РОССИЙСКАЯ ФЕДЕРАЦИЯ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Физический факультет
Кафедра астрономии и космической геодезии

АВДЮШЕВ Виктор Анатольевич

ЧИСЛЕННЫЕ МЕТОДЫ ИНТЕГРИРОВАНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

(Курс лекций)

Томск — 2009

ОГЛАВЛЕНИЕ

Введение

Почти все важные для современной практики дифференциальные уравнения, описывающие физические явления, не интегрируются аналитически. Поэтому для их решения прибегают к приближенным методам интегрирования, которые условно делят на аналитические и численные. Рассмотрим их основные принципы на примере метода малого параметра и метода разложения в ряд Тейлора.

Пусть нам необходимо решить уравнение $\mathbf{x}' = \mathbf{f}(t, \mathbf{x}, \mu)$ при условии $\mathbf{x}(0) = \mathbf{x}_0$. Здесь $\mathbf{x}$ — искомое решение; t — независимая переменная, а μ — некоторый малый параметр, причем такой, что при $\mu = 0$ уравнение имеет аналитическое решение, выраженное через элементарные функции.

Именно это решение выбирается в качестве опорного $\bar{\mathbf{x}}$ в аналитических методах, а приближенное решение представляется в виде усеченного ряда по степеням μ:

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i=1}^{p} \mathbf{x}_i \mu^i,$$

где p — порядок точности решения, а $\mathbf{x}_i$ — неизвестные коэффициенты, которые определяются путем подстановки в дифференциальное уравнение степенного ряда и уравнивания коэффициентов при одинаковых степенях μ. При этом конструкция решения не зависит от начального условия, т.е. решение можно рассматривать как общее. Известно, что оно будет пригодным на временном интервале порядка μ^{-p}. Впрочем, эти замечательные особенности аналитического решения дискредитируются существенными недостатками, которые уже в настоящее время не могут быть не приняты во внимание.

Обязательным условием для получения аналитического приближенного решения является наличие хорошего опорного, которое бы обеспечивало малость возмущений $\delta\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}$. Однако это не всегда возможно, как, например, в задачах астероидной динамики. Кроме того, для обеспечения достаточно высокой точности приближенного решения требуется большой порядок аппроксимации, что существенно усложняет решение. Именно эти недостатки становятся главной причиной, почему в современной практике все чаще прибегают к численной альтернативе в решении дифференциальных уравнений.

Задача: $\mathbf{x}' = \mathbf{f}(t, \mathbf{x}, \mu), \quad \mathbf{x}(0) = \mathbf{x}_0$

Приближенные методы	
Аналитический	Численный
Опорное решение	
$\bar{\mathbf{x}}$: $\bar{\mathbf{x}}' = \mathbf{f}(t, \bar{\mathbf{x}}, 0)$	$\bar{\mathbf{x}} = \mathbf{x}_0$
Приближенное решение	
$\mathbf{x} = \bar{\mathbf{x}} + \underbrace{\sum_i \mathbf{x}_i \mu^i}_{\delta\mathbf{x}}$	$\mathbf{x} = \bar{\mathbf{x}} + \underbrace{\sum_i \mathbf{x}_i t^i}_{\delta\mathbf{x}}$
1. Общее!	1. Частное
2. Для больших $t \sim \mu^{-p}$!	2. Для малых t
3. Не всегда $\exists\, \mathbf{x}$	3. Всегда (практич.) $\exists\, \mathbf{x}$!
4. Сложное	4. Простое!

В методе разложения в ряд Тейлора в качестве опорного $\bar{\mathbf{x}}$ выступает начальное $\mathbf{x}_0$, а решение представляется в виде усеченного степенного ряда по t:

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i=1}^{p} \mathbf{x}_i t^i,$$

где p — порядок метода, а $\mathbf{x}_i$ — коэффициенты ряда Тейлора. Несмотря на то, что численное решение пригодно лишь для малых t, эта трудность разрешается путем пошагового интегрирования: большой интервал времени делится на малые подынтервалы (шаги), на которых последовательно шаг за шагом определяются приближенные решения, при этом на каждом следующем шаге в качестве опорного выбирается решение предыдущего шага.

Отсюда главное достоинство численных методов состоит в том, что с их помощью почти всегда можно получить решение дифференциальных уравнений. Кроме того, методическая точность может быть повышена не только увеличением порядка аппроксимирующей формулы, но и уменьшением шага интегрирования. В связи с этим численное решение на шаге может быть довольно простым. Хотя необходимо заметить, что вследствие пошагового интегрирования компьютерная реализация каждого численного метода дает только частное решение.

1 Терминология

Дифференциальное уравнение первого порядка — это уравнение вида

$$x' = f(t, x) \tag{1.1}$$

с заданной функцией f. Здесь штрих означает полную производную по времени.

Функция $x = x(t)$ называется *решением* (1.1), если для всех t выполняется равенство

$$x'(t) = f(t, x(t)). \tag{1.2}$$

На самом деле уравнение (1.1) имеет не одно, а целое семейство решений с одним свободным параметром. Этот параметр определяется единственным образом, если задано *начальное условие*

$$x_0 = x(t_0). \tag{1.3}$$

Дифференциальное уравнение второго порядка имеет вид

$$x'' = f(t, x, x'). \tag{1.4}$$

Решение этого уравнения содержит уже два параметра, которые определяются из начального условия

$$x_0 = x(t_0), \quad x'_0 = x'(t_0). \tag{1.5}$$

Обычно при численном решении (1.5) вводят новые переменные $x_1 = x$ и $x_2 = x'$ и это уравнение приводят к *системе уравнений*

$$\begin{array}{ll} x'_1 = x_2, & x_1(t_0) = x_0, \\ x'_2 = f(t, x_1, x_2), & x_2(t_0) = x'_0. \end{array} \tag{1.6}$$

Следует заметить, что подобным способом можно привести любую систему, состоящую из уравнений второго порядка, к системе уравнений первого порядка. Очевидно, (1.6) является частным случаем системы уравнений первого порядка общего вида

$$\begin{array}{ll} x'_1 = f_1(t, x_1, \ldots, x_n), & x_1(t_0) = x_{10}, \\ \ldots & \\ x'_m = f_m(t, x_1, \ldots, x_m), & x_m(t_0) = x_{m0}; \end{array} \tag{1.7}$$

или

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \tag{1.8}$$

где $\mathbf{x} = (x_1, \ldots, x_m)^T$, а $\mathbf{f} = (f_1, \ldots, f_m)^T$. Дифференциальные уравнения (1.8) вместе со своими начальными условиями составляют *задачу Коши*.

В нашем курсе мы сосредоточим основное внимание именно на тех методах, которые были специально разработаны для численного интегрирования систем (1.8).

При этом мы будем полагать, что вектор-функция $\mathbf{f}$ удовлетворяет *условию Липшица*, т.е. для любых решений $\mathbf{x}_1$ и $\mathbf{x}_2$ справедливо неравенство

$$\|\mathbf{f}(t, \mathbf{x}_1) - \mathbf{f}(t, \mathbf{x}_2)\| \le L\|\mathbf{x}_1 - \mathbf{x}_2\|, \tag{1.9}$$

где L — так называемая *постоянная Липшица*.

2 Метод разложения в ряд Тейлора

Одним из старейших методов решения дифференциальных уравнений является *метод разложения в ряд Тейлора.*

Обратимся к скалярному случаю. Пусть требуется найти решение уравнения (1.1)

$$x' = f(t, x)$$

для заданного значения $t = t_0 + T$ при начальном условии $x_0 = x(t_0)$. Потребуем, чтобы f была аналитична в точке (t_0, x_0). Дифференцируя уравнение по t, будем иметь

$$x'' = f'_t + f'_x x', \quad x''' = f''_{tt} + 2f''_{tx}x' + f''_{xx}x'^2 + f'_x x'', \quad \ldots$$

Подставляя t_0 и x_0, последовательно получаем

$$x'(t_0), \quad x''(t_0), \quad x'''(t_0), \quad \ldots \tag{2.1}$$

Таким образом, при достаточно малом T приближенное решение в окрестности t_0 можно представить в виде ряда Тейлора

$$x(t_0 + T) = x_0 + \sum_{i=1}^{\infty} \frac{x^{(i)}(t_0)}{i!} T^i. \tag{2.2}$$

Оборвем (усечем) ряд (2.2) на p-м члене. Тогда получим приближенную формулу *p-порядка*

$$x(t_0 + T) \approx x_T = x_0 + \sum_{i=1}^{p} \frac{x^{(i)}(t_0)}{i!} T^i \tag{2.3}$$

с *погрешностью*

$$\Delta x = x(t_0 + T) - x_T = \sum_{i=p+1}^{\infty} \frac{x^{(i)}(t_0)}{i!} T^i = O(T^{p+1}). \tag{2.4}$$

Нетрудно видеть, что при малых T погрешность Δx (*ошибка усечения*) будет определяться главным образом первым членом ряда (2.4):

$$\frac{x^{(p+1)}(t_0)}{(p+1)!} T^{p+1}. \tag{2.5}$$

Его называют *главным членом погрешности.*

Если для $t = t_0 + T$ ряд Тейлора (2.2) расходится или ошибка приближенной формулы (2.3) довольно большая, то прибегают к так называемому *пошаговому интегрированию.* Разбивают $[t_0, t_0 + T]$ на малые отрезки $[t_n, t_{n+1}]$ $(n = 0, \ldots, N-1)$, где $t_N = t_0 + T$, и последовательно получают приближенные решения x_{n+1} до $x_N = x_T$ по формуле

$$x_{n+1} = x_n + \sum_{i=1}^{p} \frac{x^{(i)}(t_n)}{i!} h_{n+1}^i \quad (n = 0, \ldots, N-1), \tag{2.6}$$

где $h_{n+1} = t_{n+1} - t_n$ — n-й *шаг интегрирования.*

Рассмотрим частные случаи. Для $p = 1$ и $p = 2$ с постоянным шагом h имеем формулы

$$x_{n+1} = x_n + hf(t_n, x_n) \quad (\textit{метод Эйлера}), \tag{2.7}$$

$$x_{n+1} = x_n + hf(t_n, x_n) + \frac{h^2}{2}[f'_t(t_n, x_n) + f'_x(t_n, x_n)f(t_n, x_n)]. \tag{2.8}$$

Главным недостатком метода разложения в ряд Тейлора является то, что для получения производных (2.1) необходимо знать производные от f как функции t и x. Очевидно, если f — сложная функция, получение этих производных может представлять собой весьма утомительное занятие. В связи с этим методы разложения в ряд Тейлора редко используются на практике.

Впрочем, в небесной механике имеет место ряд важных с прикладной точки зрения задач, где, используя специфику дифференциальных уравнений, удается получить достаточно простые рекуррентные соотношения для временных производных в разложениях Тейлора. В 1957 г. К. Стеффенсен путем введения новых вспомогательных переменных вывел подобные соотношения для численного интегрирования уравнений планетной задачи.

Рассмотрим идею *метода Стеффенсена* на примере нормализованных уравнений задачи двух тел:

$$\mathbf{x}'' = -\frac{\mathbf{x}}{|\mathbf{x}|^3}. \tag{2.9}$$

Если ввести вспомогательные переменные $u = |\mathbf{x}|^{-3}$, $w = |\mathbf{x}|^{-2}$, $s = \mathbf{x} \cdot \dot{\mathbf{x}}$ и $\sigma = ws$, то уравнение (2.9) можно привести к дифференциально-алгебраической системе уравнений с квадратичными правыми частями

$$\mathbf{x}'' = -u\mathbf{x}, \quad u' = -3u\sigma, \quad w' = -2u\sigma, \quad s = \mathbf{x} \cdot \dot{\mathbf{x}}, \quad \sigma = ws. \tag{2.10}$$

Тогда, подставляя в (2.10) степенные ряды для $\mathbf{x}$, u, w, s и σ по малому параметру h и приравнивая коэффициенты при одинаковых степенях h, получим для них рекуррентные формулы

$$\mathbf{x}_{j+2} = -\frac{1}{(j+1)(j+2)} \sum_{i=0}^{j} \mathbf{x}_i u_{j-i}, \quad u_j = -\frac{3}{j} \sum_{i=0}^{j-1} u_i \sigma_{j-i-1}, \quad w_j = -\frac{2}{j} \sum_{i=0}^{j-1} w_i \sigma_{j-i-1},$$

$$s_j = \sum_{i=0}^{j} (i+1)\mathbf{x}_{i+1} \cdot \mathbf{x}_{j-1}, \quad \sigma_j = \sum_{i=0}^{j} w_i s_{j-1}. \tag{2.11}$$

При этом $\mathbf{x}_i = \mathbf{x}^{(i)}(t_0)/i!$ Коэффициенты нулевого порядка u_0, w_0, s_0 и σ_0 вычисляются по начальным данным $\mathbf{x}_0$ и $\dot{\mathbf{x}}_0$.

Подобный подход также применяется и для численного интегрирования более сложных дифференциальных уравнений небесной механики, в частности, задачи нескольких тел в классической постановке.

3 Методы Рунге–Кутты

В начале прошлого века К.Д.Т. Рунге, а затем К. Хойн и М.В. Кутта предложили подход, основанный на построении приближенной формулы, близкой к (2.3), не содержащей производных от f.

3.1 Первый метод Рунге

Пусть мы имеем задачу Коши

$$x' = f(t), \quad x(t_0) = x_0. \tag{3.1}$$

Для ее решения применим метод Эйлера (2.7). На первом шаге решение будет

$$x_1 = x_0 + hf(t_0). \tag{3.2}$$

Мы знаем, что метод Эйлера имеет ошибку $\Delta x = O(h^2)$. Однако порядок ошибки можно повысить, если в (3.2) функцию f вычислять не в t_0, а в $t_0+h/2$ (*правило средней точки*), т.е. если вместо (3.2) использовать схему интегрирования

$$x_1 = x_0 + hf(t_0 + h/2). \tag{3.3}$$

Действительно, разложим $f(t_0 + h/2)$ в степенной ряд Тейлора:

$$f(t_0 + h/2) = f(t_0) + \frac{h}{2}f'(t_0) + \frac{1}{2}\left(\frac{h}{2}\right)^2 f''(t_0) + \ldots;$$

и подставим в (3.3). Поскольку $x' = f$, будем иметь

$$x_1 = x(t_0) + hx'(t_0) + \frac{h^2}{2}x''(t_0) + \frac{h^3}{8}x'''(t_0) + \ldots \tag{3.4}$$

Мы видим, что (3.4) совпадает с разложением точного решения $x(t_0 + h)$ до третьего члена. Следовательно, ошибка формулы (3.3) будет порядка h^3.

Рунге (1895) поставил вопрос: нельзя ли распространить метод (3.3) на уравнение $x' = f(t, x)$? По аналогии с (3.3) первый шаг должен иметь вид

$$x_1 = x_0 + hf(t_0 + h/2, x(t_0 + h/2)).$$

Но какое значение взять для $x(t_0 + h/2)$? За неимением лучшего Рунге предлагает использовать один шаг метода Эйлера длины $h/2$. Тогда получаем

$$x_1 = x_0 + hf(t_0 + h/2, x_0 + h/2f(t_0, x_0)), \tag{3.5}$$

или

$$\begin{aligned} k_1 &= f(t_0, x_0), \\ k_2 &= f(t_0 + h/2, x_0 + h/2k_1), \\ x_1 &= x_0 + hk_2. \end{aligned} \tag{3.6}$$

Чтобы определить порядок решения (3.5), разложим его в ряд:

$$x_1 = x_0 + hf(t_0, x_0) + \frac{h^2}{2}(f'_t + f'_x f)(t_0, x_0) + \frac{h^3}{8}(f''_{tt} + 2f''_{tx}f + f''_{xx}f^2)(t_0, x_0) + \ldots; \quad (3.7)$$

и сравним с разложением точного решения $x(t_0 + h)$:

$$x(t_0+h) = x_0+hf(t_0, x_0)+\frac{h^2}{2}(f'_t+f'_x f)(t_0, x_0)+\frac{h^3}{6}(f''_{tt}+2f''_{tx}f+f''_{xx}f^2+f'_t f'_x+f'^2_x f)(t_0, x_0)+\ldots \quad (3.8)$$

Вычитая (3.7) из (3.8), получаем погрешность

$$x(t_0 + h) - x_1 = \frac{h^3}{24}(f''_{tt} + 2f''_{tx}f + f''_{xx}f^2 + 4(f'_t f'_x + f'^2_x f))(t_0, x_0) + \ldots = O(h^3). \quad (3.9)$$

Таким образом, мы получили усовершенствованный метод Эйлера второго порядка, где не требуется находить производные от f, какие необходимы в методе разложения в ряд Тейлора (2.8).

3.2 Методы Рунге–Кутты второго порядка

Методы Рунге–Кутты второго порядка имеют общий вид

$$x_1 = x_0 + h(b_1 f(t_0, x_0) + b_2 f(t_0 + hc_2, x_0 + ha_{21}f(t_0, x_0))), \quad (3.10)$$

или

$$\begin{aligned} k_1 &= f(t_0, x_0), \\ k_2 &= f(t_0 + hc_2, x_0 + ha_{21}k_1), \\ x_1 &= x_0 + h(b_1 k_1 + b_2 k_2); \end{aligned} \quad (3.11)$$

где a_{21}, b_1, b_2 и c_2 — постоянные, подлежащие определению. Нетрудно видеть, что схема Рунге (3.6) есть частный случай (3.11).

Определим постоянные схемы интегрирования так, чтобы она имела второй порядок. Разложим (3.10) в ряд:

$$x_1 = x_0 + h(b_1 + b_2)f(t_0, x_0) + h^2 b_2[c_2 f'_t + a_{21}f'_x f](t_0, x_0) + O(h^3);$$

и сравним с разложением точного решения:

$$x_1 = x_0 + hf(t_0, x_0) + h^2/2[f'_t + f'_x f](t_0, x_0) + O(h^3).$$

Следовательно, чтобы схема интегрирования имела второй порядок, постоянные должны удовлетворять уравнениям

$$b_1 + b_2 = 1, \quad b_2 c_2 = 1/2, \quad b_2 a_{21} = 1/2 \quad (\textit{условия второго порядка}). \quad (3.12)$$

Система уравнений (3.12) дает однопараметрическое семейство решений, которое можно представить в виде

$$b_1 = 1 - \alpha, \quad b_2 = \alpha, \quad a_{21} = c_2 = 1/(2\alpha),$$

где $\alpha \neq 0$ — свободный параметр. Например, при $\alpha = 1/2$ имеем *формулу Хойна*

$$x_1 = x_0 + \frac{1}{2}h[f(t_0, x_0) + f(t_0 + h, x_0 + hf(t_0, x_0))]. \tag{3.13}$$

Это *явная* формула, в которой требуются два вычисления функции f на шаге.

3.3 Явные методы Рунге–Кутты

Рунге продемонстрировал идею получения новых методов только для низких порядков. Но именно Кутта (1901) сформулировал общую схему того, что теперь называется методами Рунге–Кутты. Дадим определение этих методов применительно к системе (1.8).

Метод

$$\begin{aligned}
&\mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{x}_0),\\
&\mathbf{k}_2 = \mathbf{f}(t_0 + hc_2, \mathbf{x}_0 + ha_{21}\mathbf{k}_1),\\
&\mathbf{k}_3 = \mathbf{f}(t_0 + hc_3, \mathbf{x}_0 + h(a_{31}\mathbf{k}_1 + a_{32}\mathbf{k}_2)),\\
&\dots\\
&\mathbf{k}_s = \mathbf{f}(t_0 + hc_s, \mathbf{x}_0 + h(a_{s1}\mathbf{k}_1 + \dots + a_{s,s-1}\mathbf{k}_{s-1})),\\
&\mathbf{x}_1 = \mathbf{x}_0 + h(b_1\mathbf{k}_1 + \dots + b_s\mathbf{k}_s),
\end{aligned} \tag{3.14}$$

или

$$\mathbf{x}_1 = \mathbf{x}_0 + h\sum_{j=1}^{s} b_j\mathbf{k}_j, \quad \mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{x}_0), \quad \mathbf{k}_i = \mathbf{f}(t_0 + hc_i, \mathbf{x}_0 + h\sum_{j=1}^{i-1} a_{ij}\mathbf{k}_j) \quad (i = 2, \dots, s)$$

называется *s-этапным (s-стадийным) явным методом Рунге–Кутты* для задачи (1.8).

Обычно коэффициенты c_i удовлетворяют условиям

$$c_2 = a_{21}, \quad c_3 = a_{31} + a_{32}, \quad \dots, \quad c_s = a_{s1} + \dots + a_{s,s-1}, \quad \text{или} \quad c_i = \sum\nolimits_j a_{ij}.$$

Кроме того, поскольку согласно (3.14)

$$\mathbf{f}(t_0, \mathbf{x}_0) = \lim_{h\to 0} \frac{\mathbf{x}_1 - \mathbf{x}_0}{h} = \mathbf{f}(t_0, \mathbf{x}_0)\sum_{j=1}^{s} b_j, \quad \text{то} \quad \sum_{j=1}^{s} b_j = 1.$$

Метод Рунге–Кутты (3.14) имеет *порядок* p, если для (1.8)

$$\|\mathbf{x}(t_0 + h) - \mathbf{x}_1\| = o(h^p), \tag{3.15}$$

т.е. если ряды Тейлора для точного решения $\mathbf{x}(t_0 + h)$ и для $\mathbf{x}_1$ совпадают до члена h^p включительно.

Символически методы Рунге–Кутты (3.14) принято представлять в виде таблицы

$$\begin{array}{c|ccccc}
0 & & & & & \\
c_2 & a_{21} & & & & \\
c_3 & a_{31} & a_{32} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
c_s & a_{s1} & a_{s2} & \dots & a_{s,s-1} & \\
\hline
 & b_1 & b_2 & \dots & b_{s-1} & b_s
\end{array} \tag{3.16}$$

В качестве примера рассмотрим явный 4-этапный *классический метод Рунге–Кутты* 4-го порядка для произвольного шага $n+1$:

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(t_n, \mathbf{x}_n),\\
\mathbf{k}_2 &= \mathbf{f}(t_n + h/2, \mathbf{x}_n + h\mathbf{k}_1/2),\\
\mathbf{k}_3 &= \mathbf{f}(t_n + h/2, \mathbf{x}_n + h\mathbf{k}_2/2),\\
\mathbf{k}_4 &= \mathbf{f}(t_n + h, \mathbf{x}_n + h\mathbf{k}_3),\\
\mathbf{x}_{n+1} &= \mathbf{x}_n + \frac{h}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4).
\end{aligned} \tag{3.17}$$

Классический метод имеет табличный вид

$$\begin{array}{c|cccc}
0 & & & & \\
1/2 & 1/2 & & & \\
1/2 & 0 & 1/2 & & \\
1 & 0 & 0 & 1 & \\
\hline
 & 1/6 & 2/6 & 2/6 & 1/6
\end{array} \tag{3.18}$$

3.4 Условия порядка

Коэффициенты a_{ij}, b_i и c_i определяются из так называемых *условий порядка*. Эти условия получаются из разложения приближенного решения в ряд Тейлора и сравнения его с разложением точного решения. При этом коэффициенты определяются таким образом, чтобы получить в разложениях как можно больше попарно одинаковых первых членов.

Подобные манипуляции мы выполняли для определения коэффициентов в методах Рунге–Кутты второго порядка. В конечном итоге мы получили условия второго порядка (3.12). Таким же образом нетрудно вывести условия третьего порядка. Они будут

$$\sum_i b_i = 1, \quad 2\sum_{i,j} b_i a_{ij} = 1, \quad 3\sum_{i,j,k} b_i a_{ij} a_{ik} = 1, \quad 6\sum_{i,j,k} b_i a_{ij} a_{jk} = 1.$$

Условия для более высоких порядков чрезвычайно громоздки, и поэтому мы не будем их расматривать.

Очевидно, именно условия порядка задают связь между порядком метода p и числом этапов (числом обращения к функции правых частей) s. К сожалению, только до четвертого порядка $p = s$. Для методов более высокого порядка $p < s$.

3.5 Оценка погрешности и выбор длины шага

Локальная погрешность метода — это погрешность на одном шаге. Как мы определили ранее, метод p-порядка имеет на каждом шаге локальную погрешность

$$\|\mathbf{e}_n\| \equiv \|\mathbf{x}(t_{n-1} + h_n) - \mathbf{x}_n\| = o(h_n^p).$$

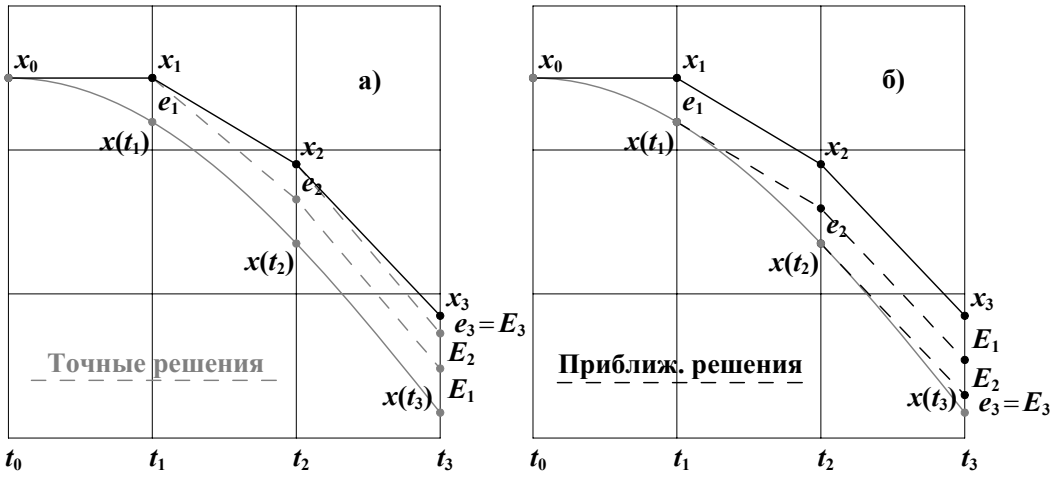

Рисунок 3.1 — Способы перенесения погрешностей

Это означает, что существует такая постоянная C, что для любых h_n справедливо неравенство

$$\|\mathbf{e}_n\| \le C h_n^{p+1}. \tag{3.19}$$

Глобальной погрешностью называется погрешность численного решения после выполнения нескольких шагов.

3.5.1 Оценка глобальной погрешности

Один из способов оценки глобальной погрешности является суммирование всех оценок для локальных погрешностей, *перенесенных* на конец интервала интегрирования. Погрешность на некотором n-м шаге переносится посредством выполнения оставшихся $N - n$ шагов с использованием либо точных решений, либо приближенных (рис. 3.1).

В способе переноса погрешностей (рис. 3.1, б) на некотором n-м шаге рассматривают еще одно (ошибочное) решение $\mathbf{z}_n$, которое отличается от решения $\mathbf{x}_n$ на локальную ошибку $\mathbf{e}_n$, т.е. $\mathbf{z}_n = \mathbf{x}_n + \mathbf{e}_n$. Для $\mathbf{z}_n$ схема интегрирования запишется как

$$\begin{aligned}
&\mathbf{l}_1 = \mathbf{f}(t_n, \mathbf{z}_n),\\
&\mathbf{l}_2 = \mathbf{f}(t_n + h_{n+1}c_2, \mathbf{z}_n + h_{n+1}a_{21}\mathbf{l}_1),\\
&\dots\\
&\mathbf{z}_{n+1} = \mathbf{z}_n + h_{n+1}(b_1\mathbf{l}_1 + \dots + b_s\mathbf{l}_s).
\end{aligned}$$

Тогда для норм разностей будем иметь

$$\begin{aligned}
&\|\mathbf{k}_1 - \mathbf{l}_1\| \le L\|\mathbf{x}_n - \mathbf{z}_n\|,\\
&\|\mathbf{k}_2 - \mathbf{l}_2\| \le L(1 + |a_{21}|h_{n+1}L)\|\mathbf{x}_n - \mathbf{z}_n\|,\\
&\dots\\
&\|\mathbf{x}_{n+1} - \mathbf{z}_{n+1}\| \le \|\mathbf{x}_n - \mathbf{z}_n\| + h_{n+1}(|b_1|\|\mathbf{k}_1 - \mathbf{l}_1\| + \dots + |b_s|\|\mathbf{k}_s - \mathbf{l}_s\|),
\end{aligned} \tag{3.20}$$

где $L \geq 0$ — постоянная Липшица для $\mathbf{f}$. Если ввести постоянную

$$\Lambda = L\left(\sum_i |b_i| + hL\sum_{i,j}|b_i a_{ij}| + h^2L^2\sum_{i,j,k}|b_i a_{ij}a_{jk}| + \dots\right) \geq 0, \quad \text{где} \quad h = \max h_n;$$

последнее неравенство (3.20) можно переписать как

$$\|\mathbf{x}_{n+1} - \mathbf{z}_{n+1}\| \leq (1 + h_{n+1}\Lambda)\|\mathbf{x}_n - \mathbf{z}_n\| \leq \exp(h_{n+1}\Lambda)\|\mathbf{x}_n - \mathbf{z}_n\|.$$

Отсюда оценка для перенесенной погрешности на n-м шаге будет

$$\|\mathbf{E}_n\| \leq \exp((t_N - t_n)\Lambda)\|\mathbf{e}_n\| \leq \exp((t_N - t_n)\Lambda)Ch^p h_n,$$

а оценка для глобальной погрешности

$$\|\mathbf{E}\| \leq \sum_{n=1}^{N}\|\mathbf{E}_n\| \leq Ch^p\sum_{n=1}^{N}\exp((t_N - t_n)\Lambda)h_n \leq Ch^p\int_{t_0}^{t_N}\exp((t_N - t)\Lambda)\mathrm{d}t.$$

Наконец, получаем

$$\|\mathbf{E}\| \leq h^p\frac{C}{\Lambda}(\exp((t_N - t_0)\Lambda) - 1), \tag{3.21}$$

или

$$\|\mathbf{E}\| = O(h^p).$$

Отсюда следует, что все явные методы Рунге–Кутты *сходящиеся*, т.е.

$$\lim_{h\to 0}\mathbf{x}_N = \mathbf{x}(t_N).$$

Оценку, подобную (3.21), для глобальной погрешности можно также получить и для неявных методов Рунге–Кутты, которые мы рассмотрим позже. Поэтому и все неявные методы — сходящиеся.

3.5.2 Практическая оценка погрешности. Экстраполяция численного решения.

Для практики оценка погрешности необходима с целью выбора оптимального шага: с одной стороны, достаточно малого для обеспечения требуемой точности, но с другой — достаточно большого во избежание бесполезной вычислительной работы.

Очевидно, полученные нами оценки погрешностей (3.19) и (3.21) не представляют практической ценности, поскольку являются качественными.

Самый старый способ, который еще использовал Рунге для практической оценки точности численного решения, состоит в повторении вычислений с уменьшенной вдвое длиной шага и сравнении результатов. Рассмотрим этот способ подробнее.

Предположим, мы используем некоторый метод Рунге–Кутты порядка p. Отправляясь от начальных значений $\mathbf{x}_0$, получим сначала численное решение $\mathbf{x}_{1(h)}$ на шаге h. Затем повторим интегрирование от $\mathbf{x}_0$, последовательно выполнив два шага длиной $h/2$. При этом получим соответственно два численных решения $\mathbf{x}_{1(h)}$ и $\mathbf{x}_{2(h/2)}$.

Разложим погрешность $\mathbf{e}_{1(h)}$ на шаге h в степенной ряд. Поскольку наш метод имеет порядок p, то разложение представимо в виде

$$\mathbf{e}_{1(h)} = \mathbf{x}(t_0 + h) - \mathbf{x}_{1(h)} = \mathbf{C}h^{p+1} + O(h^{p+2}), \tag{3.22}$$

где $\mathbf{C}$ выражается через производные $(p+1)$-порядка от $\mathbf{x}$, вычисленные в t_0.

Погрешность $\mathbf{e}_{1(h/2)}$ решения $\mathbf{x}_{1(h/2)}$, очевидно, будет иметь вид (3.22), но для $h/2$:

$$\mathbf{e}_{1(h/2)} = \mathbf{x}(t_0 + h/2) - \mathbf{x}_{1(h/2)} = \mathbf{C}(h/2)^{p+1} + O(h^{p+2}). \tag{3.23}$$

Погрешность второго шага состоит из двух частей: из перенесенной погрешности первого шага $\mathbf{E}_{1(h/2)}$ и локальной погрешности второго $\mathbf{e}_{2(h/2)}$.

Найдем перенесенную погрешность способом, представленным на рис. 3.1, а. Составим для нее дифференциальное уравнение. Рассмотрим две одинаковые системы уравнений вида (1.8), но с разными начальными условиями, отличающимися на начальную величину локальной погрешности:

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}_{1(h/2)} = \mathbf{x}(t_0 + h/2); \quad \text{и} \quad \mathbf{z}' = \mathbf{f}(t, \mathbf{z}), \quad \mathbf{z}_{1(h/2)} = \mathbf{x}_{1(h/2)} + \mathbf{e}_{1(h/2)}.$$

Следовательно, для перенесенной погрешности $\mathbf{E} \equiv \mathbf{z} - \mathbf{x}$ будем иметь дифференциальное уравнение

$$\mathbf{E}' = \mathbf{f}(t, \mathbf{z}) - \mathbf{f}(t, \mathbf{x}) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\mathbf{E} + O(\mathbf{E}^2), \quad \mathbf{E}_0 = \mathbf{e}_{1(h/2)}.$$

Заметим, что $O(\mathbf{E}^2) = O(h^{2p+2})$ — достаточно малая величина, которой здесь можно пренебречь. Решая это уравнение на шаге $h/2$ методом разложения в ряд Тейлора, получим необходимую перенесенную погрешность, представимую в виде

$$\mathbf{E}_{1(h/2)} = \left(\mathbf{I} + \frac{h}{2}\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + O(h^2)\right)\mathbf{e}_{1(h/2)} = (\mathbf{I} + O(h))\mathbf{e}_{1(h/2)}. \tag{3.24}$$

Локальная погрешность $\mathbf{e}_{2(h/2)}$ на втором шаге будет иметь вид (3.23), но с $\mathbf{C}$, вычисленной при $\mathbf{x}_{1(h/2)} = \mathbf{x}_0 + O(h)$, т.е. с $\mathbf{C} + O(h)$.

Наконец, полная (глобальная) погрешность на втором шаге будет

$$\mathbf{E}_{(h/2)} = \mathbf{E}_{1(h/2)} + \mathbf{e}_{2(h/2)} = (\mathbf{I} + O(h))\mathbf{C}(h/2)^{p+1} + (\mathbf{C} + O(h))(h/2)^{p+1} + O(h^{p+2}),$$

откуда

$$\mathbf{E}_{(h/2)} = \mathbf{x}(t_0 + h) - \mathbf{x}_{2(h/2)} = 2\mathbf{C}(h/2)^{p+1} + O(h^{p+2}). \tag{3.25}$$

Из (3.22) и (3.25) получаем константу $\mathbf{C}$, а затем и погрешность $\mathbf{e}_{1(h)}$:

$$\mathbf{e}_{1(h)} = \mathbf{x}(t_0 + h) - \mathbf{x}_{1(h)} = \frac{\mathbf{x}_{1(h)} - \mathbf{x}_{2(h/2)}}{(1/2)^p - 1} + O(h^{p+2}). \tag{3.26}$$

Нетрудно видеть, что формула (3.26) позволяет по $\mathbf{x}_{2(h/2)}$ и $\mathbf{x}_{1(h)}$ получить *экстраполяцию* численного решения

$$\hat{\mathbf{x}}_1 = \mathbf{x}_{1(h)} + \frac{\mathbf{x}_{1(h)} - \mathbf{x}_{2(h/2)}}{(1/2)^p - 1}, \tag{3.27}$$

лучше представляющую точное решение $\mathbf{x}(t_0+h)$. Действительно, согласно (3.26)

$$\mathbf{x}(t_0+h) - \hat{\mathbf{x}}_1 = O(h^{p+2}),$$

т.е. решение $\hat{\mathbf{x}}_1$ (3.27) аппроксимирует точное с порядком $p+1$, когда $\mathbf{x}_{2(h/2)}$ и $\mathbf{x}_{1(h)}$ — с порядком p.

Таким образом, формула (3.26) дает простой способ оценки погрешности, тогда как (3.27) позволяет дополнительно повысить точность численного интегрирования на один порядок.

Прежде чем переходить к следующему разделу, выведем общую формулу экстраполяции. Рассмотрим два решения $\mathbf{x}_{N(h/N)}$ и $\mathbf{x}_{M(h/M)}$, аппроксимирующие точное решение $\mathbf{x}(t_0+h)$ и полученные на разных равномерных сетках с шагами h/N и h/M соответственно. Для погрешностей этих решений можно получить оценки, аналогичные (3.25):

$$\mathbf{E}_{(h/N)} = \mathbf{x}(t_0+h) - \mathbf{x}_{N(h/N)} = N\mathbf{C}(h/N)^{p+1} + O(h^{p+2}),$$

$$\mathbf{E}_{(h/M)} = \mathbf{x}(t_0+h) - \mathbf{x}_{M(h/M)} = M\mathbf{C}(h/M)^{p+1} + O(h^{p+2}),$$

из которых после исключения $\mathbf{C}$ будем иметь

$$\mathbf{x}(t_0+h) - \mathbf{x}_{N(h/N)} = \frac{\mathbf{x}_{N(h/N)} - \mathbf{x}_{M(h/M)}}{(N/M)^p - 1} + O(h^{p+2}).$$

Следовательно, экстраполированное решение порядка $p+1$ можно представить в виде

$$\hat{\mathbf{x}}_1 = \mathbf{x}_{N(h/N)} + \frac{\mathbf{x}_{N(h/N)} - \mathbf{x}_{M(h/M)}}{(N/M)^p - 1} \quad (\textit{экстраполяция Ричардсона}). \tag{3.28}$$

Нетрудно видеть, что (3.27) есть частный случай (3.28) при $N=1$ и $M=2$.

3.5.3 Выбор шага

Как мы уже знаем, возможность использования переменного шага необходима для обеспечения допустимой величины локальной погрешности, что весьма важно при сложном поведении функции $\mathbf{f}$. Мы предлагаем следующий алгоритм выбора шага.

Пусть $\|\mathbf{e}\|_{tol} = \text{const}$ — допустимая величина локальной погрешности. Нам необходимо подобрать такой шаг интегрирования $\bar{h}$, который бы обеспечивал заданную величину локальной погрешности для некоторого метода Рунге–Кутты порядка p.

Итак, выполняем интегрирование на шаге h и на двух шагах $h/2$. В результате получаем два решения $\mathbf{x}_{1(h)}$ и $\mathbf{x}_{2(h/2)}$, по которым затем вычисляем величину локальной погрешности на шаге h, используя главный член погрешности (3.26):

$$\|\mathbf{e}\|_{cal} = \frac{\|\mathbf{x}_{1(h)} - \mathbf{x}_{2(h/2)}\|}{1-(1/2)^p}.$$

Тогда новый шаг $\bar{h}$ определяем по формуле

$$\bar{h} = h\left(\frac{\|\mathbf{e}\|_{tol}}{\|\mathbf{e}\|_{cal}}\right)^{\frac{1}{p+1}}, \tag{3.29}$$

которая фактически получается из оценок

$$\|\mathbf{e}\|_{tol} = \mathbf{C}\bar{h}^{p+1} \quad \text{и} \quad \|\mathbf{e}\|_{cal} = \mathbf{C}h^{p+1}.$$

Далее обычно применяют экстраполяцию (3.27), с тем чтобы дополнительно повысить точность численного решения.

Следует заметить, если поведение решения довольно гладкое, то оценку (3.29) можно применять не для текущего, а для следующего шага, т.е. использовать формулу

$$h_{n+1} = h_n \left(\frac{\|\mathbf{e}\|_{tol}}{\|\mathbf{e}\|_{cal}} \right)^{\frac{1}{p+1}}. \tag{3.30}$$

3.6 Вложенные методы Рунге–Кутты

Вложенные методы Рунге–Кутты основаны на ином способе оценки локальной погрешности $\|\mathbf{e}\|_{cal}$ для выбора переменного шага. Их идея состоит в том, чтобы вместо использования экстраполяции Ричардсона построить такие формулы Рунге–Кутты, которые бы кроме основного решения $\mathbf{x}_1$ содержали выражения для вспомогательного решения $\hat{\mathbf{x}}_1$ более высокого порядка. Тогда по последнему можно было бы оценить погрешность $\mathbf{x}_1$.

Следует заметить, что для алгоритма выбора переменного шага можно использовать решение $\hat{\mathbf{x}}_1$ более низкого порядка и по величине его погрешности относительно основного решения $\mathbf{x}_1$ выбирать шаг интегрирования. Хотя тогда мы не будем иметь информацию о точности основного решения $\mathbf{x}_1$.

Вложенные методы Рунге–Кутты имеют табличный вид

$$\begin{array}{c|ccccc}
0 & & & & & \\
c_2 & a_{21} & & & & \\
c_3 & a_{31} & a_{32} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
c_s & a_{s1} & a_{s2} & \dots & a_{s,s-1} & \\
\hline
 & b_1 & b_2 & \dots & b_{s-1} & b_s \\
\hline
 & \hat{b}_1 & \hat{b}_2 & \dots & \hat{b}_{s-1} & \hat{b}_s
\end{array} \tag{3.31}$$

В схеме интегрирования основное решение

$$\mathbf{x}_1 = \mathbf{x}_0 + h(b_1\mathbf{k}_1 + \ldots + b_s\mathbf{k}_s)$$

имеет порядок p, тогда как вспомогательное

$$\hat{\mathbf{x}}_1 = \mathbf{x}_0 + h(\hat{b}_1\mathbf{k}_1 + \ldots + \hat{b}_s\mathbf{k}_s)$$

— порядок $q = p+1$ (либо $q = p-1$). При этом ошибка основного решения (либо вспомогательного) оценивается как $\|\mathbf{e}\|_{cal} = \|\mathbf{x}_1 - \hat{\mathbf{x}}_1\| = h\|(b_1 - \hat{b}_1)\mathbf{k}_1 + \ldots + (b_s - \hat{b}_s)\mathbf{k}_s\|$, что дает хорошую оценку главного члена методической погрешности (2.5). Вложенные методы

обычно называют по фамилии автора с указанием порядков основного и вспомогательного решений p (q). На практике широко используются вложенные методы Мерсона, Инглэнда, Фельдберга, а также Дормана–Принса. В качестве примера приведем схему Мерсона 4 (5):

$$\begin{array}{c|ccccc}
0 & & & & & \\
1/3 & 1/3 & & & & \\
1/3 & 1/6 & 1/6 & & & \\
1/2 & 1/8 & 0 & 3/8 & & \\
1 & 1/2 & 0 & -3/2 & 2 & \\
\hline
\mathbf{x}_1 & 1/2 & 0 & -3/2 & 2 & 0 \\
\hline
\hat{\mathbf{x}}_1 & 1/6 & 0 & 0 & 2/3 & 1/6
\end{array} \tag{3.32}$$

3.7 Неявные методы Рунге–Кутты

Метод

$$\begin{aligned}
&\mathbf{k}_1 = \mathbf{f}(t_0 + hc_1, \mathbf{x}_0 + h(a_{11}\mathbf{k}_1 + \ldots + a_{1s}\mathbf{k}_s)), \\
&\ldots \\
&\mathbf{k}_s = \mathbf{f}(t_0 + hc_s, \mathbf{x}_0 + h(a_{s1}\mathbf{k}_1 + \ldots + a_{ss}\mathbf{k}_s)), \\
&\mathbf{x}_1 = \mathbf{x}_0 + h(b_1\mathbf{k}_1 + \ldots + b_s\mathbf{k}_s),
\end{aligned} \tag{3.33}$$

или

$$\mathbf{x}_1 = \mathbf{x}_0 + h\sum_{j=1}^{s} b_j\mathbf{k}_j, \quad \mathbf{k}_i = \mathbf{f}(t_0 + hc_i, \mathbf{x}_0 + h\sum_{j=1}^{s} a_{ij}\mathbf{k}_j) \quad (i = 1, \ldots, s)$$

называется *s-этапным (s-стадийным) неявным методом Рунге–Кутты* для задачи (1.8).

Простые примеры неявных методов Рунге–Кутты — это *неявный метод Эйлера* (первого порядка)

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h\mathbf{f}(t_{k+1}, \mathbf{x}_{k+1}) \tag{3.34}$$

и *метод трапеций* (второго порядка)

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{h}{2}[\mathbf{f}(t_k, \mathbf{x}_k) + \mathbf{f}(t_{k+1}, \mathbf{x}_{k+1})]. \tag{3.35}$$

В соответствии с (3.33) неявные методы будут иметь табличный вид

$$\begin{array}{c|cccc}
c_1 & a_{11} & a_{12} & \ldots & a_{1s} \\
c_2 & a_{21} & a_{22} & \ldots & a_{2s} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_s & a_{s1} & a_{s2} & \ldots & a_{ss} \\
\hline
 & b_1 & b_2 & \ldots & b_s
\end{array}$$

Главным недостатком неявных методов Рунге–Кутты является то, что величины $\mathbf{k}_i$ задаются в (3.33) неявным образом, и поэтому их нужно находить итерационным

способом. Для этого, как правило, используют *метод простых итераций*:

$$\begin{aligned}
&\mathbf{k}_1^{l+1} = \mathbf{f}(t_0 + hc_1, \mathbf{x}_0 + h(a_{11}\mathbf{k}_1^l + \ldots + a_{1s}\mathbf{k}_s^l)),\\
&\ldots\\
&\mathbf{k}_s^{l+1} = \mathbf{f}(t_0 + hc_s, \mathbf{x}_0 + h(a_{s1}\mathbf{k}_1^l + \ldots + a_{ss}\mathbf{k}_s^l)).
\end{aligned} \tag{3.36}$$

Введем $\mathbf{K} = (\mathbf{k}_1, \ldots, \mathbf{k}_s)$ и норму $\|\mathbf{K}\| = \max_i \|\mathbf{k}_i\|$. Тогда (3.36) можно записать в виде $\mathbf{K}^{l+1} = \mathbf{F}(\mathbf{K}^l)$, где $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_s)$, а $\mathbf{f}_i(\mathbf{K}) = \mathbf{f}(t_0 + hc_i, \mathbf{x}_0 + h(a_{i1}\mathbf{k}_1 + \ldots + a_{is}\mathbf{k}_s))$ $(i = 1, \ldots, s)$. Из условия Липшица (1.9) для двух произвольных $\mathbf{K}_1$ и $\mathbf{K}_2$ будет справедливо неравенство

$$\|\mathbf{F}(\mathbf{K}_1) - \mathbf{F}(\mathbf{K}_2)\| \le hL \max_i \sum\nolimits_j |a_{ij}| \|\mathbf{K}_1 - \mathbf{K}_2\|. \tag{3.37}$$

Для сходимости метода простых итераций отображение $\mathbf{F}$ должно быть сжимающим, что согласно (3.37) оказывается возможным при условии, когда

$$h < \frac{1}{L \max_i \sum_j |a_{ij}|}. \tag{3.38}$$

Очевидно, для нахождения $\mathbf{k}_i$ необходимо, по меньшей мере, две итерации. Отсюда все неявные s-этапные методы требуют вычислений функций $\mathbf{f}$ на шаге не меньше, чем $2s$. (В данном случае число s и число вычислений функций $\mathbf{f}$ не одно и то же, как в явных методах.) Таким образом, неявные методы должны работать медленнее, нежели явные с тем же числом этапов s.

Несмотря на это, привлекательность неявных методов состоит в том, что при всех s существуют такие методы, которые имеют порядок $p = 2s$. Ж. Кунцман (1961) и Дж. Бутчер (1964) показали, что такой порядок p достигается путем специального выбора коэффициентов c_i.

Рассмотрим два метода 8-го порядка: явный Дормана–Принса ($s = 13$) и неявный Кунцмана–Бутчера ($s = 4$). На практике начальные приближения $\mathbf{k}_i^{(0)}$ в неявных методах получают из соответствующих величин на предыдущем шаге. Тогда при хорошем выборе начальных приближений для сходимости итерационного процесса может потребоваться всего 2–3 итерации. Поэтому можно считать, что метод Кунцмана–Бутчера будет вычислять функцию $\mathbf{f}$ 8–12 раз на шаге, тогда как метод Дормана–Принса — 13. Следовательно, благодаря своему замечательному свойству неявный метод будет работать быстрее. Кроме того, необходимо заметить, что с повышением порядка p разность $s - p$ для явных методов увеличивается, поэтому указанное преимущество неявных методов будет только возрастать.

3.8 Порядок и шаг интегрирования при компьютерной реализации метода

Теоретически совместное увеличение порядка и уменьшение шага метода неограниченно повышают методическую точность численных результатов интегрирования. Однако при компьютерной реализации в арифметике с определенной точностью вследствие ошибок округления существуют такие значения параметров интегрирования, которые

дают предельно высокую методическую точность ввиду того, что методические ошибки становятся соизмеримыми с ошибками округления, и в этом случае не имеет смысла предпринимать какие-либо дальнейшие попытки получить более высокую точность численного интегрирования.

Получим оценки оптимальной пары порядок–шаг интегрирования применительно к численному решению дифференциальных уравнений круговой двумерной задачи двух тел:

$$\mathbf{r}' = \mathbf{v}, \quad \mathbf{v}' = -\frac{\mu}{|\mathbf{r}|^3}\mathbf{r}.$$

Здесь $\mathbf{r} = (r_1, r_2)$ и $\mathbf{v} = (v_1, v_2)$ — векторы положения и скорости соответственно, а μ — гравитационный параметр. Поскольку $|\mathbf{r}| = a = \text{const}$, будем полагать, что величина $\mu/|\mathbf{r}|^3 = \nu^2 = \text{const}$, т.е. решение $\mathbf{x} = (\mathbf{r}, \mathbf{v})$ описывается уравнениями гармонического осциллятора с частотой ν и может быть записано в виде

$$x_1 = r_1 = a\cos\nu t, \quad x_2 = r_2 = a\sin\nu t, \quad x_3 = v_1 = -a\nu\sin\nu t, \quad x_4 = v_2 = a\nu\cos\nu t.$$

Оценим методическую ошибку $|\mathbf{e}|_M$ по главному члену погрешности (2.5):

$$|\mathbf{e}|_M = \frac{a\sqrt{1+\nu^2}(\nu h)^{p+1}}{(p+1)!}, \tag{3.39}$$

где использована формула $|\mathbf{x}^{(p+1)}| = a\nu^{p+1}\sqrt{1+\nu^2}$. Согласно формулам методов Рунге–Кутты ошибку округления $|\mathbf{e}|_R$ можно оценить как

$$|\mathbf{e}|_R = \varepsilon|\mathbf{x}| = \varepsilon a\sqrt{1+\nu^2}, \tag{3.40}$$

где ε — машинный эпсилон.

Очевидно, что не имеет смысла выбирать такие шаг и порядок интегрирования, при которых методическая ошибка будет меньше ошибки округления. Из условия $|\mathbf{e}|_M = |\mathbf{e}|_R$ получим отношение между оптимальными параметрами интегрирования h и p:

$$h_\varphi^{p+1} = \varepsilon(p+1)!, \tag{3.41}$$

где $h_\varphi = \nu h$ — шаг по долготе $\varphi = \nu t$, соответствующий шагу h.

Отношение (3.41) дает нижнюю границу для шага h, в то время как для неявных методов имеет место верхняя граница, задаваемая условием (3.38). Если положить, что $\max_i \sum_j |a_{ij}| = 1$ (в действительности максимум близок к единице), то получим следующее ограничение на шаг интегрирования $h < 1/L$ или $h_\varphi = \nu/L$. Оценим постоянную Липшица L для исследуемой задачи.

Рассмотрим отношение

$$\frac{|\delta\mathbf{f}|^2}{|\delta\mathbf{x}|^2} = \frac{\nu^4|\delta\mathbf{r}|^2 + |\delta\mathbf{v}|^2}{|\delta\mathbf{r}|^2 + |\delta\mathbf{v}|^2},$$

где $\delta\mathbf{x}$, $\delta\mathbf{r}$ и $\delta\mathbf{v}$ — всевозможные разности векторов в соответствующих переменных. Принимая $|\delta\mathbf{r}| = \xi\cos\psi$ и $|\delta\mathbf{v}| = \xi\sin\psi$, где $\xi \geq 0$ и $0 \geq \psi \geq \pi/2$, будем иметь

$$\frac{|\delta\mathbf{f}|^2}{|\delta\mathbf{x}|^2} = (\nu^4 - 1)\cos^2\psi + 1.$$

Отсюда нетрудно видеть, что все значения отношения лежат между 1 и ν^4. Следовательно, согласно (1.9) в качестве постоянной Липшица можно выбрать $L = \max(1, \nu^2)$. Тогда получаем верхнюю границу шага

$$h_\varphi < \frac{\nu}{\max(1, \nu^2)} \leq 1.$$

Таким образом, в лучшем случае, а именно при $\nu = 1$, когда верхняя граница максимальна, шаг интегрирования h_φ должен удовлетворять неравенствам

$$\varepsilon(p+1)! < h_\varphi^{p+1} < 1.$$

Очевидно, условие

$$\varepsilon(p+1)! > 1 \tag{3.42}$$

означает, что порядок метода завышен и использование такого метода при вычислениях в арифметике с точностью ε не разумно в том смысле, что ту же точность результатов интегрирования можно получить с использованием методов более низких порядков. Оптимальные порядки p неявных методов Рунге–Кутты для различных ε, соответствующих одинарной, двойной, расширенной и четверной точности, представлены в таблице. Хотя следует иметь в виду, что эти порядки получены для задачи с $\nu = 1$. В ином случае они могут быть меньше.

Максимально возможные порядки p неявных методов для различных ε

ε	$1.1 \cdot 10^{-7}$	$2.2 \cdot 10^{-16}$	$1.1 \cdot 10^{-19}$	$1.9 \cdot 10^{-34}$
p	9	16	19	29

3.9 Коллокационные методы

Оказывается, что многие неявные методы Рунге–Кутты некоторым образом эквивалентны так называемым коллокационным методам. Основная идея *коллокационных методов* для решения обыкновенных дифференциальных уравнений состоит в том, чтобы в качестве приближенного решения $\mathbf{x}$ подобрать такой полином $\mathbf{g}$ порядка s, который бы удовлетворял условиям задачи Коши (1.8) в $s+1$ точках:

$$\mathbf{g}(t_0) = \mathbf{x}_0, \quad \mathbf{g}'(t_0 + c_i h) = \mathbf{f}(t_0 + c_i h, \mathbf{g}(t_0 + c_i h)) \quad (i = 1, \ldots, s), \tag{3.43}$$

где c_i — вещественные числа, выбираемые, как правило, на отрезке $[0, 1]$. В соответствии с (3.43) приближенное решение уравнения на шаге h представляется в виде $\mathbf{x}_1 = \mathbf{g}(t_0 + h)$. Например, для $s = 1$ полином должен иметь вид $\mathbf{g}(t) = \mathbf{x}_0 + (t - t_0)\mathbf{k}$ с линейным коэффициентом

$$\mathbf{k} = \mathbf{g}'(t_0 + c_1 h) = \mathbf{f}(t_0 + c_1 h, \mathbf{g}(t_0 + c_1 h)) = \mathbf{f}(t_0 + c_1 h, \mathbf{x}_0 + c_1 h \mathbf{k}).$$

Отсюда видно, что явный и неявный методы Эйлера, а также метод средней точки являются коллокационными с коэффициентами $c_1 = 0$, $c_1 = 1$ и $c_1 = 1/2$ соответственно.

Введем обозначение

$$\mathbf{k}_i = \mathbf{f}(t_0 + c_i h, \mathbf{g}(t_0 + c_i h)). \tag{3.44}$$

Тогда интерполирующая функция Лагранжа для $\mathbf{g}'(t)$ будет

$$\mathbf{g}'(t_0 + \tau h) = \sum_{j=1}^{s} \mathbf{k}_j l_j(\tau), \quad \text{где} \quad l_j(\tau) = \prod_{m \neq j} \frac{\tau - c_m}{c_j - c_m}, \tag{3.45}$$

а $\tau = (t - t_0)/h$. Отсюда интегрируя (3.45) по τ и пользуясь записью (3.44), будем иметь

$$\mathbf{x}_1 = \mathbf{x}_0 + h \sum_{j=1}^{s} \mathbf{k}_j \int_0^1 l_j(\tau) \mathrm{d}\tau, \quad \mathbf{k}_i = \mathbf{f}(t_0 + c_i h, \mathbf{x}_0 + h \sum_{j=1}^{s} \mathbf{k}_j \int_0^{c_i} l_j(\tau) \mathrm{d}\tau) \quad (i = 1, \ldots, s). \tag{3.46}$$

Нетрудно видеть, что формулы (3.46) совпадают с формулами неявного метода Рунге–Кутты (3.33), где в качестве коэффициентов a_{ij} и b_i выступают

$$a_{ij} = \int_0^{c_i} l_j(\tau) \mathrm{d}\tau, \quad b_j = \int_0^1 l_j(\tau) \mathrm{d}\tau \quad (i, j = 1, \ldots, s). \tag{3.47}$$

Методы Рунге–Кутты с коэффициентами (3.47) называются коллокационными.

Далее поскольку

$$\tau^{q-1} = \sum_{j=1}^{s} c_j^{q-1} l_j(\tau) \quad (q = 1, \ldots, s)$$

(т.е. интерполяция Лагранжа точна для всех полиномов ниже s-го порядка), то коэффициенты (3.47) должны удовлетворять системам линейных уравнений

$$\sum_{j=1}^{s} a_{ij} c_j^{q-1} = \frac{c_i^q}{q}, \quad \sum_{j=1}^{s} b_j c_j^{q-1} = \frac{1}{q} \quad (i, q = 1, \ldots, s). \tag{3.48}$$

Отметим, что коллокационные формулы очень удобны для определения приближенных значений $\mathbf{k}$ следующего шага. Обозначая их как $\bar{\mathbf{k}}$, а величину следующего шага $\bar{h}$, согласно формулам (3.45) получаем экстраполированную оценку

$$\bar{\mathbf{k}}_i = \sum_{j=1}^{s} \mathbf{k}_j l_j(1 + c_i \bar{h}/h) \quad (i = 1, \ldots, s).$$

3.10 Методы Гаусса

В общем случае метод (3.46) имеет порядок $p = s$. Однако Бутчер показал, что если коэффициенты b_j и c_j удовлетворяют первым $2s$ уравнениям во второй системе (3.48) (т.е. если они являются соответственно весами и узлами *квадратурной формулы Гаусса*), коллокационный метод Рунге–Кутты будет иметь порядок $p = 2s$.

Известно, что узловые значения квадратурной формулы Гаусса на отрезке $[0, 1]$ являются корнями смещенного полинома Лежандра, иначе говоря, они являются решениями уравнения

$$\frac{\mathrm{d}^s}{\mathrm{d}\tau^s}[\tau^s(\tau - 1)^s] = 0. \tag{3.49}$$

Поэтому для получения коллокационного метода порядка $p = 2s$ на практике удобно сначала вычислять c_i из уравнения (3.49), а затем — a_{ij} и b_i из формул (3.47).

Разбиение шага узловыми значениями c_i, удовлетворяющими уравнению (3.49), называется *разбиением Гаусса–Лежандра*. Если зафиксировать начальное узловое значение ($c_1 = 0$), либо конечное ($c_s = 1$), либо и то, и другое ($c_1 = 0$ и $c_s = 1$), то получим другие известные разбиения Гаусса: *левое и правое разбиения Радо* и *разбиение Лобатто*, которые соответственно удовлетворяют уравнениям

$$\frac{\mathrm{d}^{s-1}}{\mathrm{d}\tau^{s-1}}[\tau^s(\tau-1)^{s-1}] = 0, \quad \frac{\mathrm{d}^{s-1}}{\mathrm{d}\tau^{s-1}}[\tau^{s-1}(\tau-1)^s] = 0, \quad \frac{\mathrm{d}^{s-2}}{\mathrm{d}\tau^{s-2}}[\tau^{s-1}(\tau-1)^{s-1}] = 0.$$

Однако любой коллокационный метод, построенный на одном из этих разбиений, будет иметь порядок ниже $2s$: узловые значения Радо дают метод порядка $p = 2s-1$, а узловые значения Лобатто — $p = 2s - 2$. Методы, основанные на рассмотренных разбиениях, называют еще *методами Гаусса*.

Получим метод Кунцмана–Бутчера 4-го порядка ($s = 2$). Уравнение узлов (3.49) можно преобразовать к виду

$$6\tau^2 - 6\tau + 1 = 0, \quad \text{отсюда} \quad c_{1,2} = \frac{1}{2}\left(1 \mp \frac{1}{\sqrt{3}}\right).$$

Затем, используя формулы (3.47) с линейными функциями $l_1 = (\tau - c_2)/(c_1 - c_2)$ и $l_2 = (\tau - c_1)/(c_2 - c_1)$, находим коэффициенты

$$a_{11} = a_{22} = \frac{1}{4}, \quad a_{12,21} = \frac{1}{4}\left(1 \mp \frac{2}{\sqrt{3}}\right), \quad b_1 = b_2 = \frac{1}{2}.$$

Далее рассмотрим практическую реализацию коллокационных методов на примере интегратора Эверхарта, который широко применяется в небесной механике.

3.11 Интегратор Эверхарта

В 1973 г. Э. Эверхарт предложил интегратор, разработанный им специально для численного исследования орбит, и продемонстрировал его высокую эффективность в задачах кометной динамики. По-видимому, обнаружив в дальнейшем принадлежность своего интегратора к семейству интеграторов типа Бутчера, Эверхарт акцентировал внимание на оригинально реализованном им алгоритме интегрирования и обобщил его для численного решения любых обыкновенных дифференциальных уравнений первого и второго порядков. Тем самым ему удалось расширить область применения своего интегратора, который тем не менее остается одним из самых популярных именно в решении задач небесной механики.

3.11.1 Основные формулы

Предположим, на шаге h мы решаем задачу Коши (1.8):

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}_0 = \mathbf{x}(t_0).$$

Введем $\tau = (t - t_0)/h$ и представим правую часть уравнений $\mathbf{f}$ в виде полинома степени $s-1$:

$$\mathbf{x}' = \mathbf{x}'_\tau/h = \mathbf{f} = \mathbf{A}_1 + \mathbf{A}_2\tau + \mathbf{A}_3\tau^2 + \ldots + \mathbf{A}_s\tau^{s-1}, \tag{3.50}$$

где коэффициенты $\mathbf{A}$ пока не определены. Интегрируя (3.50) по τ, получаем решение

$$\mathbf{x} = \mathbf{x}_0 + h\left(\mathbf{A}_1\tau + \frac{1}{2}\mathbf{A}_2\tau^2 + \frac{1}{3}\mathbf{A}_3\tau^3 + \ldots + \frac{1}{s}\mathbf{A}_s\tau^s\right). \tag{3.51}$$

Перепишем (3.50) в виде интерполяционного многочлена Ньютона на сетке $\tau_1, \ldots, \tau_s$:

$$\mathbf{f} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2(\tau - \tau_1) + \boldsymbol{\alpha}_3(\tau - \tau_1)(\tau - \tau_2) + \ldots + \boldsymbol{\alpha}_s(\tau - \tau_1)\ldots(\tau - \tau_{s-1}). \tag{3.52}$$

Из соотношений

$$\begin{aligned}
&\mathbf{f}_1 = \boldsymbol{\alpha}_1,\\
&\mathbf{f}_2 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2(\tau_2 - \tau_1),\\
&\mathbf{f}_3 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2(\tau_3 - \tau_1) + \boldsymbol{\alpha}_3(\tau_3 - \tau_1)(\tau_3 - \tau_2)\\
&\ldots
\end{aligned} \tag{3.53}$$

получаем конечные разности $\boldsymbol{\alpha}$:

$$\begin{aligned}
&\boldsymbol{\alpha}_1 = \mathbf{f}_1\\
&\boldsymbol{\alpha}_2 = (\mathbf{f}_2 - \boldsymbol{\alpha}_1)/(\tau_2 - \tau_1),\\
&\boldsymbol{\alpha}_3 = ((\mathbf{f}_3 - \boldsymbol{\alpha}_1)/(\tau_3 - \tau_1) - \boldsymbol{\alpha}_2)/(\tau_3 - \tau_2).\\
&\ldots
\end{aligned} \tag{3.54}$$

В свою очередь, сравнивая (3.50) и (3.52), будем иметь

$$\begin{aligned}
\mathbf{A}_1 = \boldsymbol{\alpha}_1 + (-\tau_1)\boldsymbol{\alpha}_2 + (\tau_1\tau_2)\boldsymbol{\alpha}_3 + \ldots + (-1)^{s-1}(\tau_1\ldots\tau_{s-1})\boldsymbol{\alpha}_s &= c_{11}\boldsymbol{\alpha}_1 + c_{21}\boldsymbol{\alpha}_2 + \ldots + c_{s1}\boldsymbol{\alpha}_s,\\
\mathbf{A}_2 = \boldsymbol{\alpha}_2 + (-\tau_1 - \tau_2)\boldsymbol{\alpha}_3 + \ldots &= c_{22}\boldsymbol{\alpha}_2 + \ldots + c_{s2}\boldsymbol{\alpha}_s,\\
\ldots &\ldots\\
\mathbf{A}_s = \boldsymbol{\alpha}_s &= c_{ss}\boldsymbol{\alpha}_s.
\end{aligned} \tag{3.55}$$

Тогда обратный переход от $\mathbf{A}$ к $\boldsymbol{\alpha}$ можно представить как

$$\begin{aligned}
\boldsymbol{\alpha}_1 = &\ d_{11}\mathbf{A}_1 + d_{21}\mathbf{A}_2 + \ldots + d_{s1}\mathbf{A}_s,\\
\boldsymbol{\alpha}_2 = &\ d_{22}\mathbf{A}_2 + \ldots + d_{s2}\mathbf{A}_s,\\
\ldots &\ldots\\
\boldsymbol{\alpha}_s = &\ d_{ss}\mathbf{A}_s.
\end{aligned} \tag{3.56}$$

Коэффициенты c_{ij} и d_{ij} являются *числами Стирлинга*, которые вычисляются по формулам

$$\begin{aligned}
&c_{ii} = d_{ii} = 1, \quad c_{i0} = d_{i0} = 0 \quad (i > 0),\\
&c_{ij} = c_{i-1,j-1} - \tau_{i-1}c_{i-1,j}, \quad d_{ij} = d_{i-1,j-1} - \tau_j d_{i-1,j} \quad (i > j > 0).
\end{aligned} \tag{3.57}$$

3.11.2 Интегрирование на шаге

Величины $\boldsymbol{\alpha}$ определяются по $\mathbf{f}$, которые, в свою очередь, вычисляются по решениям $\mathbf{x}$. Согласно (3.51) эти решения будем задавать как

$$\begin{aligned}
&\mathbf{x}_1 = \mathbf{x}_0 + h\left(\mathbf{A}_1\tau_1 + \frac{1}{2}\mathbf{A}_2\tau_1^2 + \ldots + \frac{1}{s}\mathbf{A}_s\tau_1^s\right),\\
&\ldots\\
&\mathbf{x}_s = \mathbf{x}_0 + h\left(\mathbf{A}_1\tau_s + \frac{1}{2}\mathbf{A}_2\tau_s^2 + \ldots + \frac{1}{s}\mathbf{A}_s\tau_s^s\right).
\end{aligned} \tag{3.58}$$

Формулы (3.58) представляют собой неявные уравнения относительно $\mathbf{x}$, поэтому они решаются итерационным способом.

Для получения начального приближения $\bar{\boldsymbol{\alpha}}$ на следующем шаге $\bar{h}$ используется информация о коэффициентах $\mathbf{A}$ на текущем шаге h. Безразмерная независимая переменная следующего шага будет $\bar{\tau} = (t - t_h)/\bar{h}$, где $t_h = t_0 + h$. Отсюда

$$\tau = r\bar{\tau} + 1, \tag{3.59}$$

где $r = \bar{h}/h$. Согласно (3.50)

$$\mathbf{A}_1 + \mathbf{A}_2\tau + \mathbf{A}_3\tau^2 + \ldots + \mathbf{A}_s\tau^{s-1} = \bar{\mathbf{A}}_1 + \bar{\mathbf{A}}_2\bar{\tau} + \bar{\mathbf{A}}_3\bar{\tau}^2 + \ldots + \bar{\mathbf{A}}_s\bar{\tau}^{s-1}. \tag{3.60}$$

Подставляя (3.59) в (3.60) и приравнивая коэффициенты при одинаковых степенях $\bar{\tau}$, получаем

$$\begin{aligned}
\bar{\mathbf{A}}_1 = &\quad e_{11}\mathbf{A}_1 + e_{21}\mathbf{A}_2 + e_{31}\mathbf{A}_3 + \ldots + e_{s1}\mathbf{A}_s,\\
\bar{\mathbf{A}}_2 = &\quad r(e_{22}\mathbf{A}_1 + e_{32}\mathbf{A}_2 + \ldots + e_{s2}\mathbf{A}_s),\\
\bar{\mathbf{A}}_3 = &\quad r^2(e_{33}\mathbf{A}_2 + \ldots + e_{s3}\mathbf{A}_s),\\
\ldots &\quad \ldots\\
\bar{\mathbf{A}}_s = &\quad r^{s-1}e_{ss}\mathbf{A}_s,
\end{aligned} \tag{3.61}$$

где e_{ij} — *числа арифметического треугольника*, вычисляемые по рекуррентным формулам

$$e_{ii} = e_{i1} = 1, \quad e_{ij} = e_{i-1,j-1} + e_{i-1,j} \quad (i > j > 1). \tag{3.62}$$

Пользуясь соотношениями (3.56) для $\bar{\mathbf{A}}$, будем иметь начальное приближение $\bar{\boldsymbol{\alpha}}$.

Оценку $\bar{\mathbf{A}}$ для $\bar{h}$ можно существенно улучшить, если к ней добавлять поправку $\Delta\mathbf{A}$, получаемую как разность между значениями $\mathbf{A}$ после итераций и оценкой $\bar{\mathbf{A}}$ на текущем шаге h.

Каждая итерация выполняется следующим образом. Сначала определяется решение $\mathbf{x}_1$, из которого по первой формуле (3.54) находится значение $\boldsymbol{\alpha}_1$. Далее определяется $\mathbf{x}_2$, по которому улучшается $\boldsymbol{\alpha}_2$, и так до $\mathbf{x}_s$. Как правило, для получения достаточно хороших $\boldsymbol{\alpha}$ необходимо всего лишь 2 итерации, очень редко — 3.

Как только величины $\boldsymbol{\alpha}$ получены, решение на шаге h ($\tau = 1$) будет

$$\mathbf{x}_h = \mathbf{x}_0 + h\left(\mathbf{A}_1 + \frac{1}{2}\mathbf{A}_2 + \ldots + \frac{1}{s}\mathbf{A}_s\right). \tag{3.63}$$

В начале интегрирования, на первом шаге, в качестве $\bar{\boldsymbol{\alpha}}$ выбирают нулевые значения и запускается вышеописанный итерационный процесс. Если начальный шаг достаточно большой, чтобы обеспечить заданную локальную точность, то его следует уменьшить. При оптимально выбранном шаге высокая точность $\boldsymbol{\alpha}$ достигается уже на 4-й итерации.

3.11.3 Формулы интегратора как одно из представлений неявного метода Рунге–Кутты

Пользуясь (3.54), (3.55) и (3.58), нетрудно показать, что решение (3.63) представимо в виде

$$\mathbf{x}_h = \mathbf{x}_0 + h\sum_{i=1}^{s} b_i\mathbf{k}_i, \quad \text{где} \quad \mathbf{k}_i = \mathbf{f}(t_0 + h\tau_i, \mathbf{x}_0 + h\sum_{j=1}^{s} a_{ij}\mathbf{k}_j) \quad (i = 1, \ldots, s),$$

а коэффициенты a_{ij} и b_i — постоянные, зависящие только от τ_i. Таким образом, интегратор Эверхарта фактически основан на видоизмененных формулах неявного метода Рунге–Кутты, который, кроме того, является коллокационным, что очевидно следует из (3.50)–(3.55). Поэтому порядок аппроксимирующей схемы интегратора можно повысить до $p = 2s$, если ее строить на узлах τ_i, являющихся корнями полинома Лежандра, т.е. если они удовлетворяют уравнению (3.49).

В оригинальной версии интегратора Эверхарта используется левое разбиение Гаусса–Радо, и аппроксимирующая схема имеет меньший порядок $p = 2s - 1$. Несмотря на это, примечательность такого разбиения состоит в том, что оно не требует перевычислений коэффициента $\mathbf{A}_1$ при решении нелинейных уравнений в схеме интегрирования, поскольку $\tau_1 = 0$ и $\mathbf{A}_1 = \mathbf{f}_1 = \mathbf{f}_0 = \mathbf{f}(t_0, \mathbf{x}_0)$.

3.11.4 Выбор шага

В интеграторе Эверхарта контроль шага интегрирования осуществляется по величине последнего члена в (3.63). Пусть $\|\mathbf{e}\|_{tol}$ — заданная точность. Потребуем, чтобы на следующем шаге выполнялось равенство

$$\frac{\bar{h}}{s}\|\bar{\mathbf{A}}_s\| = \|\mathbf{e}\|_{tol}.$$

Отсюда, используя последнее соотношение в (3.61), получаем оценку

$$\bar{h} = hr = h\left(\frac{s}{h}\frac{\|\mathbf{e}\|_{tol}}{\|\mathbf{A}_s\|}\right)^{1/s}. \tag{3.64}$$

Очевидно, при разбиениях Гаусса недостаток такой оценки состоит в том, что шаг по ней выбирается как для решения порядка $s-1$, поэтому, вообще говоря, она не обеспечивает сохранение требуемой локальной точности.

Во избежание слишком больших (и малых) локальных ошибок на r следует наложить ограничение:

$$\frac{1}{\sigma} < r^s < \sigma. \tag{3.65}$$

Для того чтобы величина последнего члена в (3.63) была ограничена в пределах одного порядка, значение σ должно быть равно $\sqrt{10}$. Это следует из того факта, что

$$\bar{h}\|\bar{\mathbf{A}}_s\| \sim r^s.$$

Выполнение обоих неравенств проверяется лишь в начале интегрирования при выборе стартового шага: если (3.65) не выполняется, то интегрирование повторяется с новым шагом $\bar{h} = hr$ и так далее, пока не выполнится условие (3.65). Обычно для получения стартового шага требуется не более 4 итераций. В дальнейшем для ограничения r проверяется только правое неравенство: если неравенство не выполняется, то r принимает значение правого предела.

Начальное приближение стартового шага получается из оценки, подобной (3.64), для $s = 2$:

$$\bar{h} = \sqrt{\frac{2h\|\mathbf{e}\|_{tol}}{\|\mathbf{f}_2 - \mathbf{f}_1\|}}, \quad \mathbf{f}_1 = \mathbf{f}(t_0, \mathbf{x}_0), \quad \mathbf{f}_2 = \mathbf{f}(t_0 + h, \mathbf{x}_0 + h\mathbf{f}_1), \tag{3.66}$$

где h — малая величина. Если h настолько мала, что в компьютерной арифметике $\mathbf{f}_1 = \mathbf{f}_2$, она увеличивается в 10 раз и оценка повторяется снова.

3.12 A-устойчивость методов Рунге–Кутты

Свойства устойчивости численных методов интегрирования обычно исследуют на примере асимптотически устойчивых систем линейных уравнений

$$\mathbf{x}' = \mathbf{L}\mathbf{x}, \tag{3.67}$$

где $\mathbf{L}$ — квадратная матрица размера $m \times m$ с постоянными коэффициентами, собственные числа которой λ_i имеют отрицательные действительные части: $\mathrm{Re}(\lambda_i) < 0$. Введем преобразование координат

$$\mathbf{x} = \mathbf{T}\mathbf{z}, \tag{3.68}$$

которое приводит матрицу $\mathbf{L}$ к *жордановой канонической форме*:

$$\mathbf{T}^{-1}\mathbf{L}\mathbf{T} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m). \tag{3.69}$$

Тогда в новых переменных $\mathbf{z}$ система (3.67) преобразуется к виду

$$\mathbf{z}' = \mathbf{\Lambda}\mathbf{z}. \tag{3.70}$$

Поскольку $\mathbf{\Lambda}$ — диагональная матрица, то система (3.70) расщепляется на m независимых уравнений $z_i' = \lambda_i z_i$, $(i = 1, \ldots, m)$, которые соответственно имеют решения $z_i = z_{0i}e^{\lambda_i(t-t_0)}$, где компоненты начального вектора $\mathbf{z}_0$ определяются по $\mathbf{x}_0$ согласно линейному преобразованию (3.68): $\mathbf{z}_0 = \mathbf{T}^{-1}\mathbf{x}_0$. Ввиду того, что $\mathrm{Re}(\lambda_i) < 0$ для всех i, $\mathbf{z} \to \mathbf{0}$ при $t \to \infty$, следовательно, и $\mathbf{x} \to \mathbf{0}$. Естественно, для качественного описания решения системы (3.67) численным методом необходимо потребовать, чтобы

$$\mathbf{x}_n \to \mathbf{0} \quad \text{при} \quad n \to \infty. \tag{3.71}$$

Метод, обеспечивающий выполнение (3.71), называется *устойчивым*. Поскольку все методы Рунге–Кутты сходящиеся, теоретически всегда можно подобрать такой достаточно малый шаг интегрирования, при котором будет выполняться условие (3.71). В связи с этим представляет интерес определение максимально допустимого шага h, когда условие (3.71) будет нарушаться. Если условие (3.71) выполняется для любого h, говорят, что численный метод *абсолютно устойчив* либо *A-устойчив*.

Покажем теперь, что между численными решениями систем (3.67) и (3.70), полученными методом (3.33), имеет место линейное соответствие, т.е. $\mathbf{x}_n = \mathbf{T}\mathbf{z}_n$. Для систем линейных уравнений схемы интегрирования на шаге h согласно (3.33) можно записать в виде

$$\mathbf{K} = \mathbf{L}(\mathbf{X}_0 + h\mathbf{K}\mathbf{A}), \quad \mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{K}\mathbf{b}; \quad \text{и} \quad \bar{\mathbf{K}} = \mathbf{\Lambda}(\mathbf{Z}_0 + h\bar{\mathbf{K}}\mathbf{A}), \quad \mathbf{z}_1 = \mathbf{z}_0 + h\bar{\mathbf{K}}\mathbf{b}, \tag{3.72}$$

где $\mathbf{K} = (\mathbf{k}_1, \ldots, \mathbf{k}_s)$ для системы (3.67), а $\bar{\mathbf{K}}$ — соответствующая матрица коэффициентов для системы (3.70); $\mathbf{X}_0 = (\mathbf{x}_0, \ldots, \mathbf{x}_0)$ и $\mathbf{Z}_0 = (\mathbf{z}_0, \ldots, \mathbf{z}_0)$ — матрицы размера $m \times s$, тогда как

$$\mathbf{A} = \begin{pmatrix} a_{11} & \ldots & a_{s1} \\ \vdots & \ddots & \vdots \\ a_{1s} & \ldots & a_{ss} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_s \end{pmatrix}.$$

Рекуррентные формулы для $\mathbf{K}$ и $\bar{\mathbf{K}}$ в (3.72) дают

$$\mathbf{K} = \sum_{i=0}^{\infty} h^i \mathbf{L}^{i+1} \mathbf{X}_0 \mathbf{A}^i, \quad \bar{\mathbf{K}} = \sum_{i=0}^{\infty} h^i \mathbf{\Lambda}^{i+1} \mathbf{Z}_0 \mathbf{A}^i. \tag{3.73}$$

Отсюда, используя связь между матрицами $\mathbf{L}$ и $\mathbf{\Lambda}$ (3.69) и соотношение $\mathbf{X}_0 = \mathbf{T}\mathbf{Z}_0$, можно показать, что $\mathbf{K} = \mathbf{T}\bar{\mathbf{K}}$, поэтому согласно представлению решений $\mathbf{x}_1$ и $\mathbf{z}_1$ (3.72) $\mathbf{x}_1 = \mathbf{T}\mathbf{z}_1$. Тогда очевидно, что линейное соответствие будет иметь место и для других решений $\mathbf{x}_n = \mathbf{T}\mathbf{z}_n$. Следовательно, условие (3.71) будет выполняться, если $\mathbf{z}_n \to \mathbf{0}$ при $n \to \infty$ и, таким образом, устойчивость численного метода можно исследовать на примере более простой системы (3.70) с независимыми друг от друга уравнениями. Поскольку эти уравнения однотипны, исследование устойчивости фактически сводится к рассмотрению одного уравнения общего вида

$$z' = \lambda z, \tag{3.74}$$

где λ — комплексное число, причем $\mathrm{Re}(\lambda) < 0$.

Исследуем на предмет устойчивости некоторые простейшие методы Рунге–Кутты, а именно: явный (2.7) и неявный (3.34) методы Эйлера ($p = 1$), а также метод Хойна (3.13) и трапеций (3.35) ($p = 2$). Схемы интегрирования методов применительно к уравнению (3.74) можно представить в виде $z_{n+1} = S(\eta) z_n$, где $\eta = h\lambda$, а S — так называемые *функции устойчивости*, которые в порядке перечисления методов представимы как

$$S = 1 + \eta, \quad S = \frac{1}{1-\eta}, \quad S = 1 + \eta + \frac{\eta^2}{2}, \quad S = \frac{1 + \eta/2}{1 - \eta/2}.$$

Очевидно, метод будет устойчивым при заданной паре чисел h и λ, если $|S| < 1$. На рис. 3.2 показаны области устойчивости методов $|S(\eta)| < 1$ (серая заливка) на комплексной плоскости η. Поскольку $\mathrm{Re}(\lambda) < 0$ и $h > 0$, то нас интересует левая полуплоскость η. Из рисунка видно, что рассматриваемые явные методы имеют ограниченные области устойчивости, что говорит об ограничении в выборе величины шага интегрирования h с точки зрения качественного описания решения. Напротив, неявные методы не имеют никаких ограничений на шаг h, поскольку левая полуплоскость η оказывается в области устойчивости методов. Следовательно, рассматриваемые неявные методы абсолютно устойчивы.

Заметим, что из численного анализа известно, что среди явных методов нет абсолютно устойчивых. Это главным образом обусловлено тем, что функция устойчивости для любого явного метода представляет собой полином вида $S(\eta) = 1 + \eta + O(\eta^2)$, величина которого неограниченно возрастает при $\mathrm{Re}(\eta) \to -\infty$.

В то же время среди неявных методов имеется немало A-устойчивых, которые образуют класс основных методов, используемых для получения приближенных решений так называемых жестких систем. Напомним, что система линейных дифференциальных уравнений типа (3.67) называется *жесткой*, если она плохообусловлена, т.е. если отношение $\max_i |\mathrm{Re}(\lambda_i)| / \min_i |\mathrm{Re}(\lambda_i)|$ достаточно большое.

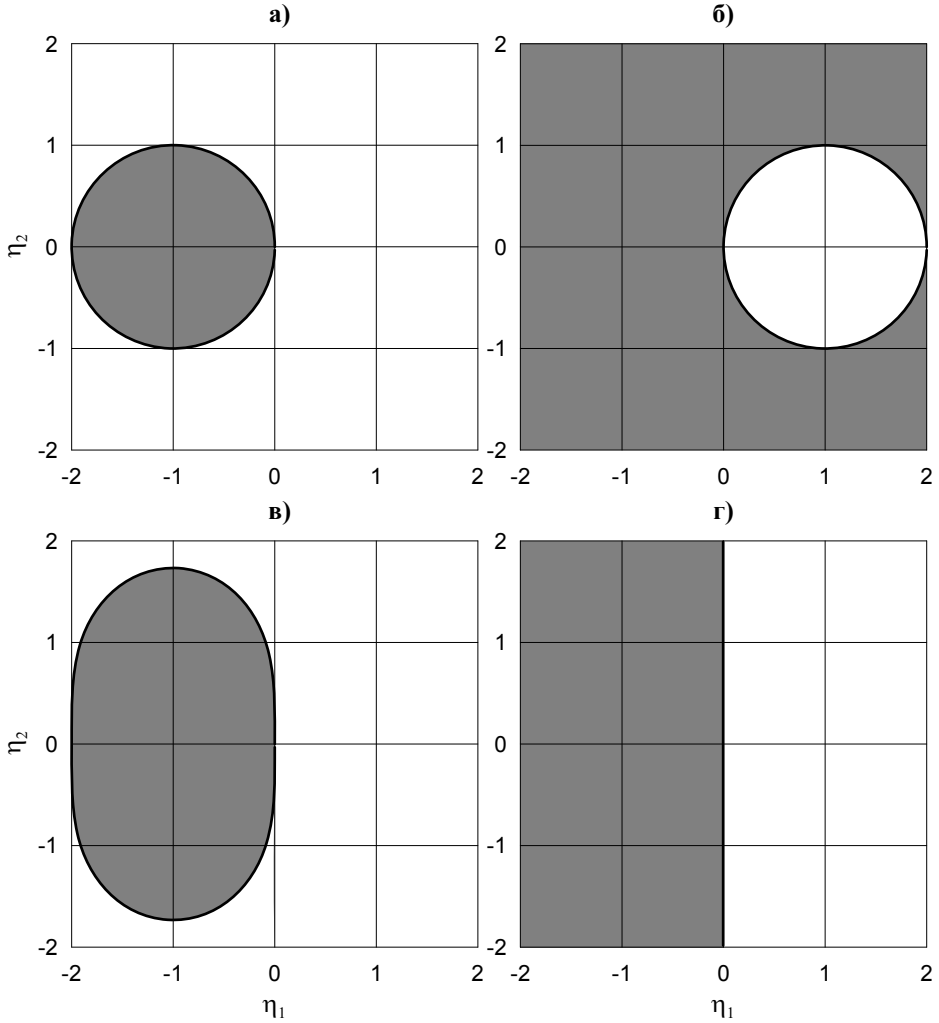

Рисунок 3.2 — Области устойчивости методов Рунге–Кутты низких порядков: а) явный метод Эйлера; б) неявный метод Эйлера; в) метод Хойна; г) метод трапеций

4 Экстраполяционные методы

Экстраполяционные методы основаны на том факте (который мы не будем доказывать), что глобальная погрешность численного решения порядка p допускает разложение в степенной ряд по h:

$$\mathbf{x}_H - \mathbf{x}(t_0 + H) = \sum_{i=p}^{\infty} \mathbf{e}_i h^i, \tag{4.1}$$

где $\mathbf{x}_H$ — приближенное решение на отрезке H, полученное на равномерной сетке с шагом h.

Формула (4.1) представляет приближенное решение на фиксированном отрезке H как функцию $\mathbf{x}_H = \mathbf{x}_H(h)$. Для этой функции

$$\lim_{h \to 0} \mathbf{x}_H(h) = \mathbf{x}(t_0 + H).$$

Теоретически это означает, что при достаточно малом шаге можно любым методом получить приближенное решение со сколь угодно высокой точностью. С практической точки зрения использование очень малых шагов не разумно, поскольку, во-первых, это увеличивает объем вычислений, что, во-вторых, сопряжено с большими ошибками округления. Тем не менее, используя идею *экстраполяции*, можно получить с высокой точностью предельное значение $\mathbf{x}_H(0)$, избегая при этом многочисленных вычислений.

4.1 Общий подход

В соответствии с (4.1) представим интерполяцию приближенного решения в виде многочлена порядка $p + s - 1$:

$$\mathbf{P}(h) = \mathbf{a}_0 + \mathbf{a}_p h^p + \ldots + \mathbf{a}_{p+s-1} h^{p+s-1}. \tag{4.2}$$

Выберем последовательность

$$N_0 < \ldots < N_s \tag{4.3}$$

и определим соответствующие длины шагов $h_i = H/N_i$ $(i = 0, \ldots, s)$. Далее для каждого шага h_i вычислим решение $\mathbf{x}_H(h_i)$ каким-либо (*опорным*) методом Рунге–Кутты порядка p. Тогда, требуя, чтобы $\mathbf{P}(h_i) = \mathbf{x}_H(h_i)$, получим следующую систему алгебраических уравнений:

$$\begin{aligned}
&\mathbf{P}(h_0) = \mathbf{a}_0 + \mathbf{a}_p h_0^p + \ldots + \mathbf{a}_{p+s-1} h_0^{p+s-1} = \mathbf{x}_H(h_0), \\
&\ldots \\
&\mathbf{P}(h_s) = \mathbf{a}_0 + \mathbf{a}_p h_s^p + \ldots + \mathbf{a}_{p+s-1} h_s^{p+s-1} = \mathbf{x}_H(h_s).
\end{aligned} \tag{4.4}$$

После определения из системы (4.4) коэффициентов интерполирующего многочлена нам остается выполнить экстраполяцию при $h \to 0$, что дает приближенное решение $\mathbf{a}_0$:

$\mathbf{P}(0) = \mathbf{a}_0$. Согласно (4.1) и (4.2) экстраполяция $\mathbf{a}_0$ будет представлять точное решение $\mathbf{x}(t_0 + H)$ с ошибкой порядка $p + s$.

При $s = 1$ интерполяционный многочлен приводится к виду

$$\mathbf{P}(h) = \mathbf{a}_0 + \mathbf{a}_p h^p,$$

коэффициенты которого будут определяться из системы

$$\mathbf{a}_0 + \mathbf{a}_p h_0^p = \mathbf{x}_H(h_0),$$
$$\mathbf{a}_0 + \mathbf{a}_p h_1^p = \mathbf{x}_H(h_1).$$

Отсюда получаем экстраполяционную формулу (3.28):

$$\mathbf{a}_0 (= \hat{\mathbf{x}}_H) = \mathbf{x}_H(h_0) + \frac{\mathbf{x}_H(h_0) - \mathbf{x}_H(h_1)}{(N_0/N_1)^p - 1}.$$

4.2 Алгоритм Эйткена–Невилла

При $p = 1$ (метод Эйлера) формулы (4.2) и (4.4) приводят к классической задаче интерполяции. В этом случае для упрощения вычислений интерполяционного многочлена удобно использовать так называемую *схему Эйткена–Невилла*.

Пусть $E_{n,n+1,\ldots,m}(h)$ — интерполяционный многочлен порядка $m-n$ с узлами интерполяции $h_n, \ldots, h_m$ для некоторой функции $g = g(h)$, в частности, $E_n(h) = g(h_n)$. Тогда справедливы равенства

$$E_{n,n+1,\ldots,m+1}(h) = \frac{E_{n+1,\ldots,m+1}(h)(h - h_n) - E_{n,\ldots,m}(h)(h - h_{m+1})}{h_{m+1} - h_n}, \tag{4.5}$$

где $E_{n,n+1,\ldots,m+1}(h)$ — интерполяционный многочлен уже порядка $m - n + 1$ на сетке $h_n, \ldots, h_{m+1}$. Фактически вычисление многочлена $E_{0,1,\ldots,s}(h)$ на сетке $h_0, \ldots, h_s$ сводится к последовательному вычислению с помощью (4.5) элементов таблицы

$$\begin{array}{lllll}
E_0(h) & & & & \\
 & E_{0,1}(h) & & & \\
E_1(h) & & E_{0,1,2}(h) & & \\
 & E_{1,2}(h) & & \ldots & E_{0,1,\ldots,s}(h) \\
E_2(h) & & \ldots & & \\
\ldots & \ldots & & & \\
E_s(h) & & & &
\end{array} \tag{4.6}$$

Для экстраполяции нам нужна будет формула (4.5) при $h = 0$:

$$E_{n,n+1,\ldots,m+1}(0) = \frac{E_{n,\ldots,m}(0)h_{m+1} - E_{n+1,\ldots,m+1}(0)h_n}{h_{m+1} - h_n}. \tag{4.7}$$

Переобозначим элементы таблицы (4.6) и представим ее как

$$\begin{array}{llll}
E_{00} & & & \\
E_{10} & E_{11} & & \\
E_{20} & E_{21} & \ldots & \\
\ldots & \ldots & \ldots & \\
E_{s0} & E_{s1} & \ldots & E_{ss}
\end{array} \tag{4.8}$$

Тогда в новых обозначениях формула Эйткена–Невилла (4.7) перепишется в виде

$$E_{i,j+1} = \frac{E_{i-1,j}h_i - E_{i,j}h_{i-j-1}}{h_i - h_{i-j-1}}. \tag{4.9}$$

Из (4.9) получаем расчетную формулу для нашей интерполяции

$$\mathbf{P}_{i,j+1} = \mathbf{P}_{i,j} + \frac{\mathbf{P}_{i,j} - \mathbf{P}_{i-1,j}}{(N_i/N_{i-j-1}) - 1} \quad (j = 0, \ldots, s-1;\ i = j, \ldots, s), \tag{4.10}$$

где $\mathbf{P}_{i,0} = \mathbf{P}(h_i) = \mathbf{x}_H(h_i)$ $(i = 0, \ldots, s)$ вычисляются по схеме метода Эйлера:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h_i\mathbf{f}(t_n, \mathbf{x}_n) \quad (n = 0, \ldots, N_i - 1).$$

В итоге будем иметь экстраполированное решение $\mathbf{a}_0 = \hat{\mathbf{x}}_H = \mathbf{P}_{ss}$ порядка s.

В (4.10) для получения узловых значений в качестве последовательности (4.3) обычно используют

$$1, 2, 4, 8, 16, 32, \ldots \quad (\textit{последовательность Ромберга})$$

либо

$$1, 2, 3, 4, 5, 6, \ldots \quad (\textit{гармоническая последовательность}).$$

4.3 Метод Грэгга

Метод называется *симметричным*, если он инвариантен относительно перестановки:

$$\mathbf{x}_0 \leftrightarrow \mathbf{x}_1, \quad h \leftrightarrow -h, \quad t_0 \leftrightarrow t_0 + h.$$

Например, неявный метод трапеций — симметричен:

$$\mathbf{x}_1 = \mathbf{x}_0 + \frac{h}{2}[\mathbf{f}(t_0, \mathbf{x}_0) + \mathbf{f}(t_1, \mathbf{x}_1)].$$

Если опорный метод порядка $p = 2q$ является симметричным, то, как известно, он будет иметь разложение глобальной ошибки по степеням h^2, т.е.

$$\mathbf{x}_H - \mathbf{x}(t_0 + H) = \sum_{i=q}^{\infty} \mathbf{e}_i h^{2i}.$$

Следовательно, для экстраполяции нужно использовать интерполирующий полином вида

$$\mathbf{P}(h) = \mathbf{a}_0 + \mathbf{a}_q(h^2)^q + \ldots + \mathbf{a}_{q+s-1}(h^2)^{q+s-1}. \tag{4.11}$$

Ввиду схожести полиномов (4.11) и (4.2) (при определенном s) вычисление их коэффициентов будет одинаково трудоемко в обоих случаях. Однако примечательно то, что в итоге интерполяция (4.11) дает экстраполированное решение $\mathbf{a}_0$ порядка $2(q + s - 1) = p + 2s - 2$, на $s - 1$ выше, нежели для (4.2).

У.Б. Грэгг (1963) предложил в качестве опорного использовать симметричный метод второго порядка ($p = 2$), основанный на *правиле средней точки*:

$$\mathbf{x}_1 = \mathbf{x}_0 + h_i\mathbf{f}(t_0, \mathbf{x}_0), \quad \mathbf{x}_{n+1} = \mathbf{x}_{n-1} + 2h_i\mathbf{f}(t_n, \mathbf{x}_n) \quad (n = 1, \ldots, N_i - 1). \tag{4.12}$$

Поскольку интерполирующий многочлен (4.11) имеет форму (4.2), для вычисления экстраполяции используется та же формула Эйткена–Невилла, где только h заменяется на h^2, а именно:

$$\mathbf{P}_{i,j+1} = \mathbf{P}_{i,j} + \frac{\mathbf{P}_{i,j} - \mathbf{P}_{i-1,j}}{(N_i/N_{i-j-1})^2 - 1} \quad (j = 0, \ldots, s-1;\ i = j, \ldots, s). \tag{4.13}$$

Здесь $\mathbf{P}_{i,0} = \mathbf{x}_H(h_i)$ $(i = 0, \ldots, s)$ получаются из (4.12). При этом последовательности (4.3) выбираются с условием, что N_i должны быть четными. Например,

$$2, 4, 8, 16, 32, \ldots$$

Метод, основанный на (4.12) и (4.13), называется *методом Грэгга* (или *Грэгга–Булирша–Штера*). Именно благодаря своим замечательным особенностям он единственный из всех экстраполяционных методов широко используется на практике.

4.4 Выбор шага

Шаг в экстраполяционных методах выбирается так же, как и во вложенных методах Рунге–Кутта. Мы знаем, что второй индекс j в решении $\mathbf{P}_{i,j}$ определяет его порядок аппроксимации. В методе Грэгга решение $\mathbf{P}_{s,s}$ имеет (наивысший) порядок $2s$, а $\mathbf{P}_{s,s-1}$ — порядок $2s - 2$. Поэтому для управления шага естественно использовать выражение

$$\|\mathbf{e}\|_{cal} = \|\mathbf{P}_{s,s} - \mathbf{P}_{s,s-1}\|.$$

5 Многошаговые методы

До сих пор мы рассматривали *одношаговые методы*, в которых численное решение получают только из дифференциальных уравнений и решения на предыдущем шаге. *Многошаговые методы*, в отличие от одношаговых, используют несколько решений, вычисленных на предыдущих шагах. В начале интегрирования, когда известно только одно решение (начальное условие), первые (стартовые) решения для многошаговых методов, как правило, вычисляют с помощью одношаговых методов Рунге–Кутты. Затем на каждом следующем шаге выполняют многошаговую процедуру интегрирования.

Мы уже знакомы с одним из таких методов (4.12), который в качестве опорного использовался в экстраполяционном методе Грэгга. Это двушаговый симметричный метод второго порядка, где для вычисления начального второго решения используется метод Эйлера.

5.1 Методы Адамса

Многошаговые методы появились гораздо раньше, чем методы Рунге–Кутты. Впервые их получил Дж.К. Адамс еще в 1855 г. Тем же способом, что и Адамс, выведем первое семейство многошаговых методов.

Предположим, для задачи Коши (1.8):

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}(t_0) = \mathbf{x}_0;$$

нам известны первые $r+1$ решения $\mathbf{x}_0, \ldots, \mathbf{x}_r$ на равномерной сетке $t_0, \ldots, t_r$ с шагом h. Тогда для шага $r+1$ формально решение можно представить как

$$\mathbf{x}(t_{r+1}) = \mathbf{x}(t_r) + \int_{t_r}^{t_{r+1}} \mathbf{f}(t, \mathbf{x}(t))\mathrm{d}t. \tag{5.1}$$

Заменим подынтегральную функцию на ее интерполирующий полином Ньютона, полученный по узловым значениям $\mathbf{f}_i = \mathbf{f}(t_i, \mathbf{x}_i)$ $(i = 0, \ldots, r)$:

$$\mathbf{p}(t) = \mathbf{f}_0 + \mathbf{f}_{0,1}(t - t_0) + \mathbf{f}_{0,1,2}(t - t_0)(t - t_1) + \ldots + \mathbf{f}_{0,1,\ldots,r}(t - t_0)\ldots(t - t_{r-1}), \tag{5.2}$$

где разделенные разности вычисляются из треугольника

$$\begin{array}{llllll}
t_0 & \mathbf{f}_0 & & & & \\
 & & \mathbf{f}_{0,1} & & & \\
t_1 & \mathbf{f}_1 & & \mathbf{f}_{0,1,2} & & \\
 & & \mathbf{f}_{1,2} & & \ldots & \mathbf{f}_{0,1,\ldots,r} \\
t_2 & \mathbf{f}_2 & & \ldots & & \\
\ldots & \ldots & \ldots & & & \\
t_r & \mathbf{f}_r & & & &
\end{array}$$

Тогда численный аналог (5.1) будет задаваться формулой

$$\mathbf{x}_{r+1} = \mathbf{x}_r + \int_{t_r}^{t_{r+1}} \mathbf{p}(t)\mathrm{d}t. \tag{5.3}$$

Согласно (5.3) для $r = 0, 1, 2$ получаем следующие явные формулы Адамса:

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{f}_0 \quad \text{(метод Эйлера)},$$
$$\mathbf{x}_2 = \mathbf{x}_1 + h\left[\frac{3}{2}\mathbf{f}_1 - \frac{1}{2}\mathbf{f}_0\right],$$
$$\mathbf{x}_3 = \mathbf{x}_2 + h\left[\frac{23}{12}\mathbf{f}_2 - \frac{16}{12}\mathbf{f}_1 + \frac{5}{12}\mathbf{f}_0\right].$$

Формулы Адамса получаются при интегрировании интерполяционного многочлена (5.2) от t_r до t_{r+1}, т.е. вне интервала интерполяции. Однако, как мы знаем, вне этого интервала интерполяционный многочлен обычно дает довольно плохое приближение. Таким образом, явные методы Адамса не очень точны. Чтобы разрешить эту проблему, Адамс предложил для интерполяции использовать значение $\mathbf{f}_{r+1}$ вместо $\mathbf{f}_0$. В итоге он получил неявные методы. Приведем первые из них для $r = 0, 1$:

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{f}_1 \quad (\textit{неявный метод Эйлера}),$$
$$\mathbf{x}_2 = \mathbf{x}_1 + h\left[\frac{1}{2}\mathbf{f}_2 + \frac{1}{2}\mathbf{f}_1\right] \quad (\textit{правило трапеций}).$$

Итак, любой s-шаговый *метод Адамса* можно представить в общем виде

$$\mathbf{x}_{n+s} = \mathbf{x}_{n+s-1} + h\sum_{i=0}^{s}\beta_i\mathbf{f}_{n+i}, \tag{5.4}$$

где β_i — постоянные метода. Если $\beta_s = 0$, метод явный, иначе — неявный. Порядок метода определяется точностью интерполирующей формулы (5.2). В общем случае явный метод (5.4) имеет порядок $p = s$, неявный — $p = s + 1$.

5.2 Формулы дифференцирования

Формулы Адамса основаны на приближенном вычислении интеграла в (5.1). Хотя многошаговые методы можно получить иным способом, а именно численно дифференцируя искомое решение.

Интерполяцию Ньютона для $\mathbf{x}$ по решениям $\mathbf{x}_0, \ldots, \mathbf{x}_s$ формально можно представить в виде

$$\mathbf{q}(t) = \mathbf{x}_0 + \mathbf{x}_{0,1}(t - t_0) + \mathbf{x}_{0,1,2}(t - t_0)(t - t_1) + \ldots + \mathbf{x}_{0,1,\ldots,s}(t - t_0)\ldots(t - t_{s-1}). \tag{5.5}$$

Последнее решение $\mathbf{x}_s$ пока неизвестно и определим его так, чтобы многочлен $\mathbf{q}$ удовлетворял дифференциальному уравнению, по крайней мере, в одной узловой точке:

$$\mathbf{q}'(t_r) = \mathbf{f}(t_r, \mathbf{x}_r). \tag{5.6}$$

Здесь индекс r принимает одно из значений $0, 1, \ldots, s$. После взятия производной в (5.6) и подстановки t_r получаем схему многошагового интегрирования для $\mathbf{x}_s$. Если $r = s$, метод неявный, иначе — явный. В качестве примера приведем неявный метод для $s = 2$:

$$\frac{3}{2}\mathbf{x}_2 - 2\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_0 = h\mathbf{f}_2.$$

Таким образом, s-шаговый метод, полученный на основе дифференцирования, представим в виде

$$\sum_{i=0}^{s} \alpha_i \mathbf{x}_{n+i} = h\mathbf{f}_{n+r}, \tag{5.7}$$

где α_i — постоянные метода. Порядок метода определяется точностью интерполирующей формулы (5.5). В общем случае метод (5.7) имеет порядок $p = s$.

5.3 Предиктор–корректор

Неявная схема интегрирования предполагает итерационное решение системы нелинейных уравнений относительно искомого решения $\mathbf{x}_{n+s}$. Для этого обычно применяют метод простых итераций. Легко показать, что в случае методов Адамса (5.4) и дифференцирования (5.7) итерации будут сходиться, если соответственно

$$h < \frac{1}{\|\beta_s \mathbf{f}'_{\mathbf{x}}(\mathbf{x}_{n+s})\|} \quad \text{и} \quad h < \frac{|\alpha_s|}{\|\mathbf{f}'_{\mathbf{x}}(\mathbf{x}_{n+s})\|}.$$

Итерационный процесс требует начального приближения решения нелинейной системы. Обычно его получают из явной схемы с точностью на порядок ниже. Явную схему называют *предиктором*, тогда как неявную — *корректором*, а итерационную процедуру в целом — *предиктором–корректором*. Например, для метода трапеций (неявного метода Адамса второго порядка) в качестве предиктора используют метод Эйлера (явный метод Адамса первого порядка) и тогда процедура предиктор–корректор (P–C) будет

$$P:\ \mathbf{x}_1^0 = \mathbf{x}_0 + h\mathbf{f}_0, \quad C:\ \mathbf{x}_1^{l+1} = \mathbf{x}_0 + h\left[\frac{1}{2}\mathbf{f}(\mathbf{x}_0) + \frac{1}{2}\mathbf{f}(\mathbf{x}_1^l)\right] \quad (l = 0, 1, \ldots).$$

Интересно, что первый шаг итерационной последовательности дает решение метода Хойна.

5.4 Линейные многошаговые методы

Таким образом, многошаговые методы можно представить в общем виде

$$\alpha_s \mathbf{x}_{n+s} + \ldots + \alpha_0 \mathbf{x}_n = h(\beta_s \mathbf{f}_{n+s} + \ldots + \beta_0 \mathbf{f}_n). \tag{5.8}$$

При этом будем считать, что выполняются условия

$$\alpha_s \neq 0, \quad |\alpha_0| + |\beta_0| > 0.$$

Обычно полагают, что $\alpha_s = 1$. Если $\beta_s = 0$, метод (5.8) называется явным, иначе — неявным.

Методы вида (5.8) еще называют *линейными многошаговыми*, поскольку решения $\mathbf{x}$ и функции $\mathbf{f}$ входят в них линейным образом.

5.5 Порядок многошаговых методов

Многошаговый метод (5.8) имеет порядок p, если невязка ρ, которая получается после подстановки точного решения в (5.8), является величиной порядка h^{p+1}, т.е.

$$\rho \equiv \sum_{i=0}^{s} \alpha_i \mathbf{x}(t_{n+i}) - h \sum_{i=0}^{s} \beta_i \mathbf{x}'(t_{n+i}) = O(h^{p+1}). \tag{5.9}$$

Разложим решения $\mathbf{x}(t_{n+i})$ в ряд Тейлора:

$$\mathbf{x}(t_{n+i}) = \mathbf{x}(t_n + ih) = \sum_{j=0}^{p+1} \frac{(ih)^j}{j!} \mathbf{x}^{(j)}(t_n) + O(h^{p+2}). \tag{5.10}$$

Тогда для производных $\mathbf{x}'(t_{n+i})$ будем иметь

$$\mathbf{x}'(t_{n+i}) = \mathbf{x}'(t_n + ih) = \sum_{j=0}^{p} \frac{(ih)^j}{j!} \mathbf{x}^{(j+1)}(t_n) + O(h^{p+1}). \tag{5.11}$$

После подстановки (5.10) и (5.11) в (5.9) может быть получено разложение невязки по степеням h:

$$\begin{gathered}\rho = \left(\sum_i \alpha_i\right) \mathbf{x}(t_n) + h \left(\sum_i i\alpha_i - \sum_i \beta_i\right) \mathbf{x}'(t_n) + \frac{h^2}{2} \left(\sum_i i^2\alpha_i - 2\sum_i i\beta_i\right) \mathbf{x}''(t_n) + \dots \\ \dots + \frac{h^p}{p!} \left(\sum_i i^p\alpha_i - p\sum_i i^{p-1}\beta_i\right) \mathbf{x}^{(p)}(t_n) + \\ + \frac{h^{p+1}}{(p+1)!} \left(\sum_i i^{p+1}\alpha_i - (p+1)\sum_i i^p\beta_i\right) \mathbf{x}^{(p+1)}(t_n) + O(h^{p+2}).\end{gathered} \tag{5.12}$$

Отсюда для того, чтобы невязка ρ имела порядок h^{p+1}, должны выполняться условия

$$\begin{aligned} &\sum_i \alpha_i = 0, \\ &\sum_i i\alpha_i - \sum_i \beta_i = 0, \\ &\dots \\ &\sum_i i^p\alpha_i - p\sum_i i^{p-1}\beta_i = 0. \end{aligned} \tag{5.13}$$

Поскольку $p \geq 1$, то первые два условия должны выполняться всегда. Они образуют так называемое *условие согласованности.*

Найдем явный двушаговый метод максимально возможного порядка. Явный двушаговый метод можно представить в общем виде

$$\mathbf{x}_{n+2} + \alpha_1 \mathbf{x}_{n+1} + \alpha_0 \mathbf{x}_n = h(\beta_1 \mathbf{f}_{n+1} + \beta_0 \mathbf{f}_n) \quad (\alpha_2 = 1).$$

Из (5.13) выпишем первые четыре соотношения:

$$1 + \alpha_1 + \alpha_2 = 0,$$
$$2 + \alpha_1 - \beta_0 - \beta_1 = 0,$$
$$4 + \alpha_1 - 2\beta_1 = 0,$$
$$8 + \alpha_1 - 3\beta_1 = 0.$$

Решая эту систему, получаем $\alpha_0 = -5$, $\alpha_1 = 4$, $\beta_0 = 2$, $\beta_1 = 4$. Пятое соотношение в (5.13):

$$16 + \alpha_1 - 4\beta_1 = 0,$$

при найденных коэффициентах уже не выполняется. Поэтому двушаговый метод с полученными нами коэффициентами:

$$\mathbf{x}_{n+2} + 4\mathbf{x}_{n+1} - 5\mathbf{x}_n = h(4\mathbf{f}_{n+1} + 2\mathbf{f}_n), \tag{5.14}$$

будет иметь третий порядок.

5.6 Устойчивость многошаговых методов

Ранее мы выяснили, что глобальная ошибка в методах Рунге–Кутты представляет собой величину порядка h^p:

$$\|\mathbf{E}\| = O(h^p).$$

Это означает, что при уменьшении шага интегрирования точность численного решения повышается. Однако для некоторых многошаговых методов это не так, несмотря на то, что согласно условиям (5.13) они имеют малую локальную ошибку (невязку) некоторого порядка h. Для таких методов уменьшение шага интегрирования приводит только к увеличению глобальной ошибки. Эта особенность многошаговых методов связана с понятием *неустойчивости*.

Важную роль играет поведение решения при $h \to 0$ при фиксированном $H = Nh$. Очевидно, при $h \to 0$ схема (5.8) приводится к формуле

$$\alpha_s \mathbf{x}_{n+s} + \ldots + \alpha_0 \mathbf{x}_n = \mathbf{0}. \tag{5.15}$$

Ее можно рассматривать как численную схему, примененную к решению уравнения

$$\mathbf{x}' = \mathbf{0}.$$

Представим решение как $\mathbf{x}_i = \mathbf{z}\zeta^i$ и подставим его в (5.15):

$$\mathbf{z}\zeta^n(\alpha_s\zeta^s + \ldots + \alpha_0) = \mathbf{0}.$$

Отсюда ζ должно удовлетворять так называемому *характеристическому уравнению*

$$\alpha_s\zeta^s + \ldots + \alpha_0 = 0. \tag{5.16}$$

Исследование характеристического уравнения усложняется, когда у него имеются кратные корни кратности m. В этом случае m решений для этого корня представляются в виде

$$\mathbf{x}_i = i^{j-1}\mathbf{z}\zeta^i \quad (j = 1, \ldots, m). \tag{5.17}$$

Таким образом, если характеристический многочлен (5.16) имеет корни $\zeta_1, \ldots, \zeta_l$ кратностей $m_1, \ldots, m_l$ соответственно, то суперпозиция решений (5.17) будет давать общее решение уравнения (5.15):

$$\mathbf{x}_i = \mathbf{p}_1(i)\zeta_1^i + \ldots + \mathbf{p}_l(i)\zeta_l^i, \tag{5.18}$$

где $\mathbf{p}_j(i)$ — многочлены степеней $m_j - 1$, коэффициенты которых определяются из начальных условий. Поэтому для ограниченности решений $\mathbf{x}_i$ необходимо потребовать, чтобы, во-первых, корни ζ лежали внутри единичной окружности ($|\zeta| < 1$) и, во-вторых, чтобы все корни на единичной окружности ($|\zeta| = 1$) были простыми. Многошаговые методы, удовлетворяющие этим требованиям, называются *устойчивыми по Далквисту* или *D-устойчивыми*.

Исследуем на устойчивость метод (5.14). Характеристический многочлен для него запишется как

$$\zeta^2 + 4\zeta - 5 = 0.$$

Его корни $\zeta_1 = -5$, $\zeta_2 = 1$. Корень $|\zeta_1| > 1$, поэтому метод (5.14) неустойчив.

Наконец, в дополнение к данному разделу отметим известный факт по поводу абсолютной устойчивости многошаговых методов. Доказано, что среди линейных многошаговых методов нет явных A-устойчивых, а порядок неявных A-устойчивых методов не превышает двух.

5.7 Наивысший достижимый порядок для устойчивых методов

В условиях порядка (5.13) содержатся $2s+1$ неизвестных ($\alpha_s = 1$). Для их определения необходимо $2s + 1 = p + 1$ условий. Следовательно, после определения неизвестных коэффициентов мы получим метод порядка $p = 2s$. Однако такие методы не имеют практической ценности, поскольку они неустойчивы. Г. Далквист (1956) показал, что порядок p устойчивого линейного s-шагового метода подчиняется ограничениям:

$$\begin{aligned} &p \le s + 2 && \text{при четных } s, \\ &p \le s + 1 && \text{при нечетных } s, \\ &p \le s && \text{при } \beta_s/\alpha_s \le 0. \end{aligned}$$

5.8 Практическая оценка локальной погрешности

Локальная погрешность многошагового метода может быть оценена в сопоставлении его решения на шаге с решением другого метода того же порядка. Согласно (5.9) и

(5.12) для метода порядка p будем иметь

$$\mathbf{x}(t_{n+s}) = \frac{1}{\alpha_s}\left(h\sum_{i=0}^{s}\beta_i\mathbf{f}(t_{n+i}) - \sum_{i=0}^{s-1}\alpha_i\mathbf{x}(t_{n+i})\right) + \frac{\rho}{\alpha_s} =$$

$$= \mathbf{x}_{n+s} + C_{p+1}\frac{h^{p+1}}{(p+1)!}\mathbf{x}^{(p+1)}(t_n) + O(h^{p+2}), \tag{5.19}$$

где

$$C_{p+1} = \frac{1}{\alpha_s}\left(\sum_i i^{p+1}\alpha_i - (p+1)\sum_i i^p\beta_i\right).$$

Решение вспомогательного метода $\hat{\mathbf{x}}_{n+s}$ того же порядка также будет удовлетворять соотношению (5.19), но со своей константой $\hat{C}_{p+1}$. Отсюда, игнорируя члены порядка h^{p+2}, получаем оценку локальной погрешности

$$\|\mathbf{e}\|_{cal} = \|\mathbf{x}_{n+s} - \mathbf{x}(t_{n+s})\| = \frac{|C_{p+1}|}{|C_{p+1} - \hat{C}_{p+1}|}\|\mathbf{x}_{n+s} - \hat{\mathbf{x}}_{n+s}\|, \tag{5.20}$$

а также уточненное решение

$$\mathbf{x}_{n+s} - \frac{C_{p+1}}{C_{p+1} - \hat{C}_{p+1}}(\mathbf{x}_{n+s} - \hat{\mathbf{x}}_{n+s}) = \mathbf{x}(t_{n+s}) + O(h^{p+2}).$$

5.9 Выбор шага

Изменение величины шага для эффективного численного интегрирования в многошаговых методах гораздо сложнее, чем в одношаговых. Это связано с тем, что при использовании многошаговой схемы интегрирования получение каждого следующего решения на новом шаге предполагает перевычисление предыдущих решений на новой равномерной сетке с соответствующим шагом. Для этого, как правило, прибегают к интерполяционным многочленам. Главный недостаток такого подхода состоит в том, что вместе с решениями необходимо перевычислять и значения правых частей уравнений. Другой подход — это построение многошаговых схем интегрирования на неравномерной сетке.

Например, для $(s+1)$-го решения при известных s предыдущих решениях обобщение явного метода Адамса на случай переменного шага дает формулу

$$\mathbf{x}_{s+1} = \mathbf{x}_s + h_{s+1}\sum_{i=0}^{s-1} g_i(s)\mathbf{\Phi}_i(s),$$

где

$$g_i(s) = \frac{1}{h_{s+1}}\int_{t_s}^{t_{s+1}}\prod_{j=0}^{i-1}\frac{t - t_{s-j}}{t_{s+1} - t_{s-j}}\mathrm{d}t, \quad \mathbf{\Phi}_i(s) = \mathbf{f}_{s,\ldots,s-i}\prod_{j=0}^{i-1}(t_{s+1} - t_{s-j}).$$

Для функций $g_i(s)$ и $\mathbf{\Phi}_i(s)$ возможно получить рекуррентные формулы. Тем не менее их перевычисление на каждом шаге чрезвычайно усложняет вычислительный процесс и это является существенным недостатком в использовании многошаговых методов с переменным шагом.

Каждый следующий шаг интегрирования определяется, как и в одношаговых методах, по формуле (3.30), где в качестве $\|\mathbf{e}\|_{cal}$ можно использовать оценку (5.20).

6 Геометрические методы

В последнее время в динамической астрономии возрос интерес к так называемым геометрическим численным методам интегрирования. Учитывая геометрические свойства дифференциальных уравнений, а точнее, их решений, эти методы позволяют качественно улучшить результаты численного интегрирования, в особенности, если оно выполняется на длительных (космогонических) интервалах времени.

Напомним сначала, что *канонической системой* уравнений называется система вида

$$\mathbf{x}' = \frac{\partial g}{\partial \mathbf{y}}, \quad \mathbf{y}' = -\frac{\partial g}{\partial \mathbf{x}}, \tag{6.1}$$

где $\mathbf{x}$ и $\mathbf{y}$ — канонические m-мерные векторы, а $g = g(\mathbf{x}, \mathbf{y})$ — *гамильтониан*, который при этом является *интегралом* канонической системы: $g(\mathbf{x}, \mathbf{y}) = \text{const}$. Отметим важное для нас свойство канонической системы: преобразование временного сдвига

$$\mathbf{x}(t) \xrightarrow{h} \bar{\mathbf{x}}(t) = \mathbf{x}(t+h), \quad \mathbf{y}(t) \xrightarrow{h} \bar{\mathbf{y}}(t) = \mathbf{y}(t+h) \tag{6.2}$$

является *симплектическим* (каноническим). Действительно, ведь новые переменные также удовлетворяют системе (6.1).

Кроме того, заметим, что все рассматриваемые нами дифференциальные уравнения *обратимы по времени*. Это означает, что выполненные последовательно прямое и обратное интегрирования задачи Коши с одним и тем же шагом h дают ее начальные условия:

$$\mathbf{x}(t_0) \xrightarrow{h} \mathbf{x}(t_0 + h) \xrightarrow{-h} \mathbf{x}(t_0). \tag{6.3}$$

6.1 Уравнения гармонического осциллятора

Рассмотрим уравнение гармонического осциллятора

$$x'' = -x, \tag{6.4}$$

которое представимо в виде системы уравнений первого порядка

$$x' = y, \quad y' = -x. \tag{6.5}$$

Отметим важные свойства системы (6.5): 1) она имеет интеграл

$$g(x, y) = \frac{1}{2}(x^2 + y^2) = \text{const}; \tag{6.6}$$

2) система канонична с гамильтонианом g; и, наконец, 3) она обратима по времени.

Таким образом, для интегрирования системы (6.5) целесообразно применять такие методы, которые бы учитывали названные свойства системы. Поскольку эти свойства тем или иным образом связаны с геометрией решений системы, то методы интегрирования, учитывающие хотя бы одно из них, обычно называют *геометрическими*.

Интересно заметить, что свойства 1)–3) характерны для многих дифференциальных уравнений небесной механики. В частности, уравнения кеплеровского движения (относительно центральной массы с гравитационным параметром μ) представимы в виде

$$\mathbf{x}' = \dot{\mathbf{x}} = \frac{\partial g}{\partial \dot{\mathbf{x}}}, \quad \dot{\mathbf{x}}' = -\mu \frac{\mathbf{x}}{|\mathbf{x}|^3} = -\frac{\partial g}{\partial \mathbf{x}}, \quad \text{где} \quad g = \frac{\dot{\mathbf{x}}^2}{2} - \frac{\mu}{|\mathbf{x}|} = \text{const};$$

и они обратимы по времени.

6.2 Методы Эйлера

Применим теперь к уравнениям (6.5) явный и неявный методы Эйлера:

$$x_{n+1} = x_n + h y_n, \quad y_{n+1} = y_n - h x_n; \quad x_{n+1} = x_n + h y_{n+1}, \quad y_{n+1} = y_n - h x_{n+1}. \tag{6.7}$$

Разрешение в неявной схеме $(n+1)$-го решения относительно n-го решения дает

$$x_{n+1} = \frac{x_n + h y_n}{1 + h^2}, \quad y_{n+1} = \frac{y_n - h x_n}{1 + h^2}.$$

Отсюда видно, что результаты интегрирования, получаемые методами Эйлера на одном шаге, совпадают с точностью до коэффициента $1/(1+h^2)$. Подставляя решения в интегральное соотношение $g(x, y)$ (6.6), получаем для явного метода

$$g(x_{n+1}, y_{n+1}) = (1 + h^2) g(x_n, y_n) = (1 + h^2)^{n+1} g(x_0, y_0);$$

и для неявного

$$g(x_{n+1}, y_{n+1}) = \frac{g(x_n, y_n)}{1 + h^2} = \frac{g(x_0, y_0)}{(1 + h^2)^{n+1}}.$$

Как видно, при пошаговом интегрировании методами Эйлера интегральное соотношение не сохраняется: в явном случае оно увеличивается, тогда как в неявном уменьшается.

Далее поскольку преобразование сдвига по времени (6.2) симплектично, то пошаговое отображение в соответствии со схемой интегрирования

$$x_n \xrightarrow{h} x_{n+1}, \quad y_n \xrightarrow{h} y_{n+1} \tag{6.8}$$

также должно быть симплектическим.

Проверим на симплектичность схемы Эйлера. Вычислим *скобки Пуассона*

$$\{x_{n+1}, y_{n+1}\} = \frac{\partial x_{n+1}}{\partial x_n} \frac{\partial y_{n+1}}{\partial y_n} - \frac{\partial x_{n+1}}{\partial y_n} \frac{\partial y_{n+1}}{\partial x_n}. \tag{6.9}$$

Тогда для явного и неявного методов соответственно получаем $\{x_{n+1}, y_{n+1}\} = 1 + h^2$ и $\{x_{n+1}, y_{n+1}\} = 1/(1+h^2)$. Вместе с тем известно, что преобразование (6.8) симплектично, если скобка Пуассона равна единице. Следовательно, методы Эйлера не симплектичны.

Обратимость по времени (6.3) уравнений (6.5) требует, чтобы схема интегрирования при ее последовательном использовании сначала с шагом h, а затем с шагом $-h$ давала исходные результаты. Однако согласно схемам Эйлера (6.7) будем иметь: в явном случае

$$x_n \xrightarrow{h} x_n + hy_n \xrightarrow{-h} (1+h^2)x_n, \quad y_n \xrightarrow{h} y_n - hx_n \xrightarrow{-h} (1+h^2)y_n;$$

и в неявном случае

$$x_n \xrightarrow{h} \frac{x_n + hy_n}{1+h^2} \xrightarrow{-h} \frac{x_n}{1+h^2}, \quad y_n \xrightarrow{h} \frac{y_n - hx_n}{1+h^2} \xrightarrow{-h} \frac{y_n}{1+h^2}.$$

Следовательно, схемы Эйлера не обратимы по времени.

Таким образом, методы Эйлера не геометрические. Хотя они могут стать таковыми после введения ряда незначительных модификаций.

Простым способом сохранения интегрального соотношения (6.6) является возвращение численного решения каждого шага на интегральную кривую, задаваемую уравнением $g(x,y) = g_0 = g(x_0, y_0)$. Как видно из (6.6), интегральные кривые уравнений гармонического осциллятора представляют собой концентрические окружности с центром в начале координат и радиусами $\sqrt{2g_0}$, определяемыми начальными условиями. Перемещения численных решений Δx_{n+1} и Δy_{n+1} будем осуществлять вдоль нормали к интегральной кривой в точке (x_{n+1}, y_{n+1}), т.е. вдоль вектора положения этой точки в фазовом пространстве. Тогда из геометрических соображений нетрудно получить поправки к решениям:

$$\Delta x_{n+1} = -\frac{\sqrt{2g(x_{n+1}, y_{n+1})} - \sqrt{2g_0}}{\sqrt{x_{n+1}^2 + y_{n+1}^2}}\, x_{n+1}, \quad \Delta y_{n+1} = -\frac{\sqrt{2g(x_{n+1}, y_{n+1})} - \sqrt{2g_0}}{\sqrt{x_{n+1}^2 + y_{n+1}^2}}\, y_{n+1}. \tag{6.10}$$

Интересно, что простой формальной заменой в схемах (6.7) первых (или вторых) частей для приближенных решений x_{n+1} (или y_{n+1}) удается получить две явные симплектические схемы интегрирования

$$y_{n+1} = y_n - hx_n, \quad x_{n+1} = x_n + hy_{n+1}; \quad x_{n+1} = x_n + hy_n, \quad y_{n+1} = y_n - hx_{n+1}. \tag{6.11}$$

Симплектичность модифицированных методов Эйлера несложно проверить, вычислив, как и ранее, скобки Пуассона (6.9), которые оказываются равными единице в обоих случаях.

Комбинируя схемы интегрирования методов Эйлера, можно получить новую, удовлетворяющую условию обратимости по времени (6.3): применим сначала явную схему Эйлера с шагом $h/2$, а затем — неявную с тем же шагом, в итоге будем иметь схему метода трапеций

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

Поскольку метод трапеций симметричен, он будет удовлетворять условию обратимости (6.3).

6.3 Проекционный метод

Для сохранения интегралов систем дифференциальных уравнений часто применяют так называемый *проекционный метод*, который, по-видимому, впервые предложил П. Накози в 1971 г. Заметим, что исследуемая нами ранее система (6.5) вместе с интегральным соотношением (6.6) являются частным случаем *дифференциально-алгебраических систем уравнений*, которые имеют вид

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad g(\mathbf{x}) = g(\mathbf{x}_0) = g_0 = \text{const},$$

где в качестве алгебраического уравнения обычно выступают интегралы системы дифференциальных уравнений либо налагаемые на нее связи. Рассмотрим проекционный метод на примере этой системы.

Идея метода состоит в том, чтобы после каждой стандартной процедуры интегрирования на шаге вносить в численное решение $\mathbf{x}_{n+1}$ поправки $\Delta\mathbf{x}_{n+1}$ за отклонение $g(\mathbf{x}_{n+1})$ относительно g_0, т.е. чтобы

$$g(\mathbf{x}_{n+1} + \Delta\mathbf{x}_{n+1}) = g_0.$$

С точностью до малых первого порядка имеем

$$g(\mathbf{x}_{n+1}) + \frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_{n+1}) \cdot \Delta\mathbf{x}_{n+1} = g_0. \tag{6.12}$$

Чтобы обеспечить наименьшую величину поправки $\Delta\mathbf{x}_{n+1}$ в соответствии с (6.12), ее направляют вдоль нормали к поверхности $g(\mathbf{x})$ в точке $\mathbf{x}_{n+1}$, представляя, таким образом, в виде

$$\Delta\mathbf{x}_{n+1} = \alpha \frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_{n+1}). \tag{6.13}$$

Коэффициент α определяется после подстановки (6.13) в (6.12). В итоге будем иметь поправку

$$\Delta\mathbf{x}_{n+1} = -\Delta g \left[\frac{\partial g}{\partial \mathbf{x}} \left(\frac{\partial g}{\partial \mathbf{x}} \cdot \frac{\partial g}{\partial \mathbf{x}} \right)^{-1} \right] (\mathbf{x}_{n+1}), \quad \text{где} \quad \Delta g = g(\mathbf{x}_{n+1}) - g_0. \tag{6.14}$$

Поскольку (6.14) получается из приближенного соотношения (6.12), то исправленное решение может еще не удовлетворять алгебраическому уравнению. В этом случае следует повторять процедуру уточнения решения (6.14) до тех пор, пока поправка $\Delta\mathbf{x}_{n+1}$ не станет достаточно малой.

Интересно заметить, что если мы выполним в (6.10) приближенную замену

$$\sqrt{2g(x_{n+1}, y_{n+1})} - \sqrt{2g_0} \approx \frac{g(x_{n+1}, y_{n+1}) - g_0}{\sqrt{2g(x_{n+1}, y_{n+1})}},$$

то получим коррекцию проекционного метода (6.14) применительно к дифференциально-алгебраической системе гармонического осциллятора (6.5) и (6.6).

6.4 Симплектические и симметрические методы

6.4.1 Простые симплектические методы

Симплектические схемы интегрирования строятся таким образом, чтобы отображение

$$\mathbf{x}_n \xrightarrow{h} \mathbf{x}_{n+1}, \quad \mathbf{y}_n \xrightarrow{h} \mathbf{y}_{n+1}$$

было каноническим. Тогда должна существовать *производящая функция*, например, $W = W(\mathbf{x}_n, \mathbf{y}_{n+1})$ такая, что

$$\mathbf{x}_{n+1} = \frac{\partial W}{\partial \mathbf{y}_{n+1}}, \quad \mathbf{y}_n = \frac{\partial W}{\partial \mathbf{x}_n}.$$

Выберем в качестве таких функций

$$W(\mathbf{x}_{n+1}, \mathbf{y}_n) = \mathbf{x}_{n+1}\mathbf{y}_n - hg(\mathbf{x}_{n+1}, \mathbf{y}_n), \quad W(\mathbf{x}_n, \mathbf{y}_{n+1}) = \mathbf{x}_n\mathbf{y}_{n+1} + hg(\mathbf{x}_n, \mathbf{y}_{n+1}),$$

тогда получаем симплектические схемы интегрирования первого порядка

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\frac{\partial g}{\partial \mathbf{y}_n}, \quad \mathbf{y}_{n+1} = \mathbf{y}_n - h\frac{\partial g}{\partial \mathbf{x}_{n+1}}; \quad \mathbf{x}_{n+1} = \mathbf{x}_n + h\frac{\partial g}{\partial \mathbf{y}_{n+1}}, \quad \mathbf{y}_{n+1} = \mathbf{y}_n - h\frac{\partial g}{\partial \mathbf{x}_n}, \tag{6.15}$$

в которых без труда узнаем первую и вторую симплектические схемы Эйлера (6.11) в случае интегрирования уравнений гармонического осциллятора, если вместо g подставить (6.6).

Для метода второго порядка выберем

$$W(\mathbf{x}_n, \mathbf{y}_{n+1}) = \mathbf{x}_n\mathbf{y}_{n+1} + hg(\mathbf{x}_n, \mathbf{y}_{n+1}) + \frac{h^2}{2}\frac{\partial g}{\partial \mathbf{y}_{n+1}}\frac{\partial g}{\partial \mathbf{x}_n}.$$

Тогда данной производящей функции будет соответствовать неявный метод

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\frac{\partial g}{\partial \mathbf{y}_{n+1}} + \frac{h^2}{2}\left(\frac{\partial^2 g}{\partial \mathbf{y}_{n+1}^2}\frac{\partial g}{\partial \mathbf{x}_n} + \frac{\partial g}{\partial \mathbf{y}_{n+1}}\frac{\partial^2 g}{\partial \mathbf{y}_{n+1}\partial \mathbf{x}_n}\right),$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n - h\frac{\partial g}{\partial \mathbf{x}_n} - \frac{h^2}{2}\left(\frac{\partial^2 g}{\partial \mathbf{y}_{n+1}\partial \mathbf{x}_n}\frac{\partial g}{\partial \mathbf{x}_n} + \frac{\partial g}{\partial \mathbf{y}_{n+1}}\frac{\partial^2 g}{\partial \mathbf{x}_n^2}\right).$$

Методы более высоких порядков можно конструировать подобным образом. Однако очевидно, что такой подход не столь эффективен, поскольку требует получения формул для вычисления частных производных высоких порядков от гамильтониана канонической системы.

6.4.2 Простые симметрические методы

Как мы уже отмечали, условию обратимости (6.3) отвечают симметрические методы. Для их построения часто прибегают к комбинированию обычных методов Рунге–Кутты с их смежными методами.

Смежным к исходному методу Рунге–Кутты (3.33) называется метод, получаемый в результате замены в исходном методе

$$\mathbf{x}_0 \leftrightarrow \mathbf{x}_1, \quad h \leftrightarrow -h, \quad t_0 \leftrightarrow t_0 + h. \tag{6.16}$$

Очевидно, повторная замена дает исходный метод, иначе говоря, смежный к смежному методу является исходным.

Смежный метод также является методом Рунге–Кутты, коэффициенты которого a^*_{ij}, b^*_i и c^*_i связаны с коэффициентами исходного метода a_{ij}, b_i и c_i как

$$a^*_{ij} = b_{s+1-j} - a_{s+1-i,s+1-j}, \quad b^*_i = b_{s+1-i}, \quad c^*_i = 1 - c_{s+1-i} \quad (i,j = 1,\ldots,s). \tag{6.17}$$

Действительно, выполнив замену (6.16) в методе (3.33), приводим его к смежному виду

$$\mathbf{x}_1 = \mathbf{x}_0 + h\sum_{j=1}^{s} b_j \mathbf{k}_j, \quad \mathbf{k}_i = \mathbf{f}(t_0 + h(1-c_i), \mathbf{x}_0 + h\sum_{j=1}^{s}(b_j - a_{ij})\mathbf{k}_j) \quad (i = 1,\ldots,s).$$

Чтобы упорядочить в формуле $\mathbf{k}_i$ по возрастанию коэффициентов c^*_i, нам остается произвести замену индексов $i \to s+1-i$ и $j \to s+1-j$. В итоге находим связь между коэффициентами (6.17).

Если исходный метод имеет порядок p, то ввиду замены (6.16) смежный метод будет аппроксимировать решение $\mathbf{x}(t_0)$ в окрестности $t_0 + h$ с ошибкой порядка h^{p+1}, следовательно, и смежный метод будет иметь порядок p.

Нетрудно видеть, что смежный к явному методу Эйлера будет неявный и наоборот. Интересно, что и симплектические методы Эйлера (6.15) также взаимосмежные. В то же время метод средней точки

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{f}\left(t_0 + \frac{h}{2}, \frac{\mathbf{x}_0 + \mathbf{x}_1}{2}\right)$$

и метод трапеций

$$\mathbf{x}_1 = \mathbf{x}_0 + \frac{h}{2}[\mathbf{f}(t_0, \mathbf{x}_0) + \mathbf{f}(t_1, \mathbf{x}_1)]$$

остаются без изменения относительно преобразований (6.16). Такие методы, как мы уже говорили (разд. 4.3), называются симметричными.

Введение смежных методов позволяет путем комбинирования их с исходными очень просто конструировать симметрические методы. Пусть $\mathbf{x}_1 = \Phi_h(\mathbf{x}_0, \mathbf{x}_1)$ — исходная схема интегрирования на шаге h, а $\mathbf{x}_1 = \Phi^*_h(\mathbf{x}_0, \mathbf{x}_1)$ — смежная. Тогда последовательное использование на половине шага сначала смежного метода, а затем исходного $\Phi_{h/2} \circ \Phi^*_{h/2}$ или наоборот $\Phi^*_{h/2} \circ \Phi_{h/2}$ дает симметричный метод. Симметричность следует из того, что $(\Phi_{h/2} \circ \Phi^*_{h/2})^* = (\Phi^{**}_{h/2} \circ \Phi^*_{h/2}) = \Phi_{h/2} \circ \Phi^*_{h/2}$. То же самое справедливо и для второго случая. Например, если мы применим на полшага сначала явный метод Эйлера, а затем неявный, то получим метод трапеций; если в обратном порядке — метод средней точки. Следует заметить, что результирующие методы имеют уже второй порядок, тогда как их составляющие — первый.

6.4.3 Методы Штермера–Верлета

Скомпануем две схемы интегрирования как

$$\Phi_{h/2} \circ \Phi^*_{h/2} \quad \text{и} \quad \Phi^*_{h/2} \circ \Phi_{h/2},$$

где в качестве Φ выберем первую симплектическую схему Эйлера (6.15). Тогда составные схемы будут

$$\begin{aligned}
&\mathbf{x}_{n+1/2} = \mathbf{x}_n + \frac{h}{2}\frac{\partial g}{\partial \mathbf{y}}(\mathbf{x}_{n+1/2}, \mathbf{y}_n),\\
&\mathbf{y}_{n+1} = \mathbf{y}_n - \frac{h}{2}\left[\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_{n+1/2}, \mathbf{y}_n) + \frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_{n+1/2}, \mathbf{y}_{n+1})\right],\\
&\mathbf{x}_{n+1} = \mathbf{x}_{n+1/2} + \frac{h}{2}\frac{\partial g}{\partial \mathbf{y}}(\mathbf{x}_{n+1/2}, \mathbf{y}_{n+1});
\end{aligned} \tag{6.18}$$

$$\begin{aligned}
&\mathbf{y}_{n+1/2} = \mathbf{y}_n - \frac{h}{2}\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_n, \mathbf{y}_{n+1/2}),\\
&\mathbf{x}_{n+1} = \mathbf{x}_n + \frac{h}{2}\left[\frac{\partial g}{\partial \mathbf{y}}(\mathbf{x}_n, \mathbf{y}_{n+1/2}) + \frac{\partial g}{\partial \mathbf{y}}(\mathbf{x}_{n+1}, \mathbf{y}_{n+1/2})\right],\\
&\mathbf{y}_{n+1} = \mathbf{y}_{n+1/2} - \frac{h}{2}\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}_{n+1}, \mathbf{y}_{n+1/2}).
\end{aligned} \tag{6.19}$$

Методы, определяемые схемами (6.18) и (6.19), называются *методами Штермера–Верлета.*

Для важного в классической механике случая $g = \mathbf{y}^2/2 + U(\mathbf{x})$ схемы перепишутся как

$$\mathbf{x}_{n+1/2} = \mathbf{x}_n + \frac{h}{2}\mathbf{y}_n, \quad \mathbf{y}_{n+1} = \mathbf{y}_n - h\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x}_{n+1/2}), \quad \mathbf{x}_{n+1} = \mathbf{x}_{n+1/2} + \frac{h}{2}\mathbf{y}_{n+1}; \tag{6.20}$$

$$\mathbf{y}_{n+1/2} = \mathbf{y}_n - \frac{h}{2}\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x}_n), \quad \mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{y}_{n+1/2}, \quad \mathbf{y}_{n+1} = \mathbf{y}_{n+1/2} - \frac{h}{2}\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x}_{n+1}). \tag{6.21}$$

Совокупность формул интегрирования (6.20) образует так называемую «схему прыгающей лягушки». Если вторую схему (6.21) записать для двух шагов, то удается исключить переменные $\mathbf{y}$ и в результате можно получить многошаговую схему

$$\mathbf{x}_{n+1} - 2\mathbf{x}_n + \mathbf{x}_{n-1} = -h^2\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x}_n),$$

которая применялась еще И. Ньютоном, Ж.Л. Даламбером и Й.Ф. Энке.

Отметим важные свойства методов Штермера–Верлета: 1) методы имеют второй порядок; 2) они симплектические и симметрические; 3) для разделенного гамильтониана $g(\mathbf{x}, \mathbf{y}) = g_1(\mathbf{x}) + g_2(\mathbf{y})$ методы явные. Таким образом, методы Штермера–Верлета геометрические, однако для долгосрочного интегрирования в задачах динамической астрономии они в оригинальном виде оказываются не востребованы, поскольку имеют низкий порядок. Тем не менее эти методы могут быть весьма полезными для получения составных геометрических методов высоких порядков.

6.4.4 Симплектические и симметрические методы высоких порядков

Составные методы. Интересный способ построения симплектических и симметрических методов высоких порядков — это последовательное использование на шаге простых схем интегрирования. Подобным образом мы уже получили методы Штермера–Верлета, а также симметрические методы средней точки и трапеций.

Особого внимания достойны составные *методы Йошиды.* Х. Йошида показал, что если симметрический метод Φ_h имеет порядок $p = 2q$, а некоторые константы c_0 и c_1 удовлетворяют условиям

$$c_0 + 2c_1 = 1, \quad c_0^{2q+1} + 2c_1^{2q+1} = 0, \tag{6.22}$$

то составной метод

$$\Psi_h = \Phi_{c_1 h} \circ \Phi_{c_0 h} \circ \Phi_{c_1 h} \tag{6.23}$$

является симметрическим и имеет порядок $p = 2q + 2$. Таким образом, это позволяет получить составные методы любого четного порядка. Допустим, мы имеем симметрический метод второго порядка $\Phi_h^{(2)}$, например, средней точки или Штермера–Верлета. Тогда согласно (6.22) и (6.23) получаем метод четвертого порядка

$$\Phi_h^{(4)} = \Phi_{c_1 h}^{(2)} \circ \Phi_{c_0 h}^{(2)} \circ \Phi_{c_1 h}^{(2)}, \quad \text{где} \quad c_0 = -\frac{1}{3}(1 + \sqrt[3]{2})^2, \quad c_1 = \frac{1}{3}(2 + \sqrt[3]{2} + 1/\sqrt[3]{2}).$$

Применяя (6.23) к $\Phi_h^{(4)}$, получаем метод шестого порядка $\Phi_h^{(6)}$ с соответствующими коэффициентами c_0 и c_1 и так далее. Очевидно, если исходный метод $\Phi_h^{(2)}$ симплектический (например, Штермера–Верлета), то и составной метод, построенный на его основе, также будет симплектическим.

Заметим, что очередное повышение порядка по схеме (6.23) утраивает применение исходного метода $\Phi_h^{(2)}$, так что в составном методе порядка $p = 2q$ число использования исходного метода будет равно $s = 3^{q-1}$. Например, $s = 27$ уже при порядке $p = 8$. В связи с этим для методов высоких порядков возникает проблема уменьшения числа s при сохранении порядка. Разрешая эту проблему, У. Кахану и Р. Ли (1997) удалось получить составной симметрический метод восьмого порядка (отличный от метода Йошиды) с числом $s = 17$.

Отметим очевидные свойства составных методов вида

$$\Psi_h = \Phi_{c_{(s-1)/2} h} \circ \ldots \circ \Phi_{c_1 h} \circ \Phi_{c_0 h} \circ \Phi_{c_1 h} \circ \ldots \circ \Phi_{c_{(s-1)/2} h},$$

такие как: 1) методу Ψ передаются свойства симметричности либо симплектичности метода Φ; 2) если метод Φ сохраняет какой-либо инвариант системы уравнений, то и составной метод Ψ также будет сохранять этот инвариант.

Методы Гаусса. В 1988 г. И.М. Санц-Серна и Ф.М. Ласани обнаружили, что если коэффициенты неявного метода Рунге–Кутты удовлетворяют условиям

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0 \quad (i, j = 1, \ldots, s), \tag{6.24}$$

то метод является симплектическим. Нетрудно убедиться, что согласно (6.24) среди одноэтапных методов Рунге–Кутты симплектическим будет только метод средней точки с коэффициентами $a_{11} = 1/2$ и $b_1 = 1$. В общем случае условиям (6.24) оказывается удовлетворяют методы Гаусса–Лежандра (см. разд. 3.10), которые, кроме того, являются симметрическими.

Многошаговые методы. Эффективными геометрическими методами высоких порядков среди многошаговых являются симметрические *методы Куинлэна–Тремейна*, которые применяются для интегрирования уравнений второго порядка $\mathbf{x}'' = \mathbf{f}(\mathbf{x})$ и имеют вид

$$\sum_{i=0}^{s} \alpha_i \mathbf{x}_{n+i} = h^2 \sum_{i=0}^{s} \beta_i \mathbf{f}(\mathbf{x}_{n+i}),$$

где $\alpha_i = \alpha_{s-i}$ и $\beta_i = \beta_{s-i}$ $(i = 0, \ldots, s)$. Авторы получили методы вплоть до 14-го порядка и продемонстрировали их высокую эффективность на примере планетной задачи.

6.4.5 Особенности в использовании симплектических методов

Очевидно, вдоль точного решения канонической системы ее гамильтониан сохраняется, тогда как вдоль приближенного решения нет. Тем не менее в силу симплектичности схем интегрирования они будут сохранять другой, близкий к оригинальному гамильтониан. Так, нетрудно убедиться, что методы Эйлера (6.11) не сохраняют интеграл гармонического осциллятора (6.6), однако их приближенные решения на каждом шаге удовлетворяют другим интегральным соотношениям

$$\bar{g}(x, y) = g(x, y) - \frac{h}{2}xy = \text{const} \quad \text{и} \quad \bar{g}(x, y) = g(x, y) + \frac{h}{2}xy = \text{const} \tag{6.25}$$

соответственно для первой и второй схем (6.11). Функции $\bar{g}$ еще называют *связанными гамильтонианами*. Соотношения (6.25) описывают эллипсы, большие полуоси которых направлены вдоль прямых $y = x$ и $y = -x$. Таким образом, приближенные решения x_{n+1} и y_{n+1} (6.11) согласно (6.25) будут лежать на замкнутых и компактных (при малых h) кривых, поэтому ошибка в g будет ограничена:

$$\Delta g = g(x_{n+1}, y_{n+1}) - g(x_0, y_0) = \pm\frac{h}{2}(x_{n+1}y_{n+1} - x_0 y_0),$$

хотя при использовании обычных (негеометрических) методов ее величина имеет тенденцию к неограниченному росту.

В случае применения методов Эйлера (6.15) для интегрирования канонической системы с произвольным гамильтонианом g известно, что связанный гамильтониан будет иметь вид

$$\bar{g} = g + \sum_{i=1}^{\infty} (\mp h)^i g_i = g + O(h), \tag{6.26}$$

$$g_1 = \frac{1}{2}\frac{\partial g}{\partial \mathbf{x}}\frac{\partial g}{\partial \mathbf{y}}, \quad g_2 = \frac{1}{12}\left[\frac{\partial^2 g}{\partial \mathbf{x}^2}\left(\frac{\partial g}{\partial \mathbf{y}}\right)^2 + \frac{\partial^2 g}{\partial \mathbf{y}^2}\left(\frac{\partial g}{\partial \mathbf{x}}\right)^2\right], \quad \ldots$$

В степенном ряде (6.26) знак минус при h соответствует первой схеме Эйлера (6.15), тогда как плюс — второй. Если используются симплектические методы p-го порядка, то связанный гамильтониан формально можно представить как

$$\bar{g} = g + O(h^p).$$

Таким образом, симплектический метод будет сохранять связанный гамильтониан $\bar{g}$, близкий к оригинальному g, причем близость гамильтонианов определяется не только величиной шага интегрирования h, но и порядком метода p. Это, в свою очередь, обеспечивает ограниченность ошибки Δg.

Следует заметить, что интегральное соотношение $\bar{g} = \mathrm{const}$ явно зависит от h, поэтому оно будет сохраняться только при постоянном шаге. Если шаг переменный, то связанный гамильтониан уже не будет интегралом и в этом случае ограниченность ошибки Δg не гарантируется. Так, при симплектическом интегрировании орбиты задачи двух тел с переменным шагом гамильтониан задачи (кеплеровская энергия) не сохраняется и ошибка в интегральном соотношении неограниченно возрастает. В связи с этим для эффективного использования симплектических методов необходимо выполнять интегрирование с постоянным шагом.

Впечатляющими оказываются результаты при использовании симплектической схемы средней точки, которая также является симметрической. Применительно к уравнениям гармонического осциллятора она точно сохраняет гамильтониан системы. Следует заметить, что в случае линейных систем дифференциальных уравнений схема средней точки совпадает с симметричной схемой трапеций, которая, вообще говоря, не симплектична. Вместе с тем во многих последних работах по геометрическим интеграторам показывается, что симметрические методы так же, как и симплектические, позволяют ограничить ошибки в гамильтониане при интегрировании канонических систем.

Литература

1. Хайрер Э., Нерсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. М.: Мир, 1990.

2. Бахвалов Н.С. Численные методы. М.: Наука, 1973.

3. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. М.: Бином, 2004.

4. Everhart E. Implicit Single Sequence Methods for Integrating Orbits // Cel. Mech. 1974. V. 10. P. 35–55.

5. Butcher J. Numerical Methods for Ordinary Differential Equations. John Wiley & Sons, 2003.

6. Hairer E., Lubich C., Wanner G. Geometric Numerical Integration Structure-Preserving Algorithms for Ordinary Differential Equations. Springer Series in Computational Mathematics. V. 31. Berlin: Springer, 2002.

7. Вержбицкий В.М. Численные методы. Математический анализ и обыкновенные дифференциальные уравнения. М.: Высш. шк., 2001.

8. Арушанян О.Б., Залеткин С.Ф. Численное решение обыкновенных дифференциальных уравнений на Фортране. М.: Изд-во МГУ, 1990.