В.Б. Андреев

ЧИСЛЕННЫЕ МЕТОДЫ

Часть II

Глава IV

Методы решения задачи Коши для обыкновенных дифференциальных уравнений

§ 15

Постановка задачи и первые примеры

15.1 Введение. Задача Коши

Рассмотрим задачу Коши для обыкновенного дифференциального уравнения первого порядка

$$\frac{d\,u}{d\,t} = f(t,u), \quad t > 0, \qquad u(0) = u_0. \tag{15.1}$$

Из курса дифференциальных уравнений известно, что для однозначной разрешимости задачи (15.1) в некоторой окрестности точки $t = 0$ достаточно, чтобы функция $f(t,u)$ была непрерывна в окрестности точки $(0,u_0)$ и удовлетворяла условию Липшица по второму аргументу. Известны примеры, иллюстрирующие отсутствие решения задачи (15.1) или его неединственность при нарушении указанных условий. Мы всегда будем предполагать, что решение задачи (15.1) существует и единственно. Для дальнейшего нам даже придется предполагать, что искомое решение достаточно гладкое.

15.2 Примеры численных методов

Приведем несколько простейших численных методов решения задачи (15.1). Для этого введем на полуоси $t \geqslant 0$ равномерную сетку, т.е. множество точек (которые назовем узлами)

$$\omega = \{t_n = n\tau, \quad n = 0,1,\dots; \quad \tau > 0\}$$

и будем искать приближенное решение задачи (15.1) в узлах ω. Величину τ будем называть шагом сетки ω. Договоримся приближенное решение в узле t_n обозначать той же буквой, что и решение задачи (15.1), но с индексом n внизу: u_n. Тем самым, мы отказываемся от часто используемого обозначения $u(t_n) = u_n$; теперь $u(t_n)$ — значение точного решения в узле t_n, а u_n — значение приближенного решения в этом узле, и, вообще говоря, $u(t_n) \neq u_n$. Наоборот, $u_n - u(t_n)$ представляет собой погрешность численного метода в узле t_n, которую нам предстоит оценивать. Данное соглашение не представляется наилучшим, однако остановимся на нем.

Для построения численных методов проинтегрируем уравнение (15.1) от t_n до t_{n+1}

$$u(t_{n+1}) - u(t_n) = \int_{t_n}^{t_{n+1}} f(t, u(t))\, dt \tag{15.2}$$

и заменим приближенно интеграл в правой части этой формулы какой-либо квадратурной формулой. Здесь мы рассмотрим четыре таких формулы.

Построенная в курсе "Введение в численные методы"квадратурная формула прямоугольников представляет интеграл произведением длины отрезка интегрирования и значения подынтегральной функции в середине этого отрезка

$$\int_a^b \varphi(x)\, dx \approx |b-a| \varphi\left(\frac{a+b}{2}\right). \tag{15.3}$$

Эта квадратурная формула точна на многочленах первой степени, и при малых $|b-a|$ ее погрешность есть $O(|b-a|^3)$.

Наряду с этой квадратурной формулой, которую мы впредь будем называть формулой *центральных прямоугольников*, можно ввести так называемые формулы *левых* и *правых прямоугольников*. Первая из них состоит в представлении интеграла произведением длины отрезка интегрирования и значения подынтегральной функции в левом конце отрезка

$$\int_a^b \varphi(x)\, dx \approx |b-a| \varphi(a), \tag{15.4}$$

а вторая — произведением длины отрезка интегрирования и значения подынтегральной функции в правом конце отрезка

$$\int_a^b \varphi(x)\, dx \approx |b-a| \varphi(b). \tag{15.5}$$

Обе эти формулы точны только на многочленах нулевой степени и имеют погрешность $O(|b-a|^2)$.

а) **Метод Эйлера**. Заменим интеграл в (15.2) формулой левых прямоугольников (15.4). В результате получим приближенное равенство

$$u(t_{n+1}) - u(t_n) \approx \tau f(t_n, u(t_n)). \tag{15.6}$$

Определим приближенное решение задачи (15.1) как такую сеточную функцию, заданную на ω, которое превращает соотношение (15.6) в равенство. Разделив полученное равенство на τ, будем иметь

$$\frac{u_{n+1} - u_n}{\tau} = f(t_n, u_n), \quad n = 0, 1, \ldots, \qquad u_0 = u(0). \tag{15.7}$$

Соотношение (15.7) позволяет рекуррентным образом найти приближенное решение во всех узлах. Численный метод решения задачи (15.1), реализуемый формулами (15.7), называется *методом Эйлера*.

б) **Неявный метод Эйлера**. Заменим теперь интеграл в (15.2) формулой правых прямоугольников (15.5). Для отыскания приближенного решения получим уравнения

$$\frac{u_{n+1}-u_n}{\tau} = f(t_{n+1}, u_{n+1}), \quad n = 0, 1, \ldots, \qquad u_0 = u(0). \tag{15.8}$$

Соотношения (15.8) коренным образом отличаются от соотношений (15.7): для отыскания приближенного решения u_{n+1} теперь нужно решать нелинейные уравнения

$$u_{n+1} - \tau f(t_{n+1}, u_{n+1}) = u_n.$$

Метод (15.8) называется *неявным методом Эйлера*. С точки зрения простоты вычислений он сильно уступает обычному методу Эйлера (15.7). Как будет показано позже, по точности оба метода сравнимы. Еще позже будет установлена существенно бóльшая устойчивость метода (15.8) по сравнению с (15.7).

в) **Метод Рунге**. Заменим интеграл в (15.2) формулой центральных прямоугольников (15.3)

$$u(t_{n+1}) - u(t_n) \approx \tau f(t_{n+1/2}, u(t_{n+1/2})). \tag{15.9}$$

Использованный нами ранее прием получения численного метода путем превращения приближенного равенства в точное за счет замены $u(t_n)$ на u_n здесь напрямую не проходит: в приближенном равенстве фигурирует значение f при u в точке $t_{n+1/2}$, которая не является узловой. Если же мы все же воспользуемся этим приемом и введем *промежуточное* значение приближенного решения в точке $t_{n+1/2}$, то нам потребуется дополнительное уравнение для определения приближенного решения в точке $t_{n+1/2}$. Обозначим промежуточное значение приближенного решения через $u_{n+1/2}$. Тогда из (15.9)

$$\frac{u_{n+1}-u_n}{\tau} = f(t_{n+1/2}, u_{n+1/2}), \tag{15.10}$$

а для отыскания $u_{n+1/2}$ напишем, например, соотношение Эйлера (15.7)

$$\frac{u_{n+1/2}-u_n}{\tau/2} = f(t_n, u_n). \tag{15.11}$$

Итак, в методе (15.10), (15.11) вычисление нового приближенного значения искомого решения u_{n+1} осуществляется поэтапно. Сначала находится промежуточное значение $u_{n+1/2}$ по формуле (15.11), а затем и само u_{n+1} из (15.10). Вычисления по обеим формулам явные. *Метод* (15.10), (15.11) был предложен немецким математиком *Рунге* и носит его имя. Будет показано, что точность метода (15.10), (15.11) выше, чем точность методов (15.7) и (15.8); для вычисления интеграла все же использована более точная квадратурная формула.

Замечание 15.1. В некоторых учебниках по численным методам метод (15.10), (15.11) называется методом предиктор-корректор (предсказывающе корректирующим).

г) **Метод трапеций.** Наконец, заменим интеграл в (15.2) формулой трапеций. В результате получим

$$\frac{u_{n+1}-u_n}{\tau}=\frac{f(t_n,u_n)+f(t_{n+1},u_{n+1})}{2},\quad n=0,1,\dots,\qquad u_0=u(0). \tag{15.12}$$

Как и в случае неявного метода Эйлера (15.8), реализация метода (15.12) требует решения нелинейного уравнения

$$u_{n+1}-\frac{\tau}{2}f(t_{n+1},u_{n+1})=F(u_n).$$

Будет показано, что точность метода (15.12) сравнима с точностью метода Рунге (15.10), (15.11), а по устойчивости он значительно превосходит последний и в этом отношении близок к неявному методу Эйлера (15.8). Метод (15.12) иногда называют *методом трапеций*.

15.3 Аппроксимация.

Определение 15.1. Сеточная функция

$$z_n=u_n-u(t_n),\quad n=1,2,\dots$$

называется погрешностью решения.

Замечание 15.2. Погрешность решения определена только в узлах основной сетки ω, но не в промежуточных узлах.

Выведем уравнение, которому удовлетворяет погрешность решения в методе Эйлера (15.7). Подставив $u_n=z_n+u(t_n)$ в (15.7), получим

$$\frac{z_{n+1}-z_n}{\tau}+\frac{u(t_{n+1})-u(t_n)}{\tau}=f(t_n,u(t_n)+z_n). \tag{15.13}$$

Преобразуем правую часть этого соотношения путем разложения по формуле Тейлора

$$f(t_n,u(t_n)+z_n)=f(t_n,u(t_n))+z_n\frac{\partial f}{\partial u}(t_n,\tilde{u}),$$

где

$$\tilde{u}=u(t_n)+\theta z_n,\qquad 0<\theta<1.$$

Подставляя это разложение в (15.13) и преобразовывая, найдем, что

$$\frac{z_{n+1}-z_n}{\tau}=\frac{\partial f}{\partial u}(t_n,\tilde{u})z_n+\psi_n, \tag{15.14}$$

где

$$\psi_n=f(t_n,u(t_n))-\frac{u(t_{n+1})-u(t_n)}{\tau}. \tag{15.15}$$

Искомое уравнение получено.

Определение 15.2. Сеточная функция ψ_n, задаваемая соотношением (15.15), называется *погрешностью аппроксимации* дифференциального уравнения (15.1) уравнением (15.7).

Замечание 15.3. Погрешность аппроксимации представляет собой разность между правой и левой частями уравнения, определяющего численный метод, если туда вместо приближенного решения подставить точное.

Оценим погрешность аппроксимации метода Эйлера. Используя формулу Тейлора и принимая во внимание уравнение (15.1), в предположении непрерывности второй производной $u(t)$, из (15.15) будем иметь

$$\psi_n = f(t_n, u(t_n)) - \frac{u(t_n) + \tau u'(t_n) + \frac{\tau^2}{2} u''(t_n + \theta\tau) - u(t_n)}{\tau} =$$
$$= [f(t_n, u(t_n)) - u'(t_n)] + \frac{\tau}{2} u''(t_n + \theta\tau) = O(\tau).$$

Тем самым, метод Эйлера имеет первый порядок аппроксимации.

Упражнение 15.1. Исследовать погрешности аппроксимации методов (15.8) и (15.12).

Указание. Для упрощения выкладок разложение по формуле Тейлора в методе (15.8) вести в точке t_{n+1}, а в методе (15.12) — в точке $t_{n+1/2}$.

§ 16

Методы Рунге-Кутты

16.1 Общая концепция

Численные методы решения уравнения

$$\frac{du}{dt} = f(t,u), \quad t > 0, \quad u(0) = u_0 \tag{16.1}$$

и систем таких уравнений, наиболее широко используемые в вычислительной практике, делятся на два больших класса: многошаговые методы и методы типа Рунге-Кутты. Все приведенные в качестве примеров численные методы относятся к методам Рунге-Кутты, хотя некоторые из них могут трактоваться и как многошаговые (одношаговые).

Сейчас мы опишем общую концепцию методов Рунге-Кутты. Для этого вновь обратимся к интегральному соотношению (15.2), на основе которого мы строили изложенные выше методы. Но прежде сделаем одно допущение относительно уравнения (16.1), которое в дальнейшем существенно облегчит нам жизнь. Будем предполагать, что правая часть f этого уравнения не зависит явным образом от t, т.е. $f \equiv f(u)$ и, следовательно,

$$\frac{du}{dt} = f(u), \quad t > 0, \quad u(0) = u_0. \tag{16.1$'$}$$

Сделанное допущение не является ограничением, ибо все численные методы, построенные для одного уравнения, допускают очевидное распространение на случай системы, т.е., вообще говоря, u можно считать вектором. Если же f зависит явным образом от t, то, обозначив например, $t = u_0(t)$ и объявив $u_0(t)$ новой неизвестной, удовлетворяющей уравнению

$$u_0'(t) = 1, \qquad u_0(0) = 0,$$

мы сведем задачу к ранее оговоренному случаю.

Итак, пусть $f = f(u)$. Перепишем для этого случая интегральное соотношение (15.2)

$$u(t_{n+1}) - u(t_n) = \int_{t_n}^{t_{n+1}} f(u)\,dt. \tag{16.2}$$

Сделаем в интеграле (16.2) замену переменной интегрирования, полагая

$$(t - t_n)/\tau = \theta. \tag{16.3}$$

Эта замена переводит отрезок $[t_n, t_{n+1}]$ в $[0, 1]$ так, что

$$u(t_{n+1}) - u(t_n) = \tau \int_0^1 f(\hat{u}(\theta))\,d\theta, \tag{16.4}$$

где

$$\hat{u}(\theta) = u(t(\theta)).$$

Пусть

$$0 \leqslant \theta_1 < \theta_2 < \cdots < \theta_s \leqslant 1 \tag{16.5}$$

суть узлы, а $b_1, b_2, \ldots, b_s$ — веса некоторой квадратурной формулы, аппроксимирующей интеграл $\int\limits_0^1 \varphi(\theta)d\theta$. Используя эту формулу для аппроксимации интеграла в (16.4), будем иметь

$$u(t_{n+1}) - u(t_n) \approx \tau \sum_{i=1}^{s} b_i f\left(\hat{u}(\theta_i)\right). \tag{16.6}$$

Чтобы получить из этого соотношения численный метод, нужно точные значения искомого решения заменить на приближенные, а приближенное равенство — наоборот на точное. Но прежде мы должны ввести дополнительные обозначения. Будем обозначать значение приближенного решения в точке t, отвечающей узлу квадратурной формулы θ_i $(t = t_n + \tau\theta_i)$ через Y_i. Тогда искомое уравнение примет вид

$$u_{n+1} = u_n + \tau \sum_{i=1}^{s} b_i f(Y_i). \tag{16.7}$$

Чтобы получить уравнение для определения Y_i, проинтегрируем (16.1′) от t_n до $t_n + \tau\theta_i$ и сделаем замену (16.3)

$$\hat{u}(\theta_i) - u(t_n) = \int_{t_n}^{t_n+\tau\theta_i} f(u(t))\,dt = \tau \int_0^{\theta_i} f(\hat{u}(\theta))\,d\theta.$$

Заменим и здесь интеграл квадратурной формулой *с теми же узлами* (16.5). Эта квадратурная формула будет несколько своеобразной, ибо не все ее узлы будут лежать

на отрезке интегрирования. Разумеется, ее веса, вообще говоря, должны быть отличны от b_j и даже быть своими для каждого i. Пусть

$$Y_i = u_n + \tau \sum_{j=1}^{s} a_{ij} f(Y_j), \quad i = \overline{1,s}. \tag{16.8}$$

Соотношения (16.7), (16.8) полностью определяют численный метод.

Итак, для того, чтобы найти приближенное решение u_{n+1} (когда u_n уже найдено), сначала нужно решить, вообще говоря, нелинейную систему (16.8) и определить Y_i, $i = \overline{1,s}$, которые затем следует подставить в (16.7).

Определение 16.1. Метод (16.7), (16.8) называется *s-этапным методом Рунге-Кутты.*

Этот метод принято записывать таблицей его коэффициентов, которая называется таблицей Бутчера

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \dots & a_{1s} \\ c_2 & a_{21} & a_{22} & \dots & a_{2s} \\ & \dots & \dots & \dots & \dots \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss} \\ \hline & b_1 & b_2 & \dots & b_s \end{array} \qquad c_i = \sum_{j=1}^{s} a_{ij}. \tag{16.9}$$

Замечание 16.1. Поскольку b_i суть весовые коэффициенты квадратурной формулы для интеграла по единичному отрезку, то $\sum\limits_{i=1}^{s} b_i = 1$. Из аналогичных соображений $c_i = \sum\limits_{j=1}^{s} a_{ij} = \theta_i$.

Определение 16.2. Если в таблице Бутчера (16.9) коэффициенты $a_{ij} = 0$ при $j \geqslant i$, то метод (16.7), (16.8) называется явным s-этапным методом Рунге-Кутты.

Определение 16.3. Если $a_{ij} = 0$ при $i > j$ и хотя бы один $a_{ii} \neq 0$, то метод (16.7), (16.8) называется диагонально неявным.

Во всех остальных случаях мы говорим о неявных методах Рунге-Кутты.

Коэффициенты в таблице Бутчера (16.9) *при заданных ограничениях* выбираются из соображений максимальной точности численного метода.

16.2 Одноэтапные методы Рунге-Кутты

Исследуем одноэтапные ($s = 1$) методы Рунге-Кутты. При $s = 1$ соотношения (16.8), (16.7) принимают вид

$$Y_1 = u_n + \tau a_{11} f(Y_1), \tag{16.10}$$

$$u_{n+1} = u_n + \tau b_1 f(Y_1) \tag{16.11}$$

Из соображений аппроксимации (квадратурная формула должна быть точной по крайней мере на const) находим, что $b_1 = 1$. Если теперь положить $a_{11} = 0$, то метод будет явным, причем $Y_1 = u_n$, а (16.11) можно переписать в виде

$$\frac{u_{n+1} - u_n}{\tau} = f(u_n).$$

Мы получили метод Эйлера. Тем самым, метод Эйлера есть *явный одноэтапный метод Рунге-Кутты*.

Если взять $a_{11} = 1$, то метод (16.10), (16.11) будет неявным. При этом правые части (16.10) и (16.11) совпадают и приводят к соотношению $Y_1 = u_{n+1}$. В этом случае система (16.10), (16.11) преобразуется к виду

$$\frac{u_{n+1} - u_n}{\tau} = f(u_{n+1}).$$

Это неявный метод Эйлера (15.8). Он также является одноэтапным методом Рунге-Кутты.

Исследуем теперь наиболее целесообразный выбор параметров b_1 и a_{11} с точки зрения минимизации погрешности аппроксимации. Чтобы найти погрешность аппроксимации, перепишем уравнение (16.11) в виде

$$\frac{u_{n+1} - u_n}{\tau} = b_1 f(Y_1) \tag{16.12}$$

(ср. с (15.7), (15.8), (15.10) и (15.12)), а решение уравнения (16.10) обозначим через $Y_1(u_n)$. Если, как и выше, $z_n = u_n - u(t_n)$, то

$$\frac{z_{n+1} - z_n}{\tau} = b_1 f(Y_1(u(t_n) + z_n)) - \frac{u(t_{n+1}) - u(t_n)}{\tau}.$$

И снова, раскладывая первое слагаемое правой части по формуле Тейлора, находим, что

$$\begin{aligned}\frac{z_{n+1} - z_n}{\tau} &= b_1 \left[f(Y_1(u(t_n))) + \frac{\partial f}{\partial u}(\tilde{u}) z_n \right] - \frac{u(t_{n+1}) - u(t_n)}{\tau} = \\ &= b_1 \frac{\partial f}{\partial Y_1} \frac{\partial Y_1}{\partial u}(\tilde{u}) z_n + \psi_n,\end{aligned}$$

где

$$\psi_n = b_1 f(Y_1(u(t_n))) - \frac{u(t_{n+1}) - u(t_n)}{\tau} \tag{16.13}$$

— погрешность аппроксимации, а $Y_1(u(t_n))$ — решение уравнения (16.10) с $u(t_n)$ вместо u_n, т.е.

$$Y_1(u(t_n)) = u(t_n) + \tau a_{11} f(Y_1(u(t_n))). \tag{16.14}$$

Замечание 16.2. Погрешность аппроксимации (16.13) представляет собой разность между правой и левой частями уравнения (16.12), если туда вместо приближенного решения подставить точное (ср. с замечанием 15.3).

Разложим погрешность аппроксимации (16.13) по степеням τ. Имеем

$$\psi_n = b_1 \left[f(Y_1) \mid_{\tau=0} + \tau \frac{d\, f(Y_1)}{d\, \tau} \mid_{\tau=0} + \frac{\tau^2}{2} \widetilde{\frac{d^2 f}{d\tau^2}} \right] - \\ - \left[u'(t_n) + \frac{\tau}{2} u''(t_n) + \frac{\tau^2}{6} \tilde{\tilde{u}}''' \right].$$

Из (16.14) находим, что $Y_1|_{\tau=0} = u(t_n)$ и, следовательно,

$$f(Y_1) \mid_{\tau=0} = f(u(t_n)).$$

Снова с использованием (16.14)

$$\frac{df(Y_1)}{d\tau} \mid_{\tau=0} = \frac{df}{dY_1} \frac{dY_1}{d\tau} \mid_{\tau=0} = \frac{df}{du}(u(t_n)) a_{11} f(u(t_n)),$$

а из уравнения (16.1′)

$$u'(t_n) = f(u(t_n)), \quad u''(t_n) = \frac{df}{dt}(u(t_n)) = \frac{df}{du}(u(t_n)) \frac{du}{dt}(t_n) = \frac{df}{du} f.$$

Поэтому

$$\psi_n = (b_1 - 1) f(u(t_n)) + \tau \left[b_1 a_{11} - \frac{1}{2} \right] f(u(t_n)) \frac{df}{du}(u(t_n)) + O(\tau^2).$$

Тем самым, для того, чтобы погрешность аппроксимации была $O(\tau^2)$, необходимо и достаточно, чтобы выполнялись условия

$$b_1 = 1, \qquad a_{11} b_1 = 1/2. \tag{16.15}$$

Отсюда находим

$$b_1 = 1, \qquad a_{11} = 1/2$$

и, следовательно, неявный одноэтапный метод Рунге-Кутты

$$\begin{aligned} Y_1 &= u_n + \frac{\tau}{2} f(Y_1), \\ u_{n+1} &= u_n + \tau f(Y_1) \end{aligned} \tag{16.16}$$

имеет второй порядок аппроксимации.

Замечание 16.3. Из первого уравнения (16.16) следует, что момент времени, на который Y_1 приближает $u(t)$, есть $t + \tau/2$, ибо для задачи $u' = 1$, $u(0) = 0$, имеющей решение $u = t$, $Y_1 = u_n + \tau/2 = t_n + \tau/2$.

Соотношения (16.16) можно преобразовать. Исключив $f(Y_1)$, найдем, что $u_{n+1} = 2Y_1 - u_n$. Выражая отсюда Y_1 и подставляя его во второе уравнение (16.16), получим

$$u_{n+1} = u_n + \tau f\left(\frac{u_{n+1} + u_n}{2}\right).$$

Замечание 16.4. Метод (16.16) очень сильно напоминает метод Рунге (15.10), (15.11). Отличие между ними состоит в том, что здесь промежуточное значение находится по неявной формуле, а в методе Рунге по явной формуле (15.11). Метод (16.16), как мы уже сказали, является одноэтапным (неявным) методом Рунге-Кутты, а метод (15.10), (15.11) — двухэтапным (явным) методом. Подчеркнем, что слову *этап* здесь мы придаем четкий математический смысл.

16.3 Методы третьего порядка аппроксимации

Выясним ограничения на коэффициенты (16.9), обеспечивающие третий порядок аппроксимации s-этапного метода Рунге-Кутты. Для этого нужно исследовать погрешность аппроксимации

$$\psi_n := \psi_n(\tau) := \sum_{i=1}^{s} b_i f(Y_i(u(t_n))) - \frac{u(t_{n+1}) - u(t_n)}{\tau},$$

где

$$Y_i(u(t_n)) = u(t_n) + \tau \sum_{j=1}^{s} a_{ij} f(Y_j(u(t_n))) =: Y_i(u(t_n); \tau). \tag{16.17}$$

Раскладывая $\psi_n(\tau)$ по τ до третьего порядка, будем иметь

$$\begin{aligned}\psi_n(\tau) = \sum_{i=1}^{s} b_i \left[f(Y_i) \Big|_{\tau=0} + \tau \frac{d\, f(Y_i)}{d\,\tau} \Big|_{\tau=0} + \frac{\tau^2}{2} \frac{d^2\, f(Y_i)}{d\,\tau^2} \Big|_{\tau=0} + O(\tau^3) \right] - \\ - \left[u'(t_n) + \frac{\tau}{2} u''(t_n) + \frac{\tau^2}{6} u'''(t_n) + O(\tau^3) \right].\end{aligned} \tag{16.18}$$

Поскольку $f(Y_i(u(t_n)))$ есть сложная функция τ, то вычислим сначала производные по τ функции $Y_i(u(t_n); \tau)$ при $\tau = 0$. Из (16.17) с учетом (16.9), находим, что

$$Y_i \Big|_{\tau=0} = u(t_n),$$

$$Y_i' \Big|_{\tau=0} = \frac{d\, Y_i}{d\,\tau} \Big|_{\tau=0} = \left[\sum_{j=1}^{s} a_{ij} f(Y_i) + \tau \sum_{j=1}^{s} a_{ij} \frac{d\, f}{d\, Y_j} Y_j' \right] \Big|_{\tau=0} = f(u(t_n)) c_i,$$

$$Y_i'' \Big|_{\tau=0} = \left[2 \sum_{j=1}^{s} a_{ij} \frac{df}{dY_j} Y_j' + \tau \sum_{j=1}^{s} a_{ij} \frac{d^2 f}{dY_j^2} (Y_j')^2 + \tau \sum_{j=1}^{s} a_{ij} \frac{df}{dY_j} Y_j'' \right] \Big|_{\tau=0} = 2 f(u(t_n)) \frac{df}{du} \sum_{j=1}^{s} a_{ij} c_j.$$

Теперь можно найти производные f:

$$f(Y_i(u(t_n)))\Big|_{\tau=0} = f(u(t_n)),$$
$$\frac{df(Y_i)}{d\tau}\Big|_{\tau=0} = \frac{df}{dY_i}Y_i'\Big|_{\tau=0} = f(u(t_n))\frac{df}{du}c_i,$$
$$\frac{d^2f(Y_i)}{d\tau^2}\Big|_{\tau=0} = \left[\frac{d^2f}{dY_i^2}(Y_i')^2 + \frac{df}{dY_i}Y_i''\right]\Big|_{\tau=0} = f^2(u(t_n))\frac{d^2f}{du^2}c_i^2 + 2f(u(t_n))\left(\frac{df}{du}\right)^2\sum_{j=1}^{s}a_{ij}c_j.$$

Далее, из (16.1′)

$$u' = f, \qquad u'' = \frac{df}{du}u' = f'f, \qquad u''' = f''u'f + (f')^2u' = f''f^2 + (f')^2f$$

и, следовательно,

$$\frac{u(t_{n+1}) - u(t_n)}{\tau} = f + \frac{\tau}{2}f'f + \frac{\tau^2}{6}\left(f''f^2 + (f')^2f\right) + O(\tau^3).$$

Подставляя теперь найденные разложения в (16.18), будем иметь

$$\begin{aligned}\psi_n = &\sum_{i=1}^{s}b_i\left[f + \tau ff'c_i + \frac{\tau^2}{2}\left(f^2f''c_i^2 + 2ff'^2\sum_{j=1}^{s}a_{ij}c_j\right)\right] - \\ &- \left[f + \frac{\tau}{2}ff' + \frac{\tau^2}{6}\left(f^2f'' + ff'^2\right)\right] + O(\tau^3).\end{aligned} \tag{16.19}$$

Отсюда, приравнивая коэффициенты при одинаковых степенях τ, находим, что условия третьего порядка аппроксимации суть

$$\begin{aligned}&\sum_{i=1}^{s}b_i = 1 \quad , \\ &\sum_{i=1}^{s}b_ic_i = \frac{1}{2} \quad ,\end{aligned} \tag{16.20}$$

$$\begin{aligned}&\sum_{i=1}^{s}b_ic_i^2 = \frac{1}{3}, \\ &\sum_{i,j=1}^{s}b_ia_{ij}c_j = \frac{1}{6}.\end{aligned} \tag{16.21}$$

При этом (16.20) суть условия второго порядка аппроксимации.

Замечание 16.5. Чтобы иметь условия четвертого порядка аппроксимации, к условиям (16.20), (16.21) нужно добавить следующие условия:

$$\begin{aligned}
&\sum_{i=1}^{s} b_i c_i^3 = \frac{1}{4},\\
&\sum_{i,j=1}^{s} b_i c_i a_{ij} c_j = \frac{1}{8},\\
&\sum_{i,j=1}^{s} b_i a_{ij} c_j^2 = \frac{1}{12},\\
&\sum_{i,j,k=1}^{s} b_i a_{ij} a_{jk} c_k = \frac{1}{24}.
\end{aligned} \tag{16.22}$$

Замечание 16.6. Условия (16.20) с учетом замечания 16.1 можно трактовать как условия точности квадратурной формулы из (16.6) на линейных функциях.[1] Добавление к этим условиям первого из соотношений (16.21), а затем и первого из соотношений (16.22) на указанную квадратурную формулу накладывает дополнительные условия точности на квадратичных и кубичных функциях.

Упражнение 16.1. Показать, что метод трапеций (15.12) является неявным двухэтапным методом Рунге-Кутты второго порядка аппроксимации. (Найти все b_i, a_{ij} и показать невыполнение хотя бы одно из условий (16.21))

Ответ:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array} \qquad \left(\frac{1}{2}0^2 + \frac{1}{2}1^2\right) \neq \frac{1}{3}.$$

Упражнение 16.2. Показать, что метод Рунге (15.10), (15.11) является явным двухэтапным методом Рунге-Кутты второго порядка.

Ответ:

$$\begin{array}{c|cc} 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array} \qquad \left[0 \cdot 0 + 1 \cdot \frac{1}{4}\right] \neq \frac{1}{3}.$$

16.4 Двухэтапные неявные методы третьего порядка

Положим в (16.20), (16.21) параметр $s = 2$. В результате система примет вид

$$\begin{aligned}
&b_1 + b_2 = 1,\\
&c_1 b_1 + c_2 b_2 = 1/2,\\
&c_1^2 b_1 + c_2^2 b_2 = 1/3,\\
&b_1(a_{11}c_1 + a_{12}c_2) + b_2(a_{21}c_1 + a_{22}c_2) = 1/6.
\end{aligned} \tag{16.23}$$

[1] Ведь $c_i = \theta_i$, т.е. координата переменной интегрирования в i-ом узле.

Эта система содержит четыре уравнения и шесть неизвестных (Если не считать c_1 и c_2, задаваемые (16.9)). Поэтому, вообще говоря, два из этих неизвестных должны остаться свободными, а остальные выразиться через них. Система (16.23) нелинейная, и нет регулярных способов ее решения. Укажем один путь, приводящий к решению этой системы.

Для отыскания решения системы (16.23) предположим сначала, что неизвестные c_1 и c_2 найдены, и рассмотрим первые три уравнения (16.23) как систему линейных уравнений относительно b_1 и b_2. Поскольку эта система переопределена, то для ее разрешимости необходимо обращение в нуль определителя расширенной матрицы

$$\begin{aligned}
&\begin{vmatrix} 1 & 1 & 1 \\ c_1 & c_2 & 1/2 \\ c_1^2 & c_2^2 & 1/3 \end{vmatrix} = \frac{1}{3}c_2 + \frac{1}{2}c_1^2 + c_1c_2^2 - c_1^2c_2 - \frac{1}{2}c_2^2 - \frac{1}{3}c_1 = \\
&= \frac{1}{3}(c_2 - c_1) - \frac{1}{2}(c_2 + c_1)(c_2 - c_1) + c_1c_2(c_2 - c_1) = \\
&= (c_2 - c_1)\left[\frac{1}{3} - \frac{c_1 + c_2}{2} + c_1c_2\right] = 0.
\end{aligned} \tag{16.24}$$

Проанализируем это соотношение. Если бы $c_1 = c_2$, то последнее уравнение (16.23) приняло бы вид

$$c_1^2 b_1 + c_2^2 b_2 = 1/6,$$

что противоречит третьему уравнению (16.23), и поэтому

$$c_1 - c_2 \neq 0. \tag{16.25}$$

Тем самым, из (16.24) следует, что

$$2 - 3(c_1 + c_2) + 6c_1c_2 = 0$$

или

$$(3 - 6c_1)c_2 = 2 - 3c_1.$$

Поскольку $c_1 = 1/2$ не удовлетворяет этому уравнению, то

$$c_1 \neq 1/2 \tag{16.26}$$

и можно найти

$$c_2 = \frac{2 - 3c_1}{3 - 6c_1}. \tag{16.27}$$

Разрешим теперь первые два уравнения (16.23) относительно b_1 и b_2 при помощи правила Крамера. Будем иметь

$$\Delta = \begin{vmatrix} 1 & 1 \\ c_1 & c_2 \end{vmatrix} = c_2 - c_1 \neq 0, \qquad \Delta_1 = \begin{vmatrix} 1 & 1 \\ 1/2 & c_2 \end{vmatrix} = c_2 - 1/2, \qquad \Delta_2 = \begin{vmatrix} 1 & 1 \\ c_1 & 1/2 \end{vmatrix} = 1/2 - c_1$$

и, следовательно,

$$b_1 = \frac{c_2 - 1/2}{c_2 - c_1} = \frac{1}{4(3c_1^2 - 3c_1 + 1)}, \qquad b_2 = \frac{1/2 - c_1}{c_2 - c_1}. \tag{16.28}$$

Из (16.27), (16.28) следует, что c_1 можно принять за параметр. В качестве второго параметра возьмем a_{12}. Тогда

$$a_{11} = c_1 - a_{12}. \tag{16.29}$$

Поскольку

$$a_{21} = c_2 - a_{22}, \tag{16.30}$$

то, подставляя эти выражения для a_{11} и a_{21} в последнее из уравнений (16.23), получим

$$b_1[(c_1 - a_{12})c_1 + a_{12}c_2] + b_2[(c_2 - a_{22})c_1 + a_{22}c_2] = 1/6.$$

Принимая во внимание (16.28), второе из уравнений (16.23) и разрешая полученное соотношение относительно a_{22}, будем иметь

$$a_{22} = \frac{1/6 - c_1/2 - a_{12}(c_2 - 1/2)}{1/2 - c_1} = \frac{(1 - 3c_1)(1 - 2c_1) - a_{12}}{3(1 - 2c_1)^2}. \tag{16.31}$$

Соотношения (16.27), (16.28), (16.29), (16.30), (16.31) задают двухпараметрическое семейство неявных двухэтапных методов Рунге-Кутты третьего порядка.

Если положить, например,

$$a_{12} = 0, \qquad c_1 \equiv a_{11} = a_{22}, \tag{16.32}$$

то для c_1 из (16.30) получим квадратное уравнение $6c_1^2 - 6c_1 + 1$ с корнями $c_1 = \gamma = \dfrac{3 \pm \sqrt{3}}{6}$. Таблица Бутчера для этого метода имеет вид

$$\begin{array}{l|ll} \theta_1 = \gamma & \gamma & 0 \\ \theta_2 = 1 - \gamma & 1 - 2\gamma & \gamma \\ \hline & 1/2 & 1/2 \end{array} \qquad \gamma = \frac{3 \pm \sqrt{3}}{6}. \tag{16.33}$$

Упражнение 16.3. Доказать, что метод (16.33) есть метод (16.27)-(16.32).

16.5 Явные двухэтапные методы

В силу определения для явного двухэтапного метода $a_{11} = a_{12} = a_{22} = 0$ и лишь $a_{21} \neq 0$. (При $a_{21} = 0$ мы получим явный одноэтапный метод.) Поскольку двухэтапные методы третьего порядка имеют лишь два свободных параметра, а мы задали три, то рассчитывать на третий порядок у явных двухэтапных методов, вообще говоря, не приходится. Мы покажем, что так оно и есть.

Принимая a_{21} за параметр, из условий второго порядка аппроксимации (16.20), которые в нашем случае принимают вид

$$b_1 + b_2 = 1, \qquad a_{21} b_2 = 1/2,$$

находим

$$b_1 = \left(1 - \frac{1}{2a_{21}}\right), \qquad b_2 = \frac{1}{2a_{21}}.$$

Тем самым, явные двухэтапные методы Рунге-Кутты второго проядка образуют однопараметрическое семейство.

Далее, поскольку в рассматриваемом случае наряду с a_{11}, a_{12}, a_{22} и $c_1 = 0$, то левая часть четвертого из условий (16.23) обращается в нуль и следовательно это условие выполненным быть не может. Мы доказали, что явных двухэтапных методов третьего порядка не существует.

Если положить, например, $a_{21} = 1$, то получим метод Хойна

$$\begin{aligned} &Y_1 = u_n, \quad Y_2 = u_n + \tau f(Y_1), \\ &u_{n+1} = u_n + \frac{\tau}{2}[f(Y_1) + f(Y_2)]. \end{aligned} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

Упражнение 16.4. Выписать все построенные методы второго порядка.

16.6 Двухэтапный метод четвертого порядка

Коэффициенты метода четвертого порядка должны удовлетворять еще четырем условиям (16.22). Хотя у двухэтапного метода третьего порядка осталось только два параметра, существует единственный двухэтапный метод четвертого порядка. Его коэффициенты суть

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array} \tag{16.34}$$

Замечание 16.7. Если бы f не зависела от u ($u' = f(t)$, $u_{n+1} = u_n + \tau \int_0^1 \hat{f}(\theta)\, d\theta$), то двухэтапный метод четвертого порядка получился бы только если квадратура в (16.7) была квадратурой Гаусса, т.е. $b_1 = b_2 = 1/2$, а θ_1 и θ_2 — сдвинутые на $[0,1]$ нули полинома Лежандра второй степени $P_2(x) = \frac{1}{2}(3x^2 - 1)$. Корнями этого полинома являются числа $x_{1,2} = \pm 1/\sqrt{3}$. Делая линейную замену, переводящую отрезок $[-1,1]$ в отрезок $[0,1]$, находим, что узлы квадратуры Гаусса на $[0,1]$ суть $\theta_{1,2} = 1/2 \mp \sqrt{3}/6$, как в (16.34). Остальные коэффициенты получаются, если проинтегрировать по $[0,\theta_1]$ и $[0,\theta_2]$ весовые функции интерполяционного полинома Лагранжа с гауссовыми узлами.

Упражнение 16.5. Доказать, что в (16.34)

$$a_{ij} = \int_0^{\theta_j} p_i(\theta)\, d\theta,$$

где $p_i(\theta)$ — линейная функция (интерполянт по двум узлам) такая, что $p_i(\theta_i) = 1$, $p_i(\theta_j) = 0$ при $i \neq j$. Убедиться в выполнении условий (16.20)-(16.22).

16.7 Явные трехэтапные методы Рунге-Кутты третьего порядка

Рассмотрим более подробно явные трехэтапные методы. В силу определения

$$a_{11} = a_{12} = a_{13} = a_{22} = a_{23} = a_{33} = 0,$$

и указанные методы задаются таблицей

$$\begin{array}{c|ccc} c_2 & a_{21} & & \\ c_3 & a_{31} & a_{32} & \\ \hline & b_1 & b_2 & b_3 \end{array}$$

Условия третьего порядка аппроксимации (16.20), (16.21) в рассматриваемом случае принимают вид

$$\begin{aligned} &b_1 + b_2 + b_3 = 1, \\ &b_2 c_2 + b_3 c_3 = 1/2, \\ &b_2 c_2^2 + b_3 c_3^2 = 1/3, \\ &b_3 a_{32} c_2 = 1/6. \end{aligned} \tag{16.35}$$

Эта система имеет два однопараметрических семейства решений и одно двухпараметрическое. Найдем их. Будем рассматривать второе и третье уравнения системы (16.35) как линейную систему относительно b_2 и b_3. Эта система может быть как вырожденной (и это приводит к двум однопараметрическим семействам решений), так и невырожденной (двухпараметрическое семейство).

Пусть эта система вырождена, т.е.

$$\begin{vmatrix} c_2 & c_3 \\ c_2^2 & c_3^2 \end{vmatrix} = c_2 c_3 (c_3 - c_2) = 0. \tag{16.36}$$

В силу последнего из уравнений (16.35) $c_2 \neq 0$, и поэтому либо

$$c_3 = 0 \tag{16.37}$$

либо

$$c_2 = c_3. \tag{16.38}$$

i) Пусть сначала имеет место (16.37). Тогда второе и третье уравнения (16.35) принимают вид

$$c_2 b_2 = 1/2, \qquad c_2^2 b_2 = 1/3$$

и, следовательно,

$$c_2 = 2/3, \qquad b_2 = 3/4.$$

Если теперь $b_3 = b$ принять за параметр, то из последнего уравнения (16.35) находим

$$a_{32} = \frac{1}{4b}.$$

Поскольку $c_3 = 0$, то

$$a_{31} = -a_{32} = -\frac{1}{4b}$$

и, наконец, из первого уравнения (16.35)

$$b_1 = \frac{1}{4} - b.$$

Собирая найденные значения, получим таблицу

$$\begin{array}{c|ccc} 2/3 & 2/3 & & \\ 0 & -(4b)^{-1} & (4b)^{-1} & \\ \hline & 1/4 - b & 3/4 & b \end{array} \tag{16.39}$$

ii) Теперь пусть имеет место (16.38). Снова из второго и третьего уравнений (16.35) находим, что

$$b_2 + b_3 = \frac{1}{2c_2} = \frac{1}{3c_2^2},$$

т.е.

$$c_3 = c_2 = 2/3, \qquad b_2 = 3/4 - b,$$

где $b = b_3$ — параметр. Из последнего уравнения (16.35)

$$a_{32} = \frac{1}{4b},$$

а из первого уравнения

$$b_1 = 1/4.$$

Таблица рассматриваемого метода имеет вид

$$\begin{array}{c|ccc} 2/3 & 2/3 & & \\ 2/3 & \dfrac{2}{3} - \dfrac{1}{4b} & \dfrac{1}{4b} & \\ \hline & 1/4 & 3/4 - b & b \end{array} \tag{16.40}$$

iii) Если соотношение (16.36) места не имеет, то из второго и третьего уравнений (16.35) находим

$$b_2 = \frac{c_3/2 - 1/3}{c_2(c_3 - c_2)}, \qquad b_3 = \frac{1/3 - c_2/2}{c_3(c_3 - c_2)}, \tag{16.41}$$

Привлекая первое уравнение (16.35), найдем, что

$$b_1 = 1 - \frac{3(c_2 + c_3) - 2}{6c_2c_3}, \tag{16.42}$$

а из четвертого

$$a_{32} = \frac{c_3(c_3 - c_2)}{c_2(2 - 3c_2)}. \tag{16.43}$$

Еще раз напомним, что в методе (16.41)-(16.43) параметры c_2 и c_3 удовлетворяет условию

$$c_2c_3(c_3 - c_2)(c_2 - 2/3) \neq 0.$$

Среди явных трехэтапных методов Рунге-Кутты третьего порядка в силу исторических причин наибольшей популярностью пользуется следующий метод из семейства (16.41)-(16.43)

$$\begin{array}{c|ccc} 1/2 & 1/2 & & \\ 1 & -1 & 2 & \\ \hline & 1/6 & 2/3 & 1/6 \end{array}. \tag{16.44}$$

16.8 Более общие методы Рунге-Кутты

Теорема 16.1. *Не существует явного s-этапного метода Рунге-Кутты порядка p, если $p > s$.*

Для $s = 1, 2$ эта теорема нами фактически доказана.

Теорема 16.2. *При $s \geqslant 5$ не существует явного s-этапного метода Рунге-Кутты порядка $p = s$ (1963г.). При $s \geqslant 8$ не существует явного s-этапного метода Рунге-Кутты порядка $p = s - 1$ (1965г.). При $s \geqslant 10$ — $p = s - 2$ (1985г.).*

Замечание 16.8. При $s = 6$ существует явный метод Рунге-Кутты порядка 5. При $s = 7$ существует явный метод порядка 6. Наивысший порядок, фактически достигнутый для явно построенных явных методов Рунге-Кутты равен 10. При этом число этапов равно 17. (Этот результат занесен в книгу рекордов Гиннеса).

Теорема 16.3. *При любом s существует единственный неявный метод Рунге-Кутты порядка $p = 2s$.*

Замечание 16.9. Для оптимального метода порядка $p = 2s$ узлы θ_j и веса b_j суть узлы и веса квадратурной формулы Гаусса, а

$$a_{ij} = \int_0^{\theta_j} p_i(\theta)\, d\theta,$$

где $p_j(\theta)$ — многочлен степени s такой, что $p_i(\theta_i) = 1$, $p_i(\theta_j) = 0$ при $i \neq j$.

В вычислительной практике широко используется следующий явный 4-этапный метод Рунге-Кутты 4-го порядка

1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	1/3	1/3	1/6.

Замечание 16.10. До недавнего времени в методах Рунге-Кутты (явных) вместо переменных Y_j фигурировали $k_j = f(Y_j)$. Поэтому, вместо (16.7), (16.8) писали

$$\begin{aligned} k_i &= f(u_n + \tau \sum_{j=1}^{s-1} a_{ij} k_j), \\ u_{n+1} &= u_n + \tau \sum_{j=1}^{s} b_j k_j. \end{aligned} \tag{16.45}$$

16.9 Сходимость методов Рунге-Кутты

Установим оценку погрешности приближенного решения, получаемого при помощи того или иного метода Рунге-Кутты.

Если, как и раньше,

$$z_n = u_n - u(t_n),$$

то из (16.7) находим, что z_{n+1} удовлетворяет уравнению

$$\frac{z_{n+1} - z_n}{\tau} = \sum_{i=1}^{s} b_i \frac{d\, f(Y_i(u))}{d\, u} \Bigg|_{u=u(t_n)+\sigma_i z_n} z_n + \psi_n. \tag{16.46}$$

Прежде чем оценивать z_{n+1}, оценим коэффициент при z_n в правой части (16.46). Будем при этом предполагать, что

$$\max_{|u|<\infty} \left| \frac{d\, f(u)}{d\, u} \right| \leqslant L. \tag{16.47}$$

Тогда

$$\max_{|u|<\infty}\left|\frac{d\,f(Y_j(u))}{d\,u}\right|=\max_{|u|<\infty}\left|\frac{d\,f}{d\,Y_j}\,\frac{d\,Y_j}{d\,u}\right|\leqslant L\max_{|u|<\infty}\left|\frac{d\,Y_j}{d\,u}\right|. \tag{16.48}$$

Оценим $\left|\,dY_j/du\right|$. Из (16.8) с u вместо u_n

$$\frac{d\,Y_i}{d\,u}=1+\tau\sum_{j=1}^{s}a_{ij}\frac{d\,f}{d\,Y_j}\,\frac{d\,Y_j}{d\,u}.$$

Пусть

$$\max_{|u|<\infty}\left|\frac{d\,Y_{i_0}}{d\,u}\right|=\max_{1\leqslant j\leqslant s}\max_{|u|<\infty}\left|\frac{d\,Y_j}{d\,u}\right|.$$

Тогда с учетом (16.47)

$$\max_{|u|<\infty}\left|\frac{d\,Y_{i_0}}{d\,u}\right|\leqslant 1+\tau\sum_{j=1}^{s}|a_{i_0j}|\,L\max_{|u|<\infty}\left|\frac{d\,Y_j}{d\,u}\right|\leqslant 1+\tau asL\max_{|u|<\infty}\left|\frac{d\,Y_{i_0}}{d\,u}\right|,$$

где

$$a=\max_{ij}\left|\,a_{ij}\,\right|, \tag{16.49}$$

и, следовательно,

$$(1-\tau asL)\max_{|u|<\infty}\left|\frac{d\,Y_{i_0}}{d\,u}\right|\leqslant 1.$$

Будем предполагать, что

$$\tau\leqslant\frac{1}{2asL},\quad\text{т.е.}\quad 1-\tau asL\geqslant\frac{1}{2}. \tag{16.50}$$

Тогда

$$\left|\frac{d\,Y_j}{d\,u}\right|\leqslant 2,\quad j=1,\ldots,s. \tag{16.51}$$

Для простоты будем предполагать, что коэффициенты b_j неотрицательны. Поскольку их сумма равна единице, то с учетом (16.48), (16.51)

$$\left|\sum_{j=1}^{s}b_j\frac{d\,f(Y_j(u))}{d\,u}\right|\leqslant 2L.$$

Принимая во внимание эту оценку, из (16.46) находим, что

$$|z_{n+1}|\leqslant(1+2\tau L)|z_n|+\tau|\psi_n|.$$

Разрешим эти неравенства. Поскольку $z_0=u_0-u(0)=0$, то

$$\begin{aligned}
&|z_1|\leqslant\tau|\psi_0|,\\
&|z_2|\leqslant(1+2\tau L)|z_1|+\tau|\psi_1|,\\
&|z_3|\leqslant(1+2\tau L)|z_2|+\tau|\psi_2|,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&|z_n|\leqslant(1+2\tau L)|z_{n-1}|+\tau|\psi_{n-1}|
\end{aligned}$$

Подставим теперь оценку $|z_1|$ в правую часть оценки $|z_2|$, а полученную оценку $|z_2|$ в свою очередь в правую часть оценки $|z_3|$ и т.д. Получим

$$|z_n| \leqslant \sum_{j=0}^{n-1} \tau(1+2\tau L)^{n-1-j}|\psi_j| \leqslant (1+2\tau L)^n \sum_{j=0}^{n-1} \tau|\psi_j| \leqslant e^{2\tau LT/\tau} T \max_j |\psi_j| = \\ = e^{2LT} T \max_j |\psi_j|. \tag{16.52}$$

Из (16.52) следует

Теорема 16.4. *Если метод Рунге-Кутты* (16.7), (16.8) *аппроксимирует исходное уравнение (16.1′)с порядком p, то при* $\tau \to 0$ *он сходится с тем же порядком.*

§ 17

Линейные многошаговые методы

При изучении методов Рунге-Кутты, используемых при решении задачи Коши

$$\frac{du}{dt} = f(u), \quad t > 0, \qquad u(0) = u_0, \tag{17.1}$$

мы не обращали особого внимания на задание начальных условий, ибо это совершенно тривиальная процедура: для того, чтобы начал работать любой из рассмотренных нами методов Рунге-Кутты, нужно задать $u_0 = u(0)$, т.е. так же как и для дифференциального уравнения. Обусловлено это тем, что в каждом уравнении связаны между собой значения приближенного решения в двух соседних узлах сетки (не считая промежуточных значений). Другой класс методов составляют так называемые многошаговые методы, в которых уравнения связывают значения приближенного решения в нескольких соседних узлах.

17.1 Методы Адамса

Наиболее известными из многошаговых методов и наиболее старыми являются методы Адамса. Опишем эти методы на примере уравнения (17.1). Вновь будем предполагать, что на отрезке интегрирования введена равномерная сетка с шагом τ, а уравнение (17.1) проинтегрировано по отрезку между узлами t_n и t_{n+1}

$$u(t_{n+1}) - u(t_n) = \int_{t_n}^{t_{n+1}} f(u(t))dt. \tag{17.2}$$

Заменим подынтегральную функцию в (17.2) интерполяционным многочленом Лагранжа по некоторым узлам сетки ω (а не по промежуточным узлам, как это было в методах Рунге-Кутты (!)). В зависимости от того, участвует ли узел t_{n+1} в интерполяции $f(u(t))$ или нет, различают неявные и явные методы Адамса.

а) **Явные методы Адамса**. Предположим, что $u(t)$ известна в k узлах сетки ω

$$t_n, t_{n-1}, \ldots, t_{n+1-k}. \tag{17.3}$$

Построим по этим узлам интерполяционный многочлен Лагранжа степени $k-1$ для подынтегральной функции $f(u(t))$ из (17.2)

$$f(u(t)) \approx L_{k-1}(t) = \sum_{j=1}^{k} p_j(t) f(u(t_{n+1-j})), \tag{17.4}$$

где, как обычно,

$$p_j(t) = \prod_{\substack{i=1 \\ i \neq j}}^{k} \frac{t - t_{n+1-i}}{t_{n+1-j} - t_{n+1-i}} \tag{17.5}$$

суть весовые функции интерполяционного полинома (многочлены степени $(k-1)$), обращающиеся в нуль при $t = t_{n+1-i}$, $i = \overline{1, j-1}, \overline{j+1, k}$ и в единицу при $t = t_{n+1-j}$). Подставляя (17.4), (17.5) в (17.2) и заменяя приближенное равенство на точное, получим следующее уравнение для определения приближенного решения

$$u_{n+1} - u_n = \tau \sum_{j=1}^{k} b_j f(u_{n+1-j}), \tag{17.6}$$

где

$$b_j = \frac{1}{\tau} \int_{t_n}^{t_{n+1}} p_j(t)dt = \int_0^1 \hat{p}_j(\theta)d\theta = \int_0^1 \prod_{\substack{i=1 \\ i \neq j}}^{k} \frac{\theta - 1 + i}{i - j} d\theta. \tag{17.7}$$

Определение 17.1. Численный метод (17.6), (17.7) называется явным k-шаговым методом Адамса (иногда его называют методом Адамса-Бэшфорта).

Примеры. 1°. $k = 1$.

$$p_1(t) = \hat{p}_1(\theta) = 1, \qquad b_1 = 1.$$

2°. $k = 2$.

$$\begin{aligned} &\hat{p}_1(\theta) = \theta + 1, \quad && b_1 = 3/2, \\ &\hat{p}_2(\theta) = -\theta, && b_2 = -1/2. \end{aligned}$$

3°. $k = 3$.

$$\begin{aligned} &\hat{p}_1(\theta) = \frac{1}{2}(\theta+1)(\theta+2), && b_1 = \frac{23}{12}, \\ &\hat{p}_2(\theta) = -\theta(\theta+2), && b_2 = -\frac{4}{3}, \\ &\hat{p}_3(\theta) = \frac{\theta(\theta+1)}{2}, && b_3 = \frac{5}{12}. \end{aligned}$$

Выпишем уравнения (17.6) для этих частных случаев

$$\begin{aligned} u_{n+1} &= u_n + \tau f(u_n), \\ u_{n+1} &= u_n + \tau \left[\frac{3}{2} f(u_n) - \frac{1}{2} f(u_{n-1}) \right], \\ u_{n+1} &= u_n + \tau \left[\frac{23}{12} f(u_n) - \frac{16}{12} f(u_{n-1}) + \frac{5}{12} f(u_{n-2}) \right]. \end{aligned} \tag{17.8}$$

Упражнение 17.1. Построить явный 4-х-шаговый метод Адамса (17.6).

Ответ.

$$u_{n+1} = u_n + \tau \left[\frac{55}{24} f(u_n) - \frac{59}{24} f(u_{n-1}) + \frac{37}{24} f(u_{n-2}) - \frac{9}{24} f(u_{n-3}) \right].$$

Замечание 17.1. Очевидно, что первое из уравнений (17.8) определяет исследованный нами ранее метод Эйлера. Тем самым, метод Эйлера может быть отнесен как к методам Рунге-Кутты, так и к методам Адамса.

Формулы (17.6) получены при интегрировании в пределах от t_n до t_{n+1}, в то время как узлы интерполяции располагались на отрезке $[t_{n+1-k}, t_n]$, т.е. вне интервала интегрирования (Для подынтегральной функции использовалась экстраполяция). В связи с этим явные методы Адамса иногда называют экстраполяционными методами.

б) **Неявные методы Адамса**. Можно построить и неявные методы Адамса. Для этого к узлам интерполяции (17.3) нужно добавить еще узел t_{n+1}. В этом случае интерполяционный многочлен (степени k) примет вид

$$L_k(t) = \sum_{j=0}^{k} p_j(t) f(u(t_{n+1-j})), \tag{17.9}$$

а соответствующим ему уравнением будет уравнение

$$u_{n+1} - u_n = \tau \sum_{j=0}^{k} b_j f(u_{n+1-j}), \tag{17.10}$$

где (ср. с (17.7))

$$b_j = \frac{1}{\tau} \int_{t_n}^{t_{n+1}} p_j(t) dt = \int_0^1 \prod_{\substack{i=0 \\ i \neq j}}^{k} \frac{\theta - 1 + i}{i - j} d\theta. \tag{17.11}$$

Определение 17.2. Численный метод (17.10), (17.11) называется неявным k-шаговым методом Адамса (Иногда его называют методом Адамса-Мултона).

Примеры. 4°. $k = 0$.

$$\hat{p}_0(\theta) = 1, \quad b_0 = 1.$$

5°. $k = 1$.

$$\begin{aligned} \hat{p}_0(\theta) &= \theta, & b_0 &= 1/2, \\ \hat{p}_1(\theta) &= -\theta + 1, & b_1 &= 1/2. \end{aligned}$$

6°. $k = 2$.

$$\begin{aligned} \hat{p}_0(\theta) &= \frac{1}{2}\theta(\theta+1), & b_0 &= \frac{5}{12}, \\ \hat{p}_1(\theta) &= -(\theta^2-1), & b_1 &= \frac{2}{3}, \\ \hat{p}_2(\theta) &= \frac{1}{2}\theta(\theta-1), & b_2 &= -\frac{1}{12}. \end{aligned}$$

Напишем уравнения (17.10) для этих частных случаев

$$\begin{aligned} &u_{n+1} = u_n + \tau f(u_{n+1}), \\ &u_{n+1} = u_n + \frac{\tau}{2}\left[f(u_{n+1}) + f(u_n)\right], \\ &u_{n+1} = u_n + \tau\left[\frac{5}{12}f(u_{n+1}) + \frac{8}{12}f(u_n) - \frac{1}{12}f(u_{n-1})\right]. \end{aligned} \tag{17.12}$$

Упражнение 17.2. Построить неявный 3-х-шаговый метод Адамса (17.10), (17.11).

Ответ.

$$u_{n+1} = u_n + \tau\left[\frac{9}{24}f(u_{n+1}) + \frac{19}{24}f(u_n) - \frac{5}{24}f(u_{n-1}) + \frac{1}{24}f(u_{n-2})\right].$$

Замечание 17.2. Очевидно, что первое из уравнений (17.12), отвечающее $k = 0$, является неявным методом Эйлера, а второе уравнение, отвечающее $k = 1$, — методом трапеций. Тем самым, эти одношаговые неявные методы Адамса являются и методами Рунге-Кутты.

17.2 Формулы дифференцирования назад

Во всех предыдущих случаях, как при построении методов Рунге-Кутты, так и при построении методов Адамса, мы получали численные методы путем интегрирования уравнения (17.1) и замены подынтегральной функции $f(u)$ в (17.2) интерполяционным многочленом или замены интеграла квадратурной формулой. А можно поступать и иначе: интерполяционным многочленом заменить $u(t)$. Тогда для построения численного метода нужно будет выражение интерполяционного многочлена подставить в

(17.1). Чтобы получился численный метод, точка t_{n+1} должна быть в числе узлов интерполяции. Пусть

$$u(t) \approx L_k(t) = \sum_{j=0}^{k} p_j(t) u(t_{n+1-j}), \tag{17.13}$$

где

$$p_j(t) = \prod_{\substack{i=0 \\ i \neq j}}^{k} \frac{t - t_{n+1-i}}{t_{n+1-j} - t_{n+1-i}}.$$

Подставляя (17.13) в (17.1), получим приближенное равенство

$$\sum_{j=0}^{k} p_j'(t) u(t_{n+1-j}) \approx f\left(\sum_{j=0}^{k} p_j(t) u(t_{n+1-j}) \right).$$

Превратим его в точное равенство в каком-либо узле. В результате получим уравнение для определения приближенного решения. Рассмотрим случай, когда указанным узлом является t_{n+1}. Будем иметь

$$\sum_{j=0}^{k} p_j'(t_{n+1}) u_{n+1-j} = f(u_{n+1}).$$

Как и раньше, сделаем локальную замену переменной $(t - t_n)/\tau = \theta$. Тогда

$$p_j'(t) = \frac{dp_j(t)}{dt} = \frac{1}{\tau} \frac{d\hat{p}_j(\theta)}{d\theta} = \frac{1}{\tau} \hat{p}_j'(\theta),$$

где

$$p_j(t) = \hat{p}_j(\theta) = \prod_{\substack{i=0 \\ i \neq j}}^{k} \frac{\theta - 1 + i}{i - j},$$

и полученный метод принимает вид

$$\sum_{j=0}^{k} \hat{p}_j'(1) u_{n+1-j} = \tau f(u_{n+1}). \tag{17.14}$$

Определение 17.3. Численные методы (17.14) называются формулами дифференцирования назад.

Примеры. 7°. $k = 1$.

$$\begin{aligned} &\hat{p}_0(\theta) = \theta, && \hat{p}_0'(1) = 1, \\ &\hat{p}_1(\theta) = -\theta + 1, && \hat{p}_1'(1) = -1. \end{aligned}$$

8°. $k = 2$.

$$\hat{p}_0(\theta) = \frac{1}{2}\theta(\theta+1), \quad \hat{p}_0'(1) = \frac{3}{2},$$
$$\hat{p}_1(\theta) = 1-\theta^2, \qquad \hat{p}_1'(1) = -2,$$
$$\hat{p}_2(\theta) = \frac{1}{2}\theta(\theta-1), \quad \hat{p}_2'(1) = \frac{1}{2}.$$

Выпишем уравнения (17.14) для этих случаев

$$u_{n+1} - u_n = \tau f(u_{n+1}), \tag{17.15}$$

$$\left(\frac{3}{2}u_{n+1} - 2u_n + \frac{1}{2}u_{n-1}\right) = \tau f(u_{n+1}). \tag{17.16}$$

Упражнение 17.3. Построить формулу (17.14), отвечающую $k = 3$.

Ответ.
$$\left(\frac{11}{6}u_{n+1} - 3u_n + \frac{3}{2}u_{n-1} - \frac{1}{3}u_{n-2}\right) = \tau f(u_{n+1}).$$

17.3 Общие линейные многошаговые методы

Методы Адамса, явные и неявные, и формулы дифференцирования назад являются частными случаями формулы

$$\sum_{j=0}^{k} \alpha_j u_{n-j} = \tau \sum_{j=0}^{k} \beta_j f(u_{n-j}), \tag{17.17}$$

где α_j и β_j — действительные числа. (Обратим внимание на то, что в этой формуле вместо нового неизвестного u_{n+1} фигурирует u_n). Будет предполагать, что

$$\alpha_0 \neq 0, \qquad |\alpha_k| + |\beta_k| \neq 0. \tag{17.18}$$

Первое из условий (17.18) обеспечивает разрешимость неявного ($\beta_0 \neq 0$) уравнения (17.17) по крайней мере, для достаточно малого шага τ. Второе из условий (17.18) всегда можно считать выполненным, уменьшив при необходимости k.

Определение 17.4. Формула (17.17) называется линейным многошаговым (k-шаговым) методом.

Метод является явным, если $\beta_0 = 0$, и неявным в противном случае.

Чтобы линейый многошаговый метод (17.17) можно было использовать для численного решения задачи (17.1), необходимо, чтобы уравнение (17.17) аппроксимировало уравнение (17.1).

Определение 17.5. Величина

$$\psi_n = \sum_{j=0}^{k} \beta_j f(u(t_{n-j})) - \frac{1}{\tau} \sum_{j=0}^{k} \alpha_j u(t_{n-j}) \tag{17.19}$$

называется погрешностью аппроксимации метода (17.17).

Выясним вопрос о порядке погрешности аппроксимации метода (17.17) при $\tau \to 0$.

Теорема 17.1. *Многошаговый метод* (17.17) *имеет погрешщность аппроксимации порядка* $p \leqslant 2R$ *тогда и только тогда, когда выполняются следующие условия*

$$\sum_{j=0}^{k} \alpha_j = 0, \quad \sum_{j=0}^{k} \left(\alpha_j j^q + q\beta_j j^{q-1}\right) = 0, \quad q = 1, \ldots, p. \tag{17.20}$$

Доказательство. Разложим $u(t)$ по формуле Тейлора в точке t_n:

$$u(t) = \sum_{q=0}^{p} \frac{(t-t_n)^q}{q!} u^{(q)}(t_n) + O((t-t_n)^{p+1}). \tag{17.21}$$

Так как $f(u) = u'(t)$, то, дифференцируя (17.21), получим

$$f(u(t)) = \sum_{q=0}^{p} q \frac{(t-t_n)^{q-1}}{q!} u^{(q)}(t_n) + O((t-t_n)^p). \tag{17.22}$$

Подставляя теперь разложения (17.21), (17.22) при $t = t_{n-j}$ в (17.19), будем иметь

$$\begin{aligned}
\psi_n = &\sum_{j=0}^{k} \beta_j \sum_{q=0}^{p} q \frac{(-j\tau)^{q-1}}{q!} u^{(q)}(t_n) - \\
&- \frac{1}{\tau} \sum_{j=0}^{k} \alpha_j \sum_{q=0}^{p} \frac{(-j\tau)^q}{q!} u^{(q)}(t_n) + O(\tau^p) = \\
&= \sum_{q=0}^{p} \frac{(-\tau)^{q-1}}{q!} u^{(q)}(t_n) \sum_{j=0}^{k} \left[\beta_j q j^{q-1} + \alpha_j j^q\right] + O(\tau^p).
\end{aligned}$$

Приравнивая нулю коэффициентыты при τ^{q-1} для $q = 0, 1, \ldots, p$, получим (17.20). Теорема доказана.

Замечание 17.3. Решение уравнения (17.17) не изменится, если его умножить на какое-либо число, отличное от нуля. Это означает, что его коэффициенты определяются с точностью до множителя (до мультипликативной постоянной). Чтобы устранить этот произвол, пронормируем их, полагая, например,

$$\sum_{j=0}^{k} \beta_j = 1. \tag{17.23}$$

Замечание 17.4. Из (17.20), (17.23) имеем $(p+2)$ уравнения для $2(k+1)$ коэффициентов метода (17.17). Тем самым, максимальный порядок аппроксимации линейного k-шагового метода есть $p = 2k$.

17.4 Погрешность аппроксимации методов Адамса

Исследуем вопрос о порядке погрешности аппроксимации методов Адамса. Для этого перепишем сначала явный метод Адамса (17.6), (17.7) в виде (17.17), т.е. заменим $n+1$ на n:

$$u_n - u_{n-1} = \tau \sum_{j=1}^{k} b_j f(u_{n-j}).$$

Сравнивая это соотношение с (17.17), находим, что

$$\alpha_0 = 1,\ \alpha_1 = -1,\ \alpha_2 = \cdots = \alpha_k = 0,\ \beta_0 = 0,\ b_j = \beta_j,\ j = \overline{1,k}.$$

Определим, для каких дифференциальных уравнений явные методы Адамса теоретически дают точное решение в узлах сетки. Это произойдет в том случае, когда интерполяционный многочлен $L_{k-1}(t)$, определяющий явный метод Адамса, совпадает с $f(u)$ или с $f(t,u)$. Пусть $f(t,u(t)) = f(t)$, т.е. f не зависит от u и является многочленом степени не выше $k-1$. Тогда $f(t)$ совпадает со своим интерподяционным многочленом $L_{k-1}(t)$, и явный метод Адамса точен для уравнений

$$u' = qt^{q-1}, \quad q = 0, \ldots, k.$$

Это означает, что погрешность аппроксимации (17.19) на решениях этих уравнений равна нулю. Подставляя решения этих уравнений $u = t^q$ в (17.19) при $n = 0$, получим

$$\psi_0 = \sum_{j=0}^{k} \left[\beta_j q(-\tau j)^{q-1} - \frac{1}{\tau}\alpha_j(-\tau j)^q \right] = 0, \quad q = 0, \ldots, k,$$

что совпадает с первыми $(k+1)$ уравнениями (17.20). Тем самым, мы доказали, что явный k-шаговый метод Адамса имеет порядок погрешности аппроксимации не ниже k. Можно показать, что его порядок аппроксимации в точности равен k.

Упражнение 17.4. Доказать, что порядок аппроксимации неявного k-шагового метода Адамса не ниже $k+1$.

Упражнение 17.5. Доказать, что порядок аппроксимации k-шаговой формулы дифференцирования назад не ниже k.

17.5 Поучительный пример

Построим двухшаговый явный метод максимального порядка аппроксимации. Согласно ранее сказанному, порядок аппроксимации этого метода должен быть равен трем. Из (17.20), (17.23) имеем

$$\begin{aligned}
\alpha_0 + \alpha_1 + \alpha_2 &= 0,\\
\alpha_1 + 2\alpha_2 &= -(\beta_0 + \beta_1 + \beta_2),\\
\alpha_1 + 4\alpha_2 &= -2(\beta_1 + 2\beta_2),\\
\alpha_1 + 8\alpha_2 &= -3(\beta_1 + 4\beta_2),\\
\beta_0 + \beta_1 + \beta_2 &= 1,\\
\beta_0 &= 0.
\end{aligned}$$

Разрешая эту линейную систему, находим, что

$$\alpha_0 = \frac{1}{6}, \quad \alpha_1 = \frac{2}{3}, \quad \alpha_2 = -\frac{5}{6}, \quad \beta_1 = \frac{2}{3}, \quad \beta_2 = \frac{1}{3}.$$

Тем самым, метод (17.17) приобретает вид

$$\left(\frac{1}{6}u_n + \frac{4}{6}u_{n-1} - \frac{5}{6}u_{n-2}\right) = \tau\left[\frac{2}{3}f_{n-1} + \frac{1}{3}f_{n-2}\right]. \tag{17.24}$$

Применим этот метод к решению уравнения (17.1) с $f(u) = \lambda u$, где λ=const. Будем при этом предполагать, что начальное значение $u_0 = 1$. В этом случае задача (17.1) примет вид

$$u'(t) = \lambda u, \quad u(0) = 1, \tag{17.25}$$

а ее решением будет функция

$$u(t) = e^{\lambda t}. \tag{17.26}$$

Отвечающий (17.25) метод (17.17) можно записать так

$$\sum_{j=0}^{k}(\alpha_j - \tau\lambda\beta_j)u_{n-j} = 0, \tag{17.27}$$

а применительно к методу (17.24)

$$\frac{1}{6}u_n + \left(\frac{4}{6} - \frac{2}{3}\tau\lambda\right)u_{n-1} + \left(-\frac{5}{6} - \frac{1}{3}\tau\lambda\right)u_{n-2} = 0. \tag{17.28}$$

Это есть линейное однородное разностное уравнение второго порядка с постоянными коэффициентами (см. §6). Найдем его решение. Для этого нужно написать характеристическое уравнение, отвечающее разностному уравнению (17.28), и найти его корни. Искомое характеристическое уравнение имеет вид

$$q^2 + 4(1 - \tau\lambda)q - (5 + 2\tau\lambda) = 0, \tag{17.29}$$

а его корни суть

$$\begin{aligned} q_1 &= -2 + 2\tau\lambda + \sqrt{9 - 6\tau\lambda + 4\tau^2\lambda^2} = 1 + \tau\lambda + O(\tau^2\lambda^2), \\ q_2 &= -2 + 2\tau\lambda - \sqrt{9 - 6\tau\lambda + 4\tau^2\lambda^2} = -5 + O(\tau\lambda). \end{aligned} \tag{17.30}$$

Упражнение 17.6. Доказать, что $q_1 - e^{\tau\lambda} = O(\tau^4\lambda^4)$.

Поскольку корни (17.30) характеристического уравнения различны, то общее решение разностного уравнения (17.28) имеет вид

$$u_n = c_1 q_1^n + c_2 q_2^n, \tag{17.31}$$

где c_1 и c_2 — произвольные постоянные.

Рассматриваемый нами метод (17.24) является двухшаговым, и одного начального условия

$$u_0 = 1 \tag{17.32}$$

для его реализации недостаточно. Поскольку точное решение нам известно, то не будем ломать голову над тем, как задать недостающее начальное условие при $n = 1$, а просто положим

$$u_1 = u(t_1) = e^{\tau\lambda}. \tag{17.33}$$

Потребуем, чтобы решение (17.31) удовлетворяло условиям (17.32), (17.33). После простых вычислений находим, что искомое решение имеет вид

$$u_n = \frac{e^{\tau\lambda} - q_2}{q_1 - q_2} q_1^n + \frac{q_1 - e^{\tau\lambda}}{q_1 - q_2} q_2^n. \tag{17.34}$$

Изучим поведение этого решения при $n \to \infty$. Пусть $t = n\tau$ фиксировано, а $\tau \to 0$. Тогда $n = t/\tau \to \infty$. С учетом (17.30) и упражнения 17.6 находим, что

$$\begin{aligned} c_1 &= \frac{e^{\tau\lambda} - q_2}{q_1 - q_2} = \frac{1 + O(\tau) + 5}{6 + O(\tau)} = 1 + O(\tau), \\ c_2 &= \frac{q_1 - e^{\tau\lambda}}{q_1 - q_2} = \frac{O(\tau^4)}{6 + O(\tau)} = O(\tau^4). \end{aligned} \tag{17.35}$$

Далее,

$$q_1^n = \left[e^{\tau\lambda} + O(\tau^4)\right]^n = e^{\lambda\tau n}(1 + O(\tau^4))^n = e^{\lambda t}(1 + O(\tau^3)). \tag{17.36}$$

Подставляя теперь (17.35), (17.36), (17.30) в (17.34), будем иметь

$$u_n = [1 + O(\tau)]\, e^{t\lambda} + O(\tau^4)\, [-5 + O(\tau)]^n .$$

Проанализируем полученный результат. Первое слагаемое аппроксимирует решение (17.26) задачи (17.25), а второе слагаемое является паразитным. Уже при не слишком больших n это слагаемое превосходит первое, ибо

$$O(\tau^4)\, [-5 + O(\tau)]^n = O\left(\left(\frac{t}{n}\right)^4\right)\left(-5 + O\left(\frac{t}{n}\right)\right)^n .$$

Метод (17.28) сходящимся не является.

§ 18

Устойчивость многошаговых методов

18.1 Нуль-устойчивость

Обратимся к разностному уравнению (17.27) и введем следующие обозначения

$$\rho(\zeta) := \sum_{j=0}^{k} \alpha_j \zeta^{k-j}, \quad \sigma(\zeta) := \sum_{j=0}^{k} \beta_j \zeta^{k-j}. \tag{18.1}$$

Определение 18.1. Многочлены $\rho(\zeta)$ и $\sigma(\varsigma)$ из (18.1) называются соответственно первым и вторым производящими многочленами линейного многошагового метода (17.17).

Как уже было отмечено, линейный многошаговый метод (17.17) для уравнения (17.25) принимает вид линейного разностного уравнения с постоянными коэффициентами (17.27). Его характеристическое уравнение есть

$$\rho(q) - \tau\lambda\sigma(q) = 0. \tag{18.2}$$

Применительно к двушаговому методу (17.24)

$$\rho(q) = q^2 + 4q - 5,$$

а корни уравнения

$$\rho(q) = 0 \tag{18.3}$$

суть

$$q_1 = 1, \qquad q_2 = -5,$$

т.е. совпадают с главными членами корней (17.30) характеристического уравнения(17.29). Именно наличие корня q_2 и привело к неустойчивости метода (17.24). Тем самым, корни уравнения (18.3) позволяют судить об устойчивости или неустойчивости метода (17.17). А они связаны с корнями характеристического уравнения (18.2). В силу (17.18), $\alpha_0 \neq 0$ и, следовательно, степени уравнений (18.2) и (18.3) совпадают.

Поэтому характеристическое уравнение (18.2) можно рассматривать как *регулярное возмущение*(при малых $\tau\lambda$) уравнения (18.3) (объяснение терминов: коэффициенты многочлена $\rho(\zeta)$ суть пределы при $\tau\lambda \to 0$ соответствующих коэффициентов характеристического многочлена, и поэтому можно говорить о возмущении; регулярность есть следствие того, что степени возмущенного и невозмущенного многочленов совпадают). Но тогда (в силу регулярности возмущения) корни уравнения (18.3) являются пределами корней уравнения (18.2) при $\tau\lambda \to 0$. Поэтому вопрос о том, будет ли решение уравнения (17.27) неограниченно возрастать при $n \to \infty$ (и фиксированном $t = n\tau$), можно решить при анализе корней уравнения (18.3). Отметим, что уравнение (18.3) является характеристическим уравнением для разностного уравнения

$$\sum_{j=0}^{k} \alpha_j u_{n-j} = 0, \tag{18.4}$$

которое, в свою очередь, получается из (5.27), если в нем положить $\lambda = 0$. Это означает, что (18.4) есть линейный многошаговый метод для уравнения

$$u' = 0. \tag{18.5}$$

Тем самым, отбраковка "плохих"(неустойчивых) методов может быть осуществлена при анализе их свойств применительно к уравнению (18.5).

Итак, наличие у уравнения (18.3) корней, модули которых превосходят единицу, приводит к неустойчивости. Однако опасность представляют не только такие корни, но и корни, равные по модулю единице, если они кратные. В самом деле, пусть q_1 — корень характеристического уравнения (18.3) кратности $s > 1$ такой, что $|q_1| = 1$. Тогда сеточная функция

$$P_{s-1}(n)q_1^n$$

будет растущим решение уравнения (18.4), в то время как решением уравнения (18.5), которое и аппроксимирует изучаемое уравнение (18.4), есть постоянная.

Определение 18.2. Говорят, что линейный многошаговый метод (17.17) удовлетворяет корневому условию, если
1) все корни первого производящего многочлена (18.1) расположены в единичном круге $|\zeta| \leqslant 1$;
2) нули $\rho(\zeta)$, расположенные на единичной окружности $|\zeta| = 1$ простые.

Определение 18.3. Линейный многошаговый метод (17.17), удовлетворяющий корневому условию, называется нуль-устойчивым (устойчивым).

Замечание 18.1. Если линейный многошаговый метод (17.17) аппроксимирует какое-либо дифференциальное уравнение, то среди нулей $\rho(\zeta)$ обязательно есть $\zeta = 1$, о чем свидетельствует первое из условий (17.20), являющее собой условие $\rho(1) = 0$.

Примеры. 1° Явный и неявный методы Адамса. В обоих случаях $\alpha_0 = 1$, $\alpha_1 = -1$, а остальные $\alpha_j = 0$. Поэтому

$$\rho(q) = q^k - q^{k-1}$$

и, следовательно,

$$q_1 = 1, \quad q_2 = \cdots = q_k = 0.$$

Методы Адамса нуль-устойчивы.

2° Двухшаговая формула дифференцирования назад (17.16).

$$\rho(q) = \frac{3}{2}q^2 - 2q + \frac{1}{2},$$
$$q_1 = 1, \quad q_2 = 1/3.$$

Метод нуль-устойчив.

3° Трехшаговая формула дифференцирования назад (17.3)

$$\rho(q) = \frac{11}{6}q^3 - 3q^2 + \frac{3}{2}q - \frac{1}{3}.$$

Хотя это и многочлен третьей степени, нули его легко находятся, ибо один из его нулей есть $q_1 = 1$. Деля $\rho(q)$ на $(q-1)$, приходим к уравнению

$$\frac{11}{6}q^2 - \frac{7}{6}q + \frac{1}{3} = 0$$

с корнями

$$q_{2,3} = \frac{7 \pm i\sqrt{39}}{22}.$$

Отсюда

$$\left|q_{2,3}\right|^2 = \frac{2}{11} < 1.$$

Метод нуль-устойчив.

Теорема 18.1 (Первый барьер Далквиста). *Порядок p устойчивого линейного k-шагового метода подчиняется следующим ограничениям:*
$p \leqslant k$ *для явных методов;*
$p \leqslant k+1$ *для неявных методов при нечетном k;*
$p \leqslant k+2$ *для неявных методов при четном k.*

В качестве иллюстрации первого утверждения теоремы может служить построенный нами в предыдущем параграфе явный двухшаговый метод максимального порядка аппроксимации $p = 3$, который оказался неустойчивым.

Упражнение 18.1. Построить общий явный устойчивый двухшаговый метод максимального порядка аппроксимации.

Ответ: α_0 — параметр метода,

$$\alpha_1 = 1 - 2\alpha_0, \qquad \alpha_2 = \alpha_0 - 1,$$
$$\beta_0 = 0, \qquad \beta_1 = \frac{1}{2} + \alpha_0, \qquad \beta_2 = \frac{1}{2} - \alpha_0.$$

Условие устойчивости: $1/2 \leqslant \alpha_0 < \infty$. При $\alpha_0 = 1$ имеем явный метод Адамса, при $\alpha_0 = 1/2$ — метод прямоугольников с шагом $\tau' = 2\tau$. При $\alpha_0 = 1/6$ метод имеет погрешность аппроксимации $O(\tau^3)$, но неустойчив.

Упражнение 18.2. Построить устойчивый двухшаговый метод максимального порядка аппроксимации.

Ответ:

$$\alpha_0 = 1/2, \quad \alpha_1 = 0, \quad \alpha_2 = -1/2,$$
$$\beta_0 = 1/6, \quad \beta_1 = 2/3, \quad \beta_2 = 1/6.$$

Этот метод иногда называется методом Симпсона (по аналогии с одноименной квадратурной формулой). Метод имеет четвертый порядок аппроксимации.

18.2 Жесткие задачи

При определении нуль-устойчивости многошагового метода мы могли ограничиться изучением простейшего дифференциального уравнения (18.5), ибо производящий многочлен $\rho(\zeta)$ из (18.1) многошагового метода (17.17), от расположения нулей которого зависит, будет ли метод устойчивым или нет, является характеристическим многочленом именно в применении к уравнению (18.5). Условие нуль-устойчивости предъявляет минимальные требования к численному методу, производя лишь грубую отбраковку абсолютно непригодных для вычислений методов. По существу, нуль-устойчивость метода обеспечивает лишь ограниченность приближенного решения для конечного временного интервала $[0, T]$ при $n \to \infty$.

Однако имеются задачи, отыскание решений которых при помощи только нуль-устойчивых методов оказывается весьма затруднительным, если не невозможным. Проще всего объяснить возникающие трудности не на примере одного уравнения, а на примере систем уравнений.

Рассмотрим однородную систему линейных дифференциальных уравнений с постоянными коэффициентами

$$\mathbf{u}' = A\mathbf{u}, \qquad \mathbf{u}(0) = \mathbf{u}_0, \tag{18.6}$$

где $\mathbf{u} = [u_1\, u_2]^T$, а

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Найдем и проанализируем решение задачи (18.6). Как обычно, будем его искать в виде

$$\mathbf{u}(t) = \boldsymbol{\xi} e^{\lambda t}, \tag{18.7}$$

где $\boldsymbol{\xi}$ — двумерный числовой вектор, а λ — постоянная. Подставляя (18.7) в (18.6), находим, что

$$\lambda \boldsymbol{\xi} e^{\lambda t} = e^{\lambda t} A \boldsymbol{\xi},$$

а, сокращая на $e^{\lambda t}$, получим следующую задачу на собственные значения:

$$A\boldsymbol{\xi} = \lambda \boldsymbol{\xi}. \tag{18.8}$$

Будем предполагать, что A — матрица простой структуры, т.е. у нее имеется полный набор собственных векторов. Тогда

$$A\boldsymbol{\xi}_1 = \lambda_1 \boldsymbol{\xi}_1, \qquad A\boldsymbol{\xi}_2 = \lambda_2 \boldsymbol{\xi}_2$$

и $\boldsymbol{\xi}_1$ и $\boldsymbol{\xi}_2$ линейно независимы.

В рассматриваемом случае общее решение системы (18.6) принимает вид

$$\mathbf{u}(t) = c_1 \boldsymbol{\xi}_1 e^{\lambda_1 t} + c_2 \boldsymbol{\xi}_2 e^{\lambda_2 t}, \tag{18.9}$$

а решение задачи Коши (18.6) получается отсюда при значениях c_1 и c_2, найденных из алгебраической системы

$$\boldsymbol{\xi}_1 c_1 + \boldsymbol{\xi}_2 c_2 = u_0. \tag{18.10}$$

Будем для простоты предполагать, что собственные числа λ_1 и λ_2 действительны. Более существенным для нас будет предположение об их отрицательности

$$\lambda_1 < 0, \qquad \lambda_2 < 0. \tag{18.11}$$

В силу сделанных предположений модули компонент u_1 и u_2 решения (18.9) будет стремиться к нулю при $t \to \infty$.

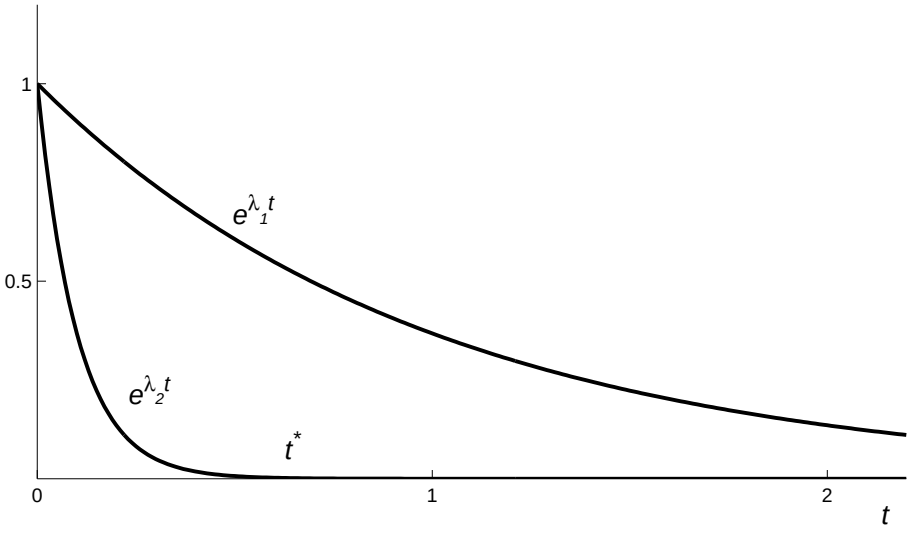

Рис. 1

Предположим теперь дополнительно, что

$$\lambda_1 = O(1), \qquad |\lambda_2| \gg |\lambda_1|. \tag{18.12}$$

Так как в этом случае $e^{\lambda_2 t}$ убывает значительно быстрее $e^{\lambda_1 t}$, то через некоторое время t^* составляющая $c_2\boldsymbol{\xi}_2 e^{\lambda_2 t}$ решения (18.9) будет практически равной нулю, и решение будет почти полностью определяться составляющей $c_1\boldsymbol{\xi}_1 e^{\lambda_1 t}$. (см. рис. 1)

В рассматриваемой ситуации естественно было бы ожидать, что и у численного решения задачи (18.6) модули компонент хотя бы не возрастали.

Применим для решения задачи (18.6) метод Эйлера

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} = A\mathbf{u}^n, \qquad \mathbf{u}^0 = \mathbf{u}_0. \tag{18.13}$$

Найдем решение задачи (18.13). Искать его будем в виде (см. (6.30))

$$\mathbf{u}^n = \boldsymbol{\xi} q^n, \qquad q = \text{const} \neq 0. \tag{18.14}$$

Подставляя (18.14) в (18.13), получим

$$q^n \frac{q-1}{\tau}\boldsymbol{\xi} = q^n A\boldsymbol{\xi},$$

а после сокращения на q^n обнаруживаем, что для отыскания $\boldsymbol{\xi}$ имеем задачу (18.8) с $\lambda = (q-1)/\tau$. Поэтому $q = 1 + \tau\lambda$, и решение задачи (18.13) есть

$$\mathbf{u}^n = c_1\boldsymbol{\xi}_1(1+\tau\lambda_1)^n + c_2\boldsymbol{\xi}_2(1+\tau\lambda_2)^n, \tag{18.15}$$

где c_1, c_2 — решение системы (18.10).

Чтобы модули компонент решения (18.15) не возрастали при $n \to \infty$, необходимо и достаточно, чтобы выполнялись условия

$$|1+\tau\lambda_1| \leqslant 1, \qquad |1+\tau\lambda_2| \leqslant 1,$$

что вместе с (18.11) и (18.12) приводит к условию

$$\tau \leqslant 2/|\lambda_2| \ll 1. \tag{18.16}$$

Ограничение (18.16), вообще говоря, является довольно жестким. Если при $t \leqslant t^*$ это ограничение вполне разумно, и даже из соображений аппроксимации и точности нужно требовать $\tau \ll 2/|\lambda_2|$, то при $t > t^*$, когда вторая составляющая каждой компоненты решения (18.15) вроде бы не должна поставлять новой информации, и желательно было бы увеличить шаг τ с той целью, чтобы сэкономить ресурсы и не воспроизводить первую составляющую с излишней точностью. Но тогда придется нарушить условие (18.16), что приведет к резкому возрастанию второй составляющей решения и полной потере точности.

Определение 18.4. Система дифференциальных уравнений (18.6) с постоянной матрицей A порядка m называется жесткой, если
1° $\operatorname{Re}\lambda_j < 0 \quad , j = 0, \ldots, m,$
2° отношение

$$S = \frac{\max\limits_j |Re\,\lambda_j|}{\min\limits_j |Re\,\lambda_j|} \gg 1. \tag{18.17}$$

Определение 18.5. Число S из (18.17) называется коэффициентом жесткости задачи (18.6).

Замечание 18.2. Для линейной системы с матрицей A, зависящей от t, коэффициент жесткости также зависит от t, и, если он велик для каких-либо t из интересующего нас интервала, то система жесткая. Для нелинейных систем жесткость определяется в окрестности какого-либо решения при помощи соответствующей матрицы Якоби.

Применим теперь для решения задачи (18.6) неявный метод Эйлера

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} = A\mathbf{u}^{n+1}.$$

Подставляя сюда (18.14), находим, что

$$q^{n+1}\frac{1-q^{-1}}{\tau}\boldsymbol{\xi} = q^{n+1}A\boldsymbol{\xi},$$

т.е. $\lambda\tau = (1 - q^{-1})$, $q = (1 - \tau\lambda)^{-1}$ и

$$\mathbf{u}^n = c_1\boldsymbol{\xi}_1(1 - \tau\lambda)^{-n} + c_2\boldsymbol{\xi}_2(1 - \tau\lambda_2)^{-n}.$$

Очевидно, что при выполнении условий (18.11) модули компонент $\mathbf{u}^n$ монотонно убывают при $n \to \infty$ при *любых* τ, и, следовательно, τ можно выбирать только из соображений точности.

Неявный метод Эйлера при решении жестких систем оказался существенно более устойчивым, чем просто метод Эйлера.

Как отобрать методы, пригодные для решения жестких задач? Ужесточить требование устойчивости.

18.3 A-устойчивость

Если при определении нуль-устойчивости основной моделью было уравнение (18.5), то теперь следует обратиться к уравнению (17.25). Многошаговый метод (17.17) в применении к линейному однородному уравнению (17.25) имеет вид (17.27), а характеристическое уравнение этого разностного уравнения задается соотношением (18.2).

Определение 18.6. Линейный многошаговый метод (17.17) в применении к уравнению (17.25) называется абсолютно устойчивым для данного λ и данного τ, если при указанном значении $\tau\lambda$ все корни характеристического уравнения (18.2) расположены внутри единичного круга.

Определение 18.7. Множество всех точек комплексной плоскости $\tau\lambda$, для которых линейный многошаговый метод (17.17) в применении к (17.25) абсолютно устойчив, называется областью абсолютной устойчивости метода.

Пример 4°. Метод Эйлера (15.7). Единственный корень характеристического уравнения $q = 1 + \tau\lambda$. Условие абсолютной устойчивости

$$|1 + \tau\lambda| \leqslant 1.$$

Областью абсолютной устойчивости является единичный круг с центром в точке $\tau\lambda = -1$. (см. рис. 2)

Пример 5°. Неявный метод Эйлера (15.8). Условие абсолютной устойчивости

$$|q| = |1 - \tau\lambda|^{-1} \leqslant 1, \quad \text{т.е.} \quad |1 - \tau\lambda| \geqslant 1.$$

Областью абсолютной устойчивости является внешность единичного круга с центром в точке $\tau\lambda = 1$.(см. рис. 3)

Определение 18.8. Линейный многошаговый метод (17.17) называется A-устойчиваым, если область его абсолютной устойчивости содержит левую полуплоскость $\mathrm{Re}(\tau\lambda) < 0$.

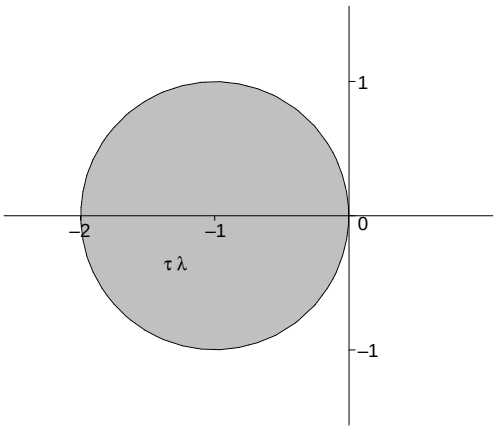

Рис. 2

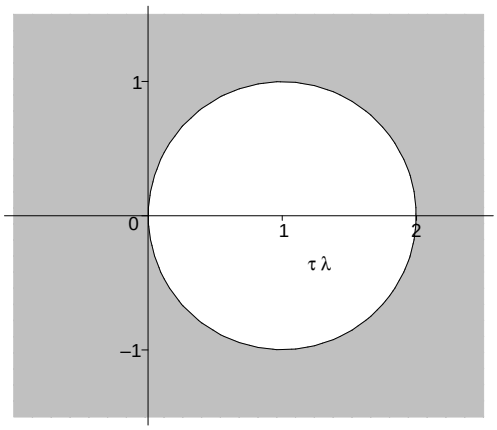

Рис. 3

Из приведенных примеров следует, что метод Эйлера не является A-устойчивым, а неявный метод Эйлера A-устойчив.

Пример 6°. Метод трапеций (15.12). Применительно к уравнению (17.25) этот метод имеет вид

$$\frac{u^{n+1}-u^n}{\tau}=\lambda\frac{u^{n+1}+u^n}{2},$$

а его характеристическое уравнение есть $(q-1)/\tau=\lambda(q+1)/2$. Отсюда находим единственный корень

$$q=\frac{1+\tau\lambda/2}{1-\tau\lambda/2}$$

и условие абсолютной устойчивости

$$|q|=\left|\frac{1+\tau\lambda/2}{1-\tau\lambda/2}\right|\leqslant 1$$

или

$$\left|1+\tau\lambda/2\right|\leqslant\left|1-\tau\lambda/2\right|.$$

Пусть $\tau\lambda=x+iy$. Тогда условие абсолютной устойчивости примет вид

$$\left|1+\frac{x}{2}+i\frac{y}{2}\right|\leqslant\left|1-\frac{x}{2}-i\frac{y}{2}\right|.$$

или

$$\left(1+\frac{x}{2}\right)^2+\frac{y^2}{4}\leqslant\left(1-\frac{x}{2}\right)^2+\frac{y^2}{4}.$$

Раскрывая скобки, находим, что условие абсолютной устойчивости есть

$$x=\mathrm{Re}(\tau\lambda)<0.$$

Областью абсолютной устойчивости метода трапеций является левая полуплоскость $\mathrm{Re}(\tau\lambda)<0$ (Рис. 4). Метод A-устойчив.

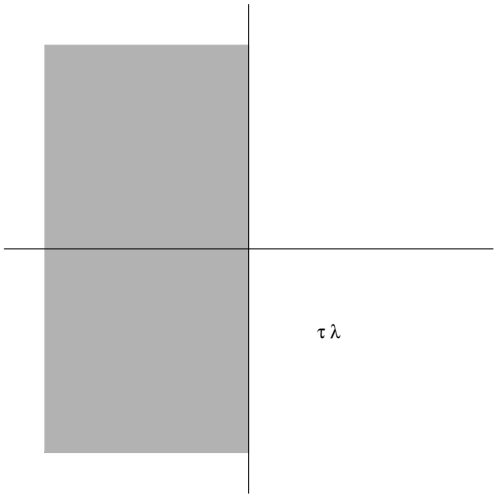

Рис. 4

Теорема 18.2. *Среди линейных многошаговых методов* (17.17) *не существует явных A-устойчивых методов.*

Теорема 18.3. *Среди неявных линейных многошаговых методов* (17.17) *не существует A-устойчивых методов, имеющих порядок точности выше второго.*

Пример 7°. Двухшаговая формула дифференцирования назад. Этот метод задается соотношением (17.16)

$$\left(\frac{3}{2}u_{n+1} - 2u_n + \frac{1}{2}u_{n-1}\right) = \tau f(u_{n+1}). \tag{18.18}$$

Характеристическое уравнение, отвечающее этому методу в применении к уравнению (17.25) есть

$$\frac{3}{2}q^2 - 2q + \frac{1}{2} - \tau\lambda q^2 = 0. \tag{18.19}$$

Определим область абсолютной устойчивости этого метода. Для этого достаточно найти ее границу, т.е. такое множество комплексной плоскости $z = \tau\lambda$, где $|q(z)| = 1$. С этой целью выразим из (18.19) $\tau\lambda$ через q

$$z = \frac{3}{2} - \frac{2}{q} + \frac{1}{2q^2}. \tag{18.20}$$

Поскольку нас интересуют значения $|q| = 1$, то пусть $q = e^{-i\varphi}$. Отсюда и из (18.20)

$$z = \frac{3}{2} - 2e^{i\varphi} + \frac{1}{2}e^{2i\varphi}. \tag{18.21}$$

При изменении аргумента φ от 0 до 2π точка z из (18.21) описывает замкнутую кривую, симметричную относительно действительной оси (функция $\sin k\varphi$ — нечетная), которая и является границей области абсолютной устойчивости.

$$\begin{aligned} z &= \frac{3}{2} - 2\cos\varphi + \frac{1}{2}\cos 2\varphi + i(-2\sin\varphi + \frac{1}{2}\sin 2\varphi) = \\ &= \frac{3}{2} - 2\cos\varphi + \cos^2\varphi - \frac{1}{2} + i(-2\sin\varphi + \sin\varphi\cos\varphi) = \\ &= (1-\cos\varphi)^2 \pm i\sqrt{1-\cos^2\varphi}(2-\cos\varphi) = \\ &= (1-t)^2 \pm i\sqrt{1-t^2}(2-t), \qquad t = \cos\varphi. \end{aligned}$$

Отсюда следует, что

$$\operatorname{Re} z = (1-t)^2 \geqslant 0,$$

и, следовательно, кривая расположена в правой полуплоскости. Построим ее. Мнимая часть $z(t)$ обращается в нуль при $t = \pm 1$. Действительная часть $z(t)$ при этих значениях параметра равна 0 и 4.

Исследования показывают, что

$$\max_{[-1,1]} \operatorname{Im} z(t) = \operatorname{Im} z\left(\frac{1-\sqrt{3}}{2}\right) = \frac{(3+\sqrt{3})\sqrt[4]{3}}{2\sqrt{2}} \approx 2.20,$$
$$\operatorname{Re} z\left(\frac{1-\sqrt{3}}{2}\right) = \frac{2+\sqrt{3}}{2} \approx 1.86.$$

Из (18.20) находим, что при

$$|q| \to \infty, \qquad z \to \frac{3}{2} \in G,$$

и, следовательно, внутренность области — область неустойчивости. Тем самым, вне G (Рис. 5) $|q| < 1$, и метод абсолютно устойчив, а, следовательно, и A-устойчив. Этот метод второго порядка точности.

Пример 8°. Трехшаговая формула дифференцирования назад. (Упражнение 17.3)

$$\frac{11}{6}u_{n+1} - 3u_n + \frac{3}{2}u_{n-1} - \frac{1}{3}u_{n-2} = \tau\lambda u_{n+1}. \tag{18.22}$$

Характеристическое уравнение этого разностного уравнения имеет вид

$$\frac{11}{6}q^3 - 3q^2 + \frac{3}{2}q - \frac{1}{3} = \tau\lambda q^3. \tag{18.23}$$

Снова положим $|q| = 1$, т.е. $q = e^{-i\varphi}$ и $\tau\lambda = z$. Тогда

$$z = \frac{11}{6} - 3e^{i\varphi} + \frac{3}{2}e^{2i\varphi} - \frac{1}{3}e^{3i\varphi}.$$

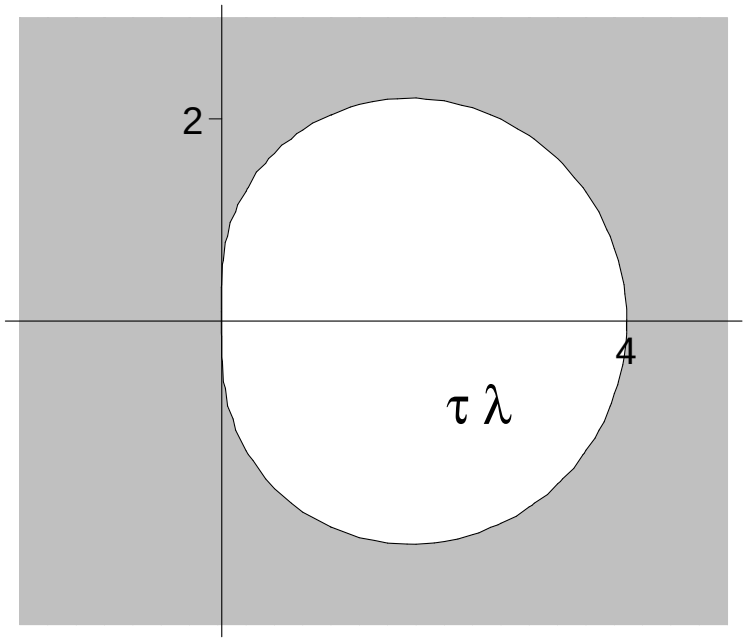

Рис. 5

Обозначая $\cos\varphi = t$, после простых вычислений находим, что

$$z = -\frac{1}{3}(t-1)^2(4t-1) \pm \frac{i}{3}\sqrt{1-t^2}(4t^2-9t+8).$$

При $t = \pm 1$ $\operatorname{Im} z = 0$, а $\operatorname{Re} z = 0$ или $20/3$. Исследования показывают, что $\operatorname{Re} z$ как функция t принимает экстремальные значения при $t = 1/2$ и $t = 1$. Значение $t = 1$ мы уже рассмотрели, а

$$\operatorname{Re} z\,(1/2) = \min \operatorname{Re}(t) = -1/12, \quad \operatorname{Im} z(1/2) = \pm 3\sqrt{3}/4 \approx \pm 1.30$$

и, следовательно, часть границы области устойчивости расположена в левой полуплоскости. Как легко видеть, мнимую ось граница устойчивости пересекает при $t = 1/4$ и

$$\operatorname{Im} z(1/4) = \pm\sqrt{15}/2 \approx \pm 1.94.$$

Экстремальные значения $\operatorname{Im} z(t)$ принимает в точке

$$t^* = -\frac{1}{2}\left[(2+\sqrt{3})^{1/3} + (2+\sqrt{3})^{-1/3} - 1\right] \approx -0.60,$$

причем

$$\operatorname{Im} z(t^*) \approx \pm 3.96, \quad \operatorname{Re} z(t^*) \approx 2.89.$$

Область устойчивости изображена на рис. 6.

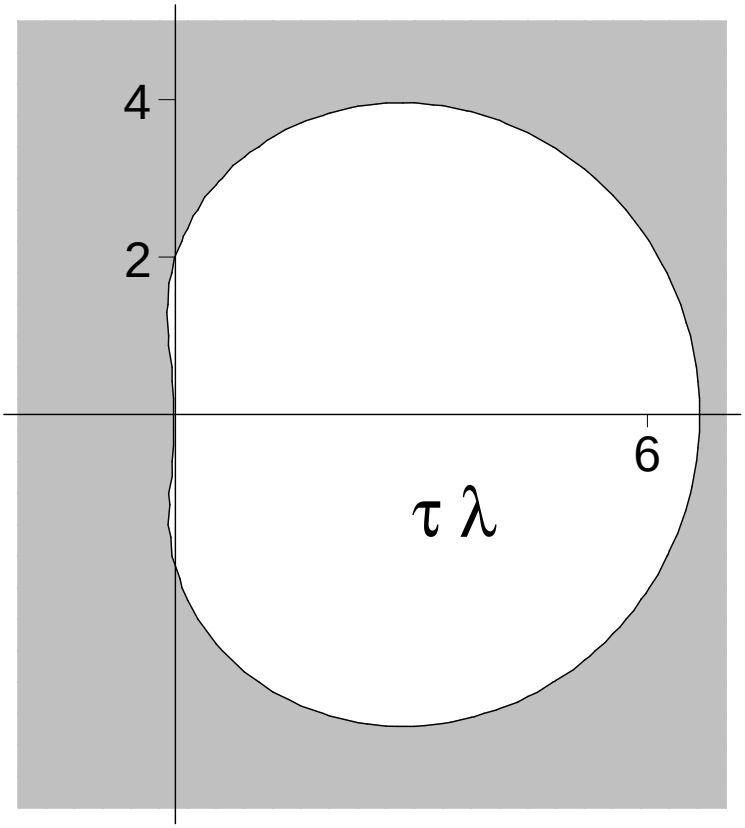

Рис. 6.

Определение 18.9. Линейный многошаговый метод называется $A(\alpha)$-устойчивым, если его область абсолютной устойчивости содержит угол

$$\big|\arg(-\tau\lambda)\big| < \alpha.$$

Замечание 18.3. $A(\pi/2)$- и A- устойчивости совпадают.

Теорема 18.4. *Существуют многошаговые методы 3-го и 4-го порядков точности $A(\alpha)$-устойчивые при любых $\alpha < \pi/2$.*

Теорема 18.5. *Явные линейные многошаговые методы не являются $A(\alpha)$ - устойчивыми ни при каких α.*

Теорема 18.6. *Методы дифференцирования назад при $k \leqslant 6$ являются $A(\alpha)$ - устойчивыми при соответствующих значениях $\alpha \neq 0$.*

Упражнение 18.3. Исследовать область абсолютной устойчивости двухшагового неявного метода Адамса.

18.4 Устойчивость методов Рунге-Кутты

Как было уже отмечено, методы (все) Рунге-Кутты являются нуль-устойчивыми. Исследуем области абсолютной устойчивости некоторых из этих методов. Рассмотрим явный трехэтапный метод третьего порядка, задаваемый таблицей (16.44), которая имеет вид

$$\begin{array}{c|ccc} 1/2 & 1/2 & & \\ 1 & -1 & 2 & \\ \hline & 1/6 & 2/3 & 1/6 \end{array}$$

Применительно к уравнению (17.25) этот метод задается следующими соотношениями

$$\begin{aligned} Y_1 &= u_n, \\ Y_2 &= u_n + \frac{\tau\lambda}{2} Y_1, \\ Y_3 &= u_n - \tau\lambda Y_1 + 2\tau\lambda Y_2, \\ u_{n+1} &= u_n + \tau\lambda \left(\frac{1}{6} Y_1 + \frac{2}{3} Y_2 + \frac{1}{6} Y_3 \right). \end{aligned}$$

Исключая из этих соотношений промежуточные величины Y_1, Y_2 и Y_3, будем иметь

$$\begin{aligned} Y_2 &= \left(1 + \frac{\tau\lambda}{2} \right) u_n, \\ Y_3 &= \left[1 - \tau\lambda + 2\tau\lambda \left(1 + \frac{\tau\lambda}{2} \right) \right] u_n = (1 + \tau\lambda + \tau^2\lambda^2) u_n, \\ u_{n+1} &= \left\{ 1 + \tau\lambda \left[\frac{1}{6} + \frac{2}{3} \left(1 + \frac{\tau\lambda}{2} \right) + \frac{1}{6} (1 + \tau\lambda + \tau^2\lambda^2) \right] \right\} u_n = \\ &= \left(1 + \tau\lambda + \frac{\tau^2\lambda^2}{2} + \frac{\tau^3\lambda^3}{6} \right) u_n. \end{aligned}$$

Это есть линейное разностное уравнение первого порядка, единственный корень характеристического уравнения которого равен

$$q = 1 + \tau\lambda + \frac{\tau^2\lambda^2}{2} + \frac{\tau^3\lambda^3}{6} = e^{\tau\lambda} + O(\tau^4\lambda^4).$$

Обозначим $\tau\lambda$ через z. Тогда

$$q = 1 + z + \frac{z^2}{2} + \frac{z^3}{6}.$$

Этот корень характеристического уравнения есть многочлен третьей степени от z и в левой полуплоскости $\operatorname{Re} z < 0$ ограниченным быть не может. Метод не является $A(\alpha)$ -устойчивым ни при каком α.

Рассмотрим теперь двухэтапный метод третьего порядка, задаваемый таблицей (16.33), которая имеет вид

$$\begin{array}{l|ll} \theta_1=\gamma & \gamma & 0 \\ \theta_2=1-\gamma & 1-2\gamma & \gamma \\ \hline & 1/2 & 1/2 \end{array} \qquad \gamma=\frac{3\pm\sqrt{3}}{6}.$$

Применительно к уравнению (17.25) этот метод записывается следующим образом

$$\begin{aligned} Y_1 &= u_n+\gamma\tau\lambda Y_1,\\ Y_2 &= u_n+\tau\lambda(1-2\gamma)Y_1+\tau\lambda\gamma Y_2,\\ u_{n+1} &= u_n+\frac{\tau\lambda}{2}(Y_1+Y_2). \end{aligned}$$

Как и в предыдущем примере, положим $\tau\lambda=z$ и исключим Y_1 и Y_2. Решая систему линейных алгебраических уравнений второго порядка относительно Y_1 и Y_2 (первые два уравнения) и подставляя результат в третье уравнение, находим, что

$$\begin{aligned} Y_1 &= \frac{u_n}{1-\gamma z}, \quad Y_2=\frac{1+(1-3\gamma)z}{(1-\gamma z)^2}u_n,\\ u_{n+1} &= \left[1+\frac{z}{2}\left(\frac{1}{1-\gamma z}+\frac{1+(1-3\gamma)z}{(1-\gamma z)^2}\right)\right]u_n. \end{aligned}$$

Отсюда следует, что единственным корнем характеристического уравнения является

$$\begin{aligned} q &= \frac{1-2\gamma z+\gamma^2z^2+z/2-\gamma z^2/2+z/2+(1-3\gamma)z^2/2}{(1-\gamma z)^2}=\\ &= \frac{1+(1-2\gamma)z+(\gamma^2-2\gamma+1/2)z^2}{1-2\gamma z+\gamma^2z^2}=\frac{P(z)}{Q(z)}. \end{aligned}$$

Этот корень является дробно-рациональной функцией, полюсом второго порядка которой является точка $z=\gamma^{-1}=(3\mp\sqrt{3})/6$, расположенная в правой полуплоскости. В левой полуплоскости эта функция аналитична и, следовательно, максимум ее модуля здесь не превосходит максимума модуля на границе, т.е. при $z=iy$. Оценим ее модуль на мнимой оси. Имеем

$$\begin{aligned} |P(iy)|^2 &= \left[1-(\gamma^2-2\gamma+1/2)y^2\right]^2+(1-2\gamma)^2y^2=\\ &= 1-2(\gamma^2-2\gamma+1/2)y^2+(\gamma^2-2\gamma+1/2)^2y^4+(1-4\gamma+4\gamma^2)y^2=\\ &= 1+2\gamma^2y^2+(\gamma^2-2\gamma+1/2)^2y^4 \end{aligned}$$

и

$$|Q(iy)|^2=(1-\gamma^2y^2)^2+4\gamma^2y^2=1+2\gamma^2y^2+\gamma^4y^4.$$

Отсюда

$$\left|\frac{P(iy)}{Q(iy)}\right|^2=\frac{1+2\gamma^2y^2+(\gamma^2-2\gamma+1/2)^2y^4}{1+2\gamma^2y^2+\gamma^4y^4}.$$

Добавим к числителю и вычтем из него $\gamma^4 y^4$, после чего выделим единицу

$$\left|\frac{P(iy)}{Q(iy)}\right|^2 = 1 + \frac{(\gamma^2 - 2\gamma + 1/2 - \gamma^2)(\gamma^2 - 2\gamma + 1/2 + \gamma^2)y^4}{(1+\gamma^2 y^2)^2}.$$

Подставим вместо γ его значения из (16.33). Найдем, что

$$-2\gamma + 1/2 = \frac{-3 \mp 2\sqrt{3}}{6},$$

а

$$2\gamma^2 - 2\gamma + 1/2 = \frac{9 + 3 \pm 6\sqrt{3}}{18} + \frac{-3 \mp 2\sqrt{3}}{6} = \frac{1}{6.}$$

Поэтому

$$\left|\frac{P(iy)}{Q(iy)}\right|^2 = 1 - \frac{3 \pm 2\sqrt{3}}{36}\frac{y^4}{(1+\gamma^2 y^2)^2}.$$

Поскольку это выражение не меньше нуля, а при $\gamma = (3+\sqrt{3})/6$ (верхний знак в коэффициенте у второго слагаемого) вычитаемое неотрицательно, то в рассматриваемом случае

$$\left|\frac{P(iy)}{Q(iy)}\right| \leqslant 1,$$

и изучаемый метод является A-устойчивым. При $\gamma = (3-\sqrt{3})/6$ вычитаемое отрицательно, и поэтому $|P(iy)/Q(iy)| > 1$ для $y \neq 0$. В этом случае метод A-устойчивым не является. Области абсолютной устойчивости этих методов изображены на рисунках 7 и 8, соответственно.

Тем самым, один из методов (16.33), именно, отвечающий $\gamma = (3+\sqrt{3})/6$, является A-устойчивым, в то время как второй таким свойством не обладает и даже не является $A(\alpha)$-устойчивым.

Упражнение 18.4. Доказать, что неявный двухэтапный метод Рунге-Кутты четвертого порядка (оптимальный двухэтапный метод) (16.34) является $A(\alpha)$-устойчивым.

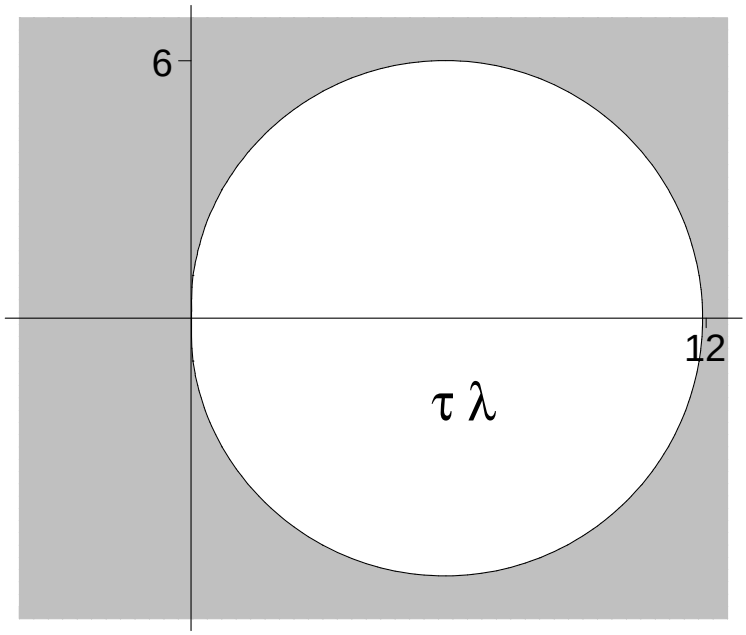

Рис. 7

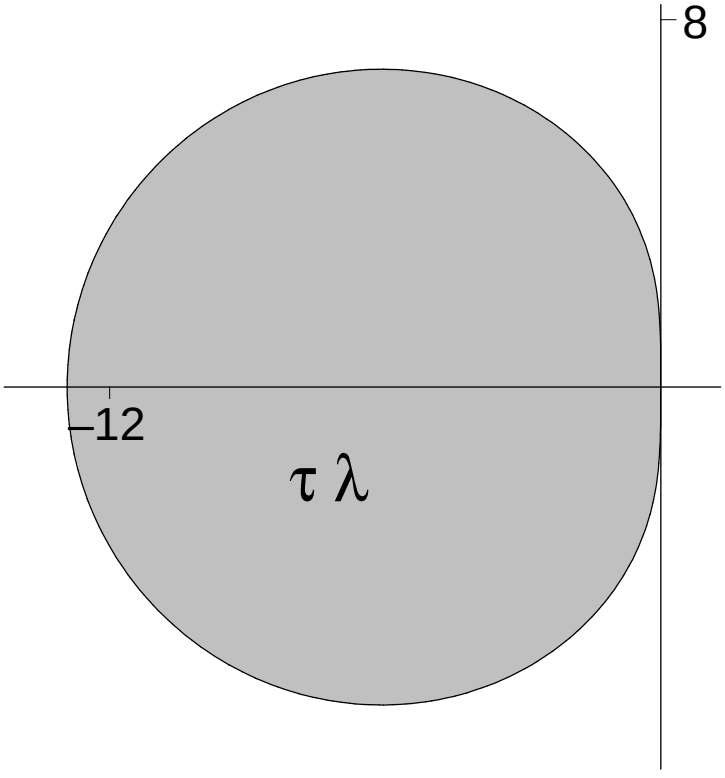

Рис. 8

Глава V

Численные методы решения краевых задач для обыкновенных дифференциальных уравнений

§ 19

Элементы теории разностных схем

19.1 Введение

Простейшим содержательным примером краевой задачи для обыкновенного дифференциального уравнения является следующий

$$-u''(x) = f(x), \qquad 0 < x < l, \tag{19.1}$$

$$u(0) = g_0, \qquad u(1) = g_1. \tag{19.2}$$

У краевой задачи, в отличие от задачи Коши, дополнительные условия, выделяющие единственное решение уравнения (19.1), задаются не в одной точке, а в нескольких (обычно в двух), и называются краевыми (или граничными) условиями. Это вносит дополнительные трудности в процесс решения задачи.

Мы будем изучать разностные методы решения краевых задач. Для этого на отрезке $[0, l]$ введем сетку

$$\overline{\omega} := \left\{ x = x_i = ih \;\middle|\; i = 0, \ldots, N \right\}.$$

Точки x_i будем называть узлами сетки, а число $h = l/N$ — ее шагом. Введенная сетка является равномерной. Если бы расстояния между узлами менялось при переходе от одного узла к другому, то сетка была бы неравномерной.

Суть разностных методов решения краевых задач для дифференциальных уравнений состоит в том, что производные, входящие в дифференциальное уравнение и граничные условия, заменяются подходящими разностными отношениями. В результате краевая задача заменяется (аппроксимируется) системой алгебраических (линейных, если исходная задача была линейной) уравнений, решение которой и принимается за приближенное решение краевой задачи.

Напомним простейшие аппроксимации первой и второй производных

$$\frac{u(x_i)-u(x_{i-1})}{h} = u'(x_i)+O(h), \tag{19.3}$$

$$\frac{u(x_{i+1})-u(x_i)}{h} = u'(x_i)+O(h), \tag{19.4}$$

$$\frac{u(x_{i+1})-u(x_{i-1})}{2h} = u'(x_i)+O(h^2), \tag{19.5}$$

$$\frac{-u(x_{i+2})+4u(x_{i+1})-3u(x_i)}{2h} = u'(x_i)+O(h^2), \tag{19.6}$$

$$\frac{u(x_{i+1})-2u(x_i)+u(x_{i-1})}{h^2} = u''(x_i)+O(h^2). \tag{19.7}$$

Для справедливости соотношений (19.3) и (19.4) достаточно, чтобы $u(x)\in C^2$, для справедливости (19.5) и (19.6) — $u(x)\in C^3$, для справедливости (19.7) — $u(x)\in C^4$. В этом можно убедиться путем разложения левых частей (19.3)-(19.7) в точке $x=x_i$ по формуле Тейлора.

Упражнение 19.1. Убедиться в справедливости (19.3)-(19.7).

Замечание 19.1. Если функцию $u(x)$ заменить интерполяционным многочленом Лагранжа первой степени по узлам x_{i-1} и x_i или x_i и x_{i+1}, а затем его продифференцировать, то получим левые части соотношений (19.3), (19.4). Заменяя $u(x)$ интерполяционным многочленом второй степени по узлам x_{i-1},x_i,x_{i+1} или x_i,x_{i+1},x_{i+2}, дифференцируя полученный интерполянт и полагая $x=x_i$, получим левые части (19.5) и (19.6), соответственно.

Воспользуемся соотношением (19.7) для замены второй производной в (19.1) разностным отношением

$$-\frac{u(x_{i+1})-2u(x_i)+u(x_{i-1})}{h^2}\approx f(x_i), \qquad x_i=h,2h,\dots,l-h.$$

Превратим приближенные равенства в точные путем замены точного решения $u(x_i)$ в узле x_i на приближенное u_i^h:

$$-\frac{u_{i+1}^h-2u_i^h+u_{i-1}^h}{h^2}=f_i, \qquad i=\overline{1,N-1}, \tag{19.8}$$

Это есть система $(N-1)$ линейных алгебраических уравнений с $N+1$ неизвестными $u_0^h,u_1^h,\dots,u_N^h$. Система (19.8) недоопределена (как и следовало ожидать). Воспользуемся граничными условиями (19.2) и положим

$$u_0^h=g_0, \qquad u_N^h=g_1. \tag{19.9}$$

Решение системы (19.8), (19.9), если оно существует, будем называть приближенным решением задачи (19.1), (19.2).

19.2 Основные понятия теории разностных схем

Обозначим дифференциальное выражение, стоящее в левой части (19.1), через Lu. Тогда дифференциальное уравнение (19.1) примет вид

$$Lu = f(x), \qquad 0 < x < l. \tag{19.10}$$

Граничные условия (19.2) запишем в виде

$$lu = g. \tag{19.11}$$

Аналогично, разностное выражение, стоящее в левой части (19.8), обозначим через $L^h u^h$. Тогда из (19.8) будем иметь

$$L^h u_i^h = f_i^h, \qquad i = \overline{1, N-1}, \tag{19.12}$$

где $f_i^h = f_i$. Граничные условия (19.9) запишем в виде, аналогичном (19.11)

$$l^h u^h = g^h. \tag{19.13}$$

Определение 19.1. Сеточная функция

$$\Psi_v(x) := L^h v - Lv, \qquad x \in \omega, \tag{19.14}$$

определенная на сетке ω, где v — достаточно гладкая функция, заданная на $[0, l]$, называется погрешностью аппроксимации дифференциального выражения Lv разностным выражением $L^h v$.

Определение 19.2. Разностное выражение $L^h v$ аппроксимирует дифферециальное выражение Lv, если погрешность аппроксимации $\Psi_v \to 0$ (в каком-нибудь смысле) при $h \to 0$.

Определение 19.3. Сеточная функция

$$z = u^h - u, \qquad x \in \overline{\omega}, \tag{19.15}$$

где u^h — решение задачи (19.12), (19.13), а u — решение задачи (19.10), (19.11), называется погрешностью решения.

Сформулируем задачу для погрешности решения z. Подставим в (19.12), (19.13) u^h, выражаемое из (19.15) через z и u: $u^h = z + u$. Будем иметь

$$L^h z = f^h - L^h u, \quad l^h z = g^h - l^h u. \tag{19.16}$$

Определение 19.4. Функция

$$\Psi = f^h - L^h u, \qquad x \in \omega, \tag{19.17}$$

являющаяся правой частью уравнения для погрешности решения (19.16), называется погрешностью аппроксимации уравнения (19.10) уравнением (19.12).

Определение 19.5. Функция

$$\psi = g^h - l^h u, \tag{19.18}$$

являющаяся правой частью в граничных условиях для погрешности решения (19.16), называется погрешностью граничных условий (19.11) граничными условиями (19.13).

Замечание 19.2. Так как в силу (19.10) $Lu - f = 0$, то, добавляя этот нуль к представлению погрешности аппроксимации (19.17), будем иметь

$$\Psi = f^h - L^h u = f^h - f - (L^h u - Lu) = (f^h - f) - \Psi_u, \tag{19.19}$$

где Ψ_u определяется соотношением (19.14). Тем самым, погрешность аппроксимации уравнения представляет собой разность между погрешностью аппроксимации правой части и погрешностью аппроксимации дифференциального выражения. Аналогичные представления имеют место и для погрешности аппроксимации граничных условий:

$$\psi = g^h - l^h u = g^h - g - (l^h u - lu) = (g^h - g) - \psi_u. \tag{19.20}$$

Определение 19.6. Задача (19.12), (19.13) аппроксимирует задачу (19.10), (19.11), если Ψ и ψ стремятся к нулю при $h \to 0$ вместе с Ψ_u и ψ_u.

Определение 19.7. Решение задачи (19.12), (19.13) сходится к решению задачи (19.10), (19.11), если $z \to 0$ (в каком-либо смысле) при $h \to 0$.

Определение 19.8. Задача (19.12), (19.13) аппроксимирует задачу (19.10), (19.11) с погрешностью порядка $n > 0$, если

$$\|\Psi_u\|_{(1)} = o(1), \quad \|\psi_u\|_{(2)} = o(1), \quad \|\Psi\|_{(1)} = O(h^n), \quad \|\psi\|_{(2)} = O(h^n)$$

Определение 19.9. Решение задачи (19.12), (19.13) сходится к решению задачи (19.10), (19.11) со скоростью $O(h^n)$, если

$$\|z\|_{(3)} = O(h^n).$$

Проиллюстрируем введенные понятия на примере задачи (19.1), (19.2). Так как в данном случае $L = -d^2v/dx^2$, а

$$L^h v = -\frac{v(x_{i+1}) - 2v(x_i) + v(x_{i-1})}{h^2},$$

то, в силу (19.7),

$$\Psi_v = O(h^2),$$

т.е. дифференциальное выражение v'' аппроксимируется разностным выражением $(v_{i+1} - 2v_i + v_{i-1})/h^2$ на функциях $v(x) \in C^4$ с погрешностью $O(h^2)$.

Далее, так как $f_i^h = f(x_i)$, то с учетом (19.19) заключаем, что дифференциальное уравнение (19.1) аппроксимируется разностным уравнением (19.8) с погрешностью $O(h^2)$, если $u(x) \in C^4[0,l]$.

Наконец,

$$lu = \{u(0), u(1)\},$$
$$l^h u = \{u_0, u_N\},$$
$$g = \{g_0, g_1\} = g^h,$$

так что

$$\psi = g^h - l^h u = 0.$$

Итак, задача (19.8), (19.9) аппроксимирует задачу (19.1), (19.2) (при $u(x) \in C^4[0,l]$) с погрешностью $O(h^2)$.

Очевидно, что, если вместо уравнения (19.1) рассмотреть уравнение

$$L_1 u := -u''(x) + q(x)u(x) = f(x), \qquad x \in (0,1) \tag{19.21}$$

и аппроксимировать его разностным уравнением

$$L_1^h u^h := -\frac{u_{i+1}^h - 2u_i^h + u_{i-1}^h}{h^2} + q(x_i)u_i^h = f(x_i), \quad i = \overline{1, N-1} \tag{19.22}$$

то задача (19.22), (19.9) будет аппроксимировать задачу (19.21), (19.2) тоже с погрешностью $O(h^2)$.

19.3 Разрешимость и сходимость

Исследуем вопрос о сходимости решения разностной задачи к решению задачи дифференциальной. Для уравнения (19.21) это сделать несколько проще, чем для уравнения (19.1). Поэтому к нему мы и обратимся. Но сначала установим существование и единственность решения задачи (19.22), (19.9).

Теорема 19.1. *Если*

$$q(x) \geqslant c_1 > 0, \quad 0 < x < 1, \tag{19.23}$$

то решение задачи (19.22), (19.9) *существует, единственно, и для него справедлива априорная оценка*

$$\max_i |u_i^h| \leqslant |g_0| + |g_1| + \max_i \frac{|f_i|}{c_1}. \tag{19.24}$$

Доказательство. Задача (19.22), (19.9) представляет собой систему линейных алгебраических уравнений с квадратной матрицей порядка $(N+1)$. Поэтому всегда существует такая правая часть $[g_0, f_1, \ldots, f_{N-1}, g_1]$ этой системы (берется первое уравнение из (19.9), затем последовательно все уравнения (19.22) и, наконец, второе уравнение (19.9)), что решение u^h существует. Например, возьмем произвольный набор чисел

$u_0^h, u_1^h, \ldots, u_N^h$ и подставим его в левые части (19.22), (19.9). Этим мы определим правые части (19.22), (19.9), при которых решение заведомо существует.

Получим априорную оценку этого решения. Пусть

$$\max_i |u_i^h| = |u_{i_0}^h|.$$

Если $i_0 = 0$ или $i_0 = N$, то в силу (19.9)

$$\max_i |u_i^h| \leqslant \max\{|g_0|, |g_1|\} \leqslant |g_0| + |g_1|, \tag{19.25}$$

что согласуется с (19.24). В противном случае максимум модуля достигается во внутреннем узле $x_{i_0} \in \omega$. Запишем уравнение (19.22) в этом узле

$$-\frac{u_{i_0-1}^h - 2u_{i_0}^h + u_{i_0+1}^h}{h^2} + q_{i_0} u_{i_0}^h = f_{i_0}.$$

Если $u_{i_0}^h \geqslant 0$, то

$$-[\underset{\substack{\wedge\\ 0}}{(u_{i_0-1}^h - u_{i_0}^h)} + \underset{\substack{\wedge\\ 0}}{(u_{i_0+1}^h - u_{i_0}^h)}] \geqslant 0$$

и, следовательно,

$$q_{i_0} u_{i_0}^h \leqslant f_{i_0}.$$

Отсюда с учетом (19.23)

$$0 \leqslant u_{i_0}^h \leqslant \frac{f_{i_0}}{q_{i_0}} \leqslant \frac{1}{c_1} \max_i |f_i|. \tag{19.26}$$

Если же $u_{i_0}^h < 0$, то

$$-[\underset{\substack{\vee\\ 0}}{(u_{i_0-1}^h - u_{i_0}^h)} + \underset{\substack{\vee\\ 0}}{(u_{i_0+1}^h - u_{i_0}^h)}] \leqslant 0 \tag{19.27}$$

и, следовательно,

$$q_{i_0} u_{i_0}^h \geqslant f_{i_0}.$$

Отсюда

$$-q_{i_0} |u_{i_0}^h| \geqslant f_{i_0}$$

и снова

$$|u_{i_0}^h| \leqslant -\frac{f_{i_0}}{q_{i_0}} \leqslant \frac{1}{c_1} \max_i |f_i|. \tag{19.28}$$

Собирая оценки (19.25), (19.26), (19.28), приходим к (19.24). Априорная оценка получена.

Докажем теперь, что решение единственно. Допустим противное, т.е. допустим существование двух решений $u_{(1)}^h$ и $u_{(2)}^h$. Очевидно, что их разность $z = u_{(1)}^h - u_{(2)}^h$ удовлетворяет однородному уравнению (19.22) и однородным граничным условиям (19.9). В силу априорной оценки (19.24)

$$\max_i |z_i| \leqslant 0.$$

Следовательно, $z_i \equiv 0$, что противоречит предположению. Мы доказали, что однородная система (19.22), (19.9) имеет лишь тривиальное решение. Следовательно, матрица этой системы невырождена, и задача (19.22), (19.9) имеет единственное решение при любых g_0, g_1 и f_i. Теорема доказана.

Теорема 19.2. *Если выполнено условие* (19.23), *и решение* $u(x)$ *задачи* (19.21), (19.2) *принадлежит* $C^4[0,l]$, *то решение* u^h *задачи* (19.22), (19.9) *сходится к решению задачи* (19.21), (19.2) *со скоростью* $O(h^2)$, *т.е.*

$$\left|u(x_i) - u_i^h\right| = O(h^2).$$

Доказательство. Напишем задачу для погрешности решения $z_i = u_i^h - u(x_i)$. Будем иметь

$$-\frac{z_{i+1} - 2z_i + z_{i-1}}{h^2} + q_i z_i = \Psi_i, \qquad z_0 = z_N = 0. \tag{19.29}$$

К задаче (19.29)применим теорему 19.1, в силу которой

$$\max_i |z_i| \leqslant \frac{1}{c_1} \max_i |\Psi_i|.$$

Но в силу вышедоказанного $\Psi_i = O(h^2)$, что и доказывает теорему.

Замечание 19.3. Более детальный анализ показывает, что

$$\max_i |u_i^h - u(x_i)| \leqslant \frac{1}{c_1} \max_{x\in[0,l]} |u^{IV}(x)| \frac{h^2}{12}.$$

Теорема 19.3 (О монотонности). *Если выполнено условие*

$$q_i \geqslant 0, \qquad i = \overline{1, N-1}, \tag{19.30}$$

а сеточная функция U_i, $i = \overline{0, N}$ *такова, что*

$$U_0 \geqslant 0, \qquad U_N \geqslant 0 \tag{19.31}$$

и

$$L_1^h U_i \geqslant 0, \qquad i = \overline{1, N-1}, \tag{19.32}$$

то

$$U_i \geqslant 0, \qquad i = \overline{1, N-1}. \tag{19.33}$$

Доказательство. Допустим противное, т.е. допустим, что функция U_i может принимать отрицательные значения. Тогда существует такой узел x_{i_0}, $i_0 \in \{1, 2, \ldots, N–1\}$, что

$$\min_i U_i = U_{i_0} < 0 \tag{19.34}$$

и в силу (19.27)

$$-(U_{i_0-1} - 2U_{i_0} + U_{i_0+1}) \leqslant 0.$$

Исследуем обе эти возможности. Если $-(U_{i_0-1} - 2U_{i_0} + U_{i_0+1}) < 0$, то с учетом (19.30) и (19.34)

$$L_1^h U_{i_0} = -U_{i_0-1} - 2U_{i_0} + U_{i_0+1} + q_{i_0}U_{i_0} < 0,$$

и мы пришли к противоречию с (19.32). Если $(U_{i_0-1} - 2U_{i_0} + U_{i_0+1}) = 0$, а $q_{i_0} \neq 0$, мы снова получаем противоречие. Для выхода из этих противоречий мы должны предположить, что $q_{i_0} = 0$ и $(U_{i_0-1} - 2U_{i_0} + U_{i_0+1}) = 0$. Но в силу (19.27), (19.34) это означает, что $U_{i_0-1} = U_{i_0} = U_{i_0+1} < 0$, и в качестве i_0 из (19.34) можно взять также $(i_0 - 1)$ или $(i_0 + 1)$. Делая этот выбор, мы теми же рассуждениями приходим к утверждению, что и $U_{i_0-2} = U_{i_0}$ (или $U_{i_0+2} = U_{i_0}$). И т.д. Поскольку в силу (19.31), (19.34) функция U_i, $i = \overline{0, N}$ не является постоянной, то существует такой узел x_{i_1}, $i_1 \in \{1, 2, \ldots, N-1\}$, что $U_{i_1} = U_{i_0}$, а U_{i_1-1} или U_{i_1+1} больше U_{i_1}. В этом узле $-(U_{i_1-1} - 2U_{i_1} + U_{i_1+1}) < 0$, и мы вернулись к уже рассмотренному случаю, который привел нас к противоречию с (19.32). Все противоречия снимаются, если мы откажемся от предположения, что U_i может принимать отрицательные значения. Теорема доказана.

Определение 19.10. Матрица A называется монотонной, если любой вектор x, для которого $Ax \geqslant 0$, является неотрицательным.

Теорема 19.4 (Принцип сравнения). *Пусть u_i^h — решение задачи* (19.22), (19.9), *а U_i — решение следующей задачи:*

$$L_1^h U_i = F_i, \quad i = \overline{1, N-1}, \quad U_0 = G_0, \quad U_N = G_1.$$

Пусть

$$|f_i| \leqslant F_i, \quad |g_0| \leqslant G_0, \quad |g_1| \leqslant G_1. \tag{19.35}$$

Тогда, если выполнено условие (19.30), *то*

$$|u_i^h| \leqslant U_i, \qquad i = \overline{1, N-1}. \tag{19.36}$$

Доказательство. Легко видеть, что функция $(U_i - u_i^h)$ является решением задачи

$$L_1^h (U - u^h)_i = F_i - f_i, \quad i = \overline{1, N-1}, \quad U_0 - u_0^h = G_0 - g_0,$$
$$U_N - u_N^h = G_1 - g_1.$$

В силу (19.35) и теоремы 19.3 заключаем, что $U_i - u_i^h \geqslant 0$. Из аналогичных соображений находим, что и $U_i + u_i^h \geqslant 0$. Тем самым, $-U_i \leqslant u_i^h \leqslant U_i$, и теорема доказана.

Замечание 19.4. Функция U_i из (19.36) называется барьером.

Теорема 19.5. *Для решения задачи* (19.22), (19.9) *при выполнении условия* (19.30) *справедлива априорная оценка*

$$\max_i |u_i^h| \leqslant |g_0| + |g_1| + \frac{l^2}{8} \max_i |f_i|.$$

Доказательство. Введем в рассмотрение функцию

$$U_i = |g_0|(1-x_i) + x_i|g_1| + c\,x_i(1-x_i) \geqslant 0, \tag{19.37}$$

где $c > 0$ — некоторая постоянная. Очевидно, что $U_0 = |u_0^h|$, $U_N = |u_N^h|$. Легко проверить, что

$$L_1^h U_i = 2c + q_i U_i =: F_i \geqslant 2c.$$

Пусть $c = 1/2 \max\limits_i |f_i|$. Тогда $|f_i| \leqslant F_i$, и мы находимся в условиях теоремы 19.4, т.е. $|u_i^h| \leqslant U_i$. Но

$$\max_i U_i \leqslant |g_0| + |g_1| + c/4.$$

Теорема доказана.

Упражнение 19.2. Сформулировать и доказать теорему о скорости сходимости разностной задачи (19.22), (19.9).

19.4 Уравнения с переменными коэффициентами

Рассмотрим общее самосопряженное уравнение второго порядка

$$-\frac{d}{d\,x}\left(p(x)\frac{d\,u}{d\,x}\right) + q(x)u = f(x), \qquad 0 < x < 1. \tag{19.38}$$

и изучим вопрос о его аппроксимации. На первый взгляд кажется вполне естественным раздифференцировать первое слагаемое левой части (19.38)

$$-p(x)\frac{d^2\,u}{d\,x^2} - p'(x)\frac{d\,u}{d\,x} + q(x)u = f(x) \tag{19.39}$$

и в этом виде заменить d^2u/dx^2 и du/dx соответствующими разностными отношениями. Но так поступать плохо в силу целого ряда причин. В частности, уравнение (19.38) является формально самосопряженным по Лагранжу (симметричным, т.е. если $Lv := -(pv')' + qv$, а $u(x)$ и $v(x)$ обращаются в нуль при $x = 0$ и $x = 1$, то $\int\limits_0^1 vLu\,d\,x = \int\limits_0^1 uLv\,d\,x$. Сравнить с симметричной матрицей $A = A^T$ — $(Ax, y) = (x, Ay)$). Если же аппроксимировать (19.39), которое эквивалентно (19.38) при гладкой $p(x)$, то аппроксимация, вообще говоря, симметричной не будет. Уравнение (19.38) нужно аппроксимировать сразу в исходном виде.

Построим аппроксимацию (19.38) при помощи интегро - интерполяционного метода (метода баланса, метода конечных объемов). Пусть $x_{i\pm1/2} = x_i \pm h/2$. Проинтегрируем уравнение (19.38) по отрезку $(x_{i-1/2}, x_{i+1/2})$. Будем иметь

$$\begin{aligned} &- p(x_{i+1/2})u'(x_{i+1/2}) + p(x_{i-1/2})u'(x_{i-1/2}) + \\ &+ \int_{x_{i-1/2}}^{x_{i+1/2}} [q(x)u(x) - f(x)]\,d\,x = 0. \end{aligned} \tag{19.40}$$

Заменим в (19.40) интеграл квадратурной формулой прямоугольников, а производные — соответствующими разностными отношениями. Именно

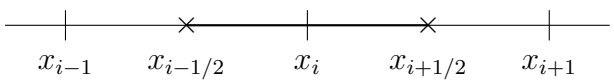

Рис. 1

$$\begin{gathered}\int\limits_{x_{i-1/2}}^{x_{i+1/2}} [q(x)u - f(x)]\,dx \approx q_i u_i h - f_i h,\\ u'_{i+1/2} \approx \frac{u_{i+1}-u_i}{h}, \quad u'_{i-1/2} \approx \frac{u_i - u_{i-1}}{h}.\end{gathered} \tag{19.41}$$

Подставляя (19.41) в (19.40), получим приближенное равенство. Заменяя приближенное равенство на точное, получим уравнение для приближенного решения. После деления на h оно примет вид:

$$-\frac{1}{h}\left[p_{i+1/2}\frac{u^h_{i+1}-u^h_i}{h} - p_{i-1/2}\frac{u^h_i - u^h_{i-1}}{h}\right] + q_i u^h_i = f_i, \quad i = \overline{1, N-1} \tag{19.42}$$

Введем следующие обозначения

$$\begin{aligned} u_x &:= u_{x,i} := \frac{u_{i+1}-u_i}{h} \text{ — правое разностное отношение,}\\ u_{\bar{x}} &:= u_{\bar{x},i} := \frac{u_i - u_{i-1}}{h} \text{ — левое разностное отношение.}\end{aligned}$$

Очевидно, что $v_{x,i} \equiv v_{\bar{x},i+1}$. Далее,

$$\begin{aligned}\frac{v_{i+1}-2v_i+v_{i-1}}{h^2} &= \frac{1}{h}\left[\frac{v_{i+1}-v_i}{h} - \frac{v_i - v_{i-1}}{h}\right] =\\ &= \frac{1}{h}\left(v_{x,i} - v_{\bar{x},i}\right) = \frac{1}{h}\left(v_{\bar{x},i+1} - v_{\bar{x},i}\right) = (v_{\bar{x}})_{x,i} =\\ &= v_{\bar{x}x,i} =: v_{\bar{x}x}.\end{aligned}$$

Используя введенные обозначения, уравнение (19.42) можно переписать так:

$$-\left(p^h u^h_{\bar{x}}\right)_{x,i} + q^h_i u^h_i = f^h_i, \quad i = \overline{1, N-1}, \tag{19.43}$$

где

$$p^h := p^h_i := p\left(x_i - \frac{h}{2}\right), \quad q^h := q^h_i := q(x_i), \quad f^h := f^h_i := f(x_i). \tag{19.44}$$

19.5 Аппроксимация граничных условий

Применим теперь интегро-интерполяционный метод для построения аппроксимации граничного условия, содержащего производную. Пусть для уравнения (19.38) в точке $x = 0$ (граничной точке) задано граничное условие

$$\alpha \frac{du(0)}{dx} + \beta u(0) = \gamma. \tag{19.45}$$

Граничное условие (19.45) содержит в себе все основные граничные условия для уравнения (19.38): именно, граничные условия первого рода ($\alpha = 0$), второго рода ($\beta = 0$) и третьего рода. Нас будут интересовать граничные условия второго и третьего рода, т.е. условия, содержащие производную. Простейшая аппроксимация условия (19.45) имеет вид

$$\alpha \frac{u_1^h - u_0^h}{h} + \beta u_0^h = \gamma. \tag{19.46}$$

Упражнение 19.3. Доказать, что погрешность аппроксимации граничного условия (19.45) граничным условием (19.46) при $\alpha \neq 0$ есть $O(h)$.

Мы не будем заниматься этой аппроксимацией из-за того, что она имеет большую погрешность. Построим другую аппроксимацию условия (19.45), но прежде его несколько преобразуем. По предположению $\alpha \neq 0$, и на этот коэффициент условие (19.45) можно разделить. Коэффициент $p(x)$ уравнения (19.38) будем предполагать строго положительным

$$p(x) \geqslant c_0 > 0, \tag{19.47}$$

и домножение (19.45) на $-p(0)$ приведет к эквивалентному уравнению. Будем вместо (19.45) рассматривать граничное условие

$$-p(0) \frac{du(0)}{dx} + \varkappa_0 u(0) = g_0, \tag{19.48}$$

которое при $\alpha = -p(0) \neq 0$, $\beta = \varkappa_0$ и $\gamma = g_0$ совпадает с (19.45). Комбинация $p(0)u'(0)$ в (19.48) хороша уже тем, что величина $-p(x)u'(x)$ имеет смысл потока и фигурирует в самом уравнении (19.38). Знак минус перед производной должен свидетельствовать о том, что производная берется по "внешней нормали": производная $du(0)/dx$ вычислена по направлению внутрь отрезка $[0, 1]$, а производная $-du(0)/dx$ — по направлению, выходящему из отрезка.

Чтобы построить аппроксимацию (19.48), проинтегрируем уравнение (19.38) по отрезку $(0, h/2)$. Будем иметь

$$-p(h/2) \frac{du(h/2)}{dx} + p(0) \frac{du(0)}{dx} + \int_0^{h/2} [q(x)u(x) - f(x)]\, dx = 0. \tag{19.49}$$

Затем выразим $p(0)du(0)/dx$ из (19.48)

$$p(0)\frac{du(0)}{dx} = \varkappa_0 u(0) - g_0, \tag{19.50}$$

аппроксимируем производную

$$\frac{du(h/2)}{dx} \approx \frac{u_1 - u_0}{h} \tag{19.51}$$

и аппроксимируем интеграл в (19.49) квадратурной формулой "левых прямоугольников"

$$\int_0^{h/2} [q(x)u(x) - f(x)]\,dx \approx [q(0)u(0) - f(0)]\frac{h}{2}. \tag{19.52}$$

Подставляя теперь (19.50)-(19.52) в (19.49), получим приближенное равенство, которое превратим в точное путем замены точного решения $u(x)$ на приближенное $u^h(x)$. Будем иметь

$$-p_{1/2}\frac{u_1^h - u_0^h}{h} + \left(\varkappa_0 + \frac{h}{2}q_0\right)u_0^h = g_0 + \frac{h}{2}f_0$$

или, принимая обозначения (19.44),

$$-p_1^h u_{\bar{x},1}^h + (\varkappa_0 + \frac{h}{2}q_0^h)u_0^h = g_0 + \frac{h}{2}f_0^h. \tag{19.53}$$

Соотношение (19.53) представляет собой искомую аппроксимацию.

19.6 Исследование погрешности аппроксимации

Исследуем погрешность аппроксимации разностной схемы (19.43). Исследуем даже более общую схему. Пусть разностная схема имеет вид

$$-\frac{1}{h}\left[b_i u_{x,i}^h - a_i u_{\bar{x},i}^h\right] + q_i^h u_i^h = f_i^h. \tag{19.54}$$

Погрешность аппроксимации этой схемы есть

$$\begin{aligned}\Psi_i &= f_i^h + \frac{1}{h}\left[b_i u_{x,i} - a_i u_{\bar{x},i}\right] - q_i^h u_i = \\ &= [f_i^h - f(x_i)] - [q_i^h - q(x_i)]u_i + \\ &+ \frac{1}{h}\left[b_i u_{x,i} - a_i u_{\bar{x},i}\right] - (pu')'_i.\end{aligned} \tag{19.55}$$

При $u(x) \in C^4[0,1]$ имеют место следующие разложения

$$\begin{aligned}u_{x,i} &= u'_i + \frac{h}{2}u''_i + \frac{h^2}{6}u'''_i + O(h^3), \\ u_{\bar{x}.i} &= u'_i - \frac{h}{2}u''_i + \frac{h^2}{6}u'''_i + O(h^3).\end{aligned}$$

Подставляя эти соотношения в (19.55), будем иметь

$$\begin{aligned}\Psi_i = \frac{1}{h}\Big[&b_i(u_i' + \frac{h}{2}u_i'' + \frac{h^2}{6}u_i''' + O(h^3)) - \\ &- a_i(u_i' - \frac{h}{2}u_i'' + \frac{h^2}{6}u_i''' + O(h^3))\Big] - \\ &- (p'u' + pu'') - \\ &- [q_i^h - q(x_i)]u_i + [f_i^h - f(x_i)] = \\ &= \left(\frac{b_i - a_i}{h} - p_i'\right)u_i' + \left(\frac{b_i + a_i}{2} - p_i\right)u_i'' + \\ &+ h\frac{b_i - a_i}{6}u_i''' + O(h^2) - (q_i^h - q_i)u_i + \\ &+ (f_i^h - f_i).\end{aligned}$$

Отсюда находим, что для аппроксимации $O(h^2)$ необходимо и достаточно выполнения условий

$$\begin{aligned}&1^\circ. \quad \frac{b_i - a_i}{h} - p_i' = O(h^2), \\ &2^\circ. \quad \frac{b_i + a_i}{2} - p_i = O(h^2), \\ &3^\circ. \quad q_i^h - q_i = O(h^2), \\ &4^\circ. \quad f_i^h - f_i = O(h^2).\end{aligned} \tag{19.56}$$

Для схемы (19.43), (19.44) условия (19.56_3) и (19.56_4) очевидны. Обратимся к (19.56_1) и (19.56_2). Имеем

$$\begin{aligned}b_i = p_{i+1/2} = p_i + \frac{h}{2}p_i' + \frac{h^2}{8}p_i'' + O(h^3), \\ a_i = p_{i-1/2} = p_i - \frac{h}{2}p_i' + \frac{h^2}{8}p_i'' + O(h^3).\end{aligned}$$

Отсюда

$$\frac{b_i - a_i}{h} = p_i' + O(h^2), \qquad \frac{b_i + a_i}{2} = p_i + O(h^2).$$

Теорема 19.6. *Если решение уравнения* (19.38) *обладает четвертыми непрерывными производными, то разностная схема* (19.43), (19.44) *имеет погрешность аппроксимации* $O(h^2)$.

Упражнение 19.4. Доказать, что разностная схема (19.43) при $b_i = a_{i+1}$ и

$$\text{а)} \quad a_i = \frac{p_i + p_{i-1}}{2}, \quad q_i^h = q_i, \quad f_i^h = f_i, \tag{19.57}$$

$$\text{б)} \quad a_i = \frac{1}{h}\int\limits_{x_{i-1}}^{x_i} p(x)\,dx, \quad q_i^h = \frac{1}{h}\int\limits_{x_{i-1}}^{x_{i+1}} q(x)(1-|x-x_i|)\,dx,$$

$$f_i^h = \frac{1}{h}\int\limits_{x_{i-1}}^{x_{i+1}} f(x)(1-|x-x_i|)\,dx \tag{19.58}$$

имеет погрешность аппроксимации $O(h^2)$.

Исследуем погрешность аппроксимации ψ_0 граничного условия (19.53). Имеем

$$\psi_0 := g_0 + \frac{h}{2}f_0 + p_{1/2}\frac{u_1-u_0}{h} - (\varkappa_0 + \frac{h}{2}q_0)u_0 =$$
$$= g_0 + \frac{h}{2}f_0 + \left(p_0 + \frac{h}{2}p_0' + O(h^2)\right)\left(u_0' + \frac{h}{2}u_0'' + O(h^2)\right) - (\varkappa_0 + \frac{h}{2}q_0)u_0 =$$
$$= (p_0u_0' - \varkappa_0 u_0 + g_0) + \frac{h}{2}\left(p_0u_0'' + p_0'u_0' - q_0u_0 + f_0\right) + O(h^2).$$

Первая скобка в этом представлении равна нулю в силу (19.48), а вторая — в силу уравнения (19.38), продолженного по непрерывности с $(0,1)$ на $[0,1)$. Тем самым, погрешность аппроксимации граничного условия (19.53) *на решении уравнения* (19.38) есть $O(h^2)$.

Упражнение 19.5. Интегро-интерполяционным методом построить аппроксимацию граничного условия

$$p(1)\frac{du(1)}{dx} + \varkappa_1 u(1) = g_1 \tag{19.59}$$

и исследовать погрешность полученной аппроксимации.

Теорема 19.7. *Пусть выполнены условия*

$$p_i^h \geqslant c_0 > 0, \quad q_i^h \geqslant c_1 > 0, \quad \varkappa_0 > 0. \tag{19.60}$$

Тогда существует единственное решение задачи (19.43), (19.53), (19.61)

$$u_N^h = g_1, \tag{19.61}$$

и для него справедлива априорная оценка

$$\max_i |u_i^h| \leqslant \frac{|g_0|}{\varkappa_0} + |g_1| + \max_i \frac{|f_i|}{c_1}. \tag{19.62}$$

Упражнение 19.6. Доказать теорему 19.7.

Теорема 19.8. *Если выполнены условия* (19.60), *и решение задачи* (19.38), (19.48), (19.63) $u(x) \in C^4[0,1]$,

$$u(1) = g_1, \tag{19.63}$$

то решение u^h задачи (19.43), (19.44), (19.53), (19.61) *сходится к решению задачи* (19.38), (19.48), (19.63) *со скоростью $O(h^2)$ равномерно по $x_1 \in \omega$, т.е.*

$$\max_i |u(x_i) - u_i^h| = O(h^2).$$

19.7 Некоторые обобщения

Для квазилинейного уравнения

$$-\frac{d}{dx}\left(p(x,u)\frac{du}{dx}\right)+q(x,u)=0 \tag{19.64}$$

разностную аппроксимацию можно взять в виде

$$\begin{aligned}&-\frac{1}{h}\left[p\left(x_{i+1/2},\frac{u_{i+1}^h+u_i^h}{2}\right)u_{x,i}^h-p\left(x_{i-1/2},\frac{u_i^h+u_{i-1}^h}{2}\right)u_{\bar{x},i}^h\right]+\\&+q(x_i,u_i^h)=0,\quad i=1,\dots,N-1.\end{aligned} \tag{19.65}$$

С равным успехом можно поступить и так:

$$\begin{aligned}&-\frac{1}{h}\left[\frac{p(x_{i+1},u_{i+1}^h)+p(x_i,u_i^h)}{2}u_{x,i}^h-\frac{p(x_i,u_i^h)+p(x_{i-1},u_{i-1}^h)}{2}u_{\bar{x},i}^h\right]+\\&+q(x_i,u_i^h)=0,\quad i=1,\dots,N-1.\end{aligned} \tag{19.66}$$

Упражнение 19.7. Выяснить порядки погрешности аппроксимации схем (19.65) и (19.66).

Если решения уравнения (19.38) не являются достаточно гладкими, то теорема 19.8 о сходимости со скоростью $O(h^2)$ может не иметь место. В этой ситуации для уменьшения погрешности аппроксимации в окрестности тех точек, где уменьшается гладкость решения, полезно использовать неравномерную сетку. Пусть

$$\widehat{\omega}=\left\{x_i \;\middle|\; x_0=0<x_1<x_2<\dots<x_{N-1}<x_N=1\right\} \tag{19.67}$$

— произвольная неравномерная сетка на $[0,1]$. Будем обозначать

$$h_i=x_i-x_{i-1},\qquad \hbar_i=\frac{h_i+h_{i+1}}{2}.$$

На сетке (19.67) для уравнения (19.38) методом баланса получим следующую аппроксимацию

$$\begin{aligned}&-\frac{1}{\hbar_i}\left[p\left(x_i+\frac{h_{i+1}}{2}\right)\frac{u_{i+1}^h-u_i^h}{h_{i+1}}-p\left(x_i-\frac{h_{i+1}}{2}\right)\frac{u_i^h-u_{i-1}^h}{h_i}\right]+\\&+q(x_i)u_i^h=f(x_i),\quad i=\overline{1,N-1}.\end{aligned} \tag{19.68}$$

Если сетка (19.67) является произвольной, то погрешность аппроксимации (19.68) есть только $O(h)$, где $h=\max h_i$. Однако можно доказать, что погрешность решения соответствующей сеточной задачи и на этой сетке будет величиной $O(h^2)$ при соответствующей гладкости решения уравнения (19.38).

Упражнение 19.8. Исследовать погрешность аппроксимации (19.68).

Выше всюду речь шла о том случае, когда коэффициенты уравнения (19.38) достаточно гладкие. В приложениях часто коэффициенты бывают кусочно-гладкие (например, кусочно-постоянные). В этом случае для аппроксимации уравнения целесообразно использовать сетку, у которой в качестве узлов присутствуют все точки разрыва коэффициентов $p(x)$, $q(x)$ и правой части $f(x)$. Такая сетка будет, как правило, неравномерной и может быть кусочно-равномерной. Указанный выбор сетки позволяет получать точность приближенного решения не ниже, чем в гладком случае.

19.8 Уравнение конвекции-диффузии

Добавим к уравнению еще (19.38) один член — первую производную искомого решения, умноженную на некоторый коэффициент

$$-(pu')' - r(x)u' + q(x)u = f. \tag{19.69}$$

Как аппроксимировать первый и последний члены левой части (19.69), мы знаем. Осталось построить аппроксимацию второго члена. С точки зрения наилучшего порядка аппроксимации следует положить

$$u'_i \approx \frac{u_{i+1} - u_{i-1}}{2h} =: u_{\overset{\circ}{x}}. \tag{19.70}$$

Тогда аппроксимация уравнения (19.69) примет вид

$$-\left(p_{i-1/2}u^h_{\bar{x}}\right)_{x,i} - r_i u^h_{\overset{\circ}{x},i} + q_i u^h_i = f_i, \quad i = 1, \ldots, N-1. \tag{19.71}$$

Теорема 19.9. *Если $u \in C^4[0,1]$, то погрешность аппроксимации разностной схемы* (19.71) $\Psi = O(h^2)$.

Дополним уравнение (19.69) граничными условиями. Пусть, например,

$$u(0) = g_0, \qquad u(1) = g_1. \tag{19.72}$$

Тогда разностные уравнения (19.71) нужно дополнить граничными условиями

$$u^h_0 = g_0, \qquad u^h_N = g_1. \tag{19.73}$$

Имеет место

Теорема 19.10. *Если коэффициенты уравнения* (19.69) *удовлетворяют условиям* (19.23), (19.47), *а сетка такова, что*

$$\max_i \frac{|r(x_i)|h}{2c_0} \leqslant 1, \tag{19.74}$$

то задача (19.71), (19.73) *имеет единственное решение, и для него справедлива априорная оценка*

$$\max_i |u_i^h| \leqslant |g_0| + |g_1| + \max_i \frac{|f_i|}{c_1}\,. \tag{19.75}$$

Доказательство. Представим

$$u_{\mathring{x}}^h = \frac{u_{i+1}^h - u_{i-1}^h}{2h} = \frac{u_{i+1}^h - u_i^h + u_i^h - u_{i-1}^h}{2h} = \frac{1}{2}u_x + \frac{1}{2}u_{\bar{x}}.$$

Подставим это представление в (19.71)

$$-\frac{1}{h}\left(p_{i+1/2}u_{x,i}^h - p_{i-1/2}u_{\bar{x},i}^h\right) - \frac{r_i}{2}\left(u_{x,i}^h + u_{\bar{x},i}^h\right) + q_i u_i^h = f_i.$$

Отсюда

$$-\left(\frac{p_{i+1/2}}{h} + \frac{r_i}{2}\right)\frac{u_{i+1}^h - u_i^h}{h} + \left(\frac{p_{i-1/2}}{h} - \frac{r_i}{2}\right)\frac{u_i^h - u_{i-1}^h}{h} + q_i u_i^h = f_i.$$

При выполнении условий (19.74) выражения в скобках неотрицательные. Этого замечания достаточно для того, чтобы завершить доказательство этой теоремы, используя те же самые рассуждения, что и при доказательстве теорем 19.7 и 19.1.

Упражнение 19.9. Завершить доказательство теоремы 19.10.

Упражнение 19.10. Сформулировать и доказать теорему о сходимости разностной задачи (19.71), (19.73).

§ 20

Сингулярно возмущенные уравнения. Негладкие решения

20.1 Осцилляции решения и сингулярно возмущенные уравнения

При исследовании разрешимости и сходимости разностной схемы (19.71) для уравнения конвекции-диффузии (19.69) мы ввели ограничение (19.74) на шаг сетки. Это ограничение в ряде случаев оказывается излишне обременительным, и тогда от аппроксимации (19.70) первой производной приходится отказываться. Обсудим этот вопрос на примере простейшего однородного уравнения с постоянными коэффициентами

$$\frac{d^2u}{dx^2} + r\frac{du}{dx} = 0, \qquad r = \text{const}. \tag{20.1}$$

Наряду с аппроксимацией (19.70) производной u' рассмотрим также ее аппроксимации односторонними разностными отношениями u_x и $u_{\bar{x}}$. Разумеется, порядок погрешности аппроксимации в этих случаях будет хуже. Будем рассматривать одновременно все три из указанных аппроксимаций u'. Для этого в разностное уравнение введем параметр σ

$$u^h_{\bar{x}x} + r\left[\sigma u^h_x + (1-\sigma)u^h_{\bar{x}}\right] = 0. \tag{20.2}$$

При $\sigma = 1/2$ имеем $u_{\overset{\circ}{x}} = (u_x + u_{\bar{x}})/2$, при $\sigma = 1$ — u_x, а при $\sigma = 0$ — $u_{\bar{x}}$. Перепишем (20.2) в поточечном виде

$$\left(\frac{1}{h^2} + \frac{\sigma r}{h}\right)u^h_{i+1} - \left(\frac{2}{h^2} + \frac{(2\sigma-1)r}{h}\right)u^h_i + \left(\frac{1}{h^2} + \frac{(\sigma-1)r}{h}\right)u^h_{i-1} = 0.$$

Это есть разностное уравнение с постоянными коэффициентами. Его характеристическое уравнение имеет вид

$$\left(\frac{1}{h} + \sigma r\right)q^2 - \left(\frac{2}{h} + (2\sigma-1)r\right)q + \left(\frac{1}{h} + (\sigma-1)r\right) = 0. \tag{20.3}$$

Поскольку сумма коэффициентов уравнения (20.3) равна нулю, то среди его корней есть корень $q_1 = 1$. Второй корень

$$q_2 = q = \frac{1 + (\sigma - 1)rh}{1 + \sigma rh}. \tag{20.4}$$

Проведем качественное сравнение решений дифференциального уравнения (20.1) и разностного уравнения (20.2). Для этого предположим, что

$$r > 0 \tag{20.5}$$

и поставим задачу для (20.1) на положительной полуоси Ox

$$u(0) = 1, \qquad u(\infty) = 0. \tag{20.6}$$

Очевидно, что решение задачи (20.1), (20.5), (20.6) имеет вид

$$u(x) = e^{-rx}. \tag{20.7}$$

Функция (20.7) положительна и монотонно убывает при $x \to \infty$.

Будем искать решение разностного уравнения (20.2), удовлетворяющее условиям

$$u_0^h = 1, \qquad \lim_{i\to\infty} u_i^h = 0. \tag{20.8}$$

Общее решение уравнения (20.2) в силу вышесказанного есть

$$u_i^h = c_1 + c_2 q^i. \tag{20.9}$$

Для того, чтобы это решение на бесконечности было хотя бы ограниченным, нужно потребовать, чтобы (см. (20.4))

$$|q| \leqslant 1, \quad \text{т.е.} \quad -1 \leqslant \frac{1 + (\sigma - 1)rh}{1 + \sigma rh} \leqslant 1. \tag{20.10}$$

Пусть $\sigma \geqslant 0$. Тогда знаменатель в (20.10) положителен, правая часть неравенства имеет место всегда, и поэтому остается только ограничение

$$-1 - \sigma rh \leqslant 1 + (\sigma - 1)rh,$$

т.е.

$$2 + (2\sigma - 1)rh \geqslant 0.$$

Если $\sigma = 1$ или $\sigma = 1/2$, то это условие выполнено со знаком строгого неравенства, и решение задачи (20.2), (20.8) при этих значениях имеет вид

$$u_i^h = q^i, \qquad i \in \mathbb{N}. \tag{20.11}$$

Наложим более сильное условие на сеточное решение. Потребуем, чтобы оно было монотонным как и решение (20.7) дифференциальной задачи. Решение (20.11) будет монотонным тогда и только тогда, когда $q \geqslant 0$, т.е если

$$1 + (\sigma - 1)rh \geqslant 0. \tag{20.12}$$

При $\sigma = 1$ это условие выполнено, а при $\sigma = 1/2$ требуется, чтобы

$$rh \leqslant 2 \tag{20.13}$$

(сравнить с (19.74)).

Итак, если $\sigma = 1$, погрешность аппроксимации уравнения (20.2) есть $O(h)$, но решение (20.11) задачи (20.2), (20.8) монотонно при любых h. Если $\sigma = 1/2$, то погрешность аппроксимации есть $O(h^2)$, но решение (20.11) монотонно только при выполнении условия (20.13). В противном случае решение (20.11) будет колебаться (см. рис. 1), меняя знак при переходе от одного узла к другому.

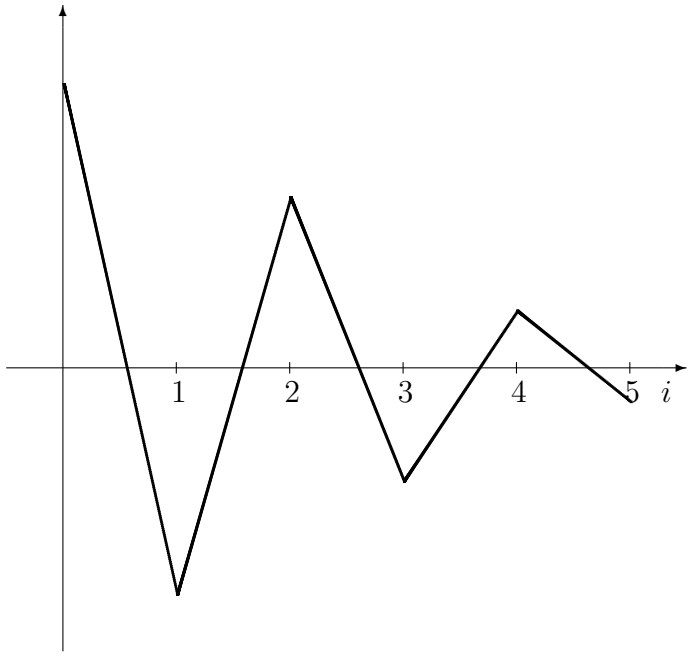

Рис. 1

Именно эти осцилляции решения разностной схемы (20.2) при $\sigma = 1/2$ и не любят прикладники.

Замечание 20.1. Проведенный анализ показал принципиальное различие между схемами (20.2) при $\sigma = 1$ и при $\sigma = 0$, хотя обе эти схемы имеют погрешность $O(h)$ и в этом смысле близки. Причина различия состоит в знаке коэффициента r. Если бы он был отрицательным, то схемы с $\sigma = 1$ и $\sigma = 0$ поменялись бы ролями.

Казалось бы, ограничение (20.13) не является слишком обременительным, чтобы всегда требовать его выполнения. Для обычных задач это так. Но есть важный класс так называемых сингулярно возмущенных уравнений, когда ограничение (20.13) оказывается весьма обременительным. Простейшим примером является уравнение

$$\varepsilon u'' + u' = 0. \tag{20.14}$$

Здесь $\varepsilon \in (0, 1]$ — малый параметр. При $\varepsilon \to 0$ дифференциальное уравнение второго порядка (20.14) переходит в уравнение первого порядка, для которого одно из двух граничных условий, выделяющих единственное решение уравнения (20.14), становится лишним. Это и является причиной непростого поведения решения соответствующей задачи для уравнения (20.14) при малых ε. Если для уравнения (20.14) поставить граничные условия

$$u(0) = 0, \qquad u(1) = 1, \tag{20.15}$$

то решением этой задачи будет функция

$$u(x) = \frac{1 - e^{-x/\varepsilon}}{1 - e^{-1/\varepsilon}} = 1 - \frac{e^{-x/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}, \tag{20.16}$$

являющаяся суммой гладкой, медленно меняющейся функции $u_0(x) := 1$ и быстро меняющейся функции $u_1(x) := (e^{-x/\varepsilon} - e^{-1/\varepsilon})/(1 - e^{-1/\varepsilon})$.

Поскольку уравнения (20.1) и (20.14) переходят одно в другое при $r = 1/\varepsilon$, то условие (20.13) применительно к разностной схеме (20.2) для уравнения (20.14) примет вид

$$h \leqslant 2\varepsilon. \tag{20.17}$$

Но в (20.14) параметр ε может принимать значения 10^{-2}, 10^{-4} или даже 10^{-8}, и ограничение (20.17) становится слишком обременительным. В этой ситуации следует либо ограничиться схемой с $\sigma = 1$, которая не накладывает никаких ограничений на шаг сетки с точки зрения осцилирования решения, и довольствоваться погрешностью аппроксимации $O(h)$, либо пытаться строить другие схемы, которые имеют погрешность $O(h^2)$ и не требуют ограничения типа (20.17).

20.2 Четырехточечная схема

Построим другую аппроксимацию уравнения (20.1). Будем аппроксимировать в (20.1) второе слагаемое при помощи соотношения

$$u'(x_i) \approx \frac{-u_{i+2} + 4u_{i+1} - 3u_i}{2h},$$

погрешность аппроксимации которого есть $O(h^2)$. Используя эту аппроксимацию, вместо (20.2) будем иметь

$$\frac{u^h_{i+1} - 2u^h_i + u^h_{i-1}}{h^2} + r\frac{-u^h_{i+2} + 4u^h_{i+1} - 3u^h_i}{2h} = 0. \tag{20.18}$$

Напишем характеристическое уравнение этого разностного уравнения с постоянными коэффициентами

$$\frac{q^2-2q+1}{h^2}+r\frac{-q^3+4q^2-3q}{2h}=0.$$

Обозначим $rh/2=\xi$ и перепишем характеристическое уравнение в виде

$$-\xi q^3+(1+4\xi)q^2-(2+3\xi)q+1=0.$$

Сумма коэффициентов этого уравнения равна нулю, и следовательно, $q=1$ есть корень этого уравнения. После деления многочлена из левой части на $(q-1)$ получим уравнение

$$-\xi q^2+(1+3\xi)q-1=0,$$

корнями которого являются числа

$$q_{2,3}=\frac{1+3\xi\pm\sqrt{1+6\xi+9\xi^2-4\xi}}{2\xi}.$$

Очевидно, что оба эти корня положительны при любых положительных ξ. Поскольку общее решение уравнения (20.18) имеет вид

$$u_i^h=c_1+c_2q_2^i+c_3q_3^i,$$

то осцилляции этого решения будут отсутствовать на любой сетке, т.е. при любых h.

Аппроксимирующим экспоненту e^{-rh} будет корень

$$\begin{aligned}q_2&=\frac{1}{2\xi}\left[1+3\xi-\sqrt{1+2\xi+9\xi^2}\right]=\\&=\frac{1}{2\xi}\left\{1+3\xi-\left[1+\frac{2\xi+9\xi^2}{2}-\frac{(2\xi+9\xi^2)^2}{8}+\frac{8}{16}\xi^3+O(\xi^4)\right]\right\}=\\&=1-2\xi+2\xi^2+O(\xi^3)=1-rh+\frac{r^2h^2}{2}+O(h^3)=e^{-rh}+O(h^3)\end{aligned}$$

Замечание 20.2. Поскольку уравнение (20.18) нельзя написать для $i=N-1$, то в этом узле должна быть написана другая аппроксимация уравнения (20.1), например, (20.2) при $i=N-1$ с любым σ (либо $\sigma=1/2$, либо $\sigma=1$).

Замечание 20.3. Мы рассмотрели случай $r>0$. Если $r<0$, то, сделав в (20.1) замену независимой переменной $1-x=t$, придем к уравнению

$$\frac{d^2u}{dt^2}-r\frac{du}{dt}=u''+|r|u'=0.$$

Отсюда следует, что при $r<0$ в исходных переменных нужно использовать аппроксимацию, зеркальную к той, которая используется при $r>0$. Именно, вместо разности вперед
$(u_{i+1}-u_i)/h$ — разность назад $(u_i-u_{i-1})/h$, а вместо
$(-u_{i+2}+4u_{i+1}-3u_i)/2h$ — аппроксимация

$$u_{\bar{x},i}+\frac{h}{2}u_{\bar{x}x,i}=\frac{u_{i-2}-4u_{i-1}+3u_i}{2h}.$$

20.3 О равномерной по ε сходимости

Исследования показывают, что какой бы метод аппроксимации уравнения (20.1) из числа рассмотренных выше мы ни избрали, в любом случае при фиксированном N и $\varepsilon \to 0$ найдутся такие узлы равномерной сетки, в которых погрешность решения будет $O(1)$. Чтобы отметить этот факт, говорят, что разностная схема не обладает свойством *равномерной по малому параметру сходимости*.

Один из путей обеспечения равномерной по малому параметру сходимости — использование сгущающихся сеток. Одна из простейших сеток, называемая сеткой Шишкина, имеет вид

$$\overline{\Omega} = \left\{ x_i \;\middle|\; x_i = ih,\ i = \overline{0, N/2},\ x_i = x_{N/2} + (i - N/2)H,\ i = \overline{N/2+1, N},\right.$$
$$\left. h = \delta/(N/2), \quad H = (1-\delta)/(N/2), \quad \delta = \min\{c\,\varepsilon \ln N, 1/2\} \right\}$$

или (см. рис. 2)

$$x_i = x(t_i), \quad \text{где} \quad t_i = i/N, \quad \text{а}$$
$$x(t) = \begin{cases} 2\delta t, & 0 \leqslant t \leqslant 1/2, \\ 1 - 2(1-\delta)(1-t), & 1/2 < t \leqslant 1 \end{cases}$$

есть кусочно-линейное непрерывное отображение отрезка $[0,1]$ на себя.

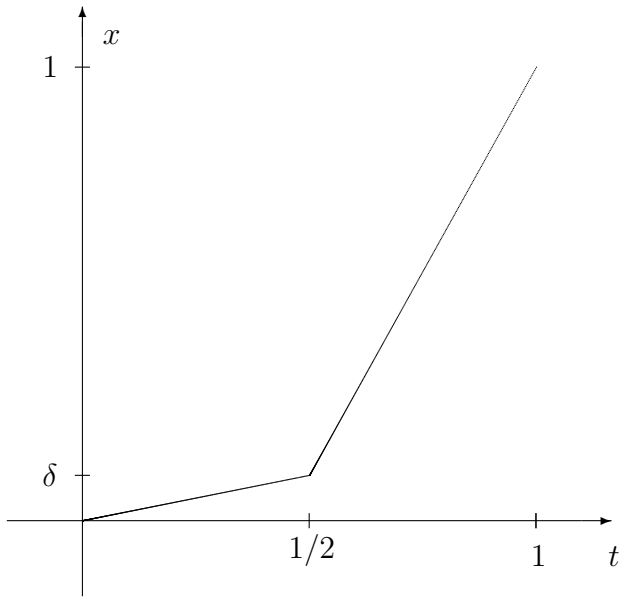

Рис. 2

Эта сетка является кусочно-равномерной с шагом $h \ll H$ на отрезке $[0, \delta]$ и с шагом H на отрезке $[\delta, 1]$.

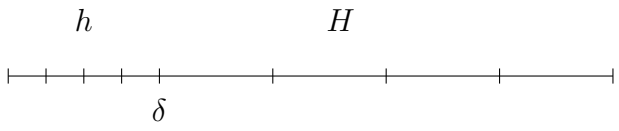

Рис. 3

Равномерная по малому параметру точность разностной схемы определяется погрешностью аппроксимации разностной схемы и величиной параметра c, который должен быть выбран таким, чтобы на длине δ быстро меняющаяся составляющая точного решения успела принять столь малое значение, которое уже не влияет на погрешность приближенного решения.

20.4 Негладкие решения

Рассмотрим следующее дифференциальное уравнение

$$-\frac{1}{x}(xu')' + \frac{\lambda^2}{x^2}u = 0, \quad 0 < x < 1. \tag{20.19}$$

Это уравнение не вкладывается в тот класс уравнений, который мы для себя выделили. Именно, коэффициент $p(x) := x \geqslant 0$, но не отрезан от нуля постоянной (на рассматриваемом отрезке). Поэтому для уравнения (20.19) в точке $x = 0$ нельзя ставить произвольное граничное условие. В самом деле, будем искать решение уравнения (20.19) в виде

$$u(x) = x^\alpha.$$

Подставляя это выражение в (20.19), находим, что для удовлетворения уравнения требуется выполнение условия

$$\alpha^2 = \lambda^2,$$

т.е. $\alpha = \pm\lambda$. Тем самым, мы нашли два фундаментальных решения уравнения (20.19), и его общее решение есть

$$u(x) = c_1 x^\lambda + c_2 x^{-\lambda}. \tag{20.20}$$

Без ограничения общности можно считать, что $\lambda > 0$. Если нас интересует ограниченное решение (что естественно с точки зрения приложений), то $c_2 = 0$ и

$$u(x) = c_1 x^\lambda.$$

Отсюда находим, что единственным допустимым граничным условием из числа классических является условие

$$u(0) = 0. \tag{20.21}$$

(именно это условие и будет выделять из (20.20) ограниченное решение). При $x = 1$ можно ставить любое граничное условие, например,

$$u(1) = 1. \tag{20.22}$$

Тогда решением задачи (20.19), (20.21), (20.22) будет функция

$$u(x) = x^\lambda. \tag{20.23}$$

Если $0 < \lambda < 1$, то уже первая производная интересующего нас решения не ограничена, не говоря уже о четвертой производной, которая фигурирует в погрешности аппроксимации. О хорошей сходимости численного решения на равномерной сетке говорить трудно. Выход из создавшегося положения можно найти на пути использования специальной сгущающейся к точке $x = 0$ сетки. Как эту сетку построить? Пусть $x = x(t)$ есть отображение отрезка $[0, 1]$ на себя. Для $t \in [0, 1]$ введем равномерную сетку с шагом $h = 1/N$. Тогда

$$x_i = x(t_i)$$

будет задавать узлы неравномерной сетки по x. На этой неравномерной сетке

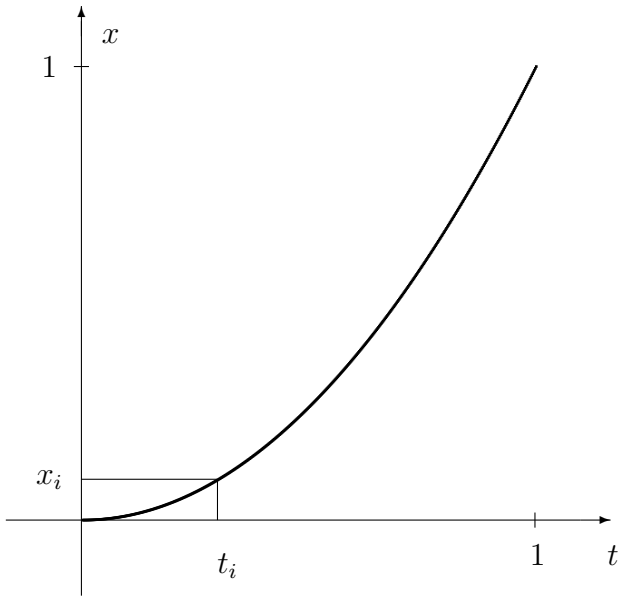

Рис. 4

и аппроксимируем уравнение (20.19). Пусть

$$h_i = x_i - x_{i-1}, \qquad \hbar_i = (h_i + h_{i+1})/2.$$

Тогда, используя, например, метод баланса (см. § 7), для уравнения (20.19) получим следующую аппроксимацию

$$\frac{1}{x_i}\frac{1}{\hbar_i}\left(x_{i+1/2}\frac{u^h_{i+1}-u^h_i}{h_{i+1}}-x_{i-1/2}\frac{u^h_i-u^h_{i-1}}{h_i}\right)-\frac{\lambda^2}{x_i^2}u^h_i=0. \tag{20.24}$$

Исследуем погрешность аппроксимации этой разностной схемы на неравномерной сетке. Используя формулу Тейлора, будем иметь

$$\begin{aligned}
\Psi_i &= \frac{1}{x_i}\frac{1}{\hbar_i}\left[x_{i+1/2}u_{x,i}-x_{i-1/2}u_{\bar{x},i}\right]-\frac{1}{x_i}(xu')'_i= \\
&= \frac{1}{x_i}\frac{1}{\hbar_i}\left[\left(x_i+\frac{h_{i+1}}{2}\right)\left(u'_i+\frac{h_{i+1}}{2}u''_i+\frac{h^2_{i+1}}{6}u'''_i+\frac{h^3_{i+1}}{24}\tilde{u}^{IV}_i\right)-\right. \\
&- \left.\left(x_i-\frac{h_i}{2}\right)\left(u'_i-\frac{h_i}{2}u''_i+\frac{h_i^2}{6}u'''_i-\frac{h_i^3}{24}\tilde{\tilde{u}}^{IV}_i\right)\right]-\frac{1}{x_i}(xu')'_i= \\
&= \frac{1}{x_i}\frac{1}{\hbar_i}\left[h^2_{i+1}\left(\frac{x_i}{6}u'''_i+\frac{1}{4}u''_i\right)+h^3_{i+1}\left(\frac{x_i}{24}\tilde{u}^{IV}_i+\frac{1}{12}u'''_i\right)+\frac{h^4_{i+1}}{48}\tilde{u}^{IV}_i-\right. \\
&- \left. h_i^2\left(\frac{x_i}{6}u'''_i+\frac{1}{4}u''_i\right)+h_i^3\left(\frac{x_i}{24}\tilde{\tilde{u}}^{IV}_i+\frac{1}{12}u'''_i\right)-\frac{h_i^4}{48}\tilde{\tilde{u}}^{IV}_i\right]= \\
&= \frac{h^2_{i+1}-h_i^2}{\hbar_i}\left(\frac{1}{6}u'''_i+\frac{1}{4}\frac{u''_i}{x_i}\right)+\frac{h^3_{i+1}}{\hbar_i}\left(\frac{1}{24}\tilde{u}^{IV}_i+\frac{1}{12}\frac{u'''_i}{x_i}\right)+ \\
&+ \frac{h_i^3}{\hbar_i}\left(\frac{1}{24}\tilde{\tilde{u}}^{IV}_i+\frac{1}{12}\frac{u'''_i}{x_i}\right)+\frac{1}{48}\frac{h^4_{i+1}}{\hbar_i}\frac{\tilde{u}^{IV}_i}{x_i}-\frac{1}{48}\frac{h_i^4}{\hbar_i}\frac{\tilde{\tilde{u}}^{IV}_i}{x_i}.
\end{aligned} \tag{20.25}$$

Подставим сюда истинное значение $u(x)$ из (20.23) и оценим вклад в погрешность решения типичной составляющей погрешности аппроксимации

$$\overset{\circ}{\psi}_i = c(x_i)h_i^2x_i^{\lambda-4}.$$

Эта составляющая представлена в погрешности аппроксимации (20.25) вторым и третьим слагаемыми. Составляющую погрешности решения, отвечающую $\overset{\circ}{\psi}_i$, обозначим через $\overset{\circ}{z}_i$. Для нее имеем уравнение

$$-\frac{1}{x_i}\frac{1}{\hbar_i}\left(x_{i+1/2}\frac{\overset{\circ}{z}_{i+1}-\overset{\circ}{z}_i}{h_{i+1}}-x_{i-1/2}\frac{\overset{\circ}{z}_i-\overset{\circ}{z}_{i-1}}{h_i}\right)+\frac{\lambda^2}{x_i^2}\overset{\circ}{z}_i=c(x_i)h_i^2x_i^{\lambda-4}.$$

Как и при доказательстве теоремы 19.1 для максимума $|\overset{\circ}{z}_i|$ получаем оценку

$$\max_i|\overset{\circ}{z}_i|\leqslant\max_i\frac{c(x_i)h_i^2x_i^{\lambda-2}}{\lambda^2}. \tag{20.26}$$

Из этой оценки следует, что, если $\lambda\geqslant 2$, то никаких проблем нет, ибо в этом случае выражение, стоящее в правой части под знаком max, имеет равномерную по x_i малость

$O(h_i^2)$, и сетку можно брать равномерной. Если же $\lambda < 2$, то равномерной по x_i малости $O(h_i^2)$ указанного выражения не гарантируется, если сетка не выбрана надлежащим образом. Поскольку $c(x_i)$ из правой части (20.26) меняется мало, выберем сетку при $\lambda < 2$ так, чтобы

$$h_i^2 x_i^{\lambda-2} \approx \text{const}.$$

Так как

$$h_i = x_i - x_{i-1} = N^{-1}x'(t_i^*), \tag{20.27}$$

то

$$h_i^2 x_i^{\lambda-2} = N^{-2}x'^2(t_i^*)x_i^{\lambda-2}.$$

Пусть

$$x'^2 x^{\lambda-2} = c,$$

где c — некоторая постоянная, или

$$x' x^{\lambda/2-1} = \sqrt{c} = c_1.$$

Интегрируя это уравнение, находим, что

$$x^{\lambda/2} = c_1 t + c_2,$$

или

$$x = (c_1 t + c_2)^{2/\lambda}.$$

Так как $x(0) = 0$, а $x(1) = 1$, то $c_2 = 0$, а $c_1 = 1$. Тем самым,

$$x = t^{2/\lambda}, \tag{20.28}$$

и, следовательно,

$$x_i = (i/N)^{2/\lambda}. \tag{20.29}$$

Если узлы сетки будут заданы по закону (20.29), то, в силу (20.27), (20.28) при $\lambda < 2$

$$h_i = 2N^{-1}\tilde{t}_i^{2/\lambda-1}/\lambda,$$

и величины шагов сетки уменьшаются при приближении к границе $x = 0$, т.е. построенная сетка является сгущающейся в окрестности $x = 0$. Если $i \sim N$, то $h_i \sim cN^{-1}$, а если $i = 1$, то

$$h_1 = \frac{2}{\lambda}N^{-2/\lambda} \ (\lambda < 2).$$

Принимая во внимание сказанное, а также (20.23), (20.28) и (20.26), легко проверить, что вклад последних четырех слагаемых погрешности аппроксимации (20.25) в погрешность решения является величиной $O(N^{-2})$.

Обратимся к первому слагаемому правой части (20.25). Используя, например, формулу Тейлора, находим, что

$$\frac{h_{i+1}^2 - h_i^2}{\hbar_i} = 2(h_{i+1} - h_i) = 2(x_{i+1} - 2x_i + x_{i-1}) = 2N^{-2}x_i''(t^*).$$

Снова принимая во внимание (20.23), (20.28) и (20.26), заключаем, что вклад и первого слагаемого погрешности аппроксимации (20.25) в погрешность решения оценивается величиной $O(N^{-2})$.

Глава VI

Численные методы для дифференциальных уравнений с частными производными

§ 21

Разностные схемы для уравнения теплопроводности

21.1 Нестационарное уравнение теплопроводности

Нестационарное уравнение теплопроводности являет собой простейший пример параболического уравнения — уравнения с частными производными. Возьмем его в виде

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t), \qquad 0 < x < 1, \quad 0 < t \leqslant T. \tag{21.1}$$

Чтобы выделить единственное решение уравнения (21.1), нужно задать дополнительные условия. Таковыми могут быть граничные условия, задаваемые при $x = 0$ и $x = 1$, и начальное условие, задаваемое при $t = 0$. Пусть, например, граничные условия имеют вид

$$u(0,t) = u(1,t) = 0, \tag{21.2}$$

а начальное условие —

$$u(x,0) = \varphi(x). \tag{21.3}$$

Как известно из курса методов математической физики, задача (21.1)-(21.3) поставлена корректно и при надлежащей гладкости $f(x,t)$ и $\varphi(x)$ имеет единственное решение.

Посмотрим на уравнение (21.1) с точки зрения краевых задач для обыкновенных дифференциальных уравнений. Для этого обозначим $\partial u/\partial t = \dot{u}$ и перепишем (21.1) в виде

$$-\frac{\partial^2 u}{\partial x^2} = f(x,t) - \dot{u} \equiv \mathcal{F}(x,t). \tag{21.4}$$

Считая $\mathcal{F}(x,t)$ в (21.4) заданной функцией, а t — параметром, мы можем условно рассматривать (21.4) как обыкновенное дифференциальное уравнение, аппроксимацию которого мы строить умеем. На $[0,1]$ введем сетку

$$\bar{\omega}^h = \left\{ x = x_i = ih \;\middle|\; i = 0, \ldots, N \right\}$$

с внутренними узлами

$$\omega^h = \left\{ x_i \in \bar{\omega}^h \;\middle|\; i = 1, \ldots, N-1 \right\}$$

и на этой сетке дифференциальное уравнение (21.4) аппроксимируем разностным уравнением

$$-u^h_{\bar{x}x,i} = \mathcal{F}^h(x_i, t), \qquad x_i \in \omega^h. \tag{21.5}$$

Теперь нужно вспомнить (21.4), в силу которого

$$\mathcal{F}(x_i, t) = f(x_i, t) - \dot{u}(x_i, t).$$

Поэтому естественно положить

$$\mathcal{F}^h(x_i, t) = f^h(x_i, t) - \dot{u}^h_i.$$

Тогда (21.5) примет вид

$$-u^h_{\bar{x}x,i} = f^h_i(t) - \dot{u}^h_i, \quad x_i \in \omega^h. \tag{21.6}$$

Это соотношение представляет собой систему $(N-1)$ обыкновенных дифференциальных уравнений первого порядка с $(N+1)$ неизвестными u^h_i, $i = 0, \ldots, N$. Воспользуемся граничными условиями (21.2) и положим

$$u^h_0(t) = u^h_N(t) = 0. \tag{21.7}$$

После исключения этих неизвестных из (21.6) будем иметь систему $(N-1)$ уравнений с $(N-1)$ неизвестными.

Перепишем теперь (21.6) по-другому, поставив на первое место производную

$$\dot{u}^h_i = u^h_{\bar{x}x,i} + f^h_i(t), \qquad i = 1, \ldots, N-1, \tag{21.8}$$

и введем обозначения

$$\begin{aligned} U &= [u^h_1 \ldots u^h_{N-1}]^T, \\ \Lambda &= \frac{1}{h^2} \begin{bmatrix} -2 & 1 & 0 & \ldots & 0 & 0 \\ 1 & -2 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & -2 \end{bmatrix}, \\ F &= [f^h_1 \ldots f^h_{N-1}]. \end{aligned} \tag{21.9}$$

Тогда система (21.8) с учетом (21.7) примет вид

$$\frac{dU}{dt} = \Lambda U + F. \tag{21.10}$$

Введем еще одно обозначение

$$\Phi = [\varphi_1 \dots \varphi_{N-1}]^T$$

и положим

$$U(0) = \Phi. \tag{21.11}$$

Соотношения (21.10), (21.11) представляют собой задачу Коши для системы обыкновенных дифференциальных уравнений первого порядка. Для приближенного решения этой задачи можно использовать уже изученные методы. Например, метод Эйлера, который приводит к соотношениям

$$\frac{U^{j+1} - U^j}{\tau} = \Lambda U^j + F^j, \qquad U^0 = \Phi, \tag{21.12}$$

или неявный метод Эйлера

$$\frac{U^{j+1} - U^j}{\tau} = \Lambda U^{j+1} + F^{j+1}, \qquad U^0 = \Phi, \tag{21.13}$$

а можно и метод трапеций

$$\frac{U^{j+1} - U^j}{\tau} = \frac{1}{2}\Lambda\left[U^{j+1} + U^j\right] + \frac{1}{2}\left(F^{j+1} + F^j\right), \qquad U^0 = \Phi. \tag{21.14}$$

Мы не будем изучать общие методы решения задачи (21.10), (21.11), а ограничимся одношаговыми, как (21.12)-(21.14), которые в теории разностных схем для параболических уравнений принято называть двухслойными.

Изучение (21.12), (21.13) и (21.14) можно проводить одновременно, если записать их единым образом за счет введения параметра σ:

$$\frac{U^{j+1} - U^j}{\tau} = \sigma\Lambda U^{j+1} + (1-\sigma)\Lambda U^j + \sigma F^{j+1} + (1-\sigma)F^j. \tag{21.15}$$

Полагая здесь $\sigma = 0$, 1 или 1/2, получим (21.12), (21.13) или (21.14), соответственно.

Посмотрим теперь на (21.15) с точки зрения аппроксимации не задачи Коши для системы обыкновенных дифференциальных уравнений (21.10), (21.11), а с точки зрения аппроксимации задачи (21.1)-(21.3). В результате двух шагов аппроксимации в области $[0,1] \times [0,T]$ образована сетка (см. рис. 1),

на которой дифференциальное уравнение (21.1) аппроксимировано системой разностных уравнений

$$\frac{u_i^{h\,j+1} - u_i^{h\,j}}{\tau} = \sigma u_{\bar{x}x,i}^{h\,j+1} + (1-\sigma)u_{\bar{x}x,i}^{h\,j} + f_i^{h\,j}, \quad i = 1, \dots, N-1, \quad j = 0, \dots, J-1, \tag{21.16}$$

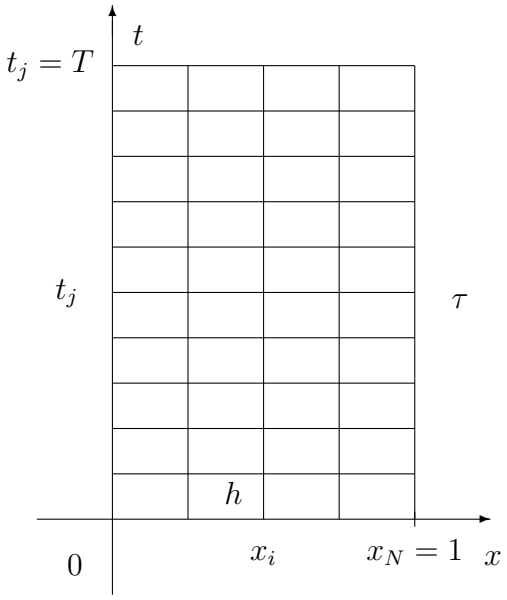

Рис. 1

а граничные (21.2) и начальное (21.3) условия — соотношениями

$$u_0^{h\,j} = 0, \quad u_N^{h\,j} = 0, \quad j = 1, \dots, J, \tag{21.17}$$

и

$$u_i^{h\,0} = \varphi_i, \quad i = 0, \dots, N, \tag{21.18}$$

соответственно. (На связи $f_i^{h\,j}$ с $f(x,t)$ мы не останавливаемся).

Введем дополнительные обозначения

$$u_i^j = u, \quad u_i^{j+1} = \hat{u}, \quad (\hat{u} - u)/\tau = u_t.$$

В новых обозначениях уравнения (21.16) примут вид

$$u_t^h = \sigma \hat{u}_{\bar{x}x}^h + (1-\sigma) u_{\bar{x}x}^h + f^h. \tag{21.19}$$

Погрешностью аппроксимации уравнения (21.1) уравнениями (21.19) будет сеточная функция

$$\Psi = f^h + \sigma \hat{u}_{\bar{x}x} + (1-\sigma) u_{\bar{x}x} - u_t,$$

где $u = u(x_i, t_j)$ — значения решения уравнения (21.1) в узлах (x_i, t_j).

Упражнение 21.1. Доказать, что при надлежащей гладкости (какой?)

$$\Psi = \begin{cases} O(\tau + h^2) & \text{при} \quad \sigma = 0, \sigma = 1, \\ O(\tau^2 + h^2) & \text{при} \quad \sigma = 1/2. \end{cases}$$

Замечание 21.1. Для написания уравнений (21.19) при $\sigma = 0$, $\sigma = 1$ или $\sigma = 1/2$ требуются следующие множества узлов

соответственно, называемые шаблонами.

21.2 Устойчивость по начальным данным

Исследуем разностную схему (21.16)-(21.18) на предмет ее устойчивости по начальным данным. Для этого будем считать, что правая часть в уравнениях (21.16) равна нулю, т.е.

$$\frac{u_i^{h\,j+1} - u_i^{h\,j}}{\tau} = \sigma u_{\bar{x}x,i}^{h\,j+1} + (1-\sigma) u_{\bar{x}x,i}^{h\,j}, \quad \begin{matrix} i = 1, \ldots, N-1, \\ j = 0, \ldots, J-1. \end{matrix} \tag{21.20}$$

Чтобы исследовать вопрос об устойчивости, найдем решение задачи (21.20), (21.17), (21.18). Решение будем искать методом разделения переменных. Будем искать частные решения уравнений (21.20) в виде

$$u_i^j = X_i T_j.$$

Тогда

$$\frac{T_{j+1} - T_j}{\tau} X_i = (\sigma T_{j+1} + (1-\sigma) T_j) X_{\bar{x}x,i}$$

или

$$\frac{(T_{j+1} - T_j)/\tau}{\sigma T_{j+1} + (1-\sigma) T_j} = \frac{X_{\bar{x}x,i}}{X_i} = -\lambda^h, \tag{21.21}$$

где λ^h — постоянная. С учетом граничных условий (21.17) для X_i из (21.21) получим задачу

$$X_{\bar{x}x,i} + \lambda^h X_i = 0, \quad i = 1, 2, \ldots, N-1, \quad X_0 = X_N = 0,$$

или, в развернутом виде,

$$-X_{i-1} + 2X_i - X_{i+1} = h^2 \lambda^h X_i, \quad i = 1, 2, \ldots, N-1, \quad X_0 = X_N = 0. \tag{21.22}$$

Но эта задача совпадает с рассмотренной нами ранее задачей (6.33), если в последней под λ понимать $h^2\lambda^h$. Поэтому, в силу (6.35)

$$X_i^{(k)} = \sqrt{2} \sin k\pi x_i, \quad i = 1, \ldots, N-1 \tag{21.23}$$

суть собственные векторы задачи (21.22), которые ортогональны в смысле скалярного произведения

$$(u, v) = \sum_{i=1}^{N-1} u_i v_i h$$

и нормированы, т.е. $\|X_i^{(k)}\|^2 = \left(X_i^{(k)}, X_i^{(k)}\right) = 1$. В силу (6.42)

$$\lambda_k^h = \frac{4}{h^2}\sin^2\frac{k\pi h}{2}, \quad k = 1, \ldots, N-1 \tag{21.24}$$

— различные собственные значения этой задачи.

Далее, из (21.21) находим, что

$$\frac{T_{j+1}^{(k)} - T_j^{(k)}}{\tau} + \lambda_k^h\left[\sigma T_{j+1}^{(k)} + (1-\sigma)T_j^{(k)}\right] = 0$$

или

$$(1 + \sigma\tau\lambda_k^h)T_{j+1}^{(k)} = (1 - (1-\sigma)\tau\lambda_k^h)T_j^{(k)}.$$

Отсюда следует, что

$$T_{j+1}^{(k)} = q_k T_j^{(k)},$$

где

$$q_k = \frac{1 - (1-\sigma)\tau\lambda_k^h}{1 + \sigma\tau\lambda_k^h} \tag{21.25}$$

и поэтому

$$T_j^{(k)} = c_k q_k^j. \tag{21.26}$$

Итак, мы нашли, что функции

$$u_i^{j(k)} = X_i^{(k)} T_j^{(k)}, \quad k = 1, \ldots, N-1,$$

где $X_i^{(k)}$ и $T_j^{(k)}$ из (21.23) и (21.26), соответственно, являются частными решениями уравнений (21.20), удовлетворяющими граничным условиям (21.17). Построим линейную комбинацию этих решений

$$u_i^{h\,j} = \sum_{k=1}^{N-1} c_k X_i^{(k)} q_k^j. \tag{21.27}$$

Полагая здесь $j = 0$, получим

$$u_i^{h\,0} = \sum_{k=1}^{N-1} c_k X_i^{(k)},$$

а принимая во внимание (21.18), заключаем, что функция (21.27) будет удовлетворять начальным условиям (21.18), если

$$\sum_{k=1}^{N-1} c_k X_i^{(k)} = \varphi_i,$$

т.е. если постоянные c_k суть коэффициенты Фурье функции φ_i при разложении по ортонормированной системе $X_i^{(k)}$

$$c_k = (\varphi, X^{(k)}) = \sum_{i=1}^{N-1} \varphi_i X_i^{(k)} h. \tag{21.28}$$

Итак, сеточная функция (21.27) с коэффициентами c_k из (21.28) удовлетворяет уравнениям (21.20), граничным условиям (21.17) и начальному условию (21.18), а поэтому является решением задачи (21.20), (21.17), (21.18).

Найдем оценку этого решения. Возводя левую и правую части (21.27) в квадрат и суммируя результат по i от 1 до $N-1$, с учетом ортогональности $X_i^{(k)}$, будем иметь

$$\|u^{h\,j}\|_{L_2^h}^2 = \sum_{i=1}^{N-1} (u_i^{h\,j})^2 h = \sum_{i=1}^{N-1} h \sum_{k,l=1}^{N-1} c_k c_l X_i^{(k)} X_i^{(l)} q_k^j q_l^j =$$

$$= \sum_{k,l=1}^{N-1} c_k c_l q_k^j q_l^j (X_i^{(k)}, X_i^{(l)}) = \sum_{k=1}^{N-1} c_k^2 q_k^{2j} \leqslant \max_k\, q_k^{2j} \sum_{k=1}^{N-1} c_k^2 = \max_k\, q_k^{2j} \|\varphi\|_{L_2^h}^2.$$

Пусть

$$|q_k| \leqslant 1. \tag{21.29}$$

Тогда

$$\|u^{h\,j}\|_{L_2^h} \leqslant \|\varphi\|_{L_2^h}, \tag{21.30}$$

т.е. L_2^h-норма решения при любом j не превосходит L_2^h-нормы начального условия.

Выясним, когда выполняется условие (21.29). С учетом (21.24), (21.25) при $\sigma \geqslant 0$ имеем

$$-\left(1 + \frac{4\tau}{h^2}\sigma \sin^2 \frac{k\pi h}{2}\right) \leqslant 1 - \frac{4\tau}{h^2}(1-\sigma)\sin^2 \frac{k\pi h}{2}$$

или, после приведения подобных членов,

$$2 - \frac{4\tau}{h^2}(1-2\sigma)\sin^2 \frac{k\pi h}{2} \geqslant 0, \quad k = 1, \ldots, N-1.$$

Отсюда вытекает условие

$$(1-2\sigma) \leqslant \frac{h^2}{2\tau} \min_k \frac{1}{\sin^2 \frac{k\pi h}{2}}.$$

Поскольку $\min\limits_k \sin^{-2} \frac{k\pi h}{2} \geqslant 1$, то (21.29) будет выполнено, если

$$(1-2\sigma) \leqslant \frac{h^2}{2\tau}$$

или, что эквивалентно,

$$\sigma \geqslant \frac{1}{2} - \frac{h^2}{4\tau}. \tag{21.31}$$

Итак, нами доказана

Теорема 21.1. *Если параметр σ схемы* (21.20) *удовлетворяет условию* (21.31), *то для решения задачи* (21.20), (21.17), (21.18) *справедлива априорная оценка*

$$\max_j \|u^{h\,j}\|_{L_2^h} \leqslant \|u^{h\,0}\|_{L_2^h}, \quad j = 1, 2, \ldots, J. \tag{21.32}$$

Определение 21.1. Говорят, что разностная схема (21.20) устойчива по начальным данным, если для решения задачи (21.20), (21.17), (21.18) справедлива оценка

$$\|u^{h\,j}\|_{(1)} \leqslant M\|u^{h\,0}\|_{(2)},$$

где $\|\cdot\|_{(1)}$ и $\|\cdot\|_{(2)}$ — некоторые нормы, а $M = \mathrm{const} > 0$ не зависит от τ и h.

Следствие 1. *Теорема 21.1 утверждает устойчивость по начальным данным схемы* (21.20) *при выполнении условий* (21.31), *когда*

$$\|\cdot\|_{(2)} = \|\cdot\|_{L_2^h}, \quad \|\cdot\|_{(1)} = \|\cdot\|_{L_\infty^h(0,T)\times L_2^h(0,1)}.$$

Обсудим условие (21.31). Если $\sigma = 1$, т.е. использован неявный метод Эйлера для системы, то (21.31) выполнено при любых τ и h. То же самое имеет место и при $\sigma = 1/2$ (схема трапеций). Если же $\sigma = 0$, то для выполнения (21.31) нужно, чтобы

$$\tau \leqslant h^2/2. \tag{21.33}$$

Про первые две схемы (при $\sigma = 1$ и $\sigma = 1/2$) говорят, что они безусловно устойчивы, а третья ($\sigma = 0$) устойчива условно (для устойчивости шаги по временно́й переменной и по пространственной связаны неравенством (21.33)).

Напомним, что все три схемы нуль-устойчивы по терминологии из обыкновенных дифференциальных уравнений, а первые две еще и A-устойчивы.

Отметим, что числа $(-\lambda_k^h)$ являются собственными числами матрицы (21.9)

$$\frac{\max\limits_k |-\lambda_k^h|}{\min\limits_k |-\lambda_k^h|} = \frac{\lambda_{N-1}^h}{\lambda_1^h} = \frac{\sin^2 \frac{(N-1)\pi h}{2}}{\sin^2 \frac{\pi h}{2}} =$$
$$= \frac{\cos^2 \frac{\pi h}{2}}{\sin^2 \frac{\pi h}{2}} = \mathrm{ctg}^2 \frac{\pi h}{2} \gg 1 \quad \text{при} \quad h \ll 1,$$

т.е. система уравнений (21.8) жесткая.

21.3 Устойчивость по правой части

Обратимся теперь к неоднородному уравнению (21.19), а вместо (21.18) поставим однородные начальные условия

$$u_i^{h\,0} = 0, \qquad i = 0, \ldots, N. \tag{21.34}$$

Теорема 21.2. *Если параметр σ схемы* (21.19) *удовлетворяет условию* (21.31), *то для решения задачи* (21.19), (21.17), (21.34)*справедлива априорная оценка*

$$\max_j \|u^{hj}\|_{L_2^h} \leqslant T \max_j \|f^{hj}\|_{L_2^h}. \tag{21.35}$$

Доказательство. Разложим u_i^{hj} и f_i^{hj} при каждом j по собственным векторам задачи (21.22)

$$u_i^{hj} = \sum_{k=1}^{N-1} T_j^{(k)} X_i^{(k)}, \qquad f_i^{hj} = \sum_{k=1}^{N-1} f_j^{(k)} X_i^{(k)}.$$

Подставляя эти разложения в (21.19) и принимая во внимание ортогональность $X_i^{(k)}$, получим

$$\frac{T_{j+1}^{(k)} - T_j^{(k)}}{\tau} + \lambda_k^h \left[\sigma T_{j+1}^{(k)} + (1-\sigma) T_j^{(k)}\right] = f_j^{(k)}.$$

Приводя подобные члены, найдем, что

$$[1 + \sigma\tau\lambda_k^h] T_{j+1}^{(k)} = [1 - (1-\sigma)\tau\lambda_k^h] T_j^{(k)} + \tau f_j^{(k)},$$

а, разрешая относительно $T_{j+1}^{(k)}$, будем иметь

$$T_{j+1}^{(k)} = q_k T_j^{(k)} + \frac{\tau}{1 + \sigma\tau\lambda_k^h} f_j^{(k)}.$$

В силу (21.29) $|q_k| \leqslant 1$, а при $\sigma \geqslant 0$ знаменатель $(1 + \sigma\tau\lambda_k^h) \geqslant 1$ и поэтому

$$\left| T_{j+1}^{(k)} \right| \leqslant \left| T_j^{(k)} \right| |q_k| + \tau \left| f_j^{(k)} \right| \leqslant \left| T_j^{(k)} \right| + \tau \left| f_j^{(k)} \right|.$$

Далее

$$\|u^{hj+1}\|_{L_2^h} = \sqrt{\sum_{k=1}^{N-1} \left(T_{j+1}^{(k)}\right)^2} \leqslant \sqrt{\sum_{k=1}^{N-1} \left(\left| T_j^{(k)} \right| + \tau \left| f_j^{(k)} \right|\right)^2} \leqslant \|u^{hj}\|_{L_2^h} + \tau \|f^{hj}\|_{L_2^h}.$$

Суммируя это неравенство по j в нужных пределах, придем к (21.35). Теорема доказана.

Теорема 21.3 (сходимости). *Если выполнено условие* (21.31), *и решение задачи* (21.1)-(21.3) $u(x,t) \in C^4[0,1] \times C^3[0,T]$, *то решение u^h задачи* (21.16)-(21.18) *при соответствующей f_i^{hj} сходится к решению и задачи* (21.1)-(21.3) *со скоростью* $O(h^2 + (\sigma - 1/2)\tau + \tau^2)$.

Доказательство. Пусть $z_i^j = u_i^{hj} - u(x_i, t_j)$ — погрешность решения. Выражая u_i^{hj} через z_i^j и $u(x_i, t_j)$ и подставляя результат в (21.16)-(21.18), для z_i^j получим задачу

$$\begin{gathered} \frac{z_i^{j+1} - z_i^j}{\tau} = \sigma z_{\bar{x}x,i}^{j+1} + (1-\sigma) z_{\bar{x}x,i}^j + \Psi_i^j, \\ z_0^j = z_N^j = 0, \qquad z_i^0 = 0. \end{gathered} \tag{21.36}$$

Для задачи (21.36) справедлива оценка, устанавливаемая теоремой 21.2, т.е.

$$\max_j \|z_i^j\|_{L_2^h} \leqslant T \max_j \|\Psi_i^j\|_{L_2^h}.$$

Используя теперь результаты упражнения 21.1, приходим к утверждению теоремы.

21.4 Устойчивость в смысле максимума модуля

Теорема 21.4. *Если выполнено условие*

$$\frac{2(1-\sigma)}{h^2}\tau \leqslant 1, \tag{21.37}$$

то для решения задачи (21.16)-(21.18) *справедлива априорная оценка*

$$\max_{ij} |u_i^{hj}| \leqslant \max_i |\varphi_i| + T \max_{ij} |f_i^{hj}|. \tag{21.38}$$

Доказательство. Перепишем уравнение (21.16) в поточечном виде

$$\frac{u_i^{hj+1} - u_i^{hj}}{\tau} = \sigma \frac{u_{i-1}^{hj+1} - 2u_i^{hj+1} + u_{i+1}^{hj+1}}{h^2} + (1-\sigma)\frac{u_{i-1}^{hj} - 2u_i^{hj} + u_{i+1}^{hj}}{h^2} + f_i^{hj}$$

и приведем подобные члены

$$\left(\frac{1}{\tau} + \frac{2\sigma}{h^2}\right) u_i^{hj+1} =$$
$$= \frac{\sigma}{h^2} u_{i-1}^{hj+1} + \frac{\sigma}{h^2} u_{i+1}^{hj+1} + \left(\frac{1}{\tau} - \frac{2(1-\sigma)}{h^2}\right) u_i^{hj} + \frac{1-\sigma}{h^2} u_{i-1}^{hj} + \frac{1-\sigma}{h^2} u_{i+1}^{hj} + f_i^{hj}.$$

Возьмем модули левой и правой частей и оценим правую часть этого соотношения через максимальные значения модулей u_i^{hj}, u_i^{hj+1} и f_i^{hj}. Будем иметь

$$\left(\frac{1}{\tau} + \frac{2\sigma}{h^2}\right) |u_i^{hj+1}| \leqslant$$
$$\leqslant \frac{2\sigma}{h^2} \max_i |u_i^{hj+1}| + \left(\left| \frac{1}{\tau} - \frac{2(1-\sigma)}{h^2} \right| + \frac{2(1-\sigma)}{h^2} \right) \max_i |u_i^{hj}| + \max_i |f_i^{hj}|.$$

Беря теперь максимум по i левой части и приводя подобные члены, после домножения на τ получим:

$$\max_i |u_i^{hj+1}| \leqslant \left(\left| 1 - \frac{2(1-\sigma)\tau}{h^2} \right| + \frac{2(1-\sigma)\tau}{h^2} \right) \max_i |u_i^{hj}| + \tau \max_i |f_i^{hj}|.$$

Изучим коэффициент при $\max_i |u_i^{h\,j}|$:

$$\left|1-\frac{2(1-\sigma)\tau}{h^2}\right|+\frac{2(1-\sigma)}{h^2}\tau=\begin{cases}1 & \text{при } \dfrac{2(1-\sigma)}{h^2}\tau\leqslant 1,\\ \dfrac{4(1-\sigma)\tau}{h^2}-1>1 & \text{при } \dfrac{2(1-\sigma)}{h^2}\tau>1\end{cases}.$$

В силу условия (21.37) теоремы реализуется первая возможность, и следовательно

$$\max_i |u_i^{h\,j+1}|\leqslant \max_i |u_i^{h\,j}|+\tau\max_i |f_i^{h\,j}|. \tag{21.39}$$

Пусть

$$\max_j \max_i |u_i^{h\,j}|=\max_i |u_i^{h\,j_0}|.$$

Тогда

$$\max_{i\,j} |u_i^{h\,j}|=\max_i |u_i^{h\,j_0}|\leqslant \max_i |u_i^{h\,j_0-1}|+\tau\max_i |f_i^{h\,j_0-1}|.$$

Прибавляя сюда все предыдущие неравенства (21.39) при $j=j_0-2,\ldots,j=0$, получим

$$\max_{i\,j} |u_i^{h\,j}|\leqslant \max_i |\varphi_i|+\sum_{j=1}^{J}\tau\max_i |f_i^{h\,j-1}|\leqslant \max_i |\varphi_i|+T\max_{i\,j} |f_i^{h\,j}|.$$

Теорема доказана.

Следствие 2. При $\sigma=1$ оценка (21.38) верна на любой сетке. При $\sigma=0$ (21.37) совпадает с (21.31), и для справедливости оценки (21.38) должно быть выполнено условие (21.33). Если же $\sigma=1/2$, то оценка (21.38) верна при

$$\tau/h^2\leqslant 1. \tag{21.40}$$

21.5 Сеточное преобразование Фурье

Пусть $\{x_m\}$ — совокупность равноотстоящих узлов на оси Ox. Будем использовать обозначение

$$v(x_m)=v_m,\qquad m\in\mathbb{Z}.$$

Будем предполагать, что $v_m\in l_2$, т.е.

$$\sum_{m\in\mathbb{Z}}|v_m|^2<\infty. \tag{21.41}$$

Определение 21.2. Будем называть 2π-периодическую функцию

$$(Fv_m)(\xi)=\sum_{m\in\mathbb{Z}}v_m e^{-im\xi}=\tilde{v}(\xi) \tag{21.42}$$

сеточным преобразованием Фурье.

Определение 21.3. Обратным сеточным преобразованием Фурье называется сеточная функция

$$(F^{-1}\tilde{v})_m = \frac{1}{2\pi}\int_0^{2\pi} \tilde{v}(\xi)e^{im\xi}d\xi = v_m. \tag{21.43}$$

Замечание 21.2. Соотношение (21.42) на самом деле представляет собой сумму ряда Фурье, коэффициентами которого являются значения рассматриваемой нами сеточной функции v_m. С этой точки зрения (21.43) есть формула для коэффициентов Фурье 2π-периодической функции $\tilde{v}(\xi)$.

Замечание 21.3. Можно было бы называть сеточным преобразованием Фурье функцию

$$Fv_m = \sum_{m\in\mathbb{Z}} hv_m e^{-i(mh)\xi/h} \equiv \tilde{v}(\xi/h), \quad \xi/h = \xi', \tag{21.44}$$

где h — расстояние между соседними узлами $h = x_m - x_{m-1}$. Тогда обратное преобразование приняло бы вид

$$F^{-1}\tilde{v} = \frac{1}{2\pi h}\int_0^{2\pi} \tilde{v}(\xi')e^{i(mh)\xi/h}d\xi = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} \tilde{v}(\xi')e^{i(mh)\xi'}d\xi'. \tag{21.45}$$

Устремляя в (21.44) и в (21.45) h к нулю, получим обычные прямое и обратное преобразование Фурье:

$$Fv(x) = \int_{-\infty}^{\infty} v(x)e^{-ix\xi}dx = \tilde{v}(\xi),$$

$$F^{-1}\tilde{v}(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{v}(\xi)e^{ix\xi}d\xi = v(x).$$

Для дальнейшего нам потребуется известное из теории рядов Фурье равенство Парсеваля

$$\int_0^{2\pi} |\tilde{v}|^2 d\xi =: \|\tilde{v}\|^2_{L_2(0,2\pi)} = \frac{1}{2\pi}\|v_m\|^2_{l_2} := \frac{1}{2\pi}\sum_{m\in\mathbb{Z}} |v_m|^2. \tag{21.46}$$

Пусть T есть оператор сдвига направо, т.е.

$$Tv_m = v_{m+1}.$$

Обратным к нему будет оператор сдвига налево

$$T^{-1}v_m = v_{m-1}.$$

Найдем сеточное преобразование Фурье этих операторов

$$F(Tv_m) = \sum_{m\in\mathbb{Z}} v_{m+1}e^{-im\xi}e^{-i\xi}e^{i\xi} = e^{i\xi}\tilde{v}(\xi).$$

Аналогично

$$F(T^{-1}v_m) = e^{-i\xi}\tilde{v}(\xi).$$

Теперь найдем преобразование Фурье разностных отношений. Имеем

$$Fv_{x,m} = \frac{1}{h}F(T-I)v_m = \frac{e^{i\xi}-1}{h}\tilde{v}(\xi), \tag{21.47}$$

$$Fv_{\bar{x},m} = \frac{1}{h}(I-T^{-1})v_m = \frac{1-e^{-i\xi}}{h}\tilde{v}(\xi), \tag{21.48}$$

$$Fv_{\bar{x}x,m} = \frac{1}{h}F(v_{x,m}-v_{\bar{x},m}) = \frac{e^{i\xi}-1-1+e^{-i\xi}}{h^2}\tilde{v}(\xi) =$$

$$= \frac{(e^{i\xi/2}-e^{-i\xi/2})^2}{h^2}\tilde{v}(\xi) = -\frac{4\sin^2\xi/2}{h^2}\tilde{v}(\xi). \tag{21.49}$$

21.6 Устойчивость по начальным данным разностной схемы для уравнения теплопроводности

Рассмотрим разностную схему (21.19) при $f^h \equiv 0$ на сетке, заданной на всей оси Ox, т.е. пусть

$$u^h_{t,m} = \sigma\hat{u}^h_{\bar{x}x,m} + (1-\sigma)u^h_{\bar{x}x,m}, \qquad m\in\mathbb{Z}. \tag{21.50}$$

Теорема 21.5. *Если параметр σ схемы* (21.50) *положителен и удовлетворяет условию*

$$\sigma \geqslant \frac{1}{2} - \frac{h^2}{4\tau} \tag{21.51}$$

то для решения (21.50) *имеет место априорная оценка*

$$\max_j \|u^{h\,j}\|_{L_2^h} \leqslant \|u^{h\,0}\|_{L_2^h}, \quad j = 1,2,\ldots. \tag{21.52}$$

Доказательство. Сделаем в (21.50) сеточное преобразование Фурье

$$\tilde{u}_t + \frac{4\sin^2\xi/2}{h^2}(\sigma\hat{\tilde{u}} + (1-\sigma)\tilde{u}) = 0.$$

Разрешая это обыкновенное разностное уравнение первого порядка относительно $\hat{\tilde{u}}$, получим

$$\hat{\tilde{u}} = q(\xi)\tilde{u}, \tag{21.53}$$

где

$$q(\xi) = \frac{1 - (1-\sigma)\dfrac{4\tau}{h^2}\sin^2\dfrac{\xi}{2}}{1 + \sigma\dfrac{4\tau}{h^2}\sin^2\dfrac{\xi}{2}}. \tag{21.54}$$

Из (21.53)

$$\|\hat{\tilde{u}}\|_{L_2(0,2\pi)} = \|q(\xi)\tilde{u}\|_{L_2(0,2\pi)} \leqslant \max_{0\leqslant\xi\leqslant 2\pi} |q(\xi)|\,\|\tilde{u}\|_{L_2(0,2\pi)}.$$

Отсюда следует, что L_2-норма образа Фурье решения не будет возрастать, если

$$|q(\xi)| \leqslant 1. \tag{21.55}$$

При этом

$$\|\hat{\tilde{u}}\|_{L_2(0,2\pi)} \leqslant \|\tilde{u}\|_{L_2(0,2\pi)} \leqslant \cdots \leqslant \|\tilde{u}^0\|_{L_2(0,2\pi)}.$$

Принимая теперь во внимание равенство Парсеваля (21.46), приходим к (21.52).

Покажем теперь, что (21.55) следует из (21.51). Так как $\sigma \geqslant 0$, то знаменатель в (21.54) положителен, и всегда $q \leqslant 1$. Осталось проверить условие $q \geqslant -1$, которое эквивалентно условию

$$2 - (1-2\sigma)\frac{4\tau}{h^2}\,\sin^2\frac{\xi}{2} \geqslant 0$$

или

$$1 - 2\sigma \leqslant \frac{h^2}{2\tau\sin^2\dfrac{\xi}{2}}.$$

Но это условие будет выполнено, если

$$1 - 2\sigma \leqslant \min_{\xi}\,\frac{h^2}{2\tau\sin^2\dfrac{\xi}{2}} = \frac{h^2}{2\tau},$$

что эквивалентно (21.51). Теорема доказана.

Упражнение 21.2. Рассмотреть неоднородное уравнение и установить оценку решения через правую часть.

§ 22

Разностные схемы для уравнения колебаний струны

22.1 Аппроксимация

Рассмотрим другой пример уравнения с частными производными — уравнение колебаний струны

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < 1, \quad 0 < t < T. \tag{22.1}$$

Это — гиперболическое уравнение. Корректной для него является смешанная задача, например,

$$u(0,t) = u(1,t) = 0, \quad u(x,0) = \bar{u}(x), \quad \frac{\partial u}{\partial t}(x,0) = \bar{\bar{u}}(x). \tag{22.2}$$

Граничные условия при $x = 0$ и $x = 1$ предполагаются однородными граничными условиями первого рода, а в качестве начальных функций взяты некоторые функции $\bar{u}(x)$ и $\bar{\bar{u}}(x)$.

Как и при построении аппроксимации уравнения теплопроводности, аппроксимируем сначала производную по пространственной переменной x. В результате получим задачу

$$\begin{aligned} &\ddot{u}_i^h(t) = u^h_{\bar{x}x,i}(t), \quad i = 1, \ldots, N-1, \\ &u_0^h(t) = u_N^h(t) = 0, \\ &u^h(x_i,0) = \bar{u}(x_i), \quad \dot{u}^h(x_i,0) = \bar{\bar{u}}(x_i), \end{aligned} \tag{22.3}$$

которая представляет собой задачу Коши для системы $(N-1)$ дифференциальных уравнений второго порядка. Теперь произведем аппроксимацию по временно́й переменной: производную $\ddot{u}(t)$ заменим вторым разностным отношением

$$u_{\bar{t}t}(t_j) \equiv [u(t_{j+1}) - 2u(t_j) + u(t_{j-1})]/\tau^2,$$

а правую часть (22.3) — линейной комбинацией ее значений при $t = t_{j-1}$, $t = t_j$ и $t = t_{j+1}$. В результате будем иметь

$$u^h_{\bar{t}t,i} = \sigma \hat{u}^h_{\bar{x}x,i} + (1-2\sigma) u^h_{\bar{x}x,i} + \sigma \check{u}^h_{\bar{x}x,i}, \quad i = 1, \ldots, N-1, \tag{22.4}$$

где наряду с уже введенным ранее обозначением $\hat{v}_i = v_i(t_{j+1})$ принято обозначение $\check{v}_i = v_i(t_{j-1})$. Правая часть (22.4) представляет собой не общую линейную комбинацию, а линейную комбинацию, симметричную относительно t_{j-1} и t_{j+1}.

К уравнениям (22.4) нужно добавить граничные и начальные условия, которые должны аппроксимировать соответственно условия (22.3)

$$u^{h\,j}_0 = \quad u^{h\,j}_N = 0, \qquad j = 0, \ldots, J, \tag{22.5}$$

$$u^{h\,0}_i = \quad \bar{u}(x_i), \qquad i = 1, \ldots, N-1. \tag{22.6}$$

Второе из начальных условий (22.3) содержит производную. Аппроксимируя ее по двум точкам, получим

$$u^{h\,0}_{t,i} = \bar{\bar{u}}(x_i), \quad i = 1, \ldots, N-1. \tag{22.7}$$

Теорема 22.1. *Если решение $u(x,t)$ уравнения (22.1) обладает непрерывными четвертыми производными, то погрешность аппроксимации разностной схемы (22.4) есть $O(\tau^2 + h^2)$.*

Доказательство.

$$\Psi^j_i = \sigma \hat{u}_{\bar{x}x,i} + (1-2\sigma) u_{\bar{x}x,i} + \sigma \check{u}_{\bar{x}x,i} - u_{\bar{t}t,i} = \sigma\tau^2 u_{\bar{x}x\bar{t}t,i} + u_{\bar{x}x,i} - u_{\bar{t}t,i} =$$
$$= \frac{\partial^2 u}{\partial x^2} + O(h^2) - \frac{\partial^2 u}{\partial t^2} + O(\tau^2) = O(\tau^2 + h^2).$$

Теорема доказана.

Найдем погрешность аппроксимации начального условия (22.7)

$$\mathring{\psi}_i = -u^0_{t,i} + \bar{\bar{u}}(x_i) = -\frac{\partial u}{\partial t}(x_i, 0) - \frac{\tau}{2}\frac{\partial^2 u}{\partial t^2}(x_i, 0) + O(\tau^2) + \bar{\bar{u}}(x_i) = -\frac{\tau}{2}\frac{\partial^2 u}{\partial t^2}(x_i, 0) + O(\tau^2). \tag{22.8}$$

Погрешность аппроксимации начального условия (22.7) есть $O(\tau)$. Построим другую аппроксимацию с погрешностью не хуже $O(\tau^2 + h^2)$. Для этого преобразуем (22.8). В силу (22.1) $\dfrac{\partial^2 u}{\partial t^2}(x,0) = \dfrac{\partial^2 u}{\partial x^2}(x,0)$, и поэтому

$$\mathring{\psi}_i = -\frac{\tau}{2}\frac{\partial^2 u}{\partial x^2}(x_i, 0) + O(\tau^2).$$

Принимая теперь во внимание (22.2), будем иметь

$$\mathring{\psi}_i = -\frac{\tau}{2}\bar{u}''(x_i) + O(\tau^2).$$

Отсюда и из (22.8) следует, что если вместо $\bar{\bar{u}}(x_i)$ в (22.7) положить $\bar{\bar{u}}(x_i) + \frac{\tau}{2}\bar{u}''(x_i)$, т.е. написать условие

$$u^{h\,0}_{t,i} = \bar{\bar{u}}_i + \frac{\tau}{2}\bar{u}''_i, \tag{22.9}$$

то погрешность этой аппроксимации будет $O(\tau^2)$. Очевидно также, что если вместо $\bar{u}''_i$ в (22.9) подставить $\bar{u}_{\bar{x}x,i}$, т.е. взять

$$u^{h\,0}_{t,i} = \bar{\bar{u}}_i + \frac{\tau}{2}\bar{u}_{\bar{x}x,i}, \tag{22.10}$$

то погрешность этой аппроксимации будет $O(\tau^2 + h^2)$.

Аппроксимация (22.10) всем хороша за исключением одного но. Именно, для аппроксимации уравнения (22.1) мы использовали однопараметрическое семейство разностных схем, среди которых имеются как явная ($\sigma = 0$), так и неявные ($\sigma \neq 0$). Аппроксимация же (22.10) всегда явная. Внесем параметр и в начальное условие. Пусть

$$u^{h\,0}_{t,i} = \bar{\bar{u}}_i + \frac{\tau}{2}[\gamma u^{h\,1}_{\bar{x}x,i} + (1-\gamma)u^{h\,0}_{\bar{x}x,i}].$$

Ясно, что погрешность этой аппроксимации снова не хуже $O(\tau^2+h^2)$. Наконец, согласуем параметр γ с параметром σ, полагая $\gamma/2 = \sigma$. Аппроксимация второго начального условия (22.2) примет следующий окончательный вид

$$u^{h\,0}_{t,i} = \tau\sigma u^{h\,1}_{\bar{x}x,i} + \tau(\frac{1}{2} - \sigma)u^{h\,0}_{\bar{x}x,i} + \bar{\bar{u}}_i. \tag{22.11}$$

22.2 Устойчивость по начальным данным

Исследуем вопрос об устойчивости схемы (22.4) по начальным данным. Ограничимся изучением задачи Коши, т.е. будем предполагать, что уравнения (22.4) и начальные условия (22.6) и (22.11) заданы для всех $i \in \mathbb{Z}$. Именно, будем рассматривать следующую задачу

$$u^h_{\bar{t}t,i} = \sigma\hat{u}^h_{\bar{x}x,i} + (1-2\sigma)u^h_{\bar{x}x,i} + \sigma\check{u}^h_{\bar{x}x,i}, \quad i \in \mathbb{Z}, \tag{22.12}$$

$$u^{h\,0}_i = \bar{u}_i, \quad u^{h\,0}_{t,i} = \tau\sigma u^{h\,1}_{\bar{x}x,i} + \tau(\frac{1}{2} - \sigma)u^{h\,0}_{\bar{x}x,i} + \bar{\bar{u}}_i, \quad i \in \mathbb{Z}. \tag{22.13}$$

Теорема 22.2. *Если параметр σ задачи* (22.12), (22.13) *неотрицателен и удовлетворяет условию*

$$\sigma \geqslant \frac{1}{4} - \frac{h^2}{4\tau^2}, \tag{22.14}$$

то для решения этой задачи справедлива априорная оценка

$$\|u^{h\,j}\|_{L_2^h} \leqslant \|\bar{u}\|_{L_2^h} + T\,\|\bar{\bar{u}}\|_{L_2^h}. \tag{22.15}$$

Доказательство. Сделаем сеточное преобразование Фурье (22.12), (22.13). Для образа Фурье $\tilde{u}^j(\xi)$ решения u_i^{hj} получим задачу

$$\begin{aligned} &\tilde{u}_{\bar{t}t} + \frac{4\sin^2\frac{\xi}{2}}{h^2}[\sigma\hat{\tilde{u}} + (1-2\sigma)\tilde{u} + \sigma\check{\tilde{u}}] = 0, \\ &\tilde{u}^0 = \tilde{\tilde{u}}, \quad \tilde{u}_t^0 + \frac{4\tau\,\sin^2\frac{\xi}{2}}{h^2}\left[\sigma\tilde{u}^1 + \left(\frac{1}{2}-\sigma\right)\tilde{u}^0\right] = \widetilde{\tilde{\tilde{u}}}. \end{aligned} \tag{22.16}$$

Умножим теперь уравнение (22.16) на τ^2 и перепишем в поточечном виде

$$\tilde{u}^{j+1} - 2\tilde{u}^j + \tilde{u}^{j-1} + \frac{4\tau^2}{h^2}\,\sin^2\frac{\xi}{2}[\sigma\tilde{u}^{j+1} + (1-2\sigma)\tilde{u}^j + \sigma\tilde{u}^{j-1}] = 0.$$

Введем обозначение

$$\frac{2\tau}{h}\,\sin\frac{\xi}{2} = \lambda \tag{22.17}$$

и напишем характеристическое уравнение разностного уравнения

$$(1+\sigma\lambda^2)q^2 - 2\left(1 + \left(\sigma - \frac{1}{2}\right)\lambda^2\right)q + (1+\sigma\lambda^2) = 0$$

или

$$q^2 - 2\frac{1+(\sigma-1/2)\lambda^2}{1+\sigma\lambda^2}q + 1 = 0.$$

Отсюда находим корни

$$\begin{aligned} q_{1,2} &= \frac{1+(\sigma-1/2)\lambda^2}{1+\sigma\lambda^2} \pm \frac{\sqrt{[1+(\sigma-1/2)\lambda^2+1+\sigma\lambda^2][1+(\sigma-1/2)\lambda^2-1-\sigma\lambda^2]}}{1+\sigma\lambda^2} = \\ &= \frac{1+(\sigma-1/2)\lambda^2 \pm \sqrt{[1+(\sigma-1/4)\lambda^2](-\lambda^2)}}{1+\sigma\lambda^2}. \end{aligned} \tag{22.18}$$

Если

$$1 + \left(\sigma - \frac{1}{4}\right)\lambda^2 > 0,$$

то корни q_1 и q_2 будут комплексными и равными по модулю 1. Если же

$$1 + \left(\sigma - \frac{1}{4}\right)\lambda^2 = 0,$$

то

$$q_{1,2} = \frac{1+(\sigma-1/2)\lambda^2}{1+\sigma\lambda^2} = \frac{-1/4\,\lambda^2}{1/4\,\lambda^2} = -1$$

и снова $|q_{1,2}| = 1$.

Выясним, когда

$$1 + \left(\sigma - \frac{1}{4}\right)\lambda^2 \geqslant 0,$$

или, что то же самое,

$$(1/4 - \sigma) \leqslant \frac{1}{\lambda^2}.$$

Это неравенство будет выполнено при всех $\xi \in (0, 2\pi]$, если (см. (22.17))

$$\frac{1}{4} - \sigma \leqslant \min_{\xi} \frac{1}{\lambda^2} = \frac{h^2}{4\tau^2}.$$

Но это условие совпадает с условием (22.14) теоремы 22.2, и, следовательно, $|q_{1,2}| = 1$.

Введем обозначение

$$q_{1,2} = e^{\pm i\varphi(\xi)} = \cos\varphi \pm i \sin\varphi,$$

где, согласно (22.18),

$$\cos\varphi = \frac{1 + (\sigma - 1/2)\lambda^2}{1 + \sigma\lambda^2}, \quad \sin\varphi = \frac{|\lambda|\sqrt{1 + (\sigma - 1/4)\lambda^2}}{1 + \sigma\lambda^2}, \tag{22.19}$$

и найдем решение задачи (22.16). Общее решение разностного уравнения (22.16) есть

$$\tilde{u}^j = c_1 \cos j\varphi + c_2 \sin j\varphi.$$

При $j = 0$

$$\tilde{u}^0 = c_1 = \widetilde{\bar{u}},$$

а при $j = 1$

$$\tilde{u}^1 = c_1 \cos\varphi + c_2 \sin\varphi.$$

Из вышесказанного следует, что

$$c_2 = \frac{\tilde{u}^1 - \widetilde{\bar{u}} \cos\varphi}{\sin\varphi}. \tag{22.20}$$

Далее, из второго начального условия (22.16)

$$\tilde{u}^1(1 + \sigma\lambda^2) = \left(1 + \left(\sigma - \frac{1}{2}\right)\lambda^2\right)\tilde{u}^0 + \tau\widetilde{\bar{\bar{u}}},$$

и, следовательно,

$$\tilde{u}^1 = \frac{1 + (\sigma - 1/2)\lambda^2}{1 + \sigma\lambda^2}\tilde{u}^0 + \frac{\tau\widetilde{\bar{\bar{u}}}}{1 + \sigma\lambda^2},$$

а с учетом (22.19)

$$\tilde{u}^1 = \cos\varphi\,\widetilde{\bar{u}} + \frac{\tau\widetilde{\bar{\bar{u}}}}{1 + \sigma\lambda^2}.$$

Подставляя это значение $\tilde{u}^1$ в (22.20), найдем, что

$$c_2 = \frac{\tau\widetilde{\bar{\bar{u}}}}{(1 + \sigma\lambda^2)\sin\varphi}.$$

Окончательно для решения задачи (22.16) получаем представление

$$\tilde{u}^j = \widetilde{\bar{u}} \cos j\varphi + \tau \frac{\widetilde{\bar{\bar{u}}}}{1+\sigma\lambda^2} \frac{\sin j\varphi}{\sin\varphi}. \tag{22.21}$$

Чтобы оценить правую часть (22.21), нам потребуется

Лемма 22.1. *При* $n \in \mathbb{N}$

$$|\sin(n\varphi)/\sin\varphi| \leqslant n.$$

Доказательство. Имеем

$$\left| \frac{e^{in\varphi} - e^{-in\varphi}}{e^{i\varphi} - e^{-i\varphi}} \right| = \left| \frac{e^{2in\varphi} - 1}{e^{2i\varphi} - 1} \frac{e^{i\varphi}}{e^{in\varphi}} \right| = \left| \frac{e^{2in\varphi} - 1}{e^{2i\varphi} - 1} \right| = \left| e^{2i(n-1)\varphi} + e^{2i(n-2)\varphi} + \cdots + 1 \right| \leqslant n.$$

Лемма доказана.

Используя лемму 22.1, из (22.21) находим, что

$$|\tilde{u}^j| \leqslant |\widetilde{\bar{u}}| + \tau j |\widetilde{\bar{\bar{u}}}| \leqslant |\widetilde{\bar{u}}| + T|\widetilde{\bar{\bar{u}}}|.$$

Отсюда

$$\|\tilde{u}^j\|_{L_2(0,2\pi)} \leqslant \|\widetilde{\bar{u}}\|_{L_2(0,2\pi)} + T\|\widetilde{\bar{\bar{u}}}\|_{L_2(0,2\pi)},$$

а с учетом равенства Парсеваля

$$\|u^j\|_{L_2^h} \leqslant \|\bar{u}\|_{L_2^h} + T\|\bar{\bar{u}}\|_{L_2^h}.$$

Теорема доказана.

Следствие 3. При $\sigma \geqslant 1/4$ условие (22.14) выполнено, и оценка (22.15) решения имеет место для любых h и τ. При $\sigma = 0$ (явная схема) условие (22.14) выполнено, если $\tau \leqslant h$, и поэтому оценка (22.15) имеет место только при указанном соотношении между τ и h.