

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

В главе VIII рассмотрены основные методы численного решения различных типов задач для обыкновенных дифференциальных уравнений. В § 1 изложены постановка и методы решения задачи с начальными условиями (задачи Коши); эти методы применяются и при решении других типов задач. В § 2 даны постановки и методы решения краевых задач, а в § 3 — задач на собственные значения.

§ 1. Задача Коши

1. Постановка задачи. Обыкновенными дифференциальными уравнениями можно описать задачи движения системы взаимодействующих материальных точек, химической кинетики, электрических цепей, сопротивления материалов (например, статический прогиб упругого стержня) и многие другие. Ряд важных задач для уравнений в частных производных также сводится к задачам для обыкновенных дифференциальных уравнений. Так бывает, если многомерная задача допускает разделение переменных (например, задачи на нахождение собственных колебаний упругих балок и мембран простейшей формы, или определение спектра собственных значений энергии частицы в сферически-симметричном поле), или если ее решение зависит только от некоторой комбинации переменных (так называемые автомодельные решения). Таким образом, решение обыкновенных дифференциальных уравнений занимает важное место среди прикладных задач физики, химии и техники.

Конкретная прикладная задача может приводить к дифференциальному уравнению любого порядка, или к системе уравнений любого порядка. Но известно, что обыкновенное дифференциальное уравнение p -го порядка

$$u^{(p)}(x) = f(x, u, u', u'', \dots, u^{(p-1)})$$

при помощи замены $u^{(k)}(x) \equiv u_k(x)$ можно свести к эквивалентной системе p уравнений первого порядка

$$\begin{aligned} u'_k(x) &= u_{k+1}(x), & 0 \leq k \leq p-2, \\ u'_{p-1}(x) &= f(x, u_0, u_1, \dots, u_{p-1}), \end{aligned}$$

где $u_0(x) \equiv u(x)$. Аналогично, произвольную систему дифференциальных уравнений любого порядка можно заменить некоторой эквивалентной системой уравнений первого порядка. Поэтому в дальнейшем мы будем, как правило, рассматривать системы уравнений первого порядка

$$u'_k(x) = f_k(x, u_1, u_2, \dots, u_p), \quad 1 \leq k \leq p, \quad (1a)$$

записывая их для краткости в векторной форме

$$\begin{aligned} u'(x) &= f(x, u(x)), \\ u &= \{u_1, u_2, \dots, u_p\}, \quad f = \{f_1, f_2, \dots, f_p\}. \end{aligned} \quad (1b)$$

Известно, что система p -го порядка (1a) имеет множество решений, которое в общем случае зависит от p параметров $c = \{c_1, c_2, \dots, c_p\}$ и может быть записано в форме $u = u(x; c)$. Для определения значений этих параметров, т. е. для выделения единственного (или нужного) решения, надо наложить p дополнительных условий на функции $u_k(x)$.

Различают три основных типа задач для обыкновенных дифференциальных уравнений: задачи Коши, краевые задачи и задачи на собственные значения.

Задача Коши (задача с начальными условиями) имеет дополнительные условия вида

$$u_k(\xi) = \eta_k, \quad 1 \leq k \leq p, \quad (2)$$

т. е. заданы значения всех функций в одной и той же точке $x = \xi$. Эти условия можно рассматривать как задание координат начальной точки $(\xi, \eta_1, \eta_2, \dots, \eta_p)$ интегральной кривой в $(p+1)$ -мерном пространстве $(x, u_1, u_2, \dots, u_p)$. Решение при этом обычно требуется найти на некотором отрезке $\xi \leq x \leq X$ (или $X \leq x \leq \xi$), так что точку $x = \xi$ [можно считать начальной точкой этого отрезка.

Напомним*), что если правые части (1) непрерывны и ограничены в некоторой окрестности начальной точки $(\xi, \eta_1, \eta_2, \dots, \eta_p)$, то задача Коши (1) — (2) имеет решение, но, вообще говоря, не единственное. Если правые части не только непрерывны, но и удовлетворяют условию Липшица по переменным u_k , то решение задачи Коши единственно и непрерывно зависит от координат начальной точки, т. е. задача корректно поставлена. Если вдобавок правые части имеют непрерывные производные по всем аргументам вплоть до q -го порядка, то решение $u(x)$ имеет $q+1$ непрерывную производную по x .

2. Методы решения. Их можно условно разбить на точные, приближенные и численные. К точным относятся методы, позволяющие выразить решение дифференциального уравнения

*) См., например, [37].

через элементарные функции, либо представить его при помощи квадратур от элементарных функций. Эти методы изучаются в курсах обыкновенных дифференциальных уравнений. Нахождение точного решения задачи (1) — (2), а тем более — общего решения системы (1) облегчает качественное исследование этого решения и дальнейшие действия с ним.

Однако классы уравнений, для которых разработаны методы получения точных решений, сравнительно узки и охватывают только малую часть возникающих на практике задач. Например, доказано, что решение несложного уравнения

$$u'(x) = x^2 + u^2(x) \quad (3)$$

не выражается через элементарные функции. А уравнение

$$u'(x) = \frac{u-x}{u+x} \quad (4)$$

можно точно проинтегрировать и найти общее решение

$$\frac{1}{2} \ln(x^2 + u^2) + \operatorname{arctg} \frac{u}{x} = \text{const.} \quad (5)$$

Однако для того, чтобы составить таблицу значений $u(x)$, надо численно решить трансцендентное уравнение (5), а это несколько не проще, чем непосредственно численно проинтегрировать уравнение (4)!

Приближенными будем называть методы, в которых решение получается как предел $u(x)$ некоторой последовательности $y_n(x)$, причем $y_n(x)$ выражаются через элементарные функции или при помощи квадратур. Ограничиваясь конечным числом n , получаем приближенное выражение для $u(x)$. Примером может служить метод разложения решения в обобщенный степенной ряд, рассматриваемый в курсах обыкновенных дифференциальных уравнений; некоторые другие приближенные методы будут изложены в этой главе. Однако эти методы удобны лишь в том случае, когда большую часть промежуточных выкладок удастся сделать точно (например, найти явное выражение коэффициентов ряда). Это выполнимо лишь для сравнительно простых задач (таких как линейные), что сильно сужает область применения приближенных методов.

Численные методы — это алгоритмы вычисления приближенных (а иногда — точных) значений искомого решения $u(x)$ на некоторой выбранной сетке значений аргумента x_n . Решение при этом получается в виде таблицы. Численные методы не позволяют найти общего решения системы (1); они могут дать только какое-то частное решение, например, решение задачи Коши (1) — (2). Это основной недостаток численных методов. Зато эти методы применимы к очень широким классам уравнений и всем типам

задач для них. Поэтому с появлением быстродействующих ЭВМ численные методы решения стали одним из основных способов решения конкретных практических задач для обыкновенных дифференциальных уравнений.

Численные методы можно применять только к корректно поставленным (или регуляризованным) задачам. Заметим, однако, что для успешного применения численных методов формальное выполнение условий корректности может оказаться недостаточным. Надо, чтобы задача была *хорошо обусловлена*, т. е. малые изменения начальных условий приводили бы к достаточно малому изменению интегральных кривых. Если это условие не выполнено, т. е. задача *плохо обусловлена (слабо устойчива)*, то небольшие изменения начальных условий или эквивалентные этим изменениям небольшие погрешности численного метода могут сильно исказить решение.

В качестве примера плохой обусловленности рассмотрим задачу

$$u'(x) = u - x, \quad 0 \leq x \leq 100, \quad (6a)$$

$$u(0) = 1. \quad (6б)$$

Общее решение уравнения (6a) содержит одну произвольную постоянную

$$u(x; c) = 1 + x + ce^x.$$

При начальном условии (6б) она равна $c = 0$, так что $u(100) = 101$. Однако небольшое изменение начального условия $\bar{u}(0) = 1,000001$ слегка меняет постоянную: $\bar{c} = 10^{-6}$; тогда $\bar{u}(100) \approx 2,7 \times 10^{37}$, т. е. решение изменилось очень сильно.

В этом параграфе рассмотрены методы решения задачи Коши. Для простоты записи мы почти всюду ограничимся случаем одного уравнения первого порядка. Алгоритмы для случая системы p уравнений (1б) легко получаются из алгоритмов, составленных для одного уравнения, формальной заменой $u(x)$ и $f(x, u)$ на $\mathbf{u}(x)$ и $\mathbf{f}(x, \mathbf{u})$.

3. Метод Пикара. Это приближенный метод решения, являющийся обобщением метода последовательных приближений (см. главу V, § 2). Рассмотрим задачу Коши для уравнения первого порядка

$$u'(x) = f(x, u(x)), \quad \xi \leq x \leq X, \quad u(\xi) = \eta. \quad (7)$$

Интегрируя дифференциальное уравнение, заменим эту задачу эквивалентным ей интегральным уравнением типа Вольтерра

$$u(x) = \eta + \int_{\xi}^x f(\tau, u(\tau)) d\tau. \quad (8)$$

Решая это интегральное уравнение методом последовательных приближений, получим итерационный процесс Пикара

$$y_s(x) = \eta + \int_{\xi}^x f(\tau, y_{s-1}(\tau)) d\tau, \quad y_0(x) \equiv \eta \quad (9)$$

(приближенное решение, в отличие от точного, мы будем обозначать через y). На каждой итерации этого процесса интегрирование выполняется либо точно, либо численными методами, описанными в главе IV.

Докажем сходимость метода, предполагая, что в некоторой ограниченной области $G(x, u)$ правая часть $f(x, u)$ непрерывна и удовлетворяет по переменной u условию Липшица $|f(x, u_1) - f(x, u_2)| \leq L |u_1 - u_2|$.

Поскольку область $G(x, u)$ ограничена, то выполняются соотношения $|x - \xi| \leq a$, $|u - \eta| \leq b$. Обозначим погрешность приближенного решения через $z_s(x) = y_s(x) - u(x)$. Вычитая (8) из (9) и используя условие Липшица, получим

$$|z_s(x)| \leq L \int_{\xi}^x |z_{s-1}(\tau)| d\tau.$$

Решая это рекуррентное соотношение и учитывая, что $|z_0(x)| = |\eta - u(x)| \leq b$, найдем последовательно

$$|z_1(x)| \leq bL(x - \xi), \quad |z_2(x)| \leq \frac{1}{2} bL^2(x - \xi)^2, \dots,$$

$$|z_s(x)| \leq \frac{1}{s!} bL^s(x - \xi)^s.$$

Отсюда следует оценка погрешности

$$|z_s(x)| \leq \frac{b}{s!} (aL)^s \approx \frac{b}{\sqrt{2\pi s}} \left(\frac{eaL}{s}\right)^s. \quad (10)$$

Видно, что $\max |z_s(x)| \rightarrow 0$ при $s \rightarrow \infty$, т. е. *приближенное решение равномерно сходится к точному во всей области $G(x, u)$.*

Пример. Применим метод Пикара к задаче Коши для уравнения (3), решение которого не выражается через элементарные функции

$$u'(x) = x^2 + u^2, \quad u(0) = 0.$$

В этом случае квадратуры (9) вычисляются точно, и мы легко получаем

$$y_0(x) = 0, \quad y_1(x) = \frac{1}{3} x^3, \quad y_2(x) = \frac{1}{3} x^3 \left(1 + \frac{1}{21} x^4\right),$$

$$y_3(x) = \frac{1}{3} x^3 \left(1 + \frac{1}{21} x^4 + \frac{2}{693} x^8 + \frac{1}{19845} x^{12}\right),$$

и т. д. Видно, что при $x \leq 1$ эти приближения быстро сходятся и позволяют вычислить решение с высокой точностью.

Из этого примера видно, что метод Пикара выгодно применять, если интегралы (9) удастся вычислить через элементарные функции. Если же правая часть уравнения (7) более сложна, так что эти интегралы приходится находить численными методами, то метод Пикара становится не слишком удобным.

Метод Пикара легко обобщается на системы уравнений способом, описанным в п. 2. Однако на практике чем выше порядок системы, тем реже удается точно вычислять интегралы в (9), что ограничивает применение метода в этом случае.

Имеется много других приближенных методов. Например, С. А. Чаплыгин предложил метод, являющийся обобщением алгебраического метода Ньютона на случай дифференциальных уравнений. Другой способ обобщения метода Ньютона предложил Л. В. Канторович в 1948 г. В обоих этих методах, так же как и в методе Пикара, итерации выполняются при помощи квадратур. Однако квадратуры в них имеют гораздо более сложный вид, чем (9), и редко берутся в элементарных функциях. Поэтому эти методы почти не применяют.

4. Метод малого параметра. Достаточно простыми оказываются вычисления методом малого параметра, предложенным Пуанкаре в 1892 г. Пусть правая часть уравнения $u' = f(x, u; \lambda)$ зависит от параметра и известно частное решение $y_0(x)$ при некотором значении параметра $\lambda = \lambda_0$. Будем искать решение в виде ряда

$$u(x) = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n y_n(x). \quad (11)$$

Подставляя этот ряд в исходное уравнение и разлагая $f(x, u; \lambda)$ по формуле Тейлора по степеням $(\lambda - \lambda_0)$, получим для определения $y_n(x)$ линейные уравнения

$$y_n'(x) = \alpha_n(x) y_n(x) + v_n(x), \quad n = 1, 2, 3, \dots \quad (12)$$

Здесь коэффициенты $\alpha_n(x)$ выражаются через производные $f(x, u; \lambda)$ при $u = y_0(x)$, $\lambda = \lambda_0$, а функции $v_n(x)$ выражаются через $y_k(x)$, $0 \leq k < n$. Тем самым нахождение $y_n(x)$ сводится к квадратурам. Достаточным условием сходимости ряда (11) является аналитичность $f(x, u; \lambda)$ по всем аргументам.

При практическом применении метода малого параметра специально вводят в правую часть уравнения (7) параметр так, чтобы при некотором его значении легко находилось частное решение; после этого действуют по описанной схеме. Например, для уравнения (3) можно прибавить к правой части член λu^2 , положив, таким образом, $f(x, u; \lambda) = x^2 + (1 + \lambda) u^2$; тогда при $\lambda_0 = -1$ сразу видно частное решение $y_0(x) = \frac{1}{3} x^3 + c$, где постоянная c определяется из начального условия.

Метод малого параметра естественно переносится на уравнения высоких порядков или на системы уравнений. При этом

вместо цепочки последовательно решаемых линейных уравнений (12) возникают цепочки систем линейных дифференциальных уравнений. Однако все выкладки становятся существенно более громоздкими.

5. Метод ломаных. Это простейший численный метод. В практике вычислений он употребляется очень редко из-за невысокой точности. Но на его примере удобно пояснить способы построения и исследования численных методов.

Рассмотрим задачу Коши (7) и выберем на отрезке $[\xi, X]$ некоторую сетку $\{x_n, 0 \leq n \leq N\}$ значений аргумента так, чтобы выполнялись соотношения $\xi = x_0 < x_1 < x_2 < \dots < x_N = X$ (сетка может быть неравномерной). Разлагая решение $u(x)$ по формуле Тейлора на интервале сетки $x_n \leq x \leq x_{n+1}$ и обозначая $u(x_n) = u_n$, получим

$$u_{n+1} = u_n + h_n u'_n + \frac{1}{2} h_n^2 u''_n + \dots, \quad h_n = x_{n+1} - x_n. \quad (13)$$

Стоящие в правой части производные можно найти, дифференцируя уравнение (7) требуемое число раз:

$$u' = f(x, u), \quad u'' = \frac{d}{dx} f(x, u) = f_x + f f_u \quad (14)$$

и т. д. В принципе, если $f(x, u)$ имеет q -е непрерывные производные по совокупности аргументов, то в разложении (13) можно удерживать члены вплоть до $O(h^{q+1})$.

Однако использовать для расчетов формулу (13) с большим числом членов невыгодно. Во-первых, даже при сравнительно простой правой части выражения для производных могут оказаться громоздкими. Во-вторых, если правая часть известна лишь приближенно, то находить ее производные нежелательно. В простейшем случае, подставляя (14) в (13) и ограничиваясь только первым членом разложения, получим *схему ломаных* *):

$$y_{n+1} = y_n + h_n f(x_n, y_n), \quad h_n = x_{n+1} - x_n. \quad (15)$$

Поскольку при такой замене можно найти только приближенные значения искомой функции в узлах, то будем обозначать эти значения через y_n в отличие от точных значений $u_n = u(x_n)$. Для численного расчета по схеме ломаных достаточно задать начальное значение $y_0 = \eta$. Затем по формуле (15) последовательно вычисляем величины y_1, y_2, \dots, y_N .

Геометрическая интерпретация этой схемы дана на рис. 41, где изображено поле интегральных кривых. Использование только первого члена формулы Тейлора означает движение не по интегральной кривой, а по касательной к ней. На каждом шаге мы

*) Она была предложена Эйлером и называется также схемой Эйлера.

заново находим касательную; следовательно, траектория движения будет ломаной линией.

Исследуем сходимость метода ломаных, предполагая правую часть $f(x, u)$ непрерывной и ограниченной вместе со своими первыми производными: $|f| \leq M_1$, $|f_x| \leq M_2$, $|f_u| \leq M_3$ (отсюда следует, что $|u''| \leq M_4 = M_2 + M_1 M_3$).

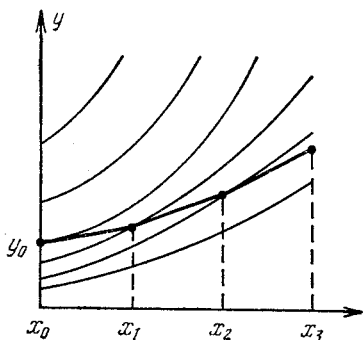


Рис. 41.

Рассмотрим погрешность приближенного решения $z_n = y_n - u_n$. Вычитая (15) из (13), получим соотношение, связывающее погрешности в соседних узлах сетки:

$$\begin{aligned} z_{n+1} &= z_n + h_n [f(x_n, y_n) - \\ &\quad - f(x_n, u_n)] - \frac{1}{2} h_n^2 u_n'' = \\ &= z_n (1 + h f_u)_n - \frac{1}{2} h_n^2 u_n'' \quad (16) \end{aligned}$$

(члены более высокого порядка малости здесь опущены). Последовательно применяя рекуррентное соотношение (16), выразим погрешность на произвольном шаге через погрешность начальных данных

$$z_m = z_0 \prod_{n=0}^{m-1} (1 + h f_u)_n - \frac{1}{2} \sum_{n=0}^{m-1} h_n^2 u_n'' \prod_{k=n+1}^{m-1} (1 + h f_u)_k. \quad (17)$$

Отсюда нетрудно дать асимптотическую оценку погрешности. Заметим, что при малых шагах сетки

$$\begin{aligned} \prod_{n=0}^{m-1} (1 + h f_u)_n &\approx \prod_{n=0}^{m-1} \exp(h f_u)_n = \\ &= \exp \left[\sum_{n=0}^{m-1} (h f_u)_n \right] \approx \exp \left[\int_{x_0}^{x_m} f_u(\tau, u(\tau)) d\tau \right], \end{aligned}$$

причем в качестве верхнего предела интеграла можно взять x_m , ибо ошибка при этом остается в пределах общей точности преобразований. Аналогично преобразуя второй член (17), получим

$$z_m = z_0 \exp \left(\int_{x_0}^{x_m} f_u d\tau \right) - \frac{1}{2} \int_{x_0}^{x_m} d\tau h(\tau) u''(\tau) \exp \left(\int_{\tau}^{x_m} f_u d\mu \right). \quad (18)$$

Здесь $h(x)$ — непрерывная функция, дающая в каждом узле x_n величину шага h_n ; в качестве такой функции можно выбрать линейный сплайн.

Рассмотрим структуру погрешности (18). Первое слагаемое справа связано с погрешностью начального значения $z_0 = y_0 - u_0$, которая умножается на ограниченную (благодаря ограниченности производных) величину. Начальное значение можно задать точно и считать, что $z_0 = 0$. Остановимся на втором слагаемом. Оно обусловлено тем членом формулы Тейлора (13), который был отброшен при выводе схемы ломаных (15). Оценим это слагаемое сверху; заменяя все функции под интегралами их модулями и вынося $\max h(x)$ за знак интеграла, получим

$$|z_m| \leq M(x_m) \max_{0 \leq n \leq m} h_n = O(\max h_n), \quad (19a)$$

где

$$M(x_m) = \frac{1}{2} \int_{x_0}^{x_m} d\tau |u''(\tau)| \exp\left(\int_{\tau}^{x_m} |f_u| d\mu\right) \leq \frac{M_4}{2M_3} e^{M_3(x_m - x_0)}. \quad (19b)$$

Таким образом, при $h \rightarrow 0$ приближенное решение сходится к точному равномерно (на ограниченном отрезке $|x - x_0| \leq a$) с первым порядком точности.

Замечание 1. Оценка погрешности (19) является мажорантной. Для функций со знакопеременными производными эта оценка может быть сильно завышена по сравнению с асимптотической оценкой (18).

Замечание 2. Экспоненциальный член в оценке (18) характеризует расхождение интегральных кривых (см. рис. 41); если он очень велик, то исходная задача Коши плохо обусловлена.

Пример. Проинтегрируем по схеме Эйлера задачу Коши для уравнения (3):

$$u'(x) = x^2 + u^2, \quad 0 \leq x \leq 1, \quad u(0) = 0.$$

В таблице 18 даны численные решения $y(x)$, полученные на сетках с шагами $h = 1, 1/2$ и $1/4$; столбик $\tilde{y}(x)$ будет пояснен в п. 10. Приведено также точное решение $u(x)$, вычисленное методом Пикара (см. пример в п. 3). Видно, что схема Эйлера для получения удовлетворительной точности требует гораздо более малого шага, чем использованный здесь.

Таблица 18

x_n	y_n			\tilde{y}_n	$u(x)$
	$h=1$	$h=0,5$	$h=0,25$	$h=0,25$	
0,00	0,000	0,000	0,000	0,000	0,000
0,25	—	—	0,000	0,008	0,005
0,50	—	0,000	0,016	0,031	0,042
0,75	—	—	0,078	0,114	0,143
1,00	0,000	0,125	0,220	0,316	0,350

Ограничимся только написанными членами, так как уже они обеспечивают четвертый порядок точности. Для вычисления решения в следующей точке запишем дифференциальное уравнение в интегральной форме

$$u_{n+1} = u_n + \int_{x_n}^{x_{n+1}} f(x, u(x)) dx = u_n + \int_{x_n}^{x_{n+1}} F(x) dx \quad (29)$$

и подставим в него интерполяционный многочлен (28). Получим формулу Адамса для переменного шага

$$\begin{aligned} y_{n+1} = & y_n + h_n F(x_n) + \frac{1}{2} h_n^2 F(x_n, x_{n-1}) + \\ & + \frac{1}{6} h_n^3 (2h_n + 3h_{n-1}) F(x_n, x_{n-1}, x_{n-2}) + \\ & + \frac{1}{12} h_n^4 (3h_n^3 + 8h_n h_{n-1} + 4h_n h_{n-2} + 6h_{n-1}^2 + 6h_{n-1} h_{n-2}) \times \\ & \times F(x_n, x_{n-1}, x_{n-2}, x_{n-3}), \text{ где } h_n = x_{n+1} - x_n. \end{aligned} \quad (30)$$

Эта формула имеет четвертый порядок точности. Если отбросить последнее слагаемое, получим формулу третьего порядка точности. Аналогично получаются формулы низших порядков. Формула первого порядка совпадает со схемой ломаных.

Чаще пользуются менее громоздким вариантом формулы (30), рассчитанным на постоянный шаг интегрирования. Вместо разделенных разностей вводят конечные разности $\Delta^p F_n = = p! F(x_n, x_{n-1}, \dots, x_{n-p})$, приблизительно равные p -й производной в точке $(x_n + x_{n-p})/2$, и получают

$$y_{n+1} = y_n + h F_n + \frac{1}{2} h^2 \Delta^1 F_n + \frac{5}{12} h^3 \Delta^2 F_n + \frac{3}{8} h^4 \Delta^3 F_n. \quad (31)$$

Остаточный член этой формулы равен $(251/750) h^5 F^{IV}(x)$.

Метод без изменений переносится на системы уравнений первого порядка типа (25).

Чтобы начать расчет методом Адамса, недостаточно знать $y(x_0)$. Для начала расчета по формуле (30) надо знать величину решения в четырех точках x_0, x_1, x_2, x_3 (а при формуле p -го порядка точности — в p точках). Поэтому надо вычислить недостающие значения y_n каким-либо другим методом — методом Рунге — Кутты, или разложением по формуле Тейлора (13) — (14) с достаточно большим числом членов. При работе на ЭВМ это вдвое увеличивает объем программы. Кроме того, формулы (30) громоздки, а несложные формулы (31) рассчитаны только на постоянный шаг и требуют нестандартных действий при смене шага: надо перейти к формулам (30), сделать по ним четыре шага и снова вернуться

(21), равномерно сходится к точному решению с погрешностью $O(\max h_n^2)$, т. е. двуэшелонная схема Рунге—Кутты имеет второй порядок точности.

Формула (21) имеет неплохую точность и нередко используется в численных расчетах. При этом обычно полагают либо

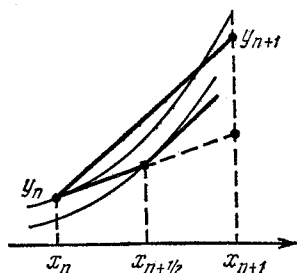


Рис. 42.

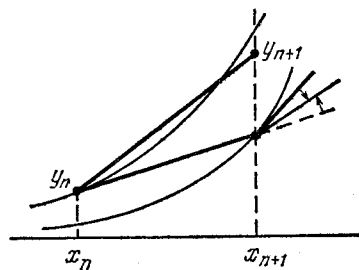


Рис. 43.

$\alpha = 1$, либо $\alpha = 1/2$. В первом случае получается схема особенно простого вида

$$y_{n+1} = y_n + hf \left(x_n + \frac{1}{2} h, y_n + \frac{1}{2} hf_n \right). \quad (22)$$

Ее смысл поясняется рис. 42. Сначала делаем половинный шаг по схеме ломаных, находя $y_{n+1/2} = y_n + \frac{1}{2} hf_n$. Затем в найденной точке определяем наклон интегральной кривой $y'_{n+1/2} = f(x_{n+1/2}, y_{n+1/2})$. По этому наклону определяем приращение функции на целом шаге $y_{n+1} = y_n + hy'_{n+1/2}$.

Геометрическая интерпретация второго случая

$$y_{n+1} = y_n + \frac{h}{2} [f(x_n, y_n) + f(x_n + h, y_n + hf_n)] \quad (23)$$

изображена на рис. 43. Здесь мы сначала грубо вычисляем по схеме ломаных значение функции $\bar{y}_{n+1} = y_n + hf_n$ и наклон интегральной кривой $\bar{y}'_{n+1} = f(x_{n+1}, \bar{y}_{n+1})$ в новой точке. Затем находим средний наклон на шаге $y'_{n+1/2} = (y'_n + \bar{y}'_{n+1})/2$ и по нему уточняем значение y_{n+1} . Схемы подобного типа нередко называют «предиктор — корректор».

Таблица 19

x_n	y_n		\tilde{y}_n	$u(x)$
	$h = 1$	$h = 0,5$	$h = 0,5$	
0,00	0,000	0,000	0,000	0,000
0,50	—	0,031	0,042	0,042
1,00	0,250	0,317	0,339	0,350

В таблице 19 приведен численный расчет по схеме (22) того же примера, который рассмотрен в таблице 18 (п.5). Из таблицы видно, что схема второго порядка точности дает существенно лучшие результаты, чем схема ломаных; уже расчет на грубой сетке с $h=0,5$ можно считать удовлетворительным.

Методом Рунге—Кутта можно строить схемы различного порядка точности. Например, схема ломаных (15) есть схема Рунге—Кутта первого порядка точности. Наиболее употребительны схемы четвертого порядка точности, образующие семейство четырехчленных схем. Приведем без вывода ту из них, которая записана в большинстве стандартных программ ЭВМ:

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

$$k_1 = f(x_n, y_n), \quad k_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} k_1\right), \quad (24)$$

$$k_3 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} k_2\right), \quad k_4 = f(x_n + h, y_n + h k_3)$$

(при величинах k_m и шаге h следует также ставить индекс сетки n , но для простоты записи мы его опускаем).

Формулы более высокого порядка точности практически не употребляются. Пятичленные формулы имеют всего лишь четвертый порядок точности; шестичленные имеют шестой порядок, но слишком громоздки. Кроме того, высокий порядок реализуется лишь при наличии у правой части непрерывных производных соответствующего порядка.

Схемы Рунге—Кутта имеют ряд важных достоинств. 1) Все они (кроме схемы ломаных) имеют хорошую точность. 2) Они являются явными, т. е. значение y_{n+1} вычисляется по ранее найденным значениям за определенное число действий по определенным формулам. 3) Все схемы допускают расчет переменным шагом; значит, нетрудно уменьшить шаг там, где функция быстро меняется, и увеличить его в обратном случае. 4) Для начала расчета достаточно выбрать сетку x_n и задать значение $y_0 = \eta$; далее вычисления идут по одним и тем же формулам. Все эти свойства схем очень ценны при расчетах на ЭВМ.

На случай систем уравнений схемы Рунге—Кутта легко переносятся, как во всех других методах, при помощи формальной замены $y, f(x, y)$ на $\mathbf{y}, \mathbf{f}(x, \mathbf{y})$. Нетрудно произвести покомпонентную запись этих схем. Например, для системы двух уравнений

$$u'(x) = f(x, u(x), v(x)),$$

$$v'(x) = g(x, u(x), v(x)), \quad (25)$$

обозначая через y, z приближенные значения функций $u(x),$

$v(x)$, запишем аналогичную (24) четырехчленную схему следующим образом:

$$\begin{aligned} y_{n+1} &= y_n + \frac{1}{6} h (k_1 + 2k_2 + 2k_3 + k_4), \\ z_{n+1} &= z_n + \frac{1}{6} h (q_1 + 2q_2 + 2q_3 + q_4), \end{aligned} \quad (26a)$$

где

$$\begin{aligned} k_1 &= f(x_n, y_n, z_n), \quad q_1 = g(x_n, y_n, z_n), \\ k_2 &= f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1, z_n + \frac{1}{2}hq_1\right), \\ q_2 &= g\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1, z_n + \frac{1}{2}hq_1\right), \\ k_3 &= f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2, z_n + \frac{1}{2}hq_2\right), \\ q_3 &= g\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2, z_n + \frac{1}{2}hq_2\right), \\ k_4 &= f(x_n + h, y_n + hk_3, z_n + hq_3), \\ q_4 &= g(x_n + h, y_n + hk_3, z_n + hq_3). \end{aligned} \quad (26б)$$

Напомним, что именно эта схема четвертого порядка точности (разумеется, записанная для системы произвольного числа уравнений) лежит в основе большинства стандартных программ численного решения задачи Коши на ЭВМ.

З а м е ч а н и е. Погрешности различных схем Рунге—Кутта связаны с максимумами модулей соответствующих производных $f(x, u)$ громоздкими выражениями типа (18)—(19). Наглядное представление о величине этих погрешностей можно получить в одном частном случае, когда $f = f(x)$. При этом дифференциальное уравнение сводится к квадратуре, а все схемы численного интегрирования переходят в квадратурные формулы. Легко убедиться, что схема (22) переходит в формулу средних (4.16), схема (23)—в формулу трапеций (4.7) с шагом h , а схема (24)—в формулу Симпсона (4.11) с шагом $h/2$. Напомним, что мажоранты остаточных членов этих формул на равномерной сетке с указанными шагами соответственно равны

$$\begin{aligned} R_{\text{сред}} &= \frac{b-a}{24} h^2 \max |f''|, \quad R_{\text{трап}} = \frac{b-a}{12} h^2 \max |f''|, \\ R_{\text{Симп}} &= \frac{b-a}{2880} h^4 \max |f^{IV}|. \end{aligned} \quad (27)$$

Численные коэффициенты в остаточных членах (27) малы; это является одной из причин хорошей точности схем Рунге—Кутта.

Какими из формул Рунге—Кутта целесообразно пользоваться в каждом конкретном случае и как выбирать шаг сетки?

Если правая часть дифференциального уравнения непрерывна и ограничена вместе со своими четвертыми производными (и эти производные не слишком велики), то хорошие результаты дает схема четвертого порядка (24) благодаря очень малому коэффициенту в остаточном члене и быстрому возрастанию точности при уменьшении шага. Если же правая часть не имеет указанных производных, то предельный порядок точности этой схемы не может реализоваться. Тогда не худшие (хотя, по-видимому, и не лучшие) результаты дают схемы меньшего порядка точности, равного порядку имеющихся производных; например, для двукратно непрерывно дифференцируемых правых частей — несложные схемы (21)—(23).

Шаг сетки следует выбирать настолько малым, чтобы обеспечить требуемую точность расчета; других ограничивающих шаг условий в методе Рунге—Кутта нет. Но выражения остаточных членов типа (18)—(19) слишком громоздки; поэтому априорными оценками точности для выбора шага в практических расчетах не пользуются. Удобнее делать расчеты со сгущением сетки, давая апостериорную оценку точности (подробнее это будет рассмотрено в п. 11).

Встречаются важные задачи, в которых функции являются достаточно гладкими, но настолько быстро меняющимися, что схемы Рунге—Кутта как низкого, так и высокого порядка точности требуют неприемлемо малого шага для получения удовлетворительного результата. Такие задачи требуют использования (а нередко — разработки) специальных методов, ориентированных на данный узкий класс задач.

7. Метод Адамса. Будем рассматривать правую часть уравнения $f(x, u)$ не на всей плоскости ее аргументов x, u , а только на определенной интегральной кривой $u(x)$, соответствующей искомому решению. Тогда она будет функцией только одного аргумента x ; обозначим ее через

$$F(x) \equiv f(x, u(x)).$$

Пусть нам уже известно приближенное решение в нескольких точках сетки: $y_n, y_{n-1}, \dots, y_{n-m}$. Тогда в этих точках известны также $F(x_k) = f(x_k, y_k)$. В окрестности этих узлов можно приближенно заменить $F(x)$ интерполяционным многочленом; запишем его для неравномерной сетки в форме Ньютона (2.8):

$$\begin{aligned} F(x) = & F(x_n) + (x - x_n) F(x_n, x_{n-1}) + \\ & + (x - x_n)(x - x_{n-1}) F(x_n, x_{n-1}, x_{n-2}) + \\ & + (x - x_n)(x - x_{n-1})(x - x_{n-2}) F(x_n, x_{n-1}, x_{n-2}, x_{n-3}) + \dots \end{aligned} \quad (28)$$

Ограничимся только написанными членами, так как уже они обеспечивают четвертый порядок точности. Для вычисления решения в следующей точке запишем дифференциальное уравнение в интегральной форме

$$u_{n+1} = u_n + \int_{x_n}^{x_{n+1}} f(x, u(x)) dx = u_n + \int_{x_n}^{x_{n+1}} F(x) dx \quad (29)$$

и подставим в него интерполяционный многочлен (28). Получим формулу Адамса для переменного шага

$$\begin{aligned} y_{n+1} = & y_n + h_n F(x_n) + \frac{1}{2} h_n^2 F(x_n, x_{n-1}) + \\ & + \frac{1}{6} h_n^3 (2h_n + 3h_{n-1}) F(x_n, x_{n-1}, x_{n-2}) + \\ & + \frac{1}{12} h_n^4 (3h_n^3 + 8h_n h_{n-1} + 4h_n h_{n-2} + 6h_{n-1}^2 + 6h_{n-1} h_{n-2}) \times \\ & \times F(x_n, x_{n-1}, x_{n-2}, x_{n-3}), \text{ где } h_n = x_{n+1} - x_n. \end{aligned} \quad (30)$$

Эта формула имеет четвертый порядок точности. Если отбросить последнее слагаемое, получим формулу третьего порядка точности. Аналогично получаются формулы низших порядков. Формула первого порядка совпадает со схемой ломаных.

Чаще пользуются менее громоздким вариантом формулы (30), рассчитанным на постоянный шаг интегрирования. Вместо разделенных разностей вводят конечные разности $\Delta^p F_n = = p! F(x_n, x_{n-1}, \dots, x_{n-p})$, приблизительно равные p -й производной в точке $(x_n + x_{n-p})/2$, и получают

$$y_{n+1} = y_n + h F_n + \frac{1}{2} h^2 \Delta^1 F_n + \frac{5}{12} h^3 \Delta^2 F_n + \frac{3}{8} h^4 \Delta^3 F_n. \quad (31)$$

Остаточный член этой формулы равен $(251/750) h^5 F^{IV}(x)$.

Метод без изменений переносится на системы уравнений первого порядка типа (25).

Чтобы начать расчет методом Адамса, недостаточно знать $y(x_0)$. Для начала расчета по формуле (30) надо знать величину решения в четырех точках x_0, x_1, x_2, x_3 (а при формуле p -го порядка точности — в p точках). Поэтому надо вычислить недостающие значения y_n каким-либо другим методом — методом Рунге — Кутты, или разложением по формуле Тейлора (13) — (14) с достаточно большим числом членов. При работе на ЭВМ это вдвое увеличивает объем программы. Кроме того, формулы (30) громоздки, а несложные формулы (31) рассчитаны только на постоянный шаг и требуют нестандартных действий при смене шага: надо перейти к формулам (30), сделать по ним четыре шага и снова вернуться

к формулам (31). Все это делает метод Адамса неудобным для расчетов на ЭВМ.

Внешне этот метод привлекателен тем, что за один шаг приходится только один раз вычислять $f(x, u)$, которая может быть очень сложной. А в четырехчленной схеме Рунге — Кутта того же порядка точности $f(x, u)$ вычисляется за шаг четыре раза. Однако коэффициент в остаточном члене (27) схемы Рунге — Кутта (24) меньше в 960 раз, чем в схеме (31)! Значит, при одинаковой точности схема Рунге — Кутта (24) позволяет брать шаг в $\sqrt[4]{960} = 5,7$ раза крупнее, т. е. фактически вычислять $f(x, u)$ даже меньшее число раз, чем в методе Адамса.

Поэтому сейчас метод Адамса и аналогичные методы (например, Милна) употребляются реже метода Рунге — Кутта.

8. Неявные схемы. Предыдущие методы были явными, т. е. значение y_{n+1} определялось за заранее известное число действий. Пример неявной схемы получим, если запишем дифференциальное уравнение в интегральной форме (29), а интеграл по одному интервалу сетки приближенно вычислим по формуле трапеций

$$y_{n+1} = y_n + \frac{1}{2} h [f(x_n, y_n) + f(x_{n+1}, y_{n+1})]. \quad (32)$$

Решая это алгебраическое уравнение, можно определить y_{n+1} , которое и будет приближенным значением искомого решения $u(x_n)$. Схема (32) имеет второй порядок точности, допускает счет неравномерным шагом, не требует специальных приемов для начала счета.

Но у этой схемы есть серьезные недостатки. Во-первых, неизвестно, имеет ли уравнение (32) вещественный корень, т. е. разрешима ли задача. Можно привести пример, когда при большом шаге корня нет. Пусть $f(x, u) = u^2$ и $u(0) = 1$; тогда на первом шаге $y_1 = 1 + \frac{1}{2} h(1 + y_1^2)$ и при $h > (1 + \sqrt{2})^{-1}$ вещественного корня нет.

Во-вторых, даже если корень есть, то как его найти? Метод Ньютона применять нежелательно, так как для этого надо дифференцировать $f(x, u)$. Метод деления пополам не обобщается на системы уравнений. Остается метод последовательных приближений

$$y_{n+1}^{(s)} = y_n + \frac{1}{2} h [f(x_n, y_n) + f(x_{n+1}, y_n^{(s-1)})]. \quad (33)$$

Однако он сходится к корню, только если $h|f_u| < 2$, т. е. при достаточно малом шаге. Если в ходе расчета f_u возрастает, то итерации (33) могут перестать сходиться.

От последней трудности можно избавиться, заодно уменьшив объем вычислений. Для этого ограничим заранее число итераций

и рассмотрим (33) как самостоятельную явную схему. Очевидно, вопроса о существовании корня при этом не возникает; y_{n+1} всегда определяется, даже если алгебраическое уравнение (32) вещественного корня не имеет.

Роль числа итераций хорошо видна на примере уравнения $u' = f(u)$. Естественное нулевое приближение есть $y_{n+1}^{(0)} = y_n$, так что первая и вторая итерации

$$y_{n+1}^{(1)} = y_n + hf(y_n).$$

$$y_{n+1}^{(2)} = y_n + \frac{1}{2} h [f(y_n) + f(y_n + hf_n)]$$

являются соответственно схемой ломаных (15) первого порядка точности и схемой Рунге — Кутта второго порядка точности (23) типа «предиктор — корректор». Дальнейшие итерации уже не увеличат порядка точности, так как он не может быть выше, чем в исходной схеме (32); они влияют только на коэффициенты в остаточном члене и увеличивают время счета.

Таким образом, неявные схемы с заданным числом итераций мало отличаются от схем Рунге — Кутта и бывают удобны лишь для некоторых нестандартных задач. Но они приводят к интересной идее ограничения числа итераций.

Есть эмпирическое правило, в общем случае не обоснованное. Пусть для решения дифференциального уравнения написана неявная схема p -го порядка точности. Разрешим ее методом последовательных приближений аналогично (33) и зададим число итераций. Тогда при одной итерации получим схему первого порядка точности, при двух — второго и так далее, при p итерациях — p -го порядка точности. Дальнейшее увеличение числа итераций уже не увеличивает порядок точности.

Это правило оказывается полезным даже для уравнений в частных производных. По существу схема с заданным числом итераций есть новая явная схема. Поэтому здесь не возникает вопроса о существовании корня или сходимости итераций, подобных (33).

9. Специальные методы. Из всех численных методов интегрирования обыкновенных дифференциальных уравнений, рассчитанных на произвольные уравнения (точнее, на классы уравнений, у которых правые части имеют определенное число непрерывных и ограниченных производных), наилучшие результаты и при расчетах на ЭВМ, и при ручных расчетах дают методы Рунге — Кутта. Поэтому, приступая к решению какой-либо конкретной задачи Коши, обычно пробуют решить ее одной из описанных в п. 6 схем.

Но выше отмечалось, что встречаются задачи с быстропеременными решениями, когда все схемы Рунге — Кутта для получения удовлетворительной точности требуют неприемлемо малого

шага. Характерным примером такой задачи является система уравнений химической кинетики.

Сначала разберем задачу химического распада одного вещества

$$u'(t) = -\alpha(t, u)u, \quad u(0) = u_0 > 0, \quad \alpha(t, u) > 0; \quad (34)$$

здесь u — концентрация вещества, t — время, α — скорость распада, которую считаем зависящей от t и u (ибо она зависит от температуры, а температура определяется выделением тепла при реакции и внешними условиями охлаждения). Запишем для уравнения (34) схему ломаных (15)

$$y_{n+1} = y_n [1 - \tau \alpha(t_n, y_n)], \quad \tau = t_{n+1} - t_n. \quad (35)$$

По смыслу задачи, концентрация вещества должна быть положительной. Но если скорость распада настолько велика, что хотя бы в одной точке $\alpha_n > 1/\tau$, то численное решение (35) будет знакопеременным, что физически бессмысленно. Применение вместо (15) схем Рунге—Кутта более высокого порядка точности лишь немного ослабляет указанное ограничение шага, не устраняя его (см. задачу 7).

Для одного уравнения (34) эта трудность несущественна: если скорость распада $\alpha(t, u)$ велика, то вещество распадается за малое время $t \sim \alpha^{-1}$, так что число шагов сетки $N = t/\tau$ будет умеренным. Но в системах уравнений химической кинетики присутствуют вещества с самыми различными константами распада или синтеза (нередко от $\alpha \sim 10^8$ сек⁻¹ до $\alpha \sim 10^{-2}$ сек⁻¹). Тогда общий промежуток времени будет определяться самой медленной реакцией ($t \sim 100$ сек), а допустимый шаг интегрирования — самой быстрой реакцией ($\tau \sim 10^{-8}$ сек). Ясно, что такой объем расчетов — около 10^{10} шагов — совершенно неприемлем*).

Для подобных задач приходится использовать специальные методы, разработанные именно для данных узких классов уравнений; для других классов уравнений эти методы обычно оказываются непригодными. Способы построения специальных методов основаны на изучении и использовании свойств общих решений исследуемого класса уравнений. Рассмотрим некоторые способы.

Большинство способов основано на том, что для исходного уравнения $u'(x) = f(x, u)$ стараются найти такое вспомогательное уравнение $v'(x) = g(x, v)$, чтобы решение последнего возможно более просто выражалось через элементарные функции, и при этом на заметном отрезке изменения аргумента выполнялось бы $u(x) \approx v(x)$. Иными словами, ищется приближенное решение, имеющее достаточно простой вид.

*) Кроме того, при таком числе шагов существенно сказывается некорректность задачи, связанная с ошибками округления.

Для нахождения приближенных решений можно применить метод Пикара или другие аналогичные методы. Нередко удается добиться успеха, слегка упрощая правую часть исходного уравнения. Например, если в задаче (34) положить $\alpha(t, u) \approx \alpha_0 = \text{const}$, тогда вспомогательное уравнение будет $v'(t) = -\alpha_0 v$, а его решением при заданном начальном условии является $v(t) = u_0 \times \exp[-\alpha_0(t - t_0)]$.

Первый способ построения специальных схем удобен для знакопеременных решений (например, быстро осциллирующих). В нем рассматривается разность $w(x) = u(x) - v(x)$. Вычитая вспомогательное уравнение из исходного, получим уравнение, которому удовлетворяет эта разность

$$w'(x) = f(x, v(x) + w) - g(x, v(x)); \quad (36)$$

здесь $v(x)$ — известная функция. Если $v(x)$ действительно является хорошим приближением к решению, то функция $w(x)$ невелика, поэтому уравнение (36) должно легко интегрироваться обычными схемами Рунге — Кутта.

Второй способ выгоден для знакопостоянных решений (например, растущих по экспоненциальному или степенному закону). В нем рассматривается отношение $w(x) = u(x)/v(x)$, а для систем уравнений — отношения $w_k(x) = u_k(x)/v_k(x)$. Нетрудно убедиться, что это отношение удовлетворяет уравнению

$$w'(x) = \frac{1}{v(x)} [f(x, wv(x)) - wg(x, v(x))], \quad (37)$$

где $v(x)$ — известное приближенное решение. Аналогично предыдущему случаю, полученное уравнение должно хорошо интегрироваться численно схемами Рунге — Кутта.

Пример. Если для уравнения распада (34) воспользоваться приближенным решением $v(t) = u_0 \exp[-\alpha_0(t - t_0)]$, то специальная схема (37) примет следующий вид:

$$w'(t) = -[\alpha_0 - \alpha(t, wv(t))] w;$$

при слабо меняющейся $\alpha(t, u)$ малость правой части очевидна.

Третий способ заключается в том, что вспомогательное уравнение рассматривается не на большом промежутке изменения аргумента, а на одном шаге сетки $x_n \leq x \leq x_{n+1}$. Берется его приближенное решение $v_n(x)$, удовлетворяющее начальному условию $v_n(x_n) = y_n \approx u(x_n)$. Поскольку интервал сетки невелик, то на нем приближенное решение будет близко к точному, поэтому можно положить $u(x_{n+1}) \approx y_{n+1} = v_n(x_{n+1})$. Этот способ означает написание такой разностной схемы, которой решение вспомогательного уравнения удовлетворяет точно, а решение исходного уравнения — приближенно, но с малой погрешностью.

Пример. Рассмотрим уравнение, возникающее в задачах так называемой дифференциальной прогонки:

$$u'(x) = -[u^2 + \rho(x)], \quad \rho(x) > 0. \quad (38)$$

Если положить $\rho(x) \approx \text{const} = \rho_{n+1/2}$ при $x_n \leq x \leq x_{n+1}$, то вспомогательное уравнение примет вид

$$v'(x) = -(v^2 + \rho), \quad \rho = \text{const}.$$

Оно интегрируется в элементарных функциях

$$\arctg \frac{v_{n+1}}{\sqrt{\rho}} - \arctg \frac{v_n}{\sqrt{\rho}} = -h \sqrt{\rho}.$$

Это соотношение явно разрешается, давая такую специальную схему:

$$y_{n+1} = \sqrt{\rho_{n+1/2}} \frac{y_n - \sqrt{\rho_{n+1/2}} \operatorname{tg}(h \sqrt{\rho_{n+1/2}})}{\sqrt{\rho_{n+1/2}} + y_n \operatorname{tg}(h \sqrt{\rho_{n+1/2}})}. \quad (39a)$$

Если можно считать $h \sqrt{\rho} \ll 1$, то схема (39a) упрощается:

$$y_{n+1} = \frac{y_n - h \rho_{n+1/2}}{1 + h y_n}. \quad (39б)$$

Схемы (39a) и (39б) дают неплохие результаты даже в тех случаях, когда условие устойчивости прогонки нарушено, а точное решение задачи (38) имеет полюсы.

При использовании третьего способа обычно удается построить схемы первого или второго порядка точности, но с малым остаточным членом (точнее, мала по величине комбинация производных, входящая множителем в остаточный член); схемы более высокого порядка точности построить этим путем трудно. Первый и второй способы позволяют использовать схемы Рунге—Кутты высокого порядка точности, но остаточный член при этом будет не очень мал, ибо решения $u(x)$ и $v(x)$ на большом отрезке изменения аргумента могут заметно отличаться, и правые части уравнений (36) или (37) становятся большими. Однако первый и второй способы также можно применить к одному интервалу сетки; на этом пути можно построить специальные схемы высокого порядка точности с малым остаточным членом. Заметим, что все эти способы по существу эквивалентны специально подобранным нелинейным интерполяциям искомого решения.

Упомянем четвертый способ, заключающийся в построении так называемых точных разностных схем, которым точно удовлетворяет решение исходной задачи. Коэффициенты таких

схем обычно являются функционалами от коэффициентов исходного уравнения (и могут зависеть также от искомого решения). Но техника построения точных схем более сложна, и мы их не будем рассматривать, отсылая читателя к монографии [30].

10. Особые точки. Решение может иметь в отдельных точках отрезка интегрирования особенности, обусловленные обращением в бесконечность правой части $f(x, u)$ или какой-нибудь ее производной. Сначала рассмотрим случай, когда начальная точка $x = x_0$ является особой. Есть три основных способа численного интегрирования таких решений. Рассмотрим их на примере задачи

$$u'(x) = \frac{1}{2\sqrt{x}} + u^2(x), \quad u(0) = 0, \quad (40)$$

где правая часть в начальной точке обращается в бесконечность; очевидно, начинать интегрирование по схеме Рунге — Кутты любого порядка точности при этом невозможно.

Первый способ — это найти такую замену переменных, которая преобразует уравнение к виду, не имеющему особенностей. Для задачи (40) достаточно сделать замену аргумента $x = t^2$; тогда эта задача принимает вид

$$\frac{du}{dt} = 1 + 2tu^2, \quad u(0) = 0,$$

который допускает применение стандартных численных методов.

Второй способ — построить в небольшой окрестности особой точки приближенное решение, выраженное через элементарные (или другие легко вычисляющиеся) функции. Например, выбирая нулевое приближение $y_0(x) \equiv 0$ и применяя к задаче (40) метод Пикара, получим

$$y_1(x) = \sqrt{x}, \quad y_2(x) = \sqrt{x} + \frac{1}{2}x^2, \dots$$

Отступим от особой точки на конечное расстояние в некоторую точку x_1 и вычислим в ней решение с требуемой точностью на основе найденного приближения. Точка x_1 уже не особая; ее можно считать первым узлом разностной сетки и вести из нее интегрирование стандартными численными методами.

Следует помнить, что если точка x_1 лежит близко к x_0 , то правая часть уравнения или ее производные еще велики в этой точке и стандартные численные методы дают заметную погрешность вблизи точки x_1 . Поэтому желательно выбирать точку x_1 подальше от x_0 . Но тогда, чтобы вычислить $u(x_1)$ с нужной точностью, необходимо строить достаточно хорошее приближенное решение: например, брать высокие приближения метода Пикара.

Третий способ — составить для данной задачи специальную схему, позволяющую вести численное интегрирование непосредственно от особой точки. Например, проинтегрируем уравнение (40) по одному интервалу сетки, и первое слагаемое в подинтегральном выражении проинтегрируем точно, а второе — по формуле прямоугольников с использованием левого конца интервала; тогда получим

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} \left[\frac{1}{2\sqrt{\xi}} + y^2(\xi) \right] d\xi \approx \\ \approx y_n + (\sqrt{x_{n+1}} - \sqrt{x_n}) + (x_{n+1} - x_n) y_n^2. \quad (41)$$

Это явная схема, напоминающая схему ломаных (15). Она построена по образцу схем первого порядка точности. Но имеет ли эта схема на самом деле точность $O(h)$ — заранее не очевидно, ибо производные правой части уравнения (40) не ограничены; этот вопрос требует дополнительного исследования.

Если решение имеет особенности во внутренних точках отрезка интегрирования, то при этом обычно нельзя сказать заранее, в каких именно точках: правая часть $f(x, u)$ зависит от решения, которое нам не известно. В этом случае целесообразно применять третий способ — составлять специальные схемы, не теряющие своей применимости вблизи особых точек. Примером является схема (39), позволяющая вести сквозной расчет даже при наличии у решения особенностей типа полюсов.

11. Сгущение сетки. Как получить требуемую точность расчета? Априорные оценки точности для этого мало полезны. Во-первых, остаточные члены выражаются через производные решения, которое до начала расчета не известно. Во-вторых, априорные оценки обычно являются мажорантными и могут во много раз превосходить фактическую ошибку расчета.

Имеются стандартные программы численного интегрирования дифференциальных уравнений с так называемым «автоматическим выбором шага». В них каждый шаг выбирается так, чтобы вносимая на нем погрешность не превышала заданной величины. Но при этом не учитывается, что эта погрешность в ходе дальнейших расчетов умножается на величину типа экспоненты в (18), т. е. может сильно возрасти. Кроме того, общее число шагов заранее не определено. В результате фактическая точность расчета по подобным программам обычно неизвестна.

Поэтому основным практическим приемом является апостериорная оценка точности. Для ее получения расчет проводят на двух или более сгущающихся сетках и применяют правило Рунге или Рунге — Ромберга (см. главу III, п. 3). Напомним, в чем оно заключается.

Вспомним априорную оценку погрешности схемы ломаных (18). Запишем ее, опуская первое слагаемое, связанное с неточным

заданием начальных данных:

$$z(x) = \frac{1}{2} \int_{x_0}^x d\tau h(\tau) u''(\tau) \exp\left(\int_{\tau}^x f_u d\mu\right); \quad (42)$$

здесь $h(x)$ есть некоторая функция, значение которой в каждом узле сетки дает величину шага. Для схем более высокого порядка точности p остаточный член имеет аналогичную структуру, но содержит $h^p(x)$ и соответствующие производные решения или правой части $f(x, u)$.

Если сетка равномерная, $h(x) = h = \text{const}$, то остаточный член типа (43) для схемы p -го порядка точности пропорционален h^p . Поэтому при сгущении равномерной сетки применима оценка точности по Рунге. Если имеются численные решения на двух сетках $y(x; h)$ и $y(x; rh)$, где $r > 1$, то погрешность решения на сетке с меньшим шагом составляет

$$\Delta y(x; h) \approx \frac{y(x; h) - y(x; rh)}{r^p - 1}. \quad (43)$$

Вместо оценки точности можно погрешность (43) прибавить к численному решению, уточнив его:

$$\tilde{y}(x; h) = y(x; h) + \frac{y(x; h) - y(x; rh)}{r^p - 1}, \quad (44)$$

но тогда вопрос о погрешности уточненного решения остается открытым.

Приведенное рассуждение справедливо и в том случае, если сетки с разным числом узлов не равномерны, но их можно описать функциями $h(x)$, отношение которых есть $h_I(x)/h_{II}(x) = r = \text{const}$. Это выполняется, например, для квазиравномерных сеток (описанных в главе III, п. 4).

При выводе оценок типа (18) старшими членами формулы Тейлора (13) пренебрегают. Если их учесть (считая правую часть уравнения непрерывно дифференцируемой достаточное число раз), то погрешность выразится суммой, где последующие слагаемые содержат более высокие степени $h(x)$ и соответствующие производные. В этом случае можно уточнять численное решение по правилу Ромберга или по рекуррентному правилу Рунге, используя расчеты на k различных сетках. Применение этих правил эквивалентно построению некоторой схемы более высокого порядка точности $q = p + k - 1$, где p — порядок точности исходной схемы. Разумеется, фактически получить точность $O(h^q)$ можно только для q раз непрерывно дифференцируемых решений $u(x)$.

Правило Рунге применимо для сеток с любым отношением шагов r . Но используют его преимущественно для целого r , когда все узлы менее подробной сетки являются узлами более

подробной; особенно удобно сгущать сетки вдвое (рис. 44). При этом как для равномерных, так и для квазиравномерных сеток условие совпадения узлов выполняется.

В тех узлах, которые являются общими для нескольких сеток, можно уточнить $y(x)$ непосредственно по правилу Рунге (44). Так, в n -м узле можно увеличить порядок точности на двойку, в $(n+2)$ -м — на единицу, а в $(n+1)$ -м — нельзя увеличить (рис. 44). Разумеется, если мы не уточняем решение, а лишь оцениваем погрешность, то достаточно найти ее по формуле (43) только в части узлов.

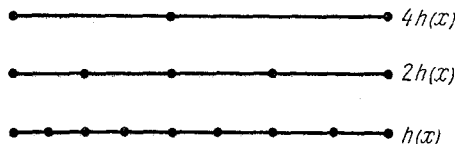


Рис. 44.

Однако можно уточнить функцию во всех узлах самой подробной сетки, если немного усложнить вычисления. Например, для двух нижних сеток на рис. 44 это делается так. Используем совпадающие узлы сеток для определения поправок к значениям функции

$$\Delta_m = [y(x_m; h) - y(x_m; rh)] / (r^p - 1), \quad m = n, n + 2. \quad (45a)$$

Значение поправок в остальных узлах найдем простейшей интерполяцией. Для равномерных или квазиравномерных сеток можно положить

$$\Delta_{n+1} = \frac{1}{2}(\Delta_n + \Delta_{n+2}). \quad (45b)$$

Затем вычислим уточненные значения

$$\tilde{y}(x_m; h) = y(x_m; h) + \Delta_m, \quad m = n, n + 1, n + 2. \quad (45b)$$

Этот способ легко обобщается на произвольное число сеток. Такое уточнение выгодно для специальных схем третьего типа, имеющих невысокий порядок точности; выполнить уточнение обычно проще, чем составить специальную схему высокого порядка точности.

Примеры применения правила Рунге даны в таблице 18 (п. 5) и таблице 19 (п. 6), содержащих численное решение задачи

$$u' = x^2 + u^2, \quad 0 \leq x \leq 1, \quad u(0) = 0.$$

В таблице 18 интегрирование выполнено по схеме ломаных (15), и для уточнения использованы сетки с $h=1$ и $h=0,5$; видно, что, несмотря на плохую точность исходной схемы, уточненное решение не сильно отличается от искомого. В таблице 19 уточнено численное решение, найденное по неплохой схеме Рунге — Кутта второго порядка точности (22); это уточнение уже близко к искомому решению, несмотря на очень грубую сетку.

§ 2. Краевые задачи

1. Постановки задач. Краевая задача — это задача отыскания частного решения системы (1а):

$$\frac{d}{dx} u_k(x) = f_k(x, u_1, u_2, \dots, u_p), \quad 1 \leq k \leq p,$$

на отрезке $a \leq x \leq b$, в которой дополнительные условия налагаются на значения функций $u_k(x)$ более чем в одной точке этого отрезка. Очевидно, что краевые задачи возможны для систем порядка не ниже второго.

Свое первоначальное название этот тип задач получил по простейшим случаям, когда часть дополнительных условий задается на одном конце отрезка, а остальная часть — на другом (т. е. только в точках $x=a$ и $x=b$). Примером является задача нахождения статического прогиба $u(x)$ нагруженной струны с закрепленными концами

$$u''(x) = -f(x), \quad a \leq x \leq b, \quad u(a) = u(b) = 0; \quad (46)$$

здесь $f(x)$ — внешняя изгибающая нагрузка на единицу длины струны, деленная на упругость струны.

Для уравнений или систем более высокого порядка, где число дополнительных условий больше двух, постановки краевых условий более разнообразны. При этом возможны случаи, когда часть условий задана во внутренних точках отрезка $[a, b]$; их нередко называют внутренними краевыми условиями. Например, статический прогиб нагруженного упругого бруска удовлетворяет уравнению четвертого порядка

$$u^{IV}(x) = f(x), \quad a \leq x \leq b; \quad (47a)$$

если этот брусок лежит в точках x_i , $1 \leq i \leq 4$, на опорах, то дополнительные условия имеют вид

$$u(x_i) = 0, \quad 1 \leq i \leq 4, \quad a \leq x_1 < x_2 < x_3 < x_4 \leq b, \quad (47b)$$

т. е. все они заданы в разных точках.

Сами дополнительные условия могут связывать между собой значения нескольких функций в одной точке (или даже в разных точках); тогда для системы p -го порядка (1) они примут вид

$$\begin{aligned} \Phi_k(u_1(\xi_k), u_2(\xi_k), \dots, u_p(\xi_k)) &= \eta_k, \\ 1 \leq k \leq p, \quad a \leq \xi_k \leq b. \end{aligned} \quad (48)$$

Существуют задачи с еще более сложными по форме дополнительными условиями, например, условиями нормировки

$$\int_a^b u_k^2(x) dx = 1, \quad (49)$$

обычными в квантовой механике, и т. д.

Несмотря на разнообразие форм краевых условий, краевые задачи решаются в основном одними и теми же численными методами, что оправдывает их объединение в один тип. Остановимся на методах решения.

Найти точное решение краевой задачи в элементарных функциях удается редко: для этого надо найти общее решение системы (1) и суметь явно определить из краевых условий значения входящих в него постоянных.

К приближенным методам решения краевых задач относятся разложение в ряды Фурье, методы Рунге и Галеркина. Ряды Фурье применяют к линейным задачам; этот метод излагается в курсах математической физики (см. [2, 40]) и здесь рассматриваться не будет. Остальные два метода применимы и к некоторым нелинейным задачам. Метод Рунге разбирается в главе VII, а метод Галеркина будет рассмотрен в этом параграфе.

Для численного решения краевых задач применяют метод стрельбы и разностный метод. Метод стрельбы основан на сведении краевой задачи к некоторой задаче Коши для той же системы уравнений. В разностном методе задача приближенно заменяется решением алгебраической системы уравнений с очень большим числом неизвестных (неизвестными являются значения решения в узлах сетки). В случае нелинейных задач оба метода являются итерационными; при этом построение хорошо сходящихся итерационных процессов само оказывается достаточно сложным.

2. Метод стрельбы (называемый также *баллистическим*). Это численный метод, заключающийся в сведении краевой задачи к некоторой задаче Коши для той же системы дифференциальных уравнений. Рассмотрим его на примере простейшей задачи для системы двух уравнений первого порядка с краевыми условиями достаточно общего вида

$$u'(x) = f(x, u, v), \quad v'(x) = g(x, u, v), \quad a \leq x \leq b, \quad (50a)$$

$$\varphi(u(a), v(a)) = 0, \quad \psi(u(b), v(b)) = 0. \quad (50b)$$

Выберем произвольно значение $u(a) = \eta$, рассмотрим левое краевое условие как алгебраическое уравнение $\varphi(\eta, v(a)) = 0$ и определим удовлетворяющее ему значение $v(a) = \zeta(\eta)$. Возьмем значения $u(a) = \eta$, $v(a) = \zeta$ в качестве начальных условий задачи Коши для системы (50a) и проинтегрируем эту задачу Коши

любым численным методом (например, по схемам Рунге — Кутта). При этом получим решение $u(x; \eta)$, $v(x; \eta)$, зависящее от η , как от параметра.

Значение ζ выбрано так, что найденное решение удовлетворяет левому краевому условию (50б). Однако правому краевому условию это решение, вообще говоря, не удовлетворяет: при его подстановке левая часть правого краевого условия, рассматриваемая как некоторая функция параметра η :

$$\bar{\psi}(\eta) = \psi(u(b; \eta) v(b; \eta)), \quad (51)$$

не обратится в нуль. Надо каким-либо способом менять параметр η , пока не подберем такое значение, для которого $\bar{\psi}(\eta) \approx 0$ с требуемой точностью. Таким образом, решение краевой задачи (50) сводится к нахождению корня одного алгебраического уравнения

$$\bar{\psi}(\eta) = 0. \quad (52)$$

Эта алгебраическая задача изучена в главе V, § 2. Рассмотрим, какие методы ее решения целесообразно применять в данном случае.

Простейшим является *метод дихотомии*. Делают пробные «выстрелы» — расчеты с наудачу выбранными значениями η_i , до тех пор, пока среди величин $\bar{\psi}(\eta_i)$ не окажется разных по знаку. Пара таких значений η_i, η_{i+1} образует «вилку». Деля ее последовательно пополам до получения нужной точности, производим «пристрелку» параметра η . Благодаря этому процессу весь метод получил название стрельбы.

Однако нахождение каждого нового значения функции $\bar{\psi}(\eta)$ требует численного интегрирования системы (50а), т. е. достаточно трудоемко. Поэтому корень уравнения (52) желательно находить более быстрым методом, чем дихотомия.

Если правые части уравнений (50а) и левые части краевых условий (50б) имеют непрерывные и ограниченные первые производные, то $\bar{\psi}(\eta)$ также будет иметь непрерывную производную*). В этом случае можно построить аналог *метода Ньютона*. Нам пока известен только способ вычисления $\bar{\psi}(\eta)$, а нужно научиться определять также производную

$$\frac{d\bar{\psi}(\eta)}{d\eta} = \frac{\partial\psi}{\partial u(b)} \frac{\partial u(b; \eta)}{\partial \eta} + \frac{\partial\psi}{\partial v(b)} \frac{\partial v(b; \eta)}{\partial \eta}. \quad (53)$$

Входящие сюда производные по параметру от решения задачи Коши можно найти, если продифференцировать по этому параметру систему (50а). Вводя обозначения

$$\mu(x; \eta) = \frac{\partial u(x; \eta)}{\partial \eta}, \quad \nu(x; \eta) = \frac{\partial v(x; \eta)}{\partial \eta} \quad (54)$$

*) Это следует из теорем о зависимости решения задач Коши от параметра (см. [37]).

и дифференцируя (50а) по параметру, получим

$$\begin{aligned}\frac{d\mu}{dx} &= f_u(x, u, v) \mu(x) + f_v(x, u, v) v(x), \\ \frac{dv}{dx} &= g_u(x, u, v) \mu(x) + g_v(x, u, v) v(x), \quad a \leq x \leq b\end{aligned}\quad (55a)$$

Одно из начальных условий для этой системы очевидно: $\mu(a) = \partial u(a)/\partial \eta = 1$; второе условие нетрудно найти, дифференцируя левое краевое условие (50б) по η . Отсюда получим

$$\mu(a) = 1, \quad v(a) = -\varphi_u(\eta, \zeta)/\varphi_v(\eta, \zeta). \quad (55б)$$

Интегрируя систему (55а) с начальными условиями (55б) совместно с задачей Коши для системы (50а), определим вспомогательные функции $\mu(x)$, $v(x)$. Подставляя их значения при $x=b$ в (53), найдем значение производной правого краевого условия по пристрелочному параметру. Новое значение параметра определяется по формуле касательных (5.28):

$$\eta_{s+1} = \eta_s - [\bar{\Psi}(\eta_s)/\bar{\Psi}'(\eta_s)]. \quad (56)$$

Однако описанный способ требует интегрирования лишней пары дифференциальных уравнений, что приводит к усложнению и двукратному увеличению трудоемкости каждой итерации. Поэтому им пользуются не часто.

Можно избежать этого усложнения, если решать уравнение (52) разностным аналогом метода Ньютона — *методом секущих*. Для этого первые два расчета делают с наудачу выбранными значениями η_0 , η_1 , а следующие значения параметра вычисляют по формуле (5.32):

$$\eta_{s+1} = \eta_s - \frac{(\eta_s - \eta_{s-1}) \bar{\Psi}(\eta_s)}{\bar{\Psi}(\eta_s) - \bar{\Psi}(\eta_{s-1})}. \quad (57)$$

Вместо этого процесса можно использовать *метод парабол*, в котором также не требуется располагать явным выражением производных, а достаточно лишь знать об их существовании. Напомним, что последние три метода быстро сходятся вблизи корня; сходимость вдали от корня зависит от того, насколько удачно выбрано нулевое приближение.

Линейные задачи решаются методом стрельбы особенно просто. Пусть система (50а) и краевые условия (50б) линейны;

$$u'(x) = \alpha_1(x) u + \beta_1(x) v + \gamma_1(x), \quad a \leq x \leq b, \quad (58a)$$

$$v'(x) = \alpha_2(x) u + \beta_2(x) v + \gamma_2(x),$$

$$p_1 u(a) + q_1 v(a) = r_1 \quad p_2 u(b) + q_2 v(b) = r_2. \quad (58б)$$

Тогда начальные условия соответствующей задачи Коши примут вид

$$u(a) = \eta, \quad v(a) = \zeta = (r_1 - p_1 \eta)/q_1. \quad (58в)$$

Нетрудно сообразить, что решение задачи Коши (58а), (58в) будет линейно зависеть от параметра η , поэтому $\bar{\Psi}(\eta)$ также

будет линейной функцией. Но линейная функция одного аргумента полностью определяется своими значениями в любых двух точках η_0 и η_1 , а ее график является прямой, т. е. совпадает со своей секущей. Значит, найденное по формуле секущих (57) значение η_2 является точным корнем уравнения (52), так что расчет с этим значением параметра даст искомое решение. Таким образом, для решения линейной краевой задачи (58а)—(58б) достаточно трижды решить задачу Коши.

З а м е ч а н и е. Для линейных задач можно несколько уменьшить объем расчетов, если воспользоваться тем, что общее решение линейной неоднородной системы равно сумме ее какого-нибудь частного решения и общего решения соответствующей однородной системы. Найдем частное решение неоднородной системы (58а), (58в), соответствующее значению $\eta_0 = 0$, и обозначим его через $u_0(x)$, $v_0(x)$. Затем рассмотрим соответствующую однородную задачу Коши

$$\begin{aligned} u'(x) &= \alpha_1(x)u + \beta_1(x)v, & v'(x) &= \alpha_2(x)u + \beta_2(x)v, \\ u(a) &= \eta_1 = 1, & v(a) &= -p_1/q_1; \end{aligned}$$

вычислим ее решение и обозначим его через $u_1(x)$, $v_1(x)$. Тогда общее решение неоднородной задачи Коши, удовлетворяющее (в силу выбора начальных условий) левому краевому условию (58б), является однопараметрическим семейством

$$u(x) = u_0(x) + cu_1(x), \quad v(x) = v_0(x) + cv_1(x). \quad (59)$$

Значение параметра c выбираем так, чтобы удовлетворить правому краевому условию (58б):

$$c = -\frac{p_2 u_0(b) + q_2 v_0(b) - r_2}{p_2 u_1(b) + q_2 v_1(b)}.$$

Затем найдем искомое решение по формуле (59), что позволяет избежать третьего интегрирования задачи Коши.

Метод стрельбы прост, применим как к линейным, так и к нелинейным задачам и позволяет использовать при численном интегрировании схемы Рунге—Кутта (или другие) высокого порядка точности. К большинству задач типа (50) он применяется успешно.

Затруднения возникают в тех случаях, когда краевая задача (50) хорошо обусловлена, а соответствующая ей задача Коши плохо обусловлена. При этом численное интегрирование задачи Коши определяет функцию $\bar{\psi}(\eta)$ с большой погрешностью, что осложняет организацию итераций.

В этом случае пробуют поставить начальные условия на другом конце отрезка $x = b$, т. е. интегрировать задачу Коши справа налево; нередко при этом устойчивость улучшается. Если изме-

нение направления интегрирования не помогает, то такую краевую задачу решают либо специальными, либо разностными методами.

Одним из специальных методов для линейных краевых задач является *дифференциальная прогонка* (ее идея предложена в [1], а подробное описание алгоритма имеется, например, в [3, 4]). Этот метод хорошо устойчив именно в том случае, когда задача Коши для исходной линейной системы плохо обусловлена; этот факт вызывал одно время большой интерес к прогонке. Однако при хорошей устойчивости линейной задачи Коши прогонка становится недостаточно устойчивой. Поэтому в настоящее время дифференциальная прогонка употребляется не часто. Обычно используются ее разностные аналоги, рассматриваемые ниже; они обеспечивают удовлетворительную устойчивость расчета в большинстве интересных случаев.

3. Уравнения высокого порядка или системы большого числа уравнений имеют соответствующее число краевых условий, и способы задания этих условий достаточно разнообразны. Поэтому к таким задачам применять метод стрельбы много труднее, чем к простейшей задаче (50).

Рассмотрим тот (сравнительно несложный) случай, когда для системы p уравнений

$$\frac{du_k(x)}{dx} = f_k(x, u_1, u_2, \dots, u_p), \quad 1 \leq k \leq p, \quad a \leq x \leq b, \quad (60a)$$

дополнительные условия заданы только на концах отрезка и имеют следующий вид:

$$\varphi_k(u_1(a), \dots, u_p(a)) = 0 \quad 1 \leq k \leq m, \quad (60б)$$

$$\varphi_k(u_1(b), \dots, u_p(b)) = 0 \quad m+1 \leq k \leq p. \quad (60в)$$

Для определенности, будем полагать $m \geq p/2$.

Выберем за исходный тот конец отрезка $[a, b]$, где задана *большая* часть краевых условий; в нашем случае это будет левый конец $x = a$. В качестве пристрелочных параметров возьмем $p - m$ каких-то функций $u_k(x)$ из полного набора, например,

$$u_q(a) = \eta_q, \quad 1 \leq q \leq p - m. \quad (61a)$$

Если подставить эти значения в левые краевые условия (60б), то эти условия образуют систему алгебраических уравнений относительно начальных значений остальных функций; решая эту систему, найдем

$$u_q(a) = \psi_q(\eta_1, \eta_2, \dots, \eta_{p-m}), \quad p - m + 1 \leq q \leq p. \quad (61б)$$

Рассмотрим задачу Коши для системы уравнений (60a) с начальными условиями (61a, б). Решение этой задачи, которое можно найти численным интегрированием, удовлетворяет левому краевому условию (60б) и зависит от параметров $\eta = \{\eta_1, \eta_2, \dots, \eta_{p-m}\}$. Подстановка этого решения в правые краевые условия (60в)

определяет вспомогательные функции параметров

$$\bar{\varphi}_k(\eta) = \varphi_k(u_1(b, \eta), \dots, u_p(b, \eta)), \quad m+1 \leq k \leq p; \quad (62a)$$

те значения параметров, которые удовлетворяют системе алгебраических уравнений

$$\bar{\varphi}_k(\eta_1, \eta_2, \dots, \eta_{p-m}) = 0, \quad m+1 \leq k \leq p, \quad (62b)$$

определяют искомое решение краевой задачи (60).

Напомним, что решение системы алгебраических уравнений высокого порядка само по себе является нелегкой задачей. Здесь оно осложняется тем, что вычисление функций $\bar{\varphi}_k(\eta)$ очень трудоемко, ибо требует численного интегрирования системы дифференциальных уравнений. Явный вид этих функций неизвестен, так что преобразовать систему (62b) к эквивалентной форме $\eta_q = \bar{\varphi}_q(\eta)$ и применять метод последовательных приближений затруднительно. А если мы захотим, как в п. 2, построить аналог метода Ньютона, то для вычисления матрицы производных $(\partial \bar{\varphi}_k / \partial \eta_q)$ надо будет дополнительно записать и численно интегрировать систему $p(p-m)$ дифференциальных уравнений.

Отсюда видно, что «пристрелка» большого числа параметров очень сложна. Поэтому для нелинейных задач метод стрельбы употребляют в основном тогда, когда $p-m=1$. Такие постановки краевых задач нередко встречаются в системах большого числа уравнений.

Линейные уравнения. В этом случае метод стрельбы сильно упрощается и позволяет легко решать задачи при любом числе параметров $p-m$. В самом деле, функции $\bar{\varphi}_k(\eta)$ будут линейными, т. е. они однозначно определяются по своим значениям в $p-m+1$ точке η^s , $1 \leq s \leq p-m+1$. Значит, выполнив $p-m+1$ интегрирование задачи Коши (60a), (61) с разными наборами параметров, можно найти искомый набор параметров η . Тогда $(p-m+2)$ -е интегрирование даст решение краевой задачи (60).

Вычисления при этом удобно вести следующим образом. Сначала возьмем некоторый набор параметров $\eta_1^0, \eta_2^0, \dots, \eta_{p-m}^0$ и обозначим полученные значения функций (62a) через $\bar{\varphi}_k^0 = \bar{\varphi}_k(\eta^0)$, $m+1 \leq k \leq p$. Затем изменим первый параметр на величину $\Delta\eta = 1$, т. е. возьмем набор $\eta_1^0 + 1, \eta_2^0, \eta_3^0, \dots, \eta_{p-m}^0$ и обозначим полученные значения функций через $\bar{\varphi}_k^1$. Затем возьмем набор $\eta_1^0, \eta_2^0 + 1, \eta_3^0, \dots, \eta_{p-m}^0$ и т. д. Выполнив полный цикл вычислений, можно записать каждую функцию в виде многомерного интерполяционного многочлена Ньютона первой степени (2.33):

$$\bar{\varphi}_k(\eta) = \bar{\varphi}_k^0 + \sum_{q=1}^{p-m} (\bar{\varphi}_k^q - \bar{\varphi}_k^0)(\eta_q - \eta_q^0), \quad m+1 \leq k \leq p.$$

Приравнивая эти функции нулю, получим систему линейных алгебраических уравнений для определения искоемых параметров η_q :

$$\sum_{q=1}^{p-m} (\bar{\varphi}_k^q - \bar{\varphi}_k^0) \eta_q = \sum_{q=1}^{p-m} (\bar{\varphi}_k^q - \bar{\varphi}_k^0) \eta_q^0 - \bar{\varphi}_k^0, \quad m+1 \leq k \leq n. \quad (63)$$

Заметим, что можно уменьшить на единицу число интегрирований системы линейных дифференциальных уравнений, если воспользоваться приемом, описанным в п. 2; но при большом значении $p-m$ это лишь незначительно сокращает общий объем вычислений, а организацию расчета усложняет.

4. Разностный метод; линейные задачи. Подробно рассмотрим разностный метод на примере простейшей краевой задачи для линейного уравнения второго порядка с краевыми условиями первого рода

$$u''(x) - p(x)u(x) = f(x), \quad a \leq x \leq b, \quad (64a)$$

$$u(a) = \alpha, \quad u(b) = \beta. \quad (64б)$$

Введем на $[a, b]$ сетку $a = x_0 < x_1 < x_2 < \dots < x_N = b$, которую для упрощения выкладок будем считать равномерной. Приближенно выразим вторую производную от решения через значения решения в узлах сетки $u_n = u(x_n)$; например, воспользуемся простейшей аппроксимацией (3.7):

$$u''(x_n) \approx \frac{1}{h^2} (u_{n-1} - 2u_n + u_{n+1}), \quad h = x_{n+1} - x_n = \text{const}. \quad (65)$$

Такую аппроксимацию можно записать в каждом внутреннем узле сетки x_n , $1 \leq n \leq N-1$. Если подставить ее в уравнение (64a), то уравнение станет приближенным; точно удовлетворять этому уравнению будет уже не искомое решение $u(x)$, а некоторое приближенное решение $y_n \approx u(x_n)$. Выполняя эту подстановку и обозначая $p_n = p(x_n)$ и $f_n = f(x_n)$, получим

$$y_{n-1} - (2 + h^2 p_n) y_n + y_{n+1} = h^2 f_n, \quad 1 \leq n \leq N-1. \quad (66a)$$

Эта система состоит из $N-1$ алгебраического уравнения, а неизвестными в ней являются приближенные значения решения в узлах сетки. Число неизвестных y_n , $0 \leq n \leq N$, равно $N+1$, т. е. оно больше, чем число уравнений (66a). Недостающие два уравнения легко получить из краевых условий (64б):

$$y_0 = \alpha, \quad y_N = \beta. \quad (66б)$$

Решая алгебраическую систему (66a, б), найдем приближенное решение.

При таком подходе возникает три вопроса. 1) Существует ли (вещественное) решение алгебраической системы типа (66)? 2) Как

фактически находить это решение? 3) Сходится ли разностное решение к точному в какой-либо норме при стремлении шага сетки к нулю?

В качестве иллюстрации проведем полное исследование рассмотренного выше примера, дополнительно требуя $p(x) > 0$.

Сначала рассмотрим вопрос о существовании разностного решения. Исходная задача (64) была линейной, разностная аппроксимация (65) — тоже линейна. Благодаря этому система (66а, б) оказалась системой линейных алгебраических уравнений. Поскольку $p_n > 0$, то в матрице этой системы диагональные элементы преобладают: в каждой строке модуль диагонального элемента больше суммы модулей остальных элементов. Как отмечалось в главе V, § 3, п. 4, при этом решение линейной системы существует и единственно.

Вычислить решение линейной системы всегда можно методом исключения Гаусса. В данном случае благодаря использованию трехточечной аппроксимации (65) система (66) имеет трехдиагональную матрицу. Поэтому решение экономично находится частным случаем метода Гаусса — методом алгебраической прогонки (см. главу V, § 1, п. 5).

Докажем утверждение: *если $p(x)$ и $f(x)$ дважды непрерывно дифференцируемы, то разностное решение равномерно сходится к точному с погрешностью $O(h^2)$ при $h \rightarrow 0$.*

При сделанном предположении $u(x)$ имеет четвертую непрерывную производную; тогда для погрешности аппроксимации (65) справедливо соотношение (3.12):

$$\frac{1}{h^2} (u_{n-1} - 2u_n + u_{n+1}) - u''(x_n) = \frac{1}{12} h^2 u^{IV}(\xi_n), \quad x_{n-1} < \xi_n < x_{n+1}.$$

Значит, точное решение удовлетворяет разностному уравнению

$$u_{n-1} - (2 + h^2 p_n) u_n + u_{n+1} = h^2 f_n + \frac{h^4}{12} u^{IV}(\xi_n), \quad 1 \leq n \leq N-1.$$

Вычитая из него уравнение (66а), получим уравнение, которому удовлетворяет погрешность $z_n = y_n - u(x_n)$; его удобно записать в следующем виде:

$$(2 + h^2 p_n) z_n = z_{n-1} + z_{n+1} + \frac{h^4}{12} u^{IV}(\xi_n), \quad 1 \leq n \leq N-1, \quad (67a)$$

$$z_0 = 0, \quad z_N = 0. \quad (67b)$$

Последние два уравнения являются очевидным следствием того, что уравнение (66б) точно передает граничное условие первого рода.

Выберем такую точку x_n , где $|z_n|$ достигает своего максимума; очевидно, это не граничная точка. Учитывая условие $p_n > 0$,

сравним в этой точке модули правой и левой частей уравнения (67а):

$$(2 + h^2 p_{n_0}) |z_{n_0}| \leq |z_{n_0-1}| + |z_{n_0+1}| + \frac{h^4}{12} |u^{IV}(\xi_{n_0})|.$$

Заменяя в правой части $|z_{n_0 \pm 1}|$ на $|z_{n_0}|$, мы только усилим неравенство и после сокращений получим оценку погрешности

$$\max |z_n| \leq \frac{h^2}{12} \left| \frac{u^{IV}(\xi_{n_0})}{p_{n_0}} \right| \leq \frac{h^2}{12} \max \left| \frac{u^{IV}(x)}{p(x)} \right|. \quad (68)$$

Утверждение доказано.

Сейчас была найдена мажорантная оценка погрешности. При некоторых дополнительных ограничениях можно получить асимптотическую оценку типа $z_n = \zeta(x_n) h^2 + o(h^2)$, где $\zeta(x)$ — некоторая функция (общая теорема о таких оценках будет доказана в главе IX). Из наличия асимптотической оценки следует возможность применения правила Рунге—Ромберга для апостериорной оценки точности или для уточнения решения при помощи расчетов на сгущающихся сетках.

Остановимся на устойчивости расчета. Если $p(x) > 0$, то задача Коши для уравнения (64а) плохо обусловлена, причем чем больше $p(x)$, тем хуже ее устойчивость. А из оценки (68) видно, что погрешность нашего разностного решения при большом $p(x)$ мала. Отсюда видно, что хорошо построенные разностные схемы нечувствительны к неустойчивости задачи Коши.

В обратном случае $p(x) < 0$ не выполняется достаточное условие устойчивости алгебраической прогонки (5.14). Однако в практике численных расчетов нарушение этого условия обычно не вызывает заметного ухудшения устойчивости. Только в редких случаях, когда определитель алгебраической системы (66) почти равен нулю, точность расчета резко падает из-за возрастания ошибок округления.

Чтобы легко опознать и исключить такую потерю устойчивости, можно провести расчет на трех (или более) сетках с различными шагами. Если при убывании h все разностные решения близки между собой и стремятся к некоторому пределу со скоростью $O(h^2)$ *, то это свидетельствует о хорошей устойчивости.

Пример. Возьмем частный случай задачи (64), соответствующий $p(x) < 0$:

$$u''(x) + u(x) = -x, \quad u(0) = u\left(\frac{\pi}{2}\right) = 0. \quad (69)$$

и воспользуемся разностной схемой (66). Если выбрать шаг сетки $h = \pi/4$, то алгебраическая система фактически будет со-

*) В общем случае — со скоростью, соответствующей порядку точности схемы.

стоять из одного уравнения, а при $h = \pi/8$ — из трех уравнений. Разностное решение y_n для этих случаев приведено в таблице 20. К этим двум решениям применено правило Рунге, и уточненное решение \tilde{y}_n тоже представлено в таблице; для сравнения дано точное решение $u(x) = \frac{\pi}{2} \sin x - x$. Из таблицы видно, что рассмотренная разностная схема дает неплохие результаты даже на сетке с большим шагом.

Таблица 20

x_n	y_n		\tilde{y}_n	$u(x)$
	$h = \pi/4$	$h = \pi/8$		
$\pi/8$	—	0,2122	0,2090	0,2084
$\pi/4$	0,3503	0,3311	0,3247	0,3253
$3\pi/8$	—	0,2778	0,2746	0,2731

Заметим, что, зная разностное решение в узлах сетки, можно интерполяцией получить приближенное решение при произвольных значениях x . Точность интерполяции целесообразно согласовывать с точностью разностного решения: например, для схемы (66) интерполировать многочленом первой степени, имеющим точность $O(h^2)$, а уточненное решение \tilde{y} интерполировать многочленом второй степени.

5. Разностный метод; нелинейные задачи. Выше была построена несложная разностная схема для простейшей задачи. Перейдем к более общим случаям.

Наибольшие трудности вызывают нелинейные задачи. Рассмотрим краевую задачу для нелинейного уравнения второго порядка

$$u''(x) = f(x, u), \quad u(a) = \alpha, \quad u(b) = \beta \quad (70)$$

с краевыми условиями первого рода. Будем предполагать, что $f(x, u)$ ограничена и непрерывна вместе со своими вторыми производными, так что существует ограниченная и непрерывная $u^{IV}(x)$. Обозначим через $M_1 = \max |f_u|$, $M_2 = \max |u^{IV}|$.

Аналогично п. 4, введем на $[a, b]$ равномерную сетку x_n и заменим вторую производную разностным выражением (65). Подставляя его в дифференциальное уравнение (70), получим систему нелинейных алгебраических уравнений

$$y_{n-1} - 2y_n + y_{n+1} = h^2 f(x_n, y_n), \quad 1 \leq n \leq N-1, \quad (71)$$

$$y_0 = \alpha, \quad y_N = \beta;$$

последние два уравнения аппроксимируют краевые условия.

Докажем сходимость разностного решения к точному, дополнительно предполагая, что $f_u \geq m_1 > 0$. Поскольку для погрешности аппроксимации производной (65) справедливо соотношение (3.12):

$$u_{n-1} - 2u_n + u_{n+1} = h^2 u''(x_n) + \frac{h^4}{12} u^{IV}(\xi_n), \quad \xi_n \in (x_{n-1}, x_{n+1}),$$

точное решение удовлетворяет разностным уравнениям

$$u_{n-1} - 2u_n + u_{n+1} = h^2 f(x_n, u_n) + \frac{h^4}{12} u^{IV}(\xi_n), \quad 1 \leq n \leq N-1, \\ u_0 = \alpha, \quad u_N = \beta.$$

Вычитая эти уравнения из (71), обозначая погрешность $z_n = y_n - u_n$ и учитывая, что $f(x_n, y_n) - f(x_n, u_n) = (f_u)_n z_n$, получим для погрешности систему уравнений

$$z_{n-1} - (2 + h^2 f_u)_n z_n + z_{n+1} = \frac{h^4}{12} u^{IV}(\xi_n), \quad 1 \leq n \leq N-1, \quad (72) \\ z_0 = 0, \quad z_N = 0.$$

Пусть x_{n_0} есть узел, в котором $|z_n|$ максимален. В этом узле перепишем соотношение (72) в форме неравенства

$$(2 + h^2 f_u)_{n_0} |z_{n_0}| \leq |z_{n_0-1}| + |z_{n_0+1}| + \frac{h^4}{12} |u^{IV}(\xi_{n_0})|.$$

Усилим это неравенство, заменяя в правой части $|z_{n_0 \pm 1}|$ на $|z_{n_0}|$; тогда получим

$$|z_{n_0}| = \|z_n\|_c \leq \frac{h^2}{12} \frac{|u^{IV}(\xi_{n_0})|}{(f_u)_{n_0}} \leq \frac{h^2 M_2}{12 m_1}. \quad (73)$$

Это означает, что при $h \rightarrow 0$ разностное решение равномерно сходится к точному со вторым порядком точности.

Займемся фактическим нахождением разностного решения. Алгебраические системы общего вида решают методами последовательных приближений или линеаризации. Однако, если взять метод последовательных приближений в естественной форме (5.44):

$$2y_n^{(s+1)} = y_{n-1}^{(s)} + y_{n+1}^{(s)} - h^2 f(x_n, y_n^{(s)}),$$

то нетрудно убедиться, что критерии сходимости этого метода (5.45) не выполняются. Положение улучшается, если придать методу последовательных приближений специфическую форму

$$y_{n-1}^{(s)} - 2y_n^{(s)} + y_{n+1}^{(s)} = h^2 f(x_n, y_n^{(s-1)}), \quad 1 \leq n \leq N-1, \quad (74) \\ y_0^{(s)} = \alpha, \quad y_N^{(s)} = \beta.$$

Тогда для определения $y_n^{(s)}$ на каждой итерации получается линейная система, решаемая алгебраической прогонкой. Исследуем сходимость итераций (74).

Рассмотрим погрешность итерации $\zeta_n^{(s)} = y_n^{(s)} - y_n$. Она удовлетворяет системе уравнений, получаемой вычитанием (71) из (74):

$$\begin{aligned} \zeta_{n-1}^{(s)} - 2\zeta_n^{(s)} + \zeta_{n+1}^{(s)} &= d_n^{(s)}, \quad 1 \leq n \leq N-1, \\ \zeta_0^{(s)} &= 0, \quad \zeta_N^{(s)} = 0, \end{aligned} \quad (75)$$

$$d_n^{(s)} = h^2 f(x_n, y_n^{(s-1)}) - h^2 f(x_n, y_n) \approx h^2 f_u(x_n, y_n) \zeta_n^{(s-1)}.$$

Решим эту трехдиагональную систему методом прогонки. Для данной системы рекуррентные соотношения (5.12) для коэффициентов прогонки нетрудно преобразовать к такому виду:

$$\xi_{n+1} = \frac{1}{2 - \xi_n} = \frac{n}{n+1}, \quad 0 \leq n \leq N-2, \quad \xi_N = 0,$$

$$\eta_{n+1} = \xi_{n+1} (\eta_n - d_n^{(s)}) = -\frac{1}{n+1} \sum_{k=1}^n k d_k^{(s)}, \quad 1 \leq n \leq N-1.$$

Формулы обратного хода прогонки (5.11) также преобразуются

$$\zeta_n^{(s)} = \xi_{n+1} \zeta_{n+1}^{(s)} + \eta_{n+1} = -n \sum_{k=n+1}^N \frac{1}{k(k-1)} \sum_{p=1}^{k-1} p d_p^{(s)} \quad (76)$$

и дают искомое решение системы (75).

Для правых частей системы (75) выполняется неравенство

$$|d_n^{(s)}| \leq q^{(s)}, \quad q^{(s)} = h^2 M_1 \| \zeta_n^{(s-1)} \|_c.$$

Подставляя его в (76), получим

$$| \zeta_n^{(s)} | \leq q^{(s)} n \sum_{k=n+1}^N \frac{1}{k(k-1)} \sum_{p=1}^{k-1} p = \frac{1}{2} q^{(s)} n (N-n) \leq \frac{1}{8} q^{(s)} N^2.$$

Отсюда следует

$$\| \zeta_n^{(s)} \|_c \leq \bar{q} \| \zeta_n^{(s-1)} \|_c, \quad \text{где} \quad \bar{q} = \frac{1}{8} N^2 h^2 M_1 = \frac{1}{8} (b-a)^2 M_1. \quad (77)$$

Это означает, что итерации (74) сходятся при выполнении условия

$$\frac{1}{8} (b-a)^2 M_1 < 1, \quad M_1 = \max \left| \frac{\partial f}{\partial u} \right|. \quad (78)$$

Из соотношения (78) следует, что сходимость линейная, т. е. довольно медленная.

Условие (78) является достаточным, но оно близко к необходимому: более сложные оценки показывают, что если $(b-a)^2 M_1 > > \pi^2$, то итерации (74) могут расходиться.

Целесообразнее решать уравнения (71) методом Ньютона. Соответствующие формулы нетрудно записать, линеаризуя правые части этих уравнений:

$$\begin{aligned} y_n^{(s+1)} &= y_n^{(s)} + \Delta_n^{(s)}, \quad 0 \leq n \leq N, \\ \Delta_{n-1}^{(s)} - (2 + h^2 f_u)_n \Delta_n^{(s)} + \Delta_{n+1}^{(s)} &= h^2 f_n^{(s)} - y_{n-1}^{(s)} + 2y_n^{(s)} - y_{n+1}^{(s)}, \\ 1 &\leq n \leq N-1, \\ \Delta_0^{(s)} &= \Delta_N^{(s)} = 0. \end{aligned} \quad (79)$$

Линеаризованную систему также решают алгебраической прогонкой. Сходимость итераций исследуют описанными выше приемами. Потребуем, чтобы $f_u \geq m_1 > 0$. Тогда сравнивая (79) и (71), можно получить для поправки $\zeta_n^{(s)} = y_n^{(s)} - y_n$ такое неравенство:

$$\|\zeta_n^{(s)}\|_c \leq \left\| \frac{f_{uu}(x, u)}{2f_u(x, u)} \right\|_{C(x)} \|\zeta_n^{(s-1)}\|_c^2.$$

Это означает, что если нулевое приближение взято не слишком далеко от корня (например, удовлетворяет условию

$$\|\zeta_n^{(0)}\|_c \leq \|2f_u/f_{uu}\|_C(x),$$

то итерации (79) сходятся, причем квадратично. Поэтому метод Ньютона обычно выгодней метода последовательных приближений, несмотря на более громоздкие формулы.

Замечание. Если итерации (79) или (74) сходятся, то в силу непрерывности и гладкости функции $f(x, u)$ они сходятся к решению системы (71). Тем самым устанавливается существование разностного решения в этих случаях.

Для нелинейных задач очень эффективна *комплексная организация расчета*, позволяющая при небольшом объеме вычислений получать высокую точность. Опишем ее.

Возьмем первую сетку с очень малым числом интервалов $N = 2 - 8$; остальные сетки получим из нее последовательным сгущением вдвое. На первой сетке начальное приближение выберем каким-либо приближенным способом: методом Галеркина, или разложением по малому параметру. Поскольку для первой сетки порядок алгебраической системы мал, качество нулевого приближения здесь малосущественно.

Когда итерации сошлись, полученное разностное решение интерполируем (например, линейно) на второй сетке и возьмем на ней в качестве нулевого приближения. Тогда итерации обычно быстро сходятся; в методе Ньютона достаточно 2—4 итераций. Интерполируем это решение на следующей сетке и т. д. Общий объем расчетов при этом невелик и примерно эквивалентен 5—8 итерациям последней сетки.

В заключение разностное решение на всех сетках уточним по рекуррентному правилу Рунге. Это настолько повышает точность, что даже в сложных задачах позволяет ограничиться небольшим числом интервалов последней сетки ($N = 32 - 128$).

Если проводится серия расчетов при варьировании параметров исходной задачи, то целесообразно результат расчета одного варианта брать в качестве нулевого приближения для первой сетки следующего варианта.

Рассмотрим некоторые другие усложнения задачи.

1) Сетка может быть неравномерной. В этом случае надо использовать соответствующую аппроксимацию производных, например,

$$u''(x_n) \approx \frac{2}{x_{n+1} - x_{n-1}} \left(\frac{u_{n+1} - u_n}{x_{n+1} - x_n} - \frac{u_n - u_{n-1}}{x_n - x_{n-1}} \right).$$

Напомним, что эта аппроксимация имеет погрешность $O(h^2)$ на квазиравномерных сетках и $O(h)$ на произвольных сетках. Исследование разностной схемы (71), проведенное выше, легко обобщается на случай неравномерной сетки.

2) Можно использовать аппроксимации, явно учитывающие вид общего решения исходного дифференциального уравнения; при этом получаются специальные схемы (см. § 1, п. 9). Составим, например, для задачи (64) с $p(x) < 0$ такую схему, чтобы она была точна при $p = \text{const}$, $f(x) = \text{const}$. При этом ограничении общее решение уравнения (64а) имеет вид

$$\tilde{u}(x) = -\frac{p}{f} + A \sin(\sqrt{|p|}x) + B \cos(\sqrt{|p|}x),$$

где A, B — произвольные постоянные. Легко проверить, что на равномерной сетке подстановка этого решения в разностную схему

$$y_{n-1} - 2y_n \cos(h\sqrt{-p_n}) + y_{n+1} = 2f_n [1 - \cos(h\sqrt{-p_n})], \quad 1 \leq n \leq N-1, \quad (80)$$

(краевые условия учитываются аналогично (71)) дает тождество. Следовательно, эта схема точна в указанном смысле. Она позволяет получать хорошую точность расчета быстро осциллирующих решений даже на грубой сетке, если $p(x)$ и $f(x)$ являются медленно меняющимися функциями.

Однако заметим, что применять правило Рунге для уточнения разностных решений, полученных по схемам типа (80), можно не всегда. Причина этого была подробно рассмотрена в связи с формулами Филона (глава IV, § 2, п. 3).

3) Дифференциальное уравнение может иметь более высокий порядок. Аппроксимация старших производных требует большего числа узлов, и каждое уравнение типа (71) или (6ба) будет содержать соответственно большее число неизвестных. Поэтому для решения алгебраической линейной (или линеаризованной) системы вместо алгебраической прогонки надо использовать несколько более трудоемкие способы. Но принципиальных осложнений это не вызывает.

4) Возможны более сложные краевые условия. Рассмотрим, например, нелинейное условие третьего рода

$$u'(a) = \varphi(u(a)). \quad (81)$$

Если подставить в него аппроксимацию $u'_0 \approx (u_1 - u_0)/h$, то ее погрешность $O(h)$ велика, что ухудшает общую точность расчета. Чтобы записать разностное краевое условие повышенной точности, рассмотрим формулу Тейлора

$$u(x_1) = u(x_0) + hu'(x_0) + \frac{1}{2}h^2u''(x_0) + \dots$$

и на основании уравнения (70) положим $u''(x_0) = f(x_0, u_0)$, а из краевого условия (81) возьмем $u'(x_0) = \varphi(u_0)$. Тогда получим

$$\frac{1}{h}(y_1 - y_0) = \varphi(y_0) + \frac{h}{2}f(x_0, y_0). \quad (82)$$

Другие способы аппроксимации краевых условий будут рассмотрены в главе IX.

Подведем итоги. Разностный метод имеет свои трудности, связанные в основном с решением алгебраической системы уравнений. Однако эти трудности успешно преодолеваются. Метод естественно переносится на уравнения высокого порядка, причем трудоемкость вычислений почти не возрастает. Его численная устойчивость обычно хорошая.

Поэтому для уравнений второго порядка разностный метод успешно конкурирует с методом стрельбы, а для уравнений более высокого порядка, особенно при сложной постановке краевых условий, оказывается выгоднее стрельбы.

6. Метод Галеркина. Краевая задача для уравнения $A(u(x)) = 0$ сводилась в главе VII, § 4 к отысканию минимума функционала типа (Au, Au) или (u, Au) . Затем решение $u(x)$ приближенно заменялось отрезком разложения по некоторой полной системе функций, а коэффициенты разложения находились из условия минимума функционала. Этот способ для функционалов первого типа называют методом наименьших квадратов, а для второго — методом Ритца.

Метод наименьших квадратов неудобен тем, что под интегралом возникают квадраты старших производных, входящих в оператор A , и вычисления становятся громоздкими. Метод Ритца имеет тот недостаток, что не для всякого оператора A удается найти эквивалентный функционал (обычно нужна самосопряженность оператора). Более удобен на практике метод Б. Г. Галеркина (или Бубнова — Галеркина), свободный от этих недостатков. Изложим этот метод.

Пусть дано уравнение с некоторыми краевыми условиями (для определенности — первого рода)

$$A(u(x)) = f(x), \quad a \leq x \leq b, \quad u(a) = \alpha, \quad u(b) = \beta. \quad (83)$$

Как и в методе Ритца (см. главу VII, § 4, п. 3), будем искать приближенное решение в виде суммы

$$u(x) \approx y_n(x) = \varphi_0(x) + \sum_{k=1}^n c_k \varphi_k(x), \quad (84)$$

где $\varphi_0(x)$ — некоторая непрерывная функция, удовлетворяющая неоднородным краевым условиям (83), а $\varphi_k(x)$, $1 \leq k < \infty$, — какая-то система линейно-независимых функций, полная в классе непрерывных функций, определенных на отрезке $[a, b]$ и обращающихся в нуль на его концах.

Докажем, что если для некоторой функции $F(x)$ и полной системы функций $\varphi_k(x)$ выполняется соотношение

$$\int_a^b F(x) \varphi_k(x) dx = 0 \quad \text{при} \quad 1 \leq k < \infty,$$

то $F(x) \equiv 0$ на $[a, b]$. Для этого из полной системы $\varphi_k(x)$ последовательной ортогонализацией построим полную ортогональную систему $\psi_k(x)$. Очевидно, тогда

$$\varphi_k(x) = \sum_{m=1}^k \zeta_{km} \psi_m(x),$$

причем $\zeta_{kk} \neq 0$, иначе $\varphi_k(x)$ были бы линейно-зависимы. Разлагая по новой системе

$$F(x) = \sum_{l=1}^{\infty} \gamma_l \psi_l(x),$$

придем к соотношению

$$0 = \int_a^b F(x) \varphi_k(x) dx = \sum_{m=1}^k \gamma_m \zeta_{km} = 0, \quad k = 1, 2, \dots$$

Полагая $k=1$, получим $\gamma_1=0$. Полагая $k=2$, получим $\gamma_2=0$ и т. д. Следовательно, все $\gamma_l=0$ и $F(x) \equiv 0$. Отметим, что если исходная система $\varphi_k(x)$ уже ортогональна, то доказательство становится тривиальным.

Таким образом, если бы мы нашли такую функцию $u(x)$, чтобы $A(u(x)) - f(x)$ было ортогонально $\varphi_k(x)$ при любых $k \geq 1$, то это означало бы, что $A(u(x)) = f(x)$ и задача (83) была бы решена*). Если же ортогональность есть только при $k \leq n$, то в разложе-

*) Ортогональности $A(u) - f$ к $\varphi_0(x)$ не требуется, ибо $\varphi_0(x)$ не входит в полную систему функций $\varphi_k(x)$.

ние $A(u) - f$ по $\phi_k(x)$ входят γ_{n+1} и более старшие коэффициенты, т. е. $A(u) \approx f$.

Возьмем вместо $u(x)$ приближенное решение в форме (84) и потребуем, чтобы

$$\int_a^b [A(y_n(x)) - f(x)] \phi_k(x) dx = 0, \quad 1 \leq k \leq n. \quad (85)$$

Это дает нам алгебраическую систему для определения коэффициентов c_k . Найдя из нее коэффициенты, получим приближенное решение (84). В этом и заключается метод Галеркина. Вопрос об условиях сходимости $y_n(x)$ при $n \rightarrow \infty$ к точному решению и о скорости сходимости здесь не рассматривается.

Если оператор $A(u)$ нелинейный, то система (85) тоже будет нелинейной. При этом больше чем 3—4 коэффициента трудно найти. Если же оператор линейный, то алгебраическая система (85) линейна и можно решать задачу с большим числом коэффициентов. Отметим, что для линейных уравнений второго порядка метод Галеркина приводит точно к тем же уравнениям, что и метод Рунца.

Пример. Рассмотрим задачу (69). Положим $\phi_0(x) = 0$ и выберем полную систему функций $\phi_k(x) = x^k ((\pi/2) - x)$, $1 \leq k < \infty$. Тогда, если ограничиться одним членом суммы (84), то легко получить, что

$$c_1 = 5\pi/(40 - \pi^2) \approx 0,521, \quad n = 1.$$

Если возьмем два члена суммы, то получим

$$c_1 \approx 0,815, \quad c_2 \approx 0,377, \quad n = 2.$$

Соответствующие приближенные решения, вычисленные в нескольких точках отрезка, приведены в таблице 21; для сравнения там же дано точное решение.

Т а б л и ц а 21

x	$y_1(x)$	$y_2(x)$	$u(x)$
$\pi/8$	0,241	0,445	0,208
$\pi/4$	0,322	0,685	0,325
$3\pi/8$	0,241	0,582	0,273

Хотя в этой задаче решение является плавно меняющейся функцией, метод Галеркина при небольшом числе членов дает неважные результаты.

Обратим внимание на то, что при увеличении n не только добавляются новые коэффициенты, но и меняются старые, что не очень удобно. Легко заметить, что если задача линейная, а система $\phi_k(x)$ ортогональная, то уже найденные c_k не будут меняться при увеличении n . Поэтому ортогональные системы обычно удобнее неортогональных.

Метод Галеркина для нелинейных задач используют лишь для нахождения грубого приближения; для линейных задач им можно

найти решение с хорошей точностью. Результат очень чувствителен к тому, насколько удачно выбрана система функций $\varphi_k(x)$ для данной задачи.

Отметим также, что при нелинейном краевом условии вида, например, $u'(a) = g(u(a))$ линейная комбинация (84) с произвольными коэффициентами c_k уже не будет удовлетворять этому краевому условию. Поэтому метод Галеркина применим только к задачам с линейными (относительно $u(x)$ и ее производных) краевыми условиями, хотя допустим и нелинейный оператор $A(u)$.

7. Разрывные коэффициенты. Во всех предыдущих пунктах явно или неявно предполагалось, что правые части рассматриваемых дифференциальных уравнений непрерывны вместе с некоторым числом своих производных. Однако в задачах о слоистых средах коэффициенты уравнений (коэффициентами являются различные свойства вещества — плотность, теплопроводность, упругость и т. д.) обычно разрывны на границах раздела двух сред, т. е. во внутренних точках $[a, b]$.

Бегло рассмотрим, как переносятся на этот случай развитые выше методы. Сделаем это на примере уравнения

$$\frac{d}{dx} \left[k(x) \frac{du}{dx} \right] - q(x) u(x) = f(x). \quad (86a)$$

Сначала обсудим характер решения. Если $q(x)$ или $f(x)$ кусочно-непрерывны, то $u''(x)$ также лишь кусочно-непрерывна. Очевидно, в точке разрыва аппроксимировать вторую производную разностным соотношением нельзя.

Еще сложнее случай разрыва $k(x)$ в некоторой точке \bar{x} . При этом решение краевой задачи становится, вообще говоря, не единственным. Существует множество обобщенных решений, каждое из которых удовлетворяет своему условию согласования в точке \bar{x} . Для выделения единственного решения требуется поставить в этой точке внутреннее краевое условие; оно выбирается из физических соображений и должно входить в полную постановку задачи.

Пусть, например, (86a) есть уравнение теплопроводности в стержне, составленном из разных материалов, а $k(x)$ — коэффициент теплопроводности. Тогда дополнительным условием будет непрерывность температуры и теплового потока $W = -ku_x$ в точке соединения

$$[u(x)]_{\bar{x}-0}^{\bar{x}+0} = 0, \quad \left[k(x) \frac{du}{dx} \right]_{\bar{x}-0}^{\bar{x}+0} = 0. \quad (86b)$$

Поэтому и в методе стрельбы, и в разностном методе все точки разрыва коэффициентов выбирают в качестве узлов сетки; такие сетки называют *специальными*.

В методе стрельбы до прихода (для определенности слева) в такую точку пользуются «левыми» значениями коэффициентов.

Придя в эту точку, при помощи внутреннего краевого условия формируют новые начальные условия. Например, в задаче (86) это будут условия

$$u(\bar{x} + 0) = u(\bar{x} - 0), \quad u_x(\bar{x} + 0) = \frac{k(\bar{x} - 0)}{k(\bar{x} + 0)} u_x(\bar{x} - 0).$$

Затем продолжают численное интегрирование, пользуясь уже «правыми» значениями коэффициентов.

В разностном методе для точки разрыва вместо аппроксимации дифференциального уравнения (86а) можно записать аппроксимацию внутреннего краевого условия (86б), или можно составить такую разностную схему, которая применима во всех точках, включая точку \bar{x}^* .

В методе Галеркина систему функций $\varphi_k(x)$ следует выбирать так, чтобы линейная комбинация (84) при любых значениях коэффициента c_k удовлетворяла внутреннему краевому условию.

§ 3. Задачи на собственные значения

1. Постановки задач. Задачи на собственные значения — это краевые задачи для системы p уравнений первого порядка

$$\begin{aligned} u'(x) &= f(x, u; \lambda_1, \lambda_2, \dots, \lambda_q), \\ [u &= \{u_1, u_2, \dots, u_p\}, \quad f = \{f_1, f_2, \dots, f_p\}, \end{aligned}$$

в которых правые части зависят от параметров λ_r , значения которых неизвестны и должны быть определены из самой задачи; число дополнительных (краевых) условий соответственно равно $p + q$. Функции $u_k(x)$, $1 \leq k \leq p$, и значения параметров λ_r , $1 \leq r \leq q$, удовлетворяющие всем уравнениям и краевым условиям, называются *собственными функциями* и *собственными значениями* задачи.

Задачи на собственные значения часто встречаются в физике и технике **). Например, определение собственных колебаний струны приводит к задаче для линейного уравнения второго порядка с одним параметром

$$\frac{d}{dx} \left[k(x) \frac{du}{dx} \right] + \lambda \rho(x) u(x) = 0,$$

а собственных колебаний упругого стержня — к линейному уравнению четвертого порядка (краевые условия зависят от способа закрепления струны или стержня). Дифференциальное уравнение

*) Это так называемые *консервативные* схемы, способ построения которых будет изложен в следующих главах.

***) Много примеров таких задач приведено в [17].

второго порядка возникает при нахождении спектра атома водорода. Нахождение уровней энергии многоэлектронного атома в приближении Хартри—Фока приводит к задаче для системы нелинейных уравнений, в которой число функций и число параметров равно числу электронов атома.

Исследование корректности постановки задачи на собственные значения еще более сложно, чем для краевых задач. Исследованы в основном линейные задачи с одним параметром. Однако в курсах теории колебаний и квантовой механики имеется немало примеров, из которых видно, что в зависимости от постановки задачи собственные значения могут существовать или не существовать, быть вещественными или комплексными; спектр собственных значений может быть дискретным, сплошным, состоящим из полос или являющимся комбинацией перечисленных случаев.

Наиболее употребительными численными методами решения задач на собственные значения являются метод стрельбы и разностный метод, подробно рассмотренные ниже. Из приближенных методов упомянем методы Ритца и Галеркина.

2. Метод стрельбы. В задачах на собственные значения имеются естественные пристрелочные параметры — величины λ ; поэтому такие задачи нередко решают методом стрельбы. Основные черты этого метода те же, что и для краевых задач; рассмотрим детали метода на двух примерах.

Простейший пример — задача для одного уравнения первого порядка с одним параметром и двумя краевыми условиями

$$u'(x) = f(x, u; \lambda), \quad u(a) = \alpha, \quad u(b) = \beta. \quad (87)$$

Если отбросить правое краевое условие и выбрать некоторое значение λ , то (87) превратится в задачу Коши. Численно интегрируя ее, получим решение $u(x; \lambda)$, удовлетворяющее левому краевому условию и зависящее от параметра λ . Вообще говоря, $u(b; \lambda) \neq \beta$, т. е. это решение не удовлетворяет правому краевому условию. Тогда будем варьировать λ до тех пор, пока не получим $u(b; \lambda) \approx \beta$ с требуемой точностью. Разумеется, при варьировании используют обычные методы нахождения корня алгебраического уравнения, как это было сделано в § 2, п. 2.

Другой пример — это классическая задача на собственные значения уравнения второго порядка при нулевых краевых условиях

$$u''(x) + p(x)u'(x) + [\lambda + q(x)]u(x) = 0, \quad u(a) = u(b) = 0. \quad (88)$$

Уравнение имеет второй порядок и содержит одно собственное значение; следовательно, задача требует трех дополнительных условий. Но в силу линейности и однородности решение определено с точностью до множителя; это и есть неявное задание третьего условия. Формально третье условие здесь удобно задать в форме $u'(a) = 1$ (что возможно, если $p(a)$ и $q(a)$ конечны).

Тогда можно взять для исходного уравнения задачу Коши с начальными условиями $u(a) = 0$, $u'(a) = 1$ и вести пристрелку параметра λ до выполнения правого краевого условия.

Заметим, что линейность уравнения и краевых условий не упрощает стрельбу, ибо зависимость $u(x; \lambda)$ от параметра все равно остается нелинейной.

Метод стрельбы удобно применять, если стрельба является однопараметрической, как это было в рассмотренных примерах. Если это требование не выполнено, то алгоритмы стрельбы сильно усложняются и становятся менее надежными; тогда выгодней использовать разностный метод.

Метод стрельбы трудно применять также в том случае, если задача Коши плохо обусловлена. Тогда малая вариация λ может резко изменить решение $u(x)$ и даже вывести его за пределы представимых на ЭВМ чисел. При этом невозможно организовать процесс решения алгебраического уравнения типа

$$u(b; \lambda) = 0.$$

Иногда, как и в краевых задачах, помогает смена направления интегрирования (но ее применяют только, если от этого не увеличивается число параметров пристрелки).

3. Фазовый метод. Классическая задача для уравнения второго порядка (88) имеет много важных физических приложений. В частности, к этому уравнению приводит квантовомеханическая задача об уровнях энергии частицы, движущейся в заданном одномерном (например, сферически-симметричном) поле. В последнем случае задача Коши для уравнения (88) оказывается очень плохо обусловленной: общее решение уравнения обращается в бесконечность на обоих концах отрезка ($x=0$ и $x=\infty$). Поэтому применять метод стрельбы трудно. Но эта задача настолько важна, что для нее разработаны специальные схемы. Рассмотрим одну из них — *фазовый метод*.

Воспользуемся тем, что качественное поведение решения известно. Решение имеет осциллирующий характер, причем амплитуда может сильно зависеть от координаты. Введем амплитуду ρ и фазу φ решения при помощи соотношения

$$u(x) = \rho(x) \sin \varphi(x). \quad (89a)$$

Это соотношение неоднозначно определяет амплитуду и фазу. Для определенности подчиним их дополнительному соотношению

$$u'(x) = \rho(x) \cos \varphi(x). \quad (89b)$$

Наглядный смысл его состоит в том, что если взять вектор с координатами u , u' , т. е. перейти в фазовую плоскость, то ρ и φ будут амплитудой и фазой этого вектора.

Дифференцируя (89а) и (89б) и сравнивая их между собой, получим соотношения

$$\begin{aligned} u'' &= \rho' \cos \varphi - \varphi' \rho \sin \varphi, \\ \rho' \sin \varphi &= (1 - \varphi') \rho \cos \varphi. \end{aligned}$$

Исключая при помощи этих соотношений и формул (89) функцию $u(x)$ и ее производные из уравнения (88), после несложных преобразований расцепим (88) на уравнения для амплитуды и фазы:

$$\rho'(x) = -\rho(x) \{p(x) \cos \varphi(x) + [q(x) + \lambda - 1] \sin \varphi(x)\} \cos \varphi(x), \quad (90)$$

$$\varphi'(x) = \cos^2 \varphi(x) + p(x) \sin \varphi(x) \cos \varphi(x) + [\lambda + q(x)] \sin^2 \varphi(x). \quad (91)$$

Граничные условия (88) при этом естественно приписываются фазе. Если надо найти решение, соответствующее квантовому числу n , т. е. имеющее n полувольт на $[a, b]$, то следует положить

$$\varphi(a) = 0, \quad \varphi(b) = n\pi. \quad (92)$$

Таким образом, мы получили задачу на собственные значения (91)—(92) только для уравнения фазы. Она легко решается методом стрельбы, поскольку задача Коши для уравнения (91) хорошо обусловлена. Важной особенностью этой задачи является то, что правому краевому условию (92) удовлетворяет только одно определенное λ_n из всего спектра исходной задачи (88). Поэтому стрельба всегда сходится именно к требуемому собственному значению.

После нахождения фазы уравнение для амплитуды легко интегрируется в квадратурах

$$\rho(x) = \rho(a) \exp \left\{ - \int_a^x [p(\xi) \cos \varphi(\xi) + (q(\xi) + \lambda - 1) \sin \varphi(\xi)] \cos \varphi(\xi) d\xi \right\}.$$

Амплитуда определена с точностью до множителя и не меняет знака, как и должно быть по смыслу задачи.

Замечание 1. Задача (88) может иметь и другие типы краевых условий. Если исходное краевое условие имеет вид $u'(b) = 0$, то для фазы надо взять условие $\varphi(b) = \pi(n - 1/2)$. Несколько сложнее асимптотическое условие $u(\infty) = 0$, возникающее в задаче на отрезке $a \leq x < \infty$; обычно в таких задачах выполняется $p(\infty) = q(\infty) = 0$. Тогда нетрудно построить асимптотику решения $u(x) \sim \exp(-\sqrt{-\lambda}x)$ при $x \rightarrow \infty$ и получить отсюда асимптотическое краевое условие для фазы

$$\cos \varphi(x) + \sqrt{-\lambda} \sin \varphi(x) \rightarrow 0 \quad \text{при} \quad x \rightarrow +\infty.$$

Замечание 2. Фаза $\varphi(x)$ может быть немонотонной функцией. Однако, если при некотором \bar{x} значение фазы $\varphi(\bar{x}) = \pi k$, то $\varphi'(\bar{x}) = 1$; поэтому каждую линию $\varphi = \pi k$ интегральная кривая пересекает лишь однажды, а немонотонность может проявляться только между этими линиями. При таком поведении интегральных кривых стрельба с использованием дихотомии надежно сходится к собственному значению, а при использовании метода Ньютона область сходимости нередко оказывается очень узкой.

Замечание 3. Для преодоления последнего недостатка предложена замена функций, несколько более сложная, чем (89), но зато делающая $\varphi(x)$ монотонной функцией. При этом стрельба с использованием метода Ньютона сходится за небольшое число итераций.

4. Разностный метод обычно используется в тех случаях, когда стрельба оказывается многопараметрической, или если задача Коши для исходного дифференциального уравнения плохо обусловлена.

Формулируется он так же, как для краевых задач. Введем на $[a, b]$ сетку $\{x_n, 0 \leq n \leq N\}$ и заменим в исходной задаче все производные некоторыми разностными соотношениями. Тогда вместо дифференциального уравнения и краевых условий получим систему алгебраических уравнений

$$F_k(x_0, x_1, \dots, x_N, y_0, y_1, \dots, y_N; \lambda) = 0, \quad 0 \leq k \leq N+1 \quad (93)$$

(для простоты записи мы ограничиваемся случаем одного собственного значения). Эта система содержит $N+2$ уравнения, и из нее надо определить такое же число неизвестных: $\lambda, y_0, y_1, \dots, y_N$.

Возникают те же вопросы, что и в краевых задачах. Имеет ли алгебраическая система (93) решение? Если имеет, то как его фактически вычислить? Если разностное решение найдено, то насколько оно близко к точному решению? Сейчас мы рассмотрим линейные задачи, для которых на эти вопросы ответить легче.

Пусть исходная задача является линейной и однородной относительно $u(x)$, как, например, задача (88). Воспользуемся линейными разностными аппроксимациями производных. Тогда система (93) будет относительно y_n линейной однородной, т. е. это будет алгебраическая задача на собственные значения матрицы. Так, для задачи (88) при простейших аппроксимациях на равномерной сетке получим систему

$$\left(1 - \frac{1}{2} h p_n\right) y_{n-1} - (2 - h^2 q_n - \lambda h^2) y_n + \left(1 + \frac{1}{2} h p_n\right) y_{n+1} = 0, \\ 1 \leq n \leq N-1, \quad (94)$$

где $y_0 = y_N = 0$ в силу краевых условий. Эта система содержит $N-1$ уравнение; из нее надо определить $\lambda, y_1, y_2, \dots, y_{N-1}$.

Задача (94) имеет спектр собственных значений, состоящий из $N - 1$ числа (по порядку матрицы). Первые собственные значения являются приближениями к первым собственным значениям λ_m из дискретного спектра исходной задачи (88). Если разностная схема составлена так, что матрица алгебраической системы (93) является эрмитовой, то приближенные собственные значения будут вещественными.

Собственные значения и собственные векторы линейной системы (93) вычисляются методами, описанными в главе VI. Поскольку во многих приложениях матрица системы трехдиагональная (реже — пятидиагональная), а нужны только несколько первых собственных значений, то выгодно применять метод Дервюдьё (см. главу VI, § 4, п. 2). При небольшом числе интервалов сетки удобно также находить корни характеристического многочлена методом парабол, вычисляя сам многочлен по рекуррентным соотношениям (см. главу VI, § 1, п. 4).

Сходимость разностного решения к точному при $h \rightarrow 0$ хорошо исследована только для задач Штурма — Лиувилля *)

$$\frac{d}{dx} \left[k(x) \frac{du}{dx} \right] + [\lambda r(x) - q(x)] u(x) = 0, \quad u(a) = u(b) = 0.$$

Оказывается, что простейшая схема (94) дает не очень хорошие, а при разрывных коэффициентах — даже неверные результаты. Следует составлять консервативные разностные схемы (они будут подробно рассмотрены в главах X и XI). Если коэффициенты уравнения непрерывны вместе со своими вторыми производными, то простейшие консервативные схемы обеспечивают равномерную сходимость y_n к $u(x)$ с погрешностью $O(h^2)$. Так называемая *наилучшая* консервативная схема обеспечивает погрешность $O(h^2)$ даже при коэффициентах, кусочно-непрерывных со своими вторыми производными, если выбраны *специальные* разностные сетки (в которых эти точки разрыва являются узлами).

Пример. Рассмотрим частный случай задачи Штурма — Лиувилля

$$u''(x) + \lambda u(x) = 0, \quad u(0) = u(1) = 0. \quad (95)$$

Точное решение этой задачи есть $\lambda_m = \pi^2 m^2$, $u_m(x) = \sin \pi m x$, $m = 1, 2, \dots$; оно нужно для сравнения с численными расчетами. Простейшей разностной схемой для этой задачи является схема (94), в которой надо положить $p_n = q_n = 0$. Эта схема имеет второй порядок точности.

Выполняя расчеты для сеток с числом интервалов $N = 2, 3, 4$, приближенно определим три первых собственных значения.

*) Это исследование и доказательства приведенных ниже утверждений см. в [30].

Они представлены в таблице 22 вместе с точными значениями λ_m . Из таблицы видно, что с малой погрешностью определяются только те собственные значения, номер которых заметно меньше N . При сгущении сетки приближенные значения быстро стремятся к точным. Очень эффективным оказывается уточнение по правилу Рунге — Ромберга, также приведенное в таблице; уточнение λ_2 по двум сеткам дает неплохую точность, а уточнение λ_1 по трем сеткам — отличную.

Таблица 22

N	2	3	4	Уточненное по Рунге	Точное
λ_1	8,00	9,00	9,37	9,88	9,87
λ_2	—	27,0	32,0	38,4	39,5
λ_3	—	—	54,6	—	88,8

На этом примере хорошо видно, что сочетание схемы невысокого (обычно второго) порядка точности с правилом Рунге выгодно: оно обеспечивает высокую точность расчета при несложном алгоритме. Схемы высокого порядка точности обычно довольно громоздки, и организация расчета по ним сложнее.

5. Метод дополненного вектора. Для разностного метода, особенно в случае сложных нелинейных задач, важным и трудным является вопрос о фактическом вычислении разностного решения, ибо алгебраическая система (93) имеет заведомо высокий порядок. Для многих задач удобно находить это решение методом *дополненного вектора*. Изложим этот метод.

Заметим сначала, что метод стрельбы (и многие конкретные разностные алгоритмы) можно схематически описать следующим образом. Выбирается некоторое приближение $\lambda^{(0)}$; затем вычисляется соответствующее ему приближение $y^{(0)}(x)$. По этой функции находится новое приближение $\lambda^{(1)}$ и т. д. При этом собственное значение и собственная функция считаются элементами разных метрических пространств.

Будем рассуждать иначе. Разностную собственную функцию $y = \{y_0, y_1, \dots, y_N\}$ можно считать вектором в $(N+1)$ -мерном пространстве. Увеличивая размерность пространства на единицу, рассмотрим собственное значение как новую компоненту этого вектора, $y_{N+1} \equiv \lambda$. Новый вектор $Y = \{y_0, y_1, \dots, y_N, y_{N+1}\}$ назовем *дополненным*. Относительно компонент дополненного вектора алгебраическая система (93) переписывается в каноническом виде

$$F_k(y_0, y_1, \dots, y_N, y_{N+1}) = 0, \quad 0 \leq k \leq N+1. \quad (96)$$

Эта система нелинейна, даже если исходная задача была линейной относительно $u(x)$, как в примере (88).

Решать систему (96) будем методом Ньютона. Линеаризуя (96), получим на каждой итерации систему уравнений

$$\sum_{p=0}^{N+1} \frac{\partial F_k(Y^{(s)})}{\partial y_p} \delta y_p^{(s)} = -F_k(Y^{(s)}), \quad 0 \leq k \leq N+1, \quad (97)$$

линейную относительно приращений неизвестных $\delta y_p^{(s)} = y_p^{(s+1)} - y_p^{(s)}$. Если искомое решение алгебраической системы (96) не особенное, т. е. в нем $\det(\partial F/\partial Y) \neq 0$, то при не слишком плохом нулевом приближении итерации (97) быстро сходятся к разностному решению. Отметим, что для линейных задач на собственные значения этот итерационный процесс совпадает с методом Дервюдьё (см. главу VI, § 4, п. 2).

Удовлетворительное нулевое приближение для итераций (97) можно найти приближенными методами (метод Галеркина, разложение по малому параметру и т. д.), а в прикладных задачах его нередко удается получить из качественных соображений. Исключительно эффективна в таких задачах *комплексная организация расчета*, подробно описанная в § 2, п. 5.

З а м е ч а н и е 1. Метод дополненного вектора особенно полезен для уравнений, у которых задача Коши плохо обусловлена: он подавляет такую неустойчивость.

З а м е ч а н и е 2. Метод легко переносится на более общие задачи вида $A(u(x), \lambda) = 0$, где оператор A может быть интегро-дифференциальным (краевые условия предполагаются включенными в определение оператора). Вводя сетку x_n и аппроксимируя разностными выражениями все производные и интегралы, входящие в оператор, получим алгебраическую систему (96) и решим ее итерационным процессом (97)*).

З а м е ч а н и е 3. Недостатком метода является то, что при неудачном выборе нулевого приближения итерации (97) могут не сойтись, или в задачах со спектром собственных значений итерации могут сойтись не к искомому собственному значению.

З а м е ч а н и е 4. В методе дополненного вектора требуется решать систему линейных уравнений (97). Это легко делать, только если матрица системы целиком помещается в оперативной памяти ЭВМ (например, на БЭСМ-6 это будет при числе неизвестных $N \lesssim 150$). Это приводит к ограничению допустимого числа интервалов сетки.

Если требуется решить задачу для системы большого числа дифференциальных уравнений (например, уравнения Хартри—Фока для многоэлектронного атома), то даже при довольно грубой сетке число узловых значений

*) Здесь обсуждается только вопрос о вычислении разностного решения. Вопрос о его сходимости к точному решению при $h \rightarrow 0$ надо рассматривать отдельно; он связан со свойствами оператора A и выбором аппроксимации.

всех функций будет велико, и метод дополненного вектора применять трудно. В подобных задачах успешно применяется так называемый *непрерывный аналог метода Ньютона*, имеющий линейную сходимость итераций, но зато позволяющий оперировать с очень большим числом неизвестных. Этот метод является специальным вариантом метода последовательных приближений, организованным так, что итерации всегда сходятся.

6. Метод Галеркина. Многие приближенные методы пригодны для нахождения собственных значений и собственных функций задач, у которых краевые условия линейны относительно функции и ее производных. Среди этих методов к наиболее простым вычислениям приводит метод Галеркина.

Метод формулируется почти так же, как для краевых задач. Ищем решение задачи $A(u(x), \lambda) = f(x)$ в виде линейной комбинации отрезка полной системы функций $\varphi_k(x)$, $k \geq 1$:

$$u(x) \approx y_n(x) = \varphi_0(x) + \sum_{k=1}^n c_k \varphi_k(x), \quad a \leq x \leq b, \quad (98)$$

выбранной так, чтобы удовлетворялись краевые условия. Потребуем, чтобы выполнялись условия ортогональности

$$\int_a^b [A(y_n(x), \lambda) - f(x)] \varphi_k(x) dx = 0, \quad 1 \leq k \leq n. \quad (99)$$

Эти условия образуют алгебраическую систему n уравнений с $n+1$ неизвестным $c_1, c_2, \dots, c_n, \lambda$. Недостающее уравнение надо получить из одного из краевых условий.

По тем же соображениям, что и в краевых задачах, удобнее пользоваться ортогональными системами функций $\varphi_k(x)$. В линейных задачах вычисления при этом заметно упрощаются.

Пример. Рассмотрим задачу (95)

$$u''(x) + \lambda u(x) = 0, \quad u(0) = u(1) = 0$$

и воспользуемся полной системой многочленов $\varphi_k(x) = x^k(1-x)$, которые заметно отличаются от точного решения задачи. Одним из дополнительных условий является условие нормировки решения. Благодаря линейности задачи его можно формулировать разными способами; для удобства вычислений зададим его в форме $y_n(0) = 1$, что означает $c_1 = 1$. Тогда, полагая $n = 1$ и 2 , легко получим первые приближения

$$\begin{aligned} n = 1, & \quad \lambda^I = 10, & y^I(x) &= x(1-x), \\ n = 2, & \quad \lambda^I = 10, & y^I &= x(1-x), \\ & \quad \lambda^{II} = 21, & y^{II} &= x(1-x) - \frac{11}{13}x^2(1-x). \end{aligned}$$

Первое собственное значение определилось с хорошей точностью, второе — с много худшей.

Методом Галеркина можно довольно хорошо находить наименьшие собственные значения. Но точность определения собственных функций обычно заметно хуже.

Обоснование метода Галеркина сложно. В частном случае, если дифференциальный оператор A линеен и однороден относительно $u(x)$, система (99) является задачей на определение собственных значений матрицы. Для задачи Штурма—Лиувилля метод Галеркина приводит к тем же самым алгебраическим уравнениям, что и метод Ритца (сходимость которого в задачах Штурма—Лиувилля доказана).

ЗАДАЧИ

1. Доказать теорему о сходимости метода Пикара, сформулированную в § 1, п. 3.
2. Вывести оценку (10) скорости сходимости метода Пикара.
3. В методе малого параметра вывести формулы для коэффициентов α_n и функций $v_n(x)$ в уравнениях (12).
4. Найти приближенное решение уравнения (3) методом малого параметра.
5. Для системы двух уравнений (25) написать схемы Рунге—Кутты второго порядка точности, аналогичные (22) и (23).
6. Для уравнения химического распада (34) составить схемы Рунге—Кутты второго и четвертого порядка точности и выяснить ограничения на шаг в этих схемах, следующие из положительности решения.
7. Составить для уравнения химического распада (34) специальную схему интегрирования по третьему способу из § 1, п. 8.
8. Вычисляя в (41) интеграл от второго слагаемого по формуле трапеций, получить неявную специальную схему; исследовать ее точность и найти ограничение на шаг сетки.
9. Написать формулы метода стрельбы применительно к краевой задаче (46) для одного дифференциального уравнения второго порядка.
10. Составить формулы метода Ньютона для нахождения корня уравнения (62б), возникающего при решении краевой задачи (60) методом стрельбы.
11. Решить краевую задачу (69) методом Галеркина, выбрав ортогональную систему функций $\varphi_k(x) = \sin 2kx$; сравнить результат с примером, приведенным в § 2, п. 6.
12. Для итерационного процесса при решении задачи на собственные значения (87) баллистическим методом составить а) формулы метода секущих, б) формулы метода Ньютона.
13. Для задачи на собственные значения (95) найти разностным методом при $N=2$ и $N=4$ первую собственную функцию и уточнить ее по правилу Рунге; ответ сравнить с точным решением.