

ISSN 0044-4669

Том 63, Номер 1

Январь 2023



# ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

[www.sciencejournals.ru](http://www.sciencejournals.ru)



---

# СОДЕРЖАНИЕ

---

Том 63, номер 1, 2023 год

---

## АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ

Вещественная нормальная форма бинарного многочлена  
в критической точке второго порядка

*А. Б. Батхин, А. Д. Брюно*

3

## ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

Функциональное суммирование рядов

*В. П. Варин*

16

Аппарат геометрической алгебры и кватернионов в системах символьных  
вычислений для описания вращений в евклидовом пространстве

*Т. Р. Велиева, М. Н. Геворкян, А. В. Демидова, А. В. Королькова, Д. С. Кулябов*

31

О структуре решений ключевого уравнения Госпера в задачах  
символьного суммирования

*Е. В. Зима*

43

Обобщение задачи о сумме подмножеств и кубические формы

*А. В. Селиверстов*

51

## ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

БАД-методология и дифференцирование сложной функции

*А. Ф. Албу, А. Ю. Горчаков, В. И. Zubov*

61

Оптимальное граничное управление распределенной неоднородной  
колебательной системой с заданными состояниями в промежуточные  
моменты времени

*В. Р. Барсегян*

74

## ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Контрпримеры к предположению о возможности продолжения усеченных решений  
усеченного линейного обыкновенного дифференциального уравнения

*С. А. Абрамов, А. А. Рябенко, Д. Е. Хмельнов*

85

Аналитическое исследование кубатурных формул на сфере в системах  
компьютерной алгебры

*Р. Э. Байрамов, Ю. А. Блинков, И. В. Левичев, М. Д. Малых, В. С. Мележик*

93

## УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Сходимость формальных решений второго члена четвертой иерархии Пенлеве  
в окрестности нуля

*В. И. Аношин, А. Д. Бекетова, А. В. Парусникова, Е. Д. Прокопенко*

102

Символьно-численное моделирование распространения адиабатической  
волноводной моды в плавном волноводном переходе

*Д. В. Диваков, А. А. Тютюнник*

112

О задачах коши для нелинейных соболевских уравнений теории сегнетоэлектричества

*М. О. Корпусов, Р. С. Шафир*

123

---

## МАТЕМАТИЧЕСКАЯ ФИЗИКА

Исследование динамики системы двух связанных тел в плоскости круговой орбиты  
с применением методов компьютерной алгебры

*С. А. Гутник, В. А. Сарычев*

145

Исследование вековых возмущений в ограниченной задаче трех тел переменной массы  
с применением компьютерной алгебры

*А. Т. Ибраимова, М. Дж. Минглибаев, А. Н. Прокопеня*

154

---

## ИНФОРМАТИКА

Обучение порт-гамильтоновым системам: алгоритмы

*Д. Лозиенко, В. Сальников, А. Фалез*

165

Квантовая марковская динамика после времени корреляции резервуара

*А. Е. Теретёнков*

175

---

---

# АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ

УДК 519.16

## ВЕЩЕСТВЕННАЯ НОРМАЛЬНАЯ ФОРМА БИНАРНОГО МНОГОЧЛЕНА В КРИТИЧЕСКОЙ ТОЧКЕ ВТОРОГО ПОРЯДКА

© 2023 г. А. Б. Батхин<sup>1,2,\*</sup>, А. Д. Брюно<sup>2,\*\*</sup><sup>1</sup> 141701, Долгопрудный, МО, Институтский переулок, 9, МФТИ, Россия<sup>2</sup> 125047 Москва, Миусская пл., 4, ИПМ им. М.В. Келдыша РАН, Россия

\*e-mail: batkhin@gmail.com

\*\*e-mail: abruno@keldysh.ru

Поступила в редакцию 25.04.2022 г.

Переработанный вариант 25.04.2022 г.

Принята к публикации 17.09.2022 г.

Рассматривается вещественный многочлен от двух переменных. Его разложения вблизи нулевой критической точки начинаются с формы третьей степени. Находятся его простейшие формы, к которым приводится этот многочлен с помощью обратимых вещественных локальных аналитических замен координат. Сначала для кубической формы с помощью линейных замен координат получены нормальные формы. Их оказалось три. Затем для полного многочлена получены три нелинейные нормальные формы. Предложено упрощение вычисления нормальной формы. Рассмотрен содержательный пример. Библ. 14. Фиг. 7.

**Ключевые слова:** кубическая форма, замена координат, нормальная форма, нелинейная нормализация.

DOI: 10.31857/S0044466923010064, EDN: LDMYKU

### 1. ВВЕДЕНИЕ

В окрестности нуля  $x = y = 0$  рассматривается вещественный многочлен

$$f(x, y) = \sum_{p, q} f_{pq} x^p y^q, \quad (1)$$

где  $0 \leq p, q \in \mathbb{Z}$ ,  $p + q \geq 3$  и  $f_{pq}$  — вещественные коэффициенты. Ищется его простейший вид в классе вещественных аналитических обратимых замен координат

$$x = \xi(w, z), \quad y = \eta(w, z), \quad (2)$$

где  $\xi(0, 0) = \eta(0, 0) = 0$ . Эта задача разбита на две.

В разд. 2 рассматривается задача упрощения вещественной кубической формы

$$f_3(x, y) = f_{30}x^3 + f_{21}x^2y + f_{12}xy^2 + f_{03}y^3 \neq 0 \quad (3)$$

с помощью обратимых линейных вещественных замен координат

$$x = Au + Bv, \quad y = Cu + Dv, \quad (4)$$

где обратимость означает, что определитель

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = AD - BC \neq 0. \quad (5)$$

**Теорема 1.** Форма (3) с помощью замен (4) всегда приводится к одной из трех нормальных форм:

$$1. \quad f_3 = u^3 + v^3; \quad (6a)$$

$$2. \quad f_3 = u^3; \quad (6b)$$

$$3. \quad f_3 = u^2v. \quad (6b)$$

В разд. 3 рассматривается задача упрощения вещественного многочлена

$$g(u, v) = \sum_{p+q \geq 3} g_{pq} u^p v^q \quad (7)$$

с начальной кубической формой теоремы 1 в классе нелинейных вещественных аналитических замен координат

$$u = A_1 w + \xi_1(w, z), \quad v = D_1 z + \eta_1(w, z), \quad (8)$$

где  $A_1$  и  $D_1$  — ненулевые постоянные и ряды  $\xi_1$ ,  $\eta_1$  начинаются с квадратичных членов.

**Теорема 2.** Многочлен (7) с помощью замен (8) всегда приводится к одной из трех нормальных форм:

1.  $g = w^3 + z^3$ ;
2.  $g = w^3 + bz^4$ ,  $b = \pm 1$ , если в форме четвертого порядка  $g_4(u, v)$  коэффициент  $g_{04} \neq 0$ ;
3.  $g = w^2 z + \tau z^m$ ,  $m > 3$ ,  $\tau \in \{-1, 0, 1\}$ .

В разд. 4 доказана

**Теорема 3.** Если многочлен  $f(u, v)$  имеет укороченный многочлен  $\hat{f}(u, v)$  одного из трех видов

1.  $\hat{f} = \alpha u^3 + \beta v^3$ ,
2.  $\hat{f} = \alpha u^3 + \beta v^4$ ,
3.  $\hat{f} = \alpha u^2 v + \beta v^m$ ,

где числа  $\alpha, \beta \neq 0$ ,  $m > 3$ , то его нормальная форма относится к случаям 1, 2, 3 теоремы 2 соответственно и получается из укороченного многочлена  $\hat{f}(x, y)$  линейной диагональной заменой  $u = Aw$ ,  $v = Dz$ , где

- $A = \sqrt[3]{\alpha}$ ,  $D = \sqrt[3]{\beta}$  для вида 1;
- $A = \sqrt[3]{\alpha}$ ,  $D = \sqrt[4]{|\beta|}$  для вида 2;
- $A = \sqrt[2m]{|\beta|/\sqrt{|\alpha|}}$ ,  $D = \text{sign } \alpha/\sqrt[2m]{|\beta|}$  для вида 3.

Определение укороченного многочлена  $\hat{f}$  см. в [1].

В разд. 5 рассматривается пример вычисления нормальной формы некоторого многочлена вблизи его особой точки второго порядка.

Отметим, что нормальные формы бинарного многочлена находят различные применения. Например, для классификации обобщенных алгебр Клиффорда, связанных с бинарными формами степени 3 [2], используется нормальная форма (6а) теоремы 1. Нормальные формы комплексных многочленов (1) в классе комплексных обратимых замен координат (2) рассмотрены в [3].

## 2. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1

Будем различать три случая.

1. В форме (3) хотя бы один из коэффициентов  $f_{30}$  и  $f_{03}$  отличен от нуля.
2. В форме (3) коэффициенты  $f_{30} = f_{03} = 0$ , но  $f_{21}$  и  $f_{12}$  отличны от нуля.
3. В форме (3) коэффициенты  $f_{30} = f_{12} = f_{03} = 0$ , но  $f_{21} \neq 0$ .

### 2.1. Случай 1

Если  $f_{30} \equiv 0$ , а  $f_{03} \neq 0$ , то  $u$  и  $v$  поменяем местами. Это линейное преобразование. Пусть  $f_{30} \neq 0$ .

Положим  $x_1 = \sqrt[3]{f_{30}}x$ , тогда

$$f_3 = x_1^3 + a_1 x_1^2 y + a_2 x_1 y^2 + a_3 y^3, \quad (9)$$

где коэффициенты  $a_j$ ,  $j = 1, 2, 3$ , выражаются через коэффициенты формы (3).

Положим  $x_1 = x_2 - \frac{1}{3}a_1y$ , тогда

$$f_3 = x_2^3 + b_2x_2y^2 + b_3y^3,$$

где коэффициенты  $b_2, b_3$  полиномиально выражаются через коэффициенты  $a_j, j = 1, 2, 3$ .

Если  $b_2 = 0$ , то при  $b_3 = 0$  имеем нормальную форму 2 (6б) из теоремы 1, а при  $b_3 \neq 0$  полагаем

$$u = x_2, \quad v = \sqrt[3]{b_3}y,$$

и получаем нормальную форму 1 (6а) теоремы 1.

Если  $b_2 \neq 0$ , то, полагая  $y_1 = \sqrt{|b_2|}y$ , получаем форму

$$f_3 = x_2^3 + \sigma x_2 y_1^2 + \gamma y_1^3, \quad \sigma = \pm 1, \quad (10)$$

где  $\sigma = \text{sign } b_2$  и  $\gamma = b_3/|b_2|^{3/2}$  – любое вещественное число.

Будем упрощать форму (10) с помощью линейных замен

$$x_2 = u + Bv, \quad y_1 = Cu + v. \quad (11)$$

Условие (5) здесь означает, что

$$BC \neq 1. \quad (12)$$

После замены (11) форма (10) принимает вид

$$f_3 = (\gamma C^3 + \sigma C^2 + 1)u^3 + (\sigma BC^2 + 3\gamma C^2 + 2\sigma C + 3B)u^2v + \\ + (2\sigma BC + 3B^2 + 3\gamma C + \sigma)uv^2 + (B^3 + \sigma B + \gamma)v^3. \quad (13)$$

Коэффициенты при мономах  $u^2v$  и  $uv^2$  аннулируются, если выполнены равенства

$$\sigma BC^2 + 3\gamma C^2 + 2\sigma C + 3B = 0, \quad (14)$$

$$2\sigma BC + 3B^2 + 3\gamma C + \sigma = 0. \quad (15)$$

Из уравнения (14) получаем

$$B = -\frac{3\gamma C^2 + 2\sigma C}{\sigma C^2 + 3}. \quad (16)$$

Подставив это выражение в уравнение (15), умножив его на  $(\sigma C^2 + 3)^2$ , приведя подобные члены и сократив на 3, получаем уравнение

$$-\gamma C^5 + (9\gamma^2 - \sigma)C^4 + 12\gamma\sigma C^3 + 2C^2 + 9C\gamma + 3\sigma = 0. \quad (17)$$

Этот многочлен распадается на множители

$$-(C^2 - 9\gamma C - 3\sigma)(\gamma C^3 + \sigma C^2 + 1). \quad (18)$$

Выделим тот исключительный случай, когда условие (12) не выполнено. Согласно (16) равенство  $BC = 1$  эквивалентно уравнению

$$2\sigma C^2 + 3\gamma C^3 + 3 + \sigma C^2 = 0,$$

т.е.

$$\gamma C^3 + \sigma C^2 + 1 = 0. \quad (19)$$

Поэтому второй множитель в (18) должен быть отличным от нуля, и уравнение (17) сводится к уравнению

$$C^2 - 9\gamma C - 3\sigma = 0. \quad (20)$$

Согласно (19) в исключительном случае имеем

$$\gamma = -\frac{\sigma C^2 + 1}{C^3}. \quad (21)$$

Подставляя это значение в уравнение (20), умножая его на  $C^2$  и приводя подобные члены, получаем уравнение

$$(C^2 + 3\sigma)^2 = 0, \quad \text{т.е.} \quad C^2 = -3\sigma.$$

Вещественное значение  $C$  возможно только при  $\sigma = -1$ , тогда  $C = \pm\sqrt{3}$ . Из (21) получаем  $\gamma = \pm\frac{2}{3\sqrt{3}}$ .

Итак, всегда, кроме исключительного случая

$$\sigma = -1, \quad \gamma = \pm\frac{2}{3\sqrt{3}}, \quad (22)$$

существует невырожденная замена (11), приводящая форму (10) к виду

$$f_3 = \alpha u^3 + \beta v^3, \quad (23)$$

где постоянные  $\alpha$  и  $\beta$  отличны от нуля. Из формулы (13) имеем

$$\alpha = \gamma C^3 + \sigma C^2 + 1, \quad \beta = B^3 + \sigma B + \gamma.$$

Согласно (19)  $\alpha = 0$  только в исключительном случае.

Аналогично, из уравнения (15) получаем

$$C = -\frac{3B^2 + \sigma}{2\sigma B + 3\gamma}.$$

Вырожденность  $BC = 1$  означает, что

$$\sigma B + 3B^3 + 3\gamma + 2\sigma B = 3(B^3 + \sigma B + \gamma) = 0.$$

Поскольку мы рассматриваем невырожденный случай, то  $\beta \neq 0$ . Замена

$$u_1 = \sqrt[3]{\alpha}u, \quad v_1 = \sqrt[3]{\beta}v$$

приводит форму (23) к нормальной форме 1 (6а) теоремы 1.

Согласно (22) форма (10) в исключительном случае раскладывается на множители

$$x_2^3 - x_2 y_1^2 \pm \frac{2}{3\sqrt{3}} y_1^3 = \left(x_2 \mp \frac{1}{\sqrt{3}} y_1\right)^2 \left(x_2 \pm \frac{2}{\sqrt{3}} y_1\right).$$

Поэтому, полагая

$$u = x_2 \mp \frac{1}{\sqrt{3}} y_1, \quad v = x_2 \pm \frac{2}{\sqrt{3}} y_1,$$

получаем нормальную форму 3 (6в) теоремы 1.

Итак, для случая 1 теорема 1 доказана.

## 2.2. Случай 2

В этом случае

$$f_3 = f_{21} x^2 y + f_{12} x y^2, \quad (24)$$

где оба коэффициента  $f_{21}$  и  $f_{12}$  отличны от нуля. Преобразование

$$x = \sqrt[3]{\frac{f_{12}}{f_{21}^2}} x_1, \quad y = \sqrt[3]{\frac{f_{21}}{f_{12}^2}} y_1$$

приводит форму (24) к виду

$$f_3 = x_1^2 y_1 + x_1 y_1^2.$$

Выполняем замену  $x_1 = x_2 - \frac{1}{2}y_1$  и получаем форму  $f_3 = x_2^2 y_1 - \frac{1}{4}y_1^3$ , которая после преобразования

$$x_2 = \frac{Y}{\sqrt[3]{2}}, \quad y_1 = -\sqrt[3]{4}X$$

приобретает вид

$$f_3 = X^3 - XY^2.$$

Но в п. 2.1 было показано, что эта форма эквивалентна нормальной форме 1 (6а) теоремы 1.

### 2.3. Случай 3

В этом случае

$$f_3 = f_{21}x^2y.$$

Линейной заменой  $u = x$ ,  $v = f_{21}y$  приводим эту форму к виду  $f_3 = u^2v$ , т.е. к нормальной форме 3 (6в) теоремы 1.

### 2.4. Неэквивалентность нормальных форм 1, 2, 3 теоремы 1

Покажем, что нормальная форма 1 (6а) и нормальная форма 2 (6б) не эквивалентны между собой.

Форма 2 есть  $f_3 = x^3$ . После линейной замены

$$x = Au + Bv, \quad y = Cu + Dv \quad (25)$$

она принимает вид

$$f_3 = A^3u^3 + 3A^2Bu^2v + 3AB^2uv^2 + B^3v^3. \quad (26)$$

Она имеет вид формы 1 (6а), только если  $A^2B = AB^2 = 0$ . Но тогда либо  $A = 0$ , либо  $B = 0$ , либо  $A = B = 0$ . В первых двух случаях форма (26) является формой вида 2 (6б). В третьем случае замена (25) вырождена.

Покажем теперь, что нормальная форма 3 (6в) не эквивалентна нормальной форме 1 (6а) и нормальной форме 2 (6б).

Форма  $f_3 = x^2y$  после замены (25) принимает вид

$$f_3 = A^2Cu^3 + (A^2D + 2ABC)u^2v + (2ABD + B^2C)uv^2 + B^2Dv^3.$$

Она имеет вид форм 1 и 2, только если

$$A^2D + 2ABC = A(AD + 2BC) = 0,$$

$$2ABD + B^2C = B(2AD + BC) = 0.$$

Эта система уравнений имеет 4 решения:

1.  $A = B = 0$ ;

2.  $A = 2AD + BC = 0$ , т.е.  $A = BC = 0$ ;

3.  $B = AD + 2BC = 0$ , т.е.  $B = AD = 0$ ;

4.  $AD + 2BC = 2AD + BC = 0$ , т.е.  $AD = BC = 0$ , ибо определитель этой однородной системы равен  $-3$ , т.е. отличен от нуля.

Для всех этих решений замена (25) вырождена.

Теорема 1 доказана.

## 2.5. Замечания к доказательству теоремы 1

**Замечание 1.** Доказательство теоремы 1 можно провести, используя современный вариант теории исключений, опирающийся на вычисление базисов Грёбнера соответствующих идеалов [4, Гл. 2, 3].

Для приведения формы (10) к виду (6а) с помощью преобразования (11) составляется идеал  $\mathcal{F}_1$  из многочленов (14) и (15). Затем для идеала  $\mathcal{F}_1$  вычисляется его базис Грёбнера с лексикографическим порядком переменных  $C \prec B$ . Величины  $\sigma$  и  $\gamma$  в этом случае выступают в качестве параметров. Первым многочленом этого базиса (с точностью до знака) является полином (17), который раскладывается на множители (18).

Если задать другой лексикографический порядок  $B \prec C$  и вычислить новый базис Грёбнера, то его первым многочленом будет полином, раскладывающийся на множители

$$(3\sigma B^2 + 9\gamma B - \sigma^2)(B^3 + \sigma B + \gamma). \quad (27)$$

Наличие кубических членов в формулах (18) и (27) гарантирует существование по крайней мере одного вещественного решения системы (14), (15). На самом деле, для всех значений параметров  $\sigma$ ,  $\gamma$ , кроме исключительного случая (22), имеется три набора решений указанной системы.

В этом легко убедиться, если вычислить дискриминанты  $D_C^{(1)} = D_B^{(1)} = 3(4\sigma + 27\gamma^2)$  квадратичных множителей формул (18) и (27) и дискриминанты кубических множителей этих же формул:  $D_C^{(2)} = D_B^{(2)} = -(4\sigma + 27\gamma^2)$ . Разные знаки этих выражений обеспечивают наличие двух вещественных корней квадратичных множителей и одного вещественного корня кубических при  $D_C^{(2)} < 0 < D_C^{(1)}$ , либо трех вещественных корней кубических множителей при  $D_C^{(1)} < 0 < D_C^{(2)}$ .

**Замечание 2.** Для преобразования многочлена (9) к нормальной форме 1 (6а) можно воспользоваться методом, который называется *преобразование Чирнгауза* (см. [5, Гл. 4, п. 19.1], [6, Ч. 1, п. 6.2]). Его идея заключается в том, что исходный многочлен преобразуется к многочлену  $F(y)$  наиболее простого вида (обычно  $F(y) = y^3 - c$ ) с помощью вспомогательного многочлена  $h(x)$ .

Рассмотрим, например, как преобразовать форму (10) к виду (6а) с помощью вспомогательного квадратного многочлена  $h = X^2 + c_1X + c_2$ , где  $c_1, c_2$  — некоторые неизвестные величины. Сначала с помощью подстановки  $x_2 = X/y_1$  перейдем к многочлену  $\tilde{f}_3 = X^3 + \sigma X + \gamma$  от одной переменной  $X$ . Многочлен  $F(Y)$  — результат преобразования Чирнгауза — можно вычислить одним из двух способов:

- либо как результат  $\text{Res}_X(\tilde{f}_3(X), y - h(X))$  многочленов  $\tilde{f}_3(X)$  и  $h(X)$  относительно переменной  $X$ ;
- либо как характеристический многочлен матрицы  $\mathfrak{B}$  — аналога матрицы Безу [6], [7], составленной из коэффициентов остатков от деления  $X^k h(X)$  на  $\tilde{f}_3(X)$ .

Вычисления удобнее выполнять в любой из систем компьютерной алгебры (СКА). Авторы использовали СКА Maple [8], [9], хотя здесь подойдет любая СКА, например, бесплатная SymPy [10]. Применяя один из указанных выше способов, получаем многочлен

$$F(Y) = Y^3 + (2\sigma - 3c_2)Y^2 + (\sigma c_1^2 + 3\gamma c_1 - 4\sigma c_2 + 3c_2^2 + 1)Y + \gamma c_1^3 - \sigma c_1^2 c_2 + \gamma \sigma c_1 - 3\gamma c_1 c_2 + 2\sigma c_2^2 - c_2^3 - \gamma^2 - c_2. \quad (28)$$

Подберем теперь коэффициенты  $c_1$  и  $c_2$  многочлена  $h(X)$  так, чтобы коэффициенты при линейном и квадратном членах многочлена  $F(Y)$  были равны нулю:

$$2\sigma - 3c_2 = 0, \quad \sigma c_1^2 + 3\gamma c_1 - 4\sigma c_2 + 3c_2^2 + 1 = 0.$$

Базис Грёбнера этого идеала имеет вид

$$\mathcal{GB} = \{3\sigma c_1^2 + 9\gamma c_1 - 1, 3c_2 - 2\sigma\}. \quad (29)$$

Очевидно, что вещественные корни идеала  $\mathcal{GBJ}$  существуют при условии, что дискриминант  $D^{(1)} = 3(27\gamma^2 + 4\sigma)$  его первого многочлена неотрицателен, что согласуется с выводами замечания 1.

Упрощая свободный член многочлена (28) по модулю базиса Грёбнера  $\mathcal{GBJ}$ , получаем для него выражение

$$\delta = \frac{1}{81} D^{(1)} (9\gamma c_1^* - 2\sigma),$$

где  $c_1^*$  — один из корней первого многочлена базиса (29).

Итак, при условии  $D^{(1)} \geq 0$  форма (10) приводится к форме  $u^3 + \delta v^3$ , которая соответствует либо нормальной форме 1 (6а) при  $\delta \neq 0$ , либо нормальной форме 2 (6б) теоремы 1 при  $\delta = 0$ .

### 3. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2

Здесь рассмотрим многочлен (7) последовательно по случаям 1, 2, 3 на стр. 4.

$$3.1. \text{ Случай } g_3 = u^3 + v^3.$$

Сделаем замену

$$u = u_1 - \frac{1}{3} g_{22} v^2.$$

Тогда получим

$$\begin{aligned} u^3 &= u_1^3 - g_{22} u_1^2 v^2 + 3g_{22} u_1 v^4 - \frac{1}{27} g_{22}^3 v^6, \\ g_{22} u^2 v^2 &= g_{22} u_1^2 v^2 - \frac{2}{3} g_{22}^2 u_1 v^4 + \frac{1}{9} g_{22}^3 v^6. \end{aligned}$$

Поэтому в новых координатах  $u_1, v$

$$g = u_1^3 + v^3 + \sum_{p+q \geq 4} \tilde{g}_{pq} u_1^p v^q, \quad (30)$$

где коэффициент  $\tilde{g}_{22} = 0$ .

Разобьем сумму (30) на две:

$$g = g^* + g^{**}, \quad (31)$$

где

$$\begin{aligned} g^* &= u_1^3 [1 + \sum_{\substack{p \geq 3 \\ q \geq 0}} \tilde{g}_{pq} u_1^{p-3} v^q] \stackrel{\text{def}}{=} u_1^3 [1 + \alpha(u_1, v)], \\ g^{**} &= v^3 [1 + \sum_{\substack{0 \leq p \leq 2 \\ q \geq 3}} \tilde{g}_{pq} u_1^p v^{q-3}] \stackrel{\text{def}}{=} v^3 [1 + \beta(u_1, v)], \end{aligned}$$

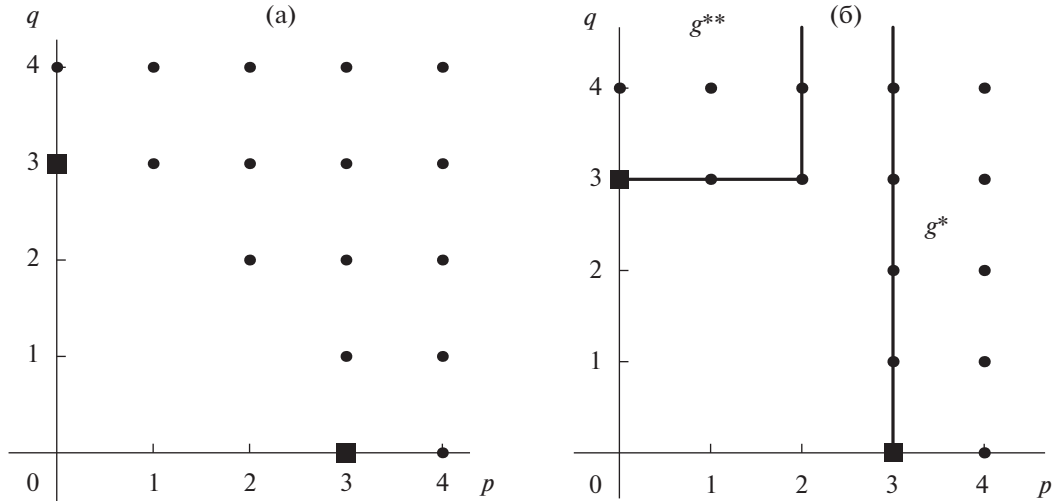
а величины  $\alpha(u_1, v)$  и  $\beta(u_1, v)$  — это многочлены без свободных членов. Поэтому замена

$$w = u_1 \sqrt[3]{1 + \alpha(u_1, v)}, \quad z = v \sqrt[3]{1 + \beta(u_1, v)}$$

аналитична. Получаем нормальную форму 1 теоремы 2.

Для многочлена (7) множество точек  $(p, q)$ , для которых  $g_{pq} \neq 0$ , называется его *носителем*.

На фиг. 1 слева показан носитель исходного многочлена (7), а справа — преобразованного многочлена (30). При этом квадратиками выделены точки носителя нормальной формы (6а) и отчеркнуты носители для сумм  $g^*$  и  $g^{**}$ .



Фиг. 1. Носители исходного (7) (а) и преобразованного (30) (б) многочленов. Случай 1.

### 3.2. Случай $g_3 = u^3$

Пусть в разложении (7) коэффициент  $g_{04} \neq 0$ . Это и есть условие невырожденности формы  $g_4(u, v)$ . Можно считать, что  $g_{04} = \pm 1 \stackrel{\text{def}}{=} b$ .

Теперь аннулируем коэффициенты  $g_{13}$ ,  $g_{22}$  и  $g_{23}$  с помощью последовательных замен. Вначале выполняем подстановку  $v = v_1 - \frac{g_{13}}{4b}u$  и получаем новые коэффициенты  $\tilde{g}_{jk}$  разложения (7). Затем делаем подстановку  $u = u_1 - \frac{1}{3}\tilde{g}_{22}v_1^2 - \frac{1}{3}\tilde{g}_{23}v_1^3$ . Тогда полученная сумма

$$g(u_1, v_1) = u_1^3 + bv_1^4 + \sum \tilde{g}_{pq}u_1^p v_1^q \quad (32)$$

разбивается на две части (31), где

$$g^* = u_1^3 \left[ 1 + \sum_{\substack{p \geq 3 \\ q \geq 0}} \tilde{g}_{pq} u_1^{p-3} v_1^q \right] \stackrel{\text{def}}{=} u_1^3 [1 + \alpha(u_1, v_1)],$$

$$g^{**} = bv_1^4 \left[ 1 + \sum_{\substack{0 \leq p \leq 2 \\ q \geq 4}} \tilde{g}_{pq} u_1^p v_1^{q-4} \right] \stackrel{\text{def}}{=} bv_1^3 [1 + \beta(u_1, v_1)],$$

где  $\alpha(u_1, v_1)$  и  $\beta(u_1, v_1)$  — это многочлены без свободных членов. Поэтому замена

$$w = u_1 \sqrt[3]{1 + \alpha(u_1, v_1)}, \quad z = v_1 \sqrt[4]{1 + \beta(u_1, v_1)}$$

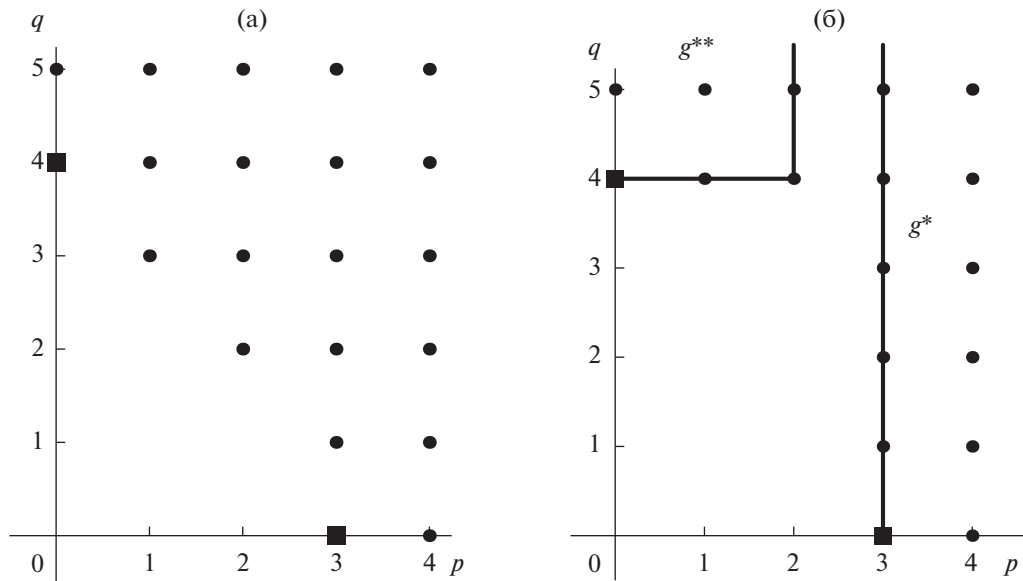
аналитична и приводит многочлен к нормальной форме 2 теоремы 2.

На фиг. 2 показаны носители исходного (7) (а) и преобразованного (32) (б) многочленов. Точки главных членов обозначены квадратами.

### 3.3. Случай $g_3 = u^2v$

Согласно [1] многоугольник Ньютона многочлена  $g$  имеет вершину  $(p, q) = (2, 1)$ , а многочлен  $g$  имеет корень в виде сходящегося степенного ряда

$$v = \sum_{k=2}^{\infty} \varphi_k u^k \stackrel{\text{def}}{=} \varphi(u).$$



Фиг. 2. Носители исходного (7) (а) и преобразованного (32) (б) многочленов. Случай 2.

Сделаем замену  $v = v_1 - \varphi(u)$ . Тогда  $g(u, v_1)$  делится на  $v_1$ . Теперь по теореме о неявной функции уравнение  $\frac{\partial g(u, v_1)}{\partial u} = 0$  имеет решение в виде сходящегося степенного ряда

$$u = \psi(v_1) \stackrel{\text{def}}{=} \sum_{k \geq 2}^{\infty} \psi_k v_1^k.$$

После замены  $u = u_1 - \psi(v_1)$  получаем

$$g(u_1, v_1) = u_1^2 v_1 + \sum \tilde{g}_{pq} u_1^p v_1^q, \quad (33)$$

где все  $\tilde{g}_{p0} = 0$  и все  $\tilde{g}_{1q} = 0$ . Поэтому сумма  $g$  разбивается на две части (31), где

$$g^* = u_1^2 v_1 [1 + \sum \tilde{g}_{pl} u_1^{p-2} v_1^{l-1}] \stackrel{\text{def}}{=} u_1^2 v_1 [1 + \alpha(u_1, v_1)], \quad (34)$$

$$g^{**} = \beta(v_1) \stackrel{\text{def}}{=} \sum_{k=m}^{\infty} \beta_k v_1^k, \quad m > 3,$$

где  $\alpha(u_1, v_1)$  и  $\beta(v_1)$  суть сходящиеся степенные ряды без свободных членов, причем ряд  $\beta(v_1)$  начинается с некоторого члена  $\beta_m v_1^m$ , либо  $\beta(v_1) \equiv 0$ .

Если  $\beta \not\equiv 0$ , то замена  $z = \sqrt[m]{\beta(v_1)/\text{sign } \beta_m}$  приводит ряды (34) к виду

$$g^* = \frac{1}{\sqrt[m]{|\beta_m|}} u_1^2 z [1 + \tilde{\alpha}(u_1, z)],$$

$$g^{**} = \tau z^m,$$

где  $\tau = \text{sign } \beta_m = \pm 1$ . Если  $\beta \equiv 0$ , то  $\tau = 0$ .

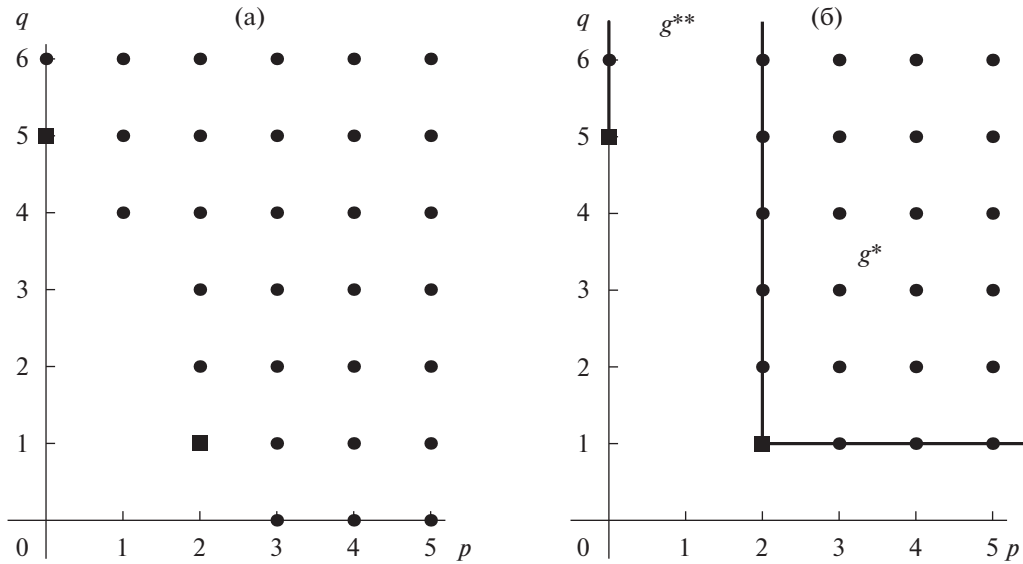
Теперь замена

$$w = u_1 \sqrt{\frac{1 + \tilde{\alpha}}{\sqrt[m]{|\beta_m|}}}$$

приводит многочлен к нормальной форме 3 теоремы 2.

На фиг. 3 показаны носители исходного (7) и преобразованного (33) многочленов.

Доказательство окончено.



Фиг. 3. Носители исходного (7) (а) и преобразованного (33) (б) многочленов. Случай 3.

#### 4. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 3

В ситуациях теоремы 3 нелинейные преобразования координат в доказательстве теоремы 2 не меняют укороченного многочлена  $\hat{f}$  и приводят исходный многочлен к укороченному. Выписанные в формулировке теоремы 3 линейные замены приводят укороченные многочлены к нормальным формам теоремы 2.

Теорема доказана.

#### 5. ПРИМЕР

Вычислим нормальную форму некоторого нетривиального многочлена вблизи его критической точки второго порядка. Кроме этого, определим расположение линий уровня ее нормальной формы вблизи этой критической точки так, как это описано в работе [11, разд. 6.2]. Все дальнейшие вычисления выполнялись в системе компьютерной алгебры (СКА) Maple [8] с использованием процедур из пакетов Groebner, PolyhedralSets и algcurves. Такие же вычисления несложно реализуются в других СКА, например, в Mathematica [12] или в SymPy [10].

В работах [13], [14] изучалась устойчивость в линейном приближении некоторой многопараметрической системы Гамильтона. В процессе исследования был получен многочлен  $g(x, y, z)$  6-го порядка от трех переменных  $x, y, z$  (см. формулу (3.7) в [14]). Здесь исследуем многочлен

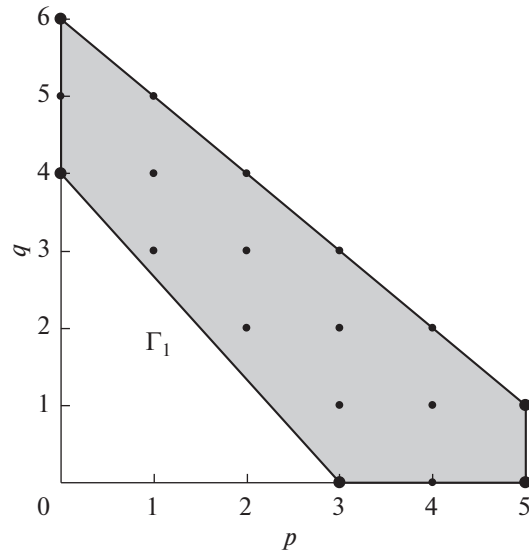
$f(x, y) \stackrel{\text{def}}{=} g(x, y, z)|_{z=0}$ , имеющий следующий вид:

$$\begin{aligned} f(x, y) = & 64x^3 + 1728yx^2 + 15552y^2x + 46656y^3 + 13x^4 + 412yx^3 + 2782y^2x^2 + 828y^3x - \\ & - 37827y^4 + 2x^5 + 82yx^4 + 740y^2x^3 + 3156y^3x^2 + 6490y^4x + 20250y^5 + 4yx^5 + \\ & + 48y^2x^4 + 344y^3x^3 + 1200y^4x^2 + 2500y^5x. \end{aligned} \quad (35)$$

Легко видеть, что многочлен (35) имеет в нуле особую точку 2-го порядка. Если составить полиномиальный идеал

$$\mathcal{F} = \left\{ f(x, y) - c, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\},$$

а затем вычислить для него базис Грёбнера с чисто лексикографическим порядком  $c \prec y \prec x$ , то можно найти все его критические точки. Для вычисления их порядков надо в них вычислить частные производные до их ненулевых значений.



Фиг. 4. Многоугольник Ньютона  $\Gamma(f)$  многочлена (37).

Вычисления показывают, что имеется две вещественных критических точки  $x_1 = y_1 = 0$  порядка 2 и  $x_2 = -9/2$ ,  $y_2 = 3/2$  порядка 1, соответствующих критическому значению  $c_1^* = 0$ , а также две вещественных критических точки порядка 1, соответствующих критическим значениям  $c_2^* \approx -476393.035$  и  $c_3^* \approx 51.504$ . Координаты двух последних критических точек являются вещественными корнями многочленов шестой степени, которые здесь не приводятся из-за их громоздкости.

Находим кубическую форму в особой точке 2-го порядка  $x = y = 0$ :

$$f_3(x, y) = 64x^3 + 1728yx^2 + 15552y^2x + 46656y^3 = 64(x + 9y)^3.$$

Тогда замена

$$u = 4(x + 9y), \quad \text{т.е.} \quad x = (u - 9y)/4 \quad (36)$$

приводит ее к нормальной форме 2 (66). В новых переменных  $(u, y)$  многочлен записывается в следующем виде (промежуточный многочлен):

$$\begin{aligned} f(u, y) = & u^3 - 34992y^4 + 3240uy^3 - \frac{253}{2}u^2y^2 - \frac{7}{8}u^3y + \frac{13}{256}u^4 + 97920y^5 - 11000uy^4 + \\ & + 528u^2y^3 - \frac{37}{4}u^3y^2 - \frac{u^4y}{32} + \frac{u^5}{512} - 97344y^6 + 13936uy^5 - 870u^2y^4 + 29u^3y^3 - \frac{33}{64}u^4y^2 + \frac{u^5y}{256}. \end{aligned} \quad (37)$$

Как описано, например, в [1], вычислим его носитель  $S(f)$ , найдем его выпуклую оболочку  $\Gamma(f)$  — многоугольник Ньютона — и выделим в нем ребро  $\Gamma_1$ , внешняя нормаль которого лежит в третьем квадранте. Описанные объекты показаны на фиг. 4.

Этому ребру соответствует укороченный многочлен

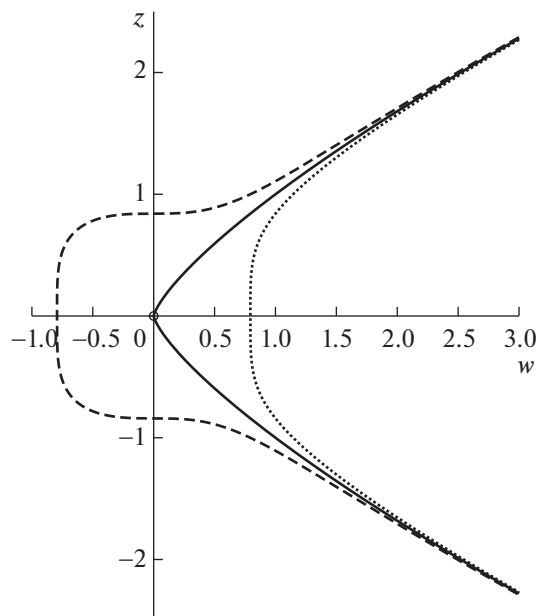
$$\hat{f}(u, y) = u^3 - 34992y^4. \quad (38)$$

Многочлен (38) соответствует виду 2 теоремы 3, следовательно, преобразованием

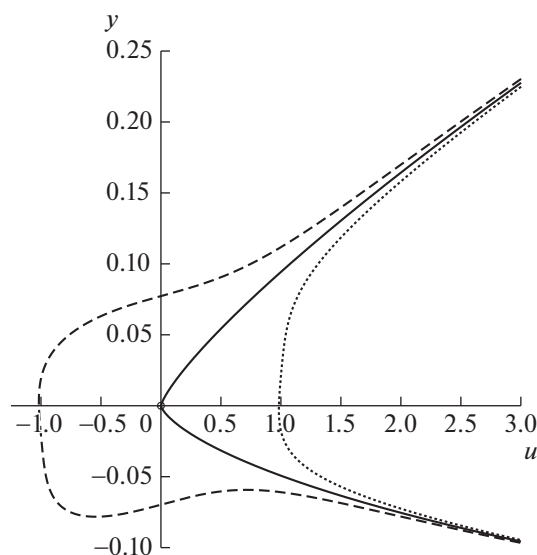
$$u = w, \quad y = \sqrt[4]{34992}z = 64^{3/4}z \quad (39)$$

он приводится к нормальной форме  $g$  вида 2 теоремы 2:

$$g(w, z) = w^3 - z^4. \quad (40)$$



Фиг. 5. Линии уровня нормальной формы (40) для значений  $c = -0.5, 0.0, 0.5$ .



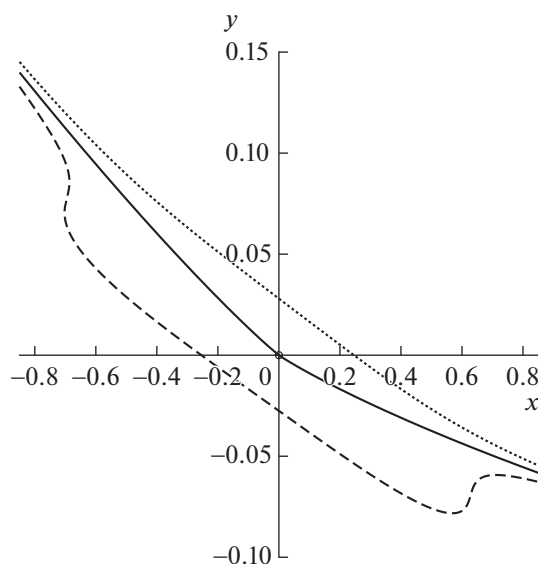
Фиг. 6. Линии уровня промежуточного многочлена (37) для значений  $c = -0.5, 0.0, 0.5$ .

Теперь построим линии уровня  $w^3 - z^4 = c$  нормальной формы (40) вблизи критической точки  $w = z = 0$  для значений  $c = -0.5$ ,  $c = 0$  и  $c = 0.5$ . На фиг. 5 они показаны соответственно штриховой, сплошной и пунктирной линиями.

Отметим, что с помощью пакета `algsurves` СКА Maple можно установить, что кривая  $f(u, y) = 0$  имеет род 0, т.е. допускает рациональную параметризацию

$$u(\tau) = \frac{(8\tau^2 + 420\tau + 8100)\tau^4}{(15\tau + 675)(\tau^2 + 30\tau + 450)^2}, \quad y(\tau) = -\frac{(\tau^2 + 45\tau + 900)\tau^3}{(2\tau + 90)(\tau^2 + 30\tau + 450)^2}.$$

Вид линий уровня  $f(u, y) = c$  для значений  $c = -0.5$ ,  $c = 0$  и  $c = 0.5$  вблизи начала координат показан соответственно штриховой, сплошной и пунктирной линиями на фиг. 6. Нетрудно ви-



Фиг. 7. Линии уровня исходного многочлена (35) для значений  $c = -0.5, 0.0, 0.5$ .

деть, что нормальная форма (40) качественно точно воспроизводит поведение линий уровня промежуточного многочлена (37).

Наконец, покажем на фиг. 7 расположение линий уровня  $f(x, y) = c$  вблизи начала координат исходного многочлена (35) для значений  $c = -0.5$ ,  $c = 0$  и  $c = 0.5$ . С учетом линейных преобразований (36) и (39) фиг. 6 качественно точно дает представление о поведении линий уровня исходного многочлена вблизи критической точки второго порядка.

### СПИСОК ЛИТЕРАТУРЫ

1. Брюно А.Д., Батхин А.Б. Алгоритмы и программы вычисления корней многочлена от одной или двух неизвестных // Программирование. 2021. № 5. С. 22–43.  
<https://doi.org/10.31857/S0132347421050046>
2. Haile D.E. On the Clifford algebra of a binary cubic form // American J. of Math. 1984. V. 106. № 6. P. 1269–1280.
3. Арнольд В.И. Нормальные формы функций в окрестности вырожденных критических точек // Успехи матем. наук. 1974. Т. 29. № 2. С. 11–49.
4. Кокс Д., Литтл Д., О'Ши Д. Идеалы, многообразия и алгоритмы. Введение в вычислительные аспекты алгебраической геометрии и коммутативной алгебры. М.: Мир, 2000. 687 с.
5. Прасолов В.В. Многочлены. 4-е изд., исправленное. М.: МЦНМО, 2014.
6. Калинина Е.А., Утешев А.Ю. Теория исключения: Учеб. пособие. СПб.: Изд-во НИИ химии СПбГУ, 2002. 72 с.
7. Батхин А.Б. Параметризация дискриминантного множества вещественного многочлена // Программирование. 2016. Т. 42. № 2. С. 8–21.
8. Maplesoft, a division of Waterloo Maple Inc. Maple. Version 2019. Waterloo, Ontario, 2019. URL <https://hadoop.apache.org>.
9. Thompson I. Understanding Maple. Cambridge University Press, 2016. 228 p.
10. Meurer A. [et al.]. SymPy: symbolic computing in Python // PeerJ Computer Science. 2017. V. 3. e103. ISSN 2376–5992.  
<https://doi.org/10.7717/peerj-cs.103>.
11. Брюно А.Д., Батхин А.Б., Хайдаров З.Х. Примеры вычисления линий уровня многочленов на плоскости // Препринты ИПМ им. М.В. Келдыша. 2021. № 98. 36 с.  
<https://doi.org/10.20948/prepr-2021-98>
12. Wolfram S. The Mathematica Book. Wolfram Media, Inc., 2003. 1488 p.
13. Брюно А.Д., Батхин А.Б. Разрешение алгебраической сингулярности алгоритмами степенной геометрии // Программирование. 2012. № 2. С. 12–30.
14. Батхин А.Б., Брюно А.Д., Варин В.П. Множества устойчивости многопараметрических гамильтоновых систем // Прикл. матем. и механ. 2012. Т. 76. № 1. С. 80–133.

## ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.16

### ФУНКЦИОНАЛЬНОЕ СУММИРОВАНИЕ РЯДОВ

© 2023 г. В. П. Варин<sup>1,\*</sup>

<sup>1</sup> 125047 Москва, Миусская пл., 4, Институт прикладной математики им. М.В. Келдыша РАН, Россия

\*e-mail: varin@keldysh.ru

Поступила в редакцию 25.04.2022 г.  
Переработанный вариант 25.04.2022 г.  
Принята к публикации 17.09.2022 г.

Рассматривается способ суммирования рядов, который сводится к решению некоторых линейных функциональных уравнений. Частичные суммы любого числового ряда удовлетворяют очевидному разностному уравнению. Это уравнение преобразуется в функциональное уравнение на интервале  $[0, 1]$  для непрерывного аргумента. Далее это уравнение либо решается явно (с точностью до произвольной константы), либо вычисляется асимптотическое разложение решения в нуле. Сумма исходного ряда определяется однозначно как константа, которая необходима для согласования асимптотического разложения решения с частичными суммами исходного ряда. Понятие предела не участвует в данной вычислительной схеме, что позволяет суммировать также расходящиеся ряды. Библ. 16. Фиг. 1.

**Ключевые слова:** суммирование рядов, функциональные уравнения, расходящиеся ряды.

**DOI:** 10.31857/S0044466923010131, **EDN:** LDWDNY

#### 1. ВВЕДЕНИЕ

Суммирование рядов является одной из старейших задач математического анализа и берет свое начало в античности. Эта задача всегда имела два аспекта: практический, т.е. вычислительный, и теоретический, т.е. концептуальный. Иными словами, прежде чем вычислять сумму бесконечного ряда, следует условиться, что следует понимать под этой суммой.

Во времена Эйлера понятие суммы ряда не было четко определено, что давало определенный простор в толковании этого понятия и позволило просуммировать многие расходящиеся в обычном понимании ряды.

С появлением строгого понятия предела многие математики отказались от использования расходящихся рядов. Достаточно вспомнить знаменитое высказывание Абеля, который усмотрел в расходящихся рядах происки дьявола.

Однако довольно скоро выяснилось, что без расходящихся рядов никак не обойтись, что привело к созданию теории расходящихся рядов и появлению методов их суммирования (см. [1]).

В настоящее время существует обширный арсенал методов суммирования рядов, многие из которых могут суммировать также и расходящиеся ряды. Однако все эти методы (за исключением формулы Эйлера–Маклорена) сводят так или иначе исходно медленно сходящийся или расходящийся процесс к сходящемуся, предел которого и называется суммой ряда.

Формула Эйлера–Маклорена, а также значительно менее известная формула Эйлера–Буле здесь стоят особняком, так как они сводят (медленно) сходящийся ряд к расходящемуся асимптотическому. То есть эти формулы являются скорее (весьма эффективными) способами вычисления суммы исходно сходящегося ряда.

Таким образом, все существующие способы суммирования рядов, как сходящихся, так и расходящихся, используют понятие предела и определяют сумму ряда как предел некоторого сходящегося процесса, который и является алгоритмом суммирования.

В то же время сумму ряда можно определить так, что понятие предела нигде не будет использовано, что дает возможность суммировать ряды, суммы которых ранее не были определены и для которых существующие методы суммирования неприменимы. Приведем пример такого ряда.

Рассмотрим формальный ряд

$$y(x) = \sum_{n=0}^{\infty} n! x^n \quad (1)$$

и определим его сумму при  $x = 1$ .

При  $x = -1$  сумма этого ряда равна константе Эйлера–Гомпертца  $\delta = e \operatorname{Ei}(1, 1)$ , где  $\operatorname{Ei}()$  – это интегральная экспонента (см. разд. 5).

Если бы этот ряд сходилась при  $x < 1$  и  $\lim_{x \rightarrow 1} y(x)$  существовал, то это был бы классический способ суммирования Абеля. Однако этот ряд всюду расходится. Причем ни один из существующих способов суммирования рядов, как то  $\varepsilon$  – алгоритм Винна, Паде-аппроксимации, метод Левина и его модификации, итерационный метод Айткена (см. обзор в [2] и ссылки там), а также  $\tau$  – метод Ланцоша (см. [3]) и т.д., не могут вычислить его сумму.

В то же время функция  $y(x)$  (формальный ряд) удовлетворяет дифференциальному уравнению

$$x^2 y' - (1 - x)y + 1 = 0,$$

которое интегрируемо. Функция

$$y(x) = \frac{1}{x} \exp\left(-\frac{1}{x}\right) \left(C - \operatorname{Ei}\left(1, -\frac{1}{x}\right)\right),$$

где  $C$  – это константа интегрирования, является общим решением этого уравнения с начальным значением  $y(0) = 1$ , т.е. все решения этого уравнения имеют это начальное значение и одно и то же асимптотическое разложение в нуле (так как  $\exp(-1/x)$  является плоской функцией).

Общее решение  $y(x)$ , вообще говоря, комплекснозначно, причем аргумент функции  $\operatorname{Ei}()$  при  $x \in [0, 1]$  расположен в точности на разрезе комплексной плоскости  $(-\infty, 0)$ , где функция  $\operatorname{Ei}()$  претерпевает скачок. Константу  $C$  можно выбрать, в определенном смысле, единственно возможным разумным образом. Возьмем эту константу чисто мнимой и так, чтобы функция  $y(x)$  принимала вещественные значения при  $x \in \mathbb{R}_+$ , т.е. возьмем в качестве решения функцию

$$y(x) = -\frac{1}{x} \exp\left(-\frac{1}{x}\right) \left(i\pi + \operatorname{Ei}\left(1, -\frac{1}{x}\right)\right).$$

Тогда в качестве суммы исходного ряда следует принять

$$y(1) = -\exp(-1) (i\pi + \operatorname{Ei}(1, -1)) \approx 0.69717488323505.$$

Заметим, что интегральный метод суммирования Бореля здесь не применим напрямую. Но если его модифицировать и вычислять интеграл в смысле главного значения, то найденная нами константа правильно вычисляется.

Данный метод суммирования, т.е. замена числового ряда на степенной, предъявление дифференциального уравнения для его суммы с последующим интегрированием ОДУ (аналитически либо численно) – является весьма мощным средством суммирования, но имеющим весьма ограниченное применение. Нужно дифференциальное уравнение в общем случае может и не существовать. Но этот пример показывает, что сумму ряда можно, в принципе, определить, не прибегая к понятию предела.

## 2. СУММА РЯДА КАК КОНСТАНТА ИНТЕГРИРОВАНИЯ

По аналогии с ОДУ решение разностных уравнений мы будем называть их интегрированием, а общее решение разностного уравнения будем называть первым интегралом разностного уравнения.

Эта аналогия хорошо заметна при построении непрерывных итераций дискретных аналитических отображений (см. [4]), где предъявление первого интеграла отображения эквивалентно построению непрерывных итераций и инвариантных кривых этого отображения (см. [5]).

При суммировании числового ряда его частичные суммы

$$s(n) = \sum_{k=1}^n a(k) \quad (2)$$

удовлетворяют очевидному разностному уравнению

$$s(n) - s(n-1) = a(n), \quad (3)$$

где  $a(x)$  — это некоторая аналитическая функция при  $x \in \mathbb{R}$ ,  $x > \text{const}$ . Это единственное ограничение, которое мы накладываем. При этом ряд может быть и знакопеременным (см. ниже).

Начальный индекс суммирования в (2) не играет никакой роли, но обычно будем брать  $k = 1$ .

Таким образом, вместо суммы (2), т.е. “интеграла”, предлагается изучать разностное уравнение (3), т.е. “ОДУ”. Аналогия здесь очевидна.

Отмеченная аналогия разностного уравнения и дифференциального на самом деле является весьма глубокой и хорошо изученной. Например, разностное уравнение (3) можно записать как уравнение

$$(\exp(D) - 1)s(n) = a(n+1),$$

где  $D$  — это оператор дифференцирования. Если двигаться по этому пути, то мы придем к методу суммирования Эйлера—Маклорена. Но мы сосредоточимся на разностном уравнении (3).

Разностное уравнение (3) иногда можно преобразовать в однородное трехзвенное линейное соотношение. Для линейных разностных уравнений с полиномиальными коэффициентами существует весьма развитая теория, которая, в частности, позволяет вычислить асимптотику решения при  $n \rightarrow \infty$ . Но в общем случае проще работать с уравнением (3).

Предположим, что мы нашли какое-то решение  $\tilde{s}(n)$  уравнения (3) или, что более реалистично, его асимптотику при  $n \rightarrow \infty$ . Функция  $\tilde{s}(x)$ , вообще говоря, определена теперь и при вещественных  $x$ , т.е. одновременно решена (по крайней мере локально) задача о непрерывном (либо аналитическом) продолжении дискретного отображения  $n \rightarrow s(n)$ .

Пусть исходный ряд (2) сходится. Тогда можно считать, что  $\tilde{s}(x) \rightarrow 0$  при  $x \rightarrow \infty$ , так как решение уравнения (3) определено лишь с точностью до константы  $C$ . В любом случае именно такую асимптотику и предлагается искать.

Таким образом, имеем два решения уравнения (3): одно дискретное (2) и одно непрерывное

$$s(x) = C + \tilde{s}(x),$$

которое, вообще говоря, может быть определено лишь асимптотически при достаточно больших  $x$ . Тогда константа интегрирования  $C$  (она же сумма ряда (2) при  $n \rightarrow \infty$ ) однозначно восстанавливается условием согласования (т.е. равенства) частичных сумм исходного ряда со значениями функции  $s(x)$  в дискретных точках  $x = n \in \mathbb{N}$ , т.е.

$$C = s(n) - \tilde{s}(n). \quad (4)$$

При этом, очевидно, достаточно установить это равенство только в одной (любой) точке  $n = 1, 2, \dots$ . Во всех остальных точках это равенство будет выполняться автоматически.

Таким образом, сумма ряда (2) при  $n \rightarrow \infty$  определена однозначно, но никакого предела при этом мы не вычисляли. Задача свелась к решению разностного уравнения (3) либо к вычислению асимптотики его решения.

Разумеется, вычисление асимптотики предполагает, строго говоря, вычисление предела на каждом шаге. Но алгоритмически это чаще всего чисто алгебраическая операция.

Возьмем в качестве примера Базельскую проблему, т.е.  $a(n) = 1/n^2$ . Предположим, что мы ничего не знаем об этой проблеме и не умеем решать разностное уравнение (3). Но мы можем тем не менее вычислить асимптотику его решения  $\tilde{s}(n)$  методом неопределенных коэффициентов, что является простой алгебраической задачей. Имеем

$$\tilde{s}(n) = -\frac{1}{n} + \frac{1}{2n^2} - \frac{1}{6n^3} + \frac{1}{30n^5} + O\left(\frac{1}{n^6}\right).$$

Заметим теперь, что асимптотические ряды такого рода хорошо изучены. В частности, известно, что величина погрешности при вычислении по таким рядам обычно оценивается последним отброшенным членом (см. [6]). Поэтому они суммируются до наименьшего по модулю члена.

В данном случае, если взять  $n = 10$ , что эквивалентно суммированию первых 10 членов исходного ряда, а затем учесть найденную асимптотическую поправку, то мы найдем константу

$C = \pi^2/6$  с шестью десятичными знаками (на самом деле, с восемью). Исходная сумма (2) при  $n = 10$  дает только два десятичных разряда.

То, что мы получили, весьма напоминает метод суммирования Эйлера—Маклорена, но только без вычисления интеграла.

Другая аналогия, которая также очевидна, — это решение краевых задач методом погранслоя, когда решение получается в результате сращивания двух различных разложений, вычисленных в противоположных концах интервала.

В случае, если ряд (2) расходится в обычном смысле, это означает, что последовательность частичных сумм  $\{s(n)\}$  не имеет предела при  $n \rightarrow \infty$ , либо стремится к бесконечности. Так же ведет себя, очевидно, и решение  $\tilde{s}(n)$  уравнения (3). Однако  $\tilde{s}(n)$  (вообще говоря) имеет некоторую асимптотику, которую и предлагается вычислять. Иногда можно найти и явное решение. Тогда асимптотика частичных сумм  $s(n)$  и асимптотика решения  $\tilde{s}(n)$  в уравнении (4) могут сократить друг друга (как разрешение неопределенности  $\infty - \infty$ ) и сумма расходящегося в обычном смысле ряда будет однозначно определена.

Отметим, что смысл указанной операции не столько в назначении числового значения сумме расходящегося в обычном смысле ряда, сколько в установлении связи частичных сумм ряда с их асимптотическим поведением в бесконечности.

Приведем пример суммирования для  $a(n) = 1/n$ , т.е. для гармонического ряда. Он также не может быть просуммирован обычными средствами.

Разностное уравнение (3) для  $a(n) = 1/n$  решается в явном виде. Но мы не будем этим пользоваться, а вычислим асимптотику его решения  $\tilde{s}(n)$  методом неопределенных коэффициентов. Имеем

$$\tilde{s}(n) = \ln n + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} + O\left(\frac{1}{n^6}\right) = \ln n - \sum_{k=1}^{\infty} \frac{B(k)}{kn^k},$$

где  $B(k)$  — это числа Бернулли.

Обозначив константу интегрирования  $C = \gamma$ , получаем из (4) знаменитое асимптотическое тождество Эйлера

$$\gamma = H(n) - \ln n + \sum_{k=1}^{\infty} \frac{B(k)}{kn^k}, \quad n \in \mathbb{N}, \quad (5)$$

где  $H(n)$  — это гармоническое число.

Метод суммирования, изложенный в этом разделе, нельзя назвать принципиально новым потому, что подобный подход уже использовался Раманужаном. Хотя Раманужан исходил из формулы Эйлера—Маклорена, но он также пришел к сумме ряда как к “константе Раманужана” (см. [7, pp. ix, 3]).

В следующих разделах мы приведем ряд примеров суммирования рядов, которые (как мы надеемся) покажут преимущества нашего метода. Но сперва преобразуем разностное уравнение (3) к функциональному уравнению на интервале  $[0, 1]$  и дадим мотивировку этого преобразования.

### 3. ФУНКЦИОНАЛЬНОЕ УРАВНЕНИЕ ДЛЯ СУММЫ РЯДА

Разностное уравнение (3) определено на интервале  $[0, \infty]$ , что может иметь иногда некоторые преимущества. Однако, как показывает опыт, функциональное уравнение на конечном интервале часто бывает удобнее.

Сделаем замену переменных в уравнении (3),  $n = 1/x$ ,  $S(x) = s(1/x)$ ,  $A(x) = a(1/x)$  (“Man muss immer umkehren.” C.G.J. Jacobi). Тогда получим функциональное уравнение

$$S(x) - S\left(\frac{x}{1-x}\right) = A(x). \quad (6)$$

Например, для Базельской проблемы, т.е. для  $A(x) = x^2$ , мы получаем уравнение, которое лишь знаком в левой части отличается от уравнения (A.7) в [9, p. 241]. Это последнее приводилось в [9] в качестве примера уравнения, решением которого является обыкновенная производящая функция для чисел Бернулли (помноженная на  $x$ ).

Уравнение (6) легко преобразуется к функциональному уравнению на интервале  $[0, 1]$ . Для этого сделаем замену  $x \rightarrow x/(1+x)$ . Тогда получим уравнение

$$S\left(\frac{x}{1+x}\right) - S(x) = A\left(\frac{x}{1+x}\right). \quad (7)$$

Одно из преимуществ функциональной формы записи разностного уравнения (3) — это возможность использовать развитый аппарат вычисления степенных рядов (и более общих разложений), реализованный в системах компьютерной алгебры (CAS).

Другое не столь очевидное преимущество появляется, если уравнение (7) решать численно, например, спектральным методом. Функцию  $S(x)$  можно представить, например, в виде разложения по смещенным полиномам Чебышёва. Так можно, в принципе, решить численно и с контролируемой точностью уравнение (7), т.е. вычислить сумму исходного ряда, даже если функция  $A(x)$  устроена весьма сложно.

Численное решение уравнения (7) являлось первоначальной мотивировкой данного исследования. Однако это является отдельной задачей, которой мы намерены посвятить отдельную статью.

Для вычисления асимптотики решения  $S(x)$  при  $x \rightarrow +0$  (которая может быть весьма сложной) уравнение (7) не имеет никаких преимуществ по сравнению с уравнением (6). Так что это последнее мы и будем далее использовать.

В случае знакопеременного ряда запишем его в виде

$$s(n) = \sum_{k=1}^n a(k) \cos \pi k. \quad (8)$$

Тогда, сделав те же замены, что и ранее, и взяв  $S(x) = U(x) \cos \pi/x$ , получим уравнение

$$U(x) + U\left(\frac{x}{1-x}\right) = A(x) \quad (9)$$

вместо уравнения (6).

Асимптотическое решение уравнения (9) имеет некоторые незначительные отличия от решения уравнения (6). Это, как мы полагаем, выгодно отличает наш подход от классического, который в случае знакопеременных рядов приводит к (весьма малоизвестной) формуле Эйлера—Буле (см. [8]).

Заметим также, что иногда вместо множителя  $\cos \pi k = (-1)^k$  следует брать множитель  $\exp(i\pi k)$ , который бывает нужен для аналитического продолжения дискретного отображения  $n \rightarrow s(n)$ .

Разностные уравнения типа (6) и (9) для достаточно простых правых частей  $A(x)$  уже встречались в литературе, но в другом контексте. Так, подобные уравнения вводились в [9] комбинаторным путем для обыкновенных производящих функций чисел Бернулли и их обобщений. Многие из таких уравнений решаются в явном виде (см. [10]), хотя в [9] были даны лишь асимптотические разложения решений.

Поскольку эти уравнения линейны, то это дает возможность комбинировать уже известные решения с целью получения новых. Приведем пример суммирования ряда с использованием известных решений уравнения (6). Рассмотрим ряд

$$\sum_{k=2}^{\infty} \left( \frac{1}{k} - \ln \left( \frac{k}{k-1} \right) \right) = \gamma - 1 \quad (10)$$

и найдем его сумму нашим методом и двумя способами. Для этого ряда уравнение (6) имеет вид

$$S(x) - S\left(\frac{x}{1-x}\right) = x + \ln(1-x). \quad (11)$$

Уравнение (11) лишь знаком левой части отличается от уравнения (A.14) в [9, р. 243], для которого мы нашли явное решение в [10]. Решение уравнения (11) имеет вид

$$S(x) = x + \ln x + \Psi\left(\frac{1}{x}\right) = \frac{1}{2}x - \frac{1}{12}x^2 + \frac{1}{120}x^4 + O(x^6).$$

Комбинируя полученное решение и формулы (4) и (10), получаем тождество

$$\gamma = \frac{n-1}{n} + \ln n - \Psi(n) + \sum_{k=2}^n \left( \frac{1}{k} - \ln \left( \frac{k}{k-1} \right) \right) = H(n-1) - \Psi(n), \quad n \in \mathbb{N}, \quad (12)$$

которое является явным решением асимптотического тождества Эйлера (5).

Заметим теперь, что правая часть уравнения (11) записывается как

$$x + \ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^{n+1}}{n+1}.$$

Решение уравнения (6) с правой частью  $x^{n+1}$ ,  $n \in \mathbb{N}_0$ , имеет вид (см. [10])

$$S(x) = \frac{(-1)^n}{n!} \Psi \left( n, 1 + \frac{1}{x} \right). \quad (13)$$

Комбинируя полученные решения и формулу (4), получаем тождество

$$\gamma = 1 + \sum_{k=2}^n \left( \frac{1}{k} - \ln \left( \frac{k}{k-1} \right) \right) + \sum_{k=1}^{\infty} \frac{(-1)^k \Psi(k, n+1)}{(k+1)!}. \quad (14)$$

Используем интегральное представление функции  $\Psi()$  в формуле (14):

$$\Psi(n, 1+x) = \int_0^1 \frac{t^x \ln^n t}{t-1} dt, \quad n \in \mathbb{N}.$$

Тогда получим обобщение известной формулы для константы Эйлера:

$$\gamma = H(n) - \ln n + \int_0^1 \left( \frac{1}{\ln t} + \frac{t}{1-t} \right) t^{n-1} dt, \quad n \in \mathbb{N}.$$

Ясно, что если суммировать не весь ряд (14) этим способом, а только его часть, то получим множество аналогов формулы Эйлера–Маклорена для гармонического ряда с остаточными членами в интегральной форме.

Пользуясь выражением гармонического числа через  $\Psi$ -функцию, получим (см. [2])

$$\Psi(n+1) = \frac{1}{n} + \ln n + \Gamma(n) \sum_{k=1}^{\infty} \frac{(-1)^k}{k \Gamma(n+k)} \sum_{j=0}^k \frac{S_1(k, j)}{j+1}, \quad \operatorname{Re}(n) > 0,$$

где  $S_1$  — это числа Стирлинга первого рода.

Из этой формулы можно получить бесконечное множество разложений константы Эйлера в (сколь угодно быстро) сходящиеся ряды рациональных чисел (например, подставить  $n = 2$ , затем  $n = 4$  и исключить  $\ln 2$ ).

Если теперь вычесть (12) из (14) и подставить туда  $n = x$ , то получим тождество

$$\ln x = \sum_{n=0}^{\infty} \frac{(-1)^n \Psi(n, x+1)}{(n+1)!}, \quad x > 0. \quad (15)$$

Заметим, что ряд Тейлора для логарифма от Гамма функции дает похожее тождество

$$\ln x = \sum_{n=0}^{\infty} \frac{\Psi(n, x)}{(n+1)!}, \quad x > 0. \quad (16)$$

Если подставить  $x = 1$  в формулы (15) и (16), то получим два известных (еще Эйлеру) представления константы Эйлера

$$\gamma = 1 - \sum_{n=1}^{\infty} \frac{\zeta(n+1) - 1}{n+1} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \zeta(n+1)}{n+1}, \quad (17)$$

причем первый из этих рядов сходится значительно быстрее.

Подстановка  $x = 2, 3, \dots$  в формулы (15) и (16) не дает, как легко проверить, ничего нового, так как все логарифмы  $\ln n$ ,  $n > 2$  там сокращаются.

Данный пример показывает, что решения уравнений (6) и (9), пользуясь их линейностью, можно составлять из уже известных решений. В следующем разделе мы приведем сводку формул для некоторых известных решений этих уравнений. Далее мы дадим ряд примеров суммирования рядов с их помощью.

#### 4. РЕШЕНИЕ ФУНКЦИОНАЛЬНЫХ УРАВНЕНИЙ

Решение уравнений

$$S(x) - S\left(\frac{x}{1-x}\right) = x^{s+1} \quad (18)$$

при  $s = n \in \mathbb{N}_0$  мы нашли в [10]. Это решение (13), что легко проверяется. Однако это решение не продолжается на вещественные либо комплексные  $s$ .

Уравнение (18) для произвольных  $s$ , а также другие подобные уравнения с нестандартными правыми частями можно решать с помощью подходящей замены переменных.

Для уравнения (18) следует сделать замену  $S(x) = x^s V(x)$ . Тогда для функции  $V(x)$  получим уравнение

$$V(x) - (1-x)^{-s} V\left(\frac{x}{1-x}\right) = x,$$

которое не имеет логарифмической особенности в нуле. Для него следует использовать биномиальные тождества.

Пользуясь известным асимптотическим разложением (см., например, [10])

$$\Psi\left(n, 1 + \frac{1}{x}\right) = (-1)^{n+1} (n-1)! \sum_{k=0}^{\infty} B(k) C(k+n-1, k) x^{n+k}, \quad n \in \mathbb{N}, \quad (19)$$

где  $B(n)$  — это числа Бернулли и  $C()$  — это биномиальный коэффициент, для произвольного комплексного  $s \neq 0$ , получим (с помощью аналитического продолжения) асимптотическое решение

$$S(x) = -\frac{x^s}{s} \sum_{n=0}^{\infty} B(n) C(n+s-1, n) x^n. \quad (20)$$

Для отрицательных целых  $s$  функция  $S(x)$  в (20) выражается через полиномы Бернулли:

$$S(x) = \frac{(-1)^n}{n} B(n, -1/x), \quad n = -s \in \mathbb{N}.$$

Для  $s = 0$  асимптотическое разложение (20) непригодно, и его следует заменить на разложение

$$S(x) = \Psi\left(1 + \frac{1}{x}\right) = -\ln x - \sum_{n=1}^{\infty} \frac{B(n)}{n} x^n. \quad (21)$$

И, наконец, все эти формулы (при  $s \neq 0$ ) объединяет решение

$$S(x) = -\zeta\left(0, s+1, 1 + \frac{1}{x}\right),$$

где  $\zeta()$  — это  $\zeta$ -функция Гурвица в обозначении, принятом в Maple, т.е. где первый аргумент — это номер производной этой функции.

Последняя формула, впрочем, может быть использована скорее для вычисления функции Гурвица с использованием предыдущих формул, чем наоборот.

Перейдем к решению уравнений (9) для знакопеременных рядов.

Решение уравнения

$$U(x) + U\left(\frac{x}{1-x}\right) = x^{s+1} \quad (22)$$

для  $s = n \in \mathbb{N}_0$  имеет вид

$$U(x) = \frac{(-1)^{n+1}}{n!} \left( \Psi \left( n, 1 + \frac{1}{x} \right) - \frac{1}{2^n} \Psi \left( n, 1 + \frac{1}{2x} \right) \right) + \begin{cases} \ln 2, & n = 0, \\ 0, & n \neq 0. \end{cases} \quad (23)$$

Пользуясь разложением (19), получаем асимптотическое решение уравнения (22) для  $s \in \mathbb{N}$ , а также (аналитическим продолжением) и для произвольного комплексного  $s \neq 0$ :

$$U(x) = \frac{x^s}{s} \sum_{n=1}^{\infty} B(n) C(n+s-1, n) (2^n - 1) x^n. \quad (24)$$

Для отрицательных целых  $s$  функция  $U(x)$  в (24) выражается через полиномы Эйлера:

$$U(x) = \frac{(-1)^n}{2} E(n, -1/x), \quad n = -s - 1 \in \mathbb{N}_0,$$

что соответствует появлению этих полиномов в формуле Эйлера–Буле для знакопеременных рядов (см. [7, р. 135] и [8]).

Для  $s = 0$  асимптотическое разложение (24) непригодно, и его следует заменить на разложение

$$U(x) = \ln 2 - \Psi \left( 1 + \frac{1}{x} \right) + \Psi \left( 1 + \frac{1}{2x} \right) = \sum_{n=1}^{\infty} \frac{B(n)(1 - 2^n)}{n} x^n.$$

И, наконец, все эти формулы (при  $s \neq 0$ ) объединяет решение

$$U(x) = \zeta \left( 0, s+1, 1 + \frac{1}{x} \right) - \frac{1}{2^s} \zeta \left( 0, s+1, 1 + \frac{1}{2x} \right).$$

Приведенные формулы позволяют суммировать обычный ряд для  $\zeta$ -функции Римана во всей комплексной плоскости, кроме особой точки. Сращивая частичные суммы с асимптотическим разложением, по формуле (4) получаем:

$$\zeta(s+1) = \sum_{n=1}^m \frac{1}{n^{s+1}} + \frac{1}{sm^s} \sum_{n=0}^{\infty} \frac{B(n) C(n+s-1, n)}{m^n}, \quad s \neq 0, \quad m \in \mathbb{N}. \quad (25)$$

Данный ряд всюду расходится, но если его суммировать до наименьшего по модулю члена, то функцию Римана можно вычислять по формуле (25) в любой точке  $s \in \mathbb{C}$ ,  $s \neq 0$  с контролируемой точностью. Причем при выбранном  $s$  чем больше  $m$ , тем больше членов ряда можно использовать.

Формула (25) аналогична той, которая получается по формуле Эйлера–Маклорена, и которая ранее была основным вычислительным инструментом для функции Римана (см. [11]).

При целых отрицательных  $s$  ряд в (25) обрывается, и мы получаем известную формулу

$$\zeta(s+1) = \sum_{n=1}^m \frac{1}{n^{s+1}} + B(-s, m+1)/s, \quad -s \in \mathbb{N},$$

которая дает  $\zeta(0) = -1/2$ , а также рациональные значения функции  $\zeta(s+1)$  в точках  $s = -2k$  и тривиальные нули в точках  $s = -2k - 1$ ,  $k \in \mathbb{N}$ .

И, наконец, при  $s \in \mathbb{N}_0$  формула (25) принимает вид

$$\zeta(s+1) = \sum_{n=1}^m \frac{1}{n^{s+1}} - \frac{(-1)^s \Psi(s, 1+m)}{s!}, \quad m \in \mathbb{N}_0,$$

где

$$\zeta(s) = \begin{cases} \gamma, & s = 1, \\ \zeta(s), & s \in \mathbb{N}, \quad s > 1. \end{cases} \quad (26)$$

Расходящийся асимптотический ряд (25) можно преобразовать в сходящийся факториальный ряд, причем двумя способами (см. [2], [10]). Однако, как мы полагаем, факториальные сходящиеся ряды в данном случае не дают никаких преимуществ.

Рассмотрим преобразование степенных рядов, которое (насколько нам известно) не описано в литературе, но может быть полезно для решения разностных уравнений, которые мы рассматриваем.

Асимптотический ряд (19) можно записать следующим образом:

$$\Psi\left(n, 1 + \frac{1}{x}\right) = (-1)^{n+1} \sum_{k=n}^{\infty} \frac{B(k-n)(k-1)!}{(k-n)!} x^k, \quad n \in \mathbb{N}, \quad (27)$$

что можно рассматривать как бесконечную треугольную матрицу, преобразующую вектор  $\langle x, x^2, \dots \rangle^T$  в вектор функций  $\Psi()$ . Эта матрица обратима для любого порядка главного минора. Можно проверить, что мы получим сходящиеся ряды

$$x^n = \frac{1}{(n-1)!} \sum_{k=n}^{\infty} \frac{(-1)^{k+1}}{(k-n+1)!} \Psi\left(k, 1 + \frac{1}{x}\right), \quad n \in \mathbb{N}, \quad \operatorname{Re}(x) > 0, \quad (28)$$

что, кстати, дает еще одно представление для чисел Бернулли в матричной форме.

Формула (28) позволяет получать необычные разложения. Например,

$$\ln(1+x) = 2 \sum_{k=1}^{\infty} \frac{1}{(2k)!} \Psi\left(2k-1, 1 + \frac{1}{x}\right), \quad \ln 2 = \sum_{k=1}^{\infty} \frac{\zeta(2k)-1}{k}.$$

Однако использование формулы (28) в (25) ( $x = 1/m$ ) дает лишь другой асимптотический ряд.

## 5. ПРИМЕРЫ СУММИРОВАНИЯ ЗНАКОПЕРЕМЕННЫХ РЯДОВ

Мы уже привели несколько примеров суммирования для сходящихся и расходящихся знакопостоянных рядов. Дадим примеры суммирования некоторых расходящихся знакопеременных рядов.

Рассмотрим ряд

$$g = \sum_{n=1}^{\infty} (-1)^n \ln n. \quad (29)$$

Суммирование этого ряда приводилось в [12] как демонстрация эффективности предложенного численного алгоритма. Заметим, что Maple также без труда находит численное значение суммы этого ряда, которое оказывается равным

$$g = \frac{1}{2} \ln\left(\frac{\pi}{2}\right) \approx 0.225791352644727.$$

Эвристически угадать это значение можно из представления функции Римана

$$(1 - 2^{1-s})\zeta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s}, \quad \operatorname{Re}(s) > 0.$$

Если продифференцировать это тождество по  $s$  и подставить туда  $s = 0$  (что, вообще-то, незаконно), то получим как раз нужное значение. Таким образом, ряд (29) “помнит” о своем происхождении.

Покажем, что ряд (29) суммируется нашим способом в явном виде (чего Maple сделать не может).

Действуя описанным в разд. 3 способом, получаем функциональное уравнение

$$U(x) + V\left(\frac{x}{1-x}\right) = -\ln x.$$

Сделаем в нем подстановку  $U(x) = V(x) + W(x) \ln x$ . Тогда получим уравнение

$$V(x) + V\left(\frac{x}{1-x}\right) - W\left(\frac{x}{1-x}\right) \ln(1-x) = -\left(W(x) + W\left(\frac{x}{1-x}\right) + 1\right) \ln x.$$

Правая часть этого уравнения обнуляется при  $W(x) \equiv -1/2$ . Поэтому получаем уравнение

$$V(x) + V\left(\frac{x}{1-x}\right) = -\frac{1}{2}\ln(1-x).$$

Заметим теперь, что последнее уравнение мы получили бы описанным в разд. 3 способом из сходящегося ряда. Поэтому

$$V(x) = \frac{1}{2} \sum_{n=1}^{\infty} (-1)^n \ln \left( 1 - \frac{x}{1+nx} \right).$$

Обычно CAS не может суммировать такие ряды. Поэтому преобразуем его к виду

$$\exp(2V(x)) = \prod_{n=1}^{\infty} \frac{(1 + (2n-1)x)^2}{(1 + (2n-2)x)(1 + 2nx)}.$$

Это произведение вычисляется в явном виде. Получаем

$$V(x) = \ln \Gamma\left(\frac{1}{2x}\right) - \ln \Gamma\left(\frac{1}{2} + \frac{1}{2x}\right) - \frac{1}{2}\ln(2x), \quad x > 0.$$

Заметим, что если бы целью являлось численное решение задачи (с неограниченной точностью), то мы могли бы вычислить асимптотическое разложение решения методом неопределенных коэффициентов, т.е.

$$V(x) = \frac{1}{4}x - \frac{1}{24}x^3 + \frac{1}{20}x^5 - \frac{17}{112}x^7 + \frac{31}{36}x^9 - \frac{691}{88}x^{11} + \dots$$

Появление числа 691 в асимптотическом разложении  $V(x)$  является явным индикатором того, что в нем присутствуют числа Бернулли и данное разложение расходится. Поэтому другой способ просуммировать такие ряды – это найти достаточно много коэффициентов асимптотического разложения, угадать их явный вид и применить средства CAS. В данном случае

$$V(x) = \sum_{n=1}^{\infty} \frac{B(2n)(4^n - 1)}{2n(2n-1)} x^{2n-1}.$$

Собирая полученные формулы и применяя формулу (4), получаем сумму ряда (29)  $g$ , как и ожидалось.

Однако, как мы отмечали, суммирование расходящегося в обычном смысле ряда имеет вполне прозаический смысл. В данном случае мы нашли частичную сумму ряда (29) в замкнутой форме (чего Maple сделать не может), а также получили аналитическую интерполяцию частичных сумм ряда (29). Эта формула имеет вид

$$\sum_{k=1}^n (-1)^k \ln k = \frac{1}{2} \ln \left( \frac{1}{2} \pi \right) + \left( \ln \Gamma\left(\frac{n}{2}\right) - \ln \Gamma\left(\frac{1}{2} + \frac{n}{2}\right) + \frac{1}{2} \ln \left( \frac{n^2}{2} \right) \right) \cos(\pi n).$$

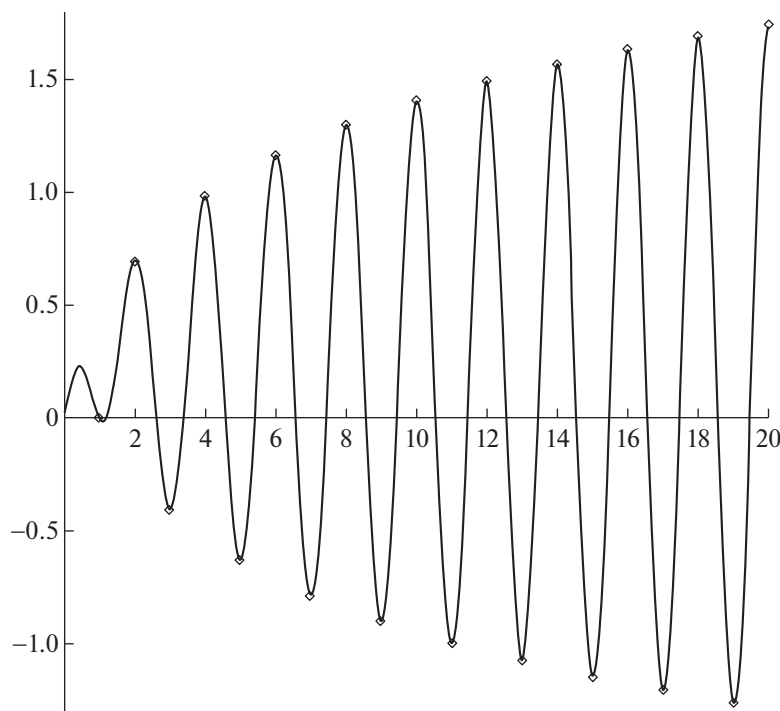
На фиг. 1 показаны график частичных сумм ряда (29) в дискретных точках и график их интерполяции по данной формуле.

В завершение этого примера отметим, что найденная нами функция  $V(x)$ , которая описывает асимптотическое поведение частичных сумм ряда (29) при  $n \rightarrow \infty$ , связана с тангенциальными числами следующей формулой:

$$xV'(x) = \frac{1}{2}T\left(\frac{1}{2}x\right),$$

где  $T(x)$  – это обыкновенная производящая функция тангенциальных чисел, для которой мы нашли аналитическое представление (см. [10]). Цепная дробь производящей функции тангенциальных чисел дана в [9, р. 257].

Нам удалось просуммировать нашим способом все ряды, приведенные в [12] в качестве примеров применения численного алгоритма в [12]. Если принять в качестве постулата, что суммирование асимптотических рядов следует производить до наименьшего по модулю члена (что требует отдельного исследования в каждом конкретном случае), то наш метод суммирует эти ряды с произвольной точностью и, по сути, алгебраическими средствами, программируемыми в CAS.



Фиг. 1. Частичные суммы ряда (29) и их интерполяция.

Применим наш метод к классическому ряду (1) при  $x = -1$ , т.е. найдем константу Эйлера–Гомпертца  $\delta = e\text{Ei}(1,1)$ .

Действуя описанным в разд. 3 способом, получаем функциональное уравнение

$$U(x) + U\left(\frac{x}{1-x}\right) = \frac{1}{x} \Gamma\left(\frac{1}{x}\right).$$

Сделаем в нем подстановку  $U(x) = \Gamma(1 + 1/x)W(x)$ . Тогда для  $W(x)$  получим уравнение

$$W(x) + xW\left(\frac{x}{1-x}\right) = 1. \quad (30)$$

Данное уравнение отличается от стандартных функциональных уравнений, которые мы рассматривали, но алгоритмически (и асимптотически) решается тем же способом. Получаем

$$W(x) = 1 - x + x^2 - x^4 - x^5 + 2x^6 + 9x^7 + 9x^8 - 50x^9 - 267x^{10} + O(x^{11}).$$

Последовательность целых коэффициентов этого разложения можно найти в энциклопедии целых последовательностей (см. [13, A000587]). Она связана с числами Белла. С помощью формулы суммирования Бореля получаем

$$W(x) = 1 - x \int_0^\infty \exp(-t) \exp(1 - \exp(tx)) dt, \quad x > 0. \quad (31)$$

Фактически мы уже решили задачу. Однако Maple не может вычислить данный интеграл в явном виде при произвольном  $x > 0$ . Поэтому применим еще один способ решения функционального уравнения (30), который обладает большей универсальностью.

Сделаем в уравнении (30) замены, обратные тем, которые мы применили в начале разд. 3. Тогда получим из (30) полиномиальное разностное уравнение

$$nw(n) + w(n-1) = n, \quad n \in \mathbb{N}. \quad (32)$$

Maple умеет решать такие уравнения в явном виде. Получаем

$$w(n) = \frac{n-1 + e\text{Ei}(n,1)}{n} + (-1)^n \frac{e\text{Ei}(1,1) - 1 + w(0)}{\Gamma(n+1)},$$

где  $w(0)$  — это начальное значение рекурсии (32).

Осциллирующая часть этого выражения удовлетворяет однородному уравнению (32). Но осциллирующей части здесь не должно быть, так как она учтена при выводе уравнения (30). Поэтому необходимо взять  $w(0) = 1 - e\text{Ei}(1,1)$ .

Таким образом, имеем еще одно решение уравнения (30):

$$W(x) = 1 - x + xe\text{Ei}\left(\frac{1}{x}, 1\right).$$

Это решение совпадает с предыдущим, т.е. мы вычислили в явном виде преобразование Лапласа в (31) (чего Maple сделать не может). Константа  $w(0)$  определилась бы также из совпадения этих решений.

Собирая полученные формулы и применяя формулу (4), получаем интерполяционную формулу для суммы ряда (1) при  $x = -1$ :

$$\sum_{k=0}^n (-1)^k k! = \delta + n! \left( 1 - \frac{1}{n} + \frac{e\text{Ei}(n,1)}{n} \right) \cos(\pi n),$$

где  $\delta = e\text{Ei}(1,1)$  — это принятая сумма данного ряда, т.е. константа Эйлера–Гомпертца.

Таким образом, как мы отмечали, сумма расходящегося ряда — это некоторая однозначно определяемая константа, необходимая для согласования дискретной последовательности частичных сумм ряда с их непрерывной (либо аналитической) интерполяцией в бесконечности. С этой точки зрения никакого различия между сходящимися и расходящимися рядами нет.

## 6. ФОРМУЛА СТИРЛИНГА В ВИДЕ СХОДЯЩЕГОСЯ РЯДА

Здесь мы рассмотрим пример суммирования расходящегося знакопостоянного ряда, который приводит к формуле Стирлинга в виде сходящегося ряда полигамма функций.

Рассмотрим ряд

$$h = \sum_{n=1}^{\infty} \ln n. \quad (33)$$

Суммирование этого ряда с помощью формулы Эйлера–Маклорена дает асимптотическую формулу Стирлинга для факториала.

Ряд (33) также “помнит” о своем происхождении, что видно, если продифференцировать ряд для функции Римана

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \text{Re}(s) > 1,$$

и подставить туда  $s = 0$  (что, разумеется, незаконно). Тогда получим

$$h = -\zeta'(0) = \frac{1}{2} \ln(2\pi) \approx 0.918938533204,$$

но мы не будем этим пользоваться.

Действуя описанным в разд. 3 способом, получаем функциональное уравнение

$$S(x) - S\left(\frac{x}{1-x}\right) = -\ln x.$$

Сделаем в нем подстановку  $S(x) = V(x) + W(x) \ln x$ . Тогда получим уравнение

$$V(x) - V\left(\frac{x}{1-x}\right) + W\left(\frac{x}{1-x}\right) \ln(1-x) = -\left(W(x) - W\left(\frac{x}{1-x}\right) + 1\right) \ln x.$$

Правая часть этого уравнения обнуляется при  $W(x) = -1/x$ . Поэтому получаем уравнение

$$V(x) - V\left(\frac{x}{1-x}\right) = \frac{(1-x)\ln(1-x)}{x} = -1 + \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)(n+2)}.$$

Далее мы можем применить асимптотические формулы (20) и (21) и собрать с их помощью асимптотическое решение для функции  $V(x)$ . Но мы применим формулу (13) и получим функцию  $V(x)$  в виде сходящегося ряда.

Комбинируя найденные решения и формулу (4), получаем сумму ряда (33)

$$h = \sum_{k=1}^n \ln k + n(1 - \ln n) - \sum_{k=0}^{\infty} \frac{(-1)^k \Psi(k, n+1)}{(k+2)!}, \quad n \in \mathbb{N},$$

откуда, подставив  $n = 1, 2, \dots$ , можно получить (с помощью Maple) численное значение константы  $h$ .

Чтобы получить точное значение, запишем эту формулу в виде

$$h - \ln \Gamma(n+1) - n(1 - \ln n) + \frac{1}{2} \Psi(n+1) = - \sum_{k=1}^{\infty} \frac{(-1)^k \Psi(k, n+1)}{(k+2)!}.$$

Правая часть этого выражения стремится к нулю при  $n \rightarrow \infty$ , а левая часть имеет асимптотическое разложение

$$h - \frac{1}{2} \ln 2 - \frac{1}{2} \ln \pi + \frac{1}{6n} - \frac{1}{24n^2} + \dots,$$

поэтому  $h = -\zeta'(0)$ , как мы нашли в начале этого раздела.

Таким образом, получаем вариант формулы Стирлинга в виде сходящегося ряда

$$\ln \Gamma(x+1) = \frac{1}{2} \ln(2\pi) - x(1 - \ln x) + \sum_{k=0}^{\infty} \frac{(-1)^k \Psi(k, x+1)}{(k+2)!}, \quad \operatorname{Re}(x) \geq 0. \quad (34)$$

Данная формула справедлива и при  $x = 0$ , если перейти к пределу. Тогда получим тождество

$$\frac{\gamma}{2} = \frac{1}{2} \ln(2\pi) - \sum_{k=1}^{\infty} \frac{\zeta(k+1)}{(k+1)(k+2)}.$$

Этот медленно сходящийся ряд можно ускорить с помощью метода Куммера.

Формулу (34) можно преобразовать в обычный асимптотический ряд Стирлинга с помощью преобразования (27), сделав в нем замену  $x \rightarrow 1/x$ . Однако существует более общее преобразование, включающее логарифм и функцию  $\Psi(0, x) = \Psi(x)$ .

Рассмотрим две бесконечные верхне-треугольные матрицы

$$P = \left| \frac{(-1)^k}{(k-n+1)!} \right|_{1 \leq n \leq k}, \quad Q = \left| \frac{(-1)^n B(k-n)}{(k-n)!} \right|_{1 \leq n \leq k}.$$

Тогда можно проверить, что  $P = Q^{-1}$  и  $Q = P^{-1}$  формально. Причем это справедливо для любого главного минора.

Рассмотрим также два бесконечных вектора

$$X = \left\langle -\ln x, \frac{1}{x}, \frac{1}{x^2}, \frac{2}{x^3}, \dots, \frac{(n-1)!}{x^n}, \dots \right\rangle^T, \quad Y = \langle \Psi(n-1, x+1) \rangle_{1 \leq n}^T.$$

Тогда можно проверить, что  $Y = Q \cdot X$  и  $X = P \cdot Y$  формально. Причем компоненты  $Y$  — это асимптотические ряды, а компоненты  $X$  — это сходящиеся ряды. См. формулы (15), (28).

Применив преобразование  $P$  к ряду (34), получим (после некоторых преобразований) формулу Стирлинга в асимптотической форме:

$$\ln \Gamma(x+1) = \frac{1}{2} \ln(2\pi) - x(1 - \ln x) + \frac{1}{2} \ln(x) + \sum_{k=1}^{\infty} \frac{B(k+1)}{k(k+1)x^k}. \quad (35)$$

Мы получили формулу Стирлинга в классической форме. Ей посвящено большое количество исследований. Расходящийся ряд в (35) — это функция Бине (Binet)  $J(x)$  (или  $\mu(x)$ ). Она имеет интегральное представление

$$J(x) = \sum_{k=1}^{\infty} \frac{B(k+1)}{k(k+1)x^k} = \frac{1}{\pi} \int_0^{\infty} \ln \left( \frac{1}{1 - \exp(-2\pi t)} \right) \frac{x dt}{x^2 + t^2}.$$

Другое интегральное представление функции Бине можно получить из формулы (34), заменив полигамма функции на их интегральные представления, как мы делали в разд. 3.

Расходящийся ряд в формуле (35) также можно преобразовать к сходящимся факториальным рядам (см. [2], [10]). Однако это не дает никаких преимуществ, так как в коэффициентах факториальных рядов не удастся выявить никаких закономерностей.

Эта же проблема возникает, если преобразовать ряд в (35) в (сходящуюся) цепную дробь, что делал еще Стилтес (см., например, [14] и ссылки там) и нашел этот процесс “très pénible”.

Если дифференцировать формулу (35) по  $x$ , то мы получим известные асимптотические разложения для полигамма функций. Однако и для них, за исключением  $\Psi(1, x+1)$  и  $\Psi(2, x+1)$ , не удастся получить цепные дроби с “квазипериодической” структурой (см. также [2]).

Для факториальных рядов та же картина, за исключением ряда для  $\gamma = -\Psi(1)$ , который мы нашли в [2] и который имеет хотя и сложную, но предсказуемую структуру.

Если же дифференцировать формулу (34) по  $x$ , то мы получим сходящиеся разложения полигамма функций. Поскольку значения этих функций в натуральных аргументах выражаются с помощью  $\zeta$ -функции Римана, то мы получаем бесконечный набор тождеств для  $\zeta$ -функции Римана. Введем вектор-функцию

$$W = \langle w(n) \rangle_{1 \leq n}^T, \quad w(n) = \frac{\zeta(n) - 1}{n}, \quad n \in \mathbb{N},$$

где  $\zeta(n)$  определена в (26). Тогда получим  $G = A \cdot W$  и  $W = A^{-1} \cdot G$ , где  $A$  и  $A^{-1}$  — это бесконечные верхне-треугольные матрицы

$$A = \left| \frac{j!}{(j-i+1)!} \right|_{1 \leq i \leq j}, \quad A^{-1} = \left| \frac{B(j-i)}{i!(j-i)!} \right|_{1 \leq i \leq j},$$

а вектор  $G$  имеет вид

$$G = \langle 0, 1, 1, 2, 6, \dots, (i-2)!, \dots \rangle^T.$$

Например, первая компонента вектора  $G$  даст первую формулу (17), следующие компоненты дадут

$$\sum_{j=2}^{\infty} (\zeta(j) - 1) = 1, \quad \sum_{j=3}^{\infty} (\zeta(j) - 1)(j-1) = 1, \quad \sum_{j=4}^{\infty} (\zeta(j) - 1)(j-1)(j-2) = 2, \dots$$

Эти формулы, вероятно, известны, но мы не смогли найти их в литературе.

## 7. ЗАКЛЮЧЕНИЕ

Мы представили метод суммирования сходящихся и расходящихся рядов, который является аналогом методов Эйлера–Маклорена и Эйлера–Буле, но проще их в реализации и выводе. Метод ориентирован на использование имеющегося весьма развитого инструментария CAS и позволяет в ряде случаев суммировать ряды чисто алгебраическими средствами.

Понятие суммы ряда получило толкование, которое не использует понятие предела, что позволяет рассматривать сходящиеся и расходящиеся (в обычном смысле) ряды с единой позиции.

Основной вычислительной задачей при этом является вычисление асимптотических разложений решений некоторых классов функциональных уравнений. Эта задача в общем случае является весьма сложной и малоисследованной, но во многом аналогична задаче вычисления асимптотических разложений решений ОДУ.

В частности, помимо обычных асимптотических рядов, которые мы получали в данной статье, могут появляться т.н.  $\psi$ -ряды (см. [15]), а также более сложные разложения, известные под названием “трансряды” (см. [16]). Развитие этой теории тесно связано с развитием CAS.

Суммирование медленно сходящихся рядов со сложно устроенным общим членом было бы, вероятно, непрактично осуществлять с помощью асимптотических разложений. Однако наш метод суммирования позволяет сделать это численно, путем численного решения соответствующих функциональных уравнений.

К этой теме мы намерены вернуться в одной из следующих публикаций.

## СПИСОК ЛИТЕРАТУРЫ

1. Харди Г. Расходящиеся ряды. М.: Ленанд, 1951.
2. Варин В.П. Преобразование последовательностей в доказательствах иррациональности некоторых фундаментальных констант // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 9. С. 3–30.
3. Lanczos C. Applied Analysis. Dover Publications. New-York, 1956.
4. Варин В.П. Об интерполяции некоторых рекуррентных последовательностей // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 6. С. 913–925.
5. Варин В.П. Инвариантные кривые некоторых дискретных динамических систем // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 2. С. 199–216.
6. Olver F.W.J. Asymptotics and special functions. Acad. Press, New-York, 1974.
7. Candelpergher B. Ramanujan Summation of Divergent Series. Lecture Notes in Math., Springer. 2017.
8. Borwein J.M., Calkin N.J., Manna D. Euler–Boole Summation Revisited // The American Mathematical Monthly. 2009. V. 116. № 5. P. 387–412.
9. Arakawa T., Ibukiyama T., Kaneko M. Bernoulli Numbers and Zeta Functions. Springer, Japan. 2014.
10. Варин В.П. Факториальное преобразование некоторых классических комбинаторных последовательностей // Ж. вычисл. матем. и матем. физ. 2018. Т. 59. № 6. С. 1747–1770.
11. Borwein J.M., Bradley D.M., Crandall R.E. Computational strategies for the Riemann zeta function // J. of Computational and Applied Mathematics. 2000. V. 121. P. 247–296.
12. Cohen H., Villegas F.R., Zagier D. Convergence Acceleration of Alternating Series // Experimental Mathematics. 2000. V. 9. № 1. P. 3–12.
13. Sloane online encyclopedia of integer sequences. (<http://oeis.org>).
14. Char B.W. On Stieltjes’ Continued Fraction for the Gamma Function // MATHEMATICS OF COMPUTATION. 1980. V. 34. № 150. P. 547–551.
15. Hille E. Ordinary differential equations in the complex domain. New-York: John Wiley & Sons. (1976).
16. Edgar G.A. Transseries for beginners // [arxiv:0801.4877v5], (2009). (<http://arxiv.org/abs/0801.4877v5>).

## ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.16

# АППАРАТ ГЕОМЕТРИЧЕСКОЙ АЛГЕБРЫ И КВАТЕРНИОНОВ В СИСТЕМАХ СИМВОЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ ОПИСАНИЯ ВРАЩЕНИЙ В ЕВКЛИДОВОМ ПРОСТРАНСТВЕ<sup>1)</sup>

© 2023 г. Т. Р. Велиева<sup>1,\*</sup>, М. Н. Геворкян<sup>1,\*\*</sup>, А. В. Демидова<sup>1,\*\*\*</sup>,  
А. В. Королькова<sup>1,\*\*\*\*</sup>, Д. С. Кулябов<sup>1,2,\*\*\*\*\*</sup>

<sup>1</sup> 117198 Москва, ул. Миклухо-Маклая, 6, Российский университет дружбы народов, Россия

<sup>2</sup> 141980 Дубна, М.о., ул. Жолио-Кюри, 6, Объединенный институт ядерных исследований, Россия

\*e-mail: velieva-tr@rudn.ru

\*\*e-mail: gevorgyan-mn@rudn.ru

\*\*\*e-mail: demidova-av@rudn.ru

\*\*\*\*e-mail: korolkova-av@rudn.ru

\*\*\*\*\*e-mail: kulyabov-ds@rudn.ru

Поступила в редакцию 15.04.2022 г.

Переработанный вариант 15.04.2022 г.

Принята к публикации 17.09.2022 г.

В физических и технических задачах достаточно распространенным в применении математическим аппаратом является тензорный формализм (и его частный случай — векторный формализм). Хотя этот формализм и считается достаточно универсальным и подходящим для описания многих пространств, порой требуется применение специального математического аппарата. Например, задача вращения в трехмерном пространстве достаточно плохо описывается в тензорном представлении, и для ее решения более целесообразно использовать формализм представлений алгебры Клиффорда, в частности, кватернионов и геометрической алгебры. В статье средствами компьютерной алгебры демонстрируется решение задачи вращения в трехмерном пространстве с использованием как формализма кватернионов, так и формализма геометрической алгебры. Показано, что при всей принципиальной схожести формализмов кватернионов и геометрической алгебры последний представляется более наглядным как при проведении вычислений, так и при интерпретации результатов. Библ. 32. Фиг. 1.

**Ключевые слова:** геометрическая алгебра, кватернионы, компьютерная алгебра, мультивектор, вращения в трехмерном пространстве.

**DOI:** 10.31857/S0044466923010143, **EDN:** LEANHI

## 1. ВВЕДЕНИЕ

Задача описания вращения твердого тела в трехмерном пространстве восходит еще к Эйлеру [1] и актуальна для многих физических и инженерных задач. Основная задача сводится к вращению трехмерного вектора вокруг начала декартовой системы координат. Можно выделить два классических подхода к ее решению. Первый подход опирается на теорию ортогональных операторов и представляет оператор вращения в виде матрицы, а второй подход основан на применении математического аппарата кватернионов.

Матричный подход имеет ряд недостатков, основным из которых является так называемый “эффект складывания рамок” (gimbal lock) [2], который приводит к вычислительной неустойчивости. Также в силу теоремы Эйлера любое вращение твердого тела может быть представлено как последовательность не более чем трех вращений вокруг координатных осей, где никакие два вращения подряд не производятся вокруг одной и той же оси координат [3].

Учитывая это утверждение, а также некоммутативность вращений, получаем, что все возможные тройки вращений насчитывают двенадцать комбинаций.

<sup>1)</sup> Работа выполнена при финансовой поддержке Программы стратегического академического лидерства РУДН.

Наличие двенадцати комбинаций приводит к тому, что в инженерной или компьютерной системе приходится реализовывать или все двенадцать вариантов, или же выбирать одну конкретную комбинацию. Первый подход очень трудоемкий, поэтому исторический прижился второй вариант, что привело к использованию разных комбинаций в разных областях с разной терминологией.

Приведем три примера. В аэронавтике [3] используется последовательность вращений  $ZYX$  (вокруг оси  $OZ$ , затем вокруг оси  $OY$  и, наконец, вокруг оси  $OX$ ) и углы носят названия крена, тангажа и рыскания. В астрономии используются несколько последовательностей. Так, последовательность углов Эйлера для орбит небесных тел [3] описывает вращение  $ZXY$  и углы именуется восхождением, склонением и широтой. Другая последовательность —  $ZYX$  носит название эфемерид [3] с названием углов долгота, широта и направление пути эфемериды. Хотя эта комбинация похожа на ту, которая используется в аэронавтике, но угол поворота вокруг оси  $OY$  (широта) берется с отрицательным знаком.

Эти недостатки привели к тому, что в инженерных задачах, в частности в области компьютерной графики, матричный подход был практически вытеснен кватернионным способом описания вращений, который свободен от эффекта складывания рамок и требует задания четырех параметров: трех компонент направляющего вектора оси вращения и угла вращения вокруг этой оси, из которых только три параметра являются независимыми, так как вектор оси вращения должен быть единичным. Заметим, что абсолютно твердое тело с одной закрепленной точкой имеет три степени свободы. Следовательно, для параметризации вращения формально достаточно задать только три параметра.

Хотя кватернионы и лишены недостатков вращений с помощью матриц с вычислительной точки зрения, они имеют некоторые недостатки с точки зрения геометрической интерпретации. Так, например, сложно дать геометрически ясное объяснение удвоению угла вращения.

В данной статье мы проводим сравнительный анализ кватернионного способа описания поворотов с альтернативным (сравнительно новым) способом, который начал развиваться лишь в XXI веке. Речь идет об описании поворотов с помощью алгебры мультивекторов, которая рассматривается в рамках геометрической алгебры. Сравнение проводится с использованием систем компьютерной алгебры.

Геометрическая алгебра основывается на работах Г.Г. Грассмана [4], У.Р. Гамильтона [3] и У.К. Клиффорда [5]. В геометрической алгебре рассматривается конкретная реализация алгебры Клиффорда, основанная на мультивекторах, которая включает в себя реализацию алгебры Грассмана в виде внешней алгебры  $p$ -векторов (контравариантных антисимметричных тензоров с операцией внешнего умножения). Алгебра кватернионов также является частным случаем алгебры мультивекторов.

Долгое время работы Клиффорда не привлекали особого внимания физиков и математиков. Алгебра Клиффорда была известна [6], но математический аппарат не был разработан. Д. Хинстейн [7], судя по всему, является первым исследователем, который формализовал разработки Клиффорда в виде современного математического аппарата. Основная волна исследований на эту тему началась уже в XXI веке.

Первоначально алгебра мультивекторов формулировалась с прицелом на использование в физике [8], [9], однако довольно быстро она обрела большую популярность у специалистов по компьютерной графике [10], [11].

В настоящее время сведения из геометрической алгебры входят во многие учебники по компьютерной графике [12], [13], а методы описания вращений с помощью бивекторов могут потенциально вытеснить кватернионы и бикватернионы [14].

В первых двух частях данной статьи авторы конспективно излагают основные формулы математического аппарата кватернионов и дают более подробное изложение алгебры мультивекторов, ограничиваясь только трехмерным пространством. В третьей части приводится конкретный пример поворота вектора на заданный угол с использованием кватернионов и алгебры мультивекторов. Вычисления для случая кватернионов проводятся с помощью пакета компьютерной алгебры SymPy [15] для языка Python [16], а в случае алгебры мультивекторов — с помощью модуля Galgebra [17], основанного на SymPy. Авторы умышленно использовали простой пример, так как он позволяет не отвлекаться на технические детали, и обсудить разницу между подходами, которая в большей степени относится к области теории и геометрической интерпретации.

## 2. КВАТЕРНИОНЫ

### 2.1. Определение и основные операции

Кватернионом называют гиперкомплексное число вида:  $q = a + bi + cj + dk$ , где  $a, b, c, d$  — действительные числа, а  $i, j, k$  — кватернионные мнимые единицы, определяемые соотношением  $i^2 = j^2 = k^2 = ijk = -1$ .

Выделяют скалярную часть кватерниона  $q_0 = a$  и векторную часть  $\mathbf{q} = bi + cj + dk$ , называемую также *чисто мнимой* или просто *чистой*. Каждому чисто мнимому кватерниону можно взаимно однозначно сопоставить вектор в трехмерном пространстве. Кватернион часто записывают в виде  $q = q_0 + \mathbf{q}$ .

Кватернионы образуют векторное пространство размерностью четыре над полем вещественных чисел. Кватернионы получаются из эллиптических комплексных чисел посредством процедуры Кэли–Диксона [18]. Согласно модели Кэли–Клайна [19]–[21], для проективных пространств задается три типа мероопределения: эллиптическое, параболическое и гиперболическое. Каждому типу мероопределения соответствует свой тип комплексных чисел [22], [23]. Наше евклидово пространство соответствует эллиптическому мероопределению углов, поэтому для его описания и используются эллиптические комплексные числа (это привычные нам комплексные числа). Например галилеево пространство имеет параболическое мероопределение углов, а пространство Минковского — гиперболическое. Для корректного описания этих пространств следует применять соответствующие комплексные числа.

Множество кватернионов часто обозначают через  $\mathbb{H}$  в честь ирландского математика У.Р. Гамильтона (1805–1865 г.), разработавшего теорию кватернионов [24]–[26].

Для двух кватернионов  $q_1 = a_1 + b_1i + c_1j + d_1k$  и  $q_2 = a_2 + b_2i + c_2j + d_2k$  сложение определяется следующим образом:

$$q_1 + q_2 = (a_1 + a_2) + (b_1 + b_2)i + (c_1 + c_2)j + (d_1 + d_2)k.$$

Для умножения кватернионов необходимо составить таблицу умножения для мнимых единиц  $i, j, k$ , которая выводится из формулы  $i^2 = j^2 = k^2 = ijk = -1$ :

|   |    |    |    |
|---|----|----|----|
|   | i  | j  | k  |
| i | -1 | k  | -j |
| j | -k | -1 | i  |
| k | j  | -i | -1 |

(1)

Так как операция умножения кватернионов в общем случае не коммутативна, следует отметить, что в строке таблицы указан первый сомножитель, а в столбце — второй.

Используя таблицу (1), можно найти результат кватернионного умножения, который удобно выразить через скалярное и векторное произведения:

$$q_1 \cdot q_2 = a_1a_2 - (\mathbf{q}_1, \mathbf{q}_2) + a_2\mathbf{q}_1 + a_1\mathbf{q}_2 + \mathbf{q}_1 \times \mathbf{q}_2. \quad (2)$$

Кватернион  $q^*$  называется *сопряженным* к кватерниону  $q = q_0 + \mathbf{q} = q_0 + q_1i + q_2j + q_3k$ , если  $q^* = q_0 - \mathbf{q} = q_0 - q_1i - q_2j - q_3k$ . Норма векторной части кватерниона  $\|\mathbf{q}\|$  совпадает с нормой вектора  $\mathbf{q}$  и вычисляется как  $\|\mathbf{q}\| = \sqrt{(\mathbf{q}, \mathbf{q})} = \sqrt{q_1^2 + q_2^2 + q_3^2}$ . Модуль кватерниона определяется формулой

$$|q| = \sqrt{qq^*} = \sqrt{q_0^2 + \|\mathbf{q}\|^2} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}.$$

Сопряженный кватернион позволяет найти обратный по умножению кватернион  $q^{-1}$  и ввести операцию деления кватернионов:

$$q^{-1} = \frac{q^*}{qq^*} = \frac{q^*}{|q|^2}, \quad \frac{q_1}{q_2} = q_1q_2^{-1} = \frac{q_1q_2^*}{|q_2|^2}.$$

## 2.2. Чистый и единичный кватернионы

Рассмотрим два чистых кватерниона  $q_1 = 0 + \mathbf{q}_1$  и  $q_2 = 0 + \mathbf{q}_2$  и перемножим их по формуле (2):

$$q_1 \cdot q_2 = (0 + \mathbf{q}_1)(0 + \mathbf{q}_2) = -(\mathbf{q}_1, \mathbf{q}_2) + \mathbf{q}_1 \times \mathbf{q}_2.$$

Кватернионное произведение чистого кватерниона на самого себя дает норму вектора  $q$  со знаком минус:

$$qq = q^2 = (0 + \mathbf{q})(0 + \mathbf{q}) = -(\mathbf{q}, \mathbf{q}) + \underbrace{\mathbf{q} \times \mathbf{q}}_{=0} = -(\mathbf{q}, \mathbf{q}) - \mathbf{q}^2.$$

Единичным или нормированным называется кватернион с модулем, равным единице:  $|q| = q_0^2 + \|\mathbf{q}\|^2 = 1$ , где  $q = q_0 + q_1i + q_2j + q_3k$ .

Для единичного кватерниона можно записать так называемую “тригонометрическую форму”. Всегда существует такое число  $\vartheta \in [0, \pi)$ , что  $q_0^2 = \cos^2 \vartheta$ ,  $\|\mathbf{q}\|^2 = \sin^2 \vartheta$ . Следовательно,  $q_0^2 + \|\mathbf{q}\|^2 = \cos^2 \vartheta + \sin^2 \vartheta = 1$  и единичный кватернион записывается в виде

$$q = \cos \vartheta + \mathbf{u} \sin \vartheta = \cos \vartheta + \sin \vartheta u_1 i + \sin \vartheta u_2 j + \sin \vartheta u_3 k, \quad \mathbf{u} = \frac{\mathbf{q}}{\|\mathbf{q}\|}.$$

Единичный чистый кватернион  $u = 0 + \mathbf{u}$  обладает свойством эллиптической мнимой единицы [22]:

$$uu = u^2 = (0 + \mathbf{u})(0 + \mathbf{u}) = -\underbrace{(\mathbf{u}, \mathbf{u})}_{\|\mathbf{u}\|^2} + \underbrace{\mathbf{u} \times \mathbf{u}}_0 - \|\mathbf{u}\|^2 = -1,$$

что дает возможность определить экспоненту от единичного чистого кватерниона по аналогии с формулой Эйлера для комплексных чисел:

$$e^{u\vartheta} = \cos \vartheta + \mathbf{u} \sin \vartheta.$$

## 2.3. Использование кватернионов для вращения

Замечательным свойством кватернионов оказалась возможность использовать их для задания оператора вращения вектора  $\mathbf{v}$  вокруг оси с единичным направляющим вектором  $\mathbf{u}$  (чистый кватернион  $u = 0 + \mathbf{u}$ ). Результатом вращения является вектор  $\mathbf{w}$  ( $w = 0 + \mathbf{w}$ ):

$$\mathbf{w} = \exp\left(\frac{\vartheta}{2}\mathbf{u}\right)\mathbf{v}\exp\left(-\frac{\vartheta}{2}\mathbf{u}\right) = q\mathbf{v}q^*, \quad (3)$$

где

$$q = \exp\left(\frac{\vartheta}{2}\mathbf{u}\right) = \cos \frac{\vartheta}{2} + \sin \frac{\vartheta}{2}\mathbf{u} \quad \text{и} \quad q^* = \exp\left(-\frac{\vartheta}{2}\mathbf{u}\right) = \cos \frac{\vartheta}{2} - \sin \frac{\vartheta}{2}\mathbf{u}.$$

Формулу (3) часто называют *сэндвич*-оператором. Доказательство формулы см., например, в работах [1]–[3].

## 3. АЛГЕБРА МУЛЬТИВЕКТОРОВ В ТРЕХМЕРНОМ ПРОСТРАНСТВЕ

Геометрическая алгебра является реализацией абстрактной алгебры Клиффорда [5], где элементами являются *мультивекторы*, а операцией “умножения” является операция *геометрического умножения*.

Мультивектор является градуированным объектом, представляющим собой линейную комбинацию разложимых  $p$ -векторов (кососимметричных ковариантных тензоров). Сами  $p$ -векторы совместно с операцией внешнего произведения  $\wedge$  являются реализацией абстрактной алгебры Грассмана [4]. Основным отличительным свойством внешнего произведения векторов является свойство *кососимметричности*:  $\mathbf{v} \wedge \mathbf{v} = 0$ . Более подробно о внешней алгебре см. [9], [12], [27].

Будем рассматривать трехмерное евклидово пространство  $\mathbb{E}$  с ортонормированным базисом  $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ . Пространство  $E$  и операция внешнего произведения  $\wedge$  порождают три внешние алгебры

ры: алгебру 1-векторов (совпадает с  $E$ ), алгебру бивекторов  $\Lambda^2(E)$  и алгебру тривекторов  $\Lambda^3(L)$ . Алгебры большого ранга (градуировки, grade) тривиальны и состоят из скалярных нулей. Заметим, что скаляр также считается  $p$ -вектором с  $p = 0$ .

В  $\Lambda^2(E)$  существует три базисных бивектора:  $\mathbf{e}_{12} = \mathbf{e}_1 \wedge \mathbf{e}_2$ ,  $\mathbf{e}_{13} = \mathbf{e}_1 \wedge \mathbf{e}_3$  и  $\mathbf{e}_{23} = \mathbf{e}_2 \wedge \mathbf{e}_3$ , и любой бивектор представим в виде:

$$\mathbf{U} = u^{12} \mathbf{e}_1 \wedge \mathbf{e}_2 + u^{23} \mathbf{e}_2 \wedge \mathbf{e}_3 + u^{13} \mathbf{e}_1 \wedge \mathbf{e}_3,$$

где  $u^{12}$ ,  $u^{13}$  и  $u^{23}$  — действительные числа, компоненты бивектора. В трехмерном пространстве любой бивектор *разложим*, т.е. может быть представлен как внешнее произведение двух векторов:  $\mathbf{U} = \mathbf{a} \wedge \mathbf{b}$ . Геометрической интерпретацией бивектора может служить ориентированная площадь параллелограмма, построенного на векторах  $\mathbf{a}$  и  $\mathbf{b}$ .

В  $\Lambda^3(E)$  существует только один базисный тривектор  $\mathbf{e}_{123} = \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$  и любые другие тривекторы имеют вид:  $u^{123} \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$ . В трехмерном пространстве любой тривектор также разложим и может быть представлен как  $\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}$ , и геометрической интерпретацией для него является ориентированный объем параллелепипеда, построенного на векторах  $\mathbf{a}$ ,  $\mathbf{b}$  и  $\mathbf{c}$ .

Операцию *геометрического умножения* для трехмерных векторов  $\mathbf{u}$  и  $\mathbf{v}$  можно определить конструктивно следующим образом:

$$\mathbf{uv} = (\mathbf{u}, \mathbf{u}) + \mathbf{u} \wedge \mathbf{v},$$

т.е. как сумму скалярного и внешнего произведений. Отметим, что геометрическое произведение обычно никаким знаком не обозначается.

Для ортогонального базиса из определения геометрического умножения выводятся следующие два важных соотношения:  $\mathbf{e}_i \mathbf{e}_j = -\mathbf{e}_j \mathbf{e}_i$ ,  $i \neq j$  и  $\mathbf{e}_i \mathbf{e}_i = 1$ . Также  $\mathbf{e}_i \mathbf{e}_j = \mathbf{e}_i \wedge \mathbf{e}_j$  и  $\mathbf{e}_i \mathbf{e}_j \mathbf{e}_k = \mathbf{e}_i \wedge \mathbf{e}_j \wedge \mathbf{e}_k$ , где  $i, j, k = 1, 2, 3$ . Это позволяет представить любой мультивектор как линейную комбинацию базисных  $p$ -векторов, записанных через геометрическое умножение. Так, в трехмерном пространстве мультивектор в общем случае записывается в следующей форме:

$$\mathbf{U} = u^0 + u^1 \mathbf{e}_1 + u^2 \mathbf{e}_2 + u^3 \mathbf{e}_3 + u^{12} \mathbf{e}_1 \mathbf{e}_2 + u^{23} \mathbf{e}_2 \mathbf{e}_3 + u^{13} \mathbf{e}_1 \mathbf{e}_3 + u^{123} \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3.$$

Можно определить обратный относительно геометрического умножения вектор в виде  $\mathbf{u}^{-1} = \frac{\mathbf{u}}{\|\mathbf{u}\|^2}$ . Легко проверить, что  $\mathbf{u} \mathbf{u}^{-1} = 1$ . Если мультивектор представлен как произведение двух векторов  $\mathbf{ab}$ , то  $(\mathbf{ab})^{-1} = \mathbf{b}^{-1} \mathbf{a}^{-1}$ .

Базисный тривектор  $\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3$  обозначим как  $\mathbf{I}$ , так как он обладает свойствами эллиптической мнимой единицы [22]:

$$\mathbf{I} = \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3 \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3 = \mathbf{e}_1 \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_2 \mathbf{e}_3 \mathbf{e}_3 = -\mathbf{e}_1 \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_2 \mathbf{e}_3 \mathbf{e}_3 = -1 \Rightarrow \mathbf{I}^2 = -1.$$

Тривектор  $\mathbf{I}$  коммутирует со всеми базисными векторами и бивекторами:  $\mathbf{I} \mathbf{e}_i = \mathbf{e}_i \mathbf{I}$ , где  $i = 1, 2, 3$ , и  $\mathbf{I} \mathbf{e}_i \mathbf{e}_j = \mathbf{e}_i \mathbf{e}_j \mathbf{I}$ , где  $i, j = 1, 2, 3$ ,  $i < j$ . Из этого следует, что произвольный мультивектор  $\mathbf{U}$  в трехмерном пространстве коммутирует с  $\mathbf{I}$ :  $\mathbf{IU} = \mathbf{UI}$ .

Также с помощью  $\mathbf{I}$  можно определить операцию дополнения (подробнее см. [12]):

$$\bar{\mathbf{u}} = \mathbf{I} \mathbf{u} \quad \text{и} \quad \bar{\mathbf{U}} = -\mathbf{I} \mathbf{U},$$

где  $\mathbf{u}$  — вектор, а  $\mathbf{U}$  — бивектор. Данная операция позволяет найти дополнение к вектору или бивектору до элемента максимального ранга в данном пространстве. Так, для вектора в трехмерном пространстве это будет бивектор, а для бивектора — вектор.

Если обозначить  $\mathbf{i} = \mathbf{e}_1\mathbf{e}_2$ ,  $\mathbf{j} = \mathbf{e}_2\mathbf{e}_3$  и  $\mathbf{k} = \mathbf{e}_1\mathbf{e}_3$ , то  $\mathbf{ijk} = -1$  и таблица умножения бивекторов превратится в таблицу умножения кватернионных мнимых единиц:

$$\begin{array}{c|ccc} & \mathbf{e}_{12} & \mathbf{e}_{13} & \mathbf{e}_{23} \\ \hline \mathbf{e}_{12} & -1 & -\mathbf{e}_{23} & \mathbf{e}_{13} \\ \mathbf{e}_{13} & \mathbf{e}_{23} & -1 & -\mathbf{e}_{12} \\ \mathbf{e}_{23} & -\mathbf{e}_{13} & \mathbf{e}_{12} & -1 \end{array} \Rightarrow \begin{array}{c|ccc} & \mathbf{i} & \mathbf{k} & \mathbf{j} \\ \hline \mathbf{i} & -1 & -\mathbf{j} & \mathbf{k} \\ \mathbf{k} & \mathbf{j} & -1 & -\mathbf{i} \\ \mathbf{j} & -\mathbf{k} & \mathbf{i} & -1 \end{array} \Rightarrow \begin{array}{c|ccc} & \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \hline \mathbf{i} & -1 & \mathbf{k} & -\mathbf{j} \\ \mathbf{j} & -\mathbf{k} & -1 & \mathbf{i} \\ \mathbf{k} & \mathbf{j} & -\mathbf{i} & -1 \end{array}$$

Таким образом в трехмерном пространстве множество мультивекторов вида

$$\mathbf{V} = v^0 + \mathbf{v}_2 = v^0 + v^{12}\mathbf{e}_{12} + v^{13}\mathbf{e}_{13} + v^{23}\mathbf{e}_{23} = v^0 + v^{12}\mathbf{i} + v^{13}\mathbf{k} + v^{23}\mathbf{j},$$

изоморфно множеству кватернионов. Стоит обратить внимание на то, что в двухмерном пространстве мультивекторы такого вида были изоморфны множеству комплексных чисел. Стоило повысить размерность пространства на единицу, как те же мультивекторы стали изоморфны кватернионам. Это фактически соответствует процедуре Кэли–Диксона [18].

### 3.1. Повороты в пространстве

Рассмотрим два вектора в трехмерном пространстве  $\mathbf{a}$  и  $\mathbf{b}$ . Угол между ними равен  $\vartheta$ . Бивектор  $\mathbf{a} \wedge \mathbf{b}$  задает плоскость вращения. Мультивектор  $\mathbf{R} = \mathbf{ba}$  называется *ротором* и определяет оператор вращения вектора  $\mathbf{v}$  по формуле

$$\mathbf{w} = \mathbf{RvR}^{-1},$$

где  $\mathbf{R} = \mathbf{ba} = (\mathbf{a}, \mathbf{b}) - \mathbf{a} \wedge \mathbf{b}$ , а  $\mathbf{w}$  — вектор, в который перешел вектор  $\mathbf{v}$  в результате вращения. У данной формулы есть наглядная интерпретация — двойное отражение вектора  $\mathbf{v}$  сначала относительно вектора  $\mathbf{a}$ , а затем относительно вектора  $\mathbf{b}$ .

Формула упрощается, если использовать нормированный мультивектор  $\mathbf{R}$ , т.е. такой, что  $\|\mathbf{R}\| = \mathbf{RR}^* = 1$ , где сопряженный ротор вычисляется как  $\mathbf{R}^* = (\mathbf{a}, \mathbf{b}) + \mathbf{a} \wedge \mathbf{b}$ . Тогда  $\mathbf{R}^{-1} = \mathbf{R}^*$  и  $\mathbf{w} = \mathbf{RvR}^*$ .

Рассмотрим произвольный мультивектор  $\mathbf{U}$ , состоящий из скалярной и бивекторной частей в трехмерном пространстве:

$$\mathbf{U} = u^0 + \mathbf{a} = u^0 + a^{12}\mathbf{e}_{12} + a^{13}\mathbf{e}_{13} + a^{23}\mathbf{e}_{23},$$

где бивектор помечен числом 2. Сопряженный мультивектор  $\mathbf{U}^*$  имеет вид

$$\mathbf{U}^* = u^0 + \mathbf{a}^* = u^0 + a^{12}\mathbf{e}_{21} + a^{13}\mathbf{e}_{31} + a^{23}\mathbf{e}_{32} = u^0 - a^{12}\mathbf{e}_{12} - a^{13}\mathbf{e}_{13} - a^{23}\mathbf{e}_{23} = u^0 - \mathbf{a}.$$

Вычислим норму мультивектора  $\|\mathbf{U}\| = \mathbf{UU}^*$ :

$$\|\mathbf{U}\| = \mathbf{UU}^* = (u^0)^2 + (a^{12})^2 + (a^{13})^2 + (a^{23})^2.$$

Если предположить, что мультивектор  $\mathbf{U}$  имеет единичную норму, то без потери общности его можно представить в тригонометрическом виде:

$$\mathbf{U} = \cos \vartheta + \sin \vartheta (u^{12}\mathbf{e}_{21} + u^{13}\mathbf{e}_{31} + u^{23}\mathbf{e}_{32}) = \cos \vartheta + \sin \vartheta \mathbf{u}_2,$$

где  $(u^{12})^2 + (u^{13})^2 + (u^{23})^2 = 1$ . По аналогии с формулой Эйлера для комплексных чисел его можно записать и в экспоненциальном виде, определив экспоненту от бивектора следующим образом:

$$\exp\left(\vartheta \mathbf{u}_2\right) = \cos \vartheta + \sin \vartheta \mathbf{u}_2.$$

Ротор  $\mathbf{R}$  также представим в тригонометрическом (или экспоненциальном) виде, если  $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$ :

$$\mathbf{R} = (\mathbf{a}, \mathbf{b}) - \mathbf{a} \wedge \mathbf{b} = \exp\left(\frac{\vartheta}{2} \mathbf{u}_2\right) = \cos \frac{\vartheta}{2} - \sin \frac{\vartheta}{2} \mathbf{u}_2,$$

где  $(\mathbf{a}, \mathbf{b}) = \|\mathbf{a}\| \|\mathbf{b}\| \cos \frac{\vartheta}{2} \cos \frac{\vartheta}{2}$ , а компоненты бивектора  $\mathbf{u}$  связаны с бивектором  $\mathbf{a} \wedge \mathbf{b}$  следующим образом:

$$\begin{aligned} a^1 b^2 - a^2 b^1 &= \sin \frac{\vartheta}{2} u^{12}, \\ a^1 b^3 - a^3 b^1 &= \sin \frac{\vartheta}{2} u^{13}, \\ a^2 b^3 - a^3 b^2 &= \sin \frac{\vartheta}{2} u^{23}. \end{aligned}$$

Если задана ось вращения с направляющим вектором  $\mathbf{u}$ , то бивектор, задающий плоскость вращения, можно задать, найдя правое или левое дополнение к вектору  $\mathbf{u}$ :

$$\bar{\mathbf{u}} = u^1 \bar{\mathbf{e}}_1 + u^2 \bar{\mathbf{e}}_2 + u^3 \bar{\mathbf{e}}_3 = u^1 \mathbf{e}_{23} - u^2 \mathbf{e}_{13} + u^3 \mathbf{e}_{12}.$$

## 4. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ВРАЩЕНИЙ С ПОМОЩЬЮ КОМПЬЮТЕРНОЙ АЛГЕБРЫ

### 4.1. Описание существующих программных модулей

Рассмотрим программную реализацию вращений с помощью кватернионов и мультивекторов. Для этого используем систему компьютерной алгебры SymPy [15].

Для языка Python существует несколько реализаций кватернионов. Можно выделить два наиболее проработанных модуля: `Quaternions` [28] и `pyquaternion` [29]. Оба этих модуля ориентированы на численные вычисления. Первый из них тесно интегрирован с библиотекой NumPy и поддерживает программный интерфейс этой библиотеки. Использование NumPy позволяет повысить производительность и легко преобразовывать одномерные массивы из четырех элементов в кватернионы и обратно.

Модуль `pyquaternion` также реализует численные вычисления для кватернионов, но реализован без использования внешних библиотек, что делает его более компактным, однако менее производительным в случае высокой интенсивности вычислений.

Для символьных вычислений с кватернионами существует встроенный в SymPy подмодуль `sympy.algebras`, в котором реализован класс `Quaternion`. В нем реализованы все алгебраические операции с кватернионами, включая нахождение модуля, нормы чистого кватерниона, сопряженного кватерниона и ряд других простых операций.

Мультивекторы и операции геометрической алгебры определены в модуле `Galgebra` [17], который необходимо установить отдельно. В своей работе он использует SymPy и хорошо интегрируется с функциями из него. При использовании оболочки Jupyter Notebook или Jupyter Lab задача результата вычислений будет осуществляться в формате LaTeX и выглядеть как обычная формула, поэтому в данной статье мы приводим результаты работы функций сразу в виде формул.

Также можно указать на смешанный, символьно-численный подход к описанию формализма геометрической алгебры, реализованный на языке Julia [30].

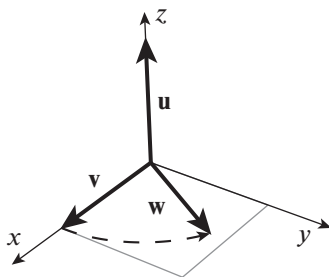
### 4.2. Пример вращения с помощью кватернионов

Рассмотрим решение задачи поворота вектора  $\mathbf{v} = (1, 0, 0)^T$  на угол  $\frac{\pi}{3}$  вокруг оси  $Oz$  [3]. Поворот изображен на фиг. 1.

Пример сознательно выбран так, чтобы можно было проверить правильность решения визуально, так как результат поворота интуитивно очевиден.

Вектору  $\mathbf{v}$  будет соответствовать чистый кватернион  $v = 0 + i$ . Единичный кватернион поворота  $q$  записывается в виде

$$q = \cos \frac{\pi}{6} + \mathbf{u} \sin \frac{\pi \sqrt{3}}{6 \cdot 2} + \frac{1}{2} \mathbf{u},$$



Фиг. 1. Поворот  $v = e_x$  вокруг  $Oz$  на угол, равный  $60^\circ$ .

где векторная часть  $\mathbf{u}$  является чистым кватернионом, задающим ось вращения  $u = 0 + \mathbf{e}_z = 0i + 0j + 1k = k$ , т.е.  $u = k$ . Следовательно,

$$q = \frac{\sqrt{3}}{2} + \frac{1}{2}k \quad \text{и} \quad q^* = \frac{\sqrt{3}}{2} - \frac{1}{2}k.$$

Для использования класса Quaternion импортируем необходимые модули:

```
import sympy as sp
from sympy.algebras import Quaternion
```

Далее зададим угол  $\vartheta$ , равный половине желаемого угла поворота, ось вращения  $\mathbf{u}$  и вектор  $\mathbf{v}$  в виде кватернионов:

```
p = sp.pi
q = p/6
u = Quaternion(0, 0, 0, 1).normalize()
v = Quaternion(0, 1, 0, 0).
```

Метод `normalize` здесь избыточен, но если бы мы задали ось вращения не в виде единичного вектора, то данный метод позволил бы автоматически нормировать кватернион для дальнейшего использования.

Далее зададим кватернион  $q$ . Можно использовать тригонометрический вид:

```
q = sp.cos(q) + sp.sin(q)*u
```

или же экспоненциальный  $q = \exp(\vartheta \mathbf{u})$ , как в формуле (3):

```
q = (q*u).exp()
```

Для вычисления результатов вращения также есть две опции. Первая опция – использование встроенного метода `rotate_point`, который принимает вращаемый вектор в виде кортежа (аргумент `pin`), а не кватерниона, и выдает результат также в виде кортежа

```
Quaternion.rotate_point(pin=(1, 0, 0), r=q)
```

$$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right).$$

Вторая опция – непосредственное использование формулы (3):

```
q*v*q.conjugate()
```

$$\frac{1}{2}i + \frac{\sqrt{3}}{2}j.$$

В результате применения формулы (3) мы получим из кватерниона  $v = i$  кватернион  $w = \frac{1}{2}i + \frac{\sqrt{3}}{2}j$ , которому соответствует вектор  $\mathbf{w} = (1/2, \sqrt{3}/2, 0)$ . Из кватерниона можно извлечь ось вращения и угол вращения с помощью соответствующих методов:

```
q.to_axis_angle()
```

$$\left((0, 0, 1), \frac{\pi}{3}\right),$$

```
q.to_rotation_matrix()
```

$$\begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Отметим также, что хотя мы задавали конкретные численные значения, они обрабатывались в символьном виде. Все числовые значения можно заменить на буквенные, используя функцию `SymPy sp.symbols`.

### 4.3. Краткое описание модуля *Galgebra*

Прежде чем переходить к вычислению поворотов с помощью компьютерной алгебры, опишем основные функции модуля *Galgebra* [17]. Так как *Galgebra* использует символьные вычисления, то для его работы нам понадобится *SymPy*:

```
import sympy as sp
from galgebra.ga import Ga
```

Будем предполагать, что работа ведется в оболочке Jupyter Notebook (или Lab). В этой оболочке весь вывод команд представляется в виде формул в формате LaTeX, поэтому непосредственно после примеров кода мы будем размещать отображаемые оболочкой формулы.

Работа с *Galgebra* начинается с задания линейного пространства и метрики на нем:

```
xy = sp.symbols('1 2', real=True)
o2d = Ga('e', g= [1, 1], coords=xy)
```

Здесь сначала мы определяем символы для индексов базисных векторов. В данном случае задается двумерное пространство, поэтому используются символы 1 и 2. Далее создается объект класса *Ga*, конструктор которого принимает первым аргументом символ для обозначения базисных векторов, а вторым аргументом — диагональную метрику. Отметим, что мы задали евклидову ортонормированную метрику, но можно также задать и метрику пространства Минковского: `g=[1, -1]`.

После создания объекта *Ga* можно приступать к созданию объектов типа мультивектор, используя функцию `mv`, где первый аргумент указывает символ для компонент, а второй — ранг объекта:

```
a_scalar = o2d.mv('a', 0)
a_vect = o2d.mv('a', 1)
a_bivect = o2d.mv('a', 2)
a_scalar, a_vect, a_bivect
```

$$(a, a^1 \mathbf{e}_1 + a^2 \mathbf{e}_2, a^{12} \mathbf{e}_1 \wedge \mathbf{e}_2).$$

Определены операции скалярного `|`, внешнего `^` и геометрического `*` умножений:

```
a_scalar ^ a_vect, a_vect * a_bivect
```

$$(aa^1 \mathbf{e}_1 + aa^2 \mathbf{e}_2, a^{12} a^2 \mathbf{e}_1 - a^1 a^{12} \mathbf{e}_2).$$

Можно создать и сложный мультивектор общего вида:

```
u = o2d.mv('u', 0) + o2d.mv('u', 1) + o2d.mv('u', 2)
```

$$u + u^1 \mathbf{e}_1 + u^2 \mathbf{e}_2 + u^{12} \mathbf{e}_1 \wedge \mathbf{e}_2.$$

В данном модуле также определен ряд операций, относящихся к анализу в рамках геометрической алгебры, т.е. есть возможность задавать поля векторов, бивекторов и т.д.

### 4.4. Пример вращения с помощью мультивектора

Рассмотрим тот же самый пример, но осуществим вращение с помощью мультивектора **R**.

Прежде всего импортируем необходимые модули:

```
import sympy as sp
from galgebra.ga import Ga
```

и зададим базис евклидова пространства, создав объект класса GA:

```
xyz = sp.symbols('1 2 3', real=True)
o3d = Ga('e_1 e_2 e_3', g= [1, 1, 1], coords=xyz).
```

Сначала зададим символы, которые будут служить индексами базисных векторов, а затем зададим евклидово пространство, указывая в аргументе *g* диагональный метрический тензор, обозначения для базисных векторов и обозначения для индексов в аргументе *coords*. Эти же индексы будут использоваться в компонентах векторов, бивекторов и тривекторов.

Затем присвоим отдельным переменным значение базисных векторов и базисного тривектора  $e_{123}$ :

```
e1, e2, e3 = o3d.mv()
I = e1*e2*e3.
```

Далее зададим угол, равный половине желаемого угла вращения, ось вращения и вращаемый вектор:

```
p = sp.pi; q = p/6; u = e3; v = e1.
```

Так как ротор  $\mathbf{R}$  является бивектором, то для его вычисления находим дополнение к вектору  $\mathbf{u}$  и бивектор  $\mathbf{e}_{12}$ , что согласуется с геометрическим представлением, так как  $\mathbf{e}_{12}$  интерпретируется как плоскость вращения  $Oxy$ . Заметим, что в модуле *Galgebra* не реализована операция дополнения в явном виде, поэтому используем умножение на  $\mathbf{I}$ :

```
U = I*u
```

Осталось составить ротор  $\mathbf{R}$ , который будет иметь вид  $\sqrt{3}/2 - \mathbf{e}_{12}/2$ :

```
R = sp.cos(q) - sp.sin(q)*U
```

и применить формулу (3.1):

```
R*v*R.inv()
```

$$\frac{1}{2}\mathbf{e}_1 + \frac{\sqrt{3}}{2}\mathbf{e}_2.$$

Мы также можем проверить нормированность мультивектора  $\mathbf{R}$  и вычислить обратный и сопряженный мультивекторы:

```
R.norm(), R.inv(), ~R
```

$$1, \frac{\sqrt{3}}{2} + \frac{1}{2}\mathbf{e}_{12}, \frac{\sqrt{3}}{2} + \frac{1}{2}\mathbf{e}_{12}.$$

Продемонстрированный пример реализации вращения является несколько упрощенным, но на наш взгляд он хорошо иллюстрирует особенности каждого формализма.

## 5. ЗАКЛЮЧЕНИЕ

В заключение укажем преимущества мультивекторного способа над кватернионным.

Во введении мы уже упомянули, что описание вращений с помощью матриц имеет ряд недостатков, которые как раз и устраняют кватернионы, а именно:

- нет явления складывания рамок и алгоритм численно стабилен;
- исходными данными для осуществления поворота являются ось вращения и угол вращения вокруг этой оси, что повышает наглядность;
- требуется меньше вычислений по сравнению с матричным подходом.

Недостатки кватернионного подхода касаются геометрической интерпретации и невозможности применять его к пространствам, отличным от трехмерного:

- кватернион, являясь четырехмерным объектом, используется для описания трехмерного вектора (с помощью чистого кватерниона);
- кватернион выступает в двух ипостасях — как вращаемый вектор и как “вращатель”;
- есть сложности в интерпретации угла вращения  $\vartheta$ , так как происходит его удвоение — в формуле задается угол  $\vartheta$ , а вращение происходит на угол  $2\vartheta$ .

При **мультивекторном подходе** сохраняются преимущества кватернионного подхода с одновременным устранением его недостатков:

- наглядная геометрическая интерпретация всех алгебраических формул:

- векторы **a** и **b** задают бивектор  $\mathbf{a} \wedge \mathbf{b}$ , который интерпретируется как параллелограмм (плоскость);
- вращение представляется как композиция двух отражений: сначала относительно прямой с направляющим вектором **a**, а затем относительно прямой с направляющим вектором **b**;
- эта интерпретация позволяет легко объяснить удвоение угла поворота  $\vartheta$ , так как это угол между векторами **a** и **b**;
- исходно задается плоскость вращения, но не составляет труда найти ось вращения, применив операцию правого дополнения; можно также задать ось вращения и найти плоскость, также применив операцию дополнения; при использовании кватернионов нет способа моделировать плоскость вращения;
- нет необходимости представлять вектор объектами другого пространства; вращается вектор (элемент линейного пространства), а вращателем выступает мультивектор специального вида (изоморфный кватернионам).

Мы провели вычисления с помощью кватернионов и мультивекторов. Оба формализма являются реализациями одной и той же алгебры Клиффорда [31]. При этом мы не рассматривали еще одну популярную реализацию алгебры Клиффорда – спиноры [32]. Каждая из реализаций имеет свою специфику, в том числе и уровень сложности формализма для исследователя, где спинорный формализм является самым сложным, а формализм геометрической алгебры – самым простым.

Применение различных представлений алгебры Клиффорда к научно-техническим расчетам имеет давнюю историю, однако для стороннего наблюдателя эти манипуляции выглядят несколько экзотично. Мы надеемся, что наша работа послужит делу популяризации такого рода формализмов, особенно формализма геометрической алгебры.

## СПИСОК ЛИТЕРАТУРЫ

1. *Kanatani K.* 3D Rotations: Parameter Computation and Lie Algebra based Optimization. Chapman and Hall/CRC, 2020.
2. *Vince J.* Rotation transforms for computer graphics. London : Springer-Verlag, 2011.
3. *Kuipers J.B.* Quaternions and rotation sequences. Princeton University Press, 1999.
4. *Grassmann H.G.* Die Mechanik nach den Principien der Ausdehnungslehre // Mathematische Annalen. 1877. Bd. 12. S. 222–240.
5. *Clifford W.K.* Applications of Grassmann's Extensive Algebra // American Journal of Mathematics. 1878. V. 1. № 4. P. 350–358.
6. *Казанова Г.* Векторная алгебра / Под ред. М.К. Поливанов. Современная математика. М.: Мир, 1979. 120 с.
7. *Hestenes D., Sobczyk G.* Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics. Fundamental Theories of Physics. Netherlands Springer, 1987.
8. *Delanghe R., Sommen F., Soucek V.* Clifford algebra and spinor-valued functions. Mathematics and Its Applications. Netherlands: Kluwer Academic Publishers, 1992.
9. *Doran C., Lasenby A.* Geometric Algebra for Physicists. Morgan Kaufmann Publishers, 2003.
10. *Dorst L., Fontijne D., Mann S.* Geometric algebra for computer science. The Morgan Kaufmann Series in Computer Graphics. Morgan Kaufmann, 2007.
11. *Vince J.* Geometric algebra for computer graphics. Springer-Verlag, 2008.
12. *Lengyel E.* Mathematics. Lincoln, California : Terathon Software LLC, 2016. Vol. 1. 195 p.
13. *Kanatani K.* Understanding Geometric Algebra. Taylor and Francis Group/CRC, 2015.
14. *ten Bosch M.* Let's remove Quaternions from every 3D Engine. Access mode: <https://marctenbosch.com/quaternions/>.
15. *SymPy.* 2022. Access mode: <http://www.sympy.org/ru/index.html>.
16. *Van Rossum G., Drake F.L.* Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009.
17. *GAAlgebra* – Symbolic Geometric Algebra/Calculus package for SymPy. 2022. Access mode: <https://galgebra.readthedocs.io/en/latest/index.html>.
18. *Dickson L.E.* On Quaternions and Their Generalization and the History of the Eight Square Theorem // The Annals of Mathematics. 1919. 3. V. 20. № 3. P. 155–171.
19. *Cayley A.* IV. A sixth memoir upon quantic // Philosophical Transactions of the Royal Society of London. 1859. 1. V. 149. P. 61–90.

20. *Klein F.C.* Ueber die sogenannte Nicht-Euklidische Geometrie // *Gauß und die Anfänge der nicht-euklidischen Geometrie*. Wien: Springer-Verlag Wien, 1985. Bd. 4 von Teubner-Archiv zur Mathematik. S. 224–238.
21. *Klein F.C.* A comparative review of recent researches in geometry // *Bulletin of the American Mathematical Society*. 1893. V. 2. № 10. P. 215–249. 0807.3161.
22. *Яглом И.М.* Комплексные числа и их применение в геометрии // *Математика, ее преподавание, приложения и история*. 1961. Т. 6 из Математическое просвещение, сер. 2. С. 61–106.
23. *Gevorkyan M.N., Korolkova A.V., Kulyabov D.S.* Approaches to the implementation of generalized complex numbers in the Julia language // *Workshop on information technology and scientific computing in the framework of the X International Conference Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems (ITTMM-2020)* / Ed. by D.S. Kulyabov, K.E. Samouylov, L.A. Sevastianov. Vol. 2639 of CEUR Workshop Proceedings. Aachen, 2020. 4. P. 141–157. arXiv: 2007.09737.
24. *Митюшов Е.А., Мисюра Н.Е., Берестова С.А.* К 175-летию открытия кватернионов // *Вестник Удмуртского университета. Матем. Мех. Компьют. науки*. Т. 28. С. 611–617.
25. *Hamilton W.R.* On a New Species of Imaginary Quantities, Connected with the Theory of Quaternions // *Proceedings of the Royal Irish Academy (1836-1869)*. 1840. V. 2. P. 424–434.
26. *Конвей Д.Х., Смит Д.А.* О кватернионах и октавах / Под ред. В.В. Доценко. М.: МЦНМО, 2009. 184 с.
27. *Кострикин А.И.* Линейная алгебра. М. : МЦНМО, 2009. Т. 2. 368 с.
28. Quaternions in numpy. 2022. Access mode: <https://quaternion.readthedocs.io>.
29. Pyquaternion. 2022. Access mode: <https://kieranwynn.github.io/pyquaternion/>.
30. *Геворкян М.Н., Демидова А.В., Велиева Т.Р., Королькова А.В., Кулябов Д.С.* Аналитико-численная реализация алгебры поливекторов на языке Julia // *Программирование*. 2022. № 1. С. 54–64.
31. *Пенроуз Р., Риндлер В.* Спиноры и пространство-время. М. : Мир, 1988. Т. 2. 573 с.
32. *Пенроуз Р., Риндлер В.* Спиноры и пространство-время. М. : Мир, 1987. Т. 1. 527 с.

## ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.161

# О СТРУКТУРЕ РЕШЕНИЙ КЛЮЧЕВОГО УРАВНЕНИЯ ГОСПЕРА В ЗАДАЧАХ СИМВОЛЬНОГО СУММИРОВАНИЯ

© 2023 г. Е. В. Зима<sup>1,\*</sup><sup>1</sup>Университет Вилфрида Лоури, Ватерлоо, Канада

\*e-mail: ezima@wlu.ca

Поступила в редакцию 10.05.2022 г.  
Переработанный вариант 01.06.2022 г.  
Принята к публикации 10.09.2022 г.

Анализируется структура полиномиальных решений ключевого уравнения Госпера, возникающего в задачах символьного суммирования. Дан метод быстрого нахождения входящих в решение множителей высокой степени. Показано, что в случаях, когда уравнению соответствует суммируемый нерациональный гипергеометрический терм, алгоритм Госпера можно ускорить, убрав несущественную зависимость времени его работы от величины дисперсии рационального сертификата. Библ. 10.

**Ключевые слова:** неопределенное гипергеометрическое суммирование, ускоренный алгоритм Госпера, факториальные полиномы, полиномиальная нормальная форма.

DOI: 10.31857/S0044466923010155, EDN: LEDEOY

## 1. ВВЕДЕНИЕ

Несмотря на то что алгоритмический подход к символьному суммированию имеет длинную историю [1], [2], стандартные алгоритмы суммирования и их реализации в системах компьютерной алгебры обладают тем недостатком, что они могут без необходимости требовать времени выполнения, которое экспоненциально зависит от размера входа. Это происходит из-за хорошо известного эффекта “разбухания промежуточных выражений”.

Одним из важных промежуточных представлений, возникающих в алгоритмах символьного суммирования, является т.н. форма Госпера–Петковшека рациональной функции [2], [3]. Современный термин для этого представления — *полиномиальная нормальная форма* (далее ПНФ [3]). Одна из проблем с ПНФ заключается в том, что степень одного из образующих эту форму полиномов может экспоненциально зависеть от размера числителя и знаменателя данной рациональной функции. Поскольку степень ПНФ определяет временную сложность выполнения алгоритмов суммирования, это напрямую влияет на эффективность их реализации в системах компьютерной алгебры.

Далее будет показано, как, используя компактное представление ПНФ, можно избежать разбухания промежуточных выражений и существенно понизить сложность алгоритма Госпера [2] для суммируемых гипергеометрических термов. В разд. 2 даются необходимые предварительные сведения и определения. В разд. 3 кратко описывается алгоритм Госпера. В разд. 4 предлагается основанная на компактном представлении модификация этого алгоритма, а в разд. 5 рассматриваются примеры, основанные на реализации предложенных алгоритмов в системе компьютерной алгебры Мэйпл [4].

## 2. ФАКТОРИАЛЬНЫЕ ПОЛИНОМЫ И ПНФ

Пусть  $\mathbb{K}$  — поле нулевой характеристики,  $x$  — независимая переменная,  $E$  — оператор сдвига относительно  $x$ , т.е.  $Ef(x) = f(x+1)$  для произвольной функции  $f(x)$ . Ненулевое выражение  $F(x)$  называется *гипергеометрическим термом* над  $\mathbb{K}$ , если существует рациональная функция  $r(x) \in \mathbb{K}(x)$  такая, что  $EF(x)/F(x) = r(x)$ . Обычно  $r(x)$  называют рациональным *сертификатом* терма  $F(x)$ .

Задача неопределенного гипергеометрического суммирования: по гипергеометрическому терму  $F(x)$  найти гипергеометрический терм  $G(x)$ , удовлетворяющий линейному разностному уравнению первого порядка

$$(E - 1)G(x) = F(x). \quad (1)$$

Если такой терм найден, результат неопределенного суммирования записывается в виде

$$\sum F(x) = G(x) + c,$$

где  $c$  — произвольная постоянная или периодическая функция  $x$  с периодом 1.

Например,  $\frac{4^x x^4}{\binom{2x}{x}}$  является гипергеометрическим термом с сертификатом  $\frac{2(x+1)^5}{x^4(2x+1)}$  и суммой

$$\sum \frac{4^x x^4}{\binom{2x}{x}} = \frac{(2x-1)(63x^4 - 140x^3 + 60x^2 + 26x - 6)4^x}{693 \binom{2x}{x}} \quad (2)$$

(произвольная постоянная обычно опускается).

Для ненулевых полиномов  $p(x)$  и  $q(x)$  *дисперсионным множеством* называется множество натуральных чисел  $h$  таких, что  $p(x)$  и  $q(x+h)$  имеют нетривиальный общий делитель ( $\deg(\text{НОД}(p(x), q(x+h))) > 0$ ). Если это множество непусто, то *дисперсией*  $p(x)$  и  $q(x)$  является его наибольший элемент, в противном случае дисперсия  $p(x)$  и  $q(x)$  полагается равной  $-\infty$  [1].

Два полинома  $u(x), v(x) \in \mathbb{K}[x]$  называются *эквивалентными по сдвигу*, если существует  $h \in \mathbb{Z}$  такое, что  $u(x+h) = v(x)$ . Установление эквивалентности по сдвигу двух полиномов не составляет труда [5]. Имея несколько различных эквивалентных по сдвигу полиномов  $u_1(x), \dots, u_t(x)$ , легко перечислить их в порядке от “крайнего левого” до “крайнего правого”, т.е. переупорядочить их как  $v_1(x), \dots, v_t(x)$  таким образом, что для любых индексов  $i, j \in \{1, \dots, t\}$ ,  $i < j$ ,  $v_j(x+h_{ij}) = v_i(x)$  и  $h_{ij} > 0$ . Мы ссылаемся на  $v_1(x)$  как на наименьший и на  $v_t(x)$  как на наибольший элемент класса сдвиговой эквивалентности, образованного полиномами  $u_1(x), \dots, u_t(x)$ .

## 2.1. Факториальные полиномы и их простейшие свойства

Следуя [6], определим *факториальный полином* (обобщение падающего факториала) для  $p(x) \in \mathbb{K}[x]$  как

$$[p(x)]_k = p(x)p(x-1)\dots p(x-k+1) \quad (3)$$

для  $k > 0$  и  $[p(x)]_0 = 1$ . Величину  $k$  будем называть *факториальной степенью*, а величину  $k \deg p$  — *полной степенью* факториального полинома  $[p(x)]_k$ .

Отметим, что левая часть (3) дает наиболее компактное представление произведения, стоящего в правой части, для больших значений  $k$ , поскольку требует  $\Theta(\log k)$  бит для представления полинома  $p(x)p(x-1)\dots p(x-k+1)$  при фиксированном  $p(x)$ . Факториальные полиномы удовлетворяют многим очевидным равенствам, отражающим их мультипликативный характер и позволяющим манипулировать с ними без полного раскрытия скобок. Перечислим некоторые из полезных равенств для иллюстрации:

$$[p(x)]_k = [p(x)]_{k-1} p(x-k+1) \quad \text{для } k > 0,$$

$$[p(x+1)]_k = p(x+1)[p(x)]_{k-1} \quad \text{для } k > 0,$$

$$[p_1(x)p_2(x)]_k = [p_1(x)]_k [p_2(x)]_k \quad \text{и т.д.}$$

На основе таких равенств легко реализуются правила “ленивых” вычислений, таких как, например,

$$A[p(x)]_k \pm B[p(x+1)]_k = \{Ap(x-k+1) \pm Bp(x+1)\}[p(x)]_{k-1}, \quad (4)$$

выполняющегося для произвольных выражений  $A$  и  $B$ . Другим примером элементарных правил для упрощения подобных выражений служат вычисления НОД и сокращения дробей. Например, для произвольных неотрицательных целых  $h, k, \ell$  и неприводимого  $p(x)$ :

$$\text{НОД}([p(x)]_k, [p(x+h)]_\ell) = \begin{cases} 1 & \text{для } \ell - h \leq 0, \\ [p(x)]_{\min(k, \ell-h)} & \text{в противном случае.} \end{cases}$$

## 2.2. Полиномиальные нормальные формы

Рассмотрим  $r(x) \in \mathbb{K}(x)$ . Если  $z \in \mathbb{K}$  и унитарные полиномы  $A(x), B(x), C(x) \in \mathbb{K}[x]$  удовлетворяют

$$(i) \quad r(x) = z \frac{A(x) EC(x)}{B(x) C(x)},$$

$$(ii) \quad A(x) \text{ и } E^k B(x) \text{ взаимно просты для всех } k \in \mathbb{N},$$

тогда  $(z, A(x), B(x), C(x))$  является *полиномиальной нормальной формой* (ПНФ) рациональной функции  $r(x)$ . Если дополнительно

(iii)  $A(x)$  взаимно прост с  $C(x)$ , а  $B(x)$  взаимно прост с  $EC(x)$ , тогда  $(z, A(x), B(x), C(x))$  является *строгой полиномиальной нормальной формой*  $r(x)$  (см. [3], [7]). (В дальнейшем под ПНФ будет всегда пониматься строгая ПНФ.)

Построение ПНФ основано на вычислении дисперсионного множества числителя  $a(x)$  и знаменателя  $b(x)$  данной рациональной функции  $r(x)$  и применении алгоритма Евклида для нахождения НОД полиномов в  $\mathbb{K}[x]$  (подробности см. в [3], [7]). Используется то, что если  $p(x) = \gcd(a(x), b(x+h))$  и  $a(x) = p(x)\tilde{a}(x)$ ,  $b(x) = p(x-h)\tilde{b}(x)$ , то  $\frac{a(x)}{b(x)} = \frac{\tilde{a}(x)}{\tilde{b}(x)} \frac{E[p(x-1)]_h}{[p(x-1)]_h}$ .

Заметим, что степень полинома  $C(x)$  в ПНФ может экспоненциально зависеть от битового размера  $r(x)$ . Например, для рациональной функции  $\frac{(x+100)(2x^2+1)}{x}$  ПНФ будет  $2, x^2 + \frac{1}{2}, 1, (x+99)(x+98)(x+97)\dots(x+1)x$  с полиномом  $C(x) = [x+99]_{100}$  полной степени 100. Привлечение компактного представления факториальных полиномов позволяет сократить как время построения ПНФ, так и ее размер.

## 3. АЛГОРИТМ ГОСПЕРА

В [2] Госпер описал процедуру суммирования гипергеометрического терма, которая широко используется в системах компьютерной алгебры. Алгоритм основан на простом наблюдении: если заданный гипергеометрический терм  $F(x)$  имеет гипергеометрическую сумму  $G(x)$ , то термы  $G(x)$  и  $F(x)$  *подобны*: существует функция  $Y(x) \in \mathbb{K}(x)$  такая, что  $G(x) = Y(x)F(x)$ . Это сводит исходную задачу суммирования к задаче нахождения рационального решения  $Y(x)$  уравнения

$$Y(x+1)r(x) - Y(x) = 1, \quad (5)$$

где  $r(x)$  — рациональный сертификат терма  $F(x)$ .

Для решения (5) рациональный сертификат преобразуется к ПНФ  $(z, A(x), B(x), C(x))$  и поиск суммы сводится к отысканию полиномиального решения  $y(x)$  *ключевого уравнения*:

$$zA(x)y(x+1) - B(x-1)y(x) = C(x). \quad (6)$$

Если  $y(x)$  найдено, то

$$G(x) = F(x) \frac{B(x-1)y(x)}{C(x)}. \quad (7)$$

Чтобы решить (6), можно найти верхнюю оценку степени решения и использовать (например) метод неопределенных коэффициентов, сведя задачу поиска к стандартной процедуре линейной алгебры. Одним из факторов, влияющих на эту оценку (и, как следствие, размер решаемой линейной системы) является степень полинома  $C(x)$  в (6).

Иногда полином  $C(x)$  из (6) называют *универсальным знаменателем*. Известная проблема с алгоритмом Госпера состоит в том, что универсальный знаменатель может иметь пессимистически завышенную степень (т.е.  $C(x)$  и  $y(x)$  могут иметь общий множитель высокой степени, который сократится после подстановки решения  $y(x)$  уравнения (6) в (7)). Это, в свою очередь, может привести к экспоненциальной зависимости времени выполнения алгоритма Госпера от размера входных данных, даже когда вход и выход малы.

Рассмотрим применение адгоритма Госпера к гипергеометрическому терму

$$F(x) = \frac{(27x^3 + 819x^2 + 246x - 194)(2x)!}{(3x + 91)(3x + 1)(x + 1)(3x + 94)(3x + 4)(x!)^2} \quad (8)$$

с рациональным сертификатом

$$r(x) = 2 \frac{(3x + 1)(3x + 91)(2x + 1)(27x^3 + 900x^2 + 1965x + 898)}{(27x^3 + 819x^2 + 246x - 194)(3x + 7)(3x + 97)(x + 2)},$$

дисперсия числителя и знаменателя которого равна 32. После построения ПНФ получаем ключевое уравнение

$$4\left(x + \frac{1}{2}\right)\left(x + \frac{1}{3}\right)y(x + 1) - \left(x + \frac{94}{3}\right)(x + 1)y(x) = \left(x^3 + \frac{91}{3}x^2 + \frac{82}{9}x - \frac{194}{27}\right)\left[x + \frac{88}{3}\right]_{28}$$

с правой частью степени 31. Это уравнение имеет полиномиальное решение степени 29

$$y(x) = \frac{1}{3}\left[x + \frac{88}{3}\right]_{29}.$$

После подстановки  $y(x)$  в (7) факториальный полином в  $C(x)$  сокращается и окончательный результат суммирования имеет вид

$$\sum \frac{(27x^3 + 819x^2 + 246x - 194)(2x)!}{(3x + 91)(3x + 1)(x + 1)(3x + 94)(3x + 4)(x!)^2} = \frac{(2x)!}{(3x + 91)(3x + 1)(x!)^2}. \quad (9)$$

Покажем, как с помощью небольших дополнительных усилий можно (после получения оценки степени решения (6)) удалить “лишние” множители в универсальном знаменателе  $C(x)$  перед решением ключевого уравнения.

#### 4. ПОНИЖЕНИЕ СТЕПЕНИ РЕШЕНИЯ КЛЮЧЕВОГО УРАВНЕНИЯ

Предположим, что заданный гипергеометрический терм  $F(x)$  не является рациональной функцией и суммируем, т.е. уравнение (6) имеет полиномиальное решение  $y(x)$ . Сокращение в (7) происходит, когда решение  $y(x)$  и правая часть  $C(x)$  уравнения (6) имеют общий множитель. Предположим, что  $C(x) = [p(x)]_k \tilde{C}(x)$  и решение  $y(x)$  также имеет множитель  $[p(x)]_k$ . Тогда подстановка

$$y(x) = \tilde{y}(x)[p(x)]_{k+1} \quad (10)$$

в (6) дает новое уравнение

$$zA(x)p(x + 1)\tilde{y}(x + 1) - B(x - 1)p(x - k)\tilde{y}(x) = \tilde{C}(x), \quad (11)$$

имеющее полиномиальное решение  $\tilde{y}(x)$ , и можно  $y(x)$  в (7) заменить на  $\tilde{y}(x)$ ,  $C(x)$  в (7) заменить на  $\tilde{C}(x)$  (по сути неявно реализуя сокращение ненужного общего множителя в числителе и знаменателе (7)). Если верхняя оценка степени решения  $y(x)$  в (6) равна  $N$ , то оценка степени решения  $\tilde{y}(x)$  в (11) уменьшится до  $N - k \deg p(x)$ .

Возвращаясь к примеру (8), если установлено, что в решение ключевого уравнения входит множитель  $\left[x + \frac{88}{3}\right]_{28}$ , то подстановка в это уравнение  $y(x) = \left[x + \frac{88}{3}\right]_{29} \tilde{y}(x)$  дает модифицированное ключевое уравнение:

$$4\left(x + \frac{1}{2}\right)\left(x + \frac{1}{3}\right)\left(x + \frac{91}{3}\right)\tilde{y}(x+1) - \left(x + \frac{94}{3}\right)\left(x + \frac{4}{3}\right)(x+1)\tilde{y}(x) = \left(x^3 + \frac{91}{3}x^2 + \frac{82}{9}x - \frac{194}{27}\right)$$

с верхней границей степени решения  $< 3$ . Решение  $\tilde{y}(x) = 1/3$  последнего уравнения дает тот же результат, что и в (9), без дополнительных затрат на сокращение числителя и знаменателя (7) на их наибольший общий делитель.

Заметим, что из-за природы ПНФ факториальные полиномы появляются только в правой части ключевого уравнения (6). Более того, число этих факториальных полиномов ограничено степенями числителя и знаменателя рационального сертификата  $r(x)$  и не зависит от величины дисперсии. С другой стороны, каждый входящий в  $C(x)$  множитель вида  $[p(x)]_k$  добавляет значение  $k \deg p(x)$  к оценке верхней границы степени решения  $y(x)$ , которая может быть такой большой, как значение дисперсии числителя и знаменателя сертификата  $r(x)$ , домноженное на  $\deg p(x)$ . Поэтому такие множители являются подходящими кандидатами на проверку их сократимости в (7), приводящей к избавлению от несущественной зависимости времени выполнения алгоритма от значения дисперсии. Рассмотрение в качестве кандидатов на сокращение других множителей не приведет к изменению асимптотической сложности алгоритма.

Покажем, как можно эффективно отыскивать те из множителей вида  $[p(x)]_k$  в правой части (6), которые также являются множителями решения  $y(x)$ . Для упрощения изложения предположим, что все полиномы, входящие в (6), разложены на неприводимые множители (позже будет показано, что это предположение не является обязательным). В дальнейшем пусть  $[p(x)]_k$  — один из множителей  $C(x)$  в (6), т.е.  $C(x) = [p(x)]_k \tilde{C}(x)$ . Наш подход основан на компактном представлении факториальных полиномов, “ленивых” вычислениях последовательных значений  $y(x)$  в (6) и простых структурных свойствах полиномиальных решений уравнения (6):

1. Выражение  $[p(x)]_k$  обращается в ноль на любом корне  $\alpha$  полинома  $p(x)$ , а также при  $x = \alpha + 1, \dots, \alpha + k - 1$ .

2. Выражения  $A(x)$  и  $B(x-1)$  не могут обращаться в ноль при  $x = \alpha, \alpha + 1, \dots, \alpha + k - 1$ . Это следует из требования (iii) в определении ПНФ.

3. Если решение  $y(x)$  уравнения (6) обращается в ноль в любой из точек  $\alpha, \alpha + 1, \dots, \alpha + k$  (где  $\alpha$  — корень  $p(x)$ ), то  $y(x)$  обращается в ноль во всех этих точках. Это значит, что полином  $[p(x)]_{k+1}$  входит как множитель в  $y(x)$  и факториальный полином  $[p(x)]_k$  из  $C(x)$  сокращается при подстановке  $y(x)$  в (7).

4. Если полиномы  $A(x)$  и  $B(x-1)$  не имеют множителей, эквивалентных по сдвигу полиному  $p(x)$ , то выражение  $[p(x)]_k$  гарантированно является множителем решения  $y(x)$  уравнения (6). Это свойство выполняется только для нерациональных гипергеометрических термов и следует из доказанного в [8] факта, что однородное уравнение

$$zA(x)y(x+1) - B(x-1)y(x) = 0,$$

соответствующее уравнению (6), имеет только тривиальное решение, если  $z, A(x)$  и  $B(x)$  получены при построении ПНФ для гипергеометрического терма, не являющегося рациональной функцией.

5. Любой множитель  $A(x)$  (соответственно  $B(x-1)$ ) из (6), эквивалентный по сдвигу полиному  $p(x)$ , дает начальное значение решения  $y(x)$  при  $x = \beta$  (соответственно, при  $x = \beta + 1$ ), где  $\beta$  является корнем этого множителя. Вычисления, необходимые для обнаружения равенства или неравенства  $y(x)$  нулю в последовательных точках, начинающиеся с  $x = \beta$  (соответственно, с  $x = \beta + 1$ ), можно выполнить “лениво”. Раскрытие скобок в  $[p(x)]_k$  или знание самого корня  $\beta$  для этого не требуется.

Поясним последнее утверждение. Пусть  $C(x) = [p(x)]_k \tilde{C}(x)$ , где  $[p(x)]_k$  — единственный факториальный полином в правой части (6), а  $b(x)$  — множитель  $B(x-1)$ , эквивалентный по сдвигу  $p(x)$  (т.е.  $p(x+h) = b(x), h > 0$ ). Принимая  $b(x) = 0$ , получаем

$$y(x+1) = z^{-1} A(x)^{-1} \tilde{C}(x) [p(x)]_k,$$

что дает значение решения  $y(x)$  при  $x = \beta + 1$ , где  $\beta$  — любой корень  $b(x)$ . Теперь из (6) можно выразить

$$y(x+2) = \{z^{-1} A(x+1)^{-1} [B(x)z^{-1} A(x)^{-1} \tilde{C}(x)p(x-k+1) + \tilde{C}(x+1)p(x+1)]\} [p(x)]_{k-1}. \quad (12)$$

Значение  $y(\beta+2)$  будет равно нулю если и только если выражение в квадратных скобках равно нулю. Для проверки этого равенства не требуется полное вычисление значения правой части. Достаточно разделить полином в квадратных скобках в (12) на полином  $b(x)$  и проверить равенство нулю полученного остатка. Продолжая вычисления по модулю полинома  $b(x)$  подобным образом, дойдем до значения  $y(x+h)$  и если оно равно нулю, то множитель  $[p(x)]_k$  входит в решение  $y(x)$  и может быть сокращен, как показано выше. Заметим, что вычисление  $A(x)^{-1}$  по модулю  $b(x)$  в квадратных скобках в (12) не представляет проблем, так как по условию (ii) из определения ПНФ полиномы  $A(x), A(x+1), \dots, A(x+h)$  обратимы по модулю полинома  $b(x)$ .

Возвращаясь к примеру (8), для установления того, что факториальный полином  $\left[x + \frac{88}{3}\right]_{28}$  является множителем решения  $y(x)$  уравнения (6), воспользуемся множителем  $(x + 94/3)$  полинома  $B(x-1)$ , эквивалентным по сдвигу полиному  $x + \frac{88}{3}$ , что даст значение  $y(x)$  для  $x = -91/3$ . Одного шага вычислений по рекуррентности (12) оказывается достаточным, чтобы определить равенство  $y(x) = 0$  при  $x = -88/3$ .

Если наряду с  $[p(x)]_k$  правая часть (6) содержит и другие факториальные полиномы  $[u(x)]_\ell, [v(x)]_m, \dots$ , то проверка сократимости  $[p(x)]_k$  будет такой же, как показано выше: просто в (12) нужно заменить  $p(x-k+1)$  на  $p(x-k+1)u(x-\ell+1)v(x-m+1)\dots$ , а  $p(x+1)$  на  $p(x+1)u(x+1)v(x+1)\dots$ , домножить правую часть (12) на  $[u(x)]_{\ell-1}[v(x)]_{m-1}\dots$ , и учитывать, что любой факториальный полином  $[q(x)]_n$  обращается в 1 при  $n \leq 0$ .

Аналогичной (12) рекуррентцией можно воспользоваться и в случае, когда  $A(x)$  в (6) содержит множитель, эквивалентный по сдвигу полиному  $p(x)$ . Если же оба полинома  $A(x)$  и  $B(x-1)$  в (6) содержат множители, эквивалентные по сдвигу полиному  $p(x)$ , можно минимизировать количество промежуточных ленивых шагов при тестировании вхождения  $[p(x)]_k$  в решение  $y(x)$ . Пусть, например,  $C(x) = [p(x)]_k \tilde{C}(x)$ ,  $b(x)$  — множитель  $B(x-1)$  эквивалентный по сдвигу  $p(x)$ , а  $a(x)$  — множитель  $A(x)$  эквивалентный по сдвигу  $p(x)$ . Отметим, что  $a(x)$  является наименьшим, а  $b(x)$  — наибольшим элементом класса сдвиговой эквивалентности, образованного полиномами  $a(x), p(x-k+1), p(x-k+2), \dots, p(x), b(x)$ . Если вычисление значений  $y(x)$  начинается в нуле полинома  $b(x)$  (соответственно в нуле полинома  $a(x)$ ), то количество промежуточных шагов для установления равенства нулю решения  $y(x)$  на нулях полинома  $[p(x)]_k$  равно дисперсии полиномов  $b(x)$  и  $p(x)$  (соответственно дисперсии полиномов  $p(x-k+1)$  и  $a(x)$ ). Меньшее из этих значений дает минимальное число промежуточных шагов.

Если  $C(x)$  в (6) содержит несколько множителей вида  $[p(x)]_k$ , то проверку вхождения этих множителей в решение  $y(x)$  можно проводить в произвольном порядке (это следует из локальных свойств ПНФ, установленных в [9]). Отметим, что если дисперсия равна  $d$ , то число промежуточных шагов при проверке не может превышать  $d-k$ , т.е., чем выше факториальная степень проверяемого множителя, тем меньшее число шагов потребуется в худшем случае. Это наблюдение позволяет избегать проверки множителей, избавление от которых не уменьшает асимптотическую сложность алгоритма (например, не проверять множители с факториальной степенью того же порядка, что и максимальная из степеней числителя и знаменателя сертификата).

## 5. РЕАЛИЗАЦИЯ И ЗАКЛЮЧЕНИЕ

Описанные выше свойства решений ключевого уравнения (6) позволяют внести простые изменения в алгоритм Госпера. Первое изменение не связано напрямую с самим алгоритмом — оно вносится в процедуру построения ПНФ и обеспечивает получение результата в компактном виде.

Второе изменение вносится в алгоритм сразу после построения ключевого уравнения (6) и действует только в том случае, когда правая часть (6) содержит факториальные полиномы. Для каждого множителя вида  $[p(x)]_k$  в правой части (6) проверяется, входит ли он в решение  $y(x)$  (как показано выше) и отбираются те факториальные полиномы, для которых получен положительный ответ. После этого для отобранных факториальных полиномов проводятся подстановки вида (10), дающие новое ключевое уравнение, которое передается на последний этап стандартной процедуры Госпера. Задействованные при этом вычисления используют те же средства, что и построение ПНФ (вычисление дисперсионного множества полиномов, НОД полиномов и т.д.). При этом часто можно воспользоваться результатами, полученными при построении ПНФ, например, для поиска множителей  $A(x)$  или  $B(x-1)$ , эквивалентных по сдвигу  $p(x)$ .

Внесенные изменения не ухудшают асимптотическую сложность процедуры суммирования, но могут привести к весьма заметному сокращению времени ее выполнения для суммируемых термов с большим значением дисперсии рационального сертификата (фактически, в этом случае модифицированный алгоритм Госпера имеет время выполнения, полиномиальное по размеру входа [10], тогда как оригинальный алгоритм имеет экспоненциальное по размеру входа время выполнения).

Реализация описанной модификации в системе Мэйпл подтверждает улучшения временных характеристик работы алгоритма. Так, например, у суммируемого гипергеометрического термина

$$\frac{(900x^5 + (1800p - 1155)x^4 + (900p^2 - 270p - 3401)x^3 + (885p^2 - 2786p - 1698)x^2 + (-165p^2 - 1576p + 898)x - 360p^2 - 24p + 844)(2x)!}{(5x+2)(5x+5p-13)(3x+3p+1)(3x+1)(3x+4)(3x+4+3p)(5x-8+5p)(5x+7)x!}$$

величина дисперсии сертификата равна  $p+2$ , а  $C(x)$  в (6) включает множители  $[x+p-2/3]_{p-2}$  и  $[x+p-18/5]_{p-3}$ . Время выполнения стандартного алгоритма Госпера в системе Мэйпл для  $p = 100, 200, 400$  и  $800$  равно, соответственно, 1.163, 7.503, 58.626 и 510.230 с, тогда как время выполнения модифицированного алгоритма для тех же значений  $p$  остается неизменным  $\approx 0.02$  с, подтверждая, что внесенная модификация убирает несущественную зависимость времени выполнения алгоритма от величины дисперсии сертификата.

В тех случаях, когда понизить оценку степени решения (6) не удастся, время выполнения модифицированного алгоритма практически не отличается от времени выполнения стандартного алгоритма Госпера. Это очевидно для случаев, когда ключевое уравнение (6) не содержит факториальных полиномов в правой части, как в примере (2): в этом случае оба алгоритма возвращают результат, затратив около 10 мс. Если же (6) содержит факториальный полином высокой степени, который не сокращается в (7), то затраты на проверку сократимости (число шагов) не превышают значения дисперсии уменьшенного на факториальную степень, решаемая система линейных уравнений имеет размер, пропорциональный значению дисперсии, и время решения системы является доминирующим. Например, для суммируемого гипергеометрического термина  $F(x+p) - F(x)$ , где  $F(x) = x!/(3x+1)$  — величина дисперсии сертификата равна  $p+1$ , а  $C(x)$  в (6) включает множитель  $[x+p-2/3]_{p-1}$ , который не входит в решение (6). Время выполнения стандартного алгоритма Госпера в системе Мэйпл для  $p = 100, 200$  и  $400$  равно, соответственно, 0.843, 5.699 и 62.856 с, тогда как время выполнения модифицированного алгоритма для тех же значений  $p$  равно, соответственно, 0.939, 5.817 и 63.001 с, подтверждая, что дополнительные затраты незначительны и проверка сократимости в (7) несущественно ухудшает общее время работы алгоритма.

Сделанное ранее предположение, что все полиномы, входящие в (6), разложены на неприводимые множители можно ослабить. Пусть полином  $p(x)$  в  $[p(x)]_k$  не является неприводимым, а полиномы  $A(x)$  и  $B(x-1)$  не разложены на множители. Построим дисперсионные множества  $S_A$  и  $S_B$  для пар полиномов  $p(x), A(x)$  и  $B(x-1), p(x)$ . Если оба эти множества пусты, то  $[p(x)]_k$  входит в решение уравнения (6) и разложения  $p(x)$  на неприводимые множители не требуется. В противном случае максимальный или минимальный элемент  $h$  множества  $S_A$  (соответственно

множества  $S_B$ ) можно использовать для получения нетривиального множителя  $q(x)$  полинома  $p(x)$  и эквивалентного ему по сдвигу множителя  $a(x)$  полинома  $A(x)$  (соответственно  $b(x)$  полинома  $B(x-1)$ ), для чего достаточно вычислить  $\text{НОД}(p(x), A(x+h))$  (соответственно  $\text{НОД}(p(x), B(x-h-1))$ ). Это выделит из  $[p(x)]_k$  множитель  $[q(x)]_k$ , для которого и проводится проверка вхождения в решение уравнения (6). Полученный полином  $q(x)$  может не быть неприводимым, но дальнейшее его разложение не требуется. Таким образом, для реализации предложенных модификаций факторизация полиномов, вообще говоря, не требуется. Заметим, однако, что используемый в системе Мэйпл алгоритм построения дисперсионного множества полиномов основан на факторизации. Поэтому попытка избежать факторизации при построении ключевого уравнения и понижении степени его решения не приводит к заметному уменьшению времени выполнения реализованных программ. Гораздо более практичным оказывается подход, при котором результаты факторизации числителя и знаменателя сертификата сохраняются и используются при построении ПНФ и ключевого уравнения. Описываемая здесь реализация включает обе возможности и позволяет модифицировать ключевое уравнение без привлечения факторизации (это сделано на случай внесения изменений в алгоритм вычисления дисперсии в системе Мэйпл).

Отметим, что предложенные здесь модификации алгоритма Госпера сохраняют свою работоспособность и для суммируемых гипергеометрических термов, зависящих от многих переменных, когда суммирование проводится по одной из них. Это показывает практичность предложенных модификаций, поскольку при использовании алгоритма Госпера в более сложных задачах, включающих несколько независимых переменных, нередко является ситуация, когда получаемый на промежуточном этапе гипергеометрический терм гарантированно является суммируемым по выделенной переменной и требуется найти его сумму.

Для установления несуммируемости гипергеометрического термина с помощью алгоритма Госпера можно использовать описанное выше понижение степени универсального знаменателя  $C(x)$  для таких множителей вида  $[p(x)]_k$ , входящих в  $C(x)$ , для которых ни  $A(x)$ , ни  $B(x-1)$  не содержит множителей, эквивалентных  $p(x)$  по сдвигу. Если (6) не имеет полиномиального решения, то и (11) не может иметь полиномиального решения, но уравнение (11) имеет меньшую степень правой части. Это означает, что все такие множители в правой части (6) избыточны и их удаление с помощью подстановки (10) не помешает установить факт отсутствия решения у ключевого уравнения.

Вопрос о том, можно ли использовать сходные методы для полного устранения несущественной зависимости времени выполнения алгоритма Госпера от значения дисперсии для несуммируемых гипергеометрических термов, требует дальнейшего изучения.

Автор выражает свою признательность П.У. Макаули (Университет Ватерлоо, Канада) за его вклад в реализацию предложенных методов, а также Х. Хуанг (Даляньский Технологический Университет, Китай) и рецензентам за полезные и конструктивные замечания по первоначальной версии настоящей работы.

## СПИСОК ЛИТЕРАТУРЫ

1. *Абрамов С.А.* О суммировании рациональных функций // Ж. вычисл. матем. и матем. физ. 1971. Т. 11. № 4. С. 1071–1075.
2. *Gosper R.W., Jr.* Decision procedure for indefinite hypergeometric summation // Proc. Nat. Acad. Sci. U.S.A., 75(1):40–42, 1978.
3. *Abramov S.A., Petkovšek M.* Rational normal forms and minimal decompositions of hypergeometric terms // J. of Symbolic Computation, 33(5):521–543, 2002.
4. *Maple User Manual.* Maplesoft, a division of Waterloo Maple Inc., 1996–2021.
5. *Абрамов С.А.* Элементы компьютерной алгебры линейных обыкновенных дифференциальных, разностных и  $q$ -разностных операторов. М: МЦМНО, 2012.
6. *Moenck R.* On computing closed forms for summations // In *Proc. of the 1977 MACSYMA Users' Conference*, pp. 225–236, 1977.
7. *Petkovšek M.* Hypergeometric solutions of linear recurrences with polynomial coefficients // J. of Symbolic Computation, 14(2):243–264, 1992.
8. *Lisoněk P., Paule P., Strehl V.* Improvement of the degree setting in Gosper's algorithm // J. of Symbolic Computation, 16 (1993), 243–258.
9. *Pirastu R., Strehl V.* Rational summation and Gosper-Petkovšek representation // J. Symb. Comput., 20(5-6):617–635, Nov. 1995.
10. *Zima E.V.* Accelerating indefinite hypergeometric summation algorithms // ACM Commun. Comput. Algebra, 52(3):96–99, 2019.

---

**ОБЩИЕ ЧИСЛЕННЫЕ  
МЕТОДЫ**

---

УДК 519.161

**ОБОБЩЕНИЕ ЗАДАЧИ О СУММЕ ПОДМНОЖЕСТВ  
И КУБИЧЕСКИЕ ФОРМЫ**© 2023 г. А. В. Селиверстов<sup>1,\*</sup><sup>1</sup> 127051 Москва, Большой Каретный пер., 19, стр. 1, ИППИ РАН, Россия

\*e-mail: slvstv@iitp.ru

Поступила в редакцию 19.04.2022 г.  
Переработанный вариант 12.05.2022 г.  
Принята к публикации 10.09.2022 г.

Предложен новый алгоритм для распознавания существования двоичного решения у системы линейных уравнений над полем нулевой характеристики, который эффективен при выполнении некоторого ограничения на систему уравнений. Это частный случай задачи целочисленного программирования. В расширенной версии задачи о сумме подмножеств вес может быть как положительным, так и отрицательным. Рассмотренная нами задача эквивалентна задаче о существовании решения для нескольких частных случаев этой задачи одновременно. Найдены новые достаточные условия, при которых вычислительная сложность почти всех частных случаев такой задачи полиномиальная. По сути, алгоритм проверяет, существует ли кубическая гиперповерхность, проходящая через каждую вершину единичного куба, но не пересекающая заданное аффинное подпространство. Ранее уже было известно несколько эвристических алгоритмов для решения этой задачи. Однако новые методы расширяют возможности для решения тех или иных задач. Хотя подробно рассмотрена лишь задача распознавания, бинарный поиск позволяет найти решение, если это возможно. Библ. 40.

**Ключевые слова:** целочисленное программирование, система линейных уравнений, сумма подмножеств, сложность в среднем.

**DOI:** 10.31857/S0044466923010118, **EDN:** LEEUWB

**1. ВВЕДЕНИЕ**

Задача о сумме подмножеств состоит в распознавании существования  $\{0,1\}$ -решения у линейного уравнения. Она служит одним из хорошо известных примеров NP-полных задач. Задача с  $n$  переменными разрешима посредством  $O(2^{n/2})$  арифметических операций (см. [1]). Общепринятая гипотеза утверждает, что для решения этой задачи в худшем случае требуется экспоненциальное время. Эта нижняя оценка была проверена для некоторых вычислительных моделей с ограничениями, включая так называемые линейные машины (см. [2], [3]) и аддитивные машины (см. [4]), а также для методов проверки разрешимости систем алгебраических уравнений, которые основаны на теореме о многочленах над вещественно замкнутым полем (Positivstellensatz) (см. [5]) или на теореме Гильберта о нулях (Nullstellensatz) (см. [6]). С другой стороны, эта задача может быть решена за псевдополиномиальное время методом динамического программирования (см. [7], [8]). Также известны эвристические алгоритмы для случая низкой плотности (см. [9], [10]). Эта задача также допускает обобщение над кольцами вычетов (модулярный случай) (см. [11]) и над мультипликативными полугруппами матриц (см. [12]).

Мы рассмотрим задачу о существовании  $\{0,1\}$ -решения у системы линейных уравнений. В худшем случае вычислительная сложность этой задачи такая же, как и для задачи о сумме подмножеств, когда рассматривается одно уравнение. Однако быстрые эвристические алгоритмы известны для систем из достаточно большого числа уравнений с ограничением на знаки коэффициентов (см. [13], [14]), с коэффициентами любого знака (см. [15]) или при выполнении некоторых других ограничений (см. [16], [17]). Обзор некоторых алгоритмов дан в следующем разделе.

Известны эвристические методы решения или аппроксимации близких задач оптимизации, включая задачу о многомерном рюкзаке (см. [18], [19]), модулярный вариант этой задачи (см. [20]) и смешанные задачи о рюкзаке и покрытии (см. [21]). Для их решения тоже применимы об-

щие методы оптимизации (см. [22], [23]). Также известны эвристические алгоритмы, основанные на локальном поиске, для проверки выполнимости пропозициональной конъюнктивной нормальной формы (КНФ) (см. [24], [25]), в частности, 3-КНФ (см. [26]). Этот же метод позволяет решать соответствующую задачу оптимизации (см. [27]). С другой стороны, известна лишь экспоненциальная верхняя оценка  $O^*(1.2989^m)$  на сложность в худшем случае задачи о максимальном числе одновременно выполнимых элементарных дизъюнкций в составе КНФ с  $m$  элементарными дизъюнкциями (см. [28]). Задача о выполнимости КНФ эквивалентна существованию  $\{0,1\}$ -решения у системы линейных неравенств от такого же числа переменных над упорядоченным кольцом целых чисел. Поэтому оценки сложности в худшем случае для этих задач связаны друг с другом.

Говоря о вычислительной сложности в среднем, мы предполагаем, что на множестве входов произвольно фиксированной длины задана вероятностная мера. Обычно каждое такое множество конечно, следовательно, все входы в нем могут быть равновероятными. Сложностью в среднем называется математическое ожидание вычислительной сложности на входах данной длины (см. [29]).

Основной результат настоящей статьи в том, что предложен эвристический алгоритм для задачи распознавания систем линейных уравнений, не имеющих  $\{0,1\}$ -решения. Он применим при более слабых ограничениях, чем было ранее. Также он может использоваться параллельно с другими методами. Результаты справедливы над произвольным полем характеристики нуль, операции в котором эффективно вычислимы. Чтобы упростить изложение, обычно рассматриваются вычисления над полем рациональных чисел или над чисто трансцендентным расширением этого поля. Однако можно эффективно проводить вычисления с алгебраическими числами (см. [30]), используя алгоритмы для работы с многочленами (см. [31]).

## 2. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Рассмотрим систему из  $m$  линейных уравнений от  $n > m$  переменных. Исключая переменные, легко получить новое уравнение от  $n - m$  переменных. Если  $n - m = O(\log_2 n)$ , задача поиска  $\{0,1\}$ -решения для исходной системы уравнений разрешима за полиномиальное время. Она сводится к поиску каждого  $\{0,1\}$ -решения для уравнения от  $O(\log_2 n)$  переменных и проверке, продолжается ли оно до  $\{0,1\}$ -решения исходной системы. Далее рассмотрены менее тривиальные случаи, когда эта задача легко разрешима.

Обозначим через  $\mathbf{b}$  столбец и через  $A$  прямоугольную  $m \times n$  матрицу с неотрицательными элементами. Все  $\{0,1\}$ -решения системы неравенств  $A\mathbf{x} \leq \mathbf{b}$  можно найти методом динамического программирования, время работы которого мало, если общее число таких решений мало. Н.Н. Кузюрин в [13], [14] показал, что при  $m > 9 \log_2 n$  и некоторых предположениях о распределении элементов матрицы  $A$  и столбца  $\mathbf{b}$  среднее число таких решений ограничено сверху многочленом от  $n$ . Следовательно, среднее время работы алгоритма также полиномиальное. Доказательство основано на оценке хвоста биномиального распределения. Зная все решения системы неравенств, среди них легко выбрать решение для системы уравнений, если оно существует. Важным ограничением применимости этого метода служит требование неотрицательности элементов матрицы  $A$ . Замена переменных типа  $y_k = 1 - x_k$  систему уравнений с произвольными коэффициентами легко свести к системе уравнений с неотрицательными коэффициентами у линейных членов, где новая и исходная системы имеют равное число  $\{0,1\}$ -решений. Но при этом искажается исходное распределение коэффициентов.

Для системы из  $m$  линейных уравнений от  $n$  переменных с положительными целыми коэффициентами  $a_k$  определим плотность по формуле

$$\varrho = \frac{n}{m \log_2 \max_k a_k}.$$

При  $m = 1$  и при условии малой плотности  $\varrho < 0.9408$  почти все частные случаи задачи о сумме подмножеств разрешимы за полиномиальное время, используя поиск кратчайшего ненулевого вектора в некоторой решетке (см. [9], [10]). Этот метод обобщается на случай систем многих уравнений с тем же ограничением на плотность (см. [16]).

Следуя работам [15] и [32], для оценки доли входов данного размера, на которых алгоритм быстро принимает правильное решение, мы используем лемму Шварца–Зиппеля (см. [33]).

**Лемма 1.** *Дан отличный от константы многочлен  $f(x_1, \dots, x_n)$  степени  $d$  над полем  $K$ . Если случайные величины  $\xi_1, \dots, \xi_n$  независимы и равномерно распределены на конечном множестве  $S \subseteq K$  мощности  $|S|$ , то выполнено неравенство*

$$\Pr[f(\xi_1, \dots, \xi_n) = 0] \leq \frac{d}{|S|},$$

где через  $\Pr[\cdot]$  обозначена вероятность выполнения условия в квадратных скобках.

Недавно этот результат был усилен для многочленов от одной переменной над полем комплексных чисел. Существует такое множество из  $2d + 1$  чисел, что любые два многочлена  $f$  и  $g$  степени  $d$  равны друг другу, когда множество значений многочлена  $f$  вложено во множество значений многочлена  $g$  на этом множестве (см. [34]).

Ранг матрицы над полем можно вычислить, используя полиномиальное число процессоров и выполнив на каждом из них лишь  $O(\log_2^2 n)$  операций над этим полем (см. [35], [36]). С другой стороны, верхние оценки сложности вычисления ранга матрицы близки к сложности матричного умножения (см. [37], [38]). В настоящей работе примеры вычислены в системе компьютерной алгебры MathPartner (см. [39]). Определитель матрицы вычисляется командой `\det()`, а ранг — командой `\rank()`. В системе Maple определитель, ранг и перманент матрицы вычислимы, например, в пакете LinearAlgebra.

### 3. РЕЗУЛЬТАТЫ

Геометрический смысл задачи состоит в проверке, проходит ли аффинное подпространство, заданное системой линейных уравнений, через некоторую вершину единичного  $n$ -мерного куба. Каждая такая вершина соответствует  $\{0, 1\}$ -решению. Метод состоит в попытке построить алгебраическую гиперповерхность малой степени, проходящую через каждую вершину единичного куба, но не пересекающую данное аффинное подпространство. Здесь гиперповерхность может быть приводимой. Существование такой гиперповерхности влечет отсутствие  $\{0, 1\}$ -решения у исходной системы уравнений. Вычислительная сложность этого метода зависит от степени гиперповерхности. Ранее было исследовано применение квадратик для случая подпространств малой размерности (см. [15]). С другой стороны, для общей гиперплоскости объединение из  $2^n$  параллельных гиперплоскостей, содержащих все вершины единичного  $n$ -мерного куба, служит исковой гиперповерхностью степени не выше  $2^n$ . Но степень может быть ниже. В частности, если гиперплоскость задана уравнением с ограниченными сверху положительными целыми коэффициентами, которое не имеет  $\{0, 1\}$ -решений, то достаточно малого числа параллельных гиперплоскостей. На этом основан метод динамического программирования для решения задачи о сумме подмножеств. Однако мы рассматриваем произвольные гиперповерхности, которые не обязательно служат объединением гиперплоскостей.

#### 3.1. Прямые на проективной плоскости

Сначала мы рассмотрим прямую на плоскости, что соответствует поиску  $\{0, 1\}$ -решения неоднородного линейного уравнения от двух переменных  $x_1$  и  $x_2$ . Его гомогенизацией служит уравнение от трех переменных  $x_0$ ,  $x_1$  и  $x_2$ . Каждое  $\{0, 1\}$ -решение соответствует вершине единичного квадрата. На проективной плоскости эти вершины имеют однородные координаты  $(1 : 0 : 0)$ ,  $(1 : 0 : 1)$ ,  $(1 : 1 : 0)$  и  $(1 : 1 : 1)$  соответственно. В этих координатах бесконечно удаленная прямая задана уравнением  $x_0 = 0$ . Проективная кривая определяется тернарной формой (однородным многочленом), которая обращается в нуль в точках кривой.

**Теорема 1.** *Дан квадрат на проективной плоскости. Прямая  $L$  не проходит через какую-либо вершину квадрата тогда и только тогда, когда существует кубическая кривая, проходящая через каждую вершину квадрата, но пересекающая прямую  $L$  лишь в точке на бесконечно удаленной прямой.*

**Доказательство.** Пусть однородные координаты вершин квадрата  $(1 : 0 : 0)$ ,  $(1 : 0 : 1)$ ,  $(1 : 1 : 0)$  и  $(1 : 1 : 1)$ . Проективная кубическая кривая определяется тернарной кубической формой. Такая форма, обращающаяся в нуль в каждой вершине квадрата, равна линейной комбинации шести форм  $x_k x_j (x_j - x_0)$ , где  $k \in \{0, 1, 2\}$  и  $j \in \{1, 2\}$ . Эти формы линейно независимы. Поэтому размерность линейного пространства кубических форм, обращающихся в нуль в каждой вершине квадрата, равна шести (см. [40]). Размерность пространства всех бинарных кубических форм равна четырём. Ограничение формы на прямую  $L$  определяет линейное отображение  $\pi$  из пространства тернарных форм, которые обращаются в нуль в каждой вершине квадрата, в пространство всех бинарных форм. Ядро отображения  $\pi$  состоит из форм, тождественно равных нулю на прямой  $L$ . Каждая форма из ядра приводима и делится на линейную форму  $\ell$ , определяющую прямую  $L$ . Поскольку по предположению прямая  $L$  не проходит через какую-либо вершину квадрата, то размерность ядра равна размерности квадратичных тернарных форм, обращающихся в нуль в каждой вершине квадрата. Согласно [40], это формы типа  $\lambda_1 x_1 (x_1 - x_0) + \lambda_2 x_2 (x_2 - x_0)$ . Поэтому размерность ядра отображения  $\pi$  равна двум. Следовательно, размерность образа отображения  $\pi$  совпадает с размерностью всех бинарных форм. Отображение  $\pi$  сюръективное. В частности, образом некоторой кубической формы  $c(x_0, x_1, x_2)$  из области определения  $\pi$  служит форма  $x_0^3$ . Уравнение  $c(x_0, x_1, x_2) = 0$  определяет кривую, которая пересекает прямую  $L$  лишь в бесконечно удаленной точке. Более того, поскольку размерность ядра равна двум, то существует однопараметрическое семейство таких кривых.

Это рассуждение существенно использует предположение, что прямая  $L$  не проходит через какую-либо вершину квадрата. Иначе ядру отображения  $\pi$  принадлежали бы произведения линейной формы  $\ell$  и некоторой квадратичной формы, которая обращается в нуль лишь в трех вершинах квадрата. Тогда размерность ядра больше двух, а отображение  $\pi$  несюръективное. Очевидно, в этом случае прямая  $L$  пересекает каждую кривую из области определения  $\pi$  в некоторой вершине квадрата. Теорема доказана.

В теореме 1 кубическую кривую нельзя заменить на конику. Действительно, тернарные квадратичные формы, обращающиеся в нуль в каждой вершине квадрата, порождают линейное пространство размерности два, но бинарные квадратичные формы порождают пространство размерности три. Поэтому аналог отображения  $\pi$  из доказательства теоремы 1 уже не будет сюръективным.

### 3.2. Линейные пространства форм

Размерность линейного пространства форм степени  $d$  от переменных  $x_0, \dots, x_s$  равна биномиальному коэффициенту  $\binom{s+d}{d}$ . Рассмотрим линейные комбинации кубических форм типа  $x_k(x_k - x_0)x_t$ . Каждая такая форма обращается в нуль в каждой точке с координатами  $x_0 = 1$  и  $x_k \in \{0, 1\}$ , где  $1 \leq k \leq n$ . Ограничение этих форм на линейное подпространство, определяемое системой уравнений  $x_j = \ell_j(x_0, \dots, x_s)$ , где  $s < j \leq n$ , равно линейной комбинации форм двух типов: либо  $x_k(x_k - x_0)x_t$ , либо  $\ell_j(\ell_j - x_0)x_t$ , где  $1 \leq k \leq s$  и  $0 \leq t \leq s$ .

**Лемма 2.** Дан набор линейных форм  $\ell_j(x_0, \dots, x_s) = a_{j0}x_0 + \dots + a_{js}x_s$ , где  $s < j \leq n$  и все коэффициенты  $a_{ji}$  — алгебраически независимые элементы чисто трансцендентного расширения поля рациональных чисел степени трансцендентности  $(n-s)(s+1)$ . Если выполнено неравенство  $(s+3)(s+2) \leq 6(n-s)$ , то каждая кубическая форма от переменных  $x_0, \dots, x_s$  равна некоторой линейной комбинации форм типа  $\ell_j^2 x_t$ , где  $s < j \leq n$  и  $0 \leq t \leq s$ .

**Доказательство.** Обозначим через  $M$  матрицу, составленную из коэффициентов форм типа  $\ell_j^2 x_t$ , где столбцы соответствуют мономам, а строки — формам. Общее число форм типа  $\ell_j^2 x_t$  должно быть не меньше, чем число мономов третьей степени. То есть должно выполняться неравенство

$$\frac{(s+3)!}{s!3!} \leq (n-s)(s+1),$$

где слева — общее число мономов третьей степени, а справа — число форм типа  $\ell_j^2 x_t$ . Оно эквивалентно неравенству  $(s+3)(s+2) \leq 6(n-s)$  из условия леммы. При выполнении этого неравенства достаточно показать, что матрица  $M$  имеет полный ранг, равный числу столбцов.

Выделим в матрице  $M$  квадратную подматрицу  $M'$  с тем же числом столбцов, которая содержит набор ненулевых элементов по одному из каждой строки и каждого столбца. Определитель равен альтернированной сумме произведений элементов из наборов по одному элементу из каждой строки и каждого столбца. Чисто трансцендентное расширение поля изоморфно полю рациональных функций. Элементы матрицы  $M'$  можно рассматривать как многочлены от переменных  $a_{jk}$ , которые упорядочены в соответствии с порядком на индексах:

$$a_{(s+1)0} > a_{(s+1)1} > \dots > a_{j0} > a_{j1} > \dots > a_{js} > \dots > a_{ns}.$$

Это определяет мономиальное упорядочение. Мы выбираем в матрице  $M'$  набор ненулевых элементов по одному из каждой строки и каждого столбца, для которого произведение максимально при этом мономиальном упорядочении. Этот максимум достигается на единственном наборе элементов из  $M'$ . Действительно, если два разных таких набора имеют равные произведения элементов, то можно найти третий набор, произведение элементов которого больше.

Поскольку  $a_{jk}$  алгебраически независимые, никакое их произведение не равно линейной комбинации других произведений. Следовательно, определитель матрицы  $M'$  отличен от нуля. Поэтому матрица  $M$  имеет полный ранг. Лемма доказана.

Элементы чисто трансцендентного расширения поля можно отождествить с функциями от независимых переменных  $a_{ji}$ . Лемма 2 справедлива для некоторых разреженных матриц, получаемых подстановкой целых значений переменных  $a_{ji}$ .

Пусть  $n = 3$  и  $s = 1$ . Выберем  $\ell_2 = x_0$  и  $\ell_3 = x_1$ . Матрица коэффициентов кубических форм типа  $\ell_j^2 x_t$  невырожденная, поскольку она равна единичной  $4 \times 4$  матрице. Поскольку размерность линейного пространства бинарных кубических форм равна четырем, формы типа  $\ell_j^2 x_t$  порождают это пространство.

Пусть  $n = 6$  и  $s = 2$ . Выберем  $\ell_3 = x_0$ ,  $\ell_4 = x_1$ ,  $\ell_5 = x_2$  и  $\ell_6 = x_1 + x_2$ . Матрица коэффициентов кубических форм типа  $\ell_j^2 x_t$  содержит двенадцать строк и десять столбцов. Ранг матрицы  $M$  равен десяти и совпадает с размерностью линейного пространства тернарных кубических форм.

Пусть  $n = 8$  и  $s = 3$ . Выберем  $\ell_4 = x_0$ ,  $\ell_5 = x_1$ ,  $\ell_6 = x_2$ ,  $\ell_7 = x_3$  и  $\ell_8 = x_0 + x_1 + x_2 + x_3$ . Матрица коэффициентов кубических форм типа  $\ell_j^2 x_t$  квадратная. Ранг равен двадцати и совпадает с размерностью линейного пространства кубических форм от четырех переменных.

**Лемма 3.** Дан набор линейных форм  $\ell_j(x_0, \dots, x_s) = a_{j0}x_0 + \dots + a_{js}x_s$ , где  $s < j \leq n$  и все коэффициенты  $a_{ji}$  — алгебраически независимые элементы чисто трансцендентного расширения поля рациональных чисел степени трансцендентности  $(n-s)(s+1)$ . Если выполнено неравенство  $(s+3)(s+2) \leq 6(n-s)$ , то каждая кубическая форма от переменных  $x_0, \dots, x_s$  равна некоторой линейной комбинации форм типа  $\ell_j(\ell_j - x_0)x_t$ , где  $s < j \leq n$  и  $0 \leq t \leq s$ .

**Доказательство.** По аналогии с доказательством леммы 2, покажем, что матрица  $N$ , состоящая из коэффициентов форм типа  $\ell_j(\ell_j - x_0)x_t$ , имеет полный ранг. Матрица  $N$  равна сумме матрицы  $M$ , состоящей из коэффициентов форм типа  $\ell_j^2 x_t$ , и матрицы  $B$ , состоящей из коэффициентов форм типа  $-x_0 \ell_j x_t$ . Однако элементы матрицы  $B$ , рассматриваемые как многочлены от переменных  $a_{ji}$ , имеют меньшую степень, чем элементы матрицы  $M$ . Поэтому ранг матрицы  $N = M + B$  не ниже ранга матрицы  $M$ . По лемме 2 матрица  $M$  имеет полный ранг. Следовательно, матрица  $N$  тоже имеет полный ранг. Лемма доказана.

---

**Алгоритм, обсуждаемый при доказательстве Теорем 2 и 3**


---

**Вход:** целые числа  $0 < m < n$  и линейные формы  $\ell_j(x_0, \dots, x_{n-m})$ , где  $n - m < j \leq n$ .

1: Элементами матрицы  $A$  служат коэффициенты формы

$$\sum_{t=0}^{n-m} \left( \sum_{k=1}^{n-m} \lambda_{tk} x_k (x_k - x_0) + \sum_{j=n-m+1}^n \lambda_{tj} \ell_j (\ell_j - x_0) \right) x_t,$$

где строки матрицы  $A$  соответствуют мономам третьей степени от переменных  $x_0, \dots, x_{n-m}$ , а столбцы — переменным  $\lambda_{tk}$  и  $\lambda_{tj}$ ;

2: Расширенная матрица  $B$  получается из матрицы  $A$  добавлением столбца, в котором единица стоит в строке, соответствующей моному  $x_0^3$ , а остальные элементы этого столбца равны нулю;

3: **if**  $\text{rank}(A) = \text{rank}(B)$  **then** вход отвергается;

4: **else** выдается уведомление о неопределенности выбора.

---

### 3.3. Алгоритм и оценки вычислительной сложности

Говоря о задачах распознавания, мы предлагаем три варианта ответа: вход может быть не только принят или отвергнут, но также возможно явное уведомление о неопределенности выбора. При этом ответ должен быть получен за конечное время и без ошибок, а уведомление о неопределенности может быть выдано лишь на малой доле входов (см. [12], [15]).

Под вычислительной сложностью подразумевается алгебраическая сложность, которая равна числу арифметических операций с числами и сравнений чисел (см. [29]). Сложность выполнения отдельной операции не учитывается.

**Теорема 2.** *Существует алгоритм распознавания с тремя вариантами ответа, получающий на вход положительные целые числа  $n$  и  $m < n$ , а также систему из  $t$  линейных форм  $\ell_j(x_0, \dots, x_{n-m})$ , где  $n - m < j \leq n$ , который при выполнении неравенства  $(n - m + 3)(n - m + 2) \leq 6t$  удовлетворяет следующим условиям:*

- алгебраическая сложность алгоритма ограничена сверху многочленом от  $n$ ;
- если вход отвергается, то система уравнений  $x_j = \ell_j(1, x_1, \dots, x_{n-m})$  для индексов  $n - m < j \leq n$  не имеет никакого  $\{0, 1\}$ -решения;
- если вход принимается, то система уравнений  $x_j = \ell_j(1, x_1, \dots, x_{n-m})$  для индексов  $n - m < j \leq n$  имеет  $\{0, 1\}$ -решение;
- для любого допустимого набора значений  $n$  и  $m$  существует такой отличный от тождественно нулевого многочлен степени не выше  $n^2(n - m + 1)^2$  от коэффициентов всех линейных форм  $\ell_j$ , что если алгоритм на некотором входе выдает уведомление о неопределенности выбора, то этот многочлен обращается в нуль для соответствующего набора значений коэффициентов.

**Доказательство.** По сути, алгоритм, приведенный в настоящей работе, проверяет существование наборов чисел  $\lambda_{tk}$  и  $\lambda_{tj}$ , для которых выполнено равенство двух кубических форм

$$\sum_{t=0}^{n-m} \left( \sum_{k=1}^{n-m} \lambda_{tk} x_k (x_k - x_0) + \sum_{j=n-m+1}^n \lambda_{tj} \ell_j (\ell_j - x_0) \right) x_t = x_0^3.$$

Здесь  $0 \leq t \leq n - m$ . Эта проверка сводится к решению системы линейных уравнений относительно  $\lambda_{tk}$  и  $\lambda_{tj}$ . Общее число переменных в этой системе не превышает  $n(n - m + 1)$ . Каждое уравнение в этой системе соответствует очередному моному третьей степени от переменных  $x_0, \dots, x_{n-m}$ . Поэтому число этих уравнений равно  $(n - m + 1)(n - m + 2)(n - m + 3)/6$ . При выполнении неравенства из условия теоремы, число уравнений этой системы не превосходит числа переменных  $\lambda_{tk}$  и  $\lambda_{tj}$ . По теореме Кронекера–Капелли решение существует при совпадении рангов двух матриц. При этом расширенная матрица получается добавлением столбца, в котором равны нулю все элементы, кроме одного. Ранг матрицы легко вычисляется. Элементами этих матриц слу-

жат многочлены степени не выше второй от коэффициентов линейных форм  $\ell_j$ . Достаточным условием существования решения служит отличие от нуля минора, равного многочлену степени не выше  $n^2(n-m+1)^2$  от коэффициентов линейных форм  $\ell_j$ . По лемме 3 этот многочлен не обращается в нуль тождественно. Теорема доказана.

Если алгоритм из теоремы 2 отвергает систему уравнений  $S$ , то он отвергает и любую систему уравнений  $S'$ , которая получена добавлением к  $S$  новых уравнений.

При сравнении теорем 1 и 2 видно, что условие в теореме 2 неоптимальное при  $n = 2$ . В частности, это связано с грубой оценкой ранга матрицы в доказательстве леммы 3.

**Теорема 3.** *Существует алгоритм распознавания с тремя вариантами ответа, получающий на вход положительные целые числа  $n$  и  $m < n$ , а также систему из  $m$  линейных форм  $\ell_j(x_0, \dots, x_{n-m})$ , где  $n-m < j \leq n$ , который при выполнении неравенства  $(n-m+3)(n-m+2) \leq 6m$  удовлетворяет следующим условиям:*

- алгебраическая сложность алгоритма ограничена сверху многочленом от  $n$ ;
- если вход отвергается, то система уравнений  $x_j = \ell_j(1, x_1, \dots, x_{n-m})$  для индексов  $n-m < j \leq n$  не имеет никакого  $\{0,1\}$ -решения;
- если вход принимается, то система уравнений  $x_j = \ell_j(1, x_1, \dots, x_{n-m})$  для индексов  $n-m < j \leq n$  имеет  $\{0,1\}$ -решение;
- для любого рационального числа  $\varepsilon$  из интервала  $0 < \varepsilon < 1$  и для любых допустимых значений  $n$  и  $m$ , если все коэффициенты форм  $\ell_j$  независимо и равномерно распределены на множестве мощностей не меньше  $(1/\varepsilon)n^2(n-m+1)^2$ , то вероятность выдачи уведомления о неопределенности выбора не превышает числа  $\varepsilon$ .

**Доказательство.** Применяем, как и при доказательстве теоремы 2, алгоритм, представленный в настоящей работе. Существует отличный от тождественно нулевого многочлен степени не выше  $n^2(n-m+1)^2$  от коэффициентов всех линейных форм  $\ell_j$ , который обращается в нуль на тех наборах, на которых выдается уведомление о неопределенности выбора. По лемме Шварца–Зиппеля (лемма 1) вероятность обращения этого многочлена в нуль не превышает числа  $\varepsilon$ . Теорема доказана.

#### 4. ЗАКЛЮЧЕНИЕ

Неравенство из теорем 2 и 3 имеет вид  $m \geq n - \sqrt{6n - o(n)}$ . Аналогичное условие в ранее опубликованном алгоритме [15], использующем не кубические, а квадратичные формы, имело вид  $m \geq n - \sqrt{2n - o(n)}$ . Поэтому новый алгоритм применим при меньших ограничениях. С другой стороны, поскольку в основе работы алгоритма лежит вычисление ранга матрицы, вычисления могут быть эффективно реализованы на многопроцессорных вычислительных устройствах.

Предлагаемый алгоритм не обеспечивает полиномиальную вычислительную сложность в худшем случае. Но для многих случаев он гораздо эффективнее полного перебора. Если коэффициенты при линейных членах уравнений положительные, то в среднем более эффективен алгоритм, который предложил Н.Н. Кузюрин (см. [13], [14]). Однако новый алгоритм может быть эффективнее в случае коэффициентов разных знаков.

Существуют различные пути сведения данной системы линейных уравнений к другой такой системе, имеющей столько же  $\{0,1\}$ -решений. Это открывает дополнительные возможности преодолеть неопределенность ответа. С другой стороны, хотя рассмотрена лишь задача распознавания, бинарный поиск позволяет искать само  $\{0,1\}$ -решение, если оно существует. В частности, доказательство отсутствия  $\{0,1\}$ -решения, соответствующего вершине какой-либо фасеты единичного куба, позволяет уменьшить число переменных в исходной задаче. То же справедливо, если вместо фасеты рассматривать пару симметрично расположенных граней коразмерности два, лежащих в одной гиперплоскости. На этом пути возможно комбинированное применение разных методов на разных шагах.

Рассмотренный в настоящей статье алгоритм никогда не принимает вход. Но используя другие известные методы, включая алгоритм Кузюрина, можно уменьшить число уведомлений об отказе. После этого новый алгоритм может принимать некоторые входы.

Хорошо известна близкая NP-полная задача о выполнимости 3-КНФ, для решения которой известны эвристические алгоритмы и их программные реализации. Хотя последняя задача также сводится за полиномиальное время к задаче о сумме подмножества, для решения задачи о выполнимости 3-КНФ новый предлагаемый алгоритм, вероятно, не даст выигрыша для практически применимых размеров входа. Однако новый алгоритм применим к более широкому набору частных задач. Это увеличивает интерес к обобщениям задачи о сумме подмножества, а не только к задаче о выполнимости 3-КНФ.

Автор благодарен М.Д. Малых и анонимному рецензенту за полезные замечания.

## СПИСОК ЛИТЕРАТУРЫ

1. *Horowitz E., Sahni S.* Computing partitions with applications to the knapsack problem // J. ACM. 1974. V. 21. № 2. P. 277–292.  
<https://doi.org/10.1145/321812.321823>
2. *Meer K.* A note on a  $P \neq NP$  result for a restricted class of real machines // J. Complexity. 1992. V. 8. № 4. P. 451–453.  
[https://doi.org/10.1016/0885-064X\(92\)90007-X](https://doi.org/10.1016/0885-064X(92)90007-X)
3. *Koiran P.* Computing over the reals with addition and order // Theoret. Comput. Sci. 1994. V. 133. № 1. P. 35–47.  
[https://doi.org/10.1016/0304-3975\(93\)00063-B](https://doi.org/10.1016/0304-3975(93)00063-B)
4. *Cucker F., Matamala M.* On digital nondeterminism // Math. Systems Theory. 1996. V. 29. P. 635–647.  
<https://doi.org/10.1007/BF01301968>
5. *Grigoriev D.* Complexity of Positivstellensatz proofs for the knapsack // Comput. Complexity. 2001. V. 10. P. 139–154.  
<https://doi.org/10.1007/s00037-001-8192-0>
6. *Margulies S., Onn S., Pasechnik D.V.* On the complexity of Hilbert refutations for partition // J. Symbolic Comput. 2015. V. 66. P. 70–83.  
<https://doi.org/10.1016/j.jsc.2013.06.005>
7. *Koiliaris K., Xu C.* Faster pseudopolynomial time algorithms for subset sum // ACM Trans. Algorithms. 2019. V. 15. № 3. Article 40. P. 1–20.  
<https://doi.org/10.1145/3329863>
8. *Polak A., Rohwedder L., Wegrzycki K.* Knapsack and subset sum with small items // In: Bansal N., Merelli E., Worrell J. (eds) 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021), Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, 2021. V. 198. № 106. P. 1–19.  
<https://doi.org/10.4230/LIPIcs.ICALP.2021.106>
9. *Lagarias J.C., Odlyzko A.M.* Solving low-density subset sum problems // J. ACM. 1985. V. 32. № 1. P. 229–246.  
<https://doi.org/10.1145/2455.2461>
10. *Coster M.J., Joux A., LaMacchia B.A., Odlyzko A.M., Schnorr C.P., Stern J.* Improved low-density subset sum algorithms // Comput. Complexity. 1992. V. 2. № 2. P. 111–128.  
<https://doi.org/10.1007/BF01201999>
11. *May A.* Solving subset sum with small space – Handling cryptanalytic Big Data // it – Information Technology. 2020. V. 62. № 3–4. P. 181–187.  
<https://doi.org/10.1515/itit-2019-0038>
12. *Рыбалов А.Н.* О генерической сложности проблемы о сумме подмножеств для полугрупп целочисленных матриц // Прикладная дискретная математика. 2020. № 50. С. 118–126.  
<https://doi.org/10.17223/20710410/50/9>
13. *Кузюрин Н.Н.* Полиномиальный в среднем алгоритм в целочисленном линейном программировании // Сиб. журн. исслед. опер. 1994. Т. 1. № 3. С. 38–48.
14. *Kuzyurin N.N.* An integer linear programming algorithm polynomial in the average case // In: Korshunov A.D. (eds.) Discrete Analysis and Operations Research. Mathematics and Its Applications. V. 355. P. 143–152. Springer, Dordrecht, 1996.  
<https://doi.org/10.1007/978-94-009-1606-7>
15. *Селиверстов А.В.* Двоичные решения для больших систем линейных уравнений // Прикладная дискретная математика. 2021. № 52. С. 5–15.  
<https://doi.org/10.17223/20710410/52/1>
16. *Pan Y., Zhang F.* Solving low-density multiple subset sum problems with SVP oracle // J. Syst. Sci. Complex. 2016. V. 29. P. 228–242.  
<https://doi.org/10.1007/s11424-015-3324-9>

17. Селиверстов А.В. О двоичных решениях систем уравнений // Прикладная Дискретная Математика. 2019. № 45. С. 26–32.  
<https://doi.org/10.17223/20710410/45/3>
18. Martins J.P., Ribas B.C. A randomized heuristic repair for the multidimensional knapsack problem // Optim. Lett. 2021. V. 15. P. 337–355.  
<https://doi.org/10.1007/s11590-020-01611-1>
19. Cacchiani V., Iori M., Locatelli A., Martello S. Knapsack problems – An overview of recent advances. Part II: Multiple, multidimensional, and quadratic knapsack problems // Comput. and Operat. Res. 2022. V. 143. № 105693. P. 1–14.  
<https://doi.org/10.1016/j.cor.2021.105693>
20. Griбанов D.V., Zolotykh N.Y. On lattice point counting in  $\Delta$ -modular polyhedra // Optim. Lett. 2022. V. 16. P. 1991–2018.  
<https://doi.org/10.1007/s11590-021-01744-x>
21. Al-Shihabi S. A novel core-based optimization framework for binary integer programs – the multidimensional knapsack problem as a test problem // Operat. Res. Perspectiv. 2021. V. 8. № 100182. P. 1–13.  
<https://doi.org/10.1016/j.orp.2021.100182>
22. Лотов А.В., Рябиков А.И. Дополненный метод стартовой площадки для аппроксимации границы Парето в задачах с многоэкстремальными критериями // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 10. С. 1734–1744.  
<https://doi.org/10.31857/S0044466921100100>
23. Жадан В.Г. Прямо-двойственный метод Ньютона с наискорейшим спуском для линейной задачи полуопределенного программирования. Ньютоновская система уравнений // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 2. С. 232–248.  
<https://doi.org/10.31857/S0044466922020132>
24. Fu H., Xu Y., Wu G., Liu J., Chen S., He X. Emphasis on the flipping variable: Towards effective local search for hard random satisfiability // Informat. Sci. 2021. V. 566. P. 118–139.  
<https://doi.org/10.1016/j.ins.2021.03.009>
25. Fu H., Liu J., Wu G., Xu Y., Sutcliffe G. Improving probability selection based weights for satisfiability problems // Knowledge-Based Systems. 2022. V. 245. № 108572. P. 1–17.  
<https://doi.org/10.1016/j.knosys.2022.108572>
26. Guo P., Zhang Y. ISSATA: An algorithm for solving the 3-satisfiability problem based on improved strategy // Applied Intelligence. 2022. V. 52. P. 1740–1751.  
<https://doi.org/10.1007/s10489-021-02493-1>
27. Cai S., Lei Z. Old techniques in new ways: Clause weighting, unit propagation and hybridization for maximum satisfiability // Artificial Intelligence. 2020. V. 287. № 103354. P. 1–14.  
<https://doi.org/10.1016/j.artint.2020.103354>
28. Li W., Xu C., Yang Y., Chen J., Wang J. A refined branching algorithm for the maximum satisfiability problem // Algorithmica. 2022. V. 84. P. 982–1006.  
<https://doi.org/10.1007/s00453-022-00938-8>
29. Абрамов С.А. Лекции о сложности алгоритмов. М: МЦНМО, 2012.
30. Алаев П.Е., Селиванов В.Л. Поля алгебраических чисел, вычислимые за полиномиальное время. I // Алгебра и логика. 2019. Т. 58, № 6. С. 673–705.  
<https://doi.org/10.33048/alglog.2019.58.601>
31. Батхин А.Б. Параметризация дискриминантного множества многочлена // Программирование. 2016. № 2. С. 8–21.
32. Селиверстов А.В. Эвристические алгоритмы распознавания некоторых кубических гиперповерхностей // Программирование. 2021. № 1. С. 65–72.  
<https://doi.org/10.31857/S0132347421010106>
33. Schwartz J. Fast probabilistic algorithms for verification of polynomial identities // J. ACM. 1980. V. 27. № 4. P. 701–717.  
<https://doi.org/10.1145/322217.322225>
34. Halbeisen L., Hungerbühler N., Schumacher S. Magic sets for polynomials of degree  $n$  // Linear Algebra Appl. 2021. V. 609. P. 413–441.  
<https://doi.org/10.1016/j.laa.2020.09.026>
35. Chistov A.L. Fast parallel calculation of the rank of matrices over a field of arbitrary characteristic // In: L. Budach (eds) Fundamentals of Computation Theory. FCT 1985. Lecture Notes in Computer Science, vol. 199. Springer, Berlin, Heidelberg, 1985. P. 63–69.  
<https://doi.org/10.1007/BFb0028792>

36. *Mulmuley K.* A fast parallel algorithm to compute the rank of a matrix over an arbitrary field // *Combinatorica*. 1987. V. 7. № 1. P. 101–104.  
<https://doi.org/10.1007/BF02579205>
37. *Переславцева О.Н.* О вычислении характеристического полинома матрицы // *Дискретная матем.* 2011. Т. 23. № 1. С. 28–45.  
<https://doi.org/10.4213/dm1128>
38. *Cheung H.Y., Kwok T.C., Lau L.C.* Fast matrix rank algorithms and applications // *J. ACM*. 2013. V. 60. № 5. Article № 31. P. 1–25.  
<https://doi.org/10.1145/2528404>
39. *Malaschonok G.I., Seliverstov A.V.* Calculation of integrals in MathPartner // *Discrete and Continuous Model. and Appl. Comput. Sci.* 2021. V. 29. № 4. P. 337–346.  
<https://doi.org/10.22363/2658-4670-2021-29-4-337-346>
40. *Seliverstov A.V., Lyubetsky V.A.* About forms equal to zero at each vertex of a cube // *J. of Communicat. Techn. and Electron.* 2012. V. 57. № 8. P. 892–895.  
<https://doi.org/10.1134/S1064226912080049>

---

## ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

---

УДК 517.23

### БАД-МЕТОДОЛОГИЯ И ДИФФЕРЕНЦИРОВАНИЕ СЛОЖНОЙ ФУНКЦИИ<sup>1)</sup>

© 2023 г. А. Ф. Албу<sup>1</sup>, А. Ю. Горчаков<sup>1,2</sup>, В. И. Zubov<sup>1,\*</sup>

<sup>1</sup> 119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия

<sup>2</sup> 141700 М. о., Долгопрудный, Институтский пер., 9, МФТИ, Россия

\*e-mail: vladimir.zubov@mail.ru

Поступила в редакцию 27.04.2022 г.  
Переработанный вариант 27.04.2022 г.  
Принята к публикации 17.09.2022 г.

Сравниваются разные подходы к вычислению градиента сложной функции многих переменных, такие как использование точных, аналитически выведенных формул; использование формул, полученных с помощью методологии быстрого автоматического дифференцирования; использование стандартных программных пакетов, реализующих идеи методологии быстрого автоматического дифференцирования. Сравнение подходов осуществляется на примере сложной функции, представляющей энергию системы атомов, потенциал взаимодействия которых – потенциал Терсоффа. В качестве критерия сравнения используется компьютерное время, необходимое для вычисления градиента функции. Результаты показывают превосходство методологии быстрого автоматического дифференцирования по сравнению с подходом, использующим аналитические формулы. Стандартные пакеты вычисляют градиент функции примерно за то же время, что и при использовании формул методологии быстрого автоматического дифференцирования. Библ. 15. Фиг. 1. Табл. 5.

**Ключевые слова:** быстрое автоматическое дифференцирование, многошаговый процесс, пакеты стандартных программ.

DOI: 10.31857/S0044466923010039, EDN: LEJFIO

### ВВЕДЕНИЕ

В настоящее время методы математического моделирования и вычислительной математики все глубже проникают в различные области науки, техники и производства. Они занимают особое место среди современных методов решения прикладных задач. В связи с интенсификацией производства в последние десятилетия особенно возрос интерес к проблемам оптимизации, большую актуальность приобрели задачи оптимального управления сложными динамическими системами.

Можно выделить несколько различных направлений в разработке численных методов решения задач оптимального управления, которые существенно отличаются друг от друга. Прежде всего следует упомянуть прямые методы, основанные на спуске в пространстве управлений. Ряд исследований связан с непрямыми методами, в которых с помощью принципа максимума Л.С. Понтрягина исходная задача оптимального управления сводится к краевой (см. [1]). В [2] Р.П. Федоренко предложил подход, который базируется на использовании идей метода линеаризации. Большой цикл работ по численным методам оптимального управления был выполнен Н.Н. Моисеевым (см. [3]) и его учениками.

Другой подход численного решения задач оптимального управления базируется на идеях нелинейного программирования: задачи оптимального управления с условиями на управляющую функцию и с фазовыми ограничениями сводятся к задачам нелинейного программирования. Такой подход оказался чрезвычайно эффективным. Обычно такие задачи решаются градиентными методами минимизации целевого функционала. Но для этого необходимо уметь точно вычислять градиент целевой функции.

---

<sup>1)</sup> Работа выполнена при частичной финансовой поддержке РФФИ (проекта № 21-71-30005).

Простейший способ вычисления градиента достаточно сложной функции — определение компонент градиента с помощью метода конечных разностей. Однако этот подход достаточно трудоемок и не обеспечивает получение точного градиента. В [4] раскрыты проблемы, возникающие при использовании метода конечных разностей для вычисления градиента целевой функции в задачах оптимального управления сложными системами.

Второй способ вычисления градиента функционала — использование сопряженной задачи. Здесь представляется естественным сформулировать исходную задачу оптимального управления в “непрерывном” пространстве, сформулировать сопряженную задачу также в “непрерывном” пространстве, определить градиент функционала в “непрерывном” случае, и только после этого осуществлять аппроксимацию функционала, прямой и сопряженной задач. В результате этих действий получают формулы для определения градиента целевого функционала в дискретном варианте. Однако здесь необходимо учитывать то, что получаемый дискретный градиент функционала неточен. Его значение зависит от способа дискретизации как прямой и сопряженной задач, так и от способа дискретизации самого градиента. Эта неточность может быть уменьшена, если уменьшить погрешности указанных аппроксимаций, или даже полностью ликвидирована за счет выбора “согласованных” аппроксимаций прямой и сопряженной задач и “согласованной” с ними аппроксимации градиента. Эти “согласованные” аппроксимации можно случайно угадать, но указать способ определения “согласованных” аппроксимаций, используя данный подход, по-видимому, невозможно. Если выбрать неправильный дискретный вариант сопряженной задачи, получается неверное значение градиента. С такими случаями приходилось встречаться на практике (см., например, [5]).

Другой путь связан с формулировкой рассматриваемой задачи оптимального управления в конечномерном пространстве и с определением градиента функционала сразу в дискретном виде. Если дискретный градиент определять с помощью метода множителей Лагранжа (сопряженных переменных), то такой путь известен как обобщенная методология быстрого автоматического дифференцирования (БАД-методология).

БАД-методология значительно превосходит по эффективности аналитическое дифференцирование (точнее, использование аналитически полученных формул) и вычисление производных с помощью метода конечных разностей. В работах Ю.Г. Евтушенко [6–8] предложен общий подход к дифференцированию сложных функций. Там показано, что БАД-методология дает возможность с единой позиции рассмотреть самые разнообразные задачи. Например, из общих формул дифференцирования, приведенных в этих работах, несложно получить формулы БАД-методологии для определения градиента функций многих переменных. При этом вычисление значения функции представляется как многошаговый процесс, в котором введены новые фазовые переменные. Эти фазовые переменные являются функциями независимых переменных, по которым ищутся производные заданной функции. Показано, что при определенных условиях использование БАД-методологии позволяет получить точное значение градиента дискретной задачи оптимального управления (дискретного целевого функционала при наличии дискретных связей) за время, не превосходящее утроенного времени вычисления самой функции.

БАД-методология эффективно использовалась при решении многих теоретически сложных прикладных задач, имеющих большое значение для народного хозяйства. Например, она была применена при решении задач из области материаловедения, при исследовании задачи моделирования и управления процессом сварки материалов, задачи управления процессом кристаллизации металла в литейном деле.

В настоящей работе проводится сравнение разных подходов к вычислению градиента сложной функции многих переменных. Сравнимаемыми подходами являются: использование точных, аналитически выведенных формул; использование формул, полученных с помощью БАД-методологии; использование стандартных программных пакетов, реализующих идеи БАД-методологии. Сравнение подходов осуществляется на примере сложной функции, представляющей энергию системы атомов, потенциал взаимодействия которых — потенциал Терсоффа.

## 1. СРАВНЕНИЕ РАЗЛИЧНЫХ МЕТОДОВ ДИФФЕРЕНЦИРОВАНИЯ ФУНКЦИЙ

Функция  $f(u)$  называется элементарной функцией, если она может быть представлена как конечная суперпозиция основных элементарных функций и арифметических действий. Будем считать, что дифференцируемая скалярная функция  $f(u)$  векторного аргумента  $u \in R^r$  задана явно. Задача состоит в вычислении частных производных функции  $f(u)$  по компонентам век-

тора  $u$ . В вычислительной практике часто встречаются функции, для которых точное вычисление частных производных представляет собой непростую задачу.

В таком случае можно воспользоваться несколькими способами нахождения частных производных функции  $f(u)$ :

1. *Приближенное численное дифференцирование.* Если  $f(u)$  — функция  $r$ -мерного вектора  $u$ , то при использовании самой простой формулы численного дифференцирования необходимо по меньшей мере  $r + 1$  раз вычислить значение самой функции  $f(u)$ , и  $r$  раз произвести вычитание и деление. Результаты таких расчетов будут точными только для линейных функций; во всех остальных случаях градиент определяется с некоторой погрешностью. Ниже будет показано, что при вычислении частных производных энергии группы атомов (потенциал взаимодействия атомов — потенциал Терсоффа) по специфичным параметрам потенциала Терсоффа с помощью этого метода необходимо проводить дополнительные исследования, связанные с выбором приращения каждого параметра потенциала, причем при каждом его новом значении.

2. *Аналитическое дифференцирование.* Здесь имеется в виду использование выведенных аналитических выражений для вычисления частных производных функции  $f(u)$ . Следует также отметить, что в современной компьютерной литературе такой подход нередко называют прямым методом автоматического дифференцирования, или аналитическим дифференцированием. Если вектор  $u$  — вектор большой размерности, то при применении этой техники возникают большие трудности.

3. *Использование формул БАД (БАД-методология)* (см. [6]–[8]). При использовании этого метода вычисление значения функции  $f(u)$  записывается в виде многошагового процесса в каноническом виде, формулируется сопряженная задача в каноническом виде и с их помощью выводятся аналитические выражения для вычисления частных производных рассматриваемой функции.

4. *Использование стандартных пакетов программ автоматического дифференцирования.*

а) Используются пакеты программ для генерации кода вычисления производных по коду вычисления функции (source-to-source AD tool). В качестве примера такой программы можно привести TAPENADE (см. [9]).

б) Используются пакеты программ, перегружающих операторы в коде вычисления функции. В результате такой перезагрузки код вычисляет не только значение функции, но и значение ее производных (Operator Overloading AD tool). В качестве примера такой программы можно привести CoDiPack и Adept (см. [10], [11]).

Ниже в этом разделе приводится сравнение дифференцирования сложной функции с помощью формул БАД-методологии (способ 3) и с помощью аналитически полученных выражений для ее частных производных (способ 2).

Для того чтобы воспользоваться техникой БАД (способ 3), необходимо представить вычисление элементарной функции  $f(u)$  как многошаговый процесс. С этой целью введем вектор  $z \in R^k$  фазовых переменных (вектор вспомогательных, промежуточных значений функции):

$$z_1 = F(1, Z_1, U_1), \quad z_2 = F(2, Z_2, U_2), \quad \dots, \quad z_k = F(k, Z_k, U_k), \quad (1.1)$$

где  $Z_i$  — набор элементов  $z_j$ , которые входят в правую часть равенства  $z_i = F(i, Z_i, U_i)$ , а  $U_i$  — набор элементов  $u_j$ , которые входят в правую часть этого равенства. Последовательность (1.1) строится таким образом, чтобы процесс был явным, и последняя компонента  $z_k$  совпала бы со значением функции  $f(u)$ , т.е.  $z_k = f(u)$ .

Предполагаем, что все функции  $F(i, Z_i, U_i)$  дифференцируемы по всем компонентам векторов  $z$  и  $u$ . Введем вектор сопряженных переменных  $p \in R^k$ , который определим как решение следующей системы линейных алгебраических уравнений:

$$p_l = \sum_{q \in \bar{Q}_l} F_{z_q}^T(q, Z_q, U_q) p_q, \quad \bar{Q}_l = \{i: z_l \in Z_i\}, \quad l = \overline{1, k}. \quad (1.2)$$

Тогда, согласно формулам БАД-методологии (см. [8]), компоненты градиента функции многих переменных  $f(u)$  вычисляются по формуле

$$\frac{\partial f}{\partial u_m} = \sum_{q \in \bar{K}_m} F_{u_m}^T(q, Z_q, U_q) p_q, \quad \bar{K}_m = \{i: u_m \in U_i\}, \quad m = \overline{1, r}. \quad (1.3)$$

Таким образом, мы сначала последовательно выполняем вычисления в соответствии с  $k$ -шаговым процессом и запоминаем значения переменных  $z$  и  $u$  (прямой ход), а потом идем в обратную сторону и рассчитываем сопряженные переменные  $p$  (обратный ход).

Что касается вычисления градиента функции методом 2, то здесь сопряженные переменные не определяются и не используются. Учитывая тот факт, что сложная функция  $f(u)$  представляет собой суперпозицию более простых функций, представим вычисление ее значения в виде  $k$ -шагового алгоритма (1.1).

Тогда производные  $\frac{\partial f}{\partial u_m}$ ,  $m = \overline{1, r}$ , вычисляются аналитически последовательно следующим образом:

$$\frac{\partial z_1}{\partial u_m}, \quad \frac{\partial z_2}{\partial u_m} = \frac{\partial z_2}{\partial z_1} \frac{\partial z_1}{\partial u_m}, \quad \frac{\partial z_3}{\partial u_m} = \frac{\partial z_3}{\partial z_2} \frac{\partial z_2}{\partial u_m}, \quad \dots, \quad \frac{\partial z_k}{\partial u_m} = \frac{\partial z_k}{\partial z_{k-1}} \frac{\partial z_{k-1}}{\partial u_m} = \frac{\partial f}{\partial u_m}, \quad m = \overline{1, r}. \quad (1.4)$$

Работу с формулами (1.2), (1.3) и (1.4) мы вначале проиллюстрируем на простом примере. Пусть задана элементарная функция  $f(u) = f(u_1, u_2) = e^{u_1 + (u_2)^2} + \sin((u_1)^3 + u_2)^2$ . Для применения формул БАД-методологии (метод 3) строим многошаговый процесс ее вычисления следующим образом:

$$z_1 = u_1 + (u_2)^2, \quad z_2 = e^{z_1}, \quad z_3 = (u_1)^3 + u_2, \quad z_4 = (z_3)^2, \quad z_5 = \sin(z_4), \quad z_6 = z_2 + z_5. \quad (1.5)$$

Используя формулы (1.2) и (1.3), имеем

$$p_1 = e^{z_1} p_2, \quad p_2 = p_6, \quad p_3 = 2z_3 p_4, \quad p_4 = \cos(z_4) p_5, \quad p_5 = p_6, \quad p_6 = 1, \quad (1.6)$$

$$\frac{\partial f}{\partial u_1} = p_1 + 3(u_1)^2 p_3, \quad \frac{\partial f}{\partial u_2} = 2u_2 p_1 + p_3. \quad (1.7)$$

Сопряженные переменные вычисляются в обратном порядке:  $p_6, p_5, \dots, p_1$ .

Для того чтобы получить частные производные этой функции без использования сопряженной задачи (с помощью формул (1.4), метод 2), введем обозначения  $\tilde{z}_i = \frac{\partial z_i}{\partial u_1}$ ,  $\tilde{\tilde{z}}_i = \frac{\partial z_i}{\partial u_2}$ . Значения  $\tilde{z}_i$  и  $\tilde{\tilde{z}}_i$  вычисляются по формулам

$$\tilde{z}_1 = 1, \quad \tilde{z}_2 = \tilde{z}_1 e^{z_1}, \quad \tilde{z}_3 = 3(u_1)^2, \quad \tilde{z}_4 = 2z_3 \tilde{z}_3, \quad \tilde{z}_5 = \tilde{z}_4 \cos z_4, \quad \tilde{z}_6 = \tilde{z}_2 + \tilde{z}_5, \quad (1.8)$$

$$\tilde{\tilde{z}}_1 = 2u_2, \quad \tilde{\tilde{z}}_2 = \tilde{\tilde{z}}_1 e^{z_1}, \quad \tilde{\tilde{z}}_3 = 1, \quad \tilde{\tilde{z}}_4 = 2z_3 \tilde{\tilde{z}}_3, \quad \tilde{\tilde{z}}_5 = \tilde{\tilde{z}}_4 \cos z_4, \quad \tilde{\tilde{z}}_6 = \tilde{\tilde{z}}_2 + \tilde{\tilde{z}}_5. \quad (1.9)$$

Тогда  $\frac{\partial f}{\partial u_1} = \tilde{z}_6$ ,  $\frac{\partial f}{\partial u_2} = \tilde{\tilde{z}}_6$ .

В случае метода 2 вместо одного многошагового процесса (1.6), который требуется для вычисления сопряженных переменных, необходимо выполнить вычисления по  $r = 2$  (количество независимых переменных) многошаговым процессам: (1.8) и (1.9). Следовательно, время, необходимое для вычисления градиента с помощью формул БАД, примерно в  $r$  раз меньше времени, которое необходимо для вычисления этого градиента с помощью формул (1.4). Эффективность использования техники БАД возрастает, если растет количество независимых переменных.

Вычислительные эксперименты, касающиеся сравнения способов определения производных функции многих переменных, были также проведены и для достаточно сложной и практически важной функции, которая определяет энергию группы атомов, потенциал взаимодействия кото-

рых — потенциал Терсоффа. В этом случае энергия  $E$  взаимодействия атомов некоторой кристаллической решетки вычисляется по формуле

$$E = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I V_{ij},$$

где

$$V_{ij} = f_c(r_{ij})(V_R(r_{ij}) - b_{ij}V_A(r_{ij})),$$

$$f_c(r) = \begin{cases} 1, & r < R - R_{\text{cut}} = R_m, \\ \frac{1}{2} \left( 1 - \sin \left( \frac{\pi(r - R)}{2R_{\text{cut}}} \right) \right), & R_m < r < R_p, \\ 0, & r > R + R_{\text{cut}} = R_p, \end{cases}$$

$$V_{ij}^R = V_R(r_{ij}) = \frac{D_e}{S-1} \exp(-\beta \sqrt{2S}(r_{ij} - r_e)), \quad V_{ij}^A = V_A(r_{ij}) = \frac{SD_e}{S-1} \exp\left(-\beta \sqrt{\frac{2}{S}}(r_{ij} - r_e)\right),$$

$$b_{ij} = \left(1 + (\gamma \varsigma_{ij})^\eta\right)^{-\frac{1}{2\eta}}, \quad \varsigma_{ij} = \sum_{\substack{k=1 \\ k \neq i, j}}^I f_c(r_{ik}) g_{ijk} \omega_{ijk}, \quad \omega_{ijk} = \exp(\lambda^3 \tau_{ijk}),$$

$$\tau_{ijk} = (r_{ij} - r_{ik})^3, \quad g_{ijk} = g(\theta_{ijk}) = 1 + \left(\frac{c}{d}\right)^2 - \frac{c^2}{d^2 + (h - \cos \theta_{ijk})^2}.$$

Здесь  $I$  — количество атомов в рассматриваемой группе;  $R$  и  $R_{\text{cut}}$  — параметры, определяющиеся из полученных экспериментально геометрических характеристик вещества;  $x_i = (x_{1i}, x_{2i}, x_{3i})$  — координаты  $i$ -го атома;  $r_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2}$  — расстояние между атомами с номерами  $i$  и  $j$ ;  $\theta_{ijk}$  — угол, образованный векторами, соединяющими атом  $i$  с атомами  $j$  и  $k$  соответственно. Косинус этого угла вычисляется по формуле  $q_{ijk} = \cos(\theta_{ijk}) = \frac{r_{ij}^2 + r_{ik}^2 - r_{jk}^2}{2r_{ij}r_{ik}}$ .

Потенциал Терсоффа зависит от десяти параметров ( $m = 10$ ), специфичных для моделируемого вещества ( $D_e, r_e, \beta, S, \eta, \gamma, \lambda, c, d, h$ ), и от координат атомов  $x = (x_1, x_2, \dots, x_I)^T$ . Указанные специфические параметры вещества, как правило, неизвестны. Для решения задач параметрической идентификации потенциалов межатомного взаимодействия в последнее время все большую актуальность приобретает применение различных методов оптимизации. В связи с этим возникает необходимость определения точного значения градиента функции энергии  $E$  по специфическим для моделируемого вещества параметрам потенциала Терсоффа. Формулы для вычисления компонент этого градиента были получены в [12] с помощью БАД-методологии.

Оптимизационные задачи решаются при определенном, фиксированном положении базисных атомов рассматриваемой кристаллической структуры. Решив задачу идентификации параметров, нет уверенности, что положения базисных атомов будут соответствовать минимуму потенциальной энергии системы. Поэтому следующим шагом исследований является оптимизация по координатам частиц, расставляющая частицы в положения, соответствующие минимуму суммарной потенциальной энергии рассматриваемой системы атомов. На этом этапе возникает необходимость вычисления градиента функции энергии по координатам атомов. Формулы для вычисления этого градиента также были получены в [12] на основе формул БАД-методологии (метод 3).

Ниже мы построим вычислительный процесс для определения указанного выше градиента с помощью метода 2 (аналитического дифференцирования) и сравним его с тем, который получен с помощью метода 3.

Функция энергии  $E$  представляет собой достаточно сложную суперпозицию более простых функций. Вычисление самой функции было представлено в [12] в виде некоторого многоступенчатого процесса.

Пусть  $\bar{u}$  и  $\bar{z}$  – векторы, имеющие координаты

$$\bar{u}^T = [u_1, u_2, \dots, u_{10}]^T, \quad \bar{z}^T = [z_1, z_2, \dots, z_{17}]^T,$$

где

$$\begin{aligned} u_1 &= D_e, \quad u_2 = r_e, \quad u_3 = \beta, \quad u_4 = S, \quad u_5 = \eta, \\ u_6 &= \gamma, \quad u_7 = \lambda, \quad u_8 = c, \quad u_9 = d, \quad u_{10} = h, \\ z_1 &= \left\{ z_1^{ijk} = \sqrt{(x_{1i} - x_{1k})^2 + (x_{2i} - x_{2k})^2 + (x_{3i} - x_{3k})^2} \right\}, \\ z_2 &= \left\{ z_2^{ijk} = \sqrt{(x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2 + (x_{3j} - x_{3k})^2} \right\}, \\ z_3 &= \left\{ z_3^{ijk} = q_{ijk} = \frac{(z_{13}^{ij})^2 + (z_1^{ijk})^2 - (z_2^{ijk})^2}{2z_{13}^{ij}z_1^{ijk}} \right\}, \\ z_4 &= \left\{ z_4^{ijk} = f_c(z_1^{ijk}) \right\}, \\ z_5 &= \left\{ z_5^{ijk} = g_{ijk} = 1 + \left( \frac{u_8}{u_9} \right)^2 - \frac{(u_8)^2}{(u_9)^2 + (u_{10} - z_3^{ijk})^2} \right\}, \\ z_6 &= \left\{ z_6^{ijk} = \tau_{ijk} = (z_{13}^{ij} - z_1^{ijk})^3 \right\}, \\ z_7 &= \left\{ z_7^{ijk} = \omega_{ijk} = \exp((u_7)^3 z_6^{ijk}) \right\}, \\ z_8 &= \left\{ z_8^{ijk} = f_c(r_{ik})g_{ijk}\omega_{ijk} = z_4^{ijk}z_5^{ijk}z_7^{ijk} \right\}, \\ z_9 &= \left\{ z_9^{ij} = \varsigma_{ij} = \sum_{k=1, k \neq i, j}^I z_8^{ijk} \right\}, \\ z_{10} &= \left\{ z_{10}^{ij} = \gamma\varsigma_{ij} = u_6 z_9^{ij} \right\}, \\ z_{11} &= \left\{ z_{11}^{ij} = (\gamma\varsigma_{ij})^\eta = (z_{10}^{ij})^{u_5} \right\}, \\ z_{12} &= \left\{ z_{12}^{ij} = b_{ij} = (1 + z_{11}^{ij})^{-\frac{1}{2u_5}} \right\}, \\ z_{13} &= \left\{ z_{13}^{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2} \right\}, \\ z_{14} &= \left\{ z_{14}^{ij} = V_{ij}^R = \frac{u_1}{u_4 - 1} \exp(-u_3 \sqrt{2u_4}(z_{13}^{ij} - u_2)) \right\}, \\ z_{15} &= \left\{ z_{15}^{ij} = V_{ij}^A = \frac{u_1 u_4}{u_4 - 1} \exp\left(-u_3 \sqrt{\frac{2}{u_4}}(z_{13}^{ij} - u_2)\right) \right\}, \\ z_{16} &= \left\{ z_{16}^{ij} = f_c(z_{13}^{ij}) \right\}, \\ z_{17} &= \left\{ z_{17}^{ij} = V_{ij} = z_{16}^{ij}(z_{14}^{ij} - z_{12}^{ij} z_{15}^{ij}) \right\}. \end{aligned}$$

Каждая компонента  $z_l$  вектора  $\bar{z}$  состоит из множества других компонент, а процесс вычисления функции энергии  $E$  строится следующим образом. Для всех фиксированных пар индексов  $(i, j)$ , таких что  $i = \overline{1, I}$ ,  $j = \overline{1, I}$  и  $j \neq i$ , вычисляются значения  $z_{13}^{ij}$ . Далее для всех таких пар индексов  $(i, j)$  и для всех индексов  $k = \overline{1, I}$ , таких что  $k \neq i$  и  $k \neq j$ , вычисляются значения  $z_1^{ijk}, z_2^{ijk}, \dots, z_8^{ijk}$ . Компоненты  $z_1^{ijk}$  используются для вычисления компонент  $z_3^{ijk}$ . При этом набор

индексов  $k$ , которые используются для вычисления компонент  $z_1^{ijk}$ , зависит от конкретного индекса  $j$  компоненты  $z_3^{ijk}$ . Из этих соображений мы присваиваем компонентам  $z_1^{ijk}$  и индекс  $j$ , хотя явно они от него не зависят. Аналогично, и компоненты  $z_2^{ijk}$  не зависят от индекса  $j$ , но мы им присваиваем этот индекс.

Далее для всех фиксированных пар индексов  $(i, j)$ , таких что  $i = \overline{1, I}$ ,  $j = \overline{1, I}$  и  $j \neq i$ , вычисляем компоненты  $z_9^{ij}, z_{10}^{ij}, z_{11}^{ij}, z_{12}^{ij}, z_{14}^{ij}, z_{15}^{ij}, z_{16}^{ij}, z_{17}^{ij}$ . Компоненты  $z_{13}^{ij}$  также не зависят от третьего индекса  $(k)$ , но они используются для вычисления компонент  $z_3^{ijk}$ , поэтому вычисляются раньше.

Наконец, энергия системы атомов вычисляется по формуле

$$E = E(u, x) = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I z_{17}^{ij}, \quad (1.10)$$

где фазовые переменные  $z_1, z_2, \dots, z_{17}$  определяются указанным выше многошаговым процессом.

Для получения аналитических выражений производных функции  $E$  по координатам атомов воспользуемся структурой функции  $E$ , которая представлена формулой (1.10). Для всех фиксированных пар индексов  $(l, m)$ , таких что  $l = 1, 2, 3$  и  $m = \overline{1, I}$ , введем следующие обозначения:

$\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$ , где

$$\tilde{z}_s = \left\{ \tilde{z}_s^{ijk} : \tilde{z}_s^{ijk} = \frac{\partial z_s^{ijk}}{\partial x_{lm}} \right\}, \quad s = \overline{1, 8}, \quad \tilde{z}_s = \left\{ \tilde{z}_s^{ij} : \tilde{z}_s^{ij} = \frac{\partial z_s^{ij}}{\partial x_{lm}} \right\}, \quad s = \overline{9, 17}.$$

Указанные выше величины вычисляются по следующим формулам:

— для всех  $l = 1, 2, 3$  и  $i, j, k, m = \overline{1, I}$

$$\begin{aligned} \tilde{z}_1^{ijk} = \frac{\partial z_1^{ijk}}{\partial x_{lm}} &= \begin{cases} \frac{x_{lm} - x_{lk}}{z_1^{mjk}}, & m = i, \quad k \neq m, \\ \frac{x_{lm} - x_{li}}{z_1^{ijm}}, & m = k, \quad i \neq m, \\ 0 & \text{в других случаях,} \end{cases} \\ \tilde{z}_2^{ijk} = \frac{\partial z_2^{ijk}}{\partial x_{lm}} &= \begin{cases} \frac{x_{lm} - x_{lk}}{z_2^{imk}}, & m = j, \quad k \neq m, \\ \frac{x_{lm} - x_{lj}}{z_2^{ijm}}, & m = k, \quad j \neq m, \\ 0 & \text{в других случаях,} \end{cases} \\ \tilde{z}_3^{ijk} &= \frac{z_1^{ijk} (z_{13}^{ij})^2 \tilde{z}_{13}^{ij} + z_{13}^{ij} (z_1^{ijk})^2 \tilde{z}_1^{ijk} - 2z_1^{ijk} z_{13}^{ij} z_2^{ijk} \tilde{z}_2^{ijk} +}{2(z_{13}^{ij})^2 (z_1^{ijk})^2} + \\ &+ \frac{-(z_1^{ijk})^3 \tilde{z}_{13}^{ij} + z_1^{ijk} (z_2^{ijk})^2 \tilde{z}_{13}^{ij} - (z_{13}^{ij})^3 \tilde{z}_1^{ijk} + z_{13}^{ij} (z_2^{ijk})^2 \tilde{z}_1^{ijk}}{2(z_{13}^{ij})^2 (z_1^{ijk})^2}; \\ \tilde{z}_4^{ijk} &= \begin{cases} 0 & \text{в других случаях,} \\ -\pi \cos \left( \frac{\pi(z_1^{ijk} - R)}{2R_{\text{cut}}} \right) \frac{1}{4R_{\text{cut}}}, & R - R_{\text{cut}} \leq z_1^{ijk} \leq R + R_{\text{cut}}, \end{cases} \end{aligned}$$

$$\tilde{z}_5^{ijk} = \frac{-2(u_8)^2 (u_{10} - z_3^{ijk}) \tilde{z}_3^{ijk}}{\left((u_9)^2 + (u_{10} - z_3^{ijk})^2\right)^2}; \quad \tilde{z}_6^{ijk} = 3(z_{13}^{ij} - z_1^{ijk})^2 (\tilde{z}_{13}^{ij} - \tilde{z}_1^{ijk});$$

$$\tilde{z}_7^{ijk} = (u_7)^3 \exp\left((u_7)^3 z_6^{ijk}\right) \tilde{z}_6^{ijk}; \quad \tilde{z}_8^{ijk} = \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk} + \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk} + \tilde{z}_4^{ijk} \tilde{z}_5^{ijk} \tilde{z}_7^{ijk};$$

$$\tilde{z}_9^{ij} = \sum_{\substack{k=1 \\ k \neq i, j}}^I \tilde{z}_8^{ijk}; \quad \tilde{z}_{10}^{ij} = \tilde{z}_9^{ij} u_6; \quad \tilde{z}_{11}^{ij} = \tilde{z}_{10}^{ij} u_5 (z_{10}^{ij})^{u_5-1}; \quad \tilde{z}_{12}^{ij} = -\frac{1}{2u_5} \tilde{z}_{11}^{ij} (1 + z_{11}^{ij})^{-\frac{1}{2u_5}-1};$$

— для всех  $l = 1, 2, 3$  и  $i, j, m = \overline{1, I}$

$$\tilde{z}_{13}^{ij} = \frac{\partial z_{13}^{ij}}{\partial x_{lm}} = \begin{cases} \frac{x_{lm} - x_{lj}}{z_{13}^{mj}}, & m = i, \quad j \neq m, \\ \frac{x_{lm} - x_{li}}{z_{13}^{im}}, & m = j, \quad i \neq m; \\ \tilde{z}_{14}^{ij} = -u_3 \sqrt{2u_4 z_{14}^{ij} \tilde{z}_{13}^{ij}}, & 0 \text{ в других случаях,} \end{cases}$$

$$\tilde{z}_{15}^{ij} = -u_3 \sqrt{2/u_4} z_{15}^{ij} \tilde{z}_{13}^{ij};$$

$$\tilde{z}_{16}^{ij} = \begin{cases} \frac{-\pi \cos\left(\frac{\pi(z_{13}^{ij} - R)}{2R_{\text{cut}}}\right)}{4R_{\text{cut}}}, & R - R_{\text{cut}} \leq z_{13}^{ij} \leq R + R_{\text{cut}}, \\ 0 & \text{в других случаях.} \end{cases}$$

$$\tilde{z}_{17}^{ij} = \tilde{z}_{16}^{ij} z_{14}^{ij} - \tilde{z}_{16}^{ij} z_{12}^{ij} z_{15}^{ij} + z_{16}^{ij} \tilde{z}_{14}^{ij} - z_{16}^{ij} \tilde{z}_{12}^{ij} z_{15}^{ij} - z_{16}^{ij} z_{12}^{ij} \tilde{z}_{15}^{ij}.$$

Порядок вычисления этих производных такой же, как и для фазовых переменных  $z_1, z_2, \dots, z_{17}$ , а формулы для вычисления частных производных функции энергии по координатам атомов имеют вид

$$\frac{\partial E(u, x)}{\partial x_{lm}} = \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I \tilde{z}_{17}^{ij}, \quad l = 1, 2, 3 \text{ и } m = \overline{1, I}.$$

Проведена большая серия вычислительных экспериментов, в которых менялись размерность материала (рассматривались трехмерная и двумерная модели) и количество атомов в фрагменте материала. Если рассматривается трехмерная модель ( $l = 1, 2, 3$ ), то для каждого  $m$ -го атома вычисления по многошаговому процессу  $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$  необходимо проводить 3 раза, а если рассматривается двумерная модель, то 2 раза. Все численные эксперименты показали, что в первом случае (трехмерная модель) градиент функции энергии по координатам атома, определенный с помощью формул БАД-методологии (см. [12]), вычисляется в 3 раза быстрее, чем с помощью многошагового процесса  $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{17}$ . Когда используется двумерная модель материала, эффективность формул БАД-методологии возрастает в 2 раза.

Возникает естественный вопрос: не проще ли при вычислении производных этой сложной функции воспользоваться приближенным численным дифференцированием? Ниже показывается, с какими трудностями приходится сталкиваться при использовании такого метода.

В [13] рассмотрена задача идентификации параметров потенциала Терсоффа применительно к однокомпонентному кристаллу кремния. Для определения начальных приближений в указанной работе допустимое множество  $X = [u, \bar{u}] = \{u \in R^{10} : \underline{u}_i \leq x_i \leq \bar{u}_i\}$  выбиралось таким образом, чтобы заведомо содержались все допустимые значения параметров, а именно:

$$\underline{u} = (0.5; 0.5; 0.5; 0.5; 0.1; 5 \times 10^{-8}; 0.5; 10000; 1; -2);$$

$$\bar{u} = (10; 5; 5; 5; 2; 3 \times 10^{-6}; 3; 200000; 30; -0.1).$$

**Таблица 1.** ( $i = 1, u_1 = 1$ )

| $\Delta$              | $10^{+0}$             | $10^{-2}$             | $10^{-4}$             | $10^{-6}$            | $10^{-8}$            | $10^{-10}$           | $10^{-12}$           | $10^{-13}$           |
|-----------------------|-----------------------|-----------------------|-----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Абсолютное отклонение | $1.0 \times 10^{-14}$ | $1.0 \times 10^{-14}$ | $5.1 \times 10^{-11}$ | $4.6 \times 10^{-9}$ | $1.3 \times 10^{-7}$ | $8.9 \times 10^{-6}$ | $9.7 \times 10^{-4}$ | $4.4 \times 10^{-2}$ |

**Таблица 2.** ( $i = 3, u_3 = 1$ )

| $\Delta$              | $10^{+0}$         | $10^{-2}$            | $10^{-4}$            | $10^{-6}$            | $10^{-7}$            | $10^{-8}$            | $10^{-10}$           | $10^{-12}$           |
|-----------------------|-------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Абсолютное отклонение | $3.1 \times 10^0$ | $3.5 \times 10^{-2}$ | $3.5 \times 10^{-4}$ | $3.5 \times 10^{-6}$ | $3.7 \times 10^{-7}$ | $5.4 \times 10^{-7}$ | $4.7 \times 10^{-5}$ | $1.7 \times 10^{-3}$ |

**Таблица 3.** ( $i = 6, u_6 = 10^{-7}$ )

| $\Delta$              | $10^{-2}$            | $10^{-4}$            | $10^{-6}$            | $10^{-8}$            | $10^{-10}$           | $10^{-11}$           | $10^{-12}$           | $10^{-13}$           |
|-----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Абсолютное отклонение | $1.3 \times 10^{+6}$ | $1.2 \times 10^{+6}$ | $3.0 \times 10^{+5}$ | $4.1 \times 10^{+3}$ | $4.0 \times 10^{+1}$ | $3.5 \times 10^{+0}$ | $7.5 \times 10^{+1}$ | $7.5 \times 10^{+2}$ |

В табл. 1–3 представлены результаты сравнения вычисления градиента потенциала Терсоффа на основе формул БАД-методологии и вычисления этого градиента с помощью метода конечных разностей для трех компонент градиента (по параметрам  $u_1 = D_e$ ,  $u_3 = \beta$ ,  $u_6 = \gamma$ ). Все входные параметры задачи выбирались из указанного выше диапазона допустимых значений. Величина  $\Delta$  в табл. 1–3 указывает на значения приращения параметра  $u_i$ , которые были использованы при вычислении градиента с помощью метода конечных разностей.

В табл. 1 сравниваются значения первой компоненты градиента энергии по параметру  $u_1 = D_e$  при  $D_e = 1$ . Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно  $\frac{\partial E(u)}{\partial u_1} = -7.079590296064$ .

В табл. 2 сравниваются значения третьей компоненты градиента энергии по параметру  $u_3 = \beta$  при  $\beta = 1$ . Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно  $\frac{\partial E(u)}{\partial u_3} = 8.117021018694$ .

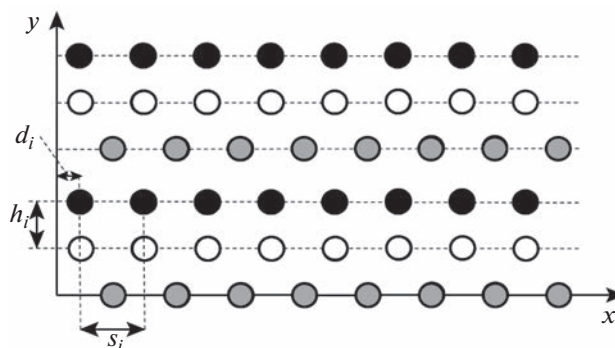
В табл. 3 сравниваются значения шестой компоненты градиента энергии по параметру  $u_6 = \gamma$  при  $\gamma = 1.0 \times 10^{-7}$ . Значение этой компоненты градиента, вычисленное с помощью БАД-методологии (точное значение), равно  $\frac{\partial E(u)}{\partial u_6} = 1.288320939087 \times 10^{+6}$ .

Как следует из табл. 1, при вычислении первой компоненты градиента наиболее подходящим является выбор  $\Delta$  из интервала  $[10^{-8}; 1.0]$ , что составляет от  $10^{-7}$  до 100% от начального приближения  $u_1 = 1.0$ .

Из табл. 2 следует, что при вычислении третьей компоненты градиента наиболее подходящим является выбор  $\Delta \approx 10^{-7}$ . Это составляет примерно  $10^{-5}\%$  от начального приближения  $u_3 = 1.0$ .

Из табл. 3 следует, что при вычислении шестой компоненты градиента наиболее подходящим является выбор  $\Delta \approx 10^{-11}$ . Это составляет примерно  $10^{-2}\%$  от начального приближения  $u_6 = 10^{-7}$ .

На основании сказанного можно сделать вывод о том, что вычисление градиента энергии системы атомов с помощью метода конечных разностей связано с определенными трудностями. В случае использования этого метода необходимо проводить исследования, связанные с выбором приращения каждого параметра потенциала, причем при каждом его новом значении. В противоположность этому, алгоритм, использующий формулы БАД, позволяет автоматически получить с машинной точностью значение этого градиента.



Фиг. 1. Двумерная модель материала.

## 2. СРАВНЕНИЕ МЕТОДОВ ОПРЕДЕЛЕНИЯ ПРОИЗВОДНЫХ ДЛЯ ПЛОСКОЙ МОДЕЛИ МНОГОСЛОЙНОГО КУСОЧНО-ОДНОРОДНОГО МАТЕРИАЛА

В качестве вычислительного эксперимента для сравнения методов определения производных с помощью указанных выше способов и с помощью стандартных пакетов программ рассмотрена предложенная в [14] плоская модель многослойного кусочно-однородного материала. Материал представляется в виде периодической кусочно-однородной многослойной структуры, в рамках которой могут отличаться типы атомов в различных слоях. Модель накладывает ряд ограничений на структуру слоев:

- каждый слой состоит из одинаковых атомов, в разных слоях могут находиться различные атомы;
- расстояния между соседними атомами в одном слое одинаковы, но в разных слоях они могут отличаться;
- в рассматриваемой системе выделяется группа из  $K$  параллельных слоев, которая периодически повторяется в направлении оси  $y$  (см. фиг. 1);
- число атомов в каждом слое и общее число слоев считаются потенциально неограниченными.

На фиг. 1 представлен пример модели, в которой повторяется группа из трех слоев. Каждый слой состоит из атомов своего вида.

В рамках данной модели положение атомов определяется следующими параметрами:

$h_i$ ,  $i = 1, \dots, K$ , — расстояние между слоем с номером  $i$  и предыдущим слоем,

$d_i$ ,  $i = 1, \dots, K$ , — смещение первого атома в слое  $i$  с положительной абсциссой относительно нулевой отметки,

$s_i$ ,  $i = 1, \dots, K$ , — расстояние между атомами в слое  $i$ .

Совокупность значений перечисленных параметров будем называть *конфигурацией*. Ниже рассматривается ситуация, когда материал находится в ненагруженном состоянии. Требуется определить конфигурацию, соответствующую минимальной энергии взаимодействия атомов, входящих в моделируемый фрагмент материала.

Оптимизационная задача состоит в минимизации энергии  $E(u, w)$  группы атомов (потенциал взаимодействия атомов — потенциал Терсоффа), расположенных на  $K$  соседних слоях. Параметрами оптимизационной задачи являются переменные  $w = (h_1, d_1, s_1, \dots, h_K, d_K, s_K)$ , составляющие вектор размерности  $r = 3K$ .

Пусть (для наглядности)  $x_i = x_{1i}$  — первая координата  $i$ -го атома двумерной структуры,  $y_i = x_{2i}$  — его вторая координата. Если  $i$ -й атом является атомом с порядковым номером  $j$  на  $k$ -м слое, то

$$x_i = d_k + (j-1)s_k; \quad y_i = \begin{cases} 0, & k = 1, \\ \sum_{m=2}^k h_m, & k = 2, \dots, K. \end{cases}$$

Соответствующие производные вычисляются по формулам (см. разд. 1)

$$\frac{\partial E}{\partial h_k} = \sum_{i=1}^I \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial h_k}, \quad \frac{\partial E}{\partial d_k} = \sum_{i=1}^I \frac{\partial E}{\partial x_i} \frac{\partial x_i}{\partial d_k}, \quad \frac{\partial E}{\partial s_k} = \sum_{i=1}^I \frac{\partial E}{\partial x_i} \frac{\partial x_i}{\partial s_k},$$

где  $\frac{\partial E}{\partial x_i}$  и  $\frac{\partial E}{\partial y_i}$  – производные функции энергии по первой и второй координатам атома соответственно.

Расчет производных энергии  $E(u, w)$  по параметрам  $w = (h_1, d_1, s_1, \dots, h_K, d_K, s_K)$  осуществлялся как с использованием аналитически выведенных формул, так и с помощью имеющихся пакетов программ.

Работа выполнялась с использованием инфраструктуры Центра коллективного пользования “Высокопроизводительные вычисления и большие данные” (ЦКП “Информатика”) ФИЦ ИУ РАН (г. Москва). Испытательный стенд состоял из следующих средств вычислительной техники:

- Intel Core i7 4770 3.4 Гц, 8 Гб RAM – 4-х ядерный процессор архитектуры Haswell с поддержкой технологий Hyper-Threading (два вычислительных протока на ядро, всего восемь логических ядер), AVX2 – Advanced Vector Extensions (в контексте решаемой задачи – обработка данных в формате с плавающей запятой в группах длиной 256 бит, т.е. для чисел с двойной точностью одновременно обрабатывается вектор из четырех чисел), и FMA – Fused Multiply-Add – выполнение совмещенной операции умножения-сложения вида  $a = a + b * c$ . Использовался компилятор GNUC++ 6.4.0.

- 2x Intel Xeon E5-2683 V4 2.1 Гц, 512 Гб RAM – двухпроцессорный сервер, каждый процессор содержит 16 ядер с поддержкой Hyper-Threading, AVX2 и FMA (всего 64 логических ядра). Использовался компилятор GNUC++ 7.2.0.

В работе использовались хорошо зарекомендовавшие себя ранее (см. [15]) пакеты CoDiPack 1.6.0 (см. [10]) и Adept 1.1 (см. [11]).

Сравнивались следующие методы вычисления градиента сложной функции  $E(u, w)$ :

1) использование выведенных аналитически выражений для вычисления частных производных этой функции (метод **Hand1**),

2) использование выведенных аналитически формул БАД-методологии для вычисления частных производных этой функции (метод **Hand2**),

3) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForwardVecGen (метод **CoDiPack\_kr**),

4) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForward (метод **CoDiPack\_rk**),

5) использование прямого метода автоматического дифференцирования, реализованного в пакете CoDiPack с типом данных RealForward, распараллеленный с помощью директивы OpenMP – #pragma omp for (метод **CoDiPack\_rk**),

6) использование обратного метода автоматического дифференцирования, реализованного в пакете Adept ([11]) (метод **Adept**).

Для методов **Hand1**, **CoDiPack\_kr**, **CoDiPack\_rk** и **CoDiPack\_rk\_omp** использовались опции оптимизации компилятора O2 и Ofast. Для методов **Hand2** и **Adept** только опция O2.

Расчеты проводились для фиксированного набора специфических параметров потенциала Терсоффа, но для различного количества слоев и атомов:

- 4 слоя и 94 атома, размерность  $r$  вектора  $w$  равна 12 (распределение атомов по слоям 28, 19, 28, 19),

- 8 слоев и 188 атомов, размерность  $r$  вектора  $w$  равна 24 (распределение атомов по слоям 28, 19, 19, 28, 28, 19, 19, 28),

- 12 слоев и 204 атома, размерность  $r$  вектора  $w$  равна 36 (распределение атомов по слоям 15, 19, 19, 15, 15, 19, 19, 15, 15, 19, 19, 15).

При расчете градиента пакетом CoDiPack применялись типы данных RealForward и RealForwardVecGen. Тип данных RealForward состоит из двух чисел – значения переменной и ее производной по выбранной переменной. Для расчета градиента функции необходимо запустить вычисление функции и производной в цикле  $r$  раз. Тип данных RealForwardVecGen состоит из  $r + 1$  чисел – значения переменной и вектора ее частных производных. При расчете градиента цикл

**Таблица 4.** Время (секунды) вычисления функции и градиента (одновременно с функцией) для прямых методов

| Компьютер                    | IntelCore i7 |        |        | IntelXeon |        |        |
|------------------------------|--------------|--------|--------|-----------|--------|--------|
| Слой*атомы                   | 4*94         | 8*188  | 12*204 | 4*94      | 8*188  | 12*204 |
| Функция O2                   | 0.25         | 0.94   | 1.17   | 0.25      | 0.98   | 1.23   |
| Функция Ofast                | 0.23         | 0.92   | 1.16   | 0.29      | 1.14   | 1.43   |
| <b>Hand1</b> O2              | 25.45        | 256.58 | 360.98 | 20.91     | 138.09 | 186.02 |
| <b>Hand1</b> Ofast           | 25.89        | 218.89 | 325.18 | 22.96     | 154.43 | 210.42 |
| <b>CoDiPack_kr</b> O2        | 7.23         | 62.85  | 122.88 | 12.19     | 99.66  | 189.23 |
| <b>CoDiPack_kr</b> Ofast     | 1.66         | 17.61  | 72.91  | 2.68      | 23.14  | 134.45 |
| <b>CoDiPack_rk</b> O2        | 7.75         | 55.71  | 99.62  | 8.76      | 66.04  | 125.05 |
| <b>CoDiPack_rk</b> Ofast     | 5.55         | 43.05  | 82.95  | 5.88      | 37.92  | 71.77  |
| <b>CoDiPack_rk_omp</b> O2    | 1.91         | 11.30  | 22.89  | 1.21      | 6.93   | 6.65   |
| <b>CoDiPack_rk_omp</b> Ofast | 1.64         | 10.00  | 20.55  | 0.72      | 5.10   | 4.75   |

**Таблица 5.** Время (секунды) вычисления функции и градиента (одновременно с функцией) для обратных методов

| Компьютер     | Intel Core i7 |       |        | Intel Xeon |       |        |
|---------------|---------------|-------|--------|------------|-------|--------|
| Слой*атомы    | 4*94          | 8*188 | 12*204 | 4*94       | 8*188 | 12*204 |
| <b>Hand 2</b> | 2.66          | 23.03 | 29.61  | 3.95       | 54.28 | 68.44  |
| <b>Adept</b>  | 1.25          | 5.88  | 8.03   | 2.43       | 10.51 | 13.60  |

по  $r$  становится внутренним, но функция рассчитывается один раз. Использование типа данных RealForwardVecGen позволяет задействовать технологию AVX2 и ускоряет вычисления от 1.5 до 4 раз без каких-либо трудозатрат; достаточно указать компилятору опцию оптимизации —Ofast.

В табл. 4 приводится сравнение прямых методов дифференцирования **Hand1**, **CoDiPack\_kr**, **CoDiPack\_rk** в однопоточном исполнении и **CoDiPack\_rk\_omp**, распараллеленный с помощью директивы OpenMP — #pragma omp for. Как видно из таблицы, время вычисления функции имеет примерно квадратичную зависимость от количества атомов.

Время вычисления производной методом **Hand1** изменяется на 10–17% в зависимости от установленных опций оптимизации, при этом для IntelCore i7 опция Ofast ускоряет расчет, а для IntelXeon наоборот замедляет.

Метод **CoDiPack\_kr** при опции O2 оказывается в 3–4 раза быстрее метода **Hand1** с той же опцией. Связано это с тем, что при реализации методов **Hand1** и **Hand2** задача оптимизации кода не ставилась, а пакетные методы оптимизировались. Сравнивая влияние опций O2 и Ofast, мы видим существенное ускорение решения задачи при  $r = 12$  и несколько уменьшающееся ускорение для больших размерностей. Для малой размерности задачи преимущества технологии AVX2 раскрываются полностью.

Метод **CoDiPack\_rk**, как можно увидеть, не так сильно чувствителен к опциям оптимизации компилятора. Вычислительно он более сложен, так как значение функции рассчитывается  $r$  раз. Но его применение целесообразно при распараллеливании процесса расчета градиента на несколько ядер. Эксперименты с методом **CoDiPack\_rk\_omp** это показывают.

В табл. 5 приведено время вычисления градиента и функции обратными методами **Hand2** и **Adept**. Как видно из таблицы, время вычисления функции и градиента мало зависит от размерности задачи. Исключение составляет **Hand2** для задачи с  $r = 12$  (4\*94). Связано это с небольшим количеством памяти, при котором вычисления выполняются в основном с использованием кэш-шей процессора. Как и в случае с прямыми методами оптимизация кода **Hand2** не проводилась, с чем и связано большее время вычислений. Также необходимо обратить внимание на тот факт, что вычисления прямым методом **CoDiPack\_rk\_omp** на компьютере IntelXeon оказываются более эффективными, чем вычисления обратным методом **Adept**, за счет параллельного исполнения.

## 3. ЗАКЛЮЧЕНИЕ

Проведенные исследования, некоторые результаты которых представлены в настоящей работе, показывают, что при вычислении градиента сложной функции наиболее эффективным оказалось использование методологии быстрого автоматического дифференцирования. При этом обе версии БАД-методологии (и “ручная”, и “пакетная”) показали свое значительное преимущество перед другими методами (при отсутствии распараллеливания). С точки зрения затрат машинного времени обе версии БАД-методологии примерно равноценны (в случае оптимизации “ручной” версии), опережая одна другую на отдельных задачах.

Преимущество “пакетной” версии БАД-методологии в том, что для ее работы необходимо задать алгоритм вычисления только самой функции. Все остальные этапы (построение алгоритма решения сопряженной задачи, построение формул для вычисления градиента) выполняются пакетом автоматически.

Что касается “ручной” версии БАД-методологии, то в этом случае необходимо вручную построить многошаговый процесс вычисления функции, с помощью канонических формул БАД-методологии построить многошаговый процесс решения сопряженной задачи и вывести формулы для вычисления градиента функции. Безусловно, все это требует серьезных интеллектуальных затрат исследователя. Однако при этом у исследователя появляется возможность учесть специфику конкретной задачи и сделать алгоритм вычисления градиента функции более эффективным. К преимуществам “ручной” версии БАД-методологии следует отнести и то, что эта версия гораздо слабее, чем “пакетная” версия, зависит от параметров используемого компьютера.

## СПИСОК ЛИТЕРАТУРЫ

1. Крылов И.А., Черноусько Ф.Л. О методе последовательных приближений для решения задач оптимального управления // Ж. вычисл. матем. и матем. физ. 1962. Т. 2. № 6. С. 669–683.
2. Федоренко Р.П. Приближенное решение задач оптимального управления. М.: Наука, 1978.
3. Моисеев Н.Н. Численные методы в теории оптимальных систем. М.: Наука, 1971.
4. Албу А.Ф., Зубов В.И. Вычисление градиента функционала в одной задаче оптимального управления, связанной с кристаллизацией металла // Ж. вычисл. матем. и матем. физ. 2009. Т. 9. № 1. С. 51–75.
5. Евтушенко Ю.Г., Засухина Е.С., Зубов В.И. О численном подходе к оптимизации решения задачи Бюргерса с помощью граничных условий // Ж. вычисл. матем. и матем. физ. 1997. Т. 37. № 12. С. 1449–1458.
6. Айда-Заде К.Р., Евтушенко Ю.Г. Быстрое автоматическое дифференцирование // Матем. моделирование. 1989. Т. 1. С. 121–139.
7. Evtushenko Y.G. Computation of exact gradients in distributed dynamic systems // Optimizat. Meth. and Software. 1998. V. 9. P. 45–75.
8. Евтушенко Ю.Г. Оптимизация и быстрое автоматическое дифференцирование. Научное издание. ВЦ им. А.А. Дородницына РАН. Москва. 2013. 144 с.
9. Hascoet L., Pascual V. The Tapenade automatic differentiation tool: principles, model, and specification // ACM Transact. on Math. Software (TOMS). 2013. T. 39. № 3. С. 1–43.
10. Albring T. et al. An aerodynamic design framework based on algorithmic differentiation // ERCOFTAC Bull. 2015. V. 102. P. 10–16.
11. Hogan R.J. Fast reverse-mode automatic differentiation using expression templates in C++ // ACM Transact. on Math. Software (TOMS). 2014. V. 40. № 4. P. 26–42.
12. Албу А.Ф. Применение быстрого автоматического дифференцирования для вычисления градиента потенциала Терсоффа // Информ. технологии и вычисл. системы. 2016. Т. № 1. С. 43–49.
13. Абгарян К.К., Посыпкин М.А. Программный комплекс для решения задач параметрической идентификации потенциалов межатомного взаимодействия // Inter. J. of Open Inform. Tech. 2014. Т. 2. № 10. С. 14–19.
14. Евтушенко Ю.Г., Лурье С.А., Посыпкин М.А., Соляев Ю.О. Применение методов оптимизации для поиска равновесных состояний двумерных кристаллов // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 12. С. 2032–2041.
15. Горчаков А.Ю. О программных пакетах быстрого автоматического дифференцирования // Информ. технологии и вычисл. системы. 2018. № 1. P. 30–36.

## ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 519.7

# ОПТИМАЛЬНОЕ ГРАНИЧНОЕ УПРАВЛЕНИЕ РАСПРЕДЕЛЕННОЙ НЕОДНОРОДНОЙ КОЛЕБАТЕЛЬНОЙ СИСТЕМОЙ С ЗАДААННЫМИ СОСТОЯНИЯМИ В ПРОМЕЖУТОЧНЫЕ МОМЕНТЫ ВРЕМЕНИ

© 2023 г. В. Р. Барсегян<sup>1, 2, \*</sup>

<sup>1</sup> 0019 Ереван, пр-т Маршала Баграмяна, 24б, Институт механики НАН Армении, Армения

<sup>2</sup> 0025 Ереван, ул. Алека Манукяна, 1, Ереванский государственный университет, Армения

\*e-mail: barseghyan@sci.am

Поступила в редакцию 25.03.2022 г.  
Переработанный вариант 16.05.2022 г.  
Принята к публикации 07.07.2022 г.

Рассмотрена задача оптимального граничного управления распределенной неоднородной колебательной системой, описываемой одномерным волновым уравнением с кусочно-постоянными характеристиками. Предположено, что время прохождения волны через каждый однородный участок одинаково. Управление осуществляется смещением на двух концах. Критерий качества задан на всем промежутке времени. Предложен конструктивный подход построения оптимального управляющего воздействия, переводящего колебания за заданный промежуток времени из начального состояния через многоточечные промежуточные состояния в конечное состояние. Полученные результаты иллюстрируются на конкретном примере. Библ. 20.

**Ключевые слова:** оптимальное управление колебаниями, оптимальное граничное управление, неоднородный колебательный процесс, волновое уравнение, кусочно-постоянные характеристики, разделение переменных

**DOI:** 10.31857/S0044466922120031, **EDN:** LEONJD

## ВВЕДЕНИЕ

К исследованию задач управления и оптимального управления распределенными колебательными системами с заданными начальными и конечными условиями, а также с многоточечными промежуточными условиями посвящены многие работы, в частности, [1]–[15]. Задачам исследования и управления разнородными распределенными составными системами посвящены, в частности, работы [8]–[18]. Одна из первых задач управления распределенной колебательной системой, состоящей из двух кусочно-однородных сред, была поставлена А.Г. Бутковским и исследована в [8]. В [9], [10] (и других работах этого же автора и его учеников) изучены задачи граничного управления колебаниями стержня, состоящего из разнородных участков. При исследовании этих задач использован метод Даламбера.

В настоящей работе рассмотрена задача оптимального граничного управления некоторой распределенной неоднородной колебательной системой с заданными состояниями в промежуточные моменты времени. Предполагается, что колебательный процесс состоит из двух участков с разными упругими свойствами и плотностями, а их длины таковы, что время прохождения волны по каждому из участков одинаково.

Цель данной статьи состоит в разработке аналитического подхода построения функции оптимального граничного управления одномерными колебательными неоднородными процессами со смещением на двух концах, переводящего колебания за определенный промежуток времени из заданного начального состояния через многоточечные промежуточные состояния в заданное конечное состояние с критерием качества, заданным на всем промежутке времени. Построение осуществляется по следующей схеме. Задача сводится к задаче оптимального управления распределенными воздействиями с нулевыми граничными условиями, далее используются метод разделения переменных и методы теории оптимального управления конечномерными системами с многоточечными промежуточными условиями.

## 1. ПОСТАНОВКА ЗАДАЧИ

Рассматриваются колебания распределенной кусочно-однородной среды, расположенной вдоль отрезка  $-l_1 \leq x \leq l$  и состоящей из двух участков:  $-l_1 \leq x \leq 0$  и  $0 \leq x \leq l$ . Скорость прохождения по участкам волны обозначим через  $a_i = \sqrt{k_i/\rho_i}$ , где  $\rho_i = \text{const}$  — плотность,  $k_i = \text{const}$  — модуль Юнга,  $i = 1, 2$ . Предполагается, что длины  $l_1$  и  $l$  участков выбраны так, что время прохождения волны по участку  $-l_1 \leq x \leq 0$  совпадает со временем прохождения волны по участку  $0 \leq x \leq l$ , т.е.

$$\frac{l_1}{a_1} = \frac{l}{a_2}. \quad (1.1)$$

Отметим, что таким колебательным неоднородным процессом могут быть продольные колебания кусочно-однородного стержня ( $\rho$  — плотность,  $k$  — модуль упругости) или поперечные колебания кусочно-однородной струны ( $\rho$  — плотность,  $k$  — натяжение струны).

Пусть состояние (продольные колебания) стержня (или поперечные колебания струны) представлено функцией  $Q(x, t)$ ,  $-l_1 \leq x \leq l$ ,  $0 \leq t \leq T$ . Отклонение от состояния равновесия описывается уравнением

$$\frac{\partial^2 Q(x, t)}{\partial t^2} = \begin{cases} a_1^2 \frac{\partial^2 Q(x, t)}{\partial x^2}, & -l_1 \leq x \leq 0, \quad 0 \leq t \leq T, \\ a_2^2 \frac{\partial^2 Q(x, t)}{\partial x^2}, & 0 \leq x \leq l, \quad 0 \leq t \leq T, \end{cases} \quad (1.2)$$

с граничными условиями

$$Q(-l_1, t) = \mu(t), \quad Q(l, t) = \nu(t), \quad 0 \leq t \leq T, \quad (1.3)$$

и с условиями сопряжения в точке  $x = 0$  соединения участков

$$Q(0-0, t) = Q(0+0, t), \quad a_1^2 \rho_1 \frac{\partial Q(x, t)}{\partial x} \Big|_{x=0-0} = a_2^2 \rho_2 \frac{\partial Q(x, t)}{\partial x} \Big|_{x=0+0}. \quad (1.4)$$

Пусть заданы начальные (при  $t = t_0 = 0$ ) и конечные (при  $t = T$ ) условия:

$$Q(x, 0) = \varphi_0(x), \quad \frac{\partial Q(x, t)}{\partial t} \Big|_{t=0} = \psi_0(x), \quad -l_1 \leq x \leq l, \quad (1.5)$$

$$Q(x, T) = \varphi_T(x), \quad \frac{\partial Q}{\partial t} \Big|_{t=T} = \psi_T(x), \quad -l_1 \leq x \leq l. \quad (1.6)$$

Пусть также в некоторые промежуточные моменты времени  $t_k$  ( $k = 1, \dots, m$ ):

$$0 = t_0 < t_1 < \dots < t_m < t_{m+1} = T,$$

заданы значения функции состояния колебания в виде

$$Q(x, t_i) = \varphi_i(x), \quad -l_1 \leq x \leq l, \quad i = 1, \dots, m. \quad (1.7)$$

В формуле (1.3) функции  $\mu(t)$  и  $\nu(t)$  — управляющие воздействия (граничные управления).

Предполагается, что функция  $Q(x, t) \in C^2(\Omega_T)$ , где  $\Omega_T = \{(x, t): x \in [-l_1, l], t \in [0, T]\}$ , а функции  $\varphi_i(x) \in C^2[-l_1, l]$  ( $i = 0, 1, \dots, m, m+1$ ),  $\psi_0(x), \psi_T(x) \in C^1[-l_1, l]$ .

Предполагается также, что все функции такие, что выполняются условия согласования

$$\begin{aligned} \mu(0) &= \varphi_0(-l_1), \quad \dot{\mu}(0) = \psi_0(-l_1), \quad \nu(0) = \varphi_0(l), \quad \dot{\nu}(0) = \psi_0(l), \\ \mu(t_i) &= \varphi_i(-l_1), \quad \nu(t_i) = \varphi_i(l), \quad i = 1, \dots, m, \\ \mu(T) &= \varphi_T(-l_1), \quad \dot{\mu}(T) = \psi_T(-l_1), \quad \nu(T) = \varphi_T(l), \quad \dot{\nu}(T) = \psi_T(l). \end{aligned} \quad (1.8)$$

Сформулируем следующую задачу оптимального граничного управления колебаниями системы (1.2) с заданными значениями в промежуточные моменты времени.

Требуется найти такие оптимальные управления  $\mu^0(t)$  и  $v^0(t)$ ,  $0 \leq t \leq T$ , (1.3), под воздействием которых колебательное движение системы (1.2) из заданного начального состояния (1.5) переходит в конечное состояние (1.6), обеспечивая выполнение условий (1.7) и минимизирующие функционал

$$\left[ \int_0^T (\mu^2(t) + v^2(t)) dt \right]^{1/2}. \quad (1.9)$$

Отметим, что так как в отдельные промежуточные моменты времени  $t_k$  ( $k = 1, \dots, m$ ) заданы только значения функции состояния (1.7), то использовать подход поэтапного решения задачи оптимального управления нецелесообразно. Поэтому в работе предлагается такой подход решения рассмотренной задачи оптимального управления, в котором учитывается специфика промежуточных условий (1.7).

## 2. СВЕДЕНИЕ ЗАДАЧИ К ЗАДАЧЕ С НУЛЕВЫМИ ГРАНИЧНЫМИ УСЛОВИЯМИ

Для решения поставленной задачи перейдем к новой переменной (см. [18])

$$\xi = \begin{cases} \frac{a_2}{a_1} x, & -l_1 \leq x \leq 0, \\ x, & 0 \leq x \leq l, \end{cases} \quad (2.1)$$

что приводит к растяжению или сжатию отрезка  $-l_1 \leq x \leq 0$  относительно точки  $x = 0$ . При этом с учетом (1.1) будем иметь, что отрезок  $-l_1 \leq x \leq 0$  переходит к отрезку  $-l \leq \xi \leq 0$ . Для функции  $Q(\xi, t)$  получим на отрезках одинаковой длины одинаковое уравнение

$$\frac{\partial^2 Q(\xi, t)}{\partial t^2} = \begin{cases} a_2^2 \frac{\partial^2 Q(\xi, t)}{\partial \xi^2}, & -l \leq \xi \leq 0, \quad 0 \leq t \leq T, \\ a_1^2 \frac{\partial^2 Q(\xi, t)}{\partial \xi^2}, & 0 \leq \xi \leq l, \quad 0 \leq t \leq T, \end{cases}$$

или

$$\frac{\partial^2 Q(\xi, t)}{\partial t^2} = a^2 \frac{\partial^2 Q(\xi, t)}{\partial \xi^2}, \quad -l \leq \xi \leq l, \quad 0 \leq t \leq T, \quad (2.2)$$

с соответствующими граничными условиями

$$Q(-l, t) = \mu(t), \quad Q(l, t) = v(t), \quad 0 \leq t \leq T, \quad (2.3)$$

с начальными условиями

$$Q(\xi, 0) = \varphi_0(\xi), \quad \left. \frac{\partial Q(\xi, t)}{\partial t} \right|_{t=0} = \psi_0(\xi), \quad -l \leq \xi \leq l, \quad (2.4)$$

промежуточными условиями

$$Q(\xi, t_i) = \varphi_i(\xi), \quad -l \leq \xi \leq l, \quad i = 1, \dots, m, \quad (2.5)$$

с конечными условиями

$$Q(\xi, T) = \varphi_T(\xi), \quad \left. \frac{\partial Q(\xi, t)}{\partial t} \right|_{t=T} = \psi_T(\xi), \quad -l \leq \xi \leq l, \quad (2.6)$$

и с условиями сопряжения в точке  $\xi = 0$  соединения участков

$$Q(0-0, t) = Q(0+0, t), \quad a_1 \rho_1 \left. \frac{\partial Q(\xi, t)}{\partial \xi} \right|_{\xi=0-0} = a_2 \rho_2 \left. \frac{\partial Q(\xi, t)}{\partial \xi} \right|_{\xi=0+0}. \quad (2.7)$$

Отметим, что для удобства после замены переменной (2.1) все функции оставлены в исходных обозначениях.

Так как граничные условия (2.3) неоднородны, решение уравнения (2.2) построим в виде суммы

$$Q(\xi, t) = V(\xi, t) + W(\xi, t), \quad (2.8)$$

где  $V(\xi, t)$  – функция с граничными условиями

$$V(-l, t) = V(l, t) = 0, \quad (2.9)$$

требуемая определения, а функция  $W(\xi, t)$  – решение уравнения (2.2) с условиями

$$W(-l, t) = \mu(t), \quad W(l, t) = \nu(t) \quad (2.10)$$

и имеет вид

$$W(\xi, t) = \frac{1}{2l}[(l - \xi)\mu(t) + (l + \xi)\nu(t)]. \quad (2.11)$$

Подстановка (2.8) в (2.2) с учетом (2.11) приводит к следующему уравнению для определения функции  $V(\xi, t)$ :

$$\frac{\partial^2 V(\xi, t)}{\partial t^2} = a_2^2 \frac{\partial^2 V(\xi, t)}{\partial \xi^2} + F(\xi, t), \quad -l \leq \xi \leq l, \quad 0 \leq t \leq T, \quad (2.12)$$

где

$$F(\xi, t) = \frac{1}{2l}[(\xi - l)\ddot{\mu}(t) - (\xi + l)\ddot{\nu}(t)]. \quad (2.13)$$

Функция  $V(\xi, t)$  удовлетворяет соответствующему (2.7) условию сопряжения в точке  $\xi = 0$  соединения участков. Из (1.8), согласно (2.1) и (1.1), будем иметь

$$\begin{aligned} \varphi_0(-l) &= \varphi_0(-l), & \varphi_i(-l) &= \varphi_i(-l), & \varphi_T(-l) &= \varphi_T(-l), \\ \psi_0(-l) &= \psi_0(-l), & \psi_T(-l) &= \psi_T(-l). \end{aligned} \quad (2.14)$$

Из начальных (2.4), промежуточных (2.5) и конечных (2.6) условий с учетом условий (1.8) и (2.14) функция  $V(\xi, t)$  должна удовлетворять следующим начальным условиям:

$$\begin{aligned} V(\xi, 0) &= \varphi_0(\xi) - \frac{1}{2l}[(l - \xi)\varphi_0(-l) + (l + \xi)\varphi_0(l)], \\ \left. \frac{\partial V(\xi, t)}{\partial t} \right|_{t=0} &= \psi_0(\xi) - \frac{1}{2l}[(l - \xi)\psi_0(-l) + (l + \xi)\psi_0(l)], \end{aligned} \quad (2.15)$$

промежуточным условиям

$$V(\xi, t_i) = \varphi_i(\xi) - \frac{1}{2l}[(l - \xi)\varphi_i(-l) + (l + \xi)\varphi_i(l)], \quad i = 1, \dots, m, \quad (2.16)$$

и конечным условиям

$$\begin{aligned} V(\xi, T) &= \varphi_T(\xi) - \frac{1}{2l}[(l - \xi)\varphi_T(-l) + (l + \xi)\varphi_T(l)], \\ \left. \frac{\partial V(\xi, t)}{\partial t} \right|_{t=T} &= \psi_T(\xi) - \frac{1}{2l}[(l - \xi)\psi_T(-l) + (l + \xi)\psi_T(l)]. \end{aligned} \quad (2.17)$$

Таким образом, решение исходной задачи сведено к задаче управления колебанием, описываемым неоднородным уравнением (2.12) с однородными граничными условиями (2.9), которая формулируется следующим образом: требуется найти такие граничные оптимальные управления  $\mu(t)$  и  $\nu(t)$ ,  $0 \leq t \leq T$ , под воздействием которых колебание, описываемое уравнением (2.12) с граничными условиями (2.9), переходит из заданного начального состояния (2.15) через промежуточные состояния (2.16) в конечное состояние (2.17) и минимизирующие функционал (1.9).

### 3. СВЕДЕНИЕ РЕШЕНИЯ ЗАДАЧИ С НУЛЕВЫМИ ГРАНИЧНЫМИ УСЛОВИЯМИ К ПРОБЛЕМЕ МОМЕНТОВ

Учитывая, что граничные условия (2.9) однородны, решение уравнения (2.12) ищем в виде

$$V(\xi, t) = \sum_{k=1}^{\infty} V_k(t) \sin \frac{\pi k \xi}{l}, \quad V_k(t) = \frac{1}{l} \int_{-l}^l V(\xi, t) \sin \frac{\pi k \xi}{l} d\xi. \quad (3.1)$$

Функции  $F(\xi, t)$ ,  $\varphi_i(\xi)$  ( $i = 0, 1, \dots, m+1$ ),  $\psi_0(\xi)$  и  $\psi_T(\xi)$  представим в виде рядов Фурье в базисе  $\left\{ \sin \frac{\pi k \xi}{l} \right\}$  ( $k = 1, 2, \dots$ ) и, подставив их значения вместе с  $V(\xi, t)$  в уравнение (2.12), в соотношение (2.13) и в условия (2.15)–(2.17), получим для каждого  $k = 1, 2, \dots$  следующее неоднородное обыкновенное дифференциальное уравнение второго порядка:

$$\ddot{V}_k(t) + \lambda_k^2 V_k(t) = F_k(t), \quad \lambda_k^2 = \left( \frac{a_2 \pi k}{l} \right)^2, \quad (3.2)$$

$$F_k(t) = \frac{a_2}{\lambda_k l} \left[ \ddot{v}(t) (2(-1)^k - 1) - \ddot{u}(t) \right], \quad (3.3)$$

с начальным

$$\begin{aligned} V_k(0) &= \varphi_k^{(0)} - \frac{a_2}{\lambda_k l} \left[ \varphi_0(-l) - \varphi_0(l) (2(-1)^k - 1) \right], \\ \dot{V}_k(0) &= \psi_k^{(0)} - \frac{a_2}{\lambda_k l} \left[ \psi_0(-l) - \psi_0(l) (2(-1)^k - 1) \right], \end{aligned} \quad (3.4)$$

промежуточным

$$V_k(t_i) = \varphi_k^{(i)} - \frac{a_2}{\lambda_k l} \left[ \varphi_i(-l) - \varphi_i(l) (2(-1)^k - 1) \right] \quad (3.5)$$

и конечным условиями

$$\begin{aligned} V_k(T) &= \varphi_k^{(T)} - \frac{a_2}{\lambda_k l} \left[ \varphi_T(-l) - \varphi_T(l) (2(-1)^k - 1) \right], \\ \dot{V}_k(T) &= \psi_k^{(T)} - \frac{a_2}{\lambda_k l} \left[ \psi_T(-l) - \psi_T(l) (2(-1)^k - 1) \right]. \end{aligned} \quad (3.6)$$

Здесь коэффициенты Фурье функций  $F(\xi, t)$ ,  $\varphi_i(\xi)$ , ( $i = 0, 1, \dots, m, m+1$ ),  $\psi_0(\xi)$  и  $\psi_T(\xi)$  обозначены через  $F_k(t)$ ,  $\varphi_k^{(i)}$  ( $i = 0, 1, \dots, m, m+1$ ),  $\psi_k^{(0)}$  и  $\psi_k^{(T)}$  соответственно.

Общее решение уравнения (3.2) с условиями (3.4) имеет вид

$$V_k(t) = V_k(0) \cos \lambda_k t + \frac{1}{\lambda_k} \dot{V}_k(0) \sin \lambda_k t + \frac{1}{\lambda_k} \int_0^t F_k(\tau) \sin \lambda_k (t - \tau) d\tau. \quad (3.7)$$

Следуя [2]–[6], [18], с учетом условий (3.5) и (3.6) из (3.7) получим, что функции управления  $\mu(t)$  и  $v(t)$  для каждого  $k$  удовлетворяют интегральным соотношениям

$$\begin{aligned} \int_0^T \mu(\tau) \sin \lambda_k (T - \tau) d\tau + E_k \int_0^T v(\tau) \sin \lambda_k (T - \tau) d\tau &= C_{1k}, \\ \int_0^T \mu(\tau) \cos \lambda_k (T - \tau) d\tau + E_k \int_0^T v(\tau) \cos \lambda_k (T - \tau) d\tau &= C_{2k}, \\ \int_0^T \mu(\tau) h_k^{(i)}(\tau) d\tau + E_k \int_0^T v(\tau) h_k^{(i)}(\tau) d\tau &= C_{1k}(t_i), \quad i = 1, \dots, m, \end{aligned} \quad (3.8)$$

где

$$\begin{aligned}
 C_{1k} &= \frac{1}{\lambda_k^2} \left[ \frac{\lambda_k l}{a_2} \tilde{C}_{1k} + X_{1k} + E_k Y_{1k} \right], \quad \tilde{C}_{1k} = \lambda_k V_k(T) - \lambda_k V_k(0) \cos \lambda_k T - \dot{V}_k(0) \sin \lambda_k T, \\
 C_{2k} &= \frac{1}{\lambda_k^2} \left[ \frac{\lambda_k l}{a_2} \tilde{C}_{2k} + X_{2k} + E_k Y_{2k} \right], \quad \tilde{C}_{2k} = \dot{V}_k(T) + \lambda_k V_k(0) \sin \lambda_k T - \dot{V}_k(0) \cos \lambda_k T, \\
 C_{1k}(t_i) &= \frac{1}{\lambda_k^2} \left[ \frac{\lambda_k l}{a_2} \tilde{C}_{1k}(t_i) + X_{1k}^{(i)} + E_k Y_{1k}^{(i)} \right], \quad \tilde{C}_{1k}(t_i) = \lambda_k V_k(t_i) - \lambda_k V_k(0) \cos \lambda_k t_i - \dot{V}_k(0) \sin \lambda_k t_i, \\
 X_{1k} &= \lambda_k \varphi_T(-l) - \psi_0(-l) \sin \lambda_k T - \lambda_k \varphi_0(-l) \cos \lambda_k T, \quad E_k = 1 - 2(-1)^k, \\
 X_{2k} &= \psi_T(-l) - \psi_0(-l) \cos \lambda_k T + \lambda_k \varphi_0(-l) \sin \lambda_k T, \\
 Y_{1k} &= \lambda_k \varphi_T(l) - \psi_0(l) \sin \lambda_k T - \lambda_k \varphi_0(l) \cos \lambda_k T, \quad Y_{2k} = \psi_T(l) - \psi_0(l) \cos \lambda_k T + \lambda_k \varphi_0(l) \sin \lambda_k T, \\
 X_{1k}^{(i)} &= \lambda_k \varphi_i(-l) - \psi_0(-l) \sin \lambda_k t_i - \lambda_k \varphi_0(-l) \cos \lambda_k t_i, \quad Y_{1k}^{(i)} = \lambda_k \varphi_i(l) - \psi_0(l) \sin \lambda_k t_i - \lambda_k \varphi_0(l) \cos \lambda_k t_i, \\
 h_k^{(i)}(\tau) &= \begin{cases} \sin \lambda_k(t_i - \tau) & \text{при } 0 \leq \tau \leq t_i, \\ 0 & \text{при } t_i < \tau \leq T. \end{cases}
 \end{aligned} \tag{3.9}$$

Таким образом, решение поставленной задачи оптимального управления сводится к нахождению таких граничных управлений  $\mu(t)$  и  $\nu(t)$ ,  $0 \leq t \leq T$ , которые для каждого  $k = 1, 2, \dots$  удовлетворяют интегральным соотношениям (3.8) и доставляют минимум функционалу (1.9).

На практике обычно выбираются несколько первых  $n$  ( $k = 1, \dots, n$ ) гармоник колебаний, и решается задача синтеза управлений, используя методы теории управления конечномерными системами. Решение задачи будем строить, придерживаясь этого подхода. Отметим, что для произвольного числа первых  $n$  ( $k = 1, \dots, n$ ) гармоник число интегральных соотношений в (3.8) равно  $(2 + m)n$ , которым должны одновременно удовлетворять искомые функции управления  $\mu(t)$  и  $\nu(t)$ . Учитывая это, из полученного соотношения (3.8) следует справедливость следующего утверждения о вполне управляемости (см. [19]).

**Утверждение.** Для произвольного числа первых  $n$  гармоник динамический процесс, описываемый (3.2) с условиями (3.4)–(3.6), вполне управляем на промежутке времени  $[0, T]$  тогда и только тогда, когда для любых значений величин  $C_{1k}$ ,  $C_{2k}$  и  $C_{1k}(t_i)$ ,  $i = 1, \dots, m$ ,  $k = 1, \dots, n$ , определяемых условиями (3.9), можно найти управления  $\mu(t)$  и  $\nu(t)$ ,  $t \in [0, T]$ , удовлетворяющие условию (3.8).

#### 4. РЕШЕНИЕ ЗАДАЧИ

Так как функционал (1.9) можно рассматривать как квадрат нормы соответствующего линейного нормированного пространства, а интегральные соотношения (3.8), порожденные функциями  $\mu(t)$  и  $\nu(t)$ ,  $0 \leq t \leq T$ , линейны, то задачу определения оптимального управления для каждого  $n$ ,  $n = 1, 2, \dots$ , можно рассматривать как проблему моментов (см. [1], [19], [20]). Следовательно, решение можно построить с помощью алгоритма решения проблемы моментов.

Поэтому построим решение задачи (1.9) и (3.8) при  $k = 1, 2, \dots, n$  с помощью алгоритма решения проблемы моментов. Для решения конечномерной (при  $k = 1, 2, \dots, n$ ) проблемы моментов (1.9) и (3.8), следуя [19], нужно найти величины  $p_k$ ,  $q_k$ ,  $\gamma_{ik}$ ,  $k = 1, \dots, n$ ,  $i = 1, \dots, m$ , связанные условием

$$\sum_{k=1}^n \left[ p_k C_{1k} + q_k C_{2k} + \sum_{i=1}^m \gamma_{ik} C_{1k}(t_i) \right] = 1, \tag{4.1}$$

для которых

$$(\rho_n^0)^2 = \min_{(4.1)} \int_0^T [h_{1n}^2(\tau) + h_{2n}^2(\tau)] d\tau, \tag{4.2}$$

где

$$h_{1n}(\tau) = \sum_{k=1}^n \left[ p_k \sin \lambda_k (T - \tau) + q_k \cos \lambda_k (T - \tau) + \sum_{i=1}^m \gamma_{ik} h_k^{(i)}(\tau) \right],$$

$$h_{2n}(\tau) = \sum_{k=1}^n E_k \left[ p_k \sin \lambda_k (T - \tau) + q_k \cos \lambda_k (T - \tau) + \sum_{i=1}^m \gamma_{ik} h_k^{(i)}(\tau) \right].$$

Для определения величин  $p_k^0, q_k^0, \gamma_{ik}^0, k = 1, \dots, n, i = 1, \dots, m$ , минимизирующих (4.2), применим метод неопределенных множителей Лагранжа. Введем функцию

$$f_n = \int_0^T \left[ (h_{1n}(\tau))^2 + (h_{2n}(\tau))^2 \right] d\tau + \beta_n \left[ \sum_{k=1}^n \left( p_k C_{1k} + q_k C_{2k} + \sum_{i=1}^m \gamma_{ik} C_{1k}(t_i) \right) - 1 \right],$$

где  $\beta_n$  — неопределенный множитель Лагранжа. На основе этого метода, вычисляя производные по  $p_k, q_k, \gamma_{ik}, k = 1, \dots, n, i = 1, \dots, m$ , функции  $f_n$  и приравнявая к нулю, получаем следующую систему интегральных соотношений:

$$\int_0^T [h_{1n}(\tau) + E_k h_{2n}(\tau)] \sin \lambda_k (T - \tau) d\tau = -\frac{\beta_n}{2} C_{1k},$$

$$\int_0^T [h_{1n}(\tau) + E_k h_{2n}(\tau)] \cos \lambda_k (T - \tau) d\tau = -\frac{\beta_n}{2} C_{2k},$$

$$\int_0^T [h_{1n}(\tau) + E_k h_{2n}(\tau)] h_k^{(i)}(\tau) d\tau = -\frac{\beta_n}{2} C_{1k}(t_i), \quad k = 1, \dots, n, \quad i = 1, \dots, m.$$

Не приводя дальнейших построений решения (так как они аналогичны построениям, приведенным в [5], [6]) отметим, что оптимальные управления  $\mu_n^0(\tau)$  и  $v_n^0(\tau)$  для любого  $n = 1, 2, \dots$  представятся в виде

$$\mu_n^0(\tau) = \begin{cases} \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=1}^m \gamma_{ik}^0 \sin \lambda_k(t_i - \tau) \right], & 0 \leq \tau \leq t_1, \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=2}^m \gamma_{ik}^0 \sin \lambda_k(t_i - \tau) \right], & t_1 < \tau \leq t_2, \\ \dots & \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \gamma_{mk}^0 \sin \lambda_k(t_m - \tau) \right], & t_{m-1} < \tau \leq t_m, \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n G_k(p_k^0, q_k^0, \lambda_k, T, \tau), & t_m < \tau \leq t_{m+1} = T, \end{cases} \quad (4.3)$$

$$v_n^0(\tau) = \begin{cases} \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n E_k \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=1}^m \gamma_{ik}^0 \sin \lambda_k(t_i - \tau) \right], & 0 \leq \tau \leq t_1, \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n E_k \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=2}^m \gamma_{ik}^0 \sin \lambda_k(t_i - \tau) \right], & t_1 < \tau \leq t_2, \\ \dots & \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n E_k \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \gamma_{mk}^0 \sin \lambda_k(t_m - \tau) \right], & t_{m-1} < \tau \leq t_m, \\ \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n E_k G_k(p_k^0, q_k^0, \lambda_k, T, \tau), & t_m < \tau \leq t_{m+1} = T, \end{cases} \quad (4.4)$$

где

$$G_k(p_k^0, q_k^0, \lambda_k, T, \tau) = p_k^0 \sin \lambda_k(T - \tau) + q_k^0 \cos \lambda_k(T - \tau).$$

Здесь величины  $p_k^0, q_k^0, \gamma_{ik}^0, k = 1, \dots, n, i = 1, \dots, m$ , являются решениями задач (4.1) и (4.2), а

$$\begin{aligned} (\rho_n^0)^2 &= \int_0^T \left[ \left( h_{1n}^0(\tau) \right)^2 + \left( h_{2n}^0(\tau) \right)^2 \right] d\tau, \\ h_{1n}^0(\tau) &= \sum_{k=1}^n \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=1}^m \gamma_{ik}^0 h_k^{(i)}(\tau) \right], \\ h_{2n}^0(\tau) &= \sum_{k=1}^n E_k \left[ G_k(p_k^0, q_k^0, \lambda_k, T, \tau) + \sum_{i=1}^m \gamma_{ik}^0 h_k^{(i)}(\tau) \right]. \end{aligned}$$

Отметим, что значения оптимальных управлений  $\mu_n^0(\tau)$  и  $v_n^0(\tau)$  на конце каждого предыдущего промежутка  $[t_{j-1}, t_j]$  совпадают со значениями в начале последующего промежутка  $(t_j, t_{j+1}]$ ,  $j = 1, \dots, m$ , и эти значения имеют следующий вид:

$$\begin{aligned} \mu_n^0(t_j) &= \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n \left[ p_k^0 \sin \lambda_k(T - t_j) + q_k^0 \cos \lambda_k(T - t_j) + \sum_{i=j+1}^m \gamma_{ik}^0 \sin \lambda_k(t_i - t_j) \right], \\ v_n^0(t_j) &= \frac{1}{(\rho_n^0)^2} \sum_{k=1}^n E_k \left[ p_k^0 \sin \lambda_k(T - t_j) + q_k^0 \cos \lambda_k(T - t_j) + \sum_{i=j+1}^m \gamma_{ik}^0 \sin \lambda_k(t_i - t_j) \right]. \end{aligned}$$

Следовательно, построенные оптимальные граничные управления  $\mu_n^0(\tau)$  и  $v_n^0(t)$  как функции от времени непрерывны на промежутке  $[0, T]$ .

Таким образом, доказана справедливость следующей теоремы.

**Теорема.** При согласовании исходных данных задачи, указанных в разд. 1, и выполнении условий вполне управляемости, задача оптимального управления (1.1)–(1.9) (или (2.12), (2.15)–(2.17), (1.9)) имеет непрерывное решение, определяемое формулами (4.3) и (4.4).

Подставляя построенные выражения для оптимальных управлений  $\mu_n^0(t)$  и  $v_n^0(\tau)$  в (3.3), а полученное для  $F_k^0(\tau)$  выражение – в (3.7), получим функцию  $V_k^0(t)$ ,  $t \in [0, T]$ ,  $k = 1, \dots, n$ . Далее, из формулы (3.1) будем иметь

$$V_n^0(\xi, t) = \sum_{k=1}^n V_k^0(t) \sin \frac{\pi k}{l} \xi, \quad (4.5)$$

а с помощью (2.8) и (2.11) получим функцию  $Q_n^0(\xi, t)$ , которую назовем *оптимальной функцией колебания* для первых  $n$  гармоник. Функция  $Q_n^0(\xi, t)$  запишется в виде

$$Q_n^0(\xi, t) = V_n^0(\xi, t) + W_n^0(\xi, t), \quad (4.6)$$

где

$$W_n^0(\xi, t) = \frac{1}{2l} \left[ (l - \xi) \mu_n^0(t) + (l + \xi) v_n^0(t) \right]. \quad (4.7)$$

Учитывая обозначения (2.1), согласно (4.5)–(4.7), оптимальная функция  $Q_n^0(x, t)$  при  $-l_1 \leq x \leq l$  представляется в виде

$$Q_n^0(x, t) = \begin{cases} \sum_{k=1}^n V_k^0(t) \sin \frac{\pi k}{l_1} x + \frac{1}{2} \left[ \left( 1 - \frac{x}{l_1} \right) \mu_n^0(t) + \left( 1 + \frac{x}{l_1} \right) v_n^0(t) \right], & -l_1 \leq x \leq 0, \quad 0 \leq t \leq T, \\ \sum_{k=1}^n V_k^0(t) \sin \frac{\pi k}{l} x + \frac{1}{2} \left[ \left( 1 - \frac{x}{l} \right) \mu_n^0(t) + \left( 1 + \frac{x}{l} \right) v_n^0(t) \right], & 0 \leq x \leq l, \quad 0 \leq t \leq T. \end{cases} \quad (4.8)$$

Функция  $Q_n^0(x, t)$  является непрерывной и можно убедиться, что для функции  $Q_n^0(x, t)$  выполняются условия сопряжения в точке  $x = 0$  соединения участков (1.4).

## 5. ПРИМЕР

Для иллюстрации вышеизложенного построения предположим, что в граничных условиях (1.3) правый конец закреплен:  $Q(l, t) = 0$ ,  $0 \leq t \leq T$  (т.е.  $v(t) = 0$ ). Рассмотрим случай  $m = 1$ , т.е. когда в одном промежуточном моменте времени  $t_1$  ( $0 < t_1 < T$ ) задано состояние колебания

$$Q(x, t_1) = \varphi_1(x), \quad -l_1 \leq x \leq l.$$

Для простоты изложения построим функцию оптимального граничного управления  $\mu_n^0(\tau)$  при  $n = 1$  (следовательно,  $k = 1$ ). В этом случае из (4.3) получим

$$\mu_1^0(t) = \begin{cases} \mu_1^{(1)0}(t) = \frac{1}{(\rho_1^0)^2} [p_1^0 \sin \lambda_1(T-t) + q_1^0 \cos \lambda_1(T-t) + \gamma_{11}^0 \sin \lambda_1(t_1-t)], & 0 \leq t \leq t_1, \\ \mu_1^{(2)0}(t) = \frac{1}{(\rho_1^0)^2} [p_1^0 \sin \lambda_1(T-t) + q_1^0 \cos \lambda_1(T-t)], & t_1 < t \leq t_2 = T, \end{cases}$$

где

$$\begin{aligned} (\rho_1^0)^2 = & \frac{T}{2} ((q_1^0)^2 + (p_1^0)^2) + \frac{t_1}{2} ((\gamma_{11}^0)^2 + 2\gamma_{11}^0 (p_1^0 \cos \lambda_1(T-t_1) - q_1^0 \sin \lambda_1(T-t_1))) + \\ & + \frac{1}{\lambda_1} \left( p_1^0 q_1^0 \sin^2 \lambda_1 T - \frac{(\gamma_{11}^0)^2}{2} \sin \lambda_1 t_1 \cos \lambda_1 t_1 + \gamma_{11}^0 (q_1^0 \sin \lambda_1 T - p_1^0 \cos \lambda_1 T) \sin \lambda_1 t_1 + \right. \\ & \left. + \frac{(q_1^0)^2 - (p_1^0)^2}{2} \sin \lambda_1 T \cos \lambda_1 T \right). \end{aligned}$$

Оптимальная функция состояния  $Q_n^0(\xi, t)$ , согласно формулам (4.5)–(4.7), будет иметь вид

$$Q_1^0(\xi, t) = V_1^0(\xi, t) + W_1^0(\xi, t) = V_1^0(t) \sin \frac{\pi}{l} \xi + \left(1 - \frac{\xi}{l}\right) \mu_1^0(t).$$

Согласно (4.8), явное выражение оптимальной функции  $Q_1^0(x, t)$  при  $-l_1 \leq x \leq l$  представляется в виде

$$Q_1^0(x, t) = \begin{cases} V_1^0(t) \sin \frac{\pi}{l_1} x + \frac{1}{2} \left(1 - \frac{x}{l_1}\right) \mu_1^0(t), & -l_1 \leq x \leq 0, \quad 0 \leq t \leq T, \\ V_1^0(t) \sin \frac{\pi}{l} x + \frac{1}{2} \left(1 - \frac{x}{l}\right) \mu_1^0(t), & 0 \leq x \leq l, \quad 0 \leq t \leq T, \end{cases}$$

или

при  $0 \leq t \leq t_1$

$$Q_1^0(x, t) = \begin{cases} V_1^0(t) \sin \frac{\pi}{l_1} x + \frac{1}{2} \left(1 - \frac{x}{l_1}\right) \mu_1^{(1)0}(t), & -l_1 \leq x \leq 0, \\ V_1^0(t) \sin \frac{\pi}{l} x + \frac{1}{2} \left(1 - \frac{x}{l}\right) \mu_1^{(1)0}(t), & 0 \leq x \leq l, \end{cases}$$

при  $t_1 < t \leq t_2 = T$

$$Q_1^0(x, t) = \begin{cases} V_1^0(t) \sin \frac{\pi}{l_1} x + \frac{1}{2} \left(1 - \frac{x}{l_1}\right) \mu_1^{(2)0}(t), & -l_1 \leq x \leq 0, \\ V_1^0(t) \sin \frac{\pi}{l} x + \frac{1}{2} \left(1 - \frac{x}{l}\right) \mu_1^{(2)0}(t), & 0 \leq x \leq l. \end{cases}$$

Из полученных явных выражений для функции  $Q_1^0(x, t)$  следует, что в точке стыка  $x = 0$  имеем

$$Q_1^0(0, t) = \mu_1^0(t), \quad 0 \leq t \leq T.$$

Из этих явных выражений также видно, что в силу непрерывности функции  $\mu_1^0(t)$  оптимальная функция колебания  $Q_1^0(x, t)$  также непрерывна.

### ЗАКЛЮЧЕНИЕ

В работе предложен конструктивный подход к построению функции оптимального граничного управления одномерными неоднородными колебательными процессами смещением на двух концах, переводящие колебания за определенный промежуток времени из заданного начального состояния через многоточечные промежуточные состояния в заданное конечное состояние с критерием качества, заданным на всем промежутке времени. Результаты могут быть использованы при проектировании оптимального граничного управления процессами разнородных колебаний в физических и технологических системах.

### СПИСОК ЛИТЕРАТУРЫ

1. Бутковский А.Г. Методы управления системами с распределенными параметрами. М.: Наука, 1975. 568 с.
2. Barseghyan V.R. Control problem of string vibrations with inseparable multipoint conditions at intermediate points in time // *Mech. of Solid.* 2019. V. 54. Iss. 8. P. 1216–1226.  
<https://doi.org/10.3103/S0025654419080120>
3. Barseghyan V.R. The problem of optimal control of string vibrations // *Inter. Appl. Mech.* 2020. V. 56. № 4. P. 471–48.  
<https://doi.org/10.1007/s10778-020-01030-w>
4. Барсегян В.Р. Задача оптимального управления колебаниями струны с неразделенными условиями на функции состояния в заданные промежуточные моменты времени // *Автоматика и телемехан.* 2020. № 2. С. 36–47; *Automation and Remote Control.* 2020. V. 81. Iss. 2. P. 226–235.  
<https://doi.org/10.31857/S0005231020020038>
5. Barseghyan V., Solodusha S. Optimal boundary control of string vibrations with given shape of deflection at a certain moment of time. *Mathematical Optimization Theory and Operations Research. MOTOR 2021* // *Lect. Not. Comp. Sci.* 2021. V. 12755. P. 299–313.  
[https://doi.org/10.1007/978-3-030-77876-7\\_20](https://doi.org/10.1007/978-3-030-77876-7_20)
6. Barseghyan V. and Solodusha S. On one problem in optimal boundary control for string vibrations with a given velocity of points at an intermediate moment of time // *Conf. Paper. Publ.: IEEE. 2021 International Russian Automation Conference (RusAutoCon)*, P. 343–349, 2021.  
<https://doi.org/10.1109/RusAutoCon52004.2021.9537514>
7. Barseghyan V.R. On the controllability and observability of linear dynamic systems with variable structure // *Proceed.s of 2016 Inter. Conf. “Stability and Oscillations of Nonlinear Control Systems” (Pyatnitskiy’s Conference)*, STAB 2016.  
<https://doi.org/10.1109/STAB.2016.7541163>
8. Львова Н.Н. Оптимальное управление некоторой распределенной неоднородной колебательной системой // *Автоматика и телемехани.* 1973. № 10. С. 22–32.
9. Ильин В.А. Оптимизация граничного управления колебаниями стержня, состоящего из двух разнородных участков // *Докл. АН.* 2011. Т. 440. № 2. С. 159–163.
10. Ильин В.А. О приведении в произвольно заданное состояние колебаний первоначально покоящегося стержня, состоящего из двух разнородных участков // *Докл. АН.* 2010. Т. 435. № 6. С. 732–735.
11. Егоров А.И., Знаменская Л.Н. Об управляемости упругих колебаний последовательно соединенных объектов с распределенными параметрами // *Тр. ИММ УрОРАН.* 2011. Т. 17. № 1. С. 85–92.
12. Егоров А.И., Знаменская Л.Н. Об управляемости колебаний сети из связанных объектов с распределенными и сосредоточенными параметрами // *Ж. вычисл. матем. и матем. физ.* 2009. Т. 49. № 5. С. 815–825. *Comput. Math. Math. Phys.* 2009. V. 49. Iss. 5. P. 786–796.  
<https://doi.org/10.1134/S0965542509050054>
13. Провоторов В.В. Построение граничных управлений в задаче о гашении колебаний системы струн // *Вестн. СПбГУ. Сер. 10.* 2012. Вып. 1. С. 62–71.  
<https://doi.org/10.2012>

14. *Amara J. Ben, Beldi E.* Boundary controllability of two vibrating strings connected by a point mass with variable coefficients // *SIAM J. Control Optim.* 2019. V. 57. № 5. P. 3360–3387.  
<https://doi.org/10.1137/16M1100496>
15. *Mercier D., Régnier V.* Boundary controllability of a chain of serially connected Euler-Bernoulli beams with interior masses // *Collectanea Math.* 2009. V. 60. № 3. P. 307–334.  
<https://doi.org/10.1007/BF03191374>
16. *Кулешов А.А.* Смешанные задачи для уравнения продольных колебаний неоднородного стержня и уравнения поперечных колебаний неоднородной струны, состоящих из двух участков разной плотности и упругости // *Докл. АН.* 2012. Т. 442. № 5. С. 594–597.
17. *Рогожников А.М.* Исследование смешанной задачи, описывающей процесс колебаний стержня, состоящего из нескольких участков с произвольными длинами // *Докл. АН.* 2012. Т. 444. С. 488–491.
18. *Холодовский С.Е., Чухрий П.А.* Задача о движении неограниченной кусочно-однородной струны // *Ученые записки Забайкальского гос. ун-та. Сер. Физика, математика, техника, технология.* 2018. Т. 13. № 4. С. 42–50.  
<https://doi.org/10.21209/2308-8761-2018-13-4-42-50>
19. *Красовский Н.Н.* Теория управления движением. М.: Наука, 1968. 476 с.
20. *Барсегян В.Р.* Управление составных динамических систем и систем с многоточечными промежуточными условиями. М.: Наука, 2016. 230 с.

# ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 517.926

## КОНТРПРИМЕРЫ К ПРЕДПОЛОЖЕНИЮ О ВОЗМОЖНОСТИ ПРОДОЛЖЕНИЯ УСЕЧЕННЫХ РЕШЕНИЙ УСЕЧЕННОГО ЛИНЕЙНОГО ОБЫКНОВЕННОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ

© 2023 г. С. А. Абрамов<sup>1,\*</sup>, А. А. Рябенко<sup>1,\*\*</sup>, Д. Е. Хмельнов<sup>1,\*\*\*</sup><sup>1</sup> 119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия

\*e-mail: sergeyabramov@mail.ru

\*\*e-mail: anna.ryabenko@gmail.com

\*\*\*e-mail: dennis\_khmelnov@mail.ru

Поступила в редакцию 25.04.2022 г.  
Переработанный вариант 01.06.2022 г.  
Принята к публикации 17.09.2022 г.

Ранее авторами были предложены алгоритмы, которые позволяют находить экспоненциально-логарифмические решения линейных обыкновенных дифференциальных уравнений с коэффициентами в виде таких степенных рядов, для которых известны только начальные члены. В решение входит конечное число степенных рядов и для них вычисляется максимально возможное число членов. Теперь к этим алгоритмам добавляется опция подтверждения того, что без дополнительной информации об уравнении невозможно получить большее число членов этих рядов: строится контрпример к предположению о возможности получения однозначно определенных дополнительных членов. В предыдущих работах авторами предлагались такого рода подтверждения для случаев лорановых и регулярных решений. Библ. 23.

**Ключевые слова:** дифференциальные уравнения, усеченные степенные ряды, системы компьютерной алгебры.

DOI: 10.31857/S0044466923010027, EDN: LEVQUY

### 1. ВВЕДЕНИЕ

В этой статье рассматриваются линейные обыкновенные дифференциальные уравнения (ЛОДУ) с коэффициентами, имеющими вид таких степенных рядов, относительно которых известны только их первые члены, а “хвосты” этих рядов неизвестны. Таким образом, о рассматриваемых уравнениях имеется лишь неполная информация. В [1]–[6] предлагались алгоритмы поиска решений таких уравнений в виде лорановых рядов, а также поиска регулярных и экспоненциально-логарифмических решений. Было доказано, что эти алгоритмы позволяют найти максимально возможное число членов тех рядов, которые входят в решения. Алгоритмы реализованы авторами в виде пакета процедур — см. [7]–[10]. Для пользователя этих процедур может оказаться желательным получить какие-то наглядные доводы в пользу максимальности числа найденных членов рядов. Такого рода наглядные средства и предлагаются ниже: описан алгоритм, который для произвольного уравнения с усеченными коэффициентами предъявляет два продолженных варианта исходного уравнения, решения которых различаются между собой в последующих (не попавших в число найденных) членах рядов, входящих в решения.

Поясним простым примером суть рассматриваемой задачи. С помощью алгоритма из [4] устанавливается, что уравнение

$$\left(x^3 + \frac{x^5}{3} + O(x^6)\right)y'(x) + (1 + 3x + O(x^3))y(x) = 0 \quad (1)$$

(здесь  $O(x^k)$  обозначает какие-то неизвестные нам члены степенного ряда с показателями степени  $x$ , не меньшими, чем  $k$ ) имеет решение

$$e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x (C + O(x)), \quad (2)$$

где  $C$  — произвольная постоянная. Можем ли мы, исходя из (1), узнать большее число членов ряда, который в (2) представлен как  $(C + O(x))$ ? Отрицательный ответ обосновывается предъявлением двух продолженных вариантов уравнения (1):

$$\left(x^3 + \frac{x^5}{3} + 4x^6 + O(x^7)\right)y'(x) + (1 + 3x + x^3 + O(x^4))y(x) = 0 \quad (3)$$

и

$$\left(x^3 + \frac{x^5}{3} - 4x^6 + O(x^7)\right)y'(x) + (1 + 3x + O(x^4))y(x) = 0. \quad (4)$$

Алгоритм из [4] находит решения этих двух уравнений:

$$e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x (C + 4Cx + O(x^2)),$$

$$e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x (C - 3Cx + O(x^2)).$$

Это показывает, что для входящего в (2) ряда  $(C + O(x))$  мы не можем найти его коэффициент при  $x$ , не используя для этого никакой дополнительной информации, касающейся уравнения (1). Пара уравнений (3), (4) образуют в терминологии этой статьи *контрпример* к предположению, что последующие члены рядов, входящих в экспоненциально-логарифмические решения уравнения (1), могут быть найдены, исходя только из этого усеченного уравнения.

В разд. 6 демонстрируется построение этого контрпримера, т.е. уравнений (3), (4), с помощью нашего алгоритма, реализованного в среде Maple.

Предварительный вариант этой работы был представлен в виде доклада [11].

## 2. УСЕЧЕННОЕ УРАВНЕНИЕ

Пусть  $K$  — алгебраически замкнутое поле характеристики 0. Для кольца полиномов от  $x$  над  $K$  будет использоваться обычное обозначение  $K[x]$ . Кольцо формальных степенных рядов от  $x$  над  $K$  обозначается через  $K[[x]]$ , поле формальных лорановых рядов — через  $K((x))$ . Очевидно,  $K[x] \subset K[[x]] \subset K((x))$ . Для принадлежащего  $K((x))$  ненулевого элемента  $a(x) = \sum a_i x^i$  его *валуация*  $\text{val} a(x)$  определена равенством  $\text{val} a(x) = \min\{i | a_i \neq 0\}$ , при этом  $\text{val} 0 = \infty$ .

Мы рассматриваем уравнения вида

$$a_r(x)y^{(r)}(x) + a_{r-1}(x)y^{(r-1)}(x) + \dots + a_0(x)y(x) = 0, \quad (5)$$

где  $y(x)$  — неизвестная функция от  $x$ . Относительно  $a_0(x), a_1(x), \dots, a_r(x)$  (*коэффициентов уравнения*) предполагается, что для каждого  $i = 0, 1, \dots, r$  коэффициент  $a_i(x)$  — это *усеченный ряд*

$$a_i(x) = \sum_{j=0}^{t_i} a_{ij} x^j + O(x^{t_i+1}), \quad (6)$$

где  $a_{ij} \in K$ ;  $t_i$  — целое, такое что  $t_i \geq -1$  (если  $t_i = -1$ , сумма в (6) полагается равной 0). Здесь и далее символ  $O(x^t)$ , встречающийся в формальных выражениях, обозначает некоторый ряд, валуация которого не меньше, чем  $t$ . Мы называем  $t_i$  *степенью усечения* ряда  $a_i(x)$ , представленного в виде (6). Заметим, что любой коэффициент в (5) может иметь вид  $O(x^m)$ ,  $m \geq 0$ .

*Продолжением* уравнения (5) будем называть любое уравнение

$$\tilde{a}_r(x)y^{(r)}(x) + \tilde{a}_{r-1}(x)y^{(r-1)}(x) + \dots + \tilde{a}_0(x)y(x) = 0,$$

для которого  $\tilde{a}_i(x) - a_i(x) = O(x^{t_i+1})$ , т.е.  $\text{val}(\tilde{a}_i(x) - a_i(x)) > t_i$ ,  $i = 0, 1, \dots, r$ .

### 3. УСЕЧЕННЫЕ РЕШЕНИЯ

Формальными экспоненциально-логарифмическими решениями уравнения

$$\left( \sum_{j=0}^{\infty} \tilde{a}_{rj} x^j \right) y^{(r)}(x) + \left( \sum_{j=0}^{\infty} \tilde{a}_{r-1,j} x^j \right) y^{(r-1)}(x) + \dots + \left( \sum_{j=0}^{\infty} \tilde{a}_{0j} x^j \right) y(x) = 0, \quad (7)$$

имеющего полностью заданные коэффициенты-ряды, называются решения вида

$$e^{Q(x^{-1/q})} x^{\lambda} w(x^{1/q}), \quad (8)$$

где  $Q$  — полином с коэффициентами из  $K$ ,  $q \in \mathbb{Z}_{>0}$ ,  $\lambda \in K$ ,

$$w(x) = \sum_{s=0}^m w_s(x) \ln^s x,$$

$m \in \mathbb{Z}_{\geq 0}$ ,  $w_s(x) \in K((x))$ ,  $s = 0, \dots, m$ , и  $w_m(x) \neq 0$ . При этом  $x^{\lambda} w(x^{1/q})$  называется *регулярной частью*,  $Q(x^{-1/q})$  — *показателем экспоненциальной части*,  $q$  — *индексом ветвления* решения (8).

Если  $q = 1$  и  $Q \in K$ , то решение (8) называется *формальным регулярным*, в противном случае — *нерегулярным*. При  $q = 1$ ,  $Q \in K$ ,  $\lambda \in \mathbb{Z}$  и  $w(x) \in K((x))$  формальное регулярное решение (8) называется *лорановым*. В обсуждениях решений уравнений слово “формальный” мы опускаем, но подразумеваем.

Пусть в уравнении (7) старший коэффициент  $\tilde{a}_r(x)$  отличен от нуля. Известно (см., например, [12, гл. V], [13]–[16]), что существует  $r$  линейно независимых над  $K$  решений вида (8) для уравнения (7). В [13]–[17] предложены алгоритмы нахождения для  $r$  линейно независимых решений вида (8) их индекса ветвления  $q$  и показателя экспоненциальной части  $Q(x^{-1/q})$ . Пусть в (7) валюация по крайней мере одного из коэффициентов равна 0. Тогда для построения индекса ветвления  $q$  и показателя экспоненциальной части  $Q(x^{-1/q})$  для всех решений достаточно знать значения  $r \text{val} \tilde{a}_r(x)$  начальных коэффициентов всех  $\tilde{a}_i(x)$ ,  $i = 0, 1, \dots, r$  (см., например, [18]). Для построения регулярной части решения с любой заданной степенью усечения входящих в  $w(x)$  рядов можно применять алгоритмы, предложенные в [12, гл. IV], [19], [20, гл. II, VIII]. Для этого построения также достаточно знать некоторое конечное число начальных коэффициентов всех  $\tilde{a}_i(x)$  (см. [21, Prop. 1]).

Пусть  $Q(x^{-1/q}) \in K[x^{-1/q}]$ ,  $q \in \mathbb{Z}_{>0}$ ,  $\lambda \in K$  и

$$w_s^{(k_s)}(x) = \sum_{j=j_s}^{k_s} w_{s,j} x^j + O(x^{k_s+1}),$$

$j_s, k_s \in \mathbb{Z}$ ,  $k_s \geq j_s$ ,  $s = 0, \dots, m$ , и  $w_{m,j_m} \neq 0$ . Для уравнения (5) с усеченными коэффициентами мы называем *решением с усеченной регулярной частью* выражение

$$e^{Q(x^{-1/q})} x^{\lambda} \sum_{s=0}^m w_s^{(k_s)}(x^{1/q}) \ln^s x, \quad (9)$$

когда скоро любое уравнение, являющееся продолжением (5), имеет решение  $e^{Q(x^{-1/q})} x^{\lambda} \tilde{w}(x^{1/q})$ , являющееся *продолжением решения* (9), т.е.  $\tilde{w}(x)$  имеет вид

$$\tilde{w}(x) = \sum_{s=0}^m \tilde{w}_s(x) \ln^s x$$

и выполняется  $\tilde{w}_s(x) - w_s^{(k_s)}(x) = O(x^{k_s+1})$ , т.е.  $\text{val}(\tilde{w}_s(x) - w_s^{(k_s)}(x)) > k_s$ ,  $s = 0, 1, \dots, m$ . Мы говорим, что усеченное решение *инвариантно* относительно любого продолжения уравнения (5).

#### 4. РЕШЕНИЯ С МАКСИМАЛЬНОЙ СТЕПЕНЬЮ УСЕЧЕНИЯ

В [1]–[4], [7] показано, что для уравнения вида (5) возможно построение всех инвариантных усеченных решений с *максимальной степенью усечения* входящих в решение рядов. Максимальность степени усечения в  $s_{\max}$  означает, что не существует инвариантного решения  $s$ , являющегося продолжением  $s_{\max}$ , такого, что степень усечения хотя бы одного ряда в  $s$  больше, чем степень усечения соответствующего ряда в  $s_{\max}$ . В этом случае мы говорим об исчерпывающем использовании информации о заданном уравнении при построении усеченных решений. В названных статьях представлены алгоритмы решения этой задачи и их реализация в Maple.

В [22], [23] мы рассматривали вопрос автоматического подтверждения такого исчерпывающего использования информации о заданном уравнении при построении лорановых и регулярных усеченных решений. Подтверждением служит контрпример, состоящий из двух различных продолжений заданного уравнения, которые приводят к появлению различных дополнительных членов в решениях.

Алгоритмы построения как самих усеченных решений, так и контрпримеров указанного типа основаны на поиске решений с *литералами*, т.е. с символическими обозначениями незадаанных коэффициентов входящих в уравнение рядов (см. [7]). Литералы обозначают коэффициенты при членах ряда, степени которых больше степени усечения ряда. Поиск решений с помощью литералов означает представление последующих (неинвариантных для всех возможных продолжений) членов ряда выражениями, содержащими литералы, т.е. незадаанные коэффициенты. Это позволяет прояснить влияние незадаанных коэффициентов на последующие члены рядов в решении.

Ниже мы расширяем полученные в [22], [23] результаты на случай экспоненциально-логарифмических решений с усеченной регулярной частью. Решается задача предъявления двух различных продолжений исходного уравнения, дающих контрпример к предположению о возможности добавления инвариантных членов к входящим в построенные усеченные решения заданного усеченного уравнения.

#### 5. ПОСТРОЕНИЕ КОНТРПРИМЕРА

Содержащее литералы продолжение уравнения (5) имеет вид

$$\left( \sum_{j=0}^{t_r} a_{rj} x^j + \sum_{j=t_r+1}^{\infty} U_{rj} x^j \right) y^{(r)}(x) + \left( \sum_{j=0}^{t_{r-1}} a_{r-1,j} x^j + \sum_{j=t_{r-1}+1}^{\infty} U_{r-1,j} x^j \right) y^{(r-1)}(x) + \dots$$

$$\dots + \left( \sum_{j=0}^{t_0} a_{0j} x^j + \sum_{j=t_0+1}^{\infty} U_{0j} x^j \right) y(x) = 0, \quad (10)$$

где  $U_{ij}$  — обозначение литерала. Алгоритм из [15], в более общем виде представленный в [17], позволит построить экспоненциальные части  $e^{Q(x^{-1/q})}$  всех решений вида (8) для уравнения (10). Нас будут интересовать только те из них, для которых индекс ветвления  $q$  и коэффициенты полинома  $Q$  не зависят от литералов. Для каждой из таких пар  $q, Q$  выполняем в (10) подстановку

$$x = t^q, \quad y(x) = e^{Q(1/t)} z(t),$$

где  $t$  — новая независимая переменная,  $z(t)$  — новая неизвестная функция. В результате подстановки и умножения уравнения на  $e^{-Q(1/t)}$  получаем новое уравнение, коэффициенты которого являются лорановыми рядами от  $t$ . Коэффициенты этих рядов являются полиномами от литералов над  $K$ . Для нового уравнения строим регулярные решения  $t^\lambda w(t)$ , используя вариант алгоритма из [3, разд. 4.2]. Этот вариант для каждого ряда, входящего в регулярное решение, вычисляет максимальное число членов, инвариантных относительно всех продолжений уравнения, и, дополнительно, еще один член, коэффициент которого зависит от литералов. Этот коэффициент является полиномом над  $K$  конечного числа литералов.

Таким образом, для экспоненциально-логарифмического решения с усеченной регулярной частью (9) мы получим конечное множество полиномов от литералов, которое можно использовать для построения контрпримера.

В [23] при рассмотрении усеченных лорановых и регулярных решений нами была доказана следующая

**Лемма 1** (см. [23, лемма 1]). *При любом целом  $m > 0$  и  $p_i(x_1, \dots, x_l) \in K[x_1, \dots, x_k] \setminus K$ ,  $i = 1, \dots, m$ , существуют такие  $\alpha_1, \dots, \alpha_l, \beta_1, \dots, \beta_l \in K$ , что*

$$p_i(\alpha_1, \dots, \alpha_l) \neq p_i(\beta_1, \dots, \beta_l), \quad i = 1, \dots, m. \quad (11)$$

Из приведенного в [23] доказательства следует, что

$$\alpha_1, \dots, \alpha_l, \beta_1, \dots, \beta_l \quad (12)$$

могут быть взяты целочисленными (в любое поле  $K$  характеристики 0 естественным образом вкладывается кольцо целых чисел). Можно перебирать все целочисленные наборы (12) до первого, удовлетворяющего (11). Это позволит фактически найти нужный набор. Здесь возможно также привлечение эвристик и случайного выбора.

На основе этого можно описать алгоритм построения контрпримера к предположению о возможности получения однозначно определенных дополнительных членов рядов, присутствующих в решениях.

**Теорема 1.** *Пусть  $\mathcal{E}$  — уравнение вида (5),  $s$  — его усеченное решение, найденное с помощью алгоритма из [4]. Тогда для  $\mathcal{E}$  существуют два различных продолжения  $\mathcal{E}_1$  и  $\mathcal{E}_2$ , имеющих усеченные решения  $s_1$  и соответственно  $s_2$ , которые являются такими продолжениями  $s$ , что любой усеченный ряд, входящий в  $s$ , имеет продолжение как в  $s_1$ , так и в  $s_2$ , и уже самые первые дополнительные члены в  $s_1$ ,  $s_2$  не совпадают.*

**Доказательство.** Каждый входящий в усеченное решение вида (9) ряд строится алгоритмом из [4] до первого содержащего литералы члена, который уже не включается в итоговое усеченное решение. Перед моментом отбрасывания членов с литералами ряд в усеченном решении может быть записан в виде

$$c_{i0} + c_{i1}x + \dots + c_{ik_i}x^{k_i} + p_i(u_1, \dots, u_l)x^{k_i+1} + O(x^{k_i+2}),$$

где

$u_1, \dots, u_l$  — некоторые из литералов, встречающихся в (10),

$c_{i0}, c_{i1}, \dots, c_{ik_i}$  — не зависящие от литералов константы,

$p_i(u_1, \dots, u_l)$  — не являющийся константой полином над  $K$  от литералов  $u_1, \dots, u_l$ ,  $i = 1, \dots, m$ .

К полиномам  $p_i(u_1, \dots, u_l)$ ,  $i = 1, \dots, m$ , можно применить лемму 1. Таким образом, существуют и могут быть найдены два различных набора (целых) значений  $\alpha_1, \dots, \alpha_l, \beta_1, \dots, \beta_l \in K$  для литералов  $u_1, \dots, u_l$ , с помощью которых строятся продолжения  $\mathcal{E}_1$  и  $\mathcal{E}_2$ , имеющие усеченные решения  $s_1$  и  $s_2$  с различающимися дополнительными членами  $p_i(\alpha_1, \dots, \alpha_l)x^{k_i+1}$  и  $p_i(\beta_1, \dots, \beta_l)x^{k_i+1}$  соответственно, не содержащими литералов. Отсюда следует утверждение теоремы.

## 6. РАСШИРЕНИЕ ВОЗМОЖНОСТЕЙ ПРОЦЕДУРЫ ПОСТРОЕНИЯ РЕШЕНИЙ

Построение контрпримера реализовано в системе компьютерной алгебры Maple 2021 как расширение возможностей процедуры *FormalSolution* из пакета *TruncatedSeries*. Этот пакет содержит наши реализации в Maple алгоритмов, представленных в [1]–[9], [22], [23]. Файлы Maple-библиотеки, содержащей пакет, и файлы Maple-сессий с примерами использования процедур пакета можно найти на странице [10].

Первый параметр процедуры *FormalSolution* — дифференциальное уравнение (5). Производная  $y(x)$  порядка  $i$  записывается стандартным для Maple образом:  $\text{diff}(y(x), x\$i)$ . Усеченные коэффициенты вида (6) записываются как  $a_i(x) + O(x^{t_i+1})$ , где  $a_i(x)$  — полином степени, не большей чем  $t_i$ , над полем алгебраических чисел.

Имя неизвестной функции задается во втором параметре процедуры.

Для работы с процедурами пакета необходимо загрузить архив *TruncatedSeries2021.zip*, расположенный на странице [10]. Этот архив содержит два файла: *maple.ind* и *maple.lib*. Необходимо

поместить эти файлы в некотором каталоге, например "/usr/userlib", и в Maple-сессии выполнить присваивание

```
> libname := " /usr/userlib", libname;
```

Следующая команда в сессии делает возможным обращение к процедурам пакета *TruncatedSeries* в короткой форме:

```
> with(TruncatedSeries);
```

Интерфейс системы Maple 2021 позволяет вводить уравнения в математической форме. Присвоим переменной *eq* выражение, обозначающее уравнение (1):

```
> eq := (1 + 3x + O(x^3)) y(x) + (x^3 + 1/3 x^5 + O(x^6)) (d/dx y(x)) = 0 :
```

В результате следующего вызова процедуры *FormalSolution* будет получено усеченное решение с максимальной степенью усечения:

```
> FormalSolution(eq, y(x), 'counterexample' = 'Eqs')
```

$$\left[ e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x^{\frac{1}{3}} (-c_1 + O(x)) \right]$$

и переменной *Eqs* будет присвоена пара уравнений, представляющая собой контрпример:

```
> Eqs[1]
```

$$(1 + 3x + x^3 + O(x^4)) y(x) + \left( x^3 + \frac{x^5}{3} + 4x^6 + O(x^7) \right) \left( \frac{d}{dx} y(x) \right) = 0$$

```
> Eqs[2]
```

$$(1 + 3x + O(x^4)) y(x) + \left( x^3 + \frac{x^5}{3} - 4x^6 + O(x^7) \right) \left( \frac{d}{dx} y(x) \right) = 0$$

Для уравнений контрпримера построим усеченные решения:

```
> FormalSolution(Eqs[1], y(x))
```

$$\left[ e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x^{\frac{1}{3}} (-c_1 + 4c_1 x + O(x^2)) \right]$$

```
> FormalSolution(Eqs[2], y(x))
```

$$\left[ e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{1}{3}} x^{\frac{1}{3}} (-c_1 - 3c_1 x + O(x^2)) \right]$$

Видно, что коэффициенты при *x* в рядах, входящих в эти решения, совпадают, только если оба эти решения нулевые.

Рассмотрим другое уравнение, второго порядка:

```
> eq := O(x^10) y(x) + (1 + 3x + O(x^3)) (d/dx y(x)) + (x^3 + x^5/3 + O(x^6)) (d^2/dx^2 y(x)) = 0 :
```

С помощью процедуры *FormalSolution* получаем экспоненциально-логарифмические решения, в которых регулярные части вычислены до максимально возможной степени:

```
> FormalSolution(eq, y(x))
```

$$\left[ -c_1 + O(x^{11}) + e^{\frac{1}{2x^2} + \frac{3}{x} - \frac{10}{3}} x^{\frac{10}{3}} (-c_2 + O(x)) \right] \quad (13)$$

Первые два слагаемых в (13), т.е.  $\_c_1 + O(x^{11})$ , означают, что все продолжения уравнения  $eq$  имеют лорановы решения с валюацией, равной 0, здесь их начальный отрезок до степени 10 равен  $\_c_1$ , где  $\_c_1$  — произвольная постоянная.

Третье слагаемое означает, что все продолжения уравнения  $eq$  имеют нерегулярные решения с показателем экспоненциальной части  $1/(2x^2) + 3/x$ , показателем  $\lambda = 10/3$ , начальным отрезком ряда  $\_c_2$ , где  $\_c_2$  — произвольная постоянная.

Если при вызове процедуры *FormalSolution* указан необязательный параметр 'output' = 'literal', то регулярные части решения вычисляются до максимальной степени и сверх этого еще добавляются слагаемые с коэффициентами, зависящими от литералов:

> *FormalSolution*(eq, y(x), 'output' = 'literal')

$$\_c_1 - \frac{U_{[0,10]}\_c_1 x^{11}}{11} + O(x^{12}) + e^{\frac{1}{2x^2} + \frac{3}{x}} x^{\frac{10}{3}} \left( \_c_2 + (-U_{[1,3]}\_c_2 + U_{[2,6]}\_c_2 - 2\_c_2)x + O(x^2) \right)$$

В нашей реализации литерал, коэффициент при  $x^k \theta^i$ , обозначается через  $U_{[i,k]}$ . Существует два таких целочисленных набора значений литералов, что выражения

$$-\frac{U_{[0,10]}\_c_1}{11} \quad \text{и} \quad -U_{[1,3]}\_c_2 + U_{[2,6]}\_c_2 - 2\_c_2$$

принимают различные значения. Этим двум наборам соответствуют два продолжения уравнения  $eq$ , которые составляют контрпример. Действительно, решения этих двух уравнений будут различными продолжениями решения (13) и продолжены будут все регулярные части решения. Очевидно, контрпримеров существует бесконечное число. В результате работы процедуры *FormalSolution* с необязательным параметром 'counterexample' = 'Eqs', переменной *Eqs* будет присвоена пара уравнений, представляющая собой искомым контрпример.

> *FormalSolution*(eq, y(x), 'counterexample' = 'Eqs') :

Для первого уравнения этого контрпримера

> *Eqs*[1]

$$(x^{10} + O(x^{11}))y(x) + (1 + 3x + 4x^3 + O(x^4))\left(\frac{d}{dx}y(x)\right) + \left(x^3 + \frac{x^5}{3} + O(x^7)\right)\left(\frac{d^2}{dx^2}y(x)\right) = 0$$

с помощью *FormalSolution* получаем усеченное решение

> *FormalSolution*(*Eqs*[1], y(x))

$$\left[ \_c_1 - \frac{\_c_1 x^{11}}{11} + O(x^{12}) + e^{\frac{1}{2x^2} + \frac{3}{x}} x^{\frac{10}{3}} \left( \_c_2 - 6\_c_2 x + O(x^2) \right) \right] \quad (14)$$

Для второго уравнения

> *Eqs*[2]

$$(-4x^{10} + O(x^{11}))y(x) + (1 + 3x + 5x^3 + O(x^4))\left(\frac{d}{dx}y(x)\right) + \left(x^3 + \frac{x^5}{3} - 2x^6 + O(x^7)\right)\left(\frac{d^2}{dx^2}y(x)\right) = 0$$

получаем

> *FormalSolution*(*Eqs*[2], y(x))

$$\left[ \_c_1 + \frac{4\_c_1 x^{11}}{11} + O(x^{12}) + e^{\frac{1}{2x^2} + \frac{3}{x}} x^{\frac{10}{3}} \left( \_c_2 - 9\_c_2 x + O(x^2) \right) \right] \quad (15)$$

Видно, что (14) и (15) являются продолжениями (13) и входящие в них усеченные ряды различны.

Авторы благодарят компанию Maplesoft (Ватерлоо, Канада) за консультации и дискуссии.

## СПИСОК ЛИТЕРАТУРЫ

1. *Abramov S., Khmel'nov D., Ryabenko A.* Laurent solutions of linear ordinary differential equations with coefficients in the form of truncated power series // Computer algebra: 3rd International Conference Materials, Moscow, June 17–21, 2019, International Conference Materials. P. 75–82.
2. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Линейные обыкновенные дифференциальные уравнения и усеченные ряды // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 10. С. 66–77.
3. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Регулярные решения линейных обыкновенных дифференциальных уравнений и усеченные ряды // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 1. С. 4–17.
4. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Усеченные ряды и формальные экспоненциально-логарифмические решения линейных обыкновенных дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 10. С. 1664–1675.
5. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Усеченные ряды // Труды XII приокской научной конференции “Дифференциальные уравнения и смежные вопросы математики”, Коломна: ГСГУ, 2020. С. 8–19.
6. *Abramov S., Khmel'nov D., Ryabenko A.* Truncated and infinite power series in the role of coefficients of linear ordinary differential equations // Lecture Notes in Computer Science. 2020. V. 12291. P. 63–76.
7. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Процедуры поиска лорановых и регулярных решений линейных дифференциальных уравнений с усеченными степенными рядами в роли коэффициентов // Труды ИСП РАН. 2019. Т. 31. № 5. С. 233–248.
8. *Abramov S., Khmel'nov D., Ryabenko A.* The TruncatedSeries package for solving linear ordinary differential equations having truncated series coefficients // In: Maple in Mathematics Education and Research, Springer Nature Switzerland. 2021. P. 19–33.
9. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Процедуры поиска усеченных решений линейных дифференциальных уравнений с бесконечными и усеченными степенными рядами в роли коэффициентов // Программирование. 2021. № 2. С. 56–65.
10. TruncatedSeries website: <http://www.ccas.ru/ca/TruncatedSeries>
11. *Abramov S., Khmel'nov D., Ryabenko A.* On truncated series involved in exponential-logarithmic solutions of truncated LODEs // In: Boulier, F., England, M., Sadykov, T.M., Vorozhtsov, E.V. (eds) Computer Algebra in Scientific Computing, CASC 2022, Lecture Notes in Computer Science. V. 13366. Springer, Cham. P. 18–28.
12. *Коддингтон Э.А., Левинсон Н.* Теория обыкновенных дифференциальных уравнений. М.: Изд-во иностр. лит., 1958.
13. *Malgrange B.* Sur la réduction formelle des équations différentielles a singularités irrégulières. Université Scientifique et Médicale de Grenoble, 1979.
14. *Tournier E.* Solutions formelles d'équations différentielles. Le logiciel de calcul formel DESIR. Étude théorique et réalisation. Thèse d'Etat, Université de Grenoble, 1987.
15. *Barkatou M.* Rational Newton algorithm for computing formal solutions of linear differential equations // Lecture Notes in Computer Science. 1989. V. 358. P. 183–195.
16. *Брюно А.Д.* Асимптотики и разложения решений обыкновенного дифференциального уравнения // Успехи матем. наук. 2004. Т. 59. Вып. 3(357). С. 31–80.
17. *Баркату М., Ришар-Жюнг Ф.* Формальные решения линейных дифференциальных и разностных уравнений // Программирование. 1997. № 2. С. 24–42.
18. *Lutz D.A., Schäfke R.* On the identification and stability of formal invariants for singular differential equations // Linear Algebra And Its Applications. 1985. V. 72. P. 1–46.
19. *Frobenius G.* Integration der linearen Differentialgleichungen mit veränder Koeffizienten // J. für die reine und angewandte Mathematik. 1873. V. 76. P. 214–235.
20. *Heffter L.* Einleitung in die Theorie der linearen Differentialgleichungen. Leipzig: Teubner, 1894.
21. *Abramov S., Barkatou M.A., Pfluegel E.* Higher-order linear differential systems with truncated coefficients // In Proc. of CASC'2011. 2011. P. 10–24.
22. *Khmel'nov D., Ryabenko A., Abramov S.* Automatic confirmation of exhaustive use of information on a given equation // Computer algebra: 4th International Conference Materials, Moscow: MAKSPress, 2021. P. 69–72.
23. *Абрамов С.А., Рябенко А.А., Хмельнов Д.Е.* Исчерпывающее использование информации о дифференциальном уравнении с усеченными коэффициентами // Программирование. 2022. № 2. С. 63–72.

# ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 519.644

## АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ КУБАТУРНЫХ ФОРМУЛ НА СФЕРЕ В СИСТЕМАХ КОМПЬЮТЕРНОЙ АЛГЕБРЫ<sup>1)</sup>

© 2023 г. Р. Э. Байрамов<sup>1</sup>, Ю. А. Блинков<sup>1,2</sup>, И. В. Левичев<sup>1</sup>,  
М. Д. Малых<sup>1,3,\*</sup>, В. С. Мележик<sup>3</sup>

<sup>1</sup> 117198 Москва, ул. Миклухо-Маклая, 6, РУДН, Россия

<sup>2</sup> 410012 Саратов, ул. Астраханская, 83, Саратовский гос. ун-т им. Н.Г. Чернышевского, Россия

<sup>3</sup> 141980 Дубна, М.о., ОИЯИ, Россия

\*e-mail: malykh\_md@pfur.ru

Поступила в редакцию 24.04.2022 г.

Переработанный вариант 24.04.2022 г.

Принята к публикации 10.09.2022 г.

Задача об отыскании весов и узлов кубатурных формул заданного порядка на единичной сфере, инвариантных относительно групп вращения икосаэдра (задача А.С. Попова) исследуется аналитически в системах компьютерной алгебры. Алгоритм Попова сведения задачи к системе нелинейных уравнений реализован в известной системе компьютерной алгебры Sage. Показано, что в Sage трудности с исследованием полученной системы нелинейных алгебраических уравнений возникают, начиная с порядка аппроксимации, равного 23. Показано также, что задача Попова при таком порядке приводит к полиномиальному идеалу, базис Грёбнера для которого содержит многочлены с экстремально большими целыми коэффициентами, что делает ее весьма трудной для исследования стандартными инструментами, реализованными в Sage. Этот базис найден в нашей системе компьютерной алгебры – GInv, новая версия которой была передана в общественный доступ одним из авторов статьи в 2021 г. Это позволило далее полностью описать множество решений задачи Попова в Sage. Проведено сравнение найденных нами точных решений с решениями, найденными Поповым численно. Обсужден потенциал использования задачи Попова как тестовой задачи для систем, специализирующихся на вычислении базиса Грёбнера. Библ. 19.

**Ключевые слова:** базис Грёбнера, инволютивный базис, кубатурные формулы, группа вращения икосаэдра.

**DOI:** 10.31857/S0044466923010052, **EDN:** LFAUXJ

### 1. ВВЕДЕНИЕ

Задача об отыскании кубатурных формул, инвариантных относительно групп симметрии правильных многогранников, была рассмотрена в серии статей А.С. Попова, где были описаны алгоритмы отыскания кубатурных формул и их численная реализация для наиболее интересных групп, а именно, для группы вращений октаэдра с и без инверсии (см. [1], [2]) и группы вращений икосаэдра с и без инверсии (см. [3], [4]). Эти результаты были подытожены автором в [5]. Поэтому задачу об отыскании кубатурных формул заданного порядка, инвариантных относительно групп симметрий многогранников, мы будем называть задачей Попова. Эта задача была также предметом недавних исследований М. Грефа (см. [6]–[8]).

Напомним кратко, в чем состоит проблема. Пусть  $S$  – единичная сфера

$$x^2 + y^2 + z^2 = 1.$$

Рассмотрим интеграл

$$U(f) = \frac{1}{4\pi} \iint_S f ds,$$

<sup>1)</sup> Работа выполнена при финансовой поддержке РНФ (проект № 20-11-20257, получатели М.Д. Малых, В.С. Мележик, сведение алгоритма Попова к системе нелинейных уравнений на языке Sage) и Программы стратегического академического лидерства РУДН (получатель Ю.А. Блинков, исследование решений задачи Попова в Sage).

где  $ds$  — элемент поверхности единичной сферы. Под кубатурной формулой для этого интеграла понимают выражение вида

$$V(f) = \sum_{i=1}^N A_i f(s_i),$$

которое можно использовать для приближенного вычисления интеграла  $U(f)$ . Здесь  $s_i$  — точки на сфере  $S$ , именуемые узлами, а  $A_i$  — положительные числа, именуемые весами. Говорят, что кубатурная формула  $V$  имеет алгебраический порядок точности  $n$ , если равенство

$$U(f) = V(f)$$

верно для всех многочленов  $f \in \mathbb{Q}[x, y, z]$ , степень которых  $\partial f \leq n$ . Говорят, что кубатурная формула  $V$  инвариантна относительно группы симметрий  $\mathfrak{G}$  сферы, если относительно этой группы инвариантно множество ее узлов и

$$\sum_{i=1}^N A_i f(Ts_i) = \sum_{i=1}^N A_i f(s_i), \quad T \in \mathfrak{G}.$$

Хотя исторически кубатурные формулы вышли из задачи о приближенном вычислении интеграла  $U(f)$  на сфере, их приложения много шире. Дело в том, что численное исследование дифференциальных уравнений в той или иной двумерной области опирается на возможность введения на ней сетки (mesh), узлы которой распределены более менее равномерно. Однако, поскольку число правильных многогранников не бесконечно, введение на сфере такой равномерной сетки представляет известные сложности. Формулы, найденные Поповым, дают такое распределение узлов и активно используются в работах В.С. Мележика (см. [9]–[11]) для построения двумерного DVR базиса, который хорошо зарекомендовал себя в аппроксимации трехмерного стационарного уравнения Шрёдингера, моделирующего спектр атома водорода в сильных скрещенных электрическом и магнитных полях, и нестационарного уравнения Шрёдингера с тремя пространственными переменными, моделирующего генерацию высоких гармоник атомом водорода в сильном эллиптически поляризованном лазерном поле.

В работах А.С. Попова решение задачи Попова сводится к исследованию системы нелинейных алгебраических уравнений, алгоритм составления которой описан в его статьях. Затем эта задача решалась численно на вычислительной технике Сибирского суперкомпьютерного центра (см. [5]). М. Греф (см. [8]) тоже занимался численным исследованием возникающих систем нелинейных уравнений. Современные системы компьютерной алгебры, вообще говоря, в состоянии предложить богатый инструментарий для составления систем алгебраических уравнений и исследования множества их решения в различных полях. В настоящей работе задача Попова исследуется в доступных и свободных системах компьютерной алгебры: популярной системе общего назначения Sage (<https://www.sagemath.org>) и созданной совсем недавно системе GInv (<https://github.com/blinkovua/GInv>), ориентированной на исследование полиномиальных идеалов.

Исследования системы нелинейных уравнений в современных системах компьютерной алгебры требуют на вычисления базисов Грёбнера идеалов, порожденных левыми частями уравнений этих систем (см. [12]). Для вычисления базиса Грёбнера в Sage используется реализация алгоритма Бухбергера (см. [12]), пришедшая в эту систему из системы Singular. Алгоритм Бухбергера — старейший, базовая версия этого алгоритма оставляет много свободы в осуществлении вычислительного процесса, значительные улучшения получаются за счет реализации критериев уменьшения числа S-многочленов, подлежащих рассмотрению. В конце 1990-х гг. в качестве альтернативы алгоритму Бухбергера были предложены инволютивные алгоритмы (см. [13]–[16]), которые реализованы в системе GInv. По инициативе В.П. Гердта и Ю.А. Блинкова в 2005 г. был основан проект GInv (<http://inv.jinr.ru>), в рамках которого было разработано программное обеспечение для вычисления инволютивных базисов, написанное в виде C++ модуля для языка Python. Недавно эта система была существенно переработана одним из авторов настоящей статьи, новая версия находится в открытом доступе по адресу <https://github.com/blinkovua/GInv>. В новую версию системы GInv добавлены механизмы динамического перераспределения памяти, позволяющие существенно, до нескольких раз, ускорить вычисления. Предлагаемый подход принципиально отличается от других алгоритмов известных под общим названием “сборка мусора” (garbage collection) (см. [17], [18]) и основан на реализации объектно-ориентированного программирования в C++.

Разумеется, мы сравнили базисы Грёбнера, полученные в Sage и GInv для задачи Попова при  $n = 19$ , и убедились в их совпадении с точностью до редукции лишних базисных элементов. Этот тест, наряду со стандартными “синтетическими” тестами, убедил нас в корректности работы новой системы.

В теории мы можем вычислить базис Грёбнера любой системы за конечное число шагов, однако на практике на эти шаги очень часто не хватает ресурсов, в первую очередь объема памяти компьютера. В настоящее время десяток неизвестных дает ту границу, после пересечения которой нельзя быть уверенным в том, что ресурсов современных РС хватит. Задача Попова при порядке аппроксимации, равном 20, как раз лежит на этой границе, что придает исследованию задачи Попова в системах компьютерной алгебры особый интерес.

## 2. КУБАТУРНЫЕ ФОРМУЛЫ, ИНВАРИАНТНЫЕ ОТНОСИТЕЛЬНО ГРУППЫ ВРАЩЕНИЙ ИКОСАЭДРА

Для отыскания кубатуры, инвариантной относительно группы вращений икосаэдра, Попов (см. [3]) зафиксировал на сфере один икосаэдр, связал с ним первые точки сети, а кубатуру переписал в виде

$$V(f) = A_0 \sum_{j=1}^{12} f(a_{0j}) + B_0 \sum_{j=1}^{20} f(b_{0j}) + C_0 \sum_{j=1}^{30} f(c_{0j}) + \sum_{i=1}^M A_i \sum_{j=1}^{60} f(a_{ij}), \quad (1)$$

где 12 точек  $a_{0j}$  лежат в вершинах вписанного в сферу икосаэдра, 20 точек  $b_{0j}$  отвечают центрам граней икосаэдра, 30 точек  $c_{0j}$  отвечают серединам ребер икосаэдра, а 60 точек  $a_{ij}$  отвечают точкам общего положения на гранях икосаэдра, точки  $a_{ij}$  получаются из  $a_{i1}$  путем вращения икосаэдра. При таком выборе условие инвариантности квадратурной формулы относительно группы вращений икосаэдра выполняется автоматически.

В силу теоремы Соболева (см. [3, р. 435]), для того чтобы кубатурная формула (1) имела порядок  $n$ , необходимо и достаточно, чтобы она была точна для всех инвариантных относительно группы вращений многочленов степени не выше  $n$ . Опираясь на результаты Клейна (см. [19]), все эти инвариантные многочлены можно описать как многочлены относительно трех многочленов, которые в работе Попова обозначены как  $u$ ,  $v$  и  $w$ . Мы не будем выписывать здесь громоздкие выражения для этих многочленов, отметим лишь, что многочлены эти не независимы, но связаны алгебраическим уравнением

$$w^2 = -16v^2 + 40u^2v - 25u^4 + 20uv^2 - 45u^3v + 27u^5 - v^3. \quad (2)$$

Поскольку  $u(a_{ij})$  не зависит от  $j$ , мы можем обозначить это значение как  $u_i$ . Приняв аналогичные обозначения для  $v$  и  $w$ , заключаем: кубатурная формула имеет порядок  $n$  тогда и только тогда, когда

$$V(u^p v^q w^r) = 12A_0 u_0^p v_0^q w_0^r + \dots + 60 \sum_{i=1}^M A_i u_i^p v_i^q w_i^r$$

для всех выражений  $u^p v^q w^r$ , степень которых относительно  $x$ ,  $y$ ,  $z$  не превосходит  $n$ . Левые части этих равенств нетрудно вычислить явно. Поэтому получается система алгебраических уравнений относительно весов и значений инвариантных многочленов в узлах сети. Таким образом, для их отыскания получается система нелинейных алгебраических уравнений, которую мы далее будем называть *системой* Попова.

На основе этих соображений Попов разработал и описал в [3] алгоритм, при реализации которого в системе Sage мы не столкнулись с трудностями, достойными упоминания. В [3] приведены результаты вычислений для порядков  $n$  с 19 до 25 в виде таблиц с десятичными дробями, случай  $n = 19$  исследован аналитически.

### 2.1. Кубатура при $n = 19$

Кубатура  $n = 19$ ,  $M = 2$ ,  $B_0 = C_0 = 0$  была исследована Поповым аналитически. Наша функция `popoff_eqs(n)` возвращает список, первым элементом которого служит список всех иско-

мых величин (весов и значений инвариантов), затем следует список всех уравнений системы. Для  $n = 19$  мы имеем:

```
sage: load("popoff.sage")
sage: popoff_eqs(19)
[[A0, A1, A2, u1, u2, v1, v2, w1, w2],
 [12*A0 + 60*A1 + 60*A2 - 1,
  60*A1*w1 + 60*A2*w2,
  60*A1*v1 + 60*A2*v2 - 256/231,
  60*A1*u1 + 60*A2*u2 - 16/21,
  60*A1*u1*v1 + 60*A2*u2*v2 - 18432/17017,
  60*A1*u1^2 + 60*A2*u2^2 - 2048/3003,
  60*A1*u1^3 + 60*A2*u2^3 - 8192/12597,
  27*u1^5 - 25*u1^4 - 45*u1^3*v1 + 40*u1^2*v1 + 20*u1*v1^2 - v1^3
- 16*v1^2 - w1^2,
  27*u2^5 - 25*u2^4 - 45*u2^3*v2 + 40*u2^2*v2 + 20*u2*v2^2 - v2^3
- 16*v2^2 - w2^2]]
```

В Sage базис Грёбнера идеала  $J$ , порожденный этими уравнениями (deglex-порядок на мономах принят по умолчанию), вычисляется без каких-либо затруднений и существенных затрат времени:

```
sage: eqs=popoff_eqs(19)
sage: K=PolynomialRing(QQ, eqs[0])
sage: J=K*eqs[1]
sage: J.groebner_basis()
[v2^2 - 3038016/1282633*v2 + 197738496/239852371,
 v2*w1 - 22131456/1115081*w1 - 20793600/1115081*w2,
 w1^2 - 282712985600/93773402523273*v2
- 17460378265190400/331341213770339443,
 v2*w2 + 20793600/1115081*w1 + 12153087936/695303689*w2,
 w1*w2 + 910295040000/16191714009097,
 w2^2 + 282712985600/93773402523273*v2 - 349759628902400/5845208757284017,
 A0 - 125/19656,
 A1 + 85936411/416549952000*v2 - 2153726/271191375,
 A2 - 85936411/416549952000*v2 - 5389889/723177000,
 u1 + 19/60*v2 - 403344/337535,
 u2 - 19/60*v2 - 416/935,
 v1 + v2 - 3038016/1282633]
```

Отметим, что в данном случае “упрощение” системы путем перехода к целым коэффициентам не только не требуется, но и несколько затягивает вычисления. Интересно отметить и огромные коэффициенты, которые появляются в базисе уже при  $n = 19$ .

Быстрота вычисления базиса Грёбнера позволяет без хлопот вычислить размерность идеала, которая равна нулю. После этого без ошибок можно применить к идеалу метод `variety` и получить множество решений системы уравнений Попова над полем вещественных алгебраических чисел  $\mathbb{A}$  в виде списка  $L$ :

```
sage: L=J.variety(AA)
```

Таким образом, при  $n = 19$  система Попова имеет четыре решения. У всех четырех вес  $A_0$  имеет одно и то же рациональное значение

$$A_0 = \frac{125}{19656},$$

указанное Поповым. Это число нетрудно найти в Sage по минимальному многочлену для алгебраического числа  $A_0$ :

```
sage: L[0][A0].minpoly()
x - 125/19656
```

Два других веса  $A_1$  и  $A_2$  являются корнями одного и того же квадратного уравнения:

```
sage: L[0][A1].minpoly()
x^2 - 1513/98280*x + 42093683/710738355600
sage: L[0][A2].minpoly()
x^2 - 1513/98280*x + 42093683/710738355600
```

Поэтому во всех четырех решениях веса имеют одни и те же значения с точностью до перестановки  $A_1$  и  $A_2$ . Эти значения указаны в таблице Попова как одно решение, сами значения совпадают до последнего знака, указанного в таблице Попова.

В данном случае кубатура (1) имеет вид

$$V(f) = A_0 \sum_{j=1}^{12} f(a_{0j}) + A_1 \sum_{j=1}^{60} f(a_{1j}) + A_2 \sum_{j=1}^{60} f(a_{2j}).$$

Поэтому из одного решения системы можно получить второе путем одновременной перестановки весов  $A_1$  и  $A_2$  и значений инвариантов  $u_1, v_1, w_1$  и  $u_2, v_2, w_2$ . Куда менее очевидно, что из одного решения системы можно получить второе путем одновременного изменения знаков у  $w_1$  и  $w_2$ . Это преобразование соответствует отражению на сфере, т.е. получающиеся при этом узлы будут совпадать с точностью до отражения. В итоге получается четыре различных решения системы Попова, которые в таблице Попова отмечены как одно.

## 2.2. Кубатура при $n = 20$

При  $n = 20$  мы имеем  $M = 2$  и  $C_0 = 0$ . Система Попова имеет вид

```
sage: popoff_eqs(20)
[[A0, B0, A1, A2, u1, u2, v1, v2, w1, w2],
 [12*A0 + 60*A1 + 60*A2 + 20*B0 - 1,
  60*A1*w1 + 60*A2*w2, 60*A1*v1 + 60*A2*v2 + 5120/81*B0 - 256/231,
  60*A1*v1^2 + 60*A2*v2^2 + 1310720/6561*B0 - 1900544/969969,
  60*A1*u1 + 60*A2*u2 + 640/27*B0 - 16/21,
  60*A1*u1*v1 + 60*A2*u2*v2 + 163840/2187*B0 - 18432/17017,
  60*A1*u1^2 + 60*A2*u2^2 + 20480/729*B0 - 2048/3003,
  60*A1*u1^3 + 60*A2*u2^3 + 655360/19683*B0 - 8192/12597,
  27*u1^5 - 25*u1^4 - 45*u1^3*v1 + 40*u1^2*v1 + 20*u1*v1^2 - v1^3
- 16*v1^2 - w1^2,
  27*u2^5 - 25*u2^4 - 45*u2^3*v2 + 40*u2^2*v2 + 20*u2*v2^2 - v2^3
- 16*v2^2 - w2^2]]
```

Над полем вещественных алгебраических чисел  $\mathbb{A}$  эта система имеет четыре решения. У всех четырех вес  $A_0$  имеет одно и то же значение, которое совпадает с указанным в таблице Попова, но, в отличие от случая  $n = 19$ , не является рациональным числом:

```
sage: L[0][A0].minpoly()
x^2 - 7375/453024*x + 1328125/20777492736
```

На самом деле, мы ожидали, что старший вес  $A_0$  всегда является рациональным числом, поскольку в сумме (1) он занимает особое положение, в то время как младшие веса  $A_1$  и  $A_2$  можно менять местами.

Второму вещественному корню этого квадратного уравнения отвечают мнимые значения инвариантов. Вообще говоря, в теории квадратур и кубатур предполагается, что веса положительные, а узлы — вещественные. Например, при замене интеграла на отрезке на сумму Дарбу выбирают узлы на этом отрезке, однако это условие не является обязательным. Мы рассматриваем задачи, в которых под интегралом стоит многочлен и в силу теоремы Коши путь интегрирования можно деформировать на комплексной плоскости. Поэтому квадратурные и кубатурные формулы с комплексными узлами имеют право на существование. Но, конечно, в задаче Попова речь шла о решениях с положительными весами и узлами на вещественной сфере.

Два других веса  $A_1$  и  $A_2$  являются корнями одного и того же уравнения 4-й степени:

```
sage: L[0][A1].minpoly()
x^4 - 11885261/482017536*x^3 + 202915177968491/1033960700403671040*x^2
- 2389552908217687/4766422781400344064000*x
- 26951199206199859/610712218135263284232192000
sage: L[0][A2].minpoly() x^4
- 11885261/482017536*x^3 + 202915177968491/1033960700403671040*x^2
- 2389552908217687/4766422781400344064000*x
- 26951199206199859/610712218135263284232192000
```

Эти уравнения имеют четыре вещественных корня, но лишь два из них дают вещественные значения для инвариантов. Структура решения над  $\mathbb{A}$  та же, что и при  $n = 19$ : из одного решения получаются другие путем перестановки индексов 1 и 2 и путем одновременной смены знаков у  $w_1$  и  $w_2$ .

### 2.3. Кубатура при $n = 23$

Обратимся теперь к случаю  $n = 23$ , когда  $M = 3$ ,  $B_0 = C_0 = 0$ , как и при  $n = 19$ . Мы ожидали, что этот случай должен быть в чем-то схож с простейшим случаем  $n = 19$ :

```
sage: load("popoff.sage")
sage: popoff_eqs(23)
[[A0, A1, A2, A3, u1, u2, u3, v1, v2, v3, w1, w2, w3],
 [12*A0 + 60*A1 + 60*A2 + 60*A3 - 1,
  60*A1*w1 + 60*A2*w2 + 60*A3*w3,
  60*A1*v1 + 60*A2*v2 + 60*A3*v3 - 256/231,
  60*A1*v1^2 + 60*A2*v2^2 + 60*A3*v3^2 - 1900544/969969,
  60*A1*u1 + 60*A2*u2 + 60*A3*u3 - 16/21,
  60*A1*u1*w1 + 60*A2*u2*w2 + 60*A3*u3*w3,
  60*A1*u1*v1 + 60*A2*u2*v2 + 60*A3*u3*v3 - 18432/17017,
  60*A1*u1^2 + 60*A2*u2^2 + 60*A3*u3^2 - 2048/3003,
  60*A1*u1^2*v1 + 60*A2*u2^2*v2 + 60*A3*u3^2*v3 - 8126464/7436429,
  60*A1*u1^3 + 60*A2*u2^3 + 60*A3*u3^3 - 8192/12597,
  27*u1^5 - 25*u1^4 - 45*u1^3*v1 + 40*u1^2*v1 + 20*u1*v1^2 - v1^3
- 16*v1^2 - w1^2,
  27*u2^5 - 25*u2^4 - 45*u2^3*v2 + 40*u2^2*v2 + 20*u2*v2^2 - v2^3
- 16*v2^2 - w2^2,
  27*u3^5 - 25*u3^4 - 45*u3^3*v3 + 40*u3^2*v3 + 20*u3*v3^2 - v3^3
- 16*v3^2 - w3^2]]
```

В системе Sage нам не удалось найти базис Грёбнера этой системы. Эффективность алгоритмов отыскания базиса существенно зависит от порядка следования неизвестных. Мы пробовали различные варианты, но, конечно, не перебрали все 13(!) вариантов.

Процессы отыскания базиса Грёбнера при  $\deglex$  порядке на мономах и

$$A_0 > A_1 > A_2 > A_3 > u_1 > u_2 > u_3 > v_1 > v_2 > v_3 > w_1 > w_2 > w_3$$

мы запустили параллельно в двух системах Sage и GInv на компьютере “Научного центра вычислительных методов в прикладной математике” РУДН. С задачей успешно справился только GInv, вернув нам базис  $B$ , состоящий из 107 многочленов небольшой степени (degree), но с огромными целыми коэффициентами порядка  $10^{161}$ . Среди них два многочлена были линейными, 42 многочлена — квадратичными и 63 многочлена — кубическими.

Поскольку на данном этапе система GInv не имеет разработанного инструментария для дальнейшего использования найденного базиса, мы передали его в Sage и определили идеал  $J$ . Работа с идеалом  $J$ , порожденным 107 многочленами с 13 переменными и огромными целыми коэффициентами, является экстремально трудной. Часто применение к этому идеалу стандартных методов приводило к долгим вычислениям, которые мы прерывали через несколько минут. Тем не менее несколько методов позволили полностью решить задачу Попова.

Прежде всего мы решили проверить, является ли найденный базис базисом Грёбнера с помощью стандартного метода `basis_is_groebner`, однако за несколько минут ответа не получили и прервали процесс. Затем мы решили вычислить базис Грёбнера идеала  $J$  при том же мономиальном порядке, что и в GInv, с помощью метода `groebner_basis` и достигли успеха буквально за несколько секунд:

```
sage: J.groebner_basis()
Polynomial Sequence with 83 Polynomials in 13 Variables
```

В Sage по умолчанию выводится редуцированный базис, поэтому уменьшения числа элементов и следовало ожидать.

Применение к идеалу метода `variety` вновь приводит к излишне затратным вычислениям, которые были прерваны. Однако нам удалось обойти эту трудность, поскольку в Sage за несколько секунд нам удалось вычислить базис Грёбнера относительно  $lex$  порядка, т.е. фактически пе-

рейти от базиса Грёбнера относительно  $\deg_{\text{lex}}$  мономиального порядка, вычисленному в  $G_{\text{Inv}}$ , к базису Грёбнера относительно  $\text{lex}$  порядка. Мы взяли такой порядок, чтобы  $A_0$  был младшим:

```
sage: K=PolynomialRing(QQ, [w1, w2, w3, u1, u2, u3, v1, v2, v3,
    A1, A2, A3, A0], order='lex')
sage: Jlex=Klex*B
```

Базис Грёбнера при таком выборе мономиального порядка имеет всего 13 элементов:

```
sage: Blex=Jlex.groebner_basis()
sage: len(Blex)
13
```

Напомним, что неизвестных у нас тоже 13.

Базис Грёбнера при  $\text{lex}$  порядке представляет собой систему многочленов, приведенную к треугольному виду. Разберем этот базис с конца. Последней элемент базиса дает многочлен 3-й степени для  $A_0$ :

```
sage: Blex[-1]
A0^3 - 453214375/25273614366*A0^2
+ 6016015625/57483863813376*A0
- 2090576171875/10608532099030914048
```

Это многочлен имеет три вещественных корня:

```
sage: QQ[A0](Blex[-1]).roots(AA)
[(0.004164880157580243?, 1),
 (0.006620018869643640?, 1),
 (0.007147414429679166?, 1)]
```

Один из них, наименьший, совпадает с точностью до последнего знака со значением веса  $A_0$  в таблице Попова.

Следующий многочлен

```
sage: Blex[-2]
A3^3 + 1/5*A3^2*A0 - 1/60*A3^2 - ...
```

содержит  $A_3$  и  $A_0$  и позволяет для любого из трех значений  $A_0$  найти по три значения  $A_3$ . При всех трех возможных значениях  $A_0$  этот многочлен относительно  $A_3$  имеет вещественные корни:

```
sage: [(A01,a),(A02,b),(A03,c)]=QQ[A0](Blex[-1]).roots(AA)
sage: AA[A3](Blex[-2].subs(A0 1)).roots()
[(0.0050776793060011?, 1),
 (0.0053495468022166?, 1),
 (0.0054064645269329?, 1)]
sage: AA[A3](Blex[-2].subs(A0 2)).roots()
[(0.001590837633796970?, 1),
 (0.006614776888519542?, 1),
 (0.007137048370421428?, 1)]
sage: AA[A3](Blex[-2].subs(A0 3)).roots()
[(0.0002487875296393600?, 1),
 (0.007210715954234012?, 1),
 (0.007777680296857461?, 1)]
```

Вообще говоря, далее следовало бы рассматривать девять случаев и из следующих двух уравнений найти  $A_2$ ,  $A_1$ . Однако задача Попова имеет естественную симметрию: если взять  $i, j > 0$  и на решении переставить

$$A_i, u_i, v_i, w_i$$

и

$$A_j, u_j, v_j, w_j,$$

то значение суммы (1) не изменится, т.е. получится опять решение задачи Попова. Как отмечалось выше, в таблице Попова указывается одно решение по модулю этих перестановок. В данном случае эта симметрия ведет к тому, что веса  $A_1$  и  $A_2$  удовлетворяют тому же уравнению 3-й степени, что и  $A_3$ , причем  $A_1$ ,  $A_2$  и  $A_3$  являются тремя различными корнями уравнения

$$A^3 + \frac{1}{5}A_0A^2 - \frac{1}{60}A^2 - \dots = 0, \quad (3)$$

где в качестве  $A_0$  может быть один из трех корней уравнения

$$A_0^3 - \frac{453214375}{25273614366} A_0^2 + \frac{6016015625}{57483863813376} A_0 - \frac{2090576171875}{10608532099030914048} = 0. \quad (4)$$

Следующие шесть многочленов, с пятого до десятого с конца, — линейные относительно  $u_1, u_2, u_3, v_1, v_2, v_3$ , они позволяют выразить эти шесть переменных через найденные выше веса.

Одиннадцатый с конца многочлен дает уравнение вида

$$w_3^2 = f(A_0, A_3).$$

Имеет ли оно вещественный корень или нет, зависит от знака многочлена  $f(A_0, A_3)$ . Подставляя вместо  $A_0$  и  $A_3$  корни выписанных выше уравнений, мы видим, что вещественное значение для  $w_3$  получается в двух случаях: когда в качестве  $A_0$  берется наименьший или наибольший корень. При среднем корне получается мнимое значение.

Первые два многочлена базиса — линейные относительно  $w_1$  и  $w_2$  соответственно, из них можно однозначно и без введения комплексных чисел найти  $w_1$  и  $w_2$ .

Таким образом, над  $\mathbb{A}$  множество решений устроено следующим образом. Вес  $A_0$  равен наибольшему или наименьшему из трех корней уравнения 3-й степени (4). Веса  $A_1, A_2, A_3$  являются тремя корнями уравнения (3). По ним однозначно определяются вещественные значения инвариантов  $u$  и  $v$ , а значения  $w$  — с точностью до знака. Всего, с точностью до перестановок, два корня: корень  $c$

$$A_0 = 0.004164880157580243,$$

указанный в таблице Попова, и корень

$$A_0 = 0.007147414429679166,$$

в таблице отсутствующий. Таким образом, в случае  $n = 23$  задача Попова допускает два решения, различающихся весом первого члена. Численно из этих решений было найдено лишь одно, отвечающее наименьшему весу.

### 3. ЗАКЛЮЧЕНИЕ

Задача Попова дает в наше распоряжение набор систем нелинейных алгебраических уравнений, сложность решения которых растет вместе с ростом порядка аппроксимации  $n$ . Мы имеем численные решения этой системы в интервале от  $n = 19$  до  $n = 25$ . Случай  $n = 19$  был исследован аналитически Поповым на руках, случай  $n = 20$  успешно исследован в системе Sage. При этом, вопреки нашим ожиданиям, старший вес  $A_0$  оказался иррациональным числом. Сравнение аналитики и численных расчетов подтверждает, что численные расчеты Попова выполнены с очень высокой точностью. Уравнения Попова при  $n = 23$  являются слишком сложными для того, чтобы их можно было исследовать стандартными средствами Sage/Singular, зато с ней справляется инволюционный алгоритм, реализованный в GInv. В нашем распоряжении имеются численные решения, поэтому задача Попова является превосходным тестом, позволяющим оценить эффективность различных алгоритмов вычисления базиса Грёбнера.

Применение инволютивных алгоритмов позволило описать все множество решений задачи Попова при  $n = 23$  и, в частности, найти ранее неизвестное решение.

Следует особо подчеркнуть, что этот тест не является искусственным, специально придуманным для тестирования систем компьютерной алгебры, а результаты сами по себе представляют научный интерес. Нельзя не отметить значительное отличие параметров, характеризующих вычисление базиса Грёбнера, при исследовании задачи Попова и стандартных тестов (<http://inv.jinr.ru/ginv/benchmark.html>): коэффициенты базисных многочленов оказываются огромными, а степени, напротив, остаются небольшими.

### СПИСОК ЛИТЕРАТУРЫ

1. *Popov A.S.* The search for the sphere of the best cubature formulae invariant under octahedral group of rotations // Siberian J. of Numer. Math. / Sib. Branch of Russ. Acad. of Sci. 2002. V. 5. № 4. P. 367–0372.

2. *Popov A.S.* The search for the best cubature formulae invariant under the octahedral group of rotations with inversion for a sphere // *Siberian J. of Numer. Math. / Sib. Branch of Russ. Acad. of Sci.* 2005. V. 8. № 2. P. 143–148.
3. *Popov A.S.* Cubature formulas on a sphere invariant under the icosahedral rotation group // *Numer. Analys. Appl.* 2008. V. 1. P. 355–361.
4. *Popov A.S.* Cubature Formulas Invariant under the Icosahedral Group of Rotations with Inversion on a Sphere // *Numer. Analys. Appl.* 2017. V. 10. P. 339–346.
5. *Popov A.S.* Cubature formulas on a sphere invariant under the symmetry groups of regular polyhedrons // *Siberian Electron. Math. Rep.* 2017. V. 14. P. 190–198.
6. *Gräf M., Potts D.* Sampling sets and quadrature formulae on the rotation group // *Numer. Funct. Anal. and Optimizat.* 2009. V. 30. № 7–8. P. 665–688.
7. *Gräf M., Kunis St., Potts D.* On the computation of nonnegative quadrature weights on the sphere // *Appl. and Comput. Harmonic Anal.* 2009. V. 27. № 1. P. 124–132.
8. *Gräf M.* Efficient Algorithms for the computation of Optimal Quadrature Points on Riemannian Manifolds : Ph. D. thesis / Manuel Gräf; Fakultät für Mathematik der Technischen Universität Chemnitz. Chemnitz, 2013. P. 2.
9. *Melezhik V.S.* New method for solving multidimensional scattering problem // *J. of Comput. Phys.* 1991. V. 92. № 1. P. 67–81. Access mode: <https://www.sciencedirect.com/science/article/pii/002199919190292S>.
10. *Shadmehri S., Saeidian Sh., Melezhik V.S.* 2D nondirect product discrete variable representation for Schrödinger equation with nonseparable angular variables // *J. of Phys. B: Atomic, Molecular and Optic. Phys.* 2020. V. 53. № 8. P. 085001.
11. *Melezhik V.S.* Improving efficiency of sympathetic cooling in atom- ion and atom-atom confined collisions // *Phys. Rev. A.* 2021. May. V. 103. P. 053109. Access mode: <https://link.aps.org/doi/10.1103/PhysRevA.103.053109>.
12. *Cox D., Little J., O'Shea D.* Ideals, varieties, and algorithms. 3 ed. Springer, 2007.
13. *Zharkov A. Yu., Blinkov Yu.A.* Involution approach to solving systems of algebraic equations // *Proceed. of the 1993 Inter. IMACS Symp. on Symbolic Comput.* Laboratoire d'Informatique Fondamentale de Lille, France, 1993. P. 11–16.
14. *Zharkov A. Yu., Blinkov Yu.A.* Solving zero-dimensional involutive systems // *Progress in Math.* Basel: Birkhauser, 1996. V. 143. P. 389–399.
15. *Blinkov Yu.A.* Division and algorithms in the ideal membership problem [Deleniye i algoritmy v zadache o pri-nadlezhnosti k idealu] // *Izvestija Saratovskogo universiteta.* 2001. V. 1. № 2. P. 156–167. In Russ.
16. *Apel J.* A Gröbner approach to involutive bases // *J. of Symbolic Comput.* 1995. V. 19. № 5. P. 441–458.
17. *McCarthy J.* Recursive functions of symbolic expressions and their computation by machine, Part I // *Communicat. of ACM.* 1960. V. 3. № 4. P. 184–195.
18. *Jones R., Hosking A., Moss J., Eliot B.* The garbage collection handbook: the a of automatic memory management. CRC Applied Algorithms and Data Structures Ser. Chapman and Hall / CRC Press / Taylor & Francis Ltd., 2011. ISBN: 978-1-4200-8279-1.
19. *Клейн Ф.* Лекции об икосаэдре и решении уравнений пятой степени. М.: Наука, 1989.

## УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.928

### СХОДИМОСТЬ ФОРМАЛЬНЫХ РЕШЕНИЙ ВТОРОГО ЧЛЕНА ЧЕТВЕРТОЙ ИЕРАРХИИ ПЕНЛЕВЕ В ОКРЕСТНОСТИ НУЛЯ<sup>1)</sup>

© 2023 г. В. И. Аношин<sup>1</sup>, А. Д. Бекетова<sup>1</sup>,  
А. В. Парусникова<sup>1,\*</sup>, Е. Д. Прокопенко<sup>1</sup>

<sup>1</sup> 123458 Москва, Таллинская ул., 34, Национальный исследовательский университет  
“Высшая школа экономики”, Россия

\*e-mail: [aparusnikova@hse.ru](mailto:aparusnikova@hse.ru)

Поступила в редакцию 04.08.2022 г.  
Переработанный вариант 04.08.2022 г.  
Принята к публикации 10.09.2022 г.

Рассматривается второй член иерархии четвертого уравнения Пенлеве. Доказана сходимость некоторых степенных асимптотических разложений в окрестности нуля. Найдены новые семейства степенных асимптотических разложений. Вычисления проводятся с использованием пакета компьютерной алгебры. Дана ссылка на код, который может быть использован для вычисления порядка Жевре формального разложения решения дифференциального уравнения второго порядка в пакете символьных вычислений. Библ. 12. Фиг. 6.

**Ключевые слова:** асимптотические разложения, порядки Жевре, уравнения Пенлеве, символьные вычисления.

**DOI:** 10.31857/S0044466923010040, **EDN:** LKQBOT

## 1. ВВЕДЕНИЕ

Существует ряд способов построения иерархий Пенлеве (см. [1], [2]) — аналогов уравнений Пенлеве более высоких порядков. В данной работе рассматривается второй член четвертой иерархии Пенлеве из [2] — обыкновенное дифференциальное уравнение четвертого порядка, продолжающее цепочку, соответствующую четвертому уравнению Пенлеве. Решения уравнений — членов иерархий уравнений Пенлеве могут, как это имеет место для исходных уравнений Пенлеве, определять новые специальные функции; впрочем, доказательство справедливости этого факта или опровергающие такое утверждение результаты на данный момент авторам неизвестны.

В [3] методами степенной геометрии [4] найдены степенные формальные асимптотические разложения решений второго члена четвертой иерархии Пенлеве. Здесь, применяя теорию из [5], [6], доказываем сходимость полученных в окрестности нуля разложений и отсутствие экспоненциальных добавок (см. [7]). Для поиска экспоненциальных добавок к решениям используется код, написанный в пакете символьных вычислений (код для реализации первых шагов методами степенной геометрии приведен в [3]). Дана ссылка на программу для символьных вычислений порядка Жевре формального разложения решения дифференциального уравнения второго порядка.

## 2. МЕТОДЫ

Рассмотрим обыкновенное дифференциальное уравнение порядка  $n$  вида

$$F(x, y, y', \dots, y^{(n)}) = 0, \quad (1)$$

<sup>1)</sup> Работа выполнена при финансовой поддержке РНФ (проект № 19-71-10003), <https://rscf.ru/project/19-71-10003/>.

где  $F$  — аналитическая функция  $n + 2$  переменных и формально удовлетворяющий ему в окрестности бесконечности степенной ряд

$$\hat{y} = \sum_{k=k_0}^{\infty} \frac{a_k}{x^k}. \quad (2)$$

Известно, что каждому формальному решению вида (2) уравнения (1) соответствует аналитическое решение  $y$  уравнения (1), асимптотически приближаемое по Жевре рядом (2) в некотором секторе с вершиной в бесконечности, т.е. на множестве  $V = \{x : |x| > R, \operatorname{Arg} x \in (\varphi_1, \varphi_2)\}$  при некотором  $R > 0$ .

Для доказательства сходимости ряда (2) можно использовать асимптотическую теорию Жевре. В случае, когда сходимость таким методом не может быть доказана, осмысленным для определения точности приближения настоящего решения уравнения (1) вновь является применение асимптотической теории Жевре, а также вычисление уточняющих экспоненциальных добавок (экспоненты, в показателях которых записан формальный степенной ряд) к решению.

Кратко опишем алгоритм поиска порядков Жевре и точности приближения по Жевре разложения (см. [5], [6]):

I.1. Вычислить производную Фреше  $L$  левой части уравнения (1) на решении  $\hat{y}$ ;

I.2. Проверить, содержит ли  $L$  оператор  $\frac{d^n}{dx^n}$  ( $n$  — порядок уравнения (1));

I.3. Выразить в  $L$  операторы  $\frac{d^k}{dx^k}$  через операторы  $D^s$ ,  $s = 1, \dots, k$ , где  $D = x \frac{d}{dx}$ , т.е. переписать  $L$  в виде

$$L = \sum_{k=0}^n a_k(x) D^k, \quad \text{где} \quad a_k \in \mathbb{C}[x][[1/x]], \quad a_k(x) = \sum_{j=j_k}^{\infty} a_{j,k} x^{-j}; \quad (3)$$

I.4. Построить многоугольник Ньютона оператора  $L$  (3) как выпуклую оболочку точек  $\{(q_1, q_2) : 0 \leq q_1 \leq k, q_2 \geq j_{k,0}\}$ ;

I.5. Определить тангенсы  $k_i$  угла наклона границы многоугольника Ньютона к оси абсцисс;

I.6. Множество  $\{0, 1/k_1, \dots, 1/k_r\}$  содержит все возможные порядки Жевре ряда (2).

Порядок Жевре 0 соответствует сходимости ряда (2). Сразу гарантирует сходимость многоугольник Ньютона с тремя сторонами, одна из которых — горизонтальный отрезок, две оставшиеся стороны — вертикальные полупрямые.

Теперь опишем алгоритм поиска экспоненциальных добавок к решению методами степенной геометрии (см. [7]):

II.1. Вычислить производную Фреше  $L$  левой части уравнения (1) на решении  $\hat{y}$ ;

II.2. Применить  $L$  к  $u$ ;

II.3. Сделать в (линейном) уравнении  $Lu = 0$  замену

$$\ln u = w, \quad (4)$$

сократить на  $u$ , получить дифференциальное уравнение  $G(x, w) = 0$ ;

II.4. Построить многоугольник Ньютона полученного уравнения по правилам степенной геометрии [4]. Поскольку подставляли ряд, многоугольник — в общем случае — не является ограниченным множеством;

II.5. Рассмотреть ребра многоугольника, внешняя нормаль  $N = (p_1, p_2)$  к которым имеет  $p_1 > 0$ ,  $p_2 > 0$ . Найти степенные разложения решений уравнения  $G(z, w) = 0$ , соответствующие укороченным уравнениям этих ребер;

II.6. Сделать обратную замену  $u = e^w$ .

**Замечание 1.** Мы рассмотрели формальный асимптотический ряд в окрестности бесконечности. В работе речь идет о формальных степенных разложениях решений ОДУ вблизи нуля. Для применения алгоритма I делаем замену  $x = 1/t$ , переводя ноль в бесконечность. При приме-

нии алгоритма II замену делать не надо, следует на шаге II.5 рассмотреть ребра многоугольника, внешняя нормаль  $N = (p_1, p_2)$  к которым имеет  $p_1 < 0$ ,  $p_2 > 0$ .

### 3. СРАВНЕНИЕ МНОГОУГОЛЬНИКОВ

В предыдущем разделе появились два многоугольника Ньютона; опишем, как они связаны. Их преобразование друг в друга корректирует сделанное в [8] замечание об их связи, поясним это.

**Утверждение 1.** Точки, используемые для построения многоугольника Ньютона, употребляемого для определения порядков Жевре по линейному дифференциальному оператору, получаются как подмножество точек носителя многоугольника Ньютона, применяемого при вычислении экспоненциальных добавок к степенным разложениям решений в бесконечности, повернутого на  $\pi/2$  по часовой стрелке.

**Доказательство.** Алгоритмы методов степенной геометрии и методов французской школы отличаются на этапе линеаризации левой части дифференциального уравнения. Итак, пусть есть линеаризация дифференциального уравнения  $n$ -го порядка, которая имеет вид

$$\sum_{k=0}^n c_k(x) u^{(k)}, \quad (5)$$

где каждая  $c_k \in \mathbb{C}[[1/x]]$  — ряд по убывающим степеням  $x$ .

1. Покажем, как преобразуются и в какие точки переходят  $u^{(k)}$  при замене (4). Видим, что  $u' = uw'$ , после сокращения на  $u$  ей соответствует точка  $(-1, 1)$ , вторая производная  $u'' = u((w')^2 + w'')$  дает точки  $(-2, 2)$  и  $(-2, 1)$  и т.д. Докажем, что функции  $u^{(n)}$  сопоставляются точки  $(-n, n)$ ,  $(-n, n-1), \dots, (-n, 1)$ .

По индукции: база проверена; пусть

$$u^{(n-1)} = u \sum_{\substack{\ell_1+2\ell_2+\dots+(n-1)\ell_{n-1}=n-1, \\ \ell_i \in \{0,1,\dots,n-1\}}} c_L(w')^{\ell_1} \dots (w^{(n-1)})^{\ell_{n-1}},$$

тогда

$$\begin{aligned} u^{(n)} &= u' \sum_{\substack{\ell_1+\dots+(n-1)\ell_{n-1}=n-1, \\ \ell_i \in \{0,1,\dots,n-1\}}} c_L(w')^{\ell_1} \dots (w^{(n-1)})^{\ell_{n-1}} + u \sum_{k=1}^{n-1} c_L \ell_k (w')^{\ell_1} \dots (w^{(k)})^{\ell_k} \dots (w^{(n-1)})^{\ell_{n-1}} = \\ &= u \sum_{\substack{\ell_1+\dots+(n-1)\ell_{n-1}=n-1, \\ \ell_i \in \{0,1,\dots,n-1\}}} c_L(w')^{\ell_1+1} \dots (w^{(n-1)})^{\ell_{n-1}} + u \sum_{k=1}^{n-2} c_L \ell_k (w')^{\ell_1} \dots (w^{(k)})^{\ell_k-1} (w^{(k+1)})^{\ell_{k+1}+1} \dots (w^{(n-1)})^{\ell_{n-1}} + \\ &\quad + u c_{(0,\dots,0,1)} \ell_{n-1} w^{(n)} = u \sum_{\substack{\ell_1+2\ell_2+\dots+(n-1)\ell_{n-1}+n\ell_n=n, \\ \ell_i \in \{0,1,\dots,n\}}} \tilde{c}_L(w')^{\ell_1} \dots (w^{(n)})^{\ell_n}. \end{aligned}$$

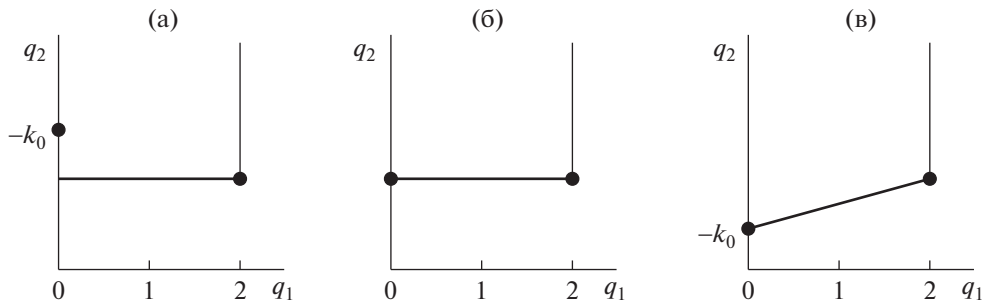
Полученной дифференциальной сумме соответствуют точки вида

$$(-\ell_1 - 2\ell_2 - \dots - (n-1)\ell_{n-1} - n\ell_n, \ell_1 + \ell_2 + \dots + \ell_{n-1} + \ell_n) = (-n, k), \quad k = 1, 2, \dots, n,$$

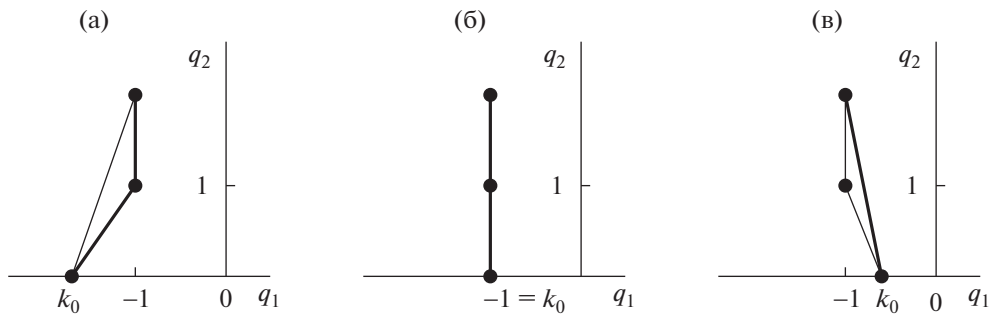
и, поскольку  $1 \leq \ell_1 + \ell_2 + \dots + \ell_{n-1} + \ell_n \leq n$ , формула доказана.

2. Теперь найдем точки, которые соответствуют  $u^{(k)}$  при построении многоугольника, по которому вычисляется порядок Жевре. Перепишем функции  $u^{(k)}$  через оператор  $D$  из предыдущего раздела. Так,  $u = Iu = D^0 u$ , и ей сопоставляется  $(0, 0)$ , функции  $u' = \frac{Du}{x}$  соответствует точка  $(1, 1)$  многоугольника, функции  $u'' = \frac{D^2 u - Du}{x^2}$  отвечают точки  $(2, 2)$ ,  $(1, 2)$ . Легко проверить по индукции, что  $u^{(k)}$  добавит в многоугольник точки  $(k, k)$ ,  $(k-1, k)$ ,  $\dots$ ,  $(1, k)$ , поскольку

$$u^{(k)} = \frac{D(D-I) \dots (D-(k-1)I)u}{x^k}.$$



Фиг. 1. Многоугольники Ньютона уравнения (6) для алгоритма I в случаях (а), (б), (в).



Фиг. 2. Многоугольники Ньютона уравнения (6) для алгоритма II в случаях (а), (б), (в).

Получили, что моному  $x^a u^{(k)}$  соответствуют точки  $(a - k, k), (a - k, n - k), \dots, (a - k, 1)$  в многоугольнике Ньютона алгоритма II и точки  $(k, k - a), (k - 1, k - a), \dots, (1, k - a)$  алгоритма I (в алгоритме I затем используется только часть точек). Имеем биективное для каждого из дифференциальных мономов отображение точек первого многоугольника на точки второго, осуществляемое с помощью поворота на  $\pi/2$  против часовой стрелки, переводящего точки второго многоугольника в точки первого по формуле

$$\begin{pmatrix} \tilde{q}_1 \\ \tilde{q}_2 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}.$$

При этом, если точка вида  $(a, k)$  появляется в многоугольнике алгоритма II на каком-то из этапов, то затем она не может исчезнуть, поскольку она соответствует выражениям вида  $c_{a,k} x^{a+k} u^{(k)}$ ,  $c_{a,k+1} x^{a+k+1} u^{(k+1)}$ , ... в операторе (3), т.е. выражениям, которые не могут сократиться после логарифмического преобразования (4) в полученном уравнении  $G(x, w) = 0$  (степень  $x$  перед мономами не меняется), а именно, мономам вида

$$\tilde{c}_{a,k} x^{a+k} (w')^k, \tilde{c}_{a,k+1} x^{a+k+1} (w')^{k-1} w'', \tilde{c}_{a,k+2} x^{a+k+2} (w')^{k-2} (w'')^2, \hat{c}_{a,k+2} x^{a+k+2} (w')^{k-1} w''', \dots$$

Это завершает доказательство утверждения.

**Пример 1.** При алгоритме I может пропасть часть точек из алгоритма II.

Рассматривается уравнение

$$xu'' + u' + x^{k_0} u = 0. \quad (6)$$

Вычисления по алгоритму I дадут точки  $(0, -k_0), (2, 1)$ , вычисления по методу II — точки  $(k_0, 0), (-1, 1), (-1, 2)$ . Таким образом, точке  $(-1, 1)$  метода II не соответствует точка по методу I.

На фиг. 1 и 2 изображены многоугольники для алгоритмов I и II для уравнения (6) в случаях (а)  $k_0 < -1$ , (б)  $k_0 = -1$  и (в)  $k_0 > -1$ .

#### 4. СХОДИМОСТЬ РАЗЛОЖЕНИЙ В НУЛЕ

Запишем второй член четвертой иерархии Пенлеве (см. [2]). Заметим, что это уравнение уже представлено в виде дифференциальной суммы (конечной суммы многочленов от  $y, y_x, \dots, y_{xxxx}$ , умноженных на  $x^\alpha$ ):

$$\begin{aligned} & (y_{xx} - 2xy - 2y^3 - \beta)y^2y_{xxxx} + (2y^2 + 8y^3y_x + 4yy_xx - y_xy_{xx} + \beta y_x)yy_{xxx} - \frac{1}{2}y^2y_{xxx}^2 - \frac{4}{3}yy_{xx}^3 + \\ & + \left(3xy^2 + 3\beta y - \frac{3}{2}y^4 + \frac{3}{2}y_x^2\right)y_{xx}^2 + (\beta y^4 - 2y_xy^2 - 12y_x^2y^3 - 2\beta^2y + 10xy^5 - 3\beta y_x^2 + 10y^7 - \\ & - 4xyy_x^2 - 4\beta xy^2)y_{xx} + 2(\beta - 4y^3)y^2y_x + \left(4\beta xy + 8xy^4 + \frac{3}{2}\beta^2 + 12\beta y^3\right)y_x^2 - \frac{10}{3}y^{10} - \\ & - 8xy^8 - 2\beta y^7 - 6x^2y^6 - 2x\beta y^5 + \left(\frac{1}{2}\beta^2 - 2 + 9\delta - \frac{4}{3}x^3\right)y^4 + x\beta^2y^2 + \frac{1}{3}\beta^3y = 0. \end{aligned} \quad (7)$$

Здесь  $x$  и  $y$  — независимая и зависимая комплексные переменные соответственно,  $\beta, \delta$  — комплексные параметры. Далее рассматриваем случай общего положения, предполагая, что оба параметра уравнения ненулевые.

##### 4.1. Разложения $W_1$ – $W_9$

Рассмотрим разложения решений уравнения (7), полученные в [3] и [9] при  $x \rightarrow 0$ .

Начнем с рассмотрения разложений семейства  $W_1$ , которые имеют вид

$$y = -\frac{\beta x^2}{10} + c_4x^4 + \frac{\beta}{80}x^5 + c_6x^6 + \sum_{k=7}^{\infty} a_{k,1}x^k.$$

Здесь и далее  $c_j$  — произвольные постоянные, коэффициенты  $a_{j,s}$  однозначно через них выражаются.

Во-первых, заметим, что линейаризация укороченного уравнения

$$\begin{aligned} & y_{xx}y^2y_{xxxx} - \beta y^2y_{xxxx} - \frac{y^2y_{xxx}^2}{2} - yy_xy_{xx}y_{xxx} + \beta yy_xy_{xxx} - \frac{4}{3}yy_{xx}^3 + 3\beta yy_{xx}^2 + \frac{3}{2}y_x^2y_{xx}^2 - \\ & - 2\beta^2yy_{xx} - 3\beta y_x^2y_{xx} + \frac{3}{2}\beta^2y_x^2 + \frac{1}{3}\beta^3y = 0, \end{aligned}$$

из которого получен первый член асимптотического разложения  $W_1$ , на решении  $w = -\beta x^2/10$  содержит оператор четвертой производной, т.е. порядок максимальной входящей в линейаризацию производной совпадает с порядком исходного уравнения (7). Для доказательства сходимости разложения можно использовать теоремы о сходимости из [10]. Разложение сходится.

Во-вторых, проверим это утверждение путем нахождения порядков Жевре и экспоненциальных добавок. Вычислим первую вариацию левой части уравнения (7) на решении  $W_1$ , применим к  $u$ , сделаем замену (4), построим многоугольник Ньютона полученного уравнения

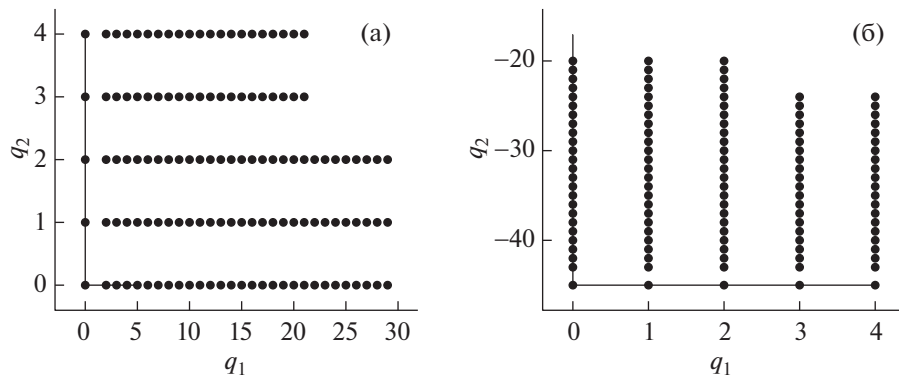
$$\frac{3\beta^3}{125} \left( -36xw_x + 3x^3w_xw_{xx} - 2x^4w_xw_{xxx} + 8x^2w_x^2 - 3x^4w_x^2w_{xx} + x^3w_x^3 - \frac{1}{2}x^4w_x^4 + \dots \right) = 0 \quad (8)$$

на фиг. 3а (в уравнении (8) выписаны только мономы с минимальными по  $x$  степенями).

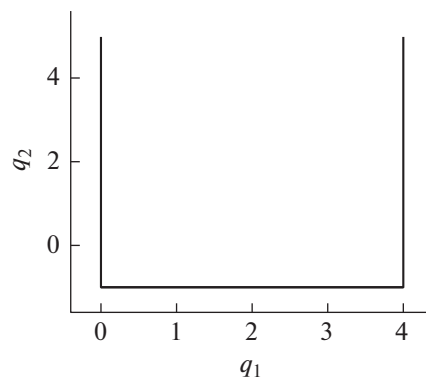
Многоугольник не имеет сторон с внешними нормальными  $N = (p_1, p_2)$ , удовлетворяющими условию  $p_1 < 0, p_2 > 0$ , поэтому разложение  $W_1$  не имеет экспоненциальных добавок.

На фиг. 3б указан многоугольник для нахождения порядков Жевре разложений. Видим, что он не имеет ненулевых тангенсов углов наклона сторон, поэтому ряд  $W_1$  сходится согласно теории из [6].

Рассуждения для семейств разложений  $W_2$ – $W_7$  из [3] идентичны проведенным выше, все эти разложения являются сходящимися, доказательство сходимости разложений семейств  $W_8, W_9$  (при  $\beta \neq -2$  и  $\beta \neq 2$  соответственно) также аналогично.



Фиг. 3. Многоугольники Ньютона алгоритмов II и I для разложения  $W_1$ .



Фиг. 4. Многоугольник Ньютона алгоритма I.

#### 4.2. Разложения, продолжающие асимптотику $y = cx$

Докажем, пользуясь алгоритмом I, сходимость разложений семейств  $W_{19} - W_{21}$ . Разложения этих семейств имеют вид  $y = cx + \beta/2x^2 + b_3x^3 + b_4x^4 + \dots$ .

Сделаем в уравнении (7) замену  $x = 1/t$ , переведя точку  $x = 0$  в  $t = \infty$ , и воспользуемся алгоритмом I построения многоугольника.

Для этого запишем линеаризацию левой части уравнения (7) (после замены переменной), выраженную через оператор  $D_t = t \frac{d}{dt}$ , получим (выписываем только старшие — ведущие степени по  $t$ )

1)  $b_3 \neq 0$ , ведущие слагаемые вида  $6c^2b_3(D_t^4 + 8D_t^3 + 19D_t^2 + 12D_t)t$ , многоугольник Ньютона которого представлен на фиг. 4. Многоугольник не имеет наклонных ребер, что говорит о сходимости разложений семейства  $W_{19}$ ;

2)  $b_3 = 0$ ,  $b_4 \neq c/6$ , ведущие слагаемые вида  $-2c^2(c - 6b_4)(D_t^4 + 9D_t^3 + 23D_t^2 + 15D_t)$ , многоугольник Ньютона получается сдвигом из многоугольника на фиг. 4. Для разложений семейств  $W_{20,21}$  коэффициент  $b_4 = c/6$  только в случае  $\delta = 0$ , такой случай не рассматриваем в данной работе, сходимость доказана.

#### 4.3. Сходимость разложений, продолжающих асимптотику $y = c$

Рассмотрим семейство разложений

$$y = c + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + \sum_{k=5}^{\infty} a_kx^k, \quad c \neq 0,$$

заметив, что разложения  $F_1 - F_6$  из [9] и найденные ниже разложения могут быть записаны как частные случаи этого семейства при различных соотношениях на коэффициенты  $c_1 - c_4$ . Вновь переводим точку  $x = 0$  в  $t = \infty$ , выписываем ведущие слагаемые при степенях оператора  $D_t$ :

$$-c^2(b + 2c^3 - 2c_2)(D_t^4 + 6D_t^3 + 11D_t^2 + 6D_t)t^4. \quad (9)$$

В случае  $b + 2c^3 - 2c_2 \neq 0$  многоугольник оператора (9) по алгоритму I получается вертикальным сдвигом из многоугольника на фиг. 4. Имеет место сходимость таких разложений. Сходимость разложений при  $b + 2c^3 - 2c_2 = 0$  должна быть рассмотрена отдельно.

#### 4.4. Разложения, продолжающие асимптотику $y = c$ при $\beta = -2c^3$

**Утверждение 2.** При  $\beta = -2c^3 \neq 0$ ,  $\delta \neq 0$  имеются следующие дополнительные семейства асимптотических разложений:

$$\begin{aligned} F_7 &= \left\{ y = c + c_2 x^2 + \sum_{k=3}^{\infty} a_{k,7} x^k \right\}, \quad c_2 \in \mathbb{C} \setminus \{0\}; \\ F_{8,9} &= \left\{ y = c + \frac{c(2 \pm 3\sqrt{2\delta})}{6} x^3 + c_4 x^4 + \sum_{k=5}^{\infty} a_{k,8,9} x^k \right\}, \quad c_4 \in \mathbb{C} \setminus \{0\}, \text{ существуют при } \delta \neq 2/9; \\ F_{10} &= \left\{ y = c + \frac{2c}{3} x^3 + c_4 x^4 + \sum_{k=5}^{\infty} a_{k,10} x^k \right\}, \quad c_4 \in \mathbb{C} \setminus \{0\}, \quad \text{при } \delta = 2/9; \\ F_{16} &= \left\{ y = c + c_4 x^4 + \sum_{k=5}^{\infty} a_{k,16} x^k \right\}, \quad c_4 \in \mathbb{C} \setminus \{0\}, \quad \text{при } \delta = 2/9; \\ F_{11} &= \left\{ y = c - \frac{c^3}{20} x^5 + \sum_{k=6}^{\infty} a_{k,11} x^k \right\}, \text{ существует при } \delta = 2/9; \\ F_{12} &= \left\{ y = c + c_1 x + c_2 x^2 + \sum_{k=3}^{\infty} a_{k,12} x^k \right\}; \\ F_{13,14} &= \left\{ y = c + c_1 x + \frac{c(2 + 6cc_1 \pm 3\sqrt{2\delta})}{6} x^3 + c_4 x^4 + \sum_{k=5}^{\infty} a_{k,13,14} x^k \right\}, \quad 6cc_1 + 2 = \pm 3\sqrt{2\delta}, \quad c_1 \neq 0; \\ F_{15} &= \left\{ y = c + c_1 x + \frac{-3c^4 c_1 - c^3 + 2c_1^3}{20} x^5 + \sum_{k=6}^{\infty} a_{k,15} x^k \right\}, \quad 6cc_1 + 2 = \pm 3\sqrt{2\delta}, \quad c_1 \neq 0. \end{aligned}$$

**Доказательство** состоит в последовательном переборе вершин и ребер многоугольников Ньютона методами степенной геометрии и следует ниже.

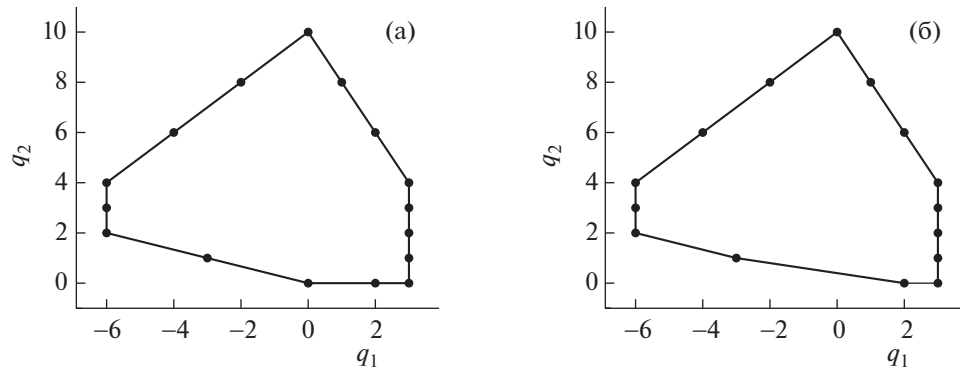
Обратимся к случаю  $\beta = -2c^3$ . Такое соотношение параметров уравнения и решения пропущено в [9]: при подстановке  $y = c + w$  в уравнение (7) у многоугольника  $\hat{\Gamma}$  может пропасть точка  $(-4, 1)$ , которой соответствует укороченное уравнение

$$-c^2(\beta + 2c^3)w_{xxxx} = 0.$$

Преобразованному уравнению (7) в этом случае соответствует многоугольник  $\hat{\Gamma}$ , изображенный на фиг. 5а. В рассмотрении в многоугольнике  $\hat{\Gamma}$  нуждаются лишь вершина  $(-6, 2)$  и ребро  $\hat{\Gamma}_1$ , соединяющее вершины  $(-6, 2)$  и  $(0, 0)$ . Эта конфигурация многоугольника возможна при наличии точки  $(0, 0)$ , т.е. в случае  $\delta \neq 2/9$ .

Вершине  $(-6, 2)$  многоугольника  $\hat{\Gamma}$  соответствует укороченное уравнение

$$c^2 \left( w_{xxxx} w_{xx} - \frac{w_{xxx}^2}{2} \right) = 0,$$

Фиг. 5. Многоугольники Ньютона  $\hat{\Gamma}$  и  $\hat{\Gamma}'$ .

в которое подставляем  $w = cx^r$ , имеем  $r = 0, r = 1$  (кратности два),  $r = 2, r = 4$  (кратности один). В конусе задачи лежат  $r = 1$  и  $r = 2$ . При  $r = 2$  асимптотика продолжается как семейство разложений  $F_7$ . Случай  $r = 1$  — кратный корень — исследуем позже, сделав в уравнении (7) замену  $w = c + c_1x + u$ .

Ребру  $\hat{\Gamma}_1$  отвечает уравнение

$$c^2 w_{xxxx} w_{xx} - \frac{1}{2} c^2 w_{xxx}^2 + 2c^3 w_{xxx} - 2c^3 x w_{xxxx} - c^4 (2 - 9\delta) = 0. \quad (10)$$

Его решение ищем в виде  $w = Ax^3$ , получаем, что  $A_{1,2} = (2c \pm 3c\sqrt{2\delta})/6$ . Имеем либо одно критическое число  $k = 4$  при  $\delta \neq 0$ , что дает разложения  $F_{8,9}$ , либо — при  $\delta = 0$  (кратном корне  $A$ ) — нулевую линейаризацию укороченного уравнения на решении. В данной работе рассматриваем случай общего положения, поэтому дальнейшее исследование нулевой линейаризации здесь не проводим.

Теперь рассмотрим многоугольник  $\hat{\Gamma}'$ , полученный для уравнения (7) после замены  $y = c + w$  в случае  $\beta = -2c^3$ ,  $\delta = 2/9$  (изображен на фиг. 5б). Рассматриваем его ребра: ребро  $\hat{\Gamma}'_1$ , соединяющее вершины  $(-6, 2)$  и  $(-3, 1)$ , и ребро  $\hat{\Gamma}'_2$ , соединяющее вершины  $(2, 0)$  и  $(-3, 1)$ , а также вершины  $(2, 0)$  (алгебраическая),  $(-6, 2)$  (встречалась ранее) и новую вершину  $(-3, 1)$ , которой отвечает укороченное уравнение

$$2c^3 (w_{xxx} - xw_{xxxx}) = 0.$$

Оно не дает подходящих степенных решений: подставляя  $w = cx^r$  в уравнение, имеем  $r \in \{0, 1, 2, 4\}$ , такие значения не подходят для продолжения разложения (не соответствуют нормальному конусу вершины).

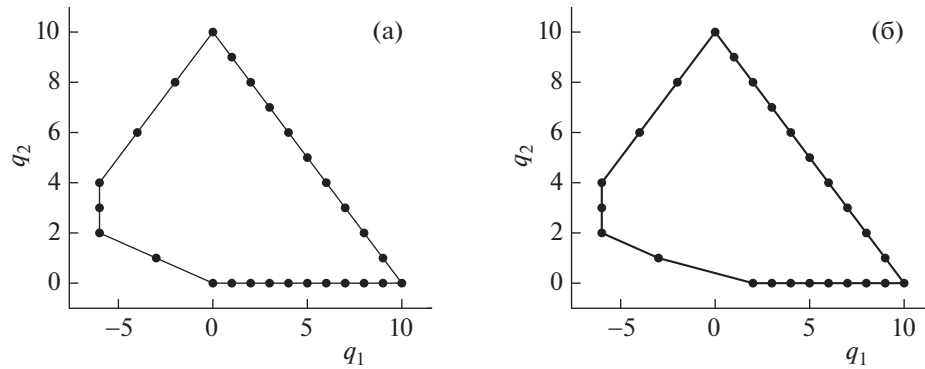
Ребру  $\hat{\Gamma}'_1$  отвечает укороченное уравнение (10) при  $\delta = 2/9$  и семейства  $F_{10}, F_{16}$ , полученные из имеющихся выше семейств  $F_{8,9}$  подстановкой.

Ребру  $\hat{\Gamma}'_2$  отвечает укороченное уравнение

$$2c^3 (w_{xxx} - xw_{xxxx}) - 6c^6 x^2 = 0, \quad (11)$$

его решение ищем, подставляя  $w = Bx^5$ . Имеем семейство разложений  $F_{11}$ .

Рассмотрим многоугольник уравнения, полученного из уравнения (7) заменой  $y = c + c_1x + u$ . В [9] не учтены разложения, отвечающие  $\beta + 2c^3 = 0$ . Опять разделяем случаи наличия (многоугольник  $\bar{\Gamma}$ , фиг. 6а) и отсутствия ( $\bar{\Gamma}'$ , фиг. 6б) точки  $(0, 0)$  в носителе многоугольника, т.е. случаи неравенства и равенства нулю выражения  $cc_1(12 + 18cc_1) - 9\delta + 2$  (заметим, что предыдущий разобранный случай вкладывается здесь как подслучай при  $c_1 = 0$ ).



Фиг. 6. Многоугольники Ньютона  $\bar{\Gamma}$  и  $\bar{\Gamma}'$ .

Уравнение вершины  $(-6, 2)$  многоугольника  $\bar{\Gamma}$  рассмотрено ранее, вершина дает семейство решений  $F_{12}$ . Вершина  $(0, 0)$  алгебраическая, не дает решений. Ребру, соединяющему вершины  $(0, 0)$  и  $(-6, 2)$ , отвечает уравнение

$$c^2 w_{xxxx} w_{xx} - \frac{1}{2} c^2 w_{xxx}^2 + (2c^3 + 6c^4 c_1) w_{xxx} - (2c^3 + 6c^4 c_1) x w_{xxxx} - c^4 (12c c_1 + 2 + 18c^2 c_1^2 - 9\delta) = 0.$$

Его решение ищем в виде  $w = Ax^3$ , получаем, что  $A_{13,14} = (2c + 6c^2 c_1 \pm 3c\sqrt{2\delta})/6$ , критическое число  $k = 4$  при  $\delta \neq 0$  и нулевую линейаризацию укороченного уравнения на решении при  $\delta = 0$ . Это дает при  $\delta \neq 0$  начальные члены разложения  $F_{13,14}$ .

Остается многоугольник  $\bar{\Gamma}'$ , в котором появляется не рассмотренное ранее ребро, соединяющее точки  $(-3, 1)$  и  $(2, 0)$ . Ему отвечает укороченное уравнение

$$2c^3(1 + 3cc_1)(u_{xxx} - xu_{xxxx}) - 6c^3(c^3 + 6c^4 c_1 + 9c^5 c_1^2 - 2c^3 - 6cc_1^4)x^2 = 0 \quad (12)$$

с подходящими решениями  $u = Bx^5$ . Заметим, что уравнение (11) получается как частный случай уравнения (12) при  $c_1 = 0$ . Имеем семейство  $F_{15}$ .

## 5. КОД ДЛЯ НАХОЖДЕНИЯ ПОРЯДКОВ ЖЕВРЕ

Имеется написанный в пакете символьных вычислений Wolfram Mathematica (см. [11]) код для вычисления порядка Жевре. На входе есть дифференциальное уравнение и формально удовлетворяющий ему ряд. Код приводится для дифференциального уравнения 2-го порядка и выложен в открытом доступе: <https://github.com/Victoranoshin/Asymptotic-expansions-of-solutions-for-equations-in-the-Wolfram-Mathematica-system>.

В команду Expand вместо многоточия подставляется левая часть дифференциального уравнения, вместо 'solution' — формальное решение этого уравнения.

Код для вычисления порядков Жевре в случае рассматриваемого уравнения 4-го порядка для конкретных формальных разложений выкладывается туда же.

## 6. О СХОДИМОСТИ ФОРМАЛЬНЫХ РЕШЕНИЙ

Пусть уравнение (1) при  $x \rightarrow 0$  имеет формальное решение  $\sum_{k=0}^{\infty} b_k x^k$ . Сделаем в уравнении (1) подстановку

$$y = \sum_{k=0}^{\infty} b_k x^k + z,$$

получим уравнение  $G(x, z) = 0$ . При этом носитель дифференциальной суммы (ряда)  $G$  не содержит точек  $(q_1, q_2)$  с  $q_2 = 0$ . Пусть  $g$  — линейная по  $z$  часть дифференциальной суммы  $G$ . Тогда

$$g(x, z) = \sum_{k=0}^{\infty} g_k(z)x^k,$$

где  $g_k$  — линейные по  $z$  дифференциальные суммы с постоянными коэффициентами. Пусть  $g_\ell$  — первая отличная от нуля такая сумма.

**Теорема** (см. [4, п. 3.4]). *Если порядки дифференцирования в  $g_\ell$  и полном уравнении (1) совпадают, то формальное решение сходится при малых  $|x|$ .*

Это условие совпадает с условием в работе [12]. Для рядов более широкого класса, чем упомянуто здесь, оно доказано в [10].

Для всех полученных вблизи нуля разложений можно проверить выполнение этого условия для доказательства сходимости.

## СПИСОК ЛИТЕРАТУРЫ

1. Пикеринг А. Иерархии Пенлеве и тест Пенлеве // Успехи матем. наук. 2003. Т. 137. № 3. С. 445–456.
2. Кудряшов Н.А. О четвертой иерархии Пенлеве // Теор. и матем. физ. 2003. Т. 134. С. 101–109.
3. Аношин В.И., Бекетова А.Д., Парусникова А.В., Романов К.В. Некоторые асимптотические разложения решений иерархии четвертого уравнения Пенлеве // Программирование. 2022. № 1. С. 34–39.
4. Брюно А.Д. Асимптотика и разложения решений обыкновенного дифференциального уравнения // Успехи матем. наук. 2004. Т. 59. Вып. 3. С. 31–80. Тр. ММО. 2010. Т. 71.
5. Рамис Ж.П. Расходящиеся ряды и асимптотическая теория. Ин-т компьютерных исследований, Москва–Ижевск, 2002.
6. Sibuya Y. Linear Differential Equations in the Complex Domain: Problems of Analytic Continuation. Providence: AMS, 1985.
7. Брюно А.Д. Экспоненциальные разложения решений обыкновенного дифференциального уравнения // Докл. АН. 2012. Т. 443. № 1. С. 539–544.
8. Parusnikova A.V. On Gevrey orders of formal power series solutions to the third and fifth Painlevé equations near infinity // Opuscula Math. 2014. V. 34. № 3. P. 591–599.
9. Аношин В.И., Бекетова А.Д., Парусникова А.В. Асимптотические разложения решений второго члена четвертой иерархии Пенлеве, продолжающие константную асимптотику при  $x \rightarrow 0$  // Дифференц. уравнения и смежные вопросы математики. Тр. XIII Приокской науч. конф. ГСГУ, 2021. С. 33–39.
10. Gontsov R., Goryuchkina I. The Maillet-Malgrange type theorem for generalized power series // Manuscripta Math. 2018. V. 156. P. 171–185.
11. Wolfram St. The Mathematica Book. Wolfram Media, Inc., 2003. 1488 p.
12. Malgrange B. Sur le théorème de Maillet-Malgrange // Asymptot. Anal. 1989. V. 2. P. 1–4.

## УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.67

### СИМВОЛЬНО-ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ АДИАБАТИЧЕСКОЙ ВОЛНОВОДНОЙ МОДЫ В ПЛАВНОМ ВОЛНОВОДНОМ ПЕРЕХОДЕ<sup>1)</sup>

© 2023 г. Д. В. Диваков<sup>1,\*</sup>, А. А. Тютюнник<sup>2,\*\*</sup>

<sup>1</sup> 117198 Москва, ул. Миклухо-Маклая, 6, Российский университет дружбы народов, Россия

<sup>2</sup> 141980 Дубна, М.о., ул. Жолио-Кюри, 6, Объединённый институт ядерных исследований, Россия

\*e-mail: divakov-dv@rudn.ru

\*\*e-mail: tyutyunnik-aa@rudn.ru

Поступила в редакцию 25.04.2022 г.  
Переработанный вариант 25.04.2022 г.  
Принята к публикации 17.09.2022 г.

В работе исследуется модель адиабатических волноводных мод средствами компьютерной алгебры. В рамках модели решение системы уравнений Максвелла приводится к форме, которая выражается через решение системы четырех обыкновенных дифференциальных уравнений и двух алгебраических уравнений для шести компонент электромагнитного поля. В случае многослойных волноводов средствами системы компьютерной алгебры уравнения модели сводятся к однородной системе линейных алгебраических уравнений, которая исследуется в работе символично. Условие нетривиальной разрешимости системы задает дисперсионное уравнение, которое решается символично-численным методом, тогда как система решается символично. В статье представлены решения, описывающие адиабатические волноводные моды в нулевом приближении с учетом малого наклона границы раздела волноводного слоя, качественно отличающиеся от решений, не учитывающих наклон границы раздела. Библ. 19. Фиг. 7.

**Ключевые слова:** символическое решение линейных уравнений, символическое решение дифференциальных уравнений, адиабатические волноводные моды, направляемые моды, плавно-нерегулярный волновод.

DOI: 10.31857/S0044466923010076, EDN: LKQVIF

## 1. ВВЕДЕНИЕ

Метод расчета электромагнитного поля волноводных мод, распространяющихся в нерегулярных волноводах с медленно меняющимися параметрами, сформировался в работах [1]–[8] и др. В русскоязычной научно-исследовательской литературе наибольшее распространение получил “метод поперечных сечений”, разработанный в трудах Б.З. Каценеленбаума [3], [4] для закрытых волноводов, и его “обобщение” для открытых волноводов, разработанное В.В. Шевченко [5], [9].

Основной сложностью постановки задачи отыскания направляемых волноводных мод в нерегулярных волноводах — волноводах с криволинейной границей — является учет условий непрерывности тангенциального поля на криволинейной границе. Понятие направляемой волноводной моды определено для планарных волноводов — волноводов с плоскими границами. Применяя классическую процедуру разделения переменных к волноводной задаче для планарных волноводов, можно сформулировать самосопряженную спектральную задачу, собственными значениями которой являются вещественные коэффициенты фазового замедления направляемых мод, собственные функции определяют соответствующие стоячие волны в поперечном сечении волновода. В случае наличия криволинейной границы переменные уже не разделяются, и понятие направляемой волноводной моды в нерегулярном волноводе строго не определено, хотя можно строить различные приближенные направляемые моды нерегулярного волновода. Как правило, методы приближенного построения направляемых мод плавно-нерегулярного волно-

<sup>1)</sup> Работа выполнена при финансовой поддержке РФФИ (код проекта 20-11-20257).

вода основаны на приближении по малому параметру, характеризующему медленное изменение геометрии структуры или поля по одной из координат.

В подходах Б.З. Каценеленбаума и В.В. Шевченко при построении приближенных мод нерегулярного волновода учтен только вклад нулевого порядка в условия сопряжения на криволинейной границе — т.е. учтена переменная толщина волноводного слоя, но отброшены вклады более высоких порядков, описывающие малый наклон криволинейной границы раздела. Подход Б.З. Каценеленбаума и В.В. Шевченко позволяет получить медленно меняющиеся вдоль волновода направляемые моды, причем полученные в каждом поперечном сечении нерегулярного волновода направляемые моды совпадают с направляемыми модами регулярного волновода аналогичной толщины. В каждом поперечном сечении спектральная задача самосопряженная, поэтому коэффициенты фазового замедления представляют собой вещественные медленно меняющиеся функции. Однако вклады первого и высших порядков малости по углу наклона границы в радианах в методе поперечных сечений не учитываются.

Авторам известен альтернативный подход к построению волноводных мод в адиабатическом приближении, который учитывает вклады более высоких порядков малости, описывающие наклон криволинейной границы раздела. В основе модели адиабатических волноводных мод (АВМ) [10]–[12] лежит адаптированное для волноводного распространения приближение коротких волн, описанное в книге В.М. Бабича и В.С. Булдырева [13]. Решение уравнений Максвелла в модели АВМ представляется в виде асимптотического ряда. В нулевом приближении асимптотического разложения модели АВМ уравнения Максвелла редуцируются к системе четырех обыкновенных дифференциальных уравнений первого порядка и двум соотношениям. Условия сопряжения на криволинейной границе в нулевом приближении асимптотического разложения формулируется с учетом наклона, который является малой величиной в волноводах с медленно меняющимися параметрами.

В работе рассматривается модель АВМ в нулевом приближении асимптотического разложения, в рамках которого рассматривается несколько первых приближений по малости наклона криволинейной границы. Приводятся результаты нулевого приближения по малости наклона криволинейной границы, а также первого и второго порядков малости, качественно отличающиеся от нулевого приближения.

## 2. МОТИВАЦИЯ ИСПОЛЬЗОВАНИЯ СИСТЕМ КОМПЬЮТЕРНОЙ АЛГЕБРЫ

Асимптотические методы удобны тем, что нулевое приближение, описывающее основной вклад решения, отыскивается, как правило, в символьном виде. Учитывая асимптотический характер решения, в модели АВМ удастся получить ряд промежуточных результатов в символьном виде, что и предопределяет использование компьютерной алгебры в качестве одного из инструментов исследования. В серии работ по модели АВМ [10]–[14] в символьном виде получены как дифференциальные уравнения модели, так и их символьные решения. Благодаря символьному решению дифференциальных уравнений модели, удастся в символьном виде сформулировать задачу отыскания направляемых мод нерегулярных волноводов. Эта задача формулируется в виде однородной системы линейных уравнений с символьной матрицей коэффициентов, которую в некоторых случаях можно решить символьно [15]. Условие разрешимости системы — равенство нулю детерминанта матрицы коэффициентов — определяет нелинейное уравнение, которому удовлетворяет коэффициент фазового замедления.

В настоящей работе для приближенного учета условий сопряжения на криволинейной границе раздела применяется разложение по малому параметру, характеризующему малость угла наклона границы. В нулевом приближении условие разрешимости системы представляет собой нелинейное уравнение с вещественными корнями, а в первом и старших приближениях корни могут быть комплексными. В работе и уравнение, и его корни представляются в виде разложений по степеням малого параметра, что позволяет получить простые формулы для первого и второго приближений. Важно отметить, что рассматриваемое уравнение представляет собой равенство нулю детерминанта, который необходимо представить в виде разложения по степеням малого параметра, что требует символьных манипуляций. Представление корней уравнения в виде разложения по степеням малого параметра, подстановка в уравнение и получение расчетных формул первых двух приближений также требуют символьных манипуляций. Поэтому основным инструментом решения задачи отыскания комплексных корней выступает система компьютерной алгебры.

После отыскания корней детерминанта в виде разложений по степеням малого параметра решается задача отыскания направляемых мод нерегулярных волноводов, которая формулируется в виде однородной системы линейных уравнений с символьной матрицей коэффициентов. В работе эта задача решается символьно с помощью алгоритма из [15].

### 3. МЕТОДЫ

#### 3.1. Модель адиабатических волноводных мод в общем случае

В работе [14] получен нулевой вклад в адиабатическое приближение волноводного решения уравнений Максвелла вида:

$$\begin{Bmatrix} \mathbf{E}(x, y, z, t) \\ \mathbf{H}(x, y, z, t) \end{Bmatrix} = \begin{Bmatrix} \mathbf{E}_0(x, y, z) \\ \mathbf{H}_0(x, y, z) \end{Bmatrix} \exp\{i\omega t - ik_0\varphi(y, z)\}, \quad (1)$$

причем

$$\varepsilon \frac{\partial E_0^y}{\partial x} = -ik_0 \left( \frac{\partial \varphi}{\partial y} \right) \left( \frac{\partial \varphi}{\partial z} \right) H_0^y - ik_0 \left( \varepsilon \mu - \left( \frac{\partial \varphi}{\partial y} \right)^2 \right) H_0^z, \quad (2)$$

$$\varepsilon \frac{\partial E_0^z}{\partial x} = ik_0 \left( \varepsilon \mu - \left( \frac{\partial \varphi}{\partial z} \right)^2 \right) H_0^y + ik_0 \left( \frac{\partial \varphi}{\partial z} \right) \left( \frac{\partial \varphi}{\partial y} \right) H_0^z, \quad (3)$$

$$\mu \frac{\partial H_0^y}{\partial x} = ik_0 \left( \frac{\partial \varphi}{\partial y} \right) \left( \frac{\partial \varphi}{\partial z} \right) E_0^y + ik_0 \left( \varepsilon \mu - \left( \frac{\partial \varphi}{\partial y} \right)^2 \right) E_0^z, \quad (4)$$

$$\mu \frac{\partial H_0^z}{\partial x} = -ik_0 \left( \varepsilon \mu - \left( \frac{\partial \varphi}{\partial z} \right)^2 \right) E_0^y - ik_0 \left( \frac{\partial \varphi}{\partial z} \right) \left( \frac{\partial \varphi}{\partial y} \right) E_0^z, \quad (5)$$

$$E_0^x = -\frac{\partial \varphi}{\partial y} \frac{1}{\varepsilon} H_0^z + \frac{\partial \varphi}{\partial z} \frac{1}{\varepsilon} H_0^y, \quad (6)$$

$$H_0^x = \frac{\partial \varphi}{\partial y} \frac{1}{\mu} E_0^z - \frac{\partial \varphi}{\partial z} \frac{1}{\mu} E_0^y. \quad (7)$$

Для тонкопленочного многослойного волновода, состоящего из оптически однородных слоев, систему уравнений (2)–(7) следует дополнить условиями сопряжения электромагнитного поля на границах раздела сред [16]. На границах раздела диэлектрических сред выполняются граничные условия сопряжения полей:

$$[n \times \mathbf{E}]_{x=h(y,z)} = \mathbf{0}, \quad [n \times \mathbf{H}]_{x=h(y,z)} = \mathbf{0}, \quad (8)$$

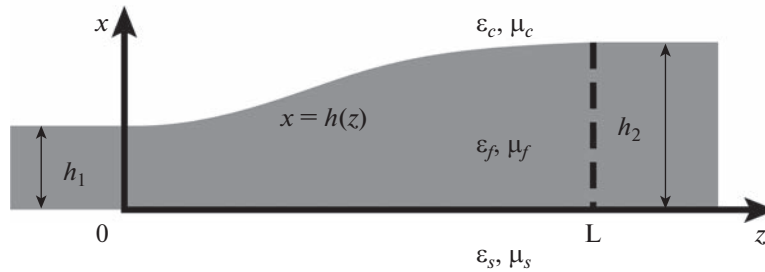
где через  $[\mathbf{f}]_{x=h(y,z)}$  обозначен скачок векторной величины  $\mathbf{f}$  на границе  $x = h(y, z)$ . Кроме того, выполняются асимптотические граничные условия на бесконечности [16]:

$$\|\mathbf{E}\| \xrightarrow{x \rightarrow \pm\infty} 0, \quad \|\mathbf{H}\| \xrightarrow{x \rightarrow \pm\infty} 0. \quad (9)$$

#### 3.2. Модель адиабатических волноводных мод для регулярных по $y$ волноводов

**3.2.1. Уравнения модели.** В работе рассматривается случай, когда от одной из горизонтальных координат не зависят ни геометрия интегрально-оптического волновода, ни решения уравнений Максвелла для АВМ, а именно случай  $\partial/\partial y \equiv 0$ . В этом случае уравнения (2)–(5) принимают более простой вид:

$$\begin{aligned} \varepsilon \frac{\partial E_0^y}{\partial x} &= -ik_0 \varepsilon \mu H_0^z, & \varepsilon \frac{\partial E_0^z}{\partial x} &= ik_0 \left( \varepsilon \mu - (\varphi'(z))^2 \right) H_0^y, \\ \mu \frac{\partial H_0^y}{\partial x} &= ik_0 \varepsilon \mu E_0^z, & \mu \frac{\partial H_0^z}{\partial x} &= -ik_0 \left( \varepsilon \mu - (\varphi'(z))^2 \right) E_0^y, \end{aligned} \quad (10)$$



**Фиг. 1.** Геометрия двумерного плавно-нерегулярного волноводного перехода между двумя регулярными волноводами.

дополнительные соотношения (6) и (7) также принимают упрощенный вид:

$$E_0^x = \frac{1}{\epsilon} \phi'(z) H_0^y, \quad H_0^x = -\frac{1}{\mu} \phi'(z) E_0^y. \quad (11)$$

**Замечание.** Систему (10) и соотношения (11) можно также сформулировать в виде уравнений второго порядка относительно компонент  $E_0^y$ ,  $H_0^y$  и дополнительных соотношений, которые разбиваются на две независимые подсистемы.

**3.2.2. Условия сопряжения.** Рассмотрим более подробно условия (8) на наклонной части границы раздела волноводных слоев. В точке  $(h(z), z)^T$  границы раздела  $x = h(z)$  условия сопряжения представляют собой непрерывность следующих величин:

$$[\mathbf{n} \times \mathbf{E}] = (h'(z) E_y; -E_z - h'(z) E_x; E_y)^T, \quad (12)$$

$$[\mathbf{n} \times \mathbf{H}] = (h'(z) H_y; -H_z - h'(z) H_x; H_y)^T, \quad (13)$$

причем экспоненциальный множитель  $\exp\{i\omega t - ik_0 \phi(z)\}$  принимает одинаковые ненулевые значения по обе стороны границы раздела слоев.

В работе используются приближенные условия сопряжения: нулевое приближение условий сопряжения, получаемое из (12) и (13) в пренебрежении малостью величины  $h'(z)$ , а также первое и второе приближения по малой величине  $h'(z)$ .

**3.2.3. Геометрия рассматриваемой структуры.** Рассматривается трехслойный плавно-нерегулярный волновод, геометрия которого представлена на фиг. 1. Параметры волновода следующие:  $\mu_c = \mu_f = \mu_s = 1$ ,  $\epsilon_c = 1$ ,  $\epsilon_f = 1.565^2$ ,  $\epsilon_s = 1.47^2$ , толшины определены как  $h_1 = 2\lambda$ ,  $h_2 = 3\lambda$ , а  $L = 100\lambda$ , где  $\lambda$  — длина волны,  $\lambda = 0.55 \text{ } [\mu\text{m}]$ . Переменная толщина  $h(z)$  определена следующим образом:

$$h(z) = 2(h_1 - h_2) \left( \frac{z}{L} \right)^3 - 3(h_1 - h_2) \left( \frac{z}{L} \right)^2 + h_1, \quad (14)$$

причем для  $h'(z)$  выполняется  $|h'(z)| \ll 1$ , т.е.  $h'(z)$  является малым параметром при каждом фиксированном  $z$ . Для описанной выше структуры вычисляем адиабатические волноводные моды.

**3.2.4. Символьный метод решения задачи.** Для вычисления адиабатических волноводных мод рассматривается система (10). Система (10) сформулирована в символьном виде, коэффициенты системы  $\epsilon, \mu$  для рассматриваемого волновода есть кусочно-постоянные функции, поэтому в каждой области их постоянства систему (10) решаем символьно в системе компьютерной алгебры Maple, используя встроенную функцию dsolve [17]. В результате применения функции dsolve в каждом слое получаем разложение решения по фундаментальной системе решений (ФСР) с неопределенными коэффициентами. Решения в полубесконечных слоях не должны нарастать на бесконечности согласно условиям (9), благодаря чему константы, стоящие перед нарастающими на бесконечности функциями ФСР, будут определены и равны нулю.

Записывая условия непрерывности (12) и (13) на границах слоев в системе компьютерной алгебры Maple [17], получаем однородную систему алгебраических уравнений вида

$$M(\gamma(z))C(z) \equiv [M_0(\gamma(z)) + h'(z)M_1(\gamma(z))]C(z) = 0, \quad (15)$$

где  $\gamma(z) = \varphi'(z)$ , вектор  $C(z)$  определяет константы разложения решения по ФСР в каждом слое при фиксированном  $z$ . Условие разрешимости системы (15) есть равенство нулю определителя системы, представляющее собой нелинейное уравнение

$$\det M(\gamma(z)) \equiv D_0(\gamma(z)) + h'(z)D_1(\gamma(z)) + h'^2(z)D_2(\gamma(z)) = 0. \quad (16)$$

Разложение (16) осуществляется в системе компьютерной алгебры Maple, символьное выражение детерминанта осуществляется с помощью функции Determinant пакета LinearAlgebra, разложение по степеням  $h'(z)$  осуществляется с помощью функции coeff [17]. Нулевое приближение представляет собой уравнение  $D_0(\gamma(z)) = 0$ , которое характеризуется вещественными корнями и при каждом фиксированном  $z$  корни отыскиваются с помощью встроенной в Maple функции RootOf [17].

Учитывая разложение уравнения (16), искомые корни также представляются в виде аналогичного разложения

$$\gamma(z) = \gamma^0(z) + h'(z)\gamma^1(z) + h'^2(z)\gamma^2(z) + \dots, \quad (17)$$

которое представляет собой разложение по малому параметру при каждом фиксированном  $z$  и определяются с помощью представленного в следующем разделе метода, реализованного в системе компьютерной алгебры Maple [17]. В работе с помощью описанного ниже метода отыскиваются первое и второе приближения корней уравнения (16).

Каждому корню уравнения (16) соответствует система линейных алгебраических уравнений (15), которая решается символьно с помощью метода из [15].

Описанная процедура вычисления корней нелинейного уравнения (16) в виде разложения (17), а также символьное решение системы (15) реализованы в системе компьютерной алгебры Maple [17].

### 3.2.5. Символьно-численный метод поиска корней нелинейного уравнения с малым параметром.

Рассмотрим нелинейное уравнение, представленное в виде разложения по степеням малого параметра  $\delta$ :

$$f(x) + \delta g(x) + \delta^2 u(x) = 0. \quad (18)$$

Искомый корень представляется в виде разложения

$$x = x_0 + \delta x_1 + \delta^2 x_2 + \dots, \quad (19)$$

которое далее подставляется в уравнение и с помощью разложения в ряд Тейлора формируются уравнения в нулевом, первом и старших приближениях по малому параметру  $\delta$ . Описанная процедура реализована в системе компьютерной алгебры Maple [17]. В результате имеем следующие уравнения:

$$\begin{aligned} \delta^0: f(x_0) &= 0, \\ \delta^1: g(x_0) + x_1 f'(x_0) &= 0, \\ \delta^2: u(x_0) + x_1 g'(x_0) + x_2 f'(x_0) + \frac{1}{2} x_1^2 f''(x_0) &= 0. \end{aligned} \quad (20)$$

В нулевом приближении необходимо решать нелинейное уравнение для отыскания значений  $x_0$ , а для отыскания поправок первого и второго порядков требуется знать производные функций  $f$  и  $g$ , которые определяются символьно с помощью функции diff:

$$x_1 = -\frac{g(x_0)}{f'(x_0)}, \quad (21)$$

$$x_2 = -\frac{x_1^2 f''(x_0) + 2x_1 g'(x_0) + 2u(x_0)}{2f'(x_0)}. \quad (22)$$

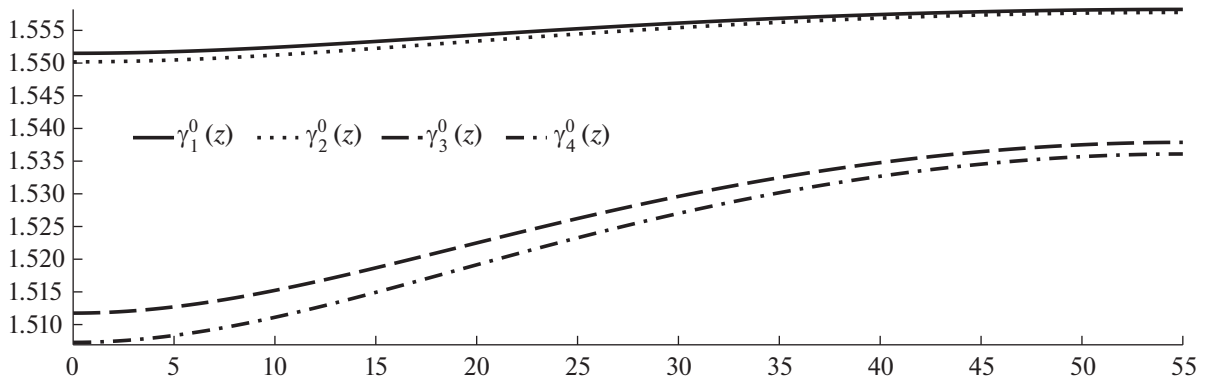
Описанный подход имеет смысл, если уравнение нулевого приближения  $f(x) = 0$  решается проще исходного уравнения  $f(x) + \delta g(x) + \delta^2 u(x) = 0$ . Такое возможно, например, если корни уравнения  $f(x) = 0$  вещественные, даже если корни исходного уравнения комплексные.

## 4. РЕЗУЛЬТАТЫ

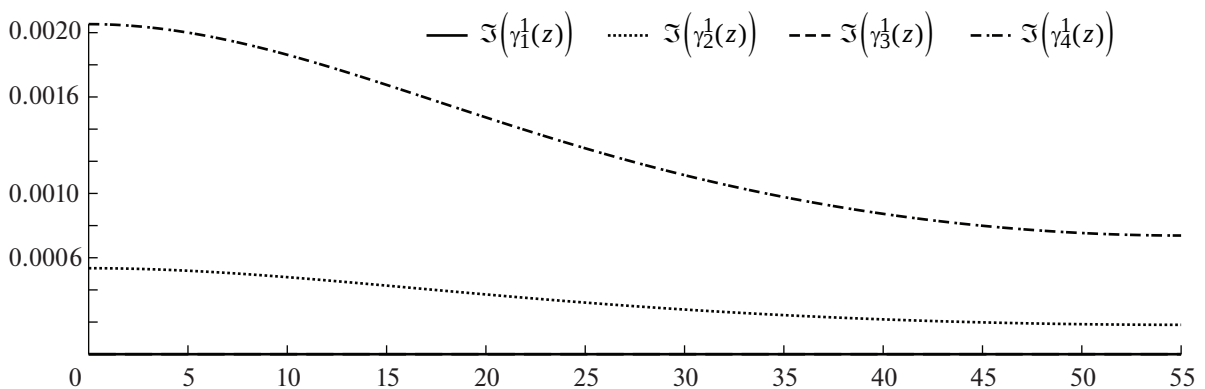
### 4.1. Приближенное вычисление корней нелинейного уравнения

Решаем для рассматриваемой структуры задачу (16). Символьное представление определителя из (16) получаем с помощью функции `Determinant` пакета `LinearAlgebra` [17]. Символьное представление детерминанта громоздкое, поэтому не приводим его в работе. Вычисляем  $\gamma^0(z)$  для рассматриваемой структуры с помощью встроенной в Maple функции `RootOf` [17]. Параметры рассматриваемой структуры приведены в п. 3.2.3. Затем, с помощью приведенного выше метода вычисляем  $\gamma^1(z)$  и  $\gamma^2(z)$ . Графики вычисленных  $\gamma_j^{0,1,2}(z)$ ,  $j = \overline{1,4}$ , приведены на фиг. 2–4.

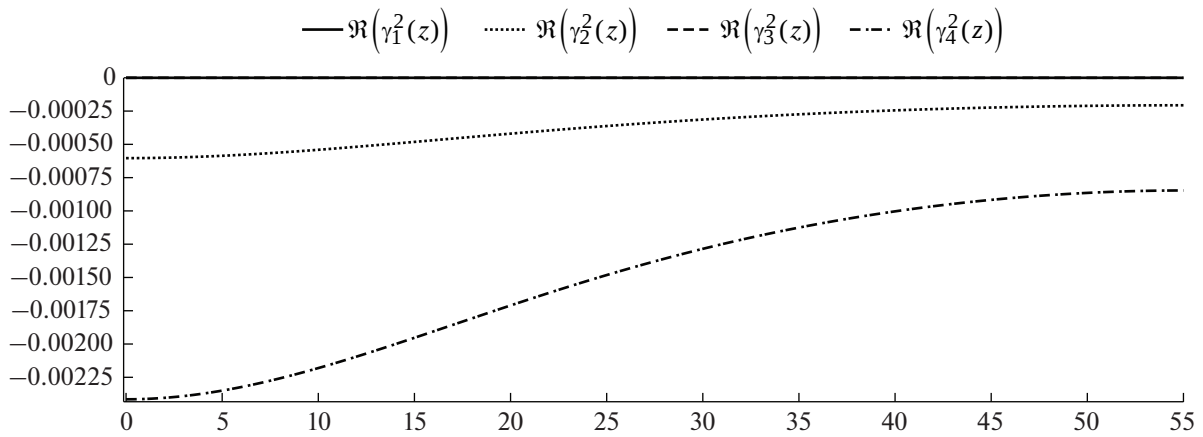
Величина  $\gamma^0(z)$  является вещественной (фиг. 2), в отличие от вкладов первого порядка малости, представленных на фиг. 3, которые являются мнимыми. Вклады второго порядка малости приведены на фиг. 4 и являются вещественными. Важно отметить, что для пары величин  $\gamma_{1,3}(z)$  вклады первого и второго порядков являются нулевыми.



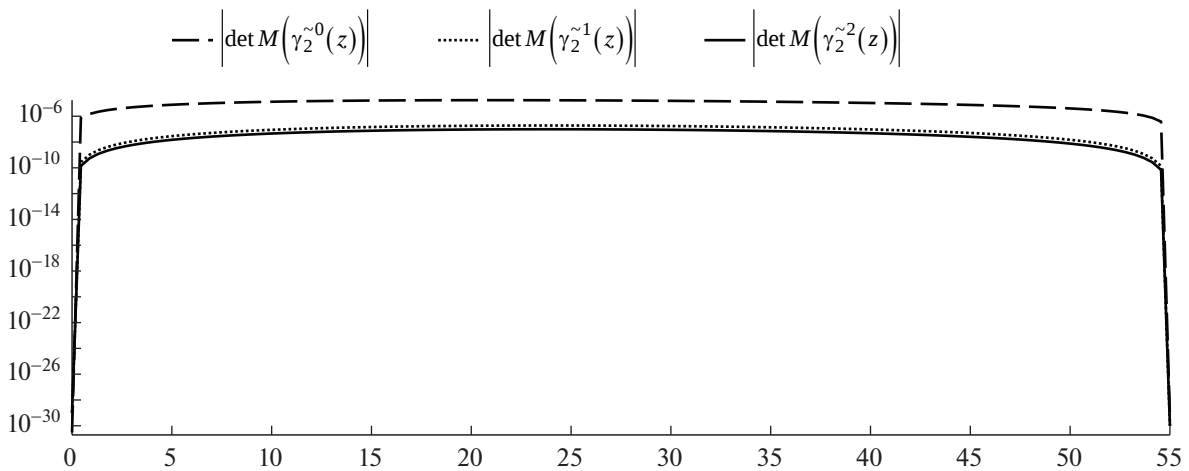
Фиг. 2. Графики величин  $\gamma_j^0(z)$ ,  $j = \overline{1,4}$ , в волноводном переходе при  $z \in [0, L]$ .



Фиг. 3. Графики величин  $\text{Im}(\gamma_j^1(z))$ ,  $j = \overline{1,4}$ , в волноводном переходе при  $z \in [0, L]$ .



Фиг. 4. Графики величин  $\operatorname{Re}(\gamma_j^2(z))$ ,  $j = \overline{1, 4}$ , в волноводном переходе при  $z \in [0, L]$ .



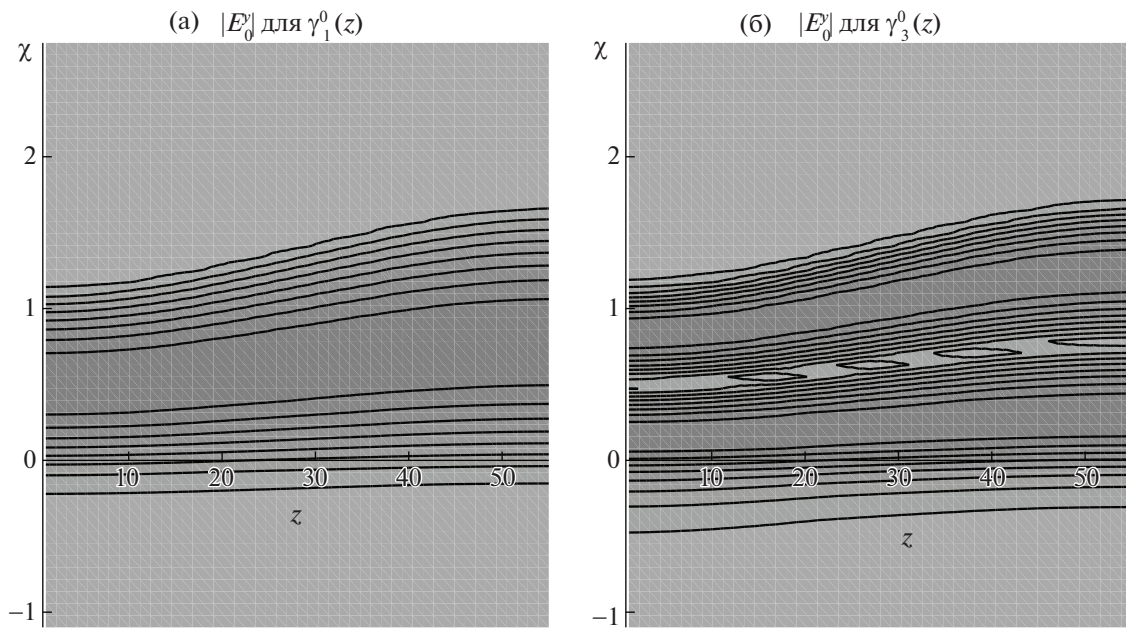
Фиг. 5. График невязки детерминанта  $|\det M(\tilde{\gamma}_2^{0,1,2}(z))|$  при  $z \in [0, L]$ .

Учитывая приближенный характер вычисленных  $\gamma_j^{0,1,2}(z)$ , приведем график невязки детерминанта  $|\det M(\tilde{\gamma}^{0,1,2}(z))|$  для нулевого, первого и второго приближений, т.е. для  $\tilde{\gamma}^0(z) = \gamma^0(z)$ ,  $\tilde{\gamma}^1(z) = \gamma^0(z) + h'(z)\gamma^1(z)$  и  $\tilde{\gamma}^2(z) = \gamma^0(z) + h'(z)\gamma^1(z) + h^2(z)\gamma^2(z)$ .

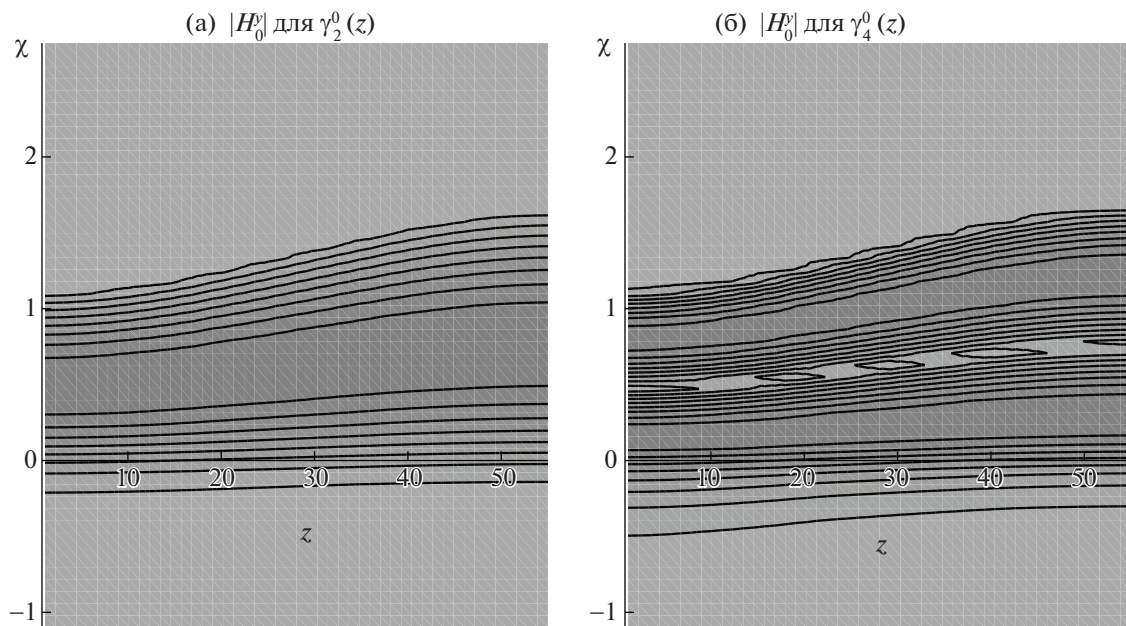
На фиг. 5 изображен график невязки детерминанта для нулевого, первого и второго приближений величины  $\gamma_2^{0,1,2}(z)$ . С увеличением порядка приближения невязка уменьшается. Аналогично выглядит график невязки детерминанта и для  $\gamma_2^{0,1,2}(z)$ , поэтому он не приводится в настоящем разделе. Невязки для  $\gamma_{1,3}(z)$ , для которых вклады первого и второго порядков отсутствуют, имеют порядок  $10^{-30}$  и также не приводятся в настоящем разделе.

#### 4.2. Приближенное решение системы граничных уравнений

Вычисление амплитудных компонент полей базируется на использовании алгоритма символического решения систем линейных уравнений из [15] для решения системы граничных уравнений (15). Уравнения системы громоздкие, поэтому в настоящем разделе не приводятся.



Фиг. 6. Графики абсолютных значений компоненты  $E_0^y$  для  $\gamma_1^0(z)$  и для  $\gamma_3^0(z)$  при  $z \in [0, L]$ .



Фиг. 7. Графики абсолютных значений компоненты  $H_0^y$  для  $\gamma_2^0(z)$  и для  $\gamma_4^0(z)$  при  $z \in [0, L]$ .

Систему (15) для рассматриваемой структуры (параметры структуры см. в п. 3.2.3) решаем символично в системе компьютерной алгебры Maple. В результате получаем символичные решения, включающие множество символьных констант, характеризующих рассматриваемую структуру. Важно заметить, что решения, отвечающие  $\gamma_{1,3}(z)$  и решения, отвечающие  $\gamma_{2,4}(z)$ , различаются между собой. В полученные решения подставляем численные значения параметров рассматриваемой структуры, а также численные значения найденных приближенно  $\gamma_j(z)$  и получаем приближенные решения системы граничных уравнений (15), используя которые мож-

но построить искомые в задаче (10) амплитудные компоненты электромагнитных полей. На фиг. 6–7 приведены амплитудные компоненты полей, построенные для  $\gamma_j^0(z)$ .

Амплитудные компоненты полей, отвечающие  $\gamma_{1,3}(z)$ , не имеют вкладов первого и второго порядков малости, так как для  $\gamma_{1,3}(z)$  таковые вклады отсутствуют. Амплитудные компоненты полей, отвечающие  $\gamma_{2,4}(z)$  характеризуются наличием вкладов первого и второго порядков малости. Графические различия амплитудных компонент полей  $H_0^y, E_0^z$ , построенных в нулевом и страших приближениях, малы и визуально не заметны, поэтому приведем далее вычисленные значения максимумов модулей разности  $H_0^y$ , посчитанной в нулевом приближении и  $H_0^y$  в первом приближении, а также в первом и втором приближениях. Аналогичные численные оценки приведем и для компоненты  $E_0^z$ . Ниже приведены расчеты для  $\gamma_2(z)$ :

$$\begin{aligned} \max |H_0^y(x, z; \tilde{\gamma}_2^1) - H_0^y(x, z; \tilde{\gamma}_2^0)| &\leq 1.5 \times 10^{-2}, & \max |E_0^z(x, z; \tilde{\gamma}_2^1) - E_0^z(x, z; \tilde{\gamma}_2^0)| &\leq 1.1 \times 10^{-3}, \\ \max |H_0^y(x, z; \tilde{\gamma}_2^2) - H_0^y(x, z; \tilde{\gamma}_2^1)| &\leq 1.3 \times 10^{-5}, & \max |E_0^z(x, z; \tilde{\gamma}_2^2) - E_0^z(x, z; \tilde{\gamma}_2^1)| &\leq 7.8 \times 10^{-7}. \end{aligned} \quad (23)$$

Далее приведены аналогичные расчеты для  $\gamma_4(z)$ :

$$\begin{aligned} \max |H_0^y(x, z; \tilde{\gamma}_4^1) - H_0^y(x, z; \tilde{\gamma}_4^0)| &\leq 1.4 \times 10^{-2}, & \max |E_0^z(x, z; \tilde{\gamma}_4^1) - E_0^z(x, z; \tilde{\gamma}_4^0)| &\leq 2.1 \times 10^{-3}, \\ \max |H_0^y(x, z; \tilde{\gamma}_4^2) - H_0^y(x, z; \tilde{\gamma}_4^1)| &\leq 2.1 \times 10^{-5}, & \max |E_0^z(x, z; \tilde{\gamma}_4^2) - E_0^z(x, z; \tilde{\gamma}_4^1)| &\leq 4.1 \times 10^{-6}. \end{aligned} \quad (24)$$

## 5. ОБСУЖДЕНИЕ

В рамках приближенного вычисления корней нелинейного уравнения вычислена функция  $\gamma_j(z)$ , описывающая переменный коэффициент фазового замедления в нулевом и старших приближениях для  $j$ -й волноводной моды. Нулевое приближение учитывает только переменную толщину волноводного слоя и не учитывает вклады более высоких порядков, описывающие малый наклон криволинейной границы раздела. Результаты, полученные в нулевом приближении, качественно соответствуют результатам метода поперечных сечений [18], [19], а именно искомая функция  $\gamma_j^0(z)$  получилась вещественнозначной (см. фиг. 2).

Более интересны результаты первого приближения: поправка первого порядка малости  $\gamma_j^1(z)$  в разложении

$$\gamma_j(z) = \gamma_j^0(z) + h'(z)\gamma_j^1(z) + h'^2(z)\gamma_j^2(z) + \dots$$

оказалась мнимой величиной для  $TM$ -моды и машинным нулем для  $TE$ -моды (см. фиг. 3). Второе приближение также не внесло вклада в  $\gamma_{1,3}(z)$ , отвечающих  $TE$ -моды (см. фиг. 4), что также подтверждается и невязкой детерминанта, который принимает значения порядка  $10^{-30}$  для  $\gamma_{1,3}^0(z)$ . Вклад второго порядка малости в  $\gamma_{2,4}(z)$ , отвечающих  $TM$ -моды, нетривиальный (см. фиг. 4), но, в отличие от вклада первого порядка, вещественный.

Фактически из непрерывности магнитной проницаемости  $\mu$  на криволинейной границе средствами компьютерной алгебры (прямой подстановкой) было получено, что в решении для  $TE$ -моды обнуляются все слагаемые, содержащие  $h'(z)$ . Поэтому и другие приближения — третьего порядка и выше — также отсутствуют.

Другими словами, даже малый наклон криволинейной границы раздела вносит нетривиальный вклад в коэффициент фазового замедления  $TM$ -моды, и не вносит вклада в коэффициент фазового замедления  $TE$ -моды —  $TE$ -мода вполне описывается в рамках нулевого приближения по малому наклону границы раздела.

Учет наклона границы раздела делает коэффициент фазового замедления для  $TM$ -моды комплексной величиной даже для идеального случая, когда показатели преломления вещественные. Напомним, что  $\gamma(z) = \varphi'(z)$  и в искомое поле вида (1) входит экспоненциальный множитель

$\exp\{i\omega t - ik_0\varphi(z)\}$ . Следовательно, искомое электромагнитное поле в волноводе будет включать множитель

$$\begin{aligned}\exp\left\{i\omega t - ik_0\int_{z_0}^z \gamma(s) ds\right\} &\approx \exp\left\{i\omega t - ik_0\int_{z_0}^z \gamma^0(s) ds + k_0\int_{z_0}^z h'(s) \operatorname{Im}(\gamma^1(s)) ds\right\} = \\ &= \exp\left\{k_0\int_{z_0}^z h'(s) \operatorname{Im}(\gamma^1(s)) ds\right\} \exp\left\{i\omega t - ik_0\int_{z_0}^z \gamma^0(s) ds\right\}.\end{aligned}$$

Первый сомножитель в полученном произведении экспонент будет описывать некую “накачку”, если подынтегральная функция положительна, и некое “поглощение”, если она отрицательна. В случае рассматриваемой структуры (см. в п. 3.2.3) имеет место  $\operatorname{Im}(\gamma^1(s)) > 0$  для  $TM$ -моды согласно фиг. 3,  $h'(z) \geq 0$  в области волноводного перехода  $z \in [0, L]$ . Отметим также, что  $h'(z) = 0$  при  $z \geq L$ , а значит, в области волноводного перехода  $z \in [0, L]$  имеет место “накачка”, которая прекратится при переходе в область регулярного волновода  $z \geq L$ .

Расчет амплитудных составляющих полей, изображенных на фиг. 6, 7 в нулевом приближении, а также вкладов первого и второго порядков (23), (24) показывает, что наличие вкладов первого и второго порядков не сильно сказывается в численном отношении.

## 6. ЗАКЛЮЧЕНИЕ

В работе построены направляемые моды плавно-нерегулярного волноводного перехода между двумя регулярными волноводами, используя модель адиабатических волноводных мод. Описан и реализован в системе компьютерной алгебры символьно-численный алгоритм решения дисперсионных уравнений с комплексными корнями. С помощью реализованного алгоритма построены дисперсионные кривые в нулевом, первом и втором приближениях по малому параметру, характеризующему малость угла наклона границы в радианах.

Модель адиабатических волноводных мод позволяет сформулировать задачу расчета направляемых мод в символьном виде и решить ее также с использованием символьных манипуляций и алгоритмов, благодаря чему метод решения реализован в системе компьютерной алгебры Maple.

Использование компьютерной алгебры позволило изучить качественно и численно адиабатические волноводные моды в нулевом приближении с учетом малого наклона границы раздела волноводного слоя для волноводного перехода. В статье представлены решения, описывающие адиабатические волноводные моды в нулевом приближении с учетом малого наклона границы раздела волноводного слоя, качественно отличающиеся от решений, не учитывающих наклон границы раздела.

## СПИСОК ЛИТЕРАТУРЫ

1. *Stevenson A.F.* General Theory of Electromagnetic Horns // J. Appl. Phys. 1951. V. 22. № 12. P. 1447.
2. *Schellkunoff S.A.* Conversion of Maxwell's equations into generalized telegraphist's equations // Bell Syst. Tech. J. 1955. V. 34. P. 995–1043.
3. *Каценеленбаум Б.З.* Теория нерегулярных волноводов с медленно меняющимися параметрами. Москва: АН СССР, 1961.
4. *Katsenelenbaum B.Z., Mercader del Rio L., Pereyaslavets M., Sorolla Ayza M., Thumm M.* Theory of Nonuniform Waveguides: the cross-section method. The Institution of Engineering and Technology, London, 1998.
5. *Шевченко В.В.* Плавные переходы в открытых волноводах: введение в теорию. М.: Наука, 1969.
6. *Свешников А.Г.* Приближенный метод расчета слабо нерегулярного волновода // Докл. АН СССР. 1956. Т. 80. № 3. С. 345–347.
7. *Свешников А.Г.* К обоснованию методов расчета нерегулярных волноводов // Ж. вычисл. матем. и матем. физ. 1963. Т. 3. № 1. С. 170–179.
8. *Fedoryuk M.V.* A justification of the method of transverse sections for an acoustic wave guide with nonhomogeneous content // U.S.S.R. Comput. Math. Math. Phys. 1973. V. 13. № 1. P. 162–173.
9. *Иванов А.А., Шевченко В.В.* Плоскопоперечный стык двух планарных волноводов // Радиотехн. и электроника. 2009. Т. 54. № 1. С. 68–77.
10. *Sevastianov L.A., Egorov A.A.* Theoretical analysis of the waveguide propagation of electromagnetic waves in dielectric smoothlyirregular integrated structures // Optics and Spectroscopy. 2008. V. 105. № 4. P. 576–584.

11. *Egorov A.A., Sevastianov L.A.* Structure of modes of a smoothly irregular integrated optical four-layer three-dimensional waveguide // *Quantum Electronics*. 2009. V. 39. № 6. P. 566–574.
12. *Egorov A.A., Lovetskiy K.P., Sevastianov A.L., Sevastianov L.A.* Simulation of guided modes (eigenmodes) and synthesis of a thin-film generalised waveguide Luneburg lens in the zero-order vector approximation // *Quantum Electronics*. 2010. V.40. № 9. P. 830–836.
13. *Бабич В.М., Булдырев В.С.* Асимптотические методы в задачах дифракции коротких волн. Метод эталонных задач. М.: Наука, 1972.
14. *Divakov D.V., Sevastianov A.L.* The Implementation of the Symbolic-Numerical Method for Finding the Adiabatic Waveguide Modes of Integrated Optical Waveguides in CAS Maple // *Lecture Notes in Computer Science*. 2019. V. 11661. P. 107–121.
15. *Divakov D.V., Tyutyunnik A.A.* Symbolic Investigation of the Spectral Characteristics of Guided Modes in Smoothly Irregular Waveguides // *Programming and Computer Software*. 2022. V. 48. № 2. P. 80–89.
16. *Adams M.J.* An Introduction to Optical Waveguides. Wiley, New York, 1981.
17. Maple homepage, <https://www.maplesoft.com/>. Last accessed 24 May 2022
18. *Gevorkyan M., Kulyabov D., Lovetskiy K., Sevastianov L., Sevastianov A.* Field calculation for the horn waveguide transition in the single-mode approximation of the cross-sections method // *Proceedings of SPIE*. 2017. V. 10337. P. 103370H.
19. *Divakov D.V., Lovetskiy K.P., Sevastianov L.A., Tiutiunnik A.A.* A single-mode model of cross-sectional method in a smoothly irregular transition between planar thin-film dielectric waveguides // *Proceedings of SPIE*. 2021. V. 11846. P. 118460T.

# УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.95

## О ЗАДАЧАХ КОШИ ДЛЯ НЕЛИНЕЙНЫХ СОБОЛЕВСКИХ УРАВНЕНИЙ ТЕОРИИ СЕГНЕТОЭЛЕКТРИЧЕСТВА

© 2023 г. М. О. Корпусов<sup>1,\*</sup>, Р. С. Шафир<sup>1,\*\*</sup><sup>1</sup> 119991 Москва, Ленинские горы, 1, МГУ им. М.В. Ломоносова, Россия

\*e-mail: korpusov@gmail.com

\*\*e-mail: romanshafir@mail.ru

Поступила в редакцию 01.08.2021 г.

Переработанный вариант 05.05.2022 г.

Принята к публикации 04.08.2022 г.

Исследуются две задачи Коши для нелинейных соболевских уравнений:  $\frac{\partial^2}{\partial t^2} \frac{\partial^2 u}{\partial x_3^2} + \Delta u = |u|^q$  и

$\frac{\partial^2}{\partial t^2} \Delta_{\perp} u + \Delta u = |u|^q$ . Найдены условия, при которых существуют слабые обобщенные локальные во времени решения задач Коши, а также происходит разрушение слабых решений этих же задач Коши. Библ. 15.

**Ключевые слова:** нелинейные уравнения соболевского типа, разрушение, blow-up, локальная разрешимость, нелинейная емкость.

DOI: 10.31857/S0044466922120092, EDN: LMBXET

### 1. ВВЕДЕНИЕ

В настоящей работе изучаются две задачи Коши. Первая задача Коши имеет вид

$$\frac{\partial^2}{\partial t^2} \frac{\partial^2 u}{\partial x_3^2} + \Delta u = |u|^q, \quad (x, t) \in \mathbb{R}^3 \otimes (0, T], \quad (1.1)$$

$$u(x, 0) = u_0(x), \quad u'(x, 0) = u_1(x), \quad x \in \mathbb{R}^3. \quad (1.2)$$

Вторая задача Коши имеет вид

$$\frac{\partial^2}{\partial t^2} \Delta_{\perp} u + \Delta u = |u|^q, \quad (x, t) \in \mathbb{R}^3 \otimes (0, T], \quad (1.3)$$

$$u(x, 0) = u_0(x), \quad u'(x, 0) = u_1(x), \quad x \in \mathbb{R}^3. \quad (1.4)$$

В формуле (1.3) под оператором  $\Delta_{\perp}$  понимается  $\Delta_{\perp} := \partial^2 / \partial x_1^2 + \partial^2 / \partial x_2^2$ .

В работе исследуются слабые и слабые обобщенные локальные во времени решения данных задач Коши. Доказывается существование слабых обобщенных решений при  $q > 4$ . При  $1 < q \leq 3$  доказано разрушение слабых решений.

Отметим работу [1], в которой исследовалась задача Коши с той же нелинейностью, но другим дифференциальным оператором, имеющим вид

$$A_{x,t}[u](x, t) := \frac{\partial^2}{\partial t^2} \Delta_{\perp} u(x, t) + \frac{\partial^2 u}{\partial x_3^2}(x, t).$$

Уравнения (1.1), (1.3) относятся к уравнениям соболевского типа (см. [2]), исследованию которых посвящено большое количество работ. Например, в работах Г.А. Свиридюка, С.А. Загребинной, А.А. Замышляевой [3]–[5] были рассмотрены в общем виде и в виде примеров начально-краевые задачи для различных типов линейных и нелинейных уравнений соболевского типа.

Теория потенциала для неклассических уравнений типа Соболева впервые была рассмотрена в работе Б.В. Капитонова [6], а в дальнейшем изучалась в работах С.А. Габова и А.Г. Свешникова [7], [8], а также в работах их учеников (см. работу Ю.Д. Плетнера [9]).

В классической работе [10] С.И. Похожаева и Э. Митидиери методом нелинейной емкости были получены глубокие результаты о роли критических показателей. Также можно отметить работы Е.И. Галахова и О.А. Салиевой [11] и [12].

Настоящая работа продолжает исследования, начатые в [13]–[15] и посвященные получению критических показателей для решений задач Коши для нелинейных уравнений соболевского типа.

## 2. ВЫВОД УРАВНЕНИЙ

Пусть  $\{O, \mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$  — некоторая декартова прямоугольная система координат в пространстве. При рассмотрении анизотропного сегнетоэлектрика вблизи температуры Кюри с анизотропией вдоль оси  $Oz$  и с учетом модельной временной дисперсии справедливы следующие уравнения:

$$D_x = \varepsilon_x E_x + \varepsilon_0 \int_0^t (t - \tau) E_x(\tau) d\tau, \quad (2.1)$$

$$D_y = \varepsilon_y E_y + \varepsilon_0 \int_0^t (t - \tau) E_y(\tau) d\tau, \quad (2.2)$$

$$D_z = \varepsilon_z E_z + \varepsilon_0 \int_0^t (t - \tau) E_z(\tau) d\tau, \quad (2.3)$$

причем

$$\max\{|\varepsilon_x|, |\varepsilon_y|\} \ll |\varepsilon_z| \quad (2.4)$$

где

$$\mathbf{D} = D_x \mathbf{e}_x + D_y \mathbf{e}_y + D_z \mathbf{e}_z. \quad (2.5)$$

Кроме того, рассмотрим электрическую часть системы уравнений Максвелла в квазистационарном приближении:

$$\operatorname{div} \mathbf{D} = -4\pi n, \quad \mathbf{E} = -\nabla \phi, \quad (2.6)$$

$$\frac{\partial n}{\partial t} = q_0 \int_0^t |\phi|^q(\tau) d\tau, \quad q > 1. \quad (2.7)$$

Из уравнений (2.1)–(2.7) получим дифференциальное следствие

$$\varepsilon_x \phi_{xxtt} + \varepsilon_y \phi_{yytt} + \varepsilon_z \phi_{zztt} + \varepsilon_0 \Delta \phi = 4\pi q_0 |\phi|^q, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (2.8)$$

С учетом выражения (2.4) приходим к выводу о том, что уравнение (2.8) можно упростить, формально положив

$$\varepsilon_x = \varepsilon_y = 0.$$

Приходим к уравнению

$$\varepsilon_z \phi_{zztt} + \varepsilon_0 \Delta \phi = 4\pi q_0 |\phi|^q. \quad (2.9)$$

Кроме того, мы в работе рассмотрим следующее уравнение:

$$\varepsilon_x \phi_{xxtt} + \varepsilon_y \phi_{yytt} + \varepsilon_0 \Delta \phi = 4\pi q_0 |\phi|^q, \quad (2.10)$$

которое получается из (2.8), если формально положить

$$\varepsilon_z = 0.$$

## 3. ОБОЗНАЧЕНИЯ

Под классом функций  $u(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$  при  $\gamma_1 \geq 0$  и  $\gamma_2 \geq 0$  мы понимаем такие функции  $u(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ , что конечна следующая норма:

$$\|u\|_T := \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} (1 + x_1^2 + x_2^2)^{\gamma_1/2} (1 + x_3^2)^{\gamma_2/2} |u(x, t)| < +\infty. \quad (3.1)$$

Можно доказать, что  $\mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$  является банаховым пространством относительно нормы (3.1).

Под классом функций  $\mathbb{C}_b^{(0,1)}(\mathbb{R}^3 \otimes [0, T])$  мы понимаем такие функции  $u(x, t)$ , что

$$u(x, t), \quad \frac{\partial u(x, t)}{\partial t} \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]). \quad (3.2)$$

Можно доказать, что это пространство банахово относительно нормы

$$\sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} \left[ |u(x, t)| + \left| \frac{\partial u(x, t)}{\partial t} \right| \right] < +\infty. \quad (3.3)$$

Под классом функций  $\mathbb{C}^{2+2}(\mathbb{R}^3 \otimes (0, T])$  мы понимаем такие функции  $u(x, t)$ , что

$$D_t^k D_x^\alpha u(x, t) \in \mathbb{C}(\mathbb{R}^3 \otimes (0, T]), \quad (3.4)$$

$$D_t^k = \partial_t^k, \quad D_x^\alpha = \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \partial_{x_3}^{\alpha_3}, \quad \alpha = (\alpha_1, \alpha_2, \alpha_3), \quad (3.5)$$

$$\alpha_j \in \{0, 1, 2\}, \quad j = 1, 2, 3,$$

$$|\alpha| := \alpha_1 + \alpha_2 + \alpha_3 \leq 2, \quad k \in \{0, 1, 2\}, \quad (3.6)$$

причем всевозможные смешанные производные вида (3.4) перестановочны.

Под классом функций  $\mathbb{C}_0^{2+2}(\mathbb{R}^3 \otimes [0, T])$  подразумеваем такие функции  $u(x, t) \in \mathbb{C}_b^{2+2}(\mathbb{R}^3 \otimes [0, T])$ , что

$$D_t^k D_x^\alpha u(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]), \quad (3.7)$$

а также  $u(x, T) = u'(x, T) = 0$  для всех  $x \in \mathbb{R}^3$ , и носитель  $\text{supp } u(x, t)$  — компакт в  $\mathbb{R}^3 \otimes [0, T]$ .

Под классом функций  $\mathbb{C}_b^{2+1}(\mathbb{R}^3 \otimes [0, T])$  мы понимаем такие функции  $u(x, t)$ , что

$$D_t^k D_x^\alpha u(x, t) \in \mathbb{C}(\mathbb{R}^3 \otimes [0, T]), \quad (3.8)$$

$$D_t^k = \partial_t^k, \quad D_x^\alpha = \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \partial_{x_3}^{\alpha_3}, \quad \alpha = (\alpha_1, \alpha_2, \alpha_3), \quad (3.9)$$

$$\alpha_j \in \{0, 1, 2\}, \quad j = 1, 2, 3,$$

$$|\alpha| := \alpha_1 + \alpha_2 + \alpha_3 \leq 2, \quad k \in \{0, 1\}, \quad (3.10)$$

причем всевозможные смешанные производные вида (3.8) перестановочны.

Под классом функций  $\mathbb{C}_b^{(0,1)}((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$  мы понимаем такие функции  $u(x, t)$ , что

$$u(x, t), \quad \frac{\partial u(x, t)}{\partial t} \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T]). \quad (3.11)$$

Можно доказать, что это пространство банахово относительно нормы

$$\sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} (1 + x_1^2 + x_2^2)^{\gamma_1/2} (1 + x_3^2)^{\gamma_2/2} \left( |u(x, t)| + \left| \frac{\partial u(x, t)}{\partial t} \right| \right) < +\infty. \quad (3.12)$$

Будем говорить, что пара функций  $\{u_0(x), u_1(x)\}$  принадлежит классу  $H_{x_3}$ , если  $u_0(x), u_1(x) \in L^1_{\text{loc}}(\mathbb{R}^3)$  и найдется такой шар  $O(x_0, R) \subset \mathbb{R}^3$  положительного радиуса  $R > 0$ , что  $u_0(x), u_1(x) \in H^2(O(x_0, R))$ , и имеет место неравенство

$$\left(\frac{\partial^2 u_0(x)}{\partial x_3^2}\right)^2 + \left(\frac{\partial^2 u_1(x)}{\partial x_3^2}\right)^2 > 0 \quad \text{для почти всех } x \in O(x_0, R). \quad (3.13)$$

Будем говорить, что пара функций  $\{u_0(x), u_1(x)\}$  принадлежит классу  $H_{x_1 x_2}$ , если  $u_0(x), u_1(x) \in L^1_{\text{loc}}(\mathbb{R}^3)$  и найдется такой шар  $O(x_0, R) \subset \mathbb{R}^3$  положительного радиуса  $R > 0$ , что  $u_0(x), u_1(x) \in H^2(O(x_0, R))$ , и имеет место неравенство

$$(\Delta_{\perp} u_0(x))^2 + (\Delta_{\perp} u_1(x))^2 > 0 \quad \text{для почти всех } x \in O(x_0, R). \quad (3.14)$$

#### 4. ФУНДАМЕНТАЛЬНЫЕ РЕШЕНИЯ И СВЯЗАННЫЕ С НИМИ ОЦЕНКИ

Рассмотрим первый дифференциальный оператор, действие которого определяется равенством

$$\mathfrak{M}_{x,t}[u](x, t) := \frac{\partial^2}{\partial t^2} \frac{\partial^2}{\partial x_3^2} u(x, t) + \Delta u(x, t), \quad (4.1)$$

где  $(x, t) = (x_1, x_2, x_3, t) \in \mathbb{R}^4_+ := \mathbb{R}^3 \otimes \mathbb{R}^1_+$ .

Найдем фундаментальное решение оператора  $\mathfrak{M}_{x,t}$ , т.е. найдем решение следующего уравнения в смысле пространства обобщенных функций:

$$\mathfrak{M}_{x,t}[\mathcal{E}](x, t) = \delta(t)\delta(x). \quad (4.2)$$

Применим преобразование Лапласа к обеим частям равенства (4.2) и получим уравнение, понимаемое в смысле обобщенных функций:

$$p^2 \frac{\partial^2}{\partial x_3^2} \bar{\mathcal{E}}(x, p) + \Delta \bar{\mathcal{E}}(x, p) = \delta(x). \quad (4.3)$$

Одним из решений данного уравнения является функция

$$\bar{\mathcal{E}}(x, p) = -\frac{1}{4\pi} \frac{1}{\sqrt{x_1^2 + x_2^2}} \frac{1}{\sqrt{p^2 + \frac{x_1^2 + x_2^2 + x_3^2}{x_1^2 + x_2^2}}}. \quad (4.4)$$

Применим обратное преобразование Лапласа и получим следующий вид фундаментального решения:

$$\mathcal{E}(x, t) = -\frac{\theta(t)}{4\pi} \frac{1}{\sqrt{x_1^2 + x_2^2}} J_0(\beta(x)t), \quad \beta(x) := \sqrt{\frac{x_1^2 + x_2^2 + x_3^2}{x_1^2 + x_2^2}}, \quad (4.5)$$

где  $\theta(t)$  — функция Хевисайда.

Для функции Бесселя справедлива оценка

$$|J_0(y)| \leq \frac{c_0}{\sqrt{|y|}} \quad \text{для всех } y \neq 0, \quad (4.6)$$

где  $c_0 > 0$  — некоторая постоянная. Тогда для функции  $\mathcal{E}(x, t)$  справедлива следующая цепочка оценок:

$$|\mathcal{E}(x, t)| \leq \frac{c_1}{(x_1^2 + x_2^2)^{1/4} (x_1^2 + x_2^2 + x_3^2)^{1/4} \sqrt{t}} \leq \frac{c_1}{|x_3|^{1/2} (x_1^2 + x_2^2)^{1/4} \sqrt{t}} \quad (4.7)$$

при  $x_3 \neq 0, \quad (x_1, x_2) \neq (0, 0), \quad t > 0, \quad c_1 = \frac{c_0}{4\pi} > 0.$

Кроме того, при  $(x_1, x_2) \neq (0, 0)$  и  $t \geq 0$  функция  $\mathcal{E}(x, t)$  дифференцируема по переменной  $t$ , причем

$$\frac{\partial \mathcal{E}(x, t)}{\partial t} = \frac{1}{4\pi} \frac{\sqrt{x_1^2 + x_2^2 + x_3^2}}{x_1^2 + x_2^2} J_1(\beta(x)t) \quad \text{при} \quad (x_1, x_2) \neq (0, 0), \quad t \geq 0. \quad (4.8)$$

При получении формулы (4.8) было использовано соотношение

$$\frac{dJ_0(x)}{dx} = -J_1(x). \quad (4.9)$$

Для функции Бесселя  $J_1(x)$  справедлива оценка, аналогичная оценке (4.6) для функции  $J_0(x)$ , поэтому из (4.6), (4.8) получаем оценку

$$\left| \frac{\partial \mathcal{E}(x, t)}{\partial t} \right| \leq c_2 \frac{(x_1^2 + x_2^2 + x_3^2)^{1/4}}{(x_1^2 + x_2^2)^{3/4} \sqrt{t}} \quad \text{при} \quad (x_1, x_2) \neq (0, 0), \quad t > 0, \quad (4.10)$$

где  $c_2 > 0$  — некоторая постоянная.

Из оценок (4.7), (4.10) следует, что

$$\mathcal{E}(x, t), \quad \frac{\partial \mathcal{E}(x, t)}{\partial t} \in L^1_{\text{loc}}(\mathbb{R}^4_+). \quad (4.11)$$

Теперь рассмотрим второй дифференциальный оператор, действие которого определяется равенством

$$\mathfrak{L}_{x,t}[u](x, t) := \frac{\partial^2}{\partial t^2} \Delta_{\perp} u(x, t) + \Delta u(x, t), \quad (4.12)$$

где  $(x, t) = (x_1, x_2, x_3, t) \in \mathbb{R}^4_+ := \mathbb{R}^3 \otimes \mathbb{R}^1_+$ .

Найдем фундаментальное решение оператора  $\mathfrak{L}_{x,t}$ , т.е. найдем решение следующего уравнения в смысле пространства обобщенных функций:

$$\mathfrak{L}_{x,t}[\mathcal{E}_1](x, t) = \delta(t)\delta(x). \quad (4.13)$$

Применим преобразование Лапласа к обеим частям равенства (4.13) и получим следующее уравнение, понимаемое в смысле обобщенных функций:

$$(p^2 + 1) \frac{\partial^2}{\partial x_1^2} \overline{\mathcal{E}_1}(x, p) + (p^2 + 1) \frac{\partial^2}{\partial x_2^2} \overline{\mathcal{E}_1}(x, p) + \frac{\partial^2}{\partial x_3^2} \overline{\mathcal{E}_1}(x, p) = \delta(x). \quad (4.14)$$

Одним из решений данного уравнения является функция

$$\overline{\mathcal{E}_1}(x, p) = -\frac{1}{4\pi|x_3|} \frac{1}{\sqrt{p^2 + 1}} \frac{1}{\sqrt{p^2 + \left(1 + \frac{x_1^2 + x_2^2}{x_3^2}\right)}}. \quad (4.15)$$

Применим обратное преобразование Лапласа и получим следующий вид фундаментального решения:

$$\mathcal{E}_1(x, t) = -\frac{\theta(t)}{4\pi|x_3|} \int_0^t J_0(\beta(x)\tau) J_0(t - \tau) d\tau, \quad \beta(x) := \sqrt{1 + \frac{x_1^2 + x_2^2}{x_3^2}}, \quad (4.16)$$

где  $\theta(t)$  — функция Хевисайда.

Снова воспользуемся оценкой (4.6) для функции Бесселя  $J_0$  и получим оценку для фундаментального решения  $\mathcal{E}_1(x, t)$ :

$$\begin{aligned} |\mathcal{E}_1(x, t)| &\leq \frac{c_1}{|x_3|^{1/2} (x_1^2 + x_2^2 + x_3^2)^{1/4}} \int_0^t \frac{1}{\sqrt{\tau} \sqrt{t - \tau}} d\tau \\ \text{при} \quad x_3 &\neq 0, \quad (x_1, x_2) \neq (0, 0), \quad t > 0, \quad c_1 > 0. \end{aligned} \quad (4.17)$$

Заметим, что

$$\int_0^t \frac{1}{\sqrt{\tau}\sqrt{t-\tau}} d\tau = \int_0^1 \frac{1}{\sqrt{s}\sqrt{1-s}} ds < +\infty \quad \text{при всех} \quad t : 0 \leq t < +\infty, \quad (4.18)$$

поэтому оценку (4.17) можно переписать следующим образом:

$$|\mathcal{E}_1(x, t)| \leq \frac{c_2}{|x_3|^{1/2} (x_1^2 + x_2^2)^{1/4}} \quad (4.19)$$

при  $x_3 \neq 0, \quad (x_1, x_2) \neq (0, 0), \quad t > 0, \quad c_2 > 0.$

Также заметим, что при  $x_3 \neq 0$  и  $t \geq 0$  функция  $\mathcal{E}_1(x, t)$  дифференцируема по переменной  $t$ , и справедливо равенство

$$\frac{\partial \mathcal{E}_1(x, t)}{\partial t} = \frac{1}{4\pi|x_3|} \int_0^t J_0(\beta(x)\tau) J_1(t-\tau) d\tau - \frac{1}{4\pi|x_3|} J_0(\beta(x)t) \quad \text{при} \quad x_3 \neq 0, \quad t \geq 0. \quad (4.20)$$

С помощью оценки (4.6) для функции Бесселя  $J_0$  и аналогичной оценке для функции Бесселя  $J_1$  нетрудно получить оценку

$$\left| \frac{\partial \mathcal{E}_1(x, t)}{\partial t} \right| \leq \frac{c_0}{|x_3|^{1/2} (x_1^2 + x_2^2)^{1/4}} \left( 1 + \frac{1}{\sqrt{t}} \right) \quad \text{при} \quad x_3 \neq 0, \quad t > 0, \quad c_0 > 0. \quad (4.21)$$

Из оценок (4.19), (4.21) следует, что  $\mathcal{E}_1(x, t), \frac{\partial \mathcal{E}_1(x, t)}{\partial t} \in L^1_{\text{loc}}(\mathbb{R}^4_+)$ .

В дальнейшем нам понадобятся две леммы.

**Лемма 1.** При  $\gamma_1 > 2$  и  $\gamma_2 > 1$  найдется такая постоянная  $M_1 = M_1(\gamma_1, \gamma_2) > 0$ , что будет выполнено неравенство

$$I := \int_0^t \int_{\mathbb{R}^3} \frac{|\mathcal{E}(x-y, t-\tau)|}{(1+y_1^2+y_2^2)^{\gamma_1/2} (1+y_3^2)^{\gamma_2/2}} dy d\tau \leq \frac{M_1 \sqrt{t}}{(1+x_1^2+x_2^2)^{1/4} (1+x_3^2)^{1/4}} \quad (4.22)$$

при  $(x, t) = (x_1, x_2, x_3, t) \in \mathbb{R}^3 \otimes [0, T]$ .

**Доказательство.** Применим оценку (4.7) для оценки интеграла  $I$ . Тогда получим

$$I \leq c_1 I_t I_x, \quad (4.23)$$

где  $c_1 > 0$ , а интегралы  $I_t, I_x$  определяются, как

$$I_t := \int_0^t \frac{d\tau}{\sqrt{t-\tau}} = 2\sqrt{t}, \quad (4.24)$$

$$I_x := \int_{\mathbb{R}^3} \frac{dy_1 dy_2 dy_3}{|x_3 - y_3|^{1/2} [(x_1 - y_1)^2 + (x_2 - y_2)^2]^{1/4} (1+y_1^2+y_2^2)^{\gamma_1/2} (1+y_3^2)^{\gamma_2/2}}. \quad (4.25)$$

Для оценки интеграла  $I_x$  нужно перейти в цилиндрическую систему координат и воспользоваться [1, леммы 1, 2]. В результате получим, что

$$I_x \leq \frac{K_1}{(1+x_1^2+x_2^2)^{1/4}} \frac{K_2}{(1+x_3^2)^{1/4}}, \quad K_1, K_2 > 0, \quad (4.26)$$

причем, постоянные  $K_1, K_2$  не зависят от  $x$ .

Из формул (4.23), (4.24), (4.26) и следует доказательство леммы.

**Лемма 2.** При  $\gamma_1 > 2$  и  $\gamma_2 > 1$  найдутся такие постоянные  $M_1 = M_1(\gamma_1, \gamma_2) > 0$ ,  $M_2 = M_2(\gamma_1, \gamma_2) > 0$ , что будут выполнены неравенства

$$I_1 := \int_0^t \int_{\mathbb{R}^3} \frac{|\mathcal{E}_1(x-y, t-\tau)|}{(1+y_1^2+y_2^2)^{\gamma_1/2}(1+y_3^2)^{\gamma_2/2}} dy d\tau \leq \frac{M_1 t}{(1+x_1^2+x_2^2)^{1/4}(1+x_3^2)^{1/4}} \quad (4.27)$$

при  $(x, t) = (x_1, x_2, x_3, t) \in \mathbb{R}^3 \otimes [0, T]$ ,

$$I_2 := \int_0^t \int_{\mathbb{R}^3} \frac{\left| \frac{\partial \mathcal{E}_1(x-y, t-\tau)}{\partial t} \right|}{(1+y_1^2+y_2^2)^{\gamma_1/2}(1+y_3^2)^{\gamma_2/2}} dy d\tau \leq \frac{M_2(t+t^{1/2})}{(1+x_1^2+x_2^2)^{1/4}(1+x_3^2)^{1/4}} \quad (4.28)$$

при  $(x, t) = (x_1, x_2, x_3, t) \in \mathbb{R}^3 \otimes [0, T]$ .

**Доказательство.** Для доказательства этой леммы достаточно воспользоваться оценками (4.19), (4.21).

## 5. СВОЙСТВА ОБЪЕМНЫХ ПОТЕНЦИАЛОВ

Рассмотрим объемный потенциал

$$U_0[\rho](x, t) := \int_0^t \int_{\mathbb{R}^3} K_{\gamma_1, \gamma_2}(x, y, t-\tau) \rho(y, \tau) dy d\tau, \quad (5.1)$$

где

$$K_{\gamma_1, \gamma_2}(x, y, t) := \frac{(1+x_1^2+x_2^2)^{1/4}(1+x_3^2)^{1/4}}{(1+y_1^2+y_2^2)^{\gamma_1/2}(1+y_3^2)^{\gamma_2/2}} \mathcal{E}(x-y, t), \quad (5.2)$$

а фундаментальное решение  $\mathcal{E}(x, t)$  определяется формулой (4.5).

**Лемма 3.** При  $\gamma_1 > 2$  и  $\gamma_2 > 1$  оператор

$$U_0: \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]), \quad (5.3)$$

причем справедливо предельное свойство

$$\lim_{t \rightarrow +0} \sup_{x \in \mathbb{R}^3} |U_0(x, t)| = 0. \quad (5.4)$$

**Доказательство. Шаг 1.** Сначала нужно доказать, что  $U_0: \mathbb{C}(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}(\mathbb{R}^3 \otimes [0, T])$ .

Пусть  $(x^1, t^1), (x^2, t^2) \in \mathbb{R}^3 \otimes [0, T]$ . Пусть также для удобства  $t^2 > t^1$ . Оценим разностное отношение

$$\begin{aligned} \left| U_0(x^2, t^2) - U_0(x^1, t^1) \right| &= \left| \int_0^{t^2} \int_{\mathbb{R}^3} K_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) \rho(y, \tau) dy d\tau - \right. \\ &\quad \left. - \int_0^{t^1} \int_{\mathbb{R}^3} K_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau) \rho(y, \tau) dy d\tau \right| \leq I_1 + I_2, \end{aligned} \quad (5.5)$$

где

$$I_1 := \int_{t^1}^{t^2} \int_{\mathbb{R}^3} |K_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau)| |\rho(y, \tau)| dy d\tau, \quad (5.6)$$

$$I_2 := \int_0^{t^1} \int_{\mathbb{R}^3} |K_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) - K_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau)| |\rho(y, \tau)| dy d\tau. \quad (5.7)$$

Введем следующие обозначения:  $x^1 := (x_1^1, x_2^1, x_3^1)$ ,  $x^2 := (x_1^2, x_2^2, x_3^2)$ ,  $x_0^1 := (x_1^1, x_2^1)$ ,  $x_0^2 := (x_1^2, x_2^2)$ ,  $x_0^{1,2}$  — середина отрезка, соединяющего точки  $x_0^1, x_0^2 \in \mathbb{R}^2$ ,  $\Delta K := K_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) - K_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau)$ .

Обозначим открытый шар с центром в точке  $x$  в  $\mathbb{R}^2$  следующим образом:

$$O(x, r) := \{(y_1, y_2) \in \mathbb{R}^2 : (y_1 - x_1)^2 + (y_2 - x_2)^2 < r^2, x = (x_1, x_2) \in \mathbb{R}^2\}. \quad (5.8)$$

Далее выберем числа  $R, \varepsilon > 0$ :  $R > d(x_0^1, x_0^2)/2 + \varepsilon$ ,  $\varepsilon < d(x_0^1, x_0^2)/2$ , где  $d(x_0^1, x_0^2)$  — это расстояние между точками  $x_0^1, x_0^2 \in \mathbb{R}^2$ . Также нам потребуется множество  $\Omega \subset \mathbb{R}^2$ :

$$\Omega := O(x_0^{1,2}, R) \setminus (O(x_0^1, \varepsilon) \cup O(x_0^2, \varepsilon)). \quad (5.9)$$

С учетом всех перечисленных выше обозначений и с учетом выбранных  $R, \varepsilon$  для интеграла  $I_2$  справедлива оценка

$$I_2 \leq I_{21} + I_{22} + I_{23}, \quad (5.10)$$

где

$$I_{21} := \int_0^{t^1} d\tau \int_{\mathbb{R}^1} dy_3 \int_{O(x_0^1, \varepsilon) \cup O(x_0^2, \varepsilon)} |\Delta K| |\rho(y, \tau)| dy_1 dy_2, \quad (5.11)$$

$$I_{22} := \int_0^{t^1} d\tau \int_{\mathbb{R}^1} dy_3 \int_{\Omega} |\Delta K| |\rho(y, \tau)| dy_1 dy_2, \quad (5.12)$$

$$I_{23} := \int_0^{t^1} d\tau \int_{\mathbb{R}^1} dy_3 \int_{\mathbb{R}^2 \setminus O(x_0^{1,2}, R)} |\Delta K| |\rho(y, \tau)| dy_1 dy_2. \quad (5.13)$$

Интеграл  $I_{22}$  можно представить следующим образом:

$$I_{22} = I_{221} + I_{222}, \quad (5.14)$$

где

$$I_{221} := \int_0^{t^1} d\tau \int_{-R_3}^{R_3} dy_3 \int_{\Omega} |\Delta K| |\rho(y, \tau)| dy_1 dy_2, \quad (5.15)$$

$$I_{222} := \int_0^{t^1} d\tau \int_{\mathbb{R}^1 \setminus [-R_3, R_3]} dy_3 \int_{\Omega} |\Delta K| |\rho(y, \tau)| dy_1 dy_2, \quad R_3 > 0. \quad (5.16)$$

Зафиксируем произвольное  $\delta > 0$ . В силу оценки (4.22) ясно, что найдутся достаточно малое  $\varepsilon > 0$  и достаточно большие  $R > 0$ ,  $R_3 > 0$ , что будут выполнены следующие неравенства:

$$I_{21} < \frac{\delta}{4}, \quad I_{23} < \frac{\delta}{4}, \quad I_{222} < \frac{\delta}{8}. \quad (5.17)$$

Для фиксированных выше  $\varepsilon, R, R_3$  найдется достаточно малое  $\eta > 0$  такое, что если выполнено неравенство

$$|x^2 - x^1|^2 + |t^2 - t^1|^2 < \eta^2, \quad (5.18)$$

то справедливы и неравенства

$$I_1 < \frac{\delta}{4}, \quad I_{221} < \frac{\delta}{8}, \quad (5.19)$$

при оценке интеграла  $I_{221}$  мы воспользовались непрерывностью подынтегральной функции на множестве  $\Omega \times [-R_3, R_3]$ .

В результате мы получили, что для любого  $\delta > 0$  найдется  $\eta > 0$  такое, что при выполнении неравенства (5.18) имеет место неравенство

$$|U_0(x^2, t^2) - U_0(x^1, t^1)| < \delta, \quad (5.20)$$

которое и означает, что  $U_0(x, t) \in \mathbb{C}(\mathbb{R}^3 \otimes [0, T])$ .

**Шаг 2.** Теперь установим, что  $U_0: \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ .

Пусть  $\rho(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ . Тогда для  $U_0[\rho](x, t)$  справедлива оценка

$$|U_0[\rho](x, t)| \leq \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| \int_0^t \int_{\mathbb{R}^3} |K_{\gamma_1, \gamma_2}(x, y, t - \tau)| dy d\tau \leq M_1 \sqrt{t} \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)|, \quad M_1 > 0, \quad (5.21)$$

для получения которой мы воспользовались оценкой (4.22). Из оценки (5.21) сразу же вытекает предельное свойство (5.4). Также из оценки (5.21) следует, что

$$\sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |U_0[\rho](x, t)| \leq M_1 \sqrt{T} \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| < +\infty. \quad (5.22)$$

Из оценки (5.22) и следует, что  $U_0: \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ . Лемма 3 доказана.

Теперь рассмотрим другой объемный потенциал:

$$V_0[\rho](x, t) := \int_0^t \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x, y, t - \tau) \rho(y, \tau) dy d\tau, \quad (5.23)$$

где

$$G_{\gamma_1, \gamma_2}(x, y, t) := \frac{(1 + x_1^2 + x_2^2)^{1/4} (1 + x_3^2)^{1/4}}{(1 + y_1^2 + y_2^2)^{\gamma_1/2} (1 + y_3^2)^{\gamma_2/2}} \mathcal{E}_1(x - y, t), \quad (5.24)$$

а фундаментальное решение  $\mathcal{E}_1(x, t)$  определяется формулой (4.16).

**Лемма 4.** При  $\gamma_1 > 2$  и  $\gamma_2 > 1$  оператор

$$V_0: \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b^{(0,1)}(\mathbb{R}^3 \otimes [0, T]), \quad (5.25)$$

причем справедливы предельные свойства

$$\lim_{t \rightarrow +0} \sup_{x \in \mathbb{R}^3} |V_0(x, t)| = 0, \quad (5.26)$$

$$\lim_{t \rightarrow +0} \sup_{x \in \mathbb{R}^3} \left| \frac{\partial V_0(x, t)}{\partial t} \right| = 0. \quad (5.27)$$

**Доказательство. Шаг 1.** Докажем сначала, что  $V_0: \mathbb{C}(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}(\mathbb{R}^3 \otimes [0, T])$ .

Пусть  $(x^1, t^1), (x^2, t^2) \in \mathbb{R}^3 \otimes [0, T]$ . Пусть также для удобства  $t^2 > t^1$ .

Оценим разностное отношение

$$\begin{aligned} |V_0(x^2, t^2) - V_0(x^1, t^1)| &= \left| \int_0^{t^2} \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) \rho(y, \tau) dy d\tau - \right. \\ &\quad \left. - \int_0^{t^1} \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau) \rho(y, \tau) dy d\tau \right| \leq I_1 + I_2, \end{aligned} \quad (5.28)$$

где

$$I_1 := \int_{t^1}^{t^2} \int_{\mathbb{R}^3} |G_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau)| |\rho(y, \tau)| dy d\tau, \quad (5.29)$$

$$I_2 := \int_0^{t^1} \int_{\mathbb{R}^3} |G_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) - G_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau)| \rho(y, \tau) dy d\tau. \quad (5.30)$$

Введем следующие обозначения:  $x^1 := (x_1^1, x_2^1, x_3^1)$ ,  $x^2 := (x_1^2, x_2^2, x_3^2)$ ,  $x_3^{1,2}$  – середина отрезка, соединяющего точки  $x_3^1, x_3^2 \in \mathbb{R}^1$ ,  $\Delta G := G_{\gamma_1, \gamma_2}(x^2, y, t^2 - \tau) - G_{\gamma_1, \gamma_2}(x^1, y, t^1 - \tau)$ . Далее выберем числа  $R, \varepsilon > 0$ :  $R > d(x_3^1, x_3^2)/2 + \varepsilon$ ,  $\varepsilon < d(x_3^1, x_3^2)/2$ , где  $d(x_3^1, x_3^2)$  – это расстояние между точками  $x_3^1, x_3^2 \in \mathbb{R}^1$ . Также нам потребуются два множества  $U_\varepsilon, U_R \subset \mathbb{R}^1$ :

$$U_\varepsilon := [x_3^1 - \varepsilon, x_3^1 + \varepsilon] \cup [x_3^2 - \varepsilon, x_3^2 + \varepsilon], \quad (5.31)$$

$$U_R := [x_3^{1,2} - R, x_3^{1,2} + R]. \quad (5.32)$$

С учетом всех перечисленных выше обозначений и с учетом выбранных  $R, \varepsilon$  для интеграла  $I_2$  следует оценка

$$I_2 \leq I_{21} + I_{22} + I_{23}, \quad (5.33)$$

где

$$I_{21} := \int_0^{t^1} d\tau \int_{U_\varepsilon} dy_3 \int_{\mathbb{R}^2} |\Delta G| \rho(y, \tau) dy_1 dy_2, \quad (5.34)$$

$$I_{22} := \int_0^{t^1} d\tau \int_{U_R \setminus U_\varepsilon} dy_3 \int_{\mathbb{R}^2} |\Delta G| \rho(y, \tau) dy_1 dy_2, \quad (5.35)$$

$$I_{23} := \int_0^{t^1} d\tau \int_{\mathbb{R}^1 \setminus U_R} dy_3 \int_{\mathbb{R}^2} |\Delta G| \rho(y, \tau) dy_1 dy_2. \quad (5.36)$$

Представим интеграл  $I_{22}$  следующим образом:

$$I_{22} = I_{221} + I_{222}, \quad (5.37)$$

где

$$I_{221} := \int_0^{t^1} d\tau \int_{U_R \setminus U_\varepsilon} dy_3 \int_{O(0, R_{12})} |\Delta G| \rho(y, \tau) dy_1 dy_2, \quad (5.38)$$

$$I_{222} := \int_0^{t^1} d\tau \int_{U_R \setminus U_\varepsilon} dy_3 \int_{\mathbb{R}^2 \setminus O(0, R_{12})} |\Delta G| \rho(y, \tau) dy_1 dy_2, \quad R_{12} > 0. \quad (5.39)$$

Зафиксируем произвольное  $\delta > 0$ . Тогда в силу оценки (4.27) найдутся достаточно малое  $\varepsilon > 0$  и достаточно большие  $R > 0$ ,  $R_{12} > 0$ , что будут выполнены неравенства

$$I_{21} < \frac{\delta}{4}, \quad I_{23} < \frac{\delta}{4}, \quad I_{222} < \frac{\delta}{8}. \quad (5.40)$$

Для фиксированных выше  $\varepsilon$ ,  $R$ ,  $R_{12}$  найдется достаточно малое  $\eta > 0$  такое, что если выполнено неравенство

$$|x^2 - x^1|^2 + |t^2 - t^1|^2 < \eta^2, \quad (5.41)$$

то справедливы и неравенства

$$I_1 < \frac{\delta}{4}, \quad I_{221} < \frac{\delta}{8}, \quad (5.42)$$

где при оценке интеграла  $I_{221}$  была использована непрерывность подынтегральной функции на соответствующем множестве.

В результате мы получаем, что для любого  $\delta > 0$  найдется  $\eta > 0$  такое, что при выполнении неравенства (5.41) имеет место неравенство

$$|V_0(x^2, t^2) - V_0(x^1, t^1)| < \delta, \quad (5.43)$$

которое и означает, что  $V_0(x, t) \in \mathbb{C}(\mathbb{R}^3 \otimes [0, T])$ .

**Шаг 2.** Теперь установим, что  $V_0 : \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ .

Пусть  $\rho(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ . Тогда для  $V_0[\rho](x, t)$  справедлива оценка

$$|V_0[\rho](x, t)| \leq \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| \int_0^t \int_{\mathbb{R}^3} |G_{\gamma_1, \gamma_2}(x, y, t - \tau)| dy d\tau \leq M_1 t \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)|, \quad M_1 > 0, \quad (5.44)$$

для получения которой мы воспользовались оценкой (4.27). Из оценки (5.44) сразу же вытекает предельное свойство (5.26). Также из оценки (5.44) следует, что

$$\sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |V_0[\rho](x, t)| \leq M_1 T \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| < +\infty. \quad (5.45)$$

Из оценки (5.45) следует, что  $V_0 : \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T]) \rightarrow \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ .

**Шаг 3.** Докажем теперь, что если  $\rho(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ , то  $V_0(x, t) \in C^{(0,1)}(\mathbb{R}^3 \otimes [0, T])$ . Пусть  $(x, t) \in \mathbb{R}^3 \otimes [0, T]$ ,  $x = (x^1, x^2, x^3)$ . Будем считать, что  $\Delta t > 0$ , тогда мы вычислим правую производную. Случай же  $\Delta t < 0$  рассматривается аналогичным образом, отличие будет состоять лишь в том, что там не возникнет выражения типа  $I_1$  (см. формулу (5.47)):

$$\left| \frac{V_0(x, t + \Delta t) - V_0(x, t)}{\Delta t} - \int_0^t \int_{\mathbb{R}^3} \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \rho(y, \tau) dy d\tau \right| \leq I_1 + I_2, \quad (5.46)$$

где

$$I_1 := \left| \frac{1}{\Delta t} \int_t^{t+\Delta t} \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x, y, t - \tau + \Delta t) \rho(y, \tau) dy d\tau \right|, \quad (5.47)$$

$$I_2 := \left| \int_0^t \int_{\mathbb{R}^3} \left( \frac{G_{\gamma_1, \gamma_2}(x, y, t - \tau + \Delta t) - G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\Delta t} - \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \right) \rho(y, \tau) dy d\tau \right|. \quad (5.48)$$

Поскольку  $\rho(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ , то существует константа  $c_0 > 0$ :  $c_0 = \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| < +\infty$ .

Введем удобные для дальнейшего обозначения:  $U_\varepsilon := [x^3 - \varepsilon, x^3 + \varepsilon]$ ,  $U_R := [x^3 - R, x^3 + R]$ ,  $0 < \varepsilon < R$ . Учтем это и оценим сначала интеграл  $I_1$ :

$$I_1 \leq \frac{c_0}{\Delta t} \int_0^{\Delta t} \int_{\mathbb{R}^3} |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)| dy ds = I_{11} + I_{12} + I_{13}, \quad (5.49)$$

где

$$I_{11} := \frac{c_0}{\Delta t} \int_0^{\Delta t} ds \int_{\mathbb{R}^2} dy_1 dy_2 \int_{U_\varepsilon} dy_3 |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)|, \quad (5.50)$$

$$I_{12} := \frac{c_0}{\Delta t} \int_0^{\Delta t} ds \int_{\mathbb{R}^2} dy_1 dy_2 \int_{U_R \setminus U_\varepsilon} dy_3 |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)|, \quad (5.51)$$

$$I_{13} := \frac{c_0}{\Delta t} \int_0^{\Delta t} ds \int_{\mathbb{R}^2} dy_1 dy_2 \int_{\mathbb{R}^1 \setminus U_R} dy_3 |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)|. \quad (5.52)$$

Интеграл  $I_{12}$  представим следующим образом:

$$I_{12} = I_{121} + I_{122}, \quad (5.53)$$

где

$$I_{121} := \frac{c_0}{\Delta t} \int_0^{\Delta t} ds \int_{O(0, R_{12})} dy_1 dy_2 \int_{U_R \setminus U_\varepsilon} dy_3 |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)|, \quad (5.54)$$

$$I_{122} := \frac{c_0}{\Delta t} \int_0^{\Delta t} ds \int_{\mathbb{R}^2 \setminus O(0, R_{12})} dy_1 dy_2 \int_{U_R \setminus U_\varepsilon} dy_3 |G_{\gamma_1, \gamma_2}(x, y, \Delta t - s)|, \quad (5.55)$$

где  $O(0, R_{12})$  — это открытый шар радиуса  $R_{12}$  с центром в нуле в  $\mathbb{R}^2$ .

Зафиксируем произвольное  $\delta > 0$ . Из формул (5.24), (4.27) следует, что найдутся достаточно малое  $\varepsilon > 0$  и достаточно большие  $R > 0$ ,  $R_{12} > 0$ , что будут выполнены неравенства

$$I_{11} < \frac{\delta}{8}, \quad I_{13} < \frac{\delta}{8}, \quad I_{122} < \frac{\delta}{8}. \quad (5.56)$$

Для фиксированных выше  $\varepsilon$ ,  $R$ ,  $R_{12}$  найдется достаточно малое  $\eta > 0$  такое, что если будет выполнено неравенство  $\Delta t < \eta$ , то будет справедливо и неравенство  $I_{121} < \delta/8$ . Для доказательства этой оценки для интеграла  $I_{121}$  достаточно воспользоваться формулой среднего значения, а затем учесть непрерывность подынтегральной функции в точке  $t = 0$ . В результате мы приходим к тому, что

$$I_1 < \frac{\delta}{2}. \quad (5.57)$$

Интеграл  $I_2$  можно оценить следующим образом:

$$I_2 \leq c_0 \int_0^t \int_{\mathbb{R}^3} \left| \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t^*)}{\partial t} - \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \right| dy d\tau, \quad \text{где } t^* \in [t - \tau, t - \tau + \Delta t]. \quad (5.58)$$

Представим интеграл  $I_2$  в виде

$$I_2 = I_{21} + I_{22} + I_{23}, \quad (5.59)$$

где

$$I_{21} := c_0 \int_0^t d\tau \int_{\mathbb{R}^2} dy_1 dy_2 \int_{U_\varepsilon} dy_3 \left| \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t^*)}{\partial t} - \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \right|, \quad (5.60)$$

$$I_{22} := c_0 \int_0^t d\tau \int_{\mathbb{R}^2} dy_1 dy_2 \int_{U_R \setminus U_\varepsilon} dy_3 \left| \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t^*)}{\partial t} - \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \right|, \quad (5.61)$$

$$I_{23} := c_0 \int_0^t d\tau \int_{\mathbb{R}^2} dy_1 dy_2 \int_{\mathbb{R}^1 \setminus U_R} dy_3 \left| \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t^*)}{\partial t} - \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \right|. \quad (5.62)$$

Далее достаточно воспользоваться оценкой (4.28) и повторить рассуждения шага 1 из доказательства данной леммы. Мы придем к тому, что

$$I_2 < \frac{\delta}{2}, \quad (5.63)$$

как только  $\Delta t < \eta$ .

В результате получаем, что для любого  $\delta > 0$  найдется  $\eta > 0$  такое, что как только  $|\Delta t| < \eta$ , то выполнено неравенство

$$\left| \frac{V_0(x, t + \Delta t) - V_0(x, t)}{\Delta t} - \int_0^t \int_{\mathbb{R}^3} \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \rho(y, \tau) dy d\tau \right| < \delta, \quad (5.64)$$

которое и означает, что  $V_0(x, t) \in \mathbb{C}^{(0,1)}(\mathbb{R}^3 \otimes [0, T])$ .

**Шаг 4.** Докажем, наконец, ограниченность производной  $\partial V_0(x, t)/\partial t$ . Пусть

$$c_0 := \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |\rho(x, t)| < +\infty.$$

Воспользуемся оценкой (4.28)

$$\left| \frac{\partial V_0(x, t)}{\partial t} \right| \leq \int_0^t \int_{\mathbb{R}^3} \left| \frac{\partial G_{\gamma_1, \gamma_2}(x, y, t - \tau)}{\partial t} \rho(y, \tau) \right| dy d\tau \leq c_0 M_2(t + t^{1/2}) \leq c_1(T + T^{1/2}), \quad (5.65)$$

$$c_1 > 0, \quad (x, t) \in \mathbb{R}^3 \otimes [0, T].$$

Выполнение предельного свойства (5.27) сразу же следует из оценки (5.65). Также из оценки (5.65) следует, что

$$\sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} \left| \frac{\partial V_0(x, t)}{\partial t} \right| \leq c_1(T + T^{1/2}) < +\infty, \quad (5.66)$$

поэтому  $V_0(x, t) \in \mathbb{C}_b^{(0,1)}(\mathbb{R}^3 \otimes [0, T])$ , как только  $\rho(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ . Лемма 4 доказана.

## 6. ОДНОЗНАЧНАЯ РАЗРЕШИМОСТЬ ДВУХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

Рассмотрим первое нелинейное интегральное уравнение:

$$u(x, t) = \int_0^t \int_{\mathbb{R}^3} \mathcal{E}(x - y, t - \tau) (|u|^q(y, \tau) + f(y, \tau)) dy d\tau, \quad (6.1)$$

где функция  $\mathcal{E}(x, t)$  определяется формулой (4.5), а функция  $f(x, t)$  будет определена далее.

Сделаем в интегральном уравнении (6.1) замену:

$$v(x, t) := (1 + x_1^2 + x_2^2)^{1/4} (1 + x_3^2)^{1/4} u(x, t), \quad (6.2)$$

$$f_0(x, t) := (1 + x_1^2 + x_2^2)^{\gamma_1/2} (1 + x_3^2)^{\gamma_2/2} f(x, t). \quad (6.3)$$

Тогда мы получим следующее интегральное уравнение:

$$v(x, t) = \int_0^t \int_{\mathbb{R}^3} K_{q/2, q/2}(x, y, t - \tau) |v|^q(y, \tau) dy d\tau + \int_0^t \int_{\mathbb{R}^3} K_{\gamma_1, \gamma_2}(x, y, t - \tau) f_0(y, \tau) dy d\tau, \quad (6.4)$$

где ядро интегрального уравнения определяется равенством

$$K_{\gamma_1, \gamma_2}(x, y, t - \tau) = \frac{(1 + x_1^2 + x_2^2)^{1/4} (1 + x_3^2)^{1/4}}{(1 + y_1^2 + y_2^2)^{\gamma_1/2} (1 + y_3^2)^{\gamma_2/2}} \mathcal{E}(x - y, t - \tau). \quad (6.5)$$

Перепишем интегральное уравнение (6.4) в виде

$$v(x, t) = U_{q/2, q/2}[|v|^q](x, t) + U_{\gamma_1, \gamma_2}[f_0](x, t), \quad (6.6)$$

где оператор  $U_{\gamma_1, \gamma_2}$  определяется равенством

$$U_{\gamma_1, \gamma_2}[\rho](x, t) := \int_0^t \int_{\mathbb{R}^3} K_{\gamma_1, \gamma_2}(x, y, t - \tau) \rho(y, \tau) dy d\tau. \quad (6.7)$$

Свойства объемного потенциала  $U_{\gamma_1, \gamma_2}$  изучались в разд. 5. Далее, воспользовавшись техникой, аналогичной технике, примененной при доказательстве теоремы 4 работы [1], можно установить следующий результат.

**Теорема 1.** Если  $q > 4$ ,  $\gamma_1 > 2$ ,  $\gamma_2 > 1$ , то для всех  $f_0(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$  найдется такое максимальное  $T_0 > 0$ , что для каждого  $T \in (0, T_0)$  существует единственное решение интегрального уравнения (6.4) в банаховом пространстве  $v(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ , причем либо  $T_0 = +\infty$ , либо  $T_0 < +\infty$ , и в последнем случае имеет место предельное свойство

$$\lim_{T \uparrow T_0} \|v\|_T = +\infty, \quad \|v\|_T = \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |v(x, t)|. \quad (6.8)$$

Непосредственным следствием данной теоремы является теорема 2.

**Теорема 2.** Если  $q > 4$ ,  $\gamma_1 > 2$ ,  $\gamma_2 > 1$ , то для всех  $f(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$  найдется такое максимальное  $T_0 > 0$ , что для каждого  $T \in (0, T_0)$  существует единственное решение интегрального уравнения (6.1) в банаховом пространстве  $u(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0, T])$ , причем либо  $T_0 = +\infty$ , либо  $T_0 < +\infty$ , и в последнем случае имеет место предельное свойство

$$\lim_{T \uparrow T_0} \|u\|_T = +\infty, \quad \|u\|_T = \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} (1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4} |u(x, t)|. \quad (6.9)$$

Рассмотрим второе нелинейное интегральное уравнение:

$$u(x, t) = F(x, t) + \int_0^t \int_{\mathbb{R}^3} \mathcal{E}_1(x - y, t - \tau) (|u|^q(y, \tau) - \rho(y, \tau)) dy d\tau, \quad (6.10)$$

где функция  $\mathcal{E}_1(x, t)$  определяется формулой (4.16), а функции  $F(x, t)$ ,  $\rho(x, t)$  будут определены далее. Сделаем в интегральном уравнении (6.10) замену:

$$v(x, t) := (1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4} u(x, t), \quad (6.11)$$

$$\rho_0(x, t) := (1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2} \rho(x, t). \quad (6.12)$$

Тогда мы приходим к следующему интегральному уравнению:

$$v(x, t) = F_0(x, t) + \int_0^t \int_{\mathbb{R}^3} G_{q/2, q/2}(x, y, t - \tau) |v|^q(y, \tau) dy d\tau - \int_0^t \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x, y, t - \tau) \rho_0(y, \tau) dy d\tau, \quad (6.13)$$

где  $F_0(x, t) = (1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4} F(x, t)$ , а ядро интегрального уравнения определено равенством

$$G_{\gamma_1, \gamma_2}(x, y, t - \tau) = \frac{(1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4}}{(1 + y_1^2 + y_2^2)^{\gamma_1/2}(1 + y_3^2)^{\gamma_2/2}} \mathcal{E}_1(x - y, t - \tau). \quad (6.14)$$

Перепишем интегральное уравнение (6.13) в операторном виде

$$v(x, t) = F_0(x, t) + V_{q/2, q/2}[|v|^q](x, t) - V_{\gamma_1, \gamma_2}[\rho_0](x, t), \quad (6.15)$$

где оператор  $V_{\gamma_1, \gamma_2}$  определяется равенством

$$V_{\gamma_1, \gamma_2}[\rho](x, t) := \int_0^t \int_{\mathbb{R}^3} G_{\gamma_1, \gamma_2}(x, y, t - \tau) \rho(y, \tau) dy d\tau. \quad (6.16)$$

Свойства объемного потенциала  $V_{\gamma_1, \gamma_2}$  изучались в разд. 5. Далее, воспользовавшись техникой, примененной при доказательстве теоремы 4 работы [1], можно установить следующий результат.

**Теорема 3.** Если  $q > 4$ ,  $\gamma_1 > 2$ ,  $\gamma_2 > 1$ , то для всяких  $\rho_0(x, t), F_0(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$  найдется такое максимальное  $T_0 > 0$ , что для каждого  $T \in (0, T_0)$  существует единственное решение интегрального уравнения (6.13) в банаховом пространстве  $v(x, t) \in \mathbb{C}_b(\mathbb{R}^3 \otimes [0, T])$ , причем либо  $T_0 = +\infty$ , либо  $T_0 < +\infty$ , и в последнем случае имеет место предельное свойство

$$\lim_{T \uparrow T_0} \|v\|_T = +\infty, \quad \|v\|_T = \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} |v(x, t)|. \quad (6.17)$$

Непосредственным следствием теоремы 3 является теорема 4.

**Теорема 4.** Если  $q > 4$ ,  $\gamma_1 > 2$ ,  $\gamma_2 > 1$ , то для всех  $F(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0, T])$ ,  $\rho(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$  найдется такое максимальное  $T_0 > 0$ , что для каждого  $T \in (0, T_0)$  существует единственное решение интегрального уравнения (6.10) в банаховом пространстве  $u(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0, T])$ , причем либо  $T_0 = +\infty$ , либо  $T_0 < +\infty$ , и в последнем случае имеет место предельное свойство

$$\lim_{T \uparrow T_0} \|u\|_T = +\infty, \quad \|u\|_T = \sup_{(x, t) \in \mathbb{R}^3 \otimes [0, T]} (1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4} |u(x, t)|. \quad (6.18)$$

## 7. СЛАБЫЕ ОБОБЩЕННЫЕ РЕШЕНИЯ ДВУХ ЗАДАЧ КОШИ

Рассмотрим первую задачу Коши:

$$\mathfrak{M}_{x,t}[u](x, t) = |u|^q + f(x, t), \quad (x, t) \in \mathbb{R}^3 \otimes (0, T], \quad (7.1)$$

$$u(x, 0) = u_0(x), \quad u'(x, 0) = u_1(x), \quad x \in \mathbb{R}^3, \quad (7.2)$$

где оператор  $\mathfrak{M}_{x,t}$  определяется равенством

$$\mathfrak{M}_{x,t}[u](x, t) := \frac{\partial^2}{\partial t^2} \frac{\partial^2}{\partial x_3^2} u(x, t) + \Delta u(x, t). \quad (7.3)$$

Дадим определение слабого локального во времени решения задачи Коши (7.1), (7.2), для которого далее будем изучать явление разрушения, а также дадим определение слабого обобщенного локального во времени решения задачи Коши с нулевыми начальными условиями, для которого будем далее исследовать разрешимость.

**Определение 1.** Слабым локальным во времени решением задачи Коши (7.1), (7.2) называется функция  $u(x, t) \in L_{\text{loc}}^q(\mathbb{R}^3 \otimes [0, T])$ , которая удовлетворяет интегральному равенству

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^3} u(x, t) \left[ \frac{\partial^2}{\partial t^2} \frac{\partial^2}{\partial x_3^2} \phi(x, t) + \Delta \phi(x, t) \right] dx dt + \int_{\mathbb{R}^3} \left[ u_0(x) \frac{\partial \phi_{x_3 x_3}}{\partial t}(x, 0) - u_1(x) \phi_{x_3 x_3}(x, 0) \right] dx = \\ = \int_0^T \int_{\mathbb{R}^3} |u(x, t)|^q \phi(x, t) dx dt + \int_0^T \int_{\mathbb{R}^3} f(x, t) \phi(x, t) dx dt \end{aligned} \quad (7.4)$$

для произвольной функции  $\phi(x, t) \in \mathbb{C}_0^{2+2}(\mathbb{R}^3 \otimes [0, T])$ . Кроме того,  $u_0(x), u_1(x) \in L_{\text{loc}}^1(\mathbb{R}^3)$ ,  $f(x, t) \in L_{\text{loc}}^1(\mathbb{R}^3 \otimes [0, T])$ .

**Определение 2.** Слабым обобщенным локальным во времени решением задачи Коши (7.1), (7.2) с нулевыми начальными условиями называется функция  $u(x, t) \in L_{\text{loc}}^q(\mathbb{R}^3 \otimes [0, T])$ , которая удовлетворяет интегральному равенству

$$\int_0^T \int_{\mathbb{R}^3} u(x, t) \left[ \frac{\partial^2}{\partial t^2} \frac{\partial^2}{\partial x_3^2} \phi(x, t) + \Delta \phi(x, t) \right] dx dt = \int_0^T \int_{\mathbb{R}^3} |u(x, t)|^q \phi(x, t) dx dt + \int_0^T \int_{\mathbb{R}^3} f(x, t) \phi(x, t) dx dt \quad (7.5)$$

для произвольной функции  $\phi(x, t) \in \mathcal{D}(\mathbb{R}^3 \otimes (0, T))$ . При этом  $f(x, t) \in L_{\text{loc}}^1(\mathbb{R}^3 \otimes [0, T])$ .

**Теорема 5.** Пусть  $u(x, t)$  — это решение нелинейного интегрального уравнения (6.1) в банаховом пространстве  $\mathbb{C}_b((1 + x_1^2 + x_2^2)^{1/4}(1 + x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0, T])$ , причем  $f(x, t) \in \mathbb{C}_b((1 + x_1^2 + x_2^2)^{\gamma_1/2}(1 + x_3^2)^{\gamma_2/2}; \mathbb{R}^3 \otimes [0, T])$ , где  $q > 4$ ,  $\gamma_1 > 2$ ,  $\gamma_2 > 1$ . Тогда функция  $u(x, t)$  является слабым обобщенным локальным во времени решением задачи Коши (7.1), (7.2) с нулевыми начальными условиями в смысле определения 2.

**Доказательство.** Пусть  $\langle \cdot, \cdot \rangle$  — это скобки двойственности между  $\mathcal{D}(\mathbb{R}^3 \otimes (0, T))$  и  $\mathcal{D}'(\mathbb{R}^3 \otimes (0, T))$ . Пусть функция  $u(x, t)$  удовлетворяет интегральному уравнению (6.1). Тогда в смысле пространства обобщенных функций  $\mathcal{D}'(\mathbb{R}^3 \otimes (0, T))$  справедливо равенство

$$u(x, t) = \mathcal{E}(x, t) * \left[ |u(x, t)|^q + f(x, t) \right]. \quad (7.6)$$

Тогда, с одной стороны,

$$\begin{aligned} \langle \mathcal{M}_{x,t}[u](x, t), \phi(x, t) \rangle &= \langle \mathcal{M}_{x,t}[\mathcal{E}(x, t) * (|u(x, t)|^q + f(x, t))], \phi(x, t) \rangle = \\ &= \langle \mathcal{M}_{x,t}[\mathcal{E}(x, t)] * (|u(x, t)|^q + f(x, t)), \phi(x, t) \rangle = \langle \delta(x, t) * (|u(x, t)|^q + f(x, t)), \phi(x, t) \rangle = \\ &= \langle |u(x, t)|^q + f(x, t), \phi(x, t) \rangle = \int_0^T \int_{\mathbb{R}^3} |u(x, t)|^q \phi(x, t) dx dt + \int_0^T \int_{\mathbb{R}^3} f(x, t) \phi(x, t) dx dt. \end{aligned} \quad (7.7)$$

Здесь нами была использована формула дифференцирования свертки обобщенных функций

$$D(f * g) = D(f) * g, \quad (7.8)$$

где  $D$  — дифференциальный оператор. Также было учтено, что

$$\mathcal{M}_{x,t}[\mathcal{E}(x, t)] = \delta(x, t) \quad (7.9)$$

в смысле равенства обобщенных функций в пространстве  $\mathcal{D}'(\mathbb{R}^3 \otimes (0, T))$ .

С другой стороны,

$$\langle \mathcal{M}_{x,t}[u](x, t), \phi(x, t) \rangle = \langle u(x, t), \mathcal{M}_{x,t}[\phi](x, t) \rangle = \int_0^T \int_{\mathbb{R}^3} u(x, t) \left[ \frac{\partial^2}{\partial t^2} \frac{\partial^2}{\partial x_3^2} \phi(x, t) + \Delta \phi(x, t) \right] dx dt. \quad (7.10)$$

Из равенств (7.7), (7.10) и следует утверждение теоремы.

Теперь получим результат, касающийся разрушения слабых локальных во времени решений задачи Коши (7.1), (7.2).

**Теорема 6.** Пусть  $1 < q \leq 3$  и  $\{u_0(x), u_1(x)\} \in H_{x_3}$ . Пусть выполнены неравенства

$$|u_0(x)| \leq \frac{A_0}{(1 + |x|^2)^{\alpha/2}}, \quad |u_1(x)| \leq \frac{A_1}{(1 + |x|^2)^{\beta/2}} \quad \text{при} \quad \alpha > 1, \quad \beta > 1. \quad (7.11)$$

Пусть, кроме того, функция  $f$ , входящая в уравнение (7.4), тождественно равна нулю.

Тогда не существует слабого локального во времени решения задачи Коши (7.1), (7.2) в смысле определения 1 ни для какого  $T > 0$ .

**Доказательство.** Пусть функция  $u(x, t)$  является слабым локальным во времени решением задачи Коши (7.1), (7.2) в смысле определения 1. Введем следующее обозначение:  $L_{x_3} := \partial^2 / \partial x_3^2$ . Также введем обозначения для интегралов из равенства (7.4):

$$I_1 := \int_0^T \int_{\mathbb{R}^3} u(x, t) \frac{\partial^2}{\partial t^2} L_{x_3} \phi(x, t) dx dt, \quad I_2 := \int_0^T \int_{\mathbb{R}^3} u(x, t) \Delta \phi(x, t) dx dt, \quad (7.12)$$

$$I_3 := \int_{\mathbb{R}^3} u_0(x) \frac{\partial}{\partial t} L_{x_3} \phi(x, 0) dx, \quad I_4 := \int_{\mathbb{R}^3} u_1(x) L_{x_3} \phi(x, 0) dx, \quad (7.13)$$

$$I := \int_0^T \int_{\mathbb{R}^3} |u(x, t)|^q \phi(x, t) dx dt. \quad (7.14)$$

Тогда равенство (7.4) можно переписать в виде

$$I_1 + I_2 + I_3 - I_4 = I. \quad (7.15)$$

Выберем пробную функцию  $\phi(x, t)$  специальным образом:

$$\phi(x, t) = \phi_1(t)\phi_2(x), \quad (7.16)$$

$$\phi_1(t) := \left(1 - \frac{t}{T}\right)^\lambda, \quad \phi_2(x) := \phi_0\left(\frac{|x|^2}{R^2}\right), \quad \lambda > 2q', \quad q' = \frac{q}{q-1}, \quad q > 1, \quad (7.17)$$

$$\phi_0(s) := \begin{cases} 1, & \text{если } 0 \leq s \leq 1, \\ 0, & \text{если } s \geq 2, \end{cases} \quad \phi_0(s) \in \mathbb{C}_0^\infty[0, +\infty), \quad (7.18)$$

причем дополнительно потребуем, чтобы функция  $\phi_0(s)$  была монотонно невозрастающей.

Оценим сначала интегралы, связанные с начальными условиями:

$$\begin{aligned} |I_3| &= \left| \int_{\mathbb{R}^3} u_0(x) L_{x_3} \frac{\partial \phi}{\partial t}(x, 0) dx \right| = \left| \int_{\mathbb{R}^3} u_0(x) L_{x_3} \frac{\partial}{\partial t} \left( \left(1 - \frac{t}{T}\right)^\lambda \phi_2(x) \right) \Big|_{t=0} dx \right| = \\ &= \frac{\lambda}{T} \left| \int_{R \leq |x| \leq \sqrt{2}R} u_0(x) L_{x_3} \phi_2(x) dx \right| = \frac{\lambda}{T} R \left| \int_{1 \leq |y| \leq \sqrt{2}} u_0(Ry) L_{y_3} \phi_0(|y|^2) dy \right| \leq \frac{A_1}{R^{\alpha-1}} \rightarrow +0 \quad \text{при } R \rightarrow +\infty, \end{aligned} \quad (7.19)$$

$$\begin{aligned} |I_4| &= \left| \int_{\mathbb{R}^3} u_1(x) L_{x_3} \phi(x, 0) dx \right| = \left| \int_{R \leq |x| \leq \sqrt{2}R} u_1(x) L_{x_3} \phi_2(x) dx \right| = R \left| \int_{1 \leq |y| \leq \sqrt{2}} u_1(Ry) L_{y_3} \phi_0(|y|^2) dy \right| \leq \\ &\leq \frac{A_2}{R^{\beta-1}} \rightarrow +0 \quad \text{при } R \rightarrow +\infty. \end{aligned} \quad (7.20)$$

Далее оценим интегралы  $I_1, I_2$ , связанные с дифференциальным оператором:

$$\begin{aligned} |I_1| &= \left| \int_0^T \int_{\mathbb{R}^3} u(x, t) \frac{\partial^2}{\partial t^2} L_{x_3} \phi(x, t) dx dt \right| \leq \\ &\leq \frac{\lambda(\lambda-1)}{T^2} \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} |u(x, t) (\phi(x, t))^{1/q}| \left(1 - \frac{t}{T}\right)^{\lambda-2} \frac{|L_{x_3} \phi_2(x)|}{(\phi(x, t))^{1/q}} dx dt \leq c_1(R) I_R^{1/q}, \end{aligned} \quad (7.21)$$

где

$$I_R := \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \phi(x, t) |u(x, t)|^q dx dt, \quad (7.22)$$

$$\begin{aligned} c_1(R) &:= \frac{\lambda(\lambda-1)}{T^2} \left[ \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \left( \left(1 - \frac{t}{T}\right)^{\lambda-2} \frac{|L_{x_3} \phi_2(x)|}{(\phi(x, t))^{1/q}} \right)^{q'} dx dt \right]^{1/q'} = \\ &= \frac{\lambda(\lambda-1)}{T^2} \left[ \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \left(1 - \frac{t}{T}\right)^{\lambda-2q'} \frac{|L_{x_3} \phi_2(x)|^{q'}}{(\phi_2(x))^{q'/q}} dx dt \right]^{1/q'} = \frac{\lambda(\lambda-1)}{T^2} \left( \frac{T}{\lambda-2q'+1} \right)^{1/q'} c_0 R^{(3-2q')/q'}, \end{aligned} \quad (7.23)$$

где

$$c_0 := \left( \int_{1 \leq |y| \leq \sqrt{2}} \frac{|L_{y_3} \phi_0(|y|^2)|^{q'}}{(\phi_0(|y|^2))^{q'/q}} dy \right)^{1/q'}, \quad (7.24)$$

$$|I_2| = \left| \int_0^T \int_{\mathbb{R}^3} u(x, t) \Delta \phi(x, t) dx dt \right| \leq \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} |u(x, t) (\phi(x, t))^{1/q}| \left(1 - \frac{t}{T}\right)^\lambda \frac{|\Delta \phi_2(x)|}{(\phi(x, t))^{1/q}} dx dt \leq c_2(R) I_R^{1/q}, \quad (7.25)$$

где интеграл  $I_R$  определяется равенством (7.22):

$$c_2(R) := \left[ \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \left( \left(1 - \frac{t}{T}\right)^\lambda \frac{|\Delta \phi_2(x)|}{(\phi(x, t))^{1/q}} \right)^{q'} dx dt \right]^{1/q'} = \left( \frac{T}{\lambda + 1} \right)^{1/q'} c_1 R^{(3-2q')/q'}, \quad (7.26)$$

где

$$c_1 := \left( \int_{1 \leq |y| \leq \sqrt{2}} \frac{|\Delta \phi_0(|y|^2)|^{q'}}{(\phi_0(|y|^2))^{q'/q}} dy \right)^{1/q'}. \quad (7.27)$$

В [10] доказано существование таких монотонно невозрастающих функций  $\phi_0(s)$ , для которых будут конечны емкости  $c_0 > 0$  и  $c_1 > 0$ .

Сначала мы рассмотрим случай  $1 < q < 3$ . Тогда  $3 - 2q' < 0$ .

Заметим, что имеет место следующая оценка:

$$I_R \leq I := \int_0^T \int_{\mathbb{R}^3} \phi(x, t) |u(x, t)|^q dx dt. \quad (7.28)$$

Из (7.15) с учетом оценок (7.21), (7.25) вытекает оценка

$$I \leq c_1(R) I_R^{1/q} + c_2(R) I_R^{1/q} + |I_3| + |I_4|. \quad (7.29)$$

Воспользуемся трехпараметрическим неравенством Юнга

$$ab \leq \varepsilon a^q + \frac{b^{q'}}{q'(q\varepsilon)^{q'/q}}, \quad a, b \geq 0, \quad \varepsilon > 0. \quad (7.30)$$

Из (7.28), (7.29) и (7.30) получаем неравенство

$$(1 - 2\varepsilon)I \leq \frac{1}{q'(q\varepsilon)^{q'/q}} (c_1^{q'}(R) + c_2^{q'}(R)) + |I_3| + |I_4|. \quad (7.31)$$

Положим в неравенстве (7.31)  $\varepsilon = 1/4$ . Тогда получим следующее неравенство:

$$I \leq 2 \left( \frac{1}{q'} \left( \frac{4}{q} \right)^{q'/q} (c_1^{q'}(R) + c_2^{q'}(R)) + |I_3| + |I_4| \right) \leq k, \quad (7.32)$$

где постоянная  $k > 0$  не зависит от  $R$ .

Положим  $R = N \in \mathbb{N}$ . Заметим, что последовательность функций

$$\chi_N(x, t) := \phi(x, t) |u(x, t)|^q = \left(1 - \frac{t}{T}\right)^\lambda \phi_0 \left( \frac{|x|^2}{N^2} \right) |u(x, t)|^q \quad (7.33)$$

удовлетворяет неравенству

$$\chi_N(x, t) \leq \chi_{N+1}(x, t) \quad \text{для всех } (x, t) \in \mathbb{R}^3 \otimes [0, T]. \quad (7.34)$$

Воспользуемся теоремой Беппо Леви и получим, что, с одной стороны,

$$I = I(N) := \int_0^T \int_{\mathbb{R}^3} \chi_N(x, t) dx dt \rightarrow \int_0^T \int_{\mathbb{R}^3} \left(1 - \frac{t}{T}\right)^\lambda |u(x, t)|^q dx dt \quad \text{при } N \rightarrow +\infty. \quad (7.35)$$

С другой стороны, из неравенства (7.32), а также оценок (7.19), (7.20), (7.23), (7.26) следует, что

$$I = I(N) \rightarrow +0 \quad \text{при } N \rightarrow +\infty. \quad (7.36)$$

Таким образом, при условии  $1 < q < 3$  получаем, что

$$\int_0^T \int_{\mathbb{R}^3} \left(1 - \frac{t}{T}\right)^\lambda |u(x, t)|^q dx dt = 0 \Leftrightarrow u(x, t) = 0 \quad \text{для почти всех } (x, t) \in \mathbb{R}^3 \otimes [0, T]. \quad (7.37)$$

Критический случай  $q = 3$  нужно рассмотреть отдельно. Если  $q = 3$ , то  $3 - 2q' = 0$ , поэтому емкости  $c_1(R)$ ,  $c_2(R)$  уже не стремятся к нулю при  $R \rightarrow +\infty$  (но являются ограниченными). Воспользуемся неравенством (7.32), которое справедливо и для случая  $q = 3$ . Тогда, снова применив теорему Беппо Леви (снова положив  $R = N \in \mathbb{N}$ ), мы придем к тому, что

$$I = I(N) \rightarrow \int_0^T \int_{\mathbb{R}^3} \left(1 - \frac{t}{T}\right)^\lambda |u(x, t)|^q dx dt < +\infty \quad \text{при } N \rightarrow +\infty. \quad (7.38)$$

А из (7.38) сразу же получаем, что

$$0 \leq I_R = \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \phi(x, t) |u(x, t)|^q dx dt \leq \int_0^T \int_{R \leq |x| \leq \sqrt{2}R} \left(1 - \frac{t}{T}\right)^\lambda |u(x, t)|^q dx dt \rightarrow +0 \quad \text{при } R \rightarrow +\infty. \quad (7.39)$$

Но тогда из оценок (7.19), (7.20), (7.29) (7.39) следует, что

$$I \rightarrow +0 \quad \text{при } R \rightarrow +\infty. \quad (7.40)$$

Из (7.38), (7.40) получаем, что

$$\int_0^T \int_{\mathbb{R}^3} \left(1 - \frac{t}{T}\right)^\lambda |u(x, t)|^q dx dt = 0 \Leftrightarrow u(x, t) = 0 \quad \text{для почти всех } (x, t) \in \mathbb{R}^3 \otimes [0, T]. \quad (7.41)$$

В результате мы приходим к тому, что при  $1 < q \leq 3$  единственным слабым локальным во времени решением задачи Коши (7.1), (7.2) является функция  $u(x, t)$ , равная нулю почти всюду, поэтому из (7.4) получаем, что

$$\int_{\mathbb{R}^3} \left[ u_0(x) L_{x_3} \frac{\partial \phi}{\partial t}(x, 0) - u_1(x) L_{x_3} \phi(x, 0) \right] dx = 0 \quad (7.42)$$

для произвольной функции  $\phi(x, t) \in \mathbb{C}_0^{2+2}(\mathbb{R}^3 \otimes [0, T])$ .

Пусть

$$\phi(x, t) = \psi_0(x) \left(1 - \frac{t}{T}\right)^\lambda, \quad \lambda > 1, \quad \psi_0(x) \in \mathbb{C}_0^\infty(\mathbb{R}^3). \quad (7.43)$$

Тогда из (7.42) и (7.43) приходим к следующему равенству:

$$\int_{\mathbb{R}^3} \left[ \frac{\lambda}{T} u_0(x) + u_1(x) \right] L_{x_3} \psi_0(x) dx = 0 \quad (7.44)$$

для всех  $\psi_0(x) \in \mathbb{C}_0^\infty(\mathbb{R}^3)$  и любого  $T > 0$ . Так как пара функций  $\{u_0(x), u_1(x)\}$  принадлежит классу функций  $H_{x_3}$ , то найдется такой шар  $O(x_0, R)$  радиуса  $R > 0$ , что  $u_0(x), u_1(x) \in H^2(O(x_0, R))$  и

$$(L_{x_3} u_0(x))^2 + (L_{x_3} u_1(x))^2 > 0 \quad \text{для почти всех } x \in O(x_0, R). \quad (7.45)$$

Выберем  $\psi_0(x) \in \mathbb{C}_0^\infty(O(x_0, R))$ . Тогда из (7.44) получим равенство

$$\int_{O(x_0, R)} \left[ \frac{\lambda}{T} L_{x_3} u_0(x) + L_{x_3} u_1(x) \right] \psi_0(x) dx = 0 \quad (7.46)$$

для каждой  $\psi_0(x) \in \mathbb{C}_0^\infty(O(x_0, R))$ . Но тогда, согласно основной лемме вариационного исчисления, мы приходим к тому, что

$$\frac{\lambda}{T} L_{x_3} u_0(x) + L_{x_3} u_1(x) = 0 \quad (7.47)$$

для почти всех  $x \in O(x_0, R)$ . Продифференцируем равенство (7.47) по параметру  $\lambda$ . Тогда получаем, что

$$L_{x_3} u_0(x) = 0, \quad \text{а значит, и} \quad L_{x_3} u_1(x) = 0, \quad (7.48)$$

что противоречит условию (7.45). Теорема доказана.

Теперь рассмотрим вторую задачу Коши:

$$\mathfrak{L}_{x,t}[u](x,t) = |u|^q, \quad (x,t) \in \mathbb{R}^3 \otimes (0,T], \quad (7.49)$$

$$u(x,0) = u_0(x), \quad u'(x,0) = u_1(x), \quad x \in \mathbb{R}^3, \quad (7.50)$$

где оператор  $\mathfrak{L}_{x,t}$  определяется равенством

$$\mathfrak{L}_{x,t}[u](x,t) := \frac{\partial^2}{\partial t^2} \Delta_{\perp} u(x,t) + \Delta u(x,t). \quad (7.51)$$

**Определение 3.** Слабым локальным во времени решением задачи Коши (7.49), (7.50) называется функция  $u(x,t) \in L^q_{\text{loc}}(\mathbb{R}^3 \otimes [0,T])$ , которая удовлетворяет равенству

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^3} u(x,t) \left[ \frac{\partial^2}{\partial t^2} \Delta_{\perp} \phi(x,t) + \Delta \phi(x,t) \right] dx dt + \int_{\mathbb{R}^3} \left[ u_0(x) \Delta_{\perp} \frac{\partial \phi}{\partial t}(x,0) - u_1(x) \Delta_{\perp} \phi(x,0) \right] dx = \\ = \int_0^T \int_{\mathbb{R}^3} |u|^q \phi(x,t) dx dt \end{aligned} \quad (7.52)$$

для всех  $\phi(x,t) \in \mathbb{C}_0^{2+2}(\mathbb{R}^3 \otimes [0,T])$ . Кроме того,  $u_0(x), u_1(x) \in L^1_{\text{loc}}(\mathbb{R}^3)$ .

**Определение 4.** Слабым обобщенным локальным во времени решением задачи Коши (7.49), (7.50) называется функция  $u(x,t) \in \mathbb{C}_b^{(0,1)}(\mathbb{R}^3 \otimes [0,T])$ , которая удовлетворяет в смысле пространства обобщенных функций  $\mathcal{D}'(\mathbb{R}^3 \otimes (0,T))$  равенству

$$\langle \mathfrak{L}_{x,t}[u](x,t), \phi(x,t) \rangle = \langle |u|^q, \phi(x,t) \rangle \quad (7.53)$$

для произвольной функции  $\phi(x,t) \in \mathcal{D}(\mathbb{R}^3 \otimes (0,T))$ , а также удовлетворяет начальным условиям (7.50) поточечно, причем  $u_0(x), u_1(x) \in \mathbb{C}_b(\mathbb{R}^3)$ .

**Теорема 7.** Пусть функция  $u(x,t)$  — это решение нелинейного интегрального уравнения (6.10) в банаховом пространстве  $\mathbb{C}_b((1+x_1^2+x_2^2)^{1/4}(1+x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0,T])$ . Пусть показатель  $q > 4$ . В уравнении (6.10) в качестве функций  $F(x,t)$ ,  $\rho(x,t)$  берутся следующие функции:

$$F(x,t) := u_0(x)\chi(t) + u_1(x)\eta(t), \quad (7.54)$$

$$\rho(x,t) := \chi(t)\Delta u_0(x) + \chi''(t)\Delta_{\perp} u_0(x) + \eta(t)\Delta u_1(x) + \eta''(t)\Delta_{\perp} u_1(x), \quad (7.55)$$

$$\chi(t), \eta(t) \in \mathbb{C}^2([0,T]), \quad \chi(0) = 1, \quad \chi'(0) = 0, \quad \eta(0) = 0, \quad \eta'(0) = 1, \quad (7.56)$$

$$u_0(x), u_1(x) \in \mathbb{C}_b((1+x_1^2+x_2^2)^{1/4}(1+x_3^2)^{1/4}; \mathbb{R}^3 \otimes [0,T]), \quad (7.57)$$

$$\Delta_{\perp} u_0(x), \Delta u_0(x), \Delta_{\perp} u_1(x), \Delta u_1(x) \in \mathbb{C}_b((1+x_1^2+x_2^2)^{\gamma_1/2}(1+x_3^2)^{\gamma_2/2}; \mathbb{R}^3), \quad \gamma_1 > 2, \quad \gamma_2 > 1. \quad (7.58)$$

Тогда функция  $u(x,t)$  является слабым обобщенным локальным во времени решением задачи Коши (7.49), (7.50) в смысле определения 4.

**Доказательство. Шаг 1.** Пусть  $u(x,t)$  — это решение интегрального уравнения (6.10). Проверим, что эта функция удовлетворяет начальным условиям (7.50):

$$u(x,0) = f(x,0) = u_0(x)\chi(0) + u_1(x)\eta(0) = u_0(x), \quad (7.59)$$

$$\begin{aligned} \frac{\partial u}{\partial t}(x, 0) &= \frac{\partial f}{\partial t}(x, 0) + \frac{\partial}{\partial t} \int_{0 \mathbb{R}^3}^t \mathcal{E}_1(x - y, t - \tau) (|u|^q(y, \tau) - \rho(y, \tau)) dy d\tau \Big|_{t=0} = \\ &= u_0(x) \chi'(0) + u_1(x) \eta'(0) = u_1(x). \end{aligned} \quad (7.60)$$

**Шаг 2.** Проверим выполнение равенства (7.53).

Заметим, что в смысле пространства обобщенных функций  $\mathcal{D}'(\mathbb{R}^3 \otimes (0, T))$  справедливо равенство

$$u(x, t) = F(x, t) + \mathcal{E}_1(x, t) * [|u(x, t)|^q - \rho(x, t)]. \quad (7.61)$$

Но тогда справедлива следующая цепочка равенств:

$$\begin{aligned} \langle \mathcal{L}_{x,t}[u](x, t), \phi(x, t) \rangle &= \langle \mathcal{L}_{x,t}[F](x, t), \phi(x, t) \rangle + \langle \mathcal{L}_{x,t}[\mathcal{E}_1(x, t) * (|u(x, t)|^q - \rho(x, t))], \phi(x, t) \rangle = \\ &= \langle \chi(t) \Delta u_0(x) + \chi''(t) \Delta_{\perp} u_0(x) + \eta(t) \Delta u_1(x) + \eta''(t) \Delta_{\perp} u_1(x), \phi(x, t) \rangle + \\ &\quad + \langle \mathcal{L}_{x,t}[\mathcal{E}_1(x, t)] * (|u(x, t)|^q - \rho(x, t)), \phi(x, t) \rangle = \\ &= \langle \rho(x, t), \phi(x, t) \rangle + \langle (|u(x, t)|^q - \rho(x, t)), \phi(x, t) \rangle = \langle |u(x, t)|^q, \phi(x, t) \rangle, \end{aligned} \quad (7.62)$$

где мы учли, что, во-первых,

$$\begin{aligned} \mathcal{L}_{x,t}[F](x, t) &= \left( \frac{\partial^2}{\partial t^2} \Delta_{\perp} + \Delta \right) (u_0(x) \chi(t) + u_1(x) \eta(t)) = \\ &= \Delta_{\perp} u_0(x) \chi''(t) + \Delta u_0(x) \chi(t) + \Delta_{\perp} u_1(x) \eta''(t) + \Delta u_1(x) \eta(t), \end{aligned} \quad (7.63)$$

причем все производные здесь понимаются в классическом смысле.

Во-вторых, нами была использована формула дифференцирования свертки обобщенных функций

$$D(f * g) = D(f) * g, \quad (7.64)$$

где  $D$  — дифференциальный оператор. В-третьих, было учтено, что

$$\mathcal{L}_{x,t}[\mathcal{E}_1(x, t)] = \delta(x, t). \quad (7.65)$$

И, в-четвертых, — что  $f(x, t) * \delta(x, t) = f(x, t)$  для всякой обобщенной функции  $f(x, t)$ . Теорема доказана.

В следующей теореме сформулирован результат о разрушении слабых локальных во времени решений задачи Коши (7.49), (7.50).

Доказательство этой теоремы проводится аналогично доказательству теоремы 6, надо только заменить оператор  $L_{x_3} = \partial^2 / \partial x_3^2$  на оператор  $L_{x_1 x_2} := \Delta_{\perp}$ .

**Теорема 8.** Пусть  $1 < q \leq 3$  и  $\{u_0(x), u_1(x)\} \in H_{x_1 x_2}$ . Пусть выполнены неравенства

$$|u_0(x)| \leq \frac{A_0}{(1 + |x|^2)^{\alpha/2}}, \quad |u_1(x)| \leq \frac{A_1}{(1 + |x|^2)^{\beta/2}} \quad \text{при} \quad \alpha > 1, \quad \beta > 1. \quad (7.66)$$

Тогда не существует слабого локального во времени решения задачи Коши (7.49), (7.50) в смысле определения 3 ни для какого  $T > 0$ .

## СПИСОК ЛИТЕРАТУРЫ

1. Корпусов М.О., Шафир Р.С. О разрушении слабых решений задачи Коши для 3 + 1-мерного уравнения дрейфовых волн в плазме // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 1. С. 124–158.
2. Al'shin A.B., Korpusov M.O., Sveshnikov A.G. Blow-up in nonlinear Sobolev type equations // De Gruyter Series in Nonlin. Anal. Appl. 2011. V. 15. P. 648.
3. Свиридюк Г.А. К общей теории полугрупп операторов // Успехи матем. наук. 1994. Т. 49. № 4. С. 47–74.
4. Загребина С.А. Начально-конечная задача для уравнений соболевского типа с сильно (L,p)–радиальным оператором // Матем. заметки ЯГУ. 2012. Т. 19. № 2. С. 39–48.

5. *Zamyshlyayeva A.A., Sviridyuk G.A.* Nonclassical equations of mathematical physics. Linear Sobolev type equations of higher order // Вестн. Южно-Ур. ун-та. Сер. Матем. Мех. Физ. 2016. V. 8. № 4. P. 5–16.
6. *Капитонов Б.В.* Теория потенциала для уравнения малых колебаний вращающейся жидкости // Матем. сб. 1979. Т. 109(151). № 4(8). С. 607–628.
7. *Габов С.А., Свешиников А.Г.* Линейные задачи теории нестационарных внутренних волн. М.: Наука, 1990. С. 344.
8. *Габов С.А.* Новые задачи математической теории волн. М.: Физматлит, 1998. С. 448.
9. *Плетнер Ю.Д.* Фундаментальные решения операторов типа Соболева и некоторые начально-краевые задачи // Ж. вычисл. матем. и матем. физ. 1992. Т. 32. № 12. С. 1885–1899.
10. *Похожаев С.И., Митидиери Э.* Априорные оценки и отсутствие решений нелинейных уравнений и неравенств в частных производных // Тр. МИАН. 2001. Т. 234. С. 3–383.
11. *Galakhov E.I.* Some nonexistence results for quasilinear elliptic problems // J. Math. Anal. Appl. 2000. V. 252. № 1. P. 256–277.
12. *Галахов Е.И., Салиева О.А.* Об отсутствии неотрицательных монотонных решений для некоторых коэрцитивных неравенств в полупространстве // СФМН. 2017. Т. 63. № 4. С. 573–585.
13. *Корпусов М.О.* Критические показатели мгновенного разрушения или локальной разрешимости нелинейных уравнений соболевского типа // Изв. РАН. Сер. матем. 2015. Т. 79. № 5. С. 103–162.
14. *Корпусов М.О.* О разрушении решений нелинейных уравнений типа уравнения Хохлова–Заболотской // ТМФ. 2018. Т. 194. № 3. С. 403–417.
15. *Korpusov M.O., Ovchinnikov A.V., Panin A.A.* Instantaneous blow-up versus local solvability of solutions to the Cauchy problem for the equation of a semiconductor in a magnetic field // Math. Meth. Appl. Sci. 2018. V. 41. № 17. P. 8070–8099.

---

МАТЕМАТИЧЕСКАЯ  
ФИЗИКА

---

УДК 629.7

## ИССЛЕДОВАНИЕ ДИНАМИКИ СИСТЕМЫ ДВУХ СВЯЗАННЫХ ТЕЛ В ПЛОСКОСТИ КРУГОВОЙ ОРБИТЫ С ПРИМЕНЕНИЕМ МЕТОДОВ КОМПЬЮТЕРНОЙ АЛГЕБРЫ

© 2023 г. С. А. Гутник<sup>1,2,\*</sup>, В. А. Сарычев<sup>3,\*\*</sup>

<sup>1</sup> 119454 Москва, пр-т Вернадского, 76, МГИМО МИД России, Россия

<sup>2</sup> 141701 Долгопрудный, М. о., Институтский пер., 9, МФТИ, Россия

<sup>3</sup> 125047 Москва, Миусская пл., 4, ИПМ им. М.В. Келдыша РАН, Россия

\*e-mail: s.gutnik@inno.mgimo.ru

\*\*e-mail: vas31@rambler.ru

Поступила в редакцию 25.04.2022 г.

Переработанный вариант 25.04.2022 г.

Принята к публикации 17.09.2022 г.

Методы компьютерной алгебры применяются для определения равновесных ориентаций системы двух тел, соединенных сферическим шарниром, движущихся в центральном ньютоновом силовом поле по круговой орбите под действием гравитационного момента. Главное внимание уделяется изучению равновесных ориентаций связки двух тел в плоскости круговой орбиты. С применением символьного дифференцирования были получены дифференциальные уравнения движения в форме уравнений Лагранжа II рода. Предложен метод преобразования системы тригонометрических уравнений, определяющих равновесия, в систему алгебраических уравнений, которая, в свою очередь, путем вычисления результата была сведена к одному алгебраическому уравнению 12 степени от одной неизвестной. Корни полученного алгебраического уравнения определяют равновесные ориентации связки двух тел в плоскости круговой орбиты. Проведена символьная факторизация полученного алгебраического уравнения на три полиномиальных множителя, каждый из которых задает определенный класс равновесных конфигураций. Классификация областей с равным числом положений равновесия выполнялась с использованием алгебраических методов построения дискриминантной гиперповерхности. Уравнения дискриминантной гиперповерхности, которая определяет границы областей с равным числом положений равновесия в пространстве параметров задачи, были получены с использованием символьных вычислений определителя матрицы результата. Число положений равновесия системы двух тел в зависимости от параметров определялось путем численного анализа действительных корней полученных алгебраических уравнений. Библ. 16. Фиг. 2.

**Ключевые слова:** система двух тел, круговая орбита, уравнения Лагранжа, положения равновесия, алгебраические уравнения, компьютерная алгебра, результат, дискриминантная гиперповерхность.

**DOI:** 10.31857/S0044466923010088, **EDN:** LMDVUO

### ВВЕДЕНИЕ

В работе представлены результаты исследования положений равновесия системы двух тел (спутник-стабилизатор), соединенных сферическим шарниром, движущейся по круговой орбите в центральном гравитационном поле с применением методов компьютерной алгебры.

Исследование равновесных ориентаций связки тел, движущихся в центральном ньютоновом силовом поле по круговой орбите, представляет значительный практический интерес для создания составных схем гравитационной системы ориентации спутников, которые могут функционировать на орбите продолжительное время без расходования энергии и рабочего тела. Действие стабилизатора на спутник позволяет получить новые положения равновесия системы двух тел и ввести диссипацию в систему, что представляет большой интерес для практики. Подробное рассмотрение динамики различных типов составных схем гравитационных систем ориентации представлено в [1].

В работе [2] главное внимание было уделено исследованию существования равновесных ориентаций системы двух тел в случаях, когда одна из главных осей инерции как первого, так и второго тела совпадает с нормалью к плоскости орбиты, радиусом-вектором или касательной к орбите. Для определения равновесных ориентаций связки двух тел проводилась декомпозиция системы 12 алгебраических уравнений на ряд подсистем. Для решения каждой из подсистем алгебраических уравнений применялись алгоритмы построения базисов Гребнера. В [2] также представлен обзор результатов последних исследований по данной теме.

В работе [3] был получен широкий класс пространственных равновесных ориентаций системы спутник-стабилизатор с использованием комбинации методов компьютерной и линейной алгебры при определенных ограничениях на параметры задачи в случае, когда сферический шарнир расположен на пересечении главных центральных осей инерции спутника и стабилизатора.

В работе [4] исходная система 12 алгебраических уравнений, определяющая равновесия системы двух тел, была решена с применением методов построения базисов Гребнера для такой комбинации значений инерциальных и геометрических параметров двух тел, при которой коэффициенты алгебраической системы, определяющей равновесные ориентации, зависят только от одного параметра.

Задача нахождения всех пространственных положений равновесия двух соединенных сферическим шарниром тел на круговой орбите в общем виде до сих пор не решена. Подробное исследование задачи о пространственных положениях равновесия системы двух связанных осесимметричных тел проведено в работе [5].

Динамика движения двух соединенных шарниром тел в плоскости круговой орбиты в случае, когда сферический шарнир расположен на пересечении главных центральных осей инерции спутника и стабилизатора, исследовалась в работах [6]–[10].

В работе [11] проведено исследование равновесных ориентаций системы двух связанных тел в плоскости перпендикулярной плоскости круговой орбиты в случае, когда сферический шарнир расположен на линии пересечения двух плоскостей, образуемых главными центральными осями инерции спутника и стабилизатора. Для исследования равновесных ориентаций применялся метод редукции полиномов с использованием символьного вычисления результата.

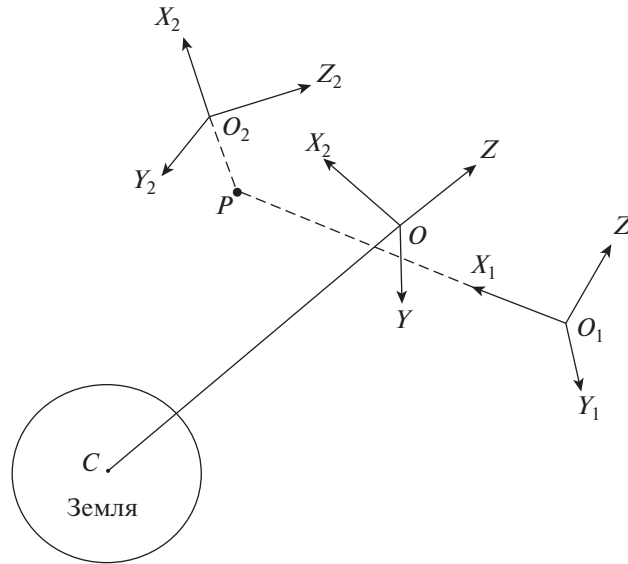
Основное внимание в представленной работе уделяется исследованию равновесных ориентаций связки двух тел в плоскости круговой орбиты, когда сферический шарнир расположен на линии пересечения двух плоскостей, образуемых главными центральными осями инерции спутника и стабилизатора. Для исследования равновесных ориентаций применялся подход, предложенный в работе [11], сводящийся к преобразованию системы тригонометрических уравнений, определяющих положения равновесия в систему алгебраических уравнений, для решения которой применялись символьные методы вычисления определителей матриц результатов.

Классификация областей с равным числом положений равновесия выполнялась с применением алгебраических методов построения дискриминантной гиперповерхности. Число положений равновесия системы двух тел в зависимости от параметров определялось путем численного анализа действительных корней полученных алгебраических уравнений.

## 1. УРАВНЕНИЯ ДВИЖЕНИЯ

Рассмотрим задачу о движении системы двух тел, соединенных сферическим шарниром, по круговой орбите [11]. Для записи уравнений движения системы двух тел введем следующие правые системы координат (фиг. 1):  $OXYZ$  – орбитальная система координат; ось  $OZ$  направлена вдоль радиуса-вектора, соединяющего центр масс  $C$  Земли и центр масс  $O$  системы двух тел; ось  $OX$  направлена вдоль вектора линейной скорости центра масс  $O$ , ось  $OY$  совпадает с нормалью к плоскости орбиты. Оси систем координат  $O_1x_1y_1z_1$  и  $O_2x_2y_2z_2$  направлены вдоль главных центральных осей инерции спутника и стабилизатора соответственно (фиг. 1). Ориентацию системы координат  $O_ix_iy_iz_i$  относительно орбитальной системы координат определим с использованием самолетных углов тангажа  $\alpha_i$ , рыскания  $\beta_i$  и крена  $\gamma_i$  (см. [1])

$$\begin{aligned} a_{11}^{(i)} &= \cos \alpha_i \cos \beta_i, \\ a_{12}^{(i)} &= \sin \alpha_i \sin \gamma_i - \cos \alpha_i \sin \beta_i \cos \gamma_i, \\ a_{13}^{(i)} &= \sin \alpha_i \cos \gamma_i + \cos \alpha_i \sin \beta_i \sin \gamma_i, \end{aligned}$$



Фиг. 1. Основные системы координат.

$$\begin{aligned}
 a_{21}^{(i)} &= \sin \beta_i, \\
 a_{22}^{(i)} &= \cos \beta_i \cos \gamma_i, \\
 a_{23}^{(i)} &= -\cos \beta_i \sin \gamma_i, \\
 a_{31}^{(i)} &= -\sin \alpha_i \cos \beta_i, \\
 a_{32}^{(i)} &= \cos \alpha_i \sin \gamma_i + \sin \alpha_i \sin \beta_i \cos \gamma_i, \\
 a_{33}^{(i)} &= \cos \alpha_i \cos \gamma_i - \sin \alpha_i \sin \beta_i \sin \gamma_i.
 \end{aligned} \tag{1.1}$$

Индекс  $i = 1$  ( $i = 2$ ) относится к телу 1 — спутнику (телу 2, стабилизатору).

Пусть  $(a_i, b_i, c_i)$  — координаты сферического шарнира  $P$  в связанной с телом системе координат  $O_i x_i y_i z_i$ ;  $A_i, B_i, C_i$  — главные центральные моменты инерции тел,  $M = M_1 M_2 / (M_1 + M_2)$ ;  $M_i$  — масса  $i$ -го тела. Рассмотрим случай, когда шарнир расположен на линии пересечения двух плоскостей, образуемых главными центральными осями инерции спутника и стабилизатора, когда  $b_1 = b_2 = 0$  и равновесные ориентации системы двух тел находятся в плоскости орбиты ( $\alpha_1 \neq 0$ ,  $\alpha_2 \neq 0$ ,  $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$ ). Тогда координаты сферического шарнира в связанных с каждым телом системе координат будут иметь значения  $(a_1, 0, c_1)$  и  $(a_2, 0, c_2)$ . В этом случае имеют место выражения для кинетической энергии

$$\begin{aligned}
 T &= \frac{1}{2} [B_1 + M(a_1^2 + c_1^2)] (\dot{\alpha}_1 + \omega_0)^2 + \frac{1}{2} [B_2 + M(a_2^2 + c_2^2)] (\dot{\alpha}_2 + \omega_0)^2 - \\
 &- M[(a_1 a_2 + c_1 c_2) \cos(\alpha_1 - \alpha_2) - (a_1 c_2 - a_2 c_1) \sin(\alpha_1 - \alpha_2)] (\dot{\alpha}_1 + \omega_0) (\dot{\alpha}_2 + \omega_0)
 \end{aligned} \tag{1.2}$$

и силовой функции

$$\begin{aligned}
 U &= \frac{3}{2} M \omega_0^2 [(a_1 \sin \alpha_1 - c_1 \cos \alpha_1) - (a_2 \sin \alpha_2 - c_2 \cos \alpha_2)]^2 - \frac{3}{2} \omega_0^2 (A_1 - C_1) \sin^2 \alpha_1 - \\
 &- \frac{3}{2} \omega_0^2 (A_2 - C_2) \sin^2 \alpha_2 + M \omega_0^2 [(a_1 a_2 + c_1 c_2) \cos(\alpha_1 - \alpha_2) - (a_1 c_2 - a_2 c_1) \sin(\alpha_1 - \alpha_2)].
 \end{aligned} \tag{1.3}$$

Используя выражения для кинетической энергии (1.2) и силовой функции (1.3), определяющей действие гравитационного поля Земли на систему двух тел, уравнения движения этой систе-

мы можно записать в форме уравнений Лагранжа II рода, используя возможности символьного дифференцирования в системе Mathematica 12.1 [12], в следующем виде:

$$\begin{aligned}
 & (B_1 + M(a_1^2 + c_1^2))\ddot{\alpha}_1 - M((a_1a_2 + c_1c_2)\cos(\alpha_1 - \alpha_2) - (a_1c_2 - a_2c_1)\sin(\alpha_1 - \alpha_2))\ddot{\alpha}_2 - \\
 & - M((a_1a_2 + c_1c_2)\sin(\alpha_1 - \alpha_2) + (a_1c_2 - a_2c_1)\cos(\alpha_1 - \alpha_2))(\dot{\alpha}_2^2 + 2\dot{\alpha}_2) + \\
 & + 3(A_1 - C_1)\sin\alpha_1\cos\alpha_1 - 3M(a_1\cos\alpha_1 + c_1\sin\alpha_1)(a_1\sin\alpha_1 - c_1\cos\alpha_1) - \\
 & - (a_2\sin\alpha_2 - c_2\cos\alpha_2)) = 0, \\
 & - M((a_1a_2 + c_1c_2)\cos(\alpha_1 - \alpha_2) - (a_1c_2 - a_2c_1)\sin(\alpha_1 - \alpha_2))\ddot{\alpha}_1 + (B_2 + M(a_2^2 + c_2^2))\ddot{\alpha}_2 + \\
 & + 3(A_2 - C_2)\sin\alpha_2\cos\alpha_2 + M((a_1a_2 + c_1c_2)\sin(\alpha_1 - \alpha_2) + (a_1c_2 - a_2c_1)\cos(\alpha_1 - \alpha_2))(\dot{\alpha}_1^2 + 2\dot{\alpha}_1) + \\
 & + 3M(a_2\cos\alpha_2 + c_2\sin\alpha_2)(a_1\sin\alpha_1 - c_1\cos\alpha_1) - (a_2\sin\alpha_2 - c_2\cos\alpha_2)) = 0.
 \end{aligned} \tag{1.4}$$

В уравнениях (1.4) точкой обозначено дифференцирование по  $\tau = \omega_0 t$ , где  $t$  – время.

## 2. ПОЛОЖЕНИЯ РАВНОВЕСИЯ

Положив в уравнениях (1.4)  $\alpha_1 = \text{const}$ ,  $\alpha_2 = \text{const}$ , получим при условиях  $A_i \neq B_i \neq C_i$  уравнения

$$\begin{aligned}
 & ((A_1 - C_1)/M)\sin\alpha_1\cos\alpha_1 - (a_1\cos\alpha_1 + c_1\sin\alpha_1)(a_1\sin\alpha_1 - c_1\cos\alpha_1) + \\
 & + (a_1\cos\alpha_1 + c_1\sin\alpha_1)(a_2\sin\alpha_2 - c_2\cos\alpha_2) = 0, \\
 & ((A_2 - C_2)/M)\sin\alpha_2\cos\alpha_2 - (a_2\cos\alpha_2 + c_2\sin\alpha_2)(a_2\sin\alpha_2 - c_2\cos\alpha_2) + \\
 & + (a_2\cos\alpha_2 + c_2\sin\alpha_2)(a_1\sin\alpha_1 - c_1\cos\alpha_1) = 0,
 \end{aligned} \tag{2.1}$$

позволяющие определить положения равновесия системы спутник-стабилизатор в орбитальной системе координат в плоскости орбиты. Тригонометрическую систему уравнений (2.1) относительно двух неизвестных самолетных углов  $\alpha_1$ ,  $\alpha_2$  аналитически решить не удастся.

Для решения системы (2.1) воспользуемся следующей универсальной заменой синусов и косинусов углов  $\alpha_1$  и  $\alpha_2$  через их тангенсы:

$$\sin\alpha_i = \frac{\text{tg}\alpha_i}{\sqrt{1 + \text{tg}^2\alpha_i}} = \frac{t_i}{\sqrt{1 + t_i^2}}, \quad \cos\alpha_i = \frac{1}{\sqrt{1 + \text{tg}^2\alpha_i}} = \frac{1}{\sqrt{1 + t_i^2}}, \tag{2.2}$$

где  $t_i = \text{tg}\alpha_i$ .

Подставляя выражения (2.2) в уравнения системы (2.1) и вводя обозначения  $d_1 = (A_1 - C_1)/M$ ,  $d_2 = (A_2 - C_2)/M$ , получаем два уравнения от двух неизвестных  $t_1$  и  $t_2$ :

$$\begin{aligned}
 d_1 t_1 - (a_1 + c_1 t_1)(a_1 t_1 - c_1) &= (a_1 + c_1 t_1)(c_2 - a_2 t_2) \frac{\sqrt{1 + t_1^2}}{\sqrt{1 + t_2^2}}, \\
 d_2 t_2 - (a_2 + c_2 t_2)(a_2 t_2 - c_2) &= (a_2 + c_2 t_2)(c_1 - a_1 t_1) \frac{\sqrt{1 + t_2^2}}{\sqrt{1 + t_1^2}}.
 \end{aligned} \tag{2.3}$$

Систему уравнений (2.3) можно упростить. Для этого сначала умножим левую (правую) часть первого уравнения (2.3) на левую (правую) часть второго уравнения, а затем разделим левую (правую) часть первого уравнения (2.3) на левую (правую) часть второго уравнения соответственно. В результате получим алгебраическую систему двух уравнений относительно двух неизвестных  $t_1$  и  $t_2$ :

$$\bar{a}_0 t_1^3 + \bar{a}_1 t_1^2 + \bar{a}_2 t_1 + \bar{a}_3 = 0, \quad \bar{b}_0 t_1^2 + \bar{b}_1 t_1 + \bar{b}_2 = 0, \tag{2.4}$$

где

$$\begin{aligned}\bar{a}_0 &= c_1(c_2(a_1^2 - a_2^2)t_2^3 + a_2(a_1^2 - a_2^2 + 2c_2^2 + d_2)t_2^2 + c_2(a_1^2 - c_2^2 + 2a_2^2 - d_2)t_2 + a_2(a_1^2 - c_2^2)), \\ \bar{a}_1 &= a_1(c_2(a_1^2 - a_2^2 - 2c_1^2 - d_1)t_2^3 + a_2(a_1^2 - a_2^2 - 2c_1^2 + 2c_2^2 - d_1 + d_2)t_2^2 + \\ &\quad + c_2(a_1^2 + 2a_2^2 - 2c_1^2 - c_2^2 - d_1 - d_2)t_2 + a_2(a_1^2 - 2c_1^2 - c_2^2 - d_1)), \\ \bar{a}_2 &= c_1(c_2(c_1^2 - a_2^2 - 2a_1^2 + d_1)t_2^3 + a_2(c_1^2 - a_2^2 - 2a_1^2 + 2c_2^2 + d_1 + d_2)t_2^2 - \\ &\quad - c_2(2a_1^2 - 2a_2^2 - c_1^2 + c_2^2 - d_1 + d_2)t_2 - a_2(2a_1^2 - c_1^2 + c_2^2 - d_1)), \\ \bar{a}_3 &= a_1(c_2(c_1^2 - a_2^2)t_2^3 - a_2(a_2^2 - c_1^2 - 2c_2^2 - d_2)t_2^2 + c_2(2a_2^2 + c_1^2 - c_2^2 - d_2)t_2 + a_2(c_1^2 - c_2^2)), \\ \bar{b}_0 &= a_1c_1d_2t_2, \quad \bar{b}_1 = a_2c_2d_1t_2^2 + ((a_2^2 - c_2^2)d_1 + (a_1^2 - c_1^2)d_2 - d_1d_2)t_2 - a_2c_2d_1, \quad \bar{b}_2 = -a_1c_1d_2t_2.\end{aligned}$$

Исключая из системы (2.4)  $t_1$  с использованием понятия результата [13], [14] системе Wolfram Mathematica алгебраическое уравнение 12 порядка относительно переменной  $t_2$ , которое после факторизации представляет собой произведение трех полиномов

$$P(t_2) = P_1(t_2)P_2(t_2)P_3(t_2) = 0, \quad (2.5)$$

где  $P_1(t_2)$  и  $P_2(t_2)$  полиномы второго порядка и имеют следующий вид:

$$P_1(t_2) = a_1c_1d_1(a_2t_2 - c_2)^2 = 0, \quad (2.6)$$

$$P_2(t_2) = a_2c_2t_2^2 + (a_2^2 - c_2^2 + d_2)t_2 + a_2c_2 = 0, \quad (2.7)$$

$P_3(t_2)$  — является полиномом 8 порядка

$$P_3(t_2) = p_0t_2^8 + p_1t_2^7 + p_2t_2^6 + p_3t_2^5 + p_4t_2^4 + p_5t_2^3 + p_6t_2^2 + p_7t_2 + p_8 = 0, \quad (2.8)$$

коэффициенты которого представляют собой полиномы от шести параметров системы  $a_1, a_2, c_1, c_2, d_1, d_2$ :

$$\begin{aligned}p_0 &= c_2^4d_1^2(a_1^2 - a_2^2)(a_2^2 - c_1^2), \\ p_1 &= 2a_2c_2^3d_1(d_2(a_1^2 - c_1^2)(a_1^2 - a_2^2 + c_1^2) + d_1(2a_2^2(a_1^2 - a_2^2) + c_1^2(c_2^2 - a_1^2 - d_2) + (2a_2^2 - a_1^2)(c_1^2 + c_2^2 + d_2))), \\ p_2 &= c_2^2(d_2^2(a_1^2 - a_2^2 + c_1^2)(a_1^2 + c_1^2)^2 + 2d_1d_2(a_1^2 - c_1^2)(3a_2^2(a_1^2 - a_2^2 + c_1^2 + c_2^2 + d_2) - \\ &\quad - (a_1^2 + c_1^2)(c_2^2 + d_2)) + d_1^2(6a_2^4(a_1^2 - a_2^2) + c_1^2(c_2^2 + d_2)^2 - a_1^2a_2^2(7c_2^2 + 6c_1^2 + 6d_2) + 2a_2^4(8c_2^2 + \\ &\quad + 3c_1^2 + 6d_2) + a_1^2(c_2^4 - 2c_1^2c_2^2 + 2c_2^2d_2 + d_2^2) - a_2^2(6c_2^4 + 7c_1^2c_2^2 + 6d_2(c_1^2 + c_2^2) + 6d_2^2))), \\ p_3 &= 2a_2c_2(d_2^2(a_1^2 - a_2^2 + c_1^2 + c_2^2 + d_2)(a_1^2 + c_1^2)^2 + d_1(d_2(a_1^2 - c_1^2)(3a_2^2a_2^2 - 3(a_2^2 - d_2)^2 + 2d_2(a_1^2 - c_1^2) - \\ &\quad - 2c_1^2(a_1^2 + c_1^2) - 3c_2^2(c_2^2 + 2d_2) + 3a_2^2(c_1^2 + 3c_2^2)) + d_1^2((c_1^2 + 2c_2^2 + 2d_2)(c_2^2 + d_2)^2 - 2a_2^6 + a_1^2(2a_2^4 - \\ &\quad - 2a_2^2c_1^2 - 4a_2^2c_2^2 - 3a_2^2d_2 - 4c_1^2c_2^2 + c_2^4 + 2c_2^2d_2 + d_2^2) - a_2^2(4c_2^2(c_1^2 + 3c_2^2) + 3d_2(c_1^2 + 6c_2^2) + 6d_2^2))), \\ p_4 &= d_2^2(a_1^2 + c_1^2)^2(a_1^2a_2^2 - (a_2^2 - c_2^2)^2 + 2d_2(a_2^2 - c_2^2) + a_2^2(c_1^2 + 2c_2^2) + c_2^2(a_1^2 + c_1^2) - d_2^2) + \\ &\quad + d_1(2d_2(a_1^2 - c_1^2)(a_2^2 - c_2^2 - d_2)(a_1^2(a_2^2 + c_2^2) + a_2^2(c_1^2 + 8c_2^2 + 2d_2) + c_1^2c_2^2 - (c_2^2 + d_2)^2 - a_2^4) + \\ &\quad + d_1^2(a_2^6(c_1^2 + 16c_2^2 + 4d_2) - a_2^8 - (c_2^4 - c_1^2c_2^2 + 2c_2^2d_2 + d_2^2)(c_2^2 + d_2)^2 - 2a_2^4((c_2^2 + d_2)(c_1^2 + 18c_2^2) + \\ &\quad + 3d_2^2) + a_2^2(4d_2^2 + 36c_2^4d_2 - 2c_2^4(c_1^2 - 8c_2^2) + d_2^2(c_1^2 + 24c_2^2)) + a_1^2(a_2^6 - a_2^4(c_1^2 + 2c_2^2 + 2d_2) + \\ &\quad + c_2^2(c_2^4 - c_1^2c_2^2 + 2c_2^2d_2 + d_2^2) + a_2^2(d_2^2 - 2c_2^4 - 12c_1^2c_2^2))), \\ p_5 &= 2a_2c_2(d_2^2(a_1^2 + a_2^2 + c_1^2 - c_2^2 - d_2)(a_1^2 + c_1^2)^2 + d_1(d_2(a_1^2 - c_1^2)(2(a_1^2 + c_1^2)(a_2^2 - d_2) + \\ &\quad + 3(a_2^2 - d_2)^2 - 3c_2^2(a_1^2 + 3a_2^2 + c_1^2 - c_2^2 - 2d_2)) + d_1^2(2a_2^6 + a_2^4(c_1^2 - 12c_2^2 - 6d_2) - \\ &\quad - (2c_2^4 - 2c_1^2c_2^2 + 4c_2^2d_2 - c_1^2d_2 + 2d_2^2)(c_2^2 + d_2) + 2a_2^2(6c_2^4 - 2c_1^2c_2^2 + 9c_2^2d_2 - c_1^2d_2 + 3d_2^2) + \\ &\quad + a_1^2(a_2^4 + d_2^2 + c_2^2(2c_2^2 - 2c_1^2 + 3d_2) - 2a_2^2(2c_1^2 + 2c_2^2 + d_2))))) ,\end{aligned}$$

$$\begin{aligned}
p_6 &= a_2^2(a_2^2(a_1^2 + c_1^2 - c_2^2)(a_1^2 + c_1^2)^2 + 2d_1d_2(a_1^2 - c_1^2)(a_1^2(a_2^2 - 3c_2^2 - d_2) + a_2^2(c_1^2 - 3c_2^2) + 3c_2^4 - \\
&\quad - 3c_1^2c_2^2 + 3c_2^2d_2 - c_1^2d_1) + d_1^2(a_1^2(a_2^4 + d_2^2 + 6c_2^2d_2 - 6c_1^2c_2^2 + 6c_2^4 - a_2^2(2c_1^2 + 7c_2^2 + 2d_2)) + \\
&\quad + a_2^2(16c_2^4 - 7c_1^2c_2^2 + 12c_2^2d_2 - 2c_1^2d_1) + a_2^4(c_1^2 - 6c_2^2) - 6c_2^2(c_2^2 + d_2)(c_2^2 - c_1^2 + d_2) + c_1^2d_2^2)), \\
p_7 &= -2a_2^3c_2d_1(d_2(a_1^2 - c_1^2)(a_1^2 + c_1^2 - c_2^2) + d_1(a_1^2(a_2^2 - d_2) + c_1^2(2a_1^2 + a_2^2 - d_2) - \\
&\quad - 2c_2^2(a_1^2 + a_2^2 + c_1^2 - c_2^2 - d_2))), \\
p_8 &= -a_2^4d_1^2(a_1^2 - c_2^2)(c_1^2 - c_2^2).
\end{aligned}$$

По определению результата каждому корню  $t_2$  уравнения (2.5) соответствует один общий корень  $t_1$  системы (2.4). Число действительных корней полученного алгебраического уравнения чётно и не превышает 12. Подставляя значение действительного корня  $t_2$  алгебраического уравнения (2.5) в уравнения системы (2.4), найдем совпадающий корень  $t_1$  этих уравнений.

Можно показать, что каждому действительному корню уравнений (2.4) и (2.5) соответствуют 4 равновесных решения исходной системы (2.1). Так как общее число действительных корней уравнения (2.5) не превышает 12, то система двух связанных тел спутник–стабилизатор в плоскости орбиты может иметь не более 48 равновесных ориентаций в орбитальной системе координат. Используя уравнения (2.4)–(2.8) для каждого набора системных параметров, мы можем определить все плоские равновесные ориентации системы спутник–стабилизатор в орбитальной системе координат.

### 3. ИССЛЕДОВАНИЕ ПОЛОЖЕНИЙ РАВНОВЕСИЯ

Уравнения (2.4)–(2.8) позволяют определить все равновесные ориентации системы спутник–стабилизатор, движущейся по круговой орбите при заданных значениях параметров  $a_1, a_2, c_1, c_2, d_1, d_2$ .

Проведем подробное исследование решений системы алгебраических уравнений (2.6)–(2.8). В случае  $P_1(t_2) = 0$  имеет место кратный корень  $\operatorname{tg} \alpha_2 = c_2/a_2$ . Величину  $\operatorname{tg} \alpha_1$  находим из квадратного уравнения (2.4), при значениях параметров, обеспечивающих условие неотрицательности его дискриминанта. Из вида коэффициентов второго уравнения системы (2.4) следует, что дискриминант этого уравнения представляет собой сумму квадратов двух величин и не отрицателен при любых значениях параметров задачи.

Уравнение (2.7)  $P_2(t_2) = 0$  при условии существования двух действительных корней определяет два значения  $\operatorname{tg} \alpha_2$ . Величину  $\operatorname{tg} \alpha_1$  находим из квадратного уравнения (2.4). Заметим, что равновесные решения в этом случае определяются только геометрическими и инерционными характеристиками второго тела.

Для исследования числа положений равновесия системы спутник–стабилизатор в случае  $P_3(t_2) = 0$  определим в пространстве 6 параметров области с одинаковым числом вещественных корней этого алгебраического уравнения 8 порядка. Для этого выделим в пространстве параметров области с одинаковым числом вещественных корней уравнения (2.7) путем построения дискриминантной гиперповерхности полинома  $P_3(t_2)$ . Эта гиперповерхность содержит компоненту коразмерности 1, которая является границей областей с одинаковым числом вещественных корней. Множество особых точек дискриминантной гиперповерхности в пространстве параметров задается следующей системой алгебраических уравнений:

$$P_3(t_2) = 0, \quad P_3'(t_2) = 0. \quad (3.1)$$

Символ “штрих” обозначает дифференцирование по переменной  $t_2$ .

Исключая переменную из системы (3.1) путем символического вычисления определителя матрицы результата 15-го порядка уравнений (3.1) с использованием матричных функций системы Wolfram Mathematica, получим выражение для гиперповерхности, которое имеет очень громоздкий вид и в данной статье не приводится.

Другим способом определить число действительных корней полинома можно путем вычисления  $i$ -х субдискриминантов [15], [16].

Рассмотрим несколько более простых случаев.

**Случай 1:**  $a_1 = a_2 = c_1 = c_2 = a$ . В этом случае, перейдя к новым параметрам в уравнении (2.4)  $d_{01} = (B_1 - A_1)/Ma^2$  и  $d_{02} = (B_2 - A_2)/Ma^2$ , получим из уравнения (2.8) алгебраическое уравнение 4 степени

$$P_4(t_2) = p_{00}t_2^4 + p_{01}t_2^3 + p_{02}t_2^2 + p_{03}t_2 + p_{04} = 0, \quad (3.2)$$

где

$$\begin{aligned} p_{00} &= 4(d_{01}d_{02} + d_{01} - d_{02})(d_{01}d_{02} + d_{01} + d_{02}), & p_{01} &= -4(d_{02} + 2)(d_{01}^2d_{02}^2 - d_{01}^2d_{02} - 2d_{01}^2 + 2d_{02}^2), \\ p_{02} &= (d_{01}^2d_{02}^4 + 4d_{02}^4 - 16d_{01}^2d_{02}^2 + 24(d_{01}^2 - d_{02}^2)), & p_{03} &= 4(d_{02} - 2)(d_{01}^2d_{02}^2 + d_{01}^2d_{02} - 2d_{01}^2 + 2d_{02}^2), \\ p_{04} &= 4(d_{01}d_{02} - d_{01} - d_{02})(d_{01}d_{02} - d_{01} + d_{02}). \end{aligned}$$

Определим теперь условия существования действительных корней уравнения (3.2). Для этого вычислим дискриминант полинома  $P_4(t_2)$ , который задает на плоскости  $d_{01}, d_{02}$ , уравнение дискриминантной кривой. Границы областей с равным числом действительных корней на плоскости параметров  $d_{01}, d_{02}$ , определяются следующей системой алгебраических уравнений:

$$P_4(t_2) = 0, \quad P_4'(t_2) = 0. \quad (3.3)$$

Исключим переменную  $t_2$  из системы (3.3) путем символического вычисления определителя результата этих уравнений и получим уравнение дискриминантной кривой в виде

$$P_5(d_{01}, d_{02}) = 1024d_{02}^{12}P_6(d_{01}, d_{02})P_7(d_{01}, d_{02})P_8(d_{01}, d_{02}) = 0. \quad (3.4)$$

Здесь

$$\begin{aligned} P_6(d_{01}, d_{02}) &= (d_{01}^2(d_{02} + 1)^2 - d_{02}^2), & P_7(d_{01}, d_{02}) &= ((d_{01}^2d_{02}^2 + 4(d_{01} - d_{02})^2)((d_{01}^2d_{02}^2 + 4(d_{01} + d_{02})^2), \\ P_8(d_{01}, d_{02}) &= (d_{01}^4(d_{02}^2 + 8)^2 + 16d_{01}^2d_{02}^2(d_{02}^2 + 10) + 64(d_{02}^4 - 108)). \end{aligned}$$

Проверим теперь, как изменяется число равновесных решений при пересечении одной из кривых (3.4). Это можно сделать численно, определив число действительных решений в одной точке из каждой области, ограниченной кривыми (3.4) на плоскости  $(d_{01}, d_{02})$ . Численный анализ показал, что только кривая  $P_8(d_{01}, d_{02})$  разделяет области с различным числом действительных корней уравнения (3.2). При пересечении границ кривых заданными уравнениями  $P_6(d_{01}, d_{02}) = 0$  и  $P_7(d_{01}, d_{02}) = 0$  смены числа действительных корней уравнения (3.2) не происходит. На фиг. 2 показаны области с одинаковым числом действительных решений уравнения (3.2), равным 4 и 2, где существуют 16 и 8 равновесных ориентаций системы двух связанных тел соответственно. Границы этих областей задаются уравнением  $P_8(d_{01}, d_{02}) = 0$ . Кривые границ областей на фиг. 2 построены с использованием функции *ContourPlot* в системе Wolfram Mathematica.

Таким образом, для случая 1  $a_1 = a_2 = b_1 = b_2 = a$  существуют только 16 и 8 равновесных ориентаций системы спутник-стабилизатор в орбитальной системе координат в плоскости орбиты.

**Случай 2:**  $a_1 \neq 0; a_2 \neq 0; c_1 \neq 0; c_2 = 0$ . В этом случае уравнения (2.1) примут следующий вид:

$$\begin{aligned} (d_1 - a_1^2 + c_1^2)\sin \alpha_1 \cos \alpha_1 + a_1c_1(\cos^2 \alpha_1 - \sin^2 \alpha_1) + a_2(a_1 \cos \alpha_1 + c_1 \sin \alpha_1)\sin \alpha_2 &= 0, \\ \cos \alpha_2[(d_2 - a_2^2)\sin \alpha_2 + a_2(a_1 \sin \alpha_1 - c_1 \cos \alpha_1)] &= 0. \end{aligned} \quad (3.5)$$

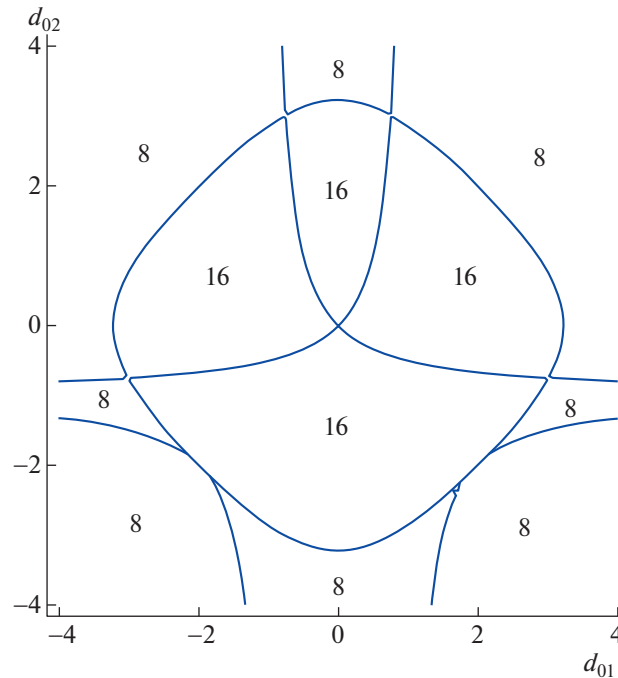
Из второго уравнения (3.5) следуют два случая:

$$\cos \alpha_2 = 0 \quad (\sin^2 \alpha_2 = 1), \quad (3.6)$$

$$\sin \alpha_2 = -\frac{a_2}{(d_2 - a_2^2)}(a_1 \sin \alpha_1 - c_1 \cos \alpha_1) \quad (\text{при } d_2 \neq a_2^2). \quad (3.7)$$

В случае (3.6) первое уравнение перепишем в виде

$$(d_1 - a_1^2 + c_1^2)\sin \alpha_1 \cos \alpha_1 + a_1c_1(\cos^2 \alpha_1 - \sin^2 \alpha_1) = \mp a_2(a_1 \cos \alpha_1 + c_1 \sin \alpha_1),$$



Фиг. 2. Области с равным числом положений равновесия в плоскости орбиты.

после чего, проведя универсальную замену (2.2) синуса и косинуса угла  $\alpha_1$  через тангенс, получим алгебраическое уравнение 4 порядка относительно  $t_1 = \tan \alpha_1$

$$P_9(t_1) = c_1^2(a_1^2 - a_2^2)t_1^4 - 2a_1c_1(a_2^2 - a_1^2 + c_1^2 + d_1)t_1^3 + ((c_1^2 - a_1^2 + d_1)^2 - a_1^2a_2^2 - 2a_1^2c_1^2 - a_2^2c_1^2)t_1^2 - 2a_1c_1(a_1^2 + a_2^2 - c_1^2 - d_1)t_1 + a_1^2(c_1^2 - a_2^2) = 0. \quad (3.8)$$

Из вида уравнения (3.8) следует, что число его действительных корней не превышает 4, а следовательно, число различных положений равновесия системы двух связанных тел не превышает, как и в предыдущем случае 16.

Рассмотрим теперь случай (3.7). Подставляя значения  $\sin \alpha_2$  из (3.7) в первое уравнение системы (3.5) получим однородное тригонометрическое уравнение, которое преобразуется в квадратное относительно  $\tan \alpha_1$ , с учетом того, что  $\cos \alpha_2 \neq 0$

$$a_1c_1d_2 \tan^2 \alpha_1 + [a_1^2d_2 + a_2^2d_1 - (d_1 + c_1^2)d_2] \tan \alpha_1 - a_1c_1d_2 = 0. \quad (3.9)$$

Из вида коэффициентов уравнения (3.9) следует, что дискриминант этого уравнения представляет собой сумму квадратов двух величин и не отрицателен при допустимых для случая (3.7) значениях параметров задачи ( $d_2 \neq a_2^2$ ). Отсюда получается, что уравнение (3.9) имеет два действительных корня, а следовательно, для случая (3.7) существует 8 различных положений равновесия системы двух связанных тел.

Другие 3 случая, когда один из параметров  $a_1$ ,  $a_2$  или  $c_1$  равны нулю, рассматриваются аналогично случаю 2 и подробно будут исследованы в отдельной статье.

**Случай 3.**  $a_1 \neq 0$ ;  $a_2 \neq 0$ ;  $c_1 = 0$ ;  $c_2 = 0$ . Уравнения (2.1) в этом случае упрощаются и их можно переписать в виде:

$$\begin{aligned} \cos \alpha_1 [((A_1 - C_1)/M) - a_1^2] \sin \alpha_1 + a_1 a_2 \sin \alpha_2 &= 0, \\ \cos \alpha_2 [((A_2 - C_2)/M) - a_2^2] \sin \alpha_2 + a_1 a_2 \sin \alpha_1 &= 0. \end{aligned} \quad (3.10)$$

Уравнения (3.10) определяют равновесные ориентации системы двух связанных тел в плоскости орбиты, когда сферический шарнир расположен на пересечении главных центральных осей инерции спутника и стабилизатора.

Если уравнения (3.10) переписать с учетом введения новых параметров  $m_1 = a_1 a_2 / [(A_1 - C_1)/M - a_1^2]$  и  $m_2 = a_1 a_2 / [(A_2 - C_2)/M - a_2^2]$ , то в результате можно получить, что система (3.10) имеет следующие 4 типа решений:

$$\begin{aligned} \sin \alpha_1 = 0, \quad \sin \alpha_2 = 0, \quad \cos \alpha_1 = 0, \quad \cos \alpha_2 = 0, \\ \cos \alpha_1 = 0, \quad (\sin \alpha_1 + m_2 \sin \alpha_2) = 0, \quad \cos \alpha_2 = 0, \quad (m_1 \sin \alpha_1 + \sin \alpha_2) = 0. \end{aligned} \quad (3.11)$$

Решения (3.11) при  $|m_1| < 1$ ,  $|m_2| < 1$ ,  $m_1 m_2 \neq 1$  в зависимости от знаков параметров определяют в каждом случае 16 различных положений равновесия системы двух связанных тел. В работе [6] были определены все положения равновесия, которые задаются уравнениями (3.10). С использованием интеграла энергии в качестве функции Ляпунова получены достаточные условия устойчивости положений равновесия. Рассмотрена возможность обеспечения асимптотической устойчивости положений равновесия при наличии диссипации.

### ЗАКЛЮЧЕНИЕ

В настоящей работе проведено исследование движения системы двух тел, соединенных сферическим шарниром, движущихся по круговой орбите под действием гравитационного момента.

Главное внимание уделено изучению равновесных ориентаций системы двух тел в плоскости круговой орбиты. Предложен алгебраический метод определения положений равновесия системы двух тел в орбитальной системе координат при условии, когда сферический шарнир расположен на линии пересечения двух плоскостей, образуемых главными центральными осями инерции спутника и стабилизатора.

Полученные результаты могут быть использованы на этапе предварительного проектирования гравитационных систем управления ориентацией искусственных спутников Земли.

### СПИСОК ЛИТЕРАТУРЫ

1. Сарычев В.А. Вопросы ориентации искусственных спутников. Итоги науки и техники. Сер. Исследование космического пространства. Т. 11. М.: ВИНТИ, 1978.
2. Гутник С.А., Сарычев В.А. Исследование стационарных движений системы двух связанных тел на круговой орбите с применением методов компьютерной алгебры // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 1. С. 80–87.
3. Гутник С.А., Сарычев В.А. Применение методов компьютерной алгебры для исследования динамики системы двух связанных тел на круговой орбите // Программирование. 2019. Т. 45. № 2. С. 32–40.
4. Gutnik S.A., Sarychev V.A. Symbolic computations of the equilibrium orientations of a system of two connected bodies moving on a circular orbit around the Earth // Math. in Comput. Scie. 2021. V. 15. № 3. P. 407–417.
5. Сарычев В.А. Положения равновесия системы двух соединенных сферическим шарниром осесимметричных тел на круговой орбите // Космические исследования. 1999. Т. 37. № 2. С. 176–181.
6. Сарычев В.А. Положения относительного равновесия двух тел, соединенных сферическим шарниром, на круговой орбите // Космические исследования. 1967. Т. 5. № 3. С. 360–364.
7. Сарычев В.А., Сазонов В.В. Оптимальные параметры пассивных систем ориентации спутников // Космические исследования. 1976. Т. 14. № 2. С. 198–208.
8. Сарычев В.А., Мирер С.А. Оптимальные параметры гравитационной системы спутник–стабилизатор // Космические исследования. 1976. Т. 14. № 2. С. 209–219.
9. Sarychev V.A., Mirer S.A., Sazonov V.V. Plane oscillations of a gravitational system satellite-stabilizer with maximal speed of response // Acta Astronautica. 1976. V. 3. № 9–10. P. 651–669.
10. Gutnik S.A., Sarychev V.A. Symbolic investigation of the dynamics of a system of two connected bodies moving along a circular orbit // In: England M., Koepe W., Sadykov T.M., Seiler W.M., Vorozhtsov E.V. Eds. CASC 2019. LNCS. Springer, Cham, 2019. V. 11661. P. 164–178.
11. Гутник С.А., Сарычев В.А. Символьные методы вычисления положений равновесия системы двух связанных тел на круговой орбите // Программирование. 2022. Т. 48. № 2. С. 16–22.
12. Wolfram Mathematica website. <http://www.wolfram.com/mathematica>
13. Курош А.Г. Курс высшей алгебры. М.: Наука, 1971. 432 с.
14. Сушкевич А.К. Основы высшей алгебры. М.: ОНТИ, 1937. 476 с.
15. Батхин А.Б. Параметризация дискриминантного множества многочлена // Программирование. 2016. № 2. С. 8–21.
16. Батхин А.Б. Параметризация множества, определяемого обобщенным дискриминантом многочлена // Программирование. 2018. № 2. С. 5–17.

---

МАТЕМАТИЧЕСКАЯ  
ФИЗИКА

---

УДК 517.93

ИССЛЕДОВАНИЕ ВЕКОВЫХ ВОЗМУЩЕНИЙ  
В ОГРАНИЧЕННОЙ ЗАДАЧЕ ТРЕХ ТЕЛ ПЕРЕМЕННОЙ МАССЫ  
С ПРИМЕНЕНИЕМ КОМПЬЮТЕРНОЙ АЛГЕБРЫ

© 2023 г. А. Т. Ибраимова<sup>1,2,\*</sup>, М. Дж. Минглибаев<sup>1,2,\*\*</sup>, А. Н. Прокопеня<sup>3,\*\*\*</sup>

<sup>1</sup> 050040 Алматы, пр-т аль-Фараби, 71, Казахский национальный университет им. аль-Фараби, Казахстан

<sup>2</sup> 050020 Алматы, Обсерватория, 23, Астрофизический институт им. В.Г. Фесенкова, Казахстан

<sup>3</sup> 02-776 SGGW Варшава, ул. Новоурсыновска, 159, Варшавский университет естественных наук, Польша

\*e-mail: [ibraimova@aphi.kz](mailto:ibraimova@aphi.kz)

\*\*e-mail: [minglibayev@gmail.com](mailto:minglibayev@gmail.com)

\*\*\*e-mail: [alexander\\_prokopenya@sggw.edu.pl](mailto:alexander_prokopenya@sggw.edu.pl)

Поступила в редакцию 04.08.2022 г.  
Переработанный вариант 04.08.2022 г.  
Принята к публикации 10.09.2022 г.

Рассматривается нестационарная ограниченная задача трех тел переменной массы с учетом реактивных сил, возникающих при неизотропном изменении масс тел. Предполагается, что тела являются сферически симметричными и взаимодействуют между собой в соответствии с законом всемирного тяготения Ньютона. На основе уравнений движения тел в относительной системе координат получены дифференциальные уравнения в оскулирующих элементах аperiодического движения по квазиконическим сечениям. В случае малых эксцентриситетов и наклонов орбит получены уравнения, определяющие вековые возмущения орбитальных элементов. Все необходимые символьные вычисления выполнены с помощью системы компьютерной алгебры Wolfram Mathematica. Библ. 21.

**Ключевые слова:** ограниченная задача трех тел, переменная масса, уравнения движения, вековые возмущения.

**DOI:** 10.31857/S004446692301009X, **EDN:** LMGDHY

## 1. ВВЕДЕНИЕ

Наблюдения за движением небесных тел показывают, что реальные космические системы являются нестационарными, их массы, размеры, форма и ряд других физических характеристик изменяются с течением времени в процессе эволюции (см. [1]–[6]). Учет нестационарности небесных тел приводит к существенному усложнению математических моделей их движения, что требует разработки новых алгоритмов символьных вычислений для их исследования.

Целью настоящей работы является получение дифференциальных уравнений, определяющих вековые возмущения орбитальных элементов двух тел  $P_1$ ,  $P_2$  переменной массы  $m_1 = m_1(t)$ ,  $m_2 = m_2(t)$ , движущихся вокруг нестационарного центрального тела  $P_0$  массой  $m_0 = m_0(t)$ , в рамках ограниченной задачи трех тел. Предполагается, что движение двух массивных тел  $P_0$ ,  $P_1$  определяется их взаимным гравитационным притяжением, а также реактивными силами, которые возникают при неизотропном изменении их масс. Третье тело  $P_2$  пренебрежимо малой массы ( $m_2 \ll m_0, m_1$ ) не влияет на движение двух массивных тел и движется в их гравитационном поле. Хотя масса третьего тела предполагается малой, ее изменение также приводит к возникновению реактивных сил и возмущению его движения. Исследование этой проблемы связано с довольно громоздкими символьными вычислениями, которые лучше всего выполнять с применением систем компьютерной алгебры (см. [7]–[11]). В данной работе все необходимые вычисления выполнены с помощью системы Wolfram Mathematica (см. [12], [13]).

## 2. ОПИСАНИЕ МОДЕЛИ

Движение тел  $P_1, P_2$  удобно рассматривать в относительной системе координат  $Ox_1y_1z_1$ , начало которой связано с центром масс тела  $P_0$ , а оси параллельны осям абсолютной инерциальной системы координат. Обозначая радиус-векторы тел через  $\mathbf{r}_j = (x_j, y_j, z_j)$ ,  $j = 1, 2$ , и используя второй закон Ньютона, уравнения движения можно записать в виде (см. [6], [9], [14])

$$\frac{d^2\mathbf{r}_1}{dt^2} + G(m_0 + m_1)\frac{\mathbf{r}_1}{r_1^3} - \frac{d^2\gamma_1}{dt^2}\frac{\mathbf{r}_1}{\gamma_1} = \mathbf{F}_1, \quad (1)$$

$$\frac{d^2\mathbf{r}_2}{dt^2} + Gm_0\frac{\mathbf{r}_2}{r_2^3} - \frac{d^2\gamma_2}{dt^2}\frac{\mathbf{r}_2}{\gamma_2} = \mathbf{F}_2. \quad (2)$$

Здесь  $G$  — гравитационная постоянная, дважды непрерывно дифференцируемые функции  $\gamma_1(t)$  и  $\gamma_2(t)$  определяются соотношениями

$$\gamma_1(t) = \frac{m_{00} + m_{10}}{m_0(t) + m_1(t)}, \quad \gamma_2(t) = \frac{m_{00}}{m_0(t)},$$

где  $m_{00} = m_0(t_0)$ ,  $m_{10} = m_1(t_0)$  — значения масс тел в начальный момент времени. Силы  $\mathbf{F}_1, \mathbf{F}_2$  в правых частях уравнений (1), (2) можно представить в виде

$$\mathbf{F}_1 = \mathbf{F}_1^{(r)} - \frac{\ddot{\gamma}_1}{\gamma_1}\mathbf{r}_1, \quad \mathbf{F}_2 = Gm_1\left(\frac{\mathbf{r}_1 - \mathbf{r}_2}{r_{12}^3} - \frac{\mathbf{r}_1}{r_1^3}\right) + \mathbf{F}_2^{(r)} - \frac{\ddot{\gamma}_2}{\gamma_2}\mathbf{r}_2, \quad (3)$$

где

$$r_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}, \quad r_1 = \sqrt{x_1^2 + y_1^2 + z_1^2},$$

а реактивные силы  $\mathbf{F}_1^{(r)}, \mathbf{F}_2^{(r)}$  определяются выражениями (см. [15])

$$\mathbf{F}_1^{(r)} = \frac{\dot{m}_1}{m_1}\mathbf{V}_1 - \frac{\dot{m}_0}{m_0}\mathbf{V}_0, \quad \mathbf{F}_2^{(r)} = \frac{\dot{m}_2}{m_2}\mathbf{V}_2 - \frac{\dot{m}_0}{m_0}\mathbf{V}_0. \quad (4)$$

Точка над символом в выражениях (3), (4) означает полную производную соответствующей функции по времени, а  $\mathbf{V}_j$ ,  $j = 0, 1, 2$ , — относительные скорости частиц, покидающих тело  $P_j$  или осаждающихся на нем, которые являются заданными функциями времени. Следует отметить,

что слагаемые  $\frac{d^2\gamma_i}{dt^2}\frac{\mathbf{r}_i}{\gamma_i}$  в левой части уравнений (1), (2), а также в выражениях (3) для сил  $\mathbf{F}_1, \mathbf{F}_2$ , вво-

дятся для того, чтобы получаемые при  $\mathbf{F}_1 = 0, \mathbf{F}_2 = 0$  уравнения были интегрируемы и можно было записать их общее решение при произвольных законах изменения масс тел. Такое решение далее используется в качестве первого приближения при построении возмущенных решений уравнений (1), (2). Законы изменения масс тел во времени  $m_j(t)$  определяются на основе наблюдений и также считаются известными, причем в общем случае массы изменяются в различных темпах, т.е. справедливо соотношение

$$\frac{\dot{m}_0}{m_0} \neq \frac{\dot{m}_1}{m_1} \neq \frac{\dot{m}_2}{m_2}.$$

Поэтому далее будем считать, что реактивные силы  $\mathbf{F}_1^{(r)}, \mathbf{F}_2^{(r)}$  являются заданными функциями времени.

Поскольку получить общее решение уравнений (1), (2) не представляется возможным даже в случае постоянных масс, воспользуемся теорией возмущений. Заметим, что при  $\mathbf{F}_1 = 0, \mathbf{F}_2 = 0$  уравнения (1), (2) являются независимыми, причем каждое из них имеет точное решение, которое описывает аperiодическое движение по квазиконическому сечению и может быть представлено в виде

$$\begin{aligned} x_j &= \gamma_j \rho_j (\cos(v_j + \omega_j) \cos \Omega_j - \sin(v_j + \omega_j) \sin \Omega_j \cos i_j), \\ y_j &= \gamma_j \rho_j (\cos(v_j + \omega_j) \sin \Omega_j + \sin(v_j + \omega_j) \cos \Omega_j \cos i_j), \end{aligned} \quad (5)$$

$$z_j = \gamma_j \rho_j \sin(v_j + \omega_j) \sin i_j, \quad j = 1, 2,$$

где  $v_j$  — истинная аномалия,

$$\rho_j = \frac{a_j(1 - e_j^2)}{1 + e_j \cos v_j}. \quad (6)$$

Постоянные  $a_j, e_j, i_j, \Omega_j, \omega_j$  в (5), (6) являются аналогами известных кеплеровских орбитальных элементов (см. [6]). Введение аналога эксцентрической аномалии  $E_j$  согласно соотношению

$$\operatorname{tg} \frac{v_j}{2} = \sqrt{\frac{1 + e_j}{1 - e_j}} \operatorname{tg} \frac{E_j}{2} \quad (7)$$

приводит к известному уравнению Кеплера

$$E_j - e_j \sin E_j = M_j. \quad (8)$$

При этом зависимости эксцентрической  $E_j$  и средней  $M_j$  аномалий в (8) от времени

$$E_j = E_j(t), \quad M_j = \frac{\sqrt{\kappa_j}}{a_j^{3/2}} (\phi_j(t) - \phi_j(\tau_j)), \quad \kappa_1 = G(m_{00} + m_{10}), \quad \kappa_2 = Gm_{00} \quad (9)$$

определяются законами изменения масс тел, поскольку функции  $\phi_j(t)$  в (9) имеют вид

$$\phi_j(t) = \int_{t_0}^t \frac{dt}{\gamma_j^2(t)}. \quad (10)$$

Через  $\tau_j$  в (9) обозначен аналог времени прохождения тела  $P_j$  через перигеум.

Отметим, что при заданных орбитальных параметрах  $a_j, e_j, i_j, \Omega_j, \omega_j, \tau_j$  каждого из тел  $P_1, P_2$ , а также известных функциях  $\gamma_1(t), \gamma_2(t)$ , которые определяются законами изменения масс всех трех тел, уравнения (9), (10) позволяют найти средние аномалии  $M_j$  как функции времени. Решая затем уравнение Кеплера (8), находим эксцентрические аномалии  $E_j$  и, используя соотношение (7), вычисляем истинные аномалии  $v_j$  и декартовы координаты тел (5). Таким образом, соотношения (5)–(10) позволяют вычислить относительные декартовы координаты тел  $P_1, P_2$  в любой момент времени и полностью описать их невозмущенное движение. Отметим также, что при постоянных массах тел ( $\gamma_1(t) = \gamma_2(t) \equiv 1$ ) решения (5), (6) сводятся к хорошо известным решениям задачи двух тел (см. [16], [17]).

### 3. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ВОЗМУЩЕННОГО ДВИЖЕНИЯ

Как и в случае постоянных масс, гравитационное взаимодействие тел  $P_1$  и  $P_2$  приводит к возмущению траектории тела  $P_2$  пренебрежимо малой массы. В отличие от классической ограниченной задачи трех тел (см. [18]), в рассматриваемом случае переменных масс учет реактивных сил приводит не только к дополнительному возмущению траектории тела  $P_2$ , но и к возмущению траектории массивного тела  $P_1$ . Это означает, что учет сил  $\vec{F}_1 \neq 0, \vec{F}_2 \neq 0$  в правых частях уравнений (1), (2) приводит к зависимости орбитальных параметров  $a_j, e_j, i_j, \Omega_j, \omega_j, \tau_j$  от времени. Используя метод вариации постоянных, решения уравнений движения (1), (2) можно представить в общем виде

$$\begin{aligned} x_j &= x_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t), \\ y_j &= y_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t), \\ z_j &= z_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t), \end{aligned} \quad (11)$$

где функции в правых частях определяются выражениями (5), в которых вместо истинных аномалий  $v_j$  используются средние аномалии  $M_j$ , а орбитальные параметры  $a_j, e_j, i_j, \Omega_j, \omega_j, \tau_j$

( $j = 1, 2$ ) каждого из тел  $P_1, P_2$  являются функциями времени, причем соотношения (8), (9) определяют неявную зависимость средних аномалий  $M_j$  от параметров  $a_j, e_j, \tau_j$ .

Частные производные функций (11) по времени совпадают с производными по времени функций (5) в отсутствие возмущений ( $F_1 \neq 0, F_2 \neq 0$ ) и определяют скорости их изменения. Используя уравнения (5)–(10), получаем

$$\begin{aligned}\dot{x}_j &= \left( \frac{\dot{\gamma}_j}{\gamma_j} + \frac{\dot{\rho}_j}{\rho_j} \right) x_j - \gamma_j \rho_j \dot{v}_j (\sin(v_j + \omega_j) \cos \Omega_j + \cos(v_j + \omega_j) \sin \Omega_j \cos i_j), \\ \dot{y}_j &= \left( \frac{\dot{\gamma}_j}{\gamma_j} + \frac{\dot{\rho}_j}{\rho_j} \right) y_j - \gamma_j \rho_j \dot{v}_j (\sin(v_j + \omega_j) \sin \Omega_j - \cos(v_j + \omega_j) \cos \Omega_j \cos i_j), \\ \dot{z}_j &= \left( \frac{\dot{\gamma}_j}{\gamma_j} + \frac{\dot{\rho}_j}{\rho_j} \right) z_j + \gamma_j \rho_j \dot{v}_j \cos(v_j + \omega_j) \sin i_j, \quad j = 1, 2,\end{aligned}\quad (12)$$

где

$$\dot{v}_j = \frac{\sqrt{\kappa_j}}{a_j^{3/2} (1 - e_j^2)^{3/2}} \frac{(1 + e_j \cos v_j)^2}{\gamma_j^2(t)}.$$

При наличии возмущений орбитальные параметры  $a_j, e_j, i_j, \Omega_j, \omega_j, \tau_j$  ( $j = 1, 2$ ) каждого из тел  $P_1, P_2$  в выражениях для производных (12) также являются функциями времени, формально их можно представить в виде

$$\begin{aligned}\dot{x}_j &= \dot{x}_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t), \\ \dot{y}_j &= \dot{y}_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t), \\ \dot{z}_j &= \dot{z}_j(a_j, e_j, i_j, \Omega_j, \omega_j, M_j(a_j, e_j, \tau_j), t),\end{aligned}\quad (13)$$

где функции в правых частях определяются выражениями (12).

Для получения дифференциальных уравнений, определяющих зависимость орбитальных параметров от времени, следует подставить выражения (11), (13) в уравнения (1), (2), что приводит к шести уравнениям второго порядка относительно 12 искоемых функций  $a_j(t), e_j(t), i_j(t), \Omega_j(t), \omega_j(t), \tau_j(t)$  ( $j = 1, 2$ ). Поскольку такая система имеет бесконечное число решений, требуется ввести дополнительные шесть уравнений для переменных  $a_j, e_j, i_j, \Omega_j, \omega_j, \tau_j$  ( $j = 1, 2$ ). Для получения таких уравнений естественным с физической точки зрения является условие равенства частных производных функций (11) по времени скоростям изменения возмущенных координат  $\dot{x}_j, \dot{y}_j, \dot{z}_j$  (см. [17], [19]). Это условие означает, что в каждый момент времени выражения для координат (11) и скоростей (13) определяются точными решениями (5), (12) задачи двух тел с переменными массами, в которых орбитальные параметры являются функциями времени. В результате получаются шесть уравнений вида (ниже мы приводим только уравнения, записанные для функций  $x_j$ , заданных уравнениями (11); уравнения для функций  $y_j, z_j$  записываются аналогичным образом)

$$\frac{\partial x_j}{\partial a_j} \frac{da_j}{dt} + \frac{\partial x_j}{\partial e_j} \frac{de_j}{dt} + \frac{\partial x_j}{\partial i_j} \frac{di_j}{dt} + \frac{\partial x_j}{\partial \Omega_j} \frac{d\Omega_j}{dt} + \frac{\partial x_j}{\partial \omega_j} \frac{d\omega_j}{dt} + \frac{\partial x_j}{\partial \tau_j} \frac{d\tau_j}{dt} = 0, \quad j = 1, 2. \quad (14)$$

Еще шесть уравнений получаем при подстановке решений (11) и (13) в уравнения движения (1), (2). Соответствующее уравнение для функций  $\dot{x}_j$ , заданных уравнениями (13), имеет вид (уравнения для функций  $\dot{y}_j, \dot{z}_j$  записываются аналогичным образом)

$$\frac{\partial \dot{x}_j}{\partial a_j} \frac{da_j}{dt} + \frac{\partial \dot{x}_j}{\partial e_j} \frac{de_j}{dt} + \frac{\partial \dot{x}_j}{\partial i_j} \frac{di_j}{dt} + \frac{\partial \dot{x}_j}{\partial \Omega_j} \frac{d\Omega_j}{dt} + \frac{\partial \dot{x}_j}{\partial \omega_j} \frac{d\omega_j}{dt} + \frac{\partial \dot{x}_j}{\partial \tau_j} \frac{d\tau_j}{dt} = F_{jx}, \quad j = 1, 2. \quad (15)$$

Решение системы (14), (15) позволяет найти выражения для производных орбитальных параметров  $\frac{da_j}{dt}, \frac{de_j}{dt}, \frac{di_j}{dt}, \frac{d\Omega_j}{dt}, \frac{d\omega_j}{dt}, \frac{d\tau_j}{dt}$  для каждого из тел  $P_1, P_2$ . Отметим, что реализация описанной процедуры получения дифференциальных уравнений для орбитальных параметров требует выполнения довольно громоздких символьных вычислений. Такие вычисления можно реализовать, используя, например, систему компьютерной алгебры Wolfram Mathematica (см. [12]) и ее встроенные функции для дифференцирования выражений, преобразования тригонометрических функций, решения алгебраических уравнений и упрощения символьных выражений. Выполняя соответствующие вычисления и преобразования, в результате получаем систему дифференциальных уравнений для определения зависимости орбитальных параметров от времени (для дальнейших вычислений вместо параметра  $\tau_j$  более удобно использовать среднюю аномалию  $M_j$ ):

$$\frac{da_j}{dt} = \frac{2a_j^{3/2}}{\sqrt{\kappa_j}} \frac{\gamma_j(t)}{1 - e_j \cos E_j} \left( \sqrt{1 - e_j^2} F_{\psi} + e_j \sin E_j F_{rj} \right), \quad (16)$$

$$\frac{de_j}{dt} = \frac{\sqrt{a_j(1 - e_j^2)}}{\sqrt{\kappa_j}} \frac{\gamma_j(t)}{1 - e_j \cos E_j} \left( F_{\psi} (2 \cos E_j - e_j - e_j \cos^2 E_j) + \sqrt{1 - e_j^2} \sin E_j F_{rj} \right), \quad (17)$$

$$\frac{di_j}{dt} = \frac{\sqrt{a_j} \gamma_j(t)}{\sqrt{\kappa_j} (1 - e_j^2)} F_{nj} \left( (\cos E_j - e_j) \cos \omega_j - \sqrt{1 - e_j^2} \sin E_j \sin \omega_j \right), \quad (18)$$

$$\frac{d\Omega_j}{dt} = \frac{\sqrt{a_j} \gamma_j(t)}{\sqrt{\kappa_j} (1 - e_j^2) \sin i_j} F_{nj} \left( (\cos E_j - e_j) \sin \omega_j + \sqrt{1 - e_j^2} \sin E_j \cos \omega_j \right), \quad (19)$$

$$\begin{aligned} \frac{d\omega_j}{dt} = & \frac{\sqrt{a_j}}{e_j \sqrt{\kappa_j} (1 - e_j \cos E_j)} \frac{\gamma_j(t)}{1 - e_j \cos E_j} \left( (2 - e_j^2 - e_j \cos E_j) \sin E_j F_{\psi} - \sqrt{1 - e_j^2} (\cos E_j - e_j) F_{rj} \right) - \\ & - \frac{\sqrt{a_j} \gamma_j(t) \cot i_j}{\sqrt{\kappa_j} (1 - e_j^2)} F_{nj} \left( (\cos E_j - e_j) \sin \omega_j + \sqrt{1 - e_j^2} \sin E_j \cos \omega_j \right), \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{dM_j}{dt} = & \frac{\sqrt{\kappa_j}}{a_j^{3/2} \gamma_j^2(t)} + \frac{\sqrt{a_j}}{e_j \sqrt{\kappa_j} (1 - e_j \cos E_j)} \frac{\gamma_j(t)}{1 - e_j \cos E_j} \left( \sqrt{1 - e_j^2} F_{\psi} \sin E_j (-2 + e_j^2 + e_j \cos E_j) + \right. \\ & \left. + F_{rj} ((1 + 3e_j^2) \cos E_j - e_j (3 + e_j^2 \cos 2E_j)) \right), \quad j = 1, 2. \end{aligned} \quad (21)$$

Силы  $F_{rj}, F_{\psi}, F_{nj}$  в правых частях уравнений (16)–(21) представляют собой соответственно радиальные, трансверсальные и нормальные составляющие сил  $\mathbf{F}_1, \mathbf{F}_2$ , определяемых выражениями (3). Использование орбитальных систем координат тел  $P_1, P_2$  связано с тем, что реактивные силы  $\mathbf{F}_1^{(r)}, \mathbf{F}_2^{(r)}$  обычно определяются именно в этих системах, а направляющие косинусы единичных векторов  $\mathbf{e}_{rj} = (e_{xj}, e_{yj}, e_{zj}), \mathbf{e}_{\psi} = (\tau_{xj}, \tau_{yj}, \tau_{zj}), \mathbf{e}_{nj} = (n_{xj}, n_{yj}, n_{zj})$  вдоль радиального, трансверсального и нормального направлений соответственно можно легко записать на основе решений (5):

$$\begin{aligned} e_{xj} &= \cos(v_j + \omega_j) \cos \Omega_j - \sin(v_j + \omega_j) \sin \Omega_j \cos i_j, \\ e_{yj} &= \cos(v_j + \omega_j) \sin \Omega_j + \sin(v_j + \omega_j) \cos \Omega_j \cos i_j, \end{aligned} \quad (22)$$

$$\begin{aligned} e_{zj} &= \sin(v_j + \omega_j) \sin i_j, \\ \tau_{xj} &= -\sin(v_j + \omega_j) \cos \Omega_j - \cos(v_j + \omega_j) \sin \Omega_j \cos i_j, \\ \tau_{yj} &= -\sin(v_j + \omega_j) \sin \Omega_j + \cos(v_j + \omega_j) \cos \Omega_j \cos i_j, \end{aligned} \quad (23)$$

$$\begin{aligned} \tau_{zj} &= \cos(v_j + \omega_j) \sin i_j, \\ n_{xj} &= \sin \Omega_j \sin i_j, \end{aligned}$$

$$n_{yj} = -\cos \Omega_j \sin i_j, \quad (24)$$

$$n_{zj} = \cos i_j, \quad j = 1, 2.$$

Используя (3), (22)–(24), получаем для первого тела

$$F_{r1} = \frac{\dot{m}_1}{m_1} V_{r1} - \frac{\dot{m}_0}{m_0} V_{r0} - \frac{\dot{\gamma}_1}{\gamma_1} r_1, \quad F_{\tau 1} = \frac{\dot{m}_1}{m_1} V_{\tau 1} - \frac{\dot{m}_0}{m_0} V_{\tau 0}, \quad F_{n1} = \frac{\dot{m}_1}{m_1} V_{n1} - \frac{\dot{m}_0}{m_0} V_{n0}, \quad (25)$$

где  $V_{r1}, V_{r0}, V_{\tau 1}, V_{\tau 0}, V_{n1}, V_{n0}$  – составляющие относительных скоростей частиц, покидающих тела  $P_1$  и  $P_0$  или осаждающихся на них, вдоль радиального, трансверсального и нормального направлений в орбитальной системе координат, связанной с телом  $P_1$ . Аналогично определяем составляющие относительных скоростей частиц, покидающих тело  $P_2$  или осаждающихся на нем, вдоль радиального, трансверсального и нормального направлений  $V_{r2}, V_{\tau 2}, V_{n2}$  в орбитальной системе координат, связанной с телом  $P_2$ . В результате радиальная, трансверсальная и нормальная составляющие силы  $\vec{F}_2$  принимают вид

$$\begin{aligned} F_{r2} &= -Gm_1 \frac{r_2}{r_{12}^3} + Gm_1 \left( \frac{r_1}{r_{12}^3} - \frac{1}{r_1^2} \right) (\mathbf{e}_{r1} \cdot \mathbf{e}_{r2}) - \frac{\dot{\gamma}_2}{\gamma_2} r_2 + Q_{r2}, \\ F_{\tau 2} &= Gm_1 \left( \frac{r_1}{r_{12}^3} - \frac{1}{r_1^2} \right) (\mathbf{e}_{r1} \cdot \mathbf{e}_{\tau 2}) + Q_{\tau 2}, \\ F_{n2} &= Gm_1 \left( \frac{r_1}{r_{12}^3} - \frac{1}{r_1^2} \right) (\mathbf{e}_{r1} \cdot \mathbf{e}_{n2}) + Q_{n2}. \end{aligned} \quad (26)$$

Поскольку относительные скорости  $V_{r0}, V_{\tau 0}, V_{n0}$  заданы в орбитальной системе координат, связанной с телом  $P_1$ , реактивные силы, действующие на тело  $P_2$ , принимают вид

$$\begin{aligned} Q_{r2} &= \frac{\dot{m}_2}{m_2} V_{r2} - \frac{\dot{m}_0}{m_0} (V_{r0} (\mathbf{e}_{r1} \cdot \mathbf{e}_{r2}) + V_{\tau 0} (\mathbf{e}_{\tau 1} \cdot \mathbf{e}_{r2}) + V_{n0} (\mathbf{e}_{n1} \cdot \mathbf{e}_{r2})), \\ Q_{\tau 2} &= \frac{\dot{m}_2}{m_2} V_{\tau 2} - \frac{\dot{m}_0}{m_0} (V_{r0} (\mathbf{e}_{r1} \cdot \mathbf{e}_{\tau 2}) + V_{\tau 0} (\mathbf{e}_{\tau 1} \cdot \mathbf{e}_{\tau 2}) + V_{n0} (\mathbf{e}_{n1} \cdot \mathbf{e}_{\tau 2})), \\ Q_{n2} &= \frac{\dot{m}_2}{m_2} V_{n2} - \frac{\dot{m}_0}{m_0} (V_{r0} (\mathbf{e}_{r1} \cdot \mathbf{e}_{n2}) + V_{\tau 0} (\mathbf{e}_{\tau 1} \cdot \mathbf{e}_{n2}) + V_{n0} (\mathbf{e}_{n1} \cdot \mathbf{e}_{n2})). \end{aligned} \quad (27)$$

При заданных реактивных силах и законах изменения масс тел уравнения (16)–(27) полностью определяют возмущенное движение тел  $P_1, P_2$ .

#### 4. ЭВОЛЮЦИОННЫЕ УРАВНЕНИЯ ДЛЯ ОРБИТАЛЬНЫХ ПАРАМЕТРОВ

Уравнения возмущенного движения (16)–(21) показывают, что при условиях  $\mathbf{F}_1 = 0, \mathbf{F}_2 = 0$  орбитальные элементы  $a_j, e_j, i_j, \Omega_j, \omega_j$  остаются постоянными, а средние аномалии  $M_j$  являются возрастающими функциями времени. При наличии возмущений орбитальные параметры изменяются с течением времени, но получить точные решения дифференциальных уравнений (16)–(21), определяющих эти изменения, не представляется возможным. Поскольку во многих практически важных случаях эксцентриситеты  $e_j$  и наклонения  $i_j$  орбит являются малыми величинами ( $e_j \ll 1, i_j \ll 1$ ), уравнения (16)–(21) можно упростить, заменяя их правые части соответствующими разложениями в ряды по малым параметрам. Получаемые при этом уравнения описывают динамику исследуемых систем на длительных временных интервалах и позволяют моделировать движение тел с требуемой точностью (см. [16]–[19]). Далее рассмотрим случай малых эксцентриситетов  $e_j \ll 1$  и наклонений  $i_j \ll 1$  и разложим правые части уравнений (16)–(21) в ряды по этим параметрам с точностью до второго порядка. Хотя такие разложения требуют выполнения довольно громоздких символьных вычислений, использование систем компьютерной алгебры позволяет весьма эффективно решать эту задачу.

При  $e_j \ll 1$  уравнение Кеплера (8) позволяет получить выражение для эксцентрической аномалии в виде сходящегося степенного ряда по  $e_j$  (см. [17])

$$E_j = M_j + e_j \sin M_j + \frac{1}{2} e_j^2 \sin(2M_j) + \dots$$

Используя это разложение и встроенную функцию Series (см. [12]), получаем

$$\begin{aligned} \cos E_j &= \cos M_j + \frac{1}{2} e_j (\cos(2M_j) - 1) + \frac{3}{8} e_j^2 (\cos(3M_j) - \cos M_j) + \dots, \\ \sin E_j &= \sin M_j + \frac{1}{2} e_j \sin(2M_j) + \frac{1}{8} e_j^2 (3 \sin(3M_j) - \sin M_j) + \dots. \end{aligned} \quad (28)$$

Разлагая в ряды по малым параметрам решения (5) с учетом (28), получаем

$$\begin{aligned} x_j &= \gamma_j a_j \left( \cos(M_j + \omega_j + \Omega_j) + \frac{1}{2} e_j (\cos(2M_j + \omega_j + \Omega_j) - 3 \cos(\omega_j + \Omega_j)) + \right. \\ &\quad \left. + \frac{1}{8} e_j^2 (\cos(M_j - \omega_j - \Omega_j) - 4 \cos(M_j + \omega_j + \Omega_j) + 3 \cos(3M_j + \omega_j + \Omega_j)) + \right. \\ &\quad \left. + s_j^2 (\cos(M_j + \omega_j - \Omega_j) - \cos(M_j + \omega_j + \Omega_j)) \right), \\ y_j &= \gamma_j a_j \left( \sin(M_j + \omega_j + \Omega_j) + \frac{1}{2} e_j (\sin(2M_j + \omega_j + \Omega_j) - 3 \sin(\omega_j + \Omega_j)) - \right. \\ &\quad \left. - \frac{1}{8} e_j^2 (\sin(M_j - \omega_j - \Omega_j) + 4 \sin(M_j + \omega_j + \Omega_j) - 3 \sin(3M_j + \omega_j + \Omega_j)) - \right. \\ &\quad \left. - s_j^2 (\sin(M_j + \omega_j - \Omega_j) + \sin(M_j + \omega_j + \Omega_j)) \right), \\ z_j &= \gamma_j \rho_j s_j (2 \sin(M_j + \omega_j) + e_j (\sin(2M_j + \omega_j) - 3 \sin \omega_j)), \quad j = 1, 2, \end{aligned} \quad (29)$$

где введенный для удобства вычислений параметр  $s_j = \sin \frac{i_j}{2}$  является малым, так как наклонения  $i_j \ll 1$ . Используя (29), находим

$$r_j = \gamma_j \rho_j (1 - e_j \cos M_j + e_j^2 \sin^2 M_j), \quad j = 1, 2. \quad (30)$$

Далее вычисляем разложение выражения для  $r_{12}^2 = r_1^2 + r_2^2 - 2\mathbf{r}_1\mathbf{r}_2$ , которое определяет силу  $\mathbf{F}_2$ :

$$\begin{aligned} r_{12}^2 &= a_1^2 \gamma_1^2 + a_2^2 \gamma_2^2 - 2a_1 a_2 \gamma_1 \gamma_2 \cos(\lambda_1 - \lambda_2) + e_1 (-2a_1^2 \gamma_1^2 \cos(\lambda_1 - \omega_1 - \Omega_1) - \\ &\quad - a_1 a_2 \gamma_1 \gamma_2 \cos(2\lambda_1 - \lambda_2 - \omega_1 - \Omega_1) + 3a_1 a_2 \gamma_1 \gamma_2 \cos(\lambda_2 - \omega_1 - \Omega_1)) + \\ &\quad + e_2 (3a_1 a_2 \gamma_1 \gamma_2 \cos(\lambda_1 - \omega_2 - \Omega_2) - a_1 a_2 \gamma_1 \gamma_2 \cos(\lambda_1 - 2\lambda_2 + \omega_2 + \Omega_2) - \\ &\quad - 2a_2^2 \gamma_2^2 \cos(\lambda_2 - \omega_2 - \Omega_2)) + \frac{e_1^2}{4} (2a_1^2 \gamma_1^2 (3 - \cos(2\lambda_1 - 2\omega_1 - 2\Omega_1)) + \\ &\quad + a_1 a_2 \gamma_1 \gamma_2 (4 \cos(\lambda_1 - \lambda_2) - 3 \cos(3\lambda_1 - \lambda_2 - 2\omega_1 - 2\Omega_1) - \\ &\quad - \cos(\lambda_1 + \lambda_2 - 2\omega_1 - 2\Omega_1))) + \frac{e_2^2}{4} (2a_2^2 \gamma_2^2 (3 - \cos(2\lambda_2 - 2\omega_2 - 2\Omega_2)) + \\ &\quad + a_1 a_2 \gamma_1 \gamma_2 (4 \cos(\lambda_1 - \lambda_2) - 3 \cos(\lambda_1 - 3\lambda_2 + 2\omega_2 + 2\Omega_2) - \\ &\quad - \cos(\lambda_1 + \lambda_2 - 2\omega_2 - 2\Omega_2))) - \frac{e_1 e_2}{2} a_1 a_2 \gamma_1 \gamma_2 (9 \cos(\omega_1 - \omega_2 + \Omega_1 - \Omega_2) + \\ &\quad + \cos(2\lambda_1 - 2\lambda_2 - \omega_1 - \Omega_1 + \omega_2 + \Omega_2) - 3 \cos(2\lambda_1 - \omega_1 - \Omega_1 - \omega_2 - \Omega_2) - \\ &\quad - 3 \cos(2\lambda_2 - \omega_1 - \Omega_1 - \omega_2 - \Omega_2)) + a_1 a_2 \gamma_1 \gamma_2 (s_1^2 (2 \cos(\lambda_1 - \lambda_2) - \end{aligned} \quad (31)$$

$$- \cos(\lambda_1 + \lambda_2 - 2\Omega_1)) + s_2^2 (2 \cos(\lambda_1 - \lambda_2) - \cos(\lambda_1 + \lambda_2 - 2\Omega_2)) + \\ + 4s_1s_2 (4 \cos(\lambda_1 + \lambda_2 - \Omega_1 - \Omega_2) - \cos(\lambda_1 - \lambda_2 - \Omega_1 + \Omega_2))) ,$$

где для удобства дальнейших вычислений вместо средней аномалии  $M_j$  введена средняя долгота  $\lambda_j = M_j + \omega_j + \Omega_j$  (см. [16–19]). Используя разложения (29), (30) и выражения для направляющих косинусов (22)–(24), находим скалярные произведения единичных векторов, входящих в выражения для составляющих силы  $\vec{F}_2$  вдоль радиального, трансверсального и нормального направлений (см. (26), (27)). Например,

$$(\mathbf{e}_{r1} \cdot \mathbf{e}_{r2}) = \cos(\lambda_1 - \lambda_2) - 2e_1 \sin(\lambda_1 - \lambda_2) \sin(\lambda_1 - \omega_1 - \Omega_1) + 2e_2 \sin(\lambda_1 - \lambda_2) \sin(\lambda_2 - \omega_2 - \Omega_2) + \\ + \frac{e_1^2}{4} \sin(\lambda_1 - \omega_1 - \Omega_1) (\sin(\lambda_2 - \omega_1 - \Omega_1) - 9 \sin(2\lambda_1 - \lambda_2 - \omega_1 - \Omega_1)) + \\ + \frac{e_2^2}{4} \sin(\lambda_2 - \omega_2 - \Omega_2) (\sin(\lambda_1 - \omega_2 - \Omega_2) + 9 \sin(\lambda_1 - 2\lambda_2 + \omega_2 + \Omega_2)) + \quad (32)$$

$$+ 4e_1e_2 \cos(\lambda_1 - \lambda_2) \sin(\lambda_1 - \omega_1 - \Omega_1) \sin(\lambda_2 - \omega_2 - \Omega_2) + 2s_1^2 \sin(\lambda_1 - \Omega_1) \sin(\lambda_2 - \Omega_1) + \\ + 2s_2^2 \sin(\lambda_1 - \Omega_2) \sin(\lambda_2 - \Omega_2) + 4s_1s_2 \sin(\lambda_1 - \Omega_1) \sin(\lambda_2 - \Omega_2),$$

$$(\vec{e}_{\tau 1} \cdot \vec{e}_{r2}) = -\sin(\lambda_1 - \lambda_2) - 2e_1 \cos(\lambda_1 - \lambda_2) \sin(\lambda_1 - \omega_1 - \Omega_1) + 2e_2 \cos(\lambda_1 - \lambda_2) \sin(\lambda_2 - \omega_2 - \Omega_2) - \\ - \frac{e_1^2}{4} \sin(\lambda_1 - \omega_1 - \Omega_1) (\cos(\lambda_2 - \omega_1 - \Omega_1) + 9 \cos(2\lambda_1 - \lambda_2 - \omega_1 - \Omega_1)) + \\ + \frac{e_2^2}{4} \sin(\lambda_2 - \omega_2 - \Omega_2) (\cos(\lambda_1 - \omega_2 - \Omega_2) + 9 \cos(\lambda_1 - 2\lambda_2 + \omega_2 + \Omega_2)) - \quad (33)$$

$$- 4e_1e_2 \sin(\lambda_1 - \lambda_2) \sin(\lambda_1 - \omega_1 - \Omega_1) \sin(\lambda_2 - \omega_2 - \Omega_2) - 2s_1^2 \cos(\lambda_1 - \Omega_1) \sin(\lambda_2 - \Omega_1) - \\ - 2s_2^2 \cos(\lambda_1 - \Omega_2) \sin(\lambda_2 - \Omega_2) + 4s_1s_2 \cos(\lambda_1 - \Omega_1) \sin(\lambda_2 - \Omega_2),$$

$$(\mathbf{e}_{n1} \cdot \mathbf{e}_{r2}) = -2s_1 \sin(\lambda_2 - \Omega_1) + 2s_2 \sin(\lambda_2 - \Omega_2) - \\ - 4e_2 (s_1 \cos(\lambda_2 - \Omega_1) - s_2 \cos(\lambda_2 - \Omega_2)) \sin(\lambda_2 - \omega_2 - \Omega_2). \quad (34)$$

Подобным образом вычисляются и остальные скалярные произведения  $(\mathbf{e}_{r1} \cdot \mathbf{e}_{\tau 2})$ ,  $(\mathbf{e}_{\tau 1} \cdot \mathbf{e}_{\tau 2})$ ,  $(\mathbf{e}_{n1} \cdot \mathbf{e}_{\tau 2})$ ,  $(\mathbf{e}_{r1} \cdot \mathbf{e}_{n2})$ ,  $(\mathbf{e}_{\tau 1} \cdot \mathbf{e}_{n2})$ ,  $(\mathbf{e}_{n1} \cdot \mathbf{e}_{n2})$ . В результате правые части дифференциальных уравнений (16)–(21) можно представить в виде разложений по степеням малых параметров  $e_1$ ,  $e_2$ ,  $s_1$ ,  $s_2$ . Коэффициенты этих разложений являются периодическими функциями средних долгот  $\lambda_1$ ,  $\lambda_2$  и представляют собой рациональные выражения, в числителях которых содержатся тригонометрические функции  $\cos(k\lambda_j)$ ,  $\sin(k\lambda_j)$ ,  $\cos(k\lambda \pm n\lambda_2)$ ,  $\sin(k\lambda \pm n\lambda_2)$  ( $k, n = 1, 2, \dots$ ). Поскольку получаемые выражения весьма громоздки, мы их не приводим. Отметим только, что при разложении выражения  $r_{12}^{-3}$  с точностью до второго порядка по малым параметрам появляются рациональные выражения вида  $1/\rho_0$ ,  $1/\rho_0^3$ ,  $1/\rho_0^5$ , где

$$\rho_0 = \left( a_1^2 \gamma_1^2 - 2a_1 a_2 \gamma_1 \gamma_2 \cos(\lambda_1 - \lambda_2) + a_2^2 \gamma_2^2 \right)^{1/2}. \quad (35)$$

Так как  $\rho_0$  является периодической функцией переменных  $\lambda_1$ ,  $\lambda_2$ , выражения  $1/\rho_0$ ,  $1/\rho_0^3$ ,  $1/\rho_0^5$  следует заменить их разложениями в ряды Фурье вида

$$\frac{1}{\rho_0} = \frac{1}{2} \sum_{k=-\infty}^{+\infty} A_k \cos(k(\lambda_1 - \lambda_2)), \\ \frac{a_1 a_2 \gamma_1 \gamma_2}{\rho_0^3} = \frac{1}{2} \sum_{k=-\infty}^{+\infty} B_k \cos(k(\lambda_1 - \lambda_2)), \quad (36)$$

$$\frac{a_1^2 a_2^2 \gamma_1^2 \gamma_2^2}{\rho_0^5} = \frac{1}{2} \sum_{k=-\infty}^{+\infty} C_k \cos(k(\lambda_1 - \lambda_2)),$$

где  $A_k$ ,  $B_k$ ,  $C_k$  — коэффициенты Лапласа, удовлетворяющие следующим рекуррентным соотношениям (см. [20]):

$$\begin{aligned} A_k &= \frac{2(k-1)}{2k-1} \left( \alpha + \frac{1}{\alpha} \right) A_{k-1} - \frac{2k-3}{2k-1} A_{k-2}, \quad k \geq 2, \\ B_k &= \frac{(2k+1)\alpha(1+\alpha^2)}{(1-\alpha^2)^2} A_k - 2\alpha^2 \frac{2k+1}{(1-\alpha^2)^2} A_{k+1}, \quad k \geq 0, \\ C_k &= \frac{(2k+3)\alpha(1+\alpha^2)}{(1-\alpha^2)^2} B_k - 2\alpha^2 \frac{2k-1}{3(1-\alpha^2)^2} B_{k+1}, \quad k \geq 0. \end{aligned}$$

Все коэффициенты Лапласа можно вычислить, используя приведенные рекуррентные соотношения и следующие выражения для  $A_0$  и  $A_1$ :

$$\begin{aligned} A_0 &= \frac{2}{\pi a_2 \gamma_2} \int_0^\pi \frac{d\lambda}{(1 + \alpha^2 - 2\alpha \cos \lambda)^{1/2}} = \frac{4}{\pi a_2 \gamma_2 (1 + \alpha)} K \left( \frac{4\alpha}{(1 + \alpha)^2} \right), \\ A_1 &= \frac{2}{\pi a_2 \gamma_2} \int_0^\pi \frac{\cos \lambda d\lambda}{(1 + \alpha^2 - 2\alpha \cos \lambda)^{1/2}} = \frac{2}{\pi a_2 \gamma_2 \alpha (1 + \alpha)} \left( (1 + \alpha^2) K \left( \frac{4\alpha}{(1 + \alpha)^2} \right) - (1 + \alpha)^2 E \left( \frac{4\alpha}{(1 + \alpha)^2} \right) \right). \end{aligned}$$

Здесь функции  $K \left( \frac{4\alpha}{(1 + \alpha)^2} \right)$  и  $E \left( \frac{4\alpha}{(1 + \alpha)^2} \right)$  обозначают полный эллиптический интеграл первого и второго рода соответственно, а параметр  $\alpha = \frac{a_1 \gamma_1}{a_2 \gamma_2} < 1$  в случае, если тело  $P_2$  является внешней планетой. Если тело  $P_2$  является внутренней планетой, то  $a_1 \gamma_1$  и  $a_2 \gamma_2$  в выражениях для  $A_0$  и  $A_1$  нужно поменять местами и параметр  $\alpha$  определить как  $\alpha = \frac{a_2 \gamma_2}{a_1 \gamma_1} < 1$ .

Выполняя описанные выше вычисления, получаем для правых частей уравнений (16)–(21) выражения в виде рядов по малым параметрам  $e_1$ ,  $e_2$ ,  $i_1$ ,  $i_2$ , коэффициенты которых являются периодическими функциями средних долгот  $\lambda_1$ ,  $\lambda_2$ . Наличие таких функций приводит к колебаниям орбитальных параметров и значительно усложняет получение численных решений уравнений (16)–(21) на больших временных интервалах. Поскольку нас интересует поведение орбитальных параметров именно на больших промежутках времени, при отсутствии соизмеримостей средних движений тел правые части уравнений (16)–(21) можно заменить их осредненными выражениями относительно средних долгот  $\lambda_1$ ,  $\lambda_2$ , что приводит к уравнениям, определяющим вековые возмущения орбитальных параметров (см. [16], [19]–[21]).

Напомним, что осреднение функции  $W(\lambda_1, \lambda_2)$  по переменным  $\lambda_1$ ,  $\lambda_2$  и переход к вековым возмущениям сводится к вычислению интеграла

$$W^{(\text{sec})} = \frac{1}{(2\pi)^2} \int_0^{2\pi} W(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2. \quad (37)$$

Поскольку мы рассматриваем ограниченную задачу трех тел, правые части уравнений (16)–(21), записанные для тела  $P_1$ , зависят только от средней долготы  $\lambda_1$ , поэтому интегрирование по  $\lambda_2$  в выражении (37) является тривиальным. При подстановке вместо функции  $W(\lambda_1, \lambda_2)$  правых ча-

стей уравнений (16)–(21), в которых выполнены разложения по малым параметрам с учетом выражений для сил (25)–(27) и разложений (28)–(36), в случае тела  $P_1$  получаем уравнения

$$\frac{da_1}{dt} = \frac{a_1^{3/2}\gamma_1(2-e_1^2)}{\sqrt{\kappa_1}} \left( \frac{\dot{m}_1}{m_1} V_{\tau 1} - \frac{\dot{m}_0}{m_0} V_{\tau 0} \right), \quad (38)$$

$$\frac{de_1}{dt} = -\frac{3\sqrt{a_1}\gamma_1 e_1}{2\sqrt{\kappa_1}} \left( \frac{\dot{m}_1}{m_1} V_{\tau 1} - \frac{\dot{m}_0}{m_0} V_{\tau 0} \right), \quad (39)$$

$$\frac{di_1}{dt} = -\frac{3\sqrt{a_1}\gamma_1 e_1}{2\sqrt{\kappa_1}} \cos \omega_1 \left( \frac{\dot{m}_1}{m_1} V_{n1} - \frac{\dot{m}_0}{m_0} V_{n0} \right), \quad (40)$$

$$\frac{d\Omega_1}{dt} = -\frac{3\sqrt{a_1}\gamma_1 e_1}{2i_1\sqrt{\kappa_1}} \sin \omega_1 \left( \frac{\dot{m}_1}{m_1} V_{n1} - \frac{\dot{m}_0}{m_0} V_{n0} \right), \quad (41)$$

$$\frac{d\omega_1}{dt} = \frac{3\sqrt{a_1}\gamma_1 e_1}{2i_1\sqrt{\kappa_1}} \sin \omega_1 \left( \frac{\dot{m}_1}{m_1} V_{n1} - \frac{\dot{m}_0}{m_0} V_{n0} \right) + \frac{\sqrt{a_1}\gamma_1}{\sqrt{\kappa_1}} \left( \frac{\dot{m}_1}{m_1} V_{r1} - \frac{\dot{m}_0}{m_0} V_{r0} \right) - \frac{3a_1^{3/2}\gamma_1 \ddot{\gamma}_1}{2\sqrt{\kappa_1}}. \quad (42)$$

Следует отметить, что интегрирование уравнений (16)–(21) по средним долготам приводит к утрате информации о положении тел на орбитах и позволяет предсказать только медленные изменения во времени орбитальных параметров  $a_j$ ,  $e_j$ ,  $i_j$ ,  $\Omega_j$ ,  $\omega_j$ . Поэтому усредненное уравнение (21) мы не приводим.

Получение подобных уравнений для орбитальных элементов тела  $P_2$  требует выполнения существенно более громоздких символьных вычислений. Однако использование системы компьютерной алгебры Wolfram Mathematica позволяет успешно выполнить все вычисления. В результате получаем уравнения для вековых возмущений орбитальных элементов тела  $P_2$ :

$$\begin{aligned} \frac{da_2}{dt} = & -\frac{Gm_1\sqrt{a_2}}{8\sqrt{\kappa_2}}(4B_2 - 3C_1 + 3C_3)i_1i_2 \sin(\Omega_1 - \Omega_2) + \frac{a_2^{3/2}\gamma_2}{\sqrt{\kappa_2}} \frac{\dot{m}_2}{m_2} V_{\tau 2}(2 - e_2^2) + \frac{Gm_1\sqrt{a_2}}{32\sqrt{\kappa_2}} e_1e_2 \times \\ & \times \sin(\omega_1 - \omega_2 + \Omega_1 - \Omega_2) \left( 6\alpha(4C_0 - 6C_1 + 4C_2 - 10C_3 + 5D_0 - 5D_4) - 24B_0 + 60B_1 + 8B_2 - 60B_3 - \right. \\ & \left. - 42C_0 - 108C_1 + 288C_2 + 12C_3 - 54C_4 - 150D_1 + 165D_3 + \frac{6}{\alpha}(12C_0 - 10C_1 - 4C_2 - 6C_3 + 5D_0 - 5D_4) \right), \end{aligned} \quad (43)$$

$$\begin{aligned} \frac{de_2}{dt} = & \frac{3\sqrt{a_2}\gamma_2}{2\sqrt{\kappa_2}} \frac{\dot{m}_0}{m_0} (V_{\tau 0}e_1 \cos(\omega_1 - \omega_2 + \Omega_1 - \Omega_2) - V_{n0}i_2 \cos \omega_2 + V_{r0}e_1 \sin(\omega_1 - \omega_2 + \Omega_1 - \Omega_2) + \\ & + V_{n0}i_1 \cos(\omega_2 - \Omega_1 + \Omega_2)) - \frac{3\sqrt{a_2}\gamma_2}{2\sqrt{\kappa_2}} \frac{\dot{m}_2}{m_2} V_{\tau 2}e_2 + \frac{Gm_1e_1}{16\sqrt{a_2}\kappa_2} \sin(\omega_1 - \omega_2 + \Omega_1 - \Omega_2) \times \\ & \times \left( 6\alpha(C_0 + C_1 + C_2 - C_3) - 6B_0 - 10B_1 + 2B_2 - 6B_3 + 3C_0 - 27C_1 + 3C_3 - 3C_4 + \frac{6}{\alpha}(3C_0 - C_2) \right), \end{aligned} \quad (44)$$

$$\frac{di_2}{dt} = -\frac{Gm_1s_1}{4\sqrt{a_2}\kappa_2} \left( B_0 + B_2 - \frac{2a_2\gamma_2}{a_1^2\gamma_1^2} \right) \sin(\Omega_1 - \Omega_2) - \frac{3\sqrt{a_2}\gamma_2e_2}{2\sqrt{\kappa_2}} \cos \omega_2 \left( \frac{\dot{m}_2}{m_2} V_{n2} - \frac{\dot{m}_0}{m_0} V_{n0} \right), \quad (45)$$

$$\frac{d\Omega_2}{dt} = -\frac{Gm_1}{8\sqrt{a_2}\kappa_2} (B_0 + B_2) \left( 1 - \frac{s_1}{s_2} \cos(\Omega_1 - \Omega_2) \right), \quad (46)$$

$$\begin{aligned} \frac{d\omega_2}{dt} = & -\frac{Gm_1}{8\sqrt{a_2}\kappa_2} (B_0 + B_2) \frac{s_1}{s_2} \cos(\Omega_1 - \Omega_2) + \frac{Gm_1}{8\sqrt{a_2}\kappa_2} \left( B_0 + 4B_1 - 7B_2 + 3C_0 + \right. \\ & + 6C_1 + 3C_2 - 6C_3 - \frac{6}{\alpha}(B_0 + 2C_1) + \frac{6}{\alpha^2}C_0 \left. \right) + \frac{Gm_1e_1}{16\sqrt{a_2}\kappa_2} \left( -6\alpha(C_0 + C_1 + C_2 - C_3) + \right. \\ & \left. + 6B_0 + 10B_1 - 2B_2 + 6B_3 - 3C_0 + 27C_1 - 3C_3 + 3C_4 - \frac{6}{\alpha}(3C_0 - C_2) \right). \end{aligned} \quad (47)$$

## ЗАКЛЮЧЕНИЕ

В настоящей работе исследуется нестационарная ограниченная задача трех тел переменной массы, взаимодействующих между собой в соответствии с законом всемирного тяготения Ньютона с учетом реактивных сил, возникающих при неизотропном изменении масс тел. Исходные уравнения движения тел в относительной системе координат записаны на основе второго закона Ньютона, что позволяет определить реактивные силы на основе уравнения Мещерского. Используя точные решения нестационарной задачи двух тел (см. [6]) и применяя метод вариации постоянных, мы получили дифференциальные уравнения возмущенного движения в оскулирующих элементах аperiодического движения по квазиконическим сечениям. В случае малых эксцентриситетов и наклонов орбит получены уравнения, определяющие вековые возмущения орбитальных элементов. Следует отметить, что полученные уравнения описывают общий случай, когда массы всех трех тел изменяются анизотропно и возникают реактивные силы. Из этих уравнений можно легко получить уравнения в различных частных случаях, представляющих практический интерес. Например, массы двух массивных тел остаются постоянными, а масса третьего тела изменяется анизотропно и на него действуют реактивные силы. Этот частный случай представляет интерес в динамике искусственных небесных тел. Далее планируется численными методами исследовать полученные эволюционные уравнения при различных законах изменения масс тел.

## СПИСОК ЛИТЕРАТУРЫ

1. *Omarov T.B.* (Ed.) Non-Stationary Dynamical Problems in Astronomy. N.Y.: Nova Sci. Publ., 2002.
2. *Bekov A.A., Omarov T.B.* The theory of orbits in non-stationary stellar systems // *Astron. Astrophys. Transact.* 2013. V. 22. № 2. P. 145–153.
3. *Черепащук А.М.* Тесные двойные звезды. Ч. II. М.: Физматлит, 2013. 572 с.
4. *Eggleton P.* Evolutionary processes in binary and multiple stars. Cambridge Univ. Press, 2006. 332 p.
5. *Luk'yanov L.G.* Dynamical evolution of stellar orbits in close binary systems with conservative mass transfer // *Astron. Rep.* 2008. V. 52. № 8. P. 680–693.
6. *Минглибаев М.Дж.* Динамика гравитирующих тел с переменными массами и размерами. LAMBERT Acad. Publ., 2012. 229 с.
7. *Прокопеня А.Н., Минглибаев М.Дж., Мамерова Г.М.* Символьные вычисления в исследованиях проблемы трех тел с переменными массами // *Программирование.* 2014. Т. 40. № 2. С. 51–59.
8. *Minglibayev M.Zh., Maymerova G.M.* Evolution of the orbital-plane orientations in the two-protoplanet three-body problem with variable masses // *Astron. Rep.* 2014. V. 58. № 9. P. 667–677.
9. *Minglibayev M.Zh., Prokopenya A.N., Maymerova G.M., Imanova Zh.U.* Three-body problem with variable masses that change anisotropically at different rates // *Math. Comp. Sci.* 2017. V. 11. № 3–4. P. 383–391.
10. *Прокопеня А.Н., Минглибаев М.Дж., Шомшекова С.А.* Применение компьютерной алгебры в исследованиях двухпланетной задачи трех тел с переменными массами // *Программирование.* 2019. Т. 45. № 2. С. 58–65.
11. *Minglibayev M., Prokopenya A., Shomsheкова S.* Computing perturbations in the two-planetary three-body problem with masses varying non-isotropically at different rates // *Math. Comp. Sci.* 2020. V. 14. № 2. P. 241–251.
12. *Wolfram S.* An Elementary Introduction to the Wolfram Language. Champaign, IL: Wolfram Media, 2015. 324 p.
13. *Прокопеня А.Н.* Решение физических задач с использованием системы Mathematica. Брест: БГТУ, 2005. 260 с.
14. *Minglibayev M.Zh., Omarov Ch.T., Ibraimova A.T.* New forms of the perturbed motion equation // *Rep. Nation. Acad. Sci. Republ. Kazakhstan.* 2020. V. 2(330). P. 5–13.
15. *Мещерский И.В.* Работы по механике тел переменной массы. М.: Гос. изд-во тех.-теор. лит-ры, 1952. 281 с.
16. *Дубошин Г.Н.* Небесная механика. Основные задачи и методы. М.: Наука, 1975. 799 с.
17. *Рой А.Э.* Движение по орбитам. М.: Мир, 1981. 544 с.
18. *Себехей В.* Теория орбит: ограниченная задача трех тел. М.: Наука, 1982. 656 с.
19. *Brouwer D., Clemence G.M.* Methods of Celestial Mechanics. N.Y.: Acad. Press, 1961. 601 p.
20. Шарлье К. Небесная механика. М.: Наука, 1966. 628 с.
21. *Murray C.D., Dermott S.F.* Solar system dynamics. Cambridge University Press, New York, 1999. 592 p.

УДК 519.72

## ОБУЧЕНИЕ ПОРТ-ГАМИЛЬТОНОВЫМ СИСТЕМАМ: АЛГОРИТМЫ<sup>1)</sup>

© 2023 г. Д. Лозиенко<sup>1,\*</sup>, В. Сальников<sup>1,\*\*</sup>, А. Фалез<sup>1,\*\*\*</sup><sup>1</sup> ЦНРС & Университет Ла-Рошели, Пр. Мишеля Крено, 17042 Ла Рошель, Франция

\*e-mail: daria.lozijenko1@univ-lr.fr

\*\*e-mail: vladimir.salnikov@univ-lr.fr

\*\*\*e-mail: antoine.falaize@univ-lr.fr

Поступила в редакцию 04.08.2022 г.  
Переработанный вариант 21.08.2022 г.  
Принята к публикации 10.09.2022 г.

Изучаются возможности определения структуры порт-гамильтоновых систем, соответствующих описывающим задачи механики системам обыкновенных дифференциальных уравнений по переменным без “определенной природы”. Предложенный алгоритм решает эту задачу в два этапа. Сначала с помощью методов машинного обучения восстанавливается структура связей системы, таким образом строится граф, характеризующийся выделенными подсистемами и взаимодействиями между ними. Затем этот граф дополняется построенными гамильтоновыми структурами каждой подсистемы, а также соответствующими портами. Описанный второй этап основывается на результатах из симплектической и пуассоновой геометрии, которые мы вкратце изложим. А конкретные решения строятся с использованием методов компьютерной алгебры и символьных вычислений. Представленный алгоритм позволяет расширить область применения порт-гамильтонова формализма до произвольных обыкновенных дифференциальных уравнений, потенциально вводя тем самым новое понятие нормальных форм обыкновенных дифференциальных уравнений. Библ. 18. Фиг. 1.

**Ключевые слова:** симплектические структуры, пуассонова геометрия, порт-гамильтоновы системы.

**DOI:** 10.31857/S0044466923010106, **EDN:** LMJWTD

### 1. ВВЕДЕНИЕ. МОТИВАЦИЯ

Настоящая статья — часть большого проекта, связанного с направлением, которое мы называем “геометризация механики”. Сам проект состоит из нескольких этапов, связанных одной целью: изучить геометрический формализм, подходящий для описания внутренней структуры дифференциальных уравнений, моделирующих задачи механики, а также для вывода этих уравнений, исходя из физических свойств системы, используя геометрические структуры (см. обзор в [1], а также работы, на которые этот обзор ссылается).

Наиболее подробно изученный пример такого подхода касается естественных механических систем, описанных в канонических переменных (координаты и импульсы), в таком случае удобным формализмом будут гамильтоновы системы, определенные на симплектических многообразиях. Это описание, действительно, классическое и восходит к Лагранжу, Лежандру, Пуанкаре, хотя в их работах формализм не был написан современным языком Сурьё и Арнольда. Важно отметить, что эта логика работает “в обоих направлениях”. С одной стороны, для конкретной симплектической структуры, описывающей свойства фазового пространства системы, зная гамильтониан, можно явно выписать уравнения движения. Это особенно важно, так как сохранение симплектической структуры потоком гарантирует сохранение энергии; данный факт дает начало целой “индустрии” симплектических интеграторов (см. [2], [3]) — численных методов, сохраняющих геометрическую структуру. С другой стороны, зная уравнения движения, можно, естественно, при выполнении некоторых условий восстановить симплектическую структуру и гамильтониан. Мы обсудим подробнее это “обратное направление”.

<sup>1)</sup> Работа поддержана проектом CNRS 80Prime “GraNum” и частично проектом PHC Procope “GraNum 2.0”.

Понятно, что класс систем, допускающих симплектическое, гамильтоново описание, очень ограничен. И даже для консервативных систем некоторые свойства (например, интегрируемость) удобнее формулировать в терминах пуассоновой геометрии. Для нее гамильтонов формализм также применим, однако вопрос построения интеграторов гораздо более деликатен. В данной статье мы не останавливаемся на численных методах, но мотивированному читателю советуем ознакомиться с новой точкой зрения на этот вопрос (см. [4]).

Допустим теперь, что в системе присутствует диссипация, или несколько систем взаимодействуют; понятно, что для описания требуется другой, более общий формализм. Один из возможных подходов — использование так называемых порт-гамильтоновых систем (ПГС). Основная идея метода заключается в “сборке” большой системы из маленьких подсистем, каждая из которых гамильтонова (а значит, консервативна), эти подсистемы затем соединяются “портами”. Порты могут либо обеспечивать взаимодействие между подсистемами, либо описывать внешние факторы, такие как диссипация или управление. Эта идея очень естественна и далеко не нова: самый ранний источник (известный нам) — статья [5], где описан очень инженерный подход к задаче. Впоследствии такие системы были в некотором смысле классифицированы, т.е. были идентифицированы различные по своей природе порты (см. [6]). Также была изучена соответствующая геометрия, а именно, структура (почти) Дирака (см. [7]). Однако это описание в терминах структур Дирака не всегда позволяет построить интеграторы, сохраняющие порт-Гамильтонову структуру: некоторые примеры были рассмотрены “вручную”, а также (достаточные) условия были получены для путей Дирака (см. [8]), общая теория до сих пор не построена. Но даже без геометрических интеграторов структура ПГС позволяет оптимизировать моделирование сложных систем с использованием “разумных” параллельных или распределенных вычислений (см. программное обеспечение, описанное в [9]).

В предыдущем абзаце мы использовали слово “сборка” неслучайно: оно подчеркивает, что в контексте ПГС есть “простое” направление: когда физическая система естественным образом явно разделена на подсистемы, между которыми известны законы взаимодействия, соответствующая ПГС выписывается явно (см. [9]). Сложности же возникают при решении обратной задачи: зная систему дифференциальных уравнений, восстановить структуру соответствующей ПГС — это и будет основной темой исследования в настоящей статье. В [10] сформулирован ряд открытых вопросов и задач, к решению которых разумно подходить с использованием методов компьютерной алгебры, в том числе и задача оптимального восстановления ПГС. Как мы объясним ниже, структура ПГС может быть задана так называемым размеченным графом — его топология восстанавливается методами машинного обучения, а “разметка” естественным образом определяется с помощью символьных вычислений и компьютерной алгебры.

В следующих разделах мы разобьем задачу восстановления структуры ПГС на более явные подзадачи и объясним идеи решения каждой из них. Вначале мы напомним классический гамильтонов формализм (разд. 2), затем введем конкретную терминологию для ПГС (разд. 3) и опишем желаемый результат решения задачи вместе с некоторыми методами его достижения (разд. 4). При написании данной статьи мы осознали, что при программной реализации каждого из шагов метода возникает некоторое количество нюансов. Но чтобы не перегружать статью техническими подробностями, мы решили вынести вопросы программирования и бенчмаркинга, а также сложные примеры в отдельный текст (см. [11]), сосредоточившись здесь на теоретических основах и математическом аппарате.

## 2. СИМПЛЕКТИЧЕСКАЯ/ПУАССОНОВА ГЕОМЕТРИЯ ДЛЯ ГАМИЛЬТОНОВЫХ СИСТЕМ

Начнем с краткого описания стандартного гамильтонова формализма. Мы позволим себе некоторые геометрические упрощения, когда они не влияют на алгоритм; тем не менее, соответствующие ссылки на подробные конструкции будут даны. Мы приводим это описание по двум причинам: с педагогической точки зрения, оно необходимо для дальнейшего обобщения до ПГС, а для конечного алгоритма оно будет одной из важных частей.

Самая простая гамильтонова система имеет вид

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad (1)$$

где  $\mathbf{q}$  — векторная переменная размера  $n$ , обозначающая (обобщенные) координаты системы, а  $\mathbf{p}$  — соответствующие импульсы. Легко проверить, что поток такой системы сохраняет поверхно-

сти уровня  $H(\mathbf{q}, \mathbf{p})$ , что физически обозначает сохранение энергии. Даже для этого простого случая “обратная” задача имеет смысл: для произвольной системы

$$\dot{\mathbf{q}} = \mathbf{f}(\mathbf{q}, \mathbf{p}), \quad \dot{\mathbf{p}} = \mathbf{g}(\mathbf{q}, \mathbf{p})$$

можно ли найти функцию  $H$ , определяющую эквивалентную гамильтонову систему? Ответ основан на условии совместности

$$\frac{\partial \mathbf{f}}{\partial \mathbf{q}} = -\frac{\partial \mathbf{g}}{\partial \mathbf{p}},$$

которое, очевидно, необходимо и (по крайней мере, локально) достаточно. Но чтобы сделать ответ конструктивным, требуется некоторая явная операция (в данном случае — интегрирование).

Хотя мы не упомянули симплектическую структуру в уравнении (1), она там неявно присутствует: систему можно переписать в более общем виде

$$\dot{\mathbf{x}} = J(\mathbf{x}) \frac{\partial H}{\partial \mathbf{x}}, \quad (2)$$

где  $J(\mathbf{x})$  — кососимметричная невырожденная матрица, удовлетворяющая некоторому дифференциальному условию. (Условие замкнутости соответствующей дифференциальной формы, см. упрощенное изложение терминологии в [10] или более подробное построение общей теории в [12].) Пример (1) соответствует  $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ , а

$$J = J_D = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \quad (3)$$

т.е. гамильтоновой системе на каноническом симплектическом пространстве, обозначаемом  $T^*Q$ . Теорема Дарбу (см., например, [13]) утверждает, что любая симплектическая структура может быть приведена к такому каноническому виду. В заданной точке это простой результат из линейной алгебры, но в ее окрестности конструкция более сложная: чтобы выбрать подходящие координаты, нужно интегрировать некоторый поток.

Систему (2) можно также рассматривать для любой кососимметричной матрицы  $J$ , не обязательно невырожденной. Если она снова удовлетворяет некоторому дифференциальному условию, то задает так называемую структуру Пуассона. Тожеству Якоби для соответствующей скобки Пуассона (см. [10]). Согласно теореме Вайнштена о расщеплении (см. [13]), пространство со структурой Пуассона представляет расслоение, слои которого симплектические. Иными словами, максимальное пространство, ограничение  $J$  на которое невырождено, допускает координаты Дарбу  $(\mathbf{q}, \mathbf{p})$ , а в трансверсальном направлении координаты  $(\mathbf{y})$  можно выбрать специальным образом, независимо от остальных:

$$J = J_W = \begin{pmatrix} J_D & 0 \\ 0 & \Phi(\mathbf{y}) \end{pmatrix},$$

где  $J_D$  — блочная матрица вида (3), а  $\Phi(\mathbf{y})$  — кососимметричная матрица, коэффициенты которой не зависят от координат  $(\mathbf{q}, \mathbf{p})$ , а также все  $\Phi_{ij}(\mathbf{0}) = 0$ .

Мы заметили, что в примере (1) поток системы сохраняет поверхности уровня энергии. Это свойство выполняется и для системы (2), оно следует, по сути, из кососимметричности матрицы  $J$ . Верно даже более сильное утверждение: сохраняются не только уровни энергии, но и вся геометрическая структура фазового пространства — симплектическая форма или соответственно пуассоново расслоение. Это более общее утверждение чуть сложнее “обратить”: на самом деле, мы хотим по заданному потоку системы подобрать гамильтониан, который порождает его для некоторой симплектической или пуассоновой структуры. Важно также заметить, что сохранение геометрической структуры некоторым потоком само по себе не гарантирует, что он является гамильтоновым. На самом деле, разница между множествами сохраняющих структуру и гамильтоновых векторных полей характеризуется некоторой когомологией фазового пространства. Не вдаваясь в геометрические подробности, дадим лишь корректные названия этим объектам. В симплектическом случае речь идет о классе первой когомологии де Рама, который характеризует топологию пространства и вполне вычислим. В случае структуры Пуассона класс тоже называется первой когомологией Пуассона, и он вычислим для большинства важных структур. Моти-

вированному читателю мы рекомендуем [13], где даны подробные определения и приведены примеры. Таким образом, разумно анализировать поток системы в два этапа: сначала убедиться, что он сохраняет структуру, а затем проверить, есть ли препятствия к его гамильтоновости. Подобный подход использовался в [4] для построения пуассоновых интеграторов.

### 3. ПОРТ-ГАМИЛЬТОНОВ ФОРМАЛИЗМ ДЛЯ ВЗАИМОДЕЙСТВУЮЩИХ И ОТКРЫТЫХ СИСТЕМ

В этом разделе мы приводим идеи порт-гамильтонова формализма и напоминаем или определяем соответствующую терминологию. Как мы уже упоминали, основная мысль состоит в изменении гамильтоновых уравнений добавлением негамильтоновых слагаемых. Например, системе (1) можно обобщить до

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} + \mathbf{F}, \quad (4)$$

где  $\mathbf{F}$  — некоторая сила, природу которой мы пока не уточняем. Подобным образом в неканоническом симплектическом или пуассоновом случае (2) система теперь может принимать вид

$$\dot{\mathbf{x}} = (J(\mathbf{x}) - R(\mathbf{x})) \frac{\partial H}{\partial \mathbf{x}} + w(\mathbf{x})\mathbf{u}, \quad (5)$$

где  $J(\mathbf{x})$  — по-прежнему кососимметричная матрица, а  $R(\mathbf{x})$ ,  $w(\mathbf{x})$  и  $\mathbf{u}$  — новые слагаемые. С качественной точки зрения  $R$  соответствует внутренним силам в системе, в то время как  $w$  и  $\mathbf{u}$  отвечают за взаимодействие с “внешним миром”. Мы различаем  $w$  и  $\mathbf{u}$ , чтобы подчеркнуть, что это взаимодействие может разделяться на части, зависящие от состояния системы и на полностью контролируемые средой. Мы записываем соответствующее слагаемое в виде произведения для простоты изложения с очевидным возможным обобщением. В [6], [7] эти слагаемые различаются более подробно, исходя из некоторой физической терминологии: они могут описывать накопление энергии, диссипацию, управление и т.д. Но для данной статьи уровень детализации (5) достаточен для объяснения предложенного подхода.

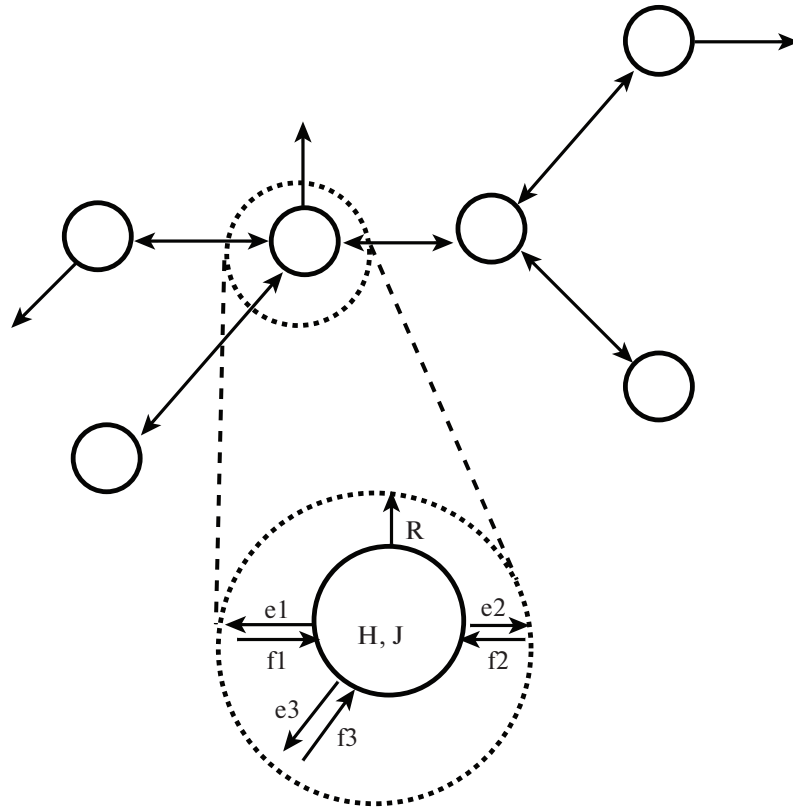
Определяющее свойство ПГС — возможность их “соединять”, что объясняет терминологию “портов”, как в электрических схемах. Иными словами, если рассмотреть несколько изначально независимых систем вида (5) и обеспечить их взаимодействие через соответствующие порты (слагаемые  $w\mathbf{u}$ ), то результат будет системой, большей, чем каждая из составляющих, но снова имеющей вид (5). Легко убедиться, что в таком случае глобальная структура этой системы будет довольно специфическая. Матрицы  $J$  и  $R$  будут блочного вида, характеризующая внутреннюю динамику подсистем, в то время как слагаемые  $w\mathbf{u}$  будут в точности обозначать взаимодействие между блоками.

ПГС удобно описывать графами. Обычно их строят из элементарных узлов, как на фиг. 1. Вершины графа соответствуют внутренним (гамильтоновым) составляющим системы, а ребра — портам (или наоборот в двойственном представлении). Некоторые из ребер идут к другим подсистемам, а некоторые заканчиваются виртуальными вершинами (т.е. идут как бы в пустоту), что соответствует диссипации или управлению. В терминах [7] портам соответствуют переменные потока (flux) и напряжения (effort), что позволяет снабдить ребра графа метками, отвечающими за негамильтонову часть системы (5). Эти переменные естественным образом двойственны друг другу, что делает связанные системы переопределенными. Таким образом, переменные будут связаны некоторыми алгебраическими соотношениями, которые определяют упомянутые ранее структуры Дирака, а физически они соответствуют балансу полной энергии системы.

Такое представление с помощью графов и используется явно, чтобы “собрать” ПГС из блоков, а затем выбрать оптимальную стратегию для численного моделирования. Этот подход был успешно реализован в программном коде с использованием распределенных вычислений (см. [9]) и применен для некоторых сложных физических систем (см. [14], [15]). Ниже мы снова рассмотрим обратную задачу по системе дифференциальных уравнений в наиболее общем виде:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad (6)$$

и построим ее порт-гамильтоново представление (5).



**Фиг. 1.** Представление части ПГС с помощью графа, подробное описание дано для одной из вершин. Граф с подробной информацией о каждой вершине мы будем называть размеченным.

#### 4. АЛГОРИТМЫ И ПРИМЕРЫ

Теперь, когда все составные части нашего подхода определены, мы можем собрать их вместе и объяснить основную мысль данной статьи, а именно, что конкретно мы подразумеваем под восстановлением порт-гамильтоновой структуры, и как это можно сделать.

Конкретный вопрос, на самом деле, несколько более деликатен, чем просто алгоритм перехода от (6) к (5). Заметим, что решение данной задачи совершенно не обязано быть единственным: мы уже упоминали, что, если соединить несколько ПГС, получится новая ПГС — понятно, что в решении обратной задачи нет никаких причин полагать, что такого слияния не происходит. Более того, для некоторых слагаемых возможен выбор: отнести их к гамильтонианам или к портам — это очень важная “степень свободы”, позволяющая значительно упростить некоторые процедуры. С учетом всех этих замечаний, нужно потребовать, чтобы процесс был в некотором смысле оптимальным, иначе можно построить совершенно неразумные решения. Рассмотрим, например, все правые части уравнений системы (6) как независимые порты — формально это решение задачи, но с большим количеством вырождений и полностью игнорирующее гамильтонову составляющую. Или, как в другом крайнем случае, если система большая, но изолированная от среды, скорее всего, ее можно полностью переписать в гамильтоновом виде, естественно, полностью теряя информацию о внутренних связях.

В свете вышесказанного, идея решения задачи становится совершенно естественной: сначала построить некоторый граф системы и оптимизировать его топологию, затем разметить его, т.е. добавить к результату остальные данные. И вопрос оптимизации становится более понятным: граф не должен получаться слишком большим, с одной стороны, и должен быть далек от полного, с другой.

##### 4.1. Определение структуры графа

В современной литературе описано множество алгоритмов анализа структуры графов — здесь мы не пытаемся ни изобрести новые, ни перечислить все существующие. Поэтому при всем бо-

гательстве выбора мы постарались повторить в программе процесс, которым человек бы решал задачу “вручную”.

Идея заключается в том, что если мы анализируем систему общего вида (6) и знаем, что она получена как ПГС, то, имея некий опыт работы с ПГС, мы можем постараться угадать ее структуру. Более конкретно, даже если уравнения, моделирующие физическую систему, записаны в странных переменных, опытный физик может распознать, какие части системы описывают взаимодействия, а какие внутреннюю динамику. Все это, естественно, разумно говорить для небольших систем, структуру которых можно изучать невооруженным глазом. При выборе алгоритма мы руководствовались именно таким подходом, но хотели сделать его масштабируемым.

Анализ разных алгоритмов показал, что задача отлично подходит для методов машинного обучения, что объясняет название статьи. С одной стороны, для обучения и тестов мы можем генерировать любое необходимое количество примеров ПГС с заданной топологией связей (см. [9]). С другой, мы четко определили набор параметров, которые нужно оптимизировать. Подобные методы известны под названием алгоритмы свертки (pooling operators), это довольно простые нейронные сети с небольшим количеством слоев.

В данной статье мы упоминаем машинное обучение в режиме “черного ящика” для полноты конечного алгоритма. Технические подробности, а также сравнение производительности методов будет подробно описано в [11]. Тем не менее уже здесь важно сделать пару замечаний для четкого понимания алгоритма. Исходный граф строится явно по правым частям системы дифференциальных уравнений (6): в первом приближении все переменные  $x_i$  объявляются независимыми вершинами, их пары связаны (направленным) ребром, если одна из них присутствует в правой части уравнения для другой. А на конечном этапе группы вершин с большим количеством попарных связей объединяются в узлы, они соответствуют гамильтоновым подсистемам, а связи между узлами объявляются портами. Некоторая “тонкая настройка” возможна на разных этапах этой оптимизации, а именно, слагаемые определенного вида в правых частях дифференциальных уравнений можно насильно учитывать внутри или снаружи узлов, мы вернемся к этой идее ниже.

#### 4.2. Разметка графа

Результат предыдущего этапа — граф, который по идее меньше, чем исходный, он описывает скопления переменных (узлы). А слагаемые в дифференциальных уравнениях разделяются на две группы: внутренние и внешние (отвечающие за взаимодействие узлов).

Разметка внутренней части — наиболее сложный этап с математической точки зрения. Изолированные (небольшие) системы дифференциальных уравнений нужно разделить на структурную часть (симплектическая/пуассонова геометрия для гамильтоновых систем, которую мы определили в разд. 2) и “порт-” часть (добавленная в разд. 3).

Даже если это, возможно, было понятно уже в разд. 2, подчеркнем еще раз, что обратная задача в общем случае не решается явно методами компьютерной алгебры и символьных вычислений. Основная трудность заключается в приведении структуры (симплектической или Пуассона) к нормальной форме, оно основано на теоремах существования, которые доказываются интегрированием определенных гамильтоновых потоков. Понятно, что возможность сделать это символьными методами — очень сильное условие, сравнимое с существованием явного аналитического решения системы. Для некоторых приложений можно решить задачу приближенно для коротких траекторий, например, с помощью преобразования Косса (аналог уравнения Гамильтона—Якоби, см. [4]). Но такой подход позволяет построить эквивалентную систему только локально, что недостаточно для ее дальнейшего использования в моделировании.

К счастью, нам и не нужно точно решать задачу, описанную в предыдущем разделе: достаточно использовать гибкость и неединственность порт-гамильтоновой формулировки. Симплектическая или пуассонова структура каждой внутренней подсистемы может быть восстановлена с точностью до некоторых остаточных слагаемых, они будут объявлены новыми портами, возможно, искусственными, но формально не запрещенными.

Более конкретно, процедура выполняется в несколько этапов. На первом составляется каталог простых симплектических и пуассоновых структур, т.е. матриц  $J$  подходящих размеров с заданным вырождением. Это могут быть симплектические структуры, записанные в координатах Дарбу, все структуры Пуассона постоянного ранга, линейные и квадратичные структуры Пуассона с некоторыми особенностями. Список можно расширить более нелинейными системами, но даже упомянутые структуры покрывают достаточный геометрический спектр. На втором эта-

не вычисляются соответствующие первые когомологии (см. разд. 2 и [13]). Для приведенных примеров они будут не очень большими, а иногда и вообще тривиальными, для некоторых структур они известны явно. На третьем этапе из правых частей уравнений выбираются слагаемые, сохраняющие построенные структуры. Те из них, которые не лежат в нетривиальном классе когомологии, интегрируются до гамильтониана. Этот последний шаг, в отличие от нормальных форм структуры, явный и может быть выполнен с помощью символьных вычислений. И даже если он оказывается технически сложным, класс гамильтонианов всегда может быть ограничен до полиномов или порожден другим конкретным набором элементарных функций. Более “академический” вариант последнего этапа — построить (один раз) список порождающих векторных полей, сохраняющих упомянутые выше структуры. Этот результат может иметь независимые геометрические приложения. А затем идентифицировать соответствующие слагаемые в правых частях уравнений. По выполнении этого этапа будет восстановлена гамильтонова часть уравнений и перечислены оставшиеся неидентифицированные слагаемые.

Эти дополнительные слагаемые тоже будут добавлены к портам, определенным ранее. И только на этом этапе их нужно будет классифицировать вручную (в терминах [7]), если классификация вообще необходима. В противном случае, результат окончательный: система разбита на меньшие подсистемы с определенной внутренней структурой, они соединены портами. Система в таком виде готова для последующего моделирования.

#### 4.3. Алгоритм

Сформулируем теперь все замечания из предыдущих подразделов в виде единого алгоритма.

Входные данные: система дифференциальных уравнений в виде  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ :

**Шаг 1.** По правым частям системы восстановить структуру графа ПГС.

В результате переменные  $x_i$  разбиты на группы (вершины графа), а слагаемые в правых частях разделены на внутренние (зависящие только от переменных “своей” группы) и внешние (остальные).

**Шаг 2.** Для каждой отдельной группы

- а) построить (или выбрать из заранее построенного каталога) все симплектические и пуассоновы структуры подходящих размеров;
- б) из внутренних переменных выбрать максимальную комбинацию слагаемых, сохраняющих одну из структур из предыдущего пункта;
- в) если когомология выбранной структуры не тривиальна, проверить выбранную комбинацию на принадлежность тривиальному классу;
- г) построить гамильтониан, соответствующий выбранной комбинации; если данный пункт приводит к чрезмерно трудоемким символьным вычислениям, повторить шаги 2а), б), заранее отбросив сильно нелинейные слагаемые.

В результате для каждой группы из внутренних переменных выделены компоненты Гамильтонова потока.

**Шаг 3.** Идентифицировать порты:

- а) слагаемые, не выбранные в шаге 2б) объявить внутренними портами, сопоставить им виртуальные вершины;
- б) внешним слагаемым из каждой группы (отвечающим за взаимодействия) сопоставить соответствующие ребра построенного графа.

В результате построен размеченный граф, как на фиг. 1.

Как уже упоминалось выше, каталог симплектических и пуассоновых структур может быть заранее дополнен порождающими векторных полей, сохраняющих эти структуры, с известными гамильтонианами из некоторого физически репрезентативного класса. В таком случае шаги 2б)—г) сводятся к идентификации подходящих слагаемых.

#### 4.4. Примеры

Приведем теперь несколько простых примеров работы различных пунктов приведенного выше алгоритма.

Для начала рассмотрим шаг 2 — идентификацию гамильтонианов, без особых подробностей приведем лишь исходные уравнения и результат.

**Осциллятор с трением.** Это, пожалуй, простейшая механическая система, где подходящая структура оказывается симплектической:

$$\ddot{x} + a\dot{x} + x = 0.$$

Она переписывается в форме системы (6):

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1 - ax_2,$$

которая, в свою очередь, имеет вид (4), т.е. (5) для

$$J = J_D = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad H = \frac{1}{2}(x_1^2 + x_2^2),$$

а оставшемуся слагаемому  $-ax_2$  соответствует диссипативный порт.

**Твердое тело с неподвижной точкой.** Данная система описывает (см. [12]) динамику твердого тела в терминах компонент кинетического момента  $\mathbf{M} \in \mathbb{R}^3$  под действием внешней силы  $\mathbf{N} \in \mathbb{R}^3$ :

$$\dot{M}_1 = a_1 M_2 M_3 + N_1, \quad \dot{M}_2 = a_2 M_3 M_1 + N_2, \quad \dot{M}_3 = a_3 M_1 M_2 + N_3.$$

Здесь сила  $\mathbf{N}$  задана явно, а коэффициенты  $a_i$  зависят от геометрии тела. Если  $I_1, I_2, I_3$  — моменты инерции тела, то

$$a_1 = \frac{I_2 - I_3}{I_2 I_3}, \quad a_2 = \frac{I_3 - I_1}{I_3 I_1}, \quad a_3 = \frac{I_1 - I_2}{I_1 I_2}.$$

Из соображений размерности ясно, что матрица  $J$  не может быть невырожденной, поэтому разумно искать решение среди структур Пуассона. Действительно, линейная структура

$$J = \begin{pmatrix} 0 & -M_3 & M_2 \\ M_3 & 0 & -M_1 \\ -M_2 & M_1 & 0 \end{pmatrix}$$

с квадратичным гамильтонианом

$$H = \frac{M_1^2}{2I_1} + \frac{M_2^2}{2I_2} + \frac{M_3^2}{2I_3}$$

восстанавливают консервативную часть системы, а сила  $\mathbf{N}$  определяет порт.

**Обобщение системы Лотки–Вольтерра.** Данная система обобщает классическую модель динамики популяций на случай произвольной размерности:

$$\dot{x}_i = \varepsilon_i x_i + \sum_{j=1}^n A_{ij} x_i x_j, \quad i = 1, \dots, n.$$

Для нечетного  $n = 2k + 1$  в отсутствие линейных правых частей ( $\varepsilon_i = 0$ ) она называется системой Богоявленского–Ито, если  $A_{ij} = 1$  для  $i + 1 \leq j \leq \min(i + k, n)$ , и  $A_{ij} = -1$  для  $\min(i + k, n) \leq j \leq n$ . Такие системы подробно изучены в [16]. В рамках описанного здесь подхода для каждого фиксированного значения  $n$  легко проверить, что система гамильтонова для квадратичной структуры Пуассона  $J_{ij}(x) = A_{ij} x_i x_j$  и линейного гамильтониана  $H = x_1 + \dots + x_n$ . Линейные слагаемые  $\varepsilon_i x_i$  могут быть добавлены как диссипативные порты.

**Динамика жидкости.** В качестве чуть более серьезного примера рассмотрим систему, выведенную в [17] в качестве упрощенной модели взаимодействия потока жидкости (или газа) с цилиндрическим твердым телом, ориентированным перпендикулярно потоку. В предложенной (линеаризованной по пространству) системе

$$a\ddot{y} + b\dot{y} + cy = dq,$$

$$\ddot{q} + k(q^2 - 1)\dot{q} + lq = m\ddot{y},$$

$u$  и  $q$  суть динамические переменные, отвечающие соответственно за положение тела и за интенсивность завихрений среды; остальные латинские буквы обозначают некоторые постоянные параметры системы. Перепишем систему в виде  $\dot{\mathbf{x}} = \mathbf{f}$ :

$$\begin{aligned}\dot{x}_1 &= x_2, \\ \dot{x}_2 &= -\frac{b}{a}x_2 - \frac{c}{a}x_1 + \frac{d}{a}x_3, \\ \dot{x}_3 &= x_4, \\ \dot{x}_4 &= -k(x_3^2 - 1)x_4 - lx_3 + m\dot{x}_2.\end{aligned}$$

Заметим, что система выше, строго говоря, еще не приведена к виду  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$  из-за зависимости правой части четвертого уравнения от  $\dot{x}_2$ . Когда мы изучали ее в [18],  $\dot{x}_2$  заменялась на соответствующие правые части в явном виде. Для текущего подхода это делать не обязательно: достаточно обозначить  $m\dot{x}_2$  новой переменной  $u(x_1, x_2, x_3)$ , не забывая, от чего она зависит для построения графа.

Первое приближение графа на шаге 1 алгоритма задается матрицей инцидентности

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Неудачная его свертка вернет один узел из всех четырех переменных, построенный вокруг  $x_4$ , а разумная — два узла:  $(x_1, x_2)$  и  $(x_3, x_4)$ , соединенные между собой. Тогда на шаге 2 внутри каждого узла гамильтонова структура соответствует гармоническому осциллятору с диссипативным портом, как в самом первом примере этого раздела. Взаимодействие между узлами, естественно, задается портом  $u$  и двойственным ему.

## 5. ОБСУЖДЕНИЕ. ПЕРСПЕКТИВЫ

Подводя итог, в этой работе мы предложили некоторый метод восстановления структуры ПГС. Он состоит из части, основанной на машинном обучении, которая выявляет связи между подсистемами, и детерминированной (символьной) части, которая описывает каждую подсистему отдельно. Вторая часть основана на результатах из области симплектической и пуассоновой геометрии, но для приложений мы представили их в форме, не требующей глубоких геометрических знаний и понимания доказательств. Для примеров на данном этапе достаточно использовать довольно простые символьные инструменты (например, пакеты для работы с многочленами и векторными полями на Python с открытым кодом), но в долгосрочной перспективе мы рассматриваем возможности использования коммерческих пакетов и систем символьных вычислений.

Алгоритм исходно задумывался для “упорядочивания” взаимодействующих мультифизических систем, но он, конечно, может использоваться и более широко. Например, для искусственных моделей с нефизическими переменными (т.е. не обязательно интерпретируемыми в терминах состояния механической системы), подход все равно позволит построить некоторое разложение и выявить внутреннюю структуру. Пример такой ситуации для упрощенной модели взаимодействия жидкости с препятствием рассмотрен выше. Некоторые подобные примеры также будут изучены в [11].

Наш подход применим не только для упорядочивания взаимодействий для физических моделей. Его можно рассматривать как новое понятие “нормальных форм” абстрактных дифференциальных уравнений. В отличие от стандартных методов, где классификация часто основана на нелинейности, наш позволяет включать сильно нелинейные эффекты на “простых” уровнях с определенной иерархией.

Мы глубоко признательны С.А. Абрамову и рецензенту за ценные замечания и комментарии, позволившие сделать эту статью гораздо более понятной.

## СПИСОК ЛИТЕРАТУРЫ

1. *Salnikov V., Hamdouni A., Lozijenko D.* Generalized and graded geometry for mechanics: a comprehensive introduction // *Math. and Mech. of Complex Syst.* 2021. V. 9. № 1. P. 59–75.
2. *Verlet L.* Computer “Experiments” on Classical Fluids // *Phys. Rev.* 1967. V. 159. P. 98–103.
3. *Yoshida H.* Construction of higher order symplectic integrators // *Phys. Lett. A.* 1990. V. 150. № 5–7. P. 262–268.
4. *Cosserat O.* Symplectic groupoids for Poisson integrators, Preprint: arXiv:2205.04838.
5. *Paynter H.M.* Analysis and Design of Engineering Systems, MIT Press, Cambridge, Massachusetts, 1961.
6. *Maschke B.M., van der Schaft A.J., Breedveld P.C.* An intrinsic Hamiltonian formulation of network dynamics: nonstandard Poisson structures and gyrators // *J. Franklin Inst.* 1992. V. 329. № 5. P. 923–966.
7. *van der Schaft A.* Port-Hamiltonian systems: an introductory survey // *Proceed. of the Inter. Congress of Math.* 2006. V. III. P. 1339–1365, Madrid.
8. *Cosserat O., Laurent-Gengoux C., Kotov A., Ryvkin L., Salnikov V.* On Dirac structures admitting a variational approach, Preprint: arXiv:2109.00313.
9. *Falaize A.* Modélisation, simulation, génération de code et correction de systèmes multi-physiques audios: Approche par réseau de composants et formulation hamiltonienne à ports, PhD thesis, Télécomm. et Électronique de Paris, Université Pierre et Marie Curie, 2016.
10. *Сальников В.Н., Хамдуни А.* Дифференциальная геометрия и механика — источник задач для компьютерной алгебры // *Программирование.* 2020. № 2. С. 60–66.
11. *Salnikov V., Falaize A., Lozijenko D.* Learning port-Hamiltonian systems — applications, готовится к печати.
12. *Арнольд В.И.* Математические методы классической механики. М.: Наука, 1974.
13. *Cannas Da Silva A., Weinstein A.* Geometric Models for Noncommutative Algebras, Am. Math. Soc. 2000.
14. *Falaize A., Hélie T.* Passive guaranteed simulation of analog audio circuits: A port-hamiltonian approach // *Applied Science, Applied Acoustics, special issue Audio Signal Process.* 2016. V. 6. P. 273.
15. *Falaize A., Hélie T.* Passive simulation of the nonlinear port-hamiltonian modeling of a rhodes piano // *J. of Sound and Vibrat.* 2016. V. 390. P. 289–309.
16. *Evrpidou C.A., Kassotakis P., Vanhaecke P.* Integrable deformations of the Bogoyavlenskij-Itoh Lotka-Volterra systems // *J. of Regular and Chaotic Dynam.* 2017. V. 22 P. 721–739.
17. *Leclercq T., de Langre E.* Vortex-induced vibrations of cylinders bent by the flow // *J. of Fluids and Structur.* 2018. V. 80. P. 77–93.
18. *Salnikov V., Hamdouni A.* Geometric integrators in mechanics — the need for computer algebra tools // *Proceed. of the Third Inter. Conf. “Computer algebra”,* 40–46, 2019, Moscow, Russia.

## КВАНТОВАЯ МАРКОВСКАЯ ДИНАМИКА ПОСЛЕ ВРЕМЕНИ КОРРЕЛЯЦИИ РЕЗЕРВУАРА<sup>1)</sup>

© 2023 г. А. Е. Теретёнков<sup>1,\*</sup>

<sup>1</sup> 119991 Москва, ул. Губкина, 8, Математический институт им. В.А. Стеклова РАН, Россия

\*e-mail: taetsu@mail.ru

Поступила в редакцию 01.06.2022 г.  
Переработанный вариант 01.06.2022 г.  
Принята к публикации 04.06.2022 г.

Показано, что динамика модели многоуровневой системы, взаимодействующей с несколькими резервуарами при нулевой температуре, переходит в марковский режим после времени корреляции резервуара. При этом учтен не только вклад спектральной плотности резервуара, приводящей к непрерывной корреляционной функции, но и вклад омической спектральной плотности, которая приводит к перенормировке как уравнений, так и начальных условий. Получен явный вид уравнения Горини–Коссаковского–Сударшана–Линдблада, описывающий динамику системы после времени корреляции резервуара, а также вид начальных условий для этого уравнения. Они не совпадают с точными начальными условиями как из-за перенормировки, связанной с омическим вкладом, так и за счет короткого начального немарковского периода до времени корреляции резервуара. Библ. 37.

**Ключевые слова:** открытые квантовые системы, квантовая динамическая полугруппа, перенормировка.

**DOI:** 10.31857/S004446692301012X, **EDN:** LMNADE

### 1. ВВЕДЕНИЕ

В работе развиваются результаты работ [1] и [2], посвященных возникновению марковской динамики на больших временах в некоторых специальных моделях открытых квантовых систем. Физическое содержание [1], [3] состоит в том, что такая марковская динамика возникает не просто на больших временах, а после небольшого периода времени (в [1] оно названо временем Зенона) порядка времени корреляции резервуара. В [1] нам удалось строго формализовать этот факт посредством перерастяжки Боголюбова–ван Хофа, суть которой состоит в том, что малый параметр  $\lambda \rightarrow +0$  вводится не только в константу связи, но и во временной масштаб как  $t \rightarrow \lambda^{-2}t$ . Это позволяет отделить масштаб времени, на котором происходит марковская динамика, от масштаба порядка времени корреляции резервуара, которому соответствует начальный временной промежуток порядка  $O(\lambda^2)$  после перерастяжки, на котором локализуются все немарковские эффекты.

Отметим, что предел Боголюбова–ван Хофа  $\lambda \rightarrow +0$  лежит в основе строгого вывода марковских кинетических уравнений вида Горини–Коссаковского–Сударшана–Линдблада (ГКСЛ) [4], [5] в пределе слабой связи [7], [6]. Но нас, подобно [1], [8], [9], интересуют не только предельные уравнения, но и пертурбативные поправки произвольного порядка к ним. Для специальной модели, рассматривавшейся в [2], [10], [11], мы покажем, что в случае выполнения некоторых ограничений на конечность моментов корреляционных функций резервуара можно получить уравнения ГКСЛ с постоянными коэффициентами для редуцированной динамики после времени корреляции с произвольной пертурбативной точностью. Обратим внимание, что существует достаточно обширное обсуждение того, как определять и характеризовать марковость (см., например, [12], [13], [15]–[17] и обзор [18]). В данном случае мы будем подразумевать под марковостью выполнение уравнения ГКСЛ с постоянными коэффициентами, так как для данной модели она обеспечивает выполнение марковских формул (а именно, регрессионной теоремы) также

<sup>1)</sup> Работа выполнена при финансовой поддержке РНФ (проект 19-11-00320).

для некоторых корреляционных функций, а не только одновременной динамики. Более подробно доводы в пользу такого подхода к марковости можно найти в [1], [13].

В работе обсуждается модель многоуровневой системы, взаимодействующая с несколькими резервуарами, предложенная в [10] и в некоторых специальных случаях исследованная в [1], [2]. Поэтому остановимся отдельно на том, что нового сделано в данной работе по сравнению с [1] и [2]. Важным новым результатом настоящей работы является учет (посредством процедуры регуляризации, развитой в [11]) омического вклада в корреляционную функцию перед дальнейшим использованием теории возмущений Боголюбова—ван Хова. Наличие такого вклада, с одной стороны, обеспечивает для ряда моделей открытых квантовых систем существование классических пределов, совпадающих с общепринятыми [14, подразд. 3.6.2], а с другой стороны, требует предварительной перенормировки. Кроме того, и сам случай произвольных порядков теории возмущений для данной многоуровневой модели даже без учета омического вклада рассматривается здесь впервые. Хотя первые два нетривиальных порядка без учета омических вкладов исследовались в [2].

Основной текст статьи имеет следующую структуру. В разд. 2 дается краткое описание модели, введенной в [10]. В ней также формулируется совокупный результат работ [2], [10], [11] в виде утверждения 1, выражающего динамику редуцированной матрицы плотности в такой модели к интегродифференциальному уравнению (6). Перенормировке, связанной с наличием омического вклада в спектральную плотность, посвящен разд. 3. Результат такой перенормировки описывается утверждением 2. Фактически он сводится к переходу от уравнения (6) к (10) и соответствующей замене начального условия. Основные результаты статьи, посвященные теории возмущений данной модели с перерастяжкой Боголюбова—ван Хова, а также приданию получившейся динамике матрицы плотности вида Горини—Коссаковского—Сударшана—Линдблада, приведены в разд. 4 и 5 соответственно. А именно, утверждение 3 дает пертурбативное описание решения уравнения (6) с учетом перехода в представление взаимодействия, описываемого леммой 1. Так как в физической литературе значительное внимание уделяется именно виду линейных кинетических уравнений за рамками нулевого порядка теории возмущений с перерастяжкой Боголюбова—ван Хова, а не самих решений, то мы также формулируем основной результат данной работы в терминах уравнений, которым удовлетворяет редуцированная матрица плотности. А именно, в теореме 1 мы утверждаем, что в условиях утверждения 3 редуцированная матрица плотности удовлетворяет уравнению ГКСЛ.

## 2. МОДЕЛЬ МНОГОУРОВНЕВОЙ СИСТЕМЫ, ВЗАИМОДЕЙСТВУЮЩЕЙ С РЕЗЕРВУАРАМИ

Исследуем модель, которая рассматривалась в [2], [10], [11], поэтому приведем только ее формулировку и основные результаты, необходимые для данной работы.

Рассмотрим гильбертово пространство

$$\mathcal{H} \equiv (\mathbb{C} \oplus \mathbb{C}^N) \otimes \bigotimes_{i=1}^N \mathfrak{F}_b(\mathcal{L}^2(\mathbb{R})),$$

где  $\mathbb{C} \oplus \mathbb{C}^N$  есть  $(N+1)$ -мерное гильбертово пространство с выделенным 1-мерным подпространством. Здесь  $\mathbb{C}^N$  соответствует возбужденным состояниям системы, а выделенное подпространство соответствует основному состоянию. Обозначим через  $|i\rangle$ ,  $i = 0, 1, \dots, N$ , некоторый фиксированный ортонормированный базис пространства  $\mathbb{C} \oplus \mathbb{C}^N$ , где  $|0\rangle$  соответствует выделенному подпространству.  $\mathfrak{F}_b(\mathcal{L}^2(\mathbb{R}))$  — бозонные фоковские пространства, соответствующие резервуарам. Пусть  $|\Omega\rangle$  — совместное вакуумное состояние этих резервуаров. Также введем операторы рождения и уничтожения, удовлетворяющие каноническим коммутационным соотношениям:  $[b_i(k), b_j^\dagger(k')] = \delta_{ij} \delta(k - k')$ ,  $[b_i(k), b_j(k')] = 0$ ,  $b_i(k)|\Omega\rangle = 0$ , где  $i, j = 1, \dots, N$ . Тот факт, что число резервуаров равно числу возбужденных уровней, связан с тем, что с физической точки зрения предполагается, что возбужденные уровни локализованы в разных точках пространства, достаточно далеких, чтобы резервуары, с которыми они связаны, можно было считать полностью независимыми.

Рассмотрим гамильтониан системы, такой, что он обнуляется на основном состоянии. А именно,

$$\hat{H}_S = \sum_i \varepsilon_i |i\rangle\langle i| + \sum_{i \neq j} J_{ij} |i\rangle\langle j| = 0 \oplus H_S, \quad i, j = 1, \dots, N, \quad (1)$$

где  $H_S \in \mathbb{C}^{N \times N}$  и  $0 \oplus H_S$  представляет собой блочную матрицу с диагональными блоками  $1 \times 1$  и  $N \times N$  и только  $N \times N$  является ненулевым и равным  $H_S$ , т.е.

$$0 \oplus H_S \equiv \begin{pmatrix} 0 & 0 \\ 0 & H_S \end{pmatrix}.$$

Подчеркнем, что здесь не предполагается, что  $|i\rangle$  являются собственными для  $\hat{H}_S$ . С физической точки зрения базис из векторов  $|i\rangle$  является локальным базисом [19].

Гамильтонианы резервуаров описываются гамильтонианами свободных бозонных полей с одинаковым дисперсионным соотношением  $\omega(k)$ . Таким образом, суммарный гамильтониан резервуаров имеет вид

$$\hat{H}_B = \sum_{i=1}^N \int \omega(k) b_i^\dagger(k) b_i(k) dk. \quad (2)$$

Взаимодействие системы и резервуаров описывается гамильтонианом дипольного взаимодействия в приближении вращающейся волны

$$\hat{H}_I = \sum_i \int (g^*(k) |0\rangle\langle i| \otimes b_{k,i}^\dagger + g(k) |i\rangle\langle 0| \otimes b_{k,i}) dk. \quad (3)$$

Мы рассматриваем унитарную эволюцию матрицы плотности

$$\rho(t) = e^{-i\hat{H}t} \rho(0) e^{i\hat{H}t}$$

с гамильтонианом  $\hat{H} = \hat{H}_S \otimes I + I \otimes \hat{H}_B + \hat{H}_I$  и факторизованными начальными условиями

$$\rho(0) = \rho_S(0) \otimes |\Omega\rangle\langle\Omega|. \quad (4)$$

Нас интересует динамика редуцированной матрицы плотности

$$\rho_{SI}(t) \equiv \text{Tr}_B \rho_I(t).$$

Данная модель и ее частные случаи широко используются как тестовые модели для многих подходов к открытым квантовым системам [20]–[23], [25]–[29]. Отметим, однако, что мы не обсуждаем здесь справедливость приближения вращающейся волны, заложенную в эту модель уже на этапе постановки задачи. Но при использовании полученных здесь результатов для описания реальных физических систем область применения приближения вращающейся волны безусловно необходимо учитывать и для этого существует широкий набор специфических методов [30]–[32], выходящий за рамки данного исследования.

Суммируем результаты [10, следствие 1], [11, теорема 1] и [2, следствие 6] в представлении Шрёдингера в виде следующего утверждения.

**Утверждение 1.** Пусть интеграл (корреляционная функция резервуара)

$$G(t) \equiv \int |g(k)|^2 e^{-i\omega(k)t} dk$$

сходится для всех  $t \in \mathbb{R}_+$  и определяет непрерывную функцию, тогда динамика редуцированной матрицы плотности может быть представлена в блочном виде

$$\rho_S(t) = \begin{pmatrix} (\rho_S(0))_{gg} + \text{Tr}((\rho_S(0))_{ee} - V(t)(\rho_S(0))_{ee} V^+(t)) & (\rho_S(0))_{ge} V^+(t) \\ V(t)(\rho_S(0))_{eg} & V(t)(\rho_S(0))_{ee} V^+(t) \end{pmatrix}, \quad (5)$$

где  $V(t) \in \mathbb{C}^{N \times N}$  – решение интегродифференциального уравнения,

$$\frac{d}{dt} V(t) = -iH_S V(t) - \int_0^t ds G(t-s) V(s) \quad (6)$$

с начальным условием  $V(0) = I$ , а  $e$  соответствует возбужденным состояниям,  $g$  — основному.

### 3. ОМИЧЕСКАЯ СПЕКТРАЛЬНАЯ ПЛОТНОСТЬ И ПЕРЕНОРМИРОВКА

Часто в физических задачах считаются заданными не форм-факторы  $g(k)$  и дисперсионные соотношения  $\omega(k)$  и не корреляционная функция резервуара  $G(t)$ , а спектральная плотность резервуара, связанная с  $G(t)$  посредством преобразования Фурье:

$$G(t) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \mathcal{J}(\omega), \quad \mathcal{J}(\omega) = \int_{-\infty}^{+\infty} G(t) e^{i\omega t} dt.$$

Нас будет интересовать влияние омической спектральной плотности:

$$\mathcal{J}_{\text{Ohmic}}(\omega) = \eta \omega, \quad \eta > 0. \quad (7)$$

Так как преобразование Фурье данной функции существует только в смысле обобщенных функций, то утверждение 1 не может быть применено непосредственно. Поэтому необходимо сформулировать ту или иную процедуру регуляризации, приблизив данную функцию такими спектральными плотностями, что соответствующие корреляционные функции уже удовлетворяют утверждению 1. Сразу подчеркнем, что для того, чтобы в результате решение (6) имело нетривиальный конечный предел, параметр регуляризации необходимо также добавить в гамильтониан системы. Данная добавка называется контр-членом [14, подразд. 3.6.1]. При этом исходный гамильтониан не будет иметь конечного предела, однако та часть его, которая не зависит от параметра регуляризации, даст вклад в предельное уравнение. Такой гамильтониан называется перенормированным, а такая процедура регуляризации называется перенормировкой. Кроме того, естественно задаться вопросом о зависимости результата от конкретной регуляризации. И уже для двухуровневых моделей системы [24] было показано, что результат существенно зависит от конкретной регуляризации. Поэтому выражение (7) даже на физическом уровне строгости не может считаться вполне строгим, так как различные физически осмысленные регуляризации данного выражения приводят к различным результатам.

Поэтому, подобно [11], остановимся на конкретной регуляризации. А именно, рассмотрим семейство спектральных плотностей с экспоненциальным обрезанием

$$\mathcal{J}_{\Omega}(\omega) = \eta \omega e^{-\frac{|\omega|}{\Omega}},$$

параметризованных частотой обрезания  $\Omega$ . В пределе  $\Omega \rightarrow +\infty$  такая спектральная плотность  $\mathcal{J}_{\Omega}(\omega)$  поточечно стремится к  $\mathcal{J}_{\text{Ohmic}}(\omega)$ . Соответствующая корреляционная функция имеет вид

$$G_{\Omega}(t) = -i\eta \frac{2t\Omega^3}{\pi(1 + (\Omega t)^2)}. \quad (8)$$

В качестве  $H_S$  также рассмотрим семейство

$$H_S(\Omega) = H_S^{(r)} + \frac{\eta\Omega}{\pi}, \quad (9)$$

где  $H_S^{(r)}$  — перенормированный гамильтониан, не зависящий от  $\Omega$ , а  $\frac{\eta\Omega}{\pi}$  играет роль контр-члена.

Подчеркнем, что несмотря на то, что в (9) этот контр-член выглядит как простое добавление константы к гамильтониану (и поэтому может показаться, что он не должен на что-либо влиять), но с учетом формулы (1) в полном гамильтониане  $\hat{H}_S(\Omega)$  контр-член сдвигает на константу только возбужденные уровни, но не основное состояние. Так как  $\Omega \rightarrow +\infty$ , то с физической точки зрения такая ситуация соответствует тому, что энергии перехода из основного состояния в возбужденное много больше, чем энергии перехода между возбужденными состояниями.

**Утверждение 2.** Пусть  $V_{\Omega}(t)$  — решение (6) с начальным условием  $V_{\Omega}(0) = I$ , где  $G(t) = G_{\Omega}(t) + G_c(t)$ ,  $G_{\Omega}(t)$  определяется по формуле (8) и зависит от  $\Omega$  как от параметра,  $G_c(t)$  — непрерывная функция,  $H_S = H_S(\Omega)$  определяется формулой (9). Пусть предел  $\lim_{\Omega \rightarrow +\infty} V_{\Omega}(t) = V^{(r)}(t)$

существует для произвольного  $t > 0$  и определяет бесконечно дифференцируемую функцию на полуоси  $(0, +\infty)$ , тогда  $V^{(r)}(t)$  является решением уравнения

$$\frac{d}{dt} V^{(r)}(t) = -i \frac{1}{1 + i \frac{\eta}{2}} H_S^{(r)} V^{(r)}(t) - \int_0^t ds \frac{1}{1 + i \frac{\eta}{2}} G_c(t-s) V^{(r)}(s) \quad (10)$$

с начальным условием  $V^{(r)}(0) = \frac{1}{1 + i \frac{\eta}{2}} I$ .

Доказательство данного утверждения является прямым обобщением [11, Утверждение 2], так как фактически матрица  $V^{(r)}(t)$  просто составлена из векторов, динамика каждого из которых описывается [11, Утверждение 2].

Перейдем в представление взаимодействия с перенормированным гамильтонианом  $\hat{H}_S^{(r)}$ , который связан с  $H_S^{(r)}$  по формуле (1), так как именно такое представление нам будет удобно в последующих разделах статьи. Для этого сформулируем следующую лемму.

**Лемма 1.** Пусть

$$\rho_{SI}(t) \equiv e^{i\hat{H}_S^{(r)}t} \rho_S(t) e^{-i\hat{H}_S^{(r)}t}$$

и

$$V_I^{(r)}(t) \equiv e^{iH_S^{(r)}t} V^{(r)}(t).$$

Тогда  $V_I^{(r)}(t)$  удовлетворяет интегродифференциальному уравнению

$$\frac{d}{dt} V_I^{(r)}(t) = -\frac{\frac{\eta}{2}}{1 + i \frac{\eta}{2}} H_S^{(r)} V_I^{(r)}(t) - \int_0^t ds \frac{1}{1 + i \frac{\eta}{2}} G_c(t-s) e^{iH_S^{(r)}(t-s)} V_I^{(r)}(s) \quad (11)$$

с начальным условием  $V_I^{(r)}(0) = V^{(r)}(0) = \frac{1}{1 + i \frac{\eta}{2}} I$ , а  $\rho_{SI}(t)$  может быть определено по формуле (5), где

вместо  $V(t)$  стоит  $V_I^{(r)}(t)$ .

#### 4. ПЕРЕРАСТЯЖКА БОГОЛЮБОВА–ВАН ХОВА

Подобно [1], [2], нас будет интересовать поведение данной модели в пределе Боголюбова–ван Хова и асимптотическое разложение по малому параметру вблизи него. Если ввести малый параметр, отвечающий за малость константы связи в исходном гамильтониане (3), заменив  $g(k) \rightarrow \lambda g(k)$ , то это приведет к замене  $G(t) \rightarrow \lambda^2 G(t)$  в (6). Таким образом, в уравнении (11) необходимо сделать замены  $G_c(t) \rightarrow \lambda^2 G_c(t)$ ,  $\eta \rightarrow \lambda^2 \eta$ . В итоге (11) примет вид

$$\frac{d}{dt} V_{I,\lambda}(t) = -\frac{\lambda^2 \frac{\eta}{2}}{1 + i \lambda^2 \frac{\eta}{2}} H_S^{(r)} V_{I,\lambda}(t) - \int_0^t ds \frac{\lambda^2}{1 + i \lambda^2 \frac{\eta}{2}} G_c(t-s) e^{iH_S^{(r)}(t-s)} V_{I,\lambda}(s). \quad (12)$$

Сделаем перерастяжку Боголюбова–ван Хова, определив функцию

$$W_\lambda(t) \equiv V_{I,\lambda}(\lambda^{-2}t).$$

Тогда с учетом (12) имеем

$$\frac{d}{dt} W_\lambda(t) = -\frac{\frac{\eta}{2}}{1 + i\lambda^2 \frac{\eta}{2}} H_S^{(r)} W_\lambda(t) - \int_0^t ds \frac{1}{1 + i\lambda^2 \frac{\eta}{2}} \frac{1}{\lambda^2} G_c\left(\frac{t-s}{\lambda^2}\right) e^{iH_S^{(r)}\left(\frac{t-s}{\lambda^2}\right)} W_\lambda(s). \quad (13)$$

Начальное условие с учетом замены  $\eta \rightarrow \lambda^2 \eta$  примет вид

$$W_\lambda(0) = \frac{1}{1 + i\lambda^2 \frac{\eta}{2}} I. \quad (14)$$

Далее мы будем исследовать асимптотическое поведение  $W_\lambda(t)$  при  $\lambda \rightarrow +0$ . Поэтому подчеркнем, что наша схема регуляризации, когда мы сначала устремили  $\Omega \rightarrow +\infty$ , а затем делаем разложение по малому параметру, является лишь одной из возможных. Хотя для моделей, подобной нашей, такой подход является распространенным [24], можно было бы рассматривать и случай частоты обрезки, явно зависящий от малого параметра  $\lambda$ , и тогда последовательное устремление  $\Omega \rightarrow +\infty$ , а затем  $\lambda \rightarrow +0$  было бы возможно, только если  $\Omega \rightarrow +\infty$  быстрее всякой степени  $\lambda$  (так как только такими членами мы пренебрежем в нашем последующем разложении по  $\lambda$  в общем случае).

Обозначим преобразования Лапласа  $W_\lambda(t)$  и  $G(t)$  как

$$\tilde{W}_\lambda(p) = \int_0^{+\infty} dt e^{-pt} W_\lambda(t), \quad \tilde{G}(p) = \int_0^{+\infty} dt e^{-pt} G(t).$$

**Лемма 2.** Преобразование Лапласа функции  $W_\lambda(t)$  имеет вид

$$\tilde{W}_\lambda(p) = \frac{1}{\left(1 + i\lambda^2 \frac{\eta}{2}\right)p + \frac{\eta}{2} H_S^{(r)} + \tilde{G}_c(-iH_S^{(r)} + \lambda^2 p)}. \quad (15)$$

**Замечание 1.** Выражение (15) необходимо понимать как функцию эрмитовой матрицы  $H_S^{(r)} = (H_S^{(r)})^\dagger$ , которая определяется стандартным образом. А именно, если диагонализировать

$$H_S^{(r)} = U \operatorname{diag}\{E_\alpha^{(r)}\} U^\dagger, \quad \alpha = 1, \dots, N,$$

посредством унитарной матрицы  $U$ , то

$$\frac{1}{\left(1 + i\lambda^2 \frac{\eta}{2}\right)p + \frac{\eta}{2} H_S^{(r)} + \tilde{G}_c(-iH_S^{(r)} + \lambda^2 p)} \equiv U \operatorname{diag} \left\{ \frac{1}{\left(1 + i\lambda^2 \frac{\eta}{2}\right)p + \frac{\eta}{2} E_\alpha^{(r)} + \tilde{G}_c(-iE_\alpha^{(r)} + \lambda^2 p)} \right\} U^\dagger.$$

**Доказательство.** В первую очередь вычислим преобразование Лапласа

$$\begin{aligned} \int_0^\infty e^{-pt} \frac{1}{\lambda^2} G\left(\frac{t}{\lambda^2}\right) e^{iH_S^{(r)}\left(\frac{t}{\lambda^2}\right)} dt &= U \operatorname{diag} \left\{ \frac{1}{\lambda^2} \int_0^\infty e^{-pt} G\left(\frac{t}{\lambda^2}\right) e^{iE_\alpha^{(r)}\frac{t}{\lambda^2}} dt \right\} U^\dagger = \\ &= U \operatorname{diag} \left\{ \tilde{G}(-iE_\alpha^{(r)} + \lambda^2 p) \right\} U^\dagger = \tilde{G}(-iH_S^{(r)} + \lambda^2 p). \end{aligned}$$

Тогда преобразование Лапласа уравнения (13) имеет вид

$$p \tilde{W}_\lambda(p) - W_\lambda(0) = -\frac{1}{1 + i\lambda^2 \frac{\eta}{2}} \left( \frac{\eta}{2} H_S^{(r)} + \tilde{G}_c(-iH_S^{(r)} + \lambda^2 p) \right) \tilde{W}_\lambda(p).$$

Откуда, с учетом начального условия (14), имеем

$$\tilde{W}_\lambda(p) = \frac{1}{p + \frac{1}{1 + i\lambda^2 \frac{\eta}{2}} \left( \frac{\eta}{2} H_S^{(r)} + \tilde{G}_c(-iH_S^{(r)} + \lambda^2 p) \right)} \frac{1}{1 + i\lambda^2 \frac{\eta}{2}} I,$$

что эквивалентно (15).

Сформулируем результаты [1, Приложение В] в виде следующей леммы.

**Лемма 3.** Пусть для функции  $\tilde{f}(p)$  верно асимптотическое разложение

$$\tilde{f}(p) = \sum_{k=0}^n \tilde{f}_k p^k + O(p^{n+1}), \quad p \rightarrow 0, \quad n \in \mathbb{Z}_+,$$

и функция

$$\tilde{x}_\lambda(p) = \frac{1}{p + \tilde{f}(\lambda^2 p)} \quad (16)$$

является преобразованием Лаласа функции  $x_\lambda(t)$ , т.е.

$$\tilde{x}_\lambda(p) = \int_0^\infty e^{-pt} x_\lambda(t) dt.$$

Тогда при каждом фиксированном  $t > 0$  и  $\lambda \rightarrow +0$

$$x_\lambda(t) = r_{2n,\lambda} e^{\tilde{p}_{2n,\lambda} t} + O(\lambda^{2n+2}), \quad (17)$$

где  $\tilde{p}_{2n,\lambda}$  определяется как полином по  $\lambda$  степени  $2n$  такой, что

$$\tilde{p}_\lambda = \tilde{p}_{2n,\lambda} + O(\lambda^{2n+2}),$$

где  $\tilde{p}_\lambda$  — решение уравнения

$$\tilde{p}_\lambda + \tilde{f}(\lambda^2 \tilde{p}_\lambda) = 0, \quad (18)$$

а  $r_{2n,\lambda}$  — полином по  $\lambda$  степени  $2n$  такой, что

$$r_{2n,\lambda} = \frac{1}{1 + \lambda^2 \tilde{f}'(\lambda^2 \tilde{p}_{2n,\lambda})} + O(\lambda^{2n+2}). \quad (19)$$

**Замечание 2.** Если функция  $\tilde{f}(p)$  разложима в ряд Маклорена при  $p \rightarrow 0$ , то (17) будет выполнено для сколь угодно большого  $n$ . По аналогии с [1, Разд. 3] это можно формально (в том смысле, что запись ниже является лишь символической и ее строгий смысл раскрывается в предыдущей лемме) записать как

$$x_\lambda(t)|_{\text{pert}} = r_\lambda e^{\tilde{p}_\lambda t}, \quad (20)$$

где  $\tilde{p}_\lambda$  — “решение” (18), а  $r_\lambda$  опять-таки формально определяется как

$$r_\lambda = \frac{1}{1 + \lambda^2 \tilde{f}'(\lambda^2 \tilde{p}_\lambda)}.$$

Здесь обозначение  $x_\lambda(t)|_{\text{pert}}$  — “пертурбативная” часть  $x_\lambda(t)$  подчеркивает тот факт, что даже в случае существования пределов  $\lim_{n \rightarrow +\infty} \tilde{p}_{2n,\lambda}$  и  $\lim_{n \rightarrow +\infty} r_{2n,\lambda}$  при некотором  $\lambda > 0$  данное выражение не совпадает с  $x_\lambda(t)$ , а отличается от него на “непертурбативные члены”, т.е. члены, стремящиеся к нулю быстрее всякой степени  $\lambda$ . Кроме того, в общем случае пределы  $\lim_{n \rightarrow +\infty} \tilde{p}_{2n,\lambda}$  и  $\lim_{n \rightarrow +\infty} r_{2n,\lambda}$  могут и не существовать ни для какого  $\lambda > 0$ . Поэтому формула (20) является лишь формальной, и ее строгий смысл раскрывается именно леммой 3 при произвольных  $n$ .

**Замечание 3.** Отметим, что простая экспоненциальная зависимость асимптотического выражения (17) от времени возникает ценой того, что данная зависимость является именно поточеч-

ной, а не равномерной по времени. Она не выполнена для малых времен  $t = O(\lambda^2)$ . В частности, поэтому  $x_\lambda(0)$ , вообще говоря, не совпадает с  $r_{2n,\lambda}$  и асимптотика (17) при  $t = 0$  не верна.

**Утверждение 3.** Пусть функция  $|G_c(t)|$  имеет конечные моменты

$$\int_0^\infty t^k |G_c(t)| dt < \infty \quad (21)$$

для  $k = 0, 1, \dots, n+1$ , тогда при  $t > 0$  и  $\lambda \rightarrow +0$  имеет место следующее асимптотическое равенство:

$$W_\lambda(t) = r_{2n,\lambda}(H_S^{(r)})e^{\tilde{p}_{2n,\lambda}(H_S^{(r)})t} + O(\lambda^{2n+2}),$$

где  $\tilde{p}_{2n,\lambda}(E)$  определяется как полином по  $\lambda$  степени  $2n$  такой, что (переменная  $E \in \mathbb{R}$  и предполагается здесь и ниже во всех асимптотических выражениях фиксированной)

$$\tilde{p}_{2n,\lambda}(E) = \tilde{p}_\lambda(E) + O(\lambda^{2n+2}),$$

где  $\tilde{p}_\lambda(E)$  является решением уравнения

$$\left(1 + i\lambda^2 \frac{\eta}{2}\right) \tilde{p}_\lambda(E) + \frac{\eta}{2} E + \tilde{G}_c(-iE + \lambda^2 \tilde{p}_\lambda(E)) = 0, \quad (22)$$

а  $r_{2n,\lambda}(E)$  — полином по  $\lambda$  степени  $2n$  такой, что

$$r_{2n,\lambda}(E) = \frac{1}{1 + i\lambda^2 \frac{\eta}{2} + \lambda^2 \tilde{G}_c(-iE + \lambda^2 \tilde{p}_{2n,\lambda}(E))} + O(\lambda^{2n+2}). \quad (23)$$

**Доказательство.** Применим лемму 3 к функции

$$f(p, E) = i\frac{\eta}{2}p + \frac{\eta}{2}E + \tilde{G}_c(-iE + p)$$

при фиксированном  $E$ . Условие (21) обеспечивает сходимость интегралов

$$\int_0^\infty t^k G_c(t) e^{iEt} dt < \infty. \quad (24)$$

С другой стороны,

$$\frac{d^k}{dp^k} \tilde{G}_c(-iE + p) = (-1)^k \int_0^\infty t^k G_c(t) e^{iEt} dt.$$

Тогда разложение  $f(p, E)$  в ряд Маклорена по  $p$  с остаточным членом в форме Пеано обеспечивает выполнение условий леммы 3 для этой функции. В результате по лемме 3 для функции  $x_\lambda(t; E)$ , преобразование Лапласа которой имеет вид

$$\tilde{x}_\lambda(p; E) = \frac{1}{p + f(\lambda^2 p, E)},$$

получим асимптотику при  $\lambda \rightarrow +0$

$$x_\lambda(t; E) = r_{2n,\lambda}(E) e^{\tilde{p}_{2n,\lambda}(E)t} + O(\lambda^{2n+2}), \quad (25)$$

где с учетом  $\frac{d}{dp} f(p, E) = i\frac{\eta}{2} + \tilde{G}_c'(-iE + p)$  по формулам (18) и (19) приходим к (22) и (23) соответственно.

По лемме 2 имеем

$$\tilde{W}_\lambda(p) = U \operatorname{diag} \left\{ \frac{1}{\left(1 + i\lambda^2 \frac{\eta}{2}\right)p + \frac{\eta}{2} E_\alpha^{(r)} + \tilde{G}_c(-iE_\alpha^{(r)} + \lambda^2 p)} \right\} U^\dagger = U \operatorname{diag} \{ \tilde{x}_\lambda(p; E_\alpha^{(r)}) \} U^\dagger.$$

Применяя асимптотическое равенство (25), имеем

$$W_\lambda(t) = U \operatorname{diag} \left\{ r_{2n,\lambda}(E_\alpha^{(r)}) e^{\tilde{p}_{2n,\lambda}(E_\alpha^{(r)})t} + O(\lambda^{2n+2}) \right\} U^\dagger = r_{2n,\lambda}(H_S^{(r)}) e^{\tilde{p}_{2n,\lambda}(H_S^{(r)})t} + O(\lambda^{2n+2}).$$

**Замечание 4.** В случае если конечны моменты (24) произвольных порядков, то можно формально записать

$$W_\lambda(t)|_{\text{pert}} = \frac{1}{1 + i\lambda^2 \frac{\eta}{2} + \lambda^2 \tilde{G}_c(-iH_S^{(r)} + \lambda^2 \tilde{p}_\lambda(H_S^{(r)}))} e^{\tilde{p}_\lambda(H_S^{(r)})t}, \quad (26)$$

с теми же оговорками о строгом смысле данной записи, что и (20).

**Замечание 5.** В частности, утверждение 3 позволяет получить выражение для  $W_\lambda(t)$  в пределе Боголюбова—ван Хова, а именно

$$W_0(t) = e^{\tilde{p}_0(H_S^{(r)})t},$$

где

$$\tilde{p}_0(E) = -\left(\frac{\eta}{2} E + \tilde{G}_c(-iE)\right).$$

Обратим внимание на то, что перенормировка начального условия (14), возникающая как следствие устремления частоты обрезки к бесконечности, в пределе Боголюбова—ван Хова исчезает.

## 5. УРАВНЕНИЕ ГОРИНИ—КОССАКОВСКОГО—СУДАРШАНА—ЛИНДБЛАДА ПОСЛЕ ВРЕМЕНИ КОРРЕЛЯЦИИ РЕЗЕРВУАРА

Используя [25, Утверждение 5], можно придать указанной выше динамике вид ГКСЛ, что может быть сформулировано в виде следующей теоремы.

**Теорема 1.** Пусть матрица плотности  $\rho_M(t)$  имеет вид (5), где вместо  $V(t)$  стоит  $r_{2n,\lambda}(H_S^{(r)}) e^{\tilde{p}_{2n,\lambda}(H_S^{(r)})t}$  в случае, когда  $-2 \operatorname{Re} \tilde{p}_{2n,\lambda}(H_S^{(r)})$  неотрицательно определен. Тогда  $\rho_M(t)$  удовлетворяет следующему уравнению, которое имеет вид ГКСЛ:

$$\frac{d}{dt} \rho_M(t) = -i[\hat{H}_M, \rho_M(t)] + \sum_{\alpha=1}^K \left( \hat{L}_\alpha \rho_M(t) \hat{L}_\alpha^\dagger - \frac{1}{2} \hat{L}_\alpha^\dagger \hat{L}_\alpha \rho_M(t) - \frac{1}{2} \rho_M(t) \hat{L}_\alpha^\dagger \hat{L}_\alpha \right), \quad (27)$$

где

$$\hat{L}_\alpha = |0\rangle \langle \hat{f}_\alpha|, \quad |\hat{f}_\alpha\rangle = 0 \oplus |f_\alpha\rangle, \quad \hat{H}_M = 0 \oplus H_M,$$

а  $H_M$  и  $|f_\alpha\rangle$ , в свою очередь, определяются формулами

$$-\operatorname{Im} \tilde{p}_{2n,\lambda}(H_S^{(r)}) = H_M, \quad -2 \operatorname{Re} \tilde{p}_{2n,\lambda}(H_S^{(r)}) = \sum_{\alpha=1}^K |f_\alpha\rangle \langle f_\alpha|.$$

**Замечание 6.** Используя спектральное разложение

$$H_S^{(r)} = \sum_{\alpha=1}^N E_\alpha^{(r)} |\alpha\rangle \langle \alpha|,$$

$H_M$  и  $|f_\alpha\rangle$  можно представить более явно как

$$H_M = -\sum_{\alpha=1}^N \operatorname{Im} \tilde{p}_{2n,\lambda}(E_\alpha^{(r)}) |\alpha\rangle \langle \alpha|, \quad |f_\alpha\rangle = \sqrt{-2 \operatorname{Re} \tilde{p}_{2n,\lambda}(E_\alpha^{(r)})} |\alpha\rangle.$$

**Замечание 7.** В случае, когда  $-2 \operatorname{Re} \tilde{p}_{2n,\lambda}(E_\alpha^{(r)}) < 0$  для некоторого  $\alpha$ , будет нарушаться положительность матрицы плотности на больших временах и мы исключим из рассмотрения такой случай, как нефизичный.

**Замечание 8.** Отметим, что начальное условие для  $\rho_M(t)$ , вообще говоря, не будет совпадать с начальным условием  $\rho_S(0)$ , а именно, подобно (5), его можно представить в блочном виде:

$$\rho_M(0) = \begin{pmatrix} (\rho_S(0))_{gg} + \operatorname{Tr}((\rho_S(0))_{ee} - r_{2n,\lambda}(H_S^{(r)})(\rho_S(0))_{ee} r_{2n,\lambda}^+(H_S^{(r)})) & (\rho_S(0))_{ge} r_{2n,\lambda}^+(H_S^{(r)}) \\ r_{2n,\lambda}(H_S^{(r)})(\rho_S(0))_{eg} & r_{2n,\lambda}(H_S^{(r)})(\rho_S(0))_{ee} r_{2n,\lambda}^+(H_S^{(r)}) \end{pmatrix}.$$

Оно также, в общем случае, не будет совпадать с начальным условием, которое возникает после перенормировки, описанной в разд. 3, и может быть вычислено по аналогичной формуле, в которой вместо  $r_{2n,\lambda}$  стоит  $W_\lambda(0)$ , определенное формулой (14). С математической точки зрения это связано с явлением начального (пограничного) слоя, известного в сингулярной теории возмущений (см. [35, Параграф 3], [36, Разд. 7.2], [37, Разд. 1.3]). А именно, в  $\rho_M(t)$  для сколь угодно большого  $n$  не будут учтены члены, которые при фиксированном (уже после перерастяжки Боголюбова—ван Хофа) времени  $t$  убывают быстрее всякой степени  $\lambda$ . Но эта асимптотика не является равномерной по времени, она не верна для небольшого начального интервала времен, длина которого имеет порядок  $O(\lambda^2)$ . И члены, не учтенные в  $\rho_M(t)$ , могут стать не малыми в этом интервале; в частности, их нужно учитывать и при постановке начального условия. Это приводит к дополнительной перенормировке начальных условий для  $\rho_M(t)$ . С физической точки зрения длина этого малого интервала времени характеризуется временем корреляции резервуара. Поэтому мы говорим, что марковская динамика, задаваемая уравнением (27), возникает после времени корреляции резервуара.

Однако, как уже было сказано в предыдущем разделе, в пределе Боголюбова—ван Хофа перенормировка начальных условий, возникшая как в результате учета омической спектральной плотности, так и в результате непосредственной неравномерности марковской по времени асимптотики, исчезают. Возможно, именно поэтому перенормировки начальных состояний, связанные с учетом омической спектральной плотности, ранее мало обсуждались в литературе.

## 6. ЗАКЛЮЧЕНИЕ

В работе были получены уравнения ГКСЛ, которые описывают динамику многоуровневой системы, взаимодействующей с несколькими резервуарами при нулевой температуре после времени корреляции резервуара. При этом был учтен вклад омической спектральной плотности, а также спектральной плотности, приводящей к непрерывной корреляционной функции резервуара. Асимптотическая точность, с которой выполнены марковские уравнения, определяется количеством конечных моментов модуля этой корреляционной функции. Как учет омической спектральной плотности, так и переход в марковский режим после времени корреляции приводят к перенормировке начальных условий. Были получены явные выражения для данной перенормировки.

В частности, было показано, что в пределе Боголюбова—ван Хофа, т.е. в нулевом порядке теории возмущений с перерастяжкой Боголюбова—ван Хофа, данные перенормировки исчезают. Этот эффект интересен в связи с тем, что, с одной стороны, в современной теории открытых квантовых систем активно обсуждаются поправки к уравнениям, возникающим в пределе слабой связи [8], [9], [24], [33], с другой стороны, основные работы в этой области до последнего времени фокусировались именно на выводе асимптотических поправок к самим уравнениям, но не к начальным условиям этих уравнений. Кроме того, результаты работы [34] позволяют смотреть на перенормированные начальные условия как на начальные условия, которые учитывают наличие корреляции между системой и резервуаром в начальный момент. В частности, есть основания полагать, что именно такие коррелированные состояния больше соответствуют реальной физике моделируемых систем, а не постулированные здесь факторизованные начальные условия. Отметим, что и в таком подходе полученные нами перенормировки важны, так как если в случае факторизованных начальных состояний в качестве начального состояния можно выбрать произвольную матрицу плотности, то множество начальных состояний, учитывающих начальную корреляцию, как раз описывается полученными перенормировками.

Автор хотел бы поблагодарить рецензента за ценные замечания, которые позволили существенно улучшить качество данной работы.

## СПИСОК ЛИТЕРАТУРЫ

1. *Teretenkov A.E.* Non-perturbative effects in corrections to quantum master equations arising in Bogolubov–van Hove limit // *J. Phys. A: Math. Theor.* 2021. V. 54. № 26. P. 265302–265302.
2. *Teretenkov A.E.* Long-time Markovianity of multi-level systems in the rotating wave approximation // *Lobachevskii J. Math.* 2021. V. 42. № 10. P. 2455–2465.
3. *Petrosky T., Barsegov V.* Quantum decoherence, Zeno process, and time symmetry breaking // *Phys. Rev. E.* 2002. V. 65. № 4. P. 046102.
4. *Gorini V., Kossakowski A., Sudarshan E.C.G.* Completely positive dynamical semigroups of N-level systems // *J. Math. Phys.* 1976. V. 17. № 5. P. 821–825.
5. *Lindblad G.* On the generators of quantum dynamical semigroups // *Comm. in Math. Phys.* 1976. V. 48. № 2. P. 119–130.
6. *Accardi L., Lu Y.G., Volovich I.* Quantum theory and its stochastic limit. Berlin: Springer, 2002.
7. *Davies E.B.* Markovian master equations // *Commun. Math. Phys.* 1974. V. 39. № 2. P. 91–110.
8. *Pechen A.N., Volovich I.V.* Quantum multipole noise and generalized quantum stochastic equations // *Quant. Prob. Rel. Top.* 2002. V. 5. № 4. P. 441–464.
9. *Pechen A.N.* On an asymptotic expansion in quantum theory // *Math. Notes.* 2004. V. 75. № 3. P. 426–429.
10. *Teretenkov A.E.* Non-Markovian Evolution of Multi-level System Interacting with Several Reservoirs. Exact and Approximate // *Lobachevskii J. Math.* 2019. V. 40. № 10. P. 1587–1605.
11. *Teretenkov A.E.* Exact Non-Markovian Evolution with Several Reservoirs, // *Physics of Particles and Nuclei.* 2020. V. 51. № 4. P. 479–484.
12. *Breuer H.-P., Laine E.M., Piilo J.* Measure for the degree of non-Markovian behavior of quantum processes in open systems // *Phys. Rev. Lett.* 2009. V. 103. № 21. P. 210401.
13. *Gullo N.L., Sinayskiy I., Busch T., Petruccione F.* Non-Markovianity criteria for open system dynamics // *arXiv:1401.1126*, 2014.
14. *Breuer H.-P., Petruccione F.* The theory of open quantum systems. Oxford: Oxford University Press, 2002.
15. *Rivas A., Huelga S.F., Plenio M.B.* Quantum non-Markovianity: characterization, quantification and detection // *Rep. Progr. in Phys.* 2014. V. 77. № 9. P. 094001.
16. *Bae J., Chruscinski D.* Operational characterization of divisibility of dynamical maps // *Phys. Rev. Lett.* 2016. V. 117. № 5. P. 050403.
17. *Haikka P., Cresser J.D., Maniscalco S.* Comparing different non-Markovianity measures in a driven qubit system // *Phys. Rev. A.* 2011. V. 83. № 1. P. 012112.
18. *Li L., Hall M.J.W., Wiseman H.M.* Concepts of quantum non-Markovianity: A hierarchy // *Phys. Rep.* 2018. V. 759. P. 1–51.
19. *Trushechkin A.S., Volovich I.V.* Perturbative treatment of inter-site couplings in the local description of open quantum networks // *EPL.* 2016. V. 113. № 3. P. 30005.
20. *Friedrichs K.O.* On the perturbation of continuous spectra // *Comm. on Pure and Applied Math.* 1948. V. 1. № 4. P. 361–406.
21. *Garraway B.M., Knight P.L.* Cavity modified quantum beats // *Phys. Rev. A.* 1996. V. 54. № 4. P. 3592.
22. *Garraway B.M.* Nonperturbative decay of an atomic system in a cavity // *Phys. Rev. A.* 1997. V. 55. № 3. P. 2290.
23. *Garraway B.M.* Decay of an atom coupled strongly to a reservoir // *Phys. Rev. A.* 1997. V. 55. № 6. P. 4636.
24. *Jang S., Cao J., Silbey R.J.* Fourth-order quantum master equation and its Markovian bath limit // *J. Chem. Phys.* 2002. V. 116. № 7. P. 2705–2717.
25. *Теретёнков А.Е.* Метод псевдомод и вибранные немарковские эффекты в светособирающих комплексах // *Труды МИАН.* 2019 Т. 306. С. 258–272.
26. *Dalton B.J., Barnett S.M., Garraway B.M.* Theory of pseudomodes in quantum optical processes // *Phys. Rev. A.* 2001. V. 64. № 5. P. 053813.
27. *Garraway B.M., Dalton B.J.* Theory of non-Markovian decay of a cascade atom in high-Q cavities and photonic band gap materials // *J. Phys. B: Atomic, Mol. Opt. Phys.* 2006. V. 39. № 15. P. S767.
28. *Luchnikov I.A., Vintskevich S.V., Ouerdane H., Filippov S.N.* Simulation complexity of open quantum dynamics: Connection with tensor networks // *Phys. Rev. Lett.* 2019. V. 122. № 16. P. 160401.
29. *Burgarth D., Facchi P., Ligabo M., Lonigro D.* Hidden non-Markovianity in open quantum systems // *Phys. Rev. A.* 2021. V. 103. № 1. P. 012203.

30. *Fleming C., Cummings N.I., Anastopoulos C., Hu B.L.* The rotating-wave approximation: consistency and applicability from an open quantum system analysis // *J. Phys.* 2010. V. 43. № 40. P. 405304.
31. *Tang N., Xu T.-T., Zeng H.-S.* Comparison between non-Markovian dynamics with and without rotating wave approximation // *Chinese Phys. B.* 2013. V. 22. № 3. P. 030304.
32. *Trubilko A.I., Basharov A.M.* Theory of relaxation and pumping of quantum oscillator non-resonantly coupled with the other oscillator // *Phys. Scr.* 2020. V. 95. № 4. P. 045106.
33. *Trushechkin A.S.* Higher-order corrections to the Redfield equation with respect to the system-bath coupling based on the hierarchical equations of motion // *Lobachevskii J. Math.* 2019. V. 40. № 10. P. 1606–1618.
34. *Трушечкин А.С.* Вывод квантового кинетического уравнения Редфилда и поправок к нему по методу Боголюбова // *Труды МИАН.* 2021. Т. 313. С. 263–274.
35. *Васильева А.Б., Бутузов В.Ф.* Асимптотические разложения решений сингулярно возмущенных уравнений. М.: Наука, 1973.
36. *Bender C.M., Orszag S.A.* Advanced mathematical methods for scientists and engineers. New York: McGraw-Hill, 1978.
37. *Lagerstrom P.A.* Matched asymptotic expansions: ideas and techniques. Vol. 76. New York: Springer, 1988.