ISSN 0044-4669

Том 64, Номер 6

Июнь 2024

Ж. ВЫЧИСЛ. МАТЕМ. И МАТЕМ. ФИЗ. • Том 64 • № 6 • 2024

ФИЦ ИУ РАН

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

НАУКА

1727

Российская академия наук
Федеральный исследовательский центр «Информатика и управление»
Российской академии наук

ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 64 Июнь № 6 2024

Выходит 12 раз в год
Основан в январе 1961 г.
академиком Анатолием Алексеевичем Дородницыным

ISSN: 0044-4669

Журнал издается под руководством
Отделения математических наук РАН

Адрес редакции: 119333 Москва,
ул. Вавилова, 44, корп. 2, ФИЦ ИУ РАН.
редакция “Журнала вычислительной математики и математической физики”
тел. 8-499-135-55-08
e-mail: comp_mat@ccas.ru

Москва
ФГБУ «Издательство «Наука»

СОДЕРЖАНИЕ

Том 64, номер 6, 2024 год

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

МАТЕМАТИЧЕСКАЯ ФИЗИКА

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 895–913

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.6

РАЦИОНАЛЬНАЯ АРИФМЕТИКА С ОКРУГЛЕНИЕМ

© 2024 г. В. П. Варин[1,*]

[1] *125047 Москва, Миусская пл., 4, Институт прикладной математики им. М. В. Келдыша РАН, Россия*

[*] *e-mail: varin@keldysh.ru*

Поступила в редакцию 16.01.2024 г.
Переработанный вариант 08.02.2024 г.
Принята к публикации 05.03.2024 г.

Вычисления на компьютере в плавающей арифметике всегда являются приближенными. Напротив, вычисления в рациональной арифметике (например, в компьютерной алгебре) всегда абсолютно точны и воспроизводимы как на других компьютерах, так и (теоретически) вручную. Поэтому такие вычисления могут быть доказательными в том смысле, что доказательство, полученное с их помощью, ничем не отличается от традиционного. Однако обычно такие вычисления в достаточно сложной задаче невозможны ввиду ограниченности ресурсов памяти и времени. Мы предлагаем механизм округления рациональных чисел при расчетах в рациональной арифметике, который решает эту проблему (ресурсов), т.е. вычисления по-прежнему могут быть доказательными, но уже не требуют неограниченных ресурсов. Приведен ряд примеров реализации стандартных численных алгоритмов в этой арифметике. Результаты имеют приложения к аналитической теории чисел. Библ. 22. Фиг. 3.

Ключевые слова: рациональная арифметика, подходящие дроби, доказательные вычисления, критерий иррациональности Бруна.

DOI: 10.31857/S0044466924060015, **EDN:** XZRESQ

1. ВВЕДЕНИЕ

Вычисления на компьютере в плавающей арифметике всегда являются приближенными, так же как и вычисления вручную с карандашом и бумагой. Однако помимо разной производительности этих вычислений между ними существует одно принципиальное отличие. Вычисления вручную контролирует сам вычислитель, т.е. человек определяет, сколько цифр сохранить в промежуточных и окончательных результатах и как при этом происходит округление. В компьютере же за эти операции отвечает, помимо прочего, тип данных, выбранных в программе, тип операционной системы, а также тип процессора, установленного на данном компьютере.

Хотя вся эта информация, в принципе, доступна пользователю, однако практическое ее использование затруднительно даже в одной элементарной операции типа $x = y + z$. Кроме того, на другом компьютере все это будет уже несколько по-другому.

Иными словами, вычисления на компьютере – это всегда обращение к “черному ящику”, “размер” которого зависит от желания и возможности пользователя изучать то, как конкретный компьютер работает на физическом уровне. В этом смысле вычисления на компьютере вполне аналогичны проведению физического эксперимента.

Вероятно, в том числе и эти соображения учитывались при создании математической дисциплины “экспериментальная математика” (см., например, [1]). Заметим также, что В. И. Арнольд полагал, что “математика – это экспериментальная наука” (см. [2]).

В то же время вычисления на компьютере могут быть абсолютно точными, в случае если это символьные вычисления в системе компьютерной алгебры (CAS). Символьные вычисления, проделанные с помощью компьютера, могут быть при этом чрезвычайно сложны и громоздки, однако они, в принципе, воспроизводимы в разных системах CAS и проверяемы вручную.

В этом смысле доказательство, полученное с помощью компьютера (computer assisted proof), ничем не отличается от традиционного и признается как таковое.

Хотя сам термин “доказательные вычисления” исторически появился значительно раньше, однако он не получил (и вряд ли получит) столь же широкое признание.

По нашему мнению, вычисления в плавающей арифметике в принципе не могут быть доказательными потому, что они не проверяемы вручную, так как человеку (как правило) неизвестен алгоритм, по которому округляются промежуточные вычисления и результат, а также потому, что на разных компьютерах это делается по-разному и дает (вообще говоря) различные результаты.

При этом с прикладной точки зрения результат может не вызывать ни малейших сомнений, как и результат физического эксперимента, однако математически это не будет доказательством.

Все это относится и к интервальной арифметике, которая одно время также претендовала на доказательность. Однако в настоящее время в контексте интервальной арифметики говорят, скорее, о достоверных вычислениях (см. [3]).

Таким образом, доказательность или бездоказательность вычислений зависят не только от качества алгоритмов и полученных решений, но и от введения в математику новых аксиом, которые будут относиться к компьютеру, и поэтому никогда не будут приняты большинством математиков.

Однако вычисления на компьютере могут быть доказательными в традиционном смысле, если они абсолютно точны, воспроизводимы на других компьютерах независимо от их архитектуры, и, главное, теоретически проверяемы вручную.

Мы привели символьные вычисления в CAS как один из примеров таких вычислений. Другой пример дают вычисления в рациональной арифметике (необязательно в CAS), которые эквивалентны вычислениям с целыми числами. Такие вычисления всегда осуществляются абсолютно точно на любом компьютере, а также всегда (теоретически) проверяемы вручную. Поэтому такие вычисления полностью укладываются в понятие “математически строгое доказательство”, если оно было получено с их помощью.

Одним из ярких примеров такого доказательства является проблема четырех красок, которая была решена с помощью вычислений на компьютере (см. [4]).

Разумеется, абсолютно точные вычисления на компьютере можно осуществлять не только в рациональной арифметике, но и, например, в квадратичных полях $\mathbb{Q}(\sqrt{n})$, $n \in \mathbb{Z}$. Такие вычисления применяются, в частности, в вычислительной геометрии (см. [5]).

При вычислениях в рациональной арифметике возникают те же проблемы с ресурсами памяти и времени, которые существуют при вычислениях в плавающей арифметике с расширенной разрядной сеткой. Теоретически вычисления в плавающей арифметике можно проводить с любым числом десятичных разрядов. Например, вполне реально вычислять с тысячами десятичных разрядов на домашнем компьютере. Однако практически ресурсы всегда ограничены.

Вычисления в рациональной арифметике, как правило, предъявляют еще бóльшие требования к ресурсам памяти и времени, чем вычисления в плавающей арифметике. К тому же, в отличие от последней, вычисления с рациональными числами значительно менее предсказуемы в плане требуемых ресурсов.

Например, итерационный алгоритм Айткена при вычислении константы Эйлера—Гомпертца уже на 10-м шаге дает приближение этой константы с погрешностью менее $1.2 \cdot 10^{-10}$. Однако суммарное количество цифр в полученном рациональном приближении равно 180980 (см. [6]).

Один из способов если не решить, то контролировать эту проблему в CAS — использование модулярной арифметики, которая широко применяется, например, в базисах Грёбнера и в алгоритмах факторизации полиномов с целыми коэффициентами.

Модулярная арифметика в таких вычислениях (с целыми числами) является аналогом вычислений в плавающей арифметике с фиксированной разрядной сеткой. Однако, в отличие от последней, вычисления в модулярной арифметике по-прежнему абсолютно точны и воспроизводимы как на других компьютерах, так и (теоретически) человеком вручную, и поэтому являются доказательными в указанном выше смысле.

Из сказанного выше следует, что сделать вычисления доказательными, просто заменив плавающую арифметику на рациональную, не получится, в силу того, что вычисления проводятся на реальном, а не на идеальном компьютере с неограниченными ресурсами.

В данной статье предлагается механизм округления рациональных чисел, который позволяет использовать рациональную арифметику в обычных вычислениях, например, для численного интегрирования ОДУ, в задачах линейной алгебры и т.п.

Вычисления в рациональной арифметике на компьютере (или вручную) эквивалентны вычислениям с целыми числами, поэтому "округленный" результат, если он получается детерминистским и предсказуемым образом, всегда (теоретически) проверяем вручную.

В качестве такого механизма округления рационального числа r (или вещественного числа x) предлагается использовать обыкновенные подходящие дроби этого числа, которые мы обозначаем через $K(r, n)$, $n \in \mathbb{N}$, или:

$$K(x, n) = a_0 + \frac{1}{a_1+}\ \frac{1}{a_2+}\ \ldots\ \frac{1}{a_n} = [a_0;\, a_1, \ldots, a_n], \quad x \in \mathbb{R}.$$

Уже сама эта запись является аналогом записи числа в виде q-адичной дроби с фиксированной разрядной сеткой. Однако здесь "цифры" a_k, $k \in \mathbb{N}$ – это натуральные числа для $k \geq 1$ и $a_0 \in \mathbb{Z}$.

В учебниках, где рассматриваются свойства цепных дробей, насколько нам известно, никогда не отмечается отличие вычислений с рациональными числами от вычислений с вещественными числами в плавающей арифметике. Между тем, если вычисления проводятся на реальном, а не на идеальном компьютере с неограниченной разрядной сеткой, алгоритм Евклида работает принципиально отличным образом для этих арифметик.

Алгоритм Евклида вычисления обыкновенной подходящей дроби рационального числа (в отличие от числа, заданного в плавающей арифметике) всегда корректно определен и всегда дает один и тот же результат на любом компьютере независимо от его архитектуры (или вручную).

Таким образом, приближенные вычисления в рациональных числах можно рассматривать, в определенном смысле, как абсолютно точные. Поэтому такие вычисления, по нашему мнению, укладываются в концепцию доказательных вычислений.

В разд. 2 мы приводим некоторые свойства функций $K(x, n)$, $x \in \mathbb{R}$, включая их аппроксимационные возможности. Оказывается, что функции $K(x, n)$, $x \in [0, 1]$, весьма быстро стремятся к x при $n \to \infty$, причем максимальное отклонение $|x_n - K(x_n, n)|$ достигается в рациональных x_n, которые стремятся к золотому сечению.

В разд. 3 приводятся примеры численных алгоритмов, реализованных с помощью этой арифметики. Показано, что такие вычисления могут давать хорошие рациональные приближения нужных величин при использовании стандартных численных методов.

В разд. 4 рассматриваются некоторые приложения этой арифметики к аналитической теории чисел.

2. ЦЕПНАЯ ДРОБЬ $K(x, n)$ КАК ФУНКЦИЯ СВОИХ АРГУМЕНТОВ

Напомним, что функции $K(x, n)$ мы определили в [6]. Они использовались для построения рациональных аппроксимаций некоторых фундаментальных постоянных (см. [6, 7]).

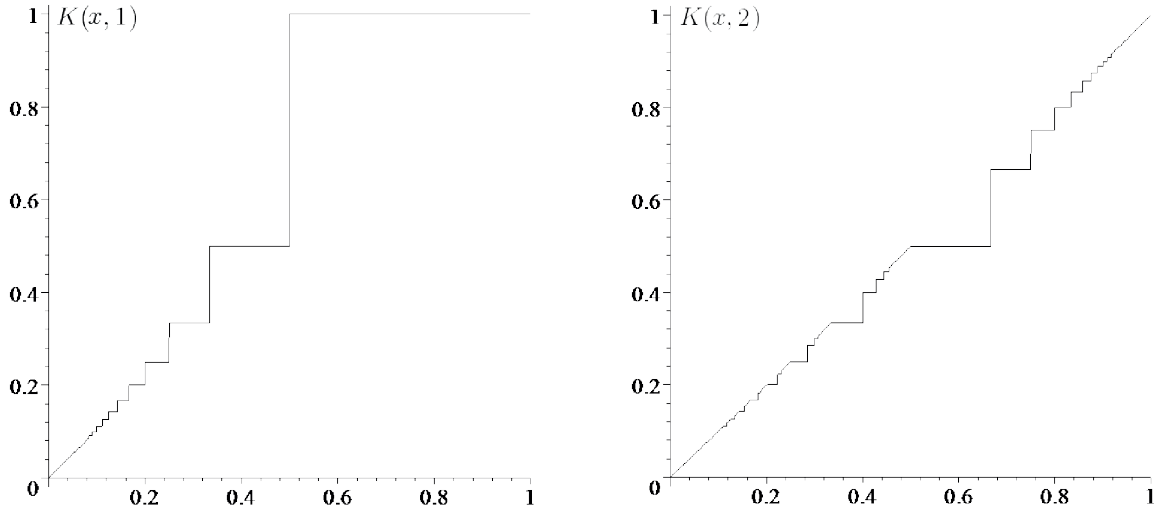
Хотя аппроксимационные свойства обыкновенных цепных дробей хорошо изучены для фиксированных x (см., например, [8]), свойства цепных дробей как функций аргумента x и порядка дроби n, насколько нам известно, нигде не изучались.

Функции $P(x, n) = K(x, n) - x$ являются, очевидно, 1-периодическими, причем

$$P(x, 2n) \leq 0, \quad P(x, 2n - 1) \geq 0,$$

по известному свойству подходящих дробей. Поэтому достаточно изучать эти функции на интервале $[0, 1]$.

На фиг. 1 приведены графики функций $K(x,1)$ и $K(x,2)$ на интервале $[0,1]$, где вертикальные отрезки – это чисто технический артефакт (удалять их было бы слишком накладно). Из фиг. 1 видно, что график функции $K(x,2)$ на отрезках $[1/2,1]$, $[1/3,1/2]$, $[1/4,1/3]$, … повторяет масштабированный график функции $K(x,1)$. Так же ведут себя графики функций $K(x,n)$, $n>2$. Таким образом, графики этих функций имеют весьма сложную квазифрактальную структуру.



Фиг. 1. Графики функций $K(x,1)$ и $K(x,2)$.

Поведение функций $K(x,n)$ на интервале $x\in[0,1]$ описывает следующее (очевидное)

Предложение 1. *Все точки разрыва функции $K(x,n)$, $n\in\mathbb{N}$, $x\in[0,1]$, даются формулой*

$$x_{k_1,\dots,k_n}=[k_1,k_2,\dots,k_n],\quad k_2,\dots,k_n\in\mathbb{N}_0,\ k_1\in\mathbb{N}. \tag{1}$$

При этом $K(x,n)=x_{k_1,\dots,k_n}$ для x, расположенных между величинами $[k_1,k_2,\dots,k_n]$ и $[k_1,k_2,\dots,k_n+t]$, $0\le t<1$, $k_1,\dots,k_n\in\mathbb{N}$.

Заметим, что выбор $k_j=0$, $j\ne 1$, в (1) означает просто, что особенности функций $K(x,m)$ наследуются функцией $K(x,n)$, $m<n$, что видно также из очевидного тождества

$$K(x,m)=K(K(x,m),n),\quad m,n\in\mathbb{N},\ m\le n.$$

Последовательности подходящих дробей $K(x,2n)$ и $K(x,2n-1)$, $n\in\mathbb{N}$, являются, соответственно, возрастающей и убывающей, т.е.

$$\dots\le K(x,2n-2)\le K(x,2n)\le x\le K(x,2n-1)\le K(x,2n-3)\le\dots,$$

причем равенства начиная с номера m здесь достигаются только для рациональных x, которые представимы в виде обыкновенной цепной дроби порядка $m\le 2n$.

Таким образом, у нас всегда имеются верхняя и нижняя оценки округляемого числа, что делает этот механизм округления схожим с интервальной арифметикой.

Другим преимуществом такого округления, т.е. замены (рационального) числа x на его подходящую дробь $K(x,n)$, является оптимальность этого представления. Как известно (см. [8]), дробь $K(x,n)$ приближает число x наилучшим образом в множестве рациональных чисел, знаменатели которых не превосходят знаменателя дроби $K(x,n)$ (а она всегда неприводима).

Пользуясь тождеством $x=1/([1/x]+\{1/x\})$, любую функцию $K(x,n)$, $0\le x\le 1$, можно выразить явно в виде композиций функций целой и дробной частей числа x. Однако это представление не имеет практического значения, так как реально вычисления проводятся по алгоритму Евклида.

Предложение 1 дает возможность записать интегралы от функций $K(x,n)$ на интервале $[0,1]$ в виде мультисумм с индексами $\{k_1,\dots,k_n\}$. Однако в явном виде такая сумма имеется только для $K(x,1)$:

$$\int_0^1 K(x,1)\,dx = \int_0^1 1/[1/x]\,dx = \frac{\pi^2}{6} - 1.$$

Согласно предложению 1, скачок, например, функции $K(x,2)$ в произвольной точке ее разрыва равен

$$\frac{1}{n+\dfrac{1}{m+1}} - \frac{1}{n+\dfrac{1}{m}} = \frac{1}{(n\,m+n+1)\,(n\,m+1)}, \quad m,n\in\mathbb{N},$$

т.е. m и n попадают в знаменатель, а числитель равен единице. Скачок любой функции $K(x,n)$, $n\in\mathbb{N}$, будет выражаться похожим образом по известному свойству подходящих дробей.

Предложение 2. *Наибольший скачок функции $K(x,n)$, $n\in\mathbb{N}$, $x\in[0,1]$, достигается в точке*

$$r_n = \frac{1}{1+}\,\frac{1}{1+}\,\cdots\,\frac{1}{1} = \mathop{\mathbf{K}}_{k=1}^{n+1}\frac{1}{1}.$$

Доказательство. Как известно, числители p_n и знаменатели q_n подходящих дробей удовлетворяют рекуррентным соотношениям

$$\begin{aligned} p_n &= k_n\,p_{n-1} + p_{n-2}, \quad p_{-1}=1,\ p_0=0,\\ q_n &= k_n\,q_{n-1} + q_{n-2}, \quad q_{-1}=0,\ q_0=1. \end{aligned} \tag{2}$$

Скачок функции $K(x,n)$ можно записать в виде

$$\frac{(k_n+1)\,p_{n-1}+p_{n-2}}{(k_n+1)\,q_{n-1}+q_{n-2}} - \frac{k_n\,p_{n-1}+p_{n-2}}{k_n\,q_{n-1}+q_{n-2}} = \frac{p_{n-1}\,q_{n-2}-p_{n-2}\,q_{n-1}}{((k_n+1)\,q_{n-1}+q_{n-2})\,(k_n\,q_{n-1}+q_{n-2})}.$$

Поэтому (см. теорему 2 в [8]) числитель последней дроби всегда равен ±1 согласно тождеству

$$\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n q_{n+1}}. \tag{3}$$

Знаменатель дроби, выражающей скачок функции $K(x,n)$, очевидно, достигает минимального значения при $k_1=k_2=\cdots=k_n=1$. Что и требовалось доказать.

Таким образом, наибольший скачок функций $K(x,n)$ достигается в дробях

$$\{r_n,\ n\in\mathbb{N}\} = \left\{\frac{1}{2},\frac{2}{3},\frac{3}{5},\frac{5}{8},\frac{8}{13},\frac{13}{21},\frac{21}{34},\frac{34}{55},\dots\right\},\ r_n\to\frac{1}{2}\,(\sqrt{5}-1)\approx 0.61803,$$

где числители и знаменатели — это не что иное, как числа Фибоначчи, F_n (см. [9]), так как формула (2) дает именно эти числа при $k_n=1$. Величина скачка при этом равна

$$\delta_n = \mathop{\mathbf{K}}_{k=1}^{n}\frac{1}{1} - \mathop{\mathbf{K}}_{k=1}^{n-1}\frac{1}{1+}\,\frac{1}{2} = \mathop{\mathbf{K}}_{k=1}^{n}\frac{1}{1} - \mathop{\mathbf{K}}_{k=1}^{n+1}\frac{1}{1} = \frac{F_n}{F_{n+1}} - \frac{F_{n+1}}{F_{n+2}}, \quad n\in\mathbb{N}.$$

По известному свойству чисел Фибоначчи, которое, впрочем, следует из формулы (3), получаем

$$\left\{\delta_n = \frac{(-1)^{n+1}}{F_{n+1}\,F_{n+2}},\ n\in\mathbb{N}\right\} = \left\{\frac{1}{2},-\frac{1}{6},\frac{1}{15},-\frac{1}{40},\frac{1}{104},-\frac{1}{273},\dots\right\}.$$

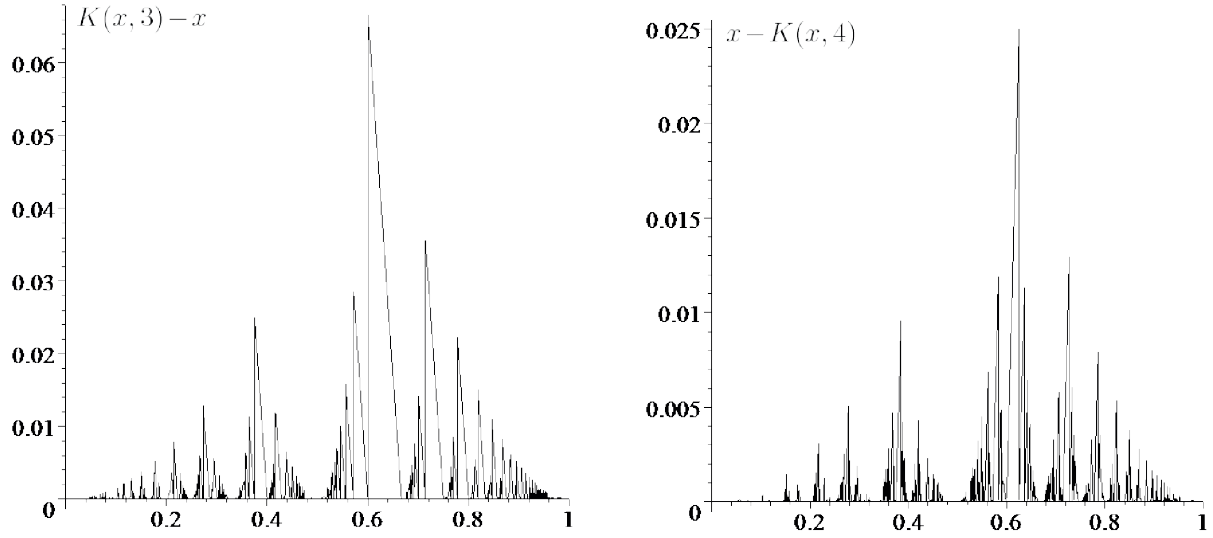
Величина $|\delta_n|$, очевидно, является максимально возможной погрешностью приближения числа $x\in[0,1]$ дробью $K(x,n)$. Например, при $n=40$ максимальная погрешность равна

$$|\delta_{40}| = \frac{1}{44361286907595736} \approx 0.225421\times10^{-16}.$$

Реальная погрешность, разумеется, может быть значительно меньше.

Таким образом, мы подтвердили известный факт: золотое сечение (или его сдвиги на целое число) приближается рациональными числами наихудшим образом.

На фиг. 2 приведены графики функций $K(x,3)-x$ и $x-K(x,4)$ на интервале $[0,1]$ (где вертикальные отрезки — это технический артефакт). Из этих рисунков (а также других для $n>4$) видно, что существуют промежутки, где вещественные числа приближаются цепными дробями в численном смысле значительно лучше, чем в других интервалах отрезка $[0,1]$. Эти "благоприятные" для аппроксимации промежутки расположены вблизи точек $1/2,\ 1/3,\ 1/4,\ldots$, а также вблизи точек $2/3,\ 3/4,\ 4/5,\ldots$.



Фиг. 2. Графики функций $K(x,3)-x$ и $x-K(x,4)$.

Этот экспериментальный факт, возможно, объясняет, почему дробно-линейное преобразование с рациональными коэффициентами иррационального числа может радикальным образом изменить скорость сходимости обыкновенной цепной дроби новой константы по сравнению с исходной (см. [7]).

В книге [8, p. 28–29] Хинчин сравнивает преимущества и недостатки представления вещественных чисел обыкновенными цепными дробями с их представлением обычными (q-адичными в современной терминологии) дробями. Согласно Хинчину, с теоретической точки зрения цепные дроби имеют явные и значительные преимущества.

С практической точки зрения цепные дроби также имеют ряд преимуществ. Хинчин отмечает лишь один, но существенный их недостаток — неудобство проведения арифметических операций с этими дробями. Однако следует вспомнить, что эта книга вышла в 1935 году, когда арифметические операции могли выполняться только вручную. При вычислениях на современных компьютерах такие ограничения во многом становятся неактуальными.

Кроме того, мы не предлагаем выполнять арифметические операции с цепными дробями. Операции выполняются исключительно с рациональными числами, т.е. с парами целых чисел. Цепные дроби нужного порядка служат только для целей округления и вычисляются только для (некоторых) промежуточных результатов. Иными словами, вычисления в приближенной рациональной арифметике не требуют неограниченных ресурсов компьютера, как это было бы при вычислениях в обычной рациональной арифметике, но по-прежнему являются абсолютно точными в указанном выше смысле.

В следующих разделах мы приведем ряд примеров реализации некоторых классических алгоритмов в этой арифметике и результаты их работы.

3. ФУНКЦИИ $K(x, n)$ И ЧИСЛЕННЫЕ МЕТОДЫ

Напомним, что символ $K(x, n)$ мы использовали как для обозначения соответствующей цепной дроби, вычисляемой по алгоритму Евклида, так и для обозначения рационального числа, выражаемого этой дробью. Далее символ $K(x, n)$ понимается только в этом последнем смысле.

Функцию $K(x, n)$ мы вычисляли, в основном, по встроенному алгоритму в CAS Maple. Однако для рациональных чисел, которые мы далее используем, алгоритм Евклида легко (и столь же эффективно) реализуется как в системах CAS, так и в обычных языках программирования.

В частности, в C++ версии 11 и выше существует встроенный тип данных, предназначенный для вычислений в рациональной арифметике. Добавление операции округления, определенной выше, позволяет использовать эту арифметику для (практически) любых численных расчетов.

Приведем два примера реализации численных алгоритмов в округленной рациональной арифметике.

Рассмотрим сначала классический алгоритм Рунге—Кутты 4-го порядка интегрирования ОДУ

$$y'(x) = f(x,\ y(x)) \tag{4}$$

на конечном интервале. В современных обозначениях (см. [10, p. 134-135]) один шаг метода Рунге—Кутты имеет вид

$$y_k = y_0 + h \sum_{i=0}^{k-1} a_{k,i}\, f(x_0 + h\, c_i,\ y_i), \quad k = 1, 2, 3,$$

затем

$$y_h = y_0 + h \sum_{i=0}^{3} b_i\, f(x_0 + h\, c_i,\ y_i), \tag{5}$$

где x_0, y_0 — это данные Коши (начальные значения), а y_h — это приближенное значение решения $y(x)$ в точке $x = x_0 + h$.

Затем следует положить $x_0 := x_0 + h$, $y_0 := y_h$ и сделать следующий шаг, и т.д. В частности, для начального значения $x_0 = 0$, N шагов данного алгоритма с шагом $h = 1/N$ проинтегрируют уравнение (4) на интервале $[0, 1]$.

Коэффициенты

$$\{a_{k,i},\ b_j,\ c_j,\ i = 0, 1, 2,\ j = 0, 1, 2, 3,\ k = 1, 2, 3\}$$

выбираются таким образом, что

$$|y(x_0 + h) - y_h| \le C\, h^5, \quad C = \mathrm{const}(x_0, f),$$

т.е. тейлоровское разложение погрешности в точке $x = x_0 + h$ начинается с члена h^5.

Набор коэффициентов a, b, c может быть выбран неединственным образом. Классический выбор:

$$\begin{array}{lll} a_{1,0} = 1/2, & a_{1,1} = 0, & a_{1,2} = 0, \\ a_{2,0} = 0, & a_{2,1} = 1/2, & a_{2,2} = 0, \\ a_{3,0} = 0, & a_{3,1} = 0, & a_{3,2} = 1; \end{array}$$

$$b_0 = \frac{1}{6},\ b_1 = \frac{1}{3},\ b_2 = \frac{1}{3},\ b_3 = \frac{1}{6}; \quad c_0 = 0,\ c_1 = \frac{1}{2},\ c_2 = \frac{1}{2},\ c_3 = 1.$$

Насколько нам известно, методы Рунге—Кутты никогда не применялись в рациональной арифметике по понятной причине: эти вычисления (даже для полиномиальной правой части (4), не говоря уже о рациональной) быстро приведут к неконтролируемому росту объема промежуточных результатов.

Рассмотрим, например, задачу Коши $y(0) = 0$ для ОДУ

$$y'(x) = \frac{1}{1 + x^2}$$

и проинтегрируем его данным выше методом на интервале [0, 1] в рациональной арифметике с шагом $h = 1/10$. Получаем приближение

$$y(1) = \frac{5988585315838311774901484536676836463}{76249036426504635203016941416552830 00}, \quad y(1) - \frac{\pi}{4} \approx -1.550 \cdot 10^{-10}.$$

Ясно, что вычисления в рациональной арифметике по этому алгоритму при измельчении шага быстро становятся нереализуемыми.

Рассмотрим теперь модификацию алгоритма Рунге—Кутты, в которой последняя операция на каждом шаге (5) заменяется на операцию

$$y_h = K(y_0 + h \sum_{i=0}^{3} b_i \, f(x_0 + h\, c_i,\, y_i), M),$$

где M — это параметр алгоритма, т.е. порядок цепной дроби, который можно менять. Тогда с тем же шагом для $M = 18$ получим аппроксимацию

$$y(1) = \frac{77072475}{98131723}, \quad y(1) - \frac{\pi}{4} \approx -1.551 \cdot 10^{-10},$$

т.е. практически с той же точностью, но уже без неконтролируемого роста объема промежуточных величин.

Рассмотрим итерационный алгоритм Айткена ускорения сходимости знакопеременного ряда (или осциллирующей последовательности) (см. [11]). В отличие от классического алгоритма Айткена, который точен на геометрически сходящихся последовательностях, итерационный алгоритм является весьма эффективным ускорителем сходимости, применимым также к расходящимся рядам. Он дается рекуррентным соотношением

$$A(n, 0) = s(n),$$
$$A(n, k+1) = A(n, k) - \frac{(A(n+1, k) - A(n, k))^2}{A(n+2, k) - 2\,A(n+1, k) + A(n, k)}, \quad k \geq 0.$$

Последовательность $\{A(n, 1),\ n = 1, 2, \dots\}$ соответствует обычному методу Айткена.

Об асимптотических свойствах итерационного алгоритма Айткена известно немного (см. ссылки в [11]). Применим этот алгоритм для суммирования расходящегося ряда, дающего константу Эйлера—Гомпертца в качестве своей обобщенной суммы.

Осциллирующая последовательность $s(n)$ для этой константы имеет вид

$$s(n) = \sum_{k=0}^{n} (-1)^k\, k!,$$

т.е. последовательность быстро расходится. Вычисление (обобщенной) суммы этого ряда, $\delta = \mathbf{e}\,\mathrm{Ei}(\mathbf{1}, \mathbf{1})$ (где $\mathrm{Ei}()$ — это интегральная экспонента), является классической задачей, впервые решенной Эйлером.

Так называемая диагональная (так как подобные итерационные схемы традиционно изображают в виде ромба) последовательность $\{A(0, n),\ n = 1, 2, \dots\}$ весьма быстро сходится к δ, но только для небольших значений n. При $n > 10$ точность стабилизируется, а затем начинает убывать. При этом не вполне ясно, является ли этот эффект свойством алгоритма как такового или результатом накопления ошибок.

Численная неустойчивость итерационного алгоритма Айткена в плавающей арифметике ранее уже отмечалась в литературе (см. [11]). Например, расчеты в обычной D-арифметике (≈ 16 десятичных разрядов) дают для константы δ точность приближения $\approx 9.2 \cdot 10^{-11}$ на 10-м шаге. Далее точность падает, а затем процесс быстро расходится.

Расчеты в QD-арифметике (≈ 64 десятичных разряда) дают аналогичную точность. Различие состоит в том, что достигнутая точность стабилизируется на более длинном отрезке последовательности итераций, но затем также начинает убывать.

Вычисления в рациональной арифметике по алгоритму Айткена могли бы прояснить вопрос о влиянии погрешностей округления на свойства его сходимости (или расходимости). Однако такие вычисления, очевидно, невозможны (см. Введение). Поэтому применим рациональное округление в этом алгоритме, т.е. вместо $A(n, k)$ будем вычислять (и хранить)

$$\tilde{A}(n, k) = K(A(n, k), M(k)),$$

где $M(k) \in \mathbb{N}$ — это некоторая функция, которую можно выбирать различным образом.

Вычислительные эксперименты показали слабую зависимость точности вычисления константы δ от способа округления (т.е. от выбора функции $M(k)$). В частности, для $M(k) = k + 10$ достигнута точность приближения $\approx 5.8 \cdot 10^{-12}$ на 18-м шаге.

Таким образом, с большой вероятностью итерационный алгоритм Айткена не сходится к константе δ, а генерирует некоторый асимптотический расходящийся процесс.

4. ВЫЧИСЛЕНИЯ С ЦЕПНЫМИ ДРОБЯМИ

Отличие вычислений в рациональной арифметике от вычислений в плавающей арифметике, по-видимому, нигде не проявляется более отчетливо, чем при вычислении обыкновенных цепных дробей.

В рациональной арифметике все вычисления сводятся к манипуляциям с целыми числами и поэтому всегда выполняются абсолютно точно. Компьютер при этом является лишь инструментом, ускоряющим работу.

Напротив, вычисление цепной дроби в плавающей арифметике требует обращения к функциям вычисления целой и дробной части числа, $[x]$ и $\{x\}$, и поэтому всегда опирается на аппаратную реализацию приближенных операций плавающей арифметики.

Разумеется, вычисления в плавающей арифметике с расширенной разрядной сеткой, необходимые для получения большого количества подходящих дробей, имеют программную, а не аппаратную реализацию. Но проблемы недетерминистского округления при этом не исчезают, а лишь отодвигаются к более поздним стадиям работы алгоритма.

Обыкновенные цепные дроби трансцендентных констант обычно вычисляются именно с использованием плавающей арифметики с расширенной разрядной сеткой (см., например, [12]). Тем более удивительным представляется факт, что для вычисления подходящих дробей алгебраических иррациональностей достаточно использовать только вычисления с целыми числами.

Алгоритм, который мы представим, не нов и принадлежит Лагранжу (см. [13, p. 560]). С тех пор он неоднократно использовался многими авторами (см., например, [14]), однако факт вычислений только с целыми числами, по-видимому, ранее не отмечался, кроме (неявно) в [15, p. 375].

Пусть $P(x)$ — это полином с целочисленными коэффициентами степени d, имеющий единственный вещественный иррациональный корень $\alpha > 0$, который и разлагается в цепную дробь. Сам корень при этом неизвестен.

На самом деле, вещественных корней у полинома $P(x)$ может быть много, но ищется положительный корень ближайший к нулю, причем следующий положительный корень должен отстоять от α более чем на единицу.

Введем комплексные величины $\{r_n,\ n = -2, -1, 0, 1, \dots\}$ (*гауссовы целые*), которые кодируют подходящие дроби p_n/q_n числа α,

$$r_{-2} = 1,\ r_{-1} = i, \quad r_n = q_n + i\,p_n.$$

Смысл данного обозначения прояснится позднее.

Целые величины $\{a_n,\ n = 0, 1, 2, \dots\}$ — это частные знаменатели разложения

$$\alpha = [a_0; a_1, a_2, \dots], \quad 0 \le a_0.$$

Дадим сначала готовую процедуру в CAS Maple, а затем необходимые пояснения:

```
cf:=proc(P,N) local d,f,n,k,j,s,sk: global a,r,x:
f:=P: d:=degree(f,x):
r[-2]:=1: r[-1]:=I:
j:=n->‘if‘(n=0,0,1):
for n from 0 to N do
s:=sign(eval(f,x=j(n))):
for k from j(n)+1 do
sk:=sign(eval(f,x=k)):
if s<>sk then a[n]:=k-1: break fi
od:
f:=expand(x^d*subs(x=a[n]+1/x,f)):
r[n]:=a[n]*r[n-1]+r[n-2]:
od
end:
```

Структура алгоритма очевидна. Для определения следующего частного знаменателя a_n вычисляются значения полинома с целочисленными коэффициентами, полученного на предыдущем шаге, в точках $k = 1, 2, \ldots$ до тех пор, пока не встретится первое отличие в знаках. Это означает, что корень этого полинома больше, чем k, т.е. $a_n = k - 1$ и т.д.

Например, для полинома $P = x^3 - 2$ вычисление 5000 подходящих дробей числа $2^{1/3}$ на нашем компьютере занимает менее 0.6 сек., а 10000 — менее 1.28 сек. Встроенная процедура Maple вычисляет 10000 частных знаменателей разложения числа $2^{1/3}$ за 8.1 сек. Выборка из первых 10000 частных знаменателей содержит следующие большие знаменатели:

$$a_{35} = 534, \quad a_{571} = 7451, \quad a_{619} = 4941, \quad a_{1990} = 12737, \quad a_{2247} = 2897, \ \ldots,$$

причем $a_{1990} = 12737$ является самым большим знаменателем из этой выборки.

Численные эксперименты показывают, что частные знаменатели подходящих дробей алгебраических иррациональностей степени больше двух неограничены. Хотя, по-видимому, не существует доказательства того, что это свойство выполняется у какой-либо конкретной иррациональности этого вида. Вообще, существует крайне мало информации о степени роста частных знаменателей конкретных алгебраических иррациональностей (см. [16]).

Необычно большие знаменатели в разложении алгебраической иррациональности в цепную дробь являются слабым местом представленного алгоритма. Это видно из строчки “`for k from j(n)+1 do`”, после которой вычисляются значения полинома на шаге n в точках $k = 1, 2, \ldots$ до смены знака. В интернете можно найти способы обойти эту проблему с использованием алгоритмов бинарного поиска, т.е., например, использовать перебор $k = 2^m$, $m = 0, 1, 2, \ldots$, и т.д.

Существует алгоритм, позволяющий вычислить следующую подходящую дробь алгебраической иррациональности по двум предыдущим, если значение приближаемого числа известно. Этот алгоритм опирается на свойства *промежуточных дробей* (см. [8]). А именно: дроби

$$\frac{p_{n-2}}{q_{n-2}}, \frac{p_{n-2} + p_{n-1}}{q_{n-2} + q_{n-1}}, \frac{p_{n-2} + 2\,p_{n-1}}{q_{n-2} + 2\,q_{n-1}}, \ldots \frac{p_{n-2} + a_n\,p_{n-1}}{q_{n-2} + a_n\,q_{n-1}} = \frac{p_n}{q_n}$$

образуют при четном n возрастающую, а при нечетном n — убывающую последовательность, расположенную по одну сторону от приближаемого числа. Поэтому если значение приближаемого числа известно, то $k = a_n$ нумерует последний член этой последовательности. Следующий будет лежать уже по другую сторону от приближаемого числа. Например, для $\alpha = 2^{1/3}$ проводится сравнение

$$(p_{n-2} + k\,p_{n-1})^3 <> 2\,(q_{n-2} + k\,q_{n-1})^3, \quad k = 1, 2, \ldots,$$

в зависимости от четности n.

Этот алгоритм (для $\alpha = 2^{1/3}$) также использует только вычисления в целых числах, но имеет тот же недостаток, что и предыдущий алгоритм, а именно: большое число проверок для больших частных знаменателей.

Независимо от того, известен корень или нет, эти проверки можно не делать, если иметь хорошее приближение следующей подходящей дроби, p_n/q_n. Для этого можно использовать метод Ньютона вычисления корня полинома. А именно: пусть $N(x)$ обозначает ньютоновскую итерацию для данного полинома, т.е.

$$N(x) = x - \frac{P(x)}{P'(x)},$$

где x — это приближенное значение корня α, $P(\alpha) = 0$.

Пусть $h = x - \alpha$, тогда

$$N(x) - \alpha = \frac{1}{2}\,\frac{P''(\alpha)}{P'(\alpha)}\,h^2 + O(h^3),$$

или

$$|\alpha - N(x)| < C\,h^2 \tag{6}$$

для некоторой константы C, которая зависит только от полинома $P(x)$ и его корня α (квадратичная сходимость метода Ньютона).

Выберем в качестве приближения α последнюю найденную подходящую дробь, $x = \frac{p_{n-1}}{q_{n-1}}$, и найдем корень k уравнения

$$\frac{p_{n-2} + k\,p_{n-1}}{q_{n-2} + k\,q_{n-1}} = N\left(\frac{p_{n-1}}{q_{n-1}}\right) \approx \alpha$$

в рациональной арифметике. Затем положим $a_n = [k]$, т.е. вычислим целую часть рационального числа k. Здесь мы используем тот факт, что p_{n-1}/q_{n-1} и p_n/q_n лежат по разные стороны от α. Поэтому целое $[k]$ и рациональное k также с большой вероятностью лежат по разные стороны от α.

На самом деле, дробь p_n/q_n почти всегда вычисляется точно этим способом, если использовать функции $K(x, n)$. Приведем сначала

Предложение 3. *Пусть дано произвольное иррациональное число $x > 0$ и его разложение в обыкновенную цепную дробь. И пусть рациональное число y находится между двумя подходящими дробями числа x, p_n/q_n и p_{n+1}/q_{n+1}, $n \in \mathbb{N}$. Тогда подходящие дроби числа y совпадают с таковыми для числа x вплоть до n-й включительно.*

Доказательство очевидно (см. [6]). Поэтому справедливо следующее

Предложение 4. *Пусть дан полином с целыми коэффициентами $P(x)$, имеющий иррациональный корень $\alpha > 0$, $P'(\alpha) \neq 0$. И пусть вычислена подходящая дробь p_{n-1}/q_{n-1} числа α с достаточно большим номером n. Тогда последующие подходящие дроби вычисляются по формуле*

$$\frac{p_n}{q_n} = K\left(N\left(\frac{p_{n-1}}{q_{n-1}}\right), n\right) \tag{7}$$

при условии, что частные знаменатели a_{n+1} или a_{n+2} не окажутся слишком большими. Например, достаточно, чтобы выполнялось условие

$$\max(a_{n+1}^3, a_{n+2}^3) < \text{const}\, q_{n-1}^2. \tag{8}$$

Доказательство. Согласно предложению 3, достаточно показать, что рациональное число $x_n = N(p_{n-1}/q_{n-1})$ лежит между числами p_n/q_n и p_{n+1}/q_{n+1}.

Подходящие дроби p_n/q_n и p_{n-1}/q_{n-1} всегда лежат по разные стороны от числа α. Всего существует четыре случая: n четно или нечетно; и функция $P(x)$ локально выпукла или вогнута в окрестности α.

В зависимости от случая достаточно, чтобы

$$|\alpha - x_n| < \left|\alpha - \frac{p_n}{q_n}\right| \quad \text{или} \quad |\alpha - x_n| < \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right|. \tag{9}$$

Но второе из этих неравенств влечет первое, поэтому рассмотрим второе.

Пусть

$$h = \left|\alpha - \frac{p_{n-1}}{q_{n-1}}\right| < \frac{1}{q_{n-1}\,q_n}. \tag{10}$$

Данное неравенство — это теорема 9 в [8]. Согласно теореме 13 в [8], имеем

$$\frac{1}{q_{n+1}\,(q_{n+1} + q_{n+2})} < \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right|.$$

Поэтому достаточно показать, что

$$|\alpha - x_n| < \frac{1}{q_{n+1}\,(q_{n+1} + q_{n+2})}.$$

Используя оценки (6) и (10), получаем неравенство

$$C < \frac{q_{n-1}^2\,q_n^2}{q_{n+1}\,(q_{n+1} + q_{n+2})},$$

достаточное для выполнения (7). Заменяя q_m по их рекуррентным формулам и загрубляя неравенство, получим (8). Что требовалось доказать.

Заметим, что для выполнения первого из неравенств (9) достаточно выполнения неравенства

$$a_{n+1} < \operatorname{const} q_{n-1}^2. \tag{11}$$

Нарушение любого из неравенств (8) или (11) означало бы появление катастрофически большого частного знаменателя у алгебраической иррациональности. Это маловероятно даже в единичных случаях, а нарушение этих условий бесконечное число раз возможно, по-видимому, только у трансцендентных чисел.

Об этом же свидетельствует теорема Рота (см. [17]), которая ограничивает скорость стремления подходящих дробей алгебраической иррациональности к своему пределу.

Например (см. [6]), положим $x_0 = 1$ и

$$x_{n+1} = K\left(x_n - \frac{x_n^3 - 2}{3\,x_n^2}, n+1\right), \quad n \in \mathbb{N}_0.$$

Тогда последовательность $\{x_n\}$ — это все подходящие дроби числа $\sqrt[3]{2}$. Во всяком случае, это справедливо для $n \le 10000$.

Алгоритм вычисления подходящих дробей алгебраических иррациональностей, изложенный в предложении 4, представляется заведомо менее оптимальным, чем алгоритм

Лагранжа, изложенный выше. Однако это, вообще говоря, не так. Рассмотрим уравнение (см. [18])

$$P(x) = x^3 - 8x - 10, \quad P(\alpha) = 0, \quad \alpha \approx 3.318628217750,$$

корень которого имеет разложение в обыкновенную цепную дробь с экстремально большими частными знаменателями, которые встречаются среди первых 200 членов разложения α в цепную дробь. Например,

$$a_{17} = 22986, \quad a_{33} = 1501790, \quad a_{121} = 16467250, \quad a_{161} = 325927.$$

В [18] показано, что это явление связано с тем, что данное уравнение имеет отношение к квадратичному полю $\mathbb{Q}(\sqrt{-163})$ и что порядок класса этого поля равен единице. Такой пример, разумеется, не единственный (см. [18]).

Алгоритм Лагранжа считает на нашем компьютере первые 100 подходящих дробей за 4.72 сек, а первые 200 — за 72.14 сек. В то время как наш алгоритм дает 200 подходящих дробей за 0.1 сек.

Правда, при нулевом приближении $x_0 = 3$ первые три подходящие дроби неправильны, но они легко восстанавливаются по полученным правильным с помощью функций $K(x, n)$ (см. ниже).

Классический критерий Дирихле иррациональности числа α имеет вид

$$|\alpha q - p| \to 0, \quad \frac{p}{q} \to \alpha, \quad p, q \in \mathbb{N}.$$

Подходящие дроби всех иррациональных чисел обладают этим свойством.

Как оказалось, для всех случаев, когда иррациональность числа удается установить по критерию Дирихле с помощью быстро сходящейся последовательности рациональных чисел (например, для π, $\ln 2$, $\zeta(2)$, $\zeta(3)$), эти последовательности стремятся к своему пределу быстрее в численном смысле, чем подходящие дроби этого предела (разумеется, это экспериментальный факт).

Это, вероятно, связано с тем, что подходящие дроби являются наилучшими диофантовыми приближениями своего предела. Поэтому другие дроби, удовлетворяющие критерию Дирихле, должны как-то компенсировать свою неоптимальность, например, за счет более быстрой сходимости. Однако данное рассуждение хотя и правдоподобно, но, вообще говоря, неверно (см. контрпример ниже).

Применение функций $K(x, n)$ позволяет генерировать подходящие дроби числа α из любой последовательности рациональных чисел p/q, если она стремится к α быстрее в обычном численном смысле, чем последовательность его подходящих дробей. А именно, будем говорить, что последовательность $s_n \to \alpha$ является *быстро сходящейся*, если выполнено условие

$$|\alpha - s_n| < \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right|, \quad \text{const} < n \in \mathbb{N}. \tag{12}$$

Тогда (см. предложение 3)

$$\frac{p_n}{q_n} = K(s_n, n), \tag{13}$$

где p_n/q_n — это обыкновенные подходящие дроби числа α. Например (см. [6]), мы сгенерировали таким образом подходящие дроби (до 100-й включительно) констант π, $\ln 2$, $\zeta(n)$, $n = 2, 3, \ldots, 9$, а также константы γ.

Возникает вопрос, как определить, является ли данная последовательность рациональных чисел быстро сходящейся, если последовательность подходящих дробей предела α неизвестна. На этот вопрос дает ответ следующее

Предложение 5. *Пусть дана сходящаяся неприводимая последовательность рациональных чисел* $t_n = p_n/q_n$. *Тогда она является* (*начиная с некоторого индекса* n) *последовательностью подходящих дробей своего предела* α *тогда и только тогда, когда*

$$t_n = K(t_n, n) = K(t_{n+2}, n), \quad \text{const} < n. \tag{14}$$

Доказательство. В одну сторону утверждение очевидно, т.е. все последовательности подходящих дробей иррациональных чисел обладают свойством (14).

Для однозначного представления рационального числа p_{n+2}/q_{n+2} в виде обыкновенной цепной дроби необходимо, чтобы $a_{n+2} \neq 1$ (см. [8]). Но в любом случае,

$$t_{n+2} = K(t_{n+2}, n+2) = [a_0; a_1, \ldots, a_n, a_{n+1}, a_{n+2}] = [a_0; a_1, \ldots, a_n + \frac{1}{a_{n+1} + \frac{1}{a_{n+2}}}].$$

Поэтому $t_n = K(t_n, n) = [a_0; a_1, \ldots, a_n]$ находится однозначно. Что требовалось доказать.

Недостающие подходящие дроби последовательности $\{t_n\}$ в начале списка однозначно восстанавливаются с помощью функций $K(x, n)$, т.е. надо положить

$$t_n = K(t_{n+2}, n), \quad n < \text{const},$$

заменив неправильные t_n в начале списка.

Наконец, частные знаменатели разложения числа α в обыкновенную цепную дробь находятся по формуле

$$\frac{r_n - r_{n-2}}{r_{n-1}} = a_n \in \mathbb{N}, \tag{15}$$

где $r_n = q_n + i\,p_n$. Иными словами, цепная дробь Эйлера, построенная по алгоритму Д. Бернулли из последовательности $\{t_n\}$ (см. [6]), совпадает с обыкновенной цепной дробью предела α, если начальный отрезок последовательности $\{t_n\}$ пересчитать с помощью функций $K(x, n)$.

Таким образом, пока выполняется условие (14), выполняется и (13), т.е. подходящие дроби, полученные с помощью функций $K(x, n)$, являются также подходящими дробями предела последовательности.

Однако доказать, что последовательность s_n, полученная, например, по алгоритму ускорения сходимости (см. [6]), обладает свойством (12) для всех $n > \text{const}$, по-видимому, невозможно. Единственный пример (квадратичные иррациональности не в счет), когда это можно установить достоверно, дает константа **e** (см. [7]).

В то же время результаты Хинчина (см. [8]) и Леви (см. [19]) дают усредненную оценку сверху роста частных знаменателей для почти всех вещественных чисел. Поэтому “в среднем” подходящие дроби иррационального числа не могут стремиться к своему пределу слишком быстро.

Это означает, что вполне реалистично ожидать, что последовательность рациональных чисел s_n, сходящаяся к пределу α быстрее, чем последовательность его подходящих дробей, может быть построена, если не требовать выполнения условия (12) всюду, а допустить его нарушение в некоторых случаях. Или даже в бесконечном числе случаев. Тогда не все подходящие дроби числа α могут быть получены по формуле (13).

С другой стороны, мы узнаем об этом из предложения 5, т.е. сгенерированные подходящие дроби $t_n = K(s_n, n)$ не будут удовлетворять свойству (14).

Однако кодирование иррационального числа его подходящими дробями обладает весьма большой избыточностью. Например, справедливо

Предложение 6. *Пусть известна последовательность четных подходящих дробей иррационального числа* $\alpha > 0$,

$$\frac{p_0}{q_0}, \frac{p_2}{q_2}, \ldots, \frac{p_{2n}}{q_{2n}}, \ldots.$$

Тогда все остальные элементы разложения числа $\alpha > 0$ *в обыкновенную цепную дробь восстанавливаются однозначно.*

Доказательство. Необходимо определить все нечетные подходящие дроби p_{2n-1}/q_{2n-1}, а также все частные знаменатели a_n числа α.

Заметим сперва, что $p_{-2} = 0$, $q_{-2} = 1$ и $p_{-1} = 1$, $q_{-1} = 0$ по определению. Поэтому $p_0 = a_0$ и $q_0 = 1$. Рассмотрим последовательность

$$p_n = a_n p_{n-1} + p_{n-2},\ q_n = a_n q_{n-1} + q_{n-2}, \quad n = 1, 2, \ldots, 2N,$$

как систему $4N$ линейных уравнений относительно $4N$ неизвестных

$$a_1, a_2, \ldots, a_{2N}, \quad p_1, p_3, \ldots, p_{2N-1},\ q_1, q_3, \ldots, q_{2N-1}.$$

Тот факт, что система линейна, можно увидеть, если разбить уравнения на группы $\{1, 2, 3, 4\}$, $\{5, 6, 7, 8\}$ и т.д., а затем исключить неизвестные a_n.

Нетрудно доказать по индукции, что эта система всегда разрешима. Например, для $N = 1$ получим решение

$$a_1 = \frac{q_2 - q_0}{p_2 q_0 - p_0 q_2},\ a_2 = \frac{p_2 q_0 - p_0 q_2}{q_0},\ p_1 = \frac{q_0 (p_2 - p_0)}{p_2 q_0 - p_0 q_2},\ q_1 = \frac{q_0 (q_2 - q_0)}{p_2 q_0 - p_0 q_2}.$$

Далее формулы становятся более громоздкими, но, вспоминая, что $q_0 = 1$, и вычисляя сперва величины с четными номерами, общее решение можно записать в виде

$$a_n = \begin{cases} p_n q_{n-2} - p_{n-2} q_n, & n \text{ четно}, \\ \dfrac{p_{n+1} q_{n-3} - p_{n-3} q_{n+1} - a_{n+1} - a_{n-1}}{a_{n+1} a_{n-1}}, & n \text{ нечетно}, \end{cases}$$

где

$$p_n = \begin{cases} p_n, & \\ \dfrac{p_{n+1} - p_{n-1}}{a_{n+1}}, & \end{cases} \qquad q_n = \begin{cases} q_n, & n \text{ четно}, \\ \dfrac{q_{n+1} - q_{n-1}}{a_{n+1}}, & n \text{ нечетно}, \end{cases}$$

что требовалось доказать.

Аналогичным образом можно восстановить четные подходящие дроби разложения числа в обыкновенную цепную дробь, если знать все нечетные подходящие дроби. Отметим, что эти операции не являются обратными операциям четного или нечетного сжатия обыкновенной цепной дроби, так как частные знаменатели здесь неизвестны.

Предложение 6, впрочем, становится почти очевидным, если использовать предложение 3 и функции $K(x, n)$: достаточно вычислить последовательность подходящих дробей

$$\frac{p_n}{q_n} = K\left(\frac{p_{2N}}{q_{2N}}, n\right), \quad n = 1, 2, \ldots, 2N - 2,$$

а затем вычислить цепную дробь Эйлера, которая даст $a_n, n = 1, \ldots, 2N - 2$ (а также и a_{2N-1}, если $a_{2N} \neq 1$).

Таким образом, один из возможных сценариев доказательства иррациональности числа α — это предъявление последовательности рациональных чисел, сходящейся к α, где условие (12) выполняется если не всюду, то достаточно часто. Однако и это ослабленное требование к скорости сходимости может оказаться слишком ограничительным. К счастью, оно не является необходимым.

Напомним, что два иррациональных числа α и β называются *эквивалентными*, если существует преобразование

$$\alpha = \frac{a\beta + b}{c\beta + d}, \quad a, b, c, d \in \mathbb{Z}, \quad a d - b c = \pm 1.$$

Тогда частные знаменатели a_n числа α и b_n числа β совпадают начиная с некоторого номера $n > N$ и $n + m$, $m \in \mathbb{Z}$, т.е. $a_n = b_{n+m}$ (см., например, [20]).

Однако, как мы уже отмечали, дробно-линейное преобразование с целыми коэффициентами $a, b, c, d \in \mathbb{Z}$, не принадлежащее модулярной группе, может радикальным образом изменить скорость сходимости подходящих дробей числа α, по сравнению с β. Например, (см. [7]),

$$\mathbf{e} - K(\mathbf{e}, 10) \approx 1.102 \cdot 10^{-7}, \quad \frac{\mathbf{e}-1}{\mathbf{e}+1} - K\left(\frac{\mathbf{e}-1}{\mathbf{e}+1}, 10\right) \approx 4.216 \cdot 10^{-26}.$$

Но всегда существует цепная дробь для α с целыми коэффициентами, которая продолжается как обыкновенная цепная дробь для β и наоборот. Например, $(\mathbf{e}-1)/(\mathbf{e}+1) = 1/(1+2/(\mathbf{e}-1))$, поэтому

$$\frac{\mathbf{e}-1}{\mathbf{e}+1} = \cfrac{1}{2+\cfrac{1}{6+\cfrac{1}{10+\cfrac{1}{14+\cfrac{1}{18+\cfrac{1}{22+\ldots}}}}}} = \cfrac{1}{1+\cfrac{2}{1+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\ldots}}}}}}},$$

т.е. очень быстро сходящаяся последовательность подходящих дробей числа $(\mathbf{e}-1)/(\mathbf{e}+1)$ может быть заменена последовательностью

$$\left\{1, \frac{1}{3}, \frac{1}{2}, \frac{5}{11}, \frac{7}{15}, \frac{6}{13}, \frac{55}{119}, \frac{67}{145}, \frac{61}{132}, \frac{799}{1729}, \ldots\right\}, \tag{16}$$

которая сходится как последовательность подходящих дробей числа $\mathbf{e}$, но не является таковой (условие (14) не выполнено).

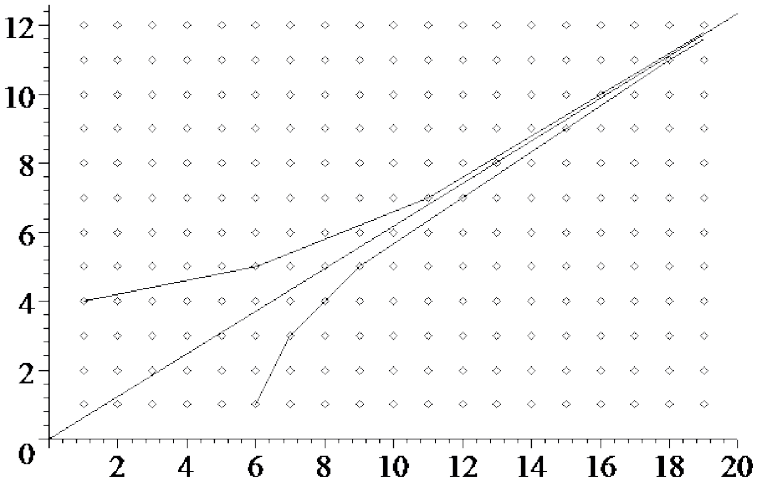
Тем не менее, последовательность (16) обладает почти всеми свойствами последовательностей обыкновенных подходящих дробей и, в частности, подразумевает иррациональность своего предела.

Монотонно возрастающие или убывающие последовательности рациональных чисел играют особую роль в установлении иррациональности своего предела. Так, четные и нечетные подходящие дроби обладают этим свойством.

Однако помимо монотонности, подходящие дроби обладают еще одним свойством, которое позволяет сразу сделать вывод об иррациональности их предела. А именно: последовательности гауссовых целых r_{2n} и r_{2n-1} всегда ведут себя так, как показано на фиг. 3.

На фиг. 3 показаны комплексные точки $r_n = q_n + i\,p_n$ для четных (снизу) и нечетных (сверху) подходящих дробей некоторого иррационального числа α, а также множество точек $\{x + i\alpha x,\ x > 0\}$, т.е. график функции $y = \alpha x$.

Монотонно возрастающая (убывающая) последовательность рациональных чисел p_n/q_n называется *выпуклой* (*вогнутой*), если последовательность $(p_{n+1} - p_n)/(q_{n+1} - q_n)$ монотонно убывает (возрастает), причем $q_{n+1} > q_n$.



Фиг. 3. Выпуклая и вогнутая последовательности.

Очевидно, это определение выпуклости (вогнутости) является дискретным аналогом обычной выпуклости (вогнутости) гладкой кривой, а также является критерием иррациональности Бруна (см. [6, 21]).

Выпуклость (вогнутость) этих ломаных влечет иррациональность α независимо от критерия Бруна (или является его доказательством), так как если бы α было рациональным, то существовала бы некоторая полоса вдоль прямой $y = \alpha x$ (или $y = \alpha x + \beta$, $\beta \in \mathbb{R}$), т.е. асимптоты этих ломаных, где все целочисленные точки лежали бы только на этой прямой либо вообще отсутствовали. В то время как существование таких выпуклых или вогнутых ломаных, имеющих целочисленные вершины, противоречит существованию такой полосы.

Как показано в [6], все последовательности четных (нечетных) подходящих дробей иррациональных чисел удовлетворяют критериям Бруна. Однако фиг. 3 указывает на существование бесконечного числа таких последовательностей для любой иррациональности. Первый такой пример дал сам Брун для числа **e**. Свойство неприводимости дробей при этом не только не является необходимым, но может препятствовать построению таких последовательностей (как для числа **e**, см. [6]).

Таким образом, имея достаточно быстро сходящуюся последовательность рациональных чисел, теоретически можно выбрать из нее (или сконструировать) выпуклую или вогнутую последовательность, что автоматически даст иррациональность предела.

Насколько нам известно, единственная работа в этом направлении — это построение выборки, удовлетворяющей критерию Бруна, из последовательности Апери для числа $\zeta(3)$ (см. [22]).

5. ЗАКЛЮЧЕНИЕ

Отметим, что вычисления в рациональной арифметике, к сожалению, пока не являются разделом численного анализа, а скорее принадлежат компьютерной алгебре.

Между тем, (помимо примеров в этой статье) существуют приложения вычислений в рациональной арифметике, относящиеся непосредственно к численному анализу (см. [7]). Так, коэффициенты разложений некоторых функций по полиномам Чебышёва или Лежандра являются рациональными числами. Квадратуры Гаусса на интервале [0, 1] от любой рациональной функции с рациональными коэффициентами также являются рациональными числами (либо неопределены). Эти и другие факты невозможно увидеть при расчетах в плавающей арифметике.

СПИСОК ЛИТЕРАТУРЫ

1. *Borwein J., Bailey D., Gigensohn R.* Experimentation in Mathematics: computational paths to discovery. A K Peters, Natick. 2004.
2. *Арнольд В. И.* Экспериментальная математика. М.: ФАЗИС, 2005.
3. *Hammer R., et.al.* Numerical Toolbox for Verified Computing I. Springer, 1993.
4. *Appel K., Haken W.* Solution of the Four Color Map Problem // Scientific American. 1977. V. 237. N 4. P. 108–121.
5. *Рухович Ф. Д.* Внешние биллиарды вне правильных многоугольников: ручной случай // Изв. РАН. Сер. матем. 2022. Т. 86. No 3. С. 105–160.
6. *Варин В. П.* Преобразование последовательностей в доказательствах иррациональности некоторых фундаментальных констант // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. N 10. С. 1587–1614.
7. *Варин В. П.* Аппроксимация дифференциальных операторов с учетом граничных условий // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. N 8. С. 1251–1271.
8. *Хинчин А. Я.* Цепные дроби. М.: Гостехиздат, 1961.
9. *Воробьев Н. Н.* Числа Фибоначчи. М.: Наука, 1984.
10. *Hairer, et. al.* Solving Ordinary Differential Equations I. Nonstiff Problems. 2nd ed. Berlin, Springer. 1993.
11. *Weniger E. J.* Nonlinear sequence transformations for the acceleration of convergence and the summation of divergent series // Comput. Phys. Rep. North-Holland. Amsterdam, 1989. V. 10. P. 189–371.
12. *Brent R. P.* Computation of the Regular Continued Fraction for Euler's Constant // Math. of Comput. 1977. V. 31. No 139. P. 771–777.
13. *Serret M. J.-A.* Oeuvres de Lagrange. Vol. 2. Gauthier-Villars. Paris. (M DCCC LXVIII).
14. *Брюно А. Д.* Разложение алгебраических чисел в цепные дроби // Ж. вычисл. матем. и матем. физ. 1964. Т. 4. No 2. С. 211–221.
15. *Knuth D. E.* The Art of Computer Programming. 3rd ed. Vol. 2. 1998. Addison Wesley Longman.
16. *Haas A.* The relative growth rate for partial quotients // New York J. Math. 2008. V. 14. P. 139–143.
17. *Roth K. F.* Rational approximations to algebraic numbers // Mathematika. 1955. V. 2. Part 1. No 3. P. 1–20.
18. *Stark H. M.* An explanation of some exotic continued fractions found by Brillhart // in: A.O.L. Atkin, B.J. Birch (ed.), Computers in Number Theory. Science Research Council Atlas Symposium N 2. Oxford, Academic Press, 1971.
19. *Lévy P.* Sur le lois de probabilité dont dépendent les qoutients complets et incomplets d'une fraction continue // Bull. Scc. Math. 1929. V. 57. P. 178–194.
20. *Borwein J., et. al.* Neverending Fractions. An Introduction to Continued Fractions. Australian Mathematical Society Lecture Series: 23. 2014.
21. *Brun V.* Ein Satz über Irrationalität // Arkiv for Mathematik og Naturvidenskab (Kristiania), 1910. V. 31. No 3. P. 3–6.
22. *Butler L. A.* A useful application of Brun's irrationality criterion // Expo. Math. 2015. V. 33. P. 121–134.

RATIONAL ARITHMETIC WITH A ROUND-OFF

V. P. Varin*

Keldysh Institute of Applied Mathematics RAS, Miusskaya sq. 4., Moscow, 125047 Russia
**e-mail: varin@keldysh.ru*

Received 16 January, 2024
Revised 08 February, 2024
Accepted 05 March, 2024

Abstract. Computer calculations in floating-point arithmetic are always approximate. In contrast, calculations in rational arithmetic (for example, in computer algebra) are always absolutely precise and reproducible both on other computers and (theoretically) manually. Therefore, such calculations can be demonstrative in the sense that the proof obtained with their help is no different from the traditional one. However, such calculations are usually impossible in a sufficiently complex problem due to limited memory and time resources. We propose a mechanism for rounding off rational numbers in calculations in rational arithmetic, which solves this problem (of resources), i.e. the calculations can still be demonstrative, but no longer require unlimited resources. A number of examples of the implementation of standard numerical algorithms in this arithmetic are given. The results have applications to analytical number theory.

Keywords: rational arithmetic, convergent fractions, proof calculations, Brun's criterion of irrationality.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 914–921

УДК 519.614

К ВОПРОСУ ОБ АСИМПТОТИКЕ СОБСТВЕННЫХ ЗНАЧЕНИЙ СЕМИДИАГОНАЛЬНЫХ ТЁПЛИЦЕВЫХ МАТРИЦ[1)]

© 2024 г. И. В. Воронин[1,*]

[1] *141700 Московская обл., г. Долгопрудный, Институтский пер., д. 9, Московский физико-технический институт (национальный исследовательский университет), Россия*

[*] *e-mail: Voronin.I@phystech.edu*

Поступила в редакцию: 02.09.2024 г.
Переработанный вариант 12.02.2024 г.
Принята к публикации 06.03.2024 г.

Построены асимптотические формулы, допускающие равномерную оценку остаточного члена для тёплицевых матриц размера n при $n \to \infty$ в случае, когда их символ $a(t)$ имеет вид $a(t) = (t-2a_0+t^{-1})^3$. Данный результат является обобщением результата работы Stukopin et al. (2021), в которой получены аналогичные асимптотические формулы для семидиагональной тёплицевой матрицы с символом аналогичного вида, когда $a_0 = 1$. Полученные формулы имеют высокую вычислительную эффективность и обобщают результаты классических работ Партера и Видома по асимптотике экстремальных собственных значений. Библ. 13. Фиг. 3. Табл. 5.

Ключевые слова: матрицы Теплица, собственные векторы, асимптотические разложения.

DOI: 10.31857/S0044466924060029, **EDN:** XZPJFN

1. ВВЕДЕНИЕ

Пусть $a(t)$ — интегрируемая по Лебегу функция, определенная на единичной окружности $T = t \in \mathbb{C} : |t| = 1$. Обозначим через $T_n(a)$ матрицу Теплица $T_n(a) := (a_{j-k})_{j,k=1}^{n-1}$, где n — натуральное число, a_l — l-й коэффициент ряда Фурье функции $a(t)$. Заметим, что матрицу Теплица можно рассматривать как оператор из конечномерного векторного пространства. Функция $a(t)$ называется символом матрицы Теплица (оператор Теплица) $T_n(a)$. Отметим, что тёплицевые матрицы, а также тесно связанные с ними тёплицевы операторы, интенсивно изучаются в течение последних ста лет(см. [1]–[6]). Важность этой темы во многом обусловлена многочисленными применениями тёплицевых матриц в численных методах, дифференциальных и интегральных уравнениях, теории вероятностей, статистической физике (см., например, [7]–[10]). Настоящая статья посвящена нахождению асимптотических формул для собственных значений семидиагональных тёплицевых матриц с символом $\tilde{a}(t) = (a_1 t - 2\tilde{a}_0 + a_1 t^{-1})^3$, $a_0, a_1 \in \mathbb{C}$, т.е. симметричный линейный многочлен Лорана, возведенный в третью степень, а размер матрицы достаточно велик. Поскольку $\tilde{a}(t) = a_1^3(t - 2(\tilde{a}_0/a_1) + t^{-1})^3$, то спектр тёплицевой матрицы с символом $\tilde{a}$ может быть легко получен из спектра тёплицевой матрицы с символом $a(t) = (t - 2a_0 + t^{-1})^3$, где $a_0 = \tilde{a}_0/a_1$. Случай, если $a_0 \notin [-1, 1]$ подпадает под случай, рассмотренный в статье [11]. В настоящей работе будет рассмотрен случай, когда $a_0 \in [-1, 1]$. Сформулируем основные результаты. В данном случае тёплицева матрица оказывается самосопряженной и, как следствие, по слабой теореме Сегё (см. [12]) спектр данной матрицы лежит на образе единичной окружности $t = e^{i\varphi}$, $\varphi \in [0, \dots, 2\pi]$ под действием данного символа, т.е. любое собственное значение можно представить в виде $\lambda = a(e^{i\varphi}) = g(\varphi)$. В силу симметрии символа достаточно рассматривать $\varphi \in [0, \dots, \pi]$. Решать данную задачу будем относительно параметра φ, при этом собственное значение может быть найденно простой подстановкой $\lambda = g(\varphi)$.

[1)] Работа поддержана грантом РНФ 21-11-00283.

2. ОСНОВНОЙ РЕЗУЛЬТАТ

Введем некоторые функции, все функции будем определять на интервале $(0, \pi)$:

$$\beta(\varphi) := \arccos\left(a_0 - (a_0 - \cos\varphi)e^{2\pi i/3}\right), \quad \gamma(\varphi) = \overline{\beta(\varphi)}, \tag{1}$$

$$C_1(\varphi) := \frac{\sin(\gamma)}{\sin(\varphi)}e^{\pi i/3}, \quad C_2(\varphi) := \frac{\sin(\beta)}{\sin(\varphi)}e^{2\pi i/3}. \tag{2}$$

Функция Arccos многозначная, $\beta(\varphi)$ — одна из ее ветвей. Для решения этой задачи необходимо решить уравнение $\det T_n(a - g(\varphi)) = 0$, $\varphi \in (0, \pi)$. Для нахождения определителя воспользуемся результатами статьи[13]. Получим

$$\begin{aligned}\det(T_{2p}(a - g(\varphi))) = \frac{1}{2^6} \times \frac{\begin{vmatrix} \cos\left(\frac{n+1}{2}\varphi\right) & \cos\left(\frac{n+1}{2}\beta\right) & \cos\left(\frac{n+1}{2}\gamma\right) \\ \cos\left(\frac{n+3}{2}\varphi\right) & \cos\left(\frac{n+3}{2}\beta\right) & \cos\left(\frac{n+3}{2}\gamma\right) \\ \cos\left(\frac{n+5}{2}\varphi\right) & \cos\left(\frac{n+5}{2}\beta\right) & \cos\left(\frac{n+5}{2}\gamma\right) \end{vmatrix}}{(\cos\gamma - \cos\beta)(\cos\gamma - \cos\varphi)(\cos\beta - \cos\varphi)} \times \\ \times \frac{\begin{vmatrix} \sin\left(\frac{n+1}{2}\varphi\right) & \sin\left(\frac{n+1}{2}\beta\right) & \sin\left(\frac{n+1}{2}\gamma\right) \\ \sin\left(\frac{n+3}{2}\varphi\right) & \sin\left(\frac{n+3}{2}\beta\right) & \sin\left(\frac{n+3}{2}\gamma\right) \\ \sin\left(\frac{n+5}{2}\varphi\right) & \sin\left(\frac{n+5}{2}\beta\right) & \sin\left(\frac{n+5}{2}\gamma\right) \end{vmatrix}}{(\cos\gamma - \cos\beta)(\cos\gamma - \cos\varphi)(\cos\beta - \cos\varphi)}.\end{aligned} \tag{3}$$

Пусть $\varphi \neq \arccos(a_0) := \varphi_0$. Тогда при $\varphi \in (0, \varphi_0) \cup (\varphi_0, \pi)$ уравнение $\det T_n(a - g(\varphi)) = 0$ эквивалентно следующим двум уравнениям:

$$\begin{vmatrix} \cos\left(\frac{n+1}{2}\varphi\right) & \cos\left(\frac{n+1}{2}\beta\right) & \cos\left(\frac{n+1}{2}\gamma\right) \\ \cos\left(\frac{n+3}{2}\varphi\right) & \cos\left(\frac{n+3}{2}\beta\right) & \cos\left(\frac{n+3}{2}\gamma\right) \\ \cos\left(\frac{n+5}{2}\varphi\right) & \cos\left(\frac{n+5}{2}\beta\right) & \cos\left(\frac{n+5}{2}\gamma\right) \end{vmatrix} = 0, \tag{4}$$

$$\begin{vmatrix} \sin\left(\frac{n+1}{2}\varphi\right) & \sin\left(\frac{n+1}{2}\beta\right) & \sin\left(\frac{n+1}{2}\gamma\right) \\ \sin\left(\frac{n+3}{2}\varphi\right) & \sin\left(\frac{n+3}{2}\beta\right) & \sin\left(\frac{n+3}{2}\gamma\right) \\ \sin\left(\frac{n+5}{2}\varphi\right) & \sin\left(\frac{n+5}{2}\beta\right) & \sin\left(\frac{n+5}{2}\gamma\right) \end{vmatrix} = 0. \tag{5}$$

Уравнение (4) эквивалентно уравнению

$$\operatorname{tan}\left(\frac{n+3}{2}\right) = f(\varphi, n), \tag{6}$$

где

$$f(\varphi, n) = C_1(\varphi)\operatorname{tg}\left(\frac{n+3}{2}\gamma\right) - C_2(\varphi)\operatorname{tg}\left(\frac{n+3}{2}\beta\right),$$

а уравнение (5) эквивалентно

$$\operatorname{ctg}\left(\frac{n+3}{2}\right) = -h(\varphi, n), \tag{7}$$

где

$$h(\varphi, n) = -C_1(\varphi)\frac{1}{\operatorname{tg}\left(\frac{n+3}{2}\gamma\right)} + C_2(\varphi)\frac{1}{\operatorname{tg}\left(\frac{n+3}{2}\beta\right)}.$$

Теорема 1. *Пусть* $\lambda = g(\varphi)$ *и* $\varphi \in (0, \varphi_0) \cup (\varphi_0, \pi)$. *Тогда уравнение* $\det T_n(a - g(\varphi)) = 0$ *эквивалентно следующему набору уравнений:*

$$\varphi = \frac{2}{n+3}\left[\pi j + \operatorname{arctg} f(\varphi, n)\right], \quad j \in \left\{1, 2, \ldots, \left[\frac{n+1}{2}\right]\right\}, \tag{8}$$

$$\varphi = \frac{2}{n+3}\left[\pi j + \frac{\pi}{2} + \operatorname{arctg} h(\varphi, n)\right], \quad j \in \left\{1, 2, \ldots, \left[\frac{n}{2}\right]\right\}, \tag{9}$$

где

$$\begin{aligned} f(\varphi, n) &= C_1(\varphi)\operatorname{tg}\left(\frac{n+3}{2}\gamma\right) - C_2(\varphi)\operatorname{tg}\left(\frac{n+3}{2}\beta\right), \\ h(\varphi, n) &= -C_1(\varphi)\frac{1}{\operatorname{tg}\left(\frac{n+3}{2}\gamma\right)} + C_2(\varphi)\frac{1}{\operatorname{tg}\left(\frac{n+3}{2}\beta\right)}. \end{aligned} \tag{10}$$

Замечание 1. Заметим, что функции $f(\varphi, n)$ и $h(\varphi, n)$ являются вещественнозначными.

Введем обозначения $d_m := \pi m/(n+3)$.

Теорема 2.

- *На каждом из интервалов* $\Delta_{2j} = (d_{2j-1}, d_{2j+1})$, *для которых* $\varphi_0 \notin \Delta_{2j}$, *уравнение* (8) *имеет хотя бы один корень.*
- *На каждом из интервалов* $\Delta_{2j+1} = (d_{2j}, d_{2j+2})$, *для которых* $\varphi_0 \notin \Delta_{2j}$, *уравнение* (9) *имеет хотя бы один корень.*

Теорема 3. *Пусть* $\varphi_0 \in \Delta_{2j}$, *тогда*

- *если* $\varphi_0 > d_{2j}$, *уравнение* (8) *имеет корень на интервале* (d_{2j-1}, d_{2j});
- *если* $\varphi_0 < d_{2j}$, *уравнение* (8) *имеет корень на интервале* (d_{2j}, d_{2j+1});
- *если* $\varphi_0 = d_{2j} > \pi/2$, *уравнение* (8) *имеет корень на интервале* (d_{2j}, d_{2j+1});
- *если* $\varphi_0 = d_{2j} < \pi/2$, *уравнение* (8) *имеет корень на интервале* (d_{2j-1}, d_{2j}).

Пусть $\varphi_0 \in \Delta_{2j+1}$, *тогда*

- *если* $\varphi_0 > d_{2j+1}$, *уравнение* (9) *имеет корень на интервале* (d_{2j}, d_{2j+1});
- *если* $\varphi_0 < d_{2j+1}$, *уравнение* (9) *имеет корень на интервале* (d_{2j+1}, d_{2j+2});
- *если* $\varphi_0 = d_{2j+1} > \pi/2$, *уравнение* (9) *имеет корень на интервале* (d_{2j+1}, d_{2j+2});
- *если* $\varphi_0 = d_{2j+1} < \pi/2$, *уравнение* (9) *имеет корень на интервале* (d_{2j}, d_{2j+1}).

Если $\varphi_0 = d_m = \pi/2$, *то* $\lambda_m = 0$ *— собственное значение.*

Замечание 2. В теореме 3 были локализованны не менее n собственных значений (точнее их прообразов), но поскольку матрица порядка n, то это — все собственные значения.

Теорема 4. *Пусть* $a(t) = (t - 2a_0 + 1/t)^3$. *Пусть для заданного* n

$$m \notin \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right),$$

тогда

1)
$$\varphi_m = \varphi_{2j-1} = d_{2j} + \frac{2u^*_{1,j}}{n+3} + \frac{4u^*_{2,j}}{(n+3)^2} + O\left(\frac{1}{n^3}\right), \tag{11}$$

где

$$u^*_{1,j} = \operatorname{sign}(d_{2j} - \varphi_0)\operatorname{arctg}\left(-i\left(C_1\left(d_{2j}\right) + C_2\left(d_{2j}\right)\right)\right)$$

и

$$u^*_{2,j} = -i\operatorname{sign}(d_{2j} - \varphi_0)\frac{C_1'(d_{2j}) + C_2'(d_{2j})}{1 + \left(-iC_1\left(d_{2j}\right) - iC_2\left(d_{2j}\right)\right)^2};$$

2)
$$\varphi_m = \varphi_{2j} = d_{2j+1} + \frac{2w^*_{1,j}}{n+3} + \frac{4w^*_{2,j}}{(n+3)^2} + O\left(\frac{1}{n^3}\right), \tag{12}$$

где

$$w^*_{1,j} = \operatorname{sign}(d_{2j+1} - \varphi_0)\operatorname{arctg}\left(-i\left(C_1\left(d_{2j}\right) + C_2\left(d_{2j}\right)\right)\right)$$

и

$$w^*_{2,j} = -i\operatorname{sign}(d_{2j+1} - \varphi_0)\frac{C_1'(d_{2j}) + C_2'(d_{2j})}{1 + \left(-iC_1\left(d_{2j}\right) - iC_2\left(d_{2j}\right)\right)^2}.$$

Введем новые функции $X_1^{(1)} = X_1^{(1)}(u_1, j, n)$, $X_2^{(1)} = X_2^{(1)}(w_1, j, n)$, $X_3^{(1)} = X_3^{(1)}(w_1, j, n)$, $R^{(1)}(u_1) = R^{(1)}(u_1, j, n)$:

$$X_1^{(1)} = C_1(d_{2j})\operatorname{tg}\left(\frac{n+3}{2}\gamma(d_{2j}) + \gamma'(d_{2j})u_1\right) - C_2(d_{2j})\operatorname{tg}\left(\frac{n+3}{2}\beta(d_{2j}) + \beta'(d_{2j})u_1\right),$$

$$X_2^{(1)} = \left[\frac{C_1(d_{2j})\gamma''(d_{2j})u_1^2}{2\cos^2(q\gamma(d_{2j}) + \gamma'(d_{2j})u_1)} + \operatorname{tg}\left(q\gamma(d_{2j}) + \gamma'(d_{2j})u_1\right)C_1'(d_{2j})u_1\right] -$$
$$- \left[\frac{C_2(d_{2j})\beta''(d_{2j})u_1^2}{2\cos^2(q\beta(d_{2j}) + \beta'(d_j)u_1)} + \operatorname{tg}\left(q\beta(d_{2j}) + \beta'(d_{2j})u_1\right)C_2'(d_{2j})u_1\right],$$

$$X_3^{(1)} = \frac{C_1(d_{2j})\gamma'(d_{2j})}{\cos^2(q\gamma(d_{2j}) + \gamma'(d_{2j})u_1)} - \frac{C_2(d_{2j})\beta'(d_{2j})}{\cos^2(q\beta(d_{2j}) + \beta'(d_{2j})u_1)},$$

$$R^{(1)}(u_1) = \frac{X_2^{(1)}}{1 + (X_1^{(1)})^2 - X_3^{(1)}},$$

а также функции $X_1^{(2)} = X_1^{(2)}(w_1, j, n)$, $X_2^{(2)} = X_2^{(2)}(w_1, j, n)$, $X_3^{(2)} = X_3^{(2)}(w_1, j, n)$, $R^{(2)}(w_1) = R^{(2)}(w_1, j, n)$:

$$X_1^{(2)} = -\frac{C_1(d_{2j})}{\operatorname{tg}\left(\frac{n+3}{2}\gamma(d_{2j}) + \gamma'(d_{2j})w_1\right)} + \frac{C_2(d_{2j})}{\operatorname{tg}\left(\frac{n+3}{2}\beta(d_{2j}) + \beta'(d_{2j})w_1\right)},$$

$$X_2^{(2)} = \left[\frac{C_1(d_{2j})\gamma''(d_{2j})w_1^2}{2\sin^2(q\gamma(d_{2j}) + \gamma'(d_{2j})w_1)} - \frac{C_1'(d_{2j})w_1}{\operatorname{tg}\left(q\gamma(d_{2j}) + \gamma'(d_{2j})w_1\right)}\right] +$$
$$+ \left[-\frac{C_2(d_{2j})\beta''(d_{2j})w_1^2}{2\sin^2(q\beta(d_{2j}) + \beta'(d_{2j})w_1)} + \frac{C_2'(d_{2j})w_1}{\operatorname{tg}\left(q\beta(d_{2j}) + \beta'(d_{2j})w_1\right)}\right],$$

$$X_3^{(2)} = \frac{C_1(d_{2j})\gamma'(d_{2j})}{\sin^2(q\gamma(d_{2j}) + \gamma'(d_{2j})w_1)} - \frac{C_2(d_{2j})\beta'(d_{2j})}{\sin^2(q\beta(d_{2j}) + \beta'(d_{2j})w_1)},$$

$$R^{(2)}(w_1) = \frac{X_2^{(2)}}{1 + (X_1^{(2)})^2 - X_3^{(2)}}.$$

Теорема 5. *Пусть* $a(t) = (t - 2 + 1/t)^3$. *Пусть для заданного* n

$$m \in \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right)$$

и $\varphi_0 \notin \Delta_m$, *тогда, начиная с некоторого* n,

1)
$$\varphi_m = \varphi_{2j-1} = d_{2j} + \frac{2u^*_{1,j}}{n+3} + \frac{4u^*_{2,j}}{(n+3)^2} + O\left(\tfrac{1}{n^3}\right), \tag{13}$$

где $u^*_{1,j}$ *— решение уравнения* $u_1 = \operatorname{arctg}\left(X_1^{(1)}(u_1)\right)$ *и* $u^*_{2,j} = R^{(1)}(u^*_{1,j})$*;*

2) $$\varphi_m = \varphi_{2j} = d_{2j+1} + \frac{2w^*_{1,j}}{n+3} + \frac{4w^*_{2,j}}{(n+3)^2} + O\left(\frac{1}{n^3}\right), \tag{14}$$

где $w^*_{1,j}$ *— решение уравнения* $w_1 = \operatorname{arctg}\left(X_1^{(2)}(w_1)\right)$ *и* $w^*_{2,j} = R^{(2)}(w^*_{1,j})$.

Замечание 3. Заметим, что в случае $\varphi_0 \in \Delta_m$, для нахождения φ_m нельзя использовать теоремы 4 и 5 однако положив $\varphi_m = d_m$, получим что $\lambda_m = g(\varphi_m) + O\left(1/n^3\right)$.

Теорема 6. *Пусть* $a(t) = (t - 2 + 1/t)^3$. *Тогда, начиная с некоторого* n*,*

$$\lambda^{(n)}_{2j-1} = g(d_{1,j}) + g'(d_{1,j})\frac{2u^*_{1,j}}{n+3} + \frac{4u^*_{2,j}g'(d_{1,j}) + 2(u^*_{1,j})^2 g''(d_{1,j})}{(n+3)^2} + O\left(\frac{1}{n^3}\right), \tag{15}$$

$$\lambda^{(n)}_{2j} = g(d_{2,j}) + g'(d_{2,j})\frac{2w^*_{1,j}}{n+3} + \frac{4w^*_{2,j}g'(d_{2,j}) + 2(w^*_{1,j})^2 g''(d_{2,j})}{(n+3)^2} + O\left(\frac{1}{n^3}\right), \tag{16}$$

$u^*_{1,j}$, $u^*_{2,j}$, $w^*_{1,j}$ *и* $w^*_{2,j}$ *определяются так же, как в теоремах 4–5, в зависимости от номера собственного значения* m*.*

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Все численные эксперименты проводились в математическом пакете Maple. Во всех расчетах все значения были найдены с точностью в 50 знаков. Под точным собственными значениями будем подразумевать собственные значения, рассчитанные с использованием встроенной функции Maple. В настоящей работе нахождение собственных значений сводилось к решению двух наборов уравнений (в зависимости от четности собственных значений). Каждое из уравнений решалось относительно параметра φ и имело единственный корень φ_m, и каждый такой корень соответствует единственному собственному значению, которое может быть найдено простой подстановкой $\lambda_m = g(\varphi_m)$. Здесь m — номера собственных значений, упорядоченных по убыванию. В работе приведены явные формулы для корней φ_m с точностью $O\left(1/n^3\right)$. Обозначим через φ^*_m приближенные корни, посчитанные по формулам (11), (12), (13), (14), $\lambda^*_m = g(\varphi^*_m)$ — соответствующие им, приближенно вычисленные собственные значения через простую подстановку, через $\bar{\lambda}_m$ — собственные значения приближенно вычисленные по формулам (15) и (16), а через φ_m и λ_m их точные значения. Введем следующие обозначения: $\Delta\varphi_m = |\varphi^*_m - \varphi_m|$, $\Delta\lambda_m = |\lambda^*_m - \lambda_m|$, $\Delta_r\lambda_m = \left|\frac{\lambda^*_m - \lambda_m}{\lambda_m}\right|$, $\Delta\bar{\lambda}_m = |\bar{\lambda}_m - \lambda_m|$, $\Delta_r\bar{\lambda}_m = \left|\frac{\bar{\lambda}_m - \lambda_m}{\lambda_m}\right|$.

Рассмотрим случай, когда $a_0 = 1/2$. На фиг. 1–3 изображены нормированные порядком остатка погрешности, т.е. $\Delta\varphi_m(n+3)^3$, $\Delta\lambda_m(n+3)^3$ и $\Delta\bar{\lambda}_m(n+3)^3$ соответственно, причем на одном изображении графики для различных размеров матрицы. Заметим, что графики наложились друг на друга, что косвенно говорит о том, что остаток действительно имеет порядок $1/(n+3)^3$.

В табл. 1 приведены максимальные отклонения корней $\Delta\varphi = \max\limits_m \Delta\varphi_m$, а также максимальные абсолютные и относительные отклонения собственных значений $\Delta\lambda = \max\limits_m \Delta\lambda_m$, $\Delta_r\lambda = \max\limits_m \Delta_r\lambda_m$ при условии

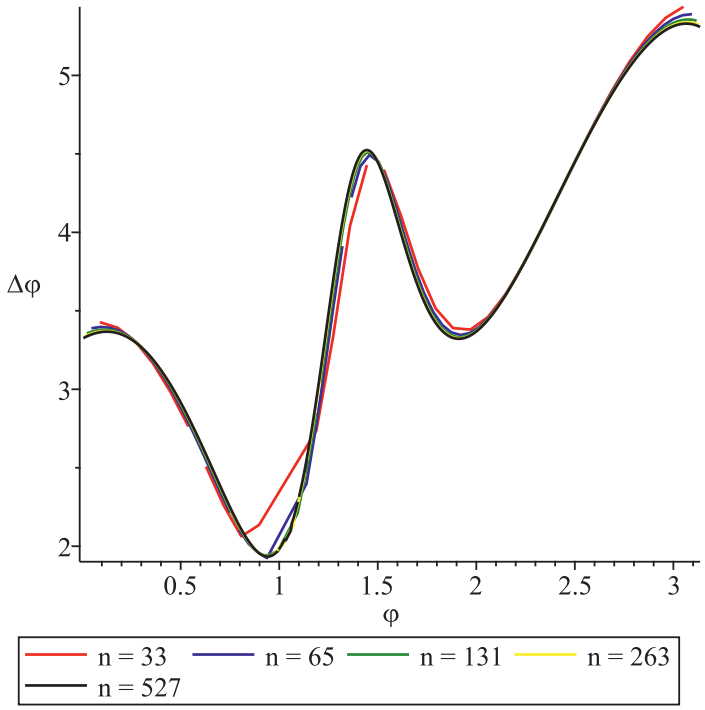
$$m \notin \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right).$$

В табл. 2 приведены аналогичные результаты только при условии $\varphi_0 \notin \Delta_m$ и

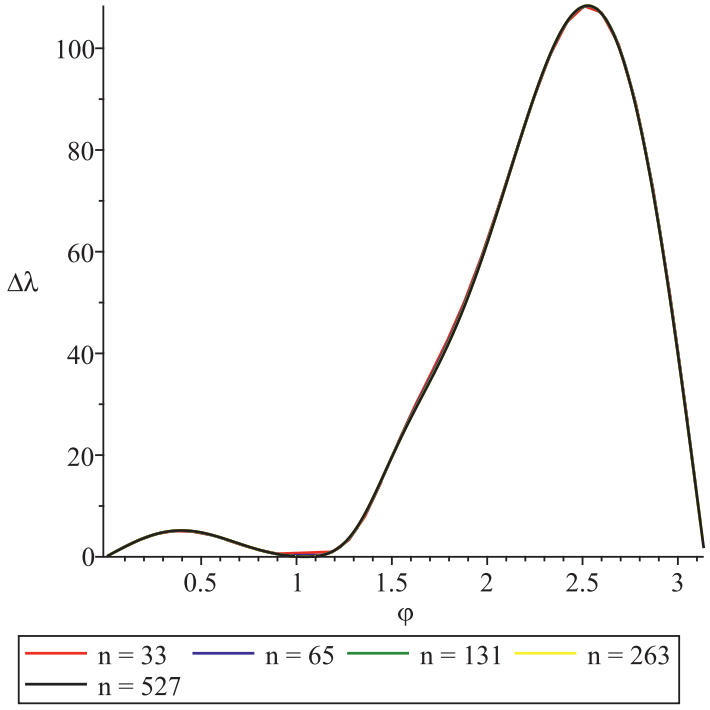
$$m \in \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right).$$

В табл. 3 приведены максимальные абсолютные и относительные отклонения собственных значений, посчитанных по формулам (15) и (16) $\Delta\bar{\lambda} = \max\limits_m \Delta\bar{\lambda}_m$, $\Delta_r\bar{\lambda} = \max\limits_m \Delta_r\bar{\lambda}_m$, при условии

$$m \notin \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right).$$



Фиг. 1. Абсолютная нормированная погрешность корней $\varphi_m(n+3)^3$.

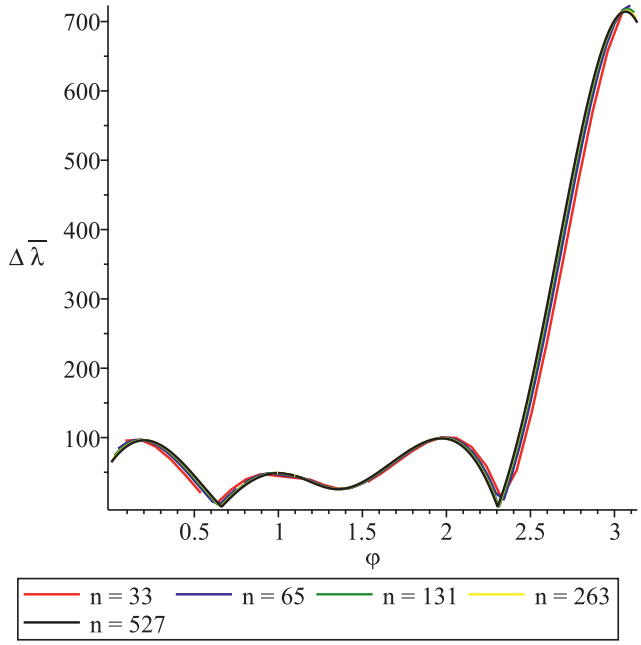


Фиг. 2. Абсолютная нормированная погрешность собственных значений $\lambda_m(n+3)^3$.

В табл. 4 приведены аналогичные результаты только при условии $\varphi_0 \notin \Delta_m$ и

$$m \in \left(\frac{\varphi_0(n+3)}{\pi} - \left(\frac{4\ln(n+3)}{\pi} + 1\right), \frac{\varphi_0(n+3)}{\pi} + \left(\frac{4\ln(n+3)}{\pi} - 1\right)\right).$$

В табл. 5 приведены максимальные погрешности, когда $\varphi_0 \in d_m$, в этом случае полагалось $\varphi_m^* = d_m$, таких m не более двух.



Фиг. 3. Абсолютная нормированная погрешность собственных значений $\lambda_m(n+3)^3$.

Таблица 1. Максимальные отклонения при использовании формул (11) и (12)

n	33	65	131	263	527
$\Delta\varphi$	1.2×10^{-4}	1.7×10^{-5}	2.2×10^{-6}	2.8×10^{-7}	3.5×10^{-8}
$\Delta\lambda$	2.3×10^{-3}	3.4×10^{-4}	4.5×10^{-5}	5.8×10^{-6}	7.3×10^{-7}
$\Delta_r\lambda$	6.1×10^{-4}	1.3×10^{-4}	2.4×10^{-5}	4.4×10^{-6}	8.9×10^{-7}

Таблица 2. Максимальные отклонения при использовании формул (13) и (14)

n	33	65	131	263	527
$\Delta\varphi$	9×10^{-5}	1.2×10^{-5}	1.2×10^{-6}	1.3×10^{-7}	1.5×10^{-8}
$\Delta\lambda$	4.8×10^{-4}	1.8×10^{-5}	4.1×10^{-7}	1.5×10^{-8}	5.6×10^{-10}
$\Delta_r\lambda$	1.3×10^{-3}	3.2×10^{-4}	8×10^{-5}	2.2×10^{-5}	5.6×10^{-6}

Таблица 3. Максимальные отклонения при использовании формул (15) и (16)

n	33	65	131	263	527
$\Delta\lambda$	1.5×10^{-2}	2.3×10^{-3}	3×10^{-4}	3.8×10^{-5}	4.8×10^{-6}
$\Delta_r\lambda$	2.1×10^{-3}	8.4×10^{-4}	7.8×10^{-4}	5.6×10^{-4}	3.9×10^{-4}

Таблица 4. Максимальные отклонения при использовании формул (15) и (16)

n	33	65	131	263	527
$\Delta\lambda$	1×10^{-3}	1.5×10^{-4}	2×10^{-5}	2.6×10^{-6}	3.3×10^{-7}
$\Delta_r\lambda$	6.8×10^{-2}	2.5×10^{-2}	2.5×10^{-2}	2.4×10^{-2}	2.4×10^{-2}

Таблица 5. Максимальные отклонения при использовании формул (15) и (16)

n	33	65	131	263	527
$\Delta\lambda$	1.5×10^{-4}	1.1×10^{-3}	1.4×10^{-4}	1.8×10^{-5}	2.3×10^{-6}

СПИСОК ЛИТЕРАТУРЫ

1. *Stukopin V., Grudsky S., Voronin I., Barrera M.* Asymptotics of the eigenvalues of seven-diagonal Toeplitz matrices of a special form // arXive. 2021. Nov. 2111.07196.
2. *Savage L. J., Grenander U., Szego G.* Toeplitz forms and their Applications // J. Am. Statistic. Associat. 1958. V. 53. N 283. P. 763.
3. *Schmidt P., Spitzer F.* The Toeplitz matrices of an arbitrary Laurent polynomial // Math. Scandinavica. 1960. V. 8. P. 15.
4. *Widom H.* Eigenvalue distribution of nonselfadjoint Toeplitz matrices and the asymptotics of Toeplitz determinants in the case of nonvanishing index // Oper. Theory Adv. Appl. 1990. V. 48.
5. *Bottcher A., Grudsky S. M.* Spectral properties of banded Toeplitz matrices // Soc. Industrial and Appl. Math. 2005.
6. *Bottcher A., Silbermann B.* Introduction to large truncated Toeplitz matrices. Springer New York, 1999.
7. *Deift P., It's A., Krasovsky I.* Toeplitz Matrices and Toeplitz determinants under the impetus of the ising model: some history and some recent results // Comm. on Pure and Appl. Math. 2013. V. 66, N 9. P. 1360–1438.
8. *Deift P., It's A., Krasovsky I.* Eigenvalues of Toeplitz matrices in the bulk of the spectrum // Bull. Inst. Math. Acad. Sin. 2012. V. 7. P. 437–461.
9. *Kadano L. P.* Spin-spin correlations in the two-dimensional ising model // Il Nuovo Cimento B Ser. 10. 1966. V. 44. N 2. P. 276–305.
10. *McCoy B., Wu T.* The Two-Dimensional Ising Model, 1973.
11. *Batalshchikov A. A., Grudsky S. M., Stukopin V. A.* Asymptotics of eigenvalues of symmetric Toeplitz band matrices // Linear Algebra and its Applications. 2015. V. 469. P. 464–486. https://www.sciencedirect.com/science/article/pii/S0024379514007691
12. *Szego G.* Ein Grenzwertsatz uber die Toeplitzschen Determinanten einer reellen positiven Funktion // Math. Annalen. 1915. V. 76. N 4. P. 490–503.
13. *Eloua M.* On a relationship between Chebyshev polynomials and Toeplitz determinants // Appl. Math. Comput. 2014. V. 229. P. 27–33.

ON THE ASYMPTOTICS OF EIGENVALUES OF SEMIDIAGONAL TOEPLITZ MATRICES

I. V. Voronin*

Moscow Institute of Physics and Technology (National Research University), Institutsky Lane., 9, Dolgoprudnyi, Moscow oblast, 141700 Russia

**e-mail: Voronin.I@phystech.edu*

Received 02 September, 2023
Revised 12 February, 2024
Accepted 06 March, 2024

Abstract. Asymptotic formulas are constructed that allow a uniform estimate of the remainder term for Toeplitz matrices of size n for $n \to \infty$ in the case when their symbol $a(t)$ has the form $a(t) = (t - 2a_0 + t^{-1})^3$. This result is a generalization of the result of Stukopin et al. (2021), in which similar asymptotic formulas were obtained for a diagonal Toeplitz matrix with a symbol of a similar form when $a_0 = 1$. The obtained formulas have high computational efficiency and generalize the results of the classical works of Parterre and Widom on the asymptotics of extreme eigenvalues.

Keywords: Toeplitz matrices, eigenvectors, asymptotic expansions.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 922–931

УДК 519.653

ФОРМУЛЫ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ НА РАВНОМЕРНОЙ СЕТКЕ ПРИ НАЛИЧИИ ПОГРАНИЧНОГО СЛОЯ[1)]

© 2024 г. А. И. Задорин[1,*]

[1] *630090 Новосибирск, пр-т Акад. Коптюга, 4, Ин-т матем. СО РАН*

[*] *e-mail: zadorin@ofim.oscsbras.ru*

Поступила в редакцию 13.10.2023 г.
Переработанный вариант 28.12.2023 г.
Принята к публикации 05.03.2024 г.

Рассматривается вопрос численного дифференцирования функций с большими градиентами. Предполагается, что для исходной функции одной переменной справедлива декомпозиция в виде суммы регулярной составляющей с ограниченными производными до некоторого порядка и пограничной составляющей, имеющей большие градиенты и известной с точностью до множителя. Такая декомпозиция, в частности, справедлива для решения сингулярно возмущенной краевой задачи. Тема исследования актуальна, так как применение к функциям с большими градиентами классических полиномиальных формул численного дифференцирования может приводить к существенным погрешностям. Оценивается погрешность формул численного дифференцирования, по построению точных на пограничной составляющей исходной функции. Приведены результаты численных экспериментов, согласующиеся с полученными оценками погрешностей. Библ. 16. Табл. 2.

Ключевые слова: функция одной переменной, большие градиенты, специальная формула численного дифференцирования, оценка погрешности.

DOI: 10.31857/S0044466924060039, **EDN:** XZEARW

1. ВВЕДЕНИЕ

Применение классических полиномиальных формул численного дифференцирования [1] при наличии пограничного слоя может приводить к значительным погрешностям [2]. В связи с этим возникает необходимость в построении формул численного дифференцирования, погрешность которых не растет из-за больших градиентов функции в области пограничного слоя. В [3] предполагается, что у исходной функции с точностью до множителя выделена составляющая, отвечающая за большие градиенты функции. Такая декомпозиция функции с выделением регулярной и пограничной составляющих обосновывалась в [4] для решения сингулярно возмущенной задачи. В [3] построена интерполяционная формула с произвольно заданным числом узлов интерполяции, точная на пограничной составляющей. Равномерная по пограничной составляющей оценка погрешности этой формулы получена в [5].

В [3] на основе дифференцирования построенного интерполянта получены формулы численного дифференцирования, точные на пограничной составляющей функции. Однако погрешность построенных формул для вычисления производных в [3] не оценена. В [6], [7] получены оценки погрешности формул из [3], равномерные по градиентам выделенной пограничной составляющей, при вычислении первой и второй производных. В [8] рассмотрен случай, когда формула из [3] содержит k узлов в сеточном шаблоне для производной. В случае $n = k - 1$ и экспоненциального пограничного слоя получена оценка погрешности, где n – порядок вычисляемой производной. При этом предполагается, что регулярная составляющая имеет ограниченные производные.

[1)] Работа выполнена в рамках государственного задания ИМ СО РАН, проект FWNF-2022-0016.

В предлагаемой работе оценивается погрешность, когда формула для производной, точная на пограничной составляющей, содержит k узлов, $k > n$. Проводится сравнение с оценкой погрешности классической формулы, основанной на многочлене Лагранжа.

Отметим, что при наличии экспоненциального пограничного слоя погрешность классических полиномиальных разностных формул для вычисления производных становится равномерной по малому параметру ε, если их применять на сетках, сгущающихся в области пограничного слоя. В [9] это обосновано в случае сетки Шишкина [10], в [11] применена сетка Бахвалова [12] с модификацией из [13].

Под C и C_j будем подразумевать положительные постоянные, не зависящие от пограничной составляющей $\Phi(x)$, ее производных и от шага сетки. В случае экспоненциального погранслоя эти постоянные не зависят от ε и h. Различные величины будем ограничивать одной постоянной C_j, если понятно по тексту. Будем подразумевать, что $f = O(g)$, если для некоторой постоянной C $|f| \le C|g|$; $f = O^*(g)$, если $f = O(g)$ и $g = O(f)$.

2. ПОСТАНОВКА ЗАДАЧИ

Пусть для достаточно гладкой функции $u(x)$ справедлива декомпозиция

$$u(x) = p(x) + \gamma\Phi(x),\ x \in [0,1], \tag{1}$$

где $p(x)$ — регулярная составляющая с ограниченными производными до некоторого порядка, $\Phi(x)$ — пограничная составляющая, являющаяся функцией общего вида и отвечающая за большие градиенты функции $u(x)$. Функция $\Phi(x)$ предполагается известной, $p(x)$ и γ не заданы, $|\gamma| \le C$ для некоторой постоянной C.

В частности, рассмотрим случай экспоненциального пограничного слоя, когда функция $u(x)$ является решением сингулярно возмущенной краевой задачи:

$$\varepsilon u''(x) + a_1(x)u'(x) - a_2(x)u(x) = f(x),\ \ u(0) = A,\ u(1) = B, \tag{2}$$

где $a_1(x) \ge \beta > 0, a_2(x) \ge 0,\ \varepsilon \in (0,1]$, функции a_1, a_2, f — достаточно гладкие.

Согласно [4], для решения задачи (2) и произвольно задаваемого n_0 справедлива декомпозиция (1), для которой

$$\Phi(x) = e^{-\alpha x/\varepsilon},\ \alpha = a_1(0), \tag{3}$$

$$|p^{(n)}(x)| \le C_0\left[1 + \frac{1}{\varepsilon^{n-1}}e^{-\beta x/\varepsilon}\right], n \le n_0,\ \gamma = -\varepsilon u'(0)/a_1(0). \tag{4}$$

В соответствии с (4), производная $p'(x)$ ограничена равномерно по параметру ε.

При наличии степенного пограничного слоя декомпозиция (1) справедлива при задании $\Phi(x) = (x+\varepsilon)^\alpha,\ 0 < \alpha < 1, 0 < \varepsilon \le 1$.

Зададим равномерную сетку интервала $[0,1]$:

$$\Omega^h = \{x_j :\ x_j = jh, j = 0,1,\dots,N,\ Nh = 1\}.$$

Предполагаем, что функция $u(x)$, обладающая декомпозицией (1), задана в узлах сетки Ω^h, $u_j = u(x_j)$. Рассмотрим вопрос численного дифференцирования такой функции на произвольном интервале $[x_m, x_{m+k-1}]$ с k узлами сетки Ω^h.

Пусть $L_k(u,x)$ — многочлен Лагранжа для функции $u(x)$ с k узлами интерполяции $x_m,\dots,x_{m+k-1}$. Производные функции $u(x)$, как известно [1], можно приближенно находить на основе дифференцирования многочлена Лагранжа: $u^{(n)}(x) \approx L_k^{(n)}(u,x), x \in [x_m, x_{m+k-1}]$. В соответствии, например, с [14] справедлива оценка погрешности:

$$|u^{(n)}(x) - L_k^{(n)}(u,x)| \le \frac{M_k(k-1)^{k-n}h^{k-n}}{(k-n)!}, \tag{5}$$

где $M_k = \max\limits_s |u^{(k)}(s)|,\ x, s \in [x_m, x_{m+k-1}]$.

Из (5) следует, что погрешность вычисления производных на основе дифференцирования многочлена Лагранжа порядка $O(h^{k-n})$, если величина M_k равномерно ограничена. Однако в случае экспоненциального пограничного слоя в соответствии с (3) $M_k = O^*(\varepsilon^{-k})$, поэтому при малых значениях ε погрешность может быть существенной.

Покажем это на примере. Пусть $u(x) = e^{-x/\varepsilon}$, $x \in [0, 1]$. Выпишем формулу для производной:

$$u'(x) \approx \frac{u_j - u_{j-1}}{h}, \ x \in [x_{j-1}, x_j]. \tag{6}$$

В данном случае производная $u'(x)$ в области пограничного слоя порядка $O(1/\varepsilon)$, поэтому в соответствии, например, с [15], [16], [9] оценивается относительная погрешность при вычислении первой производной, получаемая умножением абсолютной погрешности на малый параметр ε. В случае формулы (6) на равномерной сетке при $\varepsilon = h$ имеем

$$\varepsilon\Big|\frac{u_1 - u_0}{h} - u'(0)\Big| = e^{-1}. \tag{7}$$

Таким образом, относительная погрешность формулы (6) значительна, если $\varepsilon = h$, несмотря на малость h. Задача построения формул численного дифференцирования для функций, имеющих представление (1), актуальна.

Интерполяционная формула из [3]

$$L_{\Phi,k}(u, x) = L_{k-1}(u, x) + \frac{[x_m, \dots, x_{m+k-1}]u}{[x_m, \dots, x_{m+k-1}]\Phi}\Big[\Phi(x) - L_{k-1}(\Phi, x)\Big], \ x \in [x_m, x_{m+k-1}], \tag{8}$$

в соответствии с [5] представима в виде

$$L_{\Phi,k}(u, x) = L_k(u, x) + \frac{[x_m, \dots, x_{m+k-1}]u}{[x_m, \dots, x_{m+k-1}]\Phi}\Big[\Phi(x) - L_k(\Phi, x)\Big], \ x \in [x_m, x_{m+k-1}], \tag{9}$$

где $[x_m, \dots, x_{m+k-1}]u$ — разделенная разность [1] для функции $u(x)$, $k \geq 2$. Для корректного задания формулы (9) задаем ограничение

$$\Phi^{(k-1)}(x) \neq 0, \ x \in (x_m, x_{m+k-1}).$$

Из (9) следует, что эта формула является интерполяционной с узлами интерполяции $x_m, x_{m+1}, \dots, x_{m+k-1}$ и точной на многочленах степени $(k-2)$ и на функции $\Phi(x)$.

Дифференцируя (9), получаем формулу численного дифференцирования:

$$u^{(n)}(x) \approx L^{(n)}_{\Phi,k}(u, x) = L^{(n)}_k(u, x) + \frac{\Delta^{k-1}u_m}{\Delta^{k-1}\Phi_m}\Big[\Phi^{(n)}(x) - L^{(n)}_k(\Phi, x)\Big], \tag{10}$$

где $x \in [x_m, x_{m+k-1}]$, $n < k$, $\Delta^{k-1}u_m$ — конечная разность для $u(x)$ [1, с. 65], определяемая соотношениями $\Delta u_m = u_{m+1} - u_m$, $\Delta^j u_m = \Delta(\Delta^{j-1}u_m)$.

Отметим, что формула (10) является точной на составляющей $\Phi(x)$.

Целью работы является оценивание погрешности формулы (10), применяемой к функции $u(x)$ вида (1).

3. ОЦЕНКА ПОГРЕШНОСТИ ФОРМУЛЫ ДЛЯ ПРОИЗВОДНОЙ

Теорема. *Пусть для функции $u(x)$ справедлива декомпозиция* (1). *Пусть D_k такое, что $0 < 1/D_k \leq C_1$ и для некоторой постоянной C_2 справедлива оценка*

$$G = \frac{h^{k-2}}{D_k}\int\limits_{x_m}^{x_{m+k-1}} \big|\Phi^{(k)}(t)\big| \, dt \Big/ \big|\Delta^{k-1}\Phi_m\big| \leq C_2. \tag{11}$$

Тогда найдется C_3 такое, что справедлива оценка:

$$\frac{1}{D_k}\left|L^{(n)}_{\Phi,k}(u,x)-u^{(n)}(x)\right| \le \frac{1}{D_k}\left|L^{(n)}_{k}(p,x)-p^{(n)}(x)\right| + \frac{C_3}{h^{n-1}}\left|L_{k-1}(p,x_{m+k-1})-p(x_{m+k-1})\right|, \tag{12}$$

где $x \in [x_m, x_{m+k-1}]$, $1 \le n < k$.

Доказательство. Формула (10) является точной на $\Phi(x)$, поэтому

$$\left|L^{(n)}_{\Phi,k}(u,x)-u^{(n)}(x)\right| = \left|L^{(n)}_{\Phi,k}(p,x)-p^{(n)}(x)\right| \le \left|L^{(n)}_{k}(p,x)-p^{(n)}(x)\right| +$$

$$+\left|\frac{\Delta^{k-1}p_m}{\Delta^{k-1}\Phi_m}\left(\Phi^{(n)}(x)-L^{(n)}_{k}(\Phi,x)\right)\right|, \ x \in [x_m, x_{m+k-1}]. \tag{13}$$

В (13) оценим $\left|\Phi^{(n)}(x)-L^{(n)}_{k}(\Phi,x)\right|$. Сначала получим оценку такой погрешности в случае достаточно гладкой функции $v(x)$. Используем разложение в ряд Тейлора с остаточным членом в интегральном виде:

$$v(x) = P_k(x) + R_k(x), \tag{14}$$

где

$$P_k(x) = v(x_m) + v'(x_m)(x-x_m) + \cdots + v^{(k-1)}(x_m)\frac{(x-x_m)^{k-1}}{(k-1)!},$$

$$R_k(x) = \frac{1}{(k-1)!}\int\limits_{x_m}^{x}(x-t)^{k-1}v^{(k)}(t)\,dt.$$

Следовательно,

$$R_k(x) = \frac{1}{(k-1)!}\int\limits_{x_m}^{x_{m+k-1}}(x-t)_+^{k-1}v^{(k)}(t)\,dt, \tag{15}$$

где $(x-t)_+^{k-1} = (x-t)^{k-1}$ при $x \ge t$ и $(x-t)_+^{k-1} = 0$ при $x < t$.

Для погрешности интерполяции многочленом Лагранжа на интервале $[x_m, x_{m+k-1}]$ известна оценка [1]:

$$|v(x)-L_k(v,x)| \le \max_s |v^{(k)}(s)||w_k(x)|/k!, \ \ w_k(x) = (x-x_m)\cdots(x-x_{m+k-1}). \tag{16}$$

Учитывая (16), имеем $P_k(x) - L_k(P_k,x) = 0$. Учитывая (14), получаем

$$v(x) - L_k(v,x) = R_k(x) - L_k(R_k,x).$$

Следовательно,

$$v(x) - L_k(v,x) = R_k(x) - \sum_{j=m}^{m+k-1} R_k(x_j)\prod_{i=m,\ i\ne j}^{m+k-1}\frac{x-x_i}{x_j-x_i}. \tag{17}$$

Учитывая (15) и дифференцируя (17), для некоторой постоянной C_1 получаем

$$\left|v^{(n)}(x)-L^{(n)}_{k}(v,x)\right| \le C_1 h^{k-n-1}\int\limits_{x_m}^{x_{m+k-1}}\left|v^{(k)}(t)\right|\,dt, \ \ x \in \left[x_m, x_{m+k-1}\right]. \tag{18}$$

В случае $v(x) = \Phi(x)$ из (18) получаем

$$\left|\Phi^{(n)}(x)-L^{(n)}_{k}(\Phi,x)\right| \le C_1 h^{k-n-1}\int\limits_{x_m}^{x_{m+k-1}}\left|\Phi^{(k)}(t)\right|\,dt, \ \ x \in \left[x_m, x_{m+k-1}\right]. \tag{19}$$

Теперь из (13), (19) для некоторой постоянной C получаем

$$\frac{1}{D_k}\left|L^{(n)}_{\Phi,k}(u,x)-u^{(n)}(x)\right| \le \frac{1}{D_k}\left|L^{(n)}_{k}(p,x)-p^{(n)}(x)\right| + \\ +\frac{C}{D_k}\left|\frac{\Delta^{k-1}p_m}{\Delta^{k-1}\Phi_m}\right| h^{k-n-1}\int\limits_{x_m}^{x_{m+k-1}}\left|\Phi^{(k)}(t)\right|\,dt. \tag{20}$$

Преобразуем $\Delta^{k-1}p_m$. Учитывая соотношения [1]

$$\Delta^{k-1}p_m = (k-1)!h^{k-1}[x_m,\dots,x_{m+k-1}]p,$$

$$p(x)-L_{k-1}(p,x) = w_{k-1}(x)[x_m,\dots,x_{m+k-2},x]p,$$

где $w_{k-1}(x)$ соответствует (16), получаем

$$\left|\Delta^{k-1}p_m\right| = \left|p(x_{m+k-1})-L_{k-1}(p,x_{m+k-1})\right|. \tag{21}$$

Учитывая (11), (21) в (20), получаем (12). Теорема доказана.

Следствие. Согласно (12), оценка погрешности предложенной формулы численного дифференцирования (10) сведена к оценке погрешности многочлена Лагранжа на регулярной составляющей $p(x)$. Преобразуем эту оценку, применяя оценку (18) в случае $v(x) = p(x)$. Тогда из (12) для некоторой постоянной C получаем

$$\frac{1}{D_k}\left|L^{(n)}_{\Phi,k}(u,x)-u^{(n)}(x)\right| \le Ch^{k-n-1}\int\limits_{x_m}^{x_{m+k-1}}|p^{(k)}(t)|\,dt,\ \ x\in[x_m,x_{m+k-1}]. \tag{22}$$

Если производная $p^{(k)}(x)$ является равномерно ограниченной, то из (22) для некоторой постоянной C_1 следует

$$\frac{1}{D_k}\left|L^{(n)}_{\Phi,k}(u,x)-u^{(n)}(x)\right| \le C_1h^{k-n},\ x\in[x_m,x_{m+k-1}]. \tag{23}$$

В регулярном случае, когда производные функции $u(x)$ являются равномерно ограниченными, в соответствии с (11) можно задать $D_k = 1$. В соответствии с (23) оценка погрешности формулы (10) в этом случае такая же, как при применении классической формулы, основанной на дифференцировании многочлена Лагранжа.

4. СЛУЧАЙ ЭКСПОНЕНЦИАЛЬНОГО ПОГРАНИЧНОГО СЛОЯ

Рассмотрим применение теоремы в случае экспоненциального пограничного слоя, когда для функции $u(x)$ в декомпозиции (1) справедливы соотношения (3), (4).

Оценка погрешности классической формулы. При задании $\Phi(x)$ в соответствии с (3) для некоторой постоянной C оценка (19) принимает вид:

$$\varepsilon^{k-1}|\Phi^{(n)}(x)-L^{(n)}_k(\Phi,x)| \le Ch^{k-n-1}e^{-\alpha x_m/\varepsilon}\left(1-e^{-(k-1)\alpha h/\varepsilon}\right),\ x\in[x_m,x_{m+k-1}]. \tag{24}$$

Учитывая оценки (4) в (18) при задании $v(x) = p(x)$, получаем

$$\varepsilon^{k-1}|p^{(n)}(x)-L^{(n)}_k(p,x)| \le Ch^{k-n},\ x\in[x_m,x_{m+k-1}]. \tag{25}$$

Учитывая (1), (24) и (25), получаем оценку погрешности:

$$\varepsilon^{k-1}|u^{(n)}(x)-L^{(n)}_k(u,x)| \le C\left[h^{k-n}+h^{k-n-1}e^{-\alpha x_m/\varepsilon}\left(1-e^{-(k-1)\alpha h/\varepsilon}\right)\right]. \tag{26}$$

Из (26) при всех $x \in [x_m, x_{m+k-1}]$ имеем

$$\varepsilon^{k-1}|u^{(n)}(x) - L_k^{(n)}(u,x)| \le Ch^{k-n}\left[1 + \frac{1}{\varepsilon}e^{-\alpha x_m/\varepsilon}\right],\ (k-1)\alpha h/\varepsilon < 1, \tag{27}$$

$$\varepsilon^{k-1}|u^{(n)}(x) - L_k^{(n)}(u,x)| \le Ch^{k-n}\left[1 + \frac{1}{h}e^{-\alpha x_m/\varepsilon}\right],\ (k-1)\alpha h/\varepsilon \ge 1. \tag{28}$$

Из (28) следует, что в случае $k-1=n,\ m=0$, относительная погрешность $\varepsilon^n|u^{(n)}(x) - L_k^{(n)}(u,x)|$ может быть порядка $O(1)$, что подтверждается примером (7).

Оценка погрешности формулы (10). В случае $\Phi(x) = e^{-\alpha x/\varepsilon}$ величина G из (11) принимает вид:

$$G = \frac{h^{k-2}\alpha^{k-1}}{D_k\varepsilon^{k-1}}\left(e^{-\alpha x_m/\varepsilon} - e^{-\alpha x_{m+k-1}/\varepsilon}\right)\Big/|\Delta^{k-1}e^{-\alpha x_m/\varepsilon}|. \tag{29}$$

В соответствии с [1, с. 66], справедлива формула

$$\Delta^{k-1}\Phi_m = \sum_{j=0}^{k-1}(-1)^j C_{k-1}^j \Phi_{m+k-1-j}.$$

Учитывая (3), имеем

$$\Delta^{k-1}\Phi_m = \sum_{j=0}^{k-1}(-1)^j C_{k-1}^j e^{-\alpha x_{m+k-1-j}/\varepsilon}.$$

Тогда G из (29) принимает вид:

$$G = \frac{h^{k-2}\alpha^{k-1}}{D_k\varepsilon^{k-1}}\frac{1 - e^{-\alpha h(k-1)/\varepsilon}}{\left|\sum\limits_{j=0}^{k-1}(-1)^j C_{k-1}^j e^{-\alpha h(k-1-j)/\varepsilon}\right|} = \frac{h^{k-2}\alpha^{k-1}}{D_k\varepsilon^{k-1}}\frac{1 - e^{-\alpha h(k-1)/\varepsilon}}{\left|\sum\limits_{i=0}^{k-1}(-1)^{k-1-i} C_{k-1}^i (e^{-\alpha h/\varepsilon})^i\right|},$$

где $i = k-1-j$. Учитывая разложение для $(1-x)^{k-1}$, получаем

$$G = \frac{h^{k-2}\alpha^{k-1}}{D_k\varepsilon^{k-1}} \times \frac{1 - e^{-\alpha h(k-1)/\varepsilon}}{(1 - e^{-\alpha h/\varepsilon})^{k-1}} = \frac{h^{k-2}\alpha^{k-1}}{D_k\varepsilon^{k-1}} \times \frac{1 + e^{-\alpha h/\varepsilon} + \cdots + e^{-\alpha(k-2)h/\varepsilon}}{(1 - e^{-\alpha h/\varepsilon})^{k-2}}.$$

Следовательно, для некоторой постоянной C_4 имеем

$$G \le C_4\frac{h^{k-2}}{D_k\varepsilon^{k-1}} \times \frac{1}{(1 - e^{-\alpha h/\varepsilon})^{k-2}}.$$

Следовательно, условие (11) выполняется, если D_k задавать на основе выполнения неравенства

$$D_k \ge C\frac{h^{k-2}}{\varepsilon^{k-1}} \times \frac{1}{(1 - e^{-\alpha h/\varepsilon})^{k-2}}. \tag{30}$$

Рассмотрим два случая для значения $\alpha h/\varepsilon$.

1. Пусть $\alpha h/\varepsilon < 1$. Тогда неравенство (30) выполняется для некоторой постоянной C при задании $D_k = 1/\varepsilon$. При этом выполнено (11) и в соответствии с теоремой справедлива оценка (12), поэтому

$$\varepsilon\left|L_{\Phi,k}^{(n)}(u,x) - u^{(n)}(x)\right| \le \varepsilon\left|L_k^{(n)}(p,x) - p^{(n)}(x)\right| + \frac{C_3}{h^{n-1}}\left|L_{k-1}(p, x_{m+k-1}) - p(x_{m+k-1})\right|.$$

Умножая это неравенство на ε^{k-2} и учитывая (25) с применением $k-1$ вместо k для второго слагаемого, для некоторой постоянной C_1 получаем

$$\varepsilon^{k-1}\left|L_{\Phi,k}^{(n)}(u,x) - u^{(n)}(x)\right| \le C_1 h^{k-n},\ x \in [x_m, x_{m+k-1}],\ k > n > 0. \tag{31}$$

2. Пусть $\alpha h/\varepsilon \geq 1$. Тогда условие (30) выполнено для некоторого C, если задать

$$D_k = \frac{h^{k-2}}{\varepsilon^{k-1}}. \tag{32}$$

Тогда выполнено условие (11), и по теореме справедлива оценка (12). Применяя в этой оценке (32), получаем

$$\varepsilon^{k-1}\left|L^{(n)}_{\Phi,k}(u,x) - u^{(n)}(x)\right| \leq \varepsilon^{k-1}\left|L^{(n)}_{k}(p,x) - p^{(n)}(x)\right| + \frac{C_3 h^{k-2}}{h^{n-1}}\left|L_{k-1}(p, x_{m+k-1}) - p(x_{m+k-1})\right|. \tag{33}$$

Покажем, что в (33) для некоторой постоянной C_5

$$\left|L_{k-1}(p, x_{m+k-1}) - p(x_{m+k-1})\right| \leq C_5 h. \tag{34}$$

В соответствии с [1], для произвольного $x \in [x_m, x_{m+k-1}]$ имеем

$$L_{k-1}(p,x) = \sum_{j=m}^{m+k-2} p(x_j) \prod_{i=m, i\neq j}^{m+k-2} R_i,\ R_i = \frac{x - x_i}{x_j - x_i},\ \sum_{j=m}^{m+k-2}\prod_{i=m,i\neq j}^{m+k-2} R_i = 1.$$

Учитывая, что в (4) $|p'(x)| \leq 2C_0$, для некоторой постоянной C_5 имеем

$$\left|L_{k-1}(p,x) - p(x)\right| = \left|\sum_{j=m}^{m+k-2}\left[p(x_j) - p(x)\right]\prod_{i=m,i\neq j}^{m+k-2} R_i\right| \leq C_5 h.$$

Оценка (34) доказана. Применяя оценки (25), (34) в (33), получаем оценку (31).

Итак, для классической разностной формулы, основанной на дифференцировании многочлена Лагранжа, получены оценки погрешности (27), (28), а для формулы (10), точной на погранслойной составляющей, получена оценка погрешности (31). Из сравнения этих оценок следует, что формула подгонки к погранслойной составляющей является более точной. Как показано, погрешность классической формулы может быть порядка $O(1)$. В полученных оценках погрешностей абсолютная погрешность умножена на ε^{k-1}, и это согласуется с оценками (5), (25). Оценка (25) существенно применяется при обосновании (31).

Случай $k - 1 = n$**.** Классические формулы для вычисления производных, основанные на дифференцировании многочлена Лагранжа, в этом случае имеют вид [1]:

$$u^{(n)}(x) \approx \frac{\Delta^n u_m}{h^n},\ x \in [x_m, x_{m+n}].$$

Покажем, что формула (10) при $k - 1 = n$ существенно упрощается. Дифференцируя (8), получаем

$$u^{(n)}(x) \approx L^{(n)}_{\Phi,k}(u,x) = \frac{\Delta^n u_m}{\Delta^n \Phi_m}\Phi^{(n)}(x),\ x \in [x_m, x_{m+n}].$$

Случай $k - 1 = n$ широко применяется при аппроксимации производных, при этом полученные оценки погрешностей (27), (28), (31) переходят в оценки относительной погрешности.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

На интервале $[0, 1]$ зададим функцию

$$u(x) = \cos(\pi x) + e^{-x/\varepsilon}.$$

Здесь $\Phi(x) = e^{-x/\varepsilon}$. Сетку задаем равномерной. На каждом сеточном интервале $[x_{j-1}, x_j]$ применяем классическую формулу

$$u'(x) \approx L'_2(u,x) = \frac{u_j - u_{j-1}}{h}$$

и формулу, точную на погранслойной составляющей

$$u'(x) \approx L'_{\Phi,2}(u,x) = \frac{u_j - u_{j-1}}{\Phi_j - \Phi_{j-1}}\Phi'(x). \tag{35}$$

В табл. 1 приведена погрешность

$$\Delta_{\varepsilon,N} = \varepsilon \max_j |L'_2(u,\tilde{x}_j) - u'(\tilde{x}_j)|,$$

где $\tilde{x}_j$ — узлы сгущенной в четыре раза сетки. Согласно результатам вычислений, точность не повышается с уменьшением шага h, если $\varepsilon = h$.

В табл. 2 приведена погрешность вычислений по формуле (35)

$$\Delta_{\varepsilon,N} = \varepsilon \max_j |L'_{\Phi,2}(u,\tilde{x}_j) - u'(\tilde{x}_j)|.$$

Результаты вычислений при всех ε и N согласуются с погрешностью порядка $O(h)$, что соответствует оценке (31).

Результаты других численных экспериментов по применению формулы (10) приведены в [6]–[8].

Таблица 1. Погрешность классической формулы для вычисления первой производной

ε	N					
	16	32	64	128	256	512
1	$2.25e-3$	$5.68e-4$	$1.42e-4$	$3.57e-5$	$8.92e-6$	$2.23e-6$
16^{-1}	$6.67e-2$	$2.56e-2$	$8.14e-3$	$2.30e-3$	$6.12e-4$	$1.58e-4$
32^{-1}	$1.26e-1$	$6.67e-2$	$2.56e-2$	$8.14e-3$	$2.30e-3$	$6.12e-4$
64^{-1}	$1.32e-1$	$1.26e-1$	$6.67e-2$	$2.56e-2$	$8.14e-3$	$2.30e-3$
128^{-1}	$1.04e-1$	$1.32e-1$	$1.26e-1$	$6.67e-2$	$2.56e-3$	$8.14e-3$
256^{-1}	$6.71e-2$	$1.04e-1$	$1.32e-1$	$1.26e-1$	$6.67e-2$	$2.56e-2$
512^{-1}	$3.90e-2$	$6.71e-2$	$1.04e-1$	$1.32e-1$	$1.26e-1$	$6.67e-2$

Таблица 2. Погрешность вычисления первой производной по формуле (35)

ε	N					
	16	32	64	128	256	512
1	$2.98e-1$	$1.49e-1$	$7.43e-2$	$3.71e-2$	$1.85e-2$	$9.27e-3$
16^{-1}	$1.11e-1$	$5.17e-2$	$2.48e-2$	$1.22e-2$	$6.02e-3$	$3.00e-3$
32^{-1}	$1.23e-1$	$5.47e-2$	$2.55e-2$	$1.23e-2$	$6.01e-3$	$2.97e-3$
1024^{-1}	$1.84e-1$	$9.08e-2$	$4.39e-2$	$2.05e-2$	$9.01e-3$	$3.85e-3$
2048^{-1}	$1.59e-1$	$8.13e-2$	$3.95e-2$	$1.90e-2$	$8.83e-3$	$3.87e-3$

6. ЗАКЛЮЧЕНИЕ

Рассмотрен вопрос численного дифференцирования функций с большими градиентами на равномерной сетке. Предполагается, что для исходной функции одной переменной справедлива декомпозиция в виде суммы регулярной и погранслойной составляющих. Погранслойная составляющая отвечает за большие градиенты функции и задается как функция общего вида, известная с точностью до множителя, может отражать различные особенности дифференцируемой функции. Получены оценки погрешности классических формул, основанных на дифференцировании многочлена Лагранжа и предложенных формул, по построению точных на погранслойной составляющей. Эти оценки зависят от порядка вычисляемой производной и от числа узлов в сеточном шаблоне для этой производной. Выделен случай, когда дифференцируемая функция соответствует решению краевой задачи

при наличии экспоненциального пограничного слоя. В этом случае получены оценки погрешности, равномерные по малому параметру. Как по оценкам погрешностей, так и по результатам численных экспериментов получено преимущество в точности разработанных формул.

СПИСОК ЛИТЕРАТУРЫ

1. *Бахвалов Н. С., Жидков Н. П., Кобельков Г. М.* Численные методы. М.: Наука, 1987.
2. *Задорин А. И.* Метод интерполяции для задачи с пограничным слоем // Сиб. ж. вычисл. матем. 2007. Т. 10. N 3. С. 267–275.
3. *Zadorin A. I., Zadorin N. A.* Interpolation formula for functions with a boundary layer component and its application to derivatives calculation // Sib. Electron. Math. Rep. 2012. V. 9. P. 445–455.
4. *Kellogg R. B., Tsan A.* Analysis of some difference approximations for a singular perturbation problems without turning points // Math. Comput. 1978. V. 32. P. 1025–1039.
5. *Задорин А. И., Задорин Н. А.* Неполиномиальная интерполяция функций с большими градиентами и ее применение // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. N 2. С. 179–188.
6. *Il'in V. P., Zadorin A. I.* Adaptive formulas of numerical differentiation of functions with large gradients // J. Phys.: Conf. Ser. 2019. V. 1260. 042003.
7. *Zadorin A., Tikhovskaya S.* Formulas of numerical differentiation on a uniform mesh for functions with the exponential boundary layer // Internat. J. Numer. Anal. Model. 2019. V. 16. N 4. P. 590–608.
8. *Задорин А. И.* Формулы численного дифференцирования функций с большими градиентами // Сиб. ж. вычисл. матем. 2023. Т. 26. N 1. С. 17–26.
9. *Задорин А. И.* Анализ формул численного дифференцирования на сетке Шишкина при наличии пограничного слоя // Сиб. ж. вычисл. матем. 2018. Т. 21. N 3. С. 243–254.
10. *Шишкин Г. И.* Сеточные аппроксимации сингулярно возмущенных эллиптических и параболических уравнений. Екатеринбург: УрО РАН, 1992.
11. *Задорин А. И.* Анализ формул численного дифференцирования на сетке Бахвалова при наличии пограничного слоя // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. N 2. С. 218–226.
12. *Бахвалов Н. С.* К оптимизации методов решения краевых задач при наличии пограничного слоя // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. N 4. С. 841–890.
13. *Roos H. G.* Layer-adapted meshes: milestones in 50 years of history // Appl. Math. arXiv:1909.08273v1, 2019.
14. *Даутов Р. З., Тимербаев М. Р.* Численные методы. Приближение функций: учебное пособие. Казань: Казан. ун-т, 2021.
15. *Kopteva N. V., Stynes M.* Approximation of derivatives in a convection-diffusion two-point boundary value problem // Appl. Numer. Math. 2001. V. 39. P. 47–60.
16. *Shishkin G. I.* Approximations of solutions and derivatives for a singularly perturbed elliptic convection-diffusion equations // Math. Proc. Royal Irish Acad. 2003. V. 103A. N 4. P. 169–201.

FORMULAS FOR NUMERICAL DIFFERENTIATION ON A UNIFORM GRID IN THE PRESENCE OF A BOUNDARY LAYER

A. I. Zadorin*

Institute of Mathematics, Siberian Branch, Russian Academy of Sciences, 4 Acad. Koptyug av. Novosibirsk, 630090 Russia
**e-mail: zadorin@ofim.oscsbras.rv*

Received 10 October, 2023
Revised 28 December, 2023
Accepted 05 March, 2024

Abstract. The problem of numerical differentiation of functions with large gradients is considered. It is assumed that for the original function of one variable the decomposition is valid as the sum of a regular component with bounded derivatives up to a certain order and a boundary layer component having large gradients and known with an accuracy of up to a factor. Such a decomposition, in particular, is valid for solution of a singularly perturbed boundary value problem. The topic of the study is relevant, since the application of classical polynomial formulas of numerical differentiation to functions with large gradients can lead to significant errors. The error of the formulas of numerical differentiation, according to the construction of exact ones on the boundary layer component of the original function, is estimated. The results of numerical experiments are presented, consistent with the obtained error estimates.

Keywords: function of one variable, large gradients, special formula for numerical differentiation, error estimate.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 932–939

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 517.951

СИММЕТРИИ И ДЕКОМПОЗИЦИЯ СИСТЕМ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ И СИСТЕМ УПРАВЛЕНИЯ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ

© 2024 г. В. И. Елкин[1,*]

[1] *117333 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия*

[*] *e-mail: elk_v@mail.ru*

Поступила в редакцию: 08.08.2023 г.
Переработанный вариант 31.01.2024 г.
Принята к публикации 05.03.2024 г.

Рассматривается вопрос о симметриях уравнений с частными производными на основе использования дифференциально-геометрических и алгебраических методов теории динамических систем с управлением. Библ. 6.

Ключевые слова: группа симметрий, алгебра Ли.

DOI: 10.31857/S0044466924060048, **EDN:** XYXTRL

ВВЕДЕНИЕ

Рассматривается система дифференциальных уравнений с частными производными первого порядка

$$\Lambda_\nu(t, y, p) = 0, \quad \nu = 1, \dots, l, \tag{1}$$

После приведения этой системы к специальному виду в параметрической форме, разрешенной относительно всех производных

$$\partial_k y^i = g^i_k(t, y, u), \quad \partial_k = \partial/\partial_{t^k}, \tag{2}$$

открывается возможность применения дифференциально-геометрических и алгебраических методов теории динамических систем с управлением, используя некоторую аналогию данных объектов. Эти методы позволяют исследовать некоторые вопросы декомпозиции, построения симметрий и др. В [1] на основе понятия первого интеграла с помощью замены координат, в которую входят первые интегралы, была получена простая декомпозиция вида

$$\partial_k z^i = 0,\ i = 1, \dots, q, \quad \partial_k z^j = h^j_k(t, x, u),\ j = q+1, \dots, n. \tag{3}$$

В [2] рассматривается более общая декомпозиция, которая также имеет аналог в теории динамических систем с управлением под названием агрегирование (факторизация). Здесь исследуются симметрии, т.е. преобразования переменных, переводящие решения в решения. Кроме практического смысла тиражирования решений, симметрии определяют также некоторую декомпозицию уравнений с частными производными.

1. О СИММЕТРИЯХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Симметрии дифференциальных уравнений — это такие преобразования зависимых и независимых переменных, которые переводят решения дифференциальных уравнений в решения. Таким образом, симметрии позволяют находить новые решения по известным решениям. Но этим не ограничивается роль симметрий. Различные качественные характеристики дифференциальных уравнений определяются существованием тех или иных симметрий. Например, для обыкновенных дифференциальных уравнений симметрии помогают находить первые интегралы, которые в случае уравнений механики дают законы сохранения. Для уравнений с частными производными математической физики симметрии также поставляют законы сохранения, которые имеют более сложный вид. Знание симметрий позволяет преобразовать дифференциальные уравнения к удобному для исследований виду, например, к некоторой декомпозиции. Аппарат симметрий для дифференциальных уравнений был разработан еще в 19 веке С. Ли в его теории непрерывных групп, которые стали называть группами Ли. Л. В. Овсянников возродил эту теорию под наименованием групповой анализ [3]. Немного подробнее.

Рассмотрим систему $\mathscr{L}$ дифференциальных уравнений первого порядка (1), включающую m независимых переменных $t = (t^1, \dots, t^m)$ и n зависимых переменных $y = (y^1, \dots, y^n)$. Решением системы $\mathscr{L}$ является гладкая функция $y = \Phi(t)$, удовлетворяющая системе $\mathscr{L}$. Будем отождествлять эту функцию, заданную в области $I \subset \mathbb{R}^m$ изменения переменных t, с ее графиком, т.е. с множеством точек

$$N = \{(t, y) \in R^d : \ y = \Phi(t)\}, \quad d = m + n.$$

Этот график является m-мерным многообразием в $\mathbb{R}^d$. Рассмотрим некоторый диффеоморфизм (т.е. взаимно однозначное отображение, гладкое в обе стороны) $\psi : \ M \to M$, где M — область в $\mathbb{R}^d$, причем $N \subset M$. Под действием диффеоморфизма ψ многообразие N перейдет в некоторое многообразие $\hat{N} = \psi(N)$. Многообразие $\hat{N}$ не является, вообще говоря, графиком какой-либо однозначной функции $y = \hat{\Phi}(t)$. Понятие симметрии относится к случаю, когда $\hat{N}$ — график. Итак, будем говорить, что диффеоморфизм ψ является симметрией системы $\mathscr{L}$, если каждое решение $y = \Phi(x)$ системы, трактуемое как многообразие N, переводится этим диффеоморфизмом в многообразие $\hat{N} = \psi(N)$, являющееся графиком некоторой функции $y = \hat{\Phi}(t)$, представляющей собой решение системы $\mathscr{L}$. Естественным образом понятие симметрии обобщается для случая локального диффеоморфизма ψ.

Перейдем теперь к группам симметрий. Пусть S — локальная однопараметрическая группа диффеоморфизмов некоторой области в R^d, состоящая из преобразований s^τ вида

$$t' = h(t, y, \tau), \quad y' = g(t, y, \tau), \quad \tau \in R^1. \tag{4}$$

Под действием преобразования s^τ группы S многообразие N, являющееся графиком функции $y = \Phi(t)$, перейдет в многообразие $\hat{N}$, которое в параметрической форме (параметр t) задается уравнениями

$$t' = h(t, \Phi(t), \tau), \tag{5}$$

$$y' = g(t, \Phi(t), \tau). \tag{6}$$

Так как $h(t, y, 0) = t$, то (по крайней мере для малых τ) уравнения (5) можно разрешить относительно t и подставить в (6). В результате получим представление многообразия $\hat{M}$ в виде графика функции $y' = \hat{\Phi}(t')$. Будем говорить, что эта функция получается из функции $y = \Phi(t)$ действием преобразования s^τ. Группа S называется группой симметрий, если такие функции, получаемые действием преобразований группы на решения, снова являются решениями системы $\mathscr{L}$, т.е. преобразования s^τ — симметрии системы $\mathscr{L}$. Заметим, что симметрии часто называют допускаемыми преобразованиями, а группы симметрий — допускаемыми группами системы $\mathscr{L}$. Говорят также, что система $\mathscr{L}$ допускает преобразование или группу преобразований.

При рассмотрении вопроса о симметриях вместо нахождения конечных преобразований (5), (6) удобнее сначала отыскивать векторные поля (инфинитезимальные операторы)

$$X = \xi^k(t, y)\frac{\partial}{\partial t^k} + \eta^i(t, y)\frac{\partial}{\partial y^i}, \tag{7}$$

которые порождают однопараметрические группы симметрий (5), (6). Здесь и далее производится суммирование по повторяющемуся индексу. Более старое название для (7) — инфинитезимальные симметрии или допускаемые поля. Нахождение полей (7) осуществляется решением некоторых систем дифференциальных уравнений (так называемых определяющих уравнений). Эти уравнения находятся с помощью техники продолжения [3]. Для этой цели нужно перейти в так называемое расширенное пространство переменных, которое наряду с зависимыми и независимыми переменными содержит еще в качестве переменных производные зависимых переменных по независимым

$$p^i_k = \frac{\partial \Phi^i}{\partial t^k}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, m.$$

В качестве примера приведем определяющие уравнения для системы обыкновенных дифференциальных уравнений

$$\dot y = f(t, y), \quad (t, y) \in R^{n+1}. \tag{8}$$

Это — равенства

$$X_0\eta^i - f^i X_0\xi = X f^i, \quad i = 1, \ldots, n, \tag{9}$$

где X_0 — векторное поле(оператор полного дифференцирования в силу системы), имеющее вид

$$X_0 = \frac{\partial}{\partial t} + f^i(t, y)\frac{\partial}{\partial y^i}. \tag{10}$$

Соотношения (9) можно записать в компактной форме с помощью операции коммутирования векторных полей

$$[X_0, X] = (X_0\xi)X_0. \tag{11}$$

Перейдем к симметриям динамических систем с управлением [4].

Динамической системой с управлением называется система уравнений вида

$$\dot y^i = g^i(t, y, u), \quad i = 1, \ldots, n, \quad (t, y) \in M \subset R^{n+1}, \quad u \in U \subset R^r. \tag{12}$$

Предполагается, что функции g^i, $\partial f^i/\partial y^j$, $\partial g^i/\partial u^\alpha$ являются гладкими. Обычно называют y фазовыми переменными (состояниями), u — управлениями (внешними воздействиями). Множество M, называемое фазовым пространством, — область, U — область. Управления могут быть кусочно-непрерывными функциями $u(t)$, $t \in [t_0, t_1]$. В этом случае они называются допустимыми. Решением или фазовой траекторией системы (12) называется непрерывная кусочно-гладкая функция $y(t)$, $t \in [t_0, t_1]$, для которой существует такое допустимое управление $u(t)$, $t \in [t_0, t_1]$, что функции $y(t)$, $u(t)$ удовлетворяют соотношениям (12).

Далее будем считать, что управления являются постоянными. Дело в том, что для исследования декомпозиции этого достаточно, а в качестве приложения симметрий будет именно декомпозиция. При этом будем следовать работе [4], которая затем будет обобщена для дифференциальных уравнений с частными производными в следующем разделе. Кстати, Ю. Н. Павловский (автор), будучи учеником Л. В. Овсянникова, выполнил первые исследования по групповому анализу уравнений двумерного пограничного слоя [5]. Предполагаем, что однопараметрические группы симметрий порождаются векторными полями (инфинитезимальными) операторами

$$X = \xi(t, y)\frac{\partial}{\partial t} + \eta^i(t, y)\frac{\partial}{\partial y^i} + \omega^\alpha(t, y, u)\frac{\partial}{\partial u^\alpha}, \tag{13}$$

действующими в $(n + r + 1)$=мерном пространстве переменных (t, y, u).

Условие того, что поле X допускается системой, выглядит следующим образом:

$$X_0\eta^i - f^i X_0\xi = X f^i, \tag{14}$$

$$X_0(\omega^\alpha) = 0. \tag{15}$$

где

$$X_0 = \frac{\partial}{\partial t} + f^i \frac{\partial}{\partial y^i}$$

есть оператор полного дифференцирования.

Полагая

$$X^* = \theta^i(t, y)\frac{\partial}{\partial y^i} + \omega^\alpha \frac{\partial}{\partial u^\alpha}, \quad X = \xi X_0 + X^*, \tag{16}$$

из (14), (15) получим

$$X_0(\theta^i) = X^*(f^i), \tag{17}$$

$$X_0(\omega^\alpha) = 0. \tag{18}$$

Это можно записать с помощью коммутатора

$$[X_0, X^*] = 0 \tag{19}$$

Соотношения (17), (18) представляют собой дифференциальные уравнения относительно компонент допускаемого поля (16). После их нахождения симметрии находятся решением некоторых систем обыкновенных дифференциальных уравнений. Симметрии, как уже отмечалось, могут быть использованы для построения новых решений системы уравнений по известным решениям. Существует еще одно приложение симметрий для упрощения исследования систем уравнений, а именно, декомпозиция этих уравнений.

2. СИММЕТРИИ И ДЕКОМПОЗИЦИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ

Вернемся к дифференциальным уравнениям с частными производными (1), причем в специальной форме (2)

$$\partial_k y^i = g^i_k(t, y, u) \tag{20}$$

$$t \in I \subset R^m, \quad y \in L \subset R^n, \quad u \in U \subset R^s, \tag{21}$$

где I, L, U — некоторые области. Заметим, что если в (1) входят управляющие воздействия, т.е. это так называемая система управления с распределенными параметрами, то при переходе в специальную форму (2) управляющие воздействия переходят в разряд параметрических переменных и в математическом смысле не отличаются от них [2]. Исследуем вопрос о существовании инфенитезимальных симметрий (допускаемых векторных полей) для системы (20). Как и в [1, 2], воспользуемся аналогией систем (2) с управляемыми динамическими системами (12). Эта аналогия заключается в следующем.

Каждое решение $y(t)$ управляемой динамической системы (12) получается после подстановки в правую часть управлений $u(t)$ из некоторого класса допустимых функций. С другой стороны, решения $y(t)$ системы (20) соответствуют некоторому выбору параметрической функции $u(t)$ также из некоторого класса. Однако есть существенное различие в классах допустимых “управлений”. Для управляемых динамических систем классы допустимых управлений достаточно известны и широки: от класса кусочно-непрерывных функций до класса измеримых функций. Для систем (20) заранее задать класс допустимых “управлений” затруднительно, ибо далеко не каждый выбор функций $u(t)$ приводит к решению $y(t)$. Препятствием является, в частности, возможность несовместности полученной системы после подстановки $u(t)$. Тем не менее идеология теории управлений может быть полезна из-за возможности привлечения математического аппарата, разработанного для систем вида (12).

Далее будем считать, что параметрические переменные (“управления”) являются всевозможными постоянными. Дело в том, что (также как и для управляемых систем (12)) для исследования декомпозиции этого достаточно (см. [2]) а в качестве приложения симметрий будет именно декомпозиция. При этом будем следовать результатам работы [4], которые здесь естественным образом обобщаются на дифференциальные уравнения с частными производными. Итак, речь идет нахождении векторных полей

$$X = \xi^k(t, y, u)\frac{\partial}{\partial t^k} + \eta^i(t, y, u)\frac{\partial}{\partial y^i} + \omega^\alpha(t, y, u)\frac{\partial}{\partial u^\alpha}. \tag{22}$$

По аналогии с управляемыми динамическими системами, условие того, что поле X допускается системой, выглядит следующим образом:

$$X_l\eta^i - g^i X_l\xi = Xg^i, \quad i = 1, \dots, n, \tag{23}$$

$$X_l(\omega^\alpha) = 0, \tag{24}$$

где X_l — операторы полного дифференцирования системы (20) по переменным t^l

$$X_l = g^i_l(t, y, u)\partial/\partial y^i, \quad l = 1, \dots, m. \tag{25}$$

Это — аналоги оператора полного дифференцирования X_0 по одной независимой переменной для обыкновенных дифференциальных уравнений.

Далее при исследовании симметрий и декомпозиции рассматриваются дифференциальные уравнения с частными производными специального вида, причем для простоты и не ограничивая общности “автономные”, т.е с правыми частями, не зависящими от аргументов t.

$$\partial_k y^i = g^i_k(y, u), \tag{26}$$

$$t \in I \subset R^m, \quad y \in L \subset R^n, \quad u \in U \subset R^s, \tag{27}$$

где I, L, U — некоторые области. Речь пойдет о декомпозиции системы (26) с помощью замены зависимых переменных y, точнее о возможности преобразования системы с помощью замены переменных

$$z^i = \varphi^i(y), \quad i = 1, \dots, n, \tag{28}$$

к виду

$$\partial_k z^l = h^l_k(z^1, \dots, z^m, u), \quad l = 1, \dots, m, \tag{29}$$

$$\partial_k z^i = h^i_k(z^1, \dots, z^n, u), \quad i = m+1, \dots, n. \tag{30}$$

Если такое представление возможно, то будем говорить, что система (26) допускает декомпозицию (агрегирование) по зависимым переменным порядка $n - m$, причем первые m функций в замене переменных (28)

$$z^i = \varphi^i(y), \quad i = 1, \dots, m, \tag{31}$$

называются агрегатами, а система (29) — агрегированной системой.

Замена зависимых переменных (28) в системе (26) осуществляется следующим образом: нужно подействовать операторами полного дифференцирования (25) на функции (28)

$$X_l(\varphi^i(y) = g^j_l(y, u)\frac{\partial\varphi^i}{\partial y^j} \tag{32}$$

и выразить функционально полученные функции через (28)

$$g^j_l(y, u)\frac{\partial\varphi^i}{\partial y^j} = h^i(\varphi^1(y), \dots, \varphi^n(y), u). \tag{33}$$

Функции

$$h^i(z^1, \dots, z^n, u) \tag{34}$$

являются новыми правыми частями системы, при этом декомпозиция (29), (30) возникает, когда первые m функций (33) функционально выражаются только через агрегаты (31), т.е.

$$g_l^j(y,u)\frac{\partial \varphi^i}{\partial y^j} = h^i(\varphi^1(y), \dots, \varphi^m(y)u), \quad i = 1 \dots, m. \tag{35}$$

Непосредственное применение этого условия не дает конструктивного метода нахождения агрегатов и проверки возможности декомпозиции. Поэтому в [2] применен опосредственный подход, когда сначала с помощью решения некоторых систем дифференциальных уравнений отыскиваются полные семейства векторных полей, для которых агрегаты являются интегралами, а затем находятся агрегаты, т.е. интегралы этих семейств. При этом совместность упомянутых систем дифференциальных уравнений определяет возможность декомпозиции.

Для применения симметрий в вопросах декомпозиции вида (29), (30) в силу вышеуказанных причин (в частности, автономности) естественно рассматривать симметрии, не меняющие переменные t и параметрические переменные u. Для соответствующих полей (22) это означает, что $\xi^k = 0$, $\omega^\alpha = 0$.

Тогда для искомых векторных полей

$$X = \eta^i(t, y, u)\frac{\partial}{\partial y^i} \tag{36}$$

выражения (18), (24) можно с помощью операции коммутирования заменить компактным выражением

$$[X_l, X] = 0 \quad l = 1, \dots, m. \tag{37}$$

Векторные поля (36), удовлетворяющие (37), образуют алгебру Ли, обозначаемую через $\mathfrak{a}$, т.е. коммутаторы и линейные комбинации полей из $\mathfrak{a}$ снова принадлежат $\mathfrak{a}$. Кроме того, введем подалгебру, состоящую из полей вида

$$X = \eta^i(y)\frac{\partial}{\partial y^i}, \tag{38}$$

которую обозначим через $\mathfrak{a}_0$. Заметим, что существуют аналоги этих алгебр для управляемых динамических систем, в частности, была обнаружена глубокая связь свойств структуры управляемой динамической системы со свойствами соответствующих алгебр. Вероятно, что существуют подобные связи и для систем дифференциальных уравнений с частными производными. Здесь рассматривается применение алгебры $\mathfrak{a}_0$ для исследования возможности декомпозиции системы (26).

Напомним, что полным семейством векторных полей или операторов называется семейство полей

$$Z_a = b_a^i(y)\frac{\partial}{\partial y^i}, l = 1, \dots, p \tag{39}$$

(по повторяющемуся индексу здесь и далее производится суммирование), если

1) $$\operatorname{rank} \left\| b_a^i(y) \right\| = p,$$

т.е. векторы $Z_a(y)$, $a = 1, \dots, p$ линейно независимы в каждой точке $y \in M$ или, как еще говорят, линейно несвязанны в области M;

2) $$[Z_a, Z_c] = h_{ac}^d(y)Z_d, \quad a, c, d = 1, \dots, p, \tag{40}$$

т.е. все коммутаторы $[Z_a, Z_c]$ семейства выражаются линейно с переменными коэффициентами $h_{ac}^d(y)$ через остальные поля семейства.

Следующее утверждение определяет некоторое условие декомпозиции на основе существования симметрий.

Теорема. *Если в* $\mathfrak{a}_0$ *существует полное семейство векторных полей*

$$Z_a = b_a^i(y)\frac{\partial}{\partial y^i}, l = 1, \dots, p < n, \tag{41}$$

то система допускает декомпозицию (29), (30) *системы* (26)*, причем агрегатами являются полный набор интегралов этого семейства, где* $p = n - m$.

Доказательство. У полного семейства (41) имеется полный набор функционально независимых интегралов $\varphi^i(y)$, $i = 1, \dots, m = n - p$, т.е. $Z_a\left(\varphi^i(y)\right) = 0$ [6]. Рассмотрим выражения

$$\left[X_l, Z_a\right]\left(\varphi^i(y)\right) = X_l(Z_a\left(\varphi^i(y)\right) - Z_a\left(X_l\left(\varphi^i(y)\right)\right) = 0. \tag{42}$$

Из (42) вытекает, что функции $X_l\left(\varphi^i(y)\right)$ являются интегралами семейства (39), т.е. $Z_a\left(X_l\left(\varphi^i(y)\right)\right) = 0$ и, согласно [6], функционально выражаются через полный набор интегралов и, следовательно, $\varphi^i(y)$ являются агрегатами для системы (26).

Выражения (37) можно трактовать как систему дифференциальных уравнений

$$X_l(b_a^i) = Z_a(g^i), \tag{43}$$

$$i = 1, \dots, n, \quad a, c = 1, \dots, p, \quad u \in U,$$

относительно неизвестных компонент семейства (39). В (43) выражения $\xi_u(b_a^i)$, $Z_a(f^i)$ представляют собой действия полей ξ_u, Z_a на функции $b_a^i(y)$, g^i как операторов. После решения этой системы нужно найти полный набор интегралов $\varphi^i(y)$, $i = 1, \dots, m = n - p$, полного семейства (39), что, согласно [6, с. 12], сводится к решению некоторых систем обыкновенных дифференциальных уравнений. Для нахождения правых частей фактор-системы следует функции $g^j(y,u)\frac{\partial\varphi^k}{\partial y^j}$ выразить функционально через функции $\varphi^i(y)$, $i = 1, \dots, m = n - p$, т. е. представить их в виде

$$g^j(y,u)\frac{\partial\varphi^i}{\partial y^j} = h^p(\varphi^1(y), \dots, \varphi^m(y), u).$$

Построенные функции h^k и определяют искомую фактор-систему

$$\dot{z}^k = h^k(z,u), \quad k = 1, \dots, m.$$

Пример:

$$\begin{aligned}
\partial_k y_1 &= \quad y_1^2 + y_2^2 + y_3^2 + u_1,\\
\partial_k y_2 &= \quad 2y_1y_2 + 2y_2y_3 + u_2,\\
\partial_k y_3 &= \qquad 2y_1y_3 + u_3.
\end{aligned}$$

Убеждаемся, что компоненты полей семейства

$$\begin{aligned}
X_1 &= \quad \eta_{11}\frac{\partial}{\partial y_1} + \eta_{21}\frac{\partial}{\partial y_2} + \eta_{31}\frac{\partial}{\partial y_3},\\
X_2 &= \quad \eta_{12}\frac{\partial}{\partial y_1} + \eta_{22}\frac{\partial}{\partial y_2} + \eta_{32}\frac{\partial}{\partial y_3},
\end{aligned}$$

такие, что

$$\eta_{11} + \eta_{21} + \eta_{31} = 0, \quad \eta_{12} + \eta_{22} + \eta_{32} = 0,$$

удовлетворяют соответствующим уравнениям (37), причем интеграл семейства:

$$y_1 + y_2 + y_3.$$

Для нахождения агрегированной системы следует подействовать операторами полного дифференцирования на этот интеграл и выразить функционально полученный результат через интеграл. В полученном выражении следует заменить интеграл на новую переменную z_1. В результате, агрегированная система имеет вид

$$\partial_k z_l = z_1^2 + u_1 + u_2 + u_3.$$

СПИСОК ЛИТЕРАТУРЫ

1. *Елкин В. И.* Об одном условии управляемости систем с сосредоточенными и распределенными параметрами // Труды ИСА РАН. 2022. Т. 72. N 4. С. 11–15.
2. *Елкин В. И.* Агрегирование и декомпозиция систем дифференциальных уравнений с частными производными и систем управления с распределенными параметрами // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. N 9. С. 1575–1586.
3. *Овсянников Л. В.* Групповой анализ дифференциальных уравнений. М.: Наука, 1978. 399 с.
4. *Павловский Ю. Н.* Групповые свойства управляемых систем и фазовые организационные структуры // Ж. вычисл. матем. и матем. физ. 1974. Т. 14. N 4. С. 862–872; Т. 14. N 5. С. 1093–1103.
5. *Павловский Ю. Н.* Исследование некоторых инвариантных решений пограничного слоя // Ж. вычисл. матем. и матем. физ. 1961. Т. 1. N 2. С. 280–294.
6. *Эйзенхарт Л. П.* Непрерывные группы преобразований. М.: Изд-во иностр. лит., 1947. 359 с.

SYMMETRIES AND DECOMPOSITION OF SYSTEMS OF PARTIAL DIFFERENTIAL EQUATIONS AND DISTRIBUTED PARAMETERS CONTROL SYSTEMS

V. I. Elkin*

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Vavilova st., 44, Moscow, 119333 Russia

**e-mail: elk_v@mail.ru*

Received 08 August, 2023
Revised 31 January, 2024
Accepted 05 March, 2024

Abstract. The issue of symmetries of partial differential equations is considered based on the use of differential-geometric and algebraic methods of the theory of dynamic systems with control.

Keywords: symmetry group, Lie algebra.

УДК 519.16+519.85

ОТКАЗОУСТОЙЧИВЫЕ СЕМЕЙСТВА ПЛАНОВ ПРОИЗВОДСТВА: МАТЕМАТИЧЕСКАЯ МОДЕЛЬ, ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ И АЛГОРИТМЫ ВЕТВЕЙ И ГРАНИЦ[1)]

© 2024 г. Ю. Ю. Огородников[1,*], Р. А. Рудаков[1,], Д. М. Хачай[2,***], М. Ю. Хачай[1,****]**

[1] *620108 Екатеринбург, ул. С. Ковалевской, 16, ИММ им. Н. Н. Красовского УрО РАН, Россия*
[2] *33405 Talence, 680 cours Libération, KEDGE Business School, France*

[*] *e-mail: yogorodnikov@imm.uran.ru*
[**] *e-mail: r.a.rudakov@gmail.com*
[***] *e-mail: daniil.khachai@kedgebs.com*
[****] *e-mail: mkhachay@imm.uran.ru*

Поступила в редакцию 28.10.2023 г.
Переработанный вариант 28.11.2023 г.
Принята к публикации 05.03.2024 г.

Вопросы проектирования устойчивых к сбоям систем производства и поставок продукции составляют одно из приоритетных направлений развития современного исследования операций. Традиционный подход к моделированию таких систем основывается на привлечении вероятностных моделей, описывающих выбор возможного сценария действий в случае возникновения неполадок в производственной или транспортной сети. Наряду с рядом преимуществ данный подход обладает известным недостатком. Возникновение неполадок неизвестной природы, способных поставить под угрозу работоспособность всей моделируемой системы, существенно затрудняют его применение. В данной работе вводится в рассмотрение минимаксная задача построения отказоустойчивых планов производства (Reliable Production Process Design Problem, RPPDP), целью которой является обеспечение бесперебойного функционирования распределенной производственной системы при минимальных гарантированных издержках. Показывается, что задача RPPDP NP-трудна в сильном смысле и сохраняет труднорешаемость при достаточно специфических условиях. Для поиска точных и приближенных решений с оценками точности для данной задачи разработаны методы ветвей и границ, основанные на предложенной компактной модели смешанного целочисленного линейного программирования (Mixed Integer Linear Program, MILP) и авторской эвристике адаптивного поиска в больших окрестностях (Adaptive Large Neighborhood Search, ALNS) в рамках расширений известного MIP-солвера Gurobi. Высокая производительность и взаимодополняемость предложенных алгоритмов подтверждена результатами численных экспериментов, проведенных на разработанной авторами открытой библиотеке тестовых примеров, содержащей адаптированные постановки задач из библиотеки PCGTSPLIB. Библ. 25. Фиг. 5. Табл. 3.

Ключевые слова: задача проектирования отказоустойчивых производственных процессов, MILP-модель, метод ветвей и границ, эвристика адаптивного поиска в больших окрестностях.

DOI: 10.31857/S0044466924060058, **EDN:** XYUTPT

1. ВВЕДЕНИЕ

Проектирование распределенных производственных систем, сохраняющих работоспособность при условиях вероятных сбоев в цепях поставок комплектующих и готовой продукции, является одной из наиболее актуальных проблем современного исследования операций.

В процессе промышленного производства часто возникают непредвиденные обстоятельства, связанные с недостаточной надежностью цепей поставок, непредсказуемым задержкам в поступлении

[1)] Работа Ю. Ю. Огородникова и М. Ю. Хачая выполнена при финансовой поддержке РНФ (грант 22-21-00672).

комплектующих, недостаточными запасами критически значимых ресурсов и сбоями в транспортной сети [1]. Один из традиционных подходов к решению возникающих проблем основан на построении специализированных вероятностных моделей [2], описывающих возможные нарушения производственно-транспортной системы в контексте предопределенного набора сценариев. Несмотря на ряд преимуществ, данный подход становится малоэффективным в случае возникновения в процессе функционирования системы сбоя, неизвестного на этапе моделирования. В этой ситуации более подходящей альтернативой представляется построение минимаксных детерминированных моделей, связанных с поиском решений, оптимизирующих гарантированный результат.

Производственным процессом договоримся называть совокупность из m взаимосвязанных производственных операций, удовлетворяющих заданному отношению (частичного или линейного) порядка. Произвольная операция может быть выполнена на одном из взаимозаменяемых пунктов производства. Каждый пункт производства может выполнять единственную операцию, что приводит к образованию m производственных кластеров. В соответствии с заданным порядком продукция, произведенная в кластере-предшественнике, транспортируется в пункты производства кластера-последователя, где используется в качестве исходного ресурса.

Для описания перевозок производимой и потребляемой продукции введем в рассмотрение *транспортную сеть* — реберно-взвешенный ориентированный граф, множество вершин которого включает выделенные вершины s и t, характеризующие начало и завершение производственного процесса, кластерные вершины, соответствующие описанным выше пунктам производства, а также специализированные вершины, именуемые *транспортными хабами*. Каждый хаб обладает ценой открытия — единоразовой стоимостью включения его в производственный процесс и пропускной способностью, ограничивающей число обслуживаемых им маршрутов. Как обычно транспортные издержки моделируются весовой функцией, заданной на дугах сети.

Производственным планом мы называем подграф описанной выше транспортной сети, включающий вершины s и t, по одной вершине-представителю для каждого производственного кластера и направленные маршруты, соединяющие вершины подграфа в соответствии с заданным порядком и учетом пропускной способности транспортных хабов. Стоимость плана определяется суммарными транспортными издержками и ценами открытия используемых им хабов. По определению стоимость семейства планов совпадает с максимальной стоимостью входящих в него планов.

Цель вводимой в рассмотрение задачи проектирования отказоустойчивых производственных процессов (Reliable Production Process Design Problem, RPPDP) состоит в построении семейства минимальной стоимости, состоящего из k независимых производственных планов (пересекающихся только в начальной и конечной вершинах s и t).

2. ИЗВЕСТНЫЕ РЕЗУЛЬТАТЫ

Вероятно наиболее близкой к рассматриваемой в данной работе минимаксной задаче построения отказоустойчивых планов производства является задача о гомеоморфном вложении графа (Subgraph Homeomorphism Problem, SHP), введенная в известной работе [3]. Фактически задача SHP представляет собой семейство комбинаторных задач, параметризованное орграфами P, подлежащими вложению и именуемыми *паттернами*. В каждой задаче данного семейства требуется установить существование взаимно однозначного отображения $\chi : V(P) \to V(G)$ множества вершин паттерна P во множество вершин заданного орграфа G так, что произвольной дуге (a, b) графа P соответствует простой путь, соединяющий вершины $\chi(a)$ и $\chi(b)$ графа G, причем все построенные пути являются вершинно непересекающимися. Как показано авторами [3], труднорешаемость подклассов задачи SHP полностью определяется структурой графа P. В частности, NP-полной при каждом фиксированном $k \geqslant 2$ является задача о k вершинно непересекающихся путях (k-DP) (см., например, [4]), в которой для заданного набора адресных пар (s_i, t_i), $i = \overline{1, k}$, требуется выяснить, существует ли в орграфе G семейство попарно вершинно непересекающихся s_i-t_i-путей.

Другой близкой к рассматриваемой нами задаче является известная задача о кратчайшем пути с ограничениями (Constrained Shortest Path Tour Problem, CSPTP) [5]. Условие этой задачи описы-

вается реберно-взвешенным ориентированным графом $G = (V \cup \{s,t\}, A)$, упорядоченным множеством попарно непересекающихся подмножеств вершин $\mathfrak{T} = \{T_1, \ldots, T_k\}$ и весовыми функциями $c, q\colon A \to \mathbb{R}^+$, определяющими транспортные издержки и пропускную способность для каждой дуги графа G. Цель задачи состоит в построении кратчайшего s-t-маршрута, посещающего некоторую систему представителей подмножеств семейства $\mathfrak{T}$ в соответствии с заданным линейным порядком.

В работах [5, 6] обоснована труднорешаемость задачи CSPTP, предложен эвристический алгоритм жадного рандомизированного адаптивного поиска (GRASP) и несколько вариантов метода ветвей и границ. Авторами статьи [7] разработан алгоритм ветвей и оценок (branch-and-price), опирающийся на улучшенную модель смешанного целочисленного линейного программирования и показавший наилучшую производительность для данной задачи.

Задача о кратчайшем простом пути с k обязательными к посещению узлами (Shortest Simple Path Problem with k Must-Pass Nodes, SSPP-k-MPN), введенная в рассмотрение в статье [8], также представляется близкой к исследуемой нами задаче. Условие задачи SSPP-k-MPN задается реберно-взвешенным орграфом $G = (V, E, c)$, неупорядоченным подмножеством $T \subset V$ мощности k и выделенными начальной и конечной вершинами s и t. Задача состоит в построении кратчайшего простого s-t-пути, посещающего вершины множества T в произвольном порядке, причем каждую в точности один раз. В статье [9] показано, что задача SSPP-k-MPN NP-трудна в сильном смысле при произвольном фиксированном $k \geqslant 1$. Для данной задачи известны несколько моделей смешанно-целочисленного линейного программирования (Mixed-Integer Linear Program, MILP-моделей) [10] и специализированный метод ветвей и границ [9].

Ряд общих свойств с задачей RPPDP имеет и известная обобщенная задача коммивояжера с ограничениями предшествования (Precedence Constrained Generalized Traveling Salesman Problem, PCGTSP) [11]. Постановка PCGTSP задается реберно-взвешенным орграфом G, множество вершин которого разбито на попарно непересекающиеся подмножества-кластеры C_i, $i = \overline{1,m}$. Вспомогательный ациклический орграф П задает частичный порядок, в соответствии с которым должны посещаться кластеры произвольным допустимым маршрутом. Цель задачи состоит в построении кратчайшего простого маршрута, начинающегося и завершающегося в начальном кластере C_1, посещающего каждый кластер в единственной вершине в соответствии с заданным порядком. Задача PCGTSP NP-трудна в сильном смысле [12] в общей постановке и полиномиально разрешима при $m = O(\log n)$ [13]. Для нее разработана серия точных и приближенных алгоритмов, среди которых выделяются эффективные алгоритмы для ряда специальных случаев ограничений предшествования [14, 15, 16], эвристический солвер PCGLNS [17], схема динамического программирования и метод ветвей и границ [13], опирающиеся на эвристику PCGLNS [17] и подход [18]. На данный момент рекордным по производительности для данной задачи является метод ветвей и отсечений [19], эффективно использующий полиэдральные свойства допустимого множества задачи PCGTSP. С помощью данного подхода удалось существенно продвинуться в решении примеров известной библиотеки тестовых примеров PCGTSPLIB [20].

Как следует из приведенного обзора, имеющийся на данный момент математический аппарат позволяет эффективно находить кратчайшие маршруты в транспортных сетях при различных дополнительных ограничениях. Тем не менее, обсуждавшиеся выше комбинаторные задачи не вполне соответствуют моделированию устойчивых к сбоям производственных процессов, так как единственный оптимальный подграф (путь, маршрут и т.п.), поиск которого является их целью, может оказаться уязвимым к возможным неполадкам в транспортной сети.

В работе получены следующие результаты.

- Предложена постановка задачи RPPDP и обоснована ее труднорешаемость.
- Разработаны алгоритмы ветвей и отсечений, основанные на предложенной компактной MILP-модели и оригинальной эвристики адаптивного поиска и реализованные в виде расширений известного MIP-солвера Gurobi [21].
- Разработана открытая библиотека тестовых примеров для задачи RPPDP, расширяющая известную библиотеку PCGTSPLIB.

• Эффективность предложенных алгоритмов подтверждена результатами проведенных численных экспериментов.

Статья имеет следующую структуру. В разд. 3 вводится математическая постановка задачи RPPDP, труднорешаемость которой обосновывается в разд. 4. Раздел 5 содержит описание предлагаемой MILP-модели, обоснование ее корректности и компактности. Раздел 6 посвящен описанию предлагаемых алгоритмов. В разд. 7 обсуждаются результаты проведенного численного эксперимента. Наконец, в разд. 8 подводятся итоги работы.

3. ПОСТАНОВКА ЗАДАЧИ

Постановка задачи RPPDP задается упорядоченной тройкой (G, Π, k), в которой

• $G = (V, E, c)$ — реберно-взвешенный орграф, в котором
  – задано разбиение множества вершин $V = \mathfrak{M} \cup \mathfrak{H}$;
  – выделены вершины s и t, представляющие собой начало и конец произвольного производственного плана;
  – $\mathfrak{M} = \mathfrak{M}_0 \cup \ldots \cup \mathfrak{M}_{m+1}$, где $\mathfrak{M}_0 = \{s\}$, $\mathfrak{M}_{m+1} = \{t\}$, а кластер $\mathfrak{M}_j$, $j = \overline{1, m}$, объединяет однородные пункты производства, исполняющие идентичную операцию j;
  – $\mathfrak{H}$ — множество транспортных хабов; произвольный хаб $h \in \mathfrak{H}$ характеризуется неотрицательными пропускной способностью q_h, ограничивающей сверху число входящих (исходящих) дуг произвольного допустимого решения, инцидентных вершине h, и ценой открытия C_h;
  – весовая функция $c\colon E \to \mathbb{R}_+$ определяет транспортные издержки c_e, связанные с перевозками по дуге $e \in E$;

• $\Pi = (\{0, 1, \ldots, m+1\}, A)$ — ациклический орграф, задающий порядок на множестве производственных кластеров, минимальный и максимальный элементы которого соответствуют кластерам $\mathfrak{M}_0$ и $\mathfrak{M}_{m+1}$;

• $k \geqslant 1$ — размер искомого семейства производственных планов.

Задача состоит в построении семейства попарно вершинно-непересекающихся производственных планов $\mathscr{P} = \{P_1, P_2, \ldots, P_k\}$ таких, что

• каждый план P_r представляет собой ориентированный ациклический подграф, получаемый из графа Π путем замены произвольной дуги $(i, j) \in A$ простым маршрутом $v_i, h_1, \ldots, h_p, v_j$, начинающимся в кластере $\mathfrak{M}_i$, посещающим некоторое число хабов, и заканчивающимся в кластере $\mathfrak{M}_j$;

• стоимость каждого плана P_r состоит из суммы весов всех входящих в него дуг, а также стоимостей открытия задействованных хабов

$$\operatorname{cost}(P_r) = \sum_{e \in P_r} c_e + \sum_{h \in P_r} C_h;$$

• семейство $\mathscr{P}$ обладает минимальной возможной стоимостью, задаемой соотношением

$$\operatorname{cost}(\mathscr{P}) = \max\{\operatorname{cost}(P_r)\colon\ r = \overline{1, k}\}.$$

Фиг. 1 иллюстрирует введенную выше постановку.

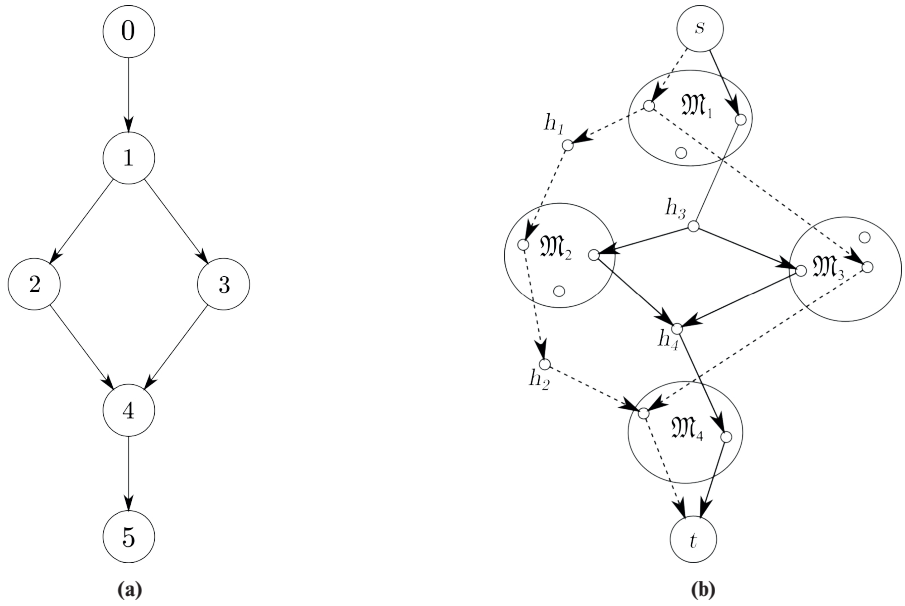
4. ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ

Раздел посвящен труднорешаемости задачи RPPDP, обоснование которой опирается на полиномиальную сводимость к данной задаче NP-полной задачи 2-DP [3].

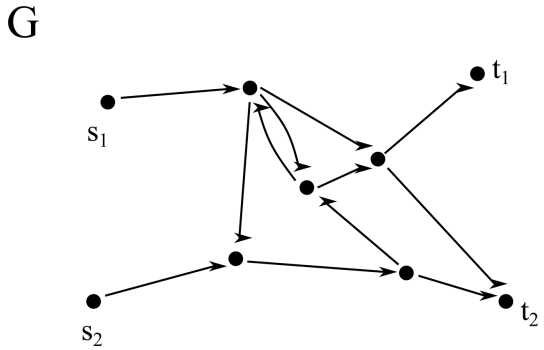
Теорема 1. *Задача RPPDP NP-трудна в сильном смысле.*

Доказательство. Рассмотрим произвольную постановку I_1 задачи 2-DP, задаваемую орграфом $G = (V, E)$ на n вершинах и адресными парами (s_1, t_1) и (s_2, t_2), где s_1, s_2, t_1 и t_2 — попарно различные вершины графа G, и требуется проверить существование вершинно непересекающихся s_1-t_1 и s_2-t_2-путей (см. фиг. 2).

Сопоставим данной постановке постановку $I_2 = (G', \Pi, 1)$ задачи RPPDP, в которой



Фиг. 1. Пример постановки RPPDP для $m = 4$ и $k = 2$. (a) граф порядка П (b) граф G, сплошными и пунктирными линиями обозначены первый и второй планы некоторого допустимого решения.



Фиг. 2. Пример постановки I_1 задачи 2-DP.

1) орграф $G' = (V', E')$ является дизъюнктным объединением двух копий графа G, размещенных в верхнем и нижнем уровнях, дополненным стартовой и завершающей вершинами s и t и вспомогательным кластером-переключателем $\mathfrak{M}'$ (см. фиг. 3б);

2) кластеры $\mathfrak{M}_0$–$\mathfrak{M}_5$ определяются соотношениями $\mathfrak{M}_0 = \{s\}$, $\mathfrak{M}_1 = \{s_1^u, s_1^l\}$, $\mathfrak{M}_2 = = \{s_2^u, s_2^l\}$, $\mathfrak{M}_3 = \{t_1^u, t_1^l\}$, $\mathfrak{M}_4 = \{t_2^u, t_2^l\}$ и $\mathfrak{M}_5 = \{t\}$;

3) произвольные вершины v_p^u и v_q^l, находящиеся на различных уровнях графа G' и не лежащие в одном кластере — смежны (соединены дугами (v_p^u, v_q^l) и (v_q^l, v_p^u));

4) транспортные издержки c_e, сопоставляемые произвольной дуге $e \in E'$, задаются соотношением

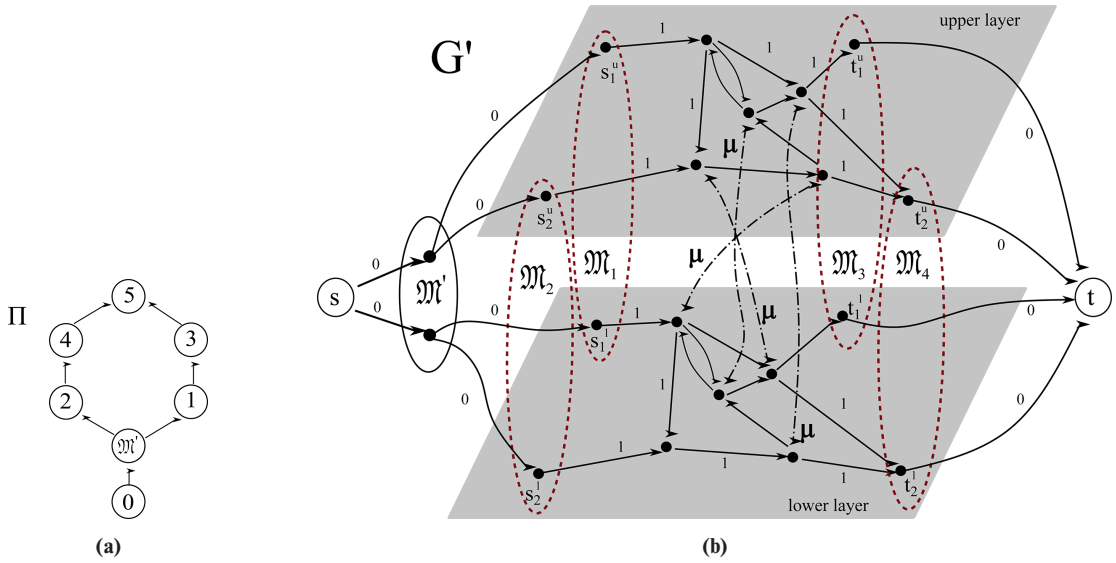
$$c_e = \begin{cases} 1, & \text{если обе вершины дуги } e \text{ лежат на одном уровне графа } G', \\ 0, & \text{если дуга } e \text{ инцидентна вершинам кластера } \mathfrak{M}' \text{ или вершине } t, \\ \mu = n^2 + 1, & \text{если } e \text{ соединяет вершины различных уровней графа } G'; \end{cases}$$

5) множество транспортных хабов задается равенством $\mathfrak{H} = (V' \setminus \mathfrak{M}') \setminus \left(\bigcup_{p=0}^{5} \mathfrak{M}_p\right)$;

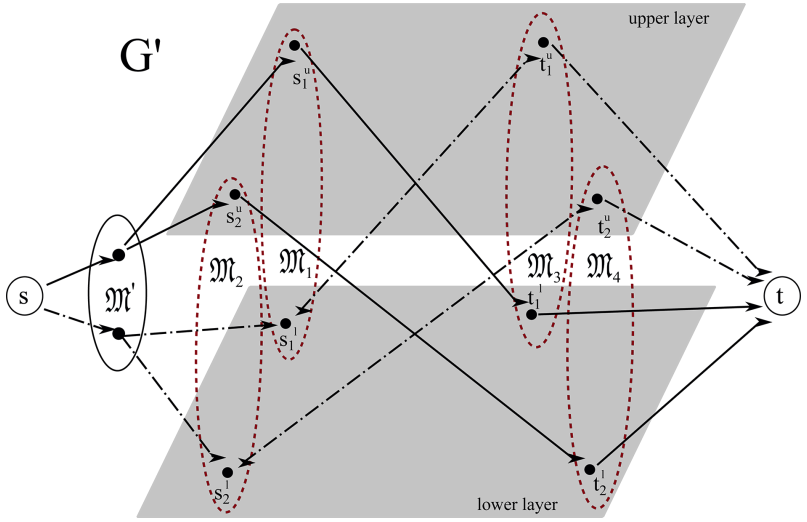
6) стоимость открытия C_h и пропускная способность q_h произвольного хаба $h \in \mathfrak{H}$ определяются равенствами $C_h = 0$ и $q_h = 1$, соответственно;

7) граф П имеет вид, представленный на фиг. 3а.

По построению, для произвольной постановки I_1 сопоставленная ей постановка I_2 обладает допустимыми решениями (см. фиг. 4) и, следовательно, разрешима.



Фиг. 3. Пример постановки I_2 задачи RPPDP.



Фиг. 4. Допустимые решения $\mathscr{P}_1 = \{P_1\}$ и $\mathscr{P}_2 = \{P_2\}$ произвольной постановки I_2. Дуги плана P_1 отображены сплошными линиями, плана P_2 – пунктиром.

Для завершения доказательства теоремы убедимся, что постановка I_1 обладает положительным ответом тогда и только тогда, когда оптимальное значение соответствующей ей постановки I_2 не превосходит n^2.

В самом деле, пусть в графе G существуют вершинно непересекающиеся s_1-t_1 и s_2-t_2-пути. Разместив копии этих путей, например в верхнем уровне графа G' и соединив стартовую вершину s с вершинами s_1^u, s_2^u путями нулевого веса, проходящими через кластер $\mathfrak{M}'$, а вершины t_1^u и t_2^u дугами нулевого веса с вершиной t, получим допустимое решение $\{P_1\}$ постановки I_2, стоимость которого не превосходит n^2. Следовательно, оптимальное значение задачи I_2 также удовлетворяет данной верхней оценке.

С другой стороны, пусть постановка I_2 обладает допустимым решением, стоимость которого не превосходит n^2. С необходимостью входящий в данное решение план не содержит дуг веса $\mu = n^2 + 1$ и, следовательно, посещает вершины лишь одного из уровней графа G'. Без ограничения общности полагаем, что данный план посещает вершины верхнего уровня графа G'. По построению, он индуцирует вершинно непересекающиеся s_1^u-t_1^u и s_2^u-t_2^u-пути, являющиеся копиями s_1-t_1 и s_2-t_2-путей в графе G, обуславливая положительный ответ в исходной постановке I_1. Теорема доказана.

Отметим, что в доказательстве теоремы 1 фактически обоснована труднорешаемость задачи RPPDP при $k = 1$, нулевых стоимостях открытия и единичной пропускной способности хабов. Данные рассуждения легко могут быть распространены на случай произвольного фиксированного $k \geqslant 1$.

5. MILP-МОДЕЛЬ

Предлагаемая нами модель смешанного целочисленного линейного программирования оперирует двумя группами индикаторных переменных:

- $x_e^{r,a} \in \{0,1\}$ $(r = \overline{1,k}, a \in A, e \in E)$ характеризуют включение дуги e графа G в маршрут, реализующий в производственном плане r дугу a графа порядка П;
- $y_v^r \in \{0,1\}$ $(r = \overline{1,k}, v \in V)$ определяют использование вершины v графа G в плане r.

Кроме того, модель включает вычисляемую неотрицательную вещественную переменную cost, значение которой совпадает со значением целевой функции.

Предлагаемая MILP-модель имеет следующий вид:

$$\min \ \text{cost} \tag{1}$$

$$\text{s.t.} \sum_{a\in A, e\in E} c_e x_e^{r,a} + \sum_{h\in\mathfrak{H}} C_h y_h^r - \text{cost} \leqslant 0 \qquad \left(r = \overline{1,k}\right), \tag{2}$$

$$\sum_{u\in\mathfrak{M}_j\cup\mathfrak{H}\,:\,(v,u)\in E} x_{(v,u)}^{r,(i,j)} - y_v^r = 0 \qquad \begin{pmatrix} r = \overline{1,k}, \\ (i,j)\in A, \\ v\in\mathfrak{M}_i \end{pmatrix}, \tag{3}$$

$$\sum_{v\in\mathfrak{M}_i\cup\mathfrak{H}\,:\,(v,u)\in E} x_{(v,u)}^{r,(i,j)} - y_u^r = 0 \qquad \begin{pmatrix} r = \overline{1,k}, \\ (i,j)\in A, \\ u\in\mathfrak{M}_j \end{pmatrix}, \tag{4}$$

$$\sum_{v\in\mathfrak{M}\cup\mathfrak{H}\,:\,(v,u)\in E} x_{(v,u)}^{r,(i,j)} = 0 \qquad \begin{pmatrix} r = \overline{1,k}, \\ (i,j)\in A, \\ u\in\mathfrak{M}\setminus\mathfrak{M}_j \end{pmatrix}, \tag{5}$$

$$\sum_{u\in\mathfrak{M}\cup\mathfrak{H}\,:\,(v,u)\in E} x_{(v,u)}^{r,(i,j)} = 0 \qquad \begin{pmatrix} r = \overline{1,k}, \\ (i,j)\in A, \\ v\in\mathfrak{M}\setminus\mathfrak{M}_i \end{pmatrix}, \tag{6}$$

$$\sum_{v\in\mathfrak{M}_i\cup\mathfrak{H}\,:\,(v,h)\in E} x_{(v,h)}^{r,(i,j)} - \sum_{u\in\mathfrak{M}_j\cup\mathfrak{H}\,:\,(h,u)\in E} x_{(h,u)}^{r,(i,j)} = 0 \qquad \begin{pmatrix} r = \overline{1,k}, \\ (i,j)\in A, h\in\mathfrak{H} \end{pmatrix}, \tag{7}$$

$$\sum_{(i,j\in A)} \sum_{v\in\mathfrak{M}_i\cup\mathfrak{H}\,:\,(v,h)\in E} x_{(v,h)}^{r,(i,j)} - q_h y_h^r \leqslant 0 \qquad \left(r = \overline{1,k}, h\in\mathfrak{H}\right), \tag{8}$$

$$\sum_{v\in\mathfrak{M}_i} y_v^r = 1 \qquad \left(r = \overline{1,k}, i = \overline{1,m}\right), \tag{9}$$

$$\sum_{r=1}^{k} y_v^r \leqslant 1 \qquad \left(v\in V\setminus\{s,t\}\right), \tag{10}$$

$$x_e^{r,a} \in \{0,1\},\ y_v^r \in \{0,1\}. \tag{11}$$

Выражения (1) и (2) задают целевую функцию, значение которой совпадает с максимальной стоимостью производственных планов, входящих в текущее решение. Уравнения (3) и (4) характеризуют начальные и конечные вершины маршрутов, реализующих дугу (i, j) графа порядка П для каждого из k планов. Уравнения (5) и (6) гарантируют, промежуточными вершинами произвольного маршрута, реализующего дугу $(i, j) \in A$ могут быть только транспортные хабы $h \in \mathfrak{H}$. Уравнение (7) является условием сохранения потока для произвольного плана r и дуги $(i, j) \in A$. Неравенство (8) задает ограничение на количество посещений хаба $h \in H$. Выражение (9) гарантирует, что каждый кластер посещается каждым планом в точности один раз. Неравенство (10) задает попарную дизъюнктность всех производственных планов.

Проведя несложные вычисления, заметим что в предложенной модели число неизвестных составляет $O(n^2m^2k)$, а число ограничений — $O(nm^2k)$. Таким образом, размер предложенной модели ограничен сверху полиномом от размера условия исходной задачи, т.е. данная модель — компактна.

Теорема 2. *Множества допустимых решений задачи RPPDP и MILP-модели* (1)–(11) *изоморфны.*

Доказательство. Рассмотрим произвольное допустимое решение $\mathscr{P} = \{P_1, \ldots, P_k\}$ задачи RPPDP. Зададим значения индикаторных переменных $x_e^{r,a}$ и y_v^r в соответствии с вхождением дуг $e \in E$ и вершин $v \in V$ в каждый план r данного решения и установим значение переменной *cost* равным стоимости $\mathrm{cost}(\mathscr{P}) = \max\{\mathrm{cost}(P_r) : r = \overline{1,k}\}$. Нетрудно убедиться, что построенный таким образом набор значений удовлетворяет каждому из ограничений MILP-модели (1)–(11).

С другой стороны, рассмотрим произвольное допустимое решение $(x, y, cost)$ задачи (1)–(11). Построим подграфы $P_1, \ldots, P_k$, относя к подграфу P_r вершины и дуги графа G, для которых выполнено соотношение $x_e^{r,a} = 1$ (для некоторой дуги $a \in A$) и $y_v^r = 1$, соответственно. Семейство $\mathscr{P} = \{P_1, \ldots, P_k\}$ является допустимым решением исходной постановки задачи RPPDP. В самом деле, справедливость ограничений (3)–(7) обеспечивает реализацию произвольной дуги $(i, j) \in A$ в каждом подграфе P_r в точности единственным маршрутом, начинающемся в некоторой вершине кластера $\mathfrak{M}_i$, посещающем некоторое количество хабов и завершающемся в вершине кластера $\mathfrak{M}_j$. Неравенство (8) гарантирует для произвольного хаба $h \in H$, что он используется маршрутами подграфа P_r только в случае его открытия и в пределах пропускной способности. В силу ограничения (9), произвольные планы P_{r_1} и P_{r_2} пересекаются только в вершинах s и t. Наконец, из истинности ограничения (2) следует, что стоимость плана P_r совпадает со суммарными транспортными издержками, ассоциированными с входящими в него дугами, дополненными суммарной стоимостью открытия транспортных хабов, используемых данным планом. Теорема доказана.

Приведенные ниже результаты численных экспериментов подтверждают высокое качество нижних оценок, предоставляемых вещественной релаксацией модели (1)–(11) и перспективность базирующихся на ней методов ветвления.

6. АЛГОРИТМЫ

Введенные в рассмотрение в данном разделе алгоритмы решения задачи RPPDP принадлежат семейству методов ветвей, границ и отсечений и основаны на реализации данного подхода в известном MIP-солвере Gurobi [21]. Алгоритм A_1 состоит в применении Gurobi со стандартными настройками внутренних параметров к MILP-модели (1)–(11). В отличие от него в алгоритме A_2 все используемые в солвере (встроенные) первичные эвристики заменены единственной специализированной эвристикой адаптивного поиска, описание которой приведено ниже.

6.1. Адаптивный поиск

Метаэвристический подход адаптивного поиска в больших окрестностях (Adaptive Large Neighborhood Search, ALNS), предложенный в работе [22], основан на последовательном повторении фаз разрушения / восстановления исходного допустимого решения с использованием заданного набора базовых эвристик и онлайн-обучения.

В данной работе мы развиваем модификацию ALNS, введенную в работе [23], использующую процедуры локального поиска и случайного возмущения, каждая из которых включает в себя как процесс

разрушения, так и процесс восстановления передаваемого ей решения. Договоримся называть предложенную эвристику *Reliable Production Process Design Adaptive Large Neighborhood Search (RPPD-ALNS)*. Псевдокод RPPD-ALNS представлен в алгоритме 1.

Algorithm 1. RPPD-ALNS :: общая схема

Входные данные: постановка RPPDP (G, Π, k)
Параметры: число запусков $i_{\max}$, банки процедур локального поиска $\mathfrak{Ls}$ и возмущения $\mathfrak{Sh}$
Выходные данные: результирующее допустимое решение $\overline{\mathscr{P}}$

for $0 < i \leqslant i_{\max}$ **do**
  получить стартовое решение $\mathscr{P}$ с использованием нескольких рандомизированных запусков жадного алгоритма;
  положить $\mathscr{R}_{\text{best}}(i) = \mathscr{P}$;
  repeat
    случайным образом выбрать процедуры возмущения и локального поиска sh и ls из текущих вероятностных распределений на множествах $\mathfrak{Sh}$ и $\mathfrak{Ls}$;
    случайное возмущение: получить новое допустимое решение $\mathscr{P}' = sh(\mathscr{P})$;
    локальный поиск: найти допустимое решение $\mathscr{P}_{\text{new}} = ls(\mathscr{P}')$
    if $\text{cost}(\mathscr{P}_{\text{new}}) < \text{cost}(\mathscr{P}')$ и не исчерпано число попыток **then**
      положить $\mathscr{P}' = \mathscr{P}_{\text{new}}$ и вернуться на шаг 7
    end if
    if $\text{cost}(\mathscr{P}_{\text{new}}) < \text{cost}(\mathscr{R}_{\text{best}}(i))$ **then**
      $\mathscr{R}_{\text{best}}(i) = \mathscr{P}_{\text{new}}$
    end if
    if *критерий принятия решения* выполнен для $\mathscr{P}_{\text{new}}$ и $\mathscr{P}$ **then**
      $\mathscr{P} = \mathscr{P}_{\text{new}}$
      запомнить улучшения, сделанные эвристиками sh и ls;
    end if
  until не выполнится *критерий остановки*
  учитывая совершенные улучшения, обновить распределения множеств $\mathfrak{Sh}$ и $\mathfrak{Ls}$
end for
return $\overline{\mathscr{P}} = \arg\min\{\text{cost}(\mathscr{R}_{\text{best}}(i)) : i \in \{1, \ldots, i_{\max}\}\}$

Основной цикл эвристики RPPD-ALNS состоит из $i_{\max}$ последовательных “холодных” запусков, каждый из которых начинается с нескольких попыток построения начального допустимого решения задач. Каждая такая попытка состоит в фиксации произвольного упорядочения множества дуг орграфа Π и применении жадного алгоритма для их маршрутизации в графе G.

На каждой итерации внутреннего цикла алгоритма последовательно применяются процедуры случайного возмущения sh и локального поиска ls, выбранные в соответствии с текущими вероятностными распределениями, заданными на встроенных банках эвристик $\mathfrak{Sh}$ и $\mathfrak{Ls}$. Каждая используемая нами процедура может рассматриваться как преобразование множества допустимых решений задачи, тем или иным способом трансформирующее подаваемое на вход решение. Так, цель процедуры sh состоит в возможном покидании окрестности локального оптимума, в то время как процедура локального поиска ls стремится улучшить качество найденного решения.

Используемое в алгоритме условие принятия найденного решения $\mathscr{P}_{\text{new}}$ традиционно опирается на метод имитации отжига (как, например, в работе [24]). При выполнении данного критерия решение $\mathscr{P}_{\text{new}}$ переводится в состояние текущего $\mathscr{P}$ и запоминаются изменения, сделанные выбранными процедурами sh и ls.

Произвольная итерация внутреннего цикла завершается проверкой критерия остановки, совпадающего в данной версии алгоритма с условием отсутствия улучшения решения на протяжении заданного числа итераций. Перед переходом на следующую итерацию обновляются вероятностные распределения на множествах $\mathfrak{Sh}$ и $\mathfrak{Ls}$ с учетом влияния конкретных эвристик на повышение качества найденных решений.

6.2. Эвристики случайного возмущения и локального поиска

Остановимся на описании множеств $\mathfrak{Sh}$ и $\mathfrak{Ls}$, содержащих адаптации известных алгоритмов локального поиска и случайного возмущения (см., например [25]). Каждая из представленных эвристик обрабатывает набор маршрутов, реализующих дуги графа порядка П. В данной работе используются следующие пять процедур возмущения текущего решения, целью которых является обеспечить выход за пределы окрестности локального оптимума:

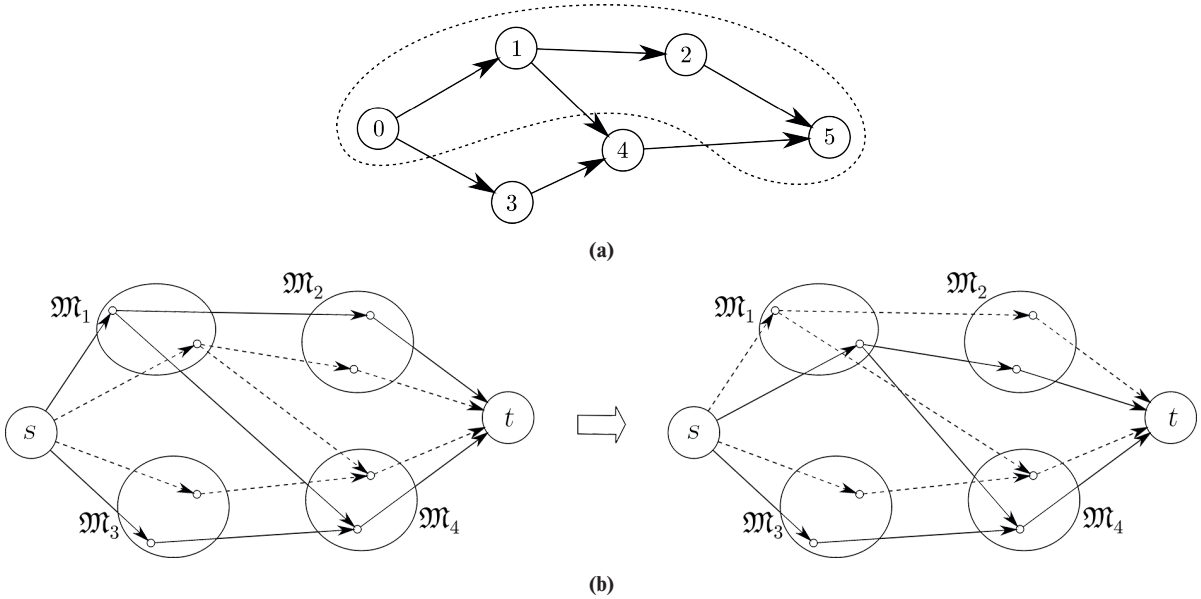
• вставка хабов случайным образом распределяет часть неиспользованных транспортных хабов между планами строящегося решения;

• обмен хабами для случайно выбранной пары планов производит обмен множествами назначенных им транспортных хабов;

• процедура переназначения кластерной вершины для случайно выбранной пары план-кластер замещает вершину кластера, включенную в заданный план, произвольной другой допустимой вершиной;

• процедура обмена кластерными вершинами для произвольно выбранной пары планов обменивает вершины, используемые данными планами в случайно выбранном кластере;

• обмен путями замещает маршрутизацию произвольного s-t-пути в графе порядка П в случайно выбранном плане маршрутизацией этого пути в произвольном другом плане (см. фиг. 5).



Фиг. 5. Пример применения обмена s-t путями для постановки RPPDP с $m = 4$; (a) граф порядка П, с выбранным путем $s \to M_1 \to M_2 \to t$; (b) планы в графе G до и после обмена маршрутами.

Каждая из данных процедур случайного возмущения неявным образом индуцирует специфические окрестности на множестве допустимых планов, что приводит к построению процедур локального поиска, основанных на схожих принципах. В данной версии алгоритма множество $\mathfrak{Ls}$ полностью определятся множеством $\mathfrak{Sh}$ и совпадает с ним по мощности.

Эвристика реализована на языке Python 3.8, с использованием библиотеки NetworkX и интегрирована в MIP-солвер Gurobi в рамках фрейморка Gurobi user callback. Исходный код размещен по адресу https://github.com/yogorodnikov/RPPDP-instances/tree/master/RPPD-ALNS.

7. ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ

В данном разделе описывается методика построения постановок задачи RPPDP и обсуждаются результаты оценивания производительности предложенных алгоритмов, предложенных в предыдущем разделе.

7.1. Библиотека тестовых примеров

Как и для большинства новых комбинаторных задач, тестирование алгоритмов для задачи RPPDP оказалось затруднено отсутствием общепризнанной библиотеки тестовых примеров и, как следствие, необходимостью создания такой библиотеки авторами данной статьи. Предложенная нами библиотека является адаптацией публичной библиотеки PCGTSPLIB [20], предназначенной для тестирования алгоритмов для задачи PCGTSP и восходящей к классической библиотеке TSPLIB.

В частности, каждая постановка задачи PCGTSP нами использована для построения рандомизированного семейства постановок задачи RPPDP по следующему принципу:

- граф G и функция транспортных издержек наследуются из исходной постановки PCGTSP;
- стартовой вершине постановки PCGTSP сопоставляется пара вершин s и t так, что все исходящие дуги этой вершины переназначаются вершине s, а входящие — вершине t;
- кластеры, содержащие не менее, чем k вершин, включаются в постановку RPPDP с сохранением отношений предшествования;
- кластеры меньшего размера подлежат расформированию, входящие в них вершины объявляются транспортными хабами;
- для каждого полученного хаба h назначаются его пропускная способность q_h и стоимость его открытия C_h, равномерно распределенные в интервалах $[1, q]$ и $[1, C]$ для заданных значений параметров q и C, соответственно.

Для каждой из полученных таким образом упорядоченных пар (G, Π) формируются постановки RPPDP, (G, Π, k) с заданным числом планов k. Доступ к сформированным примерам открыт по ссылке https://github.com/yogorodnikov/RPPDP-instances.

7.2. Параметры эксперимента

Для проведения эксперимента были отобраны тестовые примеры, обладающие следующими характеристиками:

- число планов $k = 2, 3$;
- величина стоимости открытия хабов C ограничена сверху средним значением весов дуг графа G;
- верхняя граница q пропускной способности хабов последовательно выбиралась из множества $\{1, 5\}$.

Для каждого сочетания графа G, числа планов k и параметра q для проведения эксперимента было сформировано по 10 случайных постановок.

Проведенный эксперимент состоял из трех последовательных стадий. Первая стадия была посвящена исследованию качества вещественной релаксации предложенной MILP-модели. Для каждой исследуемой постановки, в которой удавалось найти оптимальное значение задачи RPPDP, вычислялось относительное значение наблюдаемого разрыва целочисленности (Integrality Gap, IG) по формуле IG = (OPT – LB)/OPT, где OPT — оптимальное значение задачи, LB — наилучшая нижняя оценка (оптимальное значение вещественной релаксации модели). Эксперимент подтвердил высокую релаксационную способность предложенной модели для большинства постановок, что мотивировало переход к численному исследованию основанных на ней алгоритмов ветвей и границ A_1 и A_2.

На второй стадии проведено тестирование базового алгоритма A_1 на постановках RPPDP, соответствующих значениям $q = 1$ и $q = 5$. Сравнительный анализ результатов выявил увеличение как времени выполнения первого алгоритма, так и относительной погрешности получаемых решений. Поэтому третья стадия эксперимента была посвящена численному сравнению производительности обоих алгоритмов именно на этой группе постановок.

Вычисления проводились на суперкомпьютере “Уран” Института математики и механики им. Н. Н. Красовского УрО РАН http://parallel.uran.ru, Intel(R) Xeon(R) 16 core CPU E5-2697 v4 @2.30 GHz 256GB RAM. Время счета для каждого экземпляра задачи ограничивалось 5 часами (18000 сек).

7.3. Результаты и обсуждение

Таблица 1 содержит данные о структуре исследуемых постановок и результаты первой фазы эксперимента: наблюдаемые значения разрыва целочисленности и времени решения релаксированных задач для постановок RPPDP. Как следует из таблицы, максимальное значение разрыва целочисленности не превысило 7.74% для постановок с $q = 1$ и 28.35% для $q = 5$ (выделено жирным шрифтом), в то время как наибольшее время счета составило 3508 и 8406 сек. Заметим, что средние значения этих величин составили 0.34% и 1.74%, соответственно. Таким образом, полученные результаты подтверждают высокую релаксационную способность предложенной MILP-модели и перспективность ее использования в методах ветвления.

Таблица 1. Исследование релаксационной способности предложенной MILP-модели. N и M — число вершин и кластеров в исходной постановке PCGTSP; m — число полученных производственных кластеров

Iname	N	M	m	$\lvert\mathfrak{H}\rvert$	C	k	$q = 1$				$q = 5$			
							IG, %			Max time, sec	IG, %			Max time, sec
							Avg	Min	Max		Avg	Min	Max	
ESC0	39	8	7	4	420	2	0.56	0.56	0.56	0.01	0.56	0.56	0.56	0.0
						3	0.83	0.83	0.83	0.01	0.83	0.83	0.83	0.01
ESC12	65	13	11	5	222	2	0.54	0.54	0.54	0.02	0.54	0.54	0.54	0.02
						3	0.61	0.61	0.61	0.03	0.61	0.61	0.61	0.03
ESC25	133	26	23	5	438	2	3.24	0.1	**7.74**	0.02	18.14	7.27	25.31	0.06
						3	2.53	0.03	5.18	0.06	15.01	6.47	21.86	0.05
ESC47	244	48	42	11	422	2	2.03	0.12	4.33	0.11	14.46	7.16	23.04	0.1
						3	2.27	0.54	5.0	0.6	7.93	22.18	0.6	0.56
ESC63	349	64	57	10	4	2	0.12	0.0	0.21	2.2	1.66	0.46	2.82	3.72
						3	0.13	0.0	0.27	13.69	2.04	1.3	3.04	20.49
ESC78	414	79	62	28	394	2	0.01	0.0	0.05	5.15	3.12	1.51	4.43	3.02
						3	0.01	0.0	0.02	10.03	3.68	2.81	4.26	11.61
br17.10	88	17	14	6	12	2	0.59	0.0	1.52	0.04	4.46	2.06	5.84	0.03
						3	0.52	0.31	0.05	0.09	14.28	1.89	**28.35**	0.1
br17.12	92	17	16	4	14	2	0.63	0.21	1.05	0.03	1.62	1.14	2.48	0.03
						3	0.41	0.13	0.53	0.06	1.88	1.46	2.48	0.07
ft53.1	281	53	43	16	441	2	0.01	0.0	0.01	1.0	0.01	0.0	0.06	0.8
						3	0.01	0.0	0.01	2.73	0.02	0.0	0.06	3.0
ft53.2	274	53	41	18	420	2	0.01	0.0	0.01	1.09	0.02	0.0	0.05	1.07
						3	0.01	0.0	0.01	3.04	0.04	0.02	0.08	3.08
ft53.3	281	53	45	13	438	2	0.07	0.06	0.07	1.37	0.07	0.06	0.07	1.05
						3	0.09	0.08	0.09	3.18	0.09	0.08	0.13	3.11
ft53.4	275	53	45	14	494	2	0.03	0.03	0.03	0.8	0.04	0.03	0.09	0.63
						3	0.14	0.13	0.14	3.45	0.14	0.14	0.16	2.62
ft70.1	346	70	58	18	931	2	0.01	0.0	0.01	2.67	0.01	0.0	0.02	2.4
						3	0.01	0.0	0.01	7.99	0.01	0.0	0.02	9.4
ft70.2	351	70	58	19	878	2	0.01	0.0	0.01	3.4	0.01	0.0	0.01	4.95
						3	0.01	0.0	0.01	8.81	0.01	0.0	0.01	13.73
ft70.3	347	70	52	31	902	2	0.02	0.01	0.02	17.02	0.01	0.01	0.02	30.93
						3	0.04	0.04	0.04	21.58	0.04	0.04	0.04	14.0
ft70.4	353	70	50	32	876	2	0.17	0.17	0.17	37.33	0.17	0.17	0.17	36.31
						3	0.24	0.24	0.24	67.14	0.24	0.24	0.24	61.61
kro124p.1	514	100	78	30	1760	2	0.01	0.01	0.01	10.5	0.04	0.01	0.07	12.67
						3	0.05	0.01	0.1	12.67	0.07	0.01	0.11	11.95

Таблица 1 (окончание)

Iname	*N*	*M*	*m*	$\lvert\mathfrak{H}\rvert$	*C*	*k*	$q = 1$				$q = 5$			
							IG, %			Max time, sec	IG, %			Max time, sec
							Avg	Min	Max		Avg	Min	Max	
kro124p.2	524	100	76	40	1706	2	0.01	0.0	0.01	12.67	0.21	0.13	0.29	12.83
						3	0.01	0.0	0.01	61.53	0.2	0.13	0.27	52.71
kro124p.3	534	100	82	27	1702	2	0.04	0.03	0.05	9.58	0.11	0.05	0.14	6.82
						3	0.04	0.03	0.05	25.77	0.14	0.12	0.16	27.79
kro124p.4	526	100	80	32	1616	2	0.13	0.11	0.15	25.08	0.16	0.12	0.18	34.44
						3	0.13	0.11	0.15	157.09	—	—	—	146.28
p43.1	203	43	31	19	2188	2	0.01	0.0	0.07	1.18	0.36	0.27	0.47	1.32
						3	0.01	0.0	0.02	1.81	0.43	0.27	0.7	2.64
p43.2	198	43	34	16	2384	2	0.02	0.0	0.13	0.51	0.65	0.3	1.43	0.78
						3	0.02	0.0	0.13	1.28	0.71	0.34	1.09	1.73
p43.3	211	43	34	13	2384	2	0.03	0.0	0.08	0.34	0.16	0.08	0.28	0.42
						3	0.01	0.0	0.01	1.14	0.16	0.08	0.25	1.31
p43.4	204	43	33	15	2392	2	0.07	0.01	0.22	0.75	0.57	0.13	0.97	0.68
						3	0.17	0.03	0.53	3.9	0.68	0.17	1.01	3.57
prob.100	510	99	78	36	219	2	0.01	0.0	0.01	7.94	1.4	2.5	3.6	13.01
						3	0.01	0.0	0.01	24.67	1.92	1.19	2.86	48.52
prob.42	208	41	32	13	44	2	0.02	0.0	0.02	0.36	1.71	1.02	2.92	0.36
						3	0.03	0.03	0.03	0.63	1.63	0.86	2.55	1.16
ry48p.1	256	48	41	10	1001	2	0.01	0.0	0.01	0.14	0.01	0.0	0.01	0.09
						3	0.01	0.0	0.01	0.9	0.01	0.0	0.01	0.71
ry48p.2	250	48	39	15	982	2	0.01	0.0	0.01	0.7	0.01	0.0	0.01	0.45
						3	0.01	0.0	0.01	1.85	0.02	0.02	0.02	2.83
ry48p.3	254	48	38	14	1029	2	0.03	0.02	0.03	1.07	0.03	0.02	0.03	1.07
						3	0.02	0.02	0.02	3.26	0.02	0.02	0.02	3.25
ry48p.4	249	48	39	17	1027	2	0.16	0.15	0.16	1.64	0.16	0.15	0.16	1.07
						3	0.25	0.24	0.26	9.26	0.25	0.24	0.26	6.03
rbg048a	255	49	40	14	15	2	1.4	1.35	1.45	74.45	1.78	1.36	2.19	107.65
						3	—	—	—	405.73	—	—	—	428.34
rbg050c	259	51	43	14	17	2	—	—	—	82.48	—	—	—	80.45
						3	—	—	—	1275.09	—	—	—	248.35
rbg109a	573	110	86	36	16	2	—	—	—	2847.8	—	—	—	7621.46
						3	—	—	—	**3507.6**	—	—	—	**8406.4**

В табл. 2 приводятся результаты тестирования алгоритма A_1: продолжительность счета и оценки относительной погрешности найденных решений $gap = (UB - LB)/LB$, где UB — вес наилучшего найденного допустимого решения, LB — значение наилучшей нижней оценки. Видно, что для 92.5% постановок при $q = 1$ найдены оптимальные решения, в то время как для оставшейся части постановок средняя величина относительной погрешности не превысила 4.6%. При $q = 5$ доля постановок задачи, для которых удалось найти оптимальное решение снизилась до 86.3%, в оставшихся случаях максимальная величина относительной погрешности составила 22.6%.

Кроме того, средний прирост времени счета составил 53.6%. В наибольшей степени это проявилось в постановках ESC63, ESC78, kro124p.X и prob.100, соответствующих одним из наиболее труднорешаемых экземпляров задачи PCGTSP [19]. Любопытно, что для некоторых немногочисленных постановок (например, rbg048a) наблюдался обратный эффект.

Таким образом, представилось разумным провести сравнение производительности алгоритма A_1 с производительностью алгоритма A_2, опирающегося на специализированную эвристику RPPD-ALNS, причем именно на второй группе постановок с $q = 5$.

Таблица 2. Результаты тестирования алгоритма A_1

Iname	k	$q = 1$						$q = 5$					
		Time, sec			Gap, %			Time, sec			Gap, %		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
ESC07	2	0.05	0.05	0.05	0.0	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0
	3	0.07	0.06	0.1	0.0	0.0	0.0	0.1	0.08	0.13	0.0	0.0	0.0
ESC12	2	0.21	0.16	0.26	0.0	0.0	0.0	0.2	0.16	0.23	0.0	0.0	0.0
	3	0.4	0.33	0.5	0.0	0.0	0.0	0.1	0.08	0.13	0.0	0.0	0.0
ESC25	2	1.87	1.67	2.14	0.0	0.0	0.0	1.76	1.55	2.11	0.0	0.0	0.0
	3	2.72	2.55	2.94	0.0	0.0	0.0	3.56	2.53	9.0	0.0	0.0	0.0
ESC47	2	13.87	11.8	16.91	0.0	0.0	0.0	23.04	19.86	28.21	0.0	0.0	0.0
	3	22.71	18.32	35.68	0.0	0.0	0.0	38.72	33.74	47.74	0.0	0.0	0.0
ESC63	2	42.4	37.71	48.56	0.0	0.0	0.0	161.77	103.83	278.22	0.0	0.0	0.0
	3	70.74	55.53	89.36	0.0	0.0	0.0	2736.12	277.87	18000	0.05	0.0	0.47
ESC78	2	51.57	41.56	69.16	0.0	0.0	0.0	523.66	236.64	1219.75	0.0	0.0	0.0
	3	78.49	63.4	90.54	0.0	0.0	0.0	8362.78	4255.6	18000	0.2	0.0	0.6
br17.10	2	0.62	0.47	0.8	0.0	0.0	0.0	0.94	0.54	1.55	0.0	0.0	0.0
	3	1.27	0.83	2.57	0.0	0.0	0.0	2.86	0.97	6.53	0.0	0.0	0.0
br17.12	2	0.59	0.5	0.65	0.0	0.0	0.0	0.96	0.84	1.06	0.0	0.0	0.0
	3	1.06	0.79	1.31	0.0	0.0	0.0	2.7	1.62	4.54	0.0	0.0	0.0
ft53.1	2	19.51	18.53	21.47	0.0	0.0	0.0	34.4	27.5	70.9	0.0	0.0	0.0
	3	29.29	26.78	34.83	0.0	0.0	0.0	86.9	65.14	121.09	0.0	0.0	0.0
ft53.2	2	16.61	14.3	19.01	0.0	0.0	0.0	31.56	24.35	40.29	0.0	0.0	0.0
	3	23.89	20.38	28.13	0.0	0.0	0.0	105.62	69.89	153.32	0.0	0.0	0.0
ft53.3	2	17.68	16.15	22.41	0.0	0.0	0.0	30.99	27.02	37.22	0.0	0.0	0.0
	3	29.55	23.72	35.79	0.0	0.0	0.0	92.8	68.47	160.2	0.0	0.0	0.0
ft53.4	2	19.88	17.74	21.72	0.0	0.0	0.0	31.87	29.7	33.57	0.0	0.0	0.0
	3	66.67	54.5	78.75	0.0	0.0	0.0	91.17	81.15	106.04	0.0	0.0	0.0
ft70.1	2	38.17	32.3	44.43	0.0	0.0	0.0	78.38	56.58	144.66	0.0	0.0	0.0
	3	53.76	47.79	61.93	0.0	0.0	0.0	195.77	154.83	217.09	0.0	0.0	0.0
ft70.2	2	39.54	33.97	46.83	0.0	0.0	0.0	62.76	58.38	68.94	0.0	0.0	0.0
	3	53.08	48.13	60.65	0.0	0.0	0.0	167.25	139.82	202.62	0.0	0.0	0.0
ft70.3	2	76.26	66.91	90.01	0.0	0.0	0.0	244.4	133.87	440.93	0.0	0.0	0.0
	3	136.53	107.46	157.18	0.0	0.0	0.0	805.72	619.66	937.09	0.0	0.0	0.0
ft70.4	2	100.07	69.33	152.28	0.0	0.0	0.0	204.98	131.12	1473.48	0.0	0.0	0.0
	3	907.11	461.57	1563.52	0.0	0.0	0.0	4247.74	619.66	937.09	0.0	0.0	0.0
kro124p.1	2	115.6	107.5	129.8	0.0	0.0	0.0	296.72	181.35	376.27	0.0	0.0	0.0
	3	207.36	118.0	222.5	0.0	0.0	0.0	1016.04	474.89	2051.8	0.0	0.0	0.0
kro124p.2	2	147.86	116.9	232.38	0.0	0.0	0.0	597.26	425.34	1006.48	0.0	0.0	0.0
	3	271.22	244.23	363.4	0.0	0.0	0.0	4569.22	2190.44	10058.7	0.0	0.0	0.0
kro124p.3	2	171.38	118.0	205.8	0.0	0.0	0.0	421.9	333.81	509.26	0.0	0.0	0.0
	3	474.2	449.0	499.3	0.0	0.0	0.0	6502.22	3831.5	9230.8	0.0	0.0	0.0
kro124p.4	2	322.83	299.29	343.85	0.0	0.0	0.0	527.79	319.17	1164.03	0.0	0.0	0.0
	3	4006.9	1771.0	6290.2	0.0	0.0	0.0	15380.59	6526.13	18000	0.25	0.0	0.57
p43.1	2	8.47	7.6	9.94	0.0	0.0	0.0	35.12	25.47	47.81	0.0	0.0	0.0
	3	13.71	10.87	16.41	0.0	0.0	0.0	103.79	69.29	187.88	0.0	0.0	0.0
p43.2	2	7.32	6.03	8.77	0.0	0.0	0.0	17.95	12.21	22.17	0.0	0.0	0.0
	3	11.68	8.9	14.63	0.0	0.0	0.0	58.65	36.7	101.32	0.0	0.0	0.0
p43.3	2	7.06	5.9	8.34	0.0	0.0	0.0	15.9	11.13	25.75	0.0	0.0	0.0
	3	11.11	9.87	14.03	0.0	0.0	0.0	54.63	38.1	94.25	0.0	0.0	0.0
p43.4	2	6.9	5.8	8.0	0.0	0.0	0.0	22.0	14.99	39.72	0.0	0.0	0.0
	3	21.63	11.0	50.37	0.0	0.0	0.0	121.13	32.75	350.69	0.0	0.0	0.0

Таблица 2 (окончание)

Iname	*k*	$q = 1$						$q = 5$					
		Time, sec			Gap, %			Time, sec			Gap, %		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
prob.100	2	108.8	96.9	119.86	0.0	0.0	0.0	2141.03	820.13	4525.51	0.0	0.0	0.0
	3	185.9	152.4	231.56	0.0	0.0	0.0	16060.77	6659.19	18000	0.42	0.0	0.0
prob.42	2	6.8	6.35	7.48	0.0	0.0	0.0	13.99	11.26	15.89	0.0	0.0	0.0
	3	9.67	4.6	15.8	0.0	0.0	0.0	30.93	25.93	36.6	0.0	0.0	0.0
ry48p.1	2	17.41	15.8	24.37	0.0	0.0	0.0	23.93	19.89	35.84	0.0	0.0	0.0
	3	23.8	22.52	24.37	0.0	0.0	0.0	52.04	32.75	71.43	0.0	0.0	0.0
ry48p.2	2	15.6	14.2	21.0	0.0	0.0	0.0	24.11	19.09	33.42	0.0	0.0	0.0
	3	22.02	20.8	26.7	0.0	0.0	0.0	150.4	58.32	224.49	0.0	0.0	0.0
ry48p.3	2	17.5	15.9	36.9	0.0	0.0	0.0	45.6	58.32	84.49	0.0	0.0	0.0
	3	29.42	25.78	36.88	0.0	0.0	0.0	81.4	55.36	120.94	0.0	0.0	0.0
ry48p.4	2	17.85	15.6	21.1	0.0	0.0	0.0	27.25	23.25	30.78	0.0	0.0	0.0
	3	177.7	102.8	277.8	0.0	0.0	0.0	213.09	95.17	298.57	0.0	0.0	0.0
rbg048a	2	5516.52	1524.34	10078.05	0.0	0.0	0.0	2633.12	769.66	5617.9	0.0	0.0	0.0
	3	18000	18000	18000	0.71	0.38	1.12	18000	18000	18000	0.88	0.58	1.31
rbg050c	2	18000	18000	18000	1.1	0.69	1.45	18000	18000	18000	3.95	2.63	5.44
	3	18000	18000	18000	2.35	2.02	3.17	18000	18000	18000	5.16	4.94	6.31
rbg109a	2	18000	18000	18000	3.5	2.4	4.2	18000	18000	18000	6.39	2.32	10.36
	3	18000	18000	18000	4.6	2.5	8.52	18000	18000	18000	12.55	5.85	22.6

Таблица 3 отражает результаты сравнения алгоритмов A_1 и A_2 на постановках задачи с $q = 5$. Видно, что алгоритм A_2, использующий эвристику RPPD-ALNS, находит оптимальное решение в среднем в 90.9% случаев против 86.3% для алгоритма A_1. В частности, оптимальные решения постановок ESC63, ESC78, и prob.100 были найдены только алгоритмом A_2.

В 89.4% из тех случаев, где оптимальные решения были найдены обоими алгоритмами, алгоритму A_2 удавалось это сделать быстрее, в среднем на 46.8%. В частности, для постановки ft70.4 алгоритм A_2 оказался производительнее в 17 раз. В тех случаях, где оба алгоритма достигли предела времени счета, например, в задачах kro124p.4, rbg048a при $k = 3$, и rbg050c, rbg109a, алгоритму A_2 удалось найти существенно более точные приближенные решения.

Таблица 3. Результаты сравнения алгоритмов A_1 и A_2 при $q = 5$. Полужирным шрифтом выделены наилучшие полученные результаты

Iname	*k*	A_1						A_2					
		Time, sec			Gap, %			Time, sec			Gap, %		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
ESC07	2	**0.05**	0.05	0.05	0.0	0.0	0.0	0.14	0.12	0.17	0.0	0.0	0.0
	3	0.1	0.08	0.13	0.0	0.0	0.0	0.23	0.21	0.25	0.0	0.0	0.0
ESC12	2	**0.2**	0.16	0.23	0.0	0.0	0.0	0.41	0.36	0.46	0.0	0.0	0.0
	3	**0.41**	0.35	0.46	0.0	0.0	0.0	0.89	0.82	0.96	0.0	0.0	0.0
ESC25	2	**1.76**	1.55	2.11	0.0	0.0	0.0	3.19	2.54	7.52	0.0	0.0	0.0
	3	**3.56**	2.53	9.0	0.0	0.0	0.0	4.47	3.81	6.51	0.0	0.0	0.0
ESC47	2	23.04	19.86	28.21	0.0	0.0	0.0	**15.02**	13.06	19.99	0.0	0.0	0.0
	3	38.72	33.74	47.74	0.0	0.0	0.0	**30.45**	21.55	54.79	0.0	0.0	0.0
ESC63	2	161.77	103.83	278.22	0.0	0.0	0.0	**72.26**	39.53	140.1	0.0	0.0	0.0
	3	2736.12	277.87	18000	0.05	0.0	0.47	**1927.44**	525.23	10472.4	**0.0**	0.0	0.0
ESC78	2	523.66	236.64	1219.75	0.0	0.0	0.0	**338.21**	166.58	745.9	0.0	0.0	0.0
	3	8362.78	4255.6	18000	0.2	0.0	0.6	6982.45	2080.9	10472.4	**0.0**	0.0	0.0

Таблица 3 (продолжение)

Iname	*k*	A_1						A_2					
		Time, sec			Gap, %			Time, sec			Gap, %		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
br17.10	2	**0.94**	0.54	1.55	0.0	0.0	0.0	1.03	0.9	1.26	0.0	0.0	0.0
	3	2.86	0.97	6.53	0.0	0.0	0.0	**2.59**	1.79	3.84	0.0	0.0	0.0
br17.12	2	0.96	0.84	1.06	0.0	0.0	0.0	**0.67**	0.53	0.87	0.0	0.0	0.0
	3	2.7	1.62	4.54	0.0	0.0	0.0	**1.91**	1.65	2.18	0.0	0.0	0.0
ft53.1	2	34.4	27.5	40.2	0.0	0.0	0.0	**18.74**	17.9	19.86	0.0	0.0	0.0
	3	86.9	65.14	121.09	0.0	0.0	0.0	**28.8**	27.9	30.84	0.0	0.0	0.0
ft53.2	2	31.56	24.35	40.29	0.0	0.0	0.0	**16.02**	15.22	17.83	0.0	0.0	0.0
	3	105.62	69.89	153.32	0.0	0.0	0.0	**25.96**	22.97	30.35	0.0	0.0	0.0
ft53.3	2	30.99	27.02	37.22	0.0	0.0	0.0	**18.4**	16.97	21.51	0.0	0.0	0.0
	3	92.8	68.47	160.2	0.0	0.0	0.0	**32.13**	24.48	42.91	0.0	0.0	0.0
ft53.4	2	31.87	29.7	33.57	0.0	0.0	0.0	**19.8**	17.43	23.79	0.0	0.0	0.0
	3	91.17	81.15	106.04	0.0	0.0	0.0	**67.13**	55.54	103.62	0.0	0.0	0.0
ft70.1	2	78.38	56.58	144.66	0.0	0.0	0.0	**36.76**	32.35	40.87	0.0	0.0	0.0
	3	195.77	154.83	217.09	0.0	0.0	0.0	**57.03**	47.97	63.83	0.0	0.0	0.0
ft70.2	2	62.76	58.38	68.94	0.0	0.0	0.0	**36.25**	32.17	43.51	0.0	0.0	0.0
	3	167.25	139.82	202.62	0.0	0.0	0.0	**53.69**	46.78	61.91	0.0	0.0	0.0
ft70.3	2	244.4	133.87	440.93	0.0	0.0	0.0	**75.15**	55.1	96.61	0.0	0.0	0.0
	3	805.72	619.66	937.09	0.0	0.0	0.0	**122.6**	101.91	157.62	0.0	0.0	0.0
ft70.4	2	204.98	131.12	1473.48	0.0	0.0	0.0	**102.9**	80.68	137.17	0.0	0.0	0.0
	3	4247.74	619.66	937.09	0.0	0.0	0.0	**247.82**	101.91	157.62	0.0	0.0	0.0
kro124p.1	2	296.72	181.35	376.27	0.0	0.0	0.0	**131.44**	104.84	167.11	0.0	0.0	0.0
	3	1016.04	474.89	2051.8	0.0	0.0	0.0	**209.68**	179.68	279.07	0.0	0.0	0.0
kro124p.2	2	597.26	425.34	1006.48	0.0	0.0	0.0	**180.78**	137.2	228.41	0.0	0.0	0.0
	3	4569.22	2190.44	10058.7	0.0	0.0	0.0	**335.16**	252.12	219.53	0.0	0.0	0.0
kro124p.3	2	421.9	333.81	509.26	0.0	0.0	0.0	**185.7**	137.2	228.41	0.0	0.0	0.0
	3	6502.22	3831.5	9230.79	0.0	0.0	0.0	**781.82**	569.19	951.14	0.0	0.0	0.0
kro124p.4	2	527.79	319.17	1164.03	0.0	0.0	0.0	**342.18**	311.02	398.24	0.0	0.0	0.0
	3	15380.59	6526.13	18000	0.25	0.0	0.57	8268.88	4440.07	18000	**0.05**	0.0	0.28
p43.1	2	35.12	25.47	47.81	0.0	0.0	0.0	**14.28**	11.47	22.84	0.0	0.0	0.0
	3	103.79	69.29	187.88	0.0	0.0	0.0	**53.92**	35.4	73.26	0.0	0.0	0.0
p43.2	2	17.95	12.21	22.17	0.0	0.0	0.0	**12.16**	8.97	16.14	0.0	0.0	0.0
	3	58.65	36.7	101.32	0.0	0.0	0.0	**48.87**	25.03	73.67	0.0	0.0	0.0
p43.3	2	15.9	11.13	25.75	0.0	0.0	0.0	**8.62**	6.34	13.32	0.0	0.0	0.0
	3	54.63	38.1	94.25	0.0	0.0	0.0	**21.67**	13.44	31.89	0.0	0.0	0.0
p43.4	2	22.0	14.99	39.72	0.0	0.0	0.0	**17.88**	7.01	33.8	0.0	0.0	0.0
	3	121.13	32.75	350.69	0.0	0.0	0.0	**110.87**	64.17	219.0	0.0	0.0	0.0
prob.100	2	2141.03	820.13	4525.51	0.0	0.0	0.0	**1801.8**	564.67	2811.64	0.0	0.0	0.0
	3	16060.77	6659.19	18000	0.42	0.0	1.31	12403.5	7220.5	16305.5	**0.0**	0.0	0.0
prob.42	2	13.99	11.26	15.89	0.0	0.0	0.0	**10.09**	7.66	13.64	0.0	0.0	0.0
	3	30.93	25.93	36.6	0.0	0.0	0.0	**24.9**	18.28	37.88	0.0	0.0	0.0
ry48p.1	2	23.93	19.89	35.84	0.0	0.0	0.0	**11.94**	11.59	12.36	0.0	0.0	0.0
	3	52.04	32.75	71.43	0.0	0.0	0.0	**18.02**	16.09	20.3	0.0	0.0	0.0
ry48p.2	2	24.11	19.09	33.42	0.0	0.0	0.0	**11.06**	10.32	12.38	0.0	0.0	0.0
	3	150.4	58.32	224.49	0.0	0.0	0.0	**64.82**	48.33	93.95	0.0	0.0	0.0
ry48p.3	2	45.6	58.32	84.49	0.0	0.0	0.0	**29.72**	20.65	48.31	0.0	0.0	0.0
	3	81.4	55.36	120.94	0.0	0.0	0.0	**79.2**	52.18	108.58	0.0	0.0	0.0
ry48p.4	2	27.25	23.25	30.78	0.0	0.0	0.0	**12.84**	11.36	14.94	0.0	0.0	0.0
	3	213.09	95.17	298.57	0.0	0.0	0.0	**122.59**	68.73	185.67	0.0	0.0	0.0

Таблица 3 (окончание)

Iname	*k*	A_1						A_2					
		Time, sec			Gap, %			Time, sec			Gap, %		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
rbg048a	2	2633.12	769.66	5617.9	0.0	0.0	0.0	**2622.41**	518.83	7386.01	0.0	0.0	0.0
	3	18000	18000	18000	0.88	0.58	1.31	18000	18000	18000	**0.64**	0.46	0.9
rbg050c	2	18000	18000	18000	3.95	2.63	5.44	18000	18000	18000	**1.63**	1.45	1.86
	3	18000	18000	18000	5.16	4.94	6.31	18000	18000	18000	**2.6**	2.04	3.48
rbg109a	2	18000	18000	18000	6.39	2.32	10.36	18000	18000	18000	**2.02**	1.4	3.8
	3	18000	18000	18000	12.55	5.85	22.6	18000	18000	18000	**2.7**	2.13	4.83

8. ЗАКЛЮЧЕНИЕ

В статье введена в рассмотрение новая комбинаторная задача проектирования отказоустойчивых производственных процессов (Reliable Production Process Design Problem, RPPDP), моделирующая эффективное построение логистических маршрутов и цепей поставок, защищенных от широкого круга возможных сбоев в транспортных сетях. Показано, что с теоретической точки зрения введенная задача близка к нескольким известным комбинаторным задачам и NP-трудна в сильном смысле.

Для эффективного поиска оптимальных и приближенных решений с оценками точности разработаны алгоритмы ветвей и границ A_1 и A_2, основанные на предложенной компактной MILP-модели и эвристике адаптивного поиска. Для тестирования разрабатываемых алгоритмов предложена открытая библиотека постановок задачи, развивающая известную библиотеку PCGTSPLIB. Результаты численного тестирования показали, что предложенные алгоритмы успешно дополняют друг друга: с постановками малого размера успешно справляется более простой алгоритм A_1, в то время как алгоритм A_2, опирающийся на специализированную эвристику, демонстрирует более высокую производительность на постановках большего размера.

На данный момент открытыми остаются вопросы исследования аппроксимируемости задачи RPPDP в классе алгоритмов с априорными теоретическими оценками точности и проектирования алгоритмов ветвей и оценок (branch-and-price).

СПИСОК ЛИТЕРАТУРЫ

1. *Schilling L., Seuring S.* Linking the digital and sustainable transformation with supply chain practices // Int. J. Prod. Res. 2023. https://doi.org/10.1080/00207543.2023.2173502
2. *Fan Y., Schwartz F., Vob S., Woodruff D. L.* Catastrophe insurance and flexible planning for supply chain disruption management: a stochastic simulation case study // Int. J. Prod. Res. 2023. https://doi.org/10.1080/00207543.2023.2176179
3. *Fortune S., Hopcroft J., Wyllie J.* The directed subgraph homeomorphism problem // Theor. Comput. Sci. 1980. V. 10. N 2. P. 111–121.
4. *Eilam-Tzoreff T.* The disjoint shortest paths problem // Discret. Appl. Math. 1998. V. 85. N 2. P. 113–138.
5. *Ferone D., Festa P., Guerriero F., Laganà D.* The constrained shortest path tour problem // Comput. Oper. Res. 2016. V. 74. P. 64–77.
6. *Ferone D., Festa'P., Guerriero F.* An efficient exact approach for the constrained shortest path tour problem // Optim. Methods Softw. 2020. V. 35. N 1. P. 1–20.
7. *Martin S., Magnouche Y., Juvigny C., Leguay J.* Constrained shortest path tour problem: branch-and-price algorithm // Comp. Oper. Res. 2022. V. 144. P. 105819. https://doi.org/10.1016/j.cor.2022.105819
8. *Saksena J. P., Kumar S.* The routing problem with 'k' specified nodes // Oper. Res. 1966. V. 14. N 5. P. 909–913.
9. *Kudriavtsev A., Khachay D., Ogorodnikov Y., Ren J., Shao S. C., Zhang D., Khachay M.* The shortest simple path problem with a fixed number of must-pass nodes: a problem-specific branch-and-bound algorithm // LNCS. 2021. V. 12931. P. 198–210.

10. *Andrade R. C. d.* New formulations for the elementary shortest-path problem visiting a given set of nodes // Eur. J. Oper. Res. 2016. V. 254. N 3. P. 755–768.
11. *Gutin G., Punnen A. P.* The Traveling Salesman Problem and Its Variations. B.: Springer. 2007.
12. *Papadimitriou C.* Euclidean TSP is NP-complete // Theor. Comput. Sci. 1977. V. 4. P. 237–244.
13. *Khachay M., Ukolov S., Petunin A.* Problem-Specific Branch-and-Bound Algorithms for the Precedence Constrained Generalized Traveling Salesman Problem // LNCS. 2021. V. 13078. P. 136–148.
14. *Chentsov A. G., Khachai M. Y., Khachai D. M.* An exact algorithm with linear complexity for a problem of visiting megalopolises // Proc. Steklov Inst. Math. 2016. V. 295. N 1. P. 38–46.
15. *Khachai M. Y., Neznakhina E. D.* Approximation schemes for the Generalized Traveling Salesman Problem // Proc. Steklov Inst. Math. 2017. V. 299. Suppl. 1. P. 97–105.
16. *Khachay M., Neznakhina K.* Complexity and approximability of the Euclidean Generalized Traveling Salesman Problem in grid clusters // Ann. Math. Artif. Intell. 2020. V. 88. N 1. P. 53–69.
17. *Khachay M., Kudriavtsev A., Petunin A.* PCGLNS: A heuristic solver for the Precedence Constrained Generalized Traveling Salesman Problem // LNCS. 2020. V. 12422. P. 196–208.
18. *Morin T. L., Marsten R. E.* Branch-and-bound strategies for dynamic programming // Oper. Res. 1976. V. 24. N 4. P. 611–627.
19. *Khachai D., Sadykov R., Battaia O., Khachay M.* Precedence Constrained Generalized Traveling Salesman Problem: Polyhedral study, formulations, and branch-and-cut algorithm // Eur. J. Oper. Res. 2023. V. 309. N 2. P. 488–505.
20. *Salman R., Ekstedt F., Damaschke P.* Branch-and-bound for the Precedence Constrained Generalized Traveling Salesman Problem // Oper. Res. Lett. 2020. V. 48. N 2. P. 163–166.
21. Gurobi Optimization. Gurobi optimizer reference manual (2021), https://www.gurobi.com/documentation/9.5/refman/index.html
22. *Ropke S., Pisinger D.* An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows // Transp. Sci. 2006. V. 40. P. 455–472.
23. *Kalateh Ahani I., Salari M., Hosseini S. M., Iori M.* Solution of minimum spanning forest problems with reliability constraints // Comput. Ind. Eng. 2020. Vol. 142. P. 106365. https://doi.org/10.1016/j.cie.2020.106365
24. *Smith S. L., Imeson F.* GLNS: An effective large neighborhood search heuristic for the Generalized Traveling Salesman Problem // Comp. Oper. Res. 2017. V. 87. P. 1–19.
25. *Gendreau M., Potvin J.-Y.* Handbook of Metaheuristics. Cham: Springer. 2019.

FAULT-TOLERANT FAMILIES OF PRODUCTION PLANS: MATHEMATICAL MODEL, COMPUTATIONAL COMPLEXITY AND BRANCH AND BOUND ALGORITHMS

Yu. Yu. Ogorodnikov[a,*], R. A. Rudakov[a,**], D' M. Khachay[b,***], M. Yu. Khachay[a,****]

[a] *Krasovskii Institute of Mathematics and Mechanics, Ural Branch, Russian Academy of Sciences, S. Kovalevskaya st., 16, Yekaterinburg, 620108 Russia*

[b] *33405 Talence, 680 cours Libŭration, KEDGE Business School, France*

[*] *e-mail: yogorodnikov@imm.uran.ru*

[**] *e-mail: r.a.rudakov@gmail.com*

[***] *e-mail: daniil.khachai@kedgebs.com*

[****] *e-mail: mkhachay@imm.uran.ru*

Received 10 October, 2023
Revised 28 November, 2024
Accepted 05 March, 2024

Abstract. The design of fault-tolerant production and supply systems is one of the priority areas of development of modern operations research. The traditional approach to modeling such systems is based on the use of probabilistic models describing the choice of a possible scenario of actions in the event of failures in the production or transport network. Along with a number of advantages, this approach has a well-known drawback. The occurrence of failures of an unknown nature that can jeopardize the operability of the entire modeled system significantly complicates its application. In this paper, we introduce the minimax problem of constructing fault-tolerant production plans (Reliable Production Process Design Problem, RPPDP), the purpose of which is to ensure the smooth functioning of a distributed production system with minimal guaranteed costs. It is shown that the RPPDP problem is NP-hard in the strong sense and remains intractable under fairly specific conditions. To find exact and approximate solutions with accuracy estimates for this problem, branch and bound methods have been developed based on the proposed compact model of mixed integer linear programming (MILP) and the author's heuristics of adaptive large neighborhood search (ALNS) within the framework of extensions of the well-known Gurobi MIP-solver. High performance and complementarity of the proposed algorithms have been confirmed by the results of numerical experiments conducted on an open library of test examples developed by the authors, containing adapted problem statements from the PCGTSPLIB library.

Keywords: problem of designing fault-tolerant manufacturing processes, MILP model, branch and bound method, heuristics of adaptive search in large neighborhoods.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 959–972

УДК 517.977.1; 517.977.56

ОБ УПРАВЛЯЕМОСТИ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ

© 2024 г. В. К. Толстых[1,*]

[1] *283001 Донецк, ул. Университетская, 24, Донецкий государственной университет, Россия*

[*]e-mail: mail@tolstykh.com

Поступила в редакцию 09.09.2023 г.
Переработанный вариант 14.02.2024 г.
Принята к публикации 06.03.2024 г.

Рассматривается проблема управляемости для задач оптимального управления, оптимизации системами с распределенными параметрами в частных производных. Вводится понятие управляемости как корректности по А. Н. Тихонову для решения задач оптимизации. Приводится теорема с условиями управляемости для прямого решения (непосредственной минимизации целевого функционала) задач оптимизации экстремальными алгоритмами. Рассматривается тестовой пример численного решения задачи оптимизации нелинейной гиперболической системы, описывающей нестационарное течение воды в открытом русле. Демонстрируется анализ управляемости, который обеспечивает корректность решения задачи и высокую точность оптимизации распределенного коэффициента трения в уравнениях течения. Библ. 13. Фиг. 5.

Ключевые слова: управляемость, система с распределенными параметрами, оптимизация, оптимальное управление, идентификация, градиент, экстремальные методы.

DOI: 10.31857/S0044466924060067, **EDN:** XYULVG

1. ВВЕДЕНИЕ

На сегодняшний день для произвольных бесконечномерных распределенных систем (систем с распределенными параметрами) не выработано единого понятия и подхода к анализу управляемости.

В самом общем понимании управляемость — это возможность перевода системы из одного состояния в другое. В рассматриваемом случае управляемость — это совокупность условий, определяющих принципиальную возможность управления состоянием $v(\tau) \in V(\overline{\Omega})$ распределенной системы с помощью распределенного параметра $u(\tau) \in U(S)$, $S \subset \overline{\Omega}$. Здесь τ — пространственно-временная переменная, $\overline{\Omega}$ — замкнутая область функционирования системы, V — пространство (допустимое множество) состояний, S — область определения управления, U — пространство или допустимое множество управлений. Если система не управляема в принципе, то нет смысла искать оптимальное управление, оптимизировать систему.

При рассмотрении задач оптимизации систем с распределенными параметрами преобладают попытки обобщения ранее полученных результатов управляемости для систем с сосредоточенными параметрами. Но обычно, как отмечается в [1], “вопрос о существовании решения поставленной задачи для бесконечномерной системы зачастую даже не рассматривается”. Если управляемость рассматривается, то она трактуется как возможность перевода системы из заданного начального в заданное финальное состояние. Такое понятие управляемости для широкого круга задач оптимизации распределенных систем зачастую оказывается бесполезным, поскольку управляемость по заданному финальному состоянию не гарантирует управляемость по условиям, накладываемым на неизвестное финальное состояние (почти всегда заданное не на всей пространственной области системы) в целевом функционале.

Традиционное понятие и условия управляемости всегда связывают только с видом уравнений исходной задачи и не связывают с видом целевого функционала. Но такая управляемость не является достаточной для обеспечения принципиальной возможности решения задач оптимизации систем с распределенными параметрами на основе весьма разнообразных, пространственно-распределенных форм управления и целевого функционала (см. [2]).

Кроме того, существующие условия управляемости жестко связаны с предлагаемыми конкретными методами решения задач оптимального управления, например, с методом проблемы моментов, принципа максимума Понтрягина, динамического программирования. Это означает, что если задача оптимизации решается каким-либо иным методом, то для ответа на вопрос управляемости необходимо дублировать ее решение соответствующим методом оптимизации на котором основаны те или иные условия управляемости.

Представляется целесообразным, в частности, для систем с распределенными параметрами расширить понятие управляемости от “перевода системы в финальное состояние” до более практичного понятия “возможности управления системой на основе заданной цели управления”. Описываемое далее понятие управляемости основано на теории некорректных задач и связывает управляемость системы с целевым функционалом. Новое понятие управляемости не зависит от методов оптимизации, а для демонстрации его возможного применения мы выберем прямой экстремальный подход (см. [3]–[5]). Такой подход можно считать наиболее универсальным для численного решения задач оптимального управления, оптимизации с уравнениями в частных производных, где точные аналитические решения удается получить лишь в исключительных случаях.

Суть похода заключается в прямой минимизации экстремальными алгоритмами целевого функционала

$$\begin{gathered} J(u) = \int_{\omega} I(v,u)\,d\omega \to \min, \quad \omega \subset \overline{\Omega}, \\ \text{при условии } \mathbb{D}(\tau, v, u)v = 0, \quad \tau \in \overline{\Omega}. \end{gathered} \tag{1}$$

Здесь оператор $\mathbb{D}$ включает в себя не только конкретный вид дифференциальных уравнений распределенной системы на множестве Ω, но и краевые условия где-либо на границе $\partial\Omega$. Функция цели $I\big(v(\tau), u(\tau)\big)$ определена на множестве ω, а ее значение зависит от параметра v и, возможно, u. В прямом подходе не используются какие-либо промежуточные (например, необходимые) условия оптимальности, а непосредственно решается задача

$$u_* = \arg\min J(u), \tag{2}$$

где $u_*(\tau)$ — оптимальное управление (оптимум, оптималь, оптимальный параметр), доставляющий глобальный минимум функционалу $J(u)$. Все эти названия аргумента u_* в прямом экстремальном подходе являются синонимами.

Сами уравнения $\mathbb{D}v = 0$ распределенной системы в задаче (2) используются для определения направлений пошаговой минимизации $J(u)$. В подавляющем большинстве экстремальных методов направления минимизации формируются с использованием градиента $\nabla J(u)$. Чтобы найти ∇J можно, например (см. [3], [6]), взять первую вариацию задачи (1) относительно δv и δu. При этом задача линеаризуется:

$$\begin{gathered} \delta J = \langle I'_v, \delta v\rangle_{V^*(\omega)} + \langle I'_u, \delta u\rangle_{U^*(\omega)}, \\ \delta\mathbb{D}v = \mathbb{V}\delta v + \mathbb{U}\delta u = 0 \in V(\overline{\Omega}), \end{gathered}$$

где V^*, U^* — линейные сопряженные пространства состояний и управлений, $I'_v \in V^*(\omega)$, $I'_u \in U^*(\omega)$ — производные функции цели, $\mathbb{V}$, $\mathbb{U}$ — линейные операторы исходной задачи. Верхний индекс * здесь и далее означает сопряженность. После перевода вариации системы $\delta\mathbb{D}v$ в сопряженное пространство с помощью сопряженных переменных $f \in V^*(\overline{\Omega})$ и объединения ее с δJ получаем сопряженную задачу и градиент:

$$\mathbb{V}^* f + I'_v\big|_w = 0 \in V^*(\overline{\Omega}), \tag{3}$$

$$\nabla J(u) = \mathbb{U}^* f + I'_u\Big|_{\omega} \equiv \mathbb{U}^*_{\varnothing} f \in U^*(S), \tag{4}$$

где $\mathbb{V}^*$, $\mathbb{U}^*$, $\mathbb{U}^*_{\varnothing}$ — линейные сопряженные операторы. Индекс $\varnothing$ означает отсутствие нулевого ядра у неоднородного оператора $\mathbb{U}^*_{\varnothing}$, в отличие от $\mathbb{U}^*$.

Теперь могут применяться градиентные алгоритмы. Для распределенных управлений целесообразно использовать метод с регулируемым направлением спуска (см.[3], [7]):

$$u^{k+1}(\tau) = u^k(\tau) - b^k \alpha(\tau) \nabla J(u^k; \tau), \quad \tau \in S, \quad k = 0, 1, \ldots, \tag{5}$$

где b^k — шаговый множитель, задающий глубину спуска вдоль антиградиента на каждой итерации k, α — параметр регулирования направления спуска для обеспечения равномерной сходимости к функции $u_*(\tau)$ и компенсации некоторого типа помех.

Традиционные понятия управляемости связывают всегда с видом оператора $\mathbb{D}$ исходной прямой задачи и не связывают с видом критерия оптимизации J, т.е. не связывают ни с видом функции цели оптимизации I, ни с пространственно-временным множеством задания цели ω (всегда предполагают, что ω — область терминального состояния).

Мы изложим новую точку зрения на проблему управляемости распределенных систем. За основу возьмем подход к анализу идентифицируемости (см. [8]), который здесь существенно разовьем.

2. ПОНЯТИЕ УПРАВЛЯЕМОСТИ

Задачи оптимального управления являются обратными задачами. Характер отображений исходной прямой задачи с оператором $\mathbb{D}$ имеет вид

$$U(S) \to V(\overline{\Omega}), \quad S \subset \overline{\Omega}.$$

В задаче же оптимизации отображение — обратное, с условием $\min J$:

$$V(\omega) \overset{\min J}{\longrightarrow} u_*(\tau) \in U(S), \quad \omega \subset \overline{\Omega}.$$

Для решения обратных задач А. Н. Тихонов ввел отличное от классического понятие корректности (см. [9]), заключающееся в требовании ограничения пространства решений U до компактного множества $\mathscr{U}$ существования, единственности и устойчивости управлений. Для сужения U до компакта $\mathscr{U}$ используют регуляризирующие, в том числе, и экстремальные алгоритмы (см. [7]).

Предлагается трактовка управляемости как корректности по Тихонову обратной задачи условного отображения. Такое отождествление управляемости с указанной корректностью позволяет унифицировать схему выявления условий гарантирующих, либо напротив, указывающих на принципиальную невозможность определения оптимального управления той или иной системы по заданному целевому функционалу. Подытожим сказанное следующим определением.

Определение 1. Математическая модель распределенной системы $\mathbb{D}(\tau, \upsilon, u)\upsilon = 0$, $\tau \in \overline{\Omega}$ в задаче (1), характеризуемая прямым отображением

$$U(S) \to V(\overline{\Omega}), \quad S \subset \overline{\Omega},$$

управляема посредством $u(\tau) \in U(S)$ относительно целевого функционала J, когда обратная задача

$$V(\omega) \overset{\min J}{\longrightarrow} u_* \in U(S), \quad \omega \subset \overline{\Omega},$$

отображения элементов пространства $V(\omega)$ состояний модели в элемент u_* множества допустимых управлений U при условии $\min J$ является корректной по Тихонову.

Прямой подход состоит из двух практически самостоятельных подзадач. Первая — это минимизация целевого функционала J на основе ∇J, вторая — определение ∇J. При этом проблема управляемости также разбивается на две подзадачи:

1) регуляризация в алгоритмах прямой минимизации типа (5), использующих градиент ∇J;

2) корректность нахождения градиента ∇J из выражения вида (4).

Первое требование выполняется автоматически (см. [7]), если начальное приближение управления u^0 принадлежит компакту корректности $\mathcal{U}$, а последующие коррекции алгоритмом (5) осуществляются с удовлетворительными параметрами регуляризации $b^k\alpha$.

Определение 2. Параметры $b^k\alpha$, $k = 0, 1, \ldots$, являются *удовлетворительными* параметрами регуляризации в экстремальных алгоритмах, если при $u^0 \in \mathcal{U}$ последующие управления $u^{k+1} \in \mathcal{U}$, т.е. не выходят из компакта корректности $\mathcal{U}$.

Второе требование о корректности $\nabla J(u)$, как это видно из (3) и (4), распадается еще на три подзадачи:

1) обоснование корректности решения сопряженной задачи;

2) выявление области определения $V^*(\mathbf{\Omega})$, $\mathbf{\Omega} \subset \overline{\Omega}$ оператора $\mathbb{U}^*$ для отображения в область значений $U^*(S)$ градиента ∇J;

3) оценка корректности отображения линейного оператора $\mathbb{U}^*$: $V^*(\mathbf{\Omega}) \to U^*(S)$ и линейного неоднородного оператора $\mathbb{U}^*_{\varnothing}\cdot = \mathbb{U}^* \cdot + I'_u|_\omega$.

Начнем с сопряженной задачи. Ее тип и алгоритм решения определяется оператором $\mathbb{V}^*$. Она является прямой, линейной, ее дифференциальные уравнения будут того же типа и с теми же характеристиками, что и исходная задача, которая определяется оператором $\mathbb{D}$ (см. [6]). Поэтому анализ корректности решения сопряженной задачи не будет сложнее аналогичного анализа исходной задачи.

Согласно (3), возмущения сопряженного состояния f, порожденные производной I'_v, должны из области ω распространяться на такую часть $\mathbf{\Omega} \subset \overline{\Omega}$, где оператор $\mathbb{U}^*$ может реализовать требуемое отображение $V^*(\mathbf{\Omega}) \to U^*(S)$.

Определение 3. Множество сопряженных состояний $V^*(\mathbf{\Omega})$, $\mathbf{\Omega} \subset \overline{\Omega}$, корректно решенной сопряженной задачи, будем называть *областью определения* линейного оператора $\mathbb{U}^*$, если $f \in V^*(\mathbf{\Omega})$ однозначно зависит от управления u на S, и при этом *областью значений* оператора $\mathbb{U}^*$ будет множество градиентов $U^*(S)$, т.е.

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_v|_\omega I'_v|_\omega \xrightarrow{\text{однозначно}} f|_{\mathbf{\Omega}} \in V^*(\mathbf{\Omega}) \overset{\mathbb{U}^*}{\to} U^*(S).$$

При оптимизации распределенных систем почти всегда множество $\mathbf{\Omega}$ не совпадает со всей областью функционирования системы $\overline{\Omega}$. Множество $\mathbf{\Omega}$ существенным образом зависят от определения ω в целевом функционале J. Если ω задано не в том месте $\overline{\Omega}$, не тех размеров, то ожидать корректного действия оператора $\mathbb{U}^*$ не стоит. Кроме “правильного” ω необходимо учитывать вычислительные и другие помехи, включая возможную диссипацию в системе управления, которые могут нарушать однозначность указанных отображений.

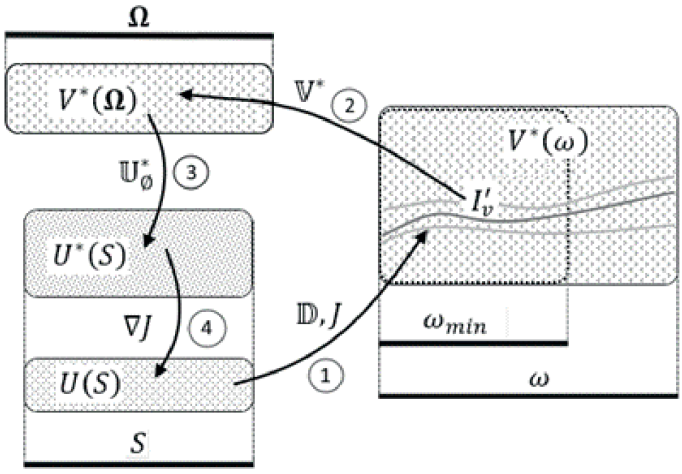
Рассмотрим корректность отображения оператора $\mathbb{U}^*$: $V^*(\mathbf{\Omega}) \to U^*(S)$. В задаче оптимизации без ограничений и при $I'_u = 0$ (функция цели не зависит явно от управления u), значение оператора $\mathbb{U}^* f = \nabla J$ должно обращаться в нуль (необходимое условие оптимальности) только при оптимальном управлении u_*. Это возможно, если линейный оператор $\mathbb{U}^*$ невырожденный, т.е. он имеет только нулевое ядро $\operatorname{Ker}\mathbb{U}^* \equiv f_{\operatorname{Ker}} = 0$. При оптимальном управлении с невырожденным $\mathbb{U}^*$ сопряженное состояние f должно быть нулевым. Наличие ненулевого ядра, когда $\mathbb{U}^*(f \neq 0) = 0$, не позволит корректно находить градиент, а следовательно, и оптимальное управление u_*.

С практической точки зрения важно не знание непосредственно множества особых решений $\{f\}_{\operatorname{Ker}}$ ненулевого ядра оператора $\mathbb{U}^*$, а знание конкретных условий, приводящих к появлению такого ядра. Возможно, придется изменить или функцию I, или множество ω так, чтобы наблюдаемое ядро исчезло (обнулилось). Только тогда можно говорить о корректном определении градиента целевого функционала.

Если целевая функция I зависит явно и от управления u, то соответствующий линейный оператор $\mathbb{U}^*_{\varnothing}\cdot = \mathbb{U}^* \cdot + I'_u|_\omega$ уже не будет однородным. Очевидно, что если $\mathbb{U}^*$ был невырожденным, то и оператор

$\mathbb{U}^*_\varnothing$ будет невырожденным. При этом произойдет всего лишь смещение ранее нулевого ядра $f_{\mathrm{Ker}} = 0$ к новому значению в новом оптимуме u_*, где градиент $\nabla J = \mathbb{U}^*_\varnothing f$ обратится в нуль. Такой оптимум не будет сопровождаться нулевым сопряженным состоянием f. Новое ядро $f_{\mathrm{Ker}} \neq 0$ неоднородного оператора $\mathbb{U}^*_\varnothing$ не делает оператор вырожденным.

На фиг. 1 символически представлены обсуждаемые отображения. Шаг 1 — это прямое отображение управления $u \in U(S)$ уравнениями $\mathbb{D}$ в состояние $v \in V^*(\overline{\Omega})$ и далее, с использованием J, в сопряженное пространство функций $I'_v \in V^*(\omega)$. Примеры таких функций изображены линиями I'_v на множестве ω.



Фиг. 1. Отображения пространств.

Шаг 2 представляет собой отображение функций I'_v (через состояние f) с помощью оператора $\mathbb{V}^*$ из множества ω в $\mathbf{\Omega}$. Заметим, что для однозначного отображения из ω во все точки $\mathbf{\Omega}$ может потребоваться не все множество ω, а лишь его часть $\omega_{\min} \subset \omega$. Эта ситуация соответствует *избыточности* ω. Попытка учета всего ω может приводить к неоправданным помехам в вычислении градиента и, соответственно, к помехам в управлении. Если же отображение I'_v из ω попадает на $\mathbf{\Omega}$ так, что область значений $\mathbb{U}^*$ оказывается определенной не на всем S, то это свидетельствует о *недостаточности* ω для формирования полноценной области определения $\mathbb{U}^*$ и, как следствие, невозможности управления на S.

Шаг 3 — это отображение пространства сопряженного состояния $V^*(\mathbf{\Omega})$ в пространство градиентов $U^*(S)$ с помощью оператора $\mathbb{U}^*_\varnothing$.

Шаг 4 — это отображение градиентов из $U^*(S)$ в множество управлений $U(S)$. Последний шаг реализуется экстремальными алгоритмами, и мы его уже обсудили.

Подытожим сказанное следующим утверждением для практического анализа управляемости в прямом экстремальном подходе.

Теорема. *Математическая модель* $\mathbb{D}(\tau, v, u)v = 0$, $\tau \in \overline{\Omega}$ *в задаче* (1) *управляема посредством* $u(\tau)$ *на* $S \subset \overline{\Omega}$ *по функционалу* J, *если*

- *существует область* $V^*(\mathbf{\Omega})$ *определения оператора* $\mathbb{U}^*$ *с его однозначными значениями в области* $U^*(S)$;
- *оператор* U^* *невырожденный;*
- *прямой экстремальный подход* (2) *при* $u^0 \in U$ *использует удовлетворительные параметры регуляризации в алгоритме* (5).

Получение конкретных *условий управляемости*, обеспечивающих выполнение требований теоремы, осуществляется в процессе анализа на корректность решения исходной и сопряженной задач, в процессе выявлении области определения $V^*(\mathbf{\Omega})$ оператора $\mathbb{U}^*$ и оценке его ядра.

Введенное понятие управляемости в виде определения 1 и следующая из него теорема позволяют сделать процедуру анализа управляемости в прямом экстремальном подходе оптимизации наглядной и универсальной.

3. ПРИМЕР АНАЛИЗА УПРАВЛЯЕМОСТИ

3.1. Постановка задачи

Наиболее наглядным примером может служить волновая система с гиперболическими уравнениями. Рассмотрим задачу идентификации оптимального распределенного коэффициента шероховатости (см.[3],[10], [11]) по экспериментальным наблюдениям за уровнем открытого потока воды в нижнем створе канала.

Нестационарное течение воды в открытом канале можно описать нелинейной квазиодномерной распределенной системой Сен—Венана:

$$\begin{aligned}&\mathbb{D}v = \frac{\partial v}{\partial t} + A\frac{\partial v}{\partial x} + F = 0, \quad \text{на } \Omega,\\ &Z(x_a, t) = Z_a(t) \text{ на } \Gamma_a = x_a \times (t_0, t_1),\\ &Q(x_b, t) = Q_b(t) \text{ на } \Gamma_b = x_b \times (t_0, t_1),\\ &v(x, t_0) = \big(Q_0(x), Z_0(x)\big) \text{ на } \Gamma_0 = [x_a, x_b] \times t_0,\end{aligned} \tag{6}$$

где матрица

$$A = \begin{pmatrix} 2w & B(c^2 - w^2) \\ 1/B & 0 \end{pmatrix},$$

вектор-столбец

$$F = \left(F_{\text{fr}} - w^2 \frac{\partial \sigma}{\partial x}\Big|_z, \quad -\frac{q}{B} \right),$$

где $\sigma(x, Z)$ — площадь живого сечения. Время $t \in [t_0, t_1]$, пространственная координата $x \in [x_a, x_b]$ размещена вдоль потока воды. Состояние системы — вектор-столбец $v = (Q, Z)$, где $Q(x, t)$ — расход воды в потоке от верхнего (левого) створа к правому (нижнему), $Z(x, t)$ — уровень свободной поверхности потока относительно заданного горизонта, $w = Q/\sigma$ — скорость потока, осредненная по живому поперечному сечению, $B(x, Z)$ — ширина потока поверху, $c = \sqrt{g\sigma/B}$ — скорость распространения малых возмущений в потоке, равная скорости звука в воде, g — ускорение свободного падения, $q(x, t)$ — боковой распределенный приток. Граничные области Γ_a, Γ_b, а также начальная Γ_0 и терминальная $\Gamma_1 = [x_a, x_b] \cdot t_1$ образуют границу $\partial\Omega$ прямоугольника Ω.

Член трения

$$F_{\text{fr}} = \frac{qQ|Q|}{\sigma R C^2}, \quad C = \frac{1}{u} R^{1/6},$$

где $R(x, Z) = \sigma/\chi$ — гидравлический радиус потока, $\chi(x, Z)$ — смоченный поперечный периметр русла, C — эмпирический коэффициент Шези.

Управлением является коэффициент шероховатости $u(x)$, $x \in (x_a, x_b) = S$, который присутствует в коэффициенте Шези и характеризует вязкое трение воды (диссипацию энергии) в пристеночной зоне русла. Можно сказать, что $S = P_x(\Omega)$, т.е. является проекцией области Ω на ось x.

Задача оптимизации формулируется следующим образом. Необходимо найти управление $u(x)$, доставляющее минимум функционалу

$$J(u) = \int_{t_0}^{t_1} I(v)|_\omega dt, \quad I(v; t)|_{\omega=\Gamma_b} = \Big(Z(\tau)|_{\Gamma_b} - Z_e(t) \Big)^2. \tag{7}$$

Функционал задается в нижнем створе канала на основе экспериментально наблюдаемого уровня воды $Z_e(t)$.

Сопряженная задача (3) формулируется относительно вектор-строки $f = (f_1, f_2)^T$ (транспонированный вектор-столбец) и принимает вид

$$\begin{aligned} &\mathbb{V}^* f = -\frac{\partial f}{\partial t} - A^T \frac{\partial f}{\partial x} + \mathbb{F}^* f = 0 \text{ на } \Omega, \\ &2wf_1 + \frac{1}{B} f_2 = 0 \text{ на } \Gamma_a, \\ &B(c^2 - w^2) f_1 + I' v|_{\omega} = 2(Z - Z_e) \text{ на } \Gamma_b, \\ &f_1 = f_2 = 0 \text{ на } \Gamma_1. \end{aligned} \tag{8}$$

Здесь

$$\mathbb{F}^* = \begin{pmatrix} \dfrac{2F_{\text{fr}}}{Q} & \dfrac{1}{B^2} \cdot \dfrac{\partial B}{\partial x} \\ -F_{\text{fr}} \left(\dfrac{B}{\sigma} + \dfrac{4}{3} \cdot \dfrac{1}{R} \cdot \dfrac{\partial B}{\partial Z} \right) + w^2 \dfrac{\partial B}{\partial x}\Big|_Z - g \dfrac{\partial \sigma}{\partial x} & \dfrac{1}{B^2} \left(q - \dfrac{\partial Q}{\partial x} \right) \dfrac{\partial B}{\partial Z} \end{pmatrix}.$$

Градиент целевого функционала $\nabla J(u; x) = \mathbb{U}^* f \in U^*(S)$. Здесь $\mathbb{U}^*_{\varnothing} = \mathbb{U}^*$, т.е. сопряженный оператор градиента — линейный, однородный:

$$\mathbb{U}^* \cdot = \int_{t_0}^{t_1} F_u'^T \cdot dt \ : \ V^*(\Omega) \to U^*(S),$$

где вектор-строка $F_u'^T = \left(\frac{\partial F_{\text{fr}}}{\partial u}, 0 \right)^T = \left(\frac{2F_{\text{fr}}}{u}, 0 \right)^T$. Оператор $\mathbb{U}^*$ реализует отображение из пространства сопряженных состояний, определенных на Ω, в пространство сопряженных управлений, определенных на S, т.е. в своем отображении он осуществляет и проецирование $P_x(\Omega)$, которое реализуется интегрированием по времени.

Таким образом, градиент

$$\nabla J(u; x) = \mathbb{U}^* f = \int_{t_0}^{t_1} \frac{2F_{\text{fr}}}{u} f_1 dt, \quad x \in (x_a, x_b). \tag{9}$$

При оптимальном управления $u = u_*$, сопряженное состояние f будет нулевым, поскольку $\mathbb{U}^*$ однородный.

3.2. Анализ управляемости (идентифицируемости)

Для анализа идентифицируемости коэффициента шероховатости воспользуемся сформулированной ранее теоремой. Сначала обоснуем корректность решения исходной и сопряженной задач. Далее уточним области определения и значений оператора $\mathbb{U}^*$, сделаем оценку его возможной вырожденности.

Исходная гиперболическая задача (6) имеет две характеристики $\xi_{1,2}(x, t)$, вдоль которых распространяются волны в русле со скоростью

$$\xi_{1,2} = \frac{d\xi_{1,2}}{dt} = w \pm c.$$

Для докритических течений волны на характеристиках первого семейства ξ_1 перемещаются вдоль потока к правому нижнему створу русла, а волны на характеристиках второго семейства ξ_2 — против течения к левому верхнему створу. Поскольку характеристики определяются собственными числами матрицы A, то сопряженная задача с матрицей A^T имеет те же характеристики $\xi_{1,2}$, что и исходная. Здесь сопряженные волны движутся по тем же характеристикам $\xi_{1,2}$, но в обратном направлении.

Исходная задача (6) и сопряженная (8) обеспечивают корректное решение для двумерных вектор-функций $v(\tau)$ и $f(\tau)$, поскольку в каждую точку $\tau \in \Omega$ попадают обе характеристики ξ_1, ξ_2, выходящие из известных краевых условий на $\partial\Omega$ (см. [12]).

К вопросу корректности следует сделать некоторые дополнения. В прямой задаче пространства существования состояния потока V и управления U определяются физически разумными, приемлемыми значениями управления. Например, при достаточно больших значениях u может быть получен уровень потока Z, идущий ниже линии дна Z_{bot}. Поэтому, ограничим U некоторым замкнутым множеством из условия

$$U = \{u : u(x) \geqslant 0 \ \ \forall x \in S, \ \ v \in V\}, \quad V = \{v : Z(\tau) > Z_{\text{bot}}(x) \ \ \forall \tau \in \overline{\Omega}\}. \tag{10}$$

Будем контролировать данное условие разумным выбором значений управления u. Приведенное условие не является ограничением для задачи идентификации и не требует его реализации в алгоритмах оптимизации.

Условие 10 означает, что прямая задача физически условно корректна. Ее решение следует искать при ограниченном изменении коэффициента шероховатости $u \in U$. Множество таких допустимых управлений u представляет собой компакт корректности $\mathcal{U}$ для решения обратной задачи идентификации

$$\mathcal{U} = U.$$

Все управления u^k, выходящие за пределы компакта $\mathcal{U}$, не позволят корректно найти ни градиент $\nabla \boldsymbol{J}^k$, ни дальнейшее управление u^{k+1}.

Из корректности исходной задачи следует, что изменение шероховатости $u(x)$ меняет уровень воды во всем канале. Поэтому, если волны, идущие вдоль ξ_1, от левого (верхнего) створа x_a доходят до правого (нижнего) створа x_b, то любые функции $u(x) \in \mathcal{U}$ будут однозначно влиять на значение производной целевой функции $I'_Z|_\omega = 2\left(Z|_{\Gamma_b} - Z_e\right)$. Другими словами имеет место левая ветка отображений в определении 3:

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_z|_\omega.$$

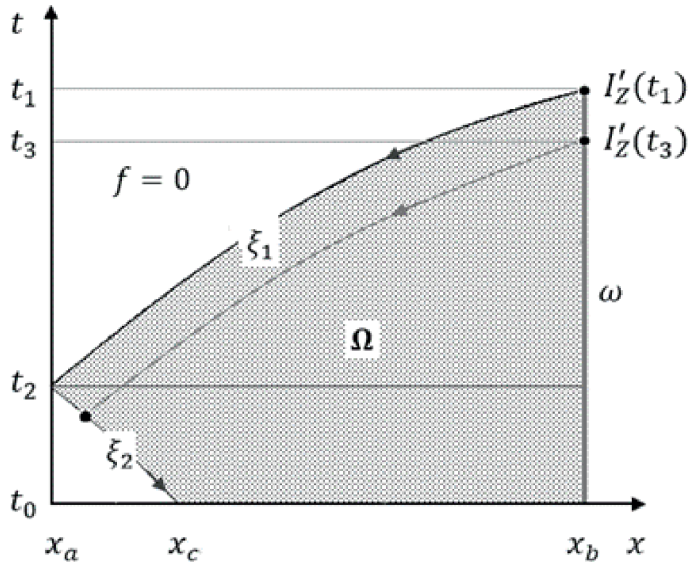
Производная $I'_z|_\omega$ является источником возмущений сопряженной задачи. Возмущения f при $t \to t_0$ распространяются в виде сопряженных волн по характеристикам первого семейства ξ_1 от правого створа x_b в сторону левого створа x_a. В левом (верхнем) створе сопряженные волны отражаются и переносятся обратно вправо по характеристикам второго семейства ξ_2. Распространение сопряженных волн и пример множества $\boldsymbol{\Omega}$ (все пространство под верхней характеристикой ξ_1) для области определения $V^*(\boldsymbol{\Omega})$ оператора $\mathbb{U}^*$ показан на фиг. 2.

Можно выделить время t_2 прихода верхней (замыкающей) характеристики ξ_1 к левому створу русла. Ниже этого времени в области $\boldsymbol{\Omega}$, под характеристикой ξ_2, начинает образовываться белый треугольник области неоднозначного влияния значений функции I'_Z на сопряженное состояние f. В каждую точку данной области будет приходить две сопряженные волны, вышедшие из разных точек ω с разными значениями I'_Z, как это показано в качестве примера для возмущений $I'_Z(t_1)$ и $I'_Z(t_3)$. При $t_0 < t_2$ применение оператора $\mathbb{U}^* f = \int_{t_0}^{t_1} F_u'^T f \, dt$ будет реализовываться в неоднозначной области определения $V^*(\boldsymbol{\Omega})$ оператора $\mathbb{U}^*$. Это может создать неприемлемые помехи в значениях $\mathbb{U}^* f$, т.е. в градиенте $\nabla \boldsymbol{J}$.

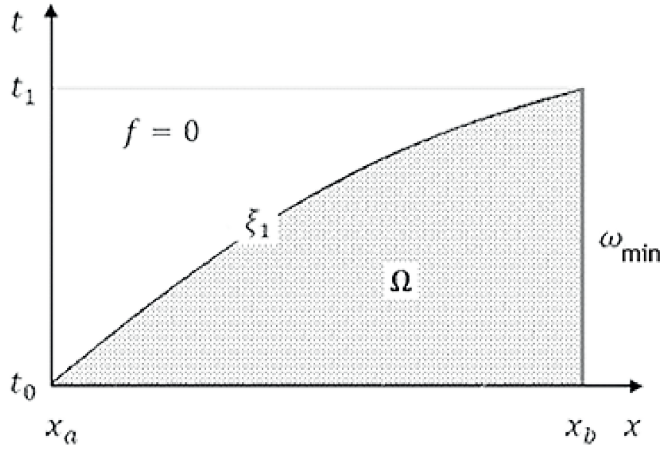
Чтобы такого не возникло, следует устранить избыточность области ω задания целевого функционала до минимально достаточной $\omega_{\min}$, как это показано на фиг. 3.

Если вычислительные и диссипативные помехи в исходной и сопряженной задачах не существенны, то однозначность области определения $\mathbb{U}^*$ обеспечивается при

$$t_1 = t_0 + \int_{x_a}^{x_b} \frac{dx}{\lambda_1}. \tag{11}$$



Фиг. 2. Пространственно-временная диаграмма области **Ω** с избыточной областью ω.



Фиг. 3. Пространственно-временная диаграмма области **Ω** с минимально достаточной областью $\omega_{\min}$.

Данное условие представляет собой условие *идентифицируемости* (*управляемости*).

Следует отметить, что наличие вычислительных помех и характерных для данной задачи диссипативных помех, может приводить в расчетах к появлению существенного затухания волн как исходной, так и сопряженной гиперболических систем. Затухание волн уровня воды Z в исходной задаче при движении слева направо будет искажать невязку $\left(Z|_{\Gamma_b} - Z_e\right)$, на основании которой решается сопряженная задача. Решение сопряженной задачи будет добавлять свое затухание волн f от источника I'_Z при движении обратно справа налево.

Это означает, что возле левого верхнего створа русла точность идентификации шероховатости может ухудшаться, а при очень длинном русле или сильно выраженной диффузорной неодномерности потока (с ростом уклона русла возрастает диссипация в потоке) вообще может стать невозможной. Поэтому требование (11), полученное из анализа области влияния ω, следует дополнить фразой “с помехами, не существенно искажающими волновые процессы в обеих задачах”.

В этих условиях выполняются оставшиеся ветки отображения в определении 3:

$$I'_Z|_{\omega} \xrightarrow{\text{однозначно}} f|_{\mathbf{\Omega}} \overset{\mathbb{U}^*}{\to} U^*(S).$$

Таким образом, область определения $V^*(\mathbf{\Omega})$ оператора $\mathbb{U}^*$ задается корректно на треугольнике **Ω** из фиг. 3.

Таким образом, первое условие теоремы “существует область $V^*(\mathbf{\Omega})$ определения оператора $\mathbb{U}^*$ с его однозначными значениями в области $U^*(S)$” выполняется.

Теперь оценим наличие возможной вырожденности оператора $\mathbb{U}^*$ в градиенте (9):

$$\nabla J = \mathbb{U}^* f = \int_{t_0}^{t_1} \frac{2F_{\text{fr}}}{u} f_1 dt \in U^*(S).$$

Необходимо выяснить, имеет ли оператор $\mathbb{U}^*$ при каких-либо условиях ядро $\operatorname{Ker}\mathbb{U}^* \neq 0$, т.е. может ли градиент ∇J при $f_1 \neq 0$, и соответственно при $u \neq u_*$, иметь нулевое значение на S или на значительной (ненулевой меры) части S.

Поскольку мы рассматриваем движущиеся потоки воды и только в одном направлении — вдоль русла к нижнему створу, то при $u \in \mathcal{U}$ член, характеризующий трение воды о ложе русла, всегда будет положительным: $F_{\text{fr}} = \frac{gQ|Q|}{\omega RC^2} > 0$ и $\frac{2F_{\text{fr}}}{u} > 0$. В то же время, в разные моменты времени волны от невязки $\left(Z|_{\Gamma_b} - Z_e\right)$, идущие вдоль характеристик первого семейства, могут быть как волнами повышения (f_1 положительное), так и волнами понижения (f_1 отрицательное). Как при всяком волновом движении суммарное по времени (интегральное в ∇J) значение таких волн (с положительным коэффициентом $\frac{2F_{\text{fr}}}{u}$ в некоторых точках $x \in S$ (только в точках, а не на всем S или на значительной части S) может обращать в нуль значение $\nabla J(u; x)$. Однако такие волны не являются стоячими, и при изменении управления от u^k на u^{k+1} ранее точечные нулевые значения градиента будут исчезать. Устойчивое нулевое значение интеграла по времени в градиенте $\nabla J(u; x)$ на всем S возможно только при оптимальном решении $u = u_*$, когда $\left(Z|_{\Gamma_b} - Z_e\right) = 0$, что соответствует $f_1 = 0$. Это означает, что оператор $\mathbb{U}^*$ имеет только нулевое ядро. Выполняется второе условие теоремы “оператор $\mathbb{U}^*$ невырожденный”.

Таким образом, согласно теореме, распределенная система (6) управляема (идентифицируема) посредством параметра $u(x)$ на $S = (x_a, x_b)$ по функционалу J (7) при условии (8) с помехами, не существенно искажающими волновые процессы в обеих задачах. Остается выбрать разумное начальное приближение $u^0 \in \mathcal{U}$ и корректно применить экстремальные алгоритмы (последнее требование теоремы).

3.3. Тестовое решение

Рассмотрим течение в тестовом канале со следующими характеристиками (характерны для Северо-Крымского канала): длина канала $[x_a, x_b] = 20$ км; трапецеидальная форма поперечного сечения с шириной по дну 30 м и откосом стенок 4 (котангенс угла наклона); ровное дно с уклоном $i = \frac{\partial Z_{\text{bot}}}{\partial x} = 0.0001$; боковой приток в тестовых расчетах отсутствует, т.е. $q = 0$.

Расчеты проводились при следующих начальных условиях: расход воды $Q_0(x) = 107$ м^3/с; уровень воды рассчитывался из условия установившегося течения при начальном значение глубины слева 3.6 м, при тестовом оптимальном управлении

$$u_*(x) = 0.025 + 0.0075 \cos\left(\pi + \frac{2\pi x}{x_b - x_a}\right).$$

Дальнейшие предварительные расчеты показали, что при управлении $u > 0.04$ уровень воды в конце русла опускается ниже линии дна, что недопустимо. Поэтому, исходя из условия (10) физической корректности решения исходной прямой задачи, будем считать компактом корректности, достаточным для решения рассматриваемой задачи оптимизации, следующее множество:

$$\mathcal{U} = [0.001, 0.04].$$

Перейдем к решению тестовой задачи оптимизации с нестационарным течением воды в канале. При оптимальной шероховатости u_* найдем уровень $Z(x_b, t)$ на правой границе русла, который далее будем считать экспериментально наблюдаемым $Z_e(t)$. Зададим начальное приближение

$$u^0(x) = 0.04 \in \mathscr{U}.$$

Теперь можно решать обратную задачу о восстановлении оптимального управления $u(x)$ по целевому функционалу (7).

Оптимизацию будем считать удовлетворительной и завершать итерационный процесс (5) при достижении точности

$$\Delta_{\max} Z = \max_t \left| Z(x_b, t) - Z_e(t) \right| \leqslant 0.01 \text{ м}.$$

Выбор шагового множителя b^k в алгоритме (5) осуществлялся по методу Носедала–Райта (см. [13]) с коэффициентами Вольфе $c_1 = 10^{-4}$ и $c_2 = 0.5$, максимальным шагом $0.1\|u^0 - u_*\|$. Если шаг увеличить, то можно ускорить сходимость к оптимуму, но изменения функции $u^k(x)$ могут быть слишком большими и выводить управление из компакта корректности $\mathbb{U}$. При этом где-либо вдоль русла могут наблюдаться $u^k < 0$ или уровень $Z^k < Z_{\text{bot}}$. Возможно даже появление сверхзвуковых течений при приближении уровня потока к линии дна, поскольку требование неразрывности потока в ситуации $Z^k \to Z_{\text{bot}}$ (сечение потока $\sigma \to 0$) значительно увеличивает расход воды Q^k, и в итоге оказывается скорость потока

$$w = \frac{Q}{\sigma} > c = \sqrt{g\sigma}B.$$

Параметр регулирования направления спуска в (5) задавался по формуле

$$\alpha(x) = \left| \frac{0.1u^0}{\nabla J(u^0; x)} \right|.$$

Описанный выбор b^k и α обеспечил удовлетворительные значения параметров регуляризации итерационных приближений управления, все $u^k \in \mathscr{U}$, $k = 1, 2, \ldots$.

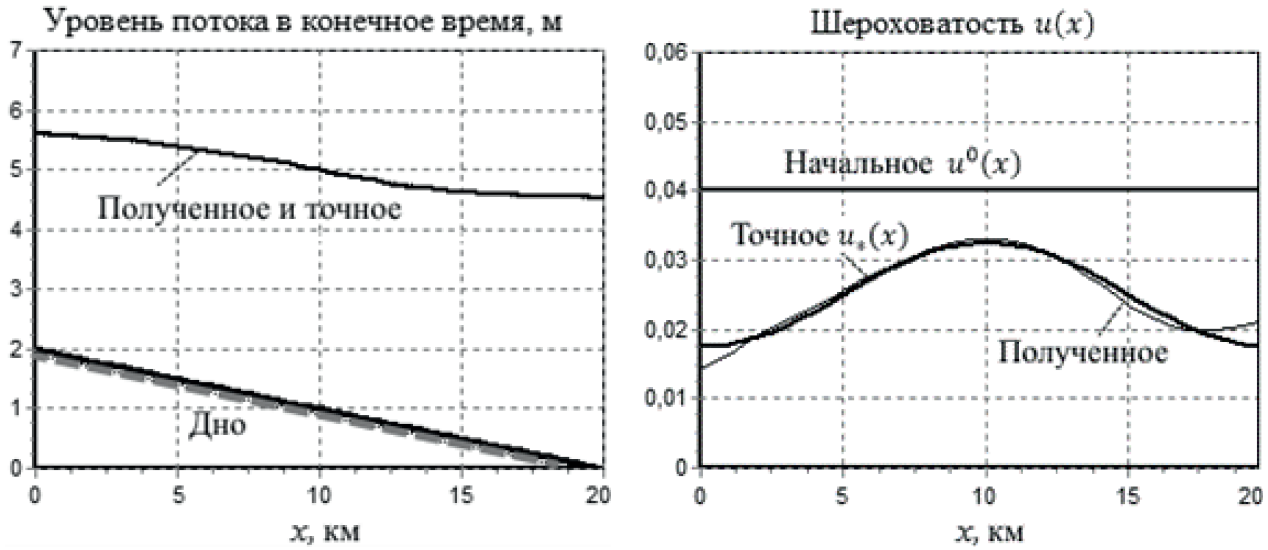
Общее время оптимизации $t_1 - t_0$ выбиралось с учетом условия управляемости (11), соответствующее времени прохождения фронта начальной волны вдоль всего русла слева направо. Для рассматриваемого потока время прохода фронта волны по характеристике ξ_1 составляет примерно 60 мин. Поскольку рассчитываемый фронт имеет некоторую протяженность (размазанность), то было выбрано множество $\omega = x_b \times (t_0 = 0, t_1 = 73 \text{ мин})$.

На фиг. 4 представлены результаты оптимизации функции $u(x)$. Заданная точность была достигнута за $k = 60$ итераций с близостью к тестовому оптимальному значению

$$\|u^{60} - u_*\|_{L_2(S)} = 0.18.$$

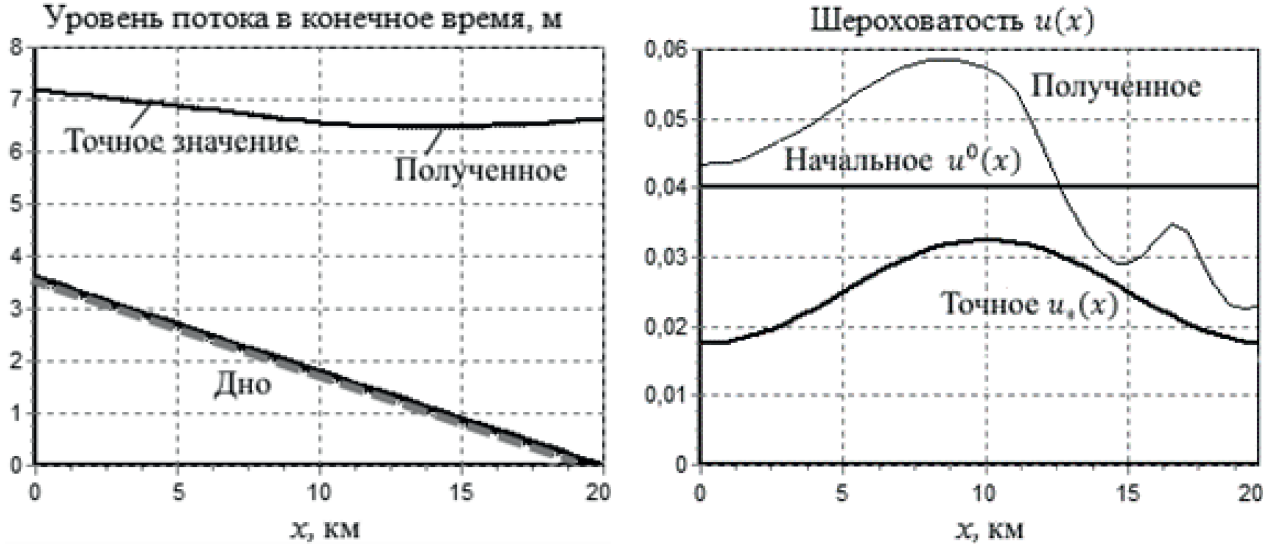
Это очень высокая точность. Заметим, что классический градиентный метод наискорейшего спуска ($\alpha = 1$) обеспечил близость последней итерации управления $\|u^{152} - u_*\|_{L_2(S)} = 0.45$.

Уменьшение времени в ω, например, до $t_1 = 30$ мин не позволило сделать ни одной разумной коррекции управления u^k. Увеличение же времени до $t_1 = 150$ мин заметно ухудшило качество восстановления оптимального управления, особенно в левой части канала, как это и предсказывалось при анализе управляемости. В обоих примерах множество ω существенно не соответствовало условию управляемости, что сделало область определения $V^*(\mathbf{\Omega})$ оператора $\mathbb{U}^*$ некорректной для однозначного отображения в область значений $U^*(S)$. В первом случае множество ω было недостаточным, во втором — избыточным (как на фиг. 2). Везде не было выполнено первое условие теоремы управляемости.



Фиг. 4. Результаты оптимизации.

Рассмотрим влияние диссипации на управляемость. На фиг. 5 показаны результаты оптимизации для русла с увеличенным уклоном дна $i = 0.00018$. На правой части рисунка видно, что искомая тестовая функция $u_*(x)$ не восстанавливается.



Фиг. 5. Результаты оптимизации с увеличенным уклоном дна.

На левой части рисунка видно, что шероховатость $u(x)$ практически не влияет на состояние потока, точное и полученное (при неправильном u) значения уровня воды мало различимы. Появившаяся нечувствительность модели (6) к параметру $u(x)$ объясняется тем, что диффузор потока (особенно в правой половине течения) существенно повышает диссипацию энергии волн. Площадь σ поперечного сечения потока в нижнем створе русла почти в 3 раза больше, чем в верхнем. При этом значение члена трения $F_{\text{fr}} = \frac{gQ|Q|}{\sigma RC^2}$ в нижнем створе канала оказывается в 5 раз выше, чем в верхнем. По сути, от волн ничего не остается.

В последнем примере, согласно определению 3, мы полностью потеряли область определения $V^*(\mathbf{\Omega})$ оператора $\mathbb{U}^*$. Наличие помех в виде существенной диссипации уничтожало зависимость целевого функционала от управления, т.е. отображение $u \nrightarrow I_v'|_\omega$, отображение исчезло. Не было выполнено первое условие теоремы управляемости.

ВЫВОДЫ

Мы получили относительно простую и универсальную процедуру контроля управляемости как возможности управлять системой с распределенными параметрами по заданному целевому функционалу. Новое понятие управляемости в виде определения 1 и теорема управляемости позволяют в прямом экстремальном подходе сделать процедуру анализа управляемости наглядной и универсальной. Она состоит из выявления области определения линейного сопряженного оператора в градиенте целевого функционала, оценке его невырожденности и выборе удовлетворительных параметров регуляризации для приближений управления к оптимуму. Все это позволило корректно и с высокой точностью решить задачу оптимизации нелинейной гиперболической системы относительно распределенного коэффициента трения по функционалу, заданному на границе русла.

СПИСОК ЛИТЕРАТУРЫ

1. *Егоров А. И., Знаменская Л. Н.* Введение в теорию управления системами с распределенными параметрами. СПб: Лань, 2017.
2. *Glowinski R., Lions J.-L., He Jiwen.* Exact and approximate controllability for distributed parameter systems. Cambridge Univer. Press, 2008.
3. *Толстых В. К.* Прямой экстремальный подход для оптимизации систем с распределенными параметрами. Донецк: Юго-Восток, 1997.
4. *Васильев Ф. П.* Методы оптимизации. Т. 2. М.: МЦНМО, 2011.
5. *Сеа Ж.* Оптимизация. Теория и алгоритмы. М.: Мир, 1973.
6. *Tolstykh V. K.* Optimization for systems governed by partial differential equations // Adv. Model. Optim. 2012. V. 14. N 3. P. 703–716.
7. *Толстых В. К.* О применении градиентного метода к задачам оптимизации систем с распределенными параметрами // Ж. вычисл. матем. и матем. физ. 1986. № 1. С. 137–140.
8. *Толстых В. К.* Идентифицируемость систем с распределенными параметрами // Автомат. и телемех. 1989. № 10. С. 49–56.
9. *Тихонов А. Н.* О решении некорректно поставленных задач и методе регуляризации // ДАН СССР. 1963. Т. 151. № 3. С. 501–504.
10. *Воронин С. Т., Толстых В. К.* Вариационный метод опредления коэффициента шероховатости открытого русла // Тр. Гидрометцентра СССР. 1986. № 283. С. 54–59.
11. *Atanov G. A, Tolstykh V. K.* Optimization problems for nonstationary wave processes // J. Math. Sci. 1995. N 77. C. 3540–3542.
12. *Рождественский Я. Н., Яненко Н. Н.* Системы квазилинейных уравнений и их приложения к газовой динамике. М.: Наука, 1979.
13. *Nocedal J., Wright S.* Numerical optimization. New York: Springer-Verlag, 1999.

ON THE CONTROLLABILITY OF DISTRIBUTED PARAMETER SYSTEMS

V. K. Tolstykh*

Donetsk State University, Universitetskaya st., 24, Donetsk, 283001 Russia
**e-mail: mail@tolstykh.com*

Received 03 September, 2023
Revised 14 February, 2024
Accepted 06 March, 2024

Abstract. The problem of controllability for optimal control problems, optimization of systems with distributed parameters in partial derivatives is considered. The concept of controllability as correctness according to A. N. Tikhonov for solving optimization problems is introduced. A theorem with controllability conditions for direct solution (direct minimization of the objective functional) of optimization problems by extremal algorithms is given. A test example of numerical solution of the optimization problem for a nonlinear hyperbolic system describing non-stationary water flow in an open channel is considered. Controllability analysis is demonstrated, which ensures correctness of the solution of the problem and high accuracy of optimization of the distributed friction coefficient in the flow equations.

Keywords: controllability, distributed parameter system, optimization, optimal control, identification, gradient, extreme methods.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 973–991

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 517.925.54

ОБ АППРОКСИМАЦИИ ПЕРВОГО СОБСТВЕННОГО ЗНАЧЕНИЯ НЕКОТОРЫХ КРАЕВЫХ ЗАДАЧ

© 2024 г. М. Ю. Ватолкин[1,*]

[1] *426069 Ижевск, ул. Студенческая, 7, Ижевский государственный технический университет им. М. Т. Калашникова, Россия*

[*] *e-mail: vmyu6886@gmail.com*

Поступила в редакцию 18.12.2023 г.
Переработанный вариант 28.02.2024 г.
Принята к публикации 05.03.2024 г.

Исследуется на предмет представления собственных функций в виде скалярных рядов двухточечная краевая задача типа $(n-1,1)$ в предположении, что существует функционал $\widetilde{\ell}$ сосредоточенный в одной точке, такой, что первые $n-1$ из исходных краевых условий и $\widetilde{\ell}\,x=1$ превращаются в условия Коши в этой точке. Собственная функция рассматриваемой краевой задачи, отвечающая собственному значению λ_*, представлена в виде ряда по степеням λ_*. Рассматривается уравнение $\Phi(\lambda)=0$, где $\Phi(\lambda)$ — сумма ряда по степеням λ, для нахождения собственных значений исходной задачи. Приведены примеры вычисления первого собственного значения некоторых краевых задач. Получены различные оценки для коэффициентов таких степенных рядов. Определяется некоторая функция двух переменных t и λ, для нее получено уравнение в частных производных и получены условия, которым она удовлетворяет. Нули "сечения" этой функции совпадают с собственными значениями исходной краевой задачи, что может быть использовано для их приближенного вычисления. Библ. 36. Табл. 1.

Ключевые слова: краевые задачи на собственные значения, собственные функции, собственные значения, функция Коши, представление собственных функций в виде сумм степенных рядов, корни уравнения, оценки для коэффициентов степенных рядов.

DOI: 10.31857/S0044466924060075, **EDN:** XYPKQO

1. ВВЕДЕНИЕ И ОБЗОР ЛИТЕРАТУРЫ

В современной спектральной теории дифференциальных операторов актуальной была и остается задача исследования свойств собственных значений и собственных функций дифференциального оператора в зависимости от гладкости коэффициентов дифференциального выражения, порождающего такой оператор. Эти вопросы достаточно полно и хорошо изложены в работе [1] и в известных монографиях (см. [2]–[6] и ссылки в этих работах). Различные свойства собственных функций и собственных значений задачи Штурма–Лиувилля с гладкими и негладкими потенциалами являются предметом исследований ведущих научных школ, занимающихся спектральной теории дифференциальных операторов уже много десятилетий. Круг этих задач на данный момент времени достаточно хорошо изучен. Авторами [10] рассмотрен оператор Штурма–Лиувилля с суммируемым потенциалом и найдены формулы для асимптотики собственных значений и собственных функций, то есть в [10] для фиксированного суммируемого потенциала получены асимптотические формулы для собственных функций и собственных значений классической задачи Штурма–Лиувилля с помощью современной трактовки метода Лиувилля–Стеклова.

В работе [7] рассматривается класс операторов Штурма–Лиувилля, порожденных симметрическими (формально самосопряженными) квазидифференциальными выражениями второго по-

рядка с локально интегрируемыми коэффициентами. Спектральная теория квазидифференциальных операторов второго порядка применяется к изучению операторов типа Штурма—Лиувилля с коэффициентами-распределениями. Основной целью работы является построение теории Титчмарша—Вейля для таких операторов. При этом вопрос о дефектных числах оператора и об условиях на коэффициенты, обеспечивающих реализацию случая предельной точки или предельного круга Вейля, — является центральным вопросом работы [7]. Изучению асимптотики решений линейных дифференциальных уравнений при различных предположениях относительно их коэффициентов и корректному определению дифференциальных уравнений в случае, когда производные понимаются в смысле теории распределений, посвящены работы [8] и [9].

Работы [11]–[13] посвящены изучению асимптотики собственных функций и собственных значений оператора Штурма—Лиувилля с сингулярным потенциалом, являющимся обобщенной функцией первого порядка, в этих работах рассмотрены операторы второго порядка с негладкими потенциалами (типа дельта-функции или потенциалами-распределениями). В работах [14], [15] на основе методики работ [11]–[13] исследовано асимптотическое поведение собственных значений операторов с потенциалом, являющимся дельта-функцией Дирака либо импульсными потенциалами (потенциалами-распределениями). В работах [16]–[18] строится аналог осцилляционной теории Штурма распределения нулей собственных функций на пространственной сети и графах. Изучению краевых задач для дифференциальных уравнений высоких порядков посвящены работы [19]–[23], в этих работах найдена асимптотика решений при больших значениях спектрального параметра при условии суммируемости потенциала.

В настоящей статье рассматриваются и изучаются краевые задачи на собственные значения для квазидифференциальных уравнений. В связи с этим приведем здесь определение квазидифференциального уравнения.

Пусть интервал $I \subseteq \mathbb{R}$ есть открытый интервал, $\mathscr{P} = (p_{ik})_0^n$ — нижняя треугольная матрица, где функции $p_{ik}(\cdot) : I \to \mathbb{R}$, такая, что функции $p_{00}(\cdot)$ и $p_{nn}(\cdot)$ измеримы, почти всюду конечны и почти всюду отличны от нуля, а функции $\dfrac{1}{p_{ii}(\cdot)}$ $(i \in 1 : n-1)$, $\dfrac{p_{ik}(\cdot)}{p_{ii}(\cdot)}$ $(i \in 1 : n, k \in 0 : i-1)$ локально суммируемы в I. Определим квазипроизводные ${}^{k}_{\mathscr{P}}x(\cdot)\,(k \in 0 : n)$ функции $x : I \to \mathbb{R}$ равенствами (см. [24], [25])

$$ {}^{0}_{\mathscr{P}}x \doteq p_{00}x, \quad {}^{k}_{\mathscr{P}}x \doteq p_{kk}\frac{d}{dt}\left({}^{k-1}_{\mathscr{P}}x\right) + \sum_{\nu=0}^{k-1} p_{k\nu}({}^{\nu}_{P}x)\,(k \in 1 : n). $$

Линейным однородным квазидифференциальным называется уравнение

$$ ({}^{n}_{\mathscr{P}})x(t) = 0, \quad t \in I. \tag{1} $$

Его решением называется всякая функция $x(\cdot) : I \to \mathbb{R}$, имеющая локально абсолютно непрерывные квазипроизводные ${}^{k}_{\mathscr{P}}x(\cdot)$ $(k \in 0 : n-1)$ и почти всюду в I удовлетворяющая этому уравнению (см. [24], [25]).

Линейным неоднородным квазидифференциальным называется уравнение

$$ (\ell x)(t) \doteq \left({}^{n}_{\mathscr{P}}x\right)(t) = f(t), \quad t \in I \quad (f : I \to \mathbb{R}). \tag{2} $$

Решением уравнения (2) называется всякая функция $x(\cdot)$, имеющая локально абсолютно непрерывные квазипроизводные до порядка $n-1$ включительно и удовлетворяющая ему почти всюду в I [24], [25]. Если функции $p_{00}(\cdot)$ и $p_{nn}(\cdot)$ измеримы, почти всюду конечны и почти всюду отличны от нуля, а функции

$$ \frac{1}{p_{\nu\nu}(\cdot)}\,(\nu \in 1 : n-1), \quad \frac{p_{\nu k}(\cdot)}{p_{\nu\nu}(\cdot)}\,(\nu \in 1 : n,\ k \in 0 : \nu-1), \quad \frac{f}{p_{nn}(\cdot)} $$

локально суммируемы в I, то задача Коши для неоднородного уравнения при начальных условиях $\left({}^{k}_{\mathscr{P}}x\right)(a) = \gamma_k \ \left(k \in 0 : n-1,\ a \in I,\ \gamma_k \in \mathbb{R}\right)$ эквивалентна задаче

$$\dot{z} = A(t)z + \tilde{f}(t), \quad z(a) = \gamma,$$

где $z(t) \doteq \left(\alpha_{\mathscr{P}}x\right)(t) \doteq \left({}^{0}_{\mathscr{P}}x(t),\ \dots,\ {}^{n-1}_{\mathscr{P}}x(t)\right)^{\mathrm{T}}$, $\tilde{f} \doteq \left(0,\ \dots,\ 0,\ \dfrac{f}{p_{nn}}\right)^{\mathrm{T}}$,

$$A = \begin{pmatrix} \dfrac{-p_{10}}{P_{11}} & \dfrac{1}{P_{11}} & 0 & \cdots & 0 \\ \dfrac{-p_{20}}{P_{22}} & \dfrac{-p_{21}}{P_{22}} & \dfrac{1}{P_{22}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \dfrac{-p_{n-1,0}}{P_{n-1,n-1}} & \dfrac{-p_{n-1,1}}{P_{n-1,n-1}} & \dfrac{-p_{n-1,2}}{P_{n-1,n-1}} & \cdots & \dfrac{1}{P_{n-1,n-1}} \\ \dfrac{-p_{n0}}{P_{nn}} & \dfrac{-p_{n1}}{P_{nn}} & \dfrac{-p_{n2}}{P_{nn}} & \cdots & \dfrac{-p_{n,n-1}}{P_{nn}} \end{pmatrix}$$

$\gamma \doteq \left(\gamma_0,\ \dots,\gamma_{n-1}\right)^{\mathrm{T}}$, т — знак транспонирования.

Последняя задача однозначно разрешима, а с нею и исходная задача Коши для уравнения (2) имеет единственное решение, компоненты которого, квазипроизводные ${}^{k}_{\mathscr{P}}x(\cdot)$ $(k \in 0 : n-1)$, локально абсолютно непрерывны в I (см. [24], [25]). Приведем один пример (см. [24], [25]). Пусть $n = 2$, $f(t) \equiv 0$, и в матрице $\mathscr{P}$ положим

$$p_{00}(t) = p_{11}(t) = p_{22}(t) \doteq \sqrt[3]{t}, \quad p_{10}(t) \doteq t^{-3/5}, \quad p_{21}(t) = p_{20}(t) \doteq 0, \quad I = (-c,\ c),$$

где $0 < c < +\infty$. Тогда решением уравнения $\left({}^{2}_{\mathscr{P}}x\right)(t) = 0$, $t \in I$, удовлетворяющим начальным условиям $\left({}^{0}_{\mathscr{P}}x\right)(0) = 1$, $\left({}^{1}_{\mathscr{P}}x\right)(0) = 0$, является функция

$$x(t) = t^{-1/3} \exp\left(-15\sqrt[15]{t}\right).$$

Найдем нулевую и первую квазипроизводные этой функции

$${}^{0}_{\mathscr{P}}x(t) = \exp\left(-15\sqrt[15]{t}\right), \quad {}^{1}_{\mathscr{P}}x(t) = t^{1/3}\left(\exp\left(-15\sqrt[15]{t}\right)\right)' + t^{-3/5}\exp\left(-15\sqrt[15]{t}\right).$$

Каждое из слагаемых в ${}^{1}_{\mathscr{P}}x(t)$ при $t = 0$ разрывно, а их сумма ${}^{1}_{\mathscr{P}}x(t) \equiv 0$, $t \in I$, является абсолютно непрерывной функцией.

Для квазидифференциального уравнения справедливы основные утверждения общей теории обыкновенного дифференциального уравнения (см. [24], [25]).

Так уравнение (1) имеет фундаментальную систему решений $\left\{u_\nu(\cdot)\right\}_0^{n-1}$, для которой определитель $W_{\mathscr{P}} \doteq \det\left({}^{\nu}_{\mathscr{P}}u_k\right)_0^{n-1}$ не обращается в нуль ни в одной точке из I. Частное решение уравнения (2), удовлетворяющее нулевым начальным условиям $\left(\alpha_{\mathscr{P}}u_*\right)(a) = 0$, имеет вид

$$u_*(t) = \int_a^t C(t,s)\left(f(s)/p_{nn}(s)\right)ds \quad (t \in I),$$

где

$$C(i,s)=\begin{vmatrix} {}^{0}_{\mathscr{P}}u_0(s) & \cdots & {}^{0}_{\mathscr{P}}u_0(s) \\ \cdots & \cdots\cdots & \\ \left({}^{n-2}_{\mathscr{P}}u_0\right)(s) & \cdots & \left({}^{n-2}_{\mathscr{P}}u_{n-1}\right)(s) \\ u_0(t) & \cdots & u_{n-1}(t) \end{vmatrix} \Big/ W_{\mathscr{P}}(s)$$

есть функция Коши уравнения (1), она по аргументу t является его решением и удовлетворяет начальным условиям

$${}^{k}_{\mathscr{P}}C(t,s)\big|_{t=s}=0 \quad (k\in 0:n-2),\ {}^{n-1}_{\mathscr{P}}C(t,s)\big|_{t=s}=1.$$

Общее решение уравнения (2) дается формулой $u=c_0u_0+\ \dots\ +c_{n-1}u_{n-1}+u_*$, где c_ν — произвольные постоянные.

Определитель $W_{\mathscr{P}}$ локально абсолютно непрерывен, он удовлетворяет уравнению

$$dW_{\mathscr{P}}\big/dt=-\sum_{\nu=0}^{n-1}\left(p_{\nu+1,\nu}(t)\big/p_{\nu+1,\nu+1}(t)\right)W_{\mathscr{P}}.$$

Имеет место аналог формулы Остроградского—Лиувилля

$$W_{\mathscr{P}}(t)=W_{\mathscr{P}}(a)\exp\int_a^t\left(-\sum_{\nu=0}^{n-1}\left(p_{\nu+1,\nu}(\tau)\big/p_{\nu+1,\nu+1}(\tau)\right)\right)d\tau.$$

Уравнение (2) обладает формально сопряженным в смысле Лагранжа уравнением (см. [24], [25])

$$\left(\ell^{+}y\right)(t)\doteq(-1)^n\left({}^{n}_{\mathscr{R}}y\right)(t)=g(t),\quad t\in I\quad (g:I\to\mathbb{R}), \tag{3}$$

где $\mathscr{R}=\left(r_{\nu k}\right)_0^n$ — нижняя треугольная матрица,

$$r_{\nu k}=(-1)^{\nu+k}\ p_{n-k,n-\nu}\,p_{n-\nu,n-\nu}\big/p_{n-k,n-k}\quad (k\in 0:\nu,\ \nu\in 0:n).$$

Это означает, что имеет место тождество Лагранжа: почти для всех $t\in I$

$$y(t)\,(\ell x)(t)-x(t)\left(\ell^{+}y\right)(t)\equiv\frac{d}{dt}\,[x,y](t),$$

где $[x,y](t)=\sum\limits_{k=1}^{n}(-1)^{n-k}\ \left({}^{k-1}_{\mathscr{P}}x\right)(t)\ \left({}^{n-k}_{\mathscr{R}}y\right)(t)$ (для всех $x(\cdot)$ и $y(\cdot)$, имеющих абсолютно непрерывные квазипроизводные до порядка $n-1$ включительно).

Пусть $C^*(t,s)$ — функция Коши сопряженного однородного уравнения. Имеет место соотношение ${}^{0}_{\mathscr{P}}C(t,s)=(-1)^{n-1}\ {}^{0}_{\mathscr{R}}C^*(s,t)$ (подробнее см. [24], [25]).

Квазидифференциальное уравнение является обобщением обыкновенного дифференциального уравнения. По-видимому, начало систематическому изучению неоднородного уравнения n-го порядка (с комплекснозначными коэффициентами) было положено работами Д. Ю. Шина (см. [26], [27]). Обстоятельства сложились таким образом, что его работы на долгое время были игнорированы и забыты, и только начиная с 1975 г. стали появляться работы A. Zettl, W. N. Everitt и их соавторов, посвященные этой тематике. Они же возродили интерес к работам Д. Ю. Шина.

В настоящее время имеется достаточно большое количество работ A. Zettl, W. N. Everitt и их учеников и соавторов по этой тематике, опубликованных в различных математических журналах, а также — опубликованных относительно недавно (см., например, работы [28]–[35], в этих работах рассматриваются и изучаются квазидифференциальные уравнения и различные краевые задачи для них).

Неоднородное квазидифференциальное уравнение позволяет с единой точки зрения рассматривать различные уравнения, которые принято называть “обобщенными”, уравнениями с особенно-

стями в коэффициентах и т.п. Обыкновенное дифференциальное уравнение с локально суммируемыми коэффициентами и его формально сопряженное в смысле Лагранжа уравнение также представляют собой частные случаи квазидифференциального уравнения.

В монографии [36] рассматривается самосопряженное квазидифференциальное уравнение четного порядка

$$(-1)^m\big(p_0(t)x^{(m)}(t)\big)^{(m)} + (-1)^{m-1}\big(p_1(t)x^{(m-1)}(t)\big)^{(m-1)} + \ldots + p_m(t)x(t) = f(t).$$

Оно получается из уравнения (2) при

$$p_{kk} = 1\ (k \in 0 : m-1), \quad p_{kk} = -1\ (k \in m+1 : n), \quad p_{i,n-i} = p_{i-m}\ (i \in m : n),$$

$$p_{ik} = 0\ (i \in 1 : n,\ k < i;\ k \neq n-i).$$

Уравнение (2), таким образом, общее и не является, вообще говоря, самосопряженным. Поэтому общепринятые простые и лаконичные обозначения квазипроизводных (см. [36]) заменены в нашем случае на более сложные, так как при переходе к сопряженному уравнению, матрица, с помощью которой строятся квазипроизводные, меняется. Но также, как и в [36], здесь рассматривается только случай вещественнозначных коэффициентов. Случай комплекснозначных коэффициентов, в отличии от работ [26] и [27] не рассматривается. Заметим, что квазидифференциальное уравнение интегрируется в квадратурах в случае, если коэффициенты $p_{\nu k} = 0\ (k \in 0 : \nu-2,\ \nu \in 2 : n)$ (см. [24], [25]).

2. О ПРЕДСТАВЛЕНИИ СОБСТВЕННЫХ ФУНКЦИЙ ДВУХТОЧЕЧНОЙ КВАЗИДИФФЕРЕНЦИАЛЬНОЙ КРАЕВОЙ ЗАДАЧИ ТИПА $(n-1, 1)$

Рассмотрим краевую задачу на собственные значения

$$\big(\overset{n}{\mathscr{P}}x\big)(t) = -\lambda\, p_{nn}(t)g(t)\big(\overset{0}{\mathscr{P}}x\big)(t) \quad (t \in J \doteq [a,b] \subset I), \tag{4}$$

$$\ell_\nu x = 0 \quad (\nu \in 1\ :\ n) \tag{5}$$

с функцией $g : J \to \mathbb{R}$ и функционалами

$$\ell_\nu x \doteq \sum_{k=0}^{n-1} \beta_{\nu k}\, \overset{k}{\mathscr{P}}x(a) \quad (\nu \in 1 : n-1), \quad \ell_n x \doteq \sum_{k=0}^{n-1} \beta_{nk}\, \overset{k}{\mathscr{P}}x(b) \tag{6}$$

в предположении, что существует функционал $\widetilde{\ell}$, сосредоточенный в точке a, такой, что краевые условия

$$\ell_\nu x = 0 \quad (\nu = 1, \ldots, n-1), \quad \widetilde{\ell}\, x = 1$$

превращаются в условия Коши в точке a.

Например, если $\det\ \big(\beta_{\nu,k-1}\big)_1^{n-1} \neq 0$, то можно положить

$$\widetilde{\ell}\, x \doteq \overset{n-1}{\mathscr{P}}x(a).$$

При $n = 2$ и таких $\ell_\nu x$ задача (4), (5) есть классическая задача Штурма–Лиувилля, при этом функционал $\widetilde{\ell}\, x$ можно выбрать следующим образом:

$$\widetilde{\ell}\, x \doteq \overset{1}{\mathscr{P}}x(a), \ \text{ если } \beta_{10} \neq 0, \quad \widetilde{\ell}\, x \doteq \overset{0}{\mathscr{P}}x(a), \ \text{ если } \beta_{10} = 0.$$

При произвольном n и функционалах вида (6) при

$$\beta_{\nu+1,k} = \delta_{\nu k}\ (\nu, k \in 0 : n-2), \ \ \beta_{nk} = \delta_{0k}\ (k \in 0 : n-1)$$

задача (4), (5) — двухточечная задача Валле Пуссена типа $(n-1, 1)$, при этом

$$\widetilde{\ell}\, x \doteq {}^{n-1}_{\mathscr{P}}x(a).$$

За начальное приближение принимаем решение задачи Коши (7), (8) для однородного уравнения (7). Далее рекуррентно находим решения задач Коши (9), (10) для неоднородных уравнений (9), в которых правая часть уравнений является известной функцией — решением предыдущей задачи Коши. Рассмотрим все это более подробно.

Следующим образом построим последовательность решений $\{x_k\}_0^\infty$: функция $x_0(\cdot)$ есть решение задачи Коши

$$\left({}^{n}_{\mathscr{P}}x\right)(t) = 0 \quad (t \in J), \tag{7}$$

$$\ell_\nu x = 0 \ (\nu \in 1 : n-1), \quad \widetilde{\ell}\, x = 1; \tag{8}$$

функции $x_k(\cdot)$ находятся рекуррентно как решения задач

$$\left({}^{n}_{\mathscr{P}}x_k\right)(t) = p_{nn}(t)g(t)\left({}^{0}_{\mathscr{P}}x_{k-1}\right)(t) \quad (t \in J), \tag{9}$$

$$\ell_\nu x_k = 0 \ (\nu \in 1 : n-1), \quad \widetilde{\ell}\, x_k = 0 \quad (k = 1, 2, \dots). \tag{10}$$

Теорема 1. *Пусть функции*

$$g(t),\ 1/p_{\nu\nu}(t) \ (\nu \in 1 : n-1),\ p_{\nu k}(t)/p_{\nu\nu}(t) \ (\nu \in 1 : n,\ k \in 0 : \nu - 1)$$

ограничены в существенном на J. Тогда справедливы следующие утверждения:

1) *ряд*

$$\sum_{k=0}^{\infty} (-\lambda)^k \ell_n x_k \tag{11}$$

сходится абсолютно и равномерно относительно λ на отрезке $[-\lambda,\ \lambda]$, при любом $\lambda > 0$;

2) *собственные значения задачи* (4), (5) *представляют собой корни уравнения $\Phi(\lambda) = 0$, где $\Phi(\cdot)$ — сумма ряда* (11)*;*

3) *функция $u(t, \lambda_*)(t) = \sum\limits_{k=0}^{\infty} \left(-\lambda_*\right)^k x_k(t)$ $(t \in J)$ есть собственная функция задачи* (4), (5)*, отвечающая собственному значению λ_*.*

Доказательство. При фиксированном λ рассмотрим ряд

$$\sum_{k=0}^{\infty} (-\lambda)^k \, {}^{0}_{\mathscr{P}}x_k(t),\ t \in J. \tag{12}$$

Из (9), (10) получаем следующее представление (с учетом того, что $x_0(t)$ есть решение задачи (7), (8))

${}^{0}_{\mathscr{P}}x_k(t) = \int\limits_a^t {}^{0}_{\mathscr{P}}C(t,s)\, g(s) {}^{0}_{\mathscr{P}}x_{k-1}(s)\, ds \quad (k = 1, 2, \dots)$, где, напомним, $C(t,s)$ — функция Коши уравнения (1).

Из определений квазипроизводных следует, что

$$\left({}^{0}_{\mathscr{P}}x_0(t)\right)' = \frac{{}^{1}_{\mathscr{P}}x_0(t)}{p_{11}(t)} - \frac{p_{10}(t)}{p_{11}(t)} {}^{0}_{\mathscr{P}}x_0(t) = \frac{1}{p_{11}(t)} \int\limits_a^t {}^{1}_{\mathscr{P}}C(t,s)\, g(s) {}^{0}_{\mathscr{P}}x_{k-1}(s) ds -$$

$$- \frac{p_{10}(t)}{p_{11}(t)} \int\limits_a^t {}^{0}_{\mathscr{P}}C(t,s)\, g(s) {}^{0}_{\mathscr{P}}x_{k-1}(s) ds \quad (k \in \mathbb{N}).$$

Обозначим

$$M \doteq \operatorname*{vraisup}_{[a,b]\times[a,b]} \left|{}_{\mathscr{P}}^{0}C(t,s)\,g(s)\right|;$$

$$C \doteq \max_{[a,b]} \left|{}_{\mathscr{P}}^{0}x_0(t)\right|, \qquad C_1 \doteq \max_{j\in 0:n-1}\ \max_{[a,b]} \left|{}_{\mathscr{P}}^{j}x_0(t)\right|;$$

$$L \doteq \max\left\{\operatorname*{vraisup}_{t\in[a,b]} \frac{1}{p_{\nu\nu}(t)}\,(\nu\in 1:n-1), \quad \operatorname*{vraisup}_{t\in[a,b]} \frac{\left|p_{\nu k}(t)\right|}{p_{\nu\nu}(t)}(\nu\in 1:n,\ k\in 0:\nu-1)\right\};$$

$$M_1 \doteq \max_{j\in 0:n-1}\ \operatorname*{vraisup}_{[a,b]\times[a,b]} \left|{}_{\mathscr{P}}^{j}C(t,s)\,g(s)\right|, \quad H \doteq n\,L\,M_1.$$

Покажем, что на J справедливы оценки

$$\left|{}_{\mathscr{P}}^{0}x_k(t)\right| \leqslant C\,(M\,(t-a))^k\,(k\,!\,)^{-1} \quad (k=0,\,1,\,2,\,\ldots), \tag{13}$$

$$\left|\left({}_{\mathscr{P}}^{0}x_0(t)\right)'\right| \leqslant n\,L\,C_1, \tag{14}$$

$$\left|\left({}_{\mathscr{P}}^{0}x_k(t)\right)'\right| \leqslant \left(H/M\right)\,(M\,(t-a))^k\,(k\,!\,)^{-1} \tag{15}$$

$$(k=1,2,\ldots).$$

Для $k=0$ они очевидны. Из справедливости (13) для некоторого $k>0$ следует, что имеют место следующие оценки

$$\left|{}_{\mathscr{P}}^{0}x_{k+1}(t)\right| \leqslant \int_a^t \left|{}_{\mathscr{P}}^{0}C(t,s)\,g(s)\right|\left|{}_{\mathscr{P}}^{0}x_k(s)\right|\,ds \leqslant C\,(M\,(t-a))^{k+1}\,((k+1)\,!\,)^{-1},$$

то есть справедливость оценки (13) для $k+1$. По индукции неравенство (13) справедливо для $k\in\mathbb{N}$. Оценка (14) очевидна.

Справедливость оценки (15) получается из следующей цепочки неравенств

$$\left|\left({}_{\mathscr{P}}^{0}x_k(t)\right)'\right| \leqslant \frac{1}{p_{11}(t)}\int_a^t \left|{}_{\mathscr{P}}^{1}C(t,s)\,g(s)\right|\left|{}_{\mathscr{P}}^{0}x_{k-1}(s)\right|\,ds +$$

$$+\ \frac{p_{10}(t)}{p_{11}(t)}\int_a^t \left|{}_{\mathscr{P}}^{0}C(t,s)\,g(s)\right|\left|{}_{\mathscr{P}}^{0}x_{k-1}(s)\right|\,ds \leqslant$$

$$\leqslant \left(H/M\right)(M\,(t-a))^k\,(k\,!\,)^{-1} \quad (k=1,\,2,\,\ldots).$$

Из оценок (13)–(15) и признака Вейерштрасса следует, что ряд, полученный почленным дифференцированием ряда (12), сходится равномерно на J. Следовательно, ряд (12) допускает почленное дифференцирование и имеет место равенство

$${}_{\mathscr{P}}^{1}\left(\sum_{k=0}^{\infty}(-\lambda)^k\,x_k(t)\right) = \sum_{k=0}^{\infty}(-\lambda)^k\,{}_{\mathscr{P}}^{1}x_k(t). \tag{16}$$

Точно так же, как мы только что доказали равенство (16), с использованием индукции по j, доказывается, что

$${}_{\mathscr{P}}^{j}\left(\sum_{k=0}^{\infty}(-\lambda)^k\,x_k(t)\right) = \sum_{k=0}^{\infty}(-\lambda)^k\,{}_{\mathscr{P}}^{j}x_k(t) \tag{17}$$

$$(j=2,\,\ldots,\,n).$$

Из оценки (13) при $\lambda \in [-\lambda,\ \lambda]$ получаем следующую оценку

$$\left|(-1)^k\ \lambda^k\ \ell_n x_k\right| \leqslant n\left(\max_{k\in 0:n-1}\ \left|\beta_{nk}\right|\ \right) M_1 L^k\ (C/M)\ \frac{(M(b-a))^k}{k!},$$

что позволяет применить к ряду (11) признак Вейерштрасса и получить первое утверждение теоремы. Учитывая равенство (17) и определение последовательности $\{x_k\}_0^\infty$, получим, что функция $u(t,\lambda)$, определяемая равенством

$$u(t,\lambda) = \sum_{k=0}^{\infty} (-\lambda)^k x_k(t), \tag{18}$$

удовлетворяет уравнению (4) и краевым условиям (5), за исключением, быть может, последнего. Очевидно, что $\ell_n u(t,\lambda) = 0$ в том и только том случае, если λ — корень уравнения $\Phi(\lambda) = 0$. Теорема доказана.

Если уравнение $\Phi(\lambda) = 0$ имеет лишь простые корни, то все собственные функции имеют представление (18).

Рассмотрим простой пример.

Пример 1:

$$x'' = -\lambda x,\quad x(a) = x(b) = 0$$

($n = 2$, $\mathscr{P}$ — единичная 3×3 матрица, $g(t)\equiv 1$, $\ell_1\ x \doteq x(a)$, $\ell_2\ x \doteq x(b)$).

Здесь

$$\Phi(\lambda) = (b-a) - \lambda\frac{(b-a)^3}{3!} + \lambda^2\frac{(b-a)^5}{5!} - \ \ldots = \frac{\sin\sqrt{\lambda(b-a)}}{\sqrt{\lambda}}.$$

Отсюда получаем собственные значения λ_k исходной задачи есть

$$\lambda_k = \frac{(\pi k)^2}{(b-a)^2},\ k = 1,\ 2,\ \ldots\ .$$

Укажем здесь на одно из возможных применений теоремы 1.

Данная теорема и формулы (7), (8), а также — (9), (10), позволяют предложить новый метод для вычисления собственных значений квазидифференциальной краевой задач (4), (5) и построить алгоритм ее численного решения на компьютере. Решалась задача

$$\begin{gathered}\left(p(t)\,x'\right)' = -\lambda g(t)\,x(t),\\ x(0) = x(1) = 0\ .\end{gathered} \tag{19}$$

Для нахождения $x_0(t)$, а также для решения задач Коши (9), (10) применялся метод Рунге—Кутта четвертого порядка. Приближения к собственным значениям найдены методом Греффе—Лобачевского как корни многочлена — нечетной частичной суммы ряда (11).

Ниже приведены результаты счета.

λ_μ^k — собственное значение, вычисленное одним из следующих методов (где нижний индекс μ у λ_μ^k означает номер собственного значения):

- $k = 1$ — рассматриваемым методом;
- $k = 2$ — методом Галеркина—Ритца (в котором взяты в качестве координатных функций: $t^2 - t$ и $t^3 - \frac{3}{2}t^2 + \frac{1}{2}t$);
- $k = 3$ — методом последовательных приближений $\left(x_0(t) = 1\right)$;
- $k = 4$ — точное значение (когда оно известно).

Таблица 1. Результаты вычислений первого и второго собственных значений задачи вида (19)*

$p(t)$	$g(t)$	λ_1^1 λ_2^1	λ_1^2 λ_2^2	λ_1^3	λ_1^4 λ_2^4
1	1	9.8667 33	10.0000 42	9.8697	9.8696 39.4784
$\sin\left(t+\frac{1}{2}\right)$	1	7.483 26	7.495 33	7.485	
1	1	18.967 68	19.19 102	18.970	18.956
$\ln^2(t+1)+0.1$	1	2.403 10	2.43 13	2.403	
$\exp(t)$	1	16.23 54	16.32 75	16.24	
t^2+3t+1	1	24.28 84	24.36 126	24.29	
$t^2\ln(t+1)+0.1$	1	2.329 9	2.39 15	2.329	

3. ТЕОРЕМЫ ОБ ОЦЕНКАХ И ПРИМЕРЫ К ТЕОРЕМАМ

Пусть $n=2$. Уравнение (1) примет вид

$$\left(\overset{2}{\mathscr{P}}x\right)(t)=0,\ \ t\in J. \tag{20}$$

Уравнение (20) называется неосцилляционным на промежутке $J\subset I$ (напомним, что здесь $J=[a,b]$, $-\infty<a<b<+\infty$), если нулевая квазипроизводная любого его нетривиального решения имеет на J не более одного нуля (см. [24], [25]), заметим, что в этих работах дается определение неосцилляции для уравнения произвольного порядка). При условии неосцилляции имеет место неравенство $\overset{0}{\mathscr{P}}C(t,s)\geqslant 0$ при $a\leqslant s\leqslant t\leqslant b$ (см. [24], [25]).

Сопряженное однородное уравнение второго порядка имеет вид

$$\left(\overset{2}{\mathscr{R}}x\right)(t)=0,\ \ t\in J, \tag{21}$$

где $\mathscr{R}=\left(r_{\nu k}\right)_0^2$ — нижняя треугольная матрица,

$$r_{\nu k}=(-1)^{\nu+k}\ p_{n-k,n-\nu}\,p_{n-\nu,n-\nu}\big/p_{n-k,n-k}\quad (k\in 0:\nu,\ \nu\in 0:2).$$

Пусть $g(t)\doteq 1\big/p_{22}(t)$ $(t\in J)$. Рассмотрим краевую задачу на собственные значения вида (4), (5), а, именно, следующую задачу

$$\left(\overset{2}{\mathscr{P}}x\right)(t)=-\lambda\left(\overset{0}{\mathscr{P}}x\right)(t)\ \ (t\in J), \tag{22}$$

$$\overset{0}{\mathscr{P}}x(a)=\overset{0}{\mathscr{P}}x(b)=0. \tag{23}$$

Решение $u(t,\lambda)$ уравнения (22), удовлетворяющее первому условию из условий (23), представимо в виде ряда (18), который в данном случае выглядит так

$$u(t,\lambda)=x_0(t)-\lambda x_1(t)+\lambda^2x_2(t)-\lambda^3x_3(t)+\ \ldots\ . \tag{24}$$

* При вычислении второго собственного числа все методы дают значительную погрешность, поэтому результаты вычислений приведены с точностью до целых.

Нулевая квазипроизводная этого решения по аргументу t представима в виде ряда

$$\overset{0}{\mathscr{P}} u(t,\lambda) = \overset{0}{\mathscr{P}} x_0(t) - \lambda \overset{0}{\mathscr{P}} x_1(t) + \lambda^2 \overset{0}{\mathscr{P}} x_2(t) - \lambda^3 \overset{0}{\mathscr{P}} x_3(t) + \ \dots \ . \tag{25}$$

Последовательность решений в (24) (см. также (25)), $\{x_k(\cdot)\}_0^\infty$, строится следующим образом: $x_0(\cdot)$ — решение задачи Коши

$$\left(\overset{2}{\mathscr{P}} x\right)(t) = 0 \ \ (t \in J), \tag{26}$$

$$\overset{0}{\mathscr{P}} x(a) = 0, \ \overset{1}{\mathscr{P}} x(a) = 1, \tag{27}$$

то есть определения (7), (8) для задачи (22), (23) принимают вид (26), (27).

Функции $x_k(\cdot)$ находятся рекуррентно как решения задач

$$\left(\overset{2}{\mathscr{P}} x_k\right)(t) = \left(\overset{0}{\mathscr{P}} x_{k-1}\right)(t) \ \ (t \in J), \tag{28}$$

$$\overset{0}{\mathscr{P}} x_k(a) = 0, \ \ \overset{1}{\mathscr{P}} x_k(a) = 0 \ \ (k = 1, 2, \dots), \tag{29}$$

то есть определения (9), (10) для задачи (22), (23) принимают вид (28), (29).

Собственные значения задачи (22), (23) представляют собой корни уравнения $\Phi(\lambda) = 0$, где $\Phi(\cdot)$ — сумма ряда (24) при $t = b$. Функция

$$u(t, \lambda_*) = \sum_{k=0}^{\infty} (-1)^k (\lambda_*)^k x_k(t) \ (t \in J)$$

есть собственная функция краевой задачи (22), (23), отвечающая собственному значению λ_*.

Предполагаем функции $p_{\nu\nu}(t)$ $(\nu \in 0 : 2)$ положительными на отрезке J.

Теорема 2. *Пусть уравнение* (20) *неосциляционно на J и $C(t, s)$ — функция Коши уравнения* (20)*, вещественные константы M_1, M_2 и функция* $\varphi(\cdot)$ *таковы, что выполняются следующие неравенства*

$$\frac{\overset{0}{\mathscr{P}} C(t,s)}{p_{22}(s)} \leqslant M_1 \frac{\varphi(t-a)}{\varphi(s-a)}(t-s)$$

при всех значениях s и t таких, что $a \leqslant s \leqslant t \leqslant b$,

$$\overset{0}{\mathscr{P}} C(t,a) \leqslant M_2 \, \varphi(t-a)(t-a)$$

при всех значениях t, таких что $a \leqslant t \leqslant b$. Пусть, далее, $M \doteq \max\{M_1, M_2\}$. Тогда в представлении (25) *имеют место точные (см. пример* 2 *ниже) оценки для коэффициентов при степенях* λ*, а, именно,*

$$0 \leqslant \overset{0}{\mathscr{P}} x_k(t) \ \leqslant \ \frac{M^{k+1}(t-a)^{2k+1}\varphi(t-a)}{(2k+1)!} \ \ \left(t \in J, \ k = 0, 1, \dots\right). \tag{30}$$

Доказательство. Докажем эту теорему методом математической индукции. При $k = 0$ оценка (30) верна. Действительно, $\overset{0}{\mathscr{P}} x_0(t) = \overset{0}{\mathscr{P}} C(t,a)$, следовательно,

$$\overset{0}{\mathscr{P}} x_0(t) = \overset{0}{\mathscr{P}} C(t,a) \leqslant M_2 \, (t-a)\varphi(t-a) \leqslant M(t-a)\varphi(t-a).$$

Пусть оценка (30) имеет место для некоторого $k \in \mathbb{N}$. Покажем ее справедливость для $k+1$. Имеет место следующая цепочка неравенств:

$$\overset{0}{\mathscr{P}} x_{k+1}(t) = \int_a^t \frac{\overset{0}{\mathscr{P}} C(t,s) \overset{0}{\mathscr{P}} x_k(s)}{p_{22}(s)} \, ds \leqslant$$

$$\leqslant \frac{M^{k+1}}{(2k+1)!}\int_a^t M_1\frac{\varphi(t-a)}{\varphi(s-a)}\varphi(s-a)(t-s)(s-a)^{2k+1}\,ds \leqslant$$

$$\leqslant \frac{M^{k+2}}{(2k+1)!}\int_a^t \frac{\varphi(t-a)}{\varphi(s-a)}\varphi(s-a)(t-s)(s-a)^{2k+1}\,ds =$$

$$= \frac{M^{k+2}\varphi(t-a)}{(2k+2)!}\int_a^t (t-s)\,d(s-a)^{2k+2} = \frac{M^{k+2}\varphi(t-a)}{(2k+2)!}\int_a^t (s-a)^{2k+2}\,ds =$$

$$= \frac{M^{k+2}\varphi(t-a)}{(2k+3)!}\int_a^t d(s-a)^{2k+3} = \frac{M^{k+2}\varphi(t-a)}{(2k+3)!}(t-a)^{2k+3} =$$

$$= \frac{M^{(k+1)+1}(t-a)^{2(k+1)+1}\varphi(t-a)}{(2(k+1)+1)!}.$$

Таким образом, получаем неравенство

$$0 \leqslant {}^{0}_{\mathscr{P}}x_{k+1}(t) \leqslant \frac{M^{(k+1)+1}(t-a)^{2(k+1)+1}\varphi(t-a)}{(2(k+1)+1)!}.$$

Следовательно, по индукции оценки (30) имеет место для всех $k\in\mathbb{N}$. Теорема доказана.

Пример 2. Рассмотрим задачу вида (22), (23)

$$x'' + tx' + \left(t^2/4 + 1/2\right)x = -(\lambda - 1/2)x, \tag{31}$$

$$x(0) = x(1) = 0. \tag{32}$$

Для задачи (31), (32) обозначения теоремы 2 примут следующий вид:

$$a = 0,\quad b = 1,\quad p_{00}(t) = 1,$$

$$p_{11}(t) = 1,\ p_{10}(t) = 0,\ p_{22}(t) = 1,\ p_{21}(t) = t,\ p_{20}(t) = \frac{t^2}{4} + \frac{1}{2};$$

$$\frac{{}^{0}_{\mathscr{P}}C(t,s)}{p_{22}(s)} = \frac{e^{-(t^2+s^2)/4}}{e^{-s^2/2}}(t-s) = \frac{e^{-t^2/4}}{e^{-s^2/4}}(t-s) = \frac{\varphi(t)}{\varphi(s)}(t-s),$$

$$M_1 = M_2 = 1,\quad \varphi(t) = e^{-t^2/4},$$

$${}^{0}_{\mathscr{P}}x_k(t) = x_k(t) = \frac{t^{2k+1}}{(2k+1)!}e^{-t^2/4}\ \left(k = 0, 1, 2, \dots\right).$$

Правые части оценок (30) функциями ${}^{0}_{\mathscr{P}}x_k(t)$ достигаются. Представление (24) (или, что то же самое, в силу того, что $p_{00}(t) = 1$, представление (25) для задачи (31), (32)) примет вид

$$u(t,\lambda) = e^{-t^2/4}\sin\left(\sqrt{\lambda - 1/2}\,t\right)\Big/\sqrt{\lambda - 1/2}.$$

Корнями уравнения $\Phi(\lambda) = 0$, где $\Phi(\lambda) = u(1,\lambda)$, и собственными значениями задачи (31), (32) являются $\lambda_k = 1/2 + \left(\pi(k+1)\right)^2$ $\left(k = 0, 1, 2, \dots\right)$. Функции

$$u(t,\lambda_k) = \sum_{m=0}^{\infty}(-1)^m(\lambda_k - 1/2)^m x_m(t) = e^{-t^2/4}\,\frac{\sin\left(\pi(k+1)\,t\right)}{\pi(k+1)}$$

являются собственными функциями задачи (31), (32), соответствующими собственным значениям λ_k задачи (31), (32).

Теорема 3. *Пусть уравнение* (20) *неосциляционно на* J *и* $C(t,s)$ *— функция Коши уравнения* (20)*, вещественные константы* M_1*,* M_2 *и функция* $\varphi(\cdot)$ *таковы, что выполняются следующие неравенства*

$$1 \leqslant \varphi(t), \quad {}^{0}_{\mathscr{P}}C(t,s) \leqslant M_1 \int_s^t \frac{\varphi(s)}{p_{22}(s)}\,ds, \quad {}^{0}_{\mathscr{P}}C(t,a) \leqslant M_2 \int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds$$

при всех значениях s *и* t *таких, что* $a \leqslant s \leqslant t \leqslant b$*. Пусть, далее,*

$$M \doteq \max\left\{M_1, M_2\right\}.$$

Тогда в представлении (25) *имеют место точные (см. примеры* 3 *и* 4 *ниже) оценки для коэффициентов при степенях* λ*,* a*, именно,*

$$0 \leqslant {}^{0}_{\mathscr{P}}x_k(t) \leqslant \frac{M^{k+1}\left(\int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds\right)^{2k+1}}{(2k+1)!} \quad \left(t \in J,\ k = 0, 1, \dots\right). \tag{33}$$

Доказательство. Докажем эту теорему методом математической индукции. При $k = 0$ оценка (33) верна. Действительно,

$${}^{0}_{\mathscr{P}}x_0(t) = {}^{0}_{\mathscr{P}}C(t,a),$$

следовательно,

$${}^{0}_{\mathscr{P}}x_0(t) = {}^{0}_{\mathscr{P}}C(t,a) \leqslant M_2 \int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds \leqslant M \int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds.$$

Пусть оценка (33) имеет место для некоторого $k \in \mathbb{N}$. Покажем ее справедливость для $k+1$. Имеет место следующая цепочка неравенств

$$\begin{aligned}
{}^{0}_{\mathscr{P}}x_{k+1}(t) &= \int_a^t \frac{{}^{0}_{\mathscr{P}}C(t,s)\,{}^{0}_{\mathscr{P}}x_k(s)}{p_{22}(s)}\,ds \leqslant \\
&\leqslant \frac{M^{k+1}}{(2k+1)!}\int_a^t \left(\frac{M_1\varphi(s)}{p_{22}(s)}\right)\left(\int_s^t \frac{\varphi(\tau)}{p_{22}(\tau)}\,d\tau\right)\left(\int_a^s \frac{\varphi(\tau_1)}{p_{22}(\tau_1)}\,d\tau_1\right)^{2k+1} ds \leqslant \\
&\leqslant \frac{M^{k+2}}{(2k+2)!}\int_a^t \left(\int_s^t \frac{\varphi(\tau)}{p_{22}(\tau)}\,d\tau\right) d\left(\int_a^s \frac{\varphi(\tau_1)}{p_{22}(\tau_1)}\,d\tau_1\right)^{2k+2} = \\
&= \frac{M^{k+2}}{(2k+2)!}\int_a^t \left(\frac{\varphi(s)}{p_{22}(s)}\right)\left(\int_a^s \frac{\varphi(\tau_1)}{p_{22}(\tau_1)}\,d\tau_1\right)^{2k+2} ds = \\
&= \frac{M^{k+2}}{(2k+3)!}\left(\int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds\right)^{2k+3} = \frac{M^{(k+1)+1}}{(2(k+1)+1)!}\left(\int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds\right)^{2(k+1)+1}.
\end{aligned}$$

Таким образом, получаем следующее неравенство

$$0 \leqslant {}^{0}_{\mathscr{P}}x_k(t) \leqslant \frac{M^{(k+1)+1}\left(\int_a^t \frac{\varphi(s)}{p_{22}(s)}\,ds\right)^{2(k+1)+1}}{(2(k+1)+1)!}.$$

По индукции оценки (33) имеет место для всех $k \in \mathbb{N}$. Теорема доказана.

Пример 3. Рассмотрим задачу вида (22), (23)

$$\frac{1}{\cos^2 t}x'' + \frac{\sin t}{\cos^3 t}x' = -\lambda x \tag{34}$$

$$x(0) = x(1) = 0. \tag{35}$$

Для задачи (34), (35) обозначения теоремы 3 примут следующий вид:

$$a = 0, \quad b = 1, \quad p_{00}(t) = 1,$$

$$p_{11}(t) = 1,\ p_{10}(t) = 0, \quad p_{22}(t) = \frac{1}{\cos^2 t}, \quad p_{21}(t) = \frac{\sin t}{\cos^3 t}, \quad p_{20}(t) = 0;$$

$$\overset{0}{\mathscr{P}} C(t,s) = C(t,s) = \int_s^t \frac{1}{\frac{1}{\cos\tau}}\,d\tau = \int_s^t \cos\tau\,d\tau = \sin t - \sin s,$$

$$M_1 = M_2 = 1, \quad \varphi(t) \equiv 1,$$

$$\overset{0}{\mathscr{P}} x_k(t) = x_k(t) = \frac{\left(\sin t\right)^{2k+1}}{(2k+1)!} \quad \left(k = 0, 1, 2, \ldots\right).$$

Правые части оценок (33) функциями $\overset{0}{\mathscr{P}} x_k(t)$ достигаются.

Представление (24) (или, что то же самое, в силу того, что $p_{00}(t) = 1$, представление (25) для задачи (34), (35)) примет вид $u(t,\lambda) = \dfrac{\sin\left(\sqrt{\lambda}\sin t\right)}{\sqrt{\lambda}}$. Корнями уравнения

$$\Phi(\lambda) = 0, \text{ где } \Phi(\lambda) = u(1,\lambda),$$

и собственными значениями задачи (34), (35) являются $\lambda_k = \left(\dfrac{\pi k}{\sin 1}\right)^2$.

Функции

$$u(t,\lambda_k) = \sum_{m=0}^{\infty}(-1)^m \lambda_k^m x_m(t) = \frac{\sin 1 \sin\left(\dfrac{\pi k \sin t}{\sin 1}\right)}{\pi k}$$

являются собственными функциями задачи (34), (35), соответствующими собственным значениям λ_k задачи (34), (35).

Пример 4. Рассмотрим задачу вида (22), (23)

$$p(t)\Big(p(t)\,x'\Big)' = -\lambda x \quad (t \in [0,1]) \tag{36}$$

(функция $p(t)$ измерима, почти всюду конечна, неотрицательна и суммируема на отрезке $[0,1]$),

$$x(0) = x(1) = 0. \tag{37}$$

Для задачи (36), (37) обозначения теоремы 3 примут следующий вид:

$$a = 0, \quad b = 1, \quad p_{00}(t) = 1,$$

$$p_{11}(t) = p(t),\ p_{10}(t) = p_{21}(t) = 0,\ p_{22}(t) = p(t),\ p_{20}(t) = 0;$$

$$\overset{0}{\mathscr{P}} C(t,s) = C(t,s) = \int_s^t \frac{d\tau}{p(\tau)}, \quad M_1 = M_2 = 1, \quad \varphi(t) \equiv 1,$$

$$ {}^{0}_{\mathscr{P}}x_k(t) = x_k(t) = \frac{\left(\int_0^t \frac{d\tau}{p(\tau)}\right)^{2k+1}}{(2k+1)!} \quad \left(k = 0, 1, 2, \dots\right). $$

Правые части оценок (33) функциями ${}^{0}_{\mathscr{P}}x_k(t)$ достигаются.

Представление (24) (или, что то же самое, в силу того, что $p_{00}(t) = 1$, представление (25) для задачи (36), (37)) примет вид

$$ u(t,\lambda) = \frac{\sin\left(\sqrt{\lambda}\left(\int_0^t \frac{d\tau}{p(\tau)}\right)\right)}{\sqrt{\lambda}}. $$

Корнями уравнения $\Phi(\lambda) = 0$, где $\Phi(\lambda) = u(1,\lambda)$, и собственными значениями задачи (36), (37) являются

$$ \lambda_k = \left(\frac{\pi k}{\int_0^1 \frac{d\tau}{p(\tau)}}\right)^2. $$

Функции

$$ u(t,\lambda_k) = \sum_{m=0}^{\infty}(-1)^m \lambda_k^m x_m(t) = \frac{\left(\int_0^1 \frac{d\tau}{p(\tau)}\right)\sin\left(\frac{\pi k \int_0^t \frac{d\tau}{p(\tau)}}{\int_0^1 \frac{d\tau}{p(\tau)}}\right)}{\pi k} $$

являются собственными функциями задачи (36), (37), соответствующими собственным значениям λ_k задачи (36), (37).

Пусть, по-прежнему, уравнение (20) неосциляционно на J. Пусть $C^*(t,s)$ есть функция Коши уравнения (21). Напомним, что функции $p_{ii}(t)\,(i \in 0 : 2)$ предполагаются положительными на J. Пусть, далее,

$$ M_1 \doteq \max_{t\in[a,b]} {}^{0}_{\mathscr{P}}C(t,a), \quad M_2 \doteq \max_{(s,t)\in[a,b]\times[a,b]} \left|{}^{1}_{\mathscr{R}}C^*(s,t)\right| $$

и $M \doteq \max\{M_1, M_2\}$, функция $\xi(t) \doteq \min\{p_{11}(t),\ p_{22}(t)\}$ при каждом значении аргумента t из J. Будем предполагать, что функция $\dfrac{1}{\xi(t)}$ суммируема на J. Определим функцию

$$ \psi(t) \doteq \int_a^t \frac{1}{\xi(s)}\, ds \ (t \in J). $$

Считаем далее также, что функция $p_{21}(t) \geqslant 0$ $(t \in J)$. Имеет место следующая теорема.

Теорема 4. *Справедливы оценки*

$$ 0 \leqslant {}^{0}_{\mathscr{P}}x_k(t) \leqslant \frac{M^{k+1}\psi^{2k}(t)}{(2k)!} \quad (t \in J,\ k = 0, 1, \dots). \tag{38} $$

Доказательство. Докажем эту теорему методом математической индукции. При $k = 0$ оценка (38) верна. Действительно, ${}^{0}_{\mathscr{P}}x_0(t) = {}^{0}_{\mathscr{P}}C(t,a)$, следовательно, ${}^{0}_{\mathscr{P}}x_0(t) \leqslant M_1 \leqslant M$. Пусть оценка (38) имеет место для некоторого $k \in \mathbb{N}$. Покажем ее справедливость для $k+1$. Выполняется следующая цепочка неравенств

$$ {}^{0}_{\mathscr{P}}x_{k+1}(t) = \int_a^t \frac{{}^{0}_{\mathscr{P}}C(t,s)\,{}^{0}_{\mathscr{P}}x_k(s)}{p_{22}(s)}\, ds \leqslant \frac{M^{k+1}}{(2k)!}\int_a^t \frac{{}^{0}_{\mathscr{P}}C(t,s)\psi^{2k}(s)}{p_{22}(s)}\, ds \leqslant \frac{M^{k+1}}{(2k)!}\int_a^t \frac{{}^{0}_{\mathscr{P}}C(t,s)\psi^{2k}(s)}{\xi(s)}\, ds = $$

$$= \Big(\text{заметим, что } {}^{0}_{\mathscr{P}}C(t,s) = -{}^{0}_{\mathscr{R}}C^*(s,t)\ \big(\text{см. [24], [25]}\big)\Big) =$$

$$= -\frac{M^{k+1}}{(2k)!}\int_a^t {}^{0}_{\mathscr{R}}C^*(s,t)\psi^{2k}(s)\,d\psi(s) = -\frac{M^{k+1}}{(2k+1)!}\int_a^t {}^{0}_{\mathscr{R}}C^*(s,t)\,d\psi^{2k+1}(s) =$$

$$= \frac{M^{k+1}}{(2k+1)!}\int_a^t \psi^{2k+1}(s)\left({}^{0}_{\mathscr{R}}C^*(s,t)\right)'_s\,ds =$$

$$= \frac{M^{k+1}}{(2k+1)!}\int_a^t \frac{p_{11}(s)\left({}^{0}_{\mathscr{R}}C^*(s,t)\right)'_s}{p_{11}(s)}\psi^{2k+1}(s)\,ds = \frac{M^{k+1}}{(2k+1)!}\times$$

$$\times\int_a^t \frac{p_{11}(s)\left({}^{0}_{\mathscr{R}}C^*(s,t)\right)'_s - \dfrac{p_{21}(s)p_{11}(s)}{p_{22}(s)}\,{}^{0}_{\mathscr{R}}C^*(s,t) + \dfrac{p_{21}(s)p_{11}(s)}{p_{22}(s)}\,{}^{0}_{\mathscr{R}}C^*(s,t)}{p_{11}(s)}\times$$

$$\times\psi^{2k+1}(s)\,ds = \Big(\text{так как имеет место следующее равенство:}$$

$${}^{1}_{\mathscr{R}}C^*(s,t) = p_{11}(s)\left({}^{0}_{\mathscr{R}}C^*(s,t)\right)'_s - \frac{p_{21}(s)p_{11}(s)}{p_{22}(s)}\,{}^{0}_{\mathscr{R}}C^*(s,t),$$

$$\text{то последнее выражение примет вид}\Big) =$$

$$= \frac{M^{k+1}}{(2k+1)!}\int_a^t \frac{{}^{1}_{\mathscr{R}}C^*(s,t) - \dfrac{p_{21}(s)p_{11}(s)}{p_{22}(s)}\,{}^{0}_{\mathscr{P}}C(t,s)}{p_{11}(s)}\psi^{2k+1}(s)\,ds \leqslant$$

$$\leqslant \left(\text{учтем, что } \frac{p_{21}(s)p_{11}(s)}{p_{22}(s)}\,{}^{0}_{\mathscr{P}}C(t,s) \geqslant 0\right) \leqslant \frac{M^{k+1}}{(2k+1)!}\int_a^t \frac{{}^{1}_{\mathscr{R}}C^*(s,t)}{p_{11}(s)}\psi^{2k+1}(s)\,ds \leqslant$$

$$\leqslant \frac{M^{k+1}}{(2k+1)!}\int_a^t \frac{\left|{}^{1}_{\mathscr{R}}C^*(s,t)\right|}{\xi(s)}\psi^{2k+1}(s)\,ds \leqslant \frac{M^{k+1}M_1}{(2k+1)!}\int_a^t \psi^{2k+1}(s)\,d\psi(s) \leqslant$$

$$\leqslant \frac{M^{k+2}}{(2k+2)!}\psi^{2k+2}(t) = \frac{M^{(k+1)+1}\psi^{2(k+1)}(t)}{(2(k+1))!}.$$

Таким образом, ${}^{0}_{\mathscr{P}}x_{k+1}(t) \leqslant \dfrac{M^{(k+1)+1}\psi^{2(k+1)}(t)}{(2(k+1))!}$. По индукции оценки (38) имеют место для всех $k\in\mathbb{N}$. Теорема доказана.

Замечание 1. Из теоремы 4 следует, что если $p_{11}(t) = p_{22}(t) \equiv 1$ на J, или $1 \leqslant p_{11}(t)$ при всех t из отрезка J и $p_{22}(t) \equiv 1$ на J (либо $1 \leqslant p_{22}(t)$ при всех t из J и $p_{11}(t) \equiv 1$ на J), то тогда оценки (38) выглядят так

$$0 \leqslant {}^{0}_{\mathscr{P}}x_k(t) \leqslant \frac{M^{k+1}(t-a)^{2k}}{(2k)!}\ (\,t\in J,\ \ k = 0,1,\dots\,). \tag{39}$$

Оценки (38) и (39), в отличие от оценок (30) и (33), не являются точными.

Ведем в рассмотрение функции $\varphi_k(t) : J \to \mathbb{R}$ с помощью следующих равенств:

$$\varphi_k(t)\frac{M^{k+1}\psi^{2k}(t)}{(2k)!} = {}^{0}_{\mathscr{P}}x_k(t),$$

где $k = 0,1,\dots$ (заметим, что $0 \leqslant \varphi_k(t) \leqslant 1$).

Определим функцию

$$v(t,\lambda) \doteq \varphi_0(b)M + \sum_{k=1}^{\infty} \frac{(-1)^k \lambda^k \varphi_k(b) M^{k+1} \psi^{2k}(t)}{(2k)!} \quad (t \in J). \tag{40}$$

Функция $v(t,\lambda)$ является как бы "сечением" функции ${}^{0}_{\mathscr{P}}u(t,\lambda)$ при $t = b$.

Теорема 5. *Функция $v(t,\lambda)$ удовлетворяет уравнению*

$$\lambda\big(v(t,\lambda)\big)''_{\sqrt{\lambda}} = \psi^2(t)\left(\xi(t)\left(\xi(t)\big(v(t,\lambda)\big)'_t\right)'_t\right) \tag{41}$$

и условиям

$$\begin{gathered} v(b,\lambda) = {}^{0}_{\mathscr{P}}u(b,\lambda), \quad \left(\xi(t)\big(v(t,\lambda)\big)'_t\right)\Big|_{t=a} = 0, \\ v(t,\lambda)\big|_{\lambda=0} = {}^{0}_{\mathscr{P}}C(b,a), \ \left(\left(\sqrt{\lambda}\, v(t,\lambda)\right)'_{\sqrt{\lambda}}\right)\Big|_{\lambda=0} = {}^{0}_{\mathscr{P}}C(b,a). \end{gathered} \tag{42}$$

Доказательство. В том, что функция $v(t,\lambda)$ удовлетворяет уравнению (41), убедимся непосредственной подстановкой правой части формулы (40) в левую и правую части уравнения (41)

$$\lambda\big(v(t,\lambda)\big)''_{\sqrt{\lambda}} =$$

$$= -\lambda\varphi_1(b)M^2\psi^2(t) + \frac{\lambda^2\varphi_2(b)M^3\psi^4(t)}{2!} - \frac{\lambda^3\varphi_3(b)M^4\psi^6(t)}{4!} + \dots,$$

$$\psi^2(t)\left(\xi(t)\left(\xi(t)\big(v(t,\lambda)\big)'_t\right)'_t\right) =$$

$$= -\lambda\varphi_1(b)M^2\psi^2 + \frac{\lambda^2\varphi_2(b)M^3\psi^4}{2!} - \frac{\lambda^3\varphi_3(b)M^4\psi^6}{4!} + \dots$$

Получили одно и то же выражение.

Подставляем правую часть формулы (40) в левые части каждого условия из четырех условий (42)

$$v(b,\lambda) = \varphi_0(b)M - \frac{\lambda\varphi_1(b)M^2\psi^2(b)}{2!} + \frac{\lambda^2\varphi_2(b)M^3\psi^4(b)}{4!} - \frac{\lambda^3\varphi_3(b)M^4\psi^6(b)}{6!} + \dots =$$

$$= p_{00}(b)\Big(x_0(b) - \lambda x_1(b) + \lambda^2 x_2(b) - \lambda^3 x_3(b) + \dots\Big) = {}^{0}_{\mathscr{P}}u(b,\lambda),$$

$$\left(\xi(t)\big(v(t,\lambda)\big)'_t\right)\Big|_{t=a} = \left(-\frac{\lambda\varphi_1(b)M^2\psi^1(t)}{1!} + \frac{\lambda^2\varphi_2(b)M^3\psi^3(t)}{3!} - \frac{\lambda^3\varphi_3(b)M^4\psi^5(t)}{5!} + \dots\right)\Big|_{t=a} =$$

$$= -\frac{\lambda\varphi_1(b)M^2\psi^1(a)}{1!} + \frac{\lambda^2\varphi_2(b)M^3\psi^3(a)}{3!} - \frac{\lambda^3\varphi_3(b)M^4\psi^5(a)}{5!} + \dots = (\psi(a) = 0) = 0,$$

$$v(t,\lambda)\big|_{\lambda=0} = \left(\varphi_0(b)M - \frac{\lambda\varphi_1(b)M^2\psi^2(t)}{2!} + \frac{\lambda^2\varphi_2(b)M^3\psi^4(t)}{4!} - \right.$$

$$\left. - \frac{\lambda^3\varphi_3(b)M^4\psi^6(t)}{6!} + \dots\right)\Big|_{\lambda=0} = {}^{0}_{\mathscr{P}}C(b,a),$$

$$\left(\left(\sqrt{\lambda}\,v(t,\lambda)\right)'_{\sqrt{\lambda}}\right)\Big|_{\lambda=0} = \left(\varphi_0(b)M\sqrt{\lambda} - \frac{(\sqrt{\lambda})^3\varphi_1(b)M^2\psi^2(t)}{2!} + \right.$$

$$\left. + \frac{(\sqrt{\lambda})^5\varphi_2(b)M^3\psi^4(t)}{4!} - \ldots\right)'_{\sqrt{\lambda}}\Big|_{\lambda=0} = \varphi_0(b)M + 0 = \overset{0}{\mathscr{P}}C(b,a).$$

Убеждаемся в том, что функция $v(t,\lambda)$ удовлетворяет условиям (42). Теорема доказана.

Уравнение (41) назовем “квазидифференциальным уравнением в частных производных”.

То, что решение задачи (41), (42), функция $v(t,\lambda)$ удовлетворяет первому условию из условий (42), означает, что собственные значения задачи (22), (23) могут быть найдены как корни уравнения

$$v(b,\lambda) = 0. \tag{43}$$

Поэтому определенный интерес представляет вопрос об аппроксимации корней уравнения (43) в тех случаях, когда его корни не могут быть найдены точно. На этом закончим рассмотрение теорем об оценках и последний раздел настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Левитан Б. М., Саргсян И. С.* Некоторые вопросы теории Штурма–Лиувилля // УМН. 1960. Т. 15. № 1(91). С. 3–98.
2. *Левитан Б. М., Саргсян И. С.* Введение в спектральную теорию. М.: Наука, 1970.
3. *Марченко В. А.* Операторы Штурма–Лиувилля и их приложения. К.: Наук. думка, 1977.
4. *Костюченко А. Г., Саргсян И. С.* Распределение собственных значений (самосопряженные обыкновенные дифференциальные операторы). М.: Наука, 1979.
5. *Садовничий В. А.* Теория операторов. М.: Изд-во МГУ, 1986.
6. *Левитан Б. М., Саргсян И. С.* Операторы Штурма–Лиувилля и Дирака. М.: Наука, 1988.
7. *Мирзоев К. А.* Операторы Штурма–Лиувилля // Тр. ММО. 2014. Т. 75. № 2. С. 335–359.
8. *Конечная Н. Н., Мирзоев К. А.* Главный член асимптотики решений линейных дифференциальных уравнений с коэффициентами-распределениями первого порядка // Матем. заметки. 2019. Т. 106. № 1. С. 74–83.
9. *Конечная Н. Н., Мирзоев К. А.* Об асимптотике решений линейных дифференциальных уравнений нечетного порядка // Вестн. Моск. ун-та. 2020. Сер. 1. Матем., мех. № 1. С. 23–28.
10. *Винокуров В. А., Садовничий В. А.* Асимптотика любого порядка собственных значений и собственных функций краевой задачи Штурма–Лиувилля на отрезке с суммируемым потенциалом // Изв. РАН, Сер. мат. 2000. Т. 64. № 4. С. 47–108.
11. *Савчук А. М., Шкаликов А. А.* Операторы Штурма–Лиувилля с сингулярными потенциалами // Матем. заметки. 1999. Т. 66. № 6. С. 897–912.
12. *Савчук А. М.* О собственных значениях и собственных функциях оператора Штурма–Лиувилля с сингулярным потенциалом // Матем. заметки. 2001. Т. 69. № 2. С. 277–285.
13. *Савчук А. М., Шкаликов А. А.* Операторы Штурма–Лиувилля с потенциалами-распределениями // Тр. Моск. матем. общ-ва. 2003. Т. 64. С. 159–212.
14. *Конечная Н. Н., Сафонова Т. А., Тагирова Р. Н.* Асимптотика собственных значений и регуляризованный след первого порядка оператора Штурма–Лиувилля с δ-потенциалом // Вестн. Сев. (Арктич.) федер. ун-та. Сер.: Естеств. науки. 2016. Вып. 1. С. 104–113.
15. *Сафонова Т. А., Рябченко С. В.* О собственных значениях оператора Штурма–Лиувилля с сингулярным потенциалом // Вестн. Сев. (Арктич.) федер. ун-та. Сер.: Естеств. науки. 2016. Вып. 2. С. 115–125.
16. *Покорный Ю. В., Прядиев В. Л.* Некоторые вопросы качественной теории Штурма–Лиувилля на пространственной сети // УМН. 2004. Т. 59. №3 (357). С. 116–150.

17. *Покорный Ю. В., Зверева М. Б., Ищенко А. С., Шабров С. А.* О нерегулярном расширении осцилляционной теории спектральной задачи Штурма–Лиувилля // Матем. заметки. 2007. Т. 82. № 4. С. 578–582.
18. *Покорный Ю. В., Зверева М. Б., Шабров С. А.* Осцилляционная теория Штурма–Лиувилля для импульсных задач // УМН. 2008. Т. 63. №1 (379). С. 111–154.
19. *Митрохин С. И.* Спектральная теория операторов: гладкие, разрывные, суммируемые коэффициенты. М.: ИНТУИТ, 2009.
20. *Митрохин С. И.* О спектральных свойствах многоточечной краевой задачи для дифференциального оператора нечетного порядка с суммируемым потенциалом // Arctic Environmental Research. 2017. Т. 17. № 4. С. 376–392.
21. *Митрохин С. И.* Асимптотика собственных значений дифференциального оператора со знакопеременной весовой функцией // Изв. вузов. Матем. 2018. № 6. С. 31–47.
22. *Митрохин С. И.* Об асимптотике собственных значений дифференциального оператора четвертого порядка со знакопеременной весовой функцией // Вестник МГУ. Сер.: “Математика, механика”. 2018. № 6. С. 46–58.
23. *Митрохин С. И.* Асимптотика спектра дифференциального оператора четного порядка с разрывной весовой функцией // Журнал СВМО. 2020. Т. 22. № 1. С. 48–70.
24. *Дерр В. Я.* Неосцилляция решений линейного квазидифференциального уравнения // Известия Института математики и информатики УдГУ. 1999. № 1(16). С. 3–105.
25. *Дерр В. Я.* Об адекватном описании сопряженного оператора // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2011. № 3. С. 43–63.
26. *Шин Д. Ю.* О решениях линейного квазидифференциального уравнения n-го порядка // Матем. сборник. 1940. Т. 7(49). № 3. С. 479–532.
27. *Шин Д. Ю.* О квазидифференциальных операторах в гильбертовом пространстве // Матем. сборник. 1943. Т. 13(55). № 1. С. 39–70.
28. *Everitt W. N., Marcus L.* Boundary value problems and symplectic algebra for ordinary differential and quasi-differential operators // Amer. Math. Soc. 1999. V. 61.
29. *Eckhardt J., Gestezy F., Nichols R., Teschl G.* Weyl–Titchmarsh theory for Sturm–Liuville operators with distributional potentials // Opuscula Mathematica. 2013. V. 33(3). P. 467–563.
30. *Everitt W. N., Race D.* The regular representation of singular second order differential expressions using quasi-derivatives // Proc. London Math. Soc. (3) 1992. V. 65(2). P. 383–404.
31. *Xiao xia Lv, Ji-jun Ao, Zettl A.* Dependence of eigenvalues of fourth-order differential equations with discontinuous boundary conditions on the problem // J. Math. Anal. Appl. 2017. V. 456(1). P. 671–685.
32. *Qinglan Bao, Jiong Sun, Xiaoling Hao, Zettl A.* Characterization of self-adjoint domains for regular even order C-symmetric differential operators // Electronic J. of Qualitative Theory of Diff. Equat. 2019. V. 62. P. 1–17.
33. *Zettl A.* Sturm-Liouville Theory. Amer. Math. Soc., 2005.
34. *Zett A.* Recent Developments in Sturm-Liouville Theory. Berlin, Boston: De Gruyter, 2021.
35. *Jianfang Qin, Kun Li, Zhaowen Zheng, Jinming Cai.* Dependence of eigenvalues of discontinuous fourth-order differential operators with eigenparameter dependent boundary conditions // J. of Nonlinear Math. Phys. 2022. V. 29(4). P. 776–793.
36. *Наймарк М. А.* Линейные дифференциальные операторы. М.: Наука, 1969.

ON THE APPROXIMATION OF THE FIRST EIGENVALUE OF SOME BOUNDARY VALUE PROBLEMS

M. Yu. Vatolkin*

Kalashnikov Izhevsk State Technical University, Studencheskaya st., 7, Izhevsk, 426069 Russia
**e-mail: vmyu6886@gmail.com*

Received 18 December, 2023
Revised 28 February, 2024
Accepted 05 March, 2024

Abstract. The paper studies the representation of eigenfunctions as scalar series for a two-point boundary value problem of the type $(n-1, 1)$ under the assumption that there exists a functional concentrated at one point such that the first $n-1$ of the original boundary conditions and $\widetilde{\ell} x = 1$ become the Cauchy conditions at this point. The eigenfunction of the boundary value problem under consideration, corresponding to the eigenvalue λ_*, is represented as a series in powers of λ_*. The equation $\Phi(\lambda) = 0$, where $\Phi(\lambda)$ is the sum of the series in powers of λ, is considered for finding the eigenvalues of the original problem. Examples of calculating the first eigenvalue of some boundary value problems are given. Various estimates are obtained for the coefficients of such power series. A certain function of two variables t and λ is defined, a partial differential equation is obtained for it, and conditions are obtained that it satisfies. The zeros of the "section" of this function coincide with the eigenvalues of the original boundary value problem, which can be used for their approximate calculation.

Keywords: boundary value problems for eigenvalues, eigenfunctions, eigenvalues, Cauchy function, representation of eigenfunctions as sums of power series, roots of the equation, estimates for the coefficients of power series.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 992–1007

УДК 517.63

АНАЛИТИКО-ЧИСЛЕННЫЙ МЕТОД РЕШЕНИЯ СПЕКТРАЛЬНОЙ ЗАДАЧИ В ОДНОЙ МОДЕЛИ ГЕОСТРОФИЧЕСКИХ ОКЕАНСКИХ ТЕЧЕНИЙ[1)]

© 2024 г. С. Л. Скороходов[1,*], Н. П. Кузьмина[2,]**

[1] *119991 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия*

[2] *117997 Москва, Нахимовский пр-т, 36, Институт океанологии им. П. П. Ширшова РАН, Россия*

[*]e-mail: sskorokhodov@gmail.com

[**]e-mail: kuzmina@ocean.ru

Поступила в редакцию 12.01.2024 г.
Переработанный вариант 09.02.2024 г.
Принята к публикации 15.02.2024 г.

Разработан новый эффективный аналитико-численный метод решения задачи для уравнения эволюции потенциального вихря в квазигеострофическом приближении с учетом вертикальной диффузии массы и импульса. Построенный метод применен к анализу малых возмущений океанских течений конечного поперечного масштаба с параболическим вертикальным профилем скорости общего вида. Для возникающей спектральной несамосопряженной задачи построены асимптотики собственных функций и собственных значений при малых значениях волнового числа k и показано существование счетного множества комплексных собственных значений с неограниченно убывающей мнимой частью. На отрезке интегрирования $z \in [-1, 1]$ введена система трех окрестностей, в каждой из которых решение строится в виде степенных разложений, гладкая сшивка которых позволяет эффективно вычислять собственные функции и собственные значения с высокой точностью. Рассчитаны траектории комплексных собственных значений для различных параметров задачи при изменении волнового числа k и показано существование двойных собственных значений. Кратко представлена сложная картина возникновения неустойчивости моделируемого течения в зависимости от физических параметров задачи. Библ. 16. Фиг. 7.

Ключевые слова: спектральная несамосопряженная задача, асимптотические разложения, высокоточный численный метод, двойные собственные значения.

DOI: 10.31857/S0044466924060088, **EDN:** XYOCBH

ВВЕДЕНИЕ

В работах [1]–[10] представлены результаты исследования устойчивых и неустойчивых малых возмущений геострофических течений с линейным и параболическим вертикальными профилями скорости. Модели основывались на уравнении эволюции потенциального вихря в квазигеострофическом приближении с учетом вертикальной диффузии массы и импульса. Вывод основных уравнений модели подробно представлен в [1]–[4]. Анализ проводился для течений с боковыми границами в конечных по вертикали слоях. Для течения с параболическим вертикальным профилем скорости были рассмотрены два частных случая: а) максимум скорости течения расположен к середине слоя (см. [1]–[5]); б) максимум скорости течения достигался на верхней или нижней границах слоя (см. [8]). Однако в океане наблюдаются течения с максимумом внутри слоя, а не только на его границах или в центре слоя. В связи с этим исследование динамики такого течения, т.е. течения с вертикальным параболическим профилем скорости общего вида, является актуальным. Такая задача была рассмотрена в работах [9], [10] с условиями равенства нулю потоков плавучести на границах слоя

[1)] Работа выполнена при участии Н. П. Кузьминой при поддержке бюджетного финансирования Института океанологии им. П. П. Ширшова РАН (тема FMWE-2024-0015).

$z = -1$ и $z = 1$. Представляется важным также исследовать динамику течения и при других граничных условиях, а именно, при условиях прилипания. Это соответствует равенству нулю возмущений горизонтальных скоростей на границах слоя, что эквивалентно, согласно формулировке задачи, равенству нулю возмущения давления на границах слоя.

Настоящая работа посвящена разработке нового эффективного подхода к аналитико-численному методу исследования устойчивости геострофического течения с параболическим профилем общего вида при различных граничных условиях. Возникающая модельная спектральная задача на отрезке $z \in [-1, 1]$ решается с помощью покрытия этого отрезка системой окрестностей, представления в каждой из окрестностей искомой собственной функции (СФ) $F(z)$ в виде степенных разложений и последующей гладкой сшивки. Предложенный метод позволяет эффективно и с высокой точностью находить собственные значения (СЗ) этой задачи. Показано, что при определенных физических параметрах могут возникать двойные СЗ. Представлены расчеты траекторий СЗ при непрерывном изменении волнового числа в задаче.

Постановка задачи. Областью исследования модельного течения является бесконечный в горизонтальном направлении слой толщины $2H$ с верхней и нижней границами $Z_1 = H$ и $Z_0 = -H$ и боковыми границами $Y_0 = 0$ и $Y_1 = L$. Декартовы координаты рассматриваемого слоя следующие: продольная переменная $X \in (-\infty, \infty)$, поперечная переменная $Y \in [0, L]$, вертикальная переменная $Z \in [-H, H]$.

Основываясь на методах анализа неустойчивости таких течений (см. работы [1]–[5]) будем представлять отклонение безразмерного давления в виде бегущей вдоль оси X волны:

$$p(X, Y, Z; T) = \sin\left(\pi n \frac{Y}{L}\right) e^{ik(X - cT)} F\left(\frac{Z}{H}\right), \qquad n \in \mathbb{N}, \tag{1}$$

где k — волновое число возмущения вдоль оси X, L/n — масштаб возмущения в направлении Y, c — комплексная фазовая скорость волны, а $F(Z/H)$ — искомый вертикальный профиль возмущения давления.

В безразмерных переменных задача исследования устойчивости течения с параболическим вертикальным профилем скорости общего вида сводится к решению спектральной задачи на отрезке $z \in [-1, 1]$. Введение безразмерных переменных (x, y, z) для этой задачи дано, например, в [3], [4]. При этом вертикальный профиль скорости основного течения в безразмерном виде будет иметь вид

$$U(z) = 1 - \alpha z^2 + \beta z, \tag{2}$$

где параметры α и β неотрицательны. Из естественного условия равенства нулю скорости $U(z)$ на нижней границе слоя $z = -1$ следует, что

$$\alpha + \beta = 1. \tag{3}$$

Постановка задачи исследования устойчивости малых возмущений течения с вертикальным профилем скорости $U(z)$ проводится аналогично задачам в [3]–[10] и сводится к решению следующей спектральной задачи на отрезке $z \in [-1, 1]$.

Задача. Найти комплекснозначные СФ $F = F(z)$ и СЗ "c", удовлетворяющие на отрезке $z \in [-1, 1]$ уравнению

$$\frac{1}{ik\mathrm{R}}\left(F^{(IV)} - \mathrm{Bu}\,\mathrm{Pr}(k^2 + \pi^2 n^2)F''\right) = (1 - \alpha z^2 + \beta z - c)\left(F'' - \mathrm{Bu}(k^2 + \pi^2 n^2)F\right) + 2\alpha F \tag{4}$$

с краевыми условиями

$$\frac{1}{ik\mathrm{R}} F'''(-1) = -c\,F'(-1) - (\beta + 2\alpha)\,F(-1), \qquad F(-1) = 0, \tag{5}$$

$$\frac{1}{ik\mathrm{R}} F'''(1) = (1 - \alpha + \beta - c)\,F'(1) - (\beta - 2\alpha)\,F(1), \qquad F(1) = 0, \tag{6}$$

причем неотрицательные параметры α и β удовлетворяют условию (3).

Здесь использованы следующие обозначения: $\mathrm{R} = \mathrm{Pe}\,H/L$, где Pe — число Пекле (аналог числа Рейнольдса), n — число полуволн в поперечном направлении ($n = 1, 2, \ldots$), Pr — число Прандтля, Bu — число Бургера, i — мнимая единица.

Краевые условия в (5), (6) с участием третьих производных задают непротекание на горизонтальных границах слоев, а условия для самих функций описывают равенство нулю возмущений давления на этих границах.

Поставленная задача является несамосопряженной, сингулярно возмущенной (для реальных течений величина $k\mathrm{R}$ может быть очень большой), а спектральный параметр "c", входит в уравнение и в краевые условия. Задача будет иметь счетное множество СФ $F_m(z)$ и соответствующих им СЗ c_m. Неустойчивые по времени возмущения давления $p(x, y, z; t)$ возникают для тех СФ, которым соответствуют СЗ c_m с положительной мнимой частью $\mathrm{Im}\,(c) > 0$, что следует из представления (1).

Исследуемая задача (4)–(6) в частном случае параметра $\alpha = 0$ обладает симметрией СЗ относительно прямой $\mathrm{Re}\,(c) = 1$.

Теорема. *Собственные значения задачи (4)–(6) при вещественных параметрах k,* R, Bu, Pr, *n и $\alpha = 0$ обладают симметрией относительно прямой* $\mathrm{Re}\,(c) = 1$*, т.е. если c_m является СЗ задачи, то и величина $\hat{c}_m = 2 - \mathrm{Re}\,(c_m) + i\mathrm{Im}\,(c_m)$ также является СЗ этой задачи.*

Доказательство совпадает с доказательством аналогичного факта в работе [6] и основано на анализе вещественной и мнимой частей СФ $F(z)$, соответствующей СЗ c_m.

1. МЕТОД РАСЧЕТА СФ И СЗ

Уравнение (4) не имеет конечных особых точек, поэтому его решение $F(z)$ является целой функцией (см., например, [11]), степенные разложения которой с центром в любой точке z_0 сходятся при всех значениях аргумента z. Метод расчета таких СФ и соответствующих СЗ был разработан ранее в [5], [6] и показал свою высокую точность и быстродействие. Суть этого подхода состоит в построении степенных разложений $F(z)$ в граничных точках $z = -1$ и $z = 1$ с тождественным удовлетворением краевым условиям и в последующей сшивке этих разложений в некоторой точке $z_* \in (-1, 1)$.

В настоящей работе использовано три степенных разложения искомого решения $F(z)$ в точках $z = -1$, $z = 0$ и $z = 1$ с последующей гладкой сшивкой первого и второго разложения в точке $z_{*,1} = -1/2$ и сшивкой второго и третьего разложения в точке $z_{*,2} = 1/2$. Сходимость таких разложений определяется скоростью убывания коэффициентов этих разложений, которая оказывается равной для всех трех случаев. Поэтому использование трех окрестностей с радиусом $r = 1/2$ для покрытия отрезка $z \in [-1, 1]$ требует гораздо меньшей длины разложений при достижении заданной точности, нежели в случае лишь двух окрестностей с радиусом $r = 1$. Помимо этого, при нахождении четных или нечетных решений $F(z)$ такой подход позволяет явно выделять соответствующие решения, благодаря построению разложения в точке $z = 0$.

Зафиксируем все параметры задачи и, в частности, спектральный параметр c; решение $F(c; z)$ уравнения (4) представим в виде разложений в точках $z = -1$, $z = 0$ и $z = 1$:

$$F(c; z) = \sum_{m=0}^{\infty} a_m(c)(z+1)^m, \quad F(c; z) = \sum_{m=0}^{\infty} b_m(c) z^m, \quad F(c; z) = \sum_{m=0}^{\infty} d_m(c)(z-1)^m, \tag{7}$$

сходящихся при любых z, $|z| < \infty$, где коэффициенты $a_m(c)$, $b_m(c)$ и $d_m(c)$ зависят от всех параметров задачи, но зависимость от c выделим отдельно, поскольку в дальнейшем этот параметр будем варьировать.

Подставляя представления (7) в уравнение (4), получаем рекуррентные соотношения для коэффициентов a_m, b_m и d_m, $m = 0, 1, \ldots$,

$$a_{m+4} = \Big\{ (m+1)(m+2)\Big[\mathrm{PrBu}(\pi^2 n^2 + k^2) - ik\mathrm{R}c\Big] a_{m+2} + ik\mathrm{R}m(m+1)(1+\alpha) a_{m+1} +$$

$$+ik\mathrm{R}\big[c\mathrm{Bu}(\pi^2n^2+k^2)-\alpha(m-2)(m+1)\big]\,a_m\,-ik\mathrm{RBu}(\pi^2n^2+k^2)(1+\alpha)a_{m-1}+$$
$$+ik\mathrm{R}\alpha\mathrm{Bu}(\pi^2n^2+k^2)a_{m-2}\Big\}\Big[(m+1)(m+2)(m+3)(m+4)\Big]^{-1}, \tag{8}$$

$$b_{m+4}=\Big\{(m+1)(m+2)\Big[\mathrm{PrBu}(\pi^2n^2+k^2)+ik\mathrm{R}(1-c)\Big]b_{m+2}+ik\mathrm{R}\beta m(m+1)b_{m+1}-$$
$$-ik\mathrm{R}\big[\alpha(m-2)(m+1)+(1-c)\mathrm{Bu}(\pi^2n^2+k^2)\big]b_m-ik\mathrm{R}\beta\mathrm{Bu}(\pi^2n^2+k^2)b_{m-1}+$$
$$+ik\mathrm{R}\alpha\mathrm{Bu}(\pi^2n^2+k^2)b_{m-2}\Big\}\Big[(m+1)(m+2)(m+3)(m+4)\Big]^{-1}, \tag{9}$$

$$d_{m+4}=\Big\{(m+1)(m+2)\Big[\mathrm{PrBu}(\pi^2n^2+k^2)+(2\beta-c)ik\mathrm{R}\Big]d_{m+2}+ik\mathrm{R}(\beta-2\alpha)m(m+1)d_{m+1}-$$
$$-ik\mathrm{R}\big[\mathrm{Bu}(\pi^2n^2+k^2)(2\beta-c)+\alpha(m-2)(m+1)\big]d_m-ik\mathrm{RBu}(\pi^2n^2+k^2)(\beta-2\alpha)d_{m-1}+$$
$$+ik\mathrm{R}\alpha\mathrm{Bu}(\pi^2n^2+k^2)d_{m-2}\Big\}\Big[(m+1)(m+2)(m+3)(m+4)\Big]^{-1}; \tag{10}$$

в этих формулах для коэффициентов a_{-2}, a_{-1}, b_{-2}, b_{-1}, d_{-2} и d_{-1} необходимо положить 0, поскольку в разложениях (7) участвуют лишь коэффициенты с неотрицательными номерами $m\geqslant 0$.

Асимптотическое поведение коэффициентов a_m, b_m и d_m с ростом номера m может быть исследовано на основе теории Пуанкаре–Перрона анализа полученных рекуррентных уравнений (8), (9) и (10) (см. [12]). Для наиболее медленно убывающих решений коэффициентов a_m, b_m и d_m это асимптотическое поведение зависит от параметра α следующим образом:

$$\frac{a_{m+1}}{a_m}\sim\frac{Q}{\sqrt{m}},\qquad \frac{b_{m+1}}{b_m}\sim\frac{Q}{\sqrt{m}},\qquad \frac{d_{m+1}}{d_m}\sim\frac{Q}{\sqrt{m}},\qquad m\to\infty,\qquad Q=(-ik\mathrm{R}\alpha)^{1/4},\qquad \alpha\neq 0,$$

$$\frac{a_{m+1}}{a_m}\sim\frac{Q}{m^{2/3}},\qquad \frac{b_{m+1}}{b_m}\sim\frac{Q}{m^{2/3}},\qquad \frac{d_{m+1}}{d_m}\sim\frac{Q}{m^{2/3}},\qquad m\to\infty,\qquad Q=(ik\mathrm{R})^{1/3},\qquad \alpha=0.$$

Такая скорость убывания коэффициентов a_m, b_m и d_m обеспечивает быструю сходимость используемых разложений (7) в соответствующих множествах $|z+1|\leqslant 1/2$, $|z|\leqslant 1/2$, $|z-1|\leqslant 1/2$.

Учет краевых условий (5) и (6) дает связь коэффициентов a_m и d_m, $m=0,1,2,3$:

$$a_0=0,\qquad a_3=-\frac{ik\mathrm{R}}{6}\big[c\,a_1\,+(1+\alpha)a_0\big], \tag{11}$$

$$d_0=0,\qquad d_3=\frac{ik\mathrm{R}}{6}\big[(2\beta-c)\,d_1\,-(\beta-2\alpha)d_0\big]. \tag{12}$$

Теперь построим линейно-независимые решения уравнения (4) в виде (7). Поскольку в точках $z=-1$ и $z=1$ задано по два условия (5) и (6), то в этих точках достаточно построить по два независимых решения, а в точке $z=0$ необходимо построить четыре независимых решения.

Сначала определим два решения $F_1(c;\,z)$ и $F_2(c;\,z)$ в виде разложений по степеням $(z+1)^m$. Для этого полагаем $a_0^{(1)}=0$ и задаем коэффициенты $a_1^{(1)}$ и $a_2^{(1)}$ решения $F_1(c;\,z)$ следующими:

$$a_1^{(1)}=1,\qquad a_2^{(1)}=0;$$

значение $a_3^{(1)}$ определим по (11), а все последующие $a_m^{(1)}$ вычислим по рекурсии (8), где полагаем $a_{-1}^{(1)}=a_{-2}^{(1)}=0$.

Для второго решения $F_2(c;\,z)$ полагаем $a_0^{(2)}=0$, коэффициенты $a_1^{(2)}$ и $a_2^{(2)}$ задаем следующими:

$$a_1^{(2)}=0,\qquad a_2^{(2)}=1,$$

а все последующие $a_m^{(2)}$ вычислим аналогично предыдущему.

Линейная комбинация $F_1(c;z)$ и $F_2(c;z)$ является общим решением $F(c;z)$ уравнения (4) с краевыми условиями (5) в точке $z=-1$:

$$F(c;z) = t_1F_1(c;z) + t_2F_2(c;z), \tag{13}$$

где t_1 и t_2 — произвольные коэффициенты.

Теперь построим четыре независимых решения $F_j(c;\,z)$, $j=\{3,4,5,6\}$, в виде разложений в точке $z=0$, задав по четыре первых коэффициента $b_0^{(j)}$, $b_1^{(j)}$, $b_2^{(j)}$ и $b_3^{(j)}$.

Для функции $F_3(c;\,z)$ полагаем

$$b_0^{(3)} = 1, \quad b_1^{(3)} = 0, \quad b_2^{(3)} = 0, \quad b_3^{(3)} = 0,$$

а все последующие коэффициенты $b_m^{(3)}$ вычислим по рекурсии (9), где полагаем $b_{-1}^{(3)} = b_{-2}^{(3)} = 0$.

Для функции $F_4(c;\,z)$ полагаем

$$b_0^{(4)} = 0, \qquad b_1^{(4)} = 1, \qquad b_2^{(4)} = 0, \qquad b_3^{(4)} = 0,$$

а все последующие коэффициенты $b_m^{(4)}$ вычислим аналогично предыдущему.

Для функции $F_5(c;\,z)$ полагаем

$$b_0^{(5)} = 0, \qquad b_1^{(5)} = 0, \qquad b_2^{(5)} = 1, \qquad b_3^{(5)} = 0,$$

а все последующие коэффициенты $b_m^{(5)}$ вычислим по рекурсии (9).

Наконец, для функции $F_6(c;\,z)$ полагаем

$$b_0^{(6)} = 0, \qquad b_1^{(6)} = 0, \qquad b_2^{(6)} = 0, \qquad b_3^{(6)} = 1,$$

а все последующие коэффициенты $b_m^{(6)}$ вычислим аналогично.

Линейная комбинация $F_j(c;z)$, $j=\{3,4,5,6\}$, также является общим решением $F(c;z)$ уравнения (4):

$$F(c;z) = t_3F_3(c;z) + t_4F_4(c;z) + t_5F_5(c;z) + t_6F_6(c;z), \tag{14}$$

где t_j, $j=\{3,4,5,6\}$, — произвольные коэффициенты.

Теперь построим два решения $F_7(c;\,z)$ и $F_8(c;\,z)$ в виде разложений по степеням $(z-1)^m$. Для этого, в соответствии с условием (12), полагаем $d_0^{(7)}=0$ и задаем коэффициенты $d_1^{(7)}$ и $d_2^{(7)}$ решения $F_7(c;\,z)$ следующими:

$$d_1^{(7)} = 1, \qquad d_2^{(7)} = 0;$$

значение $d_3^{(7)}$ определим по (12), а все последующие $d_m^{(7)}$ вычислим по рекурсии (10), где полагаем $d_{-1}^{(7)} = d_{-2}^{(7)} = 0$.

Для второго решения $F_8(c;\,z)$ полагаем $d_0^{(8)}=0$, коэффициенты $d_1^{(8)}$ и $d_2^{(8)}$ задаем следующими:

$$d_1^{(8)} = 0, \qquad d_2^{(8)} = 1,$$

а все последующие $d_m^{(8)}$ вычислим аналогично функции $F_7(c;\,z)$.

Линейная комбинация $F_7(c;z)$ и $F_8(c;z)$ является общим решением $F(c;z)$ уравнения (4) с краевыми условиями (6) в точке $z=1$:

$$F(c;z) = t_7F_7(c;z) + t_8F_8(c;z), \tag{15}$$

где t_7 и t_8 — произвольные коэффициенты.

Имея представления искомого решения $F(z)$ в трех окрестностях точек $z = -1$, $z = 0$ и $z = 1$, теперь сведем задачу нахождения СФ и соответствующих СЗ к гладкой сшивке построенных разложений в некоторых точках $z_* \in (-1, 1)$. Для этого потребуем совпадения функции (13) и ее трех производных в точке $z_{*,1} = -1/2$ с функцией (14) и ее тремя производными, т.е.

$$t_1 F_1^{(q)}(c; z_{*,1}) + t_2 F_2^{(q)}(c; z_{*,1}) - \sum_{m=3}^{6} t_m F_m^{(q)}(c; z_{*,1}) = 0, \qquad q = \{0, 1, 2, 3\}. \tag{16}$$

Аналогично этому запишем условие сшивки функции (14) в точке $z_{*,2} = 1/2$ с функцией (15):

$$\sum_{m=3}^{6} t_m F_m^{(q)}(c; z_{*,2}) - t_7 F_7^{(q)}(c; z_{*,2}) - t_8 F_8^{(q)}(c; z_{*,2}) = 0, \qquad q = \{0, 1, 2, 3\}. \tag{17}$$

Систему восьми уравнений (16), (17) относительно искомых весовых коэффициентов t_m, $m = \{1, 2, \dots, 8\}$, запишем в матричном виде

$$\mathbf{A}(c)\,\mathbf{t} = \mathbf{0}; \tag{18}$$

зависимость элементов матрицы $\mathbf{A}$ от спектрального параметра c здесь указана явно.

Первые четыре строки матрицы $\mathbf{A}(c)$ формируются в соответствии с системой уравнений (16):

$$A_{q+1,1} = F_1^{(q)}(c; z_{*,1}), \quad A_{q+1,2} = F_2^{(q)}(c; z_{*,1}), \quad A_{q+1,7} = A_{q+1,8} = 0,$$

$$A_{q+1,m} = -F_m^{(q)}(c; z_{*,1}), \quad q = \{0, 1, 2, 3\}, \qquad m = \{3, 4, 5, 6\}, \tag{19}$$

а вторые четыре строки — в соответствии с системой (17):

$$A_{q+5,1} = A_{q+5,2} = 0, \qquad A_{q+5,7} = -F_7^{(q)}(c; z_{*,2}), \qquad A_{q+5,8} = -F_8^{(q)}(c; z_{*,2}),$$

$$A_{q+5,m} = F_m^{(q)}(c; z_{*,2}), \qquad q = \{0, 1, 2, 3\}, \qquad m = \{3, 4, 5, 6\}. \tag{20}$$

Для нетривиального решения этой однородной системы (18) потребуем равенства нулю ее детерминанта

$$W(c) = \det(\mathbf{A}(c)) = 0, \tag{21}$$

что и определяет характеристическое уравнение для искомых собственных значений.

Решая уравнение (21) относительно величины c, получаем искомые СЗ — комплексные значения скорости бегущей волны, зависящие от всех параметров задачи (4)–(6). Найденные при этом коэффициенты t_m, $m = \{1, 2, \dots, 8\}$, позволяют (с точностью до произвольного множителя) найти соответствующую СФ $F(c; z)$ в виде комбинации разложений (13)–(15).

Решение уравнения (21) будем строить с помощью итерационного метода Ньютона:

$$c^{(l+1)} = c^{(l)} - \frac{W(c^{(l)})}{W'(c^{(l)})}, \quad l = 0, 1, \dots, \tag{22}$$

а начальные приближения $c^{(0)}$ будем брать на основе метода продолжения по параметру k и из полученных далее асимптотических разложений для СЗ при $k \to 0$.

Необходимая для метода Ньютона производная $W'(c^{(l)})$ находилась с помощью явного дифференцирования по спектральному параметру "c", разложений для всех производных $F_m^{(q)}(c; z_{*,1})$, $F_m^{(q)}(c; z_{*,2})$, $m = \{1, 2, \dots, 8\}$, $q = \{0, 1, 2, 3\}$, что позволило избежать использования

конечно-разностной производной $\big(W(c+\Delta c)-W(c)\big)/\Delta c$ и связанной с ней погрешности при малых $|\Delta c|$.

2. АСИМПТОТИКА СФ И СЗ ПРИ $k \to 0$

Построим асимптотическое разложение СФ и СЗ при $k \to 0$ для ненулевых параметров R, Bu, Pr и $n \in \mathbb{N}$. Метод такого построения описан, например, в [8] и применен в [9] для задачи нахождения СЗ для течений типа Куэтта или Пуазейля и в задаче с краевыми условиями $F''(\pm 1) = 0$, соответствующими равенству нулю потоков плавучести на горизонтальных границах слоя. При этих условиях возникали два конечных СЗ при $k \to 0$ и счетное число неограниченно растущих СЗ.

В настоящей работе для краевой задачи с условиями на функцию $F(\pm 1) = 0$, в отличие от результатов исследования в [9], ограниченных СЗ при $k \to 0$ не возникает, и имеется лишь счетное множество СЗ c_m, у которых мнимая часть отрицательна и по модулю неограниченно возрастает.

2.1. Вывод характеристического соотношения

Полагая все величины в задаче, кроме k, фиксированными, представим искомую СФ $F(z)$ и соответствующее ей СЗ "c", в виде разложений по степеням параметра $ik\mathrm{R}$:

$$F(z) = \varphi_0(z) + ik\mathrm{R}\,\varphi_1(z) + \cdots, \qquad c = \frac{\chi_{-1}}{ik\mathrm{R}} + \chi_0 + ik\mathrm{R}\,\chi_1 + \cdots, \qquad k \to 0. \tag{23}$$

Это приводит к цепочке краевых задач относительно функций $\varphi_m(z)$ и величин χ_{m-1}, $m = 0, 1, \ldots$. Первая из них, для функции $\varphi_0(z)$ и величины χ_{-1}, имеет вид

$$\varphi_0''''(z) + \left(\chi_{-1} - \lambda^2\right)\varphi_0''(z) - \chi_{-1}\frac{\lambda^2}{\mathrm{Pr}}\varphi_0(z) = 0, \qquad z \in [-1,\,1], \tag{24}$$

$$\varphi_0'''(-1) = -\chi_{-1}\varphi_0'(-1), \quad \varphi_0(-1) = 0, \qquad \varphi_0'''(1) = -\chi_{-1}\varphi_0'(1), \quad \varphi_0(1) = 0, \tag{25}$$

где обозначено

$$\lambda = \pi n\sqrt{\mathrm{Pr}\,\mathrm{Bu}}. \tag{26}$$

Уравнение (24) содержит производные степени 0, 2 и 4, а краевые условия (25) обладают четностью, поэтому решение $\varphi_0(z)$ этой задачи может быть либо четным, либо нечетным. Эти решения представим, соответственно, в виде

$$\varphi_{0,ev}(z) = A_1\cosh(\omega_1 z) + A_2\cosh(\omega_2 z), \qquad \varphi_{0,od}(z) = B_1\sinh(\omega_1 z) + B_2\sinh(\omega_2 z). \tag{27}$$

Подстановка любого из представлений (27) в систему (24)–(26) приводит к характеристическому уравнению для искомых величин ω_1 и ω_2, которые являются корнями биквадратного уравнения

$$\omega^4 + \left(\chi_{-1} - \lambda^2\right)\omega^2 - \chi_{-1}\frac{\lambda^2}{\mathrm{Pr}} = 0. \tag{28}$$

Четыре корня этого уравнения расположены двумя парами так, что в каждой паре корни отличаются лишь знаком. Некратные корни ω_1^2 и ω_2^2 этого уравнения возникают при условии неравенства нулю его дискриминанта, т.е.

$$\left(\chi_{-1} - \lambda^2\right)^2 + 4\chi_{-1}\frac{\lambda^2}{\mathrm{Pr}} \neq 0; \tag{29}$$

в дальнейшем рассмотрении ограничимся только этим условием.

Решение уравнения (28) запишем относительно величин ω_1^2 и ω_2^2:

$$\omega_{1,2}^2 = \frac{1}{2}\left[\lambda^2 - \chi_{-1} \pm \sqrt{\left(\lambda^2 - \chi_{-1}\right)^2 + 4\chi_{-1}\frac{\lambda^2}{\mathrm{Pr}}}\,\right]. \tag{30}$$

В последующем исследовании задачи (24), (25) представим решение $\varphi_0(z)$ в виде четной или нечетной функции, чему соответствует два следующих случая.

2.2. Четные решения $\varphi_0(z)$

Представляя решение $\varphi_{0,ev}(z)$ задачи (24), (25) в виде (27):

$$\varphi_{0,ev}(z) = A_1 \cosh(\omega_1 z) + A_2 \cosh(\omega_2 z), \tag{31}$$

получаем, что искомая величина χ_{-1} удовлетворяет трансцендентному уравнению

$$\omega_2\left(\omega_2^2 + \chi_{-1}\right) \cosh\omega_1 \sinh\omega_2 = \omega_1\left(\omega_1^2 + \chi_{-1}\right) \cosh\omega_2 \sinh\omega_1. \tag{32}$$

В частном случае $\mathrm{Pr} = 1$ находим из (30) значения $\omega_1^2 = \lambda^2$ и $\omega_2^2 = -\chi_{-1}$, подстановка которых в уравнение (32) дает соотношение $\cosh\sqrt{-\chi_{-1}} = 0$. Решение этого уравнения дает возможность получить явный вид счетного множества искомых значений $\chi_{-1,m}$:

$$\chi_{-1,m} = \pi^2\left(\frac{1}{2} + m\right)^2, \quad m = 0, 1, \ldots, \qquad \mathrm{Pr} = 1. \tag{33}$$

Теперь построим представление для величин $\chi_{-1,m}$ при малых отклонениях числа Прандля Pr от единицы. Будем искать значения $\chi_{-1,m}$ в виде

$$\chi_{-1,m} = \pi^2\left(\frac{1}{2} + m\right)^2 (1 + \Delta_m^{(ev)}). \tag{34}$$

Подстановка этого представления в уравнение (32) и учет соотношений (30) позволяет для величины $\Delta_m^{(ev)}$ построить разложение по степеням $(\mathrm{Pr} - 1)^q$, что дает

$$\Delta_m^{(ev)} = \frac{\lambda}{\lambda^2 + \pi^2(m + 1/2)^2}\left[\lambda + \frac{2\pi^2(m + 1/2)^2 \operatorname{cth}\lambda}{\lambda^2 + \pi^2(m + 1/2)^2}\right](\mathrm{Pr} - 1) + O\left((\mathrm{Pr} - 1)^2\right), \quad m = 0, 1, \ldots. \tag{35}$$

2.3. Нечетные решения $\varphi_{0,od}(z)$

Представляя решение $\varphi_{0,od}(z)$ задачи (24), (25) в виде (27):

$$\varphi_{0,od}(z) = B_1 \sinh(\omega_1 z) + B_2 \sinh(\omega_2 z), \tag{36}$$

получаем, что искомая величина χ_{-1} удовлетворяет трансцендентному уравнению

$$\omega_1\left(\omega_1^2 + \chi_{-1}\right) \cosh\omega_1 \sinh\omega_2 = \omega_2\left(\omega_2^2 + \chi_{-1}\right) \cosh\omega_2 \sinh\omega_1. \tag{37}$$

В частном случае $\mathrm{Pr} = 1$ находим из (30) значения $\omega_1^2 = \lambda^2$ и $\omega_2^2 = -\chi_{-1}$, подстановка которых в уравнение (37) дает соотношение $\sinh\sqrt{-\chi_{-1}} = 0$. Решение этого уравнения дает возможность получить явный вид счетного множества искомых значений $\chi_{-1,m}$:

$$\chi_{-1,m} = \pi^2 m^2, \quad m = 1, 2, \ldots, \qquad \mathrm{Pr} = 1. \tag{38}$$

Теперь построим представление для величин $\chi_{-1,m}$ при малых отклонениях числа Прандля Pr от единицы. Будем искать значения $\chi_{-1,m}$ в виде, аналогичном (34):

$$\chi_{-1,m} = \pi^2 m^2 (1 + \Delta_m^{(od)}). \tag{39}$$

Подстановка этого представления в уравнение (37) и учет соотношений (30) позволяет для величины $\Delta_m^{(od)}$ построить разложение по степеням $(\mathrm{Pr}-1)^q$, что дает

$$\Delta_m^{(od)} = \frac{\lambda}{\lambda^2+\pi^2m^2}\left[\lambda + \frac{2\pi^2m^2\,\mathrm{th}\,\lambda}{\lambda^2+\pi^2m^2}\right](\mathrm{Pr}-1) + O\Big((\mathrm{Pr}-1)^2\Big), \quad m = 1, 2, \ldots . \tag{40}$$

Завершая рассмотрение СФ и СЗ на основе асимптотики (23) при $k \to 0$ и результатов (34), (35) и (39), (40), убеждаемся, что все полученные СЗ имеют большие отрицательные мнимые части и соответствуют устойчивости соответствующих длинноволновых возмущений при достаточно малых k.

3. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

На основе полученных асимптотических разложений при $k \to 0$ для неограниченно растущих СЗ был разработан и протестирован алгоритм расчета СЗ c_m для заданного номера m, продемонстрировавший высокую точность результатов. Анализ эффективности этого алгоритма показал, что использование трех окрестностей разложений СФ $F(z)$ с центрами в точках $z = -1$, $z = 0$, $z = 1$ и сшивкой в точках $z_{*,1} = -1/2$ и $z_{*,2} = 1/2$ дало ускорение работы алгоритма в среднем в 7 раз по сравнению с использованием лишь двух разложений в точках $z = -1$ и $z = 1$ и сшивкой в точке $z_* = 0$. При этом относительная точность вычислений СЗ и СФ была фиксирована и достигала 10^{-20}.

Дополнительной гарантией того, что в плоскости комплексного спектрального параметра "c", не было "пропущено", ни одного СЗ исследуемой задачи, служило использование обобщенного принципа аргумента для функции $W(c)$ из (21). Поскольку функция $W(c)$ является детерминантом матрицы, все элементы которой являются значениями производных искомого решения $F(c; z)$ с различными краевыми условиями, а спектральный параметр "c", входит в эти разложения лишь степенным образом, то функция $W(c)$ не имеет конечных полюсов. Поэтому для локализации нулей функции $W(c)$ в некоторой области $\mathfrak{D}$ справедлив обобщенный принцип аргумента (см. [13]):

$$M = \frac{1}{2\pi i}\oint_{\partial\mathfrak{D}} \frac{W'(c)}{W(c)}\,dc, \qquad \sum_{j=1}^{M} c_j = \frac{1}{2\pi i}\oint_{\partial\mathfrak{D}} c\,\frac{W'(c)}{W(c)}\,dc, \tag{41}$$

где M — число комплексных нулей функции $W(c)$ внутри области $\mathfrak{D}$, а $\sum_{j=1}^{M} c_j$ — сумма координат этих нулей.

Далее опишем результаты численного анализа СЗ при различных физических параметрах задачи. Расчет траекторий $c_m(k)$ при изменении волнового числа k начинался с малых чисел k, а начальными приближениями в итерационном методе Ньютона при этом служили полученные в пп. 2.2 и 2.3 асимптотики СЗ. Затем при непрерывном возрастании k использовался метод продолжения по параметру k и метод Ньютона. При быстром изменении СЗ шаг по переменной k соответственно уменьшался, а при медленном изменении СЗ шаг по k адаптировался так, чтобы зависимость $c_m(k)$ была достаточно непрерывной.

Нумерация СЗ $c_m(k)$ проводилась так, что при $k \to 0$ все СЗ упорядочивались в соответствии с ростом модуля мнимой части $\mathrm{Im}(c_m(k))$. Учет асимптотики (23) и зависимостей (34), (35), (39), (40) позволяет расположить $c_m(k)$ следующим образом: СЗ $c_1(k)$ соответствует первой четной СФ с асимптотикой СЗ (34), (35) для $m = 0$; СЗ $c_2(k)$ соответствует первой нечетной СФ с асимптотикой СЗ (39), (40) для $m = 1$; $c_3(k)$ соответствует второй четной СФ с асимптотикой СЗ (34), (35) для $m = 1$; $c_4(k)$ соответствует асимптотике второй нечетной СФ с асимптотикой СЗ (39), (40) для $m = 2$, и т.д.

При возрастании волнового числа k от бесконечно малых положительных значений (отметим, что $k = 0$ мы не можем положить в силу нефизичности получаемого уравнения) все СЗ на комплексной плоскости выходят из особой точки $c = \infty$ с мнимой частью $\mathrm{Im}(c) = -\infty$ в соответствии с асимптотикой (23), (34), (35) либо (39), (40).

Далее, с ростом числа k, все СЗ поднимаются к вещественной оси, некоторые из них могут попадать в верхнюю полуплоскость $\mathrm{Im}(c) > 0$, что соответствует неустойчивости исследуемого течения.

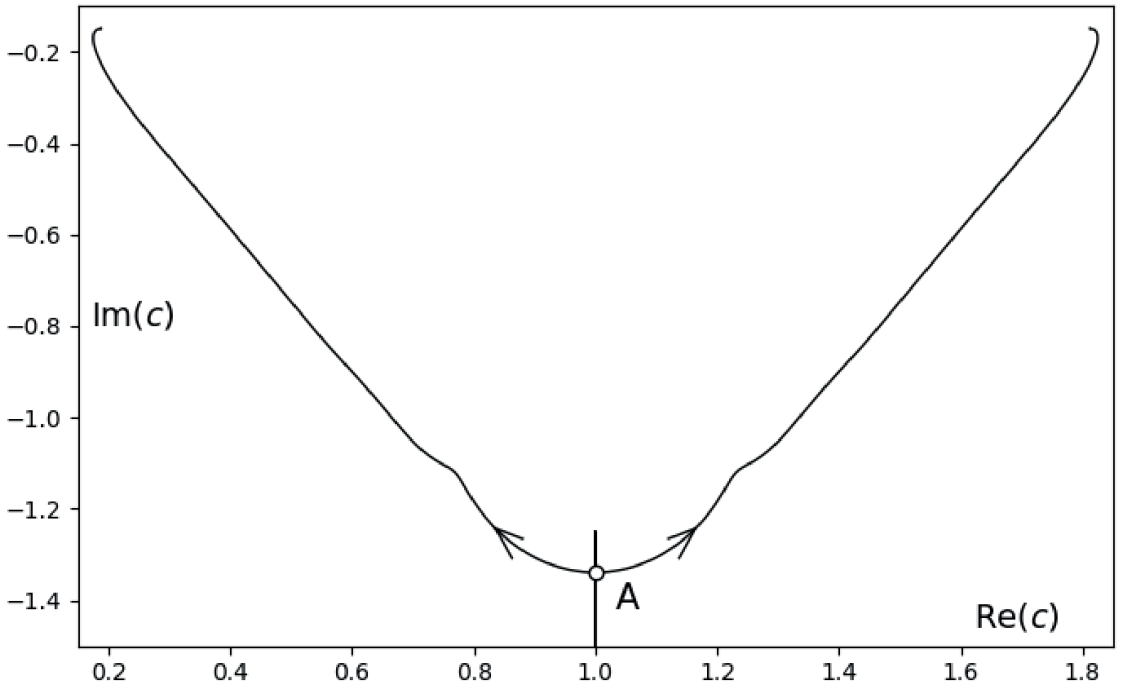
Затем эти СЗ переходят в нижнюю полуплоскость $\mathrm{Im}(c) < 0$, и далее с ростом k они расходятся направо или налево довольно сложным образом.

В частном случае линейного профиля (2) основного течения $U(z)$, т.е. когда параметр $\alpha = 0$, все СЗ, как было отмечено во Введении, располагаются на комплексной плоскости симметрично относительно оси $\mathrm{Re}(c) = 1$. Опишем динамику нескольких первых СЗ для параметров задачи $\alpha = 0$, $\mathrm{R} = 10$, $\mathrm{Pr} = 4$, $\mathrm{Bu} = 0.0001$, $n = 1$.

При возрастании числа k все СЗ поднимаются по прямой $\mathrm{Re}(c) = 1$ из точки $\mathrm{Im}(c) = -\infty$ в соответствии с асимптотикой (34), (35) либо (39), (40). Первое СЗ $c_1(k)$ при значении $k \approx 1.03$ попадает в верхнюю полуплоскость $\mathrm{Im}(c) > 0$ и в интервале значений $k \in (1.03, 112.6)$ остается в этой области, соответствующей неустойчивости исследуемого течения.

Динамика всех СЗ при росте числа k оказывается весьма сложной и заслуживает детального описания.

При значении $k \approx 2.56$ СЗ $c_3(k)$, находясь на прямой $\mathrm{Re}(c) = 1$, останавливается, затем движется вниз, а при $k_* \approx 2.945$ оно сталкивается с СЗ $c_4(k)$ и образует первое двойное СЗ $c_3(k_*) = c_4(k_*) \approx \approx 1 - 1.3387\,i$. При дальнейшем росте числа k это СЗ распадается на два простых СЗ, симметричных относительно прямой $\mathrm{Re}(c) = 1$, движущихся в нижней полуплоскости $\mathrm{Im}(c) < 0$ направо и налево и стремящихся при $k \to +\infty$ к предельным точкам $c = 0$ и $c = 2$. На фиг. 1. даны траектории этих СЗ при возрастании k на отрезке $[2.56, 2500]$; точка A соответствует первому двойному СЗ при $k_* \approx 2.945$, а стрелочки показывают направления изменения этих СЗ при росте значения k.



Фиг. 1. Образование первого двойного СЗ задачи с параметрами $\alpha = 0$, $\mathrm{R} = 10$, $\mathrm{Pr} = 4$, $\mathrm{Bu} = 0.0001$, $n = 1$ из $c_3(k)$ и $c_4(k)$ и последующая эволюция этих СЗ.

Второе двойное СЗ образуется аналогичным образом из СЗ $c_7(k)$ и $c_8(k)$ при $k_* \approx 14.5$, $c_7(k_*) = c_8(k_*) \approx 1 - 1.028\,i$. При дальнейшем росте числа k это СЗ распадается на два простых СЗ, движущихся в полуплоскости $\mathrm{Im}(c) < 0$ к предельным точкам $c = 0$ и $c = 2$, аналогично показанным траекториям СЗ на фиг. 1.

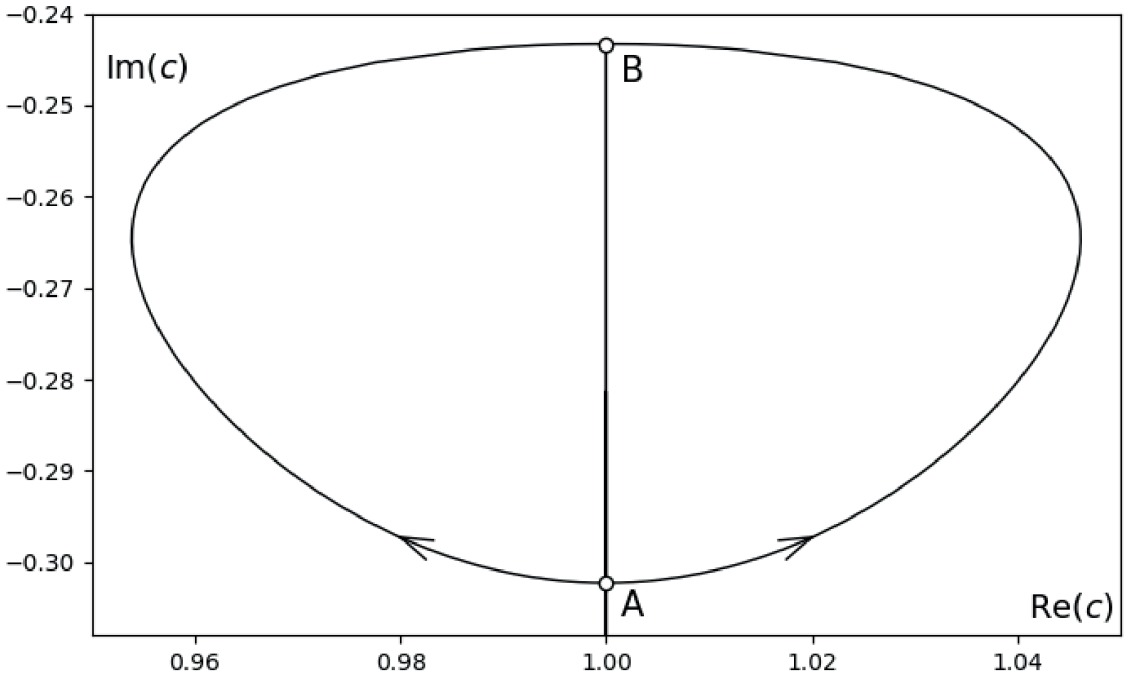
Третье двойное СЗ образуется аналогичным образом из СЗ $c_2(k)$ и $c_5(k)$ при $k_* \approx 15.68$, $c_2(k_*) = = c_5(k_*) \approx 1 - 0.145\,i$. При дальнейшем росте числа k это СЗ распадается на два простых симметричных СЗ. При изменении k в интервале $(162, 459)$ оба этих СЗ попадают в верхнюю полуплоскость

$\mathrm{Im}(c) > 0$, а затем возвращаются в полуплоскость $\mathrm{Im}(c) < 0$ и движутся к предельным точкам $c = 0$ и $c = 2$.

Четвертое двойное СЗ образуется аналогичным образом из СЗ $c_{11}(k)$ и $c_{12}(k)$ при $k_* \approx 36.2$, $c_{11}(k_*) = c_{12}(k_*) \approx 1 - 0.902\,i$. При дальнейшем росте числа k это СЗ распадается на два простых СЗ, движущихся в полуплоскости $\mathrm{Im}(c) < 0$ к предельным точкам $c = 0$ и $c = 2$.

Пятое двойное СЗ образуется аналогичным образом из СЗ $c_6(k)$ и $c_9(k)$ при $k_* \approx 42.3$, $c_6(k_*) = c_9(k_*) \approx 1 - 0.266\,i$. При дальнейшем росте числа k это СЗ распадается на два простых СЗ, движущихся в полуплоскости $\mathrm{Im}(c) < 0$ к предельным точкам $c = 0$ и $c = 2$.

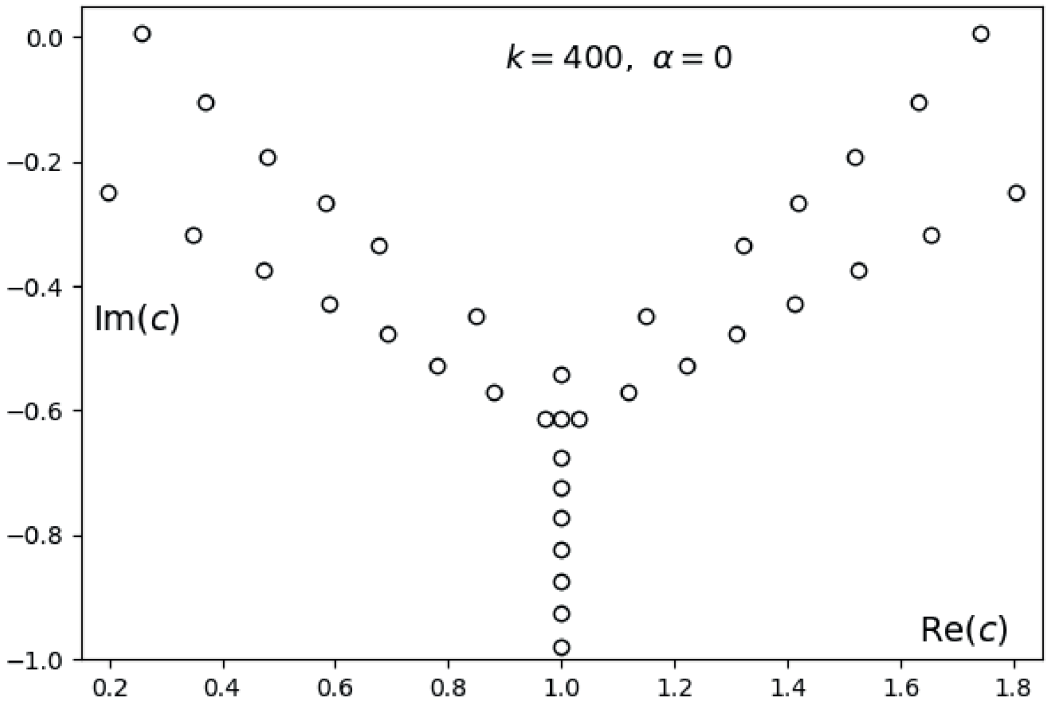
Шестое двойное СЗ образуется аналогичным образом из СЗ $c_{10}(k)$ и $c_{13}(k)$ при $k_* \approx 85.4$, $c_{10}(k_*) = c_{13}(k_*) \approx 1 - 0.302\,i$. Однако при дальнейшем росте числа k траектории этих СЗ отличаются от представленных на фиг. 1. При $k > 85.4$ это СЗ распадается на два простых СЗ, движущихся по симметричным дугам, а при значении $k_* \approx 106.75$ эти два СЗ опять сталкиваются в точке $c_{10}(k_*) = c_{13}(k_*) \approx 1 - 0.243\,i$ и образуют седьмое двойное СЗ. На фиг. 2 даны траектории этих двух СЗ при возрастании k на отрезке $[85.4, 106.75]$; точке A соответствует двойное СЗ при $k_* \approx 85.4$, а точке B – двойное СЗ при $k_* \approx 106.75$. С ростом k это двойное СЗ опять распадается на два простых, которые расходятся вверх и вниз по прямой $\mathrm{Re}(c) = 1$. Одно из этих СЗ, которое идет вниз, при значении $k_* \approx 109.66$ сталкивается с СЗ $c_{14}(k)$ и образует очередное восьмое двойное СЗ. А СЗ, образовавшееся после распада седьмого двойного СЗ и которое идет вверх, при значении $k_* \approx 113.2$ наконец сталкивается с СЗ $c_1(k)$ и образует девятое двойное СЗ $c_1(k_*) \approx 1 - 0.052\,i$. При последующем росте k это двойное СЗ распадается на два простых симметричных СЗ, движущихся в полуплоскости $\mathrm{Im}(c) < 0$ к предельным точкам $c = 0$ и $c = 2$, аналогично показанным траекториям на фиг.1.



Фиг. 2. Образование двух двойных СЗ задачи с параметрами $\alpha = 0$, $\mathrm{R} = 10$, $\mathrm{Pr} = 4$, $\mathrm{Bu} = 0.0001$, $n = 1$ из $c_{10}(k)$ и $c_{13}(k)$ при $k \in [85.4, 106.75]$.

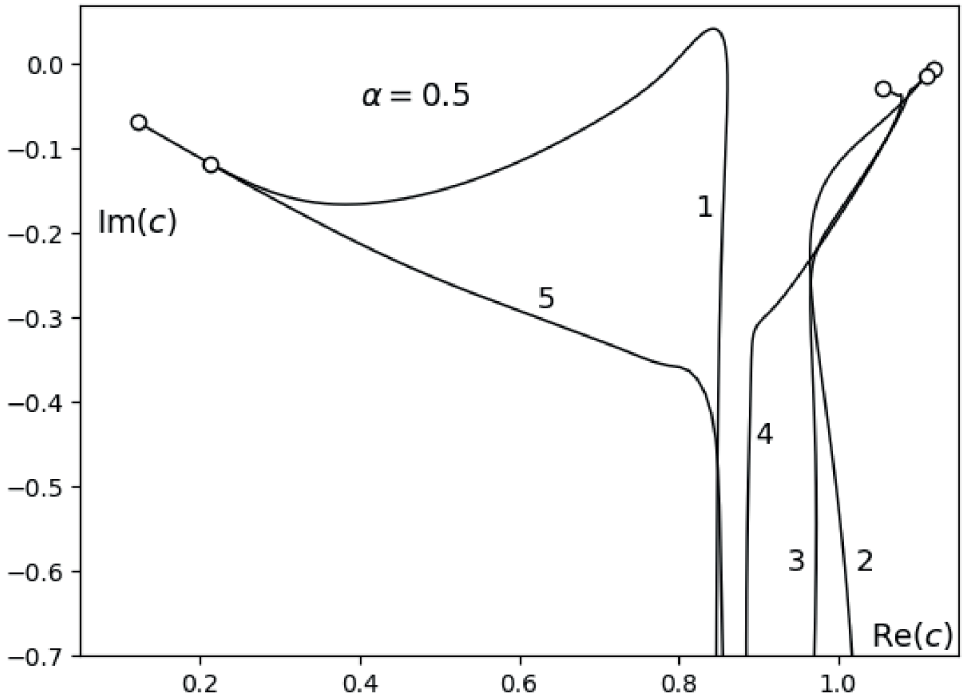
На фиг. 3 приведены первые 37 СЗ этой задачи при значении волнового числа $k = 400$, при этом лишь $c_2(k)$ и $c_5(k)$ расположены в верхней полуплоскости. Такое симметричное расположение СЗ задачи Орра–Зоммерфельда часто называют портретом “спектрального галстука”, (см., например, [14], [15]).

Если профиль скорости (2) основного течения $U(z)$ не является линейным, т.е. параметр $\alpha > 0$, то множество СЗ уже не обладает симметрией относительно прямой $\mathrm{Re}(c) = 1$, и при возрастании числа k образования двойных СЗ при проведенных расчетах уже не выявлено.



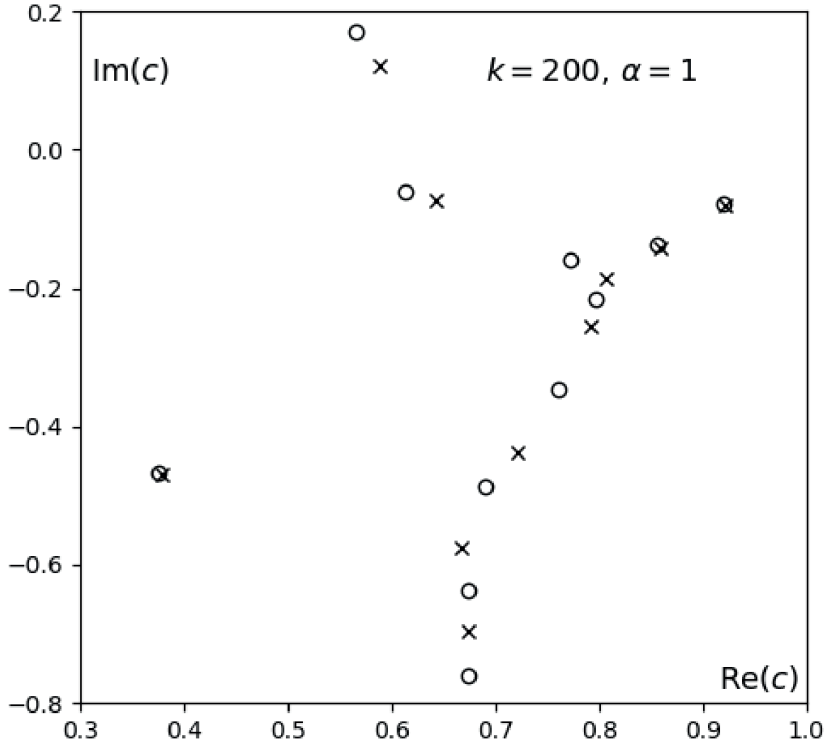
Фиг. 3. Комплексные СЗ задачи для параметров R = 10, Pr = 4, Bu = 0.0001, $n = 1$.

На фиг. 4 приведены траектории первых пяти СЗ для параметров $\alpha = 0.5$, R = 10, Pr = 1, Bu = 0.1, $n = 1$ при непрерывном возрастании k в интервале $(0, 100)$; здесь показаны участки траекторий $c_m(k)$ в области $\mathrm{Im}\, c > -0.7$. Цифры 1–5 указывают номера соответствующих траекторий, а кружочки — положения этих СЗ при $k = 100$. Отметим, что СЗ $c_1(k)$ при значениях $k \in (2.14, 5.95)$ расположено в верхней полуплоскости, что соответствует неустойчивости течения.



Фиг. 4. Траектории СЗ задачи для параметров $\alpha = 0.5$, R = 10, Pr = 1, Bu = 0.1, $n = 1$.

На фиг. 5 показаны первые 21 СЗ задачи для параметров $\alpha = 1$, $R = 10$, $Pr = 4$, $Bu = 0.0001$, $n = 1$, $k = 200$. Кружочками отмечены СЗ, соответствующие четным СФ, а крестиками – СЗ для нечетных СФ. Необходимо отметить, что СЗ при $\mathrm{Re}(c) > 0.9$ и при $\mathrm{Re}(c) < 0.4$ расположены на этой фигуре парами, так что в каждой паре расстояние между СЗ составляет порядка 10^{-3}, а с ростом значения волнового числа k это расстояние еще уменьшается.



Фиг. 5. Комплексные СЗ задачи для параметров $R = 10$, $Pr = 4$, $Bu = 0.0001$, $n = 1$.

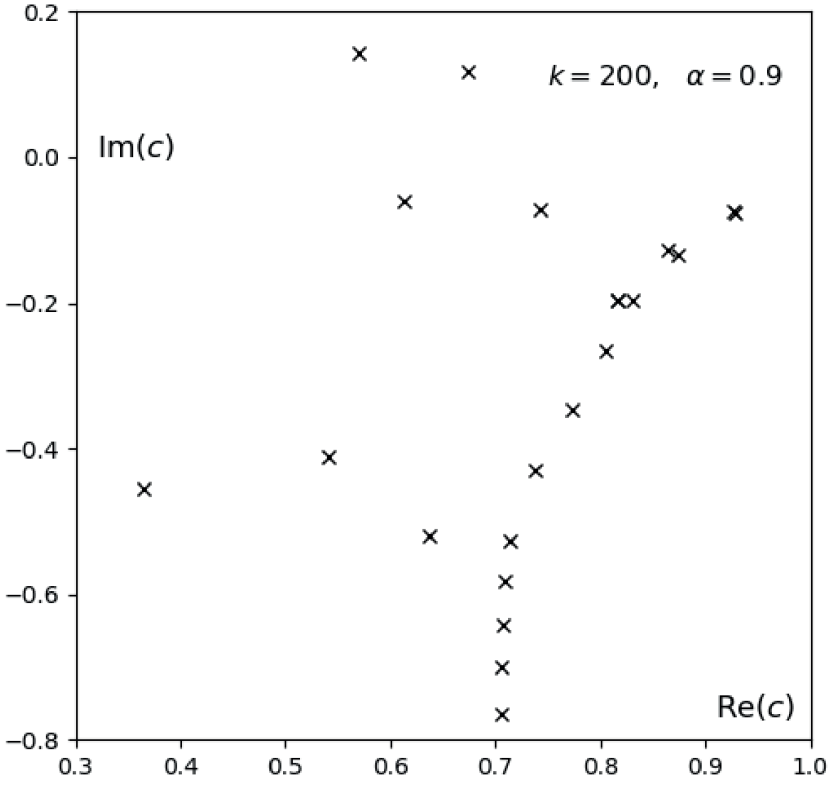
При $\alpha \neq 1$ четных и нечетных СФ уже нет, и все СЗ перестраиваются весьма сложным образом. На фиг. 6 эти СЗ показаны крестиками для $\alpha = 0.9$ и с остальными параметрами такими же, как на фиг. 5. Отметим, что расстояние порядка 10^{-3} между двумя СЗ здесь сохраняется только для двух правых СЗ при $\mathrm{Re}(c) > 0.9$.

На фиг. 7 представлены СЗ для $\alpha = 0.5$ и с остальными параметрами такими же, как на фиг. 5 и фиг. 6.

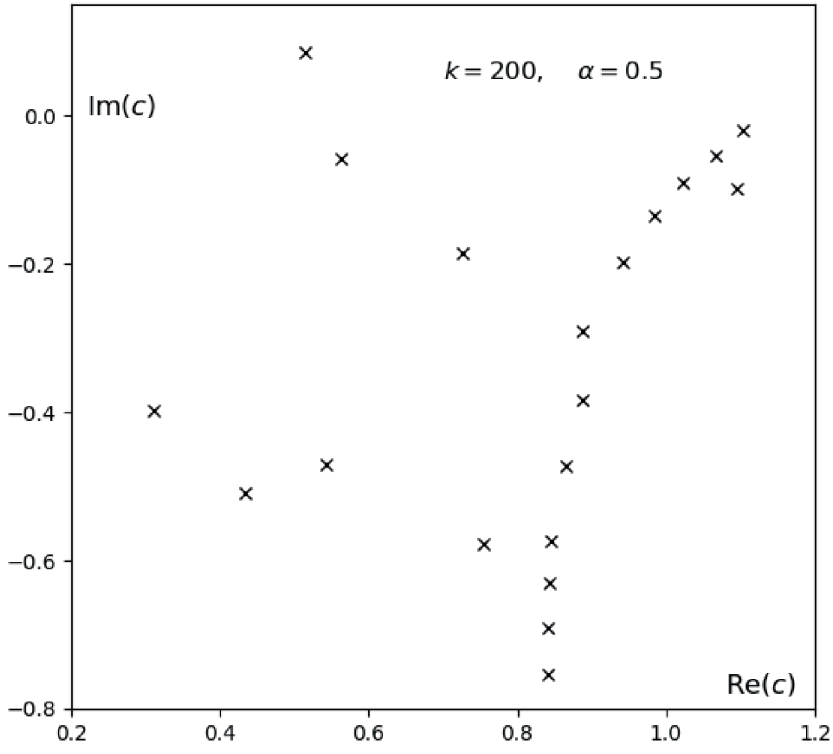
При дальнейшем уменьшении величины α такая перестройка СЗ становится еще более заметной, и при значении $\alpha = 0$ все СЗ выстраиваются симметричным образом, образуя портрет “спектрального галстука”, аналогичного показанному на фиг. 3.

ЗАКЛЮЧЕНИЕ

Разработан аналитико-численный метод решения спектральной несамосопряженной задачи, описывающей малые возмущения океанских течений с вертикальным параболическим профилем скорости общего вида. Модель основана на уравнении потенциального вихря в квазигеострофическом приближении с учетом вертикальной диффузии массы и импульса. Расчет собственных функций и собственных значений задачи проводился с помощью покрытия исходного отрезка интегрирования $z \in [-1, 1]$ системой окрестностей и использования степенных разложений в каждой окрестности с последующей гладкой сшивкой. Построенные асимптотические разложения для СЗ при малых волновых числах k являлись начальными приближениями в итерационном методе Ньютона с использованием метода продолжения. Показано, что в случае линейного профиля скорости $U(z)$ основного потока при определенных числах k_* возникают двойные СЗ на прямой $\mathrm{Re}(c) = 1$.



Фиг. 6. Комплексные СЗ задачи для параметров R = 10, Pr = 4, Bu = 0.0001, $n = 1$.



Фиг. 7. Комплексные СЗ задачи для параметров R = 10, Pr = 4, Bu = 0.0001, $n = 1$.

Предложенный метод показал свою высокую эффективность как по точности, так и по быстродействию. Относительная погрешность во всех расчетах не превышала величины 10^{-20}, а скорость вычислений оказалась в среднем в 7 раз выше скорости метода разложения решения лишь в двух граничных точках $z = -1$ и $z = 1$ (см. [3]–[10]). Это объясняется наличием множителя 2^{-m}, возникающего

в разложениях при покрытии отрезка $z \in [-1, 1]$ тремя окрестностями, по сравнению с множителем 1, участвующим в разложениях при использовании всего лишь двух окрестностей.

С помощью разработанного подхода рассчитаны траектории комплексных СЗ для различных физических параметров задачи при изменении волнового числа k. Кратко представлена сложная картина возникновения неустойчивости моделируемого течения и дано подробное описание поведения СЗ при возрастании волнового числа k.

СПИСОК ЛИТЕРАТУРЫ

1. *Кузьмина Н. П.* Об одной гипотезе образования крупномасштабных интрузий в Арктическом бассейне // Фундамент. и прикл. гидрофизика. 2016. Т. 9. № 2. С. 15–26.
2. *Kuzmina N. P.* Generation of large-scale intrusions at baroclinic fronts: An analytical consideration with a reference to the Arctic ocean // Ocean Sci. 2016. V. 12. P. 1269–1277. https://doi.org/10.5194/os-12-1269-2016
3. *Кузьмина Н. П., Скороходов С. Л., Журбас Н. В., Лыжков Д. А.* О неустойчивости геострофического течения с линейным вертикальным сдвигом скорости на масштабах интрузионного расслоения // Изв. РАН. Физика атмосферы и океана. 2018. Т. 54. № 1. С. 54–63.
4. *Кузьмина Н. П., Скороходов С. Л., Журбас Н. В., Лыжков Д. А.* Описание возмущений океанских геострофических течений с линейным вертикальным сдвигом скорости с учетом трения и диффузии плавучести // Изв. РАН. Физика атмосферы и океана. 2019. Т. 55. № 2. С. 73–85.
5. *Скороходов С. Л., Кузьмина Н. П.* Аналитико-численный метод решения задачи типа Орра–Зоммерфельда для анализа неустойчивости течений в океане // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 6. С. 1022–1039.
6. *Скороходов С. Л., Кузьмина Н. П.* Спектральный анализ модельных течений типа Куэтта применительно к океану // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 5. С. 867–888.
7. *Кузьмина Н. П., Скороходов С. Л., Журбас Н. В., Лыжков Д. А.* О влиянии трения и диффузии плавучести на динамику геострофических океанских течений с линейным вертикальным профилем скорости // Изв. РАН. Физика атмосферы и океана. 2020. Т. 56. № 6. С. 676–688.
8. *Скороходов С. Л., Кузьмина Н. П.* Спектральный анализ малых возмущений геострофических течений с параболическим вертикальным профилем скорости применительно к океану // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 12. С. 2010–2023.
9. *Скороходов С. Л., Кузьмина Н. П.* Аналитико-численный метод для анализа малых возмущений океанских геострофических течений с параболическим вертикальным профилем скорости общего вида // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 12. С. 2043–2053.
10. *Кузьмина Н. П., Скороходов С. Л., Журбас Н. В., Лыжков Д. А.* О видах неустойчивости геострофического течения с вертикальным параболическим профилем скорости // Изв. РАН. Физика атмосферы и океана. 2023. Т. 59. № 3. С. 1–10.
11. *Голубев В. В.* Лекции по аналитической теории дифференциальных уравнений. М.: ГИТТЛ, 1950.
12. *Гельфонд А. О.* Исчисление конечных разностей. М.: Наука, 1967.
13. *Лаврентьев М. А., Шабат Б. В.* Методы теории функций комплексного переменного. М.: Наука, 1973.
14. *Шкаликов А. А.* Спектральные портреты оператора Орра–Зоммерфельда при больших числах Рейнольдса // Современная математика. Фундаментальные направления. 2003. Т. 3. С. 89–112.
15. *Скороходов С. Л.* Численный анализ спектра задачи Орра–Зоммерфельда // Ж. вычисл. матем. и матем. физ. 2007. Т. 47. № 10. С. 1672–1691.

ANALYTICAL-NUMERICAL METHOD FOR SOLVING THE SPECTRAL PROBLEM IN A MODEL OF GEOSTROPHIC OCEAN CURRENTS

S. L. Skorokhodov[a,*] and N. P. Kuzmina[b,**]

[a] *Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, 119991 Russia*
[b] *Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, 117997 Russia*

[*] *e-mail: sskorokhodov@gmail.com*
[**] *e-mail: kuzmina@ocean.ru*

Received January 12, 2024
Revised February 9, 2024
Accepted February 15, 2024

Abstract. A new efficient analytical-numerical method is developed for solving a problem for the potential vorticity equation in the quasi-geostrophic approximation with allowance for vertical diffusion of mass and momentum. The method is used to analyze small perturbations of ocean currents of finite transverse scale with a general parabolic vertical profile of velocity. For the arising spectral nonself-adjoint problem, asymptotic expansions of the eigenfunctions and eigenvalues are constructed for small wave numbers and the existence of a countable set of complex eigenvalues with an unboundedly decreasing imaginary part is shown. On the integration interval , a system of three neighborhoods is introduced and a solution in each of them is constructed in the form of power series expansions, which are matched smoothly, so that the eigenfunctions and eigenvalues are efficiently calculated with high accuracy. For a varying wave number, the trajectories of complex eigenvalues are computed for various parameters of the problem and the existence of double eigenvalues is shown. The complex picture of instability developing in the simulated flow depending on physical parameters of the problem is briefly described.

Keywords: spectral non-self-adjoint problem, asymptotic expansions, high-accuracy numerical method, double eigenvalues.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1008–1015

УДК 517.911.5+517.927

СУЩЕСТВОВАНИЕ РЕШЕНИЙ НЕСАМОСОПРЯЖЕННОЙ ЗАДАЧИ ШТУРМА–ЛИУВИЛЛЯ С РАЗРЫВНОЙ НЕЛИНЕЙНОСТЬЮ[1)]

© 2024 г. О. В. Басков[1], Д. К. Потапов[1,*]

[1] *199034 Санкт-Петербург, Университетская наб., 7/9, Санкт-Петербургский государственный университет, Россия*

[*] *e-mail: d.potapov@spbu.ru*

Поступила в редакцию 20.12.2023 г.
Переработанный вариант 20.12.2023 г.
Принята к публикации 06.03.2024 г.

Рассматривается проблема существования решений задачи Штурма–Лиувилля с несамосопряженным дифференциальным оператором и разрывной по фазовой переменной нелинейностью. Для исследуемой задачи устанавливаются теоремы о существовании нетривиальных (положительных и отрицательных) решений при положительных значениях спектрального параметра. Приводятся примеры, иллюстрирующие полученные теоремы. Библ. 12. Фиг. 8.

Ключевые слова: задача Штурма–Лиувилля, несамосопряженный дифференциальный оператор, разрывная нелинейность, нетривиальные решения.

DOI: 10.31857/S0044466924060096, **EDN:** XYMHHG

1. ВВЕДЕНИЕ. ПОСТАНОВКА ЗАДАЧИ

Проблема существования решений задачи Штурма–Лиувилля для обыкновенных дифференциальных уравнений второго порядка с разрывными нелинейностями изучалась в [1]–[9]. Отметим также работу [10], посвященную решениям модельной краевой задачи для обыкновенного дифференциального уравнения второго порядка с параметром и разрывной правой частью, в которой решения выписываются в явном виде.

По сравнению с [1]–[3], [6] в настоящей статье ослаблены ограничения на множество точек разрыва и рост нелинейности на бесконечности, изучаются полуправильные решения, а в отличие от предыдущих работ авторов (см. [4], [5], [8]–[10]) дифференциальный оператор является несамосопряженным.

Рассматривается вопрос существования решений задачи Штурма–Лиувилля

$$Lu(x) \equiv -\big(p(x)u'(x)\big)' + r(x)u'(x) + q(x)u(x) = \lambda g\big(x, u(x)\big), \quad x \in (a, b), \tag{1}$$

$$u(a) = u(b) = 0 \tag{2}$$

при положительных значениях спектрального параметра λ. Здесь $p \in C_{1,\alpha}\big([a, b]\big)$, $r, q \in C_{0,\alpha}\big([a, b]\big)$, $r(x) \not\equiv 0$, $0 < \alpha < 1$, $-\infty < a < b < +\infty$. Предполагается, что $g(x, 0) = 0$ почти всюду на (a, b) и нелинейность $g(x, u)$ разрывна по фазовой переменной u.

В силу наличия у линейного дифференциального оператора L в уравнении (1) ненулевого слагаемого $r(x)u'(x)$ с производной первого порядка оператор L не является самосопряженным. Значит, вариационный метод (основной аппарат исследования задач с разрывными нелинейностями) не может

[1)] Работа выполнена при финансовой поддержке РНФ (№ 23-21-00069). https://rscf.ru/project/23-21-00069.

быть применен к изучению задачи (1), (2). Поэтому в настоящей статье используется метод верхних и нижних решений, позволяющий исследовать задачи без условия формальной самосопряженности дифференциального оператора.

Для дальнейших рассуждений потребуются следующие определения.

Определение 1. *Сильным решением* задачи (1), (2) называется функция $u \in W_1^2\big((a,b)\big) \cap \mathring{W}_2^1\big((a,b)\big)$, удовлетворяющая уравнению (1) для почти всех $x \in (a,b)$.

Определение 2. *Полуправильным решением* задачи (1), (2) называется такое сильное ее решение u, значение которого $u(x)$ для почти всех $x \in (a,b)$ является точкой непрерывности функции $g(x,\cdot)$.

Поскольку $g(x,0) = 0$ почти всюду на (a,b), то при любом значении параметра λ функция $u(x) \equiv 0$ на (a,b) является сильным решением задачи (1), (2). Нулевое (тривиальное) решение является полуправильным тогда и только тогда, когда нуль является точкой непрерывности функции $g(x,\cdot)$ для почти всех $x \in (a,b)$.

2. ТЕОРЕТИЧЕСКИЕ РЕЗУЛЬТАТЫ

Пусть нелинейность $g(x,u)$ в уравнении (1) равна разности функций $g_2(x,u)$ и $g_1(x,u)$, неубывающих по переменной u для почти всех $x \in (a,b)$, причем функция $g_2(x,u)$ суперпозиционно измерима, а функция $g_1(x,u)$ каратеодориева. Тогда имеют место следующие теоремы.

Теорема 1. *Пусть λ_1 — минимальное собственное значение дифференциального оператора L в уравнении* (1) *с граничным условием* (2) *и $\lambda_1 > 0$. Предположим также, что*

1) $\liminf\limits_{u\to 0} u^{-1} g(x,u) \geqslant k_1$ *(k_1 — положительная константа или $+\infty$) равномерно по $x \in (a,b)$;*
2) $\limsup\limits_{|u|\to+\infty} u^{-1} g(x,u) = 0$ *равномерно по $x \in (a,b)$;*
3) $g(x,0) \equiv 0$ *на (a,b);*
4) *функции $g_i(x,\cdot)$, $i = 1,2$, ограничены на отрезках числовой прямой $\mathbb{R}$ равномерно по $x \in (a,b)$.*

Тогда существует $\lambda_0 > 0$ такое, что для любого $\lambda \geqslant \lambda_0$ и произвольного $q > 1$ задача (1), (2) *имеет сильное положительное и сильное отрицательное решения на (a,b) из соболевского пространства $W_q^2\big((a,b)\big)$.*

Теорема 2. *Пусть выполнены все условия теоремы* 1*, кроме условия* 2*). Вместо него выполняется условие* 2′) $\limsup\limits_{|u|\to+\infty} u^{-1} g(x,u) \leqslant \gamma$ *равномерно по $x \in (a,b)$, где $\gamma \in (0, k_1)$ — постоянная, k_1 — константа из условия* 1) *теоремы* 1.

Тогда для любого $\lambda \in (\lambda_1 k_1^{-1}, \lambda_1 \gamma^{-1})$ и произвольного $q > 1$ задача (1), (2) *имеет сильное положительное и сильное отрицательное решения на (a,b) из соболевского пространства $W_q^2\big((a,b)\big)$.*

Пусть теперь нелинейность $g(x,u)$ в уравнении (1) суперпозиционно измерима и для некоторой положительной константы M функция $g(x,u) + Mu$ неубывающая по u на $\mathbb{R}$ для почти всех $x \in (a,b)$. Тогда справедливы следующие теоремы.

Теорема 3. *Пусть выполнены условия теоремы* 1 (*условие* 4) *для функции $g(x,\cdot)$). Тогда существует $\lambda_0 > 0$ такое, что для любого $\lambda \geqslant \lambda_0$ и произвольного $q > 1$ задача* (1), (2) *имеет полуправильное положительное и полуправильное отрицательное решения на (a,b) из соболевского пространства $W_q^2\big((a,b)\big)$.*

Теорема 4. *Пусть выполнены условия теоремы* 2*. Тогда для любого $\lambda \in \big(\lambda_1 k_1^{-1}, \lambda_1 \gamma^{-1}\big)$ и произвольного $q > 1$ задача* (1), (2) *имеет полуправильное положительное и полуправильное отрицательное решения на (a,b) из соболевского пространства $W_q^2\big((a,b)\big)$.*

Результаты, аналогичные теоремам 1–4, для эллиптических краевых задач с параметром и разрывными нелинейностями были получены в [11], [12]. В настоящей статье осуществлен перенос результатов для уравнений эллиптического типа на обыкновенные дифференциальные уравнения.

Доказательство теорем 1–4 проводится методом верхних и нижних решений аналогично доказательству следствий 4, 7 из [11] и теорем 9, 12 из [12].

3. ПРИЛОЖЕНИЯ

Приведем примеры разрывных нелинейностей $g(x,u)$, удовлетворяющих условиям теорем 1–4.

Пример 1. Пусть

$$g(x,u)\equiv g(u)=\begin{cases}-\sqrt[3]{u}-3, & \text{если } u<-1,\\ \operatorname{sgn}(u), & \text{если } |u|\leqslant 1,\\ \sqrt{u}+1, & \text{если } u>1.\end{cases} \tag{3}$$

С одной стороны, нелинейность $g(u)$ равна разности двух неубывающих на $\mathbb{R}$ функций

$$g_2(u)=\begin{cases}-4 & \text{при } u<-1,\\ g(u)-2 & \text{при } u\geqslant -1\end{cases}$$

и

$$g_1(u)=\begin{cases}-g(u)-4 & \text{при } u<-1,\\ -2 & \text{при } u\geqslant -1,\end{cases}$$

причем $g_1(u)$ непрерывна на $\mathbb{R}$, а у функции $g_2(u)$ три точки разрыва. Функция $g(u)$ неограниченная на $\mathbb{R}$ и удовлетворяет условиям 1)–3) теоремы 1.

С другой стороны, нелинейность $g(u)$ суперпозиционно измерима, функция $g(u)+u$ неубывающая на $\mathbb{R}$, $g(0)=0$. У функции $g(u)$ три точки разрыва, она не ограничена на $\mathbb{R}$, и для нее выполняются условия теоремы 3.

Отметим, что для задачи (1), (2) с такой нелинейностью функция $u(x)\equiv 0$ является сильным решением, но не является полуправильным решением.

Пример 2. Пусть

$$g(x,u)\equiv g(u)=\begin{cases}-\gamma_1 u-3, & \text{если } u<-1,\\ \operatorname{sgn}(u), & \text{если } |u|\leqslant 1,\\ \gamma u+1, & \text{если } u>1,\end{cases}$$

$\gamma>0$ – постоянная, $\gamma_1\in(0,2]$.

Нелинейность $g(u)$ равна разности двух неубывающих на $\mathbb{R}$ функций

$$g_2(u)=\begin{cases}0 & \text{при } u<-1,\\ g(u)+3-\gamma_1 & \text{при } u\geqslant -1\end{cases}$$

и

$$g_1(u)=\begin{cases}-g(u) & \text{при } u<-1,\\ 3-\gamma_1 & \text{при } u\geqslant -1,\end{cases}$$

причем $g_1(u)$ непрерывна на $\mathbb{R}$, а у функции $g_2(u)$ три точки разрыва, если $\gamma_1\in(0,2)$, и две, если $\gamma_1=2$. Отметим, что $g(u)$ имеет линейный рост на бесконечности и удовлетворяет условиям теоремы 2 и не удовлетворяет условию 2) теоремы 1.

Нелинейность $g(u)$ суперпозиционно измерима, функция $g(u)+\gamma_1 u$ неубывающая на $\mathbb{R}$, $g(0)=0$. При $\gamma_1\in(0,2)$ у функции $g(u)$ три точки разрыва, и для нее выполняются условия теоремы 4.

Далее (не ограничивая общности) положим в задаче (1), (2)

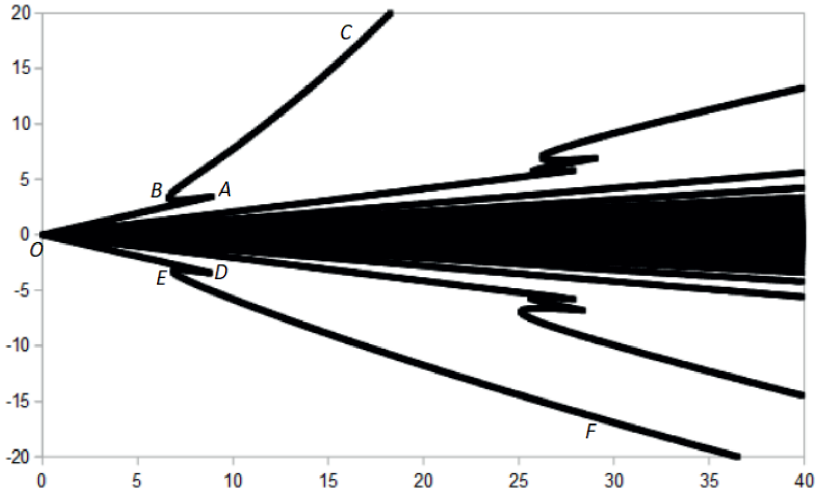
$$p(x)=r(x)=q(x)\equiv 1,\quad a=0,\quad b=1$$

и рассмотрим нелинейности $g(x,u)$ из примеров 1, 2.

Прежде всего отметим, что если при некотором значении $\lambda=\bar{\lambda}$ задача (1), (2) с нелинейностью (3) имеет решение $u=\bar{u}(x)$, целиком лежащее в полосе $-1\leqslant u\leqslant 1$, то для любого $\beta\in(0,1)$ задача (1), (2) будет иметь и решение $u=\beta\bar{u}(x)$ при значении параметра $\lambda=\beta\bar{\lambda}$. Другими словами, существование хотя бы одного решения, лежащего в полосе $-1\leqslant u\leqslant 1$ при некотором $\lambda=\bar{\lambda}$, гарантирует существо-

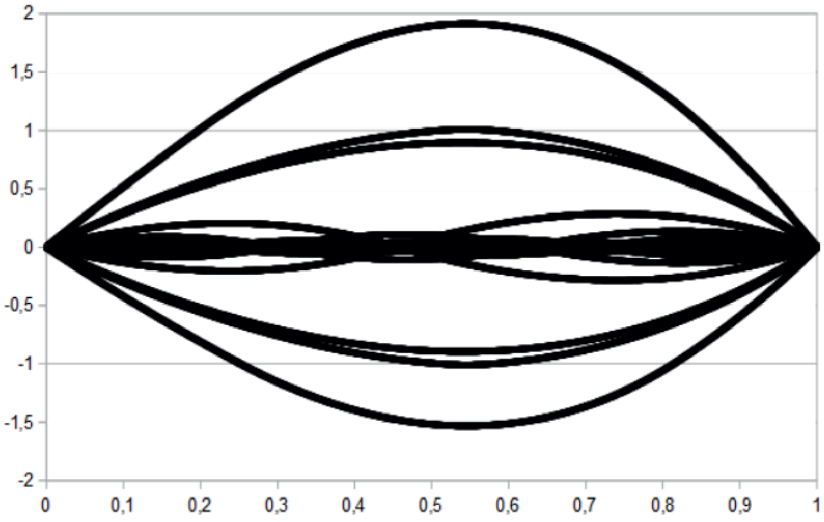
вание решений при любых $\lambda \in (0, \bar{\lambda})$. Это связано с линейностью левой части и следующим свойством правой части: $g(x, \beta u) = g(x, u) = \operatorname{sgn}(u)$ при $-1 \leqslant u \leqslant 1$.

На фиг. 1 представлена зависимость начальных данных $u'(0)$ решений задачи из примера 1 от значений параметра λ. Каждая точка на этом графике отвечает некоторому решению, поэтому он удобен для отслеживания количества решений при различных значениях λ. Проведенные рассуждения обосновывают наличие отрезков, исходящих из начала координат, которые соответствуют решениям, не выходящим за пределы полосы $-1 \leqslant u \leqslant 1$.



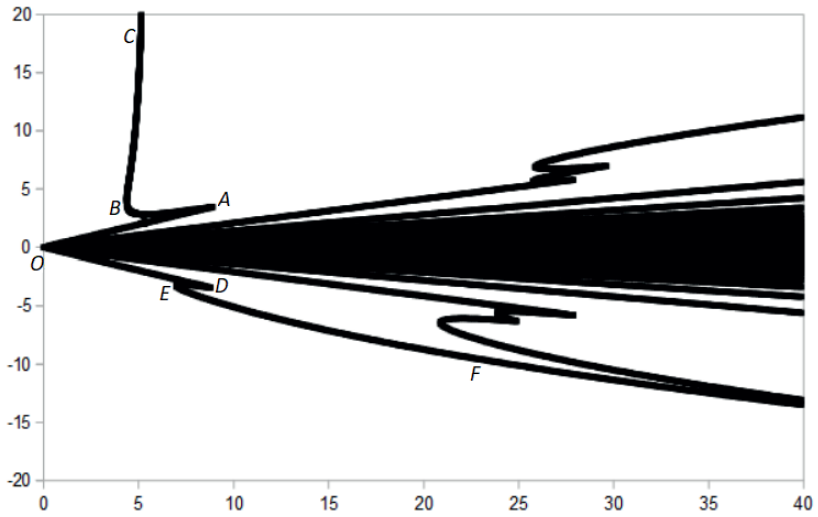
Фиг. 1. Зависимость $u'(0)$ от λ в примере 1.

Кривые $OABC$ и $ODEF$ соответствуют положительному и отрицательному решениям. Участки OA и OD суть отрезки, описывающие решения из полосы $-1 \leqslant u \leqslant 1$. Участки ABC и DEF соответствуют решениям, выходящим за пределы этой полосы. Например, на приведенных на фиг. 2 графиков решений видно, что при $\lambda = 8$ имеются три положительных и три отрицательных решения, по два из которых пересекают линии $u = \pm 1$, вдоль которых нелинейность $g(x, u)$ терпит разрыв. Кривые ABC и DEF как раз предсказываются теоремой 1: существует такое положительное λ_0 (точка E на фиг. 1), что для всех $\lambda \geqslant \lambda_0$ задача имеет сильное положительное и сильное отрицательное решения. В соответствии с теоремой 3 эти решения будут и полуправильными.



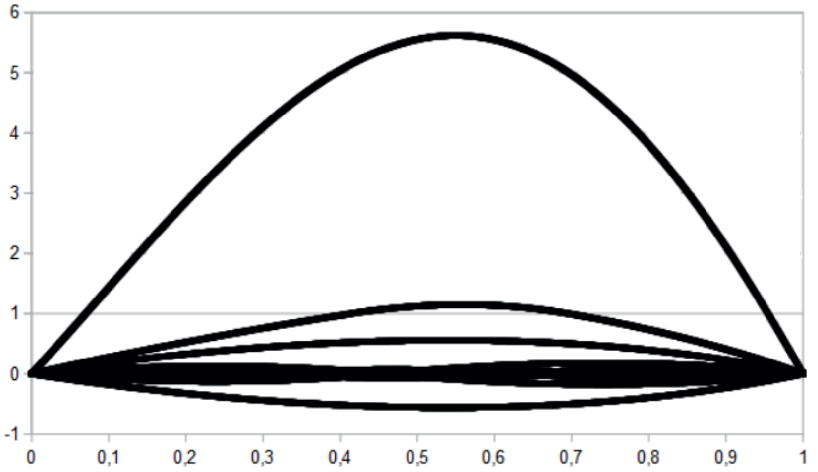
Фиг. 2. Графики решений примера 1 при $\lambda = 8$.

В примере 2 также $g(x,u) = \operatorname{sgn}(u)$ при $-1 \leqslant u \leqslant 1$, и на графике зависимости $u'(0)$ от λ, представленном на фиг. 3, вновь имеются отрезки, исходящие из начала координат. Однако теперь кривая $OABC$, соответствующая положительным решениям, уходит на бесконечность при значении $\lambda \to \pi^2/2 + 5/8$, предсказанном теоремой 2. Кривая $ODEF$ соответствует отрицательным решениям и качественно похожа на аналогичную кривую из примера 1. Таким образом, при $\lambda \in \left(0, \pi^2/2 + 5/8\right)$ имеются сильные положительное и отрицательное решения.



Фиг. 3. Зависимость $u'(0)$ от λ в примере 2 при $\gamma = 2$, $\gamma_1 = 1$.

Для иллюстрации решений было выбрано значение $\lambda = 5$. Полученные графики представлены на фиг. 4. Имеются три положительных решения, два из которых пересекают линию $u = 1$, вдоль которой правая часть $g(x,u)$ терпит разрыв. Также существует единственное отрицательное решение, лежащее в полосе $-1 \leqslant u \leqslant 0$, поскольку выбранное значение $\lambda = 5$ левее точки E на фиг. 3. Эти решения являются полуправильными в полном соответствии с теоремой 4.



Фиг. 4. Графики решений примера 2 при $\lambda = 5$, $\gamma = 2$, $\gamma_1 = 1$.

Пример 3. Рассмотрим задачу (1), (2) с $p(x) = r(x) = q(x) \equiv 1$, $a = 0$, $b = 1$ и нелинейностью

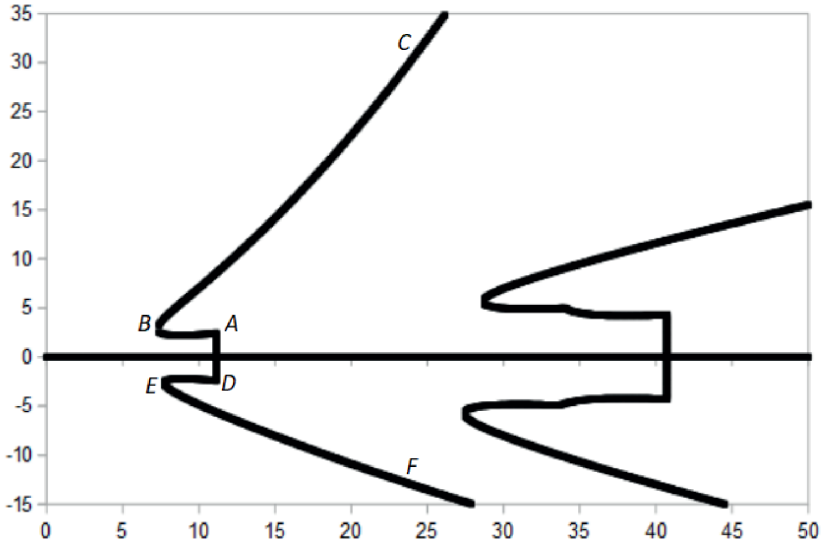
$$g(x,u) \equiv g(u) = \begin{cases} -\sqrt[3]{u} - 3, & \text{если } u < -1, \\ u, & \text{если } |u| \leqslant 1, \\ \sqrt{u} + 1, & \text{если } u > 1. \end{cases}$$

Как и в примере 1, она удовлетворяет теоремам 1 и 3, однако теперь $g(x,u)$ не имеет разрыва при $u=0$. Вследствие этого тривиальное решение $u(x)\equiv 0$ является сильным и полуправильным.

Вновь начнем с исследования ненулевых решений, не выходящих за пределы полосы $-1\leqslant u\leqslant 1$. Они будут существовать лишь при $\lambda=\lambda_k=\pi^2k^2+5/4$, $k=1,2,\ldots$, и иметь вид

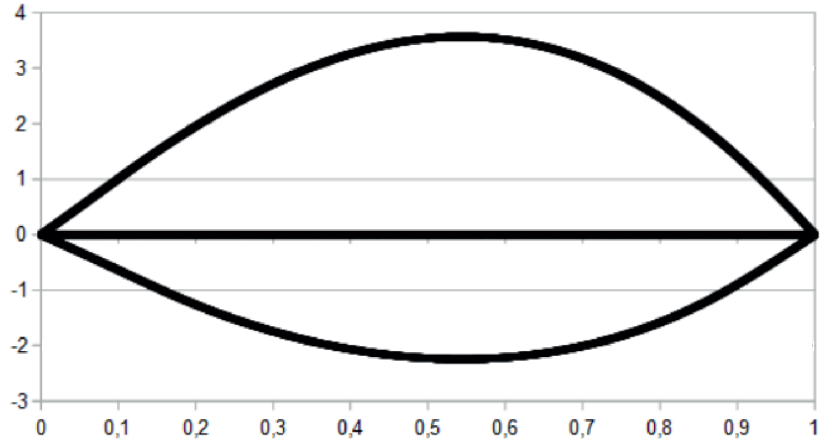
$$u=A_ke^{x/2}\sin\pi kx,\quad |A_k|\leqslant\frac{\sqrt{1+4\pi^2k^2}}{2\pi k}e^{-1/2+\operatorname{arctg}2\pi k/(2\pi k)}. \tag{4}$$

На графике зависимости $u'(0)$ от λ, приведенном на фиг. 5, этим решениям соответствуют вертикальные отрезки. Участки ABC и DEF соответствуют положительным и отрицательным решениям, выходящим за пределы полосы $-1\leqslant u\leqslant 1$. По фиг. 5 можно сделать вывод, что существует такое значение $\lambda_0>0$ (точка E), начиная с которого при каждом $\lambda\geqslant\lambda_0$ существуют положительное и отрицательное решения поставленной задачи. Данный вывод полностью согласуется с теоремой 1.



Фиг. 5. Зависимость $u'(0)$ от λ в примере 3.

Графики решений при $\lambda=12$ имеют вид, представленный на фиг. 6. Оба нетривиальных решения пересекают линии $u=\pm1$, вдоль которых правая часть терпит разрыв, по два раза. Таким образом, они являются полуправильными, что и предсказывается теоремой 3.

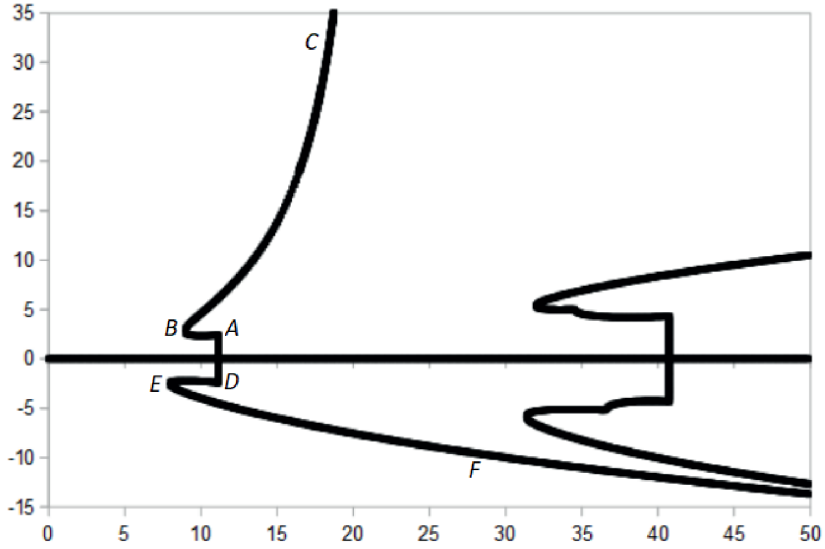


Фиг. 6. Графики решений примера 3 при $\lambda=12$.

Пример 4. Рассмотрим задачу из примера 3 с нелинейностью

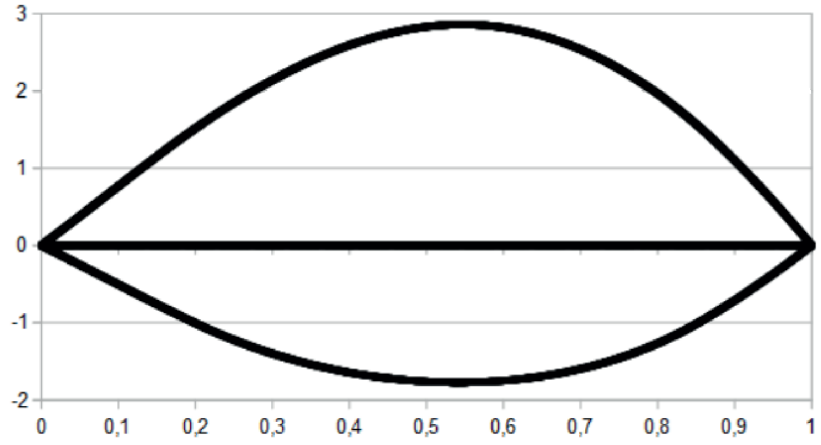
$$g(x,u) \equiv g(u) = \begin{cases} -u-3, & \text{если } u < -1, \\ u, & \text{если } |u| \leqslant 1, \\ \frac{1}{2}u + 1, & \text{если } u > 1. \end{cases}$$

Как и в примере 3, тривиальное решение является полуправильным, и при $\lambda = \lambda_k = \pi^2 k^2 + 5/4$, $k = 1, 2, \ldots$, существует семейство решений (4). Поэтому на графике зависимости $u'(0)$ для решений данной задачи от параметра λ, изображенном на фиг. 7, вновь видны вертикальные отрезки. Также есть кривая DEF, соответствующая отрицательным решениям. Однако теперь кривая ABC, отвечающая положительным решениям, выходящим за пределы полосы $-1 \leqslant u \leqslant 1$, уходит на бесконечность при $\lambda \to 2\pi^2 + 5/2$. Это согласуется с теоремой 2, утверждающей существование положительного и отрицательного решений в данном случае при $\lambda \in \left(\pi^2 + 5/4, 2\pi^2 + 5/2\right)$.



Фиг. 7. Зависимость $u'(0)$ от λ в примере 4.

Графики решений при $\lambda = 12$ приведены на фиг. 8. Как и в примере 3, нетривиальные решения дважды проходят через точки разрыва нелинейности, так что эти решения являются полуправильными в полном соответствии с теоремой 4.



Фиг. 8. Графики решений примера 4 при $\lambda = 12$.

Таким образом, полученные теоретические результаты проиллюстрированы примерами.

СПИСОК ЛИТЕРАТУРЫ

1. *Carl S., Heikkila S.* On the existence of minimal and maximal solutions of discontinuous functional Sturm–Liouville boundary value problems // J. Inequal. Appl. 2005. N 4. P. 403–412.
2. *Bonanno G., Bisci G. M.* Infinitely many solutions for a boundary value problem with discontinuous nonlinearities // Bound. Value Probl. 2009. Art. ID 670675. 20 p.
3. *Bonanno G., Buccellato S. M.* Two point boundary value problems for the Sturm–Liouville equation with highly discontinuous nonlinearities // Taiwanese J. Math. 2010. V. 14. N 5. P. 2059–2072.
4. *Потапов Д. К.* Задача Штурма–Лиувилля с разрывной нелинейностью // Дифференц. уравнения. 2014. Т. 50. № 9. С. 1284–1286.
5. *Потапов Д. К.* Существование решений, оценки дифференциального оператора и "разделяющее" множество в краевой задаче для дифференциального уравнения второго порядка с разрывной нелинейностью // Дифференц. уравнения. 2015. Т. 51. № 7. С. 970–974.
6. *Bonanno G., D'Agui G., Winkert P.* Sturm–Liouville equations involving discontinuous nonlinearities // Minimax Theory Appl. 2016. V. 1. N 1. P. 125–143.
7. *Павленко В. Н., Постникова Е. Ю.* Задача Штурма–Лиувилля для уравнения с разрывной нелинейностью // Челяб. физ.-матем. журн. 2019. Т. 4. Вып. 2. С. 142–154.
8. *Басков О. В., Потапов Д. К.* Управление и возмущение в задаче Штурма–Лиувилля с разрывной нелинейностью // Вестн. С.-Петерб. ун-та. Прикл. матем. Информ. Проц. управ. 2023. Т. 19. Вып. 2. С. 275–282.
9. *Потапов Д. К.* Аппроксимация задачи Штурма–Лиувилля с разрывной нелинейностью // Дифференц. уравнения. 2023. Т. 59. № 9. С. 1191–1198.
10. *Басков О. В., Потапов Д. К.* О решениях краевой задачи для одного дифференциального уравнения второго порядка с параметром и разрывной правой частью // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 8. С. 1296–1308.
11. *Павленко В. Н., Потапов Д. К.* Существование решений невариационной эллиптической краевой задачи с параметром и разрывной нелинейностью // Матем. тр. 2016. Т. 19. № 1. С. 91–105.
12. *Павленко В. Н., Потапов Д. К.* Существование полуправильных решений эллиптических спектральных задач с разрывными нелинейностями // Матем. сб. 2015. Т. 206. № 9. С. 121–138.

EXISTENCE OF SOLUTIONS TO THE NON-SELF-ADJOINT STURM-LIOUVILLE PROBLEM WITH DISCONTINUOUS NONLINEARITY

O. V. Baskov, D. K. Potapov*

St. Petersburg State University, 7/9 Universitetskaya Embankment, St. Petersburg, 199034 Russia

**e-mail: d.potapov@spbu.ru*

Received 20 December, 2023
Revised 20 December, 2023
Accepted 06 March, 2024

Abstract. The problem of existence of solutions of the Sturm-Liouville problem with a non-self-adjoint differential operator and non-linearity discontinuous in the phase variable is considered. Theorems on the existence of non-trivial (positive and negative) solutions for positive values of the spectral parameter are established for the problem under study. Examples illustrating the obtained theorems are given.

Keywords: Sturm-Liouville problem, non-self-adjoint differential operator, discontinuous non-linearity, non-trivial solutions.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1016–1027

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.927.25+517.988.2

ФУНКЦИОНАЛЫ СОБСТВЕННЫХ ЗНАЧЕНИЙ НА МНОГООБРАЗИИ ПОТЕНЦИАЛОВ

© 2024 г. Я. М. Дымарский[1,*]

[1] *141701 М.о., г. Долгопрудный, Институтский переулок, 9, Московский физико-технический институт*

[*] *e-mail: dymarskii@mail.ru*

Поступила в редакцию 12.12.2023 г.
Переработанный вариант 20.02.2024 г.
Принята к публикации 05.03.2024 г.

Даны аналитическое и топологическое описания функционала собственных значений на многообразии периодических потенциалов. Библ. 6.

Ключевые слова: пространство периодических краевых задач, функционал собственных значений, расслоение многообразия потенциалов.

DOI: 10.31857/S0044466924060105, **EDN:** XYGYZU

ВВЕДЕНИЕ

В работах [1] и [2] рассмотрено семейство стационарных операторов Шрёдингера с непрерывным периодическим потенциалом. Замечено, во-первых, что указанное семейство порождает многообразие соответствующих периодических собственных функций выбранной осцилляции, во-вторых, по собственной функции возможно восстановление потенциала с точностью до константы. Там же предложена параметризация многообразия периодических собственных функций выбранной осцилляции и согласованная с ней параметризация семейства операторов, получена формула соответствующих собственных значений.

В настоящей статье рассмотрено семейство операторов с периодическими потенциалами из L_2. В предварительных леммах и теореме 1 сформулированы и в некоторых случаях по-новому доказаны основные утверждения, полученные ранее для операторов с периодическими непрерывными потенциалами. В теореме 2 полностью описаны аналитические и топологические свойства функционала собственных значений. Гильбертовость семейства потенциалов позволяет доказать морсовость функционала и посчитать его индекс. Обнаружено, возможно впервые, специфическое поведение функционала собственного значения с ростом спектральной лакуны.

1. ОСНОВНЫЕ ПОНЯТИЯ И ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Рассмотрим семейство периодических самосопряженных краевых задач

$$-y'' + p(x)y = \lambda y, \tag{1}$$

$$y(0) - y(2\pi) = y'(0) - y'(2\pi) = 0 \tag{2}$$

на собственные значения λ и собственные функции y. Параметром семейства является потенциал p из гильбертова пространства $L_2(2\pi)$ 2π-периодических функций, суммируемых на периоде с квадратом. Поскольку изменение потенциала на константу не влияет на собственные функции, а лишь сдвигает все собственные значения на ту же константу, целесообразно использовать в качестве пространства параметров подпространство

$$P := \left\{ p \in L_2(2\pi) : \quad \int_0^{2\pi} p(x)dx = 0 \right\} \tag{3}$$

потенциалов с нулевым средним. В дальнейшем под пространством краевых задач мы понимаем пространство P.

Напомним свойства спектра краевой задачи (1), (2) для фиксированного потенциала $p \in P$ (см. [3]). Спектр состоит из вещественных собственных значений, которые не более чем двукратны:

$$\lambda_0(p) < \lambda_1^-(p) \leqslant \lambda_1^+(p) < \ldots < \lambda_n^-(p) \leqslant \lambda_n^+(p) < \ldots$$

Соответствующие собственные функции принадлежат соболевскому пространству $W_2^2(2\pi)$ 2π-периодических функций, которые суммируемы с квадратом и их производные вплоть до второго порядка суммируемы с квадратом на периоде (см. [4]). Заметим, что, в силу теоремы вложения, собственные функции принадлежат классу гладкости C^1. Собственные функции, отвечающие собственным значениям с нижним индексом n, имеют на полуинтервале $[0, 2\pi)$ в точности $2n$ невырожденных нулей. Отрезок $[\lambda_n^-, \lambda_n^+]$ называется n-й лакуной. Собственные функции, отвечающие разным собственным значениям, попарно ортогональны в пространстве $L_2(2\pi)$. В случае совпадения собственных значений $\lambda_n^-(p) = \lambda_n^+(p) = \lambda_n^0(p)$ (вырожденная лакуна) возникает плоскость $\Pi_n(p)$ собственных функций. Нули линейно независимых собственных функций, отвечающих общему собственному значению, перемежаются.

Таким образом, на P определен функционал λ_0, а для каждого $n \in \mathbb{N}$ и каждого выбранного знака "$\pm$" на пространстве параметров определен функционал $\lambda_n^\pm : P \to \mathbb{R}$. Очевидно, что совпадают сужения $\lambda_n^-|_{P_n^0} = \lambda_n^+|_{P_n^0} = \lambda_n^0$. Цель работы описать свойства этих функционалов.

Мы будем рассматривать собственные функции y задачи (1), (2) как элементы проективного пространства, порожденного функциональным пространством $W_2^2(2\pi)$. Введем множество всех собственных функций, отвечающих собственным значениям с натуральным нижним индексом n:

$$Y_n := \{ y \in W_2^2(2\pi) : \int_0^{2\pi} y^2 dx = 1,\ y \sim -y, \exists p \in P : \ (1), (2) \text{ верно } \text{с } \lambda = \lambda_n^\pm(p) \}.$$

Очевидно, что множество Y_n допускает аналитическое описание:

Лемма 1. *Множество Y_n ($n = 0, 1, \ldots$) состоит из всех таких функций $y \in W_2^2(2\pi)$, нормированных в $L_2(2\pi)$, что:*

1) $y(0) - y(2\pi) = y'(0) - y'(2\pi) = 0$;
2) *существуют в точности $2n$ точек $x_i \in [0, 2\pi)$, в которых $y(x_i) = 0$;*
3) $y'(x_i) \neq 0$ $(i = 1, \ldots, 2n)$;
4)

$$\int_0^{2\pi} \left(\frac{y''}{y}\right)^2 dx < \infty \ \Leftrightarrow \ \int_0^{2\pi} \left(\frac{y''}{(x - x_1)\ldots(x - x_{2n})}\right)^2 dx < \infty.$$

Исследуемое семейство краевых задач настолько узкое, что по собственной функции $y \in Y_n$ однозначно восстанавливается собственное значение и потенциал, т.е. определены отображения

$$\lambda_n : Y_n \to \mathbb{R},\ \lambda_n(y) = \lambda := -\frac{1}{2\pi}\int_0^{2\pi} \frac{y''}{y} dx; \tag{4}$$

$$\Phi_n : Y_n \to P,\ \Phi_n(y) = p := \frac{y''}{y} + \lambda_n(y). \tag{5}$$

Далее, по потенциалу p и собственному значению λ (которое, в силу определения, имеет нижний индекс n) можно определить, какой из верхних индексов $\pm$ соответствует числу λ или же это число имеет кратность два. Значит, $\lambda_n(y) = \lambda_n^\pm(\Phi_n(y))$.

До конца раздела зафиксируем нижний индекс $n \in \mathbb{N}$. Обозначим через $Y_n^{\pm} \subset Y_n$ множество всех собственных функций семейства (1)–(3), которым соответствуют простые собственные значения $\lambda_n^{\pm}$, а через $Y_n^0 \subset Y_n$ – множество всех собственных функций семейства (1)-(3), которым соответствуют двукратные собственные значения λ_n^0. Объединение $Y_n^+ \cup Y_n^- \cup Y_n^0 = Y_n$. Из определения следует, что множества $Y_n^{\pm}$, Y_n^0 с разными номерами или разными верхними индексами попарно не пересекаются.

Определение 1. 1) Пару $(y, y_r) \in Y_n^{\pm} \times Y_n^{\mp}$ собственных функций мы назовем взаимной (reciprocal), если эти функции порождены одним и тем же потенциалом p.

2) Пару $(y, y_r) \in Y_n^0 \times Y_n^0$ собственных функций мы назовем взаимной, если они порождены одним и тем же потенциалом p и ортогональны в L_2. (В первом случае взаимные функции ортогональны автоматически.)

Из предыдущих рассуждений следует, что по функции $y \in Y_n$ ($n \in \mathbb{N}$) взаимная функция $y_r \in Y_n$ восстанавливается однозначно и отношение взаимности симметрично. Значит, на множестве Y_n определена инволюция $I_n(y) := y_r$, у которой нет неподвижных точек. Обозначим сужение $I_n^0 := I_n|_{Y_n^0}$.

Через $\Delta\lambda(y) := \Lambda_n(y) - \Lambda_n(y_r)$, где $y \in Y_n$, обозначим ориентированную длину лакуны; следовательно, $\operatorname{sign}\big(\Delta\lambda(y)\big) = \pm 1$, если $y \in Y_n^{\pm}$, и $\Delta\lambda(y) = 0$, если $y \in Y_n^0$. Нас интересуют подмножества собственных функций постоянной ориентированной длины n-й лакуны:

$$Y_n(\Delta\lambda) := \big\{ y \in Y_n : \ \Lambda(y) - \Lambda(y_r) = \Delta\lambda \big\}. \tag{6}$$

Очевидно, что $I_n\big(Y_n(\Delta\lambda)\big) = Y_n(-\Delta\lambda)$. Из определения следует, что

$$Y_n^{\pm} = \bigcup_{\Delta\lambda > 0 (<0)} Y_n(\Delta\lambda), \ \ Y_n^0 = Y_n(0), \ \ Y_n = \bigcup_{\Delta\lambda \in \mathbb{R}} Y_n(\Delta\lambda). \tag{7}$$

Аналогично, нас интересуют подмножества потенциалов постоянной неориентированной длины n-й лакуны. По определению:

$$P_n(|\Delta\lambda|) := \{ p \in P : \ \lambda_n^+(p) - \lambda_n^-(p) = |\Delta\lambda| \geqslant 0 \} \subset P, \tag{8}$$

$$P_{\diagdown n} = \bigcup_{|\Delta\lambda| > 0} P_n(|\Delta\lambda|), \ \ P = \bigcup_{|\Delta\lambda| \geqslant 0} P_n(|\Delta\lambda|) = P_n(0) \bigcup P_{\diagdown n}. \tag{9}$$

Если потенциал $p \in P_n^0 := P_n(0)$, то сам потенциал, собственное значение $\lambda_n^0(p)$ и соответствующую собственную функцию $y \in Y_n^0$ мы назовем *n-вырожденными* или просто вырожденными, когда индекс n известен из контекста.

Введем в рассмотрение вспомогательные функции. Пусть $y \in Y_n$. Определим вронскиан

$$W(y, y_r)(x) := \begin{vmatrix} y(x) & y_r(x) \\ y'(x) & y_r'(x) \end{vmatrix}. \tag{10}$$

Отметим, во-первых, что знак вронскиана (10) зависит от выбора представителей элементов y, y_r и, во-вторых, только для вырожденного случая $p \in P_n(0)$ он порожден решениями одного и того же дифференциального уравнения и только в этом случае постоянен.

Далее, рассмотрим векторную функцию $\overline{y}(x) := \big(y(x), y_r(x)\big)$, значения которой принадлежат плоскости Ω_n с координатами (y, y_r); ее образом является некоторая замкнутая кривая Υ. Введем на плоскости Ω_n полярные координаты (ρ, θ). Тогда

$$y(x) = \rho(x)\cos\theta(x), \ \ y_r(x) = \rho(x)\sin\theta(x), \tag{11}$$

где значения непрерывной ветви угловой функции $\theta = \theta(x)$ выбираются с точностью до кратного π, поскольку мы отождествили собственные функции отличающиеся знаком. (Ниже мы определим вронскиан и угловую функцию однозначно.) Наконец, нам понадобится угловая скорость $\eta(x) := \theta'(x)$ и начальное значение $\varphi := \theta(0)$ угловой функции. Таким образом, по собственной функции y на-

ми определены пара чисел $(\varphi, \Delta\lambda)$ и шесть функций: $p,\ y_r,\ W,\ \rho,\ \theta,\ \eta$. Оказывается, пара чисел $(\varphi, \Delta\lambda)$ и функция η полностью определяют остальные шесть функций: $y,\ y_r,\ p,\ W,\ \rho,\ \theta$.

Леммы ниже дают описания свойств введенных функций.

Лемма 2. *Свойства вронскиана.*

1. *Производная вронскиана вычисляется по формуле*

$$W(y, y_r)'(x) = \Delta\lambda(y) \cdot y(x) \cdot y_r(x). \tag{12}$$

2. *Функция* $W(y, y_r) \in W_2^3(2\pi)$.

3. *Вронскиан нигде не обращается в нуль:* $\forall x$ *справедливо* $W(y, y_r)(x) \neq 0$.

4. *Пусть* y, y_r *— произвольные представители соответствующих элементов из* Y_n. *Из четырех возможных пар* $(\pm y, \pm y_r)$, $(\pm y, \mp y_r)$ *только одна удовлетворяет условиям* $W(y, y_r)(x) > 0$ *и* $\varphi = \theta(0) \in [0, \pi)$.

Доказательство. Первое утверждение следует непосредственно из определения вронскиана и уравнения (1). Второе утверждение следует из первого.

Чтобы доказать третье утверждение, заметим, что вронскиан принадлежит классу C^2. Вырожденный случай $\Delta\lambda = 0$ очевиден. Пусть $\Delta\lambda < 0$. Судя по формуле (12), точки экстремума вронскиана являются нулями собственных функций y и y_r. Но нули функций y и y_r перемежаются, иначе у функций y и y_r, согласно теореме Штурма, на периоде будет разное количество нулей. Пусть $y(x_1) = y_r(x_2) = 0$ и между x_1 и x_2 нет нулей функций y и y_r. Пусть, для определенности, $y'(x_1) > 0$ и $y_r(x_1) > 0$. Тогда $y(x_2) > 0$ и $y_r'(x_2) < 0$. Теперь $W(x_1) = -y_r(x_1)y'(x_1) < 0$ и $W(x_2) = y(x_2)y_r'(x_2) < 0$. Значит, в любых соседних нулях собственных функций y и y_r знак вронскиана не меняется. Следовательно, он не меняется нигде.

Последнее утверждение следует из предыдущего. Лемма доказана.

Впредь мы будем выбирать именно ту единственную пару (y, y_r), для которой $W(y, y_r) > 0$ и $\varphi = \theta(0) \in \in [0, \pi)$.

Лемма 3. *Функция полярного радиуса* ρ *строго положительна,* $\rho \in W_2^2(2\pi)$ *и справедливо тождество*

$$\eta(x) := \theta'(x) \equiv \frac{W(x)}{\rho^2(x)}.$$

Первое утверждение следует из п. 3 леммы 2, второе — из определения (11) полярного радиуса и первого утверждения. Третье утверждение следует из формулы $\operatorname{tg}\theta = y_r/y$ и второго утверждения.

Теперь очевидна

Лемма 4. *Функция угловой скорости* $\eta \in W_2^2(2\pi)$, *она строго положительна и справедливо равенство*

$$\theta(2\pi) - \theta(0) = \int_0^{2\pi} \eta(x)dx = 2\pi n. \tag{13}$$

Обозначим через $\widehat{H}_n \subset W_2^2(2\pi)$ подмножество функций η, удовлетворяющих утверждениям леммы 4, а через $H_n \subset \widetilde{H}_n$ подмножество функций η, удовлетворяющих еще двум условиям:

$$\int_0^{2\pi} \frac{\sin 2\int_0^x \eta(t)dt}{\eta(x)}\, dx = 0, \tag{14}$$

$$\int_0^{2\pi} \frac{\cos 2\int_0^x \eta(t)dt}{\eta(x)}\, dx = 0. \tag{15}$$

Лемма 5. *Множество* $\widetilde{H}_n$ *является выпуклым открытым подмножеством гиперплоскости* $Hip_n = \left\{\eta \in W_2^2(2\pi) : \int_0^{2\pi} \eta(x)dx = 2\pi n\right\}$. *Подмножество* $H_n \subset \widetilde{H}_n$ *является гомотопически тривиальным гладким (т.е. класса* C^∞*) гильбертовым подмногообразием коразмерности два.*

Лемма 5 доказана в [1] для случая $\eta \in C^2(2\pi)$ (см. там лемму 6 и п. 5.1 и 5.3). Без изменений доказательство переносится на случай $\eta \in W_2^2(2\pi)$.

2. ФОРМУЛИРОВКИ ОСНОВНЫХ УТВЕРЖДЕНИЙ

Оказывается расслоение (6), (7) многообразия Y_n ($n \in \mathbb{N}$) и стратификация (8), (9) пространства P согласованы через отображение Φ_n (см. (4) и (5)). Для тривиализации расслоения мы привлекаем переменные $(\eta, \varphi, \Delta\lambda) \in H_n \times \mathbb{R}P^1 \times \mathbb{R}$.

Теорема 1. *Отображение Φ_n ($n \in \mathbb{N}$) дополняется до коммутативной диаграммы*

$$\begin{array}{ccc} Y_n & \xrightarrow{\Phi_n} & P \\ {\scriptstyle G_n}\downarrow & & \downarrow{\scriptstyle F_n} \\ H_n \times \mathbb{R}P^1 \times \mathbb{R} & \xrightarrow[\pi_n]{} & H_n \times \mathbb{R}^2 \end{array}, \tag{16}$$

в которой:

1. *Отображение*

$$\pi_n(\eta, \varphi, \Delta\lambda) := (\eta, \Delta\lambda \cos 2\varphi, \Delta\lambda \sin 2\varphi). \tag{17}$$

2. *Отображение G_n является биекцией.*

Обратное отображение, которое индуцирует в Y_n^0 структуру гладкого банахова многообразия диффеоморфного $H_n \times \mathbb{R}P^1 \times \mathbb{R}$, задается формулой

$$G_n^{-1}(\eta, \varphi, \Delta\lambda) = y(x) := \frac{1}{\sqrt{\eta(x)}} \cdot exp\left(\frac{\Delta\lambda}{4}\int_0^x \frac{\sin 2\left(\varphi + \int_0^t \eta(\tau)d\tau\right)}{\eta(t)}\,dt\right)\cos\left(\varphi + \int_0^x \eta(t)dt\right). \tag{18}$$

3. *По значению y отображения G_n^{-1}*

(а)*: мы находим вспомогательную функцию*

$$r(\eta, \varphi, \Delta\lambda) := \frac{y''(x)}{y(x)} = -\frac{\eta''}{2\eta} + \frac{3(\eta')^2}{4\eta^2} - \eta^2 - \frac{\Delta\lambda\eta' \sin(2\theta)}{2\eta^2} + \Delta\lambda\cos(2\theta) + \frac{\Delta\lambda^2 \sin^2(2\theta)}{16\eta^2} - \frac{1}{2}\Delta\lambda; \tag{19}$$

(б)*: по формулам* (4) *и* (19) *находим собственное значение*

$$\lambda(\eta, \varphi, \Delta\lambda) = -\frac{1}{2\pi}\int_0^{2\pi} r(x)dx = \frac{1}{2\pi}\int_0^{2\pi}\left(\eta^2 - \left(\frac{\eta'}{2\eta}\right)^2 - \left(\frac{\Delta\lambda \sin 2\theta}{4\eta}\right)^2\right)dx + \frac{1}{2}\Delta\lambda. \tag{20}$$

4. *Отображение F_n является тривиализацией гладкого расслоения $\beta: P \to P_n^0$ над подмногообразием n-вырожденных потенциалов; стандартным слоем расслоения является $\mathbb{R}^2$; сужение $F_n^0: P_n^0 \to H_n$ является диффеоморфизмом на H_n, который индуцирует на $P_n^0 \subset P$ структуру гомотопически тривиального и тривиально вложено в P гладкого подмногообразия коразмерности два.*

Обратное отображение F_n^{-1} задается следующим образом:

(а)*: по точке $(\xi_1, \xi_2) \in \mathbb{R}^2$ находим длину лакуны $|\Delta\lambda| = \sqrt{\xi_1^2 + \xi_2^2} \geqslant 0$, если $|\Delta\lambda| > 0$, из системы уравнений*

$$\xi_1 = |\Delta\lambda| \cos 2\varphi, \quad \xi_2 = |\Delta\lambda| \sin 2\varphi \tag{21}$$

определяется единственный угол $\varphi \in [0, \pi)$;

(б)*: по формуле* (18) *находим собственную функцию $y = y(\eta, \varphi, \Delta\lambda) \in Y_n^+ \cup Y_n^0$, взяв $\Delta\lambda = |\Delta\lambda|$;*

(в)*: по формуле* (5) *находим потенциал*

$$p(\eta, \varphi, |\Delta\lambda|) = F_n^{-1}(\eta, \varphi, |\Delta\lambda|) = r(x) + \lambda(\eta, \varphi, |\Delta\lambda|) =$$
$$= -\frac{\eta''}{2\eta} + \frac{3(\eta')^2}{4\eta^2} - \eta^2 - \frac{|\Delta\lambda|\eta' \sin(2\theta)}{2\eta^2} + |\Delta\lambda|\cos(2\theta) + \frac{|\Delta\lambda|^2 \sin^2(2\theta)}{16\eta^2} +$$
$$+ \frac{1}{2\pi}\int_0^{2\pi}\left(\eta^2 - \left(\frac{\eta'}{2\eta}\right)^2 - \left(\frac{|\Delta\lambda|\sin 2\theta}{4\eta}\right)^2\right)dx. \quad (22)$$

5. *Функции* $\lambda_n^{\pm}(p)$ *зависимости собственных значений от потенциала в терминах переменных* $(\eta, \varphi, |\Delta\lambda|)$ (см. (21)) *имеют вид*

$$\lambda_n^{\pm}(p) = \lambda_n^{\pm}(\eta, \varphi, |\Delta\lambda|) = \frac{1}{2\pi}\int_0^{2\pi}\left(\eta^2 - \left(\frac{\eta'}{2\eta}\right)^2 - \left(\frac{|\Delta\lambda|\sin 2\left(\varphi + \int_0^x \eta(t)dt\right)}{4\eta}\right)^2\right)dx \pm \frac{|\Delta\lambda|}{2}. \quad (23)$$

6. *Отображение* Φ_n *является гладким и обладает следующими свойствами:*

(а)*: действует послойно, т. е.* $\Phi_n(Y_n(\Delta\lambda)) = P_n(|\Delta\lambda|)$*;*

(б)*: сужение* $\Phi_n^0 := \Phi_n|_{Y_n^0} : Y_n^0 \to P_n^0$ *является гладким тривиальным расслоением, слой которого есть проективная прямая; тривиализующая коммутативная диаграмма такова:*

$$\begin{array}{ccc} Y_n^0 & \xrightarrow{\Phi_n^0} & P_n^0 \\ {\scriptstyle G_n^0}\downarrow & & \downarrow{\scriptstyle F_n^0} \\ H_n \times \mathbb{R}P^1 & \xrightarrow[\pi_n^0]{} & H_n \end{array};$$

(в)*: сужения* $\Phi_n^{\pm} := \Phi_n|_{Y_n^{\pm}} : Y_n^{\pm} \to P_{\lessgtr n} \subset P$ *являются диффеоморфизмами, причем* $\Phi_n^{\pm} = \Phi_n^{\mp} \cdot I_n$.

7. *Инволюция* I_n *подобна инволюции* J_n*, которая действует в* $H_n \times \mathbb{R}P^1 \times \mathbb{R}$*:*

$$I_n = G_n^{-1} \circ J_n \circ G_n, \text{ где } J_n(\eta, \varphi, \Delta\lambda) = (\eta, \varphi + \frac{\pi}{2}, -\Delta\lambda).$$

Инволюция I_n *является гладким автоморфизмом.*

Доказательство всех утверждений теоремы приведено в работах [1], [2] для случая непрерывного 2π-периодического потенциала $p \in C(2\pi)$. Без изменений доказательство переносится на случай $p \in L_2(2\pi)$.

В пространстве двумерных вещественных симметрических матриц рассмотрим плоскость из матриц с нулевым следом. Нас интересуют собственные значения $\gamma \in \mathbb{R}$ и собственные векторы $(\cos\varphi, \sin\varphi)$ $(\varphi \in \mathbb{R}P^1)$:

$$\begin{pmatrix} \xi_1 & \xi_2 \\ \xi_2 & -\xi_1 \end{pmatrix}\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} = \gamma\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} \Leftrightarrow \begin{cases} \xi_1 = \gamma\cos 2\varphi, \\ \xi_2 = \gamma\sin 2\varphi. \end{cases} \quad (24)$$

Сравнивая отображение π_n (см. (16) и (17)) с отображением (24), мы видим, что при фиксации осцилляционного параметра $n \in \mathbb{N}$ краевая задача отличается от двумерного случая на гомотопически тривиальный сомножитель H_n. Значит, в силу формул (21), переменные $(|\Delta\lambda|, 2\varphi)$ являются полярными координатами в плоскости, которая указана в тривиализации многообразия P(см. (16)); эту плоскость мы интерпретируем как плоскость двумерных вещественных симметрических матриц с нулевым следом. Ниже в полярных координатах описаны свойства функционалов собственных значений.

Предварительно напомним, что критическая точка функционала называется невырожденной, если в ней второй дифференциал не вырожден. Коиндексом Морса невырожденной критической точки

называется точная верхняя грань размерностей линейных подпространств, на которых положительно определен второй дифференциал [6].

Теорема 2. *Рассмотрим функционалы собственных значений*

$$\lambda_0,\ \lambda_n^{\pm}\ :\ P \to \mathbb{R}\ \ (n \in \mathbb{N}).$$

Их свойства таковы:

1. *Функционал λ_0 имеет одну невырожденную критическую точку $p = 0 \in P$, это максимум $\lambda_0(0) = 0$.*
2. *Для каждого $n \in \mathbb{N}$ функционалы $\lambda_n^{\pm}$ не вырождены на открытом подмногообразии $P_{\diagdown n} \subset P$.*
3. *Сужение $\lambda_n^0\ :\ P_n^0 \to \mathbb{R}$ функционала $\lambda_n^{\pm}$ на многообразие n-вырожденных потенциалов обладает свойствами:*

(а) *В параметризации $p = (F_n^0)^{-1}(\eta)$ сужение задается интегральным функционалом*

$$\lambda_n^0\ :\ H_n \to \mathbb{R},\quad \lambda_n^0(\eta) = \frac{1}{2\pi}\int_0^{2\pi}\left(\eta^2 - \left(\frac{\eta'}{2\eta}\right)^2\right)dx. \tag{25}$$

(б) *Сужение имеет единственную критическую точку $p = 0 \in P_n^0$.*

(в) *Касательное пространство $T_0P_n^0 \subset T_0P = P \subset L_2(2\pi)$ определяется в $L_2(2\pi)$ тремя условиями ортогональности*

$$T_0P_n^0 = \{q \in L_2(2\pi)\ :\ \int_0^{2\pi} q(x)\,dx = \int_0^{2\pi} q(x)\cdot\cos 2nx\,dx = \int_0^{2\pi} q(x)\cdot\sin 2nx\,dx = 0\}.$$

Обозначим через $T_0P_n^0(\leqslant 2n-1) \subset T_0P_n^0$ линейную оболочку конечной ортогональной системы функций $\{\cos x, \sin x, \dots, \cos(2n-1)x, \sin(2n-1)x\}$, а через $T_0P_n^0(\geqslant 2n+1) \subset T_0P_n^0$ — линейную оболочку бесконечной ортогональной системы функций $\{\cos(2n+1)x, \sin(2n+1)x, \dots\}$. Тогда касательное пространство есть прямая сумма $T_0P_n^0 = T_0P_n^0(\leqslant 2n-1) \oplus T_0P_n^0(\geqslant 2n+1)$ указанных ортогональных подпространств.

(г) *Критическая точка $p = 0 \in P_n^0$ является невырожденной. Коиндекс Морса критической точки $0 \in P_n^0$ равен $2(2n-1)$. На подпространстве $T_0P_n^0(\leqslant 2n-1)$ дифференциал $d^2\lambda_n^0(n)$ положительно определен, а на подпространстве $T_0P_n^0(\geqslant 2n+1)$ — отрицательно определен.*

4. *Пусть зафиксирован вырожденный потенциал $p \in P_n^0$, пусть $\Pi_p = \beta^{-1}(p) \cong \mathbb{R}^2$ — слой расслоения β* (см. *п.* 4 *теоремы* 1)*, пусть $\eta = F_n^0(p)$. Тогда:*

(а) *На плоскости Π_p функции $\lambda_n^{\pm}$ определяются в полярных координатах $(|\Delta\lambda|, 2\varphi)$* (см. (21)) *формулами* (23)*, где $\eta = F_n^0(p)$.*

(б) *При фиксированном угле φ функции $\lambda_n^{\pm} = \lambda_n^{\pm}(|\Delta\lambda|)$ изменяются по квадратичному закону $\lambda_n^{\pm} = \lambda_n^0(\eta) - a|\Delta\lambda|^2 \pm \frac{1}{2}|\Delta\lambda|$, где $\lambda_n^0(\eta)$ определяется по формуле* (25)*, а*

$$a = a(\eta, 2\varphi) = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{\sin 2\left(\varphi + \int_0^x \eta(t)dt\right)}{4\eta}\right)^2 dx > 0;$$

функция $\lambda_n^+ = \lambda_n^+(|\Delta\lambda|)$ имеет единственный максимум в точке

$$\Delta\lambda_n^+(\eta, \varphi) = \frac{\pi}{2}\left(\int_0^{2\pi}\left(\frac{\sin\left(2\varphi + 2\int_0^x \eta(t)dt\right)}{4\eta}\right)^2 dx\right)^{-1} \tag{26}$$

(см. фиг. 1).

(в) *Функционал* $\Delta\lambda_n^+(\eta, \varphi)$ *вырождается относительно переменной* φ (*т.е. не зависит от* φ), *если выполнены условия*

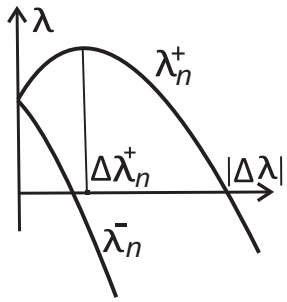
$$\eta \in H_n^* = \left\{ \eta \in H_n : \int_0^{2\pi} \frac{\cos 4\int_0^x \eta(t)dt}{\eta^2(t)} dx = 0, \int_0^{2\pi} \frac{\sin 4\int_0^x \eta(t)dt}{\eta^2(t)} dx = 0 \right\}. \tag{27}$$

Множество $H_n^* \neq \varnothing$; *в частности,* $\eta(x) \equiv n \in H_n^*$, *и в некоторой окрестности точки* $\eta(x) \equiv n$ *условия* (27) *задают гладкое подмногообразие в* H_n *коразмерности два.*

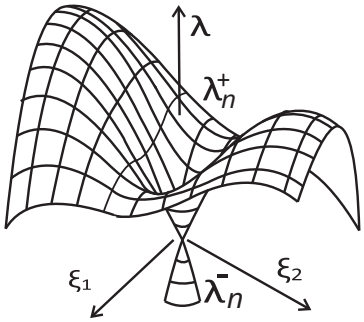
(г) *На плоскости* Π_p *функции* $\lambda_n^\pm$ *определяются в аффинных координатах* (ξ_1, ξ_2) (см. (21)) *формулами*

$$\lambda_n^\pm(\xi_1, \xi_2) = \lambda_n^0(\eta) - \frac{1}{2\pi}\int_0^{2\pi} \left(\frac{\xi_1 \sin 2\int\limits_0^x \eta(t)dt + \xi_2 \cos 2\int\limits_0^x \eta(t)dt}{4\eta} \right)^2 dx \pm \frac{1}{2}\sqrt{\xi_1^2 + \xi_2^2}. \tag{28}$$

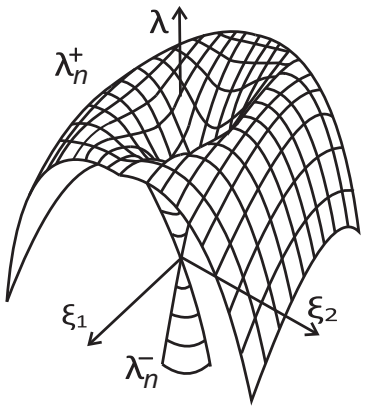
Функции $\lambda_n^\pm(\xi_1, \xi_2)$ *имеют в начале координат конические особенности. В невырожденном случае* $\eta \notin H_n^*$ *функция* $\lambda_n^+ = \lambda_n^+(\xi_1, \xi_2)$ *имеет два равных максимума и две седловых точки* (см. фиг. (2)). *В вырожденном случае* $\eta \in H_n^*$ *функция* $\lambda_n^+ = \lambda_n^+(\xi_1, \xi_2)$ *имеет окружность максимумов* $\xi_1^2 + \xi_2^2 = (\Delta\lambda_n^+(\eta))^2$ (см. фиг. 3).



Фиг. 1.



Фиг. 2.



Фиг. 3.

3. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2

3.1. Доказательство п. 1 теоремы 2

Поскольку собственное значение λ_0 всегда простое, воспользуемся формулами теории возмущения спектра в случае простого собственного значения (см. [5], гл. 5, § 13). Рассмотрим краевые задачи (1), (2) в окрестности потенциала p: пусть $\lambda = \lambda_0(p)$ — собственное значение, а $y = y_0(x) > 0$ — отвечающая ему нормированная собственная функция. Рассмотрим возмущенную краевую задачу

$$-\overline{y}'' + (p(x) + \varepsilon q(x))\overline{y} = \overline{\lambda}\overline{y}, \quad \overline{y}(0) - \overline{y}(2\pi) = \overline{y}'(0) - \overline{y}'(2\pi) = 0,$$

где $\varepsilon \in \mathbb{R}$, а $q \in L_2(2\pi)$ и $\int_0^{2\pi} q\,dx = 0$. Тогда локально возмущенное собственное значение и собственная функция допускают разложение в ряд по переменной ε:

$$\overline{\lambda} = \lambda + \varepsilon\mu + \varepsilon^2\sigma + \ldots, \quad \overline{y} = y + \varepsilon u(x) + \varepsilon^2 v(x) + \ldots .$$

Известно, что $\mu = \int_0^{2\pi} qy^2\,dx$. Возьмем $q = (yy')' \in T_pP$. Тогда

$$\mu = \int_0^{2\pi} (yy')'y^2\,dx = -2\int_0^{2\pi} (yy')^2\,dx = 0 \Leftrightarrow y(x) \equiv \text{const} = \frac{1}{\sqrt{2\pi}}.$$

Значит, если p — критическая точка функционала λ_0, то $p(x) \equiv 0$. В этом случае $\lambda = 0$.

Обратно, при $p(x) \equiv 0$ собственная функция $y = \text{const}$, поэтому при любом допустимом возмущении q верно $\mu = \int_0^{2\pi} qy^2\,dx = 0$.

В пространстве P тригонометрическая система $\{\cos kx, \sin kx\}_{k=1}^{\infty}$ образует базис. Чтобы проверить невырожденность функционала λ_0 в точке $p = 0$ достаточно рассмотреть возмущения потенциала по элементам базиса. Возьмем $q(x) = \cos kx$. Приравнивая выражения при ε и ε^2, и учитывая, что $\lambda = \mu = 0$, а $y(x) \equiv \frac{1}{\sqrt{2\pi}}$, получаем систему

$$-u'' + \cos kx \cdot \frac{1}{\sqrt{2\pi}} = 0, \quad -v'' + \cos kx \cdot u = \frac{\sigma}{\sqrt{2\pi}} \;\Rightarrow$$

$$u' = \frac{\sin kx}{\sqrt{2\pi}k}, \quad \sigma = -\int_0^{2\pi} (u'(x))2dx \;\Rightarrow\; \sigma = -\frac{1}{2k^2}.$$

3.2. Доказательство п. 2 теоремы 2

Теперь в краевой задаче (1), (2) число λ — простое собственное значение, номер которого $n \in \mathbb{N}$, а y — отвечающая ему нормированная знакопеременная собственная функция. Как и в предыдущем пункте, воспользуемся формулой теории возмущения: $\mu = \int_0^{2\pi} qy^2\,dx$ и возьмем $q = (yy')' \in T_pP$. Теперь

$$\mu = \int_0^{2\pi} (yy')'y^2\,dx = -2\int_0^{2\pi} (yy')^2\,dx < 0,$$

поскольку в окрестности нуля x_i собственной функции y верна оценка

$$y(x)y'(x) = (y'(x_i)(x - x_i) + o(x - x_i))(y'(x_i) + o(1)) > 0.$$

3.3. Доказательство пп. 3 (а) и 3 (б) теоремы 2

Согласно пп. 4 и 5 теоремы 1, исследование функционала $\lambda_n^0 = \lambda_n^0(p)$ равносильно его исследованию на многообразии H_n, т.е. сужению (25) функционала (20) при условии $\Delta\lambda = 0$.

Напомним, что подмногообразие $\widetilde{H}_n \subset W_2^2(2\pi)$ содержит положительные функции $\eta > 0$, удовлетворяющие условию (13), а подмногообразие $H_n \subset \widetilde{H}_n$ определяется еще двумя условиями (14) и (15). Заметим, что формула (25) определена и на многообразии $\widetilde{H}_n$. Оказывается, свойства функционала λ_n^0 на H_n и на $\widetilde{H}_n$ совпадают, точнее справедлива

Лемма 6. *Существует тривиальное гладкое расслоение $\chi : \widetilde{H}_n \to H_n$, слой которого диффеоморфен плоскости, и которое инвариантно относительно отображения $(F_n^0)^{-1}$, т.е. для любого $\tilde{\eta} \in \widetilde{H}_n$ вырожденный потенциал $p = (F_n^0)^{-1}(\tilde{\eta}) = (F_n^0)^{-1}(\chi(\tilde{\eta}))$. В частности, расслоение инвариантно относительно функционала λ_n^0, т.е. для любого $\tilde{\eta} \in \widetilde{H}_n$ верно $\lambda_n^0(\tilde{\eta}) = \lambda_n^0(\chi(\tilde{\eta}))$.*

Доказательство леммы приведено в [1] в Замечании 10 (там дана тривиализация расслоения χ).

Мы исследуем функционал λ_n^0 на $\widetilde{H}_n$, а затем переформулируем полученные результаты для многообразия H_n.

Итак, нас интересуют критические точки функционала (25) на открытом конусе положительных функций $\eta(x) > 0$ из пространства $W_2^2(2\pi)$ при условии $\frac{1}{2\pi}\int_0^{2\pi}(\eta(x) - n)dx = 0$. Функция Лагранжа такова:

$$L(\eta, \gamma) = \frac{1}{2\pi}\int_0^{2\pi}\left(\eta^2 - \left(\frac{\eta'}{2\eta}\right)^2 + \gamma(\eta - n)\right)dx, \quad \text{где } \gamma \in \mathbb{R}.$$

Уравнение Эйлера–Лагранжа

$$2\eta + \frac{\eta'^2}{2\eta^3} + \gamma + \frac{1}{2}\left(\frac{\eta'}{\eta^2}\right)' = 0 \tag{29}$$

имеет постоянное решение $\eta(x) \equiv -(\gamma/2)$. Значит, существует единственное постоянное решение $\eta(x) \equiv n$, а соответствующий множитель Лагранжа $\gamma = -2n$. Заметим, что мы получили решение, удовлетворяющее условиям (14) и (15), т.е. $\eta(x) \equiv n \in H_n$. Возвращаясь с помощью формулы (22), где $\Delta\lambda = 0$, к потенциалу, получаем нулевой потенциал $p(x) \equiv 0$.

Отыскивая непостоянные решения уравнения (29), замечаем, что интегрант не зависит от переменной интегрирования, поэтому уравнение Эйлера имеет интеграл:

$$\left(\eta^2 + \gamma\eta + \frac{1}{4}\left(\frac{\eta'}{\eta}\right)^2\right)' = 0 \Leftrightarrow \eta^2 + \gamma\eta + \frac{1}{4}\left(\frac{\eta'}{\eta}\right)^2 = C. \tag{30}$$

Уравнение (30) автономное, все решения которого гладкие; мы ищем замкнутые фазовые кривые в правой полуплоскости ($\eta > 0$), отвечающие решениям с периодом 2π, у которых на периоде $2n$ нулей. Пусть M, m — максимум и минимум искомого решения; тогда $M > n > m > 0$. Для M, m из (30) при условии $\eta' = 0$ получаем уравнение

$$\eta^2 + \gamma\eta - C = 0 \Leftrightarrow \gamma = -(m + M),\ C = -mM. \tag{31}$$

Поскольку уравнение (30) четное относительно производной η', период 2π содержит $2ln$ полупериодов, где натуральный параметр l предстоит определить. Будем искать решение на максимальном симметричном отрезке $[-\frac{\pi}{2ln}, \frac{\pi}{2ln}]$, на котором $\eta' > 0$ (остальные решения мы получим сдвигом аргумента), тогда

$$\eta' = 2\eta\sqrt{-\eta^2 + (M+m)\eta - Mm} \Rightarrow$$

$$2x = \frac{1}{\sqrt{Mm}}\arcsin\left(\frac{M+m}{M-m} - \frac{2Mm}{M-m}\frac{1}{\eta}\right) \Rightarrow$$

$$\eta(x) = \frac{2Mm}{(M+m) - (M-m)\sin\left(2\sqrt{Mm}x\right)},\ \text{где}\ -\frac{\pi}{4\sqrt{Mm}} \leqslant x \leqslant \frac{\pi}{4\sqrt{Mm}}.$$

Следовательно, $\frac{\pi}{4\sqrt{Mm}} = \frac{\pi}{2ln}$ и $ln = 2\sqrt{Mm}$.

С другой стороны, согласно условию (13), справедливо равенство

$$\int\limits_{-\frac{\pi}{2ln}}^{\frac{\pi}{2ln}} \frac{2Mm\,dx}{(M+m) - (M-m)\sin\left(2\sqrt{Mm}x\right)} = \frac{2\pi n}{2ln}.$$

Исключая параметр n, получаем

$$\sqrt{Mm}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{dt}{(M+m) - (M-m)\sin t} = \frac{\pi}{l} \Rightarrow l = 2 \Rightarrow Mm = n^2.$$

Таким образом, найдено однопараметрическое семейство (с параметром $m \in (0, n)$) непостоянных решений на отрезке $\left[-\frac{\pi}{4n}, \frac{\pi}{4n}\right]$:

$$\eta(x) = \frac{2n^2}{\left(\frac{n^2}{m} + m\right) - \left(\frac{n^2}{m} - m\right)\sin(2nx)}.$$

По этой же формуле определяются решения уравнения (30) при условиях (31) на $\mathbb{R}$ (проверяется прямой проверкой). Остальные решения получаются сдвигом аргумента. Наконец, прямой проверкой с помощью формулы (22), где $\Delta\lambda = 0$, убеждаемся, что полученное двупараметрическое семейство решений порождает нулевой потенциал $p(x) \equiv 0$. Пункты 3 (а) и 3 (б) теоремы 2 доказаны.

3.4. Доказательство п. 3 (в) теоремы 2

Имея в виду, что критическая точка $0 \in P_n^0$ отвечает точке $n \in H_n$, мы сначала найдем касательное пространство T_nH_n. С этой целью мы линеаризуем функционалы (13), (14), (15) в точке $\eta(x) \equiv n$. Получим

$$T_nH_n = \left\{ \nu \in W_2^2(2\pi) : \int_0^{2\pi} \nu(x)\,dx = \int_0^{2\pi} \nu(x)\cdot\cos 2nx\,dx = \int_0^{2\pi} \nu(x)\cdot\sin 2nx\,dx = 0 \right\}. \tag{32}$$

Теперь найдем производную отображения $\left(F_n^0\right)^{-1}$ в точке $n \in H_n$, т.е. линеаризуем отображение (22) по переменной η в точке $n \in H_n$ при условии $\Delta\lambda = 0$:

$$D\left(F_n^0\right)^{-1}(n)\nu = -\frac{\nu''}{2n} - 2n\nu. \tag{33}$$

Из п. 4 (в) и п. 6 теоремы 1 следует, что образ $D\left(F_n^0\right)^{-1}(n)\left(T_nH_n\right) = T_0P_n^0$. Но, согласно (32) и (33), $D\left(F_n^0\right)^{-1}(n)\left(T_nH_n\right) = T_nH_n$, что доказывает п. 3 (в) теоремы.

3.5. Доказательство п. 3 (г) теоремы 2

Доказательство осуществим в терминах многообразия H_n. В силу леммы 6 и результатов п. 4.3 о критической точке $\eta = n$, достаточно рассмотреть функцию Лагранжа $L(\eta, \gamma)$ при $\gamma = -2n$. В произвольной точке $\eta \in H_n$ второй дифференциал функции Лагранжа совпадает со вторым дифференциалом функционала λ_n^0:

$$d^2L(\eta, -2n)\nu = d^2\lambda_n^0(\eta)\nu = \frac{1}{2\pi}\int_0^{2\pi}\left(2\nu^2 - \frac{\nu'^2}{2\eta^2} + \frac{2\eta'}{\eta^3}\nu\nu' - \frac{3\eta'^2}{2\eta^4}\nu^2\right)dx.$$

В критической точке второй дифференциал равен

$$d^2\lambda_n^0(n)\nu = \frac{1}{2\pi}\int_0^{2\pi}\left(2\nu^2 - \frac{\nu'^2}{2n^2}\right)dx.$$

В силу определения (32), произвольная функция $\nu \in T_n(H_n)$ представима рядом Фурье

$$T_n(H_n) \ni \nu(x) = \sum_{k\neq 2n}^{\infty} a_k\cos kx + b_k\sin kx, \quad k \in \mathbb{N}.$$

Подставляя предыдущее выражение в формулу второго дифференциала, получаем

$$d^2\lambda_n^0(n)\nu = \frac{1}{4n^2}\sum_{k\neq 2n}^{\infty}\left((2n)^2 - k^2\right)\left(a_k^2 + b_k^2\right), \quad k \in \mathbb{N},$$

что доказывает п. 3 (г).

3.6. Доказательство п. 4 теоремы 2

Пункты 4 (а) и 4 (б) очевидны.

Вырожденность функционала (26) определяется тождеством

$$\frac{\partial}{\partial\varphi}\int_0^{2\pi}\left(\frac{\sin\left(2\varphi+2\int_0^x\eta(t)dt\right)}{4\eta}\right)^2 dx\equiv 0,$$

которое равносильно системе (27). Принадлежность $\eta(x)\equiv n\in H_n^*$ очевидна. Независимость в точке $\eta(x)\equiv n\in H_n^*$ четырех интегральных функционалов, задающих условия (14), (15) и (27) проверяется их линеаризацией в этой точке.

Доказательство п. 4 (г) сводится к исследованию функционала (28) на плоскости с координатами (ξ_1,ξ_2).

Автор благодарен А.О. Ремизову за плодотворное обсуждение и многочисленные полезные советы.

СПИСОК ЛИТЕРАТУРЫ

1. *Дымарский Я. М., Евтушенко Ю. А.* Расслоение пространства периодических краевых задач на гиперповерхности постоянной длины n-й спектральной лакуны // Матем. сб. 2016. **207**-5. С. 43–68.
2. *Дымарский Я. М., Бондарь А. А.* Многообразие собственных функций семейства периодических краевых задач // Матем. сб. 2021. **212**-9. С. 18–39.
3. *Левитан Б. М., Саргсян И. С.* Введение в спектральную теорию. М.: Наука, 1976.
4. *Ладыженская О. А.* Краевые задачи математической физики. М.: Наука, 1973.
5. *Курант Р., Гильберт Д.* Методы математической физики. Часть первая. М.-Л.: ГТТИ, 1933.
6. *Иллс Дж.* Основания глобального анализа // Успехи матем. наук. 1969. Т. XXIV, Вып. 3(147). С. 157–210.

EIGENVALUE FUNCTIONALS ON THE MANIFOLD OF POTENTIALS

Ya. M. Dymarsky*

Moscow Institute of Physics and Technology (National Research University), Institutsky Lane., 9, Dolgoprudnyi, Moscow oblast, 141701 Russia

**e-mail: dymarskii@mail.ru*

Received 12 December, 2023
Revised 20 February, 2024
Accepted 05 March, 2024

Abstract. The article gives an analytical and topological description of the eigenvalue functional on the manifold of periodic potentials.

Keywords: space of periodic boundary value problems, eigenvalue functional, bundle of potential manifolds.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1028–1041

УДК 517.956.4

О НАЧАЛЬНО-КРАЕВЫХ ЗАДАЧАХ ДЛЯ ПАРАБОЛИЧЕСКИХ СИСТЕМ В ПОЛУОГРАНИЧЕННОЙ ПЛОСКОЙ ОБЛАСТИ С ГРАНИЧНЫМИ УСЛОВИЯМИ ОБЩЕГО ВИДА

© 2024 г. С. И. Сахаров[1,*]

[1] *119991 Москва, Ленинские горы, 1, МГУ им. М. В. Ломоносова, Московский центр фундаментальной и прикладной математики, Россия*

[*] *e-mail: ser341516@yandex.ru*

Поступила в редакцию 14.12.2023 г.
Переработанный вариант 06.02.2024 г.
Принята к публикации 05.03.2024 г.

Рассмотрены начально-краевые задачи для однородных параболических систем с Дини-непрерывными коэффициентами при нулевых начальных условиях в полуограниченной плоской области с негладкой боковой границей, допускающей наличие “клювов”, на которой задаются граничные условия общего вида с переменными коэффициентами. Методом граничных интегральных уравнений доказана теорема об однозначной классической разрешимости таких задач в пространстве функций, непрерывных и ограниченных вместе со своей пространственной производной первого порядка в замыкании области. Дано представление полученных решений в виде векторных потенциалов простого слоя. Библ. 28.

Ключевые слова: параболические системы, начально-краевые задачи, негладкая боковая граница, граничные интегральные уравнения, условие Дини.

DOI: 10.31857/S0044466924060115, **EDN:** XYFZEM

ВВЕДЕНИЕ

Однозначная разрешимость начально-краевых задач для параболических систем с гёльдеровскими коэффициентами в пространстве $H^{2+\alpha,1+\alpha/2}(\overline{\Omega})$, $0 < \alpha < 1$, в областях с гладкими боковыми границами следует из общей теории параболических задач (см. [1], [2, с. 706]). В этих работах требовалось, чтобы коэффициенты в граничных условиях первого и третьего рода принадлежали пространствам $H^{1+\alpha/2}[0,T]$ и $H^{(1+\alpha)/2}[0,T]$ соответственно.

В случае параболических систем с гёльдеровскими коэффициентами первая и вторая начально-краевые задачи в плоских областях с негладкими боковыми границами из класса Жевре $H^{(1+\alpha)/2}$, допускающими, в частности, наличие “клювов”, рассматривались в [3]–[5]. В более общем случае параболических систем с коэффициентами, удовлетворяющими двойному условию Дини, в [6]–[10] доказаны теоремы о существовании и единственности решений из пространства $C^{1,0}(\overline{\Omega})$ (см. раздел 1) первой, второй и смешанной начально-краевых задач в областях с боковыми границами из класса Дини–Гёльдера $H^{1/2+\omega}$, где ω – модуль непрерывности, удовлетворяющий условию Дини.

Естественно возникает вопрос о получении аналогичных результатов и в случае начально-краевых задач для параболических систем с краевыми условиями общего вида, когда на боковых границах задаются условия как первого, так и третьего рода. Однозначная классическая разрешимость в пространстве $C^{1,0}(\overline{\Omega})$ таких задач для однородных параболических систем с Дини-непрерывными коэффициентами и нулевыми начальными условиями была доказана в [11], где рассматривалась полуограниченная область с негладкой боковой границей из класса Дини–Гёльдера $H^{1/2+\omega}$. В этой работе

предполагалось, что коэффициенты в граничных условиях являются постоянными. Было сформулировано алгебраическое условие однозначной разрешимости поставленных задач и доказана его эквивалентность известному условию дополнительности.

В настоящей работе результаты [11] обобщаются на случай граничных условий, в которых коэффициенты являются переменными функциями. При этом непрерывные на $[0,T]$ коэффициенты в граничных условиях первого рода удовлетворяют условию Дини–Гёльдера на каждом отрезке $[t_1,T]$, $t_1 \in (0,T)$, и могут не удовлетворять этому условию в окрестности $t=0$. Старшие коэффициенты в граничных условиях третьего рода являются непрерывными на отрезке $[0,T]$, а младшие коэффициенты в этих условиях могут расти определенным образом при $t \to +0$. Показано также, что условия на характер непрерывности правых частей в граничных условиях являются точными для разрешимости поставленных задач в пространстве $C^{1,0}(\overline{\Omega})$.

Задачи для параболических систем моделируют, в частности, процессы тепло- и массопереноса в многокомпонентных материалах (см., например, [12]–[19]). При этом характер негладкости боковой границы области моделирует, в частности, резкое изменение границ некоторых металлов, связанное с фазовыми переходами при изменении температуры (см., например, [20]).

Работа состоит из четырех разделов. В разделе 1 приводятся необходимые определения и формулируются основные результаты. В разделе 2 доказывается теорема об однозначной разрешимости в пространстве $C[0,T]$ систем граничных интегральных уравнений, к которым редуцируются поставленные задачи. В разделе 3 доказывается основная теорема об однозначной разрешимости рассматриваемых задач. В разделе 4 доказывается точность условий на характер непрерывности правых частей в граничных условиях.

1. НЕОБХОДИМЫЕ СВЕДЕНИЯ И ФОРМУЛИРОВКА ОСНОВНОГО РЕЗУЛЬТАТА

Следуя [21, с. 151], *модулем непрерывности* называем непрерывную, неубывающую, полуаддитивную функцию $\omega\colon [0,+\infty) \to \mathbb{R}$, для которой $\omega(0)=0$. Говорят, что модуль непрерывности ω удовлетворяет *условию Дини*, если

$$\widetilde{\omega}(z) = \int_0^z y^{-1}\omega(y)\,dy < +\infty, \qquad z>0. \tag{1}$$

Через $\mathscr{D}$ обозначим множество модулей непрерывности, которые удовлетворяют условию Дини (1).

Пусть число $T>0$ фиксировано. Через $C[0,T]$ обозначим пространство непрерывных на $[0,T]$ (вектор-)функций с нормой $\|\psi;[0,T]\|^0 = \max\limits_{t\in[0,T]}|\psi(t)|$. Положим $\overset{\circ}{C}[0,T] = \{\psi \in C[0,T]:\ \psi(0)=0\}$.

Здесь и далее для числового вектора a (числовой матрицы A) под $|a|$ (соответственно $|A|$) понимаем максимум из модулей его компонент (её элементов).

Пусть ω – некоторый модуль непрерывности. Введем пространства $H^{q+\omega}[0,T] = \Big\{\psi \in C[0,T]\colon |\psi;[0,T]\|^{q+\omega} = \|\psi;[0,T]\|^0 + \sup\limits_{\substack{t,t+\Delta t\in(0,T),\\ \Delta t\neq 0}} \frac{|\Delta_t\psi(t)|}{|\Delta t|^q\omega(|\Delta t|^{1/2})} < \infty\Big\}$, $\overset{\circ}{H}{}^{q+\omega}[0,T] = \{\psi \in H^{q+\omega}[0,T]\colon \ \psi(0)=0\}$, $q=0,1/2$, где $\Delta_t\psi(t) = \psi(t+\Delta t)-\psi(t)$.

Пусть

$$\partial^{1/2}\varphi(t) = \frac{1}{\sqrt{\pi}}\frac{d}{dt}\int_0^t (t-\tau)^{-1/2}\varphi(\tau)\,d\tau, \qquad t\in[0,T],$$

есть оператор дробного дифференцирования порядка $1/2$. Следуя [3], [4], введем пространство $\overset{\circ}{C}{}^{1/2}[0,T] = \{\psi \in \overset{\circ}{C}[0,T]:\ \partial^{1/2}\psi \in \overset{\circ}{C}[0,T],\ \|\psi;[0,T]\|^{1/2} = \|\psi;[0,T]\|^0 + \|\partial^{1/2}\psi;[0,T]\|^0 < \infty\}$.

Функция $\nu(z)$, $z>0$, называется *почти убывающей*, если для некоторой постоянной $C>0$ выполняется неравенство $\nu(z_1) \leqslant C\nu(z_2)$, $z_1 \geqslant z_2 > 0$.

В полосе $D = \{(x,t) \in \mathbb{R}^2 : \ x \in \mathbb{R},\ t \in (0,T)\}$ выделяется область $\Omega = \{(x,t) \in D : \ x > g(t)\}$, где g удовлетворяет условию

$$g \in H^{1/2+\omega_1}[0,T], \qquad \omega_1 \in \mathscr{D}, \tag{2}$$

причем для некоторого $\varepsilon_1 \in (0,1)$ функция $z^{-\varepsilon_1}\omega_1(z),\ z > 0$, почти убывает.

Через $C^0(\overline{\Omega})$ обозначим пространство непрерывных и ограниченных в $\overline{\Omega}$ (вектор-)функций с нормой $\|u;\Omega\|^0 = \sup\limits_{(x,t)\in\Omega} |u(x,t)|$.

Под значениями функций и их производных на границе области $\Omega \subset \mathbb{R}^2$ понимаем их предельные значения "изнутри" Ω.

Положим $C^{1,0}(\overline{\Omega}) = \{u \in C^0(\overline{\Omega}) : \ \partial_x u \in C^0(\overline{\Omega})\}$, $\|u;\Omega\|^{1,0} = \sum\limits_{l=0}^{1} \|\partial_x^l u;\Omega\|^0$, $\overset{0}{C}{}^{1,0}(\overline{\Omega}) = \{u \in C^{1,0}(\overline{\Omega}) : \ \partial_x^l u(x,0) = 0,\ l = 0,1\}$.

Пусть число $m \in \mathbb{N}$ фиксировано. Рассмотрим в D равномерно параболический по И. Г. Петровскому оператор

$$Lu = \partial_t u - \sum_{l=0}^{2} A_l(x,t)\partial_x^l u, \qquad u = (u_1,\dots,u_m)^{\mathrm{T}},$$

где $A_l = \|a_{jkl}\| - m \times m$-матрицы, элементами которых являются вещественные функции, определенные и ограниченные в $\overline{D}$, и выполнены условия:

a) собственные числа $\mu_r,\ r = \overline{1,m}$, матрицы A_2 подчиняются неравенствам $\mathrm{Re}\mu_r(x,t) \geqslant \delta$ для некоторого $\delta > 0$ и всех $(x,t) \in \overline{D}$;

b) $|a_{jkl}(x+\Delta x, t+\Delta t) - a_{jkl}(x,t)| \leqslant \omega_0(|\Delta x| + |\Delta t|^{1/2}),\ (x+\Delta x, t+\Delta t),\ (x,t) \in \overline{D}$, где ω_0 — модуль непрерывности, удовлетворяющий двойному условию Дини

$$\widetilde{\widetilde{\omega}}_0(z) = \int_0^z y^{-1}dy \int_0^y x^{-1}\omega_0(x)dx < +\infty, \qquad z > 0,$$

причем для некоторого $\varepsilon_0 \in (0,1)$ функция $z^{-\varepsilon_0}\omega_0(z),\ z > 0$, почти убывает.

Пусть

$$Z(x,t;A_2(\xi,\tau)) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ixy}\exp\{-y^2 t A_2(\xi,\tau)\}dy, \qquad (x,t) \in \mathbb{R}\times(0,+\infty),\ (\xi,\tau) \in \overline{D}. \tag{3}$$

Обозначим $D^* = \{(x,t;\xi,\tau) \in \overline{D}\times\overline{D} : \ t > \tau\}$. Известно (см. [22]), что при выполнении условий a) и b) у системы $Lu = 0$ существует фундаментальная матрица решений $\Gamma(x,t;\xi,\tau),\ (x,t;\xi,\tau) \in D^*$, справедливы оценки

$$|\partial_t^k\partial_x^l\Gamma(x,t;\xi,\tau)| \leqslant C(t-\tau)^{-(2k+l+1)/2}\exp(-c(x-\xi)^2/(t-\tau)),\ (x,t;\xi,\tau) \in D^*,\ 2k+l \leqslant 2, \tag{4}$$

и, кроме того, для разности

$$W(x,t;\xi,\tau) \equiv \Gamma(x,t;\xi,\tau) - Z(x-\xi,t-\tau;A_2(\xi,\tau)),\ (x,t;\xi,\tau) \in D^*, \tag{5}$$

выполнены оценки

$$|\partial_t^k\partial_x^l W(x,t;\xi,\tau)| \leqslant C\widetilde{\omega}_0\left((t-\tau)^{1/2}\right)(t-\tau)^{-(2k+l+1)/2}\exp\left(\frac{-c(x-\xi)^2}{t-\tau}\right),\ 2k+l \leqslant 2, \tag{6}$$

$$|\Delta_t\partial_x^l W(x,t;\xi,\tau)| \leqslant C(\Delta t)^{1-l/2}\widetilde{\omega}_0\left((t-\tau)^{1/2}\right)(t-\tau)^{-3/2}\exp\left(\frac{-c(x-\xi)^2}{t-\tau}\right), \tag{7}$$

$(x,t;\xi,\tau),(x,t+\Delta t;\xi,\tau)\in D^*,\ 0<\Delta t\leqslant t-\tau,\ l=0,1.$

Пусть $m_j\in\mathbb{N}\cup\{0\},\ j=0,1$, и $m_0+m_1=m$. Пусть заданы $m_0\times m$-матрица $B_0(t)=\|b_{jk0}(t)\|$ и $m_1\times m$-матрицы $B_1(t)=\|b_{jk1}(t)\|$, $\hat{B}_1(t)=\|\hat{b}_{jk1}(t)\|$, где функции b_{jk0}, b_{jk1} и $\hat{b}_{jk1}$ удовлетворяют условиям

$$b_{jk0}\in C[0,T],\ |\Delta_t b_{jk0}(t)|\leqslant\frac{|\Delta t|^{1/2}\omega_1(|\Delta t|^{1/2})}{\min\{t^{1/2},(t+\Delta t)^{1/2}\}},\ t,t+\Delta t\in(0,T], \tag{8}$$

$$b_{jk1}\in C[0,T], \tag{9}$$

$$|\hat{b}_{jk1}(t)|\leqslant\frac{\omega_2(t^{1/2})}{t^{1/2}},\ |\Delta_t\hat{b}_{jk1}(t)|\leqslant\frac{\omega_2(|\Delta t|^{1/2})}{\min\left\{t^{1/2},(t+\Delta t)^{1/2}\right\}},\ t,t+\Delta t\in(0,T], \tag{10}$$

где ω_2 — некоторый модуль непрерывности.

Рассмотрим задачу о нахождении (вектор-)функции $u\in\overset{1,0}{C_0}(\overline{\Omega})$, являющейся регулярным решением системы

$$Lu(x,t)=0,\qquad (x,t)\in\Omega, \tag{11}$$

удовлетворяющей начальному условию

$$u(x,0)=0,\qquad x\geqslant g(0), \tag{12}$$

и граничным условиям

$$B_0(t)u(g(t),t)=\psi_0(t), \tag{13}$$

$$B_1(t)\partial_x u(g(t),t)+\hat{B}_1(t)u(g(t),t)=\psi_1(t),\qquad t\in[0,T]. \tag{14}$$

Замечание 1. В состав граничных условий (13), (14) для компонент искомой (вектор-)функции u входит m_0 граничных условий первого рода и m_1 граничных условий третьего рода, причем $m_0+m_1=m$, где m — порядок матриц $A_l,\ l=0,1,2$, задающих коэффициенты оператора L.

Положим

$$A(t)=A_2(g(t),t),\quad M(t)=\frac{1}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty}\exp\{-y^2A(t)\}dy, \tag{15}$$

$$G(t)=\begin{pmatrix}B_0\\B_1M\end{pmatrix}(t),\qquad t\in[0,T]. \tag{16}$$

Заметим, что из равенства (см. [23])

$$M^2(t)=(A(t))^{-1},\qquad t\in[0,T], \tag{17}$$

следует, что

$$\det M(t)\neq 0,\qquad t\in[0,T]. \tag{18}$$

Основным результатом настоящей работы является следующая теорема.

Теорема 1. *Пусть выполнены условия а), b),* (2), (8)–(10) *и, кроме того,*

$$\det G(t)\neq 0,\qquad t\in[0,T]. \tag{19}$$

Тогда для любых $\psi_0 \in \underset{0}{C}^{1/2}[0,T]$ *и* $\psi_1 \in \underset{0}{C}[0,T]$ *существует единственное регулярное решение* $u \in \underset{0}{C}^{1,0}(\overline{\Omega})$ *задачи* (11)–(14)*, и справедлива оценка*

$$\|u;\Omega\|^{1,0} \leqslant C\{\|\psi_0;[0,T]\|^{1/2} + \|\psi_1;[0,T]\|^0\}. \tag{20}$$

При этом для решения $u \in \underset{0}{C}^{1,0}(\overline{\Omega})$ *задачи* (11)–(14) *справедливо интегральное представление в виде (векторного) параболического потенциала простого слоя*

$$u(x,t) = \int_0^t \Gamma(x,t;g(\tau),\tau)\varphi(\tau)d\tau, \qquad (x,t) \in \overline{\Omega}, \tag{21}$$

где $\varphi \in \underset{0}{C}[0,T]$ *— единственное в* $C[0,T]$ *решение системы граничных интегральных уравнений Вольтерры I и II рода*

$$B_0(t)\int_0^t \Gamma(g(t),t;g(\tau),\tau)\varphi(\tau)d\tau = \psi_0(t), \tag{22}$$

$$\begin{aligned} -\frac{1}{2}B_1(t)(A(t))^{-1}\varphi(t) + \int_0^t [B_1(t)\partial_x\Gamma(g(t),t;g(\tau),\tau) + \\ +\hat{B}_1(t)\Gamma(g(t),t;g(\tau),\tau)]\varphi(\tau)d\tau = \psi_1(t), \qquad t \in [0,T]. \end{aligned} \tag{23}$$

Здесь и далее через C, c обозначаем положительные постоянные, зависящие от чисел T, m, коэффициентов оператора L, элементов матриц B_j, $j = 0,1$, $\hat{B}_1$ и модуля непрерывности ω_1.

Замечание 2. В случае постоянных матриц $B_0(t) \equiv B_0$, $B_1(t) \equiv B_1$ и $\hat{B}_1(t) \equiv \hat{B}_1$, $t \in [0,T]$, теорема 1 доказана в [11].

Замечание 3. В [11] показано, что условие (19) эквивалентно известному (см. [2, с. 700], [24, с. 360]) условию дополнительности: для произвольно фиксированных чисел $p \in \mathbb{C}$, $\mathrm{Re}\, p > 0$, и $t \in [0,T]$

$$\det\begin{pmatrix} B_0(t) \int\limits_{\gamma^+(p,t)} (pE + y^2A(t))^{-1}dy \\ B_1(t) \int\limits_{\gamma^+(p,t)} y(pE + y^2A(t))^{-1}dy \end{pmatrix} \neq 0,$$

где $\gamma^+(p,t)$ — произвольный простой замкнутый контур, содержащийся в полуплоскости $\{\mathrm{Im}\, y > 0\}$ и охватывающий корни уравнения

$$\det\left(pE + y^2A(t)\right) = 0,$$

имеющие положительную мнимую часть (обход кривой $\gamma^+(p,t)$ предполагается направленным против часовой стрелки).

Замечание 4. Условия $\psi_0 \in \underset{0}{C}^{1/2}[0,T]$ и $\psi_1 \in \underset{0}{C}[0,T]$ являются точными для существования решения задачи (11)–(14) из пространства $\underset{0}{C}^{1,0}(\overline{\Omega})$ (см. ниже раздел 4).

2. СИСТЕМА ГРАНИЧНЫХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

В следующих двух леммах приведем известные результаты, которые будут использованы в дальнейшем.

Лемма 1 (см. [25]). *Пусть* $\omega \in \mathscr{D}$. *Тогда* $\partial^{1/2}$ *является ограниченным оператором из* $\underset{0}{H}^{1/2+\omega}[0,T]$ *в* $\underset{0}{H}^{\widetilde{\omega}}[0,T]$.

Следуя А. Н. Тихонову (см. [26]), назовем оператор $K : C[0,T] \to C[0,T]$ *вольтерровым*, если для любого $t \in [0,T]$ из равенства $\varphi_1 = \varphi_2$ на $[0,t]$ следует, что $K\varphi_1 = K\varphi_2$ на $[0,t]$.

Лемма 2 (см. [27]). *Пусть* ω *— некоторый модуль непрерывности, и* $K : C[0,T] \to \overset{\omega}{H}_0[0,T]$ *— линейный ограниченный вольтерров оператор. Тогда для любой (вектор-)функции* $\psi \in C[0,T]$ *уравнение* $\varphi + K\varphi = \psi$ *имеет единственное решение* $\varphi \in C[0,T]$*, и справедлива оценка* $\|\varphi;[0,T]\|^0 \leqslant C\|\psi;[0,T]\|^0$.

Рассмотрим систему граничных интегральных уравнений (22), (23). Полагая (см. (15))

$$N_0(t,\tau) = B_0(\tau)[\Gamma(g(t),t;g(\tau),\tau) - Z(0,t-\tau;A(\tau))] + [B_0(t) - B_0(\tau)]\Gamma(g(t),t;g(\tau),\tau), \tag{24}$$

$$N_1(t,\tau) = B_1(t)\partial_x\Gamma(g(t),t;g(\tau),\tau) + \hat{B}_1(t)\Gamma(g(t),t;g(\tau),\tau), \qquad 0 \leqslant \tau < t \leqslant T,$$

систему (22), (23) можно записать в виде

$$\int_0^t B_0(\tau)Z(0,t-\tau;A(\tau))\varphi(\tau)d\tau + \int_0^t N_0(t,\tau)\varphi(\tau)d\tau = \psi_0(t), \tag{25}$$

$$-\frac{1}{2}B_1(t)(A(t))^{-1}\varphi(t) + \int_0^t N_1(t,\tau)\varphi(\tau)d\tau = \psi_1(t), \qquad t \in [0,T]. \tag{26}$$

Пусть оператор дробного интегрирования действует на $\varphi \in C[0,T]$ по формуле

$$I^{1/2}\varphi(t) = \frac{1}{\sqrt{\pi}}\int_0^t (t-\tau)^{-1/2}\varphi(\tau)d\tau, \qquad t \in [0,T].$$

В силу (3) и (15)

$$Z(0,t-\tau;A(\tau)) = \frac{1}{2\sqrt{\pi(t-\tau)}}M(\tau), \qquad 0 \leqslant \tau < t \leqslant T,$$

поэтому уравнение (25) может быть переписано в виде

$$\frac{1}{2}I^{1/2}(B_0M\varphi)(t) + \int_0^t N_0(t,\tau)\varphi(\tau)d\tau = \psi_0(t), \qquad t \in [0,T]. \tag{27}$$

Введем операторы H_0 и H_1, действующие на функции $\varphi \in C[0,T]$ по формулам

$$(H_l\varphi)(t) = \int_0^t N_l(t,\tau)\varphi(\tau)d\tau, \qquad t \in [0,T],\ l = 0,1,$$

и перепишем систему (27), (26) в виде

$$\frac{1}{2}I^{1/2}(B_0M\varphi)(t) + (H_0\varphi)(t) = \psi_0(t), \tag{28}$$

$$-\frac{1}{2}B_1(t)(A(t))^{-1}\varphi(t) + (H_1\varphi)(t) = \psi_1(t),\ t \in [0,T]. \tag{29}$$

Пусть

$$\omega_3(z) = \sup_{\substack{t,t+\Delta t\in[0,T],\\ |\Delta t|\leqslant z}} |\Delta_t B_1(t)|, \qquad z \in [0,T],\ \omega_3(z) = \omega_3(T),\ z \geqslant T.$$

Функция ω_3 удовлетворяет определению модуля непрерывности. Положим $\omega_4(z) = \widetilde{\omega}_0(z) + \omega_1(z)$, $\omega_5(z) = \widetilde{\omega}_4(z) + \omega_2(z) + \omega_3(z)$, $z \geqslant 0$.

Лемма 3. *Пусть выполнены условия* a), b), (2) и (8). *Тогда* H_0 *является ограниченным оператором из* $C[0,T]$ *в* $\overset{0}{H}{}^{1/2+\omega_4}[0,T]$.

Доказательство. Здесь и далее полагаем $C_\varphi = C\|\varphi;[0,T]\|^0$. Достаточно доказать справедливость следующих оценок:

$$|(H_0\varphi)(t)| \leqslant C_\varphi t^{1/2}\omega_4(t^{1/2}), \tag{30}$$

$$|\Delta_t(H_0\varphi)(t)| \leqslant C_\varphi(\Delta t)^{1/2}\omega_4((\Delta t)^{1/2}), \qquad t, t+\Delta t \in [0,T],\ \Delta t > 0. \tag{31}$$

Имеет место представление (см. (24))

$$(H_0\varphi)(t) = \int_0^t B_0(\tau)[\Gamma(g(t),t;g(\tau);\tau) - Z(0,t-\tau;A(\tau))]\varphi(\tau)d\tau +$$

$$+ \int_0^t [B_0(t) - B_0(\tau)]\Gamma(g(t),t;g(\tau),\tau)\varphi(\tau)d\tau \equiv (H_{01}\varphi)(t) + (H_{02}\varphi)(t), \qquad t \in [0,T].$$

Для $H_{01}\varphi$ оценки (30), (31) следуют из результатов [28]. Докажем справедливость оценок (30), (31) для $H_{02}\varphi$.

Положим

$$N_{02}(t,\tau) = [B_0(t) - B_0(\tau)]\Gamma(g(t),t;g(\tau),\tau), \qquad 0 \leqslant \tau < t \leqslant T.$$

В силу оценок (4) и условия (8), справедливо неравенство

$$|N_{02}(t,\tau)| \leqslant C\omega_1((t-\tau)^{1/2})\tau^{-1/2}, \qquad 0 < \tau < t \leqslant T, \tag{32}$$

из которого следует оценка (**??**) для $H_{02}\varphi$:

$$|(H_{02}\varphi)(t)| \leqslant C_\varphi \int_0^t \omega_1((t-\tau)^{1/2})\tau^{-1/2}d\tau \leqslant C_\varphi t^{1/2}\omega_1(t^{1/2}), \qquad t \in [0,T].$$

Справедливость оценки (31) для $H_{02}\varphi$, в силу (30), достаточно доказать в случае $0 < \Delta t < t/2$. Положим

$$\Delta_t(H_{02}\varphi)(t) = \sum_{j=0}^{1}(-1)^{j+1}\int_{t-\Delta t}^{t+j\Delta t} N_{02}(t+j\Delta t,\tau)\varphi(\tau)d\tau +$$

$$+ \int_0^{t-\Delta t} \Delta_t N_{02}(t,\tau)\varphi(\tau)d\tau \equiv \sum_{j=0}^{2} R_j(t,\Delta t), \qquad t, t+\Delta t \in [0,T],\ 0 < \Delta t < \frac{t}{2}.$$

Используя (32), получаем оценки для интегралов $R_j,\ j = 0,1$:

$$\begin{aligned} |R_j(t,\Delta t)| &\leqslant C_\varphi \int_{t-\Delta t}^{t+j\Delta t} \omega_1((t+j\Delta t-\tau)^{1/2})\tau^{-1/2}d\tau \leqslant \\ &\leqslant C_\varphi(\Delta t)^{1/2}\omega_1\big((\Delta t)^{1/2}\big), \qquad t, t+\Delta t \in [0,T],\ 0 < \Delta t < \tfrac{t}{2}. \end{aligned} \tag{33}$$

Рассмотрим интеграл R_2. В силу условия (2) и оценок (4), имеем

$$|\Delta_t\Gamma(g(t),t;g(\tau),\tau)| \leqslant C[(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})(t-\tau)^{-1} + \Delta t(t-\tau)^{-3/2}], \tag{34}$$

$0 \leqslant \tau < t < t + \Delta t \leqslant T,\ \Delta t \leqslant t - \tau$. Отсюда, используя (4) и (8), последовательно получаем

$$|\Delta_t N_{02}(t,\tau)| \leqslant$$

$$\leqslant C\big[(\Delta t)^{1/2}\omega_1\big((\Delta t)^{1/2}\big)(t-\tau)^{-1/2}t^{-1/2} + (\Delta t)^{1/2}\omega_1\big((\Delta t)^{1/2}\big)\omega_1\big((t-\tau)^{1/2}\big)(t-\tau)^{-1/2}\tau^{-1/2} +$$

$$+(\Delta t)\omega_1\big((t-\tau)^{1/2}\big)(t-\tau)^{-1}\tau^{-1/2}\big] \leqslant C(\Delta t)^{1/2}\omega_1\big((\Delta t)^{1/2}\big)(t-\tau)^{-1/2}\tau^{-1/2},$$

$0 < \tau < t < t + \Delta t \leqslant T,\ \Delta t \leqslant t - \tau,$

$$|R_2(t,\Delta t)| \leqslant C_\varphi(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2}), \qquad t, t+\Delta t \in [0,T],\ 0 < \Delta t < \frac{t}{2}.$$

Отсюда и из (33) вытекает оценка (31) для $H_{02}\varphi$. Лемма 3 доказана.

Лемма 4. *Пусть выполнены условия* a), b), (2), (9) и (10). *Тогда* H_1 *является ограниченным оператором из* $C[0,T]$ *в* $\overset{\omega_5}{H_0}[0,T]$.

Доказательство. Достаточно доказать справедливость оценок

$$|(H_1\varphi)(t)| \leqslant C_\varphi\omega_5(t^{1/2}), \tag{35}$$

$$|\Delta_t(H_1\varphi)(t)| \leqslant C_\varphi\omega_5((\Delta t)^{1/2}), \qquad t, t+\Delta t \in [0,T],\ \Delta t > 0. \tag{36}$$

Докажем оценку (35). Используя неравенство (см. [22])

$$|\partial_x\Gamma(g(t),t;g(\tau),\tau)| \leqslant C(\widetilde{\omega}_0((t-\tau)^{1/2}) + \omega_1((t-\tau)^{1/2}))(t-\tau)^{-1}, \qquad 0 \leqslant \tau < t \leqslant T,$$

оценку (4) и условия (9), (10), последовательно получаем

$$|N_1(t,\tau)| \leqslant C\{\omega_4((t-\tau)^{1/2})(t-\tau)^{-1} + \omega_2(t^{1/2})t^{-1/2}(t-\tau)^{-1/2}\}, \qquad 0 \leqslant \tau < t \leqslant T, \tag{37}$$

$$|(H_1\varphi)(t)| \leqslant C_\varphi(\widetilde{\omega}_4(t^{1/2}) + \omega_2(t^{1/2})), \qquad t \in [0,T].$$

Справедливость оценки (36) докажем с помощью представления

$$\Delta_t(H_1\varphi)(t) = \sum_{j=0}^{1}(-1)^{j+1}\int\limits_{t-\Delta t}^{t+j\Delta t} N_1(t+j\Delta t,\tau)\varphi(\tau)d\tau +$$

$$+\int\limits_0^{t-\Delta t}\Delta_t N_1(t,\tau)\varphi(\tau)d\tau \equiv \sum_{j=0}^{2}P_j(t,\Delta t),\ t, t+\Delta t \in [0,T],\ 0 < \Delta t < t.$$

При этом, в силу (35), можно считать, что $0 < \Delta t < t$. Используя неравенство (37), получаем оценки для интегралов $P_j,\ j = 0, 1$:

$$|P_j(t,\Delta t)| \leqslant C_\varphi\int\limits_{t-\Delta t}^{t+j\Delta t}\omega_4((t+j\Delta t-\tau)^{1/2})(t+j\Delta t-\tau)^{-1} + \omega_2((t+j\Delta t)^{1/2})(t+j\Delta t)^{-1/2}(t+j\Delta t-\tau)^{-1/2}d\tau \leqslant$$

$$\leqslant C_\varphi(\widetilde{\omega}_4((\Delta t)^{1/2}) + \omega_2((\Delta t)^{1/2})), \qquad t, t+\Delta t \in [0,T],\ 0 < \Delta t < t. \tag{38}$$

Рассмотрим интеграл P_2. Имеем (см. (5))

$$N_1(t,\tau) = B_1(t)[\partial_x Z(g(t)-g(\tau),t-\tau;A(\tau)) + \partial_x W(g(t),t;g(\tau),\tau)]+$$
$$+\hat{B}_1(t)\Gamma(g(t),t;g(\tau),\tau), \qquad 0 \leqslant \tau < t \leqslant T.$$

Из условия (9) и неравенств (см. [22])

$$|\partial_x Z(g(t)-g(\tau),t-\tau;A(\tau))| \leqslant C\omega_1((t-\tau)^{1/2})(t-\tau)^{-1},$$
$$|\Delta_t\partial_x Z(g(t)-g(\tau),t-\tau;A(\tau))| \leqslant C\{(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})(t-\tau)^{-3/2} + (\Delta t)\omega_1((t-\tau)^{1/2})(t-\tau)^{-2}\} \leqslant$$
$$\leqslant C(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})(t-\tau)^{-3/2}, \qquad 0 \leqslant \tau < t < t+\Delta t \leqslant T,\ \Delta t \leqslant t-\tau,$$

следует, что

$$|\Delta_t\{B_1(t)\partial_x Z(g(t)-g(\tau),t-\tau;A(\tau))\}| \leqslant C\{\omega_3(\Delta t)\omega_1((t-\tau)^{1/2})(t-\tau)^{-1}+$$
$$+(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})(t-\tau)^{-3/2}\}, \qquad 0 \leqslant \tau < t < t+\Delta t \leqslant T,\ \Delta t \leqslant t-\tau. \tag{39}$$

Далее, в силу (2), (6), (7), (9), (10) и (34), имеем

$$|\Delta_t\{B_1(t)\partial_x W(g(t),t;g(\tau),\tau)\}| \leqslant C\{\omega_3(\Delta t)\widetilde{\omega}_0((t-\tau)^{1/2})(t-\tau)^{-1}+$$
$$+(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})\widetilde{\omega}_0((t-\tau)^{1/2})(t-\tau)^{-3/2} + (\Delta t)^{1/2}\widetilde{\omega}_0((t-\tau)^{1/2})(t-\tau)^{-3/2}\} \leqslant$$
$$\leqslant C\{\omega_3(\Delta t)\widetilde{\omega}_0((t-\tau)^{1/2})(t-\tau)^{-1} + (\Delta t)^{1/2}\widetilde{\omega}_0((t-\tau)^{1/2})(t-\tau)^{-3/2}\},$$

$$|\Delta_t\{\hat{B}_1(t)\Gamma(g(t),t;g(\tau),\tau)\}| \leqslant C\{\omega_2((\Delta t)^{1/2})t^{-1/2}(t-\tau)^{-1/2}+$$
$$+\omega_2(t^{1/2})t^{-1/2}[(\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})(t-\tau)^{-1} + \Delta t(t-\tau)^{-3/2}]\} \leqslant$$
$$\leqslant C\{(\omega_1((\Delta t)^{1/2}) + \omega_2((\Delta t)^{1/2}))t^{-1/2}(t-\tau)^{-1/2} + (\Delta t)^{1/2}\omega_2((\Delta t)^{1/2})(t-\tau)^{-3/2}\},$$

$0 \leqslant \tau < t < t+\Delta t \leqslant T,\ \Delta t \leqslant t-\tau$. Отсюда и из (39) получаем оценку

$$|P_2(t,\Delta t)| \leqslant C_\varphi\{\omega_3(\Delta t)\int_0^{t-\Delta t}\omega_4((t-\tau)^{1/2})(t-\tau)^{-1}d\tau + (\Delta t)^{1/2}\omega_1((\Delta t)^{1/2})\int_0^{t-\Delta t}(t-\tau)^{-3/2}d\tau+$$
$$+(\Delta t)^{(1-\varepsilon_0)/2}\widetilde{\omega}_0((\Delta t)^{1/2})\int_0^{t-\Delta t}(t-\tau)^{-(3-\varepsilon_0)/2}d\tau + (\omega_1((\Delta t)^{1/2}) + \omega_2((\Delta t)^{1/2}))t^{-1/2}\int_0^{t-\Delta t}(t-\tau)^{-1/2}d\tau+$$
$$+(\Delta t)^{1/2}\omega_2((\Delta t)^{1/2})\int_0^{t-\Delta t}(t-\tau)^{-3/2}d\tau\} \leqslant C_\varphi(\omega_2((\Delta t)^{1/2}) + \omega_3(\Delta t) + \omega_4((\Delta t)^{1/2})),$$

$t, t+\Delta t \in [0,T],\ 0 < \Delta t < t$, которая, вместе с (38), дает (36). Лемма 4 доказана.

Докажем, что справедлива следующая теорема.

Теорема 2. *Пусть выполнены условия* a), b), (2), (8)–(10) и (19). *Тогда для любых (вектор-)функций* $\psi_0 \in \overset{0}{C}{}^{1/2}[0,T]$ *и* $\psi_1 \in \overset{0}{C}[0,T]$ *система* (22), (23) *имеет единственное в пространстве* $C[0,T]$ *решение* $\varphi \in \overset{0}{C}[0,T]$ *и справедлива оценка*

$$\|\varphi;[0,T]\|^0 \leqslant C\{\|\psi_0;[0,T]\|^{1/2} + \|\psi_1;[0,T]\|^0\}. \tag{40}$$

Доказательство. Как показано выше, система (22), (23) может быть записана в виде (28), (29). Пусть операторы K_0 и K_1 действуют на $\varphi \in C[0,T]$ по формулам

$$(K_0\varphi)(t) = 2\partial^{1/2}(H_0\varphi)(t), \quad (K_1\varphi)(t) = -2(H_1\varphi)(t), \qquad t \in [0,T].$$

В силу лемм 1, 3 и 4, $K_j : C[0,T] \to \overset{0}{H}{}^{\omega_5}[0,T]$, $j = 0,1$, являются линейными ограниченными вольтерровыми операторами. Применяя к обеим частям уравнения (28) оператор дробного дифференцирования $\partial^{1/2}$, в силу равенств $\partial^{1/2} I^{1/2}\chi = \chi$, $I^{1/2}\partial^{1/2}\psi = \psi$, справедливых для $\chi \in C[0,T]$, $\psi \in C^{1/2}[0,T]$, для отыскания $\varphi \in C[0,T]$ получим эквивалентную (28), (29) систему уравнений II рода

$$B_0(t)M(t)\varphi(t) + (K_0\varphi)(t) = 2\partial^{1/2}\psi_0(t), \tag{41}$$

$$B_1(t)(A(t))^{-1}\varphi(t) + (K_1\varphi)(t) = -2\psi_1(t), \qquad t \in [0,T]. \tag{42}$$

Положим

$$K = \begin{pmatrix} K_0 \\ K_1 \end{pmatrix}, \ \psi = \begin{pmatrix} 2\partial^{1/2}\psi_0 \\ -2\psi_1 \end{pmatrix}.$$

В силу (17)–(19), систему (41), (42) можно переписать в виде операторного уравнения

$$\varphi + \hat{K}\varphi = \hat{\psi}, \tag{43}$$

где $\hat{K} = (GM)^{-1}K$, $\hat{\psi} = (GM)^{-1}\psi \in \overset{0}{C}[0,T]$. Из свойств операторов K_0 и K_1 вытекает, что $\hat{K} : C[0,T] \to \overset{0}{H}{}^{\omega_5}[0,T]$ – линейный ограниченный вольтерров оператор, и, следовательно, по лемме 2, уравнение (43) имеет единственное в $C[0,T]$ решение φ, причем выполнено неравенство $\|\varphi;[0,T]\|^0 \leqslant C\|\psi;[0,T]\|^0$. Отсюда получаем оценку (40). Кроме того, из вида уравнения (43) следует, что $\varphi(0) = 0$. Теорема 2 доказана.

3. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1

Сначала докажем *существование* решения задачи (11)–(14). Это решение ищем в виде потенциала простого слоя (21) с плотностью $\varphi \in C[0,T]$, подлежащей определению. Для любой $\varphi \in C[0,T]$ потенциал (21) является решением системы (11) и удовлетворяет начальному условию (12). Подставляя (21) в граничные условия (13) и (14), получаем систему интегральных уравнений Вольтерры I и II рода (22), (23). Из теоремы 2 следует, что система (22), (23) имеет единственное в $C[0,T]$ решение $\varphi \in \overset{0}{C}[0,T]$, подставляя которое в потенциал (21), получаем решение задачи (11)–(14). Из оценки (40) и свойств потенциала простого слоя (см. [22]) делаем вывод, что найденное решение принадлежит пространству $\overset{0}{C}{}^{1,0}(\overline{\Omega})$ и выполнено неравенство (20).

Далее докажем *единственность* решения задачи (11)–(14). Пусть $u \in C^{1,0}(\overline{\Omega})$ – регулярное решение задачи (11)–(14) при $\psi_j(t) = 0$, $t \in [0,T]$, $j = 0,1$. Тогда (вектор-)функция u является единственным (см. [10]) в пространстве $C^{1,0}(\overline{\Omega})$ регулярным решением второй начально-краевой задачи

$$\begin{gathered} Lu(x,t) = 0, \qquad (x,t) \in \Omega, \quad u(x,0) = 0, \quad x \geqslant g(0), \\ \partial_x u(g(t),t) = \psi(t), \qquad t \in [0,T], \end{gathered} \tag{44}$$

где $\psi \in \overset{0}{C}[0,T]$, и для нее справедливо интегральное представление (см. [7])

$$u(x,t) = \int_0^t \Gamma(x,t;g(\tau),\tau)\varphi(\tau)d\tau, \qquad (x,t) \in \overline{\omega}, \tag{45}$$

где вектор-функция $\varphi \in \overset{0}{C}[0,T]$ является решением системы интегральных уравнений Вольтерры II рода, индуцированной граничным условием (44). Подставляя выражение (45) в граничные условия (13) и (14) с нулевыми правыми частями, получим, что $\varphi \in \overset{0}{C}[0,T]$ одновременно является решением системы уравнений (22), (23) с $\psi_j(t) = 0,\ t \in [0,T],\ j = 0,1$, и, следовательно, в силу теоремы 2, $\varphi(t) = 0,\ t \in [0,T]$. Возвращаясь к представлению (45), получаем, что $u(x,t) = 0,\ (x,t) \in \overline{\Omega}$. Теорема 1 доказана.

4. О ТОЧНОСТИ ТРЕБОВАНИЙ НА ПРАВЫЕ ЧАСТИ ГРАНИЧНЫХ УСЛОВИЙ

В лемме 5, следующей ниже, покажем, что условия теоремы 1 на характер непрерывности правых частей в граничных условиях (13), (14) являются точными для разрешимости в пространстве $\overset{0}{C}{}^{1,0}(\overline{\Omega})$ поставленной задачи.

В полосе D рассмотрим параболический оператор

$$\hat{L}u = \partial_t u - A\partial_x^2 u,$$

где $A = \|a_{jk}\|_{j,k=1}^m$ — $m \times m$-матрица, элементами которой являются вещественные числа, и ее собственные числа $\mu_r,\ r = \overline{1,m}$, удовлетворяют условию $\mathrm{Re}\mu_r > 0$. Положим $D^+ = \{(x,t) \in D :\ x > 0\}$. Докажем, что справедлива

Лемма 5. *Пусть (вектор-)функция $u \in \overset{0}{C}{}^{1,0}(\overline{D}^+)$ — регулярное решение системы*

$$\hat{L}u(x,t) = 0, \qquad (x,t) \in D^+. \tag{46}$$

Тогда для (вектор-)функций $\hat{\psi}_0(t) = u(0,t),\ \hat{\psi}_1(t) = \partial_x u(0,t),\ t \in [0,T]$, справедливы включения

$$\hat{\psi}_0 \in \overset{0}{C}{}^{1/2}[0,T], \qquad \hat{\psi}_1 \in \overset{0}{C}[0,T]. \tag{47}$$

Доказательство. Выполнение (47) для $\hat{\psi}_1$ сразу следует из включения $u \in \overset{0}{C}{}^{1,0}(\overline{D}^+)$.

Докажем, что (47) имеет место для $\hat{\psi}_0$. Положим

$$Z(x,t) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} e^{ixy}\exp(-y^2 tA)dy, \quad (x,t) \in \mathbb{R}\times(0,+\infty), \qquad M = \frac{1}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty}\exp(-y^2 A)dy.$$

Из теорем о существовании (см. [7]) и единственности (см. [10]) регулярного решения из пространства $C^{1,0}(\overline{D}^+)$ второй начально-краевой задачи для системы (46) следует, что функция u может быть представлена в виде потенциала простого слоя

$$u(x,t) = -2\int\limits_0^t Z(x,t-\tau)A\hat{\psi}_1(\tau)d\tau, \qquad (x,t) \in \overline{D}^+.$$

Отсюда и из равенства (см. [4])

$$\left(\partial_t^{1/2} Z\right)(x,t) = -\partial_x Z(x,t)M^{-1}, \qquad x > 0,\ t > 0,$$

следует, что

$$\left(\partial_t^{1/2} u\right)(x,t) = -\frac{2}{\sqrt{\pi}}\partial_t \int\limits_0^t (t-\tau)^{-1/2}d\tau \int\limits_0^\tau Z(x,\tau-\eta)A\hat{\psi}_1(\eta)d\eta =$$

$$= -\frac{2}{\sqrt{\pi}}\partial_t \int_0^t \left\{ \int_\tau^t (t-\eta)^{-1/2} Z(x, \eta-\tau) d\eta \right\} A\hat{\psi}_1(\tau) d\tau = -\frac{2}{\sqrt{\pi}} \int_0^t \left\{ \partial_t \int_\tau^t (t-\eta)^{-1/2} Z(x, \eta-\tau) d\eta \right\} A\hat{\psi}_1(\tau) d\tau =$$

$$= -\frac{2}{\sqrt{\pi}} \int_0^t \left\{ \partial_t \int_0^{t-\tau} (t-\tau-\eta)^{-1/2} Z(x, \eta) d\eta \right\} A\hat{\psi}_1(\tau) d\tau = -2 \int_0^t (\partial_t^{1/2} Z)(x, t-\tau) A\hat{\psi}_1(\tau) d\tau =$$

$$= 2 \int_0^t \partial_x Z(x, t-\tau) M^{-1} A\hat{\psi}_1(\tau) d\tau, \qquad (x,t) \in D^+.$$

Поэтому, в силу равенства (17) и соотношения (см. [22])

$$\lim_{x \to +0} 2 \int_0^t \partial_x Z(x, t-\tau) M^{-1} A\hat{\psi}_1(\tau) d\tau = -M^{-1}\hat{\psi}_1(t),$$

в котором стремление к пределу равномерно по $t \in [0, T]$, получаем равенство

$$\lim_{x \to +0} \frac{1}{\sqrt{\pi}} \partial_t \int_0^t (t-\tau)^{-1/2} u(x, \tau) d\tau = -M^{-1}\hat{\psi}_1(t), \tag{48}$$

в котором стремление к пределу равномерно по $t \in [0, T]$. Заметим, кроме того, что

$$\lim_{x \to +0} \int_0^t (t-\tau)^{-1/2} u(x, \tau) d\tau = \int_0^t (t-\tau)^{-1/2} \hat{\psi}_0(\tau) d\tau, \qquad t \in [0, T]. \tag{49}$$

В силу (48) и (49) существует непрерывная дробная производная

$$(\partial^{1/2}\hat{\psi}_0)(t) = \frac{1}{\sqrt{\pi}} \frac{d}{dt} \int_0^t (t-\tau)^{-1/2} \hat{\psi}_0(\tau) d\tau = -M^{-1}\hat{\psi}_1(t), \qquad t \in [0, T],$$

и $(\partial^{1/2}\hat{\psi}_0)(0) = 0$. Следовательно, $\hat{\psi}_0 \in C_0^{1/2}[0, T]$. Лемма 5 доказана.

Автор выражает благодарность профессору Е. А. Бадерко за постановку задачи и постоянное внимание к работе.

СПИСОК ЛИТЕРАТУРЫ

1. *Солонников В. А.* О краевых задачах для линейных параболических систем дифференциальных уравнений общего вида. Тр. Матем. ин-та В. А. Стеклова АН СССР. 1965. Т. 83. С. 3–163.
2. *Ладыженская О. А., Солонников В. А., Уральцева Н. Н.* Линейные и квазилинейные уравнения параболического типа. М.: Наука, 1967. 736 с.
3. *Бадерко Е. А., Черепова М. Ф.* Первая краевая задача для параболических систем в плоских областях с негладкими боковыми границами // Докл. РАН. 2014. Т. 458. № 4. С. 379–381.
4. *Бадерко Е. А., Черепова М. Ф.* Потенциал простого слоя и первая краевая задача для параболической системы на плоскости // Дифференц. уравнения. 2016. Т. 52. № 2. С. 198–208.
5. *Коненков А. Н.* Существование и единственность классического решения первой краевой задачи для параболических систем на плоскости // Дифференц. уравнения. 2023. Т. 59. С. 904–913.
6. *Baderko E. A., Cherepova M. F.* Dirichlet problem for parabolic systems with Dini continuous coefficients Applicable Analysis. 2021. V. 100. N 13. P. 2900–2910.

7. *Зейнеддин М.* О потенциале простого слоя для параболической системы в классах Дини. Дис. … канд. физ.-мат. наук. М.: МГУ им. М. В. Ломоносова, 1992.
8. *Бадерко Е. А., Сахаров С. И.* Единственность решений начально-краевых задач для параболических систем с Дини-непрерывными коэффициентами в плоских областях // Докл. РАН. 2022. Т. 502. № 2. С. 26–29.
9. *Бадерко Е. А., Сахаров С. И.* Потенциал Пуассона в первой начально-краевой задаче для параболической системы в полуограниченной области на плоскости // Дифференц. уравнения. 2022. Т. 58. № 10. С. 1333–1343.
10. *Бадерко Е. А., Сахаров С. И.* О единственности решений начально-краевых задач для параболических систем с Дини-непрерывными коэффициентами в полуограниченной области на плоскости // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 4. С. 584–595.
11. *Сахаров С. И.* Начально-краевые задачи для однородных параболических систем в полуограниченной плоской области и условие дополнительности // Дифференц. уравнения. 2023. Т. 59. № 12. С. 1641–1653.
12. *Ворошнин Л. Г., Хусид Б. М.* Диффузионный массоперенос в многокомпонентных системах. Минск: Наука и техн., 1979. 255 с.
13. *Гуров К. П., Карташкин Б. А., Угасте Ю. Э.* Взаимная диффузия в многофазных металлических системах. М.: Наука, 1981. 350 с.
14. *Криштал М. А.* Многокомпонентная диффузия в металлах. М.: Металлургия, 1985. 177 с.
15. *Самарский А. А., Галактионов В. А., Курдюмов С. П., Михайлов А. П.* Режимы с обострением в задачах для квазилинейных параболических уравнений. М.: Наука, 1987. 480 с.
16. *Князева А. Г.* Перекрестные эффекты в твердых средах с диффузией // Прикл. механ. и техн. физ. 2003. Т. 44. № 3. С. 85–99.
17. *Дышин О. А.* Разрешимость в гёльдеровых функциях задачи нестационарной фильтрации жидкости в трещиновато-пористом кольцевом пласте // Науч. труды НИПИ Нефтегаз ГНКАР. 2012. № 2. С. 74–81.
18. *Семенов М. Ю., Смирнов А. Е., Лашнев М. М., Ступников В. В.* Математическая модель вакуумной нитроцементации теплостойкой стали ВКС-10 // Наука и образование [электронное науч.-техн. издание]. 2013. № 8. http://technomag.bmstu.ru/doc/569132.html
19. *Семенов М. Ю.* Методология разработки технологий химико-термической обработки на основе моделирования диффузионных процессов и анализа эксплуатационных свойств зубчатых передач. Дис. … докт. техн. наук. М.: МГТУ им. Н. Э. Баумана, 2015.
20. *Гуляев А. П.* Металловедение. М.: Металлургия, 1986. 544 с.
21. *Дзядык В. К.* Введение в теорию равномерного приближения функций полиномами. М.: Наука, 1977. 512 с.
22. *Зейнеддин М.* Гладкость потенциала простого слоя для параболической системы второго порядка в классах Дини. 1992. Деп. ВИНИТИ РАН. 16.04.92. № 1294–В92.
23. *Семаан Х. Д.* О решении второй краевой задачи для параболических систем на плоскости. Дис. … канд. физ-матем. наук. М.: МГУ им. М. В. Ломоносова, 1999.
24. *Эйдельман С. Д.* Параболические системы. М.: Наука, 1964. 444 с.
25. *Камынин Л. И.* Гладкость тепловых потенциалов в пространстве Дини–Гёльдера // Сиб. матем. журн. 1970. Т. 11. № 5. С. 1017–1045.
26. *Тихонов А. Н.* О функциональных уравнениях типа Volterra и их применениях к некоторым задачам математической физики // Бюлл. Моск. гос. ун-та. Секц. А. 1938. Т. 1. № 8. С. 1–25.
27. *Baderko E. A., Cherepova M. F.* Bitsadze-Samarskii problem for parabolic systems with Dini continuous coefficients. Complex Variables and Elliptic Equations. 2019. V. 64. N 5. P. 753–765.
28. *Baderko E. A., Cherepova M. F.* Mixed problems for plane parabolic systems and boundary integral equations // J. Math. Sci. 2022. V. 260. N 4. P. 418–433.

ON INITIAL-BOUNDARY VALUE PROBLEMS FOR PARABOLIC SYSTEMS IN A SEMI-BOUNDED PLANE DOMAIN WITH GENERAL BOUNDARY CONDITIONS

S. I. Sakharov*

Lomonosov Moscow State University, Moscow Center for Fundamental and Applied Mathematics, Leninskie Gory, 1, Moscow, 119991 Russia

**e-mail: ser341516@yandex.ru*

Received 14 December, 2023
Revised 02 February, 2024
Accepted 06 March, 2024

Abstract. The paper considers initial boundary value problems for homogeneous parabolic systems with Dini-continuous coefficients under zero initial conditions in a semi-bounded plane domain with a non-smooth lateral boundary that admits the presence of ”beaks”on which boundary conditions of a general type with variable coefficients are specified. Using the method of boundary integral equations, a theorem is proved on the unique classical solvability of such problems in the space of functions that are continuous and bounded together with their first-order spatial derivative in the closure of the domain. A representation of the solutions obtained is given in the form of vector single layer potentials.

Keywords: parabolic systems, initial-boundary value problems, nonsmooth lateral boundary, boundary integral equations, Dini condition.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1042–1054

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.635

ТУРБУЛЕНТНАЯ КИНЕТИЧЕСКАЯ ЭНЕРГИЯ В ПРИБЛИЖЕННОМ РЕШАТЕЛЕ ГАЗОДИНАМИЧЕСКОЙ ЗАДАЧИ РИМАНА

© 2024 г. М. И. Болдырев[1,*]

[1] *456770 Снежинск, Челябинская обл., ул. Васильева, 13, ФГУП "РФЯЦ–ВНИИТФ им. Акад. Е. И. Забабахина", Россия*

[*] *e-mail: boldyrevmi@vniitf.ru*

Поступила в редакцию 03.11.2023 г.
Переработанный вариант 03.11.2023 г.
Принята к публикации 06.03.2024 г.

Описывается учет турбулентной кинетической энергии в решении газодинамической задачи о распаде разрыва (задаче Римана) с помощью приближенного решателя HLLC. Рассматривается система уравнений Эйлера с добавлением гиперболического уравнения турбулентной кинетической энергии и учетом турбулентного давления в уравнениях баланса импульса и энергии. Находится якобиан данной системы уравнений, его собственные числа. На основе этого вносятся изменения в схему вычислений в решателе HLLC. На примере задачи Сода проверяется корректность учета турбулентной кинетический энергии в решении задачи Римана, и показывается неустойчивость схемы при большом турбулентном давлении в случае неучета турбулентности в вычислении характеристических скоростей. Библ. 19. Фиг. 7. Табл. 3.

Ключевые слова: сжимаемая газодинамика, уравнения Эйлера, турбулентная кинетическая энергия, приближенный решатель Римана, HLLC, неустойчивость Рихтмайера–Мешкова.

DOI: 10.31857/S0044466924060126, **EDN:** XYEANH

ВВЕДЕНИЕ

Для задач сжимаемой газовой динамики характерно наличие разрывов решения, что требует при численной аппроксимации специальных подходов к вычислению потоков. Одним из них является подход Годунова: использование решения задачи Римана (см. [1]). В частности, один из его вариантов – применение приближенного решения задачи Римана и вычисление потоков по схеме HLLC (см. [2]). Этот решатель, благодаря точному описанию контактной границы, успешно применяется в моделировании многокомпонентных течений (см. [3], [4]). И в настоящей работе рассматривается его применение в задачах моделирования перемешивания, вызванного действием ударных волн на границу между двумя газами.

Одним из вопросов, рассматриваемых сжимаемой газовой динамикой, является неустойчивость, развивающаяся на границе двух газов с разной молярной массой. Она возникает под действием либо постоянного или переменного во времени ускорения, тогда явление носит название неустойчивости Рэлея–Тэйлора (РТ) (см. [5]), либо мгновенного импульса ускорения, вызываемого ударной волной, тогда она называется неустойчивостью Рихтмайера–Мешкова (РМ) (см. [6]). На поздней стадии развития этих неустойчивостей течение приобретает турбулентный характер. Для его моделирования можно использовать осредненные по Рейнольдсу уравнения Навье–Стокса (RANS), для замыкания которых применяются те или иные модели турбулентности (см. [7]).

При использовании RANS-подхода совместно с моделями семейства $k - \varepsilon$ получается неполная параболическая система уравнений (см. [8]), которую для численного решения удобно расщепить на две подсистемы: гиперболическую и собственно параболическую. Проводя аналогию между молекулярными явлениями и турбулентностью, можно выделить в тензоре Рейнольдса часть, аналогичную

статическому давлению, так называемое турбулентное давление, которую можно включить в систему гиперболических уравнений. В настоящей статье сосредоточимся на том, как это включение повлияет на якобиан системы гиперболических уравнений и покажем, как нужно модифицировать решатель HLLC для учета дополнительного турбулентного давления. Подобная задача ставится в работе [9], но для точного решателя задачи Римана.

Статья разбита на следующие части: сначала кратко описывается решаемая система гиперболических уравнений, осредненных по Рейнольдсу, затем выводится якобиан для данной системы и находятся его собственные числа, после описан модифицированный алгоритм приближенного решателя HLLC с учетом влияния турбулентной кинетической энергии (ТКЭ), наконец, модифицированный решатель тестируется на задаче Сода и эксперименте Феттера и Штуртеванта с неустойчивостью Рихтмайера–Мешкова.

1. УРАВНЕНИЯ ГАЗОВОЙ ДИНАМИКИ

Рассмотрим систему уравнений сжимаемой газовой динамики для однокомпонентной системы (систему уравнений Эйлера) (см. [10]):

$$\begin{gathered}\frac{\partial \varrho}{\partial t} + \nabla \cdot \left(\varrho \vec{u}\right) = 0,\\ \frac{\partial \left(\varrho \vec{u}\right)}{\partial t} + \nabla \cdot \left(\varrho \vec{u} \otimes \vec{u}\right) + \vec{\nabla} p = 0,\\ \frac{\partial \left(\varrho E\right)}{\partial t} + \nabla \cdot \left((\varrho E + p) \cdot \vec{u}\right) = 0,\\ \varrho E = U + \varrho \vec{u} \cdot \vec{u}/2 = \varrho \cdot \left(e + \vec{u}^2/2\right),\end{gathered} \tag{1}$$

$$p = f(\varrho, \varrho\vec{u}, \varrho E).$$

Здесь E — полная энергия единицы массы, U — внутренняя энергия единицы объема, e — внутренняя энергия единицы массы. Осредним систему уравнений (1) по Рейнольдсу. При этом воспользуемся средними по Фавру, чтобы представить результат в более удобной форме. Не вдаваясь в детали (см. [11]), приведем сразу результирующую систему, $[\bar{}]$ означает среднее по Рейнольдсу, $[\tilde{}]$ — среднее по Фавру, $[\,]''$ — флуктуация по Фавру:

$$\frac{\partial \bar{\varrho}}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{\vec{u}}\right) = 0,$$

$$\frac{\partial \left(\bar{\varrho}\tilde{\vec{u}}\right)}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{\vec{u}} \otimes \tilde{\vec{u}}\right) + \vec{\nabla}\bar{p} = \nabla \cdot \mathbf{R}, \tag{2}$$

$$\frac{\partial \left(\bar{\varrho}\tilde{E}\right)}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{E} \cdot \tilde{\vec{u}} + \bar{p} \cdot \tilde{\vec{u}} - \mathbf{R} \cdot \tilde{\vec{u}}\right) = -\overline{\rho e'' \cdot \vec{u}''}, \tag{3}$$

$$\bar{\varrho}\tilde{E} = \bar{U} + \bar{\varrho}\tilde{\vec{u}} \cdot \tilde{\vec{u}}/2 + \bar{\varrho}k = \bar{\varrho} \cdot \left(\tilde{e} + \tilde{\vec{u}}^2/2 + k\right), \tag{4}$$

$$\bar{p} = f(\bar{\varrho}, \bar{\varrho}\tilde{\vec{u}}, \bar{\varrho}\tilde{E}, \bar{\varrho}k).$$

При выводе использовались, среди прочих, следующие соотношения:

$$\overline{\varrho\vec{u} \otimes \vec{u}} = \overline{\varrho \cdot \left(\tilde{\vec{u}} + \vec{u}''\right) \otimes \left(\tilde{\vec{u}} + \vec{u}''\right)},$$

$$\overline{\varrho \cdot \tilde{\vec{u}} \cdot \vec{u}''} = 0,$$

$$\mathbf{R} = -\overline{\varrho \cdot \vec{u}'' \otimes \vec{u}''},$$

$$k = \frac{1}{2}\frac{\overline{\varrho \cdot \vec{u}'' \cdot \vec{u}''}}{\bar{\varrho}},$$

$$\overline{\varrho \cdot \vec{u}'' \otimes \vec{u}'' \cdot \vec{u}''} \approx 0,$$

$$\overline{\varrho \cdot e \cdot \vec{u}''} = \overline{\rho e'' \cdot \vec{u}''}.$$

Отсюда видно, что при осреднении системы уравнений Эйлера возникают дополнительные сущности: турбулентная кинетическая энергия единицы массы k, которая является частью полной энергии системы (см. уравнение (4)), и тензор Рейнольдса $\mathbf{R}$, который служит источником импульса в уравнении (2) и создает поток полной энергии в уравнении (3).

Согласно гипотезе Буссинеска (см. [11]), тензор Рейнольдса можно представить в виде, аналогичном обобщенному тензору давления:

$$\mathbf{R} = -\frac{2}{3}\bar{\varrho}k \cdot \mathbf{I} + \mu_t \cdot \left(\vec{\nabla}\tilde{\vec{u}} + \vec{\nabla}\tilde{\vec{u}}^T - \left(\frac{2}{3}\nabla \cdot \tilde{\vec{u}}\right) \cdot \mathbf{I}\right),$$

где слагаемое $\frac{2}{3}\bar{\varrho}k$ можно интерпретировать, как “турбулентное давление” p_t.

В семействе $k - \varepsilon$ моделей сжимаемой турбулентности уравнение баланса k имеет, как правило, следующий вид:

$$\frac{\partial(\bar{\varrho}k)}{\partial t} + \nabla \cdot \left(\bar{\varrho}k \cdot \tilde{\vec{u}}\right) = -\nabla \cdot \vec{F}_D + P - \bar{\varrho}\varepsilon,$$

но в настоящей работе нас интересует только гиперболическая подсистема представленных выше уравнений, т.е.

$$\begin{gathered}
\frac{\partial\bar{\varrho}}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{\vec{u}}\right) = 0,\\
\frac{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{\vec{u}} \otimes \tilde{\vec{u}} + \left(\bar{p} + p_t\right) \cdot \mathbf{I}\right) = 0,\\
\frac{\partial\left(\bar{\varrho}\tilde{E}\right)}{\partial t} + \nabla \cdot \left(\bar{\varrho}\tilde{E} \cdot \tilde{\vec{u}} + \left(\bar{p} + p_t\right) \cdot \tilde{\vec{u}}\right) = 0,\\
\frac{\partial(\bar{\varrho}k)}{\partial t} + \nabla \cdot \left(\bar{\varrho}k \cdot \tilde{\vec{u}}\right) = 0.
\end{gathered} \tag{5}$$

Обсудим возможные пути численного решения системы (5). Практически она отличается от обыкновенной системы уравнений Эйлера присутствием турбулентного давления p_t, и если допустить, что оно много меньше давления статического, $p_t \ll p$, то одним из двух принципиальных подходов к численному решению системы (5) будет пренебрежение турбулентным давлением и использование стандартной схемы для нахождения потоков, например, HLLC. Второй принципиальный подход, соответственно, состоит в учете p_t при вычислении потоков, и в настоящей работе для этого рассматриваются три варианта.

Первый вариант — двухэтапное интегрирование по времени потоков импульса и полной энергии. На первом этапе интегрируются потоки, полученные применением стандартного решателя задачи Римана, после чего проводится дополнительный шаг интегрирования потоков, вызванных турбулентным давлением, на примере импульса:

$$\begin{gathered}
\left(\bar{\varrho}\tilde{\vec{u}}\right)^* = \left(\bar{\varrho}\tilde{\vec{u}}\right)^n - \Delta t \cdot \nabla \cdot \left(\mathbf{F}^{HLLC}_{\bar{\varrho}\tilde{\vec{u}}}\right),\\
\left(\bar{\varrho}\tilde{\vec{u}}\right)^{n+1} = \left(\bar{\varrho}\tilde{\vec{u}}\right)^* - \Delta t \cdot \nabla \cdot \left(p_t \cdot \mathbf{I}\right).
\end{gathered}$$

Второй вариант — замена статического давления при вычислении потоков в решателе HLLC на его сумму с турбулентным. Наконец, третий вариант — в дополнение к замене статического давления на

суммарное анализ влияния турбулентности на матрицу Якоби и собственные числа системы (5) с соответствующим изменением способа вычисления характеристических скоростей в решателе HLLC. Далее в работе подробно описывается последний вариант, и результаты его тестирования сравниваются с прочими упомянутыми решениями.

2. ЯКОБИАН И СОБСТВЕННЫЕ ЧИСЛА

Для применения приближенного решателя Римана HLLC к системе уравнений (5), необходимо для начала рассмотреть ее матрицу Якоби и найти собственные числа (см. [10]). Представим систему в консервативной форме

$$\frac{\partial W}{\partial t} + \nabla \cdot (F) = 0,$$

где W — это вектор консервативных переменных:

$$W = \left[\bar{\varrho}; \left(\bar{\varrho}\tilde{\vec{u}}\right); \left(\bar{\varrho}\tilde{E}\right); \left(\bar{\varrho}k\right)\right],$$

а F — вектор потоков:

$$F = \begin{bmatrix} \left(\bar{\varrho}\tilde{\vec{u}}\right); \\ \left(\bar{\varrho}\tilde{\vec{u}}\right) \otimes \left(\bar{\varrho}\tilde{\vec{u}}\right)/\bar{\varrho} + \left(\bar{p} + p_t\right) \cdot \mathbf{I}; \\ \left(\bar{\varrho}\tilde{E}\right) \cdot \left(\bar{\varrho}\tilde{\vec{u}}\right)/\bar{\varrho} + \left(\bar{p} + p_t\right) \cdot \left(\bar{\varrho}\tilde{\vec{u}}\right)/\bar{\varrho}; \\ \left(\bar{\varrho}k\right) \cdot \left(\bar{\varrho}\tilde{\vec{u}}\right)/\bar{\varrho} \end{bmatrix}.$$

Матрица Якоби определяется как производная от вектора потоков по вектору консервативных переменных ($\partial F/\partial W$). Она представлена в табл. 1.

Таблица 1. Матрица Якоби, $\partial F/\partial W$

Элемент	$/\partial\bar{\varrho}$	$/\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)$	$/\partial\left(\bar{\varrho}\tilde{E}\right)$	$/\partial\left(\bar{\varrho}k\right)$
$\partial F_1/$	0	1	0	0
$\partial F_2/$	$-\tilde{\vec{u}}^2 + \dfrac{\partial\bar{p}}{\partial\bar{\varrho}}$	$2\tilde{\vec{u}} + \dfrac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)}$	$\dfrac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)}$	$\dfrac{\partial\bar{p}}{\partial\left(\bar{\varrho}k\right)} + \dfrac{2}{3}$
$\partial F_3/$	$\tilde{\vec{u}}\left(\dfrac{\partial\bar{p}}{\partial\bar{\varrho}} - \bar{h}\right)$	$\bar{h} + \tilde{\vec{u}}\dfrac{\partial\bar{p}}{\left(\partial\bar{\varrho}\tilde{\vec{u}}\right)}$	$\tilde{\vec{u}}\left(1 + \dfrac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)}\right)$	$\tilde{\vec{u}}\left(\dfrac{\partial\bar{p}}{\partial\left(\bar{\varrho}k\right)} + \dfrac{2}{3}\right)$
$\partial F_4/$	$-k \cdot \tilde{\vec{u}}$	k	0	$\tilde{\vec{u}}$

Для краткости были введены следующие обозначения:

$$\tilde{\vec{u}} = \frac{\bar{\varrho}\tilde{\vec{u}}}{\bar{\varrho}}, \qquad k = \frac{\bar{\varrho}k}{\bar{\varrho}}, \qquad \tilde{E} = \frac{\bar{\varrho}\tilde{E}}{\bar{\varrho}}, \quad \bar{h} = \tilde{E} + \frac{\bar{p}}{\bar{\varrho}} + \frac{2}{3}k = \tilde{e} + \frac{\tilde{\vec{u}}^2}{2} + \frac{\bar{p}}{\bar{\varrho}} + \frac{5}{3}k.$$

Якобиан матрицы из табл. 1 имеет четыре собственных числа, из которых два — скорость $\tilde{\vec{u}}$, а два других имеют следующий вид:

$$\lambda_{1,4} = \tilde{\vec{u}} + \frac{1}{2}\left(\tilde{\vec{u}}\frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)} + \frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)}\right) \pm \left(\frac{1}{4}\left(\tilde{\vec{u}}\frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)} + \frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)}\right)^2 + \tilde{\vec{u}}\frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)} + h\frac{\partial\bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)} + k\frac{\partial\bar{p}}{\partial\left(\bar{\varrho}k\right)} + \frac{2}{3}k\right)^{1/2}. \quad (6)$$

Давление здесь рассматривается как функция от консервативных переменных, но из термодинамики известно, что термическое уравнение состояния — это функция от плотности и температуры, $p = f(\varrho, T)$, что вкупе с калорическим уравнением состояния, $U = f(\varrho, T)$, позволяет представить

давление как функцию плотности и внутренней энергии. Для упрощения уравнения (**??**) нам необходимо выразить производные от давления по консервативным переменным через производные по плотности и внутренней энергии объема или массы. Так как внутренняя энергия является функцией от консервативных переменных, то воспользуемся правилами дифференцирования сложной функции:

$$\frac{\partial z}{\partial x_j} = \sum_{s=1}^{m} \frac{\partial \delta}{\partial \beta_s} \cdot \frac{\partial \alpha_s}{\partial x_j},$$

где

$$z = \delta\left(\{\beta_i\}_{i=1}^{m}\right), \qquad \beta_i = \alpha\left(\{x_j\}\right).$$

Рассмотрим выражение давления через внутреннею энергию единицы объема $\bar{p} = f(\bar{\varrho}, \bar{U})$ и массы $\bar{p} = f(\bar{\varrho}, \tilde{e})$. Тогда производную давления по полной энергии можно выразить следующим образом:

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)} = \frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}} \cdot \frac{\partial\left(\left(\bar{\varrho}\tilde{E}\right) - \left(\bar{\varrho}\tilde{\vec{u}}\right)^2/(2\bar{\varrho}) - (\bar{\varrho}k)\right)}{\partial\left(\bar{\varrho}\tilde{E}\right)} = \frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}}$$

или

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}\tilde{E}\right)} = \frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \tilde{e}} \cdot \frac{\partial\left(\left(\bar{\varrho}\tilde{E}\right)/\bar{\varrho} - \left(\bar{\varrho}\tilde{\vec{u}}\right)^2/(2\bar{\varrho}^2) - (\bar{\varrho}k)/\bar{\varrho}\right)}{\partial\left(\bar{\varrho}\tilde{E}\right)} = \frac{1}{\bar{\varrho}}\frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \tilde{e}}.$$

Аналогичным образом можно вывести и остальные производные от давления в переменных $(\bar{\varrho}, \bar{U})$:

$$\frac{\partial \bar{p}}{\partial \bar{\varrho}} = \frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{\varrho}} + \frac{\tilde{\vec{u}}^2}{2} \cdot \frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}},$$

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)} = -\tilde{\vec{u}} \cdot \frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}},$$

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}k\right)} = -\frac{\partial f\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}},$$

или в переменных $(\bar{\varrho}, \tilde{e})$:

$$\frac{\partial \bar{p}}{\partial \bar{\varrho}} = \frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \bar{\varrho}} + \frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \tilde{e}} \cdot \left(-\frac{\left(\bar{\varrho}\tilde{E}\right)}{\bar{\varrho}^2} + \frac{\tilde{\vec{u}}^2}{\bar{\varrho}} + \frac{k}{\bar{\varrho}}\right),$$

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}\tilde{\vec{u}}\right)} = -\frac{\tilde{\vec{u}}}{\bar{\varrho}} \cdot \frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \tilde{e}},$$

$$\frac{\partial \bar{p}}{\partial\left(\bar{\varrho}k\right)} = -\frac{1}{\bar{\varrho}} \cdot \frac{\partial f\left(\bar{\varrho}, \tilde{e}\right)}{\partial \tilde{e}}.$$

После этих преобразований собственные числа (**??**) можно представить в существенно сокращенном виде:

$$\lambda_{1,4} = \tilde{\vec{u}} \pm \left(\frac{\partial \bar{p}\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{\varrho}} + \frac{\partial \bar{p}\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}} \cdot \frac{\bar{U} + \bar{p}}{\bar{\varrho}} + \frac{2}{3}k \cdot \left(1 + \frac{\partial \bar{p}\left(\bar{\varrho}, \bar{U}\right)}{\partial \bar{U}}\right)\right)^{1/2}$$

или

$$\lambda_{1,4} = \tilde{\bar{u}} \pm \left(\frac{\partial \bar{p}\,(\bar{\varrho}, \tilde{e})}{\partial \bar{\varrho}} + \frac{\partial \bar{p}\,(\bar{\varrho}, \tilde{e})}{\partial \tilde{e}} \cdot \frac{\bar{p}}{\bar{\varrho}^2} + \frac{2}{3} k \cdot \left(1 + \frac{1}{\bar{\varrho}} \frac{\partial \bar{p}\,(\bar{\varrho}, \tilde{e})}{\partial \tilde{e}} \right) \right)^{1/2}.$$

Первые два подкоренных слагаемых представляют собой выражение для квадрата скорости звука a (см. [13]). Тогда для случая турбулентных течений можно ввести эффективную скорость звука a^*, равную

$$a^* = \sqrt{\bar{a}^2 + \frac{2}{3} k \cdot \left(1 + \frac{1}{\bar{\varrho}} \frac{\partial \bar{p}\,(\bar{\varrho}, \tilde{e})}{\partial \tilde{e}} \right)}. \tag{7}$$

Собственные числа якобиана тогда будут

$$\lambda_{2,3} = \tilde{\bar{u}}, \qquad \lambda_{1,4} = \tilde{\bar{u}} \pm a^*,$$

что аналогично набору собственных чисел обычной системы уравнений Эйлера.

3. МОДИФИКАЦИЯ РЕШАТЕЛЯ HLLC

Приближенный решатель Римана HLLC, предложенный Торо (см. [2], [10]), является развитием решателя HLL и более подходит для моделирования турбулентных течений, развивающихся вследствие неустойчивостей РТ и РМ. По сравнению с HLL в нем введена дополнительная характеристическая скорость, соответствующая распространению контактного разрыва. Это обеспечивает меньшую численную диффузию и более точное описание контактного разрыва, что для неустойчивостей РТ и РМ особенно важно, так как в их случае контактная граница пролегает по линии раздела системы на разные вещества, и от точности ее описания зависит как минимизация ошибки в определении зоны перемешивания, так и вычисление генерации турбулентной кинетической энергии в RANS-моделях.

Исходя из вышеописанного, для применения решателя HLLC к системе уравнений (5) в него предлагается внести следующие изменения.

Во-первых, необходимо использовать формулу (7) для определения характеристических скоростей S_L, S_R одним из предложенных в литературе методов, например, по Дэвису (см. [12]):

$$\bar{S}_L = \min\{\lambda_{1,L}; \lambda_{1,R}\} = \min\{\tilde{u}_L - a^*_L; \tilde{u}_R - a^*_R\},$$

$$\bar{S}_R = \max\{\lambda_{4,L}; \lambda_{4,R}\} = \max\{\tilde{u}_L + a^*_L; \tilde{u}_R + a^*_R\}.$$

Во-вторых, включить турбулентное давление как часть давления, используемого для расчета характеристической скорости контактного разрыва и среднего давления в распадной области:

$$\bar{S}_C = \frac{\bar{p}^{\Sigma}_R - \bar{p}^{\Sigma}_L}{\bar{\rho}_L \left(\bar{S}_L - \tilde{u}_L \right) - \bar{\rho}_R \left(\bar{S}_R - \tilde{u}_R \right)} + \frac{\bar{\rho}_L \tilde{u}_L \left(\bar{S}_L - \tilde{u}_L \right) - \bar{\rho}_R \tilde{u}_R \left(\bar{S}_R - \tilde{u}_R \right)}{\bar{\rho}_L \left(\bar{S}_L - \tilde{u}_L \right) - \bar{\rho}_R \left(\bar{S}_R - \tilde{u}_R \right)},$$

$$\bar{p}_C = 0.5 \cdot \left(\bar{p}^{\Sigma}_L + \bar{p}^{\Sigma}_R + \bar{\rho}_L \cdot \left(\bar{S}_L - \tilde{u}_L \right) \times \left(\bar{S}_C - \tilde{u}_L \right) + \bar{\rho}_R \cdot \left(\bar{S}_R - \tilde{u}_R \right) \cdot \left(\bar{S}_C - \tilde{u}_R \right) \right),$$

где

$$\bar{p}^{\Sigma}_{L/R} = \bar{p}_{L/R} + \left(p_t \right)_{L/R},$$

$$\left(p_t \right)_{L/R} = \frac{2}{3} \cdot \bar{\varrho}_{L/R} \cdot k_{L/R}.$$

В-третьих, использовать суммарное давление для вычисления значений импульса и полной энергии в области распада:

$$(\bar{\varrho}\tilde{\vec{u}})_{*L/R} = \frac{(\bar{\varrho}\tilde{\vec{u}})_{L/R}\cdot(\bar{S}_{L/R} - \tilde{\vec{u}}_{L/R}\cdot\vec{n})}{(\bar{S}_{L/R} - \bar{S}_C)} - \frac{(\bar{p}^{\Sigma}_{L/R} - \bar{p}_C)\cdot\vec{n}}{(\bar{S}_{L/R} - \bar{S}_C)},$$

$$(\bar{\varrho}\tilde{E})_{*L/R} = \frac{(\bar{\varrho}\tilde{E})_{L/R}\cdot(\bar{S}_{L/R} - \tilde{\vec{u}}_{L/R}\cdot\vec{n})}{(\bar{S}_{L/R} - \bar{S}_C)} - \frac{\bar{p}^{\Sigma}_{L/R}\cdot\tilde{\vec{u}}_{L/R}\cdot\vec{n} - \bar{p}_C\cdot\bar{S}_C}{(\bar{S}_{L/R} - \bar{S}_C)},$$

где $\vec{n}$ — нормаль к грани.

В-четвертых, аналогично поступать и при вычислении потоков:

$$\mathbf{F}(\bar{\varrho}\tilde{\vec{u}})_{*L/R} = \left(\left((\bar{\varrho}\tilde{\vec{u}})_{L/R}\otimes\tilde{\vec{u}}_{L/R} + \bar{p}^{\Sigma}_{L/R}\cdot\mathbf{I}\right)\cdot\vec{n} + S_{L/R}\cdot\left((\bar{\varrho}\tilde{\vec{u}})_{*L/R} - (\bar{\varrho}\tilde{\vec{u}})_{L/R}\right)\right)\otimes\vec{n},$$

$$\vec{F}(\bar{\varrho}\tilde{E})_{*L/R} = \left(\left((\bar{\varrho}\tilde{E})_{L/R} + \bar{p}^{\Sigma}_{L/R}\right)\cdot\tilde{\vec{u}}_{L/R}\cdot\vec{n} + \bar{S}_{L/R}\cdot\left((\bar{\varrho}\tilde{E})_{*L/R} - (\bar{\varrho}\tilde{E})_{L/R}\right)\right)\cdot\vec{n}.$$

В остальном решатель HLLC остается стандартным, а для турбулентной кинетической энергии единицы объема схема нахождения распадных значений и потоков аналогична таковой для плотности.

4. ТЕСТИРОВАНИЕ

Для проверки работоспособности полученного HLLC-решателя используем обычный тест Сода (см. [14], [15]). Идея состоит в том, что турбулентное давление должно влиять на решение образом аналогичным статическому, так что если мы имеем такое начальное распределение статического и турбулентного давления в системе, что их сумма равна статическому давлению в задаче Сода без турбулентности, то решение для плотности или скорости между задачами должны совпадать.

Рассмотрим одномерную систему с положением границ $x_l = 0$ и $x_r = 1$ см. На границах задается граничное условие свободного выхода. Внутри системы задаются разрывные начальные условия плотности ($\bar{\varrho}$) и давления ($\bar{p}$) с положением границы x_I между состояниями 1 и 2, значения приведены в табл. 2, скорость $\tilde{\vec{u}}$ задана равной 0 по всей системе. Расчет ведется до момента времени 0.25 с.

Таблица 2. Начальные условия задачи Сода

x_I, см	$\bar{\varrho}_1$, г/см^3	$\bar{p}_1$, дин/см^2	$\bar{\varrho}_2$, г/см^3	$\bar{p}_2$, дин/см^2
0.5	1	1	0.125	0.1

Используется уравнение состояния идеального газа, универсальная газовая постоянная $R = 8.31446\times10^7$ эрг/(моль·K), изохорная теплоемкость $C_V = 20.78616\times10^7$ эрг/моль, молярная масса $M = 1$ г/моль. Будем рассматривать данную задачу как эталонную и сравнивать с ней расчеты следующей задачи, включающей турбулентную кинетическую энергию k, заданную так, чтобы турбулентное давление в зоне состояния 1 составляло 10% от статического давления этой области в вышеописанной задаче, само статическое давление соответственно уменьшим на ту же величину. Таким образом, начальные данные модифицированной задачи Сода будут иметь вид, приведенный в табл. 3. В остальном начальные и граничные условия совпадают.

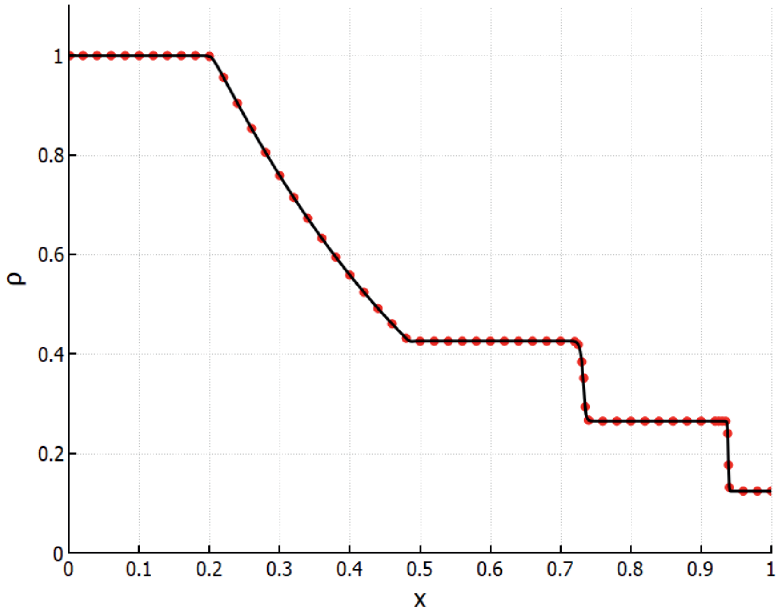
Таблица 3. Начальные условия задачи Сода с турбулентным давлением

x_I, см	$\bar{\varrho}_1$, г/см^3	$\bar{p}_1$, дин/см^2	k_1, (см/с)^2	$\bar{\varrho}_2$, г/см^3	$\bar{p}_2$, дин/см^2	k_2, (см/с)^2
0.5	1	0.9	0.15	0.125	0.1	0

Расчеты проводятся по схеме MUSCL-Hancock (см. [16]) второго порядка аппроксимации по пространству и времени с реконструкцией примитивных переменных $\bar{\varrho}$, $\tilde{\vec{u}}$, $\bar{p}$, k. Так как при стремлении ТКЭ к нулю модифицированная HLLC схема сводится к стандартной, то расчет обеих задач можно

проводить, используя один и тот же решатель, положив для первой задачи ТКЭ равную 0 во всей области расчета. В расчетах использовалась сетка на 1000 ячеек.

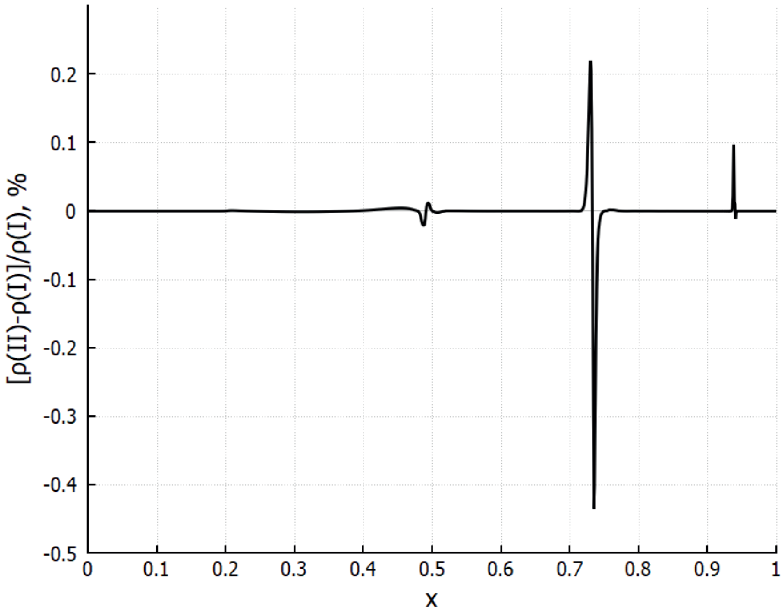
На фиг. 1 представлены графики плотности для обеих задач на конечный момент времени. Они практически совпадают, а значит, можно сделать вывод, что схема с учетом турбулентного давления корректно воспроизводит поток массы стандартной схемы. Тем не менее, детальное сравнение двух расчетов, а именно, рассмотрение относительной разности между решениями, $(\varrho(II) - \varrho(I))/\varrho(I)$ (фиг. 2), показывает небольшое, не более 0.5%, различие в расчетах, что может быть объяснено тем, что турбулентное давление, в отличие от статического, реконструируется не как единое целое, а "покомпонентно", т.е. на грани независимо интерполируются плотность и ТКЭ, и только после этого они преобразуются в турбулентное давление.



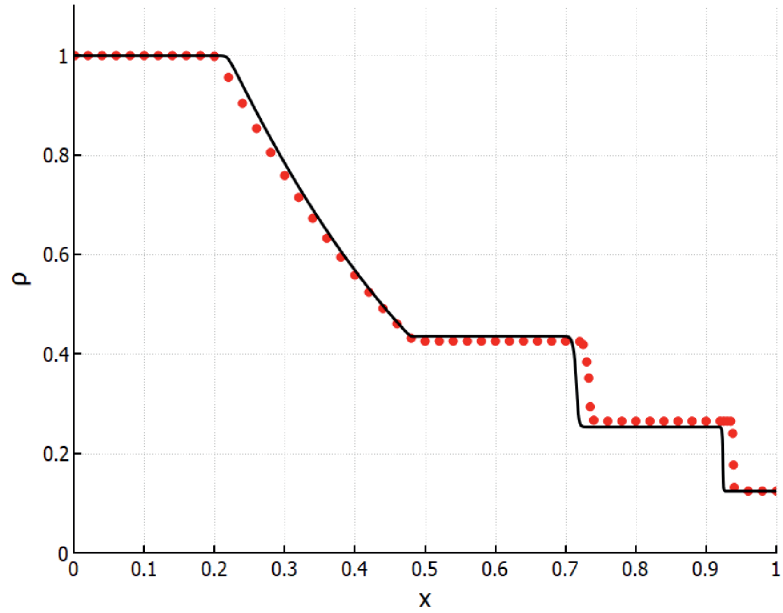
Фиг. 1. Распределение плотности на момент $t = 0.25$ с: стандартная задача Сода (•) и турбулентная задача Сода с модифицированным решателем HLLC (——).

Корректность вышеописанной схемы была показана, но возникает вопрос: насколько оправдан именно такой подход? Можно встретить утверждение (см. [11]), что турбулентная кинетическая энергия, как правило, пренебрежимо мала, и ее вклад в полную энергию и тензор Рейнольдса можно не учитывать, тем самым рассматривая ее как пассивный скаляр. Применительно к данной задаче это, очевидно, не так, что следует из фиг. 3, где сравниваются результаты стандартной задачи со схемой, где турбулентное давление игнорируется, а k рассматривается как пассивный скаляр. Видно, что решение с пассивным скаляром k отстает от эталонного. Конечно, в связи с этим возникает вопрос: в каких задачах действительно достигается такая величина k, что ею нельзя пренебрегать?

Прямолинейное введение турбулентного давления и ТКЭ в схему HLLC без анализа собственных чисел матрицы Якоби может привести к схеме, когда ТКЭ учтена в полной энергии, турбулентное давление суммируется со статическим при вычислении распадных значения и потоков, но не учитывается в вычислении волновых скоростей $\bar{S}_L, \bar{S}_R, \bar{S}_C$. Такой вариант схемы приводит к серьезным осцилляциям, возникающим за волной разряжения, как показано на фиг. 4. Это делает такую схему непригодной для расчетов. Наконец, можно учесть влияние турбулентного давления вне решателя HLLC, аппроксимировав соответствующие члены уравнений импульса, $\vec{\nabla} p_t$, и полной энергии, $\nabla \cdot (p_t \cdot \tilde{\vec{u}})$, например, линейной интерполяцией с последующим вычислением градиента и дивергенции по теореме Гаусса–Остроградского. Такой вариант схемы более устойчив, чем предыдущий, что следует из фиг. 5, наблюдаемые осцилляции весьма малы, но все же данный вариант уступает модифицированной схеме HLLC, которая сохраняет на этом участке гладкость решения.

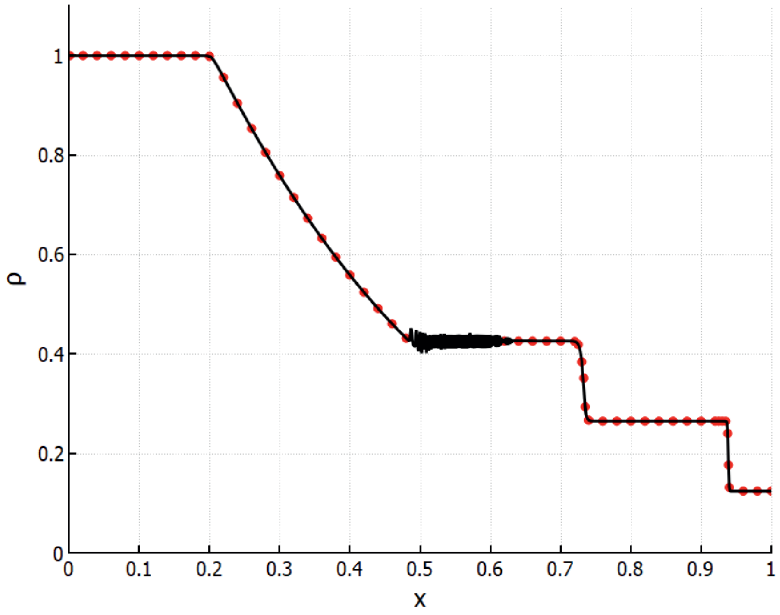


Фиг. 2. Относительная разница плотности между стандартной задачей Сода (I) и турбулентной задачей Сода с модифицированным решателем HLLC (II); $t = 0.25$ с.

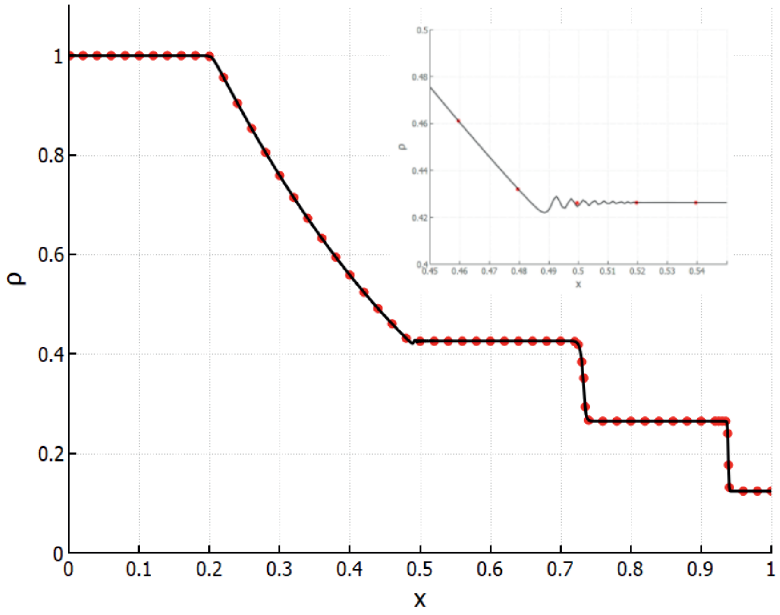


Фиг. 3. Распределение плотности на момент $t = 0.25$ с: стандартная задача Сода (●) и турбулентная задача Сода с турбулентной энергией в качестве пассивного скаляра (——).

Можно заключить, что сравнение модифицированной схемы HLLC с более простыми вариантами расчета потоков сжимаемого турбулентного течения показывает, что при значительной величине турбулентной кинетической энергии ее влияние необходимо строго учитывать в решении задачи Римана о распаде произвольного разрыва, чтобы получать точное и устойчивое решение. “Значительность” ТКЭ можно охарактеризовать соотношением турбулентного и статического давлений. Полагаем, что в случаях, когда турбулентное давление составляет порядка 1% от статического и выше, следует применять описанную в настоящей статье схему.



Фиг. 4. Распределение плотности на момент $t = 0.25$ с: стандартная задача Сода (●) и турбулентная задача Сода без учета турбулентной кинетической энергии в волновых скоростях (——).



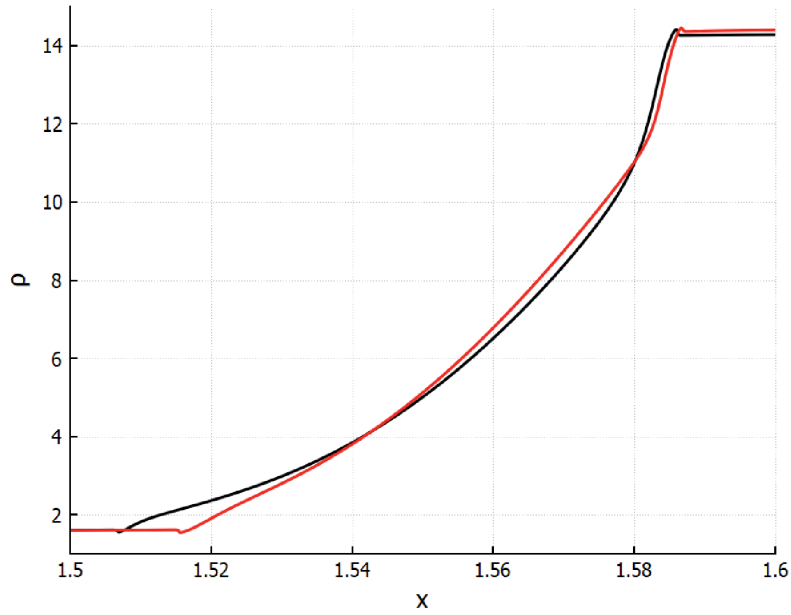
Фиг. 5. Распределение плотности на момент $t = 0.25$ с: стандартная задача Сода (●) и турбулентная задача Сода с линейной аппроксимацией p_t (——).

5. НЕУСТОЙЧИВОСТЬ РИХТМАЙЕРА–МЕШКОВА

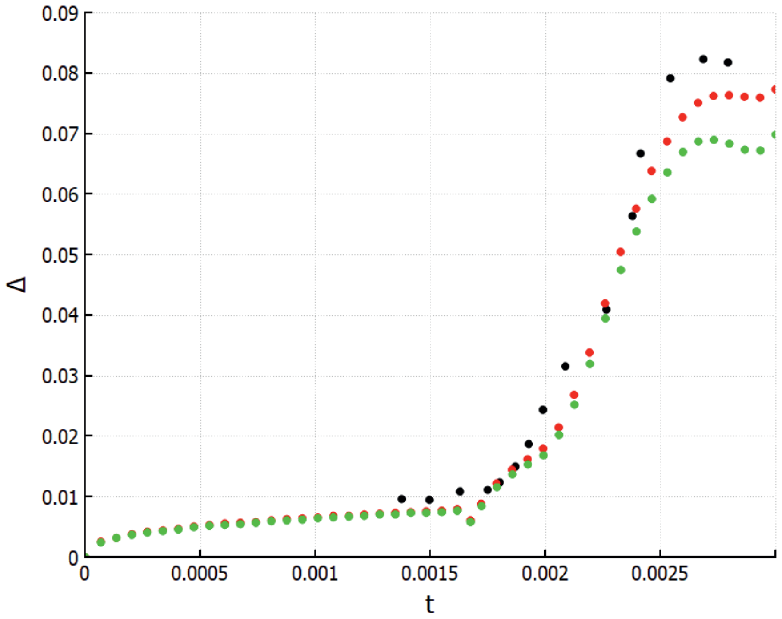
Рассмотрим эффект от учета турбулентного давления в реальной задаче моделирования турбулентного перемешивания вследствие неустойчивости Рихтмайера–Мешкова. В качестве объекта исследования возьмем эксперимент Феттера и Штуртеванта (см. [17]) по измерению ширины зоны перемешивания при прохождении ударной волны с числом Маха 1.98. Постановка задачи берется из работы [18]. Для моделирования турбулентности используется $k - \varepsilon - a - b$ модель (см. [19]), являющаяся упрощенным вариантом модели BHR.

Были выполнены расчеты этой задачи по схеме с модифицированным для учета ТКЭ решателем HLLC и по схеме с ТКЭ как пассивным скаляром. На фиг. 6 и 7 представлены сравнения: плотность по обеим схемам на момент времени 3 мс и ширина зоны перемешивания для обеих схем с экспери-

ментальными данными. Из них следует, что турбулентное давление оказывает существенное влияние на решение. На фиг. 7 ширина зоны перемешивания в конце расчета на момент времени 2.9 мс без учета ТКЭ на 10% ниже, чем при учете ТКЭ в решателе HLLC.



Фиг. 6. Эксперимент Феттера и Штуртеванта. Распределение плотности (кг/м^3) по координате x (м) на момент времени 3 мс: (——) – с учетом ТКЭ в решателе HLLC, (——) – ТКЭ рассматривается как пассивный скаляр.



Фиг. 7. Эксперимент Феттера и Штуртеванта. Ширина зоны перемешивания (м) от времени (с): (●) – экспериментальные данные, (●) – с учетом ТКЭ в решателе HLLC, (●) – ТКЭ рассматривается как пассивный скаляр.

На фиг. 6 волна разряжения в расчете с ТКЭ пассивным скаляром отстает на 1 см от волны в случае учета ТКЭ. Наибольшее влияние турбулентного давления в этой области объясняется тем, что пик турбулентной кинетической энергии на момент 3 мс располагается в районе точки $x = 1.54$ м, и турбулентное давление в нем составляет 1.85% от статического.

ЗАКЛЮЧЕНИЕ

В настоящей работе было показано, что турбулентное давление, являющееся частью тензора Рейнольдса, требует специальной модификации решателя Римана для системы уравнений сжимаемой газовой динамики, так как, если оно достаточно велико, турбулентность на среднее течение оказывает влияние не только аналогичное вязкости, но также и аналогичное давлению. На примере решателя HLLC показано, что это влияние заключается не только в простом вкладе в общее давление, но также и в изменении метода расчета волновых скоростей. Примером задачи, где подобный подход может быть применен, можно назвать моделирование турбулентного перемешивания вследствие неустойчивости Рихтмайера–Мешкова с помощью $k-\varepsilon$ и $k-L$ моделей турбулентности. При вторичном прохождении ударной волны через турбулизованную контактную границу в этих моделях происходит мощная генерация турбулентной кинетической энергии, и турбулентное давление в максимуме может достигать нескольких процентов от статического давления, что делает использование предложенной в настоящей работе модифицированной схемы HLLC в этих задачах вполне оправданным.

СПИСОК ЛИТЕРАТУРЫ

1. *Годунов С. К.* Разностный метод численного расчета разрывных решений гидродинамики // Матем. сб. 1959. Т. 47. № 3. С. 271–306.
2. *Toro E. F., Spruce M., Speares W.* Restoration of the contact surface in the HLL-Riemann solver // Shock Waves. 1994. V. 4. P. 25–34.
3. *Hu X., Adams N. A., Iaccarino G.* On the HLLC Riemann solver for interface interaction in compressible multi-fluid flow // J. Comput. Phys. 2009. V. 228. P. 6572–6589.
4. *Garrick D. P., Owkes M., Regele J. D.* A finite-volume HLLC-based scheme for compressible interfacial flows with surface tension // J. Comput. Phys. 2017. V. 339. P. 46–67.
5. *Taylor G. I.* The instability of liquid surfaces when accelerated in a direction perpendicular to their planes // Proc. Royal. Soc. Ser. A. 1950. V. 201. P. 192.
6. *Мешков Е. Е.* Неустойчивость границы раздела двух газов, ускоряемой ударной волной // Изв. АН СССР. Мех. жидкости и газа. 1969. № 5. С. 151–157.
7. *Zhou Y.* Rayleigh-Taylor and Richtmyer-Meshkov instability induced flow, turbulence, and mixing. II // Phys. Rep. 2017. V. 723–725. P. 1–160.
8. *Jakobsen H. A.* Chemical reactor modeling. Multiphase reactive flows. Berlin: Springer-Verlag, 2008.
9. *Declercq E., Forestier A., Hérard J.-C., Louis X., Poissant G.* An exact Riemann solver for multicomponent turbulent flow // Inter. J. Comput. Fluid Dyn. 2001. V. 14. P. 117–131.
10. *Toro E. F.* Riemann solvers and numerical methods for fluid dynamics. Berlin: Springer–Verlag, 2009.
11. *Mohammadi B., Pironneau O.* Analysis of the k-ε turbulence model. New York: John Wiley & Sons, 1994.
12. *Davis S. F.* Simplified second–order Godunov–tType methods // SIAM J. Sci. Stat. Comput. 1988. V. 9. P. 445–473.
13. *Pelanti M., Shyue K.-M.* A numerical model for multiphase liquid-vapor-gas flows with interfaces and cavitation // Inter. J. Mul. Flow. 2019. V. 113. P. 208–230.
14. *Sod G. A.* A survey of several finite difference methods for systems of nonlinear hyperbolic conservation Laws // J. Comput. Phys. 1978. V. 27. P. 1–31.
15. *Kamm J. R.* An exact, compressible one-dimensional Riemann solver for general, Convex Equations of State. Los Alamos National Laboratory. 2015. https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-15-21616
16. *van Leer B.* On the relation between the upwind–differencing schemes of Godunov, Enguist-Osher and Roe // SIAM J. Sci. Stat. Comput. 1985. V. 5. P. 1–20.
17. *Vetter M., Sturtevant B.* Experiments on the Richtmyere–Meshkov instability of an air/SF_6 interface // Shock Waves. 1995. V. 4. P. 247–252.
18. *Morán-López J. T., Schilling O.* Multicomponent Reynolds-averaged Naviere-Stokes simulations of reshocked RichtmyereMeshkov instability-induced mixing // High Energy Den. Phys. 2013. V. 9. P. 112–121.

19. *Schwarzkopf J. D., Livescu D., Gore R. A., Rauenzahn R. M., Ristorcelli J. R.* Application of a second-moment closure model to mixing processes involving multicomponent miscible fluids // J. of Turb. V. 12. N 49. P. 1–35.

TURBULENT KINETIC ENERGY IN AN APPROXIMATE SOLVER OF THE RIEMANN GAS DYNAMICS PROBLEM

M. I. Boldyrev*

Russian Federal Nuclear Center—Zababakhin All-Russia Research Institute of Technical Physics, Snezhinsk, Chelyabinsk oblast, 456770 Russia

**e-mail: boldyremni@vniitf.ru*

Received 03 November, 2023
Revised 03 November, 2023
Accepted 06 March, 2024

Abstract. The paper describes the consideration of turbulent kinetic energy in solving the gas-dynamic problem of discontinuity decay (Riemann problem) using the HLLC approximate solver. The system of Euler equations is considered with the addition of the hyperbolic equation of turbulent kinetic energy and consideration of turbulent pressure in the momentum and energy balance equations. The Jacobian coefficient of the system of equations and its eigenvalues are found. Based on this, changes are made to the calculation scheme in the HLLC solver. Using the Sod problem as an example, the correctness of taking into account turbulent kinetic energy in solving the Riemann problem is verified, and the instability of the scheme at high turbulent pressure is shown in the case of not taking turbulence into account in calculating the characteristic velocities.

Keywords: compressible gas dynamics, Euler equations, turbulent kinetic energy, approximate Riemann solver, HLLC, Richtmyer-Meshkov instability.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1055–1063

УДК 532.5

МОДЕЛИРОВАНИЕ ФАЗОВОГО ПЕРЕХОДА ЛЕД–ВОДА В ТРУБЕ С МАЛЫМИ ЛЕДЯНЫМИ НАРОСТАМИ НА СТЕНКЕ[1)]

© 2024 г. Р. К. Гайдуков[1,*], В. Г. Данилов[1]

[1] *Национальный исследовательский университет "Высшая школа экономики", Москва, Россия*

[*] *e-mail: roma1990@gmail.com*

Поступила в редакцию 18.09.2023 г.
Переработанный вариант 18.09.2023 г.
Принята к публикации 05.03.2024 г.

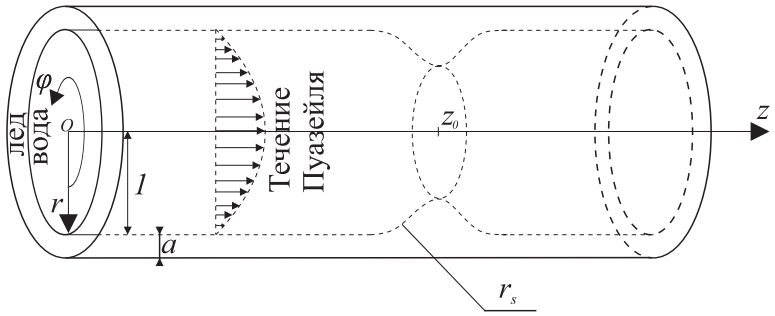
Рассмотрено математическое моделирование фазового перехода лед–вода при течении жидкости внутри трубы с малым ледяным наростом на стенке при больших числах Рейнольдса. В качестве математической модели, описывающей динамику фазового перехода, используется двухпалубная модель пограничного слоя и система фазового поля. Приведены результаты численного моделирования. Библ. 19. Фиг. 5. Табл. 1.

Ключевые слова: фазовый переход, двухпалубная структура, теплоперенос, локализованные возмущения, асимптотика, численное моделирование.

DOI: 10.31857/S0044466924060134, **EDN:** XYDHYY

1. ВВЕДЕНИЕ

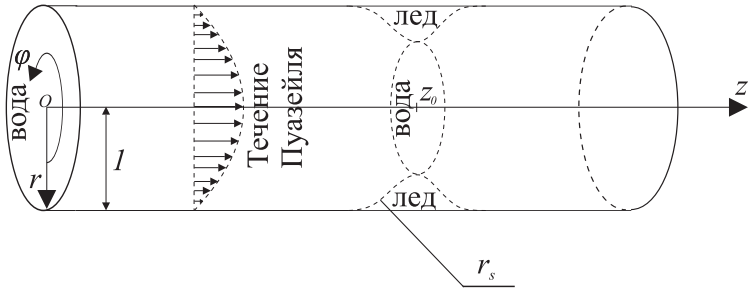
Цель работы – изучение фазового перехода "плавление–кристаллизация" (лед–вода) в задаче о течении вязкой несжимаемой теплопроводной жидкости внутри аксиально-симметричной трубы радиуса R_0, на стенке которой сформировался малый локализованный ледяной нарост (см. фиг. 1). А именно, рассматриваются две ситуации: таяние льда при течении нагретой воды внутри трубы и, наоборот, намерзание льда на стенке при течении переохлажденной воды. Задача исследуется при больших числах Рейнодьдса Re (но таких, при которых еще сохраняется течение Пуазейля в трубе) в отсутствие силы тяжести. Предполагается, что длина трубы достаточная, чтобы течение Пуазейля сформировалось, краевые эффекты не учитываются.



Фиг. 1. Геометрия рассматриваемой задачи: слоя льда с неровностью.

В работе рассматривается случай, когда внутри трубы на ее стенке сформировался ледяной слой толщиной $\hat{a}$ (здесь и далее переменные с "крышкой" являются размерными, а аналогичные переменные без них – безразмерными, процедуру обезразмеривания подробнее см. далее по тексту) с малым

[1)] Исследование выполнено за счет гранта Российского научного фонда № 22-21-00186, https://rscf.ru/project/22-21-00186/.



Фиг. 2. Предел применимости модели: кольцевой ледяной нарост финитной ширины.

ледяным наростом (фиг. 1). Отметим, что предлагаемая в настоящей работе математическая модель позволяет описывать плавление до тех пор, пока существует ледяной слой на всей стенке трубы, т.е. до ситуации, изображенной на фиг. 2. С случае, когда остался только кольцевой ледяной нарост финитной ширины, возникают точки контакта трех сред: нерасплавляемой стенки трубы, льда и воды, в которых не определены уравнения модели, и ее необходимо модифицировать (см. подробнее [1]). В случае намерзания предлагаемая модель применима до тех пор, пока высота горбика остается соизмеримой с толщиной пограничного слоя (см. подробности ниже).

Отметим, что настоящая работа (наряду с цитируемыми ниже) демонстрирует эффективность сочетания асимптотических и численных методов. В рассматриваемой задаче присутствуют два масштаба в поле скоростей течения и свободная граница (граница фазового перехода), разделяющая область течения и объемлющее пространство (собственно трубу). Непосредственное решение такой задача прямым численным интегрированием возможно, но требует серьезных вычислительных средств. В рамках предложенного в работе подхода численное решение задач, полученных в результате применения асимптотических разложений, можно построить с помощью обычного компьютера, не используя высокопроизводительные кластеры.

Ранее в работах [1], [2] были рассмотрены подобные задачи в случае другой геометрии — обтекания ледяного нароста на плоской пластине. Заметим, что не все результаты из этих работ могут быть явно перенесены на случай течения внутри трубы. Рассматриваемая в настоящей работе задача имеет свою специфику. Во-первых, таяние ледяного слоя приводит к изменению радиуса $\hat{R}(\hat{t})$ свободного ото льда пространства, в котором течет вода (далее будем называть его эффективным радиусом). Например, при таянии льда он будет изменяться от R_0 до $R_0+\hat{a}$, где $R_0 = \hat{R}(0)$. Это приводит к изменению поля скоростей течения Пуазейля $u_0(\hat{r}) = \hat{u}_c(\hat{R}(\hat{t})^2 - \hat{r}^2)/4$, где $\hat{u}_c$ — заданный градиент давления в трубе, а $0 \leqslant \hat{r} \leqslant R(t)$. Безусловно, это влияние существенно в случае $\hat{a}$, соизмеримого с R_0. Во-вторых, распределение температуры в рассматриваемой задаче отличается от распределения температуры в задаче обтекания пластинки, что также может влиять на динамику фазового перехода (см. [3]).

Проведем процедуру обезразмеривания задачи. Введем безразмерную цилиндрическую систему координат $(r, \varphi, z) = (\hat{r}, \hat{\varphi}, \hat{z})/R_0$, ось z которой сонаправлена с осью трубы (см. фиг. 1), где R_0 — характерный масштаб (расстояние от оси трубы до слоя льда в начальный момент времени, фиг. 1). Также введем характерную скорость u_∞ — максимальная скорость течения Пуазейля в трубе с ровными стенками, характерное время $t_0 = R_0/u_\infty$, характерную температуру $T_0 = l/c_l$, где c_l — удельная теплоемкость жидкости, а l — скрытая теплота плавления (см. табл. 1). Тогда безразмерный вектор скорости $\mathbf{U} = (v, w, u) = \mathbf{U}/u_\infty$, где v, w, u — радиальная, азимутальная и аксиальная его компоненты, безразмерное давление $p = \hat{p}/(\rho_l u_\infty^2)$, где ρ_l — плотность жидкости, безразмерное время $t = \hat{t}/t_0$, безразмерная температура

$$T = (\hat{T} - T_m)\big/T_0, \tag{1}$$

где T_m — температура плавления льда в случае, когда разделяющая лед и воду поверхность плоская, число Ренольдса $\mathrm{Re} = u_\infty R_0/\nu$, где ν — кинематическая вязкость рассматриваемой жидкости (для воды, например, эта величина имеет порядок 10^{-6}, см. табл. 1).

Будем считать, что неровность (ледяной нарост) находится в точке z_0, ее высота порядка $O(\varepsilon^{4/5})$, а ширина порядка $O(\varepsilon^{2/5})$ (напомним, что рассматривается аксиально-симметричный случай), где

$\varepsilon = \mathrm{Re}^{-1/2}$ — малый параметр. А именно, предполагается, что стенка трубы описывается уравнением

$$r_s = R(t) - \varepsilon^{4/5} h\big(t, (z - z_0)/\varepsilon^{2/5}\big), \tag{2}$$

где $h(t, \xi)$ — некоторая гладкая функция, такая что $h(t, \pm\infty) = 0$, а $R(t)$ — эффективный радиус трубы, учитывающий изменение толщины ледяного слоя, $R(0) = 1$. Отметим, что такие геометрические размеры малой неровности приводят к формированию пограничного слоя с двухпалубной структурой (см. [4]).

Исследуемая задача состоит из двух подзадач. Первая — гидродинамическая задача, представляющая собой нестационарную задачу о течении в аксиально-симметричной трубе с изменяемой во времени формой стенки, описывается безразмерной системой уравнений Навье–Стокса и неразрывности:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial r} + u\frac{\partial v}{\partial z} = -\frac{\partial p}{\partial r} + \varepsilon^2\Big[\Delta v - \frac{v}{r^2}\Big], \tag{3}$$

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial r} + u\frac{\partial u}{\partial z} = -\frac{\partial p}{\partial z} + \varepsilon^2\Delta u, \tag{4}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rv) + \frac{\partial u}{\partial z} = 0, \tag{5}$$

где

$$\Delta f = \frac{1}{r}\frac{\partial}{\partial r}\Big(r\frac{\partial f}{\partial r}\Big) + \frac{\partial^2 f}{\partial z^2},$$

которая дополняется граничными условиями прилипания к стенке трубы $v|_{r=r_s} = u|_{r=r_s} = 0$.

Вторая подзадача — задача о фазовом переходе в подвижной среде с криволинейной границей раздела фаз, которая исходно описывается безразмерными уравнениями теплопроводности в областях жидкой (T_l — температура жидкости) и твердой фаз (T_s — температура льда), имеющими следующий вид с учетом аксиальной симметрии:

$$\frac{\partial T_l}{\partial t} + v\frac{\partial T_l}{\partial r} + u\frac{\partial T_l}{\partial z} = \frac{k_l}{R_0 u_\infty}\Delta T_l + \frac{\nu u_\infty}{R_0 c_l T_0}\varphi, \tag{6}$$

$$\frac{\partial T_s}{\partial t} = k_s \Delta T_s, \tag{7}$$

где коэффициент теплопроводности $k_i = \lambda_i/(\rho_i c_i)$, $i = l, s$, λ_i, ρ_i, c_i — удельная теплопроводность, плотность и удельная теплоемкость (см. табл. 1), функция диссипации имеет вид

$$\varphi = 2\Big(\frac{\partial v}{\partial r}\Big)^2 + 2\Big(\frac{v}{r}\Big)^2 + 2\Big(\frac{\partial u}{\partial z}\Big)^2 + \Big(\frac{\partial v}{\partial z} + \frac{\partial u}{\partial r}\Big)^2.$$

Уравнения теплопроводности (6), (7) дополняются краевыми условиями

$$\frac{\partial T_l}{\partial r}\Big|_{r=0} = 0;\ T_s\big|_{r=1+a} = T_{\text{pipe}} = \text{const} \text{ или} \frac{\partial T_s}{\partial r}\Big|_{r=1+a} = F_{\text{pipe}} = \text{const} \tag{8}$$

(заданной температурой стенки трубы T_{pipe} или заданным тепловым потоком со стенки F_{pipe}), некоторыми заданными начальными условиями $T_l\big|_{t=0} = T_{\text{water}}$, $T_s\big|_{t=0} = T_{\text{ice}}$, условиями на свободной границе

$$T_l\big|_{r=r_s(t)} = T_s\big|_{r=r_s(t)} = \bar{T}_m, \tag{9}$$

безразмерным условием Стефана (см. [2], [5])

$$v_n = \frac{k_l}{u_\infty R_0}\Big[\frac{\partial T}{\partial n}\Big]\Big|_{r=r_s(t)}, \tag{10}$$

где v_n — нормальная скорость точек свободной границы, а $[f]$ — скачок функции f по нормали n, направленной из твердой фазы в жидкую. Температура $\bar{T}_m$ совпадает с температурой плавления T_m на плоской границе раздела фаз. Однако широко известно (см. [6]–[8]), что в случае криволинейной границы раздела фаз температура плавления отличается от равновесной температуры T_m на плоской границе раздела фаз, и $\bar{T}_m$ определяется условием Гиббса–Томсона (см. [6])

$$\bar{T}_m - \frac{T_m}{T_0} + \frac{\sigma T_m}{l\rho_l R_0 T_0}\mathscr{K} + \frac{u_\infty}{T_0 R_0 m} v_n = 0, \tag{11}$$

где $\mathscr{K}$ — средня кривизна поверхности фазового перехода, m — кинетический коэффициент роста, а σ — коэффициент поверхностного натяжения (подробности см. ниже). Значения всех физически постоянных приведены в табл. 1.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Гидродинамическая часть задачи рассмотрена в работе [4] для случая неизменной во времени формы малой неровности (2), но эти результаты тривиально переносятся на случай изменяемой во времени (в силу процессов плавления–кристаллизации) формы обтекаемой поверхности (см. также [9]), рассматриваемый в настоящей работе. А именно, в работе [4] (см. также [10]) получено, что течение имеет двухпалубную структуру пограничного слоя, и интересующее в данной работе течение вблизи неровности имеет следующий вид:

$$v = \varepsilon^{6/5} v^*(t, \xi, \bar{\theta}), \quad u = \varepsilon^{4/5} u^*(t, \xi, \bar{\theta}), \tag{12}$$

где $\bar{\theta}$ — радиальный пространственный масштаб, ξ — аксиальный пространственный масштаб, имеющие следующий вид:

$$\bar{\theta} = \frac{r_s - r}{\varepsilon^{4/5}}, \quad \xi = \frac{z - z_0}{\varepsilon^{2/5}}. \tag{13}$$

Отметим, что масштаб $\bar{\theta}$ введен таким образом, чтобы обтекаемая поверхность стала плоской в переменных $(\xi, \bar{\theta})$, а начало координат перенеслось на границу раздела фаз. Функции v^* и u^* являются решением начально-краевой задачи для системы уравнений Прандтля с самоиндуцированным давлением:

$$\varepsilon^{-2/5}\frac{\partial u^*}{\partial t} + u^*\left(\frac{\partial u^*}{\partial \xi} + \frac{\partial h}{\partial \xi}\frac{\partial u^*}{\partial \bar{\theta}}\right) - v^*\frac{\partial u^*}{\partial \bar{\theta}} + g(t)v^*\big|_{\bar{\theta}\to\infty} - \frac{\partial^2 u^*}{\partial \bar{\theta}^2} - \varepsilon^{-2/5}\frac{\partial h}{\partial t}\frac{\partial u^*}{\partial \bar{\theta}} + \varepsilon^{-6/5}R'(t)\frac{\partial u^*}{\partial \bar{\theta}} = 0, \tag{14}$$

$$\frac{\partial u^*}{\partial \xi} + \frac{\partial h}{\partial \xi}\frac{\partial u^*}{\partial \bar{\theta}} - \frac{\partial v^*}{\partial \bar{\theta}} = 0, \tag{15}$$

$$u^*\big|_{\bar{\theta}=0} = v^*\big|_{\bar{\theta}=0} = 0,\ u^*\big|_{\xi\to\pm\infty} = \bar{\theta}g(t),\ v^*\big|_{\xi\to\pm\infty} = 0,\ \frac{\partial u^*}{\partial \bar{\theta}}\Big|_{\bar{\theta}\to\infty} = g(t),\ u^*\big|_{t=0} = U^*(\xi, \bar{\theta}), \tag{16}$$

где $g(t) = u_0'(0) = u_c R^2(t)/4$, где u_c — заданный (безразмерный) градиент давления, а $R(t)$ — введенный выше эффективный радиус, $R(0) = 1$. Отметим, что наличие коэффициента $\varepsilon^{-2/5}$ перед производной по времени фактически означает рассмотрение течения вблизи поверхности на малых временах $t \sim \varepsilon^{2/5}$ при рассмотрении всей задачи (уравнений Навье–Стокса (3)–(5), а также уравнений теплопроводности (6), (7)) на конечных временах $t \sim 1$. Забегая вперед, заметим, что рассмотрение теплопереноса на малых временах, особенно в задаче о фазовом переходе (скорость плавления довольно мала, порядка 10^{-6} м/с, см. [11]), не представляет интереса, и задачу вблизи поверхности следует рассматривать на конечных временах $t \sim 1$, что означает рассмотрение всей задачи на больших временах $t \sim \varepsilon^{-2/5}$. Для этого нужно формально сделать замену $t_1 = t\varepsilon^{2/5}$, но для упрощения дальнейших обозначений мы ее не будем делать, а лишь опустим коэффициент перед $\partial/\partial t$. Отметим, что это также следует из исследования теплопереноса в течении Пуазейля без фазовых переходов (см. [3]).

Введем масштабы (12), (13) в уравнение теплопроводности (6), положив $\theta = (R(t) - r)/\varepsilon^{4/5}$ ($\theta = \bar{\theta} + h$ — радиальный масштаб с криволинейной границей обтекаемой поверхности):

$$\varepsilon^{2/5}\frac{\partial T_l}{\partial t} - \varepsilon^{2/5}v^*\frac{\partial T_l}{\partial \theta} + \varepsilon^{2/5}u^*\frac{\partial T_l}{\partial \xi} = \frac{k_l}{R_0 u_\infty}\left[\varepsilon^{-8/5}\frac{\partial^2 T_l}{\partial \theta^2} - \frac{1}{R(t) - \varepsilon^{4/5}\theta}\varepsilon^{-4/5}\frac{\partial T_l}{\partial \theta} + \varepsilon^{-4/5}\frac{\partial^2 T_l}{\partial \xi^2}\right] +$$
$$+\frac{\nu u_\infty}{R_0\rho_l c_l T_0}\left[2\left(-\varepsilon^{2/5}\frac{\partial v^*}{\partial \theta}\right)^2 + 2\left(\varepsilon^{6/5}\frac{v^*}{R(t) - \varepsilon^{4/5}\theta}\right)^2 + 2\left(\varepsilon^{2/5}\frac{\partial u^*}{\partial \xi}\right)^2 + \left(\varepsilon^{4/5}\frac{\partial v^*}{\partial \xi} - \frac{\partial u^*}{\partial \theta}\right)^2\right].$$

Отбрасывая малые слагаемые и разделив на $\varepsilon^{2/5}$, получим

$$\frac{\partial T_l}{\partial t} - v^*\frac{\partial T_l}{\partial \theta} + u^*\frac{\partial T_l}{\partial \xi} = \frac{k_l}{R_0 u_\infty}\varepsilon^{-2}\frac{\partial^2 T_l}{\partial \theta^2}. \tag{17}$$

Аналогично уравнение (7) с учетом введенных масштабов примет вид

$$\frac{\partial T_l}{\partial t} = k_s\varepsilon^{-2}\frac{\partial^2 T_s}{\partial \theta^2}. \tag{18}$$

Условия Стефана (10) и Гиббса–Томсона (11) с учетом введенных масштабов примут вид

$$v_n^* = \frac{k_l}{u_\infty R_0}\varepsilon^{-2}\left[\frac{\partial T}{\partial n^*}\right]\Big|_{\theta = h(t,\xi)}, \tag{19}$$

$$\bar{T}_m - \frac{T_m}{T_0} + \frac{\sigma T_m}{l\rho_l R_0 T_0}\varepsilon^{-4/5}\mathscr{K}^* + \varepsilon^{6/5}\frac{u_\infty}{T_0 R_0}v_n^* = 0, \tag{20}$$

где n^* — вектор нормали к границе фазового перехода $\theta = h(t, \xi)$, $n^* = (-h'_\xi, 1)$, а вектор ∇T в соответствии с введенными масштабами имеет вид

$$\nabla T = \left(\varepsilon^{6/5}\,\partial T/\partial \xi, \partial T/\partial \theta\right).$$

Отметим, что толщина ледяного слоя a с учетом введенного масштаба $\bar{\theta}$ станет довольно большой, $\varepsilon^{-4/5}a$, и это фактически означает, что влияние температуры на стенке на распределение температур в окрестности фазового перехода довольно мало. Это означает, что краевое условие (8) при $r = 1 + a$ фактически можно заменить условием при $\theta \to -\infty$. Далее мы будем рассматривать теплоизолированную стенку трубы, т.е. $(\partial T_s/\partial n)|_{\theta\to-\infty} = 0$. Также отметим, что распределение температуры в течении Пуазейля хорошо изучено (см. [3]), и из результатов этой работы видно, что вблизи обтекаемой поверхности температура мало меняется. Это означает, что в тонком слое около обтекаемой поверхности она мало меняется, и с учетом введенных масштабов (13) можно считать $\partial T_l|_{\theta\to\infty} = T_{\text{water}} = \text{const}$.

Однако задача Стефана–Гиббса–Томсона (17)–(20) сложна в вычислительном плане: требуется применение различных ресурсоемких численных методов для точного определения положения свободной границы. Существует другой подход, основанный на введении функции порядка $\varphi_\zeta = \varphi_\zeta(t, \xi, \theta)$, такой, что

$$\varphi_\zeta = \begin{cases} +1, & \theta > h(t,\xi), \\ -1, & \theta < h(t,\xi), \end{cases} \tag{21}$$

т.е. внутри твердой фазы принимает значение -1, внутри жидкой $-+1$, а ζ — параметр регуляризации. В ζ-окрестности границы фазового перехода функция φ_ζ быстро меняется от -1 до $+1$. В рамках этого подхода температура T во всей области и функция порядка φ_ζ определяются из системы уравнений фазового поля (см. [6], [7], [12]), которая имеет следующий вид с учетом введенных выше масштабов:

$$\frac{\partial T}{\partial t} + A\left[u^*\frac{\partial T}{\partial \xi} - v^*\frac{\partial T}{\partial \theta}\right] - \varepsilon^{-2}\frac{\partial}{\partial \theta}\left(k\frac{\partial T}{\partial \theta}\right)\Big) = -\frac{1}{2}\frac{\partial \varphi_\zeta}{\partial t}, \tag{22}$$

$$\zeta^2\alpha\frac{\partial\varphi_\zeta}{\partial t} = \zeta^2\beta\Delta_{\xi,\theta}\varphi_\zeta + \varphi_\zeta\left(1-\varphi_\zeta^2\right) + \zeta\left(1-\varphi_\zeta^2\right)\gamma T/\sqrt{2\beta}, \tag{23}$$

где разрывный безразмерный коэффициент k совпадает с коэффициентами перед вторыми производными в уравнениях (17), (7) в жидкой и твердой фазах соответственно, α β, γ – коэффициенты в условии Гиббса–Томсона (20) перед скоростью, кривизной, температурой:

$$\alpha = \varepsilon^{6/5}\frac{u_\infty}{T_0R_0},\ \beta = \frac{\sigma T_m}{l\rho_l R_0 T_0}\varepsilon^{-4/5},\ \gamma = 1,$$

$A = 1$ в жидкой фазе и $A = 0$ в твердой (слагаемые с коэффициентом A непрерывны на границе раздела фаз в силу того, что скорости можно гладко продолжить нулем в область твердой фазы, см. (16)). Отметим, что система фазового поля (22), (23) является регуляризацией задачи Стефана–Гибса–Томсона (17)–(20) при $\zeta \to +0$ (см. [13], [14]). Граница раздела фаз при таком подходе определяется как линия нулевого уровня функции порядка

$$\{\theta = h(t,\xi)\} = \{(\xi,\theta) : \varphi_\zeta = 0\}. \tag{24}$$

Важно отметить, что система фазового поля (22), (23) эффективно решается численно с помощью обычных разностных схем с постоянным шагом, что позволяет эффективно проводить моделирование всей задачи на обычном персональном компьютере.

3. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

В данном разделе представим некоторые качественные результаты численного решения задачи (14)–(16), (22), (23). Граничные условия для уравнения Алена–Кана (23) – условия Неймана на границе области для функции порядка φ_ζ. Условия для температуры были определены выше. Отметим, что для численного решения этой задачи не требуется каких-либо нетривиальных численных методов – она эффективно решается численно с помощью разностных схем (детали не приводятся ввиду их тривиальности). Параметры, используемые для численного моделирования, следующие: пространственная область $\Omega = \{-50 \leqslant \xi \leqslant 50;\ -50 \leqslant \theta \leqslant 50\}$, время моделирования $0 \leqslant t \leqslant 100$, шаги разностной схемы $h_\xi = h_\theta = 5\times10^{-2}$, $h_t = 5\times10^{-4}$. Отметим, что ошибка, обусловленная таким обрезанием исходной неограниченной области $\bar{\Omega} = \{\xi \in \mathbb{R},\ \theta \in R\}$, будет довольно малой, фактически того же порядка, что и ошибка аппроксимации разностной схемы (см. [9], [15], [16]). Все используемые физические постоянные, а также характерные величины, приведены в табл. 1. Параметр регуляризации $\zeta = 5h_\theta$.

Таблица 1. Физические постоянные (см. [8], [11], [17]–[19]) и характерные масштабы

Физическая величина	Вода	Лед
Удельная теплоемкость	$c_l = 4.2\times10^3$ Дж/(кг · К)	$c_s = 2.05\times10^3$ Дж/(кг · К)
Удельная теплопроводность	$\lambda_l = 0.56$ Вт/(м · К)	$\lambda_s = 2.25$ Вт/(м · К)
Плотность	$\rho_l = 10^3$ кг/м^3	$\rho_s = 0.9\times10^3$ кг/м^3
Поверхностное натяжение	$\sigma = 3.3$ Н/м	
Кинематическая вязкость	$\nu = 1.79\times10^{-6}$ м^2/с	
Кинетический коэффициент роста	$m = 7.3\times10^{-3}$ м/(с · К)	
Температура плавления	$T_m = 273$ К	
Скрытая теплота плавления	$l = 3.3\times10^5$ Дж/кг	
Характерный масштаб	$R_0 = 1$ м	
Характерная скорость	$u_\infty = 1$ м/с	

В качестве начальной формы границы фаз $h(t,\xi)$ выберем

$$h\big|_{t=0} = 5e^{-\xi^2/4}.$$

Отметим, что в работах [4], [10] ранее было получено для случая $h(\xi) = Be^{-\xi^2/4}$, что при превышении $|B|$ некоторого критического значения B^* в ламинарном потоке формируется зона отрывного течения (более точно, на этот процесс, помимо амплитуды, влияет величина угла наклона боковых стенок неровности: чем они круче, тем меньше B^*). Рассматриваемый нами случай $B = 5 > B^*$.

В качестве начальных данных для поля скоростей выберем ламинарное обтекание неровности

$$u^*\big|_{t=0} = \begin{cases} g(0)\bar{\theta}(1 + 0.2h(0,\xi)), & \bar{\theta} \leqslant 5, \\ g(0)(\bar{\theta} + h(0,\xi)), & \bar{\theta} > 5, \end{cases}, \quad \bar{\theta} = \theta - h(0,\xi). \tag{25}$$

Для функции порядка φ примем

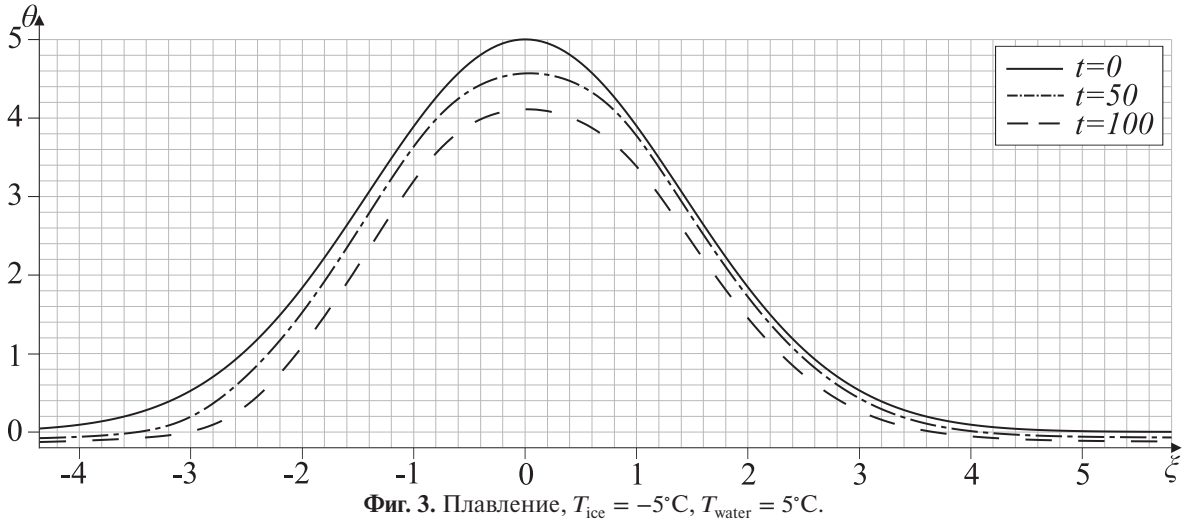
$$\varphi_\zeta\big|_{t=0} = \operatorname{th}\left(\frac{\theta - h(0,\xi)}{\zeta\sqrt{2\beta}}\right).$$

Начальное условие для температуры

$$T\big|_{t=0} = \frac{1}{2}\left[(T_{\text{water}} - T_{\text{ice}})\operatorname{th}\left(\frac{\theta - h(0,\xi)}{4}\right) + (T_{\text{water}} + T_{\text{ice}})\right],$$

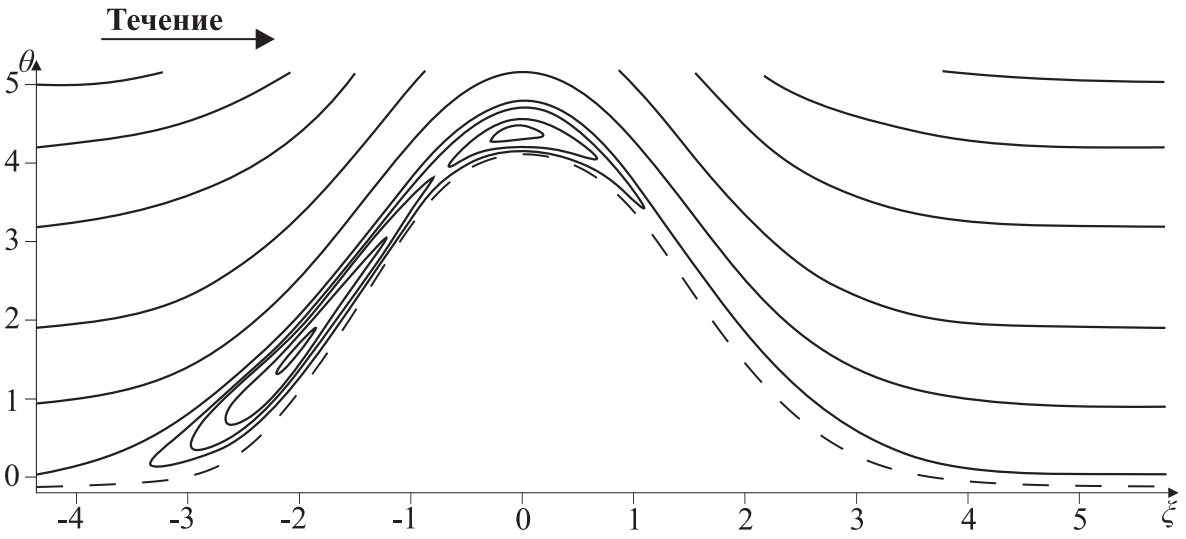
где T_{water} и T_{ice} — постоянные температуры воды и льда на границе области.

Перейдем к результатам численного моделирования. Далее для удобства восприятия будем приводить температуру в градусах Цельсия. На фиг. 3 показано характерное изменение границы раздела фаз с течением времени для случая плавления, начальные температуры воды и льда — $\pm 5^\circ$C. Видно влияние кривизны обтекаемой поверхности и наличия потока: плавление происходит неравномерно — вершина горбика плавится быстрее ровной поверхности, а также наблюдается асимметрия — левая стенка плавится быстрее правой, что обусловлено наличием потока слева направо. Также отметим, что такое асимметричное искажение формы обтекаемой поверхности приводит к образованию зоны отрыва пограничного слоя слева от обтекаемой поверхности (см. фиг. 4).

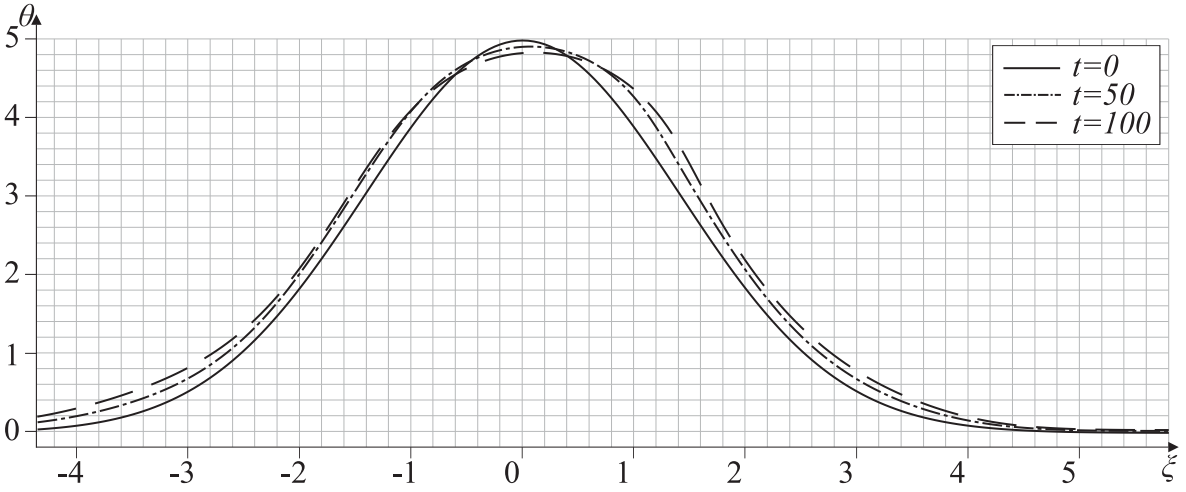


Фиг. 3. Плавление, $T_{\text{ice}} = -5^\circ$C, $T_{\text{water}} = 5^\circ$C.

На фиг. 5 показано характерное изменение границы раздела фаз с течением времени для случая намерзания, обусловленного течением переохлажденной воды ($T_{\text{water}} = -1^\circ$C). Можно заметить, что вершина горбика немного плавится, а намерзание происходит на ровной поверхности и на стенках горбика. Это обуславливается влиянием кривизны поверхности раздела фаз. Также видна асимметрия — на правой стенке намерзает больше, чем на левой, что обуславливается наличием потока слева направо.



Фиг. 4. Плавление, линии тока при $t = 100$.



Фиг. 5. Намерзание, $T_{ice} = -5^{\circ}C$, $T_{water} = -1^{\circ}C$.

4. ЗАКЛЮЧЕНИЕ

В работе предложена математическая модель фазового перехода с учетом наличия течения в жидкой фазе, позволяющая эффективно численно моделировать процессы плавления–намерзания. С помощью построенной модели исследована динамика границы раздела фаз при обтекании малой неровности на ледяной поверхности внутри трубы с масштабами двухпалубной структуры пограничного слоя. Приведены характерные результаты численного моделирования, качественно показывающие динамику границы раздела фаз с учетом влияния кривизны и наличия потока.

СПИСОК ЛИТЕРАТУРЫ

1. *Danilov V. G., Gaydukov R. K.* Ice-water phase transition on a substrate // Rus. J. Math. Phys. 2023. V. 30. P. 165–175.
2. *Гайдуков Р. К., Данилов В. Г., Фонарева А. В.* Моделирование таяния-намерзания льда в задаче обтекания жидкостью малой неровности // Изв. РАН. Механ. жидкости и газа, в печати, 2023.
3. *Aydın O., Avcı M.* Viscous-dissipation effects on the heat transfer in a poiseuille flow, 2006.
4. *Danilov V. G., Gaydukov R. K.* Double-deck structure of the boundary layer in the problem of flow in an axially symmetric pipe with small irregularities on the wall for large Reynolds numbers // Rus. J. Math. Phys. 2017. V. 24. P. 1–18.
5. *Meirmanov A. M.* The Stefan Problem. De Gruyter, 1992.

6. *Caginalp G.* An analysis of a phase field model of a free boundary // Arch. Ration. Mech. Anal. 1986. V. 92. P. 205–245.
7. *Caginalp G.* Stefan and hele-shaw type models as asymptotic limits of the phase field equations // Phys. Rev. A. 1989. V. 39. P. 5887–5896.
8. *Медведев Д. А., Ершов А. П.* Моделирование намерзания льда на подводной трубе газопровода // Вестн. НГУ. Сер. матем., мех., информ. 2013. Т. 13. С. 96–101.
9. *Gaydukov R. K.* Double-deck structure in the fluid flow problem over plate with small irregularities of time-dependent shape // Europ. J. Mech. B/Fluid. 2021. V. 89. P. 401–410.
10. *Fonareva A. V., Gaydukov R. K.* Nonstationary double-deck structure of boundary layers in compressible flow problem inside a channel with small irregularities on the walls // Rus. J. Math. Phys. 2021. V. 28. P. 224–243.
11. *Fernandez R., Barduhn A. J.* The growth rate of ice crystals // Desalinat. 1967. V. 3. P. 330–342.
12. *Plotnikov P. I., Starovoitov V. N.* The stefan problem with surface tension as a limit of the phase field model // Differ. Equat. 1993. V. 29. P. 395–404.
13. *Danilov V. G., Omel'yanov G. A., Radkevich E. V.* Hugoniot-type conditions and weak solutions to the phase-field system // Europ. J. Appl. Math. 1999. V. 10. P. 55–77.
14. *Danilov V. G., Omel'yanov G. A., Shelkovich V. M.* Weak asymptotics method and interaction of nonlinear waves, 2003.
15. *Roache P. J.* Computational fluid dynamics. Hermosa Publ., Albuquerque, 1976.
16. *Yapalparvi R.* Double-deck structure revisited // Europ. J. Mech. B/Fluid. 2012. V. 31. P. 53–70.
17. *Кикоин И. К.* Таблицы физических величин. Справочник. М.: Атомиздат, 1976.
18. *Григорьев И. С., Мейлихов Е. З.* Физические величины. Справочник. М.: Энергоатомиздат, 1991.
19. *Кузнецов В. В., Усть-Качкинцев В. Ф.* Физическая и коллоидная химия: Уч. пособие. М.: Высш. школа, 1976.

MODELING OF ICE-WATER PHASE TRANSITION IN A PIPE WITH SMALL ICE BUILDUPS ON THE WALL

R. K. Gaidukov*, V. G. Danilov

National Research University—Higher School of Economics (HSE University), Moscow, 109028 Russia

**e-mail: romal990@gmail.com*

Received 18 September, 2023
Revised 18 September, 2023
Accepted 05 March, 2024

Abstract. The mathematical modeling of the ice-water phase transition during fluid flow inside a pipe with a small ice buildup on the wall at high Reynolds numbers is considered. As a mathematical model describing the dynamics of the phase transition, a double-deck boundary layer model and a phase field system are used. The results of numerical simulation are presented.

Keywords: phase transition, double-deck structure, heat transfer, localized disturbances, asymptotics, numerical simulation.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1064–1081

УДК 517.958

ЗАДАЧИ ОПРЕДЕЛЕНИЯ КВАЗИСТАЦИОНАРНЫХ ЭЛЕКТРОМАГНИТНЫХ ПОЛЕЙ В СЛАБОНЕОДНОРОДНЫХ СРЕДА[1)]

© 2024 г. А. В. Калинин[1,2,*], А. А. Тюхтина[1,], С. А. Малов[1]**

[1] *603022 Нижний Новгород, пр-т Гагарина, 23, ННГУ им. Н. И. Лобачевского, Россия*
[2] *603950 Нижний Новгород, ул. Ульянова, 66, ИПФ РАН, Россия*

[*] *e-mail: avk@mm.unn.ru*
[**] *e-mail: tyukhtina@iee.unn.ru*

Поступила в редакцию 23.10.2023 г.
Переработанный вариант 28.12.2023 г.
Принята к публикации 05.03.2024 г.

Рассматриваются постановки начально-краевых задач для системы уравнений Максвелла в различных квазистационарных приближениях в однородных и неоднородных проводящих средах. В случае слабонеоднородных сред формулируются и обосновываются асимптотические разложения решений рассматриваемых начально-краевых задач по параметру, характеризующему степень неоднородности среды. Показано, что построение асимптотического разложения для квазистационарного электромагнитного приближения приводит к последовательному решению независимых задач для квазистационарного электрического и квазистационарного магнитного приближения в однородной среде. Приведены условия на начальные данные, при которых асимптотические ряды являются сходящимися. Библ. 32.

Ключевые слова: система уравнений Максвелла, квазистационарное электромагнитное приближение, проводимость, неоднородные среды, асимптотическое разложение.

DOI: 10.31857/S0044466924060144, **EDN:** XYBGFT

1. ВВЕДЕНИЕ

Во многих прикладных задачах для описания относительно медленных электромагнитных процессов используются различные квазистационарные приближения для системы уравнений Максвелла [1]–[3]. Вопросы иерархии квазистационарных приближений обсуждаются в работах [4]–[7]. Наряду с классическими квазистационарными приближениями (квазистационарное электрическое приближение [8], квазистационарное магнитное приближение [9]), значительное внимание уделяется обобщающему их квазистационарному электромагнитному приближению, которое также называется приближением Дарвина [10]. Это приближение основано на разложении электрического поля на потенциальную и вихревую составляющие и сохранении в системе уравнений Максвелла потенциальной составляющей тока смещения. Обсуждению применимости квазистационарного электромагнитного приближения при построении физических моделей квазистационарных процессов посвящены работы [6], [11]–[13]. Вопросы построения численных алгоритмов решения задач для квазистационарного электромагнитного приближения рассматриваются в работах [14]–[20].

Одной из важнейших областей применения квазистационарных приближений является исследование электромагнитных явлений в атмосфере Земли. В зависимости от величины проводимости при моделировании переходных процессов могут использоваться как квазистационарное электрическое [21]–[23], так и квазистационарное магнитное приближения [24]. Для единого описания физи-

[1)] Работа выполнена при финансовой поддержке РНФ (код проекта 23-21-00440), https://rscf.ru/project/23-21-00440.

ческих полей в различных слоях атмосферы естественно использовать квазистационарное электромагнитное приближение [25].

Математическому исследованию внутренних и внешних задач для системы уравнений Максвелла в квазистационарном электромагнитном приближении посвящены работы [4], [10], [26]–[29]. В этих работах для однородных сред и заданной объемной плотности тока были доказаны теоремы о корректности постановок рассматриваемых задач, построены и исследованы асимптотические разложения, обосновывающие применимость квазистационарного электромагнитного приближения. С использованием метода ортогонального проектирования доказана возможность декомпозиции исходной задачи на задачу определения потенциальной составляющей электрического поля, соответствующую квазистационарному электрическому приближению, и задачу определения магнитного поля и вихревой составляющей электрического поля, соответствующую квазистационарному магнитному приближению.

В общем случае, когда объемную плотность тока и напряженность электрического поля связывает обобщенный закон Ома, в средах с неоднородной проводимостью такая декомпозиция невозможна. Соответствующие задачи при общих условиях на коэффициенты были исследованы в работах [7], [25], [30], где были получены результаты о математической корректности постановок и приведены оценки, обосновывающие применимость различных квазистационарных приближений в зависимости от безразмерных параметров, характеризующих проводимость и масштабы неоднородности среды.

В настоящей работе рассматривается система уравнений Максвелла в квазистационарном электромагнитном приближении в случае, когда объемную плотность тока и напряженность электрического поля связывает обобщенный закон Ома, в предположении о слабой неоднородности среды. Показано, что если удельная проводимость является постоянной величиной, возможна декомпозиция начально-краевой задачи, приводящая к задаче определения потенциального электрического поля, соответствующей электрическому приближению, и задаче определения магнитного поля и вихревого электрического поля, соответствующей магнитному приближению. Построены и обоснованы асимптотические разложения решений начально-краевых задач по малому параметру, характеризующему степень неоднородности среды. Показано, что построение асимптотического разложения для квазистационарного электромагнитного приближения приводит к последовательному решению независимых задач для квазистационарного электрического и квазистационарного магнитного приближения в однородной среде. Приведены условия на начальные данные, при которых асимптотические ряды являются сходящимися.

2. КВАЗИСТАЦИОНАРНЫЕ ПРИБЛИЖЕНИЯ ДЛЯ СИСТЕМЫ УРАВНЕНИЙ МАКСВЕЛЛА

В разделе обсуждаются постановки начально-краевых задач для системы уравнений Максвелла в различных квазистационарных приближениях в однородных и неоднородных средах. В гауссовой системе единиц нестационарная система уравнений Максвелла имеет вид [1]

$$\operatorname{rot}\mathbf{H}(x,t)=\frac{4\pi}{c}\mathbf{J}(x,t)+\frac{1}{c}\frac{\partial\mathbf{D}(x,t)}{\partial t}, \tag{1}$$

$$\operatorname{rot}\mathbf{E}(x,t)=-\frac{1}{c}\frac{\partial\mathbf{B}(x,t)}{\partial t}, \tag{2}$$

$$\operatorname{div}\mathbf{B}(x,t)=0, \tag{3}$$

$$\operatorname{div}\mathbf{D}(x,t)=4\pi\rho(x,t), \tag{4}$$

где $(x,t)\in\Omega\times(0,T)$, $\Omega\subset\mathrm{R}^3$, $T>0$. Предполагается, что векторные поля $\mathbf{H}$, $\mathbf{J}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$ удовлетворяют линейным материальным соотношениям

$$\mathbf{D}=\varepsilon\mathbf{E},\ \mathbf{B}=\mu\mathbf{H},\ \mathbf{J}=\sigma\mathbf{E}+\mathbf{J}^{\mathrm{ст}}, \tag{5}$$

позволяющим формулировать задачи определения двух неизвестных функций $\mathbf{E}$ и $\mathbf{H}$, описывающих электрическое и магнитное поле соответственно. При изучении задач в проводящих средах уравнение (4) служит для определения функции ρ.

В работе предполагается, что $\Omega \subset \mathrm{R}^3$ – открытая ограниченная односвязная область с липшицнепрерывной границей Γ, состоящей из компонент связности Γ_1, Γ_2, гомеоморфных сферам в R^3. В почти каждой точке $x \in \Gamma$ определен единичный вектор внешней нормали $\boldsymbol{\nu}(x)$. Подобная пространственная область рассматривается при описании электромагнитных процессов в атмосфере, Γ_1 соответствует в этом случае поверхности Земли, Γ_2 – условной границе атмосферы с ионосферой. Результаты настоящей работы естественным образом обобщаются на случай односвязных областей, граница которых состоит из конечного числа замкнутых поверхностей, гомеоморфных сферам.

Далее предполагается, что $\varepsilon = \mu = 1$, σ – измеримая в Ω функция, при почти всех $x \in \Omega$ удовлетворяющая условиям $\sigma_1 \leqslant \sigma(x) \leqslant \sigma_2$, где $\sigma_i > 0$ $(i = 1, 2)$ – заданные числа.

Квазистационарные приближения для системы (1)–(5) будут рассматриваться при однородном граничном условии

$$\mathbf{E}(x,t) \times \boldsymbol{\nu}(x) = 0, \ (x,t) \in \Gamma \times (0,T). \tag{6}$$

При исследовании задач атмосферного электричества граничные условия (6) соответствуют предположению о том, что поверхности Γ_1 и Γ_2 являются идеальными проводниками.

Квазистационарное электрическое приближение для системы уравнений Максвелла [3, 8] формально заключается в пренебрежении слагаемым $\partial/c\partial t\mathbf{B}$ в уравнении (2). Система уравнений Максвелла с учетом материальных соотношений в этом приближении имеет вид

$$\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\sigma\mathbf{E} + \frac{4\pi}{c}\mathbf{J}^{\mathrm{ст}} + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}, \ \operatorname{rot}\mathbf{E} = 0, \ \operatorname{div}\mathbf{H} = 0. \tag{7}$$

Система (7) рассматривается при граничных условиях (6) и начальных условиях

$$\mathbf{E}(x,0) = \mathbf{e}(x), \ x \in \Omega. \tag{8}$$

Квазистационарное магнитное приближение [2] для системы уравнений Максвелла заключается в пренебрежении током смещения, т.е. в уравнении (1) можно положить $\partial/c\partial t\mathbf{D} \approx 0$. С учетом материальных соотношений система уравнений Максвелла в этом приближении имеет вид

$$\operatorname{rot}\mathbf{H} = \frac{4\pi}{c}\sigma\mathbf{E} + \frac{4\pi}{c}\mathbf{J}^{\mathrm{ст}}, \ \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t}\mathbf{H}, \ \operatorname{div}\mathbf{H} = 0. \tag{9}$$

Система (9) будет рассматриваться при граничных условиях (6) и начальных условиях

$$\mathbf{H}(x,0) = \mathbf{h}(x), \ x \in \Omega. \tag{10}$$

В квазистационарном электромагнитном приближении [4]–[7] уравнение (1) содержит только потенциальную часть тока смещения. Пусть $\mathbf{E} = \mathscr{E} - \operatorname{grad}\varphi$, где $\operatorname{div}\mathscr{E} = 0$. Тогда система уравнений Максвелла с учетом материальных соотношений имеет в этом приближении вид

$$\operatorname{rot}\mathbf{H}(x,t) = \frac{4\pi}{c}\sigma(x)\mathbf{E}(x,t) + \frac{4\pi}{c}\mathbf{J}^{\mathrm{ст}}(x,t) - \frac{1}{c}\frac{\partial}{\partial t}\operatorname{grad}\varphi(x,t), \tag{11}$$

$$\operatorname{rot}\mathbf{E}(x,t) = -\frac{1}{c}\frac{\partial}{\partial t}\mathbf{H}(x,t). \tag{12}$$

Система будет рассматриваться при граничных условиях (6) и начальных условиях

$$\mathbf{H}(x,0) = \mathbf{h}(x), \ \operatorname{grad}\varphi(x,0) = \operatorname{grad}\varphi_0(x). \tag{13}$$

При постановке рассматриваемых начально-краевых задач используются следующие гильбертовы пространства вектор-функций [31], [32]:

$$H(\operatorname{div};\Omega)=\{\mathbf{u}\in\{L_2(\Omega)\}^3:\ \operatorname{div}\mathbf{u}\in L_2(\Omega)\},\ K(\operatorname{div};\Omega)=\{\mathbf{u}\in\{L_2(\Omega)\}^3:\ \operatorname{div}\mathbf{u}=0\},$$

$$(\mathbf{u},\mathbf{v})_{\operatorname{div}}=(\mathbf{u},\mathbf{v})_{2,\Omega}+(\operatorname{div}\mathbf{u},\operatorname{div}\mathbf{v})_{2,\Omega},$$

$$H(\operatorname{rot};\Omega)=\{\mathbf{u}\in\{L_2(\Omega)\}^3:\ \operatorname{rot}\mathbf{u}\in\{L_2(\Omega)\}^3\},\ K(\operatorname{rot};\Omega)=\{\mathbf{u}\in\{L_2(\Omega)\}^3:\ \operatorname{rot}\mathbf{u}=0\},$$

$$(\mathbf{u},\mathbf{v})_{\operatorname{rot}}=(\mathbf{u},\mathbf{v})_{2,\Omega}+(\operatorname{rot}\mathbf{u},\operatorname{rot}\mathbf{v})_{2,\Omega},$$

где через $(\cdot,\cdot)_{2,\Omega}$ обозначено скалярное произведение в $L_2(\Omega)$ и в $\{L_2(\Omega)\}^3$.

Через $H_0(\operatorname{rot};\Omega)$, $H_0(\operatorname{div};\Omega)$ обозначается замыкание множества пробных вектор-функций $\{\mathscr{D}(\Omega)\}^3$ соответственно в $H(\operatorname{rot};\Omega)$ и $H(\operatorname{div};\Omega)$, $K_0(\operatorname{rot};\Omega)=K(\operatorname{rot};\Omega)\cap H_0(\operatorname{rot};\Omega)$, $K_0(\operatorname{div};\Omega)=$ $=K(\operatorname{div};\Omega)\cap H_0(\operatorname{div};\Omega)$.

Пусть $\gamma_\nu:\ H(\operatorname{div};\Omega)\to H^{-1/2}(\Gamma)$ — оператор следа, $\gamma_\nu\mathbf{u}=\mathbf{u}\cdot\boldsymbol{\nu}$,

$$K(\Omega)=\{\mathbf{u}\in K(\operatorname{div};\Omega):\ \langle\gamma_\nu\mathbf{u},1\rangle_{\Gamma_i}=0,i=1,2\},$$

$$H(\Omega)=\{\psi\in H^1(\Omega):\ \psi(x)=0,x\in\Gamma_1,\ \psi(x)=\text{const},\ x\in\Gamma_2\},$$

$$U_1(\Omega)=K(\Omega)\cap H_0(\operatorname{rot};\Omega),\ U_2(\Omega)=K_0(\operatorname{div};\Omega)\cap H(\operatorname{rot};\Omega).$$

Справедливы следующие утверждения [8, 32].

Лемма 1. *Для любой функции* $\mathbf{u}\in K(\operatorname{rot};\Omega)$ *найдется функция* $p\in H^1(\Omega)$ *такая, что* $\mathbf{u}=\operatorname{grad}p$. *Если* $\mathbf{u}\in K_0(\operatorname{rot};\Omega)$, *можно выбрать* $p\in H(\Omega)$.

Лемма 2. *Ортогональное дополнение к* $K(\Omega)$ *в* $\{L_2(\Omega)\}^3$ *совпадает с* $K_0(\operatorname{rot};\Omega)$. *Ортогональное дополнение к* $K_0(\operatorname{div};\Omega)$ *в* $\{L_2(\Omega)\}^3$ *совпадает с* $K(\operatorname{rot};\Omega)$.

2.1. Задачи для системы уравнений Максвелла в однородных средах

Рассмотрим начально-краевые задачи для квазистационарных приближений в однородной проводящей среде, т.е. в предположении, что σ — заданное положительное число. Покажем, что в этом случае возможна декомпозиция задач на задачи определения функций $\mathbf{H}$, $\mathscr{E}$ и задачи определения функции $\operatorname{grad}\varphi$.

Пусть $\mathbf{J}^{\text{ст}}\in L_2(0,T,\{L_2(\Omega)\}^3)$, $\mathbf{h}\in K_0(\operatorname{div};\Omega)$, $\varphi_0\in H(\Omega)$ — заданные функции.

Обобщенным решением начально-краевой задачи (11)–(13), (6) для системы уравнений Максвелла в квазистационарном электромагнитном приближении называются функции $\mathbf{H}\in L_2\big(0,T,U_2(\Omega)\big)$, $\mathscr{E}\in L_2\big(0,T,K(\Omega)\big)$, $\varphi\in L_2\big(0,T,H(\Omega)\big)$ такие, что равенства (11)–(13) выполнены в смысле теории распределений.

Обозначим через P_1 и P_2 операторы ортогонального проектирования из $\{L_2(\Omega)\}^3$ на пространства $K_0(\operatorname{rot};\Omega)$ и $K(\Omega)$ соответственно. Для $\mathbf{u}\in\{L_2(\Omega)\}^3$ функции $P_1\mathbf{u}\in K_0(\operatorname{rot};\Omega)$, $P_2\mathbf{u}\in K(\Omega)$ удовлетворяют при всех $\mathbf{q}\in K_0(\operatorname{rot};\Omega)$, $\mathbf{v}\in K(\Omega)$ равенствам

$$(P_1\mathbf{u},\mathbf{q})_{2,\Omega}=(\mathbf{u},\mathbf{q})_{2,\Omega},\ (P_2\mathbf{u},\mathbf{v})_{2,\Omega}=(\mathbf{u},\mathbf{v})_{2,\Omega}.$$

Теорема 1. *Для всех* $\mathbf{h}\in K_0(\operatorname{div};\Omega)$, $\varphi_0\in H(\Omega)$, $\mathbf{J}^{\text{ст}}\in L_2(0,T,\{L_2(\Omega)\}^3)$ *существует единственное решение* $\mathbf{H}\in L_2(0,T,U_2(\Omega))$, $\mathscr{E}\in L_2(0,T,K(\Omega))$, $\varphi\in H^1(0,T,H(\Omega))$ *задачи* (11)–(13), (6). *При этом* $\mathbf{H}$, $\mathscr{E}$ *— обобщенное решение задачи*

$$\operatorname{rot}\mathbf{H}=\frac{4\pi}{c}\sigma\mathscr{E}+\frac{4\pi}{c}P_2\mathbf{J}^{\text{ст}},\ \operatorname{rot}\mathscr{E}=-\frac{1}{c}\frac{\partial}{\partial t}\mathbf{H},\ \mathbf{H}(0)=\mathbf{h},\tag{14}$$

$\mathbf{H} \in C\left([0,T], K_0(\mathrm{div}\,;\Omega)\right)$ *и справедливы оценки*

$$\|\mathrm{rot}\,\mathbf{H}\|_{2,Q} \leqslant \frac{4\pi}{c}\|P_2\mathbf{J}^{\mathrm{ст}}\|_{2,Q} + \frac{2\sqrt{\pi\sigma}}{c}\|\mathbf{h}\|_{2,\Omega}, \tag{15}$$

$$\|\mathbf{H}\|^2_{C([0,T],\{L_2(\Omega)\}^3)} \leqslant \frac{4\pi}{\sigma}\|P_2\mathbf{J}^{\mathrm{ст}}\|^2_{2,Q} + \|\mathbf{h}\|^2_{2,\Omega}, \tag{16}$$

$$\|\mathscr{E}\|^2_{2,Q} \leqslant \frac{1}{\sigma^2}\|P_2\mathbf{J}^{\mathrm{ст}}\|^2_{2,Q} + \frac{1}{4\pi\sigma}\|\mathbf{h}\|^2_{2,\Omega}. \tag{17}$$

Функция $\mathrm{grad}\,\varphi$ *является обобщенным решением задачи*

$$\frac{\partial}{\partial t}\mathrm{grad}\,\varphi + 4\pi\sigma\mathrm{grad}\,\varphi = 4\pi P_1\mathbf{J}^{\mathrm{ст}},\ \mathrm{grad}\,\varphi(0) = \mathrm{grad}\,\varphi_0 \tag{18}$$

и удовлетворяет неравенству

$$\|\mathrm{grad}\,\varphi\|^2_{2,Q} \leqslant \frac{1}{\sigma^2}\|P_1\mathbf{J}^{\mathrm{ст}}\|^2_{2,Q} + \frac{1}{4\pi\sigma}\min\{1, 4\pi\sigma T\}\|\mathrm{grad}\,\varphi_0\|^2_{2,\Omega}. \tag{19}$$

Если $\mathbf{h} \in U_2(\Omega)$ *и* $\mathbf{E}^{\mathrm{ст}} \in L_2\left(0,T,H_0(\mathrm{rot}\,;\Omega)\right)$*, то* $\partial/\partial t\mathbf{H} \in L_2\left(0,T,\{L_2(\Omega)\}^3\right)$.

Справедливость теоремы 1 вытекает из более общих результатов, полученных в [7], [25]. При этом задача (14) может быть сформулирована как задача определения функции $\mathbf{H} \in L_2\left(0,T,U_2(\Omega)\right)$, удовлетворяющей начальному условию и такой, что для всех $\mathbf{v} \in U_2(\Omega)$

$$\frac{1}{c}\frac{d}{dt}(\mathbf{H},\mathbf{v})_{2,\Omega} + \frac{c}{4\pi}(\sigma^{-1}\mathrm{rot}\,\mathbf{H}, \mathrm{rot}\,\mathbf{v})_{2,\Omega} = (\sigma^{-1}P_2(\mathbf{J}^{\mathrm{ст}}), \mathrm{rot}\,\mathbf{v})_{2,\Omega}. \tag{20}$$

Решение задачи (18), функция $\varphi \in H^1(0,T,H(\Omega))$, при всех $\psi \in H(\Omega)$ удовлетворяет равенству

$$\frac{d}{dt}(\mathrm{grad}\,\varphi, \mathrm{grad}\,\psi)_{2,\Omega} + 4\pi(\sigma\mathrm{grad}\,\varphi, \mathrm{grad}\,\psi)_{2,\Omega} = 4\pi(\mathbf{J}^{\mathrm{ст}}, \mathrm{grad}\,\psi)_{2,\Omega}. \tag{21}$$

Задача (9), (10) для системы уравнений Максвелла в квазистационарном магнитном приближении с использованием проектирования на подпространства разбивается на задачу определения функций $\mathbf{H}$, $\mathscr{E}$, имеющую тот же вид, что и задача (14), и равенство

$$\mathrm{grad}\,\varphi = \sigma^{-1}P_1(\mathbf{J}^{\mathrm{ст}}). \tag{22}$$

Таким образом, магнитное поле и вихревая составляющая электрического поля в случае однородной среды определяются одинаково при использовании квазистационарного магнитного приближения и при использовании квазистационарного электромагнитного приближения.

Начально-краевая задача (7), (8) для системы уравнений Максвелла в квазистационарном электрическом приближении расщепляется на задачу определения функции $\mathrm{grad}\,\varphi \in H^1\left(0,T,K_0(\mathrm{rot}\,;\Omega)\right)$, совпадающую с (18), и систему

$$\mathrm{rot}\,\mathbf{H} = \frac{4\pi}{c}P_2\mathbf{J}^{\mathrm{ст}},\ \mathrm{div}\,\mathbf{H} = 0. \tag{23}$$

Следовательно, потенциальная составляющая электрического поля одинаково определяется в случае однородной среды квазистационарным электрическим и квазистационарным электромагнитным приближением.

Проектируя первое уравнение нестационарной системы уравнений Максвелла на подпространство $K_0(\mathrm{rot}\,;\Omega)$, получим уравнение (18), т.е. потенциальная компонента электрического поля не меняется при переходе к электрическому и к электромагнитному приближениям. Из этого, в частности, следует, что для электрического и электромагнитного приближений остается справедливым закон со-

хранения электрического заряда

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{J} = 0.$$

Таким образом, можно говорить о том, что квазистационарное электромагнитное приближение охватывает классические квазистационарные электрическое и магнитное приближения и занимает промежуточное положение между ними и нестационарной системой уравнений Максвелла.

2.2. Задачи для квазистационарных приближений в неоднородных средах

Приводятся результаты о существовании и единственности решений начально-краевых задач в предположении, что σ — измеримая функция. Декомпозиция задач, рассмотренная в предыдущем пункте, в этом случае невозможна.

Квазистационарное электрическое приближение. Пусть $\mathbf{E} = -\operatorname{grad}\varphi$, $\mathbf{e} = -\operatorname{grad}\varphi_0$. Исключая из системы (7) $\mathbf{H}$, получим задачу определения скалярного электрического потенциала φ:

$$\frac{\partial}{\partial t}\Delta\varphi + 4\pi \operatorname{div}(\sigma \operatorname{grad}\varphi) = 4\pi \operatorname{div}\mathbf{J}^{\text{ст}}, \tag{24}$$

$$\int_{\gamma_1} ((\operatorname{grad}\frac{\partial \varphi}{\partial t} + 4\pi\sigma \operatorname{grad}\varphi - 4\pi\mathbf{J}^{\text{ст}}) \cdot \boldsymbol{\nu}) d\gamma = 0, \tag{25}$$

$$\varphi(x,0) = \varphi_0(x),\ x \in \Omega, \tag{26}$$

$$\varphi(x,t)|_{x\in\Gamma_1} = 0,\ \varphi(x,t)|_{x\in\Gamma_2} = V(t),\ t \in (0,T). \tag{27}$$

Уравнение (24) в исследованиях атмосферного электричества называется уравнением глобальной электрической цепи [21].

Пусть $\varphi_0 \in H(\Omega)$. Обобщенным решением задачи (24)–(27) называется функция $\varphi \in \in H^1\big(0,T,H(\Omega)\big)$, удовлетворяющая условию $\varphi(0) = \varphi_0$ и такая, что для всех $\psi \in H(\Omega)$ имеем

$$\frac{d}{dt}(\operatorname{grad}\varphi, \operatorname{grad}\psi)_{2,\Omega} + 4\pi(\sigma\operatorname{grad}\varphi, \operatorname{grad}\psi)_{2,\Omega} = 4\pi(\mathbf{J}^{\text{ст}}, \operatorname{grad}\psi)_{2,\Omega}. \tag{28}$$

Теорема 2. *Для всех* $\varphi_0 \in H(\Omega)$, $\mathbf{J}^{\text{ст}} \in L_2\big(0,T,\{L_2(\Omega)\}^3\big)$ *существует единственное решение* $\varphi \in H^1\big(0,T,H(\Omega)\big)$ *задачи* (28). *При этом найдется единственная функция* $\mathbf{F} = \operatorname{rot}\mathbf{H} \in L_2\big(0,T,K(\Omega)\big)$ *такая, что выполнено первое равенство в* (7)*, где* $\mathbf{E} = -\operatorname{grad}\varphi$. *Справедлива оценка*

$$\|\operatorname{grad}\varphi\|_{2,Q} \leqslant \frac{1}{\sigma_1}\|\mathbf{J}^{\text{ст}}\|_{2,Q} + \sqrt{T}\|\operatorname{grad}\varphi_0\|_{2,\Omega}. \tag{29}$$

Утверждение теоремы вытекает из результатов, полученных в [8].

Напряженность магнитного поля $\mathbf{H}(x,t)$ может быть найдена как решение в каждый момент t задачи

$$\operatorname{rot}\mathbf{H}(x,t) = \mathbf{F}(x,t),\ \operatorname{div}\mathbf{H}(x,t) = 0, x \in \Omega,\ \mathbf{H}(x,t)\cdot\boldsymbol{\nu}(x) = 0,\ x \in \Gamma. \tag{30}$$

Обобщенным решением задачи (30) называется функция $\mathbf{H} \in L_2\big(0,T,U_2(\Omega)\big)$, которая при всех $\mathbf{v} \in L_2\big(0,T,U_2(\Omega)\big)$ удовлетворяет равенству

$$(\operatorname{rot}\mathbf{H}, \operatorname{rot}\mathbf{v})_{2,Q} = (\mathbf{F}, \operatorname{rot}\mathbf{v})_{2,Q}.$$

Согласно лемме Лакса-Мильграма, эта задача имеет единственное решение.

Квазистационарное магнитное приближение. Пусть $\mathbf{J}^{\text{ст}} = \sigma\mathbf{E}^{\text{ст}}$. Обобщенным решением задачи (9), (10), (6) называется функция $\mathbf{H} \in L_2\big(0,T,U_2(\Omega)\big)$, удовлетворяющая условию $\mathbf{H}(0) = \mathbf{h}$, такая,

что при всех $\mathbf{v} \in U_2(\Omega)$

$$\frac{1}{c}\frac{d}{dt}(\mathbf{H}, \mathbf{v})_{2,\Omega} + \frac{c}{4\pi}(\sigma^{-1}\operatorname{rot}\mathbf{H}, \operatorname{rot}\mathbf{v})_{2,\Omega} = (\mathbf{E}^{\text{ст}}, \operatorname{rot}\mathbf{v})_{2,\Omega}. \tag{31}$$

Теорема 3. *Для всех* $\mathbf{h} \in K_0(\operatorname{div};\Omega)$, $\mathbf{E}^{\text{ст}} \in \{L_2(Q)\}^3$ *существует единственное решение* $\mathbf{H} \in L_2\big(0,T,U_2(\Omega)\big)$ *задачи* (31). *Для функций* $\mathbf{H}$ *и* $\mathbf{E} \in L_2\big(0,T,\{L_2(\Omega)\}^3\big)$, *определяемой соотношением*

$$\mathbf{E} = \frac{c}{4\pi}\sigma^{-1}\operatorname{rot}\mathbf{H} - \mathbf{E}^{\text{ст}},$$

второе равенство в (9) *выполнено в смысле распределений на* $Q = \Omega \times (0,T)$. *Справедливы оценки*

$$\|\operatorname{rot}\mathbf{H}\|_{2,Q} \leqslant \frac{4\pi\sigma_2}{c}\|\mathbf{E}^{\text{ст}}\|_{2,Q} + \frac{2}{c}\sqrt{\pi\sigma_2}\|\mathbf{h}\|_{2,\Omega}, \tag{32}$$

$$\|\mathbf{E}\|^2_{2,Q} \leqslant \frac{\sigma_2}{\sigma_1}\|\mathbf{E}^{\text{ст}}\|^2_{2,Q} + \frac{1}{4\pi\sigma_1}\|\mathbf{h}\|^2_{2,\Omega}. \tag{33}$$

Если $\mathbf{h} \in U_2(\Omega)$ *и* $\mathbf{E}^{\text{ст}} \in L_2\big(0,T,H_0(\operatorname{rot};\Omega)\big)$, *то* $\partial/\partial t\mathbf{H} \in \{L_2(Q)\}^3$.

Квазистационарное электромагнитное приближение. Обобщенным решением задачи (11)–(13) называются функции $\mathbf{H} \in L_2\big(0,T,\{L_2(\Omega)\}^3\big)$, $\varphi \in L_2\big(0,T,H(\Omega)\big)$ и $\boldsymbol{\mathscr{E}} \in L_2\big(0,T,K(\Omega)\big)$ такие, что для всех $\mathbf{u} \in H(\operatorname{rot};\Omega)$, $\psi \in H(\Omega)$, $\mathbf{v} \in U_1(\Omega)$

$$\frac{1}{c}\frac{d}{dt}(\mathbf{H}, \mathbf{u})_{2,\Omega} + (\boldsymbol{\mathscr{E}}, \operatorname{rot}\mathbf{u})_{2,\Omega} = 0, \tag{34}$$

$$\frac{d}{dt}(\operatorname{grad}\varphi, \operatorname{grad}\psi)_{2,\Omega} - 4\pi(\sigma\boldsymbol{\mathscr{E}}, \operatorname{grad}\psi)_{2,\Omega} + 4\pi(\sigma\operatorname{grad}\varphi, \operatorname{grad}\psi)_{2,\Omega} = 4\pi(\mathbf{J}^{\text{ст}}, \operatorname{grad}\psi)_{2,\Omega}, \tag{35}$$

$$(\sigma\boldsymbol{\mathscr{E}}, \mathbf{v})_{2,\Omega} - (\sigma\operatorname{grad}\varphi, \mathbf{v})_{2,\Omega} - \frac{c}{4\pi}(\mathbf{H}, \operatorname{rot}\mathbf{v})_{2,\Omega} = -(\mathbf{J}^{\text{ст}}, \mathbf{v})_{2,\Omega} \tag{36}$$

и выполнены начальные условия

$$\mathbf{H}(0) = \mathbf{h},\ \varphi(0) = \varphi_0. \tag{37}$$

Теорема 4. *Для любых* $\mathbf{h} \in \{L_2(\Omega)\}^3$, $\mathbf{J}^{\text{ст}} \in L_2\big(0,T,\{L_2(\Omega)\}^3\big)$ *существует единственное решение* $\mathbf{H} \in C\big(0,T,H(\operatorname{rot};\Omega)\big)$, $\varphi \in H^1\big(0,T,H(\Omega)\big)$, $\boldsymbol{\mathscr{E}} \in L_2\big(0,T,K(\Omega)\big)$ *задачи* (34)–(37). *Справедливо неравенство*

$$\|\mathbf{E}\|_{2,Q} \leqslant \frac{1}{\sigma_1}\|\mathbf{J}^{\text{ст}}\|_{2,Q} + \frac{1}{\sqrt{4\pi\sigma_1}}(\|\mathbf{h}\|_{2,\Omega} + \|\operatorname{grad}\varphi_0\|_{2,\Omega}). \tag{38}$$

Если $\mathbf{h} \in U_2(\Omega)$, $\mathbf{J}^{\text{ст}} \in H^1\big(0,T,\{L_2(\Omega)\}^3\big)$, *то* $\partial/\partial t\mathbf{H} \in L_2\big(0,T,\{L_2(\Omega)\}^3\big)$, $\varphi \in H^1\big(0,T,H(\Omega)\big)$, $\boldsymbol{\mathscr{E}} \in L_2\big(0,T,U_1(\Omega)\big)$ *и справедливы соотношения* (11), (12).

3. АСИМПТОТИЧЕСКИЙ АНАЛИЗ

Для задач в слабонеоднородных средах обосновываются асимптотические разложения решений начально-краевых задач для системы уравнений Максвелла в различных квазистационарных приближениях в ряды по параметру, характеризующему степень неоднородности среды.

Перейдем к безразмерным переменным, заменяя x на $\Delta x \cdot x'$, t на $\Delta t \cdot t'$, где Δx — характерный пространственный масштаб, Δt — характерный временной масштаб, $(x',t') \in Q' = \Omega' \times (0,T')$, и вводя обозначения $\sigma = \sigma^*\sigma_0$, $\sigma_{01} \leqslant \sigma_0(x') \leqslant \sigma_{02}$,

$$\gamma = 4\pi\Delta t\sigma^*,\ \beta = \frac{\Delta x}{c\Delta t},\ \mathbf{J}^{\text{ст}} = \sigma^*\sigma_0\mathbf{E}^{\text{ст}},$$

где σ^* — характерное значение удельной проводимости.

Система уравнений Максвелла принимает вид

$$\operatorname{rot}\mathbf{H} = \gamma\beta\sigma_0\mathbf{E} + \gamma\beta\sigma_0\mathbf{E}^{\text{ст}} + \beta\frac{\partial}{\partial t'}\mathbf{E},\ \operatorname{rot}\mathbf{E} = -\beta\frac{\partial}{\partial t'}\mathbf{H}.$$

Далее опускаем штрихи при безразмерных переменных (x', t') и области их изменения $Q' = \Omega' \times (0, T')$.

Пусть $\operatorname{grad}\sigma_0 \in \{L_\infty(\Omega)\}^3$, $\sigma_0 = 1 + \eta\tilde{\sigma}$, где $\|\operatorname{grad}\tilde{\sigma}\|_{\infty,\Omega} = 1$, $\|\tilde{\sigma}\|_{\infty,\Omega} \leqslant \tilde{\sigma}^*$, $\eta < (\tilde{\sigma}^*)^{-1}$.

Предполагаем, $\mathbf{E}^{\text{ст}} = \boldsymbol{\mathscr{E}}^{\text{ст}} - \operatorname{grad}\psi^{\text{ст}}$ не зависит от η, начальные функции допускают асимптотические разложения по степеням η:

$$\mathbf{h} = \sum_{k=0}^{\infty}\eta^k\mathbf{h}_k,\ \varphi_0 = \sum_{k=0}^{\infty}\varphi_{0k}, \tag{39}$$

$$\|\mathbf{h} - \mathbf{h}^N\|_{2,\Omega} \leqslant C_{1,N}\eta^{N+1},\ \|\operatorname{grad}\varphi_0 - \operatorname{grad}\varphi_0^N\|_{2,\Omega} \leqslant C_{2,N}\eta^{N+1},\ N = 0, 1, ...,$$

где $\mathbf{h}^N = \sum_{k=0}^N \eta^k\mathbf{h}_k$, $\operatorname{grad}\varphi_0^N = \sum_{k=0}^N \eta^k\operatorname{grad}\varphi_{0k}$, постоянные $C_{1,N}$, $C_{2,N}$ не зависят от η.

Для решений $\mathbf{H}$, $\operatorname{grad}\varphi$, $\boldsymbol{\mathscr{E}}$ задач определения квазистационарных электромагнитных полей получим асимптотические разложения

$$\mathbf{H} = \mathbf{H}^0 + \sum_{k=1}^{\infty}\eta^k\mathbf{H}_k,\ \operatorname{grad}\varphi = \operatorname{grad}\varphi^0 + \sum_{k=1}^{\infty}\eta^k\operatorname{grad}\varphi_k,\ \boldsymbol{\mathscr{E}} = \boldsymbol{\mathscr{E}}^0 + \sum_{k=1}^{\infty}\eta^k\boldsymbol{\mathscr{E}}_k. \tag{40}$$

Используются обозначения

$$\mathbf{H}^N = \mathbf{H}^0 + \sum_{k=1}^{N}\eta^k\mathbf{H}_k,\ \operatorname{grad}\varphi^N = \operatorname{grad}\varphi^0 + \sum_{k=1}^{N}\eta^k\operatorname{grad}\varphi_k,\ \boldsymbol{\mathscr{E}}^N = \boldsymbol{\mathscr{E}}^0 + \sum_{k=1}^{N}\eta^k\boldsymbol{\mathscr{E}}_k,\ N \geqslant 1.$$

Квазистационарное электрическое приближение. Функции $\vec{H} \in L_2\big(0, T, U_2(\Omega)\big)$, $\varphi \in H^1\big(0, T, H(\Omega)\big)$ — решение задачи

$$\operatorname{rot}\mathbf{H} = \beta\gamma(1 + \eta\tilde{\sigma})(-\operatorname{grad}\varphi + \mathbf{E}^{\text{ст}}) - \beta\frac{\partial}{\partial t}\operatorname{grad}\varphi,\ \varphi(0) = \varphi_0, \tag{41}$$

которое, согласно теореме 2, существует и единственно при $\eta < (\tilde{\sigma}^*)^{-1}$.

С использованием операторов проектирования P_1 и P_2 на $K_0(\operatorname{rot};\Omega)$ и $K(\Omega)$ соответственно уравнение принимает вид

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi + \gamma\operatorname{grad}\varphi = -\gamma\operatorname{grad}\psi^{\text{ст}} + \gamma\eta P_1(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)),$$

$$\operatorname{rot}\mathbf{H} = \beta\gamma\mathbf{E}^{\text{ст}} + \beta\gamma\eta P_2(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)).$$

Компоненты асимптотического разложения, функции $\mathbf{H}^0$, $\mathbf{H}_k \in L_2(0, T, U_2(\Omega))$ и $\operatorname{grad}\varphi^0$, $\operatorname{grad}\varphi_k \in H^1(0, T, H(\Omega))$, $k = 1, 2, \ldots$ — решения соответствующих задач

$$\operatorname{rot}\mathbf{H}^0 = \beta\gamma(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^0) - \beta\frac{\partial}{\partial t}\operatorname{grad}\varphi^0,\ \varphi^0(0) = \varphi_{00}, \tag{42}$$

$$\operatorname{rot}\mathbf{H}_1 = -\beta\gamma\operatorname{grad}\varphi_1 + \beta\gamma\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^0) - \beta\frac{\partial}{\partial t}\operatorname{grad}\varphi_1,\ \varphi_1(0) = \varphi_{01}, \tag{43}$$

$$\operatorname{rot}\mathbf{H}_k = -\beta\gamma\operatorname{grad}\varphi_k - \beta\gamma\tilde{\sigma}\operatorname{grad}\varphi_{k-1} - \beta\frac{\partial}{\partial t}\operatorname{grad}\varphi_k,\ \varphi_k(0) = \varphi_{0k},\ k = 2, 3, ... \tag{44}$$

Из (42), проектируя на ортогональные подпространства, получаем

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi^0 + \gamma\operatorname{grad}\varphi^0 = -\gamma\operatorname{grad}\psi^{\text{ст}},\ \operatorname{rot}\mathbf{H}^0 = \beta\gamma\mathscr{E}^{\text{ст}},$$

т. е. нулевое приближение соответствует случаю однородной среды. Далее,

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi_1 + \gamma\operatorname{grad}\varphi_1 = \gamma P_1(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^0)),$$

$$\operatorname{rot}\mathbf{H}_1 = \beta\gamma P_2(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^0)),$$

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi_k + \gamma\operatorname{grad}\varphi_k = -\gamma P_1(\tilde{\sigma}\operatorname{grad}\varphi_{k-1}),$$

$$\operatorname{rot}\mathbf{H}_k = -\beta\gamma P_2(\tilde{\sigma}\operatorname{grad}\varphi_{k-1}),\ k = 2, 3, ...$$

Существование и единственность решений рассматриваемых задач вытекает из теоремы 1.

Теорема 5. *Пусть* $\eta \leqslant \eta^* < (\tilde{\sigma}^*)^{-1}$. *Справедливы оценки*

$$\|\operatorname{grad}(\varphi - \varphi^N)\|_{2,Q} \leqslant E_N\eta^{N+1},\ \|\operatorname{rot}(\mathbf{H} - \mathbf{H}^N)\| \leqslant \beta\gamma\tilde{\sigma}^* E_N\eta^N,\ N = 0, 1, ..., \tag{45}$$

где постоянные $E_N > 0$ *не зависят от* η, β, γ.

Доказательство. Для всех $\psi \in H(\Omega)$ справедливо равенство

$$\frac{d}{dt}(\operatorname{grad}(\varphi - \varphi^0), \operatorname{grad}\psi)_{2,\Omega} + \gamma(\operatorname{grad}(\varphi - \varphi^0), \operatorname{grad}\psi)_{2,\Omega} =$$

$$= -\gamma\eta(\tilde{\sigma}\operatorname{grad}\varphi, \operatorname{grad}\psi)_{2,\Omega} + \gamma\eta(\tilde{\sigma}\mathbf{E}^{\text{ст}}, \operatorname{grad}\psi)_{2,\Omega},$$

из которого следует, что

$$\|\operatorname{grad}(\varphi - \varphi^0)\|_{2,Q} \leqslant \eta(\|\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)\|_{2,Q}^2 + C_{2,0}^2 T)^{1/2}.$$

Функции φ^N, $N \geqslant 1$, — решения задач

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi^N + \gamma\operatorname{grad}\varphi^N = -\gamma\operatorname{grad}\psi^{\text{ст}} + \gamma\eta P_1(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^{N-1})),\ \varphi^N(0) = \varphi_0^N, \tag{46}$$

функции $\varphi - \varphi^N$ удовлетворяют равенствам

$$\frac{\partial}{\partial t}\operatorname{grad}(\varphi - \varphi^N) + \gamma\operatorname{grad}(\varphi - \varphi^N) = \gamma\eta P_1(\tilde{\sigma}(\operatorname{grad}\varphi - \operatorname{grad}\varphi^{N-1})).$$

Из (19) следуют оценки

$$\|\operatorname{grad}(\varphi - \varphi^N)\|_{2,Q} \leqslant \eta((\tilde{\sigma}^*)^2\|\operatorname{grad}(\varphi - \varphi^{N-1})\|_{2,Q}^2 + C_{2,N}^2 T\eta^{2N})^{1/2}.$$

По индукции получаем

$$\|\operatorname{grad}(\varphi - \varphi^N)\|_{2,Q} \leqslant \eta^{N+1}(\tilde{\sigma}^*)^N\left(\|\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)\|_{2,Q}^2 + T(C_{2,0}^2 + \frac{C_{2,1}^2}{(\tilde{\sigma}^*)^2} + ... + \frac{C_{2,N}^2}{(\tilde{\sigma}^*)^{2N}})\right)^{1/2} \leqslant$$

$$\leqslant \eta^{N+1}(\tilde{\sigma}^*)^N\left(\|\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)\|_{2,Q} + \sqrt{T}(C_{2,0} + \frac{C_{2,N}}{\tilde{\sigma}^*} + ... + \frac{C_{2,N}}{(\tilde{\sigma}^*)^N})\right).$$

Для всех $\mathbf{u} \in U_2(\Omega)$ справедливо равенство

$$(\operatorname{rot}(\mathbf{H} - \mathbf{H}^0), \operatorname{rot}\mathbf{u})_{2,\Omega} = \beta\gamma\eta(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi), \operatorname{rot}\mathbf{u})_{2,\Omega},$$

следовательно,

$$\|\operatorname{rot}(\mathbf{H} - \mathbf{H}^0)\|_{2,Q} \leqslant \beta\gamma\eta\|\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)\|_{2,Q}.$$

Пусть $N \geqslant 1$. Тогда

$$\operatorname{rot}\mathbf{H}^N = \beta\gamma(\mathbf{E}^{\text{ст}} + \eta P_2(\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi^{N-1}))),$$

для всех $\mathbf{u} \in U_2(\Omega)$ справедливы равенства

$$(\operatorname{rot}(\mathbf{H} - \mathbf{H}^N), \operatorname{rot}\mathbf{u})_{2,\Omega} = -\beta\gamma\eta(\tilde{\sigma}\operatorname{grad}(\varphi - \varphi^{N-1}), \operatorname{rot}\mathbf{u})_{2,\Omega}.$$

Таким образом,

$$\|\operatorname{rot}(\mathbf{H} - \mathbf{H}^N)\|_{2,Q} \leqslant \beta\gamma\eta\|\tilde{\sigma}\operatorname{grad}(\varphi - \varphi^{N-1})\|_{2,Q},\ N = 1, 2, \ldots.$$

Из теоремы 2 следует, что

$$\|\tilde{\sigma}(\mathbf{E}^{\text{ст}} - \operatorname{grad}\varphi)\|_{2,Q} \leqslant \tilde{\sigma}^*\left(\frac{2}{\sqrt{1 - \eta^*\tilde{\sigma}^*}}\|\mathbf{E}^{\text{ст}}\|_{2,Q} + \sqrt{T}\|\operatorname{grad}\varphi_0\|_{2,\Omega}\right).$$

Таким образом, справедливы оценки (45), где

$$E_N = (\tilde{\sigma}^*)^N\left(\frac{2\tilde{\sigma}^*}{1 - \eta^*\tilde{\sigma}^*}\|\mathbf{E}^{\text{ст}}\|_{2,Q} + \sqrt{T}\left(\tilde{\sigma}^*\|\operatorname{grad}\varphi_0\|_{2,\Omega} + \sum_{k=0}^{N}\frac{C_{2,k}}{(\sigma^*)^k}\right)\right).$$

Теорема доказана.

Квазистационарное магнитное приближение. Функции $\mathbf{H} \in L_2(0, T, U_2(\Omega))$, $\mathbf{E} = \boldsymbol{\mathscr{E}} - \operatorname{grad}\varphi \in$ $\in L_2\big(0, T, \{L_2(\Omega)\}^3\big)$, где $\boldsymbol{\mathscr{E}} \in L_2\big(0, T, K(\Omega)\big)$, $\varphi \in L_2\big(0, T, H(\Omega)\big)$, — обобщенное решение задачи

$$\operatorname{rot}\mathbf{H} = \beta\gamma(1 + \eta\tilde{\sigma})(\boldsymbol{\mathscr{E}} - \operatorname{grad}\varphi + \mathbf{E}^{\text{ст}}),\ \operatorname{rot}\boldsymbol{\mathscr{E}} = -\beta\frac{\partial}{\partial t}\mathbf{H},\ \mathbf{H}(0) = \mathbf{h}.$$

Для нулевого приближения получаем задачу

$$\operatorname{rot}\mathbf{H}^0 = \beta\gamma(\mathbf{E}^{\text{ст}} + \mathbf{E}^0),\ \operatorname{rot}\mathbf{E}^0 = -\beta\frac{\partial}{\partial t}\mathbf{H}^0,\ \mathbf{H}^0(0) = \mathbf{h}_0, \tag{47}$$

функции $\mathbf{H}_k \in L_2(0, T, U_2(\Omega))$, $\mathbf{E}_k = \boldsymbol{\mathscr{E}}_k - \operatorname{grad}\varphi_k \in L_2(0, T, \{L_2(\Omega)\}^3)$ — решения задач

$$\operatorname{rot}\mathbf{H}_1 = \beta\gamma\mathbf{E}_1 + \beta\gamma\tilde{\sigma}(\mathbf{E}^{\text{ст}} + \mathbf{E}_0),\ \operatorname{rot}\mathbf{E}_1 = -\beta\frac{\partial}{\partial t}\mathbf{H}_1,\ \mathbf{H}_1(0) = \mathbf{h}_1, \tag{48}$$

$$\operatorname{rot}\mathbf{H}_k = \beta\gamma\mathbf{E}_k + \beta\gamma\tilde{\sigma}\mathbf{E}_{k-1},\ \operatorname{rot}\mathbf{E}_k = -\beta\frac{\partial}{\partial t}\mathbf{H}_k,\ \mathbf{H}_k(0) = \mathbf{h}_k,\ k = 2, 3, \ldots. \tag{49}$$

Задачи (47)–(49) соответствуют задачам для системы уравнений Максвелла в квазистационарном магнитном приближении в однородной среде. Проектирование на ортогональные подпространства приводит к задачам вида (14) определения функций $\mathbf{H}_k \in L_2\big(0, T, U_2(\Omega)\big)$, $\boldsymbol{\mathscr{E}}_k \in L_2\big(0, T, K(\Omega)\big)$, и равенствам вида (22) для определения функций $\varphi_k \in L_2\big(0, T, H(\Omega)\big)$:

$$\operatorname{rot}\mathbf{H}^0 = \beta\gamma(\boldsymbol{\mathscr{E}}^{\text{ст}} + \boldsymbol{\mathscr{E}}^0),\ \operatorname{rot}\boldsymbol{\mathscr{E}}^0 = -\beta\frac{\partial}{\partial t}\mathbf{H}^0,\ \mathbf{H}^0(0) = \mathbf{h}_0, \tag{50}$$

$$\operatorname{grad}\varphi^0 = -\operatorname{grad}\psi^{\text{ст}}, \tag{51}$$

$$\operatorname{rot}\mathbf{H}_1 = \beta\gamma\boldsymbol{\mathscr{E}}_1 + \beta\gamma P_2(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E}_0)),\ \operatorname{rot}\boldsymbol{\mathscr{E}}_1 = -\beta\frac{\partial}{\partial t}\mathbf{H}_1,\ \mathbf{H}_1(0) = \mathbf{h}_1, \tag{52}$$

$$\operatorname{grad}\varphi_1 = P_1(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E}_0)), \tag{53}$$

$$\operatorname{rot}\mathbf{H}_k = \beta\gamma\boldsymbol{\mathscr{E}}_k + \beta\gamma P_2(\tilde{\sigma}\mathbf{E}_{k-1}),\ \operatorname{rot}\boldsymbol{\mathscr{E}}_k = -\beta\frac{\partial}{\partial t}\mathbf{H}_k,\ \mathbf{H}_k(0) = \mathbf{h}_k, \tag{54}$$

$$\operatorname{grad}\varphi_k = P_1(\tilde{\sigma}\mathbf{E}_{k-1}),\ k = 2, 3, ... \tag{55}$$

Существование и единственность решения поставленных задач вытекает из теоремы 1.

Теорема 6. *Пусть* $\eta \leqslant \eta^* < (\tilde{\sigma}^*)^{-1}$, $\mathbf{E}^N = \boldsymbol{\mathscr{E}}^N - \operatorname{grad}\varphi^N$, $N = 0, 1, \ldots$. *Справедливы оценки*

$$\|\operatorname{grad}\varphi - \operatorname{grad}\varphi^N\|_{2,Q} \leqslant \eta^{N+1}(M_N^1 + \frac{1}{\sqrt{\gamma}}M_N^0),\ \|\mathbf{E} - \mathbf{E}^N\|_{2,Q} \leqslant \eta^{N+1}(M_N^1 + \frac{1}{\sqrt{\gamma}}M_N^2), \tag{56}$$

$$\|\mathbf{H} - \mathbf{H}^N\|_{C(0,T,\{L_2(\Omega)\}^3)} \leqslant \eta^{N+1}(\sqrt{\gamma}M_N^1 + M_N^2),\ \|\operatorname{rot}\mathbf{H} - \operatorname{rot}\mathbf{H}^N\|_{2,Q} \leqslant \eta^{N+1}\beta(\gamma M_N^1 + \sqrt{\gamma}M_N^2), \tag{57}$$

где постоянные M_N^0, M_N^1, M_N^2 *не зависят от* η, β, γ.

Доказательство. Из равенств (52), (54) получаем

$$\operatorname{rot}\mathbf{H}^N = \beta\gamma(\boldsymbol{\mathscr{E}}^N + \boldsymbol{\mathscr{E}}^{\mathrm{CT}}) + \beta\gamma\eta P_2(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E}^{N-1})), \tag{58}$$

$$\operatorname{rot}\boldsymbol{\mathscr{E}}^N = -\beta\frac{\partial}{\partial t}\mathbf{H}^N,\ \mathbf{H}^N(0) = \mathbf{h}^N\ N = 1, 2, \ldots. \tag{59}$$

Следовательно, функции $\mathbf{H} - \mathbf{H}^N \in L_2\big(0, T, U_2(\Omega)\big)$, $\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N \in L_2(0, T, K(\Omega))$, $N = 0, 1, \ldots,$ — обобщенные решения задач

$$\operatorname{rot}\left(\mathbf{H} - \mathbf{H}^0\right) = \beta\gamma\left(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^0\right) + \beta\gamma\eta P_2\left(\tilde{\sigma}(\mathbf{E} + \mathbf{E}^{\mathrm{CT}})\right),$$

$$\operatorname{rot}\left(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^0\right) = -\beta\frac{\partial}{\partial t}\left(\mathbf{H} - \mathbf{H}^0\right),\ \left(\mathbf{H} - \mathbf{H}^0\right)(0) = \mathbf{h} - \mathbf{h}_0,$$

$$\operatorname{rot}\left(\mathbf{H} - \mathbf{H}^N\right) = \beta\gamma\left(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N\right) + \beta\gamma\eta P_2\left(\tilde{\sigma}(\mathbf{E} - \mathbf{E}^{N-1})\right),$$

$$\operatorname{rot}\left(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N\right) = -\beta\frac{\partial}{\partial t}\left(\mathbf{H} - \mathbf{H}^N\right),\ \left(\mathbf{H} - \mathbf{H}^N\right)(0) = \mathbf{h} - \mathbf{h}^N.$$

Для решений этих задач справедливы безразмерные аналоги оценок (15)–(17), т. е.

$$\|\mathbf{H} - \mathbf{H}^0\|^2_{C(0,T,\{L_2(\Omega)\}^3)} \leqslant \eta^2\left(\gamma\|\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E})\|^2_{2,Q} + C^2_{1,0}\right),$$

$$\|\operatorname{rot}(\mathbf{H} - \mathbf{H}^0)\|^2_{2,Q} \leqslant \eta^2\beta^2\left(\gamma^2\|\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E})\|^2_{2,Q} + C^2_{1,0}\gamma\right),$$

$$\|\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^0\|^2_{2,Q} \leqslant \eta^2\left(\|P_2(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E}))\|^2_{2,Q} + \frac{C^2_{1,0}}{\gamma}\right),$$

$$\|\mathbf{H} - \mathbf{H}^N\|^2_{C(0,T,\{L_2(\Omega)\}^3)} \leqslant \eta^2\gamma(\tilde{\sigma}^*)^2\|\mathbf{E} - \mathbf{E}^{N-1}\|^2_{2,Q} + C^2_{1,N}\eta^{2N+2},$$

$$\|\operatorname{rot}(\mathbf{H} - \mathbf{H}^N)\|^2_{2,Q} \leqslant \eta^2\beta^2\gamma^2(\tilde{\sigma}^*)^2\|\mathbf{E} - \mathbf{E}^{N-1}\|^2_{2,Q} + \beta^2\gamma C^2_{1,N}\eta^{2N+2},$$

$$\|\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N\|^2_{2,Q} \leqslant \eta^2\|P_2(\tilde{\sigma}(\mathbf{E} - \mathbf{E}^{N-1})\|^2_{2,Q} + \frac{C^2_{1,N}}{\gamma}\eta^{2N+2} \leqslant \eta^2(\tilde{\sigma}^*)^2\|\mathbf{E} - \mathbf{E}^{N-1}\|^2_{2,Q} + \frac{C^2_{1,N}}{\gamma}\eta^{2N+2}.$$

Используя равенства (51), (55), получаем оценки

$$\|\operatorname{grad}\left(\varphi - \varphi^0\right)\|_{2,Q} \leqslant \eta\|P_1\left(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}} + \mathbf{E})\right)\|_{2,\Omega},\ \|\operatorname{grad}\left(\varphi - \varphi^N\right)\|_{2,Q} \leqslant \eta\|P_1\left(\tilde{\sigma}(\mathbf{E} - \mathbf{E}^{N-1})\right)\|_{2,Q}.$$

Следовательно, справедливы неравенства

$$\|\mathbf{E}-\mathbf{E}^0\|^2_{2,Q}\leqslant\eta^2\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{C^2_{1,0}}{\gamma}\right),$$

$$\|\mathbf{E}-\mathbf{E}^N\|^2_{2,Q}\leqslant\eta^2(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+\frac{C^2_{1,N}}{\gamma}\eta^{2N+2}.$$

По индукции получаем

$$\|\mathbf{E}-\mathbf{E}^N\|_{2,Q}\leqslant\eta^{N+1}(\tilde\sigma^*)^N\left(\|\tilde\sigma(\mathbf{E}+\mathbf{E}^{\mathrm{CT}})\|_{2,Q}+\frac{1}{\sqrt{\gamma}}(C_{1,0}+\frac{C_{1,1}}{\tilde\sigma^*}+\ldots+\frac{C_{1,N}}{(\tilde\sigma^*)^N})\right),$$

$$\|\mathbf{H}-\mathbf{H}^N\|_{C([0,T],\{L_2(\Omega)\}^3)}\leqslant\eta^{N+1}(\tilde\sigma^*)^N\left(\sqrt{\gamma}\|\tilde\sigma(\mathbf{E}+\mathbf{E}^{\mathrm{CT}})\|_{2,Q}+C_{1,0}+\frac{C_{1,1}}{\tilde\sigma^*}+\ldots+\frac{C_{1,N}}{(\tilde\sigma^*)^N}\right),$$

$$\|\mathrm{rot}\,(\mathbf{H}-\mathbf{H}^N)\|_{2,Q}\leqslant\beta\eta^{N+1}(\tilde\sigma^*)^N\left(\gamma\|\tilde\sigma(\mathbf{E}+\mathbf{E}^{\mathrm{CT}})\|_{2,Q}+\sqrt{\gamma}(C_{1,0}+\frac{C_{1,1}}{\tilde\sigma^*}+\ldots+\frac{C_{1,N}}{(\tilde\sigma^*)^N})\right),$$

$$\|\mathrm{grad}\,(\varphi-\varphi^N)\|_{2,Q}\leqslant\eta^{N+1}(\tilde\sigma^*)^{N+1}\|\mathbf{E}^{\mathrm{CT}}+\mathbf{E}\|_{2,Q}.$$

Из (33) следует оценка

$$\|\mathbf{E}+\mathbf{E}^{\mathrm{CT}}\|_{2,Q}\leqslant\frac{1}{\sqrt{1-\eta\tilde\sigma^*}}\left(2\|\mathbf{E}^{\mathrm{CT}}\|_{2,Q}+\frac{1}{\sqrt{\gamma}}\|\mathbf{h}\|_{2,\Omega}\right).$$

Таким образом, справедливы оценки (56), (57), где

$$M^1_N=\frac{2(\tilde\sigma^*)^{N+1}}{\sqrt{1-\eta^*\tilde\sigma^*}}\|\mathbf{E}^{\mathrm{CT}}\|_{2,Q},\ M^0_N=\frac{(\tilde\sigma^*)^{N+1}}{\sqrt{1-\eta^*\tilde\sigma^*}}\|\mathbf{h}\|_{2,\Omega},\ M^2_N=M^0_N+(\tilde\sigma^*)^N\sum_{k=0}^{N}\frac{C_{1,k}}{(\tilde\sigma^*)^k}.$$

Теорема доказана.

Квазистационарное электромагнитное приближение. Пусть $\mathbf{H}\in L_2\left(0,T,U_2(\Omega)\right)$, $\varphi\in H^1\left(0,T,H(\Omega)\right)$, $\mathscr{E}\in L_2\left(0,T,K(\Omega)\right)$ — обобщенное решение задачи

$$\mathrm{rot}\,\mathbf{H}=\beta\gamma(1+\eta\tilde\sigma)\left(\mathscr{E}-\mathrm{grad}\,\varphi+\mathbf{E}^{\mathrm{CT}}\right)-\beta\frac{\partial}{\partial t}\mathrm{grad}\,\varphi,\tag{60}$$

$$\mathrm{rot}\,\mathscr{E}=-\beta\frac{\partial}{\partial t}\mathbf{H},\ \mathbf{H}(0)=\mathbf{h},\ \varphi(0)=\varphi_0.\tag{61}$$

Подставляя разложения (40) в (60), (61), получаем для определения компонент разложения задачи, имеющие тот же вид, что и задача для квазистационарного электромагнитного приближения в однородной среде:

$$\mathrm{rot}\,\mathbf{H}^0=\beta\gamma\left(\mathscr{E}^0-\mathrm{grad}\,\varphi^0+\mathbf{E}^{\mathrm{CT}}\right)-\beta\frac{\partial}{\partial t}\mathrm{grad}\,\varphi^0,$$

$$\mathrm{rot}\,\mathscr{E}^0=-\beta\frac{\partial}{\partial t}\mathbf{H}^0,\ \mathbf{H}^0(0)=\mathbf{h}_0,\ \varphi^0(0)=\varphi_{00},$$

$$\mathrm{rot}\,\mathbf{H}_1=\beta\gamma(\mathscr{E}_1-\mathrm{grad}\,\varphi_1)+\beta\gamma\tilde\sigma(\mathbf{E}^0+\mathbf{E}^{\mathrm{CT}})-\beta\frac{\partial}{\partial t}\mathrm{grad}\,\varphi_1,$$

$$\mathrm{rot}\,\mathscr{E}_1=-\beta\frac{\partial}{\partial t}\mathbf{H}_1,\ \mathbf{H}_1(0)=\mathbf{h}_1,\ \varphi_1(0)=\varphi_{01},$$

$$\operatorname{rot}\mathbf{H}_k = \beta\gamma(\boldsymbol{\mathscr{E}}_k - \operatorname{grad}\varphi_k) + \beta\gamma\tilde{\sigma}\mathbf{E}_{k-1} - \beta\frac{\partial}{\partial t}\operatorname{grad}\varphi_k,$$

$$\operatorname{rot}\boldsymbol{\mathscr{E}}_k = -\beta\frac{\partial}{\partial t}\mathbf{H}_k,\ \mathbf{H}(0)_k = \mathbf{h}_k,\ \varphi_k(0) = \varphi_{0k},\ k = 2, 3, \ldots,$$

где $\mathbf{E}^0 = \boldsymbol{\mathscr{E}}^0 - \operatorname{grad}\varphi^0$, $\mathbf{E}_k = \boldsymbol{\mathscr{E}}_k - \operatorname{grad}\varphi_k$, $k = 1, 2, \ldots$.

Методом ортогонального проектирования поставленные задачи расщепляются на задачи определения функций $\mathbf{H}^0$, $\mathbf{H}_k \in L_2(0, T, U_2(\Omega))$, $\boldsymbol{\mathscr{E}}^0$, $\boldsymbol{\mathscr{E}}_k \in L_2(0, T, K(\Omega))$ и соответствующие задачи определения функций φ^0, $\varphi_k \in H^1(0, T, H(\Omega))$:

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi^0 + \gamma\operatorname{grad}\varphi^0 = -\gamma\operatorname{grad}\psi^{\text{ст}},\ \varphi^0(0) = \varphi_{00},$$

$$\operatorname{rot}\mathbf{H}^0 = \beta\gamma(\boldsymbol{\mathscr{E}}^{\text{ст}} + \boldsymbol{\mathscr{E}}^0),\ \operatorname{rot}\boldsymbol{\mathscr{E}}^0 = -\beta\frac{\partial}{\partial t}\mathbf{H}^0,\ \mathbf{H}^0(0) = \mathbf{h}_0,$$

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi_1 + \gamma\operatorname{grad}\varphi_1 = \gamma P_1(\tilde{\sigma}(\mathbf{E}^{\text{ст}} + \mathbf{E}^0)),\ \varphi_1(0) = \varphi_{01},$$

$$\operatorname{rot}\mathbf{H}_1 = \beta\gamma\boldsymbol{\mathscr{E}}_1 + \beta\gamma P_2(\tilde{\sigma}(\mathbf{E}^{\text{ст}} + \mathbf{E}^0)),$$

$$\operatorname{rot}\boldsymbol{\mathscr{E}}_1 = -\beta\frac{\partial}{\partial t}\mathbf{H}_1,\ \mathbf{H}_1(0) = \mathbf{h}_1,$$

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi_k + \gamma\operatorname{grad}\varphi_k = \gamma P_1(\tilde{\sigma}\mathbf{E}_{k-1}),\ \varphi_k(0) = \varphi_{0k},$$

$$\operatorname{rot}\mathbf{H}_k = \beta\gamma\boldsymbol{\mathscr{E}}_k + \beta\gamma P_2(\tilde{\sigma}\mathbf{E}_{k-1}),$$

$$\operatorname{rot}\boldsymbol{\mathscr{E}}_k = -\beta\frac{\partial}{\partial t}\mathbf{H}_k,\ \mathbf{H}_k(0) = \mathbf{h}_k,\ k = 2, 3, \ldots$$

Существование и единственность решений этих задач вытекает из теоремы 1.

Теорема 7. Пусть $\eta \leqslant \eta^* < (\tilde{\sigma}^*)^{-1}$. Справедливы неравенства

$$\|\operatorname{grad}\varphi - \operatorname{grad}\varphi^N\|_{2,Q} \leqslant \eta^{N+1}\left(K_N^1 + \frac{1}{\sqrt{\gamma}}K_N^2\right),\ \|\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N\|_{2,Q} \leqslant \eta^{N+1}\left(K_N^1 + \frac{1}{\sqrt{\gamma}}K_N^2\right), \tag{62}$$

$$\|\mathbf{H} - \mathbf{H}^N\|_{C(0,T,\{L_2(\Omega)\}^3)} \leqslant \eta^{N+1}\left(\sqrt{\gamma}K_N^1 + K_N^2\right),\ \|\operatorname{rot}\mathbf{H} - \operatorname{rot}\mathbf{H}^N\|_{2,Q} \leqslant \eta^{N+1}\beta\left(\gamma K_N^1 + \sqrt{\gamma}K_N^2\right), \tag{63}$$

где постоянные K_N^1, K_N^2 ($N \geqslant 0$) не зависят от η, β, γ.

Доказательство. Функции $\mathbf{H}^N \in L_2\left(0, T, U_2(\Omega)\right)$, $\boldsymbol{\mathscr{E}}^N \in L_2(0, T, K(\Omega))$, $\varphi^N \in H^1(0, T, H(\Omega))$, $N = 1, 2, \ldots$ — обобщенные решения задач

$$\operatorname{rot}\mathbf{H}^N = \beta\gamma\left(\boldsymbol{\mathscr{E}}^N + \boldsymbol{\mathscr{E}}^{\text{ст}}\right) + \beta\gamma\eta P_2\left(\tilde{\sigma}(\mathbf{E}^{\text{ст}} + \mathbf{E}^{N-1})\right), \tag{64}$$

$$\operatorname{rot}\boldsymbol{\mathscr{E}}^N = -\beta\frac{\partial}{\partial t}\mathbf{H}^N,\ \mathbf{H}^N(0) = \mathbf{h}^N, \tag{65}$$

$$\frac{\partial}{\partial t}\operatorname{grad}\varphi^N + \gamma\operatorname{grad}\varphi^N = \gamma\eta P_1(\tilde{\sigma}(\mathbf{E}^{\text{ст}} + \mathbf{E}^{N-1}) - \gamma\operatorname{grad}\psi^{\text{ст}},\ \varphi^N(0) = \varphi_0^N, \tag{66}$$

где $\mathbf{E}^N = \boldsymbol{\mathscr{E}}^N - \operatorname{grad}\varphi^N$.

Следовательно, функции $\mathbf{H} - \mathbf{H}^N$, $\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^N$, $\varphi - \varphi^N$, $N = 0, 1, \ldots$ — обобщенные решения задач

$$\operatorname{rot}(\mathbf{H} - \mathbf{H}^0) = \beta\gamma(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^0) + \beta\gamma\eta P_2\left(\tilde{\sigma}(\mathbf{E} + \mathbf{E}^{\text{ст}})\right),$$

$$\operatorname{rot}(\boldsymbol{\mathscr{E}} - \boldsymbol{\mathscr{E}}^0) = -\beta\frac{\partial}{\partial t}(\mathbf{H} - \mathbf{H}^0),\ (\mathbf{H} - \mathbf{H}^0)(0) = \mathbf{h} - \mathbf{h}_0,$$

$$\frac{\partial}{\partial t}\mathrm{grad}\,(\varphi-\varphi^0)+\gamma\mathrm{grad}\,(\varphi-\varphi^0)=\beta\gamma\eta P_1\left(\tilde\sigma(\mathbf{E}+\mathbf{E}^{\mathrm{CT}})\right),\ (\varphi-\varphi^0)(0)=\varphi_0-\varphi_{00},$$

$$\mathrm{rot}\,(\mathbf{H}-\mathbf{H}^N)=\beta\gamma(\boldsymbol{\mathscr{E}}-\boldsymbol{\mathscr{E}}^N)+\beta\gamma\eta P_2\left(\tilde\sigma(\mathbf{E}-\mathbf{E}^{N-1})\right),$$

$$\mathrm{rot}\,(\boldsymbol{\mathscr{E}}-\boldsymbol{\mathscr{E}}^N)=-\beta\frac{\partial}{\partial t}(\mathbf{H}-\mathbf{H}^N),\ (\mathbf{H}-\mathbf{H}^N)(0)=\mathbf{h}-\mathbf{h}^N,$$

$$\frac{\partial}{\partial t}\mathrm{grad}\,(\varphi-\varphi^N)+\gamma\mathrm{grad}\,(\varphi-\varphi^N)=\beta\gamma\eta P_1\left(\tilde\sigma(\mathbf{E}-\mathbf{E}^{N-1})\right),\ (\varphi-\varphi^N)(0)=\varphi_0-\varphi_0^N.$$

Для решений этих задач справедливы безразмерные аналоги оценок (15)–(17), (19), из которых получаем

$$\|\mathbf{H}-\mathbf{H}^0\|^2_{C(0,T,\{L_2(\Omega)\}^3)}\leqslant\eta^2\left(\gamma\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+C^2_{1,0}\right),$$

$$\|\mathrm{rot}\,(\mathbf{H}-\mathbf{H}^0)\|^2_{2,Q}\leqslant\eta^2\beta^2\left(\gamma^2\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+C^2_{1,0}\gamma\right),$$

$$\|\boldsymbol{\mathscr{E}}-\boldsymbol{\mathscr{E}}^0\|^2_{2,Q}\leqslant\eta^2\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{C^2_{1,0}}{\gamma}\right),$$

$$\|\mathrm{grad}\,(\varphi-\varphi^0)\|^2_{2,Q}\leqslant\eta^2\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{C^2_{2,0}}{\gamma}\right),$$

$$\|\mathbf{E}-\mathbf{E}^0\|^2_{2,Q}\leqslant\eta^2\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{C^2_{1,0}+C^2_{2,0}}{\gamma}\right),$$

$$\|\mathbf{H}-\mathbf{H}^N\|^2_{C(0,T,\{L_2(\Omega)\}^3)}\leqslant\eta^2\gamma(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+C^2_{1,N}\eta^{2N+2},$$

$$\|\mathrm{rot}\,(\mathbf{H}-\mathbf{H}^N)\|^2_{2,Q}\leqslant\eta^2\beta^2\gamma^2(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+\beta^2\gamma C^2_{1,N}\eta^{2N+2},$$

$$\|\boldsymbol{\mathscr{E}}-\boldsymbol{\mathscr{E}}^N\|^2_{2,Q}\leqslant\eta^2(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+\frac{C^2_{1,N}}{\gamma}\eta^{2N+2},$$

$$\|\mathrm{grad}\,(\varphi-\varphi^N)\|^2_{2,Q}\leqslant\eta^2(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+\eta^{2N+2}\frac{C^2_{2,N}}{\gamma},$$

$$\|\mathbf{E}-\mathbf{E}^N\|^2_{2,Q}\leqslant\eta^2(\tilde\sigma^*)^2\|\mathbf{E}-\mathbf{E}^{N-1}\|^2_{2,Q}+\frac{C^2_{1,N}+C^2_{2,N}}{\gamma}\eta^{2N+2}.$$

По индукции доказывается неравенство

$$\|\mathbf{E}-\mathbf{E}^N\|^2_{2,Q}\leqslant\eta^{2N+2}(\tilde\sigma^*)^{2N}\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{1}{\gamma}\sum_{k=0}^{N}\frac{C^2_{1,k}+C^2_{2,k}}{(\tilde\sigma^*)^{2k}}\right),$$

из которого следуют оценки

$$\|\mathrm{grad}\,(\varphi-\varphi^N)\|^2_{2,Q}\leqslant\eta^{2N+2}(\tilde\sigma^*)^{2N}\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{1}{\gamma}\sum_{k=0}^{N-1}\frac{C^2_{1,k}+C^2_{2,k}}{(\tilde\sigma^*)^{2k}}+\frac{1}{\gamma}\frac{C^2_{2,N}}{(\tilde\sigma^*)^{2N}}\right),$$

$$\|\boldsymbol{\mathscr{E}}-\boldsymbol{\mathscr{E}}^N\|^2_{2,Q}\leqslant\eta^{2N+2}(\tilde\sigma^*)^{2N}\left(\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\frac{1}{\gamma}\sum_{k=0}^{N-1}\frac{C^2_{1,k}+C^2_{2,k}}{(\tilde\sigma^*)^{2k}}+\frac{1}{\gamma}\frac{C^2_{1,N}}{(\tilde\sigma^*)^{2N}}\right),$$

$$\|\mathbf{H}-\mathbf{H}^N\|^2_{C(0,T,\{L_2(\Omega)\}^3)}\leqslant\eta^{2N+2}(\tilde\sigma^*)^{2N}\left(\gamma\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\sum_{k=0}^{N-1}\frac{C^2_{1,k}+C^2_{2,k}}{(\tilde\sigma^*)^{2k}}+\frac{C^2_{1,N}}{(\tilde\sigma^*)^{2N}}\right),$$

$$\|\mathrm{rot}\,(\mathbf{H}-\mathbf{H}^N)\|^2_{2,Q} \leqslant \eta^{2N+2}(\tilde\sigma^*)^{2N}\beta^2\left(\gamma^2\|\tilde\sigma(\mathbf{E}^{\mathrm{CT}}+\mathbf{E})\|^2_{2,Q}+\gamma\sum_{k=0}^{N-1}\frac{C^2_{1,k}+C^2_{2,k}}{(\tilde\sigma^*)^{2k}}+\gamma\frac{C^2_{1,N}}{(\tilde\sigma^*)^{2N}}\right).$$

Из (38) вытекает неравенство

$$\|\tilde\sigma(\mathbf{E}+\mathbf{E}^{\mathrm{CT}})\|_{2,Q} \leqslant \frac{2\tilde\sigma^*}{1-\eta^*\tilde\sigma^*}\|\mathbf{E}^{\mathrm{CT}}\|_{2,Q}+\frac{\tilde\sigma^*}{\sqrt{\gamma(1-\eta^*\tilde\sigma^*)}}(\|\mathbf{h}\|_{2,\Omega}+\|\mathrm{grad}\,\varphi_0\|_{2,\Omega}).$$

Таким образом, справедливы оценки (62), (63), где

$$K^1_N=\frac{2(\tilde\sigma^*)^{N+1}}{1-\eta^*\tilde\sigma^*}\|\mathbf{E}^{\mathrm{CT}}\|_{2,Q},\ K^2_N=(\tilde\sigma^*)^N\left(\frac{\tilde\sigma^*}{1-\eta^*\tilde\sigma^*}(\|\mathbf{h}\|_{2,\Omega}+\|\mathrm{grad}\,\varphi_0\|_{2,\Omega})+\sum_{k=0}^{N}\frac{C_{1,k}+C_{2,k}}{(\tilde\sigma^*)^k}\right).$$

Теорема доказана.

Замечание 1. При совпадении исходных данных $\mathbf{E}^{\mathrm{CT}}$, $\mathbf{h}$, $\mathrm{grad}\,\varphi_0$, нулевые приближения для магнитного поля и вихревой составляющей электрического поля, функции $\mathbf{H}^0$ и $\mathbf{E}^0$, совпадают для квазистационарного магнитного приближения и квазистационарного электромагнитного приближения, нулевые приближения потенциальной составляющей электрического поля, функции $\mathrm{grad}\,\varphi^0$, совпадают для квазистационарного электрического приближения и квазистационарного электромагнитного приближения.

Для всех $N \geqslant 1$ задачи (64), (65) определения приближений $\mathbf{H}^N$, $\mathscr{E}^N$ для квазистационарного электромагнитного приближения имеют тот же вид, что и соответствующие задачи (58), (59) для квазистационарного магнитного приближения. Аналогично, совпадают задачи (66) и (46) определения функций $\mathrm{grad}\,\varphi^N$ для квазистационарного электромагнитного приближения и квазистационарного электрического приближения соответственно.

Замечание 2. Из доказательств теорем 5–7 следует, что сходимость полученных асимптотических рядов зависит от свойств рядов $\sum_{k=0}^{\infty}C_{1,k}(\tilde\sigma^*)^{-k}$, $\sum_{k=0}^{\infty}C_{2,k}(\tilde\sigma^*)^{-k}$. В частности, асимптотические ряды сходятся, если начальные функции $\mathbf{h}$ и $\mathrm{grad}\,\varphi_0$ не зависят от η. Справедливо также следующее утверждение.

Лемма 3. *Предположим, начальные данные задач удовлетворяют условию согласования*

$$\mathrm{rot}\,\mathbf{h}=-\frac{4\pi}{c}\sigma\mathrm{grad}\,\varphi_0+\frac{4\pi}{c}\mathbf{J}^{\mathrm{CT}}(0)$$

или, в безразмерных переменных,

$$\mathrm{rot}\,\mathbf{h}=-\beta\gamma\sigma_0\mathrm{grad}\,\varphi_0+\beta\gamma\sigma_0\mathbf{E}^{\mathrm{CT}}(0),$$

позволяющему избежать эффекта пограничного слоя по времени. Тогда асимптотические ряды (40) *для решений начально-краевых задач для системы уравнений Максвелла в квазистационарных приближениях сходятся при* $\eta \leqslant \eta^* < (\tilde\sigma^*)^{-1}$.

Доказательство. Пусть $\sigma_0 = 1+\eta\tilde\sigma$. Для компонент разложения (39) справедливы равенства

$$\mathrm{rot}\,\mathbf{h}_0=-\beta\gamma\mathrm{grad}\,\varphi_{00}+\beta\gamma\mathbf{E}^{\mathrm{CT}}(0),$$

$$\mathrm{rot}\,\mathbf{h}_1=-\beta\gamma\mathrm{grad}\,\varphi_{01}+\beta\gamma\tilde\sigma(\mathbf{E}^{\mathrm{CT}}(0)-\mathrm{grad}\,\varphi_{00}),$$

$$\mathrm{rot}\,\mathbf{h}_k=-\beta\gamma\mathrm{grad}\,\varphi_{0k}-\beta\gamma\tilde\sigma(\mathrm{grad}\,\varphi_{0,k-1}).$$

Следовательно,

$$\mathrm{grad}\,\varphi_{00}=-\mathrm{grad}\,\psi^{\mathrm{CT}},\ \mathrm{rot}\,\mathbf{h}_0=\beta\gamma\mathscr{E}^{\mathrm{CT}}(0),$$

$$\mathrm{grad}\,\varphi_{01}=P_1(\tilde\sigma(\mathbf{E}^{\mathrm{CT}}(0)-\mathrm{grad}\,\varphi_{00})),\ \mathrm{rot}\,\mathbf{h}_1=\beta\gamma P_2(\tilde\sigma\mathbf{E}^{\mathrm{CT}}(0)-\mathrm{grad}\,\varphi_{00}),$$

$$\operatorname{grad}\varphi_{0k} = P_1(\tilde{\sigma}(\operatorname{grad}\varphi_{0,k-1})),\ \operatorname{rot}\mathbf{h}_k = -\beta\gamma P_2(\tilde{\sigma}\operatorname{grad}\varphi_{0,k-1}).$$

Таким образом,

$$\operatorname{grad}(\varphi_0 - \varphi_{00}) = \eta P_1\left(\tilde{\sigma}(\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0)\right),\ \operatorname{rot}(\mathbf{h} - \mathbf{h}_0) = \beta\gamma\eta P_2\left(\tilde{\sigma}\left(\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0\right)\right),$$

$$\operatorname{grad}\varphi_0 - \operatorname{grad}\varphi_0^N = -\eta P_1\left(\tilde{\sigma}\left(\operatorname{grad}\varphi_0 - \operatorname{grad}\varphi_0^{N-1}\right)\right),$$

$$\operatorname{rot}\mathbf{h} - \operatorname{rot}\mathbf{h}_N = -\beta\gamma\eta P_2\left(\tilde{\sigma}\left(\operatorname{grad}\varphi_0 - \operatorname{grad}\varphi_0^{N-1}\right)\right).$$

По индукции получаем,

$$\|\operatorname{grad}\varphi_0 - \sum_{k=1}^{N}\eta^k\operatorname{grad}\varphi_{0k}\|_{2,\Omega} \leqslant \eta^{N+1}(\tilde{\sigma}^*)^{N+1}\|\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0\|_{2,\Omega},$$

$$\|\operatorname{rot}\mathbf{h} - \sum_{k=1}^{N}\eta^k\operatorname{rot}\mathbf{h}_k\|_{2,\Omega} \leqslant \eta^{N+1}\beta\gamma(\tilde{\sigma}^*)^{N+1}\|\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0\|_{2,\Omega}.$$

Следовательно,

$$\sum_{k=0}^{N}\frac{C_{2,k}}{(\tilde{\sigma}^*)^k} \leqslant (N+1)\tilde{\sigma}^*\|\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0\|_{2,\Omega},\ \sum_{k=0}^{N}\frac{C_{1,k}}{(\tilde{\sigma}^*)^k} \leqslant \beta\gamma(N+1)\tilde{\sigma}^*A(\Omega)\|\mathbf{E}^{\mathrm{CT}}(0) - \operatorname{grad}\varphi_0\|_{2,\Omega},$$

где постоянная $A(\Omega)$ зависит только от области Ω. Таким образом, правые части установленных в теоремах 5–7 оценок для остаточных сумм асимптотических рядов стремятся к нулю при $N \to \infty$.

Лемма доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Ландау Л. Д., Лифшиц Е. М.* Теоретическая физика. Том 8. Электродинамика сплошных сред. М.: Наука, Физматлит, 1982.
2. *Тамм И. Е.* Основы теории электричества. М.: Наука, 1989.
3. *Толмачев В. В., Головин А. М., Потапов В. С.* Термодинамика и электродинамика сплошной среды. М.: Изд-во МГУ, 1988.
4. *Raviart P.-A., Sonnendrücker E.* A hierarchy of approximate models for the Maxwell equations // Numer. Math. 1996. V. 73. P. 329–372.
5. *Larsson J.* Electromagnetics from a quasistatic perspective // Am. J. Phys. 2007. V. 75. N 3. P. 230–239.
6. *Kruger S. E.* The three quasistatic limits of the Maxwell equations // arXiv:1909.11264, 2019.
7. *Kalinin A. V., Tyukhtina A. A.* Hierarchy of models of quasi-stationary electromagnetic fields // MMST 2020, Revised Selected Papers. CCIS, v. 1413. Springer, 2021. P. 77–92.
8. *Kalinin A. V., Slyunyaev N. N.* Initial-boundary value problems for the equations of the global atmospheric electric circuit // J. Math. Anal. Appl. 2017. V. 450. N 1. P. 112–136.
9. *Alonso Rodriguez A., Valli A.* Eddy current approximation of Maxwell equations. Theory, algorithms and applications. Milan: Spriner-Verlag Italia, 2010.
10. *Degond P., Raviart P.-A.* An analysis of the Darwin model of approximation to Maxwell's equations // Forum Math. 1992. V. 4. P. 13–44.
11. *Kaufman A. N., Rostler P. S.* The Darwin model as a tool for electromagnetic plasma simulation // Phys. Fluids. 1971. V. 14. N 2. P. 446–448.
12. *Hewett D. W., Boyd J. K.* Streamlined Darwin simulation of nonneutral plasmas // J. Comput. Phys. 1987. V. 70. P. 166–181.
13. *Krause T. B., Apte A., Morrison P. J.* A unified approach to the Darwin approximation // Phys. Plasmas 2007. V. 14. 102112.

14. *Sonnendrücker E., Ambrosiano J. J., Scott T. Brandon S. T.* A finite element formulation of the Darwin PIC model for use on unstructured grids // J. Comp. Phys. 1995. V. 121. N 2. P. 281–297.
15. *Ciarlet P. Jr., Zou J.* Finite element convergence for the Darwin model to Maxwell's equations // Math. Modeling Numer. Anal. 1997. V. 31. N 2. P. 213–249.
16. *Fang N., Liao C., Ying L. A.* Darwin Approximation to Maxwell's Equations // ICCS 2009. Lecture Notes in Computer Science. V. 5544. Berlin: Springer, 2009.
17. *Koch S., Schneider H., Weiland T.* A low-frequency approximation to the Maxwell equations simultaneously considering inductive and capacitive phenomena // IEEE Trans. Magn. 2012. V. 48. P. 511–514.
18. *Badics Z., Pavo J., Bilicz S., Gyimothy S.* Subdomain perturbation finite element method for quasi-static Darwin approximation // IEEE Trans. Magn. 2020. V. 56. N 1. Art no. 7503304.
19. *Yan S., Tang Z., Henneron T., Ren Z.* Structure-preserved reduced-order modeling for frequency-domain solution of the Darwin model with a gauged potential formulation // IEEE Trans. Magn. 2020. V. 56. N 1. Art no. 7500404.
20. *Clemens M., Kasolis F., Henkel M.-L., Kähne B., Günther M.* A two-step Darwin model time-domain formulation for quasi-static electromagnetic field calculations // IEEE Trans. Magn. 2021. V. 57. N 6. P. 1–4.
21. *Мареев Е. А.* Достижения и перспективы исследований глобальной электрической цепи // Успехи физ. наук. 2010. Т. 180. N 5. С. 527–534.
22. *Калинин А. В., Слюняев Н. Н., Мареев Е. А., Жидков А. А.* Стационарные и нестационарные модели глобальной электрической цепи: корректность, аналитические соотношения, численная реализация // Известия РАН. Физика атмосферы и океана. 2014. Т. 50. N 3. С. 314–322.
23. *Slyunyaev N. N., Kalinin A. V., Mareev E. A.* Thunderstorm generators operating as voltage sources in global electric circuit models // J. Atm. Solar-Terr. Phys. 2019. V. 183. P. 99–109.
24. *Shalimov S. L., Bösinger T.* An alternative explanation for the ultra-slow tail of sprite-associated lightning discharges // J. Atm. and Solar-Terr. Phys. 2006. V. 68. N 7. P. 814–820.
25. *Калинин А. В., Тюхтина А. А.* Некоторые математические задачи атмосферного электричества // Итоги науки и техники. Совр. мат. прил. 2022. Т. 207. С. 48–60.
26. *Raviart P.-A., Sonnendrücker E.* Approximate models for the Maxwell equations // J. Comput. Appl. Math. 1994. V. 63. P. 69–81.
27. *Ciarlet P., Sonnendrücker E.* A Decomposition of the electromagnetic field – application to the Darwin model // Math. Mod. Meth. Appl. Sci. 1997. V. 07. N 8. P. 1085–1120.
28. *Fang N., Ying L.* Three dimensional exterior problem of the Darwin model and its numerical computation // Math. Mod. Meth. Appl. Sci. 2008. V. 18. N 10. P. 1673–1701.
29. *Liao C., Ying L.* An analysis of the Darwin model of approximation to Maxwell equations in 3-D unbounded domains // Comm. Math. Sci. 2008. V. 6. N 3. P. 695–710.
30. *Калинин А. В., Тюхтина А. А.* Приближение Дарвина для системы уравнений Максвелла в неоднородных проводящих средах // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. N 8. С. 121–134.
31. *Темам Р.* Уравнения Навье–Стокса. Теория и численный анализ. М.: Мир, 1981.
32. *Girault V., Raviart P.* Finite element methods for Navier–Stokes equations. N.Y.: Springler-Verlag, 1986.

PROBLEMS OF DETERMINING QUASI-STATIONARY ELECTROMAGNETIC FIELDS IN WEAKLY INHOMOGENEOUS MEDIA

A. V. Kalinin[a,b,*], A. A. Tyukhtina[a,**], S. A. Malov[a]

[a] *National Research Lobachevsky State University of Nizhny Novgorod, Gagarin Ave., 23, Nizhny Novgorod, 603022 Russia*
[b] *Institute of Applied Physics, Russian Academy of Sciences, Ulyanov St., 66, Nizhny Novgorod, 603950 Russia*

[*] *e-mail: avk@mm.unn.ru*
[**] *e-mail: tyukhtina@iee.unn.ru*

Received 23 October, 2023
Revised 28 December, 2023
Accepted 05 March, 2024

Abstract. Statements of initial-boundary value problems for the system of Maxwell equations in various quasi-stationary approximations in homogeneous and inhomogeneous conducting media are considered. In the case of weakly inhomogeneous media, asymptotic expansions of solutions of the initial-boundary value problems under consideration in a parameter characterizing the degree of inhomogeneity of the medium are formulated and substantiated. It is shown that the construction of an asymptotic expansion for a quasi-stationary electromagnetic approximation leads to a sequential solution of independent problems for a quasi-stationary electric and quasi-stationary magnetic approximation in a homogeneous medium. Conditions on the initial data are given for which the asymptotic series are convergent.

Keywords: Maxwell's system of equations, quasi-stationary electromagnetic approximation, conductivity, inhomogeneous media, asymptotic expansion.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 6, с. 1082–1094

УДК 519.635

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ КОНВЕКТИВНЫХ ТЕЧЕНИЙ В ТОНКОМ СЛОЕ ЖИДКОСТИ В УСЛОВИЯХ БОЛЬШИХ ЧИСЕЛ РЕЙНОЛЬДСА[1)]

© 2024 г. Е. В. Ласковец[1,*]

[1] *656049 Барнаул, пр-т Ленина, 61, Алтайский государственный университет, Институт математики и информационных технологий*

[*] *e-mail: katerezanova@mail.ru*

Поступила в редакцию 21.11.2023 г.
Переработанный вариант 23.02.2024 г.
Принята к публикации 05.03.2024 г.

Предложена математическая модель, описывающая течение тонкого слоя жидкости по наклонной, неравномерно нагретой подложке. В качестве определяющих уравнений используются система Навье–Стокса для вязкой несжимаемой жидкости и соотношения, представляющие собой обобщенные кинематическое, динамическое и энергетическое условия на границе раздела для случая испарения. Постановка приводится в двумерном случае для больших чисел Рейнольдса. Решение задачи осуществляется в рамках длинноволнового приближения. Проведен параметрический анализ задачи, получено эволюционное уравнение для нахождения толщины жидкого слоя. Предложен алгоритм численного решения для задачи о периодическом стекании жидкости по наклонной подложке. Изучено влияние гравитационных эффектов и характера нагрева твердой подложки на течение жидкого слоя. Библ. 24. Фиг. 4. Табл. 2.

Ключевые слова: термокапиллярное течение жидкости, обобщенные условия на границе раздела, испарение, эволюционное уравнение, численное решение.

DOI: 10.31857/S0044466924060156, **EDN:** XYAAYT

ВВЕДЕНИЕ

Потребность в теоретическом изучении задач, связанных с течением тонких слоев жидкостей, как правило, связана с их широкой применимостью в наукоемкой промышленности. Технологии, использующие в качестве рабочих сред испаряющиеся жидкости, встречаются, например, в системах термостабилизации и при нанесении покрытий.

Часто течения жидких пленок сопровождаются газовыми потоками, оказывающими влияние на характеристики течений. Изучению динамики тонких слоев жидкостей, сопровождаемых спутным потоком газа, посвящен ряд экспериментальных работ (см., например, [1, 2, 3]).

Одним из наиболее важных вопросов при изучении течений со свободными границами и границами раздела является формулировка граничных условий. Большое количество эффектов, влияющих на характер процессов, существенно затрудняет математическое моделирование подобных течений. В работах [4, 5, 6, 7, 8, 9] проводится математическое моделирование испаряющихся пленок. В работе [6] принимаются во внимание дополнительные силы на границе раздела, учитывающие перенос

[1)] Работа выполнена при финансовой поддержке проекта “Современные модели гидродинамики для задач природопользования, индустриальных систем и полярной механики” (2024-26) (гос. задание FZMW-2024-0003).

энергии на газожидкостной границе и эффективное давление. В статье [10] обсуждается влияние эффекта Марангони, гравитационного эффекта, степени неравновесности и динамики пара на неустойчивость течения пленки жидкости, сопровождаемой потоком газа. Подробный вывод условий на свободной границе на основе законов сохранения массы, импульса и энергии с учетом дополнительных гипотез проведен в [11]. Для получения граничных условий с учетом испарения в [12] использовались интегральные законы сохранения без предположения о неразрывности касательных скоростей и температуры. В [13] аналогичные условия выведены в предположении о диффузионном потоке пара на границе раздела.

Часто течения тонких слоев жидкостей моделируются с помощью уравнений Навье–Стокса [14, 15] или Обербека–Буссинеска [8]. В указанных работах проведен параметрический анализ задачи, что позволяет выявить эффекты, оказывающие наибольшее влияние на характер течения. В статьях [8, 14, 15] моделирование осуществлялось для случая умеренных чисел Рейнольдса порядка $O(1)$.

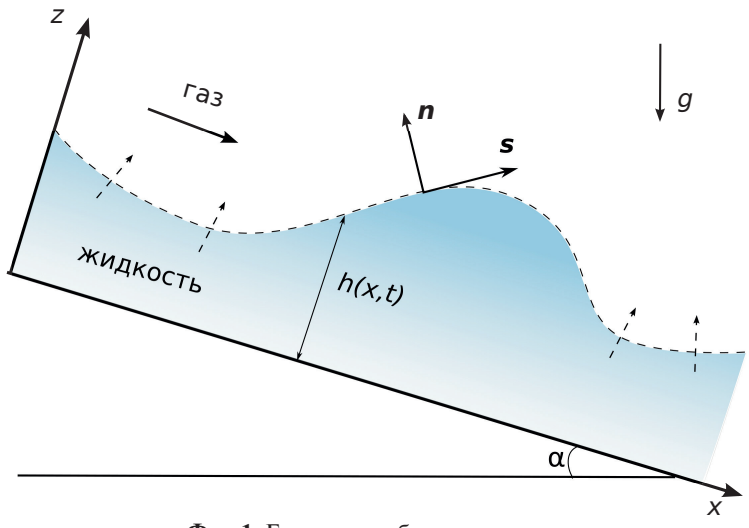
В настоящей работе предложена односторонняя математическая модель, описывающая течение тонкого слоя жидкости по наклонной неравномерно нагреваемой подложке. При моделировании приняты во внимание гравитационный, капиллярный и термокапиллярный эффекты, испарение и действие дополнительных касательных напряжений со стороны сопутствующего потока газа. Моделирование проводится на основе длинноволнового приближения системы уравнений Навье–Стокса и переноса тепла, кинематического, динамического и энергетического условий на границе раздела сред, обобщенных для случая ненулевого потока пара [13, 16–18]. Кинетическое уравнение Герца–Кнудсена используется для определения зависимости локального потока массы пара от температуры. На твердой непроницаемой подложке, подверженной неоднородному нагреву, выполняются условия прилипания. Для случая больших чисел Рейнольдса порядка $O(1/\varepsilon)$ построены точные решения для главных и первых членов разложения по степеням малого параметра, проведен параметрический анализ задачи [19]. Получено эволюционное уравнение, определяющее положение границы раздела сред, для главных членов разложения. Схема численного решения реализована для случая периодического стекания тонкого слоя жидкости. На примере системы “этанол – азот” показано, что гравитационные эффекты и характер нагрева подложки оказывают существенное влияние на структуру течения жидкости.

1. ПОСТАНОВКА ЗАДАЧИ О ТЕЧЕНИИ ТОНКОГО СЛОЯ ВЯЗКОЙ НЕСЖИМАЕМОЙ ЖИДКОСТИ С УЧЕТОМ ИСПАРЕНИЯ

Рассматривается тонкий слой вязкой несжимаемой жидкости, стекающий по наклонной, неравномерно нагретой твердой непроницаемой подложке в условиях спутного потока газа и испарения на термокапиллярной границе раздела сред. В данной постановке динамические процессы в газе не принимаются во внимание (т.е. строится односторонняя модель). Тем не менее, касательные напряжения, индуцируемые газовым потоком, могут учитываться на границе раздела.

Пусть $\mathbf{v} = (u, w)$ — вектор скорости жидкости, p — давление, T — температура, ν и χ — коэффициенты кинематической вязкости и температуропроводности, соответственно, ρ — плотность жидкости. Рассматривается движение жидкости по твердой подложке при наличии деформируемой границы раздела. Твердая непроницаемая подложка наклонена под углом α к линии горизонта. Система координат выбрана таким образом, что ось Ox направлена вдоль твердой границы, определяемой уравнением $z = 0$. Положение границы раздела задается с помощью соотношения $z = h(x, t)$ (см. фиг. 1). Тогда вектор силы тяжести $\mathbf{g}$ имеет вид $\mathbf{g} = (g_1, g_2) = (g\sin\alpha, -g\cos\alpha)$, $g = |\mathbf{g}|$.

В данной постановке характерная длина деформации свободной поверхности существенно превосходит амплитуду деформации. Таким образом, в задаче имеется два различных масштаба длины: l — продольная характерная длина, d — поперечная характерная длина, причем $l \gg d$. Пусть $\varepsilon = d/l$ — малый параметр системы. Отметим, что характерные продольная и поперечная скорости u_* и w_* также связаны между собой: $w_* = \varepsilon u_*$, а характерное время процесса t_* связано с другими параметрами задачи следующим образом: $l = u_* t_*$. Характерное давление может быть задано выражением $p_* = \rho u_*^2$.



Фиг. 1. Геометрия области течения.

Такой выбор характерных параметров позволяет моделировать течения при достаточно больших значениях чисел Рейнольдса.

В рассматриваемой задаче в качестве математической модели течения тонкого слоя жидкости используется система уравнений Навье–Стокса и уравнение переноса тепла. С учетом введенных обозначений в безразмерном виде уравнения записываются следующим образом:

$$\mathrm{Re}\varepsilon^2(u_t + uu_x + wu_z) - \varepsilon^2 u_{xx} = u_{zz} - \mathrm{Re}\varepsilon^2 p_x + \frac{\gamma}{\mathrm{Re}\varepsilon}\sin\alpha, \tag{1}$$

$$\mathrm{Re}\varepsilon^4(w_t + uw_x + ww_z) - \varepsilon^4 w_{xx} - \varepsilon^2 w_{zz} = -\mathrm{Re}\varepsilon^2 p_z - \frac{\gamma}{\mathrm{Re}}\cos\alpha, \tag{2}$$

$$u_x + w_z = 0, \tag{3}$$

$$\mathrm{Re}\,\mathrm{Pr}\varepsilon^2(T_t + uT_x + wT_z) - \varepsilon^2 T_{xx} = T_{zz}, \tag{4}$$

где $\mathrm{Re} = u_* l/\nu$ – число Рейнольдса, $\mathrm{Pr} = \nu/\chi$ – число Прандтля, $\gamma = gd^3/\nu^2$ – число Галилея.

Положение границы раздела сред в безразмерной постановке задачи также определяется в виде $z = h(x,t)$. Вектор нормали $\mathbf{n}$ и касательный вектор $\mathbf{s}$ к этой границе имеют координаты (n_1, n_2) и $(n_2, -n_1)$, соответственно. Здесь $n_1 = -\varepsilon h_x/\sqrt{1+\varepsilon^2 h_x^2}$, $n_2 = 1/\sqrt{1+\varepsilon^2 h_x^2}$. Кривизна свободной границы и скорость ее перемещения по направлению внешней нормали задаются соотношениями:

$$2H = \varepsilon h_{xx}/\sqrt{\left(1+\varepsilon^2 h_x^2\right)^3}, \quad D_n = -\varepsilon h_t/\sqrt{1+\varepsilon^2 h_x^2}.$$

Пусть характерная скорость в системе определяется следующим образом: $u_* = \nu/d$. Тогда получим, что число Рейнольдса имеет порядок $O(1/\varepsilon)$. Таким образом, дальнейшее моделирование осуществляется для случая больших чисел Рейнольдса.

На границе раздела $z = h(x,t)$ кинематическое, динамическое и энергетическое условия обобщены для случая ненулевого потока пара [13, 18, 19]. Представим кинематическое условие в виде:

$$-\varepsilon(h_t + h_x u - w)\frac{1}{\sqrt{1+\varepsilon^2 h_x^2}} = J_{ev}\bar{J}. \tag{5}$$

Здесь J_{ev} – величина локального потока массы пара на границе раздела. Данный параметр определяется с помощью уравнения Герца–Кнудсена (см. [5]). В безразмерной форме уравнение принимает вид

$$J_{ev} = \alpha_J T|_{z=h(x,t)}, \tag{6}$$

где коэффициент α_J определяется в виде:

$$\alpha_J = \alpha\rho_s\lambda_U\frac{T_*}{J_*}\left(\frac{M}{2\pi R_g T_s^3}\right)^{1/2}$$

(см. [5, 18]). Здесь α — коэффициент аккомодации, ρ_s — плотность пара, M — молекулярный вес, R_g — универсальная газовая постоянная, λ_U — скрытая теплота парообразования, T_s — температура насыщенного пара.

Проекции динамического условия в безразмерной форме имеют вид:

$$-p + \frac{2\varepsilon^2}{1+\varepsilon^2 h_x^2}\left[\varepsilon^2 h_x^2 u_x + w_z - h_x\left(u_z + \varepsilon^2 w_x\right)\right] = -p^g + \frac{\bar\rho\bar\nu\bar v}{\bar h}\frac{2\varepsilon}{1+\varepsilon^2 h_x^2}\left[\varepsilon^2 h_x^2 u_x^g + w_z^g - \varepsilon h_x\left(u_z^g + w_x^g\right)\right] + \\ + \left(1 - \frac{1}{\bar\rho}\right) J_{ev}^2 \bar J^2 + \sigma\frac{\varepsilon^2}{\mathrm{Ca}}\frac{h_{xx}}{\sqrt{(1+\varepsilon^2 h_x^2)^3}}, \tag{7}$$

$$\frac{2}{1+\varepsilon^2 h_x^2}\Big[-\varepsilon^2 h_x u_x + \varepsilon^2 h_x w_z - \frac{1}{2}\left(1-\varepsilon^2 h_x\right)\left(u_z + \varepsilon^2 w_x\right)\Big] - \frac{\bar\rho\bar\nu\bar v}{\bar h}\cdot\frac{2\varepsilon}{1+\varepsilon^2 h_x^2}\times \\ \times\Big[-\varepsilon h_x u_x^g + \varepsilon h_x w_z^g + \frac{1}{2}\left(1-\varepsilon^2 h_x^2\right)\left(u_z^g + w_x^g\right)\Big] = -\frac{\mathrm{Ma}}{\mathrm{Pr}}\frac{\varepsilon^2}{\sqrt{1+\varepsilon^2 h_x^2}}(T_x + h_x T_z). \tag{8}$$

Здесь введены следующие обозначения: $\bar v = u_*^g/u_*$ — отношение характерной продольной скорости газа к характерной скорости жидкости, $\bar\nu$, $\bar\rho$ — отношение коэффициентов кинематической вязкости и плотностей газа и жидкости, соответственно, $\bar h$ — отношение характерного размера слоя газа к l, p^g — давление в газе, $\mathrm{Ma} = \sigma_T T_* l/(\rho\nu\chi)$ — число Марангони, $\mathrm{Ca} = u_*\rho\nu/\sigma_0$ — капиллярное число. Параметр $\bar J$ определяется с помощью следующего соотношения: $\bar J = J_*^{ev}/(\rho u_*)$, характерная величина потока массы пара J_*^{ev} вычисляется как $\kappa T_*/(d\lambda_U)$, κ — коэффициент теплопроводности жидкости. Полагается, что коэффициент поверхностного натяжения σ линейно зависит от температуры. В безразмерной форме данная зависимость имеет вид $\sigma = 1 - \alpha_\sigma T$, где $\alpha_\sigma = \mathrm{Ma}\,\mathrm{Ca}\varepsilon/\mathrm{Pr}$, σ_0 — значение коэффициента поверхностного натяжения при некотором относительном значении температуры, σ_T — температурный коэффициент поверхностного натяжения, T_* — характерный перепад температуры.

Энергетическое условие в безразмерной форме представимо следующим образом:

$$\frac{\partial T}{\partial n} + \beta_2\left\{T\mathrm{div}_\Gamma v\right\} = \beta_3\bar J J_{ev} + \beta_4\bar J J_{ev}\Big\{-p + \frac{2\varepsilon^2}{1+\varepsilon^2 h_x^2}\left[\varepsilon^2 h_x^2 u_x + w_z - h_x\left(u_z + \varepsilon^2 w_x\right)\right]\Big\} + \\ + \frac{1}{2}\beta_5\bar J^3 J_{ev}^3 + \beta_6\bar J\varepsilon\sigma\frac{h_{xx}}{\sqrt{\left(1+\varepsilon^2 h_x^2\right)^3}}J_{ev}. \tag{9}$$

Здесь $\dfrac{\partial T}{\partial n}$ и $\mathrm{div}_\Gamma\mathbf{v}$ вычисляются согласно формулам

$$\frac{\partial T}{\partial n} = \frac{1}{\sqrt{1+\varepsilon^2 h_x^2}}\left(-\varepsilon^2 h_x T_x + T_z\right),$$

$$\mathrm{div}_\Gamma\mathbf{v} = \sum_{i=1}^{2}\frac{\partial v_i}{\partial x_i} - \sum_{i=1}^{2} n_i(\mathbf{n}\cdot\nabla v_i) = (u_x + w_z)|_\Gamma - \left\{\frac{\varepsilon^2 h_x^2}{1+\varepsilon^2 h_x^2}u_x - \frac{\varepsilon h_x}{1+\varepsilon^2 h_x^2}u_z - \frac{\varepsilon h_x}{1+\varepsilon^2 h_x^2}w_x + \frac{1}{1+\varepsilon^2 h_x^2}w_z\right\}.$$

Коэффициенты β_i $(i = 2, \ldots, 6)$ имеют следующий вид: $\beta_2 = \frac{\mathrm{Ma}\varepsilon^2}{\mathrm{Pr}E\overline{U}}$, $\beta_3 = \frac{1}{E}$, $\beta_4 = \frac{(1/\overline{\rho})-1}{E\overline{U}}$, $\beta_5 = \frac{\left(1-(1/\overline{\rho})\right)^2}{E\overline{U}}$, $\beta_6 = \frac{1-(1/\overline{\rho})\varepsilon}{\mathrm{Ca}E\overline{U}}$, $\overline{U} = \frac{\lambda_U}{u_*^2}$, $E = \frac{\kappa T_*}{\lambda_U \rho \nu}$ — параметр испарения [4]. Первое слагаемое в левой части условия (9) определяет дефект тепла при его переносе через границу раздела, второе отвечает за затраты энергии для преодоления деформации поверхности термокапиллярными силами вдоль этой поверхности. Правая часть соотношения (9) определяет затраты тепла на деформацию свободной поверхности в результате испарения и пропорциональна скорости потока испаряющейся массы. Первое слагаемое в правой части задает расход тепла на парообразование, второе — на деформацию границы, третье — на изменение кинетической энергии вещества при фазовом переходе, четвертое — на совершаемую веществом жидкости при массопереносе работу вследствие изменения удельного объема [13, 20].

На твердой подложке $z = 0$ полагаются выполненными условия прилипания

$$u|_{z=0} = 0, \quad w|_{z=0} = 0, \tag{10}$$

и задано распределение температуры

$$T|_{z=0} = \Theta_0(x, t). \tag{11}$$

2. ПОСТРОЕНИЕ УРАВНЕНИЯ, ОПРЕДЕЛЯЮЩЕГО ТОЛЩИНУ ЖИДКОГО СЛОЯ

В рамках сформулированной задачи (1)–(4), (5)–(11) компоненты скорости u, w, температура T, давление p и толщина жидкого слоя h определяются в длинноволновом приближении. Решение ищется в виде разложений по степеням малого параметра ε. Система уравнений (1)–(4), записанная для главных членов разложения, имеет следующий вид:

$$u^0_{zz} = -\gamma \sin\alpha, \tag{12}$$

$$p^0_z = -\gamma \cos\alpha, \tag{13}$$

$$u^0_x = -w^0_z, \tag{14}$$

$$T^0_{zz} = 0. \tag{15}$$

На твердой наклонной подложке должны быть выполнены следствия условий прилипания (10)

$$u^0|_{z=0} = 0, \quad w^0|_{z=0} = 0,$$

и задана функция, определяющая нагрев подложки (11)

$$T^0|_{z=0} = \Theta_0.$$

Для проведения параметрического анализа задачи определены значения безразмерных комплексов, порядок значений представлен в табл. 1, 2 для системы типа “этанол – азот”. При этом физические характеристики этанола и азота таковы: $\rho = 0.79$ г/см^3, $\nu = 0.015$ см^2/сек, $\kappa = 4 \cdot 10^{-4}$ кал/(сек см K), $\chi = 0.89 \cdot 10^{-3}$ см^2/сек, $\sigma_0 = 22$ дин/см, $\sigma_T = 0.08$ дин/(см K), $\lambda_U = 217$ кал/г, $\alpha = 0.01$, $M = 46$ г/моль, $\rho_s = 1.6 \cdot 10^{-3}$ г/см^3, $\rho_g = 1.2 \cdot 10^{-3}$ г/см^3, $\nu_g = 0.15$ см^2/сек, $\kappa_g = 0.65 \cdot 10^{-4}$ кал/(сек см K), $\chi_g = 0.3$ см^2/сек. Характерный перепад температуры T_* выбран 10 K (см. также [17, 18]).

С учетом проведенного параметрического анализа выпишем следствия условий (5)–(9) для главных членов разложений:

$$p^0 = p^g - \alpha_{Ca} h_{xx}(1 - \alpha_\sigma \Theta^0) + \alpha_D \alpha_J^2 (\Theta^0)^2,$$

$$u^0_z = \alpha_{Ma} \tilde{\Theta},$$

$$-T^0_z + \beta_2\{\Theta^0(u^0_x)\} = \overline{\beta}_3 J_0 - \bar{\beta}_4 p^0 J_0 + \bar{\beta}_6 h_{xx} J_0.$$

Таблица 1. Значения параметров α в системе “этанол – азот”

α – Параметр	Значения ($T_* = 10K$)
$\alpha_\sigma = \dfrac{\mathrm{MaCa}\varepsilon}{\mathrm{Pr}}$	$10^{-1}\varepsilon$
$\alpha_{Ca} = \dfrac{\varepsilon^2}{\mathrm{Ca}}$	$10^{4}\varepsilon^2$
$\alpha_D = \left(\dfrac{1}{\bar\rho} - 1\right)\bar J^2$	10^{-1}
$\alpha_\tau = \bar\rho\,\bar\nu\,\varepsilon$	$10^{-2}\varepsilon$
$\alpha_{Ma} = \dfrac{\varepsilon^2\mathrm{Ma}}{\mathrm{Pr}}$	$10^{3}\varepsilon^2$

Таблица 2. Значения параметров β в системе “этанол – азот”

β – Параметр	Значения ($T_* = 10\ K$)
$\beta_2 = \dfrac{\mathrm{Ma}\varepsilon^2}{\mathrm{Pr}E\bar U}$	$10^{5}\varepsilon^2$
$\bar\beta_3 = \beta_3\bar J$	1
$\bar\beta_4 = \beta_4 E$	10^{-1}
$\bar\beta_5 = \beta_5\bar J^3$	10^{-3}
$\bar\beta_6 = \varepsilon\beta_6 E$	$-\varepsilon^2 10^3$

Здесь $\Theta^0 = T^0|_\Gamma$, $\tilde\Theta = \left(T^0_x + h_x T^0_z\right)|_\Gamma$, $J_0 = \alpha_J\Theta^0$, а также $\alpha_{Ca} = \varepsilon^2/\mathrm{Ca}$, $\alpha_D = ((1/\bar\rho) - 1)\,\bar J^2$, $\alpha_\tau = \bar\rho\bar\nu\bar v\varepsilon/\bar h$, $\alpha_{Ma} = \varepsilon^2\mathrm{Ma}/\mathrm{Pr}$, $\bar\beta_3 = \beta_3\bar J$, $\bar\beta_4 = \beta_4 E$, $\bar\beta_5 = \beta_5\bar J^3$, $\bar\beta_6 = \varepsilon\beta_6 E$. Параметр α_J есть величина порядка 10.

Искомые функции u^0, w^0, p^0, T^0 определяются в ходе интегрирования уравнений (12)–(15):

$$u^0 = -\gamma\sin\alpha\frac{z^2}{2} + C_1 z, \tag{16}$$

$$w^0 = -(C_1)_x\frac{z^2}{2}, \tag{17}$$

$$p^0 = -\gamma\cos\alpha z + C_0(x,t), \tag{18}$$

$$T^0 = A(x,t)z + \Theta_0(x,t). \tag{19}$$

Вследствие граничных условий коэффициенты $C_0(x,t)$, $C_1(x,t)$, $A(x,t)$ удовлетворяют следующим соотношениям:

$$C_0(x,t) = p^g - \alpha_{Ca}h_{xx}\left(1 - \alpha_\sigma\left(Ah + \Theta_0\right)\right) + \gamma\cos\alpha h +$$
$$+\alpha_D\alpha_J^2\left(Ah + \Theta_0\right)^2,$$
$$C_1(x,t) = \alpha_{Ma}\left(A_x h + (\Theta_0)_x + h_x A\right) + \gamma\sin\alpha h,$$
$$A(x,t) = \frac{[-\beta_2(C_1)_x h + \alpha_J(\bar\beta_3 + \bar\beta_6 h_{xx})]\Theta_0}{\beta_2(C_1)_x h^2 + 1 - \alpha_J h(\bar\beta_3 + \bar\beta_6 h_{xx})}.$$

Аналогичным образом получена постановка задачи для первых членов разложения по степеням малого параметра ε. Искомые функции удовлетворяют следующей системе уравнений:

$$u^1_{zz} = p^0_x + u^0_t + u^0u^0_x + w^0u^0_z,$$

$$p_z^1 = w_{zz}^0, \quad w_z^1 = -u_x^1,$$
$$T_{zz}^1 = Pr(T_t^0 + u^0 T_x^0 + w^0 T_z^0).$$

На твердой границе $z = 0$ выполняются условия прилипания:

$$u^1|_{z=0} = 0, \quad w^1|_{z=0} = 0, \tag{20}$$

а из формул (11) и (15) следует соотношение

$$T^1|_{z=0} = 0. \tag{21}$$

На границе раздела сред $z = h(x,t)$ должны быть выполнены следующие условия:

$$p^1 = \alpha_{Ca}\alpha_\sigma h_{xx}\Theta^1 + \alpha_J^2\alpha_D\Theta^0\Theta^1, \tag{22}$$

$$u_z^1 = \alpha_{Ma}\tilde{\tilde{\Theta}}, \tag{23}$$

$$-T_z^1 + \beta_2\left\{\Theta^0\left[u_x^1 + h_x\left(u_z^0 + w_x^0\right)\right] + \Theta^1 u_x^0\right\} = 0. \tag{24}$$

Здесь $\Theta^1 = T^1|_\Gamma$, $\tilde{\tilde{\Theta}} = \left(T_x^1 + h_x T_z^1\right)|_\Gamma$.

Тогда, принимая во внимание условия на твердой границе (20), (21), получим аналитические решения для первых членов разложения неизвестных функций:

$$u^1 = (C_0)_x\frac{z^2}{2} + (C_1)_t\frac{z^3}{6} + (\bar{C}_3)z + (C_1)_x C_1\frac{z^4}{24}, \tag{25}$$

$$w^1 = -(C_0)_{xx}\frac{z^3}{6} - (C_1)_{tx}\frac{z^4}{24} - ((C_1)_x^2 + C_1(C_1)_{xx})\frac{z^5}{120} - (\bar{C}_3)_x\frac{z^2}{2}, \tag{26}$$

$$p^1 = -(C_1)_x z + \bar{C}_2(x,t), \tag{27}$$

$$\begin{aligned} T^1 = Pr\Bigg\{ &\left[A_t + C_1(\Theta_0)_x\right]\frac{z^3}{6} + (\Theta_0)_t\frac{z^2}{2} + \left[-\gamma\sin\alpha\frac{(\Theta_0)_x}{2} + C_1 A_x - \frac{A}{2}(C_1)_x\right]\times \\ &\times\frac{z^4}{12} - \gamma\sin\alpha A_x\frac{z^5}{40}\Bigg\} + \tilde{A}z. \end{aligned} \tag{28}$$

Функции $\bar{C}_2$, $\bar{C}_3$, $\tilde{A}$ удовлетворяют соотношениям:

$$\bar{C}_2(x,t) = (C_1)_x h + \Theta^1\left(\alpha_{Ca}\alpha_\sigma h_{xx} + +2\alpha_J^2\alpha_D(Ah + \Theta_0)\right),$$

$$\bar{C}_3(x,t) = -\alpha_{Ma}\tilde{\tilde{\Theta}} - (\bar{C}_0)_x h - (C_1)_t\frac{h^2}{2} - C_1(C_1)_x\frac{h^3}{6},$$

$$\tilde{A} = F(A, \Theta_0, C_1, h).$$

Функция $F(A, \Theta_0, C_1, h)$ определяется с помощью соотношения (24).

С учетом формул (16), (17), (25), (26) для выражения функций u и w можно получить уравнение для определения толщины слоя жидкости. Представим уравнение для определения положения границы раздела, ограничиваясь главными членами разложения (lubrication approximation):

$$h_t + h_x\left\{-\gamma\sin\alpha\frac{h^2}{2} + C_1 h + (C_1)_x\frac{h^2}{2}\right\} + EJ_{ev} = 0, \tag{29}$$

где

$$J_{ev} = \alpha_J\left[A(x,t)h + \Theta_0(x,t)\right].$$

Отметим также, что для замыкания постановки задачи необходимо определить начальное положение термокапиллярной границы раздела $h(x,0) = h_0(x)$, а также условия на бесконечности. Когда будет определена функция $h(x,t)$, распределение скоростей, давление и температура также будут найдены с учетом формул (16)–(19) и (25)–(28).

3. АЛГОРИТМ ЧИСЛЕННОГО РЕШЕНИЯ

Задача о периодическом стекании часто используется для тестирования задачи о стекании слоя жидкости [21]. Рассмотрим периодическую задачу об определении положения границы раздела h, удовлетворяющей уравнению (29) на некотором промежутке $[-L;L]$. Пусть выполнены следующие периодические условия на границах $x = \pm L$ рассматриваемой области:

$$h|_{x=-L} = h|_{x=L}, \qquad h_x|_{x=-L} = h_x|_{x=L}. \tag{30}$$

Начальное положение термокапиллярной границы имеет вид $h_0(x) = 1-\delta_0 \cos kx$. Неравномерный нагрев подложки определяется с помощью заданной функции Θ_0.

С учетом соотношения, определяющего функцию $C_1(x,t)$, уравнение (29) может быть записано в следующем виде:

$$h_t + A_2 h_{xx} + A_1 h_x + A_0 h + D = 0. \tag{31}$$

Здесь коэффициенты A_2, A_1, A_0, D представляют собой функции, зависящие от A, h, Θ_0 и их производных:

$$A_2 = \frac{h^2}{2}\alpha_{Ma}A, \qquad A_1 = 2h^2\alpha_{Ma}A_x + h\alpha_{Ma}(\Theta_0)_x + h_x h\alpha_{Ma}A + h^2\gamma\sin\alpha,$$

$$A_0 = \frac{h^2}{2}\alpha_{Ma}A_{xx} + \frac{h}{2}\alpha_{Ma}(\Theta_0)_{xx} + E\alpha_J A, \qquad D = E\alpha_J\Theta_0.$$

Для численного решения уравнения (31) используется неявная конечно-разностная схема следующего вида:

$$\frac{h^{k+1}-h^k}{\tau} + A_2^k h_{xx}^{k+1} + A_1^k h_x^{k+1} + A_0^k h^{k+1} + D^k = 0. \tag{32}$$

Для реализации неявной схемы введем равномерную разностную сетку по пространственной переменной x: $x_1, x_2, \ldots, x_{N+1}$, $x_n = -L + (n-1)\Delta x$, $n = 1,2,\ldots,N+1$, с шагом $\Delta x = 2L/N$. Для всех производных по x, входящих в уравнение (32) используются конечно-разностные аналоги второго порядка аппроксимации.

Конечно-разностную схему (32) можно записать в виде системы линейных алгебраических уравнений:

$$b_1^k h_1^{k+1} + c_1^k h_2^{k+1} = d_1^k, \quad n = 1;$$

$$a_n^k h_{n-1}^{k+1} + b_n^k h_n^{k+1} + c_n^k h_{n+1}^{k+1} = d_n^k, \quad n = 2,3,...,N;$$

$$a_{N+1}^k h_N^{k+1} + b_{N+1}^k h_{N+1}^{k+1} = d_{N+1}^k, \quad n = N+1. \tag{33}$$

Коэффициенты $a_n^k, b_n^k, c_n^k, d_n^k$ зависят от A_2, A_1, A_0, D. Для реализации периодических условий (30) на границах $x = \pm L$ также используются конечно-разностные аналоги второго порядка аппроксимации (см. [22]).

Таким образом, задача сводится к решению системы линейных алгебраических уравнений (33) с помощью метода трехточечной прогонки и прогонки с параметром. Поиск значений h_n осуществляется в виде $h_n = \alpha_n h_{n+1} + \beta_n h_N + \gamma_n$. Формулы для прогоночных коэффициентов α_n, β_n, γ_n имеют следующий вид:

$$\alpha_n = \frac{-c_n}{a_n\alpha_{n-1}+b_n}, \qquad \beta_n = \frac{-a_n\beta_{n-1}}{a_n\alpha_{n-1}+b_n}, \qquad \gamma_n = \frac{d_n - a_n\gamma_{n-1}}{a_n\alpha_{n-1}+b_n}, \qquad n = 2,3,...,N.$$

Стартовые значения коэффициентов α_1, β_1, γ_1 определяются с помощью первого уравнения системы (33) и граничного условия. В роли неизвестного параметра выступает значение толщины жидкого слоя на торцах исследуемой области h_{N+1}. В качестве обратного хода прогонки используется соотношение $h_n = \tilde{\alpha}_n h_{N+1} + \tilde{\beta}_n$.

Численное решение задачи о периодическом стекании тонкого слоя жидкости по наклонной подложке состоит из следующих этапов.

1. Расчет на новом временном слое $k+1$ начинается с вычисления значений функции A, при этом используются значения функции h, определяющей положение границы раздела, с предыдущего временного слоя. Здесь

$$(h_{xx})_n = \frac{h_{n+1} - 2h_n + h_{n-1}}{\Delta x^2}, \qquad (\Theta_0)_n = 1 + \delta\cos(kx_n), \qquad n = 2, ..., N.$$

Для аппроксимации вторых производных по переменной x на границах исследуемой области $[-L, L]$ используются следующие соотношения:

$$\frac{h_3 - 2h_2 + h_1}{\Delta x^2}, \qquad \frac{h_{N1} - 2h_N + h_{N-1}}{\Delta x^2}.$$

2. С помощью полученных для A_n значений насчитываются значения функции $(C_1)_n$ и ее производной.

3. Определяются значения толщины жидкости h на новом временном слое с помощью схемы Кранка–Николсона для переменной по времени.

Для замыкания постановки задачи необходимо задать условия на введенных торцах исследуемой области $x = \pm L$. В качестве таких условий в данном случае используются условия периодичности значений функции h и ее первой производной по координате x [24].

4. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ИССЛЕДОВАНИЯ

Проведено численное исследование для случая периодического стекания жидкости типа этанол. В качестве газа, сопровождающего течение, рассматривается азот. Значения безразмерных параметров задачи представлены в табл. 1, 2 (см. также [17, 18, 23, 24]). Характерное время процесса t_* равно $0.7 \cdot 10^2$ сек.

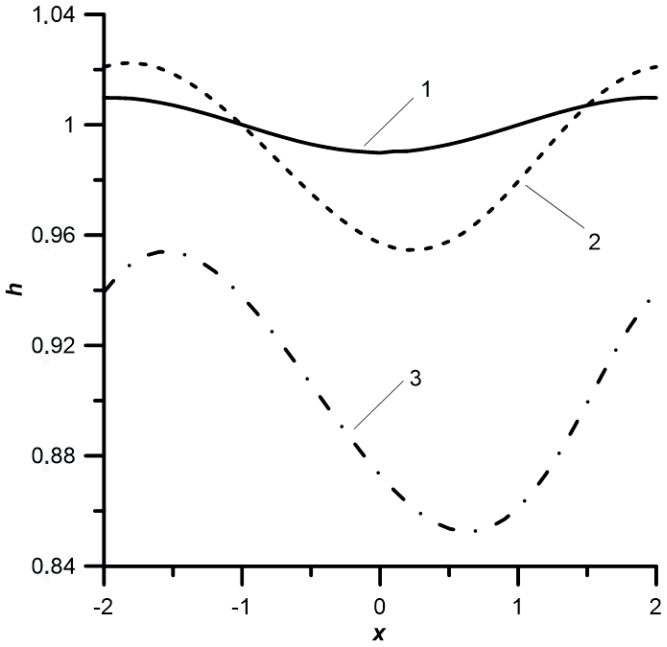
Положение границы раздела сред в начальный момент времени задано в виде следующего соотношения: $h_0 = 1 - \delta_0 \cos kx$. Неравномерный нагрев подложки определяется с помощью формулы

$$\Theta_0 = 1 + \delta_1 \cos k_1 x \cdot \cos k_2 t.$$

Значения параметров δ_0 и δ_1 здесь полагаются равными 0.01 и 0.25, соответственно; $k = k_1 = \pi/2$, $k_2 = 10$. Угол наклона твердой подложки относительно горизонта составляет $\pi/8$. Представленные результаты получены с использованием энергетического условия на границе $y = h(x, t)$ в классической постановке, т.е. безразмерные параметры $\bar{\beta}_3$ и $\bar{\beta}_6$ принимались равными 0.

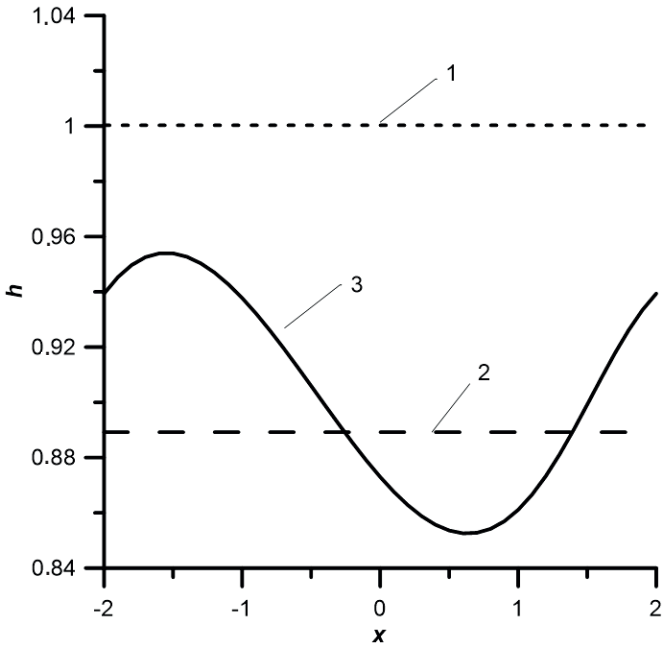
На фиг. 2 проиллюстрирован процесс испарения жидкого слоя со временем в условиях нормальной гравитации ($g = 9.8$ м/с^2). Твердая подложка подвержена неоднородному нестационарному нагреву. С течением времени наблюдается снижение толщины жидкого слоя относительно начального положения (сплошная линия). Амплитуда волны при этом существенно увеличивается, наблюдается смещение локального минимума. В случае продолжения рассматриваемого процесса во времени может наблюдаться образование сухих пятен.

Фигура 3 демонстрирует влияние характера нагрева наклонной подложки на структуру течения слоя жидкости. Приведены положения границы раздела сред в момент времени $t = 10^{-2}$ в условиях нормальной гравитации. Сплошная линия на фиг. 3 демонстрирует толщину жидкого слоя в случае неоднородного нестационарного нагрева (соответствует линии *3* на фиг. 2). В случае, когда рассматривается случай однородного распределения температуры относительно пространственной ко-



Фиг. 2. Изменение характера течения тонкого слоя жидкости со временем: *1* — начальное положение границы раздела; *2* — положение границы раздела в момент времени $t = 10^{-3}$; *3* — положение границы раздела в момент времени $t = 10^{-2}$.

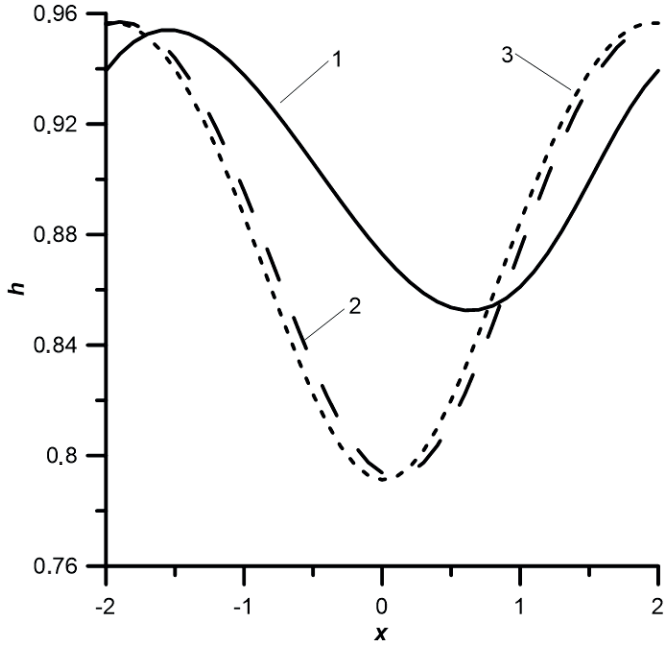
ординаты x (т.е. $k_1 = 0$), наблюдается выравнивание границы раздела сред (см. штриховые линии на фиг. 3). Линия *2* демонстрирует снижение толщины жидкого слоя со временем в условиях нестационарного нагрева подложки. В случае, когда функция Θ_0 постоянна как по пространству, так и по времени, наблюдается установление стационарной картины течения без изменения толщины жидкого слоя (линия *3* на фиг. 3).



Фиг. 3. Изменение течения тонкого слоя жидкости в зависимости от характера нагрева подложки, $t = 10^{-2}$: *1* — однородное стационарное распределение температуры; *2* — однородное нестационарное распределение температуры; *3* — неоднородное, нестационарное распределение температуры.

На фиг. 4 представлены результаты изучения влияния уровня гравитации на характер стекания тонкого слоя жидкости в случае неоднородного нестационарного нагрева подложки. Наиболее слабое снижение толщины жидкого слоя и наименьшая амплитуда волны наблюдаются в условиях нор-

мальной гравитации, когда $g = 9.8$ м/с^2 (сплошная линия *1*). При уменьшении уровня гравитации до $g = 9.8 \cdot 10^{-1}$ м/с^2 (линия *2*) и $g = 9.8 \cdot 10^{-2}$ м/с^2 (линия *3*) имеет место снижение толщины жидкого слоя. При этом смещение локального минимума положения границы раздела относительно начального положения увеличивается с ростом уровня гравитации.



Фиг. 4. Изменение течения тонкого слоя жидкости в зависимости от уровня гравитации, $t = 10^{-2}$: *1* – нормальная гравитация; *2* – слабая гравитация ($g \cdot 10^{-1}$); *3* – слабая гравитация ($g \cdot 10^{-2}$).

ЗАКЛЮЧЕНИЕ

Предложенная в работе математическая модель позволяет учитывать капиллярный, термокапиллярный, гравитационный эффекты, массоперенос на границе жидкости и газа, угол наклона и характер нагрева подложки, а также действие дополнительных касательных напряжений, возникающих под действием сопутствующего потока газа, на характер течения тонкого слоя жидкости. Параметрический анализ задачи позволяет определить эффекты, оказывающие наибольшее влияние на изучаемые процессы. В рамках рассматриваемой математической модели построены точные (аналитические) решения для главных и первых членов разложения искомых функций по степеням малого параметра задачи.

Для решения полученного эволюционного уравнения, определяющего толщину жидкого слоя, построена численная схема. В рамках задачи о периодическом стекании тонкого слоя жидкости изучено влияние гравитационного эффекта на структуру течения. Проведено численное исследование процесса стекания жидкости с различной картиной распределения температуры по твердой подложке. Показано, что наиболее интенсивное уменьшение толщины жидкого слоя наблюдается с увеличением уровня гравитации для случаев неоднородного нагрева подложки.

СПИСОК ЛИТЕРАТУРЫ

1. *Kabov O. A., Zaitsev D. V., Cheverda V. V.* Evaporation and flow dynamics of thin, shear-driven liquid films // Experim. Thermal Fluid Sci. 2011. V. 35. № 5. P. 825–831.
2. *Реутов В. П., Езерский А. Б., Рыбушкина Г. В., Чернов В. В.* Конвективные структуры в тонком слое испаряющейся жидкости, обдуваемом воздушным потоком // ПМТФ. 2007. Т. 48. № 4. С. 3–14.
3. *Frank M. A., Kabov O. A.* Thermocapillary structure formation in a falling film: Experiment and calculations // Phys. of Fluids. 2006. № 18. P. 032107-1.

4. *Oron A., Davis S. H., Bankoff S. G.* Long-scale evolution of thin liquid films // Rev. of Modern Phys. 1997. Vol. 69. № 3. P. 931–980.
5. *Miladinova S., Slavtchev S., Lebon G., Legros J.-C.* Long-wave instabilities of non-uniformly heated falling films // J. Fluid Mech. 2002. Vol. 453. P. 153–175.
6. *Shklyaev O., Fried E.* Stability of an evaporating thin liquid film // J. of Fluid Mech. 2007. Vol. 584. P. 157–183.
7. *Кабов О. А., Кабова Ю. О., Кузнецов В. В.* Испарение неизотермической пленки жидкости в микроканале при спутном потоке газа // ДАН. 2012. Vol. 446. № 5. С. 522–526.
8. *Гончарова О. Н., Резанова Е. В., Тарасов Я. А.* Математическое моделирование термокапиллярных течений в тонком слое жидкости с учетом испарения // Известия АлтГУ. 2014. № 81 (1/1). С. 47–52.
9. *Kuznetsov V. V., Fominykh E. Yu.* Evaporation of a liquid film in a microchannel under the action of a co-current dry gas flow // Microgravity Science and Technology. 2020. V. 32. P. 245–258.
10. *Liu R., Liu Q.* The Convective Instabilities in a Liquid–Vapor System with a Non-equilibrium Evaporation Interface // Microgravity Science and Technology. 2009. V. 21. P. 233–240.
11. *Андреев В. К., Гапоненко Ю. А., Гончарова О. Н., Пухначев В. В.* Современные математические модели конвекции. М.: Наука, 2008. 368 с.
12. *Das K. S., Ward C. A.* Surface thermal capacity and its effects on the boundary conditions at fluid-fluid interface // Phys. Rev. E. 2007. V. 75. P. 065303-1–065303-4.
13. *Кузнецов В. В.* Тепломассообмен на поверхности раздела жидкость – пар // Известия РАН. МЖГ. 2011. № 5. С. 97–107.
14. *Laskovets E. V.* Numerical modeling of an inclined thin liquid layer flow based on generalized boundary conditions // J. of Mathematical Science. 2022. Vol. 267. № 4. P. 501–510.
15. *Laskovets E. V.* Mathematical Modeling of the Thin Liquid Layer Runoff Process Based on Generalized Conditions at the Interface: Parametric Analysis and Numerical Solution // J. of SFU. Math. and Phys. 2023. Vol. 16, № 1. P. 56–65.
16. *Iorio C. S., Goncharova O. N., Kabov O. A.* Study of evaporative convection in an open cavity under shear stress flow // Microgravity Sci. Technol. 2009. V. 21. № 1. P. 313–320.
17. *Iorio C. S., Goncharova O. N., Kabov O. A.* Heat and mass transfer control by evaporative thermal pattering of thin liquid layers // Comput. Thermal Sci. 2011. № 3(4). P. 333–342.
18. *Гончарова О. Н.* Моделирование течений в условиях тепло- и массопереноса на границе // Известия АлтГУ. 2012. № 73 (1/2). С. 12–18.
19. *Гончарова О. Н., Резанова Е. В.* Математическая модель течений тонкого слоя жидкости с учетом испарения на термокапиллярной границе раздела // Известия АлтГУ. 2014. № 81 (1/2). С. 21–25.
20. *Бекежанова В. Б., Гончарова О. Н.* Задачи испарительной конвекции (обзор) // ПММ. 2018. Т. 82. Вып. 2. С. 219–260.
21. *Копбосынов Б. К., Пухначев В. В.* Термокапиллярное движение в тонком слое жидкости // Сб. научн. тр. Гидромеханика и процессы переноса в невесомости. АН СССР, Ур. научн. центр. 1983. С. 116–125.
22. *Самарский А. А.* Методы решения сеточных уравнений. М.: Наука, 1978. 592 с.
23. Краткий справочник физико-химических величин / *Под ред. Равделя А. А., Пономаревой А. М.* СПб.: Специальная литература, 1998. 232 с.
24. *Резанова Е. В.* Моделирование конвективных течений с учетом тепломассопереноса на границах раздела. Дис. … канд. физ.-матем. наук. Барнаул: АлтГУ, 2018.

NUMERICAL SIMULATION OF CONVECTIVE FLOWS IN A THIN LIQUID LAYER UNDER CONDITIONS OF LARGE REYNOLDS NUMBERS

E. V. Laskovets*

Altai State University, Institute of Mathematics and Information Technology, Lenin Ave., 61, Barnaul, 656049 Russia

**e-mail: katerezanova@mail.ru*

Received 21 November, 2023
Revised 23 February, 2024
Accepted 05 March, 2024

Abstract. A mathematical model is proposed that describes the flow of a thin layer of liquid on an inclined, non-uniformly heated substrate. The Navier-Stokes system for a viscous incompressible liquid and relations representing generalized kinematic, dynamic and energy conditions at the interface for the case of evaporation are used as governing equations. The statement is given in a two-dimensional case for large Reynolds numbers. The problem is solved within the framework of the long-wave approximation. A parametric analysis of the problem is carried out, an evolutionary equation is obtained for finding the thickness of the liquid layer. An algorithm for a numerical solution is proposed for the problem of periodic flow of liquid down an inclined substrate. The influence of gravitational effects and the nature of heating of a solid substrate on the flow of a liquid layer is studied.

Keywords: thermocapillary flow of liquid, generalized conditions at the interface, evaporation, evolution equation, numerical solution.

Свидетельство о регистрации средства массовой информации № 0110141 от 4 февраля 1993 г.,
выдано Министерством печати и информации Российской Федерации

Подписано к печати 06.12.2024. Дата выхода в свет 20.12.2024. Формат 60 x 88 $^{1}/_{8}$.
Усл. печ. л. 25,3. Уч.-изд. л. 25,3. Тираж 72 экз. Заказ 1604. Цена свободная.

Учредители: Российская академия наук, Федеральный исследовательский центр
«Информатика и управление» Российской академии наук

Издатель: Российская академия наук, 119991 Москва, Ленинский просп., 14
Исполнитель по контракту № 4У-ЕП-037-24 ФГБУ «Издательство «Наука»
121099, г. Москва, Шубинский пер., д. 6, стр. 1.
Отпечатано в ФГБУ «Издательство «Наука»
121099, г. Москва, Шубинский пер., д. 6, стр. 1

16+

Индекс 39324

◆◆◆◆ РОССИЙСКАЯ АКАДЕМИЯ НАУК ◆◆◆◆

Журналы РАН, выходящие в свет на русском языке

Автоматика и телемеханика
Агрохимия
Азия и Африка сегодня
Акустический журнал
Астрономический вестник. Исследования Солнечной системы
Астрономический журнал
Биологические мембраны
Биология внутренних вод
Биология моря
Биоорганическая химия
Биофизика
Биохимия
Ботанический журнал
Вестник Дальневосточного отделения Российской академии наук
Вестник древней истории
Вестник Российской академии наук
Вестник российской сельскохозяйственной науки
Водные ресурсы
Вопросы истории естествознания и техники
Вопросы ихтиологии
Вопросы языкознания
Вулканология и сейсмология
Высокомолекулярные соединения. Серия А
Высокомолекулярные соединения. Серия Б
Высокомолекулярные соединения. Серия С
Генетика
Геология рудных месторождений
Геомагнетизм и аэрономия
Геоморфология и палеогеография
Геотектоника
Геохимия
Геоэкология. Инженерная геология. Гидрогеология. Геокриология
Государство и право
Дефектоскопия
Дифференциальные уравнения
Доклады Российской академии наук. Математика, информатика, процессы управления
Доклады Российской академии наук. Науки о жизни
Доклады Российской академии наук. Науки о Земле
Доклады Российской академии наук. Физика, технические науки
Доклады Российской академии наук. Химия, науки о материалах
Журнал аналитической химии
Журнал высшей нервной деятельности им. И.П. Павлова
Журнал вычислительной математики и математической физики
Журнал неорганической химии
Журнал общей биологии
Журнал общей химии
Журнал органической химии
Журнал прикладной химии
Журнал физической химии
Журнал эволюционной биохимии и физиологии
Журнал экспериментальной и теоретической физики
Записки Российского минералогического общества
Зоологический журнал
Известия Российской академии наук. Механика жидкости и газа
Известия Российской академии наук. Механика твердого тела
Известия Российской академии наук. Серия биологическая
Известия Российской академии наук. Серия географическая
Известия Российской академии наук. Серия литературы и языка
Известия Российской академии наук. Серия физическая
Известия Российской академии наук. Теория и системы управления
Известия Российской академии наук. Физика атмосферы и океана
Известия Российской академии наук. Энергетика
Известия Русского географического общества
Исследование Земли из космоса
Кинетика и катализ
Коллоидный журнал
Координационная химия
Космические исследования
Кристаллография
Латинская Америка
Лёд и Снег
Лесоведение
Литология и полезные ископаемые
Мембраны и мембранные технологии
Металлы
Микология и фитопатология
Микробиология
Микроэлектроника
Молекулярная биология
Нейрохимия
Неорганические материалы
Нефтехимия
Новая и новейшая история
Общественные науки и современность
Общество и экономика
Океанология
Онтогенез
Палеонтологический журнал
Паразитология
Петрология
Письма в Астрономический журнал
Письма в Журнал экспериментальной и теоретической физики
Поверхность. Рентгеновские, синхротронные и нейтронные исследования
Почвоведение
Приборы и техника эксперимента
Прикладная биохимия и микробиология
Прикладная математика и механика
Проблемы Дальнего Востока
Проблемы машиностроения и надежности машин
Проблемы передачи информации
Программирование
Психологический журнал
Радиационная биология. Радиоэкология
Радиотехника и электроника
Радиохимия
Расплавы
Растительные ресурсы
Российская археология
Российская история
Российская сельскохозяйственная наука
Российский физиологический журнал им. И.М. Сеченова
Русская литература
Русская речь
Сенсорные системы
Славяноведение
Современная Европа
Социологические исследования
Стратиграфия. Геологическая корреляция
США & Канада: экономика, политика, культура
Теоретические основы химической технологии
Теплофизика высоких температур
Успехи современной биологии
Успехи физиологических наук
Физика Земли
Физика и химия стекла
Физика металлов и металловедение
Физика плазмы
Физикохимия поверхности и защита материалов
Физиология растений
Физиология человека
Химическая физика
Химия высоких энергий
Химия твердого топлива
Цитология
Человек
Экология
Экономика и математические методы
Электрохимия
Энтомологическое обозрение
Этнографическое обозрение
Ядерная физика

ISSN 0044-4669 Журнал вычислительной математики и математической физики, 2024, том 64, № 6