

Версия 0.4 beta  
26 апреля 2022

# Введение в высокоточные методы численного решения систем законов сохранения.

Часть 1:  
общие вопросы  
анализа линейных схем

## Содержание

|   |                                                                      |    |
|---|----------------------------------------------------------------------|----|
| 1 | Предисловие . . . . .                                                | 3  |
| 2 | Введение . . . . .                                                   | 4  |
|   | 2.1 Некоторые сведения из алгебры и анализа . . . . .                | 4  |
|   | 2.2 Точное и численное решение . . . . .                             | 6  |
|   | 2.3 Сеточная сходимость . . . . .                                    | 10 |
| 3 | Полностью дискретные схемы . . . . .                                 | 13 |
|   | 3.1 Схема “левый уголок” . . . . .                                   | 13 |
|   | 3.2 Аппроксимация, устойчивость, сходимость . . . . .                | 16 |
|   | 3.3 Спектральный анализ устойчивости . . . . .                       | 19 |
|   | 3.4 Метод характеристик . . . . .                                    | 28 |
|   | 3.5 Устойчивость в $\  \cdot \ _2$ на неравномерных сетках . . . . . | 29 |
|   | 3.6 Точность на неравномерных сетках . . . . .                       | 31 |
| 4 | Полудискретные схемы . . . . .                                       | 34 |
|   | 4.1 Метод линий . . . . .                                            | 34 |
|   | 4.2 Аппроксимация и устойчивость полудискретной схемы . . . . .      | 36 |
|   | 4.3 Спектральный анализ устойчивости . . . . .                       | 39 |
|   | 4.4 Решение ОДУ . . . . .                                            | 41 |
|   | 4.5 Устойчивость полностью дискретной схемы . . . . .                | 43 |

## 1. Предисловие

Настоящая брошюра является первой частью курса “Введение в высокоточные методы численного решения законов сохранения”. Как и в большинстве подобных курсов, изложение строится от простого к сложному:

- от линейных уравнений к нелинейным;
- от одного дифференциального уравнения к системам уравнений;
- от гладких решений к разрывным;
- от одномерных задач к многомерным;
- от равномерных сеток к неравномерным и неструктурированным;
- от базовых методов к методам высокой точности.

Чем сложнее решаемая задача и используемые численные методы, тем хуже они поддаются теоретическому анализу. В конечном счёте, для многомерных задач газовой динамики с разрывами нет ни одного численного метода, для которого была бы доказана сходимости. Но даже частичное знание о свойствах численных методов часто позволяет предугадать, как поведут себя эти методы в той или иной задаче, и предложить решение проблемы, если используемый метод по той или иной причине даёт неудовлетворительный результат.

Цель всего курса – пройти путь от схемы “уголок” для решения линейного уравнения переноса до современных методов высокой точности для решения уравнений газовой динамики в трёхмерном пространстве. В этой части мы будем рассматривать неравномерные расчётные сетки и методы высокой точности, но ограничимся рассмотрением задачи Коши для одномерного линейного уравнения переноса

$$\frac{\partial u}{\partial t} + \mu \frac{\partial u}{\partial x} = 0, \quad x_{\min} < x < x_{\max}, \quad 0 < t < t_{\max}, \quad (1.1)$$

с постоянным коэффициентом  $\mu \in \mathbb{R}$  и начальными данными  $u(0, x) = u_0(x)$ . Обычно будем использовать периодические граничные условия  $u(t, x_{\min}) = u(t, x_{\max})$ . Решением этой задачи является функция  $u(t, x) = \tilde{u}_0(x - \mu t)$ , где  $\tilde{u}_0$  – периодическое продолжение функции  $u_0$  на  $x \in \mathbb{R}$ . Будем предполагать, что это решение достаточно гладкое; класс гладкости будем уточнять в каждом конкретном случае.

Поскольку решение рассматриваемой задачи нам известно, задача разработки численного метода для её решения бессмысленна. Наша задача – исследовать поведение на этой задаче тех численных методов, которые используются для более сложных систем уравнений.

## 2. Введение

**2.1. Некоторые сведения из алгебры и анализа.** Будем использовать следующие стандартные обозначения:  $\mathbb{N}$  – множество натуральных чисел,  $\mathbb{Z}$  – множество целых чисел,  $\mathbb{R}$  – множество действительных чисел,  $\mathbb{C}$  – множество комплексных чисел.

Напомним определения линейного нормированного пространства.

*Абелевой группой* называется множество  $X$ , для элементов которого определены операции сложения, удовлетворяющие следующим аксиомам:

- для всех  $x, y, z \in X$  выполняется  $x + (y + z) = (x + y) + z$ ;
- для всех  $x, y \in X$  выполняется  $x + y = y + x$ ;
- существует  $0 \in X$ , такой что для каждого  $x \in X$  выполняется  $x + 0 = x$ ;
- для каждого  $x \in X$  существует  $-x \in X$ , такой что  $x + (-x) = 0$ .

*Действительным линейным пространством* называется множество  $X$ , для элементов которого определены операции сложения и умножения на действительное число, удовлетворяющие следующим аксиомам:

- $X$  является абелевой группой относительно сложения;
- для всех  $a, b \in \mathbb{R}$  и всех  $x \in X$  выполняется  $a(bx) = (ab)x$ ;
- для всех  $x \in X$  выполняется  $1x = x$ ;
- для всех  $x, y \in X$  и  $a \in \mathbb{R}$  выполняется  $a(x + y) = ax + ay$ ;
- для всех  $x \in X$  и  $a, b \in \mathbb{R}$  выполняется  $(a + b)x = ax + bx$ .

Определение комплексного линейного пространства аналогично.

Элементы линейного пространства называются *линейно зависимыми*, если существует их нетривиальная линейная комбинация, равная нулю. Максимальное число линейно независимых элементов называется *размерностью* пространства; если для любого  $n \in \mathbb{N}$  существует линейно независимая система из  $n$  элементов, такое пространство называется бесконечномерным.

*Нормой* на линейном пространстве  $X$  называется неотрицательная функция его элементов, удовлетворяющая следующим условиям:

- $\|x\| = 0$  равносильно  $x = 0$ ;
- для любого  $x \in X$  и любого действительного или комплексного (в случае комплексного линейного пространства) числа  $a$  выполняется  $\|ax\| = |a| \|x\|$ ;
- для любых  $x, y \in X$  выполняется неравенство треугольника:

$$\|x + y\| \leq \|x\| + \|y\|.$$

Линейное пространство с определённой на нём нормой называется *нормированным пространством*.

Пусть  $X$  – линейное пространство,  $\|\cdot\|_1$  и  $\|\cdot\|_2$  – некоторые нормы на нём. Эти нормы называются *эквивалентными*, если существуют константы  $c_1, c_2 > 0$ , такие что для всех  $x \in X$  выполняется  $c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$ .

**Теорема 2.1.** *В конечномерном пространстве любые две нормы являются эквивалентными.*

Пусть  $X$  и  $Y$  – линейные нормированные пространства. Тогда через  $\mathcal{L}(X, Y)$  обозначается пространство линейных отображений из  $X$  в  $Y$  с нормой

$$\|L\| = \sup_{x \neq 0} \frac{\|Lx\|_Y}{\|x\|_X}. \quad (2.1)$$

*Скалярным произведением* на комплексном линейном пространстве  $X$  называется функционал, сопоставляющий двум элементам  $x \in X$  и  $y \in X$  число  $\langle x, y \rangle \in \mathbb{C}$ , такой что выполняются следующие свойства:

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$  для всех  $x, y \in X$ ;
- $\langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle$  для всех  $x_1, x_2, y \in X$  и  $\alpha_1, \alpha_2 \in \mathbb{C}$ ;
- $\langle x, x \rangle > 0$  для всех  $x \in X, x \neq 0$ .

Скалярное произведение на действительном линейном пространстве определяется аналогично с тем отличием, что  $\langle x, y \rangle \in \mathbb{R}$ .

Линейное пространство с определённым на нём скалярным произведением называется *пространством со скалярным произведением* или *предгильбертовым пространством*. Если не оговорено обратное, на предгильбертовом пространстве подразумевается использование нормы

$$\|x\| = (\langle x, x \rangle)^{1/2}.$$

Пусть  $X$  – предгильбертово пространство. Тогда для любых  $x, y \in X$  справедливо *неравенство Коши – Буняковского* (также называемое *неравенством Шварца* или *неравенством Коши – Шварца*)

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \quad (2.2)$$

Его частными случаями являются неравенства

$$\left| \sum_{j=1}^N x_j y_j \right| \leq \left( \sum_{j=1}^N |x_j|^2 \right)^{1/2} \left( \sum_{j=1}^N |y_j|^2 \right)^{1/2}, \quad (2.3)$$

верное для любых  $x_j, y_j \in \mathbb{C}$ , и

$$\left| \int_a^b f(x) g(x) dx \right| \leq \left( \int_a^b |f(x)|^2 dx \right)^{1/2} \left( \int_a^b |g(x)|^2 dx \right)^{1/2}, \quad (2.4)$$

верное, в частности, для любых непрерывных функций  $f$  и  $g$ .

Числовая (действительно- или комплекснозначная) функция  $f$ , определённая на отрезке  $[a, b]$ , называется *непрерывной по Липшицу* (или просто *липшицевой*), если существует  $L > 0$ , такое что для любых  $x, y \in [a, b]$  выполняется

$$|f(x) - f(y)| \leq L|x - y|.$$

При этом  $L$  называется *константой Липшица* функции  $f$ . Для липшицевых функций справедлив частный случай теоремы Лебега.

**Теорема 2.2.** Пусть функция  $f$  липшицева на  $[a, b] \subset \mathbb{R}$ . Тогда функция  $f$  дифференцируема почти всюду на  $(a, b)$  (то есть мера Лебега множества точек этого интервала, на которой она не дифференцируема, равна нулю), её производная является измеримой по Лебегу и для всех  $x \in (a, b)$  справедливо равенство

$$f(x) = f(a) + \int_a^x f'(x) dx.$$

**2.2. Точное и численное решение.** Численная схема – это конечный набор арифметических операций, поэтому результатом её применения является конечный набор чисел (численное решение). С другой стороны, целью численной схемы является получение решения, близкого к искомой функции, т. е. к решению дифференциальной задачи. Как определить близость между двумя совершенно разными объектами? Есть два пути. Первый – отобразить искомую функцию на конечномерное пространство и рассмотреть разность между её образом и численным решением. Второй – отобразить численное решение на функциональное пространство и рассмотреть разность между образом и искомой функцией.

Чтобы упростить изложение, рассмотрим вначале краевую задачу для обыкновенного дифференциального уравнения:

$$u''(x) - u(x) = f(x), \quad x_{\min} < x < x_{\max};$$

$$u(x_{\min}) = u(x_{\max}), \quad u'(x_{\min}) = u'(x_{\max}),$$

где  $f$  – функция, непрерывная на  $[x_{\min}, x_{\max}]$ , такая что  $f(x_{\min}) = f(x_{\max})$ . Здесь штрихом обозначена производная по  $x$ .

В настоящем курсе будем называть *расчётной сеткой*<sup>1</sup> невырожденное разбиение отрезка  $[x_{\min}, x_{\max}]$ , то есть набор чисел  $\{x_j, j = 0, \dots, N_x\}$ , такой

---

<sup>1</sup> Данное определение не является универсальным. Множество когда-либо использованных технических приёмов настолько разнообразно, что подведение их под общую классификацию и единую систему понятий едва ли возможно. Мы будем использовать терминологию, удобную для изложения.

что

$$x_{\min} = x_0 < x_1 < \dots < x_{N_x} = x_{\max}.$$

Числа  $x_j$  будем называть *узлами* сетки. Если величины  $x_{j+1} - x_j$  одинаковы при всех  $j = 0, \dots, N_x - 1$ , такую сетку будем называть *равномерной*. Доопределим координаты сеточных узлов на  $j \in \mathbb{Z}$  равенством

$$x_{j+kN_x} = x_j + k(x_{\max} - x_{\min}), \quad \forall k \in \mathbb{Z}.$$

В конечно-разностных методах<sup>2</sup> под сеточной функцией обычно понимается набор, содержащий по одному значению на сеточный узел. При использовании периодических условий значения сеточной функции в граничных узлах должны совпадать. Поэтому удобно под сеточной функцией понимать либо набор значений в узлах  $j = 0, \dots, N_x - 1$ , либо  $N_x$ -периодическую последовательность чисел. В обоих случаях сеточные функции образуют линейное пространство размерности  $N_x$ . Пусть  $u$  – некоторая непрерывная функция на  $[x_{\min}, x_{\max}]$ , значения которой на концах совпадают. Сопоставим ей сеточную функцию

$$\Pi_h u = \{u_j, j = 0, \dots, N_x - 1\}, \quad u_j = u(x_j). \quad (2.5)$$

Другими словами, сеточный образ непрерывной функции – это набор её точечных значений в сеточных узлах. Если на этой сетке задано некоторое численное решение  $v_h = \{v_j\}$ , то близость между  $u$  и  $v_h$  можно определить как норму разности между  $\Pi_h u$  и  $v_h$ . Например,

$$\|\Pi_h u - v_h\|_{\infty} = \max_{j=0, \dots, N_x-1} |u(x_j) - v_j| \quad (2.6)$$

или

$$\|\Pi_h u - v_h\|_2 = \left( \sum_{j=0}^{N_x-1} \tilde{h}_j |u(x_j) - v_j|^2 \right)^{1/2}, \quad (2.7)$$

где

$$\tilde{h}_j = (x_{j+1} - x_{j-1})/2.$$

В конечно-объёмных схемах под сеточной функцией обычно понимается набор, содержащий по одному значению на сеточный элемент. В одномерном случае сеточным элементом является отрезок, соединяющий соседние сеточные узлы. Таким образом, сеточные функции образуют линейное пространство

---

<sup>2</sup>Читатель наверняка слышал слова “конечно-разностный”, “конечно-объёмный” и “конечно-элементный”. Они описывают не свойства схемы, а подходы к их построению. Одна и та же схема может быть получена разными подходами. Также отметим, что под численным методом может пониматься как численная схема, так и способ её построения, не обязательно однозначно её определяющий.

размерности  $N_x$ . Интегрируемой функции  $u$  сопоставляется сеточная функция

$$\Pi_h u = \{u_j, j = 0, \dots, N_x - 1\}, \quad u_j = \frac{1}{x_{j+1} - x_j} \int_{x_j}^{x_{j+1}} u(x) dx.$$

Элементы сеточной функции при желании можно нумеровать полуцелыми индексами. Близость численного решения к решению дифференциальной задачи оценивается аналогично конечно-разностному случаю.

В настоящем курсе конечно-объёмные схемы будут рассматриваться на *дуальной сетке*  $\{x_{j+1/2}, j \in \mathbb{Z}\}$ , где

$$x_{j+1/2} = \frac{x_j + x_{j+1}}{2}, \quad j = 0, \dots, N_x - 1.$$

Интервалы  $(x_{j-1/2}, x_{j+1/2})$  называются *контрольными объёмами* узлов, а интегрируемой функции  $u$  сопоставляется сеточная функция, состоящая из интегральных средних от  $u$  по контрольным объёмам:

$$\Pi_h u = \{u_j, j = 0, \dots, N_x - 1\}, \quad u_j = \frac{1}{h_j} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x) dx.$$

Принципиальной разницы с предыдущим случаем нет.

В *стандартном методе Галёркина* (standard Galerkin method, SG) под сеточной функцией понимается функция, принадлежащая некоторому конечномерному подпространству непрерывных функций. Например, пространству непрерывных функций, на каждом сеточном элементе являющихся многочленом порядка  $p$  (число  $p \in \mathbb{N}$  является параметром метода). В узлах сетки такие функции могут иметь разрыв производной; пример приведён на рис. 1. Говоря о многочлене порядка  $p$ , всюду будем подразумевать, что коэффициент при старшей степени может быть нулевым. Многочлен порядка  $p$  однозначно определяется своими значениями в  $p + 1$  точках. Две такие точки удобно выбрать в концах сеточного элемента, тогда условие непрерывности сведётся к тому, чтобы в каждом из сеточных узлов совпадали значения двух многочленов, относящихся к смежным с этим узлом элементам. Если считать, что значения сеточной функции при  $x = x_{\min}$  и  $x = x_{\max}$  должны совпадать, то пространство сеточных функций имеет размерность  $pN_x$ .

Поскольку численное решение является непрерывной функцией, чтобы определить близость численного решения  $v_h$  к решению дифференциальной задачи, нам уже не нужно отображение: мы можем положить

$$\|u - v_h\|_{\infty} = \max_{x \in [x_{\min}, x_{\max}]} |u(x) - v_h(x)| \quad (2.8)$$

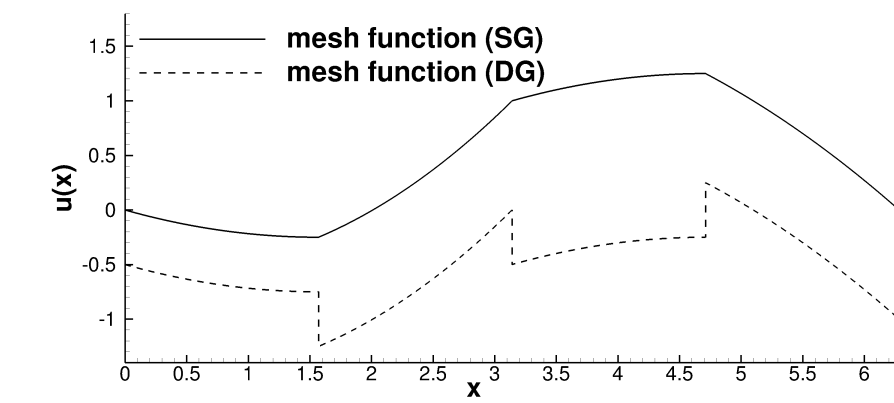


Рис. 1. Примеры сеточных функций для стандартного (сплошная линия) и разрывного (штриховая линия) метода Галёркина при  $p = 2$

или

$$\|u - v_h\|_2 = \left( \int_{x_{\min}}^{x_{\max}} |u(x) - v_h(x)|^2 dx \right)^{1/2}. \quad (2.9)$$

В методе Галёркина с разрывными базисными функциями (разрывном методе Галёркина, discontinuous Galerkin method, DG) под сеточной функцией понимается функция, на каждом элементе являющаяся многочленом порядка  $p$  (число  $p \in \mathbb{N} \cup \{0\}$  является параметром метода). Отличие от предыдущего случая заключается в том, что сеточная функция, вообще говоря, не является непрерывной. Пространство сеточных функций имеет размерность  $(p + 1)N_x$ . Близость численного решения к решению дифференциальной задачи можно оценивать по нормам (2.8) или (2.9).

Читатель может спросить, почему приведены именно эти нормы. Дело заключается в том, что именно для этих норм получаются основные теоретические результаты. Для задач с гладкими решениями оценки точности можно получать анализом аппроксимации и устойчивости. Аппроксимация легко анализируется в любой норме, поэтому оценка получается в той норме, в которой известна устойчивость. Спектральный анализ (см. разделы 3.3 и 4.3) позволяет установить устойчивость в  $\|\cdot\|_2$  (обычно говорят “в  $L_2$ ”), что даёт оценку на (2.7). Для линейных монотонных схем для линейного уравнения переноса легко доказывается устойчивость в  $\|\cdot\|_\infty$ , что даёт оценку на (2.6) (см. раздел 3.1). В следующей части курса нам также встретится норма  $\|\cdot\|_1$ . Для конечно-элементных методов легче всего установить оценку на (2.9). Для них также известны оценки, учитывающие не только само численное решение, но и его производные. Такие результаты представляют интерес, но в настоящем курсе не рассматриваются.

Теперь вернёмся к уравнению в частных производных. В настоящем курсе

будем рассматривать только случай статической (неизменной во времени) расчётной сетки. При этом пространственная расчётная сетка  $\{x_j, j = 0, \dots, N_x\}$  (см. выше) дополняется дискретизацией по времени  $\{t_n, n = 0, \dots, N_t\}$ , такой что

$$0 = t_0 < t_1 < \dots < t_{N_t} = t_{\max}.$$

Индекс  $n$  называют номером временного слоя.

В конечно-разностных методах численным решением является набор, содержащий по одному значению на сеточный узел для каждого момента времени и, таким образом, содержащий  $(N_t + 1)N_x$  чисел. Под *сеточной функцией* может пониматься как весь этот набор, так и (чаще) набор из  $N_x$  чисел, соответствующих определённому временному слою. Непрерывной функции  $u \equiv u(t, x)$  на  $n$ -м временном слое сопоставляется сеточная функция

$$\Pi_h u(t_n, \cdot) = \{u_j^n, j = 0, \dots, N_x - 1\}, \quad u_j^n = u(t_n, x_j).$$

Аналогично, в конечно-объёмных методах функции  $u$  обычно сопоставляется сеточная функция с компонентами

$$u_j^n = \frac{1}{x_{j+1} - x_j} \int_{x_j}^{x_{j+1}} u(t_n, x) dx.$$

В конечно-элементных методах сеточной функцией обычно является набор, в котором каждому временному слою соответствует функция, являющаяся многочленом заданного порядка на каждом сеточном элементе (с условием непрерывности или без него). Чтобы оценить близость между решением дифференциальной задачи и сеточной функцией, решение дифференциальной задачи берётся на некоторый фиксированный момент времени, а сеточная функция – на соответствующем временном слое, после чего сравнение проводится одним из описанных выше способов.

Методы, в которых сеточные функции являются функциями не только пространственной переменной, но и времени, также существуют, но в настоящем курсе не будут рассматриваться.

**2.3. Сеточная сходимость.** Целью численных методов является получение решения, близкого к решению поставленной задачи. Только что мы обсудили, как точность численного метода на конкретной задаче (то есть эту самую “близость”) можно выразить той или иной количественной характеристикой. Чтобы сравнить разные численные схемы, можно провести расчёты по каждой из них на одной и той же сетке, после чего сопоставить их точность и вычислительную стоимость (количество арифметических операций или затраты машинного времени).

Если точность на некоторой сетке оказывается недостаточной, то используют более подробную расчётную сетку. Ошибка не обязана монотонно уменьшаться с измельчением сетки, но важно, чтобы измельчением сетки можно было добиться любой наперёд заданной точности. Или, говоря точнее, ошибка решения должна стремиться к нулю при измельчении сетки. Это свойство численного метода называется *сходимостью*.

Сказанное, однако, не годится в качестве формального определения. Выше мы говорили о (неравномерной) расчётной сетке как о наборе узлов  $x_0, \dots, x_{N_x}$ . Положим

$$h_{j+1/2} = x_{j+1} - x_j, \quad j = 0, \dots, N_x - 1.$$

Под мелкостью сетки обычно понимается величина

$$h_{\max} = \max_{j=0, \dots, N_x-1} \{h_{j+1/2}\}.$$

Введём ещё одну характеристику сетки:

$$(\Delta h)_{\max} = \max_{j=1, \dots, N_x-1} \{|h_{j+1/2} - h_{j-1/2}|\}.$$

Неравномерная сетка обычно возникает как результат гладкого отображения равномерной сетки, то есть  $x_j = f(j/N_x)$ , где  $f$  – гладкая строго монотонная функция, равная 0 в нуле и равная 1 в единице. При измельчении сетки  $f$  не меняется. В этом случае

$$\begin{aligned} h_{j+1/2} - h_{j-1/2} &= (x_{j+1} - x_j) - (x_j - x_{j-1}) = \\ &= \left( f\left(\frac{j+1}{N_x}\right) - f\left(\frac{j}{N_x}\right) \right) - \left( f\left(\frac{j}{N_x}\right) - f\left(\frac{j-1}{N_x}\right) \right) = \\ &= \frac{1}{N_x} f'\left(\frac{j+\xi_1}{N_x}\right) - \frac{1}{N_x} f'\left(\frac{j+\xi_2}{N_x}\right) = \frac{1}{N_x^2} f''\left(\frac{j+\xi}{N_x}\right), \end{aligned}$$

где  $\xi_1 \in [-1, 0]$ ,  $\xi_2 \in [0, 1]$ ,  $\xi \in [-1, 1]$ . Последние два равенства получены по теореме Лагранжа. А поскольку  $h_{\max} \geq 1/N_x$  (и равенство достигается на равномерной сетке), получаем, что

$$|h_{j+1/2} - h_{j-1/2}| \leq h_{\max}^2 \sup_{x \in [0, 1]} |f''(x)|$$

и, следовательно,

$$(\Delta h)_{\max} \leq c h_{\max}^2,$$

где  $c = \sup |f''(x)|$  не меняется при измельчении сетки.

Допустим, для какого-то метода мы получили оценку, что ошибка решения, взятая в какой-то норме, не превосходит  $O((\Delta h)_{\max}/h_{\max})$ . Можно ли говорить, что этот метод сходится? Или, допустим, метод обладает оценкой ошибки

$O((\Delta h)_{\max})$ . Это метод первого порядка точности или второго? Если схема будет обладать оценкой ошибки  $O(h_{\max}^2/h_{\min})$ , где  $h_{\min} = \min\{x_{j+1} - x_j\}$ , можно ли говорить, что этот метод сходится при измельчении сетки?

При решении уравнений в частных производных сходимость исследуется при измельчении сетки по всем переменным (например, по пространству и по времени). Но из стандартного курса численных методов читателю должно быть известно, что у каждой явной схемы, аппроксимирующей (1.1) при  $\mu \neq 0$ , есть ограничение на шаг по времени, при котором она является устойчивой (если, конечно, эта схема вообще бывает устойчивой). Если, например, оценка на ошибку решения имеет вид  $O(h_{\max})$ , но верна только при  $\tau < h_{\min}^2$ , можно ли считать, что этот метод сходится?

Если схема записывается для решения задачи с начальными и/или граничными условиями, заданными в общем виде, то сходимость может зависеть от того, какой гладкостью они обладают или какой гладкостью обладает получающееся решение. В многомерных задачах сходимость может зависеть от того, сохраняется ли при измельчении сетки отношение шагов по разным направлениям. Наконец, выше мы уже обсуждали, что саму ошибку решения можно понимать по-разному и оценивать в разных нормах.

Читатель, конечно, может придумать какое-то определение сходимости и исходя из него разрешить все поставленные “вопросы”. Но на самом деле никаких вопросов здесь нет. В научных работах, посвящённых теоретическому исследованию численных методов, понятия сходимости, аппроксимации и устойчивости либо специально уточняются в формулировках теорем, либо не используются в них вовсе. То есть эти понятия имеют не формальный, а скорее разговорный характер. Этому же стандарту будем следовать и мы в настоящем курсе.

### 3. Полностью дискретные схемы

Будем рассматривать задачу Коши для уравнения переноса

$$\frac{\partial u}{\partial t} + \mu \frac{\partial u}{\partial x} = 0, \quad t \in (0, t_{\max}), \quad x \in (0, 2\pi), \quad (3.1)$$

$$u(t, 0) = u(t, 2\pi), \quad u(0, x) = u_0(x), \quad (3.2)$$

где  $u_0(x)$  такая, что её периодическое продолжение на  $x \in \mathbb{R}$  непрерывно дифференцируемо.

**3.1. Схема “левый уголок”.** В этом разделе будем предполагать, что периодическое продолжение  $u_0$  на  $x \in \mathbb{R}$  дважды непрерывно дифференцируемо. Будем использовать неравномерную расчётную сетку с узлами  $x_j$ ,  $j = 0, \dots, N_x$ , периодически продолженную на  $j \in \mathbb{Z}$  (см. раздел 2.2). Обозначим

$$h_{j-1/2} = x_j - x_{j-1}, \quad j = 1, \dots, N_x; \quad (3.3)$$

$$h_{\min} = \min\{h_{j-1/2}\}, \quad h_{\max} = \max\{h_{j-1/2}\}. \quad (3.4)$$

Для дискретизации по времени будем использовать равномерную<sup>3</sup> сетку  $\{t_n = t_{\max}n/N_t, n = 0, \dots, N_t\}$  и обозначим  $\tau = t_{\max}/N_t$ . Под сеточной функцией будем понимать  $N_x$ -периодическую последовательность действительных чисел. Запишем численную схему

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_j^n - v_{j-1}^n}{h_{j-1/2}} = 0, \quad j \in \mathbb{Z}, \quad n = 0, \dots, N_t - 1; \quad (3.5)$$

$$v_j^0 = u_0(x_j), \quad j \in \mathbb{Z}. \quad (3.6)$$

Поскольку периодическое граничное условие эквивалентно продолжению решения  $u(t, x \pm 2\pi) = u(t, x)$ , удобно аналогичным образом продолжать численное решение и расчётную сетку, то есть для всех  $k \in \mathbb{Z}$  доопределять  $v_{j+kN_x}^n = v_j^n$  и  $x_{j+kN_x} = x_j + ka$ , где  $a$  – период задачи (в нашем случае равный  $2\pi$ ). При этом (3.5) и (3.6) автоматически становятся справедливыми для всех  $j \in \mathbb{Z}$ . В этом примере периодическое продолжение не требуется, но мы будем активно пользоваться им в других примерах.

Пусть  $u(t, x)$  – точное решение (3.1) с заданными начальными и граничными условиями. Введём ошибку решения

$$\varepsilon_j^n = v_j^n - u(t_n, x_j). \quad (3.7)$$

---

<sup>3</sup>Равномерность дискретизации по времени в этом примере не существенна.

Подставляя  $v_j^n = u(t_n, x_j) + \varepsilon_j^n$  в численную схему (3.5)–(3.6), получаем  $\varepsilon_j^0 = 0$  и

$$\frac{\varepsilon_j^{n+1} - \varepsilon_j^n}{\tau} + \mu \frac{\varepsilon_j^n - \varepsilon_{j-1}^n}{h_{j-1/2}} = -\epsilon_j^n, \quad (3.8)$$

где

$$\epsilon_j^n = \frac{u(t_n + \tau, x_j) - u(t_n, x_j)}{\tau} + \mu \frac{u(t_n, x_j) - u(t_n, x_{j-1})}{h_{j-1/2}}. \quad (3.9)$$

То есть численная ошибка подчиняется тому же уравнению, что и численное решение, только в правой части стоит минус ошибка аппроксимации.

Вначале получим оценку на величину аппроксимационной ошибки. Разлагая слагаемые в правой части (3.9) в ряд Тейлора, получаем

$$\begin{aligned} \epsilon_j^n = & \left( \frac{\partial u}{\partial t}(t_n, x_j) + \frac{1}{2}\tau \frac{\partial^2 u}{\partial t^2}(t_n + \theta_1\tau, x_j) \right) + \\ & + \mu \left( \frac{\partial u}{\partial x}(t_n, x_j) + \frac{1}{2}h \frac{\partial^2 u}{\partial x^2}(t_n, x_j + \theta_2 h_{j-1/2}) \right), \end{aligned}$$

где  $0 \leq \theta_1 \leq 1$  и  $-1 \leq \theta_2 \leq 0$ . Поскольку  $u$  является решением (3.1), производные по времени можно выразить через производные по пространству. Поэтому слагаемые, пропорциональные первым производным, сокращаются, и

$$\epsilon_j^n = \frac{1}{2}\tau\mu^2 \frac{\partial^2 u}{\partial x^2}(t_n + \theta_1\tau, x_j) + \frac{1}{2}\mu h \frac{\partial^2 u}{\partial x^2}(t_n, x_j + \theta_2 h_{j-1/2}),$$

откуда с очевидностью следует оценка

$$|\epsilon_j^n| \leq \frac{1}{2}(\mu^2\tau + |\mu|h_{j-1/2}) \|u_{xx}\|_\infty, \quad u_{xx} \equiv \frac{\partial^2 u}{\partial x^2},$$

где норма  $\|\cdot\|_\infty$  определяется как максимум модуля функции по  $t \in [0, t_{\max}]$  и  $x \in [0, 2\pi]$ . Через  $\epsilon^n$  и  $\varepsilon^n$  будем обозначать сеточные функции с компонентами, соответственно,  $\epsilon_j^n$  и  $\varepsilon_j^n$ . Используя для сеточной функции  $f$  норму  $\|f\|_\infty = \max_j \{|f_j|\}$ , получаем

$$\|\epsilon^n\|_\infty \leq \frac{1}{2}(\mu^2\tau + |\mu|h_{\max}) \|u_{xx}\|_\infty. \quad (3.10)$$

Таким образом, схема “левый уголок” обладает первым порядком аппроксимации по времени и пространству без ограничений на соотношение шага по времени и шагов по пространству.

Перепишем (3.8) в виде

$$\varepsilon_j^{n+1} = \left(1 - \mu \frac{\tau}{h_{j-1/2}}\right) \varepsilon_j^n + \mu \frac{\tau}{h_{j-1/2}} \varepsilon_{j-1}^n - \tau \epsilon_j^n.$$

При выполнении условий  $\mu \geq 0$  и  $\tau \leq h_{\min}/\mu$  имеем

$$\begin{aligned} |\varepsilon_j^{n+1}| &\leq \left(1 - \mu \frac{\tau}{h_{j-1/2}}\right) |\varepsilon_j^n| + \mu \frac{\tau}{h_{j-1/2}} |\varepsilon_{j-1}^n| + \tau |\epsilon_j^n| \leq \\ &\leq \left(1 - \mu \frac{\tau}{h_{j-1/2}}\right) \|\varepsilon^n\|_\infty + \mu \frac{\tau}{h_{j-1/2}} \|\varepsilon^n\|_\infty + \tau \|\epsilon^n\|_\infty = \|\varepsilon^n\|_\infty + \tau \|\epsilon^n\|_\infty, \end{aligned}$$

откуда получаем

$$\|\varepsilon^{n+1}\|_\infty \equiv \max\{|\varepsilon_j^{n+1}|\} \leq \|\varepsilon^n\|_\infty + \tau \|\epsilon^n\|_\infty.$$

Теперь применяя полученное равенство последовательно для всех временных слоёв от  $n = 0$  до заданного значения, получаем

$$\|\varepsilon^n\|_\infty \leq \|\varepsilon^0\|_\infty + \tau n \max_{k=0, \dots, n-1} \|\epsilon^k\|_\infty. \quad (3.11)$$

Пользуясь оценкой (3.10) и тем, что  $\varepsilon^0 = 0$ , получаем оценку на ошибку решения

$$\|\varepsilon^n\|_\infty \leq \frac{1}{2} n \tau \|u_{xx}\|_\infty (\mu^2 \tau + \mu h_{\max}) \leq n \tau \|u_{xx}\|_\infty \mu h_{\max}.$$

Последнее неравенство написано в силу наложенного условия  $\tau \mu \leq h_{\min} \leq h_{\max}$ . Взятием максимума по  $n$  от 0 до  $N_t$  мы доказываем следующее утверждение.

**Теорема 3.1.** Пусть  $u(t, x)$  – решение задачи Коши, поставленной в начале этого раздела. Пусть  $\{x_j\}$  – невырожденное разбиение  $[0, 2\pi]$ ,  $h_{\min}$  и  $h_{\max}$  определены (3.4),  $\tau = t_{\max}/N_t \leq h_{\min}/\mu$ ,  $\mu > 0$ . Пусть  $v_h = \{v_j^n\}$  – решение по схеме (3.5)–(3.6). Тогда ошибка решения  $\varepsilon$ , компоненты которой определены равенством (3.7), удовлетворяет оценке

$$\max_{n=0, \dots, N_t} \|\varepsilon^n\|_\infty \leq \mu t_{\max} h_{\max} \|u_{xx}\|_\infty. \quad (3.12)$$

Коротко говоря, мы доказали сходимость численного решения к точному при  $h_{\max} \rightarrow 0$  с первым порядком точности по любой последовательности сеток при условии  $\tau \leq h_{\min}/\mu$ . Но, формулируя результат таким образом, мы упускаем из виду часть условий или уточнений: что понимается под ошибкой решения? В какой норме она оценивается? Поэтому предпочтительно формулировать результат, например, так, как это сделано в теореме 3.1.

**Упражнение 1.** Доказать, что утверждение теоремы 3.1 станет неверным, если заменить в ней  $v_h = \{v_j^n\}$  (решение по схеме (3.5)–(3.6)) на решение по схеме  $v_j^0 = u_0(x_j)$ ,

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_{j+1}^n - v_j^n}{h_{j-1/2}} = 0, \quad j \in \mathbb{Z}, \quad n = 0, \dots, N_t - 1.$$

**3.2. Аппроксимация, устойчивость, сходимость.** Доказательство теоремы 3.1 заняло у нас две страницы и на первый взгляд может показаться громоздким и сложным. На самом деле оно простое, поскольку построено по стандартной схеме, рассмотрением которой мы сейчас и займёмся.

Под сеточной функцией будем понимать  $N_x$ -периодическую последовательность действительных чисел. Пусть  $A_h$  – линейный оператор на пространстве сеточных функций, определяемый равенством

$$(A_h f)_j = \frac{f_j - f_{j-1}}{h_{j-1/2}},$$

а  $\Pi_h$  – оператор, сопоставляющей  $2\pi$ -периодической непрерывной функции  $f$  сеточную функцию с компонентами  $(\Pi_h f)_j = f(x_j)$ . Тогда схему (3.5)–(3.6) можно переписать в виде

$$\frac{v^{n+1} - v^n}{\tau} + A_h v^n = 0, \quad n = 0, \dots, N_t - 1; \quad (3.13)$$

$$v^0 = \Pi_h u_0. \quad (3.14)$$

Здесь  $\tau > 0$ ; введём обозначение  $t_n = n\tau$ . Отвлечёмся от конкретного вида операторов  $A_h$  и  $\Pi_h$  и рассмотрим схему общего вида (3.13)–(3.14). Введём ошибку решения как

$$\varepsilon^n = v^n - \Pi_h u(t_n, \cdot).$$

Подставим  $v^n = \varepsilon^n + \Pi_h u(t_n, \cdot)$  в (3.13)–(3.14). При этом получим  $\varepsilon^0 = 0$  и

$$\frac{\varepsilon^{n+1} - \varepsilon^n}{\tau} + A_h \varepsilon^n = -\epsilon^n, \quad n = 0, \dots, N_t - 1, \quad (3.15)$$

где

$$\epsilon^n = \frac{\Pi_h u(t_{n+1}, \cdot) - \Pi_h u(t_n, \cdot)}{\tau} + A_h \Pi_h u(t_n, \cdot). \quad (3.16)$$

**Определение 1.** Схема (3.13)–(3.14) называется устойчивой с константой устойчивости  $K \geq 1$  в норме  $\|\cdot\|$ , если на любой расчётной сетке при всех  $n = 0, \dots, N_t$  любое решение (3.13) удовлетворяет оценке

$$\|v^n\| \leq K \|v^0\|.$$

Схема называется устойчивой в норме  $\| \cdot \|$ , если она устойчива в этой норме с какой-то константой устойчивости.

Другими словами, схема называется устойчивой с константой  $K$ , если  $\|(I - \tau A)^n\| \leq K$ , где  $\| \cdot \|$  – операторная норма, порождённая нормой на пространстве сеточных функций (см. приложение). В общем случае  $K$  может зависеть от  $t_{\max}$ . Как правило, устойчивость имеет место при некоторых ограничениях: на расчётную сетку, на соотношение  $\tau/h_{\min}$  и т. д., но всегда подразумевается, что эти ограничения допускают последовательность сеток с  $h_{\max} \rightarrow 0$ , такую что  $K$  при этом не меняется.

Заметим, что способ задания начальных данных не участвует в определении устойчивости схемы.

**Лемма 3.2.** Пусть схема (3.13)–(3.14) устойчива с константой  $K$  в некоторой норме  $\| \cdot \|$ . Тогда для любого решения (3.15) справедлива оценка

$$\|\varepsilon^n\| \leq K\|\varepsilon^0\| + Kn\tau \max_{k=0,\dots,n-1} \|\epsilon^k\|. \quad (3.17)$$

*Доказательство.* В этом доказательстве индекс, соответствующий номеру временного слоя, будем писать внизу, чтобы не путать со степенями. Разрешим (3.15) относительно  $\varepsilon_{n+1}$ :

$$\varepsilon_{n+1} = (I - \tau A)\varepsilon_n + \tau\epsilon_n.$$

Здесь  $I$  – тождественный оператор. Подставим  $n$  вместо  $n + 1$  и продолжим равенство:

$$\begin{aligned} \varepsilon_n &= (I - \tau A)\varepsilon_{n-1} + \tau\epsilon_{n-1} = \\ &= (I - \tau A)^2\varepsilon_{n-2} + (I - \tau A)\tau\epsilon_{n-2} + \tau\epsilon_{n-1} = \dots = \\ &= (I - \tau A)^n\varepsilon_0 + \tau \sum_{k=0}^{n-1} (I - \tau A)^{n-k}\epsilon_k. \end{aligned}$$

Поскольку  $w_n = (I - \tau A)^n\varepsilon_0$  и  $w_n = (I - \tau A)^n\epsilon_k$  являются решениями (3.13), в силу устойчивости имеем

$$\|(I - \tau A)^n\varepsilon_0\| \leq K\|\varepsilon_0\|, \quad \|(I - \tau A)^n\epsilon_k\| \leq K\|\epsilon_k\|,$$

откуда по неравенству треугольника получаем (3.17). □

Оценка величины ошибки аппроксимации  $\epsilon_n^j$  обычно получается разложением  $u(t, x)$  в ряд Тейлора около точки  $(t_n, x_j)$ , так же, как мы получили её для схемы “левый угол”. Члены при младших производных зануляются в силу

того, что  $u(t, x)$  является решением (3.1), а члены при старших производных пропорциональны соответствующим степеням шага.

Анализ аппроксимационной ошибки можно свести к проверке точности схемы на многочленах, но для этого требуются дополнительные предположения на  $A_h$  и  $\Pi_h$ . Причина в том, что многочлены порядка выше нуля не являются периодическими функциями, поэтому применить к ним  $\Pi_h$  невозможно. Поэтому нам понадобятся предположения о локальности этих операторов.

А именно, предположим, что  $A_h$  имеет вид

$$(A_h f)_j = \sum_{k=j-S, \dots, j+S} a_{jk} f_k, \quad (3.18)$$

где  $S \in \mathbb{N}$ . При этом множество значений  $k$ , по которому ведётся суммирование, называется *шаблоном* схемы. Другими словами, мы предположили, что  $A_h$  имеет конечный шаблон. Также предположим, что  $(\Pi_h f)_j$  определяется только поведением функции  $f$  на отрезке  $[x_{j-S}, x_{j+S}]$ , а число узлов в сетке много больше ширины шаблона ( $N_x > 4(S+1)$ ).

Будем говорить, что схема (3.13)–(3.14) *точна на многочленах* порядка  $p$  (в смысле отображения  $\Pi_h$ ), если для каждого  $j \in \mathbb{Z}$  и каждого многочлена  $P(x)$  степени не выше  $p$  для  $2\pi$ -периодической функции  $g(t, x)$ , совпадающей с  $P(x - \mu t)$  на  $[t_n, t_{n+1}] \times [x_{j-2S}, x_{j+2S}]$ , выполняется

$$\frac{\Pi_h g(t_{n+1}, \cdot) - \Pi_h g(t_n, \cdot)}{\tau} + A_h \Pi_h g(t_n, \cdot) = 0.$$

Предположим теперь, что функция *точна на многочленах* степени  $p$ , и получим оценку на  $\epsilon_j^n$ . Предположим, что  $u(t, x)$  является  $p+1$  раз непрерывно дифференцируемой функцией. Зафиксируем значения  $j$  и  $n$ . Разложим решение в ряд Тейлора:

$$\begin{aligned} u(t, x) &= u_0(x - \mu t) = \\ &= u_0(x_j - \mu t_n) + \sum_{m=1}^p ((x - \mu t) - (x_j - \mu t_n))^m \frac{d^m u_0}{dx^m}(x_j - \mu t_n) + \\ &\quad + ((x - \mu t) - (x_j - \mu t_n))^{p+1} \frac{d^{p+1} u_0}{dx^{p+1}}(\xi), \end{aligned}$$

где  $\xi \in \mathbb{R}$ . Выберем в качестве  $P(y)$  следующий многочлен порядка  $p$ :

$$P(y) = u_0(x_j - \mu t_n) + \sum_{m=1}^p (y - (x_j - \mu t_n))^m \frac{d^m u_0}{dx^m}(x_j - \mu t_n),$$

и пусть  $g(t, x)$  –  $2\pi$ -периодическая функция, совпадающая с  $P(x - \mu t)$  на  $[t_n, t_{n+1}] \times [x_{j-2S}, x_{j+2S}]$ . Перепишем (3.16) в виде

$$\begin{aligned} \epsilon_j^n &= \left[ \frac{\Pi_h g(t_{n+1}, \cdot) - \Pi_h g(t_n, \cdot)}{\tau} + A_h \Pi_h g(t_n, \cdot) \right]_j + \\ &+ \left[ \frac{\Pi_h (u - g)(t_{n+1}, \cdot) - \Pi_h (u - g)(t_n, \cdot)}{\tau} + A_h \Pi_h (u - g)(t_n, \cdot) \right]_j. \end{aligned}$$

Выражение в квадратных скобках в первой строке равно нулю по условию точности на многочленах порядка  $p$ .

Обозначим через  $h_{\max,j}$  и  $h_{\min,j}$  максимальный и минимальный шаги в пределах шаблона схемы. Выражение во второй скобке оценивается покомпонентно с учётом того, что величина  $u - g$  не превосходит  $C_0 \|d^{p+1}u/dx^{p+1}\|_\infty h_{\max,j}^{p+1}$ . Если коэффициенты  $a_{jk}$  в (3.18) ограничены величиной  $C_1 h_{\min,j}^{-1}$ , а  $\|\Pi_h f\|_\infty \leq C_2 \|f\|_\infty$ , откуда получаем оценку

$$|\epsilon_j^n| \leq (2S + 1)C_0 C_1 C_2 \frac{h_{\max}^{p+1}}{h_{\min}} \left\| \frac{d^{p+1}u}{dx^{p+1}} \right\|_\infty \quad (3.19)$$

для всех  $j$  и  $n$ . Из (3.19) очевидным образом получается оценка на  $\|\epsilon^n\|_2$  и  $\|\epsilon^n\|_\infty$ .

Заметим, что вместо “честного” порядка аппроксимации  $p$  мы получили оценку, пропорциональную  $h_{\max,j}/h_{\min,j}$ . Сделанных предположений недостаточно, чтобы улучшить эту оценку.

Повторим ещё раз схему доказательства теоремы 3.1, опираясь на введённые определения.

1. Получаем оценку на аппроксимационную ошибку вида  $O(\tau + h_{\max})$ . Мы можем получить оценку явно (см. доказательство оценки (3.10)) или сделать этот вывод из на линейной функции, но в последнем случае оценка будет доказана при дополнительном предположении, что отношение соседних сеточных шагов не превосходит константы, не меняющейся при измельчении сетки.
2. Замечаем, что (3.5) переписывается в виде

$$v_j^{n+1} = \left(1 - \frac{\mu\tau}{h_{j-1/2}}\right) v_j^n + \frac{\mu\tau}{h_{j-1/2}} v_{j-1}^n.$$

При  $\tau\mu \leq h_{\min}$  множители при  $v_j^n$  и  $v_{j-1}^n$  и в сумме дают единицу, поэтому выполняется  $\|v^{n+1}\|_\infty \leq \|v^n\|_\infty$ , то есть схема устойчива в  $\|\cdot\|_\infty$  с  $K = 1$

3. Пользуемся леммой 3.2 для получения итоговой оценки.

**3.3. Спектральный анализ устойчивости.** Оценка аппроксимационной ошибки обычно получается разложением точного решения в ряд Тейлора. Выше это было проделано для схемы “левый угол”. Этот же метод применим для более сложных схем и в случае нелинейных задач. Устойчивость, напротив, удаётся исследовать лишь в отдельных частных случаях. Одним из них является случай линейной монотонной схемы, фактически единственный полезный пример которой мы рассмотрели в разделе 3.1. В этом разделе мы рассмотрим

другой случай, когда сетка является равномерной и схема “одинакова во всех узлах сетки”.

Будем понимать под расчётной сеткой набор узлов  $\{x_j = jh, j = 0, \dots, N_x\}$ , такой что  $N_x h = 2\pi$  (таким образом, будем допускать только такие значения  $h$ , которые являются делителями  $2\pi$ ). Под сеточной функцией будем понимать  $N_x$ -периодическую последовательность комплексных чисел:  $v_h = \{v_j \in \mathbb{C}, j \in \mathbb{Z}\}$ ,  $v_{j+N_x} = v_j$ . Такие сеточные функции образуют комплексное линейное пространство размерности  $N_x$ . Введём на нём норму

$$\|v_h\|_2 = \left( \sum_{j=0}^{N_x-1} h |v_j|^2 \right)^{1/2}. \quad (3.20)$$

Эта норма отличается от обычной векторной нормы на множитель  $\sqrt{h}$ ; его смысл в том, чтобы норма (3.20) совпадала со стандартной  $L_2$ -нормой кусочно-постоянной функции, равной  $v_j$  на интервале  $((j - 1/2)h, (j + 1/2)h)$ .

Рассмотрим схему следующего вида:

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \frac{\mu}{h} \sum_{k=-S}^S a_k v_{j+k}^n = 0, \quad j \in \mathbb{Z}, \quad n = 0, \dots, N_t - 1; \quad (3.21)$$

$$v^0 = \Pi_h u_0. \quad (3.22)$$

Здесь  $S \in \mathbb{N}$  и  $a_k \in \mathbb{R}$  – некоторые константы. Будем считать, что при измельчении сетки отношение  $\tau/h$  не меняется. Совокупность значений  $v_j^n, j \in \mathbb{Z}$ , будем обозначать через  $v^n$ .

Описанный ниже метод применим также для неявных и/или многослойных схем, но мы не будем на этом останавливаться. Очевидно, что решение (3.21)–(3.22) при каждом  $n$  является  $N_x$ -периодической последовательностью, поэтому эта схема имеет вид (3.13)–(3.14).

Примером схемы вида (3.21)–(3.22) является схема “левый уголок”. В ней (3.21) принимает вид

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_j^n - v_{j-1}^n}{h} = 0, \quad (3.23)$$

при этом  $a_0 = 1, a_{-1} = -1$ , а остальные коэффициенты равны нулю. Вид (3.21)–(3.22) также имеет схема Лакса – Вендроффа. В ней (3.21) принимает вид

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_{j+1/2}^n - v_{j-1/2}^n}{h} = 0, \quad (3.24)$$

где

$$v_{j+1/2}^n = \frac{v_j^n + v_{j+1}^n}{2} - \frac{\mu\tau}{h} \frac{v_{j+1}^n - v_j^n}{2}. \quad (3.25)$$

Мы ввели сеточную функцию как  $N_x$ -периодическую последовательность  $f_j, j \in \mathbb{Z}$ ; будем отождествлять её с набором из  $f_0, \dots, f_{N_x-1}$ . Тогда линейный оператор  $A_h$  на множестве сеточных функций будет представим матрицей<sup>4</sup>.

**Определение 2.** Квадратная матрица, каждая строка которой (кроме первой) получается циклическим сдвигом предыдущей строки вправо на один элемент, называется циклической матрицей или циркулянтном.

Например, если представить равенство (3.23) (описывающее схему “левый уголок”) в виде (3.13), то  $A_h$  будет иметь вид

$$A_h = \frac{\mu}{h} \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & -1 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \quad (3.26)$$

В этом разделе нам будет удобно нумеровать строки и столбцы матрицы с нуля. Введём матрицу  $S_N$  размера  $N \in \mathbb{N}$  с компонентами

$$s_{jk} = N^{-1/2} \exp(2\pi i j k / N), \quad j, k = 0, \dots, N-1 \quad (3.27)$$

(здесь  $i$  – мнимая единица). Матрица  $S_N$  имеет вид

$$S_N = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \exp(2\pi i / N) & \exp(4\pi i / N) & \dots & \exp(2\pi i (N-1) / N) \\ 1 & \exp(4\pi i / N) & \exp(8\pi i / N) & \dots & \exp(4\pi i (N-1) / N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \exp(2\pi i (N-1) / N) & \exp(4\pi i (N-1) / N) & \dots & \exp(2\pi i (N-1)^2 / N) \end{pmatrix}.$$

Можно говорить, что  $S_N^{-1}$  является матричным представлением дискретного преобразования Фурье. Очевидно, что  $S_N = S_N^T$  (здесь транспонирование без комплексного сопряжения, поэтому  $S_N$  не эрмитова).

**Лемма 3.3.** Матрица  $S_N$  является унитарной, т. е.  $S_N^{-1} = \bar{S}_N^T$ .

<sup>4</sup>Либо можно ввести на пространстве сеточных функций базис из функций  $e_k, k = 0, \dots, N_x - 1$ , компоненты которых равны  $(e_k)_j = 1, j = k \pmod{N_x}$  и  $(e_k)_j = 0$  иначе. Тогда мы будем иметь дело с матрицей, представляющей оператор в этом базисе.

*Доказательство.*

$$(S_N \bar{S}_N^T)_{jk} = \frac{1}{N} \sum_{l=0}^{N-1} \exp\left(2\pi i \frac{jl}{N}\right) \exp\left(-2\pi i \frac{kl}{N}\right) = \frac{1}{N} \sum_{l=0}^{N-1} \exp\left(2\pi i (j-k) \frac{l}{N}\right).$$

При  $j = k$  сразу имеем  $(S_N \bar{S}_N^T)_{jk} = 1$ . В противном случае последняя сумма является частичной суммой геометрической прогрессии с показателем  $q = \exp(2\pi i (j-k)/N) \neq 1$ , поэтому

$$(S_N \bar{S}_N^T)_{jk} = \frac{1}{N} \frac{1 - q^N}{1 - q} = \frac{1}{N} \frac{1}{1 - q} (1 - \exp(2\pi i (j-k))) = 0.$$

Унитарность матрицы  $S_N$  доказана.  $\square$

**Следствие 3.4.** Для любой сеточной функции  $v$  верно  $\|S_N v\|_2 = \|v\|_2$ . Как следствие,  $\|S_N\|_2 = \|S_N^{-1}\|_2 = 1$ .

*Доказательство.* Для сеточной функции  $v$  имеем

$$\|S_N v\|_2^2 = h \overline{(S_N v)}^T (S_N v) = h \bar{v}^T \bar{S}_N^T S_N v = h \bar{v}^T S_N^{-1} S_N v = h \bar{v}^T v = \|v\|_2^2.$$

Равенство  $\|S_N\|_2 = 1$  отсюда следует по определению матричной нормы.  $\|S_N^{-1}\|_2 = 1$  доказывается аналогично. Отметим, что множитель  $h$  в цепочке равенств возникал в силу выбранного определения нормы (3.20).  $\square$

**Лемма 3.5.** Пусть  $A$  – циркулянт размера  $N$ . Тогда  $A = S_N \Lambda S_N^{-1}$ , где  $\Lambda$  – диагональная матрица.

*Доказательство.* Докажем вначале это утверждение для матрицы

$$T = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Пусть  $e_j$  –  $j$ -й столбец матрицы  $S_N$  ( $j$  пробегает  $0, \dots, N-1$ ). Тогда

$$\begin{aligned} T e_j &= \frac{1}{\sqrt{N}} T \begin{pmatrix} 1 \\ \exp(2\pi i j/N) \\ \vdots \\ \exp(2\pi i j(N-1)/N) \end{pmatrix} = \\ &= \frac{1}{\sqrt{N}} \begin{pmatrix} \exp(2\pi i j/N) \\ \exp(2\pi i (2j)/N) \\ \vdots \\ 1 \end{pmatrix} = \exp(2\pi i j/N) e_j, \end{aligned}$$

то есть  $e_j$  является собственным вектором матрицы  $T$ , соответствующим собственному значению  $\lambda_j = \exp(2\pi i j/N)$ . Следовательно,  $S_N$  является матрицей правых собственных векторов матрицы  $T$ , откуда для  $A = T$  следует искомое утверждение.

Рассмотрим теперь циркулянт общего вида. Обозначим коэффициенты первой строки матрицы  $A$  через  $a_j$ ,  $j = 0, \dots, N-1$ , и доопределим их на  $j \in \mathbb{Z}$  периодическим образом:  $a_{j+N} = a_j$ . В этих обозначениях матрица  $A$  имеет вид

$$A = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_{-2} & a_{-1} \\ a_{-1} & a_0 & a_1 & \dots & a_{-3} & a_{-2} \\ a_{-2} & a_{-1} & a_0 & \dots & a_{-4} & a_{-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 & a_2 & a_3 & \dots & a_{-1} & a_0 \end{pmatrix}.$$

Легко удостовериться, что

$$A = \sum_{q=0}^{N-1} a_q T^q.$$

Следовательно, каждый  $e_j$  является собственным вектором матрицы  $A$ , соответствующим собственному значению

$$\lambda_j = \sum_{q=0}^{N-1} a_q \exp\left(2\pi i \frac{qj}{N}\right). \quad (3.28)$$

То есть  $AS_N = \Lambda S_N$ , где  $\Lambda$  – диагональная матрица с элементами  $\lambda_j$ .  $\square$

При доказательстве леммы 3.5 мы не только установили, что циркулянт приводится к диагональному виду, но и привели явные выражения для собственных векторов и собственных значений. Важно также, что все циркулянты одного размера имеют один и тот же набор собственных векторов.

**Лемма 3.6.** *Если  $A = S_N \Lambda S_N^{-1}$ , где  $\Lambda$  – диагональная матрица, то  $A$  – циркулянт.*

*Доказательство.* Пусть  $X$  – линейное пространство матриц, представимых в виде  $S_N \Lambda S_N^{-1}$ , где  $\Lambda$  – диагональная матрица. Диагональные матрицы размера  $N$  также образуют линейное пространство размерности  $N$ . Поскольку  $S_N$  – невырожденная матрица,  $X$  также имеет размерность  $N$ . По лемме 3.5 пространство циркулянтов размера  $N$  является подпространством  $X$ , то так как оно тоже имеет размер  $N$ , то оно совпадает с  $X$ . Это и означает, что все элементы  $X$  являются циркулянтами.  $\square$

**Упражнение 2.** Доказать, что сумма циркулянтов есть циркулянт, произведение циркулянтов есть циркулянт и обратная к циркулянту матрица, если существует, то тоже является циркулянтом.

Воспользуемся теперь полученным разложением для циркулянта для схемы (3.21)–(3.22). Перепишем (3.21) в виде

$$v^{n+1} = Gv^n, \quad G = I - \frac{\mu\tau}{h}A.$$

Матрица  $A$  является циркулянтом. Если число узлов в сетке много больше ширины шаблона схемы (точнее,  $N_x > 2S + 1$ ), то первая строка матрицы  $A$  имеет вид

$$(a_0, \dots, a_S, 0, \dots, 0, a_{-S}, \dots, a_{-1}).$$

Читатель может для простоты предполагать, что условие  $N_x > 2S + 1$  выполняется, но все проводимые ниже рассуждения остаются справедливыми и без наложения этого условия.

Матрица  $G$  также является циркулянтом. Пользуясь выражением (3.28) для собственных значений циркулянта, получаем выражение для собственных значений матрицы  $G$ :

$$\lambda_k = 1 - \frac{\mu\tau}{h} \sum_{q=-S}^S a_q \exp\left(2\pi i \frac{qk}{N_x}\right). \quad (3.29)$$

Введём функцию

$$\lambda(\phi) = 1 - \frac{\mu\tau}{h} \sum_{q=-S}^S a_q \exp(iq\phi), \quad \phi \in [0, 2\pi). \quad (3.30)$$

Тогда  $\lambda_k = \lambda(2\pi k/N_x)$ . При измельчении сетки ( $N_x \rightarrow \infty$ ) набор собственных значений заполняет весь контур  $\lambda(\phi)$ .

Теперь всё готово, чтобы сформулировать критерий устойчивости схемы (3.21)–(3.22).

**Теорема 3.7.** Пусть  $\tau/h$  неизменно при измельчении сетки. Схема (3.21)–(3.22) устойчива (в смысле определения 1) тогда и только тогда, когда выполняется

$$\sup_{\phi \in [0, 2\pi)} |\lambda(\phi)| \leq 1, \quad (3.31)$$

где  $\lambda(\phi)$  определено (3.30).

*Доказательство.* Поскольку  $G$  – циркулянт, имеет место

$$v^{n+1} = S_{N_x} \Lambda S_{N_x}^{-1} v^n,$$

где  $\Lambda$  – диагональная матрица с элементами  $\lambda_k = \lambda(2\pi k/N_x)$ . Отсюда

$$v^n = S_{N_x} \Lambda^n S_{N_x}^{-1} v^0.$$

С учётом унитарности  $S_{N_x}$  получаем

$$\sup_{v^0 \neq 0} \frac{\|v^n\|_2}{\|v^0\|_2} = \sup_{v^0 \neq 0} \frac{\|S_{N_x} \Lambda^n S_{N_x}^{-1} v^0\|_2}{\|v^0\|_2} = \sup_{v^0 \neq 0} \frac{\|\Lambda^n S_{N_x}^{-1} v^0\|_2}{\|S_{N_x}^{-1} v^0\|_2} = \|\Lambda^n\|_2. \quad (3.32)$$

Матрица  $\Lambda$  диагональная, поэтому  $\|\Lambda\|_2$  равна максимуму модулей её элементов и  $\|\Lambda^n\|_2 = \|\Lambda\|_2^n$ .

Предположим, что условие (3.31) выполняется. Тогда для любых  $N_x \in \mathbb{N}$  и  $k = 0, \dots, N_x - 1$  выполняется  $|\lambda_k| = |\lambda(2\pi k/N_x)| \leq 1$ , поэтому  $\|\Lambda\|_2 \leq 1$  и из (3.32) имеем

$$\|v^n\|_2 \leq \|\Lambda\|_2^n \|v^0\|_2 = \|v^0\|_2,$$

то есть схема (3.21)–(3.22) устойчива в  $\|\cdot\|_2$  с  $K = 1$ .

Обратно, пусть схема устойчива в  $\|\cdot\|_2$  с некоторой константой  $K$ . Тогда для любого  $v^0$  имеем  $\|v^{N_t}\|_2 \leq K \|v^0\|_2$ . Подставляя в качестве  $v^0$   $k$ -й собственный вектор матрицы  $G$  ( $k$ -й столбец матрицы  $S_{N_x}$ ), получаем

$$|\lambda_k|^{N_t} \leq K.$$

Взяв максимум по  $k$  и возводя в степень  $1/N_t$ , получаем

$$\max\{|\lambda_k|\} \leq K^{1/N_t}.$$

При стремлении  $N_x \rightarrow \infty$  в силу непрерывности  $\lambda(\phi)$  величина в левой части стремится к  $\sup |\lambda(\phi)|$ , а величина в правой части стремится к единице (по условию  $\tau/h = \text{const}$ , а, значит,  $N_t/N_x = \text{const}$ ). Таким образом, приходим к (3.31).  $\square$

Отметим, что коэффициенты матрицы  $A$  могут зависеть от  $\tau/h$  (как это имеет место в схеме Лакса – Вендроффа (3.24)–(3.25)).

Проиллюстрируем поведение неустойчивых схем (3.21)–(3.22) на следующем примере. Пусть выполняется  $\lambda(\pi) = m$ ,  $|m| > 1$ . Тогда при всех чётных  $N_x$  при  $k = N_x/2$  имеет место  $\lambda_k = \lambda(\pi) = m$ . Собственным вектором циркулянта, соответствующим этому собственному значению, является двухточечная

пила:  $(-1)^j/\sqrt{N_x}$ . Подставляя этот вектор в качестве  $v^0$ , получаем  $v^n = m^n v^0$  и  $\|v^n\|_2 = |m|^n \|v^0\|_2$ . Подставляя  $n = t_n/\tau$  и  $\tau = ch$ , получаем

$$\|v^n\|_2 = \exp\left(\frac{t_n}{ch} \ln(|m|)\right) \|v^0\|_2.$$

Таким образом, чем дальше измельчается сетка, тем быстрее растут со временем высокочастотные возмущения.

Если начальные данные гладкие, то высокочастотные компоненты  $v^0$  будут уменьшаться с измельчением сетки, и поэтому генерация высокочастотного возмущения, которое будет дальше экспоненциально расти, может происходить не столько за счёт начальных данных, сколько за счёт ошибок округления.

**Упражнение 3.** Рассмотрим схему вида (3.21)–(3.22), коэффициенты которой не меняются при измельчении сетки. Предположим, что при некотором  $p \in [1, \infty]$  любое ( $N_x$ -периодическое) решение (3.21) удовлетворяет условию

$$\|v^n\|_p \leq K \|v^0\|_p,$$

где

$$\begin{aligned} \|f\|_p &= \left( \sum_{j=0}^{N_x-1} h |f_j|^p \right)^{1/p}, \quad 1 \leq p < \infty; \\ \|f\|_\infty &= \max_{j=0, \dots, N_x-1} |f_j|. \end{aligned} \quad (3.33)$$

Доказать, что эта схема устойчива в норме  $\|\cdot\|_2$  с константой  $K = 1$ .

**Упражнение 4.** Рассмотрим семейство схем вида

$$\sum_{k=-S}^S m_k \frac{v_{j+k}^{n+1} - v_{j+k}^n}{\tau} + \frac{\mu}{h} \sum_{k=-S}^S a_k v_{j+k}^n = 0, \quad (3.34)$$

$v_j^0 = u(0, jh)$ , таких что собственные значения циркулянта  $M$  размерности  $N$  с элементами  $\{0, \dots, 0, m_{-S}, \dots, m_S, 0, \dots, 0\}$  не равны нулю при любом  $N$  и отделены от нуля при  $N \rightarrow \infty$ . Получить критерий устойчивости в норме  $\|\cdot\|_2$  для схем этого семейства.

Наличие периодических условий существенно для того, чтобы теорема 3.7 была применима. Покажем это на примере. Рассмотрим начально-краевую задачу

$$\frac{\partial u}{\partial t} + \mu \frac{\partial u}{\partial x} = 0, \quad t \in (0, t_{\max}), \quad x \in (0, 2\pi), \quad (3.35)$$

$$u(0, x) = u_0(x), \quad u(t, 0) = 0. \quad (3.36)$$

Запишем для этой задачи схему “левый уголок”:

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_j^n - v_{j-1}^n}{h} = 0, \quad j = 1, \dots, N_x, \quad n = 0, \dots, N_t - 1; \quad (3.37)$$

$$v_0^n = 0, \quad n = 1, \dots, N. \quad (3.38)$$

$$v_j^0 = v_0(jh), \quad j = 0, \dots, N_x. \quad (3.39)$$

Положим число Куранта  $\sigma = \mu\tau/h$  равным  $3/2$ . Подставляя  $\tau = 3h/(2\mu)$ , равенство (3.37) переписывается в виде

$$v^{n+1} = Av^n, \quad A = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 3/2 & -1/2 & 0 & \dots & 0 & 0 \\ 0 & 3/2 & -1/2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 3/2 & -1/2 \end{pmatrix}.$$

Матрица  $A$  имеет собственное значение 0 кратности 1 и собственное значение  $-1/2$  кратности  $N_x$ . Таким образом, все собственные значения матрицы  $A$  лежат в единичном круге. Однако мы знаем, что в случае периодических условий схема “левый уголок” устойчива только при числе Куранта  $\sigma = \mu\tau/h \leq 1$ , и естественно предположить, что замена граничного условия не должна сделать её применимой.

**Упражнение 5.** Доказать, что схема (3.37)–(3.39) для задачи (3.35)–(3.36) при  $\mu\tau/h = 3/2$  для любого  $K > 0$  допускает решения, такие что  $\|v^n\| > K\|v^0\|$ .

**Упражнение 6.** Пусть  $0 < \delta \leq 1$ ,  $\sigma = |\mu|\tau/h = \text{const}$ . Рассмотрим схему

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \frac{(\mu v)_{j+1/2}^n - (\mu v)_{j-1/2}^n}{h} = 0, \quad v_j^0 = u(0, jh),$$

с периодическими условиями, где

$$(\mu v)_{j+1/2}^n = \mu \frac{v_j^n + v_{j+1}^n}{2} + \delta |\mu| \frac{v_{j+1}^n - v_j^n}{2}.$$

Получить критерий устойчивости в  $\|\cdot\|_2$  для этой схемы (рассмотреть случаи  $\mu > 0$  и  $\mu < 0$ ).

### 3.4. Метод характеристик.

Метод характеристик является самым простым подходом для построения полностью дискретных схем для уравнения переноса.

Рассмотрим задачу Коши для уравнения переноса (3.1) с периодическими граничными условиями. Будем предполагать, что её решение продолжено периодическим образом. Решение уравнение переноса постоянно вдоль характеристики, то есть равенство

$$u(t, x) = u(t - \tau, x - \mu\tau) \quad (3.40)$$

выполняется для любых значений  $x, t, \tau$ , при которых оно определено. Проведём характеристику через точку  $(t_{n+1}, x_j)$ , то есть применим (3.40) для  $t = t_{n+1}, x = x_j$  и подставим в качестве  $\tau$  шаг по времени. Получим

$$u(t_{n+1}, x_j) = u(t_n, x_j - \mu\tau).$$

Опишем теперь подход к построению численной схемы. Определим какой-либо оператор интерполяции по значениям в точках  $x_j$  в точку  $x_j - \mu\tau$ , то есть определим коэффициенты  $\alpha_{jk}$ , такие что

$$\left| f(x_j - \mu\tau) - \sum_{k=0}^{N_x} \alpha_{jk} f(x_k) \right| \leq Ch_{\max}^p,$$

где  $C > 0$  и  $p \in \mathbb{N}$  – некоторые константы. Тогда запишем схему

$$u_j^{n+1} = \sum_{k=0}^{N_x} \alpha_{jk} u_k^n$$

и приведём её к привычному виду

$$\frac{u_j^{n+1} - u_j^n}{\tau} = \frac{1}{\tau} \left( -u_j^n + \sum_{k=0}^{N_x} \alpha_{jk} u_k^n \right). \quad (3.41)$$

Очевидно, что при условии  $\tau \leq ch_{\max}$  равенство (3.41) аппроксимирует уравнение (3.1) с порядком  $p - 1$ .

Конкретная схема такого вида определяется выбором способа интерполяции. Если положить

$$\begin{aligned} f(x_j - \mu\tau) &\approx f(x_j) \left( 1 - \frac{\mu\tau}{h_{j-1/2}} \right) + f(x_{j-1}) \left( \frac{\mu\tau}{h_{j-1/2}} \right), \quad \mu \geq 0; \\ f(x_j - \mu\tau) &\approx f(x_j) \left( 1 + \frac{\mu\tau}{h_{j+1/2}} \right) - f(x_{j+1}) \left( \frac{\mu\tau}{h_{j+1/2}} \right), \quad \mu < 0, \end{aligned}$$

то мы получим хорошо известные схемы “левый уголок” или “правый уголок”

$$\begin{aligned} \frac{u_j^{n+1} - u_j^n}{\tau} + \mu \frac{u_j^n - u_{j-1}^n}{h_{j-1/2}} &= 0, \quad \mu \geq 0; \\ \frac{u_j^{n+1} - u_j^n}{\tau} + \mu \frac{u_{j+1}^n - u_j^n}{h_{j+1/2}} &= 0, \quad \mu < 0. \end{aligned}$$

в зависимости от знака скорости переноса. При  $\tau < h_{\min}/|\mu|$  эта схема устойчива.

Вопрос устойчивости схем более высокого порядка, построенных на основе характеристического метода, мы рассматривать не будем.

Основная проблема такого подхода к построению схем заключается в использовании явного вида решения. Элементы этого подхода можно изредка встретить в литературе.

**3.5. Устойчивость в  $\|\cdot\|_2$  на неравномерных сетках.** В разделе 3.3 мы рассмотрели конечно-разностные схемы на равномерной сетке и вывели критерий их устойчивости в  $\|\cdot\|_2$ . В случае неравномерных сеток (или схем, меняющихся от узла к узлу) этот критерий неприменим. Тем не менее, для некоторых схем можно доказать невозрастание *некоторой*  $L_2$ -нормы от сеточной функции. Мы покажем это на примере схемы “левый уголок” (3.5)–(3.6).

Введём на пространстве сеточных функций норму

$$\|v\|_2 = \left( \sum_{j=1}^{N_x} h_{j-1/2} |v_j|^2 \right)^{1/2}.$$

**Лемма 3.8.** При  $\mu \geq 0$ ,  $\tau \leq h_{\min}/\mu$  любое решение системы (3.5) удовлетворяет оценке

$$\|v^n\|_2 \leq \|v^0\|_2. \quad (3.42)$$

*Доказательство.* Перепишем равенство (3.5) в виде

$$v_j^{n+1} = v_j^n - \frac{\mu\tau}{h_{j-1/2}}(v_j^n - v_{j-1}^n).$$

Возведём его в квадрат, домножим на  $h_{j-1/2}$  и просуммируем. Получим

$$\begin{aligned} \sum_{j=1}^{N_x} h_{j-1/2} |v_j^{n+1}|^2 &= \sum_{j=1}^{N_x} h_{j-1/2} |v_j^n|^2 - \\ &- 2 \sum_{j=1}^{N_x} \mu\tau (v_j^n - v_{j-1}^n) v_j^n + \sum_{j=1}^{N_x} \mu\tau \frac{\mu\tau}{h_{j-1/2}} (v_j^n - v_{j-1}^n)^2. \end{aligned}$$

Левая часть и первое слагаемое в правой части равны квадратам норм  $v^{n+1}$  и  $v^n$ , соответственно. Введём обозначение  $\sigma_{j-1/2} = \mu\tau/h_{j-1/2}$  и во втором слагаемом

воспользуемся равенством

$$\sum_{j=1}^{N_x} (v_j^n)^2 = \sum_{j=1}^{N_x} (v_{j-1}^n)^2 = \sum_{j=1}^{N_x} \frac{(v_j^n)^2 + (v_{j-1}^n)^2}{2},$$

которое следует из  $v_0^{n+1} = v_{N_x}^{n+1}$ . Тогда

$$(\|v^{n+1}\|_2)^2 = (\|v^n\|_2)^2 - \mu\tau \sum_{j=1}^{N_x} ((v_j^n)^2 - 2v_{j-1}^n v_j^n + (v_{j-1}^n)^2) + \mu\tau \sum_{j=1}^{N_x} \sigma_{j-1/2} (v_j^n - v_{j-1}^n)^2.$$

Или, объединяя теперь последние два слагаемых,

$$(\|v^{n+1}\|_2)^2 = (\|v^n\|_2)^2 - \mu\tau \sum_{j=1}^{N_x} (1 - \sigma_{j-1/2}) (v_j^n - v_{j-1}^n)^2.$$

Поскольку  $\sigma_{j-1/2} \leq 1$ , вычитаемое неотрицательно, и мы приходим к неравенству  $(\|v^{n+1}\|_2)^2 \leq (\|v^n\|_2)^2$ . По индукции отсюда очевидно следует искомое неравенство.  $\square$

Для схемы (3.5)–(3.6) доказательство устойчивости в  $\|\cdot\|_2$  оказалось сложнее, чем его доказательство в  $\|\cdot\|_\infty$  (последняя на неравномерной сетке определяется тем же равенством  $\|f\|_\infty = \max |f_j|$ , что и на равномерной сетке). Но для более сложных схем ситуация противоположная. Рассмотрим, например, обобщение схемы из упражнения 6 на неравномерную сетку:

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \frac{(\mu v)_{j+1/2}^n - (\mu v)_{j-1/2}^n}{\tilde{h}_j} = 0, \quad v_j^0 = u(0, x_j),$$

$$(\mu v)_{j+1/2}^n = \frac{v_j^n + v_{j+1}^n}{2} + \delta|\mu| \frac{v_{j+1}^n - v_j^n}{2}$$

(начальные и периодические условия задаются обычным образом). Здесь  $0 < \delta \leq 1$  и  $\tilde{h}_j = (x_{j+1} - x_{j-1})/2$ . Введём норму

$$\|f\|_2 = \left( \sum_{j=0}^{N_x-1} \tilde{h}_j |v_j|^2 \right)^{1/2}.$$

Доказательство неравенства  $\|v^{n+1}\|_2 \leq \|v^n\|_2$  при некотором ограничении на шаг по времени для этой схемы ненамного сложнее, чем для схемы (3.5). А условие  $\|v^{n+1}\|_\infty \leq \|v^n\|_\infty$  уже, вообще говоря, выполняться не будет. Проверка же существования такого  $K \geq 1$ , не зависящего от  $n$  и  $v^0$ , чтобы выполнялось неравенство  $\|v^n\|_\infty \leq K\|v^0\|_\infty$ , представляет собой сложную научную проблему.

**3.6. Точность на неравномерных сетках.** Выше мы показали, что решение, полученное по устойчивой схеме  $p$ -го порядка аппроксимации, будет сходиться к точному с порядком  $p$ . На равномерной сетке для устойчивой схемы вида (3.21)–(3.22) верно и обратное: если схема не обладает порядком аппроксимации  $p$ , то не будет и сходимости решения к точному с порядком  $p$ . Но на неравномерной сетке схема может обладать более высоким порядком точности, чем предсказывает анализ её ошибки аппроксимации. Поясним это на примере.

Рассмотрим задачу Коши для уравнения переноса (3.1) при  $\mu > 0$  с периодическими условиями. Рассмотрим расчётную сетку  $\{x_j, j = 0, \dots, N_x\}$  на  $[0, 2\pi]$  и равномерную сетку по времени. Будем пользоваться обозначениями (3.3), (3.4). Рассмотрим схему “левый уголок” (3.5)–(3.6):

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_j^n - v_{j-1}^n}{h_{j-1/2}} = 0, \quad v_j^0 = u(0, x_j). \quad (3.43)$$

Наряду с ней рассмотрим модифицированную схему

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_j^n - v_{j-1}^n}{\tilde{h}_j} = 0, \quad v_j^0 = u(0, x_j). \quad (3.44)$$

Схема (3.43) переписывается в виде

$$v_j^{n+1} = \left(1 - \mu \frac{\tau}{h_{j-1/2}}\right) v_j^n + \mu \frac{\tau}{h_{j-1/2}} v_{j-1}^n,$$

и при выполнении условия  $\mu\tau/h_{\min}$  решение по ней удовлетворяет неравенству  $\|v^{n+1}\|_\infty \leq \|v^n\|_\infty$ . Аналогичным образом это неравенство доказывается для решения по схеме (3.44) (заметим, что  $\min\{\tilde{h}_j\} \geq h_{\min}$ ).

Ошибка аппроксимации схемы (3.43), обозначим её через  $\epsilon$ , удовлетворяет оценке (3.10):

$$\|\epsilon^n\|_\infty \leq \frac{1}{2} \|u_{xx}\|_\infty (\mu^2 \tau + \mu h_{\max}).$$

Поэтому решение будет сходиться к точному по любой последовательности сеток, такой что  $h_{\max} \rightarrow 0$  и  $\tau \leq h_{\min}/\mu$ . Выше мы сформулировали это в виде теоремы 3.1.

Если же аналогичным образом рассмотреть схему (3.44), то мы увидим, что условие  $\|\epsilon^n\|_\infty \rightarrow 0$  при  $h_{\max} \rightarrow 0$  не выполняется. Если измельчать сетку так, чтобы выполнялось условие

$$\frac{(\Delta h)_{\max}}{h_{\max}} \rightarrow 0 \quad (3.45)$$

(то есть если сетка “разглаживается” при измельчении), то условие  $\|\epsilon^n\|_\infty \rightarrow 0$  выполняться будет. Тем не менее, сходимость при  $h_{\max} \rightarrow 0$  имеет место и без условия (3.45), а решение (3.44) по точности почти не уступает решению (3.43). Дело здесь в следующем.

Напомним, что ошибка аппроксимация и сходимость зависят от того, каким образом мы сопоставляем функции её сеточный аналог. Мы определяли сеточный образ непрерывной функции формулой (2.5)<sup>5</sup>

$$\Pi_h u = \{u_j, j = 0, \dots, N_x - 1\}, \quad u_j = u(x_j),$$

то есть как совокупность её точечных значений в сеточных узлах. Вместо этого будем сопоставлять непрерывной функции сеточную функцию

$$\tilde{\Pi}_h u = \{u_j, j = 0, \dots, N_x - 1\}, \quad u_j = u(x_{j+1/2}),$$

где  $x_{j+1/2} = (x_j + x_{j+1})/2$ .

Обозначим  $\tilde{\epsilon}^n = v^n - \tilde{\Pi}_h u(t_n, \cdot)$ , то есть  $\tilde{\epsilon}_j^n = v_j^n - u(t_n, x_{j+1/2})$ . Подставляя это равенство в (3.44), получаем

$$\frac{\tilde{\epsilon}_j^{n+1} - \tilde{\epsilon}_j^n}{\tau} + a \frac{\tilde{\epsilon}_j^n - \tilde{\epsilon}_{j-1}^n}{\tilde{h}_j} = \tilde{\epsilon}_j^n, \quad \tilde{\epsilon}_j^0 = u(0, x_j) - u(0, x_{j+1/2}),$$

где  $\tilde{\epsilon}$  – ошибка аппроксимации схемы (3.44) в смысле отображения  $\tilde{\Pi}_h$ . Повторяя доказательство теоремы 3.1, мы получаем оценки

$$\|\tilde{\epsilon}^n\|_\infty \leq \frac{1}{2} \|u_{xx}\|_\infty (\mu^2 \tau + \mu h_{\max})$$

(см. (3.10)) и, при  $\tau \leq h_{\min}/\mu$ ,

$$\|\tilde{\epsilon}^n\|_\infty \leq \|\tilde{\epsilon}^0\|_\infty + \tau n \max_{k=0, \dots, n-1} \|\tilde{\epsilon}^k\|_\infty$$

(см. (3.11)). Остаётся заметить, что для всех  $n$  выполняется

$$\|\tilde{\epsilon}^n - \epsilon^n\|_\infty \leq \frac{1}{2} h_{\max} \|u_x\|_\infty,$$

что, в частности, влечёт

$$\|\tilde{\epsilon}^0\|_\infty = \|\tilde{\epsilon}^0 - \epsilon^0\|_\infty \leq \frac{1}{2} h_{\max} \|u_x\|_\infty.$$

---

<sup>5</sup>В отличие от (2.5), в нашем случае индекс  $j$  меняется от 0 до  $N_x - 1$ , поскольку заданы периодические условия.

В итоге получаем оценку

$$\max_{n=0,\dots,N_t} \|\varepsilon^n\|_\infty \leq \frac{1}{2} \mu t_{\max} \|u_{xx}\|_\infty (\mu\tau + h_{\max}) + h_{\max} \|u_x\|_\infty. \quad (3.46)$$

Таким образом, схема (3.44) обладает первым порядком точности при  $h_{\max} \rightarrow 0$ ,  $\tau \leq h_{\min}/\mu$ .

Рассмотренный пример показывает, что на неравномерных сетках порядок точности может быть выше порядка аппроксимации. Этот же эффект может иметь место при рассмотрении схем, меняющихся от ячейки к ячейке, схем с определением нескольких переменных на ячейку (например, метода Галёркина с разрывными базисными функциями) и в случае уравнений с негладкими коэффициентами. В то же время, на равномерных сетках для линейного уравнения с постоянным коэффициентом точность не зависит от того, каким образом определять проекцию на сетку.

**Упражнение 7.** Пусть сетка равномерная в пространстве и времени. Есть два способа отображения решения на сетку: поточечный  $u_n^j = u(t_n, x_j)$  и интегральный

$$\tilde{u}_j^n = \int_{-1/2}^{1/2} u(t_n, (j + \xi)h) d\xi.$$

Пусть схема (3.21),  $v_j^0 = u_j^0$  при любых  $p+1$  раз непрерывно дифференцируемых начальных данных обладает порядком аппроксимации  $p$  в смысле поточечного отображения, то есть выполняется

$$\left| \frac{u(t_{n+1}, x_j) - u(t_n, x_j)}{\tau} + \frac{\mu}{h} \sum_{k=-S}^S a_k u(t_n, x_{j+k}) \right| \leq C \left\| \frac{\partial^{p+1} u}{\partial x^{p+1}} \right\|_\infty h^p.$$

Доказать, что схема (3.21),  $v_j^0 = \tilde{u}_j^0$  будет обладать порядком аппроксимации  $p$  в смысле интегрального отображения.

## 4. Полудискретные схемы

**4.1. Метод линий.** Пусть есть некоторое дифференциальное уравнение в частных производных, решением которого является функция двух переменных – пространства и времени. Если зафиксированы расчётные сетки в пространстве и времени, то численная схема представляет собой систему алгебраических (т. е. не дифференциальных) соотношений.

*Методом линий* (или методом прямых) называют способ построения численных методов для решения эволюционных уравнений в частных производных, заключающийся в последовательном выполнении двух этапов.

1. Замена уравнения в частных производных обыкновенным дифференциальным уравнением (ОДУ), в котором переменной является время, а неизвестной – зависящая от него сеточная функция. Если в пространстве сеточных функций выбрать базис, то это уравнение эквивалентно системе ОДУ относительно коэффициентов разложения сеточной функции по выбранному базису. Задача Коши для такого ОДУ (или системы ОДУ) называется *полудискретной схемой*.
2. Применение к этой системе ОДУ какого-либо метода численного решения систем ОДУ. Процесс численного решения этой системы ОДУ называется *интегрированием по времени*.

Для интегрирования по времени чаще всего используются явные методы Рунге – Кутты и формулы дифференцирования назад (backward difference formulas) BDF1 и BDF2. Неявное интегрирование по времени в этом курсе мы рассматривать не будем.

В рамках конечно-разностного подхода, полудискретная схема получается из исходного дифференциального уравнения заменой производных по пространственным переменным операторами, действующими на пространстве сеточных функций, и постановкой начальных условий для этого уравнения. Проиллюстрируем этот метод применительно задаче Коши для уравнения переноса

$$\frac{\partial u}{\partial t} + \mu \frac{\partial u}{\partial x} = 0, \quad t \in (0, t_{\max}), \quad x \in (0, 2\pi), \quad (4.1)$$

$\mu > 0$ , с периодическими граничными условиями  $u(t, 0) = u(t, 2\pi)$  и начальными условиями  $u(0, x) = u_0(x)$ , предполагая, что решение достаточно гладкое.

Рассмотрим расчётную сетку  $x_j, j = 0, \dots, N_x$ . Под сеточной функцией будем понимать набор из  $N_x$  действительных или комплексных чисел (в зависимости от контекста). Тогда, аппроксимируя  $\mu \partial u / \partial x$  некоторым оператором  $A_h$ , действующим в пространстве сеточных функций, можно записать

$$\frac{dv}{dt} + A_h(v) = 0, \quad v(0) = \Pi_h u_0. \quad (4.2)$$

Здесь  $\Pi_h$  – некоторое отображение некоторого функционального пространства (например, пространства непрерывных  $2\pi$ -периодических функций) на пространство сеточных функций. Таким образом, мы получили задачу Коши для системы из  $N_x$  дифференциальных уравнений. Если  $A_h v$  является непрерывной по Липшицу функцией  $v$  (в частности, если  $A_h$  линейный), задача (4.2) имеет единственное решение.

В частности, если пространственную производную аппроксимировать разностью назад, то есть положить

$$(A_h v)_j = \frac{\mu}{h}(v_j - v_{j-1}), \quad j = 0, \dots, N_x - 1,$$

где  $v_{-1} \equiv v_{N_x-1}$ , то система (4.2) превращается систему ОДУ

$$\frac{dv_j}{dt} + \frac{\mu}{h}(v_j(t) - v_{j-1}(t)) = 0, \quad j = 0, \dots, N_x - 1. \quad (4.3)$$

Дополним её начальными условиями

$$v_j(0) = u(0, x_j). \quad (4.4)$$

Равенства (4.3) и (4.4) образуют полудискретную схему вида (4.2), где оператор  $A_h$  линейный и имеет матричное представление (3.26).

Если применить для решения задачи Коши (4.3), (4.4) явный метод Рунге – Кутты 1-го порядка, то получим схему “левый уголок” (3.5)–(3.6).

В случае полностью дискретных схем мы могли продолжить начальные данные периодическим образом и сказать, что равенство (3.5) верно при всех  $j \in \mathbb{Z}$ . Решение полученной бесконечной системы автоматически получалось периодическим на каждом временном слое. В полудискретном случае мы также можем считать, что сеточная функция определена для всех  $j \in \mathbb{Z}$  и периодична с периодом  $N_x$ . При этом (4.3) будет верно при всех  $j \in \mathbb{Z}$ . Но рассматривать (4.3),  $j \in \mathbb{Z}$ , просто как счётный набор дифференциальных уравнений нежелательно, поскольку придётся отдельно доказывать единственность решения такой системы. Вместо этого можно рассматривать (4.3) как ОДУ на пространстве  $N_x$ -периодических числовых последовательностей (конечномерное пространство размерности  $N_x$ ) или как ОДУ для пространстве ограниченных числовых последовательностей (бесконечномерное пространство, полное относительно нормы  $\|\cdot\|_\infty$ ). В этих случаях единственность (и  $N_x$ -периодичность) решения гарантируется.

**4.2. Аппроксимация и устойчивость полудискретной схемы.** При использовании метода линий ошибка решения складывается из ошибки решения по полудискретной схеме и ошибки интегрирования по времени. В этом и следующем разделах мы займёмся первой составляющей.

Рассмотрим линейную полудискретную схему (т. е. схему вида (4.2) с линейным оператором  $A_h$ ):

$$\frac{dv}{dt} + A_h v = 0, \quad v(0) = \Pi_h u_0. \quad (4.5)$$

Под ошибкой решения будем понимать величину  $\varepsilon(t) = v(t) - \Pi_h u(t, \cdot)$ . Подставляя  $v(t) = \Pi_h u(t, \cdot) + \varepsilon(t)$  в (4.5), получаем

$$\frac{d\varepsilon}{dt} + A_h \varepsilon = -\epsilon(t), \quad \varepsilon(0) = 0, \quad (4.6)$$

где

$$\epsilon(t) = \frac{d(\Pi_h u(t, \cdot))}{dt} + A_h \Pi_h u(t, \cdot). \quad (4.7)$$

Величину  $\epsilon(t)$  будем называть ошибкой аппроксимации.

**Определение 3.** Будем говорить, что схема (4.5) устойчива с константой  $K(t) \geq 1$  в норме  $\|\cdot\|$ , если на любой расчётной сетке все решения уравнения

$$\frac{dv}{dt} + A_h v = 0$$

при всех  $t > 0$  удовлетворяют оценке  $\|v(t)\| \leq K(t)\|v(0)\|$ . Будем говорить, что схема устойчива, если она устойчива с какой-то  $K(t)$ .

Для полудискретных схем справедлив аналог леммы 3.2. Мы приведём её короткое доказательство, опирающееся на понятие матричной экспоненты; читатель, не знакомый с этим понятием, может пропустить доказательство.

**Лемма 4.1.** Пусть схема устойчива с константой  $K(t)$ , где  $K(t)$  не убывает с ростом  $t$ . Тогда любое решение неоднородной системы

$$\frac{dv(t)}{dt} + A_h v(t) = y(t) \quad (4.8)$$

удовлетворяет оценке

$$\|v(t)\| \leq K(t)\|v(0)\| + K(t)t \sup_{0 \leq t' \leq t} \|y(t')\|.$$

*Доказательство.* Устойчивость по начальным данным равносильна выполнению неравенства  $\|\exp(-At)\| \leq K(t)$  для всех  $t \geq 0$ . Решение (4.8) имеет вид

$$v(t) = \exp(-At)v(0) + \int_0^t \exp(-At') y(t-t') dt,$$

откуда искомая оценка очевидна.  $\square$

Пользуясь леммой 4.1, для полудискретной схемы (4.5) мы получаем оценку ошибки

$$\|\varepsilon(t)\| \leq tK(t) \sup_{0 \leq t' \leq t} \|\epsilon(t')\|. \quad (4.9)$$

То есть, как и в случае полностью дискретных схем, если схема устойчива по начальным данным, то оценка на ошибку решения легко получается из оценки аппроксимационной ошибки.

Для дальнейшего рассуждения нам понадобится следующая лемма.

**Лемма 4.2.** Пусть  $f_j(\xi)$ ,  $j = 1, \dots, N$ , – дифференцируемы в  $\xi = 0$ . Пусть  $g(\xi) = \max\{|f_j(\xi)|\}$  дифференцируема в нуле. Тогда для всех  $j$ , таких что  $|f_j(0)| = g(0)$ , справедливо  $g'(0) = f'_j(0) \operatorname{sign} f_j(0)$ .

*Доказательство.* Докажем это утверждение для  $N = 2$ , общий случай рассматривается аналогично.

- Если  $|f_1(0)| > |f_2(0)|$ , то в окрестности  $\xi = 0$  имеем  $g(\xi) = |f_1(\xi)| > 0$  и  $g'(0) = f'_1(0) \operatorname{sign} f_1(0)$ .
- Если  $|f_2(0)| > |f_1(0)|$ , то в окрестности  $\xi = 0$  имеем  $g(\xi) = |f_2(\xi)| > 0$  и  $g'(0) = f'_2(0) \operatorname{sign} f_2(0)$ .
- Если  $|f_1(0)| = |f_2(0)| \neq 0$ , то  $|f_1(\xi)|$  и  $|f_2(\xi)|$  дифференцируемы в нуле и  $|f_j'|_{\xi=0} = f'_j(0) \operatorname{sign} f_j(0)$ . Пользуясь определением дифференцируемости, легко показать, что  $g = \max\{|f_1|, |f_2|\}$  дифференцируема в нуле только если  $|f_1'|_{\xi=0} = |f_2'|_{\xi=0}$ , при этом  $g'(0)$  равно каждой из этих производных.
- Если  $|f_1(0)| = |f_2(0)| = 0$ , то  $g$  дифференцируема в нуле только если  $f'_1(0) = f'_2(0) = 0$ , при этом  $g'(0) = 0$ .

Во всех случаях искомое равенство доказано в предположении дифференцируемости  $g$  в нуле.  $\square$

Рассмотрим неравномерную расчётную сетку  $\{x_j\}$  и полудискретную схему с разностью назад

$$\frac{dv_j}{dt} + \mu \frac{v_j - v_{j-1}}{h_{j-1/2}} = 0, \quad j = 1, \dots, N_x; \quad (4.10)$$

$$v_0 = v_{N_x}, \quad v(0) = \Pi_h u_0 \equiv \{u_0(x_j)\}. \quad (4.11)$$

**Лемма 4.3.** При  $\mu > 0$  для любого решения (4.10) верно

$$\|v(t)\|_{\infty} \leq \|v(0)\|_{\infty}.$$

*Доказательство.* Пусть при некотором  $t' \in [0, t_{\max}]$  максимум по  $j = 1, \dots, N_x$  величины  $|v_j(t')|$  достигается в некотором узле  $j$ . Тогда, если в этой точке достигается максимум, то

$$\frac{dv_j(t')}{dt} = -\mu \frac{v_j(t') - v_{j-1}(t')}{h_{j-1/2}} \leq 0.$$

Аналогично, если в этой точке достигается минимум, то  $dv_j(t')/dt \geq 0$ . Во все моменты времени, в которые производная  $d\|v(t)\|_{\infty}/dt$  существует, в силу леммы 4.2 она равна

$$\frac{d\|v(t)\|_{\infty}}{dt} = \frac{dv_j(t)}{dt} \text{sign } v_j(t)$$

для всех  $j$ , в которых достигается максимум  $|v_j(t)|$ . Следовательно  $d\|v(t)\|_{\infty}/dt$  неположительно. Остаётся заметить, что  $\|v(t)\|_{\infty}$  является непрерывной по Липшицу функцией времени; по теореме Лебега (теореме 2.2) она дифференцируема почти всюду и представима в виде интеграла Лебега

$$\|v(t)\|_{\infty} = \|v(0)\|_{\infty} + \int_0^t \frac{d\|v(t')\|_{\infty}}{dt'} \Big|_{t=t'} dt',$$

откуда следует искомое неравенство. □

**Теорема 4.4.** Пусть  $u(t, x)$  – решение задачи Коши для уравнения переноса (4.1) при  $\mu > 0$  с периодическими условиями. Пусть  $\{x_j\}$  – неравномерная расчётная сетка,  $\Pi_h$  – оператор, сопоставляющий непрерывной функции  $f$  набор её значений в сеточных узлах,  $v(t) = \{v_j(t)\}$  – решение системы (4.10)–(4.11). Тогда справедливо неравенство

$$\|v(t) - \Pi_h u(t, \cdot)\|_{\infty} \leq \frac{1}{2} t \mu h_{\max} \|u_{xx}\|_{\infty}.$$

*Доказательство.* Покажем вначале, что имеет место первый порядок аппроксимации. Раскрывая (4.7), имеем

$$\epsilon_j(t) = \frac{\partial u}{\partial t}(t, x_j) + \mu \frac{u(t, x_j) - u(t, x_{j-1})}{h_{j-1/2}}. \quad (4.12)$$

Разлагая  $u(t, x_{j-1})$  в ряд Тейлора около точки  $x = x_j$  и пользуясь уравнением (4.1), получаем оценку

$$\|\epsilon(t)\|_\infty = \max_j |\epsilon_j(t)| \leq \frac{\mu h_{\max}}{2} \left\| \frac{\partial^2 u}{\partial x^2} \right\|_\infty. \quad (4.13)$$

В силу леммы 4.3 схема устойчива с константой  $K = 1$ , и остаётся применить неравенство (4.9).  $\square$

**4.3. Спектральный анализ устойчивости.** Продолжим рассмотрение численных методов для решения задачи Коши для уравнения переноса (4.1) с периодическими условиями. В этом разделе будем рассматривать только равномерные сетки.

Рассмотрим полудискретную схему вида

$$\frac{dv_j}{dt} + \frac{\mu}{h} \sum_{k=-S}^S a_k v_{j+k} = 0, \quad j = 0, \dots, N_x - 1; \quad (4.14)$$

$$v(0) = \Pi_h u_0, \quad (4.15)$$

где  $v_j$  продолжено периодическим образом на  $j \in \mathbb{Z}$ . Здесь  $a_k \in \mathbb{R}$  – некоторые константы. Как и для полностью дискретных схем вида (3.21)–(3.22), устойчивость в  $\|\cdot\|_2$  рассматриваемой схемы можно исследовать методом спектрального анализа. При  $\mu = 0$  устойчивость с  $K = 1$  очевидна, поэтому далее будем рассматривать только случай  $\mu \neq 0$ .

Перепишем (4.14) в матричном виде

$$\frac{dv(t)}{dt} + \frac{\mu}{h} A_h v(t) = 0. \quad (4.16)$$

Матрица  $A_h$  является циркулянтном. В разделе 3.3 было показано, что все циркулянты размера  $N_x$  имеют одну и ту же матрицу собственных векторов  $S_{N_x}$ , а собственные значения  $\lambda_k$ ,  $k = 0, \dots, N_x - 1$ , выражаются формулой (3.28).

**Лемма 4.5.** *Полудискретная схема (4.14)–(4.15) устойчива по начальным данным тогда и только тогда, когда при всех  $\phi \in \mathbb{R}$  выполняется  $\mu \operatorname{Re} \lambda(\phi) \geq 0$ , где*

$$\lambda(\phi) = \sum_{q=-S}^S a_q \exp(iq\phi). \quad (4.17)$$

*Доказательство.* Домножая (4.16) слева на матрицу  $S_{N_x}$ , получаем

$$\frac{d}{dt}(S_{N_x} v) + \frac{\mu}{h} \text{diag}\{\lambda_k\}(S_{N_x} v) = 0,$$

где  $\text{diag}\{\lambda_k\}$  – диагональная матрица с элементами  $\lambda_k = \lambda(2\pi k/N_x)$  на диагонали. Решение этой системы записывается в виде

$$S_{N_x} v(t) = E(t) S_{N_x} v(0), \quad E(t) = \text{diag}\{\exp(-\lambda_k \mu t/h)\}.$$

Поскольку  $E(t)$  – диагональная матрица, её норма равна максимуму модулей её элементов, то есть

$$\|E(t)\|_2 = \max_k \left\{ \left| \exp \left( -\lambda_k \mu \frac{t}{h} \right) \right| \right\} = \exp \left( \max_k \{\text{Re}(-\mu \lambda_k)\} \frac{t}{h} \right).$$

Пусть  $\mu \text{Re} \lambda(\phi) \geq 0$  при всех  $\phi$ . Поскольку  $\lambda_k = \lambda(2\pi k/N_x)$ , для всех  $k$  имеем  $\mu \text{Re} \lambda_k \geq 0$  и, следовательно,  $\|E(t)\|_2 \leq 1$ . Пользуясь унитарностью матрицы  $S_{N_x}$ , получаем оценку

$$\|v(t)\|_2 = \|S_{N_x} v(t)\|_2 \leq \|E(t)\|_2 \|S_{N_x} v(0)\|_2 \leq \|S_{N_x} v(0)\|_2 = \|v(0)\|_2,$$

то есть схема устойчива по начальным данным с  $K = 1$ .

Обратно, пусть схема устойчива. По определению устойчивости выполняется  $\|E(t)\|_2 \leq K(t)$ , то есть

$$\exp \left( \max_k \{\text{Re}(-\mu \lambda_k)\} \frac{t}{h} \right) \leq K(t).$$

Взяв логарифм и поделив на  $t/h$ , получаем

$$\max_k \{\text{Re}(-\mu \lambda_k)\} \leq h \frac{\ln K(t)}{t}.$$

При фиксированном времени устремив  $N_x$  к бесконечности (или, что то же самое,  $h$  к нулю), получаем

$$\sup_{\phi \in [0, 2\pi]} \text{Re}(-\mu \lambda(\phi)) \leq 0,$$

то есть для всех  $\phi \in [0, 2\pi)$  выполняется  $\mu \text{Re} \lambda(\phi) \geq 0$ . □

Доказательство неустойчивости проще всего понять на следующем примере, уже рассмотренном в полностью дискретном случае. Пусть  $N_x$  чётное, а  $\text{Re} \lambda(\pi) = -r < 0$ . Собственным вектором, соответствующим  $\phi = \pi$ , является

двухточечная пила:  $v_j = (-1)^j$ . Если подставить такую пилу в качестве начальных данных, решение системы (4.14) будет зависеть только от отношения  $t/h$ , но не от  $t$  и  $h$  по отдельности. Значит, если пила усиливается со временем, то не может быть никакой оценки её скорости роста, зависящей только от  $t$ .

На неравномерных сетках, как и в случае полностью дискретной схемы, нет универсальной процедуры для анализа устойчивости в  $\|\cdot\|_2$ . Для схемы с направленной разностью устойчивость в  $\|\cdot\|_2$  может быть доказана аналогично устойчивости схемы “левый угол” (раздел 3.5). В следующей части курса мы докажем устойчивость в  $\|\cdot\|_2$  классического и разрывного методов Галёркина.

**4.4. Решение ОДУ.** Методы решения обыкновенных дифференциальных уравнений подробно рассматриваются в стандартных курсах численных методов. Поэтому в этом разделе мы напомним только некоторые факты.

Рассмотрим задачу Коши для системы ОДУ

$$\frac{du}{dt} = f(u, t), \quad t \in (0, t_{\max}); \quad u(0) = u_0 \in \mathbb{C}^m, \quad (4.18)$$

где  $m \in \mathbb{N}$ , а  $f$  непрерывна по  $t$  и непрерывна по Липшицу по  $u$ . Будем предполагать, что решение этой задачи в классе непрерывно дифференцируемых на  $[0, t_{\max}]$  функций существует.

Явным методом Рунге – Кутты для решения задачи Коши (4.18) называется численная схема, которая может быть записана в следующей форме:

$$\begin{aligned} v_0 &= u_0; \\ k_{n,1} &= f(v_n, t_n + \alpha_1 \tau_n); \\ k_{n,2} &= f(v_n + \tau_n \beta_{2,1} k_{n,1}, t_n + \alpha_2 \tau_n); \\ &\dots \\ k_{n,r} &= f\left(v_n + \tau_n \sum_{s=1}^{r-1} \beta_{r,s} k_{n,s}, t_n + \alpha_r \tau_n\right); \\ v_{n+1} &= v_n + \tau_n \sum_{s=1}^r \gamma_s k_{n,s}. \end{aligned} \quad (4.19)$$

Здесь использовано обозначение  $\tau_n = t_{n+1} - t_n$ . При этом  $r$  называется числом *стадий* этого метода. Простейшим методом такого вида является явный метод Эйлера:

$$v_0 = u_0; \quad v_{n+1} = v_n + \tau_n f(v_n, t_n).$$

Метод (4.19) можно кратко записать в виде

$$\frac{v_{n+1} - v_n}{\tau_n} = F(v_n, t_n, \tau_n), \quad (4.20)$$

$$v_0 = u_0. \quad (4.21)$$

Например, для 1-стадийного метода  $F(v, t, \tau) = \gamma_1 f(v, t + \alpha_1 \tau)$ , а для 2-стадийного

$$F(v, t, \tau) = \gamma_1 f(v, t + \alpha_1 \tau) + \gamma_2 f(v + \tau \beta_{2,1} f(v, t + \alpha_1 \tau), t + \alpha_2 \tau).$$

Говорят, что метод (4.20)–(4.21) обладает порядком аппроксимации  $p$  для решения (4.18), если при всех  $t$  и  $\tau$ , таких что  $0 \leq t < t + \tau \leq t_{\max}$  выполняется

$$\left\| \frac{u(t + \tau) - u(t)}{\tau} - F(u(t), t, \tau) \right\| \leq C \tau^p,$$

где  $u$  – решение (4.18), а  $C$  не зависит от  $t$  и  $\tau$ . Порядок аппроксимации метода, вообще говоря, зависит от выбранной задачи. Высоким порядком аппроксимации схемы обычно обладают только для достаточно гладких решений. Также порядок аппроксимации может зависеть от того, есть ли у функции  $f(u, t)$  фактическая зависимость от второго аргумента, является ли она линейной по  $u$  и т. д.

Следующая теорема устанавливает устойчивость методов Рунге – Кутты.

**Теорема 4.6.** Пусть  $f(u, t)$  является непрерывной по Липшицу функцией  $u$ . Тогда для решения  $v$  системы (4.20) и любого (их много, т. к. начальные данные не фиксированы) решения  $\tilde{v}$  системы

$$\frac{\tilde{v}_{n+1} - \tilde{v}_n}{\tau} = F(\tilde{v}_n, t_n, \tau_n) + \epsilon_n$$

выполняется оценка

$$\|v_n - \tilde{v}_n\| \leq e^{Ct_n} \left( \|v_0 - \tilde{v}_0\| + \frac{2}{C} \max_{k=0, \dots, n-1} \|\epsilon_k\| \right),$$

где  $C$  – некоторая константа, зависящая только от  $F$ .

Доказательство этой теоремы и более точные оценки см. в базовых курсах численных методов<sup>6</sup>. Подстановкой  $\tilde{v}_n = u(t_n)$  легко доказывается следующий результат.

<sup>6</sup>Например, здесь: Петров И. Б., Лобанов А. И., Лекции по вычислительной математике. М.: Бином, 2006.

**Теорема 4.7.** Пусть  $u$  – решение задачи Коши (4.18). Пусть явный метод Рунге – Кутты (4.19) обладает порядком аппроксимации  $p$ . Тогда для некоторой константы  $C$ , не зависящей от  $\tau_n$ , выполняется

$$\max_{n=0,\dots,N_t} \|v_n - u(t_n)\| \leq C \left( \max_{n=0,\dots,N_t-1} \tau_n \right)^p.$$

При решении начально-краевых задач с неоднородным граничным условием  $u(t,0) = u_b(t)$  может возникнуть желание на  $s$ -й итерации метода Рунге – Кутты в (4.19) положить первый аргумент функции  $f$  в граничном узле равным  $u_b(t + \alpha_s \tau)$ . Несмотря на кажущуюся естественность такого подхода, он, вообще говоря, приводит к возникновению ошибки порядка  $O(\tau^2)$ . Этой ошибки можно избежать, если для граничного узла записывать уравнение  $dv_0/dt = u'_b(t)$  и решать его в рамках используемого метода Рунге – Кутты вместе с уравнениями для всех остальных компонент сеточной функции. Но этот метод имеет свои недостатки: во-первых, он очевидно некорректен, если  $u'_b(t)$  имеет разрыв; во-вторых, погрешность в граничном узле будет накапливаться со временем и при длительном расчёте (например, при расчёте до достижения стационарного решения) решение может получиться неудовлетворительным.

**4.5. Устойчивость полностью дискретной схемы.** Предположим, что мы построили полудискретную схему, решение которой сходится к точному решению решаемой задачи. Каким бы методом решения ОДУ мы не воспользовались (например, явным методом Эйлера), выбрав при каждом фиксированном  $h$  достаточно мелкий шаг по времени  $\tau(h)$ , мы получим полностью дискретное решение, сходящееся к точному при  $h \rightarrow 0$ ,  $\tau = \tau(h)$ . Но всегда ли можно ли добиться этого, выбирая  $\tau = ch$ , где  $c > 0$  не зависит от  $h$ ? Ответ на этот вопрос отрицательный.

Действительно, рассмотрим полудискретную схему на равномерной сетке

$$\frac{dv_j}{dt} + \mu \frac{v_{j+1} - v_{j-1}}{2h} = 0, \quad v_j(0) = u(0, jh). \quad (4.22)$$

Вычисляя  $\lambda(\phi)$  по формуле (4.17), получаем

$$\lambda(\phi) = \frac{\mu}{2h} e^{i\phi} - \frac{\mu}{2h} e^{-i\phi} = i \frac{\mu}{h} \sin \phi.$$

Поскольку  $\operatorname{Re} \lambda(\phi) = 0$ , в силу леммы 4.17 схема устойчива<sup>7</sup>. В том, что (4.22) аппроксимирует уравнение переноса, легко убедиться; следовательно, решение будет сходиться к точному.

<sup>7</sup>Установить, что  $\operatorname{Re} \lambda(\phi) = 0$ , можно и не вычисляя  $\lambda(\phi)$ . Для этого достаточно переписать (4.22) в виде

Рассмотрим теперь полностью дискретную схему

$$\frac{v_j^{n+1} - v_j^n}{\tau} + \mu \frac{v_{j+1} - v_{j-1}}{2h} = 0 \quad (4.23)$$

и исследуем её устойчивость по начальным данным. Поскольку она имеет вид  $v_j^{n+1} = Gv_j^n$ , где  $G$  – циркулянт, имеем

$$\|G^n\| = (\max\{|\lambda_k|\})^n,$$

а собственные значения  $\lambda_k$  даны формулой (3.29). При измельчении сетки собственные значения занимают весь контур  $\lambda(\phi)$ , данный (3.30), что при подстановке коэффициентов даёт

$$\lambda(\phi) = 1 - \frac{\mu\tau}{h} \left( \frac{1}{2} \exp(i\phi) - \frac{1}{2} \exp(-i\phi) \right) = 1 - i \frac{\mu\tau}{h} \sin \phi.$$

Максимум модуля достигается при  $\sin \phi = \pm 1$  (что соответствует 4-точечной пиле  $v_j = (\pm i)^j$ ) и равен  $(1 + (\mu\tau/h)^2)^{1/2}$ . Если теперь учесть, что  $n \leq t_{\max}/\tau$ , то получаем

$$\|G^n\| \approx (1 + (\mu\tau/h)^2)^{t_{\max}/(2\tau)} = \exp \left( \frac{t_{\max}}{2\tau} \ln(1 + (\mu\tau/h)^2) \right).$$

Отсюда видно, что при  $\tau = ch$  схема неустойчива. Рассматривая случай  $\tau \ll h$ , имеем

$$\|G^n\| \approx \exp \left( \frac{t_{\max}}{2\tau} (\mu\tau/h)^2 \right) = \exp \left( \frac{\mu^2 t_{\max} \tau}{2h^2} \right).$$

Таким образом, при  $\tau \leq ch^2$ , где  $c > 0$  – любая константа, мы будем иметь оценку на  $\|G^n\|$ , равномерную по  $h$  (хотя и не равномерную по  $t_{\max}$ ). Формально говоря, решение по такой схеме получить можно, но на практике полученное ограничение на шаг по времени будет слишком обременительным. Называть ли схему (4.23) устойчивой при  $\tau \leq ch^2$  – это вопрос исключительно определения понятия “устойчивость”, не имеющий практического смысла.

Как же определить, будет ли схема устойчивой при некотором ограничении  $\tau \leq ch$ ? Для этого нам понадобится вспомнить понятие области абсолютной устойчивости для методов решения ОДУ.

---

(4.16) и заметить, что матрица  $A_h$  кососимметрическая. Поскольку матрица  $iA_h$  эрмитова, её собственные значения действительные. Следовательно, собственные значения  $A_h$  чисто мнимые. Также этот факт следует из того, что  $d\|v\|_2/dt = 0$ .

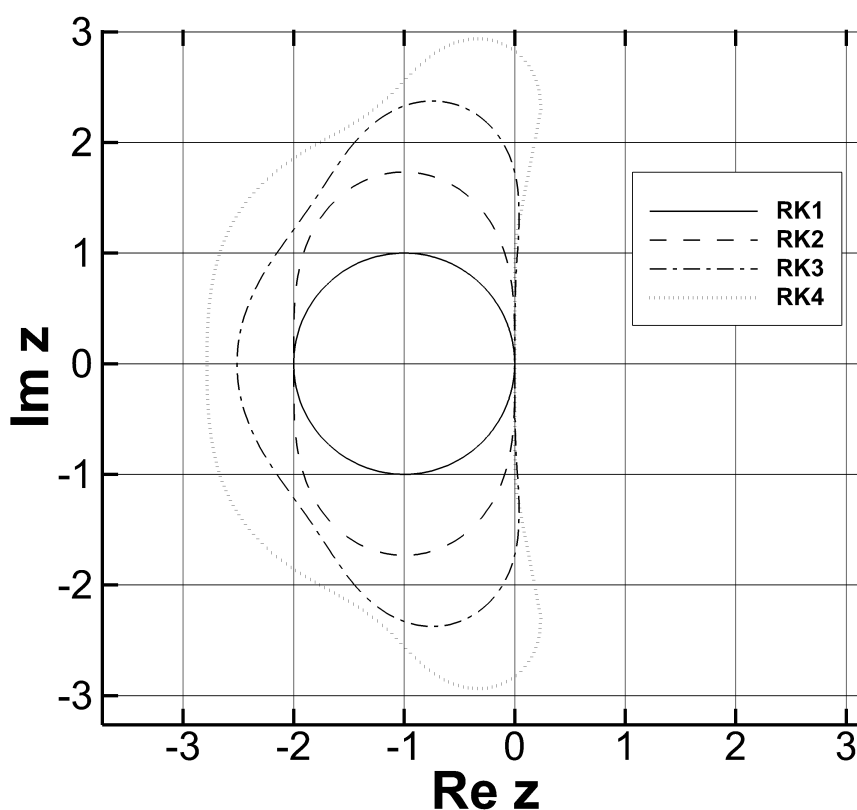


Рис. 2. Области устойчивости явных методов Рунге – Кутты

**Определение 4.** Пусть численный метод для решения задачи Коши

$$\frac{du}{dt} = \lambda u, \quad t \in (0, t_{\max}); \quad u(0) = u_0, \quad (4.24)$$

где  $\lambda \in \mathbb{C}$ , может быть представлен в виде

$$v_0 = u_0; \quad v_{n+1} = R(\lambda\tau)v_n. \quad (4.25)$$

Тогда функция  $R(z)$ ,  $z \in \mathbb{C}$ , называется функцией устойчивости численного метода, а множество  $z \in \mathbb{C}$ , таких что  $|R(z)| \leq 1$ , называется его областью абсолютной устойчивости.

Легко убедиться, что для задачи (4.24) условие  $k$ -го порядка аппроксимации равносильно условию

$$\left. \frac{d^j R(z)}{dz^j} \right|_{z=0} = 1, \quad j = 0, \dots, k.$$

Для  $r$ -стадийного явного метода Рунге – Кутты функция устойчивости является многочленом порядка  $r$ . Области абсолютной устойчивости  $r$ -стадийных явных методов Рунге – Кутты порядка  $r$  изображены на рис. 2.

**Упражнение 8.** Доказать, что для явного метода Рунге – Кутты область абсолютной устойчивости является ограниченным множеством.

**Упражнение 9.** Доказать, что в классе методов

$$\begin{aligned} v_0 &= u_0; \\ k_{n,1} &= f(v_n, t_n); \\ k_{n,2} &= f(v_n + \tau_n \beta_{2,1} k_{n,1}, t_n + \alpha_2 \tau_n); \\ &\dots \\ k_{n,r} &= f(v_n + \tau_n \beta_{r,1} k_{n,1}, t_n + \alpha_r \tau_n); \\ v_{n+1} &= v_n + \tau_n \sum_{s=1}^r \gamma_s k_{n,s} \end{aligned}$$

нет методов, обладающих третьим порядком точности для всех задач Коши вида (4.18) с гладким решением.

Докажем следующее утверждение.

**Теорема 4.8.** Рассмотрим полудискретную схему (4.14)–(4.15) и некоторый явный метод Рунге – Кутты (4.19). Пусть  $\sigma > 0$ . Полностью дискретная схема, полученная их комбинацией, будет устойчива при условии  $\tau = \sigma h/|\mu|$  тогда и только тогда, когда контур  $\lambda_R(\phi) = -(\mu\tau/h)\lambda(\phi)$ , где  $\lambda(\phi)$  задаётся (4.17), целиком лежит в области устойчивости метода Рунге – Кутты.

*Доказательство.* Перепишем полудискретную схему в матричном виде (4.16):

$$\frac{dv}{dt} + \frac{\mu}{h} A_h v(t) = 0, \quad v(0) = \Pi_h u(0).$$

Тогда полностью дискретная схема будет иметь вид

$$v^0 = \Pi_h u(0), \quad v^{n+1} = R\left(-\frac{\mu\tau}{h} A_h\right) v^n$$

(поскольку  $R$  – многочлен, его применение к матрице определено). Матрица  $R(-\mu\tau A_h/h)$  является циркулянтном; её собственные значения лежат на контуре  $\lambda_R(\phi)$  и заполняют этот контур в пределе при  $N_x \rightarrow \infty$ .

По теореме 3.7 полностью дискретная схема устойчива тогда и только тогда, когда  $\sup_\phi |\lambda_R(\phi)| \leq 1$ . А это и есть определение того, что контур  $\lambda_R(\phi)$  целиком лежит в области абсолютной устойчивости.  $\square$

Вернёмся к рассмотрению полудискретной схемы (4.22) и поймём, почему применение к ней явного метода Эйлера дало схему, не являющуюся устойчивой ни при каком ограничении  $\tau \leq ch$ . Если её записать в виде (4.16), то матрица  $A_h$  будет кососимметрической, поэтому все её собственные значения чисто

мнимые. Поскольку область абсолютной устойчивости явного метода Эйлера не содержит никакого отрезка мнимой оси, домножение спектра матрицы  $A_h$  на любую положительную константу не может поместить его внутрь области абсолютной устойчивости.

В то же время, если применить для аппроксимации (4.22), например, 3-стадийный явный метод Рунге – Кутты 3-го порядка или 4-стадийный явный метод Рунге – Кутты 4-го порядка, то схема получится устойчивой при  $\tau \leq ch$  при некотором  $c > 0$ .

Возникает вопрос, а для каких методов Рунге – Кутты область абсолютной устойчивости содержит некоторый отрезок с концами  $0, iy$ , где  $y > 0$ . Для линейных автономных (т. е. в которые явно не входит время) уравнений для каждого  $r \in \mathbb{N}$  существует  $r$ -стадийный явный метод  $r$ -го порядка. Для таких методов ответ даётся следующим утверждением.

**Лемма 4.9.** *Рассмотрим  $r$ -стадийный явный метод Рунге – Кутты, имеющий порядок аппроксимации  $r$  для задачи (4.24). Если  $r = 4k + 3$  или  $r = 4k$ , то область абсолютной устойчивости содержит отрезок, соединяющий  $z = 0$  с некоторой точкой мнимой оси; если  $r = 4k + 1$  или  $r = 4k + 2$ , то никакого отрезка этого вида область абсолютной устойчивости не содержит.*

*Доказательство.* Функция устойчивости рассматриваемого метода имеет вид

$$R(z) = \sum_{k=0}^r \frac{r^k}{k!}.$$

Рассмотрим выражение  $|R(it)|^2$  при  $t \in \mathbb{R}$ :

$$|R(it)|^2 = R(it)\overline{R(it)} = \sum_{k=0}^r \sum_{l=0}^r \frac{(it)^k}{k!} \frac{(-it)^l}{l!} = \sum_{k=0}^r \sum_{l=0}^r \frac{t^{k+l}}{k! l!} i^{k-l}.$$

Выполняя замену переменных  $s = k + l$ , получаем

$$|R(it)|^2 = \sum_{s=0}^{2r} \frac{t^s}{s!} \sum_{l=0}^{\min\{s,r\}} C_s^l (-1)^l i^s,$$

где  $C_s^l$  – биномиальные коэффициенты. Слагаемые при  $0 < s \leq r$  равны нулю: действительно,

$$\sum_{l=0}^s C_s^l (-1)^l = (1 - 1)^s = 0. \quad (4.26)$$

Также можно вычеркнуть все нечётные значения  $s$ , поскольку выражение  $|R(it)|^2$  действительное при любом  $t$ . Таким образом, полагая  $s = 2q$ ,

$$|R(it)|^2 = 1 + \sum_{q=\lceil r/2 \rceil}^r \frac{t^{2q}}{(2q)!} \sum_{l=0}^r C_{2q}^l (-1)^{l+q},$$

где  $\lceil \dots \rceil$  есть взятие целой части с округлением вверх.

Наличие или отсутствие отрезка мнимой оси в области абсолютной устойчивости определяется знаком старшей производной от  $|R(it)|^2$  по  $t$ : если она отрицательная, такой отрезок есть, а если положительная – то такого отрезка нет. Эта производная равна

$$Y_{2q} = \sum_{l=0}^r C_{2q}^l (-1)^{l+q}. \quad (4.27)$$

При нечётном  $r$  подставим  $q = \lceil r/2 \rceil$ . Значит,  $r = 2q - 1$ , поэтому запишем

$$Y_{2q} = \sum_{l=0}^{2q-1} C_{2q}^l (-1)^{l+q} = (-1)^q \sum_{l=0}^{2q} C_{2q}^l (-1)^l - C_{2q}^{2q} (-1)^{3q} = (-1)^{q+1}.$$

Первое равенство является подстановкой  $r = 2q - 1$  в (4.27), для записи второго мы добавили и вычли член суммы при  $l = 2q$ , в последнем мы использовали (4.26). Таким образом, при чётных  $q$ , что соответствует  $r = 4k + 3$ , имеем  $Y_{2q} < 0$ , а при нечётных  $q$ , что соответствует  $r = 4k + 1$ , имеем  $Y_{2q} > 0$ .

При чётном  $r$ , подставляя  $q = \lceil r/2 \rceil$ , имеем  $r = 2q$ , в силу (4.26) получим, что производная равна нулю:

$$Y_{2q} = \sum_{l=0}^{2q} C_{2q}^l (-1)^{l+q} = 0.$$

Равным образом равна нулю и  $Y_{2\lceil r/2 \rceil + 1}$ , потому что в сумме есть только чётные степени. Значит, нужно смотреть на величину  $Y_{2q}$ , где  $q = \lceil r/2 \rceil + 1$ . Подставляя  $r = 2q - 2$  и действуя аналогично случаю нечётных  $r$ , получаем

$$\begin{aligned} Y_{2q} &= \sum_{l=0}^{2q-2} C_{2q}^l (-1)^{l+q} = (-1)^q \sum_{l=0}^{2q} C_{2q}^l (-1)^l - C_{2q}^{2q-1} (-1)^{3q-1} - C_{2q}^{2q} (-1)^{3q} = \\ &= -2q(-1)^{q-1} - (-1)^q = (r+1)(-1)^q. \end{aligned}$$

При чётных  $q$ , что соответствует  $r = 4k + 2$ , имеем  $Y_{2q} > 0$ , а при нечётных  $q$ , что соответствует  $r = 4k$ , имеем  $Y_{2q} < 0$ .  $\square$