)}{(x_2 - x_0)(x_1 - x_0)} = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)(x_1 - x_0 - x_1 + x_2)}{(x_1 - x_0)(x_2 - x_0)(x_1 - x_2)} + \\ &+ \frac{f(x_2)}{(x_2 - x_1)(x_2 - x_0)} = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \\ &+ \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)} = \sum_{k=0}^2 \frac{f(x_k)}{\omega'_3(x_k)}. \end{aligned}$$

По аналогии можно записать

$$\begin{aligned} f(x_n, x_{n-1}, \dots, x_0) &= \\ &= \sum_{k=0}^n \frac{f(x_k)}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1})(x_k - x_{k+2}) \dots (x_k - x_n)} \end{aligned}$$

или

$$f(x_n, x_{n-1}, \dots, x_0) = \sum_{k=0}^n \frac{f(x_k)}{\omega'_{n+1}(x_k)}.$$

Рассмотрим случай, когда расстояния $x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1}$ равны между собой. Обозначим каждое из них через h . Выражение $f(x_1) - f(x_0)$ обозначим через Δy_0 и назовем конечной разностью первого порядка. Конечные разности второго и более высоких порядков записываются следующим образом:

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0,$$

$$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0,$$

$$\dots \dots \dots$$

$$\Delta^n y_0 = \Delta^{n-1} y_1 - \Delta^{n-1} y_0.$$

Составим таблицу конечных разностей для функции, заданной пятью значениями:

x_0	y_0	Δy_0			
x_1	y_1	Δy_1	$\Delta^2 y_0$		
x_2	y_2	Δy_2	$\Delta^2 y_1$	$\Delta^3 y_0$	
x_3	y_3	Δy_3	$\Delta^2 y_2$	$\Delta^3 y_1$	$\Delta^4 y_0$
x_4	y_4				

Легко установить связь между конечными и разделенными разностями для случая равных промежутков

$$f(x_1, x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{\Delta y_0}{1! h};$$

$$f(x_2, x_1, x_0) = \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0} = \frac{\Delta y_1 - \Delta y_0}{1! h \ 2h} = -\frac{\Delta^2 y_0}{2! h^2}$$

и аналогично получить формулу для разделенной разности n -го порядка

$$f(x_n, x_{n-1}, \dots, x_0) = \frac{\Delta^n y_0}{n! h^n}.$$

4. Интерполяционный полином Ньютона

Используя понятие разделенных разностей, можно получить интерполяционные полиномы в самых различных формах, приспособленных для решения задач интерполирования. Кроме того, можно определить остаточный член, погрешность, полученную при представлении функции в виде полинома.

Рассмотрим тождество

$$L_n(x) = L_0(x) + [L_1(x) - L_0(x)] + [L_2(x) - L_1(x)] + \dots + [L_n(x) - L_{n-1}(x)]. \quad (1.28)$$

Вычислим выражения, стоящие в скобках

$$L_1(x) - L_0(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 - y_0 =$$

$$= y_0 \frac{x - x_0}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0} = f(x_1, x_0) (x - x_0),$$

$$L_2(x) - L_1(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 +$$

$$+ \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2 - \frac{x - x_1}{x_0 - x_1} y_0 - \frac{x - x_0}{x_1 - x_0} y_1 =$$

$$= f(x_0, x_1, x_2) (x - x_0) (x - x_1),$$

.....

$$L_n(x) - L_{n-1}(x) = (x - x_0) \dots (x - x_{n-1}) f(x_0, x_1, \dots, x_n).$$

Подставив полученное в формулу (1.28), имеем

$$L_n(x) = f(x_0) + f(x_0, x_1) (x - x_0) + \dots + (x - x_0) \dots (x - x_{n-1}) f(x_0, \dots, x_n). \quad (1.29)$$

Формула (1.29) определяет полином Ньютона, который запишем для равных промежутков, используя переменную t , введенную выше, а также формулы, связывающие конечные и разделенные разности. После простых преобразований получим

$$P_n(x_0 + ht) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!} \Delta^2 y_0 + \dots + \\ + \frac{t(t-1) \dots (t-n+1)}{n!} \Delta^n y_0.$$

Интерполяционный полином Ньютона обладает рядом свойств, необходимых для нахождения значения функции в точках промежутка (x_0, x_1) . Конструкция полинома такова, что слагаемые убывают с увеличением номера, причем наибольшие слагаемые — первое и второе — включают в себя элементы нулевого и первого узлов. При расчете останавливаются на той конечной разности, величина которой имеет порядок заданной погрешности расчета. Кроме того, очень легко включить в расчет дополнительное слагаемое.

При этом достаточно вычислить одно дополнительное слагаемое, остальные слагаемые не изменятся.

Найдем погрешность интерполяционного полинома. Для этого рассмотрим разделенную разность $f(x, x_0, x_1, \dots, x_n)$ и сделаем простейшие преобразования

$$f(x, x_0, \dots, x_n) = - \frac{f(x_0, \dots, x_n)}{x - x_n} + \frac{f(x, x_0, \dots, x_{n-1})}{x - x_n} = \\ = - \frac{f(x_0, x_1, \dots, x_n)}{x - x_n} - \frac{f(x_0, x_1, \dots, x_{n-1})}{(x - x_n)(x - x_{n-1})} - \dots - \\ - \frac{f(x_0, x_1)}{(x - x_n) \dots (x - x_1)} - \frac{f(x_0)}{\omega_{n+1}(x)} + \frac{f(x)}{\omega_{n+1}(x)}.$$

Отсюда получим

$$f(x) = f(x_0) + f(x_0, x_1)(x - x_0) + \dots + (x - x_0) \dots (x - x_{n-1}) \times \\ \times f(x_0 - x_n) + (x - x_0) \dots (x - x_n) f(x, x_0, \dots, x_n).$$

Выражение

$$\omega_{n+1}(x) f(x, x_0, \dots, x_n) = R_n(x)$$

является погрешностью или остаточным членом интерполяционного полинома Ньютона. Исходя из однозначности решения задачи интерполирования, можно заметить, что полученная формула определяет погрешность любого интерполяционного полинома.

Можно показать, что в промежутке (x_0, x_n) найдется такая точка ξ , при которой справедливо равенство

$$R_n(x) = \omega_{n+1}(x) \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Пример. Используя формулу Ньютона, найти значение функции $y=f(x)$ в точке $x=0,1265$, если функция задана табл. 4.

Таблица 4

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0,1	1,656832				
0,2	1,694888	0,038056			
0,3	1,728606	0,033718	-0,004338		
0,4	1,758030	0,029424	-0,004294	0,000044	
0,5	1,783225	0,025195	-0,004229	0,000065	0,000021
0,6	1,804279	0,021054	-0,004141	0,000088	0,000023

$$f(x) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!} \Delta^2 y_0 + \frac{t(t-1)(t-2)}{3!} \Delta^3 y_0 + \\ + \frac{t(t-1)(t-2)(t-3)}{4!} \Delta^4 y_0;$$

$$t = \frac{x - x_0}{h} \cong 0,2654;$$

$$f(0,1265) \cong 1,6685.$$

§ 4. Численное дифференцирование

Пусть функция $y=f(x)$ представлена с помощью интерполяционного полинома Лагранжа

$$f(x) = P_n(x) + R_n(x),$$

где остаточный член $R_n(x)$ записан в любом виде. Для того чтобы найти производную любого порядка от функции $f(x)$ дифференцируют интерполяционный полином $P_n(x)$. В зависимости от

степени полинома n мы получаем целый ряд формул для производной любого порядка

$$f^{(k)}(x) \cong P_n^{(k)}(x).$$

Погрешность можно вычислить, дифференцируя остаточный член $R_n(x)$. К сожалению, вид остаточного члена не дает возможности отыскать, точно его производную, так как включает значения функции, вычисленной в точке ξ , положение которой в промежутке (x_0, x_1) не определено.

Покажем на примерах, как пользоваться другим способом определения погрешности при численном дифференцировании. Этот способ предполагает использование рядов Тейлора.

Пример 1. Пусть функция $y=f(x)$ задана в точках x_0 и x_1 , как y_0 и y_1 . Ее интерполяционный полином имеет вид

$$L_1(x) = (t-1)y_0 + ty_1; \quad t = \frac{x-x_0}{h}; \quad h = x_1 - x_0.$$

Тогда

$$y'_x = L'_1 t'_x = \frac{y_1 - y_0}{h}.$$

Очевидно мы получили y'_x постоянной в нашем интервале, и значения ее в точках x_0 и x_1 , которые мы обозначим: $y'_0 = y'(x_0)$, $y'_1 = y'(x_1)$ определяются формулами

$$y'_0 = \frac{y_1 - y_0}{h}; \quad (1.30)$$

$$y'_1 = \frac{y_1 - y_0}{h} = \frac{y_0 - y_1}{-h}. \quad (1.31)$$

Формула (1.30) носит название формулы для дифференцирования «вперед», а формула (1.31) — «назад».

Для определения погрешности формул (1.30) и (1.31) разложим функцию $y_1 = f(x_1)$ в ряд Тейлора в окрестности точки x_0 . Получим

$$y_1 = y_0 + hy'_0 + \frac{h^2}{2!} y''(\xi).$$

Отсюда

$$\frac{y_1 - y_0}{h} = y'_0 + \frac{h}{2} y''(\xi)$$

или

$$y'_0 = \frac{y_1 - y_0}{h} - \frac{h}{2} y''(\xi), \quad x_0 < \xi < x_1.$$

Величина $-\frac{h}{2}y''(\xi)$ является погрешностью формулы (1.30).

Для отыскания погрешности формулы (1.31) следует разложить по формуле Тейлора функцию $y(x_0)=y_0$ в окрестности точки x_1 и аналогично получить

$$y'_1 = \frac{y_1 - y_0}{h} + \frac{h}{2} y''(\xi).$$

Пример 2. Рассмотрим теперь функцию, определенную в трех точках x_0, x_1, x_2 и принимающую в них значения y_0, y_1 и y_2 . Эта функция может быть представлена в виде полинома

$$y \cong L_2(x_0 + ht) = \frac{(t-1)(t-2)}{2} y_0 - t(t-2) y_1 + \frac{t(t-1)}{2} y_2.$$

Дифференцируем построенный интерполяционный полином два раза:

$$y'_x \cong \frac{1}{h} \left[\frac{(2t-3)}{2} y_0 - 2(t-1) y_1 + \frac{(2t-1)}{2} y_2 \right];$$

$$y''_x \cong \frac{1}{h^2} (y_0 - 2y_1 + y_2).$$

Вычислим данные производные в средней точке x_1

$$y'_1 \cong \frac{y_2 - y_0}{2h}; \quad (1.32)$$

$$y''_1 \cong \frac{y_0 - 2y_1 + y_2}{h^2}. \quad (1.33)$$

Найдем погрешность формул (1.32) и (1.33). Для этого разложим функции $y_0=y(x_0)$ и $y_2=y(x_2)$ по формуле Тейлора в окрестности точки x_1

$$y_2 = y_1 + hy'_1 + \frac{h^2}{2!} y''_1 + \frac{h^3}{3!} y'''_1 + \dots$$

$$y_0 = y_1 - hy'_1 + \frac{h^2}{2!} y''_1 - \frac{h^3}{3!} y'''_1 + \dots$$

Отсюда

$$y'_1 = \frac{y_2 - y_0}{2h} - y'''(\xi_1) \frac{h^2}{3!}, \quad x_0 < \xi_1 < x_2;$$

$$y''_1 = \frac{y_0 - 2y_1 + y_2}{h^2} - y^{IV}(\xi_2) \frac{h^2}{12}, \quad x_0 < \xi_2 < x_2.$$

По аналогии можно построить формулу для производной любого порядка с любой степенью точности.

Пример 3. Отыщем численные производные в случае, если рассматривается функция двух переменных и точки, в которых функция определена, лежат не на осях координат.

Для прогноза перемещения приземных циклов выведены формулы для определения скорости и направления центра циклона. В эти формулы входят производные $\frac{\partial u}{\partial x}$ и $\frac{\partial u}{\partial y}$ от функции

$u = u(x, y)$, заданных в точках, лежащих на окружности радиуса h и в центре окружности (рис. 4).

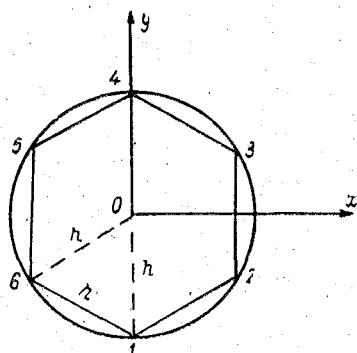


Рис. 4. Треугольная сетка

Нужно вычислить указанные производные в центре окружности в точке (x_0, y_0) . Применять интерполяционные формулы в нашем случае неудобно, поэтому сразу же применим метод, определяющий погрешность, и разложим функции u_1, u_3, u_4, u_5, u_6 , где $u_h = u(x_h, y_h)$ в ряды Тейлора в окрестности точки (x_0, y_0)

$$u_1 = u_0 - hu'_{y0} + \frac{h^2}{2} u''_{yy0} + \dots$$

$$u_4 = u_0 + hu'_{y0} + \frac{h^2}{2} u''_{yy0} + \dots$$

$$u_3 = u_0 + h \left(\frac{\sqrt{3}}{2} u'_{x0} + \frac{1}{2} u'_{y0} \right) + \frac{h^2}{2} \left(\frac{3}{4} u''_{xx0} + \frac{\sqrt{3}}{2} u''_{xy0} + \frac{1}{4} u''_{yy0} \right) + \dots$$

$$u_5 = u_0 + h \left(-\frac{\sqrt{3}}{2} u'_{x0} + \frac{1}{2} u'_{y0} \right) + \frac{h^2}{2} \left(\frac{3}{4} u''_{xx0} - \frac{\sqrt{3}}{2} u''_{xy0} + \frac{1}{4} u''_{yy0} \right) + \dots$$

$$u_2 = u_0 + h \left(\frac{\sqrt{3}}{2} u'_{x0} - \frac{1}{2} u'_{y0} \right) + \frac{h^2}{2} \left(\frac{3}{4} u''_{xx0} - \frac{\sqrt{3}}{2} u''_{xy0} + \frac{1}{4} u''_{yy0} \right) + \dots$$

$$u_6 = u_0 + h \left(-\frac{\sqrt{3}}{2} u'_{x0} - \frac{1}{2} u'_{y0} \right) + \frac{h^2}{2} \left(\frac{3}{4} u''_{xx0} + \right. \\ \left. + \frac{\sqrt{3}}{2} u''_{xy0} + \frac{1}{4} u''_{yy0} \right) + \dots$$

Комбинируя полученные выражения, мы можем построить искомые формулы с разной степенью точности.

Приведем формулы, определяющие производные с погрешностью порядка h^2 :

$$u'_{x0} = \frac{1}{2\sqrt{3}h} (u_3 - u_5 + u_2 - u_6) + O(h^2);$$

$$u'_{y0} = \frac{1}{6h} (u_4 - u_1 + 2u_3 - 2u_2 + 2u_5 - 2u_6 + O(h^2)).$$

Рекомендуем читателям самим построить формулы с погрешностью порядка h .

§ 5. Численное интегрирование

1. Формулы Ньютона—Котеса

Рассмотрим определенный интеграл

$$\int_a^b f(x) dx,$$

где $f(x)$ — непрерывная функция, заданная в промежутке (a, b) . Представим $f(x)$ с помощью ее интерполяционного полинома $L_n(x)$, принимая за узлы интерполирования точки $x_h = a + hk$, где $h = \frac{b-a}{n}$. При этом перейдем к переменной t :

$$\int_a^b f(x) dx = \int_0^n t(t-1) \dots (t-n) \sum_{k=0}^n \frac{f(x_h) (-1)^{n-k}}{(t-k)(n-k)! k!} h dt. \quad (1.34)$$

В формуле (1.34) переменим порядок суммирования и интегрирования и введем обозначения

$$H_k^{(n)} = \int_0^n \frac{t(t-1) \dots (t-n)}{t-k} \frac{(-1)^{n-k}}{(n-k)! k! n} dt.$$

Величины $H_k^{(n)}$ называются коэффициентами Котеса. В указанных обозначениях формула (1.34) примет вид

$$\int_a^b f(x) dx = (b-a) \sum_{k=0}^n H_k^{(n)} f(x_k).$$

Погрешность формул Ньютона—Котеса может быть получена интегрированием остаточного члена интерполяционного полинома. Трудности здесь будут, так же как и при численном дифференцировании, связаны с множителем $f(\xi)$, входящим в состав остаточного члена, так как положение точки ξ не может быть определено.

Разберем простейшие случаи.

1. При $n=1$, $H_0^{(1)} = H_1^{(1)} = \frac{1}{2}$

$$\int_{x_0}^{x_1} f(x) dx = [f(x_0) + f(x_1)] \frac{h}{2}.$$

Погрешность вычисляется с помощью интегрирования $R_1(x_0+ht)$

$$r_1 = \int_0^1 R_1(x_0+ht) h dt = \int_0^1 f''(\xi) \frac{t(t-1)}{2} h^3 dt = f''(\xi_1) \frac{h^3}{3},$$

где ξ_1 — точка, полученная по теореме о среднем.

2. Для случая $n=2$ формула Ньютона — Котеса примет вид

$$\int_{x_0}^{x_2} f(x) dx = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)].$$

Погрешность дадим в готовом виде из-за трудностей расчета

$$r_2 = \frac{h^5}{90} f^{IV}(\xi_2),$$

где точка ξ_2 лежит в промежутке (x_0, x_2) .

Рекомендуем читателям самим построить формулу Ньютона—Котеса при $n=3$.

Приведем теперь погрешности формул Ньютона—Котеса при $n=1, 2, \dots, 6$ и обозначим их через r_n :

$$\begin{aligned} r_1 &= c_1 h^3 f''(\xi_1), & r_4 &= c_4 h^7 f^{VI}(\xi_4) \\ r_2 &= c_2 h^5 f^{IV}(\xi_2), & r_5 &= c_5 h^7 f^{VI}(\xi_5), \\ r_3 &= c_3 h^5 f^{IV}(\xi_3), & r_6 &= c_6 h^9 f^{VIII}(\xi_6). \end{aligned}$$

Эти выражения показывают, что у двух соседних формул — предыдущей с четным номером и последующей с нечетным — погрешности совпадают с точностью до множителя c_h . Так как формула с меньшим номером содержит меньше слагаемых, то выгоднее использовать формулы с четными номерами. Включенный в оценку погрешности множитель $f^{(k)}(\xi_i)$ позволяет получить новый подход к определению точности формул численного интегрирования. Будем говорить, что формула имеет алгебраическую точность порядка n , если при интегрировании с ее помощью полинома n -й степени получаем не приближенный, а точный результат. Очевидно, формулы Ньютона—Котеса должны иметь алгебраическую точность, совпадающую с их номером, так как они построены с помощью точного интегрирования соответствующих интерполяционных полиномов. Но, рассматривая погрешности формул с четными значениями n , замечаем, что они содержат производные, порядок которых выше номера формулы на две единицы. Если предположить, что формула точна, то ее остаточный член равен нулю, а это может быть только тогда, когда $f^{(k)}(x) = 0$. Такой результат справедлив в случае, если $f(x)$ является полиномом $(k-1)$ -й степени. Следовательно, для четных номеров алгебраическая точность оказывается на единицу выше предполагаемой.

В практике применяются две первые формулы Ньютона—Котеса. Чтобы уменьшить погрешность, разбивают сначала промежуток (a, b) на m частей, к каждой части применяют соответствующую формулу, а результат суммируют.

Первая называется формулой трапеций

$$\int_a^b f(x) dx = \frac{h}{2} \{ f(x_0) + f(x_m) + 2[f(x_1) + \dots + f(x_{m-1})] \}.$$

За погрешность принимается сумма погрешностей для каждого промежутка

$$\frac{mM_2}{12} h^3 = \frac{(b-a)^3 M_2}{12m^2},$$

где $M_2 = \max_{a \leq x \leq b} |f''(x)|$.

Вторая — формула Симпсона

$$\int_a^b f(x) dx = \frac{h}{3} \{ f(x_0) + f(x_{2m}) + 2[f(x_2) + f(x_4) + \dots + f(x_{2m-2})] + 4[f(x_1) + f(x_3) + \dots + f(x_{2m-1})] \}.$$

В этой формуле $2m$ точек, так как каждый малый промежуток длины $\frac{b-a}{m}$ содержит дополнительную точку, делящую его пополам.

Погрешность формулы Симпсона оценивают по формуле

$$\frac{mh^5}{90} M_4 = (b-a) \frac{h^4}{180m} M_4,$$

где $M_4 = \max_{a \leq x \leq b} |f^{IV}(x)|$.

Пример. Вычислить с помощью формулы Симпсона интеграл

$$I = \int_{2,2310}^{3,3420} \frac{(x+1)^3}{(0,985x^2 + 0,763x + 1,25)^2} dx.$$

Примем $m=5$, $h=0,1111$.

Приведем решение в виде табл. 5.

Таблица 5

k	x_k	$y_k = f(x_k)$	q_k	$y_k q_k$
0	2,2310	0,54667	1	0,54667
1	2,3421	0,52403	4	2,09612
2	2,4532	0,50280	2	1,00560
3	2,5643	0,48396	4	1,93584
4	2,6754	0,46422	2	0,92844
5	2,7865	0,44670	4	1,78686
6	2,8976	0,43025	2	0,86050
7	3,0087	0,41475	4	1,65900
8	3,1198	0,40025	2	0,80050
9	3,2309	0,38655	4	1,54620
10	3,3420	10,37365	1	10,37365

$$I = 0,50142.$$

2. Квадратурные формулы типа Гаусса

Квадратурные формулы типа Гаусса строятся исходя из требования получить формулу наивысшей алгебраической точности. Записывая приближенно квадратурную формулу

$$\int_a^b f(x) dx \cong \sum_{k=1}^n A_k f(x_k),$$

мы будем иметь $2n$ параметров A_k и x_k . Априори можно предположить, что наивысшая алгебраическая точность подобной формулы должна быть $(2n-1)$, так как $2n$ параметров могут определить полином $(2n-1)$ -й степени.

Введем некоторые новые понятия и докажем ряд теорем, которые позволят наипростейшим образом построить искомые квадратурные формулы.

Рассмотрим определенный интеграл вида

$$\int_a^b p(x)f(x)dx,$$

где $p(x)$ называется весовой функцией и удовлетворяет следующим условиям: 1) $\int_a^b f(x)dx$ — существует; 2) $p(x) > 0$ во всем промежутке интегрирования.

Кроме того, будем иметь в виду, что $\int_a^b p(x)f(x)dx$ существует.

Определение. Формула численного интегрирования

$$\int_a^b p(x)f(x)dx = \sum_{k=1}^n A_k^{(n)} f(x_k) \quad (1.35)$$

называется интерполяционной, если коэффициенты $A_k^{(n)}$ определяются следующей формулой:

$$A_k^{(n)} = \int_a^b \frac{\omega_n(x)p(x)}{\omega_n'(x_k)(x-x_k)} dx, \quad (1.36)$$

где $\omega_n(x) = (x-x_1) \dots (x-x_n)$ — полином n -й степени.

Определение делается понятным, если представить себе, что в этом случае функция $f(x)$ заменена приближенно интерполяционным полиномом $(n-1)$ -й степени

$$L_{n-1}(x) = \sum_{k=1}^n \frac{\omega_n(x)}{\omega_n'(x_k)(x-x_k)} f(x_k).$$

Замечание: в обозначениях этого параграфа полином $\omega_n(x)$ рассматривается как полином степени n -й, в то время как в предыдущем параграфе в него входит еще один множитель: $(x-x_0)$.

Докажем теперь ряд вспомогательных теорем.

Теорема 1. Для того чтобы формула численного интегрирования была интерполяционной, необходимо и достаточно, чтобы она была точной для полинома $(n-1)$ -й степени.

Доказательство.

Необходимость. Доказательство очевидно, так как в этом случае коэффициенты $A_k^{(n)}$ определяются формулой (1.36) и, следовательно, функция $f(x)$ представлена полиномом $(n-1)$ -й степени.

Достаточность. Пусть формула (1.35) точна для полинома $(n-1)$ -й степени. Применим ее к полиному

$$Q_{n-1}^{(m)}(x) = \frac{\omega_n(x)}{\omega_n'(x_m)(x-x_m)}.$$

Очевидно, $Q_{n-1}^{(m)}(x)$ обращается в ноль во всех узлах интерполирования, исключая x_m

$$Q_{n-1}^{(m)}(x_k) = \begin{cases} 0, & k \neq m \\ 1, & k = m. \end{cases}$$

Следовательно,

$$\int_a^b Q_{n-1}^{(m)}(x) p(x) dx = \sum_{k=1}^n A_k^{(n)} Q_{n-1}^{(m)}(x_k) = A_m^{(n)} Q_{n-1}^{(m)}(x_m) = A_m^{(n)}$$

или, подставив значение $Q_{n-1}^{(m)}$, получим

$$A_m^{(n)} = \int_a^b \frac{p(x) \omega_n(x)}{\omega_n'(x_m)(x-x_m)} dx.$$

Таким образом, формула (1.35) интерполяционная.

Замечание. Формулы Ньютона—Котеса, описанные в п. 1, являются интерполяционными.

Теорема 2. Для того чтобы формула (1.35) — (1.36) имела алгебраическую точность $(2n-1)$, необходимо и достаточно, чтобы полином $\omega_n(x)$ был ортогонален любому полиному $Q_{n-1}(x)$ с весом $p(x)$ в промежутке (a, b) :

$$\int_a^b p(x) Q_{n-1}(x) \omega_n(x) dx = 0. \quad (1.37)$$

Доказательство. Необходимость. Пусть формула (1.35) — (1.36) точна для полинома $(2n-1)$ -й степени. Построим такой полином:

$$\omega_n(x) Q_{n-1}(x) = P_{2n-1}(x),$$

где $\omega_n(x) = (x-x_1) \dots (x-x_n)$; $Q_{n-1}(x)$ произвольный полином $(n-1)$ -й степени.

Проинтегрируем функцию $P_{2n-1}(x)$ в промежутке (a, b) с помощью формул (1.35), (1.36):

$$\int_a^b p(x) Q_{n-1}(x) \omega_n(x) dx = \sum_{k=1}^n A_k^{(n)} \omega_n(x_k) Q_{n-1}(x_k) = 0.$$

Полученный результат доказывает теорему.

Достаточность. Пусть справедлива формула (1.37). Покажем, что при этом для любого полинома $(2n-1)$ -й степени $P_{2n-1}(x)$ формула (1.35) — (1.36) точна. Представим полином $P_{2n-1}(x)$ в виде

$$P_{2n-1}(x) = \omega_n(x) Q_{n-1}(x) + R_{n-1}(x). \quad (1.38)$$

Формулу (1.38) получили, разделив полином $P_{2n-1}(x)$ на полином $\omega_n(x)$. Частное от деления $Q_{n-1}(x)$, остаток $R_{n-1}(x)$ — полиномы $(n-1)$ -й степени. Домножим левую и правую части равенства (1.38) на $p(x)$:

$$P_{2n-1}(x)p(x) = \omega_n(x)Q_{n-1}(x)p(x) + R_{n-1}(x)p(x). \quad (1.39)$$

Проинтегрируем (1.39) в промежутке (a, b) :

$$\int_a^b P_{2n-1}(x)p(x)dx = \int_a^b \omega_n(x)Q_{n-1}(x)p(x)dx + \int_a^b R_{n-1}(x)p(x)dx.$$

Первое слагаемое правой части по условию теоремы равно нулю. Следовательно, имеем

$$\int_a^b P_{2n-1}(x)p(x)dx = \int_a^b R_{n-1}(x)p(x)dx. \quad (1.40)$$

Вычислим теперь левую и правую часть формулы (1.39) в точках x_k , домножим каждое из полученных равенств на $A_k^{(n)}$ и результаты сложим

$$\sum_{k=1}^n A_k^{(n)} P_{2n-1}(x_k) = \sum_{k=1}^n A_k^{(n)} \omega_n(x_k) Q_{n-1}(x_k) + \sum_{k=1}^n A_k^{(n)} R_{n-1}(x_k).$$

Так как x_k — корни полинома $\omega_n(x)$, то первая сумма правой части равна нулю. Отсюда

$$\sum_{k=1}^n A_k^{(n)} P_{2n-1}(x_k) = \sum_{k=1}^n A_k^{(n)} R_{n-1}(x_k). \quad (1.41)$$

Сравнивая формулы (1.40) и (1.41), учтем, что по теореме 1 правые части формул равны между собой. Значит, равны и левые части

$$\int_a^b P_{2n-1}(x)p(x)dx = \sum_{k=1}^n A_k^{(n)} P_{2n-1}(x_k).$$

и теорема доказана.

Используя эти теоремы, можно по данному весу подобрать нужные нам полиномы $\omega_n(x)$, т. е. найти их корни, которые будут служить точками дробления интервала (a, b) , и вычислить $A_k^{(n)}$ по формуле (1.36). Для практического выполнения этой задачи чаще всего рассматривается промежуток $(-1, +1)$. Целый ряд формул типа Гаусса (1.35), (1.36) рассмотрен в книге по численному интегрированию В. И. Крылова [12]. Приведем здесь только две для примера.

Самым простым частным случаем является $p(x) \equiv 1$ и $(a, b) = (-1, +1)$

$$\int_{-1}^1 f(z) dz = \sum_{k=1}^n A_k^{(n)} f(z_k); \quad (1.42)$$

$$A_k^{(n)} = \int_{-1}^1 \frac{\omega_n(z)}{\omega_n'(z_k)(z-z_k)} dz. \quad (1.43)$$

Формулы (1.42), (1.43) называются формулами Гаусса. Предполагается, что подинтегральная функция $f(z)$ не имеет особенностей на промежутке (a, b) . В этом случае за промежутки (a, b) принимают промежуток $(-1, +1)$ и используют полиномы Лежандра, которые обладают свойствами, указанными в теореме 2. Корни полинома Лежандра n -й степени вещественны, различны и лежат в промежутке $(-1, +1)$. Полином Лежандра ортогонален любому полиному $(n-1)$ -й степени и ниже в промежутке $(-1, +1)$.

Корни полинома Лежандра при $n = 2, 3, 4$ и соответственные значения коэффициентов сведем в табл. 6.

Таблица 6

n	k	$z_k = z_{n+1-k}$	$A_k^{(n)} = A_{n+1-k}^{(n)}$
2	1	0,5774	1,0000
3	1	0,7746	0,5556
	2	0,0000	0,8889
4	1	0,8611	0,3479
	2	0,3400	0,6521

Очевидно, легко с помощью замены переменных перейти от промежутка $(-1, +1)$ к произвольному промежутку (a, b)

$$\begin{aligned} \int_a^b f(x) dx &= \frac{b-a}{2} \int_{-1}^1 f\left(\frac{b-a}{2}z + \frac{b+a}{2}\right) dz = \\ &= \frac{b-a}{2} \sum_{k=1}^n A_k^{(n)} f\left(\frac{b-a}{2}z_k + \frac{b+a}{2}\right). \end{aligned}$$

В качестве второго примера рассмотрим интеграл вида $\int_0^1 x^\alpha f(x) dx$, где α — целое. При этом за $\omega_n(x)$ примем полиномы Якоби.

Значения x_k и $A_k^{(n)}$ для $n = n, 3, 4$ приведены в табл. 7.

Таблица 7

n	k	x_k	$A_k^{(n)}$
2	1	0,3505	0,1820
	2	0,8449	0,3180
3	1	0,2134	0,0698
	2	0,5905	0,2292
	3	0,9114	0,2009
4	1	0,1398	0,0312
	2	0,4164	0,1298
	3	0,7232	0,2035
	4	0,9429	0,1355

Запишем без вывода погрешность формул типа Гаусса

$$r_n = \frac{f^{(2n)}(\xi)}{2n!} \int_a^b p(x) \omega^2(x) dx.$$

Пример. Вычислить методом Гаусса интеграл

$$I = \int_0^{1,047} e^{\frac{x}{2}} (x+1,25)^{-0,5} dx.$$

Все результаты вычислений представим в виде табл. 8.

Таблица 8

k	z_k	x_k	$f(x_k)$	$A_k^{(n)}$	$A_k^{(n)} \frac{f(x_k)}{2}$
1	-0,949118	0,0266421	0,8096905	0,129484	0,058068
2	-0,741531	0,135308	0,909097	0,279706	0,127140
3	-0,405845	0,311040	0,935053	0,331830	0,178516
4	0	0,523500	0,979982	0,417960	0,204796
5	0,405845	0,735910	1,025233	0,381830	0,195732
6	0,791531	0,911692	1,073001	0,279706	0,150062
7	0,949108	1,0203581	1,105410	0,129484	0,071567

$$I \cong 1,032217.$$

§ 6. Численное решение задачи Коши для обыкновенных дифференциальных уравнений

Рассмотрим задачу Коши для уравнения первого порядка

$$y' = f(x, y); \quad (1.44)$$

$$y|_{x=x_0} = y_0 \quad (1.45)$$

Покажем, что задача Коши для уравнений n -го порядка сведется к задаче Коши для системы n уравнений.

Пусть имеется следующая задача Коши:

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)});$$

$$y|_{x=x_0}=y_0, \quad y'|_{x=x_0}=y'_0, \quad \dots, \quad y^{(n-1)}|_{x=x_0}=y_0^{(n-1)}.$$

Введем новые функции по формулам

$$y' = y_1, \quad y'' = y_2, \quad \dots, \quad y^{(n-1)} = y_{n-1}.$$

Для новых функций получим систему уравнений

$$\left. \begin{aligned} y'_{n-1} &= f(x, y, y_1, \dots, y_{n-1}), \\ y'_{n-2} &= g_{n-1}, \\ &\vdots \\ y'_1 &= u_1 \end{aligned} \right\} \quad (1.46)$$

со следующими начальными условиями:

$$y|_{x=x_0}=y_0, \quad y_1|_{x=x_0}=y'_0, \quad \dots, \quad y_{n-1}|_{x=x_0}=y_0^{(n-1)}. \quad (1.47)$$

Таким образом, мы будем изучать только задачу Коши (1.46), (1.47) или ее простейший вид (1.44), (1.45).

1. Метод Эйлера

Вернемся к задаче (1.44), (1.45) и опишем простейший численный метод. Рассмотрим тот промежуток (x_0, x_n) , где нужно искать решение, и разобьем этот промежуток на n частей

$$\frac{x_n - x_0}{n} = h.$$

Примем, что на промежутке h искомая интегральная кривая $y = \varphi(x)$ может быть заменена касательной, проведенной через крайнюю левую точку промежутка. Для первого промежутка (x_0, x_1) уравнение касательной имеет вид

$$y - y_0 = (x - x_0)f(x_0, y_0).$$

Обозначим $f(x_0, y_0)$ через f_0 . Получим координату y_1 , точки пересечения касательной и прямой $x=x_1$

$$y_1 = y_0 + hf_0.$$

Приняв эту точку за точку интегральной кривой, получим следующие по рекуррентной формуле

$$y_k = y_{k-1} + hf_{k-1}.$$

Таким образом, интегральную кривую заменяет ломаная с вершинами $M_k(x_k, y_k)$.

Описанный метод дает очень грубое приближенное решение задачи Коши и используется в тех случаях, когда надо получить примерное представление о решении на небольшом участке.

Можно показать, что если функция $f(x, y)$ на рассматриваемом участке непрерывна и удовлетворяет условию Липшица

$$|f(x, y_i) - f(x, y_j)| \leq K|y_i - y_j|; \quad \left| \frac{df}{dx} \right| \leq N,$$

где $K = \text{const}$, $N = \text{const}$, то погрешность $\xi_k = y_k - y(x_k)$ удовлетворяет условию

$$|\xi_k| \leq \frac{hN}{2K} [(1+hK)^k - 1].$$

Погрешность убывает с уменьшением h и метод на любом конечном отрезке сходится, ибо при $h \rightarrow 0$ $\xi \rightarrow 0$.

Метод Эйлера можно усовершенствовать, выбрав дополнительные точки внутри промежутка.

2. Метод Рунге—Кутты

Если рассматривать на каждом шаге формулу Эйлера

$$y_{n+1} = y_n + hf'_n = y_n + hy'_n$$

как часть формулы Тейлора, то естественным обобщением будет добавление слагаемых, соответствующих высшим производным

$$y_{n+1} - y_n \approx hy'_n + \frac{h^2}{2} y''_n + \frac{h^3}{3!} y'''_n. \quad (1.48)$$

Чем больше слагаемых мы используем, тем точнее будет решение. Затруднением является то, что производные $y_n^{(i)}$ должны выражаться через производные функции $f(x, y)$ из уравнения (1.44).

Эти формулы с увеличением порядка производной все более и более усложняются. Так:

$$y'' = f'_x + f'_y f;$$

$$y''' = f''_{xx} + 2f''_{xy} f + f''_y f^2 + f'_y (f'_x + f'_y f).$$

Очевидно, вести расчет по формуле (1.48) неудобно. Чтобы избежать этого, заменим правую часть формулы линейной комбинацией

$$y_{n+1} - y_n = \alpha k_1 + \beta k_2 + \gamma k_3 + \delta k_4 + \dots, \quad (1.48')$$

где $k_i = hf(x_i, y_i)$, $x_i \in [x_n, x_{n+1}]$.

Рассмотрим случай, когда формула Тейлора для y_{n+1} имеет вид

$$y_{n+1} - y_n = hy'_n + \frac{h^2}{2!} y''_n, \quad (1.49)$$

и возьмем в формуле (1.48') два слагаемых

$$y_{n+1} - y_n = \alpha k_1 + \beta k_2, \quad (1.50)$$

$$k_1 = hf(x_n, y_n),$$

$$k_2 = hf(x_n + h, y_n + k_1). \quad (1.51)$$

Чтобы найти коэффициенты α и β надо приравнять правые части формул (1.49) и (1.50). При этом в формуле (1.49) производные y'_n и y''_n следует выразить через производные от функции $f(x, y)$, а в формуле (1.50) разложить функцию $f(x_n + h, y_n + k_1)$ по формуле Тейлора в окрестности точки (x_n, y_n) :

$$y_{n+1} - y_n = \Delta y_n = hf_n + \frac{h}{2} (f'_{xn} + f'_{yn} f_n), \quad (1.52)$$

где $f'_{xn} = f'_x(x_n, y_n)$, $f'_{yn} = f'_y(x_n, y_n)$, $f_n = f(x_n, y_n)$;

$$\Delta y_n = \alpha hf_n + \beta h [f_n + (f'_{xn} h + f'_{yn} k_1)] = \alpha hf_n + \beta h [f_n + f'_{xn} h + f'_{yn} f_n h]. \quad (1.53)$$

Приравнявая коэффициенты при одинаковых степенях h в формулах (1.52) и (1.53), получим $\alpha + \beta = 1$, $\beta = \frac{1}{2}$. Следовательно, $\alpha = \beta = \frac{1}{2}$, и расчетная формула примет вид

$$\Delta y_n = \frac{1}{2} k_1 + \frac{1}{2} k_2,$$

$$k_1 = hf(x_n, y_n),$$

$$k_2 = hf(x_n + h, y_n + k_1).$$

Приведем формулы Рунге—Кутта для четвертого порядка точности:

$$\Delta y_n = \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4),$$

$$k_1 = hf(x_n, y_n),$$

$$k_2 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}\right),$$

$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}\right),$$

$$k_4 = hf(x_n + h, y_n + k_3).$$

Расчет по методу Рунге—Кутта удобно располагать по табл. 9 (здесь дан расчет одного шага).

Таблица 9

x	y	$f(x, y)$	$k = hf(x, y)$	
x_0	y_0	$f(x_0, y_0)$	k_1	Δy
$x_0 + \frac{h}{2}$	$y_0 + \frac{k_1}{2}$	$f\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right)$	k_2	
$x_0 + \frac{h}{2}$	$y_0 + \frac{k_2}{2}$	$f\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right)$	k_3	
$x_0 + h$	$y_0 + k_3$	$f(x_0 + h, y_0 + k_3)$	k_4	
x_1	$y_1 = y_0 + \Delta y$			

Пример. Решить задачу Коши

$$y' = \frac{1-y^2}{2(1-x^2)} + 1, \quad y|_{x=0} = 1, \quad h = 0,1.$$

Покажем, как проводится расчет для двух шагов в виде табл. 10.

Таблица 10

x	y	$f(x, y)$	h	Δy
0	1	1,0000	0,1000	
0,05	1,050	1,05266	0,10526	
0,05	1,05263	0,94460	0,09446	
0,1	1,09446	0,900078	0,090008	0,098240
0,1	1,098240	0,89589	0,089589	
0,15	1,14304	0,84321	0,084321	
0,15	1,1404	0,84629	0,084629	
0,2	1,18287	0,79209	0,079209	0,084450
0,2	1,18269			

Аналогичным способом по методу Рунге — Кутта решаются системы уравнений. Рассмотрим схему для случая системы двух уравнений.

Пусть имеем систему

$$y' = f_1(x, y, z),$$

$$z' = f_2(x, y, z),$$

$$y|_{x=x_0} = y_0,$$

$$z|_{x=x_0} = z_0.$$

Расчетные формулы для одного шага имеют вид

$$\Delta y_n = \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4);$$

$$\Delta z_n = \frac{1}{6} (p_1 + 2p_2 + 2p_3 + p_4),$$

где

$$k_1 = hf_1(x_n, y_n, z_n),$$

$$k_2 = hf_1\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}, z_n + \frac{p_1}{2}\right),$$

$$k_3 = hf_1\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}, z_n + \frac{p_2}{2}\right),$$

$$k_4 = hf_1(x_n + h, y_n + k_3, z_n + p_3);$$

$$p_1 = hf_2(x_n, y_n, z_n),$$

$$p_2 = hf_2\left(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}, z_n + \frac{p_1}{2}\right),$$

$$p_3 = hf_2\left(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}, z_n + \frac{p_2}{2}\right),$$

$$p_4 = hf_2(x_n + h, y_n + k_3, z_n + p_3).$$

Схема расчета дана в табл. 11.

Таблица 11

x	y	z	$f_1 h = k$	$f_2 h = p$	Δy	Δz
x_n	y_n	z_n	k_1	p_1	Δy_n	Δz_n
$x_n + \frac{h}{2}$	$y_n + \frac{k_1}{2}$	$z_n + \frac{p_1}{2}$	k_2	p_2		
$x_n + \frac{h}{2}$	$y_n + \frac{k_2}{2}$	$z_n + \frac{p_2}{2}$	k_3	p_3		
$x_n + h$	$y_n + k_3$	$z_n + p_3$	k_4	p_4		
$x_{n+1} = x_n + h$	$y_{n+1} = y_n + \Delta y_n$	$z_{n+1} = z_n + \Delta z_n$				



Глава II

ЧИСЛЕННЫЕ МЕТОДЫ ЛИНЕЙНОЙ АЛГЕБРЫ

§ 1. Элементы теории определителей

I. Определители первого порядка

Решение систем линейных уравнений является одной из основных задач линейной алгебры. Так как решение линейных систем тесно связано с теорией определителей, то нам придется начать главу с этого вопроса.

Пусть имеется система двух линейных уравнений

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1, \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned} \right\} \quad (2.1)$$

Исключим сначала неизвестное x_2 . Для этого умножим первое уравнение системы на a_{22} , а второе на a_{12} и вычтем из первого второе. Тогда

$$(a_{11}a_{22} - a_{21}a_{12})x_1 = b_1a_{22} - b_2a_{12}.$$

Подобным же образом, исключая x_1 , получим

$$(a_{11}a_{22} - a_{21}a_{12})x_2 = b_2a_{11} - b_1a_{21}.$$

Если $a_{11}a_{22} - a_{21}a_{12} \neq 0$, то легко получить решение системы в виде

$$\left. \begin{aligned} x_1 &= \frac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{21}a_{12}}, \\ x_2 &= \frac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{21}a_{12}}. \end{aligned} \right\} \quad (2.2)$$

Запись формул можно сделать более удобной, если ввести ряд новых понятий.

Запишем коэффициенты системы (2.1) в виде таблицы, которую будем называть матрицей A :

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Сопоставим матрице число $a_{11}a_{22} - a_{21}a_{12}$. Будем называть это число определителем матрицы A . Обозначим его $|A|$. Тогда

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} = |A|.$$

Очевидно, выражение $|A|$ получается по следующему правилу: нужно перемножить элементы a_{11} и a_{22} , стоящие на так называемой главной диагонали матрицы A , и вычесть из полученного произведения элементов a_{21} , a_{12} , стоящих на второй диагонали. Пользуясь указанным правилом, можно легко составить определители, соответствующие выражениям, стоящим в числителях формул (2.2):

$$b_1a_{22} - b_2a_{12} = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} = |A_1|,$$

$$b_2a_{11} - b_1a_{21} = \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} = |A_2|.$$

Эти формулы принимают тогда вид

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}. \quad (2.3)$$

По аналогии с матрицами и определителями, относящимися к системе двух уравнений и содержащими две строчки и два столбца, можно построить матрицу и определитель, содержащие n строчек и n столбцов. Для определения их следует ввести ряд новых понятий.

2. Перестановки, инверсии, транспозиции

Количество всевозможных перестановок из n элементов $1, 2, \dots, n$, как известно, есть $n!$ Например, из трех чисел $1, 2, 3$ можно составить $3! = 6$ перестановок: $1, 2, 3$; $1, 3, 2$; $2, 3, 1$; $2, 1, 3$; $3, 1, 2$; $3, 2, 1$. Можно отличать перестановки друг от друга количеством инверсий — беспорядков. Если числа в перестановке расположены в натуральном порядке — $1, 2, 3$, — то инверсий нет. Если, например, число 3 стоит перед числом 2 , то такое положение называется инверсией. Инверсии в перестановке легко сосчитать. Рассмотрим перестановку $3, 2, 1$. Сначала будем рассматривать инверсии первой цифры 3 с остальными: $3 \rightarrow 1$, $3 \rightarrow 2$ дадут

две инверсии. Инверсия цифры $2(2 \rightarrow 1)$ даст еще одну. Всего перестановка 3, 2, 1 содержит три инверсии. Подсчитаем количество инверсий в перестановках, составленных из чисел 1, 2, 3: 1, 2, 3 — нет инверсий, 1, 3, 2 — одна инверсия, 2, 1, 3 — одна инверсия, 2, 3, 1 — две инверсии, 3, 1, 2 — две инверсии, 3, 2, 1 — три инверсии.

По количеству инверсий перестановки делятся на четные и нечетные. В нашем примере три четных перестановки и три нечетных.

Если в перестановке из n элементов: $i_1, i_2, \dots, i_j, \dots, i_k, \dots, i_n$ поменять местами только два элемента i_j и i_k , то получится новая перестановка: $i_1, i_2, \dots, i_k, \dots, i_j, \dots, i_n$.

Такая операция называется транспозицией и обозначается символом (i_k, i_j) .

С помощью некоторого числа транспозиций можно перейти последовательно от одной перестановки к другой. Например, от перестановки 1, 2, 3 можно перейти к перестановке 3, 1, 2 двумя транспозициями: $(1, 2, 3) \rightarrow (1, 3); (3, 2, 1) \rightarrow (2, 1); (3, 1, 2)$.

Свяжем понятие четности перестановки с транспозициями.

Теорема. От одной транспозиции четность перестановки меняется.

Доказательство. Пусть сначала для простоты транспонируемые элементы находятся рядом

$$i_1, i_2, \dots, i_k, i_{k+1}, \dots, i_n \rightarrow i_1, i_2, \dots, i_{k+1}, i_k, \dots, i_n.$$

Элементы i_k и i_{k+1} после перестановки сохраняют с остальными элементами то же количество инверсий, какое было до транспозиции. Если сами элементы i_k, i_{k+1} до перестановки составляли инверсию, то после они не будут ее составлять, и наоборот. Следовательно, общее количество инверсий при такой транспозиции или увеличится или уменьшится на одну, т. е. изменится четность перестановки.

Теперь рассмотрим такие транспонируемые элементы i_k, i_{k+m} , между которыми расположены $m-1$ других элементов:

$$i_1, i_2, \dots, i_k, i_{k+1}, \dots, i_{k+m}, \dots, i_n.$$

Будем перемещать их постепенно, каждый раз меняя местами только соседние элементы. При этом каждый раз четность перестановки будет меняться. Таких транспозиций мы получим $2m-1$, так как, переставляя i_k на место i_{k+m} , мы производим m транспозиций, а переставляя i_{k+m} на место i_k , — еще $m-1$. Нечетное число изменений четности даст в конечном итоге четность, противоположную исходной.

Следствия.

1. Чтобы изменить четность перестановки, нужно нечетное число транспозиций.

2. Из n цифр получится $n!$ перестановок; $\frac{n!}{2}$ четных и $\frac{n!}{2}$ нечетных.

3. Определители n -го порядка

Рассмотрим таблицу из n^2 элементов

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

и назовем ее квадратной матрицей n -го порядка. Сопоставим матрице определитель $|A|$ или Δ .

Определителем n -го порядка называется алгебраическая сумма $n!$ слагаемых. Каждое слагаемое представляет из себя произведение n элементов, взятых по одному из каждой строчки и каждого столбца матрицы. Знак слагаемого определяется как $(-1)^t$, где t — число инверсий во вторых индексах, если первые расположены в натуральном порядке

$$\Delta = \sum (-1)^t a_{1j_1} a_{2j_2} \dots a_{nj_n}. \quad (2.4)$$

Можно показать, что справедлива также формула

$$\Delta = \sum (-1)^s a_{i_1 1} a_{i_2 2} \dots a_{i_n n}, \quad (2.5)$$

где s — количество инверсий в перестановке первых индексов

4. Свойства определителей

Свойство 1. Если переменить местами строчки и столбцы определителя, то он не изменит своего значения.

Доказательство очевидно.

З а м е ч а н и е. Все последующие свойства, сформулированные для строчек, будут относиться также к столбцам.

Свойство 2. Если переставить местами две строчки определителя, то он изменит знак.

Доказательство. Рассмотрим определитель Δ и поменяем местами i -ю и k -ю строчки. Назовем полученный определитель Δ_1 . Тогда

$$\Delta = \sum (-1)^t a_{1j_1} a_{2j_2} \dots a_{ij_i} \dots a_{kj_k} \dots a_{nj_n};$$

$$\Delta_1 = \sum (-1)^{t_1} a_{1j_1} a_{2j_2} \dots a_{kj_k} \dots a_{ij_i} \dots a_{nj_n}.$$

Очевидно, абсолютная величина слагаемых одинакова, а знаки различны, так как при одной транспозиции перестановка меняет четность. Следовательно, $\Delta = -\Delta_1$.

Свойство 3. Определитель, у которого две строки совпадают, равен нулю.

Доказательство. В определителе Δ поменяем совпадающие строчки местами. По свойству 2 получим: $\Delta_1 = -\Delta$, но так как строчки равны, то $\Delta_1 = \Delta$. Отсюда $\Delta = \Delta_1 = 0$.

Свойство 4. Если все элементы одной из строчек определителя равны нулю, то и определитель равен нулю.

Доказательство. По определению следует в каждом слагаемом брать множители по одному из каждой строчки и из каждого столбца. Очевидно, в нашем случае каждое слагаемое будет содержать нулевой множитель и, следовательно, $\Delta = 0$.

Свойство 5. Если все элементы одной строчки обладают общим множителем, то этот множитель можно вынести за знак определителя.

Доказательство. Пусть элементы i -й строчки имеют общий множитель a . Тогда в каждое слагаемое определителя будет обязательно входить множитель a . Вынеся его за знак суммы, мы докажем свойство.

Свойство 6. Определитель, две строчки которого пропорциональны, равен нулю. Указанное свойство является очевидным следствием свойств 3 и 5.

Свойство 7. Если каждый элемент одной из строчек определителя Δ можно представить как сумму двух слагаемых, то Δ равен сумме двух определителей того же порядка, у которых элементами соответствующих строк являются указанные слагаемые.

Доказательство очевидно.

Следствие. Определитель не изменится, если к элементам одного столбца прибавить элементы другого, умноженные на один и тот же множитель.

Доказательство следует из свойств 3 и 5.

5. Миноры, алгебраические дополнения.

Разложение определителей по строчкам или столбцам

Определение 1. Минором M_{ij} элемента a_{ij} определителя Δ n -го порядка называется определитель $(n-1)$ -го порядка, который получен из исходного вычеркиванием i -й строчки и j -го столбца:

$$M_{ij} = \begin{vmatrix} a_{11} & \dots & a_{1j-1} & a_{1j+1} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i-11} & \dots & a_{i-1j-1} & a_{i-1j+1} & \dots & a_{i-1n} \\ a_{i+11} & \dots & a_{i+1j-1} & a_{i+1j+1} & \dots & a_{i+1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots & a_{nj-1} & a_{nj+1} & \dots & a_{nn} \end{vmatrix}.$$

Определение 2. Алгебраическим дополнением A_{ij} элемента a_{ij} определителя Δ называется минор M_{ij} , взятый со знаком $(-1)^{i+j}$:

$$A_{ij} = (-1)^{i+j} M_{ji}.$$

Например, найдем алгебраическое дополнение A_{34} элемента a_{34} определителя

$$\Delta_4 = \begin{vmatrix} 1 & 4 & 5 & 6 & 7 \\ 2 & 1 & 1 & 3 & 3 \\ 3 & 1 & 4 & 4 & 1 \\ 5 & 6 & 1 & 3 & 2 \\ 7 & 1 & 1 & 1 & 3 \end{vmatrix};$$

$$a_{34} = 4; \quad A_{34} = - \begin{vmatrix} 1 & 4 & 5 & 7 \\ 2 & 1 & 1 & 3 \\ 3 & 6 & 1 & 2 \\ 7 & 1 & 1 & 3 \end{vmatrix}.$$

Свойство 8. Если в определителе n -го порядка Δ все элементы i -й строчки, кроме a_{ij} , равны нулю, то определитель можно вычислить по формуле

$$\Delta = a_{ij} A_{ij}.$$

Доказательство. Предположим сначала, что $i = j = 1$, т. е. определитель

$$\Delta = \begin{vmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Так как по определению каждое слагаемое определителя состоит из элементов, взятых по одному из каждой строчки и из каждого столбца, то определитель Δ будет иметь целый ряд нулевых слагаемых. Не равные нулю элементы будут иметь вид

$$(-1)^t a_{11} a_{2 i_2} a_{3 i_3} \dots a_{n i_{n-1}},$$

где t — число инверсий в перестановке $i, i_1, \dots, i_{n-1}$ (из элементов $1, 2, \dots, n$).

Очевидно, то же число инверсий будет в перестановке из $(n-1)$ элементов. Также очевидно, что все слагаемые имеют общий множитель a_{11} . Вынесем его за знак суммы и получим

$$\Delta = a_{11} \sum (-1)^t a_{2 i_2} \dots a_{n i_n}.$$

Сумма по определению дает определить $(n-1)$ -го порядка, который назван нами M_{11} — минором элемента a_{11} . Но $A_{11} = (-1)^2 M_{11}$, и таким образом $\Delta = a_{11} A_{11}$.

Рассмотрим случай, когда не нулевой элемент находится в i -й строчке и j -м столбце. Переставим эту строчку и столбец на первые места. При этом знак поменяется $i+j$ раз. Это можно учесть, введя множитель $(-1)^{i+j}$. Таким образом, имеем

$$\Delta = a_{ij} M_{ij} (-1)^{i+j} = a_{ij} A_{ij}.$$

Свойство 9. Произведение элементов любой строчки на их алгебраические дополнения равно определителю, а произведение элементов одной строчки на алгебраические дополнения элементов другой строчки равно нулю, т. е.

$$a_{i1} A_{i1} + a_{i2} A_{i2} + \dots + a_{in} A_{in} = \Delta; \quad (2.6)$$

$$a_{i1} A_{j1} + a_{i2} A_{j2} + \dots + a_{in} A_{jn} = 0. \quad (2.7)$$

Доказательство. Запишем определитель Δ в следующем виде:

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{i1} + 0 + \dots + 0 & 0 + a_{i2} + 0 + \dots + 0 & \dots & 0 + 0 + \dots + a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Мы представим таким образом элементы i -й строчки как сумму слагаемых и по свойству 7 определитель Δ в виде суммы n определителей

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ 0 & a_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \dots +$$

$$+ \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix};$$

по свойству 8 имеем

$$\Delta = a_{i1} A_{i1} + a_{i2} A_{i2} + \dots + a_{in} A_{in}.$$

Это равенство доказывает первую часть свойства. Для доказательства второй части построим дополнительно определитель Δ_1 , отличающийся от Δ только тем, что его k -я строка заменена i -й, а i -я оставлена без изменения. Определитель Δ имеет две одинаковые строки на i -м и на k -м месте и по свойству 3 Δ_1 равен нулю. Разложим его по k -й строчке и получим

$$a_{i1} A_{k1} + a_{i2} A_{k2} + \dots + a_{in} A_{kn} = 0.$$

Построенная формула доказывает вторую часть свойства.

В заключение параграфа приведем без доказательства правило умножения определителей.

Произведением двух определителей $|A|$ и $|B|$ является определитель $|C|$, элементы которого определяются по формуле

$$c_{ij} = a_{i1} b_{1j} + a_{i2} b_{2j} + \dots + a_{in} b_{nj}.$$

Пример. Вычислитель определитель V_n , носящий имя Вандермонда:

$$V_n = \begin{vmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & a_n & a_n^2 & \dots & a_n^{n-1} \end{vmatrix}.$$

Для этого вычтем из каждого столбца предыдущий, умноженный на a_1 :

$$\begin{aligned} V_n &= \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & a_2 - a_1 & a_2(a_2 - a_1) & \dots & a_2^{n-2}(a_2 - a_1) \\ \dots & \dots & \dots & \dots & \dots \\ 1 & a_n - a_1 & a_n(a_n - a_1) & \dots & a_n^{n-2}(a_n - a_1) \end{vmatrix} = \\ &= 1 \begin{vmatrix} a_2 - a_1 & a_2(a_2 - a_1) & \dots & a_2^{n-2}(a_2 - a_1) \\ \dots & \dots & \dots & \dots \\ a_n - a_1 & a_n(a_n - a_1) & \dots & a_n^{n-2}(a_n - a_1) \end{vmatrix} = \\ &= (a_2 - a_1)(a_3 - a_1) \dots (a_n - a_1) \begin{vmatrix} 1 & a_2 & \dots & a_2^{n-2} \\ 1 & a_3 & \dots & a_3^{n-2} \\ \dots & \dots & \dots & \dots \\ 1 & a_n & \dots & a_n^{n-2} \end{vmatrix}. \end{aligned}$$

Таким образом, мы получили рекуррентную формулу, которую проще записать в виде

$$V_n = (a_n - a_1) \dots (a_n - a_{n-1}) V_{n-1}.$$

Но так как

$$V_2 = \begin{vmatrix} 1 & a_{n-1} \\ 1 & a_{n-2} \end{vmatrix} = (a_{n-2} - a_{n-1}),$$

то

$$V_n = \prod_{j>i} (a_j - a_i).$$

Вычисленный определитель используется в теории интерполирования.

§ 2. Матрицы и их простейшие свойства

Прямоугольной матрицей называется совокупность чисел, расположенных в виде прямоугольной таблицы, содержащей n строк и m столбцов:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}.$$

Мы будем в основном рассматривать квадратные матрицы, у которых $m = n$, матрицы-столбцы, где $m = 1$, и матрицы-строки, где $n = 1$. Матрицы-столбцы и матрицы-строки будем называть также векторами.

Элементы a_{ii} составляют главную диагональ матрицы. Квадратная матрица, у которой на главной диагонали стоят единицы, а остальные элементы — нули, называется единичной и обозначается E :

$$E = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix}.$$

Если все элементы квадратной матрицы равны нулю, то матрица называется нулевой [0]. Две матрицы A и B равны между собой, если равны соответственно все их элементы: $a_{ij} = b_{ij}$.

Если у матрицы A переставить строки и столбцы, то полученная матрица называется транспонированной и обозначается A^T . Если $A^T = A$, то матрица называется симметричной. Если все элементы матрицы при этом вещественны, то такая матрица называется также самосопряженной. Определитель, элементы которого совпадают с элементами квадратной матрицы, называется определителем матрицы и обозначается $|A|$. Если $|A| \neq 0$, то матрица называется неособой.

1. Действия над матрицами

Для того чтобы сложить или вычесть две матрицы, надо сложить или вычесть соответствующие элементы: $A \pm B = C$, если $a_{ij} + b_{ij} = c_{ij}$ для всевозможных i и j .

Для того чтобы матрицу умножить на число, надо умножить на это число все элементы матрицы: $B = \alpha A$, если $b_{ij} = \alpha a_{ij}$ для всевозможных i и j .

Введенные действия обладают следующими свойствами:

$$1) A + (B + C) = (A + B) + C;$$

$$2) A + B = B + A;$$

$$3) A + [0] = A;$$

$$4) (\alpha + \beta) A = \alpha A + \beta A;$$

$$5) \alpha(A + B) = \alpha A + \alpha B;$$

$$6) EA = A;$$

$$7) \alpha(\beta A) = \alpha\beta A.$$

Произведением двух матриц A и B называется матрица C , элементы которой построены по правилу:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}. \quad (2.8)$$

Очевидно, перемножать матрицы можно только в том случае, если число столбцов матрицы A равно числу строк матрицы B .

Из существования произведения AB не следует существование произведения BA , и $AB \neq BA$, даже если оба произведения существуют. В случае квадратных матриц перемножение возможно для матриц одного порядка.

В некоторых отдельных случаях произведение матриц коммутативно. Например: $AE = EA = A$; $A^T A = A A^T$.

Рассмотрим еще один пример произведения. Если перемножить матрицу-строку на матрицу-столбец, то получим матрицу, содержащую один элемент — число

$$X = (x_1, x_2, \dots, x_n), Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; XY = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

Если и матрицу-строку и матрицу-столбец рассматривать как вектор, имеющий n составляющих, обобщение трехмерного вектора, то такое произведение естественно назвать скалярным произведением двух векторов и обозначить

$$(XY) = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

Если $(XY) = 0$, то будем говорить, что векторы X и Y ортогональны.

В дальнейшем мы будем чаще всего рассматривать матрицы-столбцы как вектора.

Приведем некоторые свойства произведения матриц:

- 1) $(AB)^T = B^T A^T$;
- 2) $|AB| = |A| |B|$;
- 3) $\alpha(AB) = A(\alpha B) = (\alpha A)B$;
- 4) $(A+B)C = AC + BC$;
- 5) $C(A+B) = CA + CB$.

Обратной к данной матрице A называется матрица A^{-1} , для которой выполняются равенства

$$AA^{-1} = A^{-1}A = E.$$

Теорема. Для того чтобы существовала обратная матрица A^{-1} , необходимо и достаточно, чтобы матрица A была неособая.

Доказательство.

Достаточность: пусть $|A| \neq 0$. Введем присоединенную матрицу $\tilde{A}$ по формуле

$$\tilde{A} = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & \dots & \dots & A_{nn} \end{pmatrix},$$

где A_{ij} — алгебраическое дополнение элемента a_{ij} .

Рассмотрим произведение $A\tilde{A}$, обозначив его D . По правилу умножения матриц элементы матрицы D , d_{ij} определяются с помощью формулы

$$d_{ij} = a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn}.$$

Но на основании свойства определителей 9, d_{ij} является при $i=j$ результатом разложения определителя $|A|$ по i -му столбцу и, следовательно, $d_{ii} = |A|$. При $i \neq j$ по тому же свойству $d_{ij} = 0$. Отсюда

$$d_{ij} = \begin{cases} |A| & \text{при } i=j \\ 0 & \text{при } i \neq j \end{cases}; \quad D = \begin{vmatrix} |A| & 0 & \dots & 0 \\ 0 & |A| & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |A| \end{vmatrix} = |A| E$$

или $A\tilde{A} = |A| E$, но $AA^{-1} = E$ и поэтому $A^{-1} = \frac{\tilde{A}}{|A|}$.

Все описанные преобразования можно было бы повторить для $\tilde{A}$.

Необходимость: пусть существует A^{-1} , тогда $AA^{-1} = E$ и $|A| |A^{-1}| = |E|$, но $|E| = 1$ и, следовательно, $|A| \neq 0$.

Простейшие свойства обратной матрицы определяются формулами

$$1) |A^{-1}| = \frac{1}{|A|};$$

$$2) \quad (A^{-1})^{-1} = A;$$

$$3) \quad (AB)^{-1} = B^{-1}A^{-1}.$$

Матрицы A и B называются подобными, если существует, неособая матрица C , такая, что $B = C^{-1}AC$.

Пример. Рассмотрим линейную систему:

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n. \end{aligned} \right\}$$

Левые части системы можно представить как матрицу-столбец, являющуюся произведением матриц

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}; \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix};$$

$$AX = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{pmatrix}$$

Правые части системы представляют матрицу-столбец

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Тогда система может быть записана так:

$$AX = B, \quad (2.9)$$

Как было указано, матрицы X и B можно называть векторами. Домножим формулу (2.9) слева на матрицу A^{-1}

$$A^{-1}AX = A^{-1}B.$$

Используя формулу для обратной матрицы, получим

$$X = \frac{\tilde{A}B}{|A|}.$$

Развернем полученное в правой части произведение

$$\begin{aligned}\tilde{A}B &= \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \\ &= \begin{pmatrix} A_{12}b_1 + A_{21}b_2 + \dots + A_{n1}b_n \\ A_{12}b_1 + A_{22}b_2 + \dots + A_{n2}b_n \\ \dots \\ A_{1n}b_1 + A_{2n}b_2 + \dots + A_{nn}b_n \end{pmatrix}.\end{aligned}$$

Обозначим: $A_{1k}b + A_{2k}b_2 + \dots + A_{nk}b_n = |A_k|$.

Это выражение является разложением определителя по k -му столбцу, причем элементами этого столбца являются правые части нашей системы, а остальные столбцы определителя те же, что и у основного определителя системы $|A|$. Таким образом, решение системы можно представить в виде

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}, \quad \dots, \quad x_n = \frac{|A_n|}{|A|}.$$

Полученные формулы называются формулами Крамера.

Приведем частный случай системы, когда правые части равны нулю: $B = [0]$, т. е. однородную систему. В этом случае, если определитель $|A|$ не равен нулю, мы имеем единственное нулевое решение. Для того чтобы однородная система имела решения, отличные от нулевого, необходимо и достаточно, чтобы ее основной определитель был равен нулю.

2. Положительно определенные матрицы

Рассмотрим сначала функцию, называемую квадратичной формой

$$\begin{aligned}F(x_1, x_2, \dots, x_n) &= a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{1n}x_1x_n + \\ &+ a_{21}x_2x_1 + a_{22}x_2^2 + \dots + a_{2n}x_2x_n + \\ &+ \dots + \dots + \dots + \dots + \\ &+ a_{n1}x_nx_1 + a_{n2}x_nx_2 + \dots + a_{nn}x_n^2.\end{aligned}$$

Составим квадратичной форме симметричную матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Запишем квадратичную форму с помощью следующих формул:

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

$$\begin{aligned} X^TAX &= (x_1, x_2, \dots, x_n) \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{pmatrix} = \\ &= F(x_1, x_2, \dots, x_n), \end{aligned}$$

где $X^T = (x_1, x_2, \dots, x_n)$ — матрица-строка.

Квадратичная форма называется положительно определенной, если при всех вещественных значениях переменных, входящих в форму, кроме случая $x_1 = x_2 = \dots = x_n = 0$, квадратичная форма принимает только строго положительные значения. Матрица A , сопоставленная положительно определенной форме, называется положительно определенной. Простейшим примером положительно определенной формы и соответствующей матрицы является

$$F(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2$$

и единичная матрица E .

Рассмотрим, как изменяются коэффициенты квадратичной формы при линейном преобразовании переменных. Введем замену переменных по формулам:

$$x_1 = b_{11}y_1 + \dots + b_{1n}y_n,$$

$$x_2 = b_{21}y_1 + \dots + b_{2n}y_n,$$

$$\dots$$

$$x_n = b_{n1}y_1 + \dots + b_{nn}y_n,$$

или в матричной форме

$$X = BY.$$

Проведем эту замену в квадратичной форме $F(x_1, x_2, \dots, x_n) = X^TAX$,

$$X^TAX = (BY)^T A (BY) = Y^T B^T A B Y.$$

Таким образом, квадратичной форме в новых переменных отвечает матрица $B^T A B$. Эта матрица симметрична. Действительно

$$(B^T A B)^T = (A B)^T (B^T)^T = B^T A^T B = B^T A B, \quad A^T = A.$$

Заметим также, что при замене переменных квадратичная форма сохраняет свойство положительной определенности. Следовательно, если матрица A была положительно определенной, то, если B не особая, $B^T A B$ тоже положительно определенная. Так как матрица E положительно определенная, то $A^T E A$ или, что то же самое, $A^T A$ — положительно определенная матрица.

Если матрица A системы $A X = B$ неособая, то преобразованием Гаусса, заключающимся в умножении обеих частей системы слева на A , получим систему $A^T A X = A^T B$ с положительно определенной матрицей $A^T A$.

3. Собственные числа и собственные векторы матрицы

Рассмотрим систему

$$A X = \lambda X, \quad (2.10)$$

где λ — постоянное число.

Эта система может быть записана в виде

$$(A - \lambda E) X = 0.$$

Система имеет не нулевое решение, если ее основной определитель $|A - \lambda E|$ равен нулю. Введем обозначение: $|A - \lambda E| = D(\lambda)$.

Очевидно, $D(\lambda)$ есть полином n степени от λ и может быть представлен так:

$$D(\lambda) = (-1)^n (\lambda^n - p_1 \lambda^{n-1} - \dots - p_n).$$

Это характеристический полином матрицы A . Приравнявая $D(\lambda)$ нулю, мы получим характеристическое или вековое уравнение

$$\lambda^n - p_1 \lambda^{n-1} - \dots - p_n = 0.$$

Корни этого уравнения называются собственными числами матрицы, а ненулевые решения системы (2.10) — собственными векторами матрицы A . Очевидно, каждая квадратная неособая матрица должна иметь n собственных чисел. Среди них могут быть различные вещественные, комплексные сопряженные и кратные. Данная и транспонированная матрицы, а также данная и подобная матрицы имеют одинаковые собственные числа. Для транспонированной это свойство очевидно, а для подобной легко проверяется.

Рассмотрим характеристический определитель подобной матрицы:

$$|C^{-1}AC - \lambda E| = |C^{-1}AC - C^{-1}\lambda EC| = |C^{-1}| |A - \lambda E| |C|.$$

Собственные векторы данной и транспонированной матриц ортогональны, т. е. если $X_1, X_2, \dots, X_n$ — собственные векторы матрицы A , а $Y_1, Y_2, \dots, Y_n$ — собственные векторы матрицы A^T , то имеет место соотношение

$$(X_i, Y_j) = \sum_{k=1}^n x_{ik} y_{jk} = 0.$$

Собственные векторы симметричной и положительно определенной матриц составляют ортогональные системы. Кроме того, собственные числа положительно определенных матриц вещественны и положительны, а для симметричных матриц — вещественны.

Соотношение Кели-Гамильтона. Если $D(\lambda)$ — характеристический полином матрицы A , то $D(A) = [0]$, т. е. матрица является корнем своего характеристического полинома.

Доказательство. Построим матрицу B , элементы которой являются алгебраическими дополнениями транспонированной матрицы $(A - E\lambda)^T$. Каждый элемент этой матрицы можно представить в виде

$$b_{ik} = b_{ik}^{(0)} + \lambda b_{ik}^{(1)} + \dots + b_{ik}^{(n-1)} \lambda^{n-1},$$

и поэтому саму матрицу тоже можно представить в виде полинома $(n-1)$ -й степени от λ с матричными коэффициентами

$$B = B_0 + B_1\lambda + \dots + B_{n-1}\lambda^{n-1}.$$

Из теоремы об обратной матрице известно, что присоединенная матрица B удовлетворяет соотношению

$$(A - \lambda E)B = |A - \lambda E| E.$$

Подставив сюда выражение для B , напишем

$$\begin{aligned} (A - \lambda E)(B_0 + B_1\lambda + \dots + B_{n-1}\lambda^{n-1}) &= \\ &= (-1)^n (\lambda^n - p_1\lambda^{n-1} - \dots - p_0)E. \end{aligned}$$

Приравняем коэффициенты при одинаковых степенях λ :

$$\left. \begin{array}{l} \lambda^0 \\ \lambda^1 \\ \dots \\ \lambda^n \end{array} \right\} \begin{array}{l} AB_0 = (-1)^n (-p_n)E, \\ -B_0 + AB_1 = (-1)^n (-p_{n-1})E, \\ \dots \\ -B_{n-1} = (-1)^n E. \end{array}$$

Домножив слева первое уравнение системы на E , второе на A , третье на A^2 и т. д., получим

$$\begin{aligned} AB_0 &= (-1)^n (-p_n) E, \\ -AB_0 + A^2 B_1 &= (-1)^n (-p_{n-1}) A, \\ -A^2 B_1 + A^3 B_2 &= (-1)^n (-p_{n-2}) A^2, \\ &\vdots \\ -A^n B_{n-1} &= (-1)^n A^n. \end{aligned}$$

Сложив полученные равенства, имеем

$$[0] = D(A).$$

§ 3. Линейные нормированные пространства

Линейным вещественным пространством называют совокупность математических объектов-векторов, для которых определены два действия: сложение и умножение на любые вещественные числа. Эти действия удовлетворяют следующим свойствам-аксиомам:

- 1) $\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X}$ — переместительный закон;
- 2) $(\mathbf{X} + \mathbf{Y}) + \mathbf{Z} = \mathbf{X} + (\mathbf{Y} + \mathbf{Z})$ — сочетательный закон;
- 3) существует $\mathbf{0}$ — нулевой вектор, такой что $\mathbf{X} + \mathbf{0} = \mathbf{X}$;
- 4) существует $-\mathbf{X}$, такой, что $\mathbf{X} + (-\mathbf{X}) = \mathbf{0}$;
- 5) $1 \cdot \mathbf{X} = \mathbf{X}$;
- 6) $(a + b)\mathbf{X} = a\mathbf{X} + b\mathbf{X}$,
- 7) $a(\mathbf{X} + \mathbf{Y}) = a\mathbf{X} + a\mathbf{Y}$;
- 8) $a(b\mathbf{X}) = ab\mathbf{X}$.

Векторы $X_1, X_2, \dots, X_n$ — называются линейно независимыми, если тождество

$$c_1 \mathbf{X}_1 + c_2 \mathbf{X}_2 + \dots + c_n \mathbf{X}_n = 0,$$

где $c_1, c_2, \dots, c_n$ — постоянные, справедливо только при $c_1 = c_2 = \dots = c_n = 0$.

Рассмотрим различные линейные комбинации линейно-независимых векторов

$$\mathbf{X} = c_1 \mathbf{X}_1 + c_2 \mathbf{X}_2 + \dots c_n \mathbf{X}_n. \quad (2.11)$$

Совокупность таких комбинаций образует линейное пространство R с размерностью n .

Всякая совокупность n линейно независимых векторов линейного пространства называется его базисом. Естественно принять в качестве базиса ортогональные векторы, так как в таком базисе легче искать коэффициенты c_k , которые следует рассматривать как

составляющие вектора в данном базисе. Естественным образом принять за базис векторы

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \dots e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Эти векторы в случае $n=3$ являются ортами i, j, k .

Пусть $U_1, U_2, \dots, U_n$ — какой-нибудь базис пространства R . Тогда любой вектор X можно представить в виде

$$X = x_1 U_1 + x_2 U_2 + \dots + x_n U_n,$$

где $x_1, x_2, \dots, x_n$ — составляющие X в данном базисе

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

При переходе от базиса U_h к базису U_h получим связь между координатами векторов X и X' :

$$X = CX',$$

где C — матрица, столбцы которой состоят из координат векторов U' в базисе U .

Построенное равенство дает формулу для замены базиса, т. е. замены координат. Таким образом, координаты вектора X в базисе U переходят в координаты вектора X' в базисе U' .

Приведем линейное преобразование

$$Y = AX,$$

где A — неособая матрица этого преобразования; X и Y определены своими координатами в исходном базисе. Рассмотрим, как меняется матрица преобразования при переходе от одного базиса к другому. Пусть $X = CX'$ и $Y = CY'$ определяют векторы X и Y в новом базисе. Тогда

$$CX' = ACX', \quad C^{-1}CY' = C^{-1}ACX', \quad Y' = C^{-1}ACX'.$$

Следовательно, роль матрицы преобразования в новом базисе будет играть матрица $B = C^{-1}AC$, которую мы назвали подобной A .

Собственные числа и собственные векторы матрицы A называются также собственными числами и собственными векторами линейного преобразования. Если линейное преобразование имеет

в каком-нибудь базисе диагональную матрицу, то этот базис состоит из собственных векторов, а диагональные элементы являются собственными числами. И наоборот, если существует базис, состоящий из собственных векторов, то в этом базисе матрица преобразования будет диагональной.

В линейном пространстве R можно сопоставить каждому вектору число — аналог длины вектора. Это неотрицательное число называется нормой, обозначается через $\|X\|$ и удовлетворяет следующим требованиям:

- 1) $\|X\| > 0$, при $X \neq 0$ $\|0\| = 0$;
- 2) $\|\alpha X\| = |\alpha| \|X\|$;
- 3) $\|X + Y\| \leq \|X\| + \|Y\|$.

Будем рассматривать три вида норм, отвечающих указанным требованиям:

$$\|X\|_I = \max_i |x_i|; \quad (2.12)$$

$$\|X\|_{II} = \sum_{i=1}^n |x_i|; \quad (2.13)$$

$$\|X\|_{III} = \sqrt{\sum_{i=1}^n x_i^2}. \quad (2.14)$$

Всякое линейное пространство, в котором введена норма, называется линейным нормированным пространством. Если в линейном конечномерном нормированном пространстве определено скалярное произведение векторов, то такое пространство называется Эвклидовым.

Скалярное произведение двух векторов определено в предыдущем параграфе: $(XY) = x_1y_1 + x_2y_2 + \dots x_ny_n$.

Приведем свойства скалярного произведения:

- 1) $(XX) = 0$ при $X \neq 0$; $(00) = 0$;
- 2) $(XY) = (YX)$;
- 3) $(\alpha X)Y = \alpha(XY)$, α — число;
- 4) $(X + X_1)Y = (XY) + (X_1Y)$.

Опираясь на эти свойства, можно показать справедливость неравенства, которое носит имя Коши-Буняковского:

$$|XY| \leq \|X\|_{III} \|Y\|_{III}.$$

Рассмотрим матрицу линейного преобразования

$$Y = AX.$$

Для матрицы можно ввести понятие нормы, как нижнюю грань чисел M , для которых выполняется неравенство

$$\|AX\| \leq M\|X\|$$

и удовлетворяющих требованиям:

$$1) \|A\| > 0, \quad A \neq 0 \quad \|0\| = 0;$$

$$2) \|CA\| = |C| \|A\|;$$

$$3) \|A+B\| \leq \|A\| + \|B\|;$$

$$4) \|AB\| \leq \|A\| \|B\|.$$

Приведем нормы матриц, отвечающих выбранным нормам векторов:

$$\|A\|_I = \max_i \sum_j |a_{ij}|, \quad (2.15)$$

$$\|A\|_{II} = \max_j \sum_i |a_{ij}|, \quad (2.16)$$

$$\|A\|_{III} = \sqrt{\lambda},$$

где λ — наибольшее собственное число матриц $A^T A$.

Предел последовательности матриц

Рассмотрим последовательность матриц: $A_1, A_2, A_3, \dots, A_k, \dots$. Говорят, что последовательность матриц A_k имеет пределом матрицу A , если последовательности $a_{ij}^{(k)}$ элементов матриц A_k сходятся к соответственным элементам a_{ij} матрицы A .

Для того чтобы $\lim_{k \rightarrow \infty} A_k = A$, необходимо и достаточно, чтобы $\lim_{k \rightarrow \infty} \|A_k - A\| = 0$, где норма понимается в любом из введенных выше смысле.

Приведем ряд теорем о пределах последовательностей матриц.

Теорема 1. Для того чтобы $A^k \rightarrow 0$, необходимо и достаточно, чтобы все собственные числа матрицы A были по модулю строго меньше единицы.

Доказательство приведем только для случая, когда матрица A может быть приведена к диагональному виду:

$$A = C^{-1} B C; \quad B = \begin{vmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ 0 & & \ddots & \\ & & & \lambda_n \end{vmatrix} = (\lambda_1, \lambda_2, \dots, \lambda_n).$$

Тогда $A^m = C^{-1} B^m C$, $B^m = (\lambda_1^m, \dots, \lambda_n^m)$ и теорема доказана.

Теорема 2. Для того чтобы $A^k \rightarrow 0$, достаточно, чтобы хотя бы одна из норм матрицы A была строго меньше единицы.

Доказательство. По свойству норм

$$\|A^k\| \leq \|A\|^k, \quad \|A\|^k \xrightarrow{k \rightarrow \infty} 0.$$

Теорема 3. Модуль каждого собственного числа матрицы A не превосходит любую из ее норм.

Теорема 4. Для того чтобы ряд, составленный из матриц

$$E + A + A^2 + \dots + A^k + \dots,$$

сходился, необходимо и достаточно, чтобы $A^k \xrightarrow{k \rightarrow \infty} 0$.

Доказательство. Необходимость этого условия очевидна. Достаточность. Рассмотрим тождество

$$(E + A + \dots + A^k)(E - A) = E - A^{k+1}.$$

Умножив его справа на $(E - A)^{-1}$, получим

$$E + A + \dots + A^k = (E - A)^{-1} + A^{k+1}(E - A)^{-1}.$$

Отсюда следует, что, так как $A^{k+1} \rightarrow 0$ при $k \rightarrow \infty$, то

$$E + A + \dots + A^k + \dots = (E - A)^{-1}.$$

§ 4. Точные методы линейной алгебры. Метод Гаусса

Название параграфа имеет в виду, что погрешность описываемых методов отсутствует. Но при этом не следует забывать, что неустранимая погрешность, полученная за счет погрешности исходных данных и погрешности округлений, безусловно будет иметь место, и чем больше арифметических действий будет содержать метод, тем больше будет указанная погрешность.

Метод Гаусса представляет из себя метод последовательных исключений неизвестных и имеет много вычислительных схем. Приведем схемы, чаще всего используемые при машинном счете, а также те схемы, которые потребуются нам для решения ряда задач математической физики.

Большую часть схем изложим на примере системы трех уравнений с тремя неизвестными. Будем также иметь в виду, что основной определитель этих систем не равен нулю, решение системы существует и единственно. Отдельно придется рассмотреть системы, у которых определитель близок к нулю. Эти системы называются плохо обусловленными. Для их решения следует применять специальные приемы.

1. Схема единственного деления

Рассмотрим схему на примере системы трех уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = a_{14}; & (2.17) \end{cases}$$

$$\begin{cases} a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = a_{24}; & (2.18) \end{cases}$$

$$\begin{cases} a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = a_{34}. & (2.19) \end{cases}$$

Предположим, что $a_{11} \neq 0$, и разделим на него коэффициенты уравнения (2.17).

Обозначим $\frac{a_{ik}}{a_{11}} = b_{ik}^{(1)}$, тогда

$$x_1 + b_{12}^{(1)} x_2 + b_{13}^{(1)} x_3 = b_{14}^{(1)}. \quad (2.20)$$

Чтобы исключить из системы x_1 , умножим (2.20) на a_{21} и вычтем из уравнения (2.18). Затем умножим (2.20) на a_{31} и вычтем из (2.19). В результате получим

$$a_{22}^{(1)} x_2 + a_{23}^{(1)} x_3 = a_{24}^{(1)};$$

$$a_{32}^{(1)} x_2 + a_{33}^{(1)} x_3 = a_{34}^{(2)},$$

где

$$a_{ik}^{(1)} = a_{ik} - a_{11} b_{ik}^{(1)}.$$

Таким образом, сделан один шаг. Процесс повторяют до тех пор, пока не будут исключены все неизвестные, кроме одного. В нашем случае он кончится уравнением

$$x_3 = b_{34}^{(3)},$$

где

$$b_{34}^{(3)} = \frac{a_{34}^{(2)}}{a_{33}^{(2)}}, \quad a_{34}^{(2)} = a_{34}^{(1)} - a_{32} b_{23}^{(2)}.$$

Можно привести общий вид формул

$$b_{ik}^{(j+1)} = \frac{a_{ik}^{(j)}}{a_{ii}^{(j)}}; \quad a_{ik}^{(j)} = a_{ik}^{(j-1)} - a_{ij}^{(j-1)} b_{jk}^{(j)}.$$

Исходная система, таким образом, сводится к системе с треугольной матрицей

$$\left. \begin{aligned} x_1 + b_{12}^{(1)} x_2 + b_{13}^{(1)} x_3 &= b_{14}^{(1)}, \\ x_2 + b_{23}^{(2)} x_3 &= b_{24}^{(2)}, \\ x_3 &= b_{34}^{(3)}. \end{aligned} \right\}$$

Из этой системы находятся неизвестные x_1, x_2, x_3 . Результаты вычислений записываются в табл. 12.

a_{i1}	a_{i2}	a_{i3}	a_{ik}	Контроль
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
1	$b_{12}^{(1)}$	$b_{13}^{(1)}$	$b_{14}^{(1)}$	$b_{15}^{(1)}$
	$a_{22}^{(1)}$	$a_{23}^{(1)}$	$a_{24}^{(1)}$	$a_{25}^{(1)}$
	$a_{32}^{(1)}$	$a_{33}^{(1)}$	$a_{34}^{(1)}$	$a_{35}^{(1)}$
	1	$b_{23}^{(2)}$	$b_{24}^{(2)}$	$b_{25}^{(2)}$
		$a_{23}^{(2)}$	$a_{24}^{(2)}$	$a_{25}^{(2)}$
		1	$b_{34}^{(2)}$	$b_{35}^{(2)}$
		x_3		$\tilde{x}_3$
	x_2			$\tilde{x}_2$
x_1				$\tilde{x}_1$

Кроме описанных действий существуют еще контрольные, для которых отведен последний столбец табл. 12. Первые три числа, входящие в него, a_{15} , a_{25} , a_{35} являются суммами по строке: $a_{i5} = a_{i1} + a_{i2} + a_{i3} + a_{i4}$. Следующие величины входящие в контрольный столбец вычисляются по тому же правилу что и величины, стоящие в одной с ней строке. Например, $a_{15}^{(j)} = a_{15}^{(j-1)} - a_{1j}^{(j-1)} b_{j5}^{(j)}$.

Контроль заключается в том, что полученные числа должны снова дать сумму по строке. При нахождении неизвестных элементы контрольного столбца заменяют правые части

$$\begin{aligned}\tilde{x}_1 + b_{12}^{(1)} \tilde{x}_2 + b_{13}^{(1)} \tilde{x}_3 &= b_{15}^{(1)}, \\ \tilde{x}_2 + b_{23}^{(2)} \tilde{x}_3 &= b_{25}^{(2)}, \\ \tilde{x}_3 &= b_{35}^{(3)}.\end{aligned}$$

Контроль заключается в том, что неизвестные x_i связаны с x_i соотношением $x_i + 1 = x_i$.

Пример. Найти решение системы

$$\begin{aligned} 22,14x_1 + 7,42x_2 - 3,13x_3 + 6,22x_4 &= 3,46, \\ -2,94x_1 - 15,97x_2 + 11,29x_3 + 5,81x_4 &= -32,70, \\ 7,27x_1 + 7,88x_2 - 14,95x_3 + 4,73x_4 &= 8,85, \\ -3,71x_1 + 5,73x_2 + 4,39x_3 - 12,58x_4 &= 21,76 \end{aligned}$$

с помощью схемы единственного деления.

Решение приводится в виде табл. 13.

Таблица 13

a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}	Контроль
22,14	- 7,42	- 3,13	6,22	3,40	21,21
-2,94	-15,97	11,29	5,81	-32,70	-34,51
7,27	7,88	-14,95	4,73	8,85	13,78
-3,71	5,73	4,39	-12,58	21,76	15,59
1	-0,335514	-0,14137	0,28094	0,15357	0,95799
0	-16,95531	10,87437	6,63596	-32,24850	-31,69351
0	10,31617	-13,92224	2,68757	7,73955	6,81541
0	4,48663	3,86552	-11,53771	22,32974	19,14414
	1	-0,64135	-0,39138	1,90197	1,86924
	0	-7,30577	6,72523	-11,88807	-12,46855
	0	6,74302	-9,78173	13,79630	10,75755
		1	-0,92055	1,62722	1,70667
		0	-3,57444	2,82392	-0,75056
		1	1	-0,79003	0,20997
		1		0,89996	1,89996
	1			2,16996	3,16996
1				1,22999	2,22999

2. Схема главного элемента

Схема главного элемента построена так, чтобы погрешность арифметических действий была минимальной. Известно, что относительная погрешность резко возрастает при вычитании близких друг к другу чисел. В прямом ходе схемы Гаусса каждый шаг заканчивается вычитанием, поэтому можно предположить, что наименьшая относительная погрешность окажется у коэффициента, наибольшего по абсолютной величине. Выберем этот коэффициент и на следующем шаге будем делить на него остальные коэффициенты строки. Очевидно, погрешность этого коэффициента войдет как составная часть в погрешность других коэффициентов преобразованной строки, а также во все коэффициенты системы, полученной после проведения очередного шага прямого хода. Таким образом, по сравнению со схемой единственного деления, мы изменяем лишь выбор номера коэффициента и номер уравнения, так как надо выбирать наибольший по модулю коэффициент всей матрицы, полученной на каждом шаге, начиная с исходной системы.

Рассмотрим метод сначала на элементарном примере. Пусть дана система

$$x_1 + 2x_2 - x_3 = 2,$$

$$x_1 + x_2 + 3x_3 = 5,$$

$$2x_1 + x_2 + x_3 = 4.$$

Так же как и в предыдущей схеме, составим табл. 14 из пяти столбцов, впишем коэффициенты, свободные члены и суммы по строчке. Дополнительно, кроме указанных столбцов, введем слева шестой и обозначим его m_i .

Выберем среди коэффициентов a_{i1} , a_{i2} , a_{i3} наибольший по модулю $a_{23}=3$ и разделим на него коэффициенты, стоящие в том же столбце. Результаты деления, взятые с обратным знаком, $m_1 = -0,333$, $m_2 = -0,33$ впишем в дополнительный столбец. Умножим коэффициенты строчки, содержащей наибольший главный элемент, на m_1 и m_2 и запишем результат в две последующие строчки. Складывая полученные коэффициенты и коэффициентами первой и третьей строчки, исключаем из системы x_3 и получаем систему двух уравнений

$$1,333x_1 + 2,333x_2 = 3,667,$$

$$1,667x_1 + 0,667x_2 = 2,333.$$

Выберем за главный элемент коэффициент при x_2 в первом уравнении 2,333 и повторим описанный процесс.

m_i	a_{i1}	a_{i2}	a_{i3}	a_{i4}	Контроль
0,333	1,000	2,000	-1,000	2,000	4,000
	1,000	1,000	3,000	5,000	10,000
-0,333	2,000	1,000	1,000	4,000	8,000
	0,333	0,333	1,000	1,667	3,333
	-0,333	-0,333	-1,000	-1,667	-3,333
	1,333	2,333	0	3,667	7,333
-0,286	1,667	0,667	0	2,333	4,667
	-0,381	-0,667	0	-1,049	-2,110
	1,286	0	0	1,286	2,262
	$x_1 = 1$				$\bar{x}_1 = 2$
		$x_2 = 1$			$\bar{x}_2 = 2$
			$x_3 = 1$		$\bar{x}_3 = 3$

3. Схема прогонки

Рассмотрим систему с трехдиагональной матрицей

$$a_k x_{k-1} + b_k x_k + c_k x_{k+1} = f_k, \\ k = 2, 3, \dots, n, \quad a_1 = c_n = 0.$$

Такие системы часто встречаются при решении различных краевых задач, к которым сводятся нестационарные задачи теории теплопроводности и задачи, связанные с различными волновыми процессами. Применим к системе схему единственного деления, для удобства рассматривая систему четырех уравнений

$$\left. \begin{aligned} b_1 x_1 + c_1 x_2 &= f_1, \\ a_2 x_1 + b_2 x_2 + c_2 x_3 &= f_2, \\ a_3 x_2 + b_3 x_3 + c_3 x_4 &= f_3, \\ a_4 x_3 + b_4 x_4 &= f_4. \end{aligned} \right\} \quad (2.21)$$

Построим табл. 15, как описано в пункте первом.

Таблица 15

1	2	3	4	f	Контроль	№ строчки
b_1	c_1			f_1		(1)
a_2	b_2	c_2		f_2		(2)
	a_3	b_3	c_3	f_3		(3)
		a_4	b_4	f_4		(4)
1	c'_1			f'_1		(5)
	$\bar{b}_2$	c_2		$\bar{f}_2$		(6)
	1	c'_2		f'_2		(7)
		$\bar{b}_3$	c_3	$\bar{f}_3$		(8)
		1	c'_3	f'_3		(9)
			$\bar{b}_4$	$\bar{f}_4$		(10)
			1	$\bar{f}'_4$		(11)

Номера, стоящие в скобках справа, определяют строчки схемы. Опишем еще один раз схему в применении к нашему случаю.

Разделим элементы строчки (1) на b_1 , получим строчку (5), где $c'_1 = \frac{c_1}{b_1}$ и $f'_1 = \frac{f_1}{b_1}$. Домножим элементы строчки (3) на a_2 и вычтем полученную строчку из второй. Результатом является строчка (6), где $\bar{b}_2 = b_2 - a_2 c'_1$; $\bar{f}_2 = f_2 - a_2 f'_1$. Повторяя описанный процесс, получим строчки (7) и (8), где

$$c'_2 = \frac{c_2}{b_2 - a_2 c'_1}, \quad f'_2 = \frac{f_2 - a_2 f'_1}{b_2 - a_2 c'_1}.$$

Аналогично

$$c'_3 = \frac{c_3}{b_3 - a_3 c'_2}, \quad f'_3 = \frac{f_3 - a_3 f'_2}{b_3 - a_3 c'_2}.$$

И, наконец, разделив элементы строчки (10) на $\bar{b}_4$, построим последнюю (11), где

$$f'_4 = \frac{f_4 - a_4 f'_3}{b_4 - a_4 c'_3}.$$

Таким образом, система (2.21) приведена к виду

$$\left. \begin{aligned} x_1 + c'_1 x_2 &= f'_1, \\ x_2 + c'_2 x_3 &= f'_2, \\ x_3 + c'_3 x_4 &= f'_3, \\ x_4 &= f'_4, \end{aligned} \right\}$$

откуда можно найти обратным ходом неизвестные:

$$x_4 = f'_4, \quad x_3 = f'_3 - c'_3 x_4, \quad x_2 = f'_2 - c'_2 x_3, \quad x_1 = f'_1 - c'_1 x_2.$$

Описанная схема в случае большого числа уравнений решается по формулам, которые легко выписать:

$$\begin{aligned} f'_1 &= \frac{f_1}{b_1}, & c'_1 &= \frac{c_1}{b_1}, & x_n &= f'_n, \\ f'_k &= \frac{f_k - a_k f'_{k-1}}{b_k - a_k c'_{k-1}}, & c'_k &= \frac{c_k}{b_k - a_k c'_{k-1}}, & x_k &= f'_k - c'_k x_{k+1}, \end{aligned} \quad (2.21')$$

а запись результатов можно, например, предложить следующую:

$$\left| \begin{array}{cccc} a_2 & a_3 & a_4 & \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & \\ f_1 & f_2 & f_3 & f_4 \end{array} \right| \left| \begin{array}{cccc} 1 & 1 & 1 & \\ c'_1 & c'_2 & c'_3 & 1 \\ f'_1 & f'_2 & f'_3 & f'_4 \end{array} \right| \left| \begin{array}{cccc} x_4 & x_3 & x_2 & x_1 \end{array} \right|$$

Следует дополнительно отметить, что весь описанный метод может быть применен к векторным системам с коэффициентами из матриц. Так, система

$$A_k X_{k-1} + B_k X_k + C_k X_{k+1} = F_k, \quad (2.22)$$

где $k = 1, 2, \dots, m$; $A_1 = C_m = 0$;

$$\begin{aligned} A_k &= \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \dots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \dots & a_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ a_{n1}^{(k)} & a_{n2}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix}; & X_k &= \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{pmatrix}; \\ B_k &= \begin{pmatrix} b_{11}^{(k)} & b_{12}^{(k)} & \dots & b_{1n}^{(k)} \\ b_{21}^{(k)} & b_{22}^{(k)} & \dots & b_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ b_{n1}^{(k)} & b_{n2}^{(k)} & \dots & b_{nn}^{(k)} \end{pmatrix}; & F_k &= \begin{pmatrix} f_1^{(k)} \\ f_2^{(k)} \\ \vdots \\ f_n^{(k)} \end{pmatrix}; \\ C_k &= \begin{pmatrix} c_{11}^{(k)} & c_{12}^{(k)} & \dots & c_{1n}^{(k)} \\ c_{21}^{(k)} & c_{22}^{(k)} & \dots & c_{2n}^{(k)} \\ \dots & \dots & \dots & \dots \\ c_{n1}^{(k)} & c_{n2}^{(k)} & \dots & c_{nn}^{(k)} \end{pmatrix}. \end{aligned}$$

может быть с помощью метода прогонки приведена к виду

$$X_k + C'_k X_{k+1} = F'_k, \quad k = 1, 2, \dots, m;$$

$$C'_1 = B_1^{-1} C_1; \quad F'_1 = B^{-1} F_1;$$

где

$$C'_k = (B_k - A_k C'_{k-1})^{-1} C_k, \quad F'_k = (B_k - A_k C'_{k-1})^{-1} (F_k - A_k F'_{k-1}).$$

Значения искомых векторов легко получить с помощью обратного хода:

$$X_m = F'_m, \quad X_k = F'_k - C'_k X'_{k+1}. \quad (2.23)$$

Полученные в данном пункте формулы будут использоваться в дальнейшем.

4. Схема без обратного хода

Схема единственного деления может быть модернизирована так, чтобы нахождение неизвестных могло производиться непосредственно после прямого хода. Если схема единственного деления преобразует квадратную матрицу в треугольную, то схема без обратного хода преобразует ее к единичной, т. е. система $AX = B$ сводится к системе $X = F$.

Опишем схему снова на примере трех уравнений

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = a_{14},$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = a_{24},$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = a_{34}.$$

Выпишем схему в виде табл. 16, снабдив строчки номерами.

Первые шесть строчек табл. 16 получены так же, как соответствующие строчки схемы единственного деления. После получения строчки (6) выбираем строчку (5) и делим все ее элементы на $a_{22}^{(1)}$. Получаем строчку (8) и записываем ее на втором месте второго шага, пропуская строчку с номером (7). Затем умножаем строчку (8) на $b_{12}^{(1)}$ и вычитаем ее из строчки (4). Получаем строчку (7).

Аналогично, умножив строчку (8) на $a_{32}^{(1)}$ и вычтя результат из строчки (6), имеем строчку (9). Последний третий шаг схемы проводится, начиная с девятого уравнения, третьего в третьем шаге. Из схемы видно, что на каждом шаге строится один столбец единичной матрицы. Следовательно, в третьем шаге мы получим последний третий столбец единичной матрицы. Столбец правых частей в последнем шаге дает окончательный результат:

$$x_1 = b_{14}^{(3)}, \quad x_2 = b_{24}^{(3)}, \quad x_3 = b_{34}^{(3)}.$$

Таблица 16

a_{i1}	a_{i2}	a_{i3}	a_{i4}	Контроль	№ строчки
a_{11}	a_{12}	a_{13}	a_{14}		(1)
a_{21}	a_{22}	a_{23}	a_{24}		(2)
a_{31}	a_{32}	a_{33}	a_{34}		(3)
1	$b_{12}^{(1)}$	$b_{13}^{(1)}$	$b_{14}^{(1)}$		(4)
0	$a_{22}^{(1)}$	$a_{23}^{(1)}$	$a_{24}^{(1)}$		(5)
0	$a_{32}^{(1)}$	$a_{33}^{(1)}$	$a_{34}^{(1)}$		(6)
1	0	$b_{13}^{(2)}$	$b_{14}^{(2)}$		(7)
0	1	$b_{23}^{(2)}$	$b_{24}^{(2)}$		(8)
0	0	$a_{33}^{(2)}$	$a_{34}^{(2)}$		(9)
1	0	0	$b_{14}^{(3)}$		(10)
0	1	0	$b_{21}^{(3)}$		(11)
0	0	1	$b_{34}^{(3)}$		(12)

Легко указать рекуррентные формулы, по которым ведется расчет

$$a_{ik}^{(j)} = a_{ik}^{(j-1)} - b_{ij}^{(j)} a_{jk}^{(j-1)}, \quad i = m+1, \dots, n;$$

$$b_{ik}^{(j)} = a_{ik}^{(j-1)} - b_{ij}^{(j)} b_{jk}^{(j-1)}, \quad i = 1, 2, \dots, m-1;$$

$$b_{mk}^{(j)} = \frac{a_{mk}^{(j-1)}}{a_{mm}^{(j-1)}}.$$

С помощью схемы без обратного хода можно также решать системы с одинаковыми правыми частями и различными левыми. При этом следует только ввести дополнительные столбцы для всех правых частей соответственно количеству систем и проводить преобразования, как было описано выше. После окончания процесса мы получим в этих столбцах решения для наших систем.

Пример. Решить систему

$$-0,61x_1 - 1,04x_2 - 0,97x_3 + 0,69x_4 = -1,04,$$

$$0,57x_1 - 0,73x_2 + 1,05x_3 + 1,08x_4 = 1,24,$$

$$-1,15x_1 - 0,83x_2 - 0,65x_3 - 0,74x_4 = 1,13,$$

$$1,00x_1 - 1,00x_2 + 1,00x_3 - 1,00x_4 = 1,09.$$

Решение приведем в виде табл. 17.

Таблица 17

a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}
-0,61	-1,04	-0,97	0,69	-1,04
0,57	-0,73	1,05	1,08	1,21
-1,15	-0,83	-0,65	-0,74	1,13
-1,00	-1,00	1,00	-1,00	1,09
1	1,70492	1,59016	-1,13135	1,70492
0	-1,70180	0,14361	1,72476	0,23820
0	1,13066	1,17868	-2,04082	3,09066
0	0,70492	2,59016	-2,13115	2,79492
1	0	1,73405	0,59677	1,94356
0	1	-0,08439	-1,01349	-0,13997
0	0	1,27410	-0,89491	3,24892
0	0	2,64966	-1,41672	2,89359
1	0	0	1,81475	-2,47822
0	1	0	-1,07276	0,07522
0	0	1	-0,702239	2,54997
0	0	0	0,44437	-3,86296
1	0	0	0	13,29762
0	1	0	0	-9,25050
0	0	1	0	-3,55599
0	0	0	1	-8,69312

5. Обращение матрицы

Будем искать обратную матрицу методом Гаусса. Для этого напомним основное равенство, определяющее обратную матрицу: $AA^{-1} = E$.

Обозначим элементы j -го столбца искомой матрицы через $x_1^{(j)}$, $x_2^{(j)}$, ..., $x_n^{(j)}$ и выпишем j -й столбец этого матричного равен-

ства, где правыми частями являются элементы j -го столбца единичной матрицы

$$\begin{aligned} a_{11} x_1^{(j)} + a_{12} x_2^{(j)} + \dots + a_{1n} x_n^{(j)} &= 0, \\ a_{21} x_1^{(j)} + a_{22} x_2^{(j)} + \dots + a_{2n} x_n^{(j)} &= 0, \\ \dots & \\ a_{j1} x_1^{(j)} + a_{j2} x_2^{(j)} + \dots + a_{jn} x_n^{(j)} &= 1, \\ \dots & \\ a_{n1} x_1^{(j)} + a_{n2} x_2^{(j)} + \dots + a_{nn} x_n^{(j)} &= 0. \end{aligned}$$

Очевидно, для нахождения неизвестных $x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)}$ следует решить построенную систему. Аналогично элементы других столбцов матрицы A^{-1} получаются, как решение n систем с одинаковыми левыми частями, и с правыми частями, являющимися столбцами единичной матрицы. Приведем расчетную схему на примере матрицы с тремя столбцами и с тремя строчками

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} \\ x_3^{(1)} & x_3^{(2)} & x_3^{(3)} \end{pmatrix}$$

Все три системы, соответствующие трем столбцам равенства $AA^{-1} = E$, могут быть сведены в табл. 18.

Табл. 18 соответствует схеме без обратного хода. Мы не будем повторять ее описание, а приведем только рекуррентные формулы для расчета обратной матрицы порядка n :

$$b_{ik}^{(j)} = \frac{a_{ik}^{(j-1)}}{a_{ii}^{(j-1)}}, \quad j = 1, 2, \dots, n; \quad k \leq n + j; \quad i = j;$$

$$b_{ik}^{(j)} = \frac{1}{a_{ii}^{(j-1)}}, \quad j = 1, 2, \dots, n; \quad k = n + j; \quad i = j;$$

$$b_{ik}^{(j)} = b_{ik}^{(j-1)} - b_{jj}^{(j-1)} b_{ik}^{(j)}, \quad j = 2, 3, \dots, n; \quad k < n + j; \quad i < j;$$

$$b_{ik}^{(j)} = -b_{ij}^{(j-1)} b_{jk}^{(j)}, \quad j = 2, 3, \dots, n; \quad k = n + j; \quad i < j;$$

$$a_{ik}^{(j)} = a_{ik}^{(j-1)} - a_{ij}^{(j-1)} b_{jk}^{(j)}, \quad j = 1, 2, \dots, n-1; \quad i > j; \quad k < j + n;$$

$$a_{ik}^{(j)} = -a_{ij}^{(j-1)} b_{jk}^{(j)}, \quad j = 1, 2, \dots, n-1; \quad i > j; \quad k = j + n.$$

Кроме того, укажем на необходимость при машинном счете опускать при записи столбцы единичной матрицы, что сократит вдвое таблицу.

Таблица 18

a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}	a_{i6}
a_{11}	a_{12}	a_{13}	1	0	0
a_{21}	a_{22}	a_{23}	0	1	0
a_{31}	a_{32}	a_{33}	0	0	1
1	$b_{12}^{(1)}$	$b_{13}^{(1)}$	$b_{14}^{(1)}$	0	0
0	$b_{22}^{(1)}$	$b_{23}^{(1)}$	$b_{24}^{(1)}$	1	0
0	$b_{32}^{(1)}$	$b_{33}^{(1)}$	$b_{34}^{(1)}$	0	1
1	0	$b_{13}^{(2)}$	$b_{14}^{(2)}$	$b_{15}^{(2)}$	0
0	1	$b_{23}^{(2)}$	$b_{24}^{(2)}$	$b_{25}^{(2)}$	0
0	0	$b_{33}^{(2)}$	$b_{34}^{(2)}$	$b_{35}^{(2)}$	1
1	0	0	$b_{14}^{(3)}$	$b_{15}^{(3)}$	$b_{16}^{(3)}$
0	1	0	$b_{24}^{(3)}$	$b_{25}^{(3)}$	$b_{26}^{(3)}$
0	0	1	$b_{34}^{(3)}$	$b_{35}^{(3)}$	$b_{36}^{(3)}$

6. Вычисление определителя методом Гаусса

Рассмотрим определитель n -го порядка

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Используя свойства определителя, применим для его вычисления следующие преобразования. Вынесем за знак определителя a_{11} , разделив на него все элементы первой строчки. Затем

последовательно вычтем из второй строчки первую, умноженную на a_{21} , из третьей строчки — вторую, умноженную на a_{31} , и т. д. В результате получим определитель, у которого первый столбец содержит только один не нулевой элемент — единицу, стоящую в первой строчке. Разложим определитель по первому столбцу и получим один определитель $(n-1)$ -го порядка

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} & \dots & \frac{a_{1n}}{a_{11}} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} =$$

$$= a_{11} \begin{vmatrix} 1 & b_{12}^{(1)} & \dots & b_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \dots & \dots & \dots & \dots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} \\ a_{32}^{(1)} & a_{33}^{(1)} & \dots & a_{3n}^{(1)} \\ \dots & \dots & \dots & \dots \\ a_{n2}^{(1)} & a_{n3}^{(1)} & \dots & a_{nn}^{(1)} \end{vmatrix};$$

$$b_{ik}^{(1)} = \frac{a_{ik}}{a_{11}}, \quad k = 2, 3, \dots, n; \quad i = 1;$$

$$a_{ik}^{(1)} = a_{ik} - a_{i1} b_{1k}^{(1)}, \quad k = 2, 3, \dots, n; \quad i = 2, 3, \dots, n.$$

Повторяя преобразования, получим:

$$\Delta = a_{11} a_{22}^{(1)} \dots a_{nn}^{(n-1)};$$

$$a_{ii}^{(j)} = a_{ii}^{(j-1)} - a_{ij}^{(j-1)} b_{ji}^{(j)}, \quad j = 1, 2, \dots, n-1; \quad i = 2, 3, \dots, n;$$

$$b_{ik}^{(j)} = \frac{a_{ik}^{(j-1)}}{a_{ii}^{(j-1)}}, \quad j = 1, 2, \dots, n-1; \quad i = 2, 3, \dots, n;$$

$$k = 2, 3, \dots, n.$$

Все преобразования повторяют прямой ход схемы единственного деления и могут быть сведены в табл. 19 (для случая $n=3$).

Таблица 19

a_{i1}	a_{i2}	a_{i3}
a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}
1	$b_{12}^{(1)}$	$b_{13}^{(1)}$
	$a_{22}^{(1)}$	$a_{23}^{(1)}$
	$a_{32}^{(1)}$	$a_{33}^{(1)}$
	1	$b_{23}^{(2)}$
		$a_{33}^{(2)}$

§ 5. Итерационные методы

1. Метод простой итерации

Метод итерации предполагает решение линейной системы $AX = B$ с помощью рекуррентной формулы

$$X_k = F_k(X_0, X_1, \dots, X_{k-1}),$$

где X_0 — произвольное значение вектора X , а $X_1, X_2, \dots, X_k$ — последовательные приближения к решению системы. При применении этого метода возникают два вопроса: как построить рекуррентную формулу и будет ли последовательность $X_0, X_1, \dots, X_k \dots$ сходиться к решению системы?

Рассмотрим один из простейших приемов построения. Пусть система $AX = B$ преобразована к виду: $X = KX + F$. Такое преобразование можно провести, например, следующим образом:

$$(E - E + A)X = B, \quad X = (E - A)X + B.$$

Или, переходя к обычной записи системы, использовать такой прием:

[illegible]

В первом случае матрица K имеет вид

$$K = \begin{pmatrix} 1-a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & 1-a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & 1-a_{nn} \end{pmatrix},$$

а во втором —

$$K = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n}}{a_{22}} \\ \dots & \dots & \dots & \dots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & 0 \end{pmatrix}.$$

Выбор способа построения матрицы K определяется требованиями наиболее быстрой сходимости последовательности $X_0, X_1, \dots, X_k, \dots$ к решению системы. Легко показать, что если данная последовательность имеет предел, то этот предел и является решением системы.

Метод простой итерации очень прост в применении. Выбирают начальное приближение X_0 , чаще всего за него принимают свободные

члены системы. Подставляют эти значения в правые части системы (2.24) и получают в левых частях новые значения неизвестных — первые приближения. Описанный процесс повторяется до тех пор, пока максимальное значение разностей $|x_i^{(k+1)} - x_i^{(k)}|$ станет меньше принятой погрешности. Процесс простой итерации описывается формулами:

$$\begin{aligned} X_1 &= KX_0 + F; \\ X_2 &= KX_1 + F; \\ &\dots \dots \dots \\ X_n &= KX_{n-1} + F. \end{aligned}$$

Построим формулу, выражающую n -ю итерацию через нулевую:

$$\begin{aligned} X_1 &= KX_0 + F; \\ X_2 &= KX_1 + F = K^2X_0 + (K + E)F; \\ &\dots \dots \dots \\ X_n &= K^nX_0 + (K^{n-1} + K^{n-2} + \dots + E)F. \end{aligned} \quad (2.25)$$

На основании формулы (2.25) можно предложить другой способ расчета по методу простой итерации. Примем, как предлагалось выше, за нулевое приближение правые части системы. В этом случае формула (2.25) запишется в виде:

$$X_n = (K^n + K^{n-1} + \dots + E)F$$

или

$$\begin{aligned} X_1 - X_0 &= KF = \Delta X_1; \\ X_2 - X_1 &= K^2F = K(KF) = \Delta X_2; \\ &\dots \dots \dots \\ X_n - X_{n-1} &= K(K^{n-1}F) = \Delta X_n. \end{aligned}$$

Следовательно, можно вычислять не итерации, а их разности ΔX_n , причем каждая последующая разность получается из предыдущей умножением слева на матрицу K . Расчет заканчивается тогда, когда составляющие векторов ΔX_n не будут превосходить погрешность счета. Окончательный ответ получается по формуле

$$X_n = X_0 + \Delta X_1 + \dots + \Delta X_n.$$

Рассмотрим теперь ряд теорем, определяющих условия сходимости метода простой итерации, скорость сходимости и оценку погрешности.

Теорема 1. Для того чтобы процесс простой итерации сходился при любом начальном приближении и любой правой части необходимо и достаточно, чтобы все собственные числа матрицы K были по модулю меньше единицы.

Доказательство. Рассмотрим формулу, определяющую n -ю итерацию:

$$X_n = K^n X_0 + (K^{n-1} + K^{n-2} + \dots + E) F. \quad (2.26)$$

Чтобы последовательность $X_0, X_1, \dots, X_n$ имела предел, необходимо и достаточно, чтобы сходился ряд: $E + K + K^2 + \dots + K^n + \dots$, так как при этом первое слагаемое формулы (2.26) стремится к нулю. Применим теорему 3 § 3, которая гласит, что необходимым и достаточным условием сходимости указанного ряда является требование, чтобы все собственные числа матрицы K были по модулю меньше единицы. Таким образом, теорема доказана.

Аналогично доказывается следующая теорема.

Теорема 2. Для того чтобы итерационный метод в применении к системе $X = KX + F$ сходился при любых F и X_0 достаточно, чтобы хотя бы одна из норм матрицы K была меньше единицы.

К сожалению, практически удобно применять только достаточное условие, так как собственные числа матрицы отыскивать сложно. Укажем два признака сходимости итерационного процесса:

$$1) \max_{(i)} \sum_{j=1}^n |\alpha_{ij}| < 1;$$

$$2) \max_{(j)} \sum_{i=1}^n |\alpha_{ij}| < 1.$$

Приведем теперь две формулы, дающие оценку скорости сходимости итерационного процесса и оценку погрешности. Рассмотрим систему в виде, удобном для итерирования:

$$X = KX + F,$$

и формулу, определяющую n -ю итерацию:

$$X_n = KX_{n-1} + F.$$

Вычтем из первого уравнения второе

$$X - X_n = K(X - X_{n-1}).$$

Откуда, оценивая по норме, получим

$$\|X - X_n\| \leq \|K\| \|X - X_{n-1}\|.$$

Очевидно, чем меньше норма матрицы K , тем скорее сходимость процесса простой итерации.

Используем теперь формулу, определяющую n -ю итерацию через нулевую:

$$X_n = K^n X_0 + (E + K + \dots + K^{n-1}) F,$$

и рассмотрим решение системы, как сумму ряда

$$X = (E + K + \dots + K^n + \dots) F. \quad (2.27)$$

В таком случае разность между точным значением решения и n -й итерацией запишется в виде

$$X - X_n = -K^n X_0 + (K^n + K^{n+1} + \dots) F,$$

где второе слагаемое является остатком ряда (2.27). Оценивая результат по норме, имеем

$$\|X - X_n\| \leq \|K\|^n \|X_0\| + (\|K\|^n + \|K\|^{n+1} + \dots) \|F\|$$

или, суммируя ряд из норм, получим окончательную формулу, оценивающую погрешность на каждом шаге итерации:

$$\|X - X_n\| \leq \|K\|^n \|X_0\| + \frac{\|K\|^n \|F\|}{1 - \|K\|}.$$

Можно заметить, что если сходимость зависит только от матрицы K , то погрешность связана и с правыми частями и с начальным приближением.

В заключение описания метода простой итерации покажем, как можно преобразовать систему, чтобы метод итерации стал сходящимся. Рассмотрим систему в виде $AX = B$ и применим левое преобразование Гаусса

$$A^T A X = A^T B. \quad (2.28)$$

Очевидно, матрица $A^T A$ положительно определенная и все ее собственные числа положительны. Пусть они лежат в промежутке $(0, \mu)$, где за μ можно принять первую или вторую норму матрицы $A^T A$. Это справедливо, так как третья норма является наименьшей. Умножим систему (2.28) на постоянный множитель m , величину которого мы подберем впоследствии,

$$mA^T A X = mA^T B.$$

Преобразуем систему к виду, удобному для итерирования,

$$X = (E - mA^T A) X + mA^T B.$$

Если принять $m = \frac{2}{\mu}$, то собственные числа матрицы $E - \frac{2}{\mu} A \bar{A}$ будут лежать в промежутке $(-1, +1)$ и условие сходимости выполняется. Следует, однако, заметить, что указанное преобразование практически применяется редко из-за большого числа дополнительных арифметических действий, которые сильно увеличивают погрешность.

2. Метод Зейделя

Метод Зейделя отличается от метода простой итерации только тем, что при вычислении компоненты $x_i^{(k+1)}$ учитываются уже вычисленные раньше компоненты $x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_{i-1}^{(k+1)}$, т. е. вычисления ведутся по формуле

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

Для примера впишем $(k+1)$ -ю итерацию всех составляющих вектора X в случае системы трех уравнений

$$x_1^{(k+1)} = a_{11} x_1^{(k)} + a_{12} x_2^{(k)} + a_{13} x_3^{(k)},$$

$$x_2^{(k+1)} = a_{21} x_1^{(k+1)} + a_{22} x_2^{(k)} + a_{23} x_3^{(k)},$$

$$x_3^{(k+1)} = a_{31} x_1^{(k+1)} + a_{32} x_2^{(k+1)} + a_{33} x_3^{(k)}.$$

Бывают случаи, когда метод итерации сходится, а метод Зейделя расходится, и наоборот: метод Зейделя сходится, а метод итерации расходится. Чтобы сравнить эти методы, установим их связь. Для этого преобразуем систему $X = KX + F$ следующим образом:

$$X = (K_1 + K_2) X + F,$$

где

$$K_1 = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}; \quad K_2 = \begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & 0 \end{pmatrix}.$$

В этом случае $(n+1)$ -я итерация, подсчитанная по методу Зейделя, определяется формулой

$$X_{n+1} = K_1 X_n + K_2 X_{n+1} + F.$$

Отсюда

$$(E - K_2) X_{n+1} = K_1 X_n + F$$

или

$$X_{n+1} = (E - K_2)^{-1} K_1 X_n + (E - K_2)^{-1} F.$$

Зейделевский процесс, примененный к системе с матрицей $K = K_1 + K_2$, эквивалентен итерационному, примененному к системе с матрицей $(E - K_2)^{-1} K_1$. Если рассматривать только вопросы сходимости, то можно не обращать внимания на правые части.

Приведем две теоремы, определяющие сходимость зейделевского процесса непосредственно.

Теорема 3. Если матрица A системы $AX = F$ положительно определена, то метод Зейделя для этой системы сходится в случае, если преобразование системы проведено по формулам (2.24).

Теорема 4. Если система уравнений преобразована к виду $X = KX + F$ и первая норма матрицы K меньше единицы, то зейделевский процесс сходится, причем быстрее итерационного. При этом имеет место неравенство

$$\|X - X_k\|_1 \leq \mu \|X - X_{k-1}\|_1,$$

где $0 < \mu \leq \|K\|_1 < 1$.

Доказательство. Рассмотрим k -ю итерацию, записанную для i -го уравнения системы

$$x_i^{(k)} = \sum_{j=1}^{i-1} \alpha_{ij} x_j^{(k)} + \sum_{j=i}^n \alpha_{ij} x_j^{(k-1)} + \beta_i.$$

Обозначим:

$$\sum_{j=1}^{i-1} |\alpha_{ij}| = \delta_i, \quad \sum_{j=i}^n |\alpha_{ij}| = \gamma_i.$$

По условию теоремы $\delta_i + \gamma_i < 1$. Выпишем уравнение, определяющее i -ю неизвестную

$$x_i = \sum_{j=1}^{i-1} \alpha_{ij} x_j + \sum_{j=i}^n \alpha_{ij} x_j + \beta_i,$$

и вычтем из него уравнение, определяющее k -ю итерацию x_i :

$$x_i - x_i^{(k)} = \sum_{j=1}^{i-1} \alpha_{ij} (x_j - x_j^{(k)}) + \sum_{j=i}^n \alpha_{ij} (x_j - x_j^{(k-1)}).$$

Оценим полученное по модулю

$$|x_i - x_i^{(k)}| \leq \sum_{j=1}^{i-1} |\alpha_{ij}| |x_j - x_j^{(k)}| + \sum_{j=i}^n |\alpha_{ij}| |x_j - x_j^{(k-1)}|.$$

Так как

$$\begin{aligned} |x_j - x_j^{(k)}| &\leq \|X - X_k\|_1, \\ |x_j - x_j^{(k-1)}| &\leq \|X - X_{k-1}\|_1, \end{aligned}$$

то справедливо неравенство

$$|x - x_i^{(k)}| \leq \delta_j \|X - X_k\|_1 + \gamma_i \|X - X_{k-1}\|_1.$$

Выберем из всех уравнений то, для которого составляющая $|x_i - x_i^{(k)}|$ будет наибольшей. Пусть это соответствует номеру $i = i_0$. Тогда

$$|x_{i_0} - x_{i_0}^{(k)}| = \|X - X_k\|_1.$$

Отсюда получим

$$\|X - X_k\|_1 \leq \|X - X_k\|_1 \delta_{i_0} + \|X - X_{k-1}\|_1 \gamma_{i_0}.$$

Преобразуя неравенство, имеем

$$\|X - X_k\|_1 \leq \frac{\gamma_{i_0}}{1 - \delta_{i_0}} \|X - X_{k-1}\|_1.$$

Обозначим через μ выражение $\frac{\gamma_i}{1 - \delta_i}$, взятое при том номере i , при котором оно принимает наибольшее значение. Легко проверить, что при этом

$$\frac{\gamma_i}{1 - \delta_i} \leq \gamma_i + \delta_i.$$

Чтобы проверить этот факт, преобразуем интересующее нас неравенство

$$\frac{\gamma_i}{1 - \delta_i} - \gamma_i - \delta_i \leq 0;$$

$$\frac{\delta_i(1 - \gamma_i - \delta_i)}{1 - \delta_i} \geq 0.$$

Очевидно, все три множителя: δ_i , $1 - \gamma_i - \delta_i$, и $1 - \delta_i$ не отрицательны при любом i .

Заменяя выражение $\gamma_i + \delta_i$ на $\|K\|_1$, мы только усилим неравенство. Таким образом

$$\|X - X_k\|_1 \leq \mu \|X - X_{k-1}\|_1,$$

$$\mu = \max_i \frac{\gamma_i}{1 - \delta_i}, \quad \mu \leq \|K\|_1 < 1,$$

что и доказывает теорему.

Из теоремы 4 следует, что если выполняется первое условие сходимости для итерационного метода (по первой норме), то выгоднее применять не метод простой итерации, а метод Зейделя. Но при этом надо помнить, что метод Зейделя не всегда лучше метода простой итерации.

Пример 1. Дана система, для которой метод итерации сходится, а метод Зейделя расходится. Основная матрица системы имеет вид

$$K = \begin{pmatrix} 2,3 & -5 \\ 1 & -2,3 \end{pmatrix}.$$

Построим матрицы:

$$K_1 = \begin{pmatrix} 2,3 & 5 \\ 0 & -2,3 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$$E - K_2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad (E - K_2)^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

$$(E - K_2)^{-1} K_1 = \begin{pmatrix} 2,3 & -5 \\ 2,3 & -7,3 \end{pmatrix}.$$

Легко определить, что собственные числа матрицы K : $\lambda_{1,2} = \pm 0,25$, а собственные числа матрицы $(E - K_2)^{-1} K_1$: $\lambda_1 \cong -5,7$; $\lambda_2 \cong 0,7$. По доказанной теореме следует, что итерационный метод, примененный к системе с матрицей K , сходится, а метод Зейделя — расходится, так как матрица $(E - K_2)^{-1} K_1$ имеет собственное число по модулю, большее единицы.

Пример 2. Решить систему методом Зейделя

$$x_1 = 1,388 + 0,078x_2 + 0,071x_3 + 0,002x_4 + 0,003x_5,$$

$$x_2 = 2,447 + 0,004x_1 + 0,010x_3 + 0,083x_4 + 0,051x_5,$$

$$x_3 = 3,342 + 0,014x_1 + 0,080x_2 + 0,002x_4 + 0,020x_5,$$

$$x_4 = 1,857 + 0,030x_1 + 0,001x_2 + 0,224x_3 + 0,014x_5,$$

$$x_5 = 4,623 + 0,019x_1 + 0,054x_2 + 0,099x_3 + 0,034x_4.$$

Решение представим в виде табл. 20. Ее первые пять столбцов содержат произведения коэффициентов α_{ih} на соответствующие итерации x_i^j . Шестой столбец содержит свободный член β_i , а седьмой — сумму по строчке, т. е. следующую итерацию неизвестного x_i^{j+1} , соответствующего данной строчке. По горизонтали табл. 20 разделена на итерации. Каждая содержит пять строчек, соответствующих искомым неизвестным: x_1, x_2, x_3, x_4, x_5 . Нулевая итерация принята для всех неизвестных:

$$x_1^{(0)} = 0; \quad x_2^{(0)} = 0; \quad x_3^{(0)} = 0; \quad x_4^{(0)} = 0; \quad x_5^{(0)} = 0.$$

Таблица 20

	$\alpha_{1k} x_k^{(j)}$	$\alpha_{2k} x_k^{(j)}$	$\alpha_{3k} x_k^{(j)}$	$\alpha_{4k} x_k^{(j)}$	$\alpha_{5k} x_k^{(j)}$	β_i	$X^{(j+1)}$
$x_1^{(1)}$	—	0	0	0	0	1,388	1,388
$x_2^{(1)}$	0,006	—	0	0	0	2,447	2,453
$x_3^{(1)}$	0,019	0,196	—	0	0	3,342	3,557
$x_4^{(1)}$	0,042	0,002	0,797	—	0	1,857	2,698
$x_5^{(1)}$	0,026	0,132	0,352	0,092	—	4,623	5,225
$x_1^{(2)}$	—	0,191	0,252	0,005	0,016	1,388	1,852
$x_2^{(2)}$	0,007	—	0,036	0,224	0,226	2,447	2,980
$x_3^{(2)}$	0,026	0,238	—	0,005	0,104	3,342	3,715
$x_4^{(2)}$	0,056	0,003	0,832	—	0,073	1,857	2,821
$x_5^{(2)}$	0,035	0,161	0,368	0,096	—	4,623	5,283
$x_1^{(3)}$	—	0,233	0,264	0,006	0,016	1,388	1,906
$x_2^{(3)}$	0,008	—	0,037	0,234	0,296	2,447	2,995
$x_3^{(3)}$	0,027	0,240	—	0,006	0,106	3,342	3,721
$x_4^{(3)}$	0,057	0,003	0,834	—	0,074	1,857	2,825
$x_5^{(3)}$	0,036	0,162	0,368	0,096	—	4,623	5,284
$x_1^{(4)}$	—	0,233	0,264	0,006	0,016	1,388	1,907
$x_2^{(4)}$	0,008	—	0,037	0,234	0,270	2,447	2,916
$x_3^{(4)}$	0,227	0,240	—	0,006	0,06	3,342	3,721
$x_4^{(4)}$	0,057	0,003	0,834	—	0,074	1,857	2,825
$x_5^{(4)}$	0,036	0,162	0,368	0,096	—	4,623	5,285
$x_1^{(5)}$	—	0,233	0,264	0,006	0,016	1,388	1,907
$x_2^{(5)}$	0,008	—	0,037	0,234	0,270	2,447	2,996
$x_3^{(5)}$	0,027	0,240	—	0,003	0,116	3,342	3,721
$x_4^{(5)}$	0,057	0,003	0,834	—	0,074	1,857	2,825
$x_5^{(5)}$	0,036	0,162	0,368	0,096	—	4,623	5,285

Окончательный результат

$$x_1 = 1,907; x_2 = 2,996; x_3 = 3,721; x_4 = 2,825; x_5 = 5,285.$$

Табл. 20 удобна для счета на настольных машинах.

3. Метод Рундсона

Рассмотрим еще один итерационный метод, применяющийся в геофизике. Для этого систему $AX = F$ преобразуем к виду $X = X + v(AX - F)$, где v — любая постоянная.

Итерационный процесс ведется по формуле

$$X_{k+1} = X_k + v_k (AX_k - F).$$

Исследуем выбор v_k и условие сходимости метода для случая, когда A — положительно определенная матрица. В этом случае она имеет n вещественных положительных собственных чисел $\lambda_1, \lambda_2, \dots, \lambda_n$ и n взаимно ортогональных собственных векторов. Разложим вектор $X_0 - A^{-1}F$ по собственным векторам $U_1, U_2, \dots, U_n$:

$$X_0 - A^{-1}F = \sum_{i=1}^n c_i U_i$$

и преобразуем вычислительную формулу

$$\begin{aligned} X_{k+1} - A^{-1}F &= X_k - A^{-1}F + v_k A (X_k - A^{-1}F) = \\ &= (E + v_k A) (X_k - A^{-1}F). \end{aligned}$$

Применяя полученное выражение последовательно, начиная с первой итерации, имеем

$$X_1 - A^{-1}F = (E + v_0 A) (X_0 - A^{-1}F) = \sum_{i=1}^n c_i (1 + v_0 \lambda_i) U_i;$$

$$X_2 - A^{-1}F = (E + v_1 A) (X_1 - A^{-1}F) = \sum_{i=1}^n c_i (1 + v_0 \lambda_i) (1 + v_1 \lambda_i) U_i;$$

.....

$$X_{k+1} - A^{-1}F = (E + v_k A) (X_k - A^{-1}F) = \sum_{i=1}^n c_i \prod_{j=0}^k (1 + v_j \lambda_i) U_i.$$

Используя полученное, вычислим квадрат третьей нормы вектора $X_{k+1} - A^{-1}F$, т. е. его скалярный квадрат

$$(X_{k+1} - A^{-1}F, X_{k+1} - A^{-1}F)$$

и оценим по норме

$$\|X_{k+1} - A^{-1}F\|_{\Pi}^2 \leq M_k \|X_0 - A^{-1}F\|_{\Pi}^2,$$

где

$$M_k = \max_i \prod_{j=0}^k (1 + v_j \lambda_i).$$

Фиксируя количество итераций k , будем подбирать v_k так, чтобы M_k было возможно меньшим. Пусть собственные числа матрицы $A - \lambda_1, \lambda_2, \dots, \lambda_n$ лежат в промежутке (a, b) . Заменим отрезок (a, b) отрезком $(-1, +1)$ и многочлен

$$P_{k+1}(\lambda) = \prod_{j=0}^k (1 + v_j \lambda), \quad P_{k+1}(0) = 1$$

перейдет в многочлен $Q_{k+1}(t)$, где

$$t = \frac{2\lambda}{b-a} - \frac{b+a}{b-a}, \quad Q_{k+1}\left(\frac{b+a}{b-a}\right) = 1.$$

Нашу задачу решает многочлен

$$Q_{k+1}(t) = \frac{T_{k+1}(t)}{T_{k+1}\left(\frac{b+a}{b-a}\right)},$$

где $T_n(t)$ — многочлен Чебышева: $T_n(t) = \cos n \arccos t$, корни которого определяются по формуле

$$t_i = \cos \frac{(2i-1)\pi}{2n}.$$

Следовательно, принимая $v_i = 2[(a-b)t_i - (a+b)]^{-1}$, будем иметь при этом $M_k < 1$ и процесс будет сходиться.

Расчет ведется при фиксированных k . Если выбранного числа итераций недостаточно для нужной точности, то процесс повторяется. Например, в задачах динамической метеорологии встречается система вида

$$u_{i-1j} + u_{i+1j} + u_{ij-1} + u_{ij+1} - 4u_{ij} = \sigma_{ij},$$

$$i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n,$$

причем $u_{i0}, u_{in}, u_{0j}, u_{mj}, \sigma_{ij}$ — заданные величины, а остальные u_{ij} — неизвестные. Применяя метод Ричардсона, эту систему решают по формуле

$$u_{ij}^{(k+1)} = u_{ij}^{(k)} + \frac{1}{4}(u_{ij}^{(k)} + u_{i-1j}^{(k)} + u_{i+1j}^{(k)} + u_{ij+1}^{(k)} - 4u_{ij}^{(k)} - \sigma_{ij}),$$

принимая, таким образом, за v постоянное значение $\frac{1}{4}$.

Указанный процесс сходится медленно из-за выбора ν . Правило выбора ν , описанное в теории, практически редко используется в связи с трудностью отыскания собственных чисел произвольной матрицы A .

§ 6. Численные методы отыскания собственных чисел и собственных векторов матрицы

1. Итерационный метод отыскания наибольшего по модулю собственного числа матрицы и соответствующего ему собственного вектора

Будем предполагать при выводе, что данная матрица A имеет n вещественных собственных чисел и n различных собственных векторов, которые могут определить базис n -мерного линейного пространства. Расположим собственные числа матрицы A в порядке убывания по модулю

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$$

и будем предполагать дополнительно, что λ_1 не кратно. В этих упрощающих предположениях достаточно просто описать метод итераций для отыскания наибольшего по модулю собственного числа.

Рассмотрим произвольный вектор $Y^{(0)}$ того же порядка, что и матрица A , и построим последовательность векторов по правилу

$$\begin{aligned} Y^{(1)} &= AY^{(0)}, \\ Y^{(2)} &= AY^{(1)} = A^2Y^{(0)}, \\ &\dots \dots \dots \\ Y^{(k+1)} &= AY^{(k)} = A^kY^{(0)}. \end{aligned}$$

Разложим вектор $Y^{(0)}$ по собственным векторам матрицы A :

$$Y^{(0)} = \sum_{i=1}^n c_i U_i.$$

Отсюда легко получить

$$\begin{aligned} Y^{(1)} &= AY^{(0)} = c_1 A U_1 + c_2 A U_2 + \dots + c_n A U_n = \\ &= c_1 \lambda_1 U_1 + c_2 \lambda_2 U_2 + \dots + c_n \lambda_n U_n, \\ Y^{(2)} &= c_1 \lambda_1^2 U_1 + c_2 \lambda_2^2 U_2 + \dots + c_n \lambda_n^2 U_n, \\ &\dots \dots \dots \\ Y^{(k)} &= c_1 \lambda_1^k U_1 + c_2 \lambda_2^k U_2 + \dots + c_n \lambda_n^k U_n. \\ Y^{(k+1)} &= c_1 \lambda_1^{k+1} U_1 + c_2 \lambda_2^{k+1} U_2 + \dots + c_n \lambda_n^{k+1} U_n, \end{aligned}$$

Если ввести компоненты векторов $U_1 (x_{11}, x_{12}, \dots, x_{1n}), U_2 (x_{21}, x_{22}, \dots, x_{2n}), \dots, U_n (x_{n1}, x_{n2}, \dots, x_{nn}), Y^{(k)} (y_1^{(k)}, y_2^{(k)}, \dots, y_n^{(k)}), Y^{(k+1)} (y_1^{(k+1)}, y_2^{(k+1)}, \dots, y_n^{(k+1)})$, то k -ю и $(k+1)$ -ю итерации можно представить для i -й компоненты в виде

$$y_i^{(k)} = c_1 \lambda_1^k x_{1i} + c_2 \lambda_2^k x_{2i} + \dots + c_n \lambda_n^k x_{ni};$$

$$y_i^{(k+1)} = c_1 \lambda_1^{k+1} x_{1i} + c_2 \lambda_2^{k+1} x_{2i} + \dots + c_n \lambda_n^{k+1} x_{ni}.$$

Обозначая $c_1 x_{1i} = a_{1i}, c_2 x_{2i} = a_{2i}, \dots, c_n x_{ni} = a_{ni}$, получим

$$y_i^{(k)} = a_{1i} \lambda_1^k + a_{2i} \lambda_2^k + \dots + a_{ni} \lambda_n^k;$$

$$y_i^{(k+1)} = a_{1i} \lambda_1^{k+1} + a_{2i} \lambda_2^{k+1} + \dots + a_{ni} \lambda_n^{k+1}.$$

Предполагая, что $a_{1i} \neq 0$, преобразуем построенные формулы и рассмотрим отношение

$$\frac{y_i^{(k+1)}}{y_i^{(k)}} = \lambda_1 \left[\frac{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^{k+1} \frac{a_{2i}}{a_{1i}} + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^{k+1} \frac{a_{ni}}{a_{1i}}}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k \frac{a_{2i}}{a_{1i}} + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^k \frac{a_{ni}}{a_{1i}}} \right].$$

Начиная с некоторого номера k отношениями $\left(\frac{\lambda_2}{\lambda_1}\right)^k, \dots, \left(\frac{\lambda_n}{\lambda_1}\right)^k; \left(\frac{\lambda_2}{\lambda_1}\right)^{k+1}, \dots, \left(\frac{\lambda_n}{\lambda_1}\right)^{k+1}$ можно пренебречь по сравнению с единицей и, начиная с этого номера,

$$\lambda_1 \approx \frac{y_i^{(k+1)}}{y_i^{(k)}}.$$

Кроме этого

$$Y^{(k)} = \lambda^k c_1 \left[U_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k \frac{c_2}{c_1} U_2 + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^k \frac{c_n}{c_1} U_n \right].$$

При том же значении k собственный вектор U_1 , соответствующий собственному числу λ_1 , может быть принят в виде

$$U_1 \approx Y^{(k)}. \quad (2.29)$$

При практическом вычислении итераций следует вычислять отношение $\frac{y_i^{(k+1)}}{y_i^{(k)}}$ для нескольких компонент. Хорошее совпадение этих отношений показывает, что итерационный процесс закончен. Быстрота сходимости процесса определяется величиной $\left| \frac{\lambda_2}{\lambda_1} \right|$

и может быть медленной, если это отношение близко к единице. Чтобы избежать роста компонентов, целесообразно каким-нибудь способом нормировать на каждом шаге получаемые векторы. Иногда случается, что начальная итерация $\mathbf{Y}^{(0)}$ выбрана неудачно, так что коэффициент c_1 равен или близок нулю. В этом случае не будет ясной картина сходимости и следует изменить начальное приближение.

2. Метод Крылова

Рассмотрим теперь один из методов, с помощью которого определяются все собственные числа и собственные векторы квадратной матрицы A порядка n . Сохраним указанные в начале параграфа предположения относительно A . Выберем, так же как и в итерационном методе, произвольный вектор $Y^{(0)}$ который проще всего принять в виде

$$\mathbf{Y}^{(0)} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Построим векторы $\mathbf{Y}^{(k)}$ ($k = 1, 2, \dots, n$) по формулам

$$\begin{aligned} \mathbf{Y}^{(1)} &= A\mathbf{Y}^{(0)}, \\ \mathbf{Y}^{(2)} &= A\mathbf{Y}^{(1)} = A^2\mathbf{Y}^{(0)}, \\ &\dots\dots\dots \\ \mathbf{Y}^{(k)} &= A\mathbf{Y}^{(k-1)} = A^k\mathbf{Y}^{(0)} \end{aligned}$$

Первой частью нашей задачи является отыскание коэффициентов характеристического уравнения

$$\lambda^n - p_1 \lambda^{n-1} - \dots - p_n = 0. \quad (2.30)$$

По соотношению Кели-Гамильтона матрица A является корнем своего характеристического уравнения

$$A^n - p_1 A^{n-1} - \dots - p_n E = [0].$$

Домножим равенство справа на $Y^{(0)}$ и преобразуем его. Тогда

$$A^n \mathbf{Y}^{(0)} = p_1 A^{n-1} \mathbf{Y}^{(0)} + p_2 A^{n-2} \mathbf{Y}^{(0)} + \dots + p_n \mathbf{Y}^{(0)}$$

ИЛИ

$$\mathbf{Y}^{(n)} = p_1 \mathbf{Y}^{(n-1)} + p_2 \mathbf{Y}^{(n-2)} + \dots + p_n \mathbf{Y}^{(0)}.$$

Запишем эти уравнения в проекциях:

[illegible]

$$\mathbf{Y}^{(k)}(y_1^{(k)}, y_2^{(k)}, \dots, y_n^{(k)}).$$

Из системы (2.31) получим неизвестные коэффициенты характеристического уравнения $p_1, p_2, \dots, p_n$. Систему обычно решают с помощью схем единственного деления. Подбор $\mathbf{Y}^{(0)}$ позволяет построить систему с неособой матрицей.

Второй частью описываемого метода является решение уравнения n -й степени и нахождение корней характеристического уравнения. Для этого применяется обычно метод Лобачевского, описанный в § 2 гл. 1.

Последней, третьей частью будет отыскание собственных векторов матрицы A .

Опишем этот процесс в применении к собственному вектору X_k , отвечающему собственному числу λ_k .

Рассмотрим построенный нами в первой части характеристический полином

$$\lambda^n - p_1 \lambda^{n-1} - \dots - p_n$$

и разделим его на $(\lambda - \lambda_k)$. Получим полином $(n-1)$ -й степени с известными коэффициентами

$$\varphi_k(\lambda) = \lambda^{n-1} + b_1\lambda^{n-2} + \dots + b_{n-1}.$$

Рассмотрим полином

$$\varphi_k(A) = A^{n-1} + b_1 A^{n-2} + \dots + b_{n-1}.$$

Умножим его справа на $Y^{(0)}$. Согласно обозначениям, введенным в этом параграфе, имеем

$$\Phi_k(A) \mathbf{Y}^{(0)} = \mathbf{Y}^{(n-1)} + b_1 \mathbf{Y}^{(n-2)} + \dots + b_{n-1} \mathbf{Y}^{(0)} \quad (2.32)$$

Правая часть этого соотношения вычислена в первой части описываемого метода.

Разложим $Y^{(0)}$ по собственным векторам матрицы A . Тогда

$$\varphi_k(A) \mathbf{Y}^{(0)} = \varphi_k(A) \sum_{i=1}^n c_i \mathbf{X}_i = \sum_{i=1}^n \varphi_k(A) c_i \mathbf{X}_i.$$

Используя равенство, определяющее собственный вектор,

$$AX_i = \lambda_i X_i,$$

и следующее из него равенство

$$A^k X_i = \lambda_i^k X_i,$$

получим

$$\varphi_h(A) Y^{(0)} = \sum_{i=1}^n \varphi_h(\lambda_i) c_i X_i;$$

$\varphi_h(\lambda)$ имеет своими корнями все собственные числа матрицы A , кроме λ_h , поэтому $\sum_{i=1}^n \varphi_h(\lambda_i) = \varphi_h(\lambda_h)$. Отсюда следует $\varphi_h(A) Y^{(0)} = \varphi_h(\lambda_h) c_h X_h$, и, с точностью до постоянной $\varphi_h(\lambda_h) c_h$, $X_h = \varphi_h(A) Y^{(0)}$.

Собственный вектор X_h определяется правой частью формулы (2.32)

$$X_h = Y^{(n-1)} + b_1 Y^{(n-1)} + \dots + b_{n-1} Y^{(0)}. \quad (2.33)$$

Рассмотрим описанный метод сначала на простейшем примере

Пример 1. Дана квадратная матрица второго порядка

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

Найдем сначала ее собственные числа и собственные векторы непосредственно по определению. Характеристическое уравнение этой матрицы: $\lambda^2 - 5\lambda + 2 = 0$.

Корни этого уравнения: $\lambda_1 \cong 4,56$; $\lambda_2 \cong 0,44$.

Собственные векторы вычисляются с помощью систем

$$\left. \begin{aligned} -1,56x_1^{(1)} + 4x_2^{(1)} &= 0 \\ x_1 - 2,56x_2 &= 0 \end{aligned} \right\}, \quad \left. \begin{aligned} 2,56x_1^{(2)} + 4x_2^{(2)} &= 0 \\ x_1^{(2)} + 1,56x_2^{(2)} &= 0 \end{aligned} \right\}.$$

Отсюда

$$X_1 \cong \begin{pmatrix} 2,56 \\ 1 \end{pmatrix}, \quad X_2 \cong \begin{pmatrix} -1,56 \\ 1 \end{pmatrix}.$$

Применим метод итерации, сведя результаты в табл. 21.

Таблица 21

X_1	$Y^{(0)}$	$Y^{(1)}$	$Y^{(2)}$	$Y^{(3)}$	$Y^{(3)}$	$Y^{(4)}$	$Y^{(4)}$
x_1	1	3	13	59	2,56	11,70	2,56
x_2	0	1	5	23	1	4,56	1

С точностью до сотых из данных табл. 21 $\lambda_1 = 4,56$ и $X_1 = \begin{pmatrix} 2,56 \\ 1 \end{pmatrix}$.

Применим к этой же матрице метод Крылова. Так как $Y^{(0)}$, $Y^{(1)}$ и $Y^{(2)}$ получены, то систему уравнений для определения p_1 и p_2 составить очень просто:

$$13 = 3p_1 + p_2;$$

$$5 = p_1.$$

Характеристическое уравнение имеет вид

$$\lambda^2 - 5\lambda + 2 = 0.$$

Очевидно, оно совпадает с полученным точно, корни его: $\lambda_1 \cong 4,56$; $\lambda_2 \cong 0,44$.

Найдем собственные векторы нашей матрице: $\varphi_1(\lambda) = \lambda - 0,44$; $\varphi_2(\lambda) = \lambda - 4,56$.

Следовательно, по уравнению (2.33) получим

$$X_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} - 0,44 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2,56 \\ 1 \end{pmatrix};$$

$$X_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} - 4,56 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1,56 \\ 1 \end{pmatrix}.$$

Пример 2. Найти наибольшее по модулю собственное число и соответствующий ему собственный вектор матрицы A

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

Введем начальное приближение $Y^{(0)}$ и построим $Y^{(k)}$ по формуле

$$Y^{(k)} = AY^{(k-1)}, \quad k = 1, 2, \dots, 7.$$

Результаты сведем в табл. 22. При этом будем при ряде значений делить составляющие $y_i^{(k)}$ на одно и то же число. В этих случаях в табл. 22 два столбца с одним номером.

Таблица 22

X_1	0	1	2	3	3 ¹	4	5	5 ¹	6	6 ¹	7
x_1	1	4	24	180	18	176	1212	121	1012	101	842
x_2	0	2	20	170	17	141	1170	117	947	95	801
x_3	0	2	22	200	20	173	1471	147	1241	124	1041

Наибольшее собственное число $\lambda_1 = 8,4$, а соответствующий ему собственный вектор $X_1 = \begin{pmatrix} 0,81 \\ 0,77 \\ 1 \end{pmatrix}$.

Пример 3. Найти с помощью метода Крылова собственные числа и собственные векторы матрицы

$$A = \begin{pmatrix} 1,00 & 0,42 & 0,54 & 0,66 \\ 0,42 & 0,00 & 0,32 & 0,44 \\ 0,54 & 0,32 & 1,00 & 0,22 \\ 0,66 & 0,44 & 0,22 & 1,00 \end{pmatrix}.$$

Построим выражения $Y^{(k)}$ по формуле

$$Y^{(k)} = AY^{(k-1)}, \quad k = 1, 2, 3, 4$$

и сведем результаты в табл. 23.

Таблица 23

$Y^{(0)}$	$Y^{(1)}$	$Y^{(2)}$	$Y^{(3)}$	$Y^{(4)}$
1	1,00	1,9036	4,2567	9,8066
0	0,42	1,3032	3,2522	7,7029
0	0,54	1,3596	3,1618	7,3267
0	0,66	1,6236	3,7525	8,6885

Составим систему, определяющую коэффициенты характеристического уравнения:

$$4,2567p_1 + 1,9036p_2 + p_3 + p_4 = +9,8066,$$

$$3,2522p_1 + 1,3032p_2 + 0,42p_3 = +7,7029,$$

$$3,1618p_1 + 1,3596p_2 + 0,54p_3 = +7,3267,$$

$$3,7525p_1 + 1,6236p_2 + 0,66p_3 = 8,6886.$$

Решая систему с помощью схемы единственного деления, получим значения p_1, p_2, p_3, p_4 и по ним построим характеристический полином

$$\lambda^4 - 4,01\lambda^3 + 4,78\lambda^2 - 2,12\lambda + 0,29 = 0.$$

Применяя метод Лобачевского, находим корни характеристического уравнения: $\lambda_1 = 2,32$; $\lambda_2 = 0,85$, $\lambda_3 = 0,63$, $\lambda_4 = 0,24$.

Найдем соответственные собственные векторы матрицы. Для этого построим описанные в теории полиномы

$$\varphi_1(\lambda) = \lambda^3 - 1,72\lambda^2 + 0,891\lambda + 0,1285,$$

$$\varphi_2(\lambda) = \lambda^3 - 3,19\lambda^2 + 2,170\lambda + 0,3508,$$

$$\varphi_3(\lambda) = \lambda^3 - 3,41\lambda^2 + 2,743\lambda + 0,4757,$$

$$\varphi_4(\lambda) = \lambda^3 - 3,80\lambda^2 + 3,979\lambda + 1,2431.$$

Следовательно,

$$\begin{aligned} X_1 = & \begin{pmatrix} 4,2567 \\ 3,2522 \\ 3,1618 \\ 3,7525 \end{pmatrix} - 1,72 \begin{pmatrix} 1,9036 \\ 1,3032 \\ 1,3596 \\ 1,6236 \end{pmatrix} + 0,891 \begin{pmatrix} 1,00 \\ 0,42 \\ 0,54 \\ 0,61 \end{pmatrix} + \\ & + 0,1285 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1,7417 \\ 1,3835 \\ 1,9026 \\ 1,5666 \end{pmatrix}, \end{aligned}$$

Аналогично вычисляются остальные собственные векторы.



Глава III

МЕТОД СЕТОК В ПРИМЕНЕНИИ К ЗАДАЧАМ МАТЕМАТИЧЕСКОЙ ФИЗИКИ

§ 1. Некоторые общие понятия метода сеток

В этой главе мы будем изучать метод сеток в применении к граничным задачам для уравнений в частных производных. Рассматривать будем основные уравнения математической физики: уравнения Лапласа и Пуассона, уравнение теплопроводности, волновое уравнение; кроме того, уравнения гидродинамики. Затем рассмотрим небольшое число примеров, соответствующих задачам океанологии и динамической метеорологии. Следует иметь в виду, что все граничные задачи, входящие в данную главу, используются как составные части в задачах гидрометеорологии.

Рассмотрим сначала общие понятия метода сеток. Пусть задана плоская или пространственная область D . За одну из осей координат может быть принята ось $0t$, где t — время. Границу области обозначим через S . Пусть требуется найти решение граничной задачи:

$$Lu = f;$$

$$lu = \varphi,$$

где L — дифференциальный оператор; l — оператор граничных (или начальных) условий; f — непрерывная функция, заданная в области D ; φ — непрерывная функция, заданная на границе S .

Например, для задачи Дирихле:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0; \quad u|_S = \varphi(N), \quad L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},$$

$$f \equiv 0, \quad l \equiv 1, \quad \varphi \equiv \varphi(N),$$

где N — текущая точка границы S .

При решении методом сеток следует прежде всего указать, в каких точках области мы будем искать решение. Естественно построить эти точки на пересечении прямых, параллельных осям координат. Прямые эти представляют сетку, а точки пересечения — ее узлы. Среди построенных узлов граничными называются те, которые или лежат на истинной границе, или находятся от нее на расстоянии более $\frac{h}{2}$, где h — расстояние между прямыми сетки.

Точки, имеющие четыре соседних, принадлежащих области, или граничных, расположенных прямым крестом, называются внутренними. Граничные точки составляют сеточную границу S_h , внутренние — сеточную область D_h .

Построив сеточную область и границу, перейдем к преобразованию дифференциальных уравнений и граничных условий. Дифференциальное уравнение заменяют — аппроксимируют — конечно-разностным или, что то же самое, системой алгебраических уравнений, где неизвестными являются значения функции u в узлах сетки. Решение с этой позиции можно рассматривать как вектор, имеющий столько составляющих, сколько точек содержат D_h и S_h (рис. 5). Для аппроксимации дифференциального уравнения конечно-разностным используют формулы численного дифференцирования. Напомним их

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} \cong \frac{u_{i+1} - u_i}{h}; \quad (3.1)$$

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} \cong \frac{u_{i+1} - u_{i-1}}{2h}; \quad (3.2)$$

$$\left. \frac{\partial^2 u}{\partial x^2} \right|_{x=x_i} \cong \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2}. \quad (3.3)$$

Формула (3.1) имеет погрешность порядка h , а формулы (3.2) и (3.3) погрешность h^2 . Формулы численного дифференцирования могут строиться с помощью полиномов различных степеней, таким образом учитывая различное количество узлов. Такие формулы, как (3.1) и (3.3), используют минимальное число точек: две для производной первого порядка и три для производной второго порядка. Формула (3.2) определяет производную первого порядка с помощью трех узлов и является примером многоточечной. Можно привести пример пятиточечной формулы

$$\left. \frac{\partial u}{\partial x} \right|_{x=x_i} \cong \frac{4/3(u_{i+1} - u_{i-1}) - 1/6(u_{i+2} - u_{i-2})}{2h}.$$

Эта формула имеет погрешность порядка h^4 . Рекомендуем читателю самому получить данную формулу.

Как показано в § 4 гл. I, посвященном численному дифференцированию, кроме непосредственного дифференцирования интерполяционных полиномов можно строить формулы численного дифференцирования данной степени точности с помощью подбора коэффициентов при значениях функции φ в узлах.

При аппроксимации граничных данных затруднением служит то обстоятельство, что в случае криволинейной границы области часть точек сеточной границы не лежит на истинной. В этих случаях или приближенно сносят граничные данные в точки сетки, или применяют различные приемы исправлений граничных данных. Когда точки сеточной границы лежат на истинной, значения граничных функций просто вычисляются в точках сетки, а операторы граничных данных аппроксимируются по формулам численного дифференцирования.

После аппроксимации мы получим конечно-разностную схему

$$L_h u_h = f_h;$$

$$l_h u_h = \varphi_h.$$

В заключение параграфа следует указать на одну особенность сеточного метода, которая очень затрудняет его использование. При расчете по этому методу случается, что небольшие ошибки в начальных и граничных условиях и в правых частях начинают катастрофически расти. В таких случаях говорят, что схема неустойчива. Очевидно, нужно ввести требования, аналогичные требованиям корректности при постановке задач математической физики. Это условие называется корректностью конечно-разностной схемы.

Требования корректности таковы: 1) устойчивость — чтобы малые изменения в граничных и начальных данных, а также в правой части приводили к малым изменениям в решении; 2) существование и единственность решения конечно-разностной схемы; 3) сходимость решения конечно-разностной схемы к решению исходной задачи. Проверка требований корректности конечно-разностной схемы представляет большие теоретические трудности. В то же время расчет по неустойчивой схеме приводит часто к неверным результатам. В последующих параграфах мы опишем ряд способов, которыми можно пользоваться при определении корректности.

§ 2. Пример применения метода сеток.

Решение задачи Дирихле для уравнения Пуассона

Рассмотрим решение граничной задачи для уравнения эллиптического типа — задачу Дирихле для уравнения Пуассона методом сеток. Пусть в области D с границей S нужно найти решение граничной задачи:

$$\Delta u = -f(M);$$

$$u|_S = \varphi(N).$$

На рис. 5 приведена сеточная область D_h и сеточная граница S_h , а также истинные область D и граница S . Очевидно, мы можем принять $\Delta x = \Delta y = h$ в связи с тем, что эти переменные симметрично входят в уравнение. Сетка определителя прямыми $x = x_i$, $y = y_j$, $x_i = x_0 + hi$, $y_j = y_0 + hj$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$.

Крайние прямые сетки $x = x_0$, $x = x_0 + hn$, $y = y_0$, $y = y_0 + hm$ должны или касаться границы S или проходить от нее на расстоянии не большем $\frac{h}{2}$. На рис. 5 указаны точки, входящие в сеточную

границу, и точки, составляющие сеточную область D_h . Пронумеруем все точки сетки двойной нумерацией $i = 1, 2, \dots$ вдоль

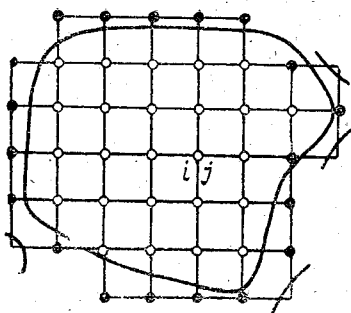


Рис. 5. Сеточная область

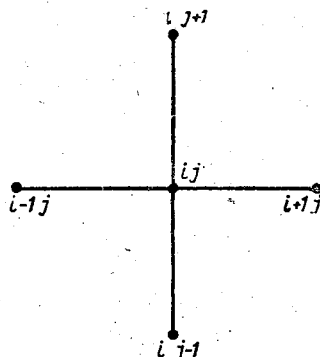


Рис. 6. Пятиточечная схема для уравнения Лапласа

оси Ox и $j = 1, 2, \dots$ вдоль оси Oy . Выберем пять внутренних точек, расположенных прямым крестом (рис. 6). Мы таким образом используем минимальную схему, так как для аппроксимации вторых производных по x и по y необходимо по три точки по вертикали и по горизонтали. Запишем разностный оператор Лапласа для точки с номером i, j :

$$\Delta_h u = \frac{u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} - 4u_{i,j}}{h^2}. \quad (3.4)$$

Погрешность аппроксимации определяется с помощью разложения в ряд Тейлора функций $u_{i-1,j}$, $u_{i+1,j}$, $u_{i,j-1}$, $u_{i,j+1}$ в окрестности точки i, j , как это было показано в гл. I:

$$\Delta_h u - \Delta u = Mh^2, \quad M = \text{const.}$$

Рассмотрим аппроксимацию граничных условий. Разделим все точки S_h на два типа. К первому типу отнесем точки N_k , лежащие на истинной границе S . В этих точках значения функции вычисляются по формуле $u_k = \varphi(N_k)$ и погрешность метода отсутствует. Ко второму типу отнесем остальные граничные точки.

Те из них, которые лежат на расстоянии, значительно меньшем, чем h , от истинной границы выделяются. В этих точках граничные условия приближенно сносят с истинной границы на сеточную, и погрешность не учитывают. В случае, когда не требуется очень большая точность расчета, этот метод применяется для всех точек N_h , не лежащих на истинной границе S . В случае, когда нужна большая точность или точки расположены на расстоянии близком к $\frac{h}{2}$ от точной границы S , проводят исправление граничных условий по Коллатцу, т. е. используют линейное интерполирование. Например, в случае, изображенном на рис. 7,

$$u(B) = u(A) + \frac{u(C) - u(A)}{h + \delta} \delta.$$

Составим для всех внутренних точек соотношение

$$u_{ij} = \frac{u_{i-1j} + u_{i+1j} + u_{ij+1} + u_{ij-1}}{4} - f_{ij}h^2. \quad (3.5)$$

Так как M — число точек — велико, то естественнее всего для решения системы применить один из итерационных методов,

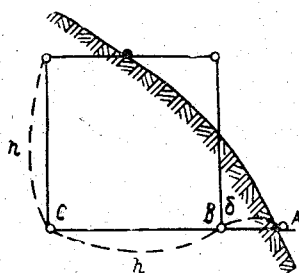


Рис. 7. Исправление граничных данных

а именно метод Зейделя. Нулевую итерацию обычно задают исходя из известных примерных значений u , или с помощью линейной интерполяции граничных условий по горизонтали или по вертикали, или как среднее арифметическое двух последних. К сожалению, для нашей системы не выполняются достаточные условия сходимости метода Зейделя. Преобразовывать матрицу системы не имеет смысла, потому что она содержит много нулей. Следовательно, нам придется отдельно доказывать сходимость метода Зейделя, примененного к системе (3.5). Расчет проводится с помощью таблиц-шаблонов, каждая клеточка которых соответствует точке сетки. В каждой клеточке первого шаблона записываются нулевые итерации. В последующих записываются итерации по Зейделю. Процесс прекращают, когда две соседние итерации совпадают с избранной степенью точности. Следует заметить, что процесс сходится медленно.

Перейдем теперь к исследованию устойчивости и сходимости. Докажем сначала конечно-разностный аналог принципа максимума.

Теорема.

1. Если $\Delta_h u_h \geq 0$ и $u_h \neq \text{const}$, то u_h во всех внутренних точках сетки не принимает своего наибольшего значения.

2. Если $\Delta_h u_h \leq 0$ и $u_h \neq \text{const}$, то u_h во всех внутренних точках сетки не принимает своего наименьшего значения.

3. Если $\Delta_h u_h = 0$ и $u_h \neq \text{const}$, то u_h во всех внутренних точках сетки не принимает ни наибольшего ни наименьшего значения.

Доказательство приведем для первого пункта. Второй доказывается аналогично, а третий следует из первых двух. Рассмотрим произвольную внутреннюю точку сетки (i, j) и предположим, что в этой точке функция u_h достигает наибольшего значения. Так как $\Delta_h u_h \geq 0$, то

$$u_{ij} \leq \frac{u_{i+1j} + u_{i-1j} + u_{ij-1} + u_{ij+1}}{4}.$$

В соответствии с нашим предположением выполняется только знак равенства. Следовательно, в четырех соседних с (i, j) точках значение u_h постоянно и равно значению в самой точке (i, j) . Повторяя процесс, мы доберемся до любой точки нашей области и докажем, что значение u_h во всех точках одинаково. Поэтому в нашем случае наибольшее значение u_h достигается в граничных точках.

Следствие. Разностная краевая задача $\Delta_h u_h = 0$, $u_h|_{S_h} = 0$ имеет своим решением $u_h = 0$.

Перейдем теперь к доказательству корректности конечно-разностной схемы.

1. Конечно-разностная схема

$$\left. \begin{aligned} \Delta_h u_h &= -f_h, \\ u_h|_{S_h} &= \varphi_h \end{aligned} \right\} \quad (3.6)$$

сводится к линейной системе уравнений. Если сравнить эту задачу с задачей, описанной в следствии к теореме, то ясно, что матрицы коэффициентов системы совпадают. Следствие утверждает, что решение однородной системы единственное нулевое. Значит определитель системы не равен нулю, а отсюда решение неоднородной системы существует и единственно.

2. Решение разностной граничной задачи (3.6) устойчиво по граничным данным, т. е.

$$\begin{aligned} \max_{D_h + S_h} |\delta u_h| &\leq \max_{S_h} |\delta \varphi_h|, \\ \text{где} \quad \delta u_h &= u_h^{(1)} - u_h^{(2)}; \quad \delta \varphi_h = \varphi_h^{(1)} - \varphi_h^{(2)}; \\ \Delta_h u_h^{(1)} &= -f_h; \quad \Delta_h u_h^{(2)} = -f_h; \\ u_h^{(1)}|_{S_h} &= \varphi_h^{(1)}; \quad u_h^{(2)}|_{S_h} = \varphi_h^{(2)}. \end{aligned}$$

Доказательство. Легко убедиться, что δu_h удовлетворяет следующей разностной граничной задаче:

$$\Delta_h \delta u_h = 0, \quad \delta u_h|_{S_h} = \delta \varphi_h,$$

и на основании принципа максимума

$$\max_{D_h + S_h} |\delta u_h| \leq \max_{S_h} |\delta \varphi_h|.$$

3. Решение разностной граничной задачи устойчиво по правой части, т. е.

$$\max_{D_h + S_h} |\delta u_h| \leq \frac{\rho^2}{4} \max_{D_h + S_h} \|\delta f_h\|,$$

где

$$\delta u_h = u_h^{(1)} - u_h^{(2)}; \quad \delta f_h = f_h^{(1)} - f_h^{(2)};$$

$$\Delta_h u_h^{(1)} = -f_h^{(1)}; \quad \Delta_h u_h^{(2)} = f_h^{(2)};$$

$$u_h^{(1)}|_{S_h} = \varphi_h; \quad u_h^{(2)}|_{S_h} = \varphi_h;$$

ρ — радиус круга (наименьшего), содержащего $D_h + S_h$;
 α, β — координаты его центра.

Доказательство. Очевидно, δu_h удовлетворяет следующей разностной граничной задаче:

$$\Delta_h \delta u_h = -\delta f_h, \quad \delta u_h|_{S_h} = 0. \quad (3.7)$$

Построим вспомогательную функцию

$$\eta_{1,2} = \pm \delta u_h - \max_{D_h + S_h} \frac{|\delta f_h|}{4} [\rho^2 - (x - \alpha)^2 + (y - \beta)^2]. \quad (3.8)$$

Известно, что $\Delta_h[(x - \alpha)^2 + (y - \beta)^2] = 4$, следовательно,

$$\left. \begin{aligned} \Delta_h \eta_{1,2} &= \pm \Delta_h \delta u_h + \max_{D_h + S_h} |\delta f_h| \geq 0, \\ \eta_{1,2}|_{S_h} &= [\pm \delta u_h]|_{S_h} - \max_{D_h + S_h} \frac{|\delta f_h|}{4} [\rho^2 - (x - \alpha)^2 - (y - \beta)^2] \leq 0. \end{aligned} \right\} \quad (3.9)$$

Тогда по принципу максимума во всей области $\eta_{1,2} \leq 0$ и

$$\max_{D_h + S_h} |\delta u_h| \leq \max_{D_h + S_h} |\delta f_h| \frac{\rho^2}{4}. \quad (3.10)$$

Неравенство (3.10) доказывает наше утверждение.

4. Разностная схема сходится, т. е. имеет место неравенство

$$|u - u_h| \leq M_1 h^2.$$

Для доказательства этого положения применим оператор Δ_h к разности $u - u_h$. Получим

$$\Delta_h(u - u_h) = \Delta_h u - \Delta_h u_h.$$

Прибавим и отнимем Δu , вычисленный в произвольной точке сетки

$$\begin{aligned} \Delta_h(u - u_h) &= \Delta_h u - \Delta u + \Delta u - \Delta_h u_h = \\ &= h^2 M + (f - f_h) = h^2 M. \end{aligned}$$

Если граничные данные не имеют погрешности аппроксимации, то $u - u_h = 0$ в любой точке границы.

Применим к разности $u - u_h$ результаты, полученные при доказательстве устойчивости. Тогда, так как

$$\Delta_h(u - u_h) = h^2 M,$$

$$(u - u_h)|_{S_h} = 0,$$

то

$$|u - u_h| \leq h^2 M \frac{\rho^2}{4}$$

или

$$|u - u_h| \leq M_1 h^2.$$

5. В заключение докажем, что итерационный процесс для описанной задачи сходится. В данном случае проще доказать это непосредственно, чем проверять необходимые и достаточные условия сходимости.

Докажем, что последовательность итераций $u_h^{(0)}, u_h^{(1)}, \dots, u_h^{(k)}, \dots$, сходится к точному решению разностной задачи u_h .

Разобьем все точки D_h на классы. К первому классу отнесем те, которые имеют хотя бы одну граничную точку соседней (по горизонтали или по вертикали). Ко второму классу отнесем те, которые имеют хотя бы одну соседнюю, принадлежащую к первому классу, и т. д. Рассмотрим разность k -й итерации и точного решения системы

$$u_h^{(k)} - u_h = V_h^{(k)}.$$

Следовательно,

$$V_h^{(k)}|_{S_h} = 0, \quad k = 0, 1, \dots$$

Выберем нулевую итерацию $V_h^{(0)}$ так, чтобы удовлетворялось неравенство $|V_h^{(0)}| \leq A$.

Получим первую итерацию для точек первого класса:

$$|V_{h1}^{(1)}| \leq \left(1 - \frac{1}{4}\right) A.$$

Для точек второго класса первая итерация удовлетворяет неравенству

$$|V_{h2}^{(1)}| \leq \left(1 - \frac{1}{4^2}\right) A,$$

и аналогично

$$|V_{hm}^{(1)}| \leq \left(1 - \frac{1}{4^m}\right) A = \alpha A,$$

где $\alpha = \left(1 - \frac{1}{4^m}\right) < 1$.

Отсюда для любой итерации k и для точки любого класса p имеем

$$|V_{hp}^{(k)}| \leq \alpha^k A,$$

откуда следует

$$u_h^{(k)} - u_h \rightarrow 0 \text{ при } k \rightarrow \infty,$$

что и доказывает сходимость итерационного процесса.

Все изложенное позволяет сделать вывод, что конечно-разностная схема, предложенная в данном параграфе, устойчива и сходится. К сожалению, зейделевский процесс в этом случае сходится медленно.

Пример. Рассмотрим граничную задачу Дирихле для уравнения Лапласа в прямоугольнике

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0;$$

$$u|_{y=0} = x^2 + 0,1; \quad u|_{y=0,75} = 0,2x + 0,6719;$$

$$u|_{x=0,5} = y^3 + 0,35; \quad u|_{y=1,5} = -1,8375y + 2,35.$$

При решении методом сеток примем $h = 0,25$. Разностная схема имеет вид

$$u_{ij} = \frac{u_{i-1j} + u_{i+1j} + u_{ij-1} + u_{ij+1}}{4}.$$

Решение дадим в виде табл. 24. В десяти граничных точках значения функции вычисляются без погрешности аппроксимации и записываются в десяти крайних квадратах. В четырех угловых точках значения функции вычисляются также, но при дальнейших расчетах не используются. Во внутренних точках и соответственно во всех внутренних квадратах значения функции определяются по методу Зейделя. В первой строчке каждого внутреннего квад-

рата записаны значения функции полученные с помощью линейного интерполирования граничных значений, — слева записан результат интерполирования по вертикали, а справа — по горизонтали. Во второй строчке стоит среднее арифметическое двух первых значений. Это значение функции принято за начальную итерацию. В следующих строчках записаны последующие итерации по Зейделю. Расчет начинается с левого верхнего угла.

Таблица 24

0,7719	0,8219	0,8719	0,9219	0,9719
0,4750	0,7687 0,7140	0,9479 0,9530	1,1688 1,192	1,4312
	0,7414	0,9504	1,1804	
	0,7443	0,9382	1,1959	
	0,7434	0,9676	1,1934	
	0,7426	0,9665	1,1926	
	0,7424	0,9656	1,1924	
	0,7423	0,9655	1,1922	
	0,7422	0,9655	1,1922	
0,3656	0,7156 0,7468	1,03229 1,128	1,4157 1,5092	1,8906
	0,7312	1,0760	1,4624	
	0,7122	1,0607	1,4524	
	0,7083	1,0572	1,4509	
	0,7072	1,0562	1,4505	
	0,7067	1,0558	1,4503	
	0,7065	0,0556	1,4502	
	0,7065	1,0556	1,4502	
0,3500	0,6625	1,100	1,6625	2,3500

§ 3. Аппроксимация дифференциальных уравнений разностными

Прежде чем рассматривать конкретные примеры аппроксимации, следует заметить что каждое дифференциальное уравнение может быть аппроксимировано различными конечными разностными уравнениями, в зависимости от того, какую формулу численного дифференцирования мы выбираем. Погрешность аппроксима-

ции уменьшается для многоточечных схем, когда интерполяционный полином взят более высокого порядка. Надо также, как всегда, иметь в виду, что при этом увеличивается вычислительная погрешность, так как растет число арифметических действий. Самое же существенное — это то, чтобы выбранная нами аппроксимация давала устойчивую, сходящуюся схему, что всего труднее проверить. Этот вопрос мы оставим пока открытым и вернемся к нему в § 7.

1. Начнем с различных схем для одномерного нестационарного уравнения теплопроводности

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = f(x, t), \quad (3.11)$$

1. Четырехточечная минимальная явная схема. Расположение точек указано на рис. 8.

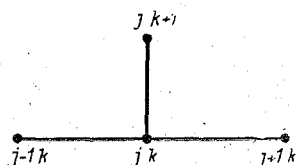


Рис. 8. Явная схема для уравнения теплопроводности

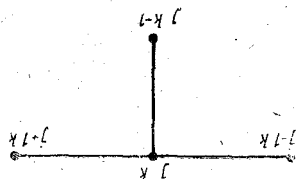


Рис. 9. Неявная схема для уравнения теплопроводности

Значения производных вычисляются для точек (j, k) :

$$\frac{u_{jk+1} - u_{jk}}{\tau} - \frac{u_{j+1k} - 2u_{jk} + u_{j-1k}}{h^2} = f_{jk}.$$

При этом $\Delta x = h$, $\Delta t = \tau$.

Погрешности формул численного дифференцирования имеют соответственно порядки h^2 и τ . Таким образом, разностное уравнение имеет вид

$$u_{jk+1} = u_{jk} \left(1 - 2 \frac{\tau}{h^2} \right) + \frac{\tau}{h^2} (u_{j-1k} + u_{j+1k}) + f_{jk}\tau, \quad (3.12)$$

где $f_{jk} = f(x_j, t_k)$.

Погрешность аппроксимации $O(\tau + h^2)$. Если принять зависимость $\tau = \omega h^2$, то получим:

$$u_{jk+1} = u_{jk} (1 - 2\omega) + \omega (u_{j-1k} + u_{j+1k}) + f_{jk}\tau. \quad (3.13)$$

В частных случаях при $\omega = \frac{1}{2}$ имеем

$$u_{jk+1} = \frac{1}{2} (u_{j+1k} + u_{j-1k}), \quad (3.13')$$

а при $\omega = \frac{1}{6}$

$$u_{jk+1} = \frac{1}{6} (u_{j-1k} + 4u_{jk} + u_{j+1k}). \quad (3.13'')$$

2. Неявная четырехточечная минимальная схема. Расположение точек указано на рис. 9.

Расчет ведется также для точки (j, k) , но производная по времени берется «назад».

$$\frac{u_{j-1k} - u_{jk}}{\tau} - \frac{u_{j-1k} - 2u_{jk} + u_{j+1k}}{h^2} = f_{jk} \quad (3.14)$$

или

$$u_{j-1k}\tau - u_{jk}(h^2 + 2\tau) + u_{j+1k}\tau = -f_{jk}h^2\tau - h^2u_{jk-1}.$$

Построенная схема имеет погрешность порядка $(\tau + h^2)$. Если, как и в предыдущей схеме принять $\tau = \omega h^2$, то конечно-разностное уравнение примет вид

$$u_{j-1k}\omega - u_{jk}(1 + 2\omega) + u_{j+1k}\omega = -f_{jk}\omega h^2 - u_{jk-1}, \quad (3.15)$$

а погрешность очевидно будет иметь порядок h^2 .

3. Пятиточечная схема. Расположение точек на рис. 10.

Ищем производные в точке (j, k) и аппроксимируем производную по времени «на середину»

$$\left. \frac{\partial u}{\partial t} \right|_{jk} = \frac{u_{jk+1} - u_{jk-1}}{2\tau};$$

$$\frac{u_{jk+1} - u_{jk-1}}{2\tau} - \frac{u_{j-1k} - 2u_{jk} + u_{j+1k}}{h^2} = f_{jk}.$$

При $\tau = \frac{h^2}{2}$

$$u_{jk+1} = u_{j-1k} + u_{j+1k} + u_{jk+1} - u_{jk} + f_{jk}h^2. \quad (3.16)$$

4. Семиточечная схема. Расположение точек указано на рис. 11. Производные ищут в точке (j, k) по формулам

$$\left. \frac{\partial u}{\partial t} \right|_{jk} = \frac{u_{j, k + \frac{1}{2}} - u_{j, k - \frac{1}{2}}}{\tau};$$

$$\left. \frac{\partial^2 u}{\partial x^2} \right|_{jk} = \frac{u_{j+1, k + \frac{1}{2}} - 2u_{j, k + \frac{1}{2}} + u_{j-1, k + \frac{1}{2}}}{2h^2} +$$

$$+ \frac{u_{j+1, k - \frac{1}{2}} - 2u_{j, k - \frac{1}{2}} + u_{j-1, k - \frac{1}{2}}}{2h^2}.$$

Погрешность имеет порядок $(\tau^2 + h^2)$. Рекомендуем читателю самому проверить погрешность аппроксимации с помощью разложения функций по формуле Тейлора в окрестности точки (j, k) .

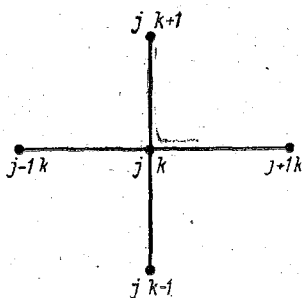


Рис. 10. Пятиточечная схема для уравнения теплопроводности

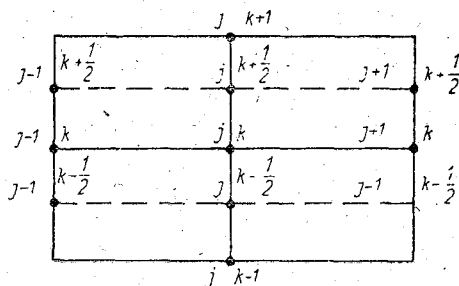


Рис. 11. Семиточечная схема для уравнения теплопроводности

II. Для волнового уравнения

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = f(x, t) \quad (3.17)$$

тоже можно построить ряд аппроксимаций. Мы рассмотрим только одну минимальную явную схему. Расположение точек указано на рис. 10.

Вычисляя производные в точке (j, k) , получим

$$\frac{u_{jk+1} - 2u_{jk} + u_{jk-1}}{\tau^2} - \frac{u_{j+1k} - 2u_{jk} + u_{j-1k}}{h^2} = f_{jk}.$$

Порядок погрешности $(h^2 + \tau^2)$. Если принять $\tau = \omega h$, то

$$u_{jk+1} = 2u_{jk}(1 - \omega^2) + \omega^2(u_{j+1k} - u_{j-1k}) - u_{jk-1} + f_{jk}h^2\omega^2. \quad (3.18)$$

III. Рассмотрим аппроксимацию нелинейного слагаемого, входящего в уравнения движения гидромеханики: $u \frac{\partial u}{\partial x}$.

1. С помощью трехточечной схемы получим

$$u_j(u_{j+1} - u_{j-1}) \frac{1}{2h}. \quad (3.19)$$

Расположение точек очевидно. Погрешность имеет порядок h^2 .

2. Используя тождественное преобразование

$$u \frac{\partial u}{\partial x} = \frac{1}{2} \frac{\partial}{\partial x} u^2,$$

можно получить схему той же точности в виде

$$\frac{1}{2h} \left[\frac{(u_{j+1} + u_j)^2}{2} - \frac{(u_j + u_{j-1})^2}{2} \right].$$

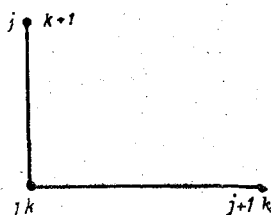


Рис. 12. Трехточечная схема для уравнения первого порядка

IV. Возьмем теперь уравнение первого порядка

$$\frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0.$$

1. Простейшая минимальная трехточечная схема

$$\frac{u_{jk+1} - u_{jk}}{\tau} - u_{jk} \frac{u_{j+1k} - u_{jk}}{h} = 0.$$

Расположение точек указано на рис. 12. Погрешность аппроксимации имеет порядок $(h + \tau)$.

2. Более точной является пятиточечная схема

$$\frac{u_{jk+1} - u_{jk-1}}{2\tau} + u_{jk} \frac{u_{j+1k} - u_{j-1k}}{2h} = 0. \quad (3.20)$$

Расположение точек указано на рис. 10. Погрешность аппроксимации $(h^2 + \tau^2)$.

Укажем ещё одну многоточечную формулу, аппроксимирующую производную по времени:

$$\left. \frac{\partial u}{\partial t} \right|_k = \frac{1}{3} \left[\frac{u_{k+1} - \left(3u_{k-1} - \frac{3}{2}u_k - \frac{1}{2}u_{k-2} \right)}{\tau} \right]. \quad (3.21)$$

Порядок погрешности τ^3 .

Можно уточнить понятие аппроксимации, указав более точно, как непрерывные функции сравниваются с функциями сеточными, которые можно рассматривать как векторы. Для этого введем два оператора. Оператор $[]_h$, примененный к функции u , сопоставляет непрерывной функции $u(x, t)$ функцию $[u]_h$, определенную только в точках сетки. Проще всего определить $[u]_h$, задав ее в точках сетки, как $u(x_j, t_k)$. Аналогично оператор $\{ \}_h$ сопоставляет функции $f(x, t)$, составляющую правую часть уравнения, функцию $\{f\}_h$, определенную только в точках сетки. Зададим ее в точке (j, k) , как $f(x_j, t_k)$.

Введем норму сеточных функций. Например, пространство U_h всевозможных функций u_h можно нормировать, введя $\|u_h\|$ как наибольшее значение $|u_h|$ по всем точкам сетки:

$$\|u_h\| = \max_{j, k} |u_{jk}|.$$

Говорят, что разностное уравнение $L_h u_h = f_h$ аппроксимирует дифференциальное уравнение $Lu = f$ с порядком m , если выполняются неравенства:

$$\|L_h[u]_h - \{Lu\}_h\| \leq M_1 h^m;$$

$$\|\{f\}_h - f_h\| \leq M_2 h^m.$$

Например:

$$Lu = \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2};$$

$$L_h[u]_h = \frac{u_{jk+1} - u_{jk}}{\tau} - \frac{u_{j+1k} - 2u_{jk} + u_{j-1k}}{h^2};$$

$$\|L_h[u]_h - \{Lu\}_h\| \leq M h^2;$$

$$\tau = \omega h^2;$$

$$f_h - \{f\}_h = 0.$$

Следовательно, мы имеем аппроксимацию порядка 2. Следует заметить, что и раньше мы практически проверили точность аппроксимации тем же способом.

§ 4. Аппроксимация начальных и граничных данных для различных задач

Аппроксимация начальных и граничных данных зависит больше всего от двух обстоятельств. Первое — все ли точки сеточной границы лежат на истинной и второе — дифференциальные или алгебраические соотношения представляют сами условия.

Рассмотрим различные граничные условия тех типов, которые чаще всего встречаются в граничных задачах.

I. Начнем с простейшего случая — задачи Коши. Так как мы имеем дело в основном с уравнениями включающими первую или вторую производные по времени, то соответственно будем иметь одно или два начальных условия.

1. В случае уравнения первого порядка относительно производной по времени получим начальное условие в виде $u|_{t=0} = \varphi_1(x)$.

Аппроксимация проводится точно, так как граница прямая и точки S_h лежат на истинной границе S .

2. В случае начального условия $\left. \frac{\partial u}{\partial t} \right|_{t=0} = \varphi_2(x)$ следует аппроксимировать производную $\frac{\partial u}{\partial t}$ по одной из формул численного дифференцирования. Используем формулу, аппроксимирующую $\frac{\partial u}{\partial t}$ с порядком погрешности Δt . Рассмотрим два ряда точек сетки: граничный — нулевой и соседний с ним — с номером первым

$$\left. \frac{\partial u}{\partial t} \right|_{t=0} \simeq \frac{u_{j1} - u_{j0}}{\tau}, \quad \tau = \Delta t.$$

Значения $\varphi_2(x)$ берутся для точек граничного слоя.

Второй способ использует формулу, аппроксимирующую производную с порядком $(\Delta t)^2$, причем приходится вводить дополнительный фиктивный слой $k = -1$ ($t = -\tau$). В этом случае

$$\left. \frac{\partial u}{\partial t} \right|_{t=0} = \frac{u_{j1} - u_{j-1}}{2\Delta t}.$$

Правые части по-прежнему вычисляются в точках нулевого ряда. Значения u_{j-1} впоследствии не используются, хотя и могут быть вычислены при решении задачи.

II. Рассмотрим теперь различные граничные условия. Они могут быть трех типов:

$$\text{а) } u|_S = \varphi_1(N), \quad \text{б) } \left. \frac{\partial u}{\partial n} \right|_S = \varphi_2(N), \quad \text{в) } \left. \frac{\partial u}{\partial n} \right|_S = \alpha(N) u|_S + \beta(N),$$

где N — текущая точка границы, n — внешняя нормаль к границе.

Условия типа (а) аппроксимируются элементарно — вычислением в граничных точках. Если мы решаем одномерную нестационарную задачу в плоскости (x, t) , то точки S_h лежат на истинной границе, и погрешность аппроксимации отсутствует. Если же граница области криволинейна, мы получаем погрешность за счет переноса граничных данных, как было описано в § 2 данной главы. Аппроксимацию граничных условий типа (в) покажем на примере прямоугольной области. Тип (б) при этом можно рассматривать как частный случай. Пусть границами области являются прямые $x = 0$ и $x = l$. В этом случае граничные условия примут вид

$$\left. \frac{\partial u}{\partial x} \right|_{x=0} = \alpha(t) u|_{x=0} + \beta(t);$$

$$\left. \frac{\partial u}{\partial x} \right|_{x=l} = \gamma(t) u|_{x=l} + \delta(t).$$

Покажем прием, который часто применяют, чтобы погрешность аппроксимации граничных данных имела порядок h^2 . Введем сетку, сдвинутую на $\frac{h}{2}$ по оси Ox , и соответственно введем два столбца дополнительных точек, соответствующих $x = -\frac{h}{2}$, $x = l + \frac{h}{2}$. Сетку составляют прямые: $x = \left(j + \frac{1}{2}\right)h$, $j = -1, 2, \dots, n$, $t = k\tau$, $k = 0, 1, \dots, m$, $h = \frac{l}{n}$.

Граничные условия в этом случае аппроксимируются с погрешностью h^2 формулами

$$\frac{u_{-1k} - u_{0k}}{h} = \alpha_k \frac{u_{-1k} + u_{0k}}{2} + \beta_k, \text{ где } \alpha_k = \alpha(t_k), \beta_k = \beta(t_k);$$

$$\frac{u_{nk} - u_{n-1k}}{h} = \gamma_k \frac{u_{nk} + u_{n-1k}}{2} + \delta_k, \text{ где } \gamma_k = \gamma(t_k), \delta_k = \delta(t_k).$$

В заключение параграфа дадим определение порядка аппроксимации граничных условий, используя введенные в § 3 этой главы операторы выборки. Рассмотрим граничные условия в общем виде: $lu = \varphi$. Их аппроксимируют разностные соотношения $l_h u_h = \varphi_h$. Тогда для сравнения $u(x, t)$ и u_h используем оператор выборки $[]_h$:

$$[u]_h = u(x_j, t_k),$$

и для сравнения $\varphi(x, t)$ с φ_h — оператор $\{ \}_h$:

$$\{\varphi\}_h = \varphi(x_j, t_k)$$

Используем также введенные выше нормы сеточных функций.

Говорят, что разностные соотношения $l_h \varphi_h = \varphi_h$ аппроксимируют граничные условия $lu = \varphi$ с порядком m , если выполнены неравенства:

$$\| \{lu\}_h - l[u]_h \| \leq N_1 h^{m_1};$$

$$\| \{\varphi\}_h - \varphi_h \| \leq N_2 h^{m_2},$$

где $m = \min(m_1, m_2)$.

Заметим, что аппроксимация граничных данных может быть проведена и иным способом. Например, аппроксимация начальных условий $u|_{t=0} = \varphi(x)$ для дифференциального уравнения

$$\frac{\partial u}{\partial t} - v \frac{\partial u}{\partial x} = f(x, t)$$

при аппроксимирующей схеме

$$\frac{u_{jk+1} - u_{jk-1}}{2\tau} - v_{jk} \frac{u_{j+1k} - u_{j-1k}}{2h} = f_{jk}$$

усложняется, так как мы должны определить значения u_h не только на нулевом слое, но и на первом. Для этого можно использовать следующее преобразование:

$$u_{j0} = \varphi_{j0};$$

$$u(jh, \tau) = u(jh, 0) + \tau \frac{\partial u}{\partial t}(jh, 0) + O(\tau^2);$$

$$\frac{\partial u}{\partial t} = v \frac{\partial u}{\partial x} + Lu;$$

$$u_{j1} = u_{j0} + \tau \frac{\partial u_{j0}}{\partial x} + \tau Lu(jh, 0) + O(\tau^2);$$

$$u(jh, 0) = \varphi_h, \quad \frac{\partial u}{\partial x}(jh, 0) = \varphi'(x_j);$$

$$Lu(jh, 0) = f(jh, 0).$$

Таким образом, аппроксимация граничного условия проводится по формулам

$$u_{j0} = \varphi_j;$$

$$u_{j1} = \varphi_j + \varphi'_j + \tau f_{j0}.$$

§ 5. Решения некоторых разностных схем

1. Уравнение теплопроводности. Явная схема

Рассмотрим граничную задачу

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = f(x, t);$$

$$u|_{t=0} = \varphi(x), \quad u|_{x=0} = \psi_1(t), \quad u|_{x=l} = \psi_2(t).$$

В § 3 и 4 этой главы показано, что такая задача может быть аппроксимирована с помощью следующей явной схемы:

$$\left. \begin{aligned} u_{j0} &= \varphi_j, \\ u_{jk+1} &= u_{jk}(1 - 2\omega) + \omega(u_{j-1k} + u_{j+1k}) + f_{jk}\tau, \\ u_{0k} &= \psi_{1k}; \quad u_{nk} = \psi_{2k}, \end{aligned} \right\} \quad (3.22)$$

где $j = 1, 2, \dots, n-1$; $k = 1, 2, \dots, m$.

Область ограничена прямыми: $x = 0$, $x = l$, $t = 0$, $t = T$. Построена сетка: $x = jh$, $t = k\tau$, $h = \frac{l}{n}$, $\tau = \omega h^2$. Расчет ведется «по-слоино». Слоями называют набор точек, лежащих на прямых $t = \tau k$. В точках нулевого слоя $t = 0$ значения функции известны из начальных условий $u_{j0} = \varphi_j$.

Расчет в первом слое ведется по формулам:

$$u_{01} = \psi_{10};$$

$$u_{j1} = u_{j0}(1 - 2\omega) + \omega(u_{j-10} + u_{j+10}) + f_{j1}\tau;$$

$$u_{n1} = \psi_{21}.$$

Когда все значения u_{j1} вычислены, переходят ко второму слою.

2. Уравнение теплопроводности. Неявная схема

Рассмотрим граничную задачу, описанную в предыдущем пункте, и используем аппроксимацию вида (3.14)

$$u_{j-1k}\omega - u_{jk}(1 - 2\omega) + u_{j+1k}\omega = f_{jk}\omega h - u_{jk+1}. \quad (3.23)$$

Система уравнений (3.23) может быть разбита на отдельные системы также по слоям. Фиксируем k и начинаем с $k=1$, так как

4. Волновое уравнение. Явная схема

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = f(x, t);$$

$$u|_{t=0} = \varphi_1(x); \quad \left. \frac{\partial u}{\partial t} \right|_{t=0} = \varphi_2(x);$$

$$u|_{x=0} = \psi_1(x); \quad u|_{x=l} = \psi_2(x).$$

Эта граничная задача описывает различные волновые процессы: акустические волны, волны понижения и повышения уровня жидкости в открытых руслах и т. д.

Используем описанную в § 3 этой главы минимальную пятиточечную схему

$$u_{jh+1} = \omega(u_{j-1h} + u_{j+1h}) + 2(1 - \omega)u_{jh} - u_{jh-1} + h^2 f_{jh}. \quad (3.29)$$

Границами области являются прямые $x = 0$, $x = l$, $t = 0$, $t = T$. Сетку определяют $x = jh$, $t = k\tau$, $\tau = \omega h$, $k = 0, 1, 2, \dots, m$; $j = 0, 1, \dots, n$.

Граничные данные аппроксимируются соотношениями:

$$\left. \begin{aligned} u_{j0} &= \varphi_{1j}, \\ u_{j1} &= \tau \varphi_{2j} + \varphi_{1j}, \\ u_{0k} &= \psi_{1k}, \\ u_{nk} &= \psi_{2k}. \end{aligned} \right\} \quad (3.30)$$

Построенная схема является явной и рассчитывается послойно, начиная с $k = 2$.

Пример. Решить уравнение теплопроводности

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0$$

при следующих граничных и начальных данных:

$$u|_{t=0} = 0, \quad u|_{x=0} = 0, \quad u|_{x=1} = 4t.$$

Для решения используем явную схему $\omega = \frac{1}{2}$:

$$u_{jh+1} = \frac{u_{j-1h} + u_{j+1h}}{2}, \quad \Delta x = 0,2, \quad \Delta t = 0,01.$$

Решение записывается в виде табл. 25. Расчет ведется от $t = 0$ до $t = 0,08$.

Таблица 25

$k \backslash j$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0,04	0	0	0	0	0
2	0,08	0,02	0	0	0	0
3	0,12	0,04	0,01	0	0	0
4	0,16	0,065	0,02	0,005	0	0
5	0,20	0,090	0,350	0,10	0,003	0
6	0,24	0,118	0,05	0,018	0,005	0
7	0,28	0,145	0,068	0,038	0,009	0
8	0,32	0,174	0,092	0,039	0,019	0

§ 6. Метод матричной прогонки

Будем рассматривать смешанную задачу для уравнения Пуассона

$$\Delta u = f(x, y),$$

$$\left. \frac{\partial u}{\partial n} \right|_S = \alpha(x, y) u|_S + \beta(x, y).$$

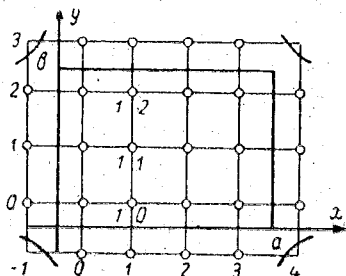


Рис. 13. Сеточная область в случае смешанной задачи

угольную область с двадцатью точками, ограниченную прямыми $x = 0$, $x = a$, $y = 0$, $y = b$, представленную на рис. 13. Координаты точек сетки определяются формулами

$$x = \left(i + \frac{1}{2}\right)h, \quad y = \left(j + \frac{1}{2}\right)h.$$

Приведем указанную конечно-разностную схему

$$u_{i-1j} + u_{i+1j} + u_{ij-1} + u_{ij+1} - 4u_{ij} = h^2 f_{ij}. \quad (3.31)$$

Граничные условия в применении к четырем границам нашей области имеют вид

$$-\frac{\partial u}{\partial x} \Big|_{x=0} = \alpha(0, y) u|_{x=0} + \beta(0, y);$$

$$-\frac{\partial u}{\partial y} \Big|_{y=0} = \alpha(x, 0) u|_{y=0} + \beta(x, 0);$$

$$\frac{\partial u}{\partial x} \Big|_{x=a} = \alpha(a, y) u|_{x=a} + \beta(a, y);$$

$$\frac{\partial u}{\partial y} \Big|_{y=b} = \alpha(x, b) u|_{y=b} + \beta(x, b).$$

Они аппроксимируются следующим образом:

$$\left. \begin{aligned} \frac{-u_{0j} + u_{-1j}}{h} + \alpha(0, y_j) \frac{u_{0j} + u_{-1j}}{2} + \beta(0, y_j), \\ \frac{u_{4j} - u_{3j}}{h} + \alpha(a, y_j) \frac{u_{4j} + u_{3j}}{2} + \beta(a, y_j); \end{aligned} \right\} \quad (3.32)$$

$$\left. \begin{aligned} \frac{-u_{i0} + u_{i-1}}{h} = \alpha(x_i, 0) \frac{u_{i0} + u_{i-1}}{2} + \beta(x_i, 0), \\ \frac{u_{i3} - u_{i2}}{h} = \alpha(x_i, b) \frac{u_{i3} + u_{i2}}{2} + \beta(x_i, b). \end{aligned} \right\} \quad (3.33)$$

Описанный в § 5 гл. II итерационный процесс в применении к данной задаче сходится очень медленно. В этом параграфе мы сведем конечно-разностную схему к обобщенной (или векторной) трехдиагональной системе, причем за коэффициенты примем матрицы, а за неизвестные — векторы, разбив таким образом основную систему на подсистемы. К построенной векторной системе будем применять метод прогонки, преобразованный для нашего случая.

Можно заметить, что в каждом уравнении основной системы используются значения функций, вычисленные в пяти точках. Но точки эти располагаются на трех вертикальных или горизонтальных прямых сетки. Примем значения функции u , лежащие на одной прямой u_{i0}, u_{i1}, u_{i2} , за составляющие вектора U_i . При

этом мы не включаем две крайние точки $(i, -1)$, $(i, 3)$. Действуя таким образом, мы получим шесть векторов

$$\begin{aligned} \mathbf{U}_{-1} &= \begin{pmatrix} u_{-10} \\ u_{-11} \\ u_{-12} \end{pmatrix}, \quad \mathbf{U}_0 = \begin{pmatrix} u_{00} \\ u_{01} \\ u_{02} \end{pmatrix}, \quad \mathbf{U}_1 = \begin{pmatrix} u_{10} \\ u_{11} \\ u_{12} \end{pmatrix} \\ \mathbf{U}_2 &= \begin{pmatrix} u_{20} \\ u_{21} \\ u_{22} \end{pmatrix}, \quad \mathbf{U}_3 = \begin{pmatrix} u_{30} \\ u_{31} \\ u_{32} \end{pmatrix}, \quad \mathbf{U}_4 = \begin{pmatrix} u_{40} \\ u_{41} \\ u_{42} \end{pmatrix}. \end{aligned}$$

Далее мы можем записать граничные условия в векторном виде следующим образом. Выпишем эти условия на прямой $x = 0$ в трех точках

$$\left. \begin{aligned} -\frac{u_{00} + u_{-10}}{h} &= \alpha(0, y_0) \frac{u_{00} + u_{-10}}{2} + \beta(0, y_0), \\ -\frac{u_{01} + u_{-11}}{h} &= \alpha(0, y_1) \frac{u_{01} + u_{-11}}{2} + \beta(0, y_1), \\ -\frac{u_{02} + u_{-12}}{h} &= \alpha(0, y_2) \frac{u_{02} + u_{-12}}{2} + \beta(0, y_2). \end{aligned} \right\} \quad (3.34)$$

Вводя матрицу α_0 и вектор β_0 по формулам

$$\alpha_0 = \begin{pmatrix} \alpha(0, y_0) & 0 & 0 \\ 0 & \alpha(0, y_1) & 0 \\ 0 & 0 & \alpha(0, y_2) \end{pmatrix}, \quad \beta_0 = \begin{pmatrix} \beta(0, y_0) \\ \beta(0, y_1) \\ \beta(0, y_2) \end{pmatrix},$$

будем рассматривать уравнения (3.34) как составляющие векторного равенства

$$-\frac{\mathbf{U}_0 + \mathbf{U}_{-1}}{h} = \alpha_0 \frac{\mathbf{U}_0 + \mathbf{U}_{-1}}{2} + \beta_0. \quad (3.35)$$

Аналогично, рассмотрев граничные условия в точках прямой $x = a$ и введя обозначения

$$\alpha_a = \begin{pmatrix} \alpha(a, y_0) & 0 & 0 \\ 0 & \alpha(a, y_1) & 0 \\ 0 & 0 & \alpha(a, y_2) \end{pmatrix}, \quad \beta_a = \begin{pmatrix} \beta(a, y_0) \\ \beta(a, y_1) \\ \beta(a, y_2) \end{pmatrix},$$

легко выписать второе граничное условие (3.32) в векторном виде

$$\frac{U_4 - U_3}{h} = \alpha_a \frac{U_4 + U_3}{2} + \beta_a. \quad (3.36)$$

Рассмотрим теперь конечно-разностное уравнение, аппроксимирующее уравнение Пуассона. Запишем это соотношение для точек $(1, 0)$, $(1, 1)$ и $(1, 2)$:

$$\left. \begin{aligned} u_{00} + u_{20} - 4u_{10} + u_{11} &= h^2 f_{10} - u_{1-1}, \\ u_{01} + u_{21} + u_{10} - 4u_{11} + u_{12} &= h^2 f_{11}, \\ u_{02} + u_{22} + u_{11} - 4u_{12} &= h^2 f_{12} - u_{13}. \end{aligned} \right\} \quad (3.37)$$

Учтем дополнительно граничные условия для точек $(1, 0)$, $(1, -1)$, $(1, 3)$ и $(1, 2)$:

$$-\frac{u_{10} + u_{1-1}}{h} = \alpha(x_1, 0) \frac{u_{10} + u_{1-1}}{2} + \beta(x_1, 0),$$

$$\frac{u_{13} - u_{12}}{h} = \alpha(x_1, b) \frac{u_{13} + u_{12}}{1} + \beta(x_1, b),$$

преобразовав их к виду

$$\left. \begin{aligned} u_{1-1} &= u_{10} \frac{2 + h\alpha_{10}}{2 - h\alpha_{10}} + \frac{2h\beta_{10}}{2 - h\alpha_{10}}, \\ u_{13} &= u_{12} \frac{2 + h\alpha_{1b}}{2 - h\alpha_{1b}} + \frac{2h\beta_{1b}}{2 - h\alpha_{1b}}, \end{aligned} \right\} \quad (3.38)$$

где $\alpha_{1b} = \alpha(x_1, b)$, $\alpha_{10} = \alpha(x_1, 0)$, $\beta_{1b} = \beta(x_1, b)$, $\beta_{10} = \beta(x_1, 0)$.

Подставив соотношения (3.37) в систему (3.38), получим

$$u_{00} + u_{20} + \left(-4 + \frac{2 + h\alpha_{10}}{2 - h\alpha_{10}} \right) u_{10} + u_{11} = h^2 f_{10} - \frac{2h\beta_{10}}{2 - h\alpha_{10}};$$

$$u_{01} + u_{21} + u_{10} - 4u_{11} + u_{12} = h^2 f_{11};$$

$$u_{02} + u_{22} + u_{11} + \left(-4 + \frac{2 + h\alpha_{1b}}{2 - h\alpha_{1b}} \right) u_{12} = h^2 f_{12} - \frac{2h\beta_{1b}}{2 - h\alpha_{1b}}.$$

Введя дополнительно матрицу B_1 и вектор F_1 по формулам

$$B_1 = \begin{pmatrix} -4 + \frac{2h\alpha_{10}}{2 - h\alpha_{10}} & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & -4 + \frac{2 + h\alpha_{1b}}{2 - h\alpha_{1b}} \end{pmatrix},$$

$$F_1 = \begin{pmatrix} h^2 f_{10} - \frac{2h\beta_{10}}{2 - h\alpha_{10}} \\ h^2 f_{11} \\ h^2 f_{12} - \frac{2h\beta_{1b}}{2 - h\alpha_{1b}} \end{pmatrix},$$

можно записать нашу систему в векторном виде

$$U_0 + B_1 U_1 + U_2 = F_1.$$

И, вообще, вводя матрицу B_i и векторы F_i , $i = 0, 1, 2, 3$

$$B_i = \begin{pmatrix} -4 + \frac{2h\alpha_{i0}}{2 - h\alpha_{i0}} & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & -4 + \frac{2 + h\alpha_{ib}}{2 - h\alpha_{ib}} \end{pmatrix}, \quad (3.39)$$

$$F_i = \begin{pmatrix} h^2 f_{i0} - \frac{2h\beta_{i0}}{2 - h\alpha_{i0}} \\ h^2 f_{i1} \\ h^2 f_{i2} - \frac{2h\beta_{ib}}{2 - h\alpha_{ib}} \end{pmatrix}, \quad (3.40)$$

где $\alpha_{i0} = \alpha(x_i, 0)$, $\beta_{i0} = \beta(x_i, 0)$, $\alpha_{ib} = \alpha(x_i, b)$, $\beta_{ib} = \beta(x_i, b)$, мы можем записать конечно-разностное соотношение для всех внутренних точек сетки в векторном виде, собрав их по «столбцам» в векторные соотношения

$$U_{i-1} + B_i U_i + U_{i+1} = F_i, \quad i = 0, 1, 2, 3.$$

Если добавить к этому граничные условия (3.35) и (3.36)

$$\left. \begin{aligned} (2E - h\alpha_0) U_{-1} - (2E + h\alpha_0) U_0 &= 2h\beta_0, \\ -(2E + h\alpha_a) U_3 + (2E - h\alpha_a) U_4 &= 2h\beta_a, \end{aligned} \right\} \quad (3.41)$$

то получим трехдиагональную систему

$$\left. \begin{aligned} B_{-1}U_{-1} + C_{-1}U_0 &= F_{-1}, \\ U_{-1} + B_0 U_0 + U_1 &= F_0, \\ U_0 + B_1 U_1 + U_2 &= F_1, \\ U_1 + B_2 U_2 + U_3 &= F_2, \\ U_2 + B_3 U_3 + U_4 &= F_3, \\ A_4 U_3 + B_4 U_4 &= F_4, \end{aligned} \right\} \quad (3.42)$$

где $B_{-1} = 2E - h\alpha_0$, $C_{-1} = -2E - h\alpha_0$, $F_{-1} = 2h\beta_0$, $A_4 = -2E - h\alpha_a$, $B_4 = 2E - h\alpha_a$, $F_4 = 2h\beta_a$; B_i , F_i ; $i = 0, 1, 2, 3$ определяются формулами (3.39) и (3.40).

Систему (3.42) решают методом прогонки. Используем формулы, полученные в § 4 гл. II:

$$\left. \begin{aligned} C'_{i-1} &= B_{i-1}^{-1} C_{i-1}; \quad F'_{i-1} = B_{i-1}^{-1} F_{i-1}; \quad C'_i = (B_i - C'_{i-1})^{-1}; \\ F'_i &= (B_i - C'_{i-1})^{-1} (F_i - F'_{i-1}), \quad i = 0, 1, 2, 3; \\ F'_4 &= (B_4 - A_4 C'_3)^{-1} (F_4 - A_4 F'_3). \end{aligned} \right\} \quad (3.43)$$

Из этой системы с помощью обратного хода найдем решение задачи в векторной форме

$$U_4 = F'_4, \quad U_{i-1} = F'_i - C'_{i-1} U_i \quad \text{при } i = 4, 3, 2, 1, 0. \quad (3.44)$$

Самым трудоемким в описанной схеме является процесс отыскания обратных матриц, который проводится дважды на каждом шаге прогонки. Поэтому при прямоугольной области следует выбирать строчки более длинными, чем столбцы, чтобы порядок векторов и матриц был возможно меньшим. Это уменьшит количество арифметических действий и соответственно уменьшит вычислительную погрешность. Очевидно, метод матричной прогонки относится к точным методам линейной алгебры, и погрешности в этом случае появляются за счет аппроксимации, за счет погрешности исходных данных и за счет округления.

Пример. Рассмотрим решение смешанной задачи для уравнения Пуассона с помощью матричной прогонки. Граничная задача имеет вид

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= 6x + 6y, \\ -\frac{\partial u}{\partial x} \Big|_{x=0} &= u|_{x=0} + y - y^3, \\ \frac{\partial u}{\partial x} \Big|_{x=0,8} &= u|_{x=0,8} + 1,418 + 0,2y - y^3, \\ -\frac{\partial u}{\partial y} \Big|_{y=0} &= u|_{y=0} + x - x^3, \\ \frac{\partial u}{\partial y} \Big|_{y=0,6} &= u|_{y=0,6} - x^3 + 0,4x + 0,48.\end{aligned}$$

Сетка представлена на рис. 13.

Из граничных условий получим

$$\alpha_0 = \alpha_{0,8} = E,$$

$$\beta_0 = \begin{pmatrix} 0,099 \\ 0,273 \\ 0,375 \end{pmatrix}, \quad \beta_{0,8} = \begin{pmatrix} 1,437 \\ 1,451 \\ 1,393 \end{pmatrix}.$$

Уравнения (3.35) и (3.36) запишем следующим образом.

$$\begin{aligned}9U_{-1} - 11U_0 &= 2\beta_0, \\ 9U_4 - 11U_3 &= 2\beta_{0,8}.\end{aligned}$$

Матрицы B_i определяются по (3.39) в виде

$$B_i = \begin{pmatrix} -2,777 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & -2,777 \end{pmatrix}, \quad i = 0, 1, 2, 3.$$

Правые части уравнений определяются следующими значениями:

$$F_0 = \begin{pmatrix} 0,026 \\ 0,547 \\ 0,750 \end{pmatrix}, \quad F_1 = \begin{pmatrix} 0,035 \\ 0,144 \\ 0,065 \end{pmatrix},$$

$$F_2 = \begin{pmatrix} 0,061 \\ 0,192 \\ 0,117 \end{pmatrix}, \quad F_3 = \begin{pmatrix} 0,113 \\ 0,240 \\ 0,195 \end{pmatrix}.$$

В результате прямого хода прогонки мы получим систему (3.43), где матрицы C'_k и векторы F'_k вычислены в виде:

$$C'_{-1} = -1,222E,$$

$$C'_0 = \begin{pmatrix} 3,223 & -0,978 & 0 \\ -0,978 & 2,924 & -0,978 \\ 0 & -0,978 & 3,223 \end{pmatrix},$$

$$C'_1 = \begin{pmatrix} 1,111 & -0,115 & 0 \\ -0,115 & 1,221 & -0,115 \\ 0 & -0,115 & 1,111 \end{pmatrix},$$

$$C'_2 = \begin{pmatrix} 0,969 & -0,063 & 0 \\ -0,063 & 1,033 & -0,063 \\ 0 & -0,063 & 0,969 \end{pmatrix},$$

$$C'_3 = \begin{pmatrix} -1,273 & -0,081 & 0 \\ -0,081 & 1,162 & -0,081 \\ 0 & -0,081 & 1,273 \end{pmatrix},$$

$$F'_{-1} = \begin{pmatrix} 0,022 \\ 0,061 \\ 0,083 \end{pmatrix}, \quad F'_0 = \begin{pmatrix} -0,021 \\ 0,083 \\ 0,018 \end{pmatrix},$$

$$F'_1 = \begin{pmatrix} 0,055 \\ 0,063 \\ 0,045 \end{pmatrix}, \quad F'_2 = \begin{pmatrix} -0,002 \\ 0,128 \\ 0,062 \end{pmatrix},$$

$$F'_3 = \begin{pmatrix} 0,137 \\ 0,113 \\ 0,160 \end{pmatrix}, \quad F'_4 = \begin{pmatrix} 0,208 \\ 0,190 \\ 0,208 \end{pmatrix}.$$

И, наконец, решение задачи запишем в виде векторов, рассчитанных по формулам (3.34):

$$\begin{aligned} \mathbf{U}_4 &= \begin{pmatrix} 0,208 \\ 0,190 \\ 0,208 \end{pmatrix}, \quad \mathbf{U}_3 = \begin{pmatrix} -0,113 \\ -0,074 \\ -0,090 \end{pmatrix}, \quad \mathbf{U}_2 = \begin{pmatrix} -0,106 \\ 0,065 \\ -0,023 \end{pmatrix}, \\ \mathbf{U}_1 &= \begin{pmatrix} -0,070 \\ 0,157 \\ 0,012 \end{pmatrix}, \quad \mathbf{U}_0 = \begin{pmatrix} -0,401 \\ 0,601 \\ -0,174 \end{pmatrix}, \quad \mathbf{U}_{-1} = \begin{pmatrix} 0,511 \\ -0,673 \\ 0,295 \end{pmatrix}. \end{aligned}$$

§ 7. Некоторые простейшие способы исследования устойчивости конечно-разностных схем

Вернемся к понятию устойчивости, введенному в § 1 данной главы. Рассмотрим различные способы проверки устойчивости конечно-разностных схем.

1. ε -схема. Простейшим способом исследования устойчивости конечно-разностных схем является непосредственная проверка поведения единичной ошибки при расчете по схеме. Такая проверка может быть применена только к явной схеме и к тому же не дает окончательного ответа на вопрос об устойчивости.

Проведем на примере двух схем для уравнения теплопроводности проверку на устойчивость по ε -схеме. Используем схемы (3.13') и (3.16). Напомним полученные формулы

$$u_{jh+1} = \frac{1}{2} (u_{j-1k} + u_{j+1k}); \quad (3.13')$$

$$u_{jh+1} = u_{j-1k} + u_{j+1k} + u_{jh-1} - 2u_{jh}. \quad (3.16)$$

Запишем решение граничной задачи для уравнения теплопроводности без правой части нулевыми граничными и начальными условиями

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0,$$

$$u|_{t=0} = 0, \quad u|_{x=0} = 0, \quad u|_{x=l} = 0.$$

Очевидно, такая задача и решенная точно, и по формулам (3.13') и (3.16) должна дать в ответе тождественный ноль. Предположим, что из-за неточности в начальных данных или из-за округления в одной точке сетки появилась ошибка ε . Рассмотрим, как изменится при этом решение. В табл. 26 дан расчет ошибки по формуле (3.13'). Имеется в виду, что каждая точка

таблицы заменяет точку сетки. Расчет ведется сверху вниз. Значения в верхней строчке известны. Расчет, как описано в § 5 этой главы, ведется послойно.

Таблица 26

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	ε	0	0	0	0	0
2	0	0	0	$\frac{\varepsilon}{2}$	0	$\frac{\varepsilon}{2}$	0	0	0	0
3	0	0	$\frac{\varepsilon}{4}$	0	$\frac{3\varepsilon}{8}$	0	$\frac{\varepsilon}{4}$	0	0	0
4	0	$\frac{\varepsilon}{8}$	0	$\frac{5\varepsilon}{16}$	0	$\frac{5\varepsilon}{16}$	0	$\frac{\varepsilon}{8}$	0	0

Погрешность, как показывает расчет, распространяется внутри треугольника, но остается все время ограниченной исходным значением. ε -схема показывает, что разностная схема (3.13') возможно устойчива.

В табл. 27 проведен расчет по формуле (3.16). В этом случае считаются известными значения решения в первых двух строчках.

Таблица 27

	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	ε	0	0	0	0	0
3	0	0	0	0	ε	-2ε	ε	0	0	0	0
4	0	0	0	ε	-4ε	7ε	-4ε	ε	0	0	0
5	0	0	ε	-6ε	17ε	-24ε	17ε	-6ε	ε	0	0
6	0	ε	-8ε	31ε	-68ε	89ε	-68ε	31ε	-8ε	ε	0

Легко заметить, что, распространяясь в таком же треугольнике, как и при расчете по формуле (3.13'), погрешность неограниченно растет. Можно сделать вывод, что схема (3.16) неустойчива.

2. Дадим более точное определение устойчивости разностной схемы. Пользуясь операторами выборки $[]_h$ и $\{ \}_h$, будем рассматривать нормированное пространство сеточных функций, где нормой будем считать наибольшее по модулю значение u_{jk} . Таким образом, таблица решений разностной задачи может рассматриваться как множество составляющих вектора u_h . Тогда указанная

норма соответствует определенной в § 3 гл. II первой норме. В определениях не будем указывать тип нормы, так как можно использовать и другие нормы.

Введем следующее определение. Разностная схема

$$\left. \begin{aligned} L_h u_h &= f_h, \\ l_h u_h &= \varphi_h \end{aligned} \right\} \quad (3.45)$$

устойчива в области D_h , если при любых h имеет место неравенство

$$\|u_h\| \leq N_1 \|f_h\| + N_2 \|\varphi_h\|, \quad (3.46)$$

где N_1 и N_2 — постоянные, а f_h и φ_h , так же как и u_h , рассматриваются в виде векторов. Нормы в левой и правой части должны быть согласованы.

Проведем доказательство устойчивости явной конечно-разностной схемы для уравнения теплопроводности, используя введенное определение и «расслоение», описанное в § 5 данной главы при решении по этой схеме. Схема определяется формулой

$$u_{jh+1} = u_{jk} (1 - 2\omega) + \omega (u_{j-1k} + u_{j+1k}) + f_{jk\tau} \quad (3.47)$$

при $j = 1, 2, \dots, n-1$, $k = 1, 2, \dots, m$, $u_{j0} = \varphi_{1j}$, $u_{0k} = \psi_{1k}$, $u_{nh} = \psi_{2k}$.

Введем вектор $U_k \{u_{0k}, u_{1k}, \dots, u_{nk}\}$.

Как было показано, мы можем вычислить все составляющие вектора U_k , если известны составляющие вектора U_{k-1} .

Введем норму множества векторов u_h так, как мы указывали, т. е.

$$\|u_h\| = \max_j |u_{jk}|.$$

Введем дополнительно пространство U_h сеточных функций u_h , включающих как составляющие вектора U_k , и определим их норму формулой

$$\|u_h\| = \max_k \|u_k\|.$$

Для доказательства устойчивости запишем конечно-разностную схему в виде

$$U_{k+1} = R_h U_k + \tau \rho_k = U_{k+1}^{(1)} + U_{k+1}^{(2)},$$

где R_h — оператор перехода от слоя к слою, который каждому вектору U_k ставит в соответствие вектор U_{k+1} по формуле

$$u_{jk+1}^{(1)} = (1 - 2\omega) u_{jk} + \omega (u_{j-1k} + u_{j+1k}),$$

а ρ_k определяется составляющими

$$\rho_k (f_{1k}, f_{2k}, \dots, f_{nk}).$$

При этом U_0 задан определением начальных условий $(\varphi_0, \varphi_1, \dots, \varphi_n)$. Граничные значения входят в u_h .

Нормы $\|f_h\|$, $\|\varphi_h\|$ определены в нашем случае формулами

$$\|f_h\| = \max_k \max_j |f_{jk}| = \max_k \|f_k\|;$$

$$\|\varphi_h\| = \max_j |\varphi_j|.$$

Рассмотрим оператор R_h и оценим его. При $\omega < \frac{1}{2}$ имеет место неравенство

$$\begin{aligned} |u_{jh}(1 - 2\omega) + \omega(u_{j+1h} + u_{j-1h})| &\leq \\ &\leq |(1 - 2\omega) + 2\omega| \max_j |u_{jk}| = \|u_h\|. \end{aligned}$$

Откуда следует

$$|u_{jh+1}| \leq \|u_h\| + \tau \|f_h\|.$$

Так как правая часть неравенства не зависит от j , то неравенство справедливо и для того j , где u_{jh+1} наибольшее, т. е.

$$\|u_{h+1}\| \leq \|u_h\| + \tau \|f_h\|.$$

Построим соответствующие неравенства для всех k :

$$\|u_1\| \leq \|u_0\| + \tau \|f_h\|,$$

$$\|u_2\| \leq \|u_1\| + \tau \|f_h\|,$$

$$\dots \dots \dots$$

$$\|u_{h+1}\| \leq \|u_h\| + \tau \|f_h\|.$$

Складывая почленно полученные неравенства, имеем

$$\|u_{h+1}\| \leq \|u_0\| + (k+1)\tau \|f_h\|.$$

Введя постоянную $T > (k+1)\tau$, получим

$$\|u_{h+1}\| \leq \|u_0\| + T \|f_h\|. \quad (3.48)$$

Неравенство (3.48) не зависит от k и, следовательно, оно справедливо и при том k , где $\|u_h\|$ достигает наибольшего значения

$$\|u_h\| \leq \|\varphi_h\| + T \|f_h\|. \quad (3.49)$$

Неравенство (3.49) определяет устойчивость.

Описанный способ доказательства устойчивости может быть применен в тех случаях, когда возможно расслоение схемы. Он носит название: устойчивость как ограниченность степеней оператора перехода. Укажем ограничения, которым должна удовлетворять устойчивая схема в общем случае.

Если конечно-разностная схема

$$L_h u_h = f_h,$$

$$l_h u_h = \varphi_h,$$

может быть сведена к форме

$$u_{h+1} = R_h u_h + \tau \rho_h$$

(т. е. расслаивается) и выполняются условия

$$1) \|u_0\| \leq k_1 \|f_h\| + k_2 \|\varphi_h\|,$$

$$2) \|\rho_h\| \leq k_3 \|f_h\| + k_4 \|\varphi_h\|,$$

а оператор перехода ограничен с любой степенью n

$$\|R_h^n\| \leq N,$$

то схема устойчива. Очевидно, в примере рассматривался очень простой частный случай, где $N = 1$:

$$\|R_h^n\| \leq \|R_h\|^n \leq 1,$$

так как $\|u_{k+1}^{(1)}\| \leq \|u_k\|$. Кроме того, $k_1 = 0$, $k_2 = 1$, $k_3 = 1$, $k_4 = 0$.

3. Введенные нами определения аппроксимации и устойчивости по нормам, позволяют легко доказать важную для исследования конечно-разностных схем теорему.

Теорема. Если для конечно-разностной схемы имеет место аппроксимация порядка k :

$$\left. \begin{aligned} \| \{Lu\}_h - L_h[u]_h \|_{F_h} &\leq M_1 h^k, \\ \| \{lu\}_h - l_h[u]_h \|_{\Phi_h} &\leq M_2 h^k, \\ \| f_h - \{f\}_h \|_{F_h} &\leq M_3 h^k, \\ \| \varphi_h - \{\varphi\}_h \|_{\Phi_h} &\leq M_4 h^k \end{aligned} \right\} \quad (3.50)$$

и устойчивость

$$\|u_h\| \leq N_1 \|f_h\| + N_2 \|\varphi_h\|,$$

то схема сходится, т. е. выполняется неравенство

$$\| [u]_h - u_h \| \leq Kh^k.$$

Замечание. Следует обратить внимание на то, что при доказательстве теоремы мы будем повторять в общем виде все действия, которые производили, доказывая в § 2 этой главы сходимость конечно-разностной схемы для задачи Дирихле.

Доказательство. Обозначим $[u]_h - u_h = v_h$. Применим оператор L_h к сеточной функции

$$L_h v_h = L_h [u]_h - L_h u_h.$$

Прибавляя и отнимая $\{Lu\}_h$, получим

$$L_h v_h = L_h [u]_h - \{Lu\}_h + \{Lu\}_h - L_h u_h.$$

Оценивая по норме, имеем

$$\|L_h v_h\| \leq \|L_h [u]_h - \{Lu\}_h\| + \|\{Lu\}_h - L_h u_h\|$$

или, используя (3.50)

$$\|L_h v_h\| \leq M_1 h^k + M_3 h^k.$$

Аналогично

$$\|L_h v_h\| \leq M_2 h^k + M_4 h^k.$$

Подставляя полученное в неравенство, определяющее устойчивость, имеем

$$\|v_h\| \leq (M_1 + M_3) N_1 h^k + (M_2 + M_4) N_2 h^k = K h^k,$$

и окончательно

$$\|[u]_h - u_h\| \leq K h^k.$$

Это и доказывает сходимость.

4. Рассмотрим теперь другой прием проверки конечно-разностной схемы на устойчивость. Этот прием часто применяется в задачах геофизики. Опишем сначала критерий Курсанта—Фридрихса—Леви. Этот метод будем называть условно методом Фурье, так как в этом случае разностные уравнения решают с помощью рядов Фурье.

Рассмотрим, например, уравнение

$$\frac{\partial u}{\partial t} = -v \frac{\partial u}{\partial x},$$

аппроксимация которого описана в § 3 этой главы. Используем пятиточечную схему

$$u_{jk+1} - u_{jk-1} = -v \frac{\tau}{h} (u_{j+1k} - u_{j-1k}). \quad (3.51)$$

Будем искать решение уравнения в виде

$$u_{jh} = \varphi^k e^{imhj}, \quad (i^2 = -1). \quad (3.52)$$

Подставим (3.52) в (3.51)

$$\varphi^{k+1} - \varphi^k = -2iv \frac{\tau}{h} \sin mh, \quad (3.53)$$

так как

$$\begin{aligned} e^{imh(j+1)} - e^{imh(j-1)} &= 2ie^{imh} \frac{1}{2i} (e^{imh} - \\ &- e^{-imh}) = 2ie^{imh} \sin mh. \end{aligned}$$

Уравнение (3.53) легко преобразовать:

$$\varphi^2 - 2iU\varphi - 1 = 0, \quad (3.54)$$

$$U = v \frac{\tau}{h} \sin mh.$$

Решив квадратное уравнение (3.54), получим $\varphi_{1,2} = -iU \pm \sqrt{1 - U^2}$.

Если $U < 1$, то, обозначив $U = \sin \Theta$, $\sqrt{1 - U^2} = \cos \Theta$, получим

$$\varphi_{1,2} = \pm \cos \Theta - i \sin \Theta$$

или

$$\varphi_1 = e^{-i\Theta} = e^{-i \arcsin U}, \quad \varphi_2 = -e^{i \arcsin U},$$

и тогда

$$\varphi^k = A\varphi_1^k + B\varphi_2^k = Ae^{i\Theta k} + B(-1)^k e^{i\Theta k}.$$

В случае $U > 1$ и при $i = e^{\frac{\pi}{2}i}$

$$\varphi^k = \{A[-U + \sqrt{U^2 - 1}]^k + B[-U - \sqrt{U^2 - 1}]^k\} e^{\frac{\pi}{2}ik}.$$

Очевидно, величина $(-U - \sqrt{U^2 - 1})^k$ растёт с ростом k и, следовательно, условием устойчивости является $v \frac{\tau}{h} \sin mh < 1$ или $\frac{\tau}{h} < v$.

Такой же прием можно применить и к описанной в § 5 этой главы неявной схеме для уравнения теплопроводности (3.14)

$$u_{jh+1} = u_{jh} + \frac{h^2}{\tau} (u_{j+1, h+1} - 2u_{jh+1} + u_{j-1, h+1}).$$

Ищем решение в виде

$$u_{jh} = \varphi^h e^{ijmh}.$$

Отсюда

$$\varphi^{k+1} e^{imjh} = \varphi^k e^{imjh} \left[1 + \frac{h^2}{\tau} \varphi (e^{ihm} + e^{-ihm} - 2) \right]$$

или

$$\varphi = 1 - 2 \frac{h^2}{\tau} \varphi (1 - \cos mh).$$

Проводя простейшие преобразования, получим

$$\varphi = 1 - 4 \frac{h^2}{\tau} \varphi \sin^2 \frac{mh}{2}.$$

Откуда

$$\varphi = \frac{1}{1 + 4 \frac{h^2}{\tau} \sin^2 \frac{mh}{2}}.$$

Очевидно, при любых соотношениях между τ и h^2 имеет место устойчивость, так как $|\varphi| < 1$ и φ^k не растет при $k \rightarrow \infty$.

5. Критерий Неймана. Опишем этот критерий на примере простейшего уравнения (задача Коши)

$$\frac{\partial u}{\partial t} = Au, \quad u|_{t=0} = u_0.$$

Простейшая двухточечная схема имеет вид

$$u_{k+1} = G_h u_k, \quad G_h = 1 + A\tau \quad \text{и} \quad u_{k+1} = G_h^{k+1} u_0,$$

где G_h^k должно быть ограничено для устойчивости схемы.

Если же мы рассмотрим трехточечную схему, то, введя матрицы

$$U_k = \begin{pmatrix} u_k \\ u_{k-1} \end{pmatrix}, \quad G = \begin{pmatrix} 2\tau A & 1 \\ 1 & 0 \end{pmatrix} \quad \text{и} \quad u_k = u_k + 0u_{k-1},$$

получим $U_{k+1} = GU_k = G^k U_0$.

Критерием устойчивости в этом случае является ограниченность матрицы перехода по норме

$$\|G^k\| \leq \left(1 + \frac{BT}{k}\right)^k < e^{BT} \quad \text{при} \quad B = \text{const}, \quad T = k\tau.$$

В более сложных задачах рассматриваются матрицы перехода для коэффициентов Фурье. В геофизической литературе используется необходимое условие устойчивости Неймана, Лакса, Рихтмайера в виде требований, чтобы модуль наибольшего собствен-

ного числа матрицы перехода удовлетворял условию $|\lambda| \leq 1 + O(\Delta t)$. В случае симметричной матрицы требование $|\lambda| \leq 1$ является и достаточным условием устойчивости.

В заключение укажем на сходство критерия устойчивости, введенном в иностранной литературе [19, 27]; описанного в пункте пятом, с критерием устойчивости, введенным Годуновым [6], описанным в пункте втором. Очевидно, различны только формы. В геофизической практике чаще всего применяют критерий Неймана для матрицы перехода коэффициентов Фурье, как функций времени.

Рассмотрим два примера применения этих критериев в задачах геофизики.

§ 8. Примеры применения метода сеток в задачах океанологии

Задача о штормовых нагонах (описана Вольцингером и Пясковским [4]).

1. Рассмотрим сначала простейшую постановку задачи: линейную одномерную для невязкой жидкости. В этом случае мы будем решать следующую систему уравнений:

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + gh \frac{\partial \xi}{\partial x} &= \tau(x, t), \\ \frac{\partial \xi}{\partial t} + \frac{\partial u}{\partial x} &= 0, \end{aligned} \right\} \quad (3.55)$$

— где x — декартова координата сечения, перпендикулярного основному потоку; t — время; ξ — превышение уровня над его равновесным положением; $h + \xi$ — глубина моря; $\tau(x, t)$ — функция, учитывающая силы трения.

Рассмотрим задачу Коши с начальными условиями

$$\left. \begin{aligned} u(x, 0) &= \varphi_1(x), \\ \xi(x, 0) &= \varphi_2(x). \end{aligned} \right\} \quad (3.56)$$

Системы (3.55), (3.56) аппроксимируются с помощью четырех-точечной схемы в виде

$$\left. \begin{aligned} \frac{u_{j,k+1} - \frac{u_{j+1,k} + u_{j-1,k}}{2}}{\Delta t} + gh \frac{\xi_{j+1,k} - \xi_{j-1,k}}{2\Delta x} &= \tau_{j,k}, \\ \frac{\xi_{j,k+1} - \frac{\xi_{j+1,k} + \xi_{j-1,k}}{2}}{\Delta t} + \frac{u_{j+1,k} - u_{j-1,k}}{2\Delta x} &= 0. \end{aligned} \right\} \quad (3.57)$$

Аппроксимация начальных данных не имеет погрешности. Аппроксимация дифференциальных уравнений имеет порядок $O(\Delta x^2 + \Delta t)$.

Расчет по схеме (3.57) аналогичен расчету по явной схеме, описанному в § 5 данной главы. Отличием является то, что на каждом шаге следует определить две функции: u_{jh+1} , ξ_{jh+1} . Расчет идет послойно, начиная с первого. Значения в нулевом слое известны из начальных условий.

Покажем, что при выполнении условия $\frac{\Delta t}{\Delta x} < gh$ конечно-разностная схема (3.57) устойчива. Для простоты проверки сделаем сначала ряд преобразований. Введем новые функции

$$p = u + \frac{\xi}{gh}; \quad \psi(x) = \varphi_1(x) + \frac{\varphi_2(x)}{gh}.$$

Подставив (3.57) в (3.55), (3.56), получим

$$\frac{\partial p}{\partial t} + gh \frac{\partial p}{\partial x} = \tau, \quad p(x, 0) = \psi(x). \quad (3.58)$$

Разностная схема, аппроксимирующая граничную задачу (3.58), имеет вид

$$\left. \begin{aligned} p_{jh+1} &= \frac{1}{2} \left(1 - \frac{\Delta t}{\Delta x} gh \right) p_{j+1k} + \frac{1}{2} \left(1 + \frac{\Delta t}{\Delta x} \right) p_{j-1k} + \tau_{jk} \Delta t, \\ p_{j0} &= \psi_j. \end{aligned} \right\} \quad (3.59)$$

Рассмотрим нормированное пространство функций p_h , введя норму формулой

$$\|p_h\| = \max_{j,k} |p_{jk}|.$$

Аналогично

$$\|\tau_h\| = \max_{j,k} |\tau_{jk}|,$$

$$\|\psi_h\| = \max_j |\varphi_{1j}| + \max_j \frac{1}{gh} |\varphi_{2j}|.$$

Если потребовать выполнения условия

$$\frac{\Delta x}{\Delta t} < gh, \quad (3.60)$$

то можно доказать устойчивость построенной разностной схемы (3.59), пользуясь методом, изложенным в предыдущем параграфе.

Из формулы (3.59) при условии (3.60) следует

$$|p_{j\,k+1}| \leq \max_j |p_{j\,k}| + \Delta t \max_{j,k} |\tau_{j\,k}|,$$

или

$$\max_j |p_{j\,k+1}| \leq \max_j |p_{j\,k}| + \Delta t \|\tau_h\|. \quad (3.61)$$

Применяя неравенство (3.61) ко всем слоям, начиная с $k=0$ до $k=m$, и суммируя полученные неравенства, имеем

$$\|p_h\| \leq \|\psi_h\| + m\Delta t \|\tau_h\|,$$

что и доказывает устойчивость схемы.

II. Рассмотрим разностную схему, предложенную Фишером [27] для интегрирования линейной системы уравнений гидромеханики, которую принято называть примитивной. Система описывает некоторое одномерное движение жидкости, но содержит сла-

гаемое $A \frac{\partial^2 p}{\partial x^2}$, не соответствующее обычным физическим условиям, введенное для симметрии. Система имеет вид

$$\left. \begin{aligned} \frac{\partial u}{\partial t} &= -v \frac{\partial u}{\partial x} - \gamma \frac{\partial p}{\partial x} + A \frac{\partial^2 u}{\partial x^2}, \\ \frac{\partial p}{\partial t} &= -v \frac{\partial p}{\partial x} - \gamma \frac{\partial u}{\partial x} + A \frac{\partial^2 p}{\partial x^2}, \end{aligned} \right\} \quad (3.62)$$

где v, γ, A — постоянные.

Системе соответствует разностная схема (рис. 14):

$$\begin{aligned} u_{j\,k+1} &= u_{j\,k} - \left[v \frac{\Delta t}{\Delta x} \frac{1}{2} (u_{j+2k} - u_{j-2k}) - \frac{1}{2} v^2 \left(\frac{\Delta t}{\Delta x} \right) (u_{j+2k} + \right. \\ &\quad \left. + u_{j-2k} - 2u_{j\,k}) \right] - \frac{\gamma \Delta t}{\Delta x} (p_{j+1k+\frac{1}{2}} - p_{j-1k+\frac{1}{2}}) + \\ &\quad + A \frac{\Delta t}{(\Delta x)^2} (u_{j+2k} + u_{j-2k} - 2u_{j\,k}); \\ p_{j\,k+\frac{3}{2}} &= p_{j\,k+\frac{1}{2}} - \left[\frac{v \Delta t}{2 \Delta x} (p_{j+2k+\frac{1}{2}} - p_{j-2k+\frac{1}{2}} - \frac{1}{2} v^2 \left(\frac{\Delta t}{\Delta x} \right)^2 (p_{j+2k+\frac{1}{2}} + \right. \\ &\quad \left. + p_{j-2k+\frac{1}{2}} - 2p_{j\,k+\frac{1}{2}}) \right] - \frac{\gamma \Delta t}{\Delta x} (u_{j+1k+1} - u_{j-1k+1}) + \\ &\quad + A \frac{\Delta t}{(\Delta x)^2} (p_{j+2k+\frac{1}{2}} + p_{j-2k+\frac{1}{2}} - 2p_{j\,k+\frac{1}{2}}). \end{aligned}$$

Погрешность аппроксимации имеет порядок $\Delta x^2 + \Delta t$. Схема рассчитывается послойно. Сначала вычисляются значения u_j в слое $(k+1)\Delta t$, а потом — p_j в слое $(k+\frac{3}{2})\Delta t$. Расчет начинается с нулевого слоя для u_j и со слоя $\frac{1}{2}\Delta t$ для p_j , где заданы начальные значения этих функций. Применим метод Фурье для проверки устойчивости.

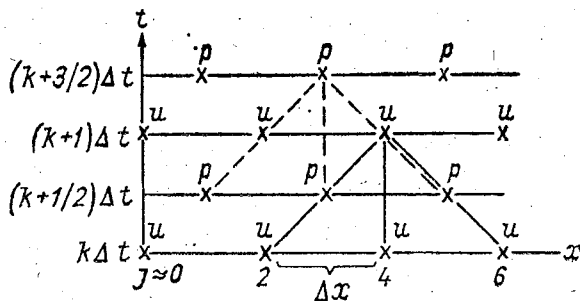


Рис. 14. Расположение точек в плоскости

Всякое решение системы (3.62) определяется видом начальных волн, амплитуда которых уменьшается. Для того чтобы решение разностной схемы было устойчиво, оно должно состоять из волн, амплитуда которых не растет со временем. Рассмотрим общий член ряда Фурье

$$\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} u_m \\ p_m \end{pmatrix} e^{\frac{2\pi i x}{L}},$$

предполагая, что решение периодическое и L — период. При переходе от слоя к слою, мы получим связь между коэффициентами Фурье

$$\begin{pmatrix} u_m^{(k)} \\ p_m^{k+\frac{1}{2}} \end{pmatrix} = G^k \begin{pmatrix} u_m^{(0)} \\ p_m^{(\frac{1}{2})} \end{pmatrix}.$$

Чтобы схема была устойчива, матрица G должна быть ограничена по норме и, следовательно, собственные числа матрицы не должны по модулю превосходить единицу. G — матрица второго порядка и ее собственные числа можно определить непосредственно, но выражение для них получается таким громоздким, что исследование устойчивости в общем виде не производится.

Рассмотрим частные случаи.

1. Если $v = 0$, то с помощью непосредственных преобразований получим

$$|\lambda_{1,2}| = 1 - 4 \frac{A \Delta t}{(\Delta x)^2} \sin \frac{2\pi \Delta x}{L} \leq 1$$

при условии

$$A \frac{\Delta t}{(\Delta x)^2} + \frac{\gamma^2 (\Delta t)^2}{(\Delta x)^2} \leq 1.$$

2. Если $\sin \frac{\pi \Delta x}{L} \ll 1$ (случай длинных волн) и если

$$A \frac{\Delta t}{(\Delta x)^2} \sin \frac{\pi \Delta x}{L} \ll 1,$$

$$\gamma \frac{\Delta t}{\Delta x} \sin \frac{\pi \Delta x}{L} \ll 1,$$

$$v \frac{\Delta t}{\Delta x} \sin \frac{\pi \Delta x}{L} \ll 1,$$

то

$$|\lambda_{1,2}| \cong 1 - 4 \left(\frac{\pi \Delta x}{L} \right)^2 \left(2 \frac{A \Delta t}{(\Delta x)^2} \pm \frac{\gamma \Delta t}{\Delta x} v \frac{\Delta t}{\Delta x} \right).$$

Если дополнительно пренебречь трением, то $|\lambda_1| > 1$ при $v > 0$ и схема неустойчива.

Если же $A \neq 0$, т. е. трение учитывается, то условием устойчивости является

$$2A > |v| \gamma \Delta t.$$

Это условие, как обычно в явной схеме, накладывает ограничение на шаг по времени.

Замечание. При решении реальных задач исследование устойчивости оказывается более сложным. Существует еще один косвенный способ проверки устойчивости разностных схем. Одновременно с основными уравнениями — движения, неразрывности и т. д. рассчитывается уравнение, описывающее изменение полной энергии системы. Неограниченное возрастание этой величины с течением времени указывает на неустойчивость разностной схемы, так как изменение полной энергии во всех реальных задачах ограничено.

ЛИТЕРАТУРА

1. Березин И. С. и Жидков Н. П. Методы вычислений, т. 1, 2. М., Физматгиз, 1959. 464 с.
2. Бут Э. Д. Численные методы. М., Физматгиз, 1959. 239 с.
3. Вазов В. и Форсайт Дж. Разностные методы решения дифференциальных уравнений в частных производных. М., ИЛ, 1963. 487 с.
4. Вольцингер Н. Е. и Пясковский Р. В. Основные океанологические задачи теории мелкой воды. Л., Гидрометеиздат, 1968. 300 с.
5. Гавурин М. К. Лекции по методам вычислений. М., «Наука», 1971. 248 с.
6. Годунов С. К. и Рябенский В. С. Введение в теорию разностных схем. М., Физматгиз, 1962. 340 с.
7. Годунов С. К. Уравнения математической физики. М., «Наука», 1971. 416 с.
8. Канторович Л. В. и Крылов В. И. Приближенные методы высшего анализа. М., Физматгиз, 1962. 708 с.
9. Коллатц Л. Численные методы решения дифференциальных уравнений. М., ИЛ, 1953.
10. Коллатц Л. Функциональный анализ и вычислительная математика. М., «Мир», 1969. 447 с.
11. Крылов В. И. Приближенное вычисление интегралов. М., Физматгиз, 1959. 328 с.
12. Марчук Г. И. Численные методы в прогнозе погоды. Л., Гидрометеиздат, 1967. 356 с.
13. Марчук Г. И. Методы вычислительной математики. М., «Наука», 1973. 352 с.
14. Марчук Г. И. Численное решение задач динамики атмосферы и океана. Л., Гидрометеиздат, 1974. 303 с.
15. Математический практикум. Под редакцией Г. Н. Положего. М., Физматгиз, 1960. 512 с.
16. Милн В. Э. Численное решение дифференциальных уравнений. М., ИЛ, 1955.
17. Михлин С. Г. Численная реализация вариационных методов. М., «Наука», 1966. 432 с.
18. Михлин С. Г. и Смолицкий Х. Л. Приближенные методы решения дифференциальных и интегральных уравнений. М., «Наука», 1965. 383 с.
19. Рихтмайер Р. Д. Разностные методы решения краевых задач. М., ИЛ, 1960. 263 с.

20. Рябенский В. С. и Филиппов А. Ф. Об устойчивости разностных уравнений. М., Гостехиздат, 1956. 171 с.
21. Скарборо Дж. Численные методы математического анализа. М.—Л., ГТТИ, 1934. 440 с.
22. Сушкевич А. К. Основы высшей алгебры. М., ОНТИ, 1937. 418 с.
23. Фаддеев Д. К. и Фаддеева В. Н. Вычислительные методы линейной алгебры. М., Физматгиз, 1960. 656 с.
24. Хаузхолдер А. С. Основы численного анализа. М., ИЛ, 1956.
25. Хемминг Р. В. Численные методы. М., «Наука», 1972. 400 с.
26. Численные методы расчета штормовых нагонов. Сборник статей. Л., Гидрометеоздат, 1964. 223 с.
27. Численные методы решения задач динамики атмосферы и океана. Сборник статей. Л., Гидрометеоздат, 1968. 367 с.
-

ОГЛАВЛЕНИЕ

	Стр.
Предисловие	3
Введение	4

Глава I. ПРОСТЕЙШИЕ ЧИСЛЕННЫЕ МЕТОДЫ

§ 1. Численные решения алгебраических и трансцендентных уравнений	10
§ 2. Решение алгебраических уравнений по методу Лобачевского	15
§ 3. Алгебраическое интерполирование	24
§ 4. Численное дифференцирование	31
§ 5. Численное интегрирование	35
§ 6. Численное решение задачи Коши для обыкновенных дифференциальных уравнений	44

Глава II. ЧИСЛЕННЫЕ МЕТОДЫ ЛИНЕЙНОЙ АЛГЕБРЫ

§ 1. Элементы теории определителей	50
§ 2. Матрицы и их простейшие свойства	58
§ 3. Линейные нормированные пространства	66
§ 4. Точные методы линейной алгебры. Метод Гаусса	70
§ 5. Итерационные методы	84
§ 6. Численные методы отыскания собственных чисел и собственных векторов матрицы	96

Глава III. МЕТОД СЕТОК В ПРИМЕНЕНИИ К ЗАДАЧАМ МАТЕМАТИЧЕСКОЙ ФИЗИКИ

§ 1. Некоторые общие понятия метода сеток	104
§ 2. Пример применения метода сеток. Решение задачи Дирихле для уравнения Пуассона	106
	149

§ 3. Аппроксимация дифференциальных уравнений разностными	113
§ 4. Аппроксимация начальных и граничных данных для различных задач	119
§ 5. Решения некоторых разностных схем	122
§ 6. Метод матричной прогонки	126
§ 7. Некоторые простейшие способы исследования устойчивости конечно-разностных схем	134
§ 8. Примеры применения метода сеток в задачах океанологии	142
Литература.	147



Балуева Александра Сергеевна

**Простейшие численные методы решения задач гидрометеорологии
(учебное пособие)**

Редактор *Ю. П. Андрейков*

М-20480	Подп. к печати 31.07.75 г.	Объем 9,5 п. л.	Зак. 2
Типография ВОК ВМФ	Тираж 1000	Цена 70 коп.	

