Санкт-Петербургский государственный университет
Факультет прикладной математики — процессов управления

А. С. Еремин, И. В. Олемской, О. С. Фирюлина

Практикум на ЭВМ по численным методам

Тема 8. Решение задачи Коши для обыкновенных дифференциальных уравнений

Методические указания

Санкт-Петербург
2016

Содержание

Постановка задачи

В области $D = \{x_0 \leqslant x \leqslant x_f,\ |y^i - y_0^i| \leqslant \bar{y}_i,\ i = \overline{1,m}\} \in \mathbb{R}^{n+1}$ определена функция $f : D \to \mathbb{R}^m$,

$$f \equiv f(x, y^1, \ldots, y^m), \quad (x, y^1, \ldots, y^m) \in D.$$

Обозначим $y = (y^1, \ldots, y^m)^T$, и будем считать, что f — тоже вектор-столбец длины m.

Необходимо найти решение системы обыкновенных дифференциальных уравнений (ОДУ)

$$\frac{dy}{dx} = f(x, y), \tag{1}$$

удовлетворяющее начальному условию

$$y(x_0) = y_0. \tag{2}$$

Используя метод последовательных приближений Пикара[1], можно получить точное решение $y(x)$ задачи Коши[2] (1), (2) как предел последовательности

$$y_0(x),\ y_1(x),\ \ldots,\ y_k(x),\ \ldots, \tag{3}$$

где

$$y_k(x) = y(x_0) + \int_{x_0}^{x} f(x, y_{k-1}(x))dx, \quad k = 1, 2, \ldots \tag{4}$$

Для сходимости этой последовательности необходимо выполнение следующих условий:

[1] *Шарль Эми́ль Пика́р (фр. Charles Émile Picard)* (1856–1941), французский математик. Известен фундаментальными результатами в области математического анализа. Внес существенный вклад также в теорию дифференциальных уравнений, теорию функций, топологию, теорию групп. Для линейных дифференциальных уравнений разработал аналог теории Галуа. Часть его трудов посвящена истории и философии математики.

[2] *Огюсте́н Луи́ Коши́ (фр. Augustin Louis Cauchy)* (1789–1857), французский математик и механик. Впервые дал строгое определение основным понятиям математического анализа — пределу, непрерывности, производной, дифференциалу, интегралу, сходимости ряда и т. д. В комплексном анализе создал теорию интегральных вычетов. В математической физике изучал краевую задачу с начальными условиями, которая с тех пор называется «задачей Коши». Также занимался механикой сплошных сред, оптикой, астрономией и другими областями естествознания.

- $f(x,y)$ непрерывна в D,
- $f(x,y)$ удовлетворяет условию Липшица[3] по аргументу y

$$\|f(x,y^*) - f(x,y^{**})\| \leqslant L\|y^* - y^{**}\| \tag{5}$$

для всех $x \in [x_0, x_f]$ и всех компонент векторов y^* и y^{**}.

При этих предположениях $y_k(x)$ равномерно сходится к точному решению задачи Коши, поэтому для достаточно больших k отклонение $\|y(x)-y_k(x)\|$ не превышает заданной величины. Таким образом, в качестве искомого решения можно взять $y_k(x)$. Практическая реализация этого метода затруднена по причине того, что для сложной функции $f(x,y)$ интеграл не берется в квадратурах и решение нельзя получить в аналитическом виде.

Обсуждаемые ниже численнные методы известны как дискретные, т. е. такие методы, посредством которых вычисляется последовательность приближений $y_n \approx y(x_n)$ к решению на множестве точек $x_{n+1} = x_n + h_{n+1}$, $n = 0, 1, ..., N-1$. Причем $x_N = x_f$. Значение $h_n > 0$ называется n-м *шагом* сетки. В большинстве рассматриваемых методов будем считать шаги постоянными, т. е. $\forall n = 1, ..., N$ $h_n = h$, $h = \text{const} > 0$.

Для простоты изложения будем предполагать $m = 1$. В рамках данного пособия применение изучаемых методов к системе производится формальной заменой скаляров на векторы.

[3] *Ру́дольф О́тто Си́гизмунд Ли́пшиц (нем. Rudolf Otto Sigismund Lipschitz)* (1832–1903), немецкий математик. В основном работал в области математического анализа, теории дифференциальных уравнений, теоретической механики и алгебры. Константа Липшица играет важную роль в численных методах.

Одношаговые методы

Одношаговые методы — методы, которые дают последовательные приближения y_{n+1} к значениям точного решения $y(x_{n+1})$ в узлах сетки x_{n+1} на основе ранее вычисленных (или заданных начальными условиями) приближений y_n к решению в точках x_n. В общем виде их можно представить как

$$y_{n+1} = F(f, x_{n+1}, y_{n+1}, x_n, y_n). \tag{6}$$

Как можно заметить, в правой части в общем виде содержится искомое значение y_{n+1}. В том случае, когда такой зависимости нет, метод называют *явным*

$$y_{n+1} = F(f, x_{n+1}, x_n, y_n). \tag{7}$$

Именно такие методы рассмотрены в настоящем пособии. Методы общего вида (6) называют *неявными*, так как в них значение y_{n+1} не выражено напрямую и для его нахождения приходится решать алгебраическое уравнение.

Методы разложения в ряд Тейлора

Предположим, что правая часть $f(x, y)$ дифференциального уравнения (1) имеет непрерывные частные производные до порядка p. Тогда искомое решение $y(x)$ имеет непрерывные производные до $p+1$-го порядка включительно. Точное значение решения в узле x_{n+1}, если известно точное значение решения в точке x_n, запишем по формуле Тейлора[4]:

$$y(x_{n+1}) = y(x_n) + hy'(x_n) + \ldots + \frac{h^p}{p!}y^{(p)}(x_n) + \frac{h^{p+1}}{(p+1)!}y^{(p+1)}(\xi) =$$
$$= y(x_n) + h\Delta(x_n, y(x_n), h), \quad h = x_{n+1} - x_n, \quad \xi \in (x_n, x_{n+1}).$$

[4] *Брук Тéйлор (англ. Brook Taylor)* (1685—1731), английский математик. Занимался задачами по весьма разнообразным темам: о центре качаний, о полете снарядов, о взаимодействии магнитов, о капиллярных явлениях, о сцеплении между жидкостями и твердыми телами. Помимо формулы, выражающей значение голоморфной функции через значения всех ее производных в одной точке, в трактате 1715–1717 гг. представил теорию колебания струн, в которой он пришел к тем же самым результатам, к которым впоследствии пришли Даламбер и Лагранж.

Если теперь этот ряд оборвать, ограничиться только первыми $p+1$ членами (до h^p) и заменить точное значение $y(x_n)$ его приближением y_i, то получим формулу

$$y_{n+1} = y_n + h\varphi(x_n, y_n, h) = y_n + hy'_n + \frac{h^2}{2}y''_n + \ldots + \frac{h^p}{p!}y_n^{(p)} \qquad (8)$$

для нахождения приближенного решения. Производные, входящие в правую часть (8), могут быть фактически найдены последовательным дифференцированием:

$$\begin{aligned} y'_n &= f(x_n, y_n), \\ y''_n &= \{f'_x + ff'_y\}\big|_{x_n}, \\ y'''_n &= \{f''_{xx} + 2ff''_{xy} + f^2 f''_{yy} + (f'_x + ff'_y)f'_y\}\big|_{x_n} \end{aligned} \qquad (9)$$

и т. д. Так для $p = 1$ и $p = 2$ получим расчетные схемы

$$y_{n+1} = y_n + hf(x_n, y_n) \quad \text{(явный метод Эйлера)} \qquad (10)$$

и

$$y_{n+1} = y_n + h\left[f(x_n, y_n) + \frac{h}{2}\left(f'_x(x_n, y_n) + f(x_n, y_n)f'_y(x_n, y_n)\right)\right], \quad (11)$$

по которым можно последовательно получать приближенное решение $\{y_i\}$. Такие формулы не требуют вычисления дополнительных начальных условий и позволяют легко менять шаг интегрирования. Недостатком расчетных схем метода разложения в ряд Тейлора является то, что их практическое применение ограничено лишь задачами, для которых легко вычисляются полные производные высшего порядка.

Явные методы Рунге — Кутты

Рунге[5], Хойн[6] и Кутта[7] предложили подход, основанный на построении приращения $\varphi(x_n, y_n, h)$, которое окажется достаточно близко к $\Delta(x_n, y(x_n), h)$, но не содержит производных от функции $f(x, y)$. Способ нахождения приближения к решению y_{n+1}, основанный на использовании приближенных рядов Тейлора, и называется методом Рунге — Кутты.

Общая схема явных методов была впервые выписана Куттой, хотя первые из таких методов были представлены Рунге и Хойном.

Определение 1. Пусть s — целое положительное число, которое мы будем называть числом «этапов», или «стадий», и a_{21}, a_{31}, ..., a_{s1}, a_{s2}, ..., $a_{s,s-1}$, b_1, ..., b_s, c_1, ..., c_s — вещественные коэффициенты. Тогда метод нахождения приближения в точке $x_1 = x_0 + h$

$$\begin{aligned} y(x_1) \approx y_1 = y_0 + h\sum_{i=1}^{s} b_i K_i, \text{ где} \\ K_i = f\Big(x_0 + c_i h,\ y_0 + h\sum_{j=1}^{i-1} a_{ij} K_j\Big), \quad i = 1, ..., s, \end{aligned} \tag{12}$$

называется *s-этапным явным методом Рунге — Кутты* (ЯМРК) для задачи Коши (1), (2).

Определение 2. Говорят, что ЯМРК (12) имеет *порядок p* (*порядок точности на шаге*, или *локальный порядок*), если для достаточно гладких задач (1), (2) и достаточно малого шага h

$$\|y(x_1) - y_1\| \leqslant Ch^{p+1}. \tag{13}$$

[5] *Карл Давúд Тольмé Рýнге (нем. Carl David Tolme Runge)* (1856–1927), немецкий математик, физик и спектроскопист. Внес существенный вклад в численный анализ, в частности, явился одним из разработчиков методов решения обыкновенных дифференциальных уравнений, носящих теперь общее название методов типа Рунге — Кутты, а также практического метода оценки погрешности приближенного интегрирования.

[6] *Карл Хойн (или Гойн) (нем. Karl Heun)* (1859–1929), немецкий математик. Занимался теорией дифференциальных уравнений, специальных функций и численных методов.

[7] *Мáртин Вильгéльм Кýтта (нем. Martin Wilhelm Kutta)* (1867–1944), немецкий математик. Помимо методов Рунге — Кутты, известен благодаря работе над аэродинамической поверхностью, которой в России занимался Николай Егорович Жуковский. В зарубежной литературе теорема Жуковского называется теоремой Кутты — Жуковского.

Иначе говоря, если ряды Тейлора для точного решения $y(x_0 + h)$ и полученного приближения к нему y_1 совпадают до члена h^p включительно.

Все последующие шаги для нахождения приближений в точках x_2, x_3 и далее выполняются по тем же самым формулам, но вместо (x_0, y_0) начальными данными выступают (x_1, y_1), (x_2, y_2) и т. д.

Для представления одношаговых методов типа Рунге — Кутты удобно использовать табличное представление коэффициентов. Следующая таблица носит название *таблицы Бутчера*[8]

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

Можно считать, что на свободных местах стоят нули. При записи их опускают, так как эта часть таблицы в явных методах не используется.

Любой ЯМРК характеризуется числом этапов, которое дает представление о требуемых вычислительных затратах, так как самым трудоемким считается вычисление функции f, и *порядком*.

ЯМРК очень просты в реализации и потому широко используются для практических вычислений: они не требует вычисления дополнительных начальных значений и позволяют легко менять шаг интегрирования (в соотвествующем разделе пособия рассматривается, зачем и как). Причем в отличие от метода Тейлора здесь не требуется вычисления полных производных точного решения. И приращение ищется в виде линейной комбинации вычислений правой части исходного дифференциального уравнения на шаге интегрирования.

Построение методов Рунге — Кутты

Вполне закономерно возникают вопросы, как же подобрать коэффициенты метода для обеспечения желаемого порядка p, какое мини-

[8] *Джон Ча́рльз Бу́тчер (англ. John Charles Butcher)* (род. 1933), новозеландский математик. Работает в области численного решения обыкновенных дифференциальных уравнений, в частности, множество работ посвящено методам Рунге — Кутты и общим многошаговым методам.

мальное число этапов s для этого нужно и какого порядка можно добиться при заданном s.

Ответим на первый вопрос. Для изложения алгоритма построения s-этапного ЯМРК порядка p введем в рассмотрение функцию

$$\Psi(h) = y(x_1) - y_1 = y(x_1) - y(x_0) - h\sum_{i=1}^{s} b_i K_i, \tag{14}$$

которую в дальнейшем будем называть *локальной погрешностью* одношагового метода.

Предполагаем, что в рассматриваемой области функция $f(x, y)$ имеет непрерывные частные производные до некоторого порядка p. Тогда искомое решение будет иметь непрерывные производные до порядка $p + 1$. Выберем параметры метода b_i, c_i, a_{ij} так, чтобы разложение методической погрешности (14) по степеням h в ряд Тейлора

$$\Psi(h) = \sum_{v=0}^{p} \frac{\Psi^{(v)}(0)}{v!} h^v + \frac{\Psi^{(p+1)}(\vartheta h)}{(p+1)!} h^{p+1}, \quad 0 < \vartheta \leqslant 1, \tag{15}$$

начиналось со степени $p + 1$ при произвольной функции $f(x, y)$ и произвольном шаге h, т. е.

$$\Psi(h) = \frac{\Psi^{(p+1)}(\vartheta h)}{(p+1)!} h^{p+1}, \quad 0 < \vartheta \leqslant 1. \tag{16}$$

Это возможно тогда и только тогда, когда параметры метода b_i, c_i, a_{ij} обеспечивают выполнение равенств

$$\Psi(0) = \Psi'(0) = \Psi''(0) = \cdots = \Psi^{(p)}(0) = 0. \tag{17}$$

Понятно, что для метода порядка p всегда найдется некоторая гладкая функция $f(x, y)$, для которой

$$\Psi^{(p+1)}(0) \neq 0.$$

По построению метода ясно, что условие $\Psi(0) = 0$ выполняется всегда. Условия же $\Psi^{(v)}(0) = 0$, $v = 1, ..., p$ означают выполнение равенств

$$y^{(v)}(x_0) = v \sum_{i=1}^{s} b_i K_i^{(v-1)} \Bigg|_{h=0}, \quad v = 1, ..., p. \tag{18}$$

Производные $y^{(v)}$ могут быть вычислены указанным ранее способом (9). Проблема заключается в вычислении производных $K_i^{(v)}$.

Введем обозначения

$$X_i = x_0 + c_i h, \quad Y_i = y_0 + h\sum_{j=1}^{i-1} a_{ij} K_j.$$

Тогда

$$K_1 = f(x_0, y_0), \quad K_1^{(v)} \equiv 0, \quad v \geqslant 1,$$

а для $i \geqslant 2$ получим

$$K_i = f(X_i, Y_i), \qquad K_i|_{h=0} = f(x_0, y_0),$$

$$K_i' = f_x'(X_i, Y_i)c_i + f_y'(X_i, Y_i)\sum_{j=1}^{i-1} a_{ij}(K_j + hK_j'),$$

$$K_i'|_{h=0} = \left\{ f_x' c_i + f f_y' \sum_{j=1}^{i-1} a_{ij} \right\}\Bigg|_{(x_0,y_0)},$$

$$K_i'' = f_{xx}''(X_i, Y_i)c_i^2 + 2f_{xy}''(X_i, Y_i)c_i \sum_{j=1}^{i-1} a_{ij}(K_j + hK_j') +$$

$$+ f_{yy}''(X_i, Y_i)\left(\sum_{j=1}^{i-1} a_{ij}(K_j + hK_j')\right)^2 +$$

$$+ f_y'(X_i, Y_i)\left(\sum_{j=1}^{i-1} a_{ij}(2K_j' + hK_j'')\right),$$

$$K_i''|_{h=0} = \left\{ f_{xx}'' c_i^2 + 2f f_{xy}'' c_i \sum_{j=1}^{i-1} a_{ij} + f^2 f_{yy}'' \left(\sum_{j=1}^{i-1} a_{ij}\right)^2 + \right.$$

$$\left. + \; 2f_y' \sum_{j=1}^{i-1} a_{ij}\left(f_x' c_j + f f_y' \sum_{\nu=1}^{j-1} a_{j\nu}\right)\right\}\Bigg|_{(x_0,y_0)} \quad \text{и т. д.}$$

Используя эти формулы и формулы (9), видим, что в уравнении (18) правая часть равна $f(x_0, y_0)$, а в левой стоит $\sum_{i=1}^{s} b_i f(x_0, y_0)$.

Таким образом, чтобы метод имел порядок 1, необходимо, чтобы

$$\sum_{i=1}^{s} b_i = 1. \tag{19}$$

Уравнения для более высоких порядков существенно упрощаются, если ввести дополнительные связи между параметрами метода

$$\sum_{j=1}^{i-1} a_{ij} = c_j, \quad i = 1, ..., s. \tag{20}$$

Тогда порядок 2 обеспечивается только одним новым условием

$$\sum_{i=1}^{s} b_i c_i = \frac{1}{2}, \tag{21}$$

а порядок 3 — еще двумя:

$$\sum_{i=1}^{s} b_i c_i^2 = \frac{1}{3}, \tag{22}$$

$$\sum_{i=1}^{s} b_i \sum_{j=1}^{i-1} a_{ij} c_j = \frac{1}{6}. \tag{23}$$

Выбирая число этапов s, мы по сути задаем количество подлежащих определению параметров метода. Желаемый порядок метода p посредством (17) определяет систему нелинейных алгебраических уравнений, которой эти параметры должны удовлетворять. Уравнения входящие в эту систему называются *условиями порядка*. Иногда к условиям порядка добавляют и другие ограничения, связанные с дополнительными критериями: минимальный объем требуемой оперативной памяти, уменьшение среднего числа вычислений правой части, фиксированные узлы, повышение численной устойчивости метода, обеспечение геометрических свойств и др.

Само собой, что при заданных p и s у системы условий порядка может как быть несколько решений (семейство), так и не быть решений вовсе. Ответы на вопросы о связи числа этапов и порядка следует из разрешимости этой системы.

Двухэтапные методы второго порядка

Построим метод второго порядка. Очевидно, что в силу равенства $c_1 = 0$ условие (21) не может быть удовлетворено при $s = 1$. Таким образом, находим, что не существует одноэтапного явного метода второго порядка.

Выберем $s = 2$. Условия порядка примут вид

$$\begin{aligned} &b_1 + b_2 = 1, \\ &b_2 c_2 = \frac{1}{2}, \\ &a_{21} = c_2 \quad (\text{или } b_2 a_{21} = \frac{1}{2}, \text{ если не учитывать (20)}). \end{aligned} \tag{24}$$

В этой системе три уравнения и четыре неизвестных. Ее решение представляет собой однопараметрическое семейство. Выберем, например, c_2 в качестве свободного параметра. Получим

$$a_{21} = c_2, \quad b_2 = \frac{1}{2c_2}, \quad b_1 = 1 - \frac{1}{2c_2}, \quad c_2 \neq 0. \tag{25}$$

Обычно выбирают такие коэффициенты, которые дают удобные для вычислений расчетные формулы, и такие, которые обеспечивают подавление того или иного слагаемого в главном члене погрешности (выражении, содержащем h^{p+1}).

Например, полагая $c_2 = 1/2$ получим формулу, построенную Рунге в 1895 году, как усовершенствование метода Эйлера[9] — первый метод Рунге — Кутты:

$$y(x_0 + h) \approx y_0 + hf\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f(x_0, y_0)\right). \tag{26}$$

Его называют *явным методом средней точки*. Если подставить значения коэффициентов в формулы для K_i'', можно выразить локальную погрешность решения с точностью до h^3:

$$y(x_1) - y_1 = \frac{h^3}{24}\Big(f''_{xx} + 2ff''_{xy} + f^2 f''_{yy} + 4\big(f'_y f'_x + f(f'_y)^2\big)\Big)\Big|_{(x_0, y_0)} + O(h^4).$$

[9] *Леона́рд Э́йлер (нем. Leonhard Euler)* (1707–1783), швейцарский, немецкий и российский математик и механик, внесший фундаментальный вклад в развитие этих наук. Автор трудов по математическому анализу, дифференциальной геометрии, теории чисел, приближенным вычислениям, небесной механике, математической физике, оптике, астрономии, баллистике, кораблестроению, теории музыки и другим областям. Основание натуральных логарифмов e называется константой Эйлера и обозначается первой буквой его фамилии.

Если правая часть исходного уравнения (1) не зависит от y, то задача сводится к численному интегрированию: все слагаемые, содержащие производные по y исчезают, и метод Рунге становится равносилен квадратурной формуле средних прямоугольников (первый этап становится не нужным).

Значение $c_2 = 1$ дает *метод Хойна*, или *явный метод трапеций*, или *усовершенствованный метод Эйлера*:

$$y(x_0 + h) \approx y_0 + \frac{h}{2}\Big(f(x_0, y_0) + f(x_0 + h, y_0 + hf(x_0, y_0))\Big). \qquad (27)$$

Методическая погрешность метода Хойна имеет вид

$$y(x_1) - y_1 = -\frac{h^3}{12}\Big(f''_{xx} + 2ff''_{xy} + f^2 f''_{yy} - 2\big(f'_y f'_x + f(f'_y)^2\big)\Big)\Big|_{(x_0,y_0)} + O(h^4).$$

Если правая часть исходного дифференциального уравнения не зависит от y, то применение метода Хойна становится равносильно использованию квадратурной формулы трапеций.

Таблицы Бутчера для рассмотренных методов имеют вид

0	
$\frac{1}{2}$	$\frac{1}{2}$
	$0 \quad 1$

Метод Рунге

0	
1	1
	$\frac{1}{2} \quad \frac{1}{2}$

Метод Хойна

Методы третьего порядка с тремя этапами

Из-за условия (23) явные методы с двумя этапами не могут иметь третий порядок (проверьте!), поэтому положим $s = 3$. В этом случае система из шести уравнений с восемью неизвестными дает двухпараметрическое семейство решений. Приведем два наиболее популярных метода.

Первый метод Рунге—Кутты третьего порядка построил Хойн в 1900 году. Его формула

$$y(x_0 + h) \approx y_0 + h\Big(\frac{1}{4}K_1 + \frac{3}{4}K_3\Big), \qquad (28)$$

где

$$K_1 = f(x_0, y_0),$$
$$K_2 = f\Big(x_0 + \frac{1}{3}h, y_0 + \frac{1}{3}hK_1\Big),$$
$$K_3 = f\Big(x_0 + \frac{2}{3}h, y_0 + \frac{2}{3}hK_2\Big).$$

Второй является обобщением квадратурной формулы Симпсона[10] на случай дифференциальных уравнений и имеет вид

$$y(x_0 + h) \approx y_0 + h\Big(\frac{1}{6}K_1 + \frac{4}{6}K_2 + \frac{1}{6}K_3\Big), \tag{29}$$

где

$$K_1 = f(x_0, y_0),$$
$$K_2 = f\Big(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}hK_1\Big),$$
$$K_3 = f(x_0 + h, y_0 - hK_1 + 2hK_2).$$

У этого метода есть неприятная особенность. Наличие отрицательного коэффициента a_{31} и коэффициента a_{32}, превышающего 1, во-первых, сказывается на росте погрешности в связи с округлением, а во-вторых, в некоторых задачах из-за этого может потребоваться вычислить функцию f вне зоны ее определения по второму аргументу. Тем не менее, в подавляющем большинстве случаев этот метод работает не хуже, чем метод Хойна (28).

Таблицы Бутчера для рассмотренных методов имеют вид

0			
$\frac{1}{3}$	$\frac{1}{3}$		
$\frac{2}{3}$	0	$\frac{2}{3}$	
	$\frac{1}{4}$	0	$\frac{3}{4}$

Метод Хойна
порядка 3

0			
$\frac{1}{2}$	$\frac{1}{2}$		
1	-1	2	
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Метод «Симпсона»
порядка 3

[10] *То́мас Си́мпсон (англ. Thomas Simpson)* (1710–1761), английский математик, наиболее известный по выведенному им и названному в его честь правилу вычисления определенных интегралов, в котором подынтегральная кривая заменяется на приближающую ее параболу.

Методы четвертого порядка

Для построения методов четвертого порядка необходимо использовать четыре этапа. Таких методов так же, как и трехэтапных методов третьего порядка, существует бесконечно много. Самым широко известным и применяемым является так называемый *«классический» метод Рунге – Кутты* (по-английски его называют *"The" Runge–Kutta method*). Он также как и метод (29) основан на квадратурной формуле Симпсона, но сохраняет ее повышенную точность (для чего требуется четвертый этап).

$$y(x_0+h) \approx y_0 + h\Big(\frac{1}{6}K_1 + \frac{1}{3}K_2 + \frac{1}{3}K_3 + \frac{1}{6}K_4\Big), \tag{30}$$

$$\begin{aligned}
K_1 &= f(x_0, y_0),\\
K_2 &= f\Big(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}hK_1\Big),\\
K_3 &= f\Big(x_0 + \frac{1}{2}h, y_0 + \frac{1}{2}hK_2\Big),\\
K_4 &= f(x_0 + h, y_0 + hK_3).
\end{aligned}$$

Другим примером метода четвертого порядка может служить обобщение квадратурной формулы 3/8.

$$y(x_0+h) \approx y_0 + h\Big(\frac{1}{8}K_1 + \frac{3}{8}K_2 + \frac{3}{8}K_3 + \frac{1}{8}K_4\Big), \tag{31}$$

$$\begin{aligned}
K_1 &= f(x_0, y_0),\\
K_2 &= f\Big(x_0 + \frac{1}{3}h, y_0 + \frac{1}{3}hK_1\Big),\\
K_3 &= f\Big(x_0 + \frac{2}{3}h, y_0 - \frac{1}{3}hK_1 + hK_2\Big),\\
K_4 &= f(x_0 + h, y_0 + hK_1 - hK_2 + hK_3).
\end{aligned}$$

В отличие от ситуации с квадратурными формулами, количество вычислений функции f у метода (30) и у метода 3/8 совпадает. При этом, второй оказывается точнее в том смысле, что коэффициенты при главных членах погрешности у него меньше.

Их таблицы Бутчера

0				
$\frac{1}{2}$	$\frac{1}{2}$			
$\frac{1}{2}$	0	$\frac{2}{2}$		
1	0	0	1	
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

«Классический» метод Рунге — Кутты

0				
$\frac{1}{3}$	$\frac{1}{3}$			
$\frac{2}{3}$	$-\frac{1}{3}$	1		
1	1	-1	1	
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Правило 3/8

Бутчер показал, что явные методы пятого порядка должны иметь не меньше шести этапов, шестого порядка — не меньше семи, а седьмого порядка — не меньше девяти.

Сходимость явных одношаговых методов

Поскольку задачей применения численных методов решения начальной задачи является не совершение одного шага, а получение приближения к решению на всем интервале, необходимо изучить, при каких условиях приближенное решение стремится к точному и с какой скоростью. Свойство численного метода обеспечивать стремление приближения y_N к точному решению $y(x_N)$ в точке x_N при стремлении максимальной длины шага h к нулю называется *сходимостью метода*. Методы, не обладающие свойством сходимости, практически не пригодны.

Здесь легко провести аналогию с составными квадратурными формулами. В случае простого интегрирования мы тоже изучали как быстро стремилась к нулю методическая погрешность интегрирования при уменьшении максимального шага.

Для оценки сходимости изучаемых методов проведем анализ погрешностей, которые возникают в ходе решения дифференциального уравнения.

Погрешности решения задачи Коши

Рассмотрим погрешности, возникающие при решении задачи (1), (2).

Во-первых, начальное условие $y(x_0) = y_0$ может быть вычислено (задано) с некоторой погрешностью. В этом случае вместо задачи (1), (2) решается задача

$$\begin{cases} \bar{y}'(x) = f(x, \bar{y}(x)), \\ \bar{y}(x_0) = \bar{y}_0 \end{cases} \tag{32}$$

с измененным начальным условием

$$\bar{y}_0 = y_0 + R_0.$$

Решение $\bar{y}(x)$ задачи (32) не совпадает с решением исходной задачи (1), (2).

Определение 3. Разность

$$\xi_n = y(x_n) - \bar{y}(x_n)$$

называется *неустранимой погрешностью решения* $\bar{y}(x)$ (в точке x_n).

Во-вторых, применение численного метода (6) для нахождения решения в точке x_1 по формуле $\bar{y}_1 = F(f, x_1, x_0, \bar{y}_0)$ порождает *локальную методическую погрешность* первого шага (см. (14))

$$\varrho_1 = \Psi(x_1 - x_0) = \bar{y}(x_1) - \bar{y}_1.$$

И дополнительно к этому, вследствие ошибок округления и приближенного вычисления правой части $f(x, y)$ дифференциального уравнения вычисление значения $\bar{y}_1$ выполняется неточно. Фактически, найденное значение $\hat{y}_1$ удовлетворяет соотношению

$$F(f, x_1, x_0, \bar{y}_0) - \hat{y}_1 = \delta_1.$$

Определение 4. Невязка δ_1 называется *погрешностью округления.*

Определение 5. Разность $\eta_1 = \bar{y}_1 - \hat{y}_1$ называется *вычислительной погрешностью.*

Таким образом, после совершения первого шага, у нас есть приближение $\hat{y}_1$ к решению $\bar{y}(x_1)$, причем

$$\bar{y}(x_1) = \hat{y}_1 + \varrho_1 + \delta_1.$$

На втором шаге мы будем решать новую начальную задачу

$$\begin{cases} \bar{y}'_{(2)}(x) = f(x, \bar{y}_{(2)}(x)), \\ \bar{y}_{(2)}(x_1) = \hat{y}_1. \end{cases} \tag{33}$$

Приближение $\hat{y}_2$ к ее решению $\bar{y}_{(2)}(x_2)$ в точке x_2 также будет содержать методическую и вычислительную погрешности, т. е.

$$\bar{y}_{(2)}(x_2) = \hat{y}_2 + \varrho_2 + \delta_2.$$

Подобное накопление погрешностей будет происходить на каждом шаге. На n-м шаге мы решаем задачу

$$\begin{cases} \bar{y}'_{(n)}(x) = f(x, \bar{y}_{(n)}(x)), \\ \bar{y}_{(n)}(x_{n-1}) = \hat{y}_{n-1}. \end{cases} \tag{34}$$

Для универсальности (34) полагаем $\bar{y}_{(1)}(x) \equiv \bar{y}(x)$.

Определение 6. Разность между значением решения $\bar{y}(x_n)$ задачи (32) и его приближенным значением $\hat{y}_n$, вычисленным последовательным применением формулы $\hat{y}_i = F(f, x_i, x_{i-1}, \hat{y}_{i-1})$, $i = 1, ..., n$ одношагового метода,

$$\varepsilon_n = \bar{y}(x_n) - \hat{y}_n$$

называется *глобальной погрешностью метода* после n шагов. Если совершен всего один шаг, то глобальная погрешность ε_1 равна сумме локальной погрешности метода на первом шаге и вычислительной погрешности.

Замечание. Часто наличие вычислительной погрешности в процессе образования глобальной не учитывают. Для оценки методической погрешности и изучения сходимости метода вычислительную погрешность полагают равной нулю. В этом случае глобальная погрешность формируется только из суммы перенесенных в точку x_n локальных погрешностей всех предыдущих шагов.

Определение 7. Разность между точным решением $y(x_n)$ задачи (1), (2) и приближенным фактически найденным решением $\hat{y}_n$

$$R_n = y(x_n) - \hat{y}_n = \xi_n + \varepsilon_n \tag{35}$$

называется *полной погрешностью приближенного решения* после n шагов. Полная погрешность приближенного решения складывается из неустранимой погрешности, погрешности метода и вычислительной погрешности.

Мажорантная оценка полной погрешности

Рассмотрим, насколько большой может быть полная погрешность при некоторых известных значениях локальных погрешностей.

Из теории дифференциальных уравнений известно, что разность между точными решениями $y(x)$ и $z(x)$ дифференциального уравнения (1), удовлетворяющими начальным условиям $y(x_0) = y_0$ и $z(x_0) = z_0$ соответственно, ограничена:

$$\|y(x) - z(x)\| \leqslant \|y_0 - z_0\| \mathrm{e}^{L(x-x_0)}, \tag{36}$$

где L — константа Липшица по второму аргументу f (см. (5)).

С использованием (36) и учетом $m = 1$ (модуль вместо нормы)

$$|\xi_n| \leqslant |R_0| \mathrm{e}^{L(x_n - x_0)} \tag{37}$$

и

$$|\varepsilon_n| \leqslant \sum_{i=1}^{n} (|\varrho_i| + |\delta_i|) \mathrm{e}^{L(x_n - x_i)},$$

так как каждый шаг мы решаем новую начальную задачу, решение которой отличается от точного решения предыдущей задачи, но с каждым шагом оставшийся отрезок решения все меньше.

Теперь можно показать, что полная погрешность после n шагов

$$|R_n| \leqslant |R_0|\mathrm{e}^{L(x_n-x_0)} + \sum_{i=1}^{n}(|\varrho_i| + |\delta_i|)\mathrm{e}^{L(x_n-x_i)} \leqslant$$
$$\leqslant \mathrm{e}^{L(x_n-x_0)}\left(|R_0| + \sum_{i=1}^{n}(|\varrho_i| + |\delta_i|)\right). \tag{38}$$

Применяя метод порядка p, можем использовать оценку его локальной погрешности (13)

$$|\varrho_i| \leqslant C|x_i - x_{i-1}|^{p+1}, \quad i = 1, ..., n,$$

а ограничивая длину шага и определяя верхнюю границу вычислительной погрешности

$$h = \max_{i=1,..,n}|x_i - x_{i-1}|, \qquad \delta = \max_{i=1,..,n}|\delta_i|,$$

запишем

$$\sum_{i=1}^{n}|\varrho_i| \leqslant \sum_{i=1}^{n} C|x_i - x_{i-1}|h^p = C(x_n - x_0)h^p,$$

откуда

$$|R_n| \leqslant \mathrm{e}^{L(x_n-x_0)}(|R_0| + C(x_n - x_0)h^p + n\delta). \tag{39}$$

Замечание. Для систем обыкновенных дифференциальных уравнений мажорантная оценка полной погрешности будет иметь вид

$$\|R_n\| \leqslant \mathrm{e}^{mL(x_n-x_0)}(\|R_0\| + C(x_n - x_0)h^p + n\delta) \tag{40}$$

с заменами модуля на норму в определениях порядка и δ.

Из (39) следует, что приближенное решение задачи Коши, полученное с использованием одношагового метода порядка точности p сходится к точному решению задачи при $h \to 0$, если

$$|R_0| \to 0 \quad \text{и} \quad \delta \to 0. \tag{41}$$

Рассмотрим смысл каждого из трех слагаемых в (39).

Первое слагаемое — неустранимая погрешность, которая распространяется на все узлы. Ее вклад в полную погрешность метода ограничен формулой (37).

Второй член показывает вклад методической погрешности, возникающей из-за того, что находится не точное решение задачи, а приближение к нему по формуле одношагового метода. Глобальная методическая погрешность метода порядка p пропорциональна h^p.

Последнее слагаемое возникает за счет ошибок округления. Его величина не превосходит $\mathrm{e}^{L(x_n-x_0)}\delta/h$. Если значение δ ограничено снизу $0 < \delta_0 \leqslant \delta$, а в вычислительной практике так и бывает (в силу конечности представления чисел в машинной памяти), и при этом длина шага h слишком мала (а значит число шагов очень велико), то вычислительная погрешность может достигать больших значений.

Погрешность метода может быть сделана сколь угодно малой за счет уменьшения шага. Вычислительная же погрешность может быть снижена за счет увеличения разрядной сетки машины (но эти возможности ограничены). Неустранимую погрешность можно снизить только за счет более точного определения начальных условий.

Правильная организация вычислительного процесса — это баланс между всеми параметрами влияющими на погрешность: требуемой точностью решения задачи, точностью задания начальных условий, порядком численного метода, величиной шага интегрирования, используемой длиной разрядной сетки.

Отдельные составляющие, входящие в полную погрешность R_n могут давать отклонения от точного решения в разные стороны. Поэтом оценки (39) и (40) являются завышенными. На практике они не используются для определения точности окончательного результата.

Асимптотическая оценка погрешности метода

Рассмотрим как ведет себя глобальная прогрешность метода в отсутствие неустранимой и вычислительной прогрешностей. В этом разделе и далее мы опустим использование черты над y, поскольку более не требуется проводить различие между исходной задачей (1), (2) и задачей с измененным начальным условием (32). Аналогичным образом опустим «крышку», так как будем считать, что фактически вычисленные значения $\hat{y}_n$ совпадают с точными результатами применения метода, которые обозначим y_n.

Итак, пусть сделан $n-1$ шаг и получено приближение y_{n-1} в точке x_{n-1}. Решается начальная задача (34)

$$\begin{cases} y'_{(n)}(x) = f(x, y_{(n)}(x)), \\ y_{(n)}(x_{n-1}) = y_{n-1}. \end{cases} \tag{42}$$

Если правая часть $f(x,y)$ исходного дифференциального уравнения имеет непрерывные частные производные до порядка $p+2$ включительно, то для локальной методической погрешности одношагового метода порядка точности p на n-м шаге длиной $h = x_n - x_{n-1}$ справедливо представление

$$y_{(n)}(x_n) - y_n = \Phi(x_{n-1}, y_{n-1})h^{p+1} + O(h^{p+2}), \tag{43}$$

где

$$\Phi(x_{n-1}, y_{n-1}) = \frac{1}{(p+1)!} \left. \frac{d^{p+1}\Big(y_{(n)}(x_n) - F(f, x_n, x_{n-1}, y_{n-1})\Big)}{dh^{p+1}} \right|_{h=0}.$$

Асимптотическое представление глобальной погрешности после n шагов будет иметь вид

$$\varepsilon_n = \zeta(x_n)h^p + O(h^{p+1}), \tag{44}$$

где коэффициент главного члена погрешности

$$\zeta(x_n) = \int_{x_0}^{x_n} \Phi(\xi, y(\xi)) \exp\left(\int_{\xi}^{x_n} \frac{\partial f}{\partial y}(\tau, y(\tau)) d\tau \right) d\xi.$$

При достаточно малых h главный погрешности $\varepsilon_n \approx \zeta(x_n)h^p$. Это значение достаточно хорошо отражает и полную погрешность R_n из (35) в том случае, когда вклад неустранимой и вычислительной погрешностей, а также членов порядка $O(h^{p+1})$, входящих в (44), мал по сравнению с главным членом. Формально эти требования можно записать в виде

$$R_0 = O(h^{p+1}), \quad \delta = O(h^{p+2}). \tag{45}$$

Таким образом, если выполняются условия (45), при $h \to 0$ для полной погрешности приближенного решения справедливо асимптотическое разложение

$$R_n = \zeta(x_n)h^p + O(h^{p+1}) \approx \zeta(x_n)h^p. \tag{46}$$

Само собой, эта оценка также не является удобной для практического применения.

Практическая реализация ЯМРК

При реализации методов необходимо оценивать глобальную и локальную погрешности, с одной стороны, чтобы обеспечить длину шага h, достаточно малую для достижения требуемой точности вычисляемых результатов, а с другой — чтобы гарантировать достаточно большую длину шага во избежание бесполезной вычислительной работы.

Мы будем считать, что в вычислительном процессе неустранимой погрешностью и погрешностью округления можно пренебречь. Обратите внимание, что в этом разделе черта над y используется в другом смысле по сравнению с полным анализом погрешностей.

Метод Рунге оценки полной погрешности

Полагаем, что в точке x_n по s-этапному методу p-го порядка точности (12) с постоянным шагом h вычислено приближенное $\bar{y}_n$ исходной задачи Коши. С учетом (46) справедливо равенство

$$y(x_n) - \bar{y}_n = \zeta(x_n)h^p + O(h^{p+1}).$$

Используя ту же расчетную формулу с шагом $\frac{h}{2}$, вычислим в той же точке x_n другое значение решения $\tilde{y}_{2n}$ (мы совершим $2n$ шагов, потому такой индекс). При достаточном малом h глобальная погрешность приближенного решения в этом случае может быть представлена как

$$y(x_n) - \tilde{y}_{2n} = \zeta(x_n)\left(\frac{h}{2}\right)^p + O(h^{p+1})$$

с тем же коэффицентом $\zeta(x_n)$.

Решая эту систему, найдем

$$\bar{R}_n = y(x_n) - \bar{y}_n = \frac{\tilde{y}_{2n} - \bar{y}_n}{1 - 2^{-p}} + O(h^{p+1}), \tag{47}$$

$$\tilde{R}_{2n} = y(x_n) - \tilde{y}_{2n} = \frac{\tilde{y}_{2n} - \bar{y}_n}{2^p - 1} + O(h^{p+1}). \tag{48}$$

В качестве решения в точке x_n имеет смысл принять значение $\tilde{y}_{2n}$ как более точное по сравнению с $\bar{y}_i$. Его оценка погрешности (48),

однако можно дополнительно увеличить порядок точности приближения, если выразить его из (47) или (48)

$$y(x_n) = \bar{y}_n + \bar{R}_n = \tilde{y}_{2n} + \tilde{R}_{2n} = \tilde{y}_{2n} + \frac{\tilde{y}_{2n} - \bar{y}_n}{2^p - 1} + O(h^{p+1}). \qquad (49)$$

Как видно, погрешность этого приближения имеет порядок $O(h^{p+1})$.

При расчетах всегда задается какая-то граница допустимой погрешности $tol > 0$ (от англ. *tolerance* — допуск). Оценка погрешности может быть по абсолютной величине как больше, так и меньше допуска. В первом случае встает вопрос о выборе меньшей длины шага, дающей оценку погрешности, не превышающую tol, во втором — об увеличении длины шага до максимального значения, допускаемого границей точности. В обоих случаях можно воспользоваться формулой

$$h_{tol} \approx h \left(\frac{tol}{|\bar{R}_n|} \right)^{\frac{1}{p}} = \frac{h}{2} \left(\frac{tol}{|\tilde{R}_{2n}|} \right)^{\frac{1}{p}}. \qquad (50)$$

Метод Рунге для оценки локальной погрешности

Применение оценки глобальной погрешности на практике оказывается слишком трудоемким из-за необходимости пересчитывать решение с новым постоянным шагом на всем интервале $[x_0, x_f]$ в случае, если текущее приближение не удовлетворяет требованиям на точность. Кроме того, в подавляющем большинстве задач можно выделить интервалы, на которых решение ведет себя по разному и вклад шагов на которых в глобальную погрешность существенно различается. Так, локальная погрешность шагов на итервале, где решение ведет себя достаточно полóго (мала константа Липшица) и производные $p+1$-го порядка, входящие в локальную погрешность, также невелики, будет сравнительно мала и ее вклад в глобальную погрешность будет куда меньше, чем вклад ошибок на интервалах, где решение резко изменяется. Считая с постоянной длиной шага, мы будем вынуждены ориентироваться как раз на погрешности «крутых» шагов, и будем совершать много ненужных шагов в «пологой» части решения.

Альтернативой такому подходу является решение с переменным шагом. Чем «круче» решение, тем меньше мы будем делать шаг, так же, как бегун по пересеченной местности делает более мелкие шаги на резких спусках и крутых подъемах. Однако для этого нам

придется оценивать не глобальную погрешность, которую мы уже не сможем так легко приблизить, как в случае постоянного шага. Мы будем оценивать локальную погрешность и потребуем, чтобы на каждом шаге именно локальная погрешность не превосходила некоторого ограничения $tol > 0$. Сведений о глобальной погрешности и ее оценок в этом случае у нас нет, кроме мажорантной оценки (39) или (40) (на самом деле она может быть переписана под использование величины tol, а не максимальной длины шага h, но мы не будем приводить здесь эту формулу). На практике считается, что почти всегда реальная глобальная погрешность значительно меньше мажорантной оценки, и потому достаточно обеспечить на каждом шаге ограниченность локальной погрешности.

Каким же образом оценивается локальная погрешность (43)? Ее, точно так же как и глобальную, можно оценивать по правилу Рунге. Просто теперь вместо n шагов длиной h и $2n$ шагов длиной $\frac{h}{2}$ для исходной задачи (1), (2) мы совершаем один и два шага длинами $h_n = x_n - x_{n-1}$ и $\frac{h_n}{2}$ соответственно, решая задачу (42). Получаем два приближения в точке x_n: $\bar{y}_n$ с шагом h_n и $\bar{\bar{y}}_n$ после двух шагов $\frac{h_n}{2}$ (к тому же получаем еще приближение $\bar{y}_{n-\frac{1}{2}}$ в точке $x_{n-1} + \frac{h_n}{2}$ после первого половинного шага, но оно не используется в оценке погрешности). Правило Рунге принимает вид

$$\bar{\varrho}_n = y_{(n)}(x_n) - \bar{y}_n = \bar{r}_n + O(h_n^{p+1}), \quad \bar{r}_n = \frac{\bar{\bar{y}}_n - \bar{y}_n}{1 - 2^{-p}}, \tag{51}$$

$$\bar{\bar{\varrho}}_n = y_{(n)}(x_n) - \bar{\bar{y}}_n = \bar{\bar{r}}_n + O(h_n^{p+1}), \quad \bar{\bar{r}}_n = \frac{\bar{\bar{y}}_n - \bar{y}_n}{2^p - 1}, \tag{52}$$

где $\bar{r}_n$ и $\bar{\bar{r}}_n$ — оценки погрешностей $\bar{\varrho}_n$ и $\bar{\bar{\varrho}}_n$ соответственно.

Само собой, мы можем увеличить точность приближения аналогично тому, как мы сделали для глобальной погрешности (49):

$$y_{(n)}(x_n) \approx \bar{y}_n + \bar{r}_n = \bar{\bar{y}}_n + \bar{\bar{r}}_n = \bar{\bar{y}}_n + \frac{\bar{\bar{y}}_n - \bar{y}_n}{2^p - 1}. \tag{53}$$

Следует обратить внимание, что на каждом шаге для применения правила Рунге требуется вычислить $3s - 1$ значение правой части, что почти втрое сложнее простого совершения шага длиной h_n.

Упражнение. Выведите формулы (51)–(53) по-другому, пользуясь только асимптотическим разложением локальной погрешности (43) (*Указание: разложите локальную погрешность на втором половинном шаге в ряд в той же точке* (x_{n-1}, y_{n-1})*, что и две других*).

Автоматический выбор шага интегрирования

Теперь, пользуясь оценкой локальной погрешности (51) или (52) мы можем составить алгоритм выбора длины нового шага (или того же самого, если его предыдущая длина была слишком большой), которая будет обеспечивать требуемую величину локальной погрешности. При этом следует учитывать, что в силу приближенного характера всех оценок слишком большое изменение длины шага в любую сторону может привести к качественному изменению поведения решения, поэтому не рекомендуется сильно увеличивать или уменьшать ее.

Правило Рунге, как отмечалось выше, является весьма дорогостоящим: требуется $3s-1$ этап. На стр. 26 приведем алгоритм, который наиболее естественно связан с этим способом оценки локальной погрешности и использует все получаемые на шаге приближения $\bar{y}_n$, $\bar{\bar{y}}_n$ и $\bar{y}_{n-\frac{1}{2}}$. В случае 1 алгоритма, когда шаг не принимается, его пересчет требует только $2s-1$ новое вычисление функции $f(x,y)$. Отметим, что чаще всего, исходя из соображений точности или свойств решаемой начальной задачи, известна некоторая максимальная допустимая длина шага h_{max}.

Помимо правила Рунге существуют другие способы оценки локальной погрешности, не использующие вычисления на половинном шаге. Для них половинное деление и удвоение шага неестественно. К тому же изменение длины шага всего в два раза не всегда позволяет сразу же подойти близко к границе допустимой точности. Потому выбор новой длины шага логичнее делать по формуле, аналогичной формуле (50). Однако рассмотрение этих вопросов выходит за рамки настоящего методического пособия.

Упражнение. Модифицируйте алгоритм управления длиной шага на стр. 26, используя сравнение оценки локальной погрешности более точного приближения $\bar{\bar{r}}_n$ с допуском tol.
Каким образом можно повысить точность вычислений, без увеличения затрат?

Алгоритм удвоения и деления шага пополам

Задана локальная точность $tol > 0$. Используется четыре варианта принятия решения в зависимости от величины оценки локальной погрешности $\bar{r}_n$, получаемой на n-м шаге длиной h_n.

1. $\|\bar{r}_n\| > tol \cdot 2^p$

   Ни $\bar{y}_n$, ни $\bar{\bar{y}}_n$ не обеспечивают необходимую точность. Нужно повторить n-й шаг еще раз с меньшей длиной. Выбираем

   $$h_n := \frac{h_n}{2}, \quad \bar{y}_n := \bar{y}_{n-\frac{1}{2}}$$

   и пересчитываем значения $\bar{y}_{n-\frac{1}{2}}$ и $\bar{\bar{y}}_n$ с новым шагом.

2. $tol < \|\bar{r}_n\| \leqslant tol \cdot 2^p$

   Приближение $\bar{\bar{y}}_n$ дает нужную точность. Можем принять результаты сделанного шага, однако велика вероятность, что новый шаг той же длины приведет к слишком большой величине погрешности, поэтому принимаем

   $$h_{n+1} := \frac{h_n}{2}, \quad y_n := \bar{\bar{y}}_n$$

   и переходим к новому шагу.

3. $tol \cdot \frac{1}{2^{p+1}} \leqslant \|\bar{r}_n\| \leqslant tol$

   Приближение $\bar{y}_n$ тоже дает нужную точность. Принимаем сделанный шаг и оставляем длину без изменений. Полагаем

   $$h_{n+1} := h_n, \quad y_n := \bar{y}_n$$

   и переходим к новому шагу.

4. $\|\bar{r}_n\| < tol \cdot \frac{1}{2^{p+1}}$

   Приближение $\bar{y}_n$ оказывается слишком точным, так что даже двойной шаг был бы достаточно точен. Выбираем

   $$h_{n+1} := \min(2h_n, h_{max}), \quad y_n := \bar{y}_n$$

   и переходим к новому шагу.

Алгоритм выбора начального шага

Алгоритм автоматического управления шагом может начать работу с любой начальной длиной шага. Однако, если она выбрана произвольным образом, может потребоваться слишком много отброшенных вычислений, пока она будет уменьшена, или будет совершено слишком много маленьких шагов, пока длина не будет увеличена до подходящего по точности значения. Поэтому имеет смысл автоматизировать выбор начального шага.

1. Вычислить $f(x_0, y_0)$;
2. Найти
$$\Delta = \left(\frac{1}{\max\left(|x_0|, |x_k|\right)}\right)^{p+1} + |f(x_0, y_0)|^{p+1};$$
3. Выбрать начальный шаг
$$h_1 = \left(\frac{tol}{\Delta}\right)^{\frac{1}{p+1}}.$$

Часто начальные условия находятся в особом положении, где большинство компонент вектора $f(x_0, y_0)$ — нули. В этом случае необходимо добавить еще два шага:

4. Сделать один шаг методом Эйлера с длиной шага h_1, полученной в п. 3, получив приближение u_1 в точке $x_0 + h_1$;
5. Повторить шаги 1–3 алгоритма, взяв точку $(x_0 + h_1, u_1)$ вместо (x_0, y_0), получить приближение h_1' и положить
$$h_1 := \min(h_1, h_1').$$

Использование различных характеристик точности

Чаще в алгоритмах выбора шага используют не абсолютную погрешность ϱ, а относительную — $\varrho/|y|$, так как более значимым является получение нужного числа верных цифр всех компонентов решения,

насколько бы они ни различались по абсолютной величине. Однако в тех случаях, когда абсолютные значения становятся очень малы или, тем более, обращаются в нуль, все же проверяется малость абсолютной погрешности.

Обычно задают различные допуски на относительную и абсолютную погрешности. Обозначим их *rtol* и *atol* (от *relative* и *absolute*) соответственно. Можно удобно объединить в одной формуле проверку относительной и абсолютной погрешностей и одновременно избежать деления на малую величины (если $|y|$ мало). Считаем, что точность удовлетворена, если

$$|\varrho| \leqslant rtol \cdot |y| + atol. \tag{54}$$

Поскольку на практике *atol* на порядки меньше чем *rtol*, то в случае, когда значение $|y|$ достаточно велико, величина правой части определяется первым слагаемым и мы проверяем малость относительной погрешности. Если же $|y|$ достаточно мало, то начинает доминировать *atol* и проверяется абсолютная погрешность.

Например, пусть мы работаем с точностью *double*, которая позволяет хранить примерно шестнадцать десятичных цифр в записи числа. Пусть $rtol = 10^{-6}$, то есть мы требуем, чтобы ответ содержал шесть верных десятичных знаков, но было бы наивно надеяться, что можно получить шесть верных цифр, если сама величина близка к нулю. Потому, зададим $atol = 10^{-12}$ и тем самым будем проверять только те цифры, которые соответствуют разрядам бóльшим чем 10^{-12}.

Так если

$$y = 0.033\,166\,247\,903\,554,$$

то

$$rtol \cdot |y| + atol = 0.000\,000\,033\,167\,248.$$

Таким образом при проверке погрешность должна быть порядка 10^{-8}, что дает шесть верных цифр в y.

Если же

$$y = 0.000\,000\,003\,316\,625,$$

то

$$rtol \cdot |y| + atol = 0.000\,000\,000\,001\,0033$$

и погрешность должна быть порядка 10^{-12}, что означает лишь три верных десятичных знака для y. Если же $|y| < atol$, то мы вообще

не проверяем его погрешность, считая его равным вычислительному нулю.

Для систем ОДУ часто приходится накладывать разные ограничения на погрешности для разных компонент решения, например, если одна часть неизвестных соответствует координатам объекта, а другая — скоростям их изменения, или одна — зарядам на узлах электрической цепи, а другая — токам через эти узлы. Характерные величины разных по физической природе искомых функций могут сильно различаться, и добиваться одинаковой точности (абсолютной или относительной) становится нецелесообразно. Весь вектор решения разделяется тогда на подвекторы w_i, для каждого из которых погрешность σ_i сравнивается со своими собственными $rtol_i$ и $atol_i$ (в предельном случае все подвекторы являются отдельными компонентами).

Может быть и так, что для некоторых из подвекторов решения мы вообще не ведем контроль погрешности. Для всех остальных w^i проверяем малость нормы погрешности

$$\|\sigma_i\| \leqslant rtol_i \cdot \|w_i\| + atol_i. \tag{55}$$

Часто используются нормы

$$\begin{aligned} \|\sigma\|_\infty &= \max_{1\leqslant i\leqslant m} |\sigma^i| \quad \text{(норма-максимум)}, \\ \|\sigma\|_1 &= \sum_{i=1}^{m} |\sigma^i|, \\ \|\sigma\|_2 &= \left(\sum_{i=1}^{n} (\sigma^i)^2\right)^{1/2} \quad \text{(евклидова норма)}. \end{aligned} \tag{56}$$

Качество алгоритма

> Бросая в воду камешки, смотри на круги, ими образуемые; иначе такое бросание будет пустою забавою.
>
> Козьма Прутков

Рассмотрим в этом разделе некоторые критерии качества алгоритма решения задачи Коши: надежность, точность и объем вычислений.

Надежность. При практической реализации методов стоит задача регулировать величину локальной погрешности посредством измене-

ния длины шага. Метод можно считать надежным, если он удачно с этим справляется, то есть достаточно быстро адаптирует длину шага и сохраняет погрешность меньшей допуска, но близкой к нему. Для оценки надежности решают задачу с известным аналитическим решением и на каждом шаге вычисляют отношение истинной локальной погрешности на шаге ϱ_n к ее оценке r_n

$$\eta_n = \left| \frac{\varrho_n}{r_n} \right|. \tag{57}$$

Способ оценки погрешности и соответствующее управление длиной шага основываются на приближении выражения порядка $O(h^{p+1})$, которое можно представить как $Ch^{p+1} + O(h^{p+2})$. Они выполняются достаточно хорошо лишь тогда, когда при изменении длины шага коэффициент C при главном члене этого выражения остается почти неизменным, что обеспечивается в некотором интеревале изменения длины шага. Если при совершении нового шага η_n мало меняется, значит погрешность оценивается почти так же точно, как и на предыдущем шаге, и мы можем так же управлять длиной шага.

Если же происходит резкое изменение уровня η_n, это значит, что либо поведение самого решения сильно изменилось и теперь хорошая оценка погрешности возможна в другом интервале изменения длины шага, либо (если известно, что в тестовой задаче поведение решения не изменялось) плоха сама практическая реализация метода.

Точность. Проверка точности метода проводится путем анализа поведения глобальной погрешности при различных допусках tol на локальную погрешность. Нас интересует,

- непрерывно ли уменьшается полная погрешность в конечной точке x_f (или нескольких точках внутри интервала решения) при уменьшении задаваемой допустимой погрешности tol, и
- правильно ли работает алгоритм, когда требуется высокая точность?

Объем вычислений. Для одношаговых методов такой характеристикой служит количество вычислений правой части $\Theta(tol)$ системы ОДУ на интервале интегрирования $[t_0, t_f]$ для каждой заданной точности tol. Само собой, эта величина зависит не только от числа этапов метода, но и от его порядка точности (чем он выше, тем с большим шагом можно идти), и от надежности управления шагом (чем меньше отброшенных шагов, тем лучше).

Недостатки явных методов Рунге — Кутты

- Зависимость порядка точности метода от количества вычислений правой части,
- Оценка локальной погрешности требует дополнительных вычислительных затрат, сравнимых по объему с вычислительными затратами на получение приближенного решения.

Задание для самостоятельной работы

1. Используя условия порядка для двухэтапного ЯМРК второго порядка, постройте расчетную схему второго порядка при значении параметра c_2, указанном в варианте.
2. В варианте задания найдите метод-оппонент, с которым будет проводиться сравнение построенного метода.
3. Для обоих методов постройте и программно реализуйте алгоритм решения задачи Коши

$$\begin{cases} y_1'(x) = 2x\big(y_2(x)\big)^{\frac{1}{B}} y_4(x), \\ y_2'(x) = 2Bx \exp\left(\dfrac{B}{C}(y_3(x) - A)\right) y_4(x), \\ y_3'(x) = 2Cx y_4(x), \\ y_4'(x) = -2x \ln y_1(x), \end{cases} \tag{58}$$

с начальным условием $y_1(0) = y_2(0) = y_4(0) = 1$, $y_3(0) = A$, на отрезке $x = [0, 5]$.

Точное решение задачи имеет вид

$$y_1(x) = \mathrm{e}^{\sin x^2}, \qquad y_3(x) = C \sin x^2 + A,$$
$$y_2(x) = \mathrm{e}^{B \sin x^2}, \qquad y_4(x) = \cos x^2.$$

Решение проводите с постоянным шагом.

- Постройте график зависимости нормы точной полной погрешности в конце отрезка от длины шага в двойной логарифмической шкале для длин шагов $1/2^k$, $k = 0, ..., 6$. Постройте на том же графике прямую с наклоном два

(и с наклоном, равным порядку метода-оппонента, если он отличен от двух). Убедитесь, что реализованный метод показывает сходимость правильного порядка.

- Оценивая полную погрешность по правилу Рунге, найдите длину оптимального шага h_{opt}, обеспечивающего погрешность не превосходящую $tol = 10^{-5}$. Постройте график зависимости нормы точной полной погрешности от независимой переменной при решении с h_{opt}.

4. Для обоих методов реализуйте алгоритм с автоматическим выбором длины шага, основанный на оценке локальной погрешности по правилу Рунге. Проверьте надежность двух алгоритмов, их экономичность.

   - Используйте значения $rtol = 10^{-6}$ и $atol = 10^{-12}$. Постройте график решения.
   - Постройте график зависимости длины шага от независимой переменной. На графике отметьте разными символами отброшенные и принятые шаги.
   - Для того же решения постройте график зависимости нормы точной полной погрешности от независимой переменной.
   - Проведите расчеты для значений $rtol = 10^{-4}$, 10^{-5}, 10^{-6}, 10^{-7} и 10^{-8} ($atol = 10^{-12}$ во всех случаях). Постройте зависимость числа обращений к правой части от $rtol$ в двойной логарифмической шкале.

Варианты

1. $c_2 = 0.05$, $A = 3$, $B = 3$, $C = -3$, оппонент — метод (26).

2. $c_2 = 0.10$, $A = -3$, $B = 2$, $C = 1$, оппонент — метод (27).

3. $c_2 = 0.15$, $A = -2$, $B = -2$, $C = 2$, оппонент — метод (28).

4. $c_2 = 0.20$, $A = 2$, $B = -1$, $C = -1$, оппонент — метод (29).

5. $c_2 = 0.25$, $A = 1$, $B = 1.5$, $C = -2$, оппонент — метод (30).

6. $c_2 = 0.30$, $A = -1$, $B = -3$, $C = 3$, оппонент — метод (31).

7. $c_2 = 0.35$, $A = 3$, $B = -3$, $C = 3$, оппонент — метод (26).

8. $c_2 = 0.40$, $A = -3$, $B = 2.5$, $C = 1$, оппонент — метод (27).

9. $c_2 = 0.45$, $A = -2$, $B = 2$, $C = -2$, оппонент — метод (28).

10. $c_2 = 0.55$, $A = 2$, $B = -1$, $C = 2$, оппонент — метод (29).

11. $c_2 = 0.60$, $A = 1$, $B = 3$, $C = 3$, оппонент — метод (30).

12. $c_2 = 0.65$, $A = -1$, $B = -2$, $C = -1$, оппонент — метод (31).

13. $c_2 = 0.70$, $A = 3$, $B = 1.5$, $C = -1$, оппонент — метод (26).

14. $c_2 = 0.75$, $A = -3$, $B = -2$, $C = -2$, оппонент — метод (27).

15. $c_2 = 0.80$, $A = -2$, $B = 3$, $C = -1$, оппонент — метод (28).

16. $c_2 = 0.85$, $A = 2$, $B = -2$, $C = 3$, оппонент — метод (29).

17. $c_2 = 0.90$, $A = 1$, $B = -3$, $C = 2$, оппонент — метод (30).

18. $c_2 = 0.95$, $A = -1$, $B = -1$, $C = -3$, оппонент — метод (31).

19. $c_2 = \frac{\sqrt{5}}{2} - \frac{1}{2}$, $A = 3$, $B = -1$, $C = -2$, оппонент — метод (28).

20. $c_2 = \frac{3}{2} - \frac{\sqrt{5}}{2}$, $A = -3$, $B = 2.5$, $C = 2$, оппонент — метод (30).

Литература

1. Хайрер Э., Нерсетт С. П., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи — М.: Мир, 1990. 512 с.

2. Вержбицкий В. М. Основы численных методов: учебник для вузов — М.: Директ-Медиа, 2013. 847 с.

3. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы, 8-е изд. — М.: БИНОМ. Лаборатория знаний, 2015. 639 с.

Биографические сведения взяты из Википедии (ru.wikipedia.org).