СОДЕРЖАНИЕ

Том 62, номер 9, 2022 год

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

О числе решений диофантова уравнения и проблеме Фробениуса

Э. Н. Гордеев, В. К. Леонтьев 1447

Применение мозаично-скелетонных аппроксимаций матриц
в методе физической оптики для задач электромагнитного рассеяния

А. В. Сетуха, С. Л. Ставцев, Р. М. Третьякова 1458

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

Оптимальное управление инвестициями в коллективной модели
пенсионного страхования: исследование сингулярных нелинейных задач
для интегродифференциальных уравнений

Т. А. Белкина, Н. Б. Конюхова, С. В. Курочкин 1473

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Трехмерные стационарные сферически симметричные модели звездной динамики,
зависящие от локальной энергии

Ю. Батт, Э. Йорн, А. Л. Скубачевский 1491

Метод континуальных теорем сложения и интегральные соотношения
между функциями Кулона и функцией Аппеля F_1

И. А. Шилин, Дж. Чой 1522

МАТЕМАТИЧЕСКАЯ ФИЗИКА

Тестирование бикомпактных схем для одномерных уравнений Максвелла
в слоистых средах

А. А. Белов, Ж. О. Домбровская 1532

Массовый поток и распределение давления газа в длинном концентрическом
кольцевом канале при неполной аккомодации молекул газа

О. В. Гермидер, В. Н. Попов 1551

A Fictitious Time Lie-Group Integrator for the Brinkman–Forchheimer Momentum
Equation Modeling Flow of Fully Developed Forced Convection

S. Abbasbandy, M. Hajiketabi 1563

ИНФОРМАТИКА

Регулярные аппроксимации наибыстрейшего движения мобильного робота
при ограниченных фазовых координатах

*А. Н. Дарьина, А. И. Дивеев, Д. Ю. Карамзин, Ф. Л. Перейра,
Е. А. Софронова, Р. А. Чертовских* 1564

EDN: OGHMLF
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1447–1457

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.16

О ЧИСЛЕ РЕШЕНИЙ ДИОФАНТОВА УРАВНЕНИЯ И ПРОБЛЕМЕ ФРОБЕНИУСА[1)]

© 2022 г. Э. Н. Гордеев[1,*], В. К. Леонтьев[2,]**

[1] *105005 Москва, 2-я Бауманская ул., 5, стр. 1, МГТУ им. Н.Э. Баумана, Россия*
[2] *119133 Москва, ул. Вавилова, 40, ВЦ РАН ФИЦ ИУ РАН, Россия*
**e-mail: werhorn@yandex.ru*
***e-mail: vkleontiev@yandex.ru*

Поступила в редакцию 10.09.2021 г.
Переработанный вариант 28.02.2022 г.
Принята к публикации 11.04.2022 г.

Рассматриваются вопросы, касающиеся разрешимости и числа решений линейного диофантова уравнения. Наряду с общим случаем внимание уделяется комбинаторным характеристикам числа решений и среднего числа решений уравнений специального вида. Один тип уравнения представляет разбиения натурального числа на натуральные слагаемые. Другой тип – это линейные уравнения с двумя переменными, обычно исследуемые в связи с проблемой Фробениуса. Основное внимание уделено трем аспектам. Первый касается исследования наличия и числа решений диофантова уравнения при параметризации задачи по правым частям. Даются формулы и оценки для подсчета этого числа как в общем, так и в частных случаях. Второй аспект посвящен задаче о разбиении. Третий касается известной проблемы Фробениуса. Библ. 31.

Ключевые слова: диофантово уравнение, разбиения, проблема Фробениуса, булевы уравнения, число Фробениуса.

DOI: 10.31857/S0044466922090046

1. ВВЕДЕНИЕ

Линейные диофантовы уравнения и неравенства являются стандартным объектом для различного рода математических моделей, относящихся к целочисленной оптимизации, защите информации, теории чисел, геометрии и т.д.

Это уравнение имеет вид

$$\sum_{i=1}^{n} a_i x_i = b, \tag{1}$$

где компоненты n-мерного вектора $\mathbf{x} = (x_1, \ldots, x_n)$, а также коэффициенты $b, a_1, \ldots, a_n$ – неотрицательные целые числа.

Вопросы разрешимости этого уравнения, нахождения решения и числа всех решений этого уравнения – частные случаи известной в области исследования операций и комбинаторной оптимизации задачи целочисленного линейного программирования. Эта задача или ее обобщения и сужения занимают ключевое место среди задач дискретной оптимизации, как это показано, например, в [1]. То, что задача о рюкзаке и задача целочисленного линейного программирования в общем случае являются NP-полными, было установлено в числе первых результатов подобных исследований (см., например, [1], [2]). Кроме того, в классической публикации [2] можно найти многочисленные примеры известных задач, которые к ним сводятся, и, наоборот, задача целочисленного линейного программирования (или ее булев вариант) сводится к той или иной прикладной проблеме.

С другой стороны, дифантово уравнение и его частные случаи – предмет исследования в таких областях, как алгебра, теория чисел, криптография и др. Примерами здесь могут служить работы [3]–[6].

[1)]Работа выполнена при финансовой поддержке РФФИ (код проекта 20-01-00645).

Данная статья в одной из своих частей является продолжением исследований авторов, опубликованных в [7], [8], где речь шла про задачу о рюкзаке. Ключевую роль в получении результатов там сыграл аппарат производящих функций (см. [9]), который используется и в настоящей статье. Как видно, в частности, из подробной монографии [10], где приводится обширный обзор результатов, связанных с задачей о рюкзаке, данный подход позволил в [7] и [8] получить ряд новых свойств и соотношений, касающихся числа решений этой задачи и множества ее решений.

С прикладной точки зрения, изучаемая проблематика затрагивалась авторами данной статьи в [11] и [12] в связи с криптографическими объектами: аннигиляторами и алгебраической иммунностью. Одно из ключевых утверждений публикации [11], посвященной аннигиляторам и алгебраической иммунности, базируется на анализе совместимости системы уравнений и сводится к нахождению комбинаторной характеристики (аналога ранга) матрицы.

Исследования в той же области, но другими методами проводились, например, в [13] и [14].

В данной статье уделено внимание линейному диофантову уравнению специального вида, связанного с проблемой Фробениуса.

Пусть $A = \{a_1, ..., a_k\}$ — возрастающая последовательность натуральных чисел, $k > 1$, $\langle A \rangle$ — аддитивная полугруппа, порожденная множеством A. Полугруппа $\langle A \rangle$ состоит из всех линейных комбинаций чисел $a_1, ..., a_k$ с целыми неотрицательными коэффициентами.

Множество A называется примитивным, если НОД$(a_1, ..., a_k) = 1$.

Для случая частного случая $k = 2$ известен следующий результат (см. [15], [16]).

Теорема Сильвестра. *Порожденная взаимно простыми числами a и b полугруппа содержит все целые числа, начиная с* $N(a, b) = (a - 1)(b - 1)$.

Добавляя образующие, из этой теоремы легко вывести общий результат для любого k.

Теорема 1. *Если множество A примитивное, то найдется* $N(a_1, ..., a_k)$ *такое, что* $t \in \langle A \rangle$ *при любом натуральном* $t \geq N(a_1, ..., a_k)$.

В большинстве публикаций на эту тему именно $N(a_1, ..., a_k)$ называется числом Фробениуса (см., например, [17]). Однако заметим, что есть и разночтения в терминологии. В некоторых источниках, например в [3], числом Фробениуса называют величину $N(a_1, ..., a_k) - 1$, т.е. максимальное $t \notin \langle A \rangle$.

Мы в нашем тексте будем придерживаться первого варианта.

Пусть N – множество всех натуральных чисел. Обозначим через C(A) множество всех чисел t таких, что $t \notin N / \langle A \rangle$.

Задача нахождения числа $N(a_1, ..., a_k)$ известна как диофантова проблема Фробениуса (ПФ). Нахождение же всего множества C(A) обычно называется расширенной проблемой Фробениуса (РПФ).

ПФ и РПФ – популярная тематика исследований алгебраистов, специалистов в теории чисел, криптографов, а в последние десятилетия она привлекает внимание специалистов в области защиты информации (см., например, [3]–[5]).

В [5] опубликован достаточно подробный обзор основных результатов по этим проблемам, полученных до 2005 г.

Как было уже сказано выше, ПФ и РПФ для $k = 2$ были решены еще в 1884 г. (см. [15], [16]). Для произвольного случая изучались асимптотика и оценки числа Фробениуса, например, в [18], [19].

Алгоритм решения РПФ при $k = 3$ получен в [20], оценена сложность алгоритма.

Формула решения ПФ при $k > 2$ не была получена, уже при $k = 3$ в [21] доказаны утверждения, объясняющие принципиальные затруднения, связанные с этой проблемой.

Точные формулы имеются лишь для частных случаев. Различные частные случаи для $k = 3$ изучаются во многих работах, например, в статьях [22] и [23]. В [22] наряду с собственными результатами дается и обзор некоторых аспектов состояния проблемы на 2017 г.

Существуют и специфические постановки, которые выглядят как обобщение ПФ. В качестве примера можно привести [24].

С алгебраической точки зрения можно наложить определенные ограничения на полугруппу $\langle A \rangle$ и решать ПФ для полученного частного случая, как это делается в статьях [25]–[27].

Проблема исследовалась и с алгоритмической точки зрения. Например, в [28] представлен теоретико-графовый алгоритм определения $N(a_1, ..., a_k)$ со сложностью $O(a_1(k + \log a_1))$. Здесь

ПФ сведена к поиску определенного вида наибольшего кратчайшего пути в орграфе с a_1 вершинами и с ka_1 дугами, где из каждой вершины исходит k дуг весов $a_1, \ldots, a_k$ соответственно.

В [29] для РПФ и ПФ предложена редукция множества А к собственному подмножеству, снижающая сложность задачи в ряде случаев. Алгоритмы не дают аналитической формулы для числа Фробениуса.

Верхние оценки для числа Фробениуса тоже представляют прикладной интерес, как это, например, указано в учебнике по криптографическим методам защиты информации [3]. В [3], [4] и [5] приведены примеры результатов на эту тему.

Настоящая работа является продолжением исследований о разрешимости и нахождению числа решений линейных дифантовых уравнений, систем таких уравнений, неравенств и систем неравенств.

Данная статья в одной из своих частей является продолжением исследований авторов, опубликованных в [7], [8], где речь шла про задачу о рюкзаке, и непосредственным продолжением работы [30].

Статья состоит из введения, трех разделов и заключения. В разд. 2 исследуются линейное диофантово уравнение общего вида и один частный случай. Разд. 3 посвящен проблеме разбиения. В разд. 4 исследуются задачи, связанные с двухмерной проблемой Фробениуса.

2. О ЧИСЛЕ РЕШЕНИЙ ДИОФАНТОВА БУЛЕВА УРАВНЕНИЯ

Обозначим через $t_b(a_1,\ldots,a_n)$ число решений уравнения (1). Ясно, например, что положительность этого числа влечет за собой разрешимость уравнения.

Если на область определения переменных наложены ограничения, то, чем шире область определения, тем “вероятнее” разрешимость уравнения (1).

Заметим, что нахождение числа решений уравнения (1) тесно связано с известной задачей о разбиениях, т.е. с нахождением числа решений уравнения:

$$x_1 + x_2 + \ldots + x_k = n.$$

Если на $x_1, x_2, \ldots, x_k$ нет никаких ограничений, то число решений уравнения – это число разбиений n на натуральные слагаемые. Этот факт формально может быть выражен в терминах преобразования вектора $\mathbf{x} = (x_1, \ldots, x_k)$ в вектор $\mathbf{y} = (y_1, \ldots, y_n)$, где y_r, $r = 1, \ldots, n$, – это число координат вектора $\mathbf{x}$, равных натуральному r.

Таким образом, нахождения числа разбиений P_n для фиксированного n эквивалентно нахождению числа решений линейного диофантова уравнения

$$n = \sum_{i=1}^{k} x_i = \sum_{r=1}^{n} r y_r.$$

Пусть $L(x_1, \ldots, x_n)$ – линейная форма:

$$L(x_1,\ldots,x_n) = \sum_{i=1}^{n} a_i x_i. \tag{2}$$

Для случая булевых переменных через $L^*(x_1, \ldots, x_n)$ обозначим множество значений этой формы. Тогда вопрос о разрешимости уравнения (1) эквивалентен вопросу о принадлежности числа b этому множеству.

Рассмотрим производящую функцию последовательности $\{t_b(a_1,\ldots,a_n)\}$:

$$F_{a_1,\ldots,a_n}(z) = \sum_{b=0}^{\infty} z^b t_b(a_1,\ldots,a_n). \tag{3}$$

Для нее известно соотношение

$$F_{a_1,\ldots,a_n}(z) = \sum_{b=0}^{\infty} z^b t_b(a_1,\ldots,a_n) = \prod_{k=1}^{n} \frac{1}{1 - z^{a_k}}. \tag{4}$$

Оно является простым следствием определения, представленного соотношением (3). Действительно, из (3) следует

$$F_{a_1,\ldots,a_n}(z) = \sum_{b=0}^{\infty} z^b t_b(a_1,\ldots,a_n) = \sum_{\{x_1,\ldots,x_n\}} z^{a_1x_1+\ldots+a_nx_n} = \prod_{i=1}^{n}\sum_{x_i=0}^{\infty} z^{a_ix_i} = \prod_{k=1}^{n}\frac{1}{1-z^{a_k}}.$$

Утверждение 1. *Справедливо соотношение*

$$t_b(a_1,\ldots,a_n) = \frac{1}{2\pi i}\oint_{|u|=\rho}\frac{F_{a_1,\ldots,a_n}(z)}{z^{b+1}}dz, \quad \rho < 1.$$

Очевидно, что эти формулы позволяют найти $t_b(a_1,\ldots,a_n)$ путем сравнения коэффициентов в левой и правой части. Кроме того, с их помощью можно найти оценки для $t_b(a_1,\ldots,a_n)$ и некоторые характеристики этой величины.

Пример 1. Если все $a_i = 1$, $i = 1, \ldots, n$, то $F_{1,\ldots,1}(z) = (1-z)^{-n}$ и $t_b(1,\ldots,1) = (-1)^b C_b^{-n} = C_b^{n+b-1} = C_{n-1}^{n+b-1}$.

Это хорошо известная формула для числа разбиений натурального b в сумму не более n слагаемых.

Пример 2. Если $n = 2$, то поведение функции $t_b(a_1,a_2)$ в значительной мере определяется упомянутой выше теоремой Сильвестра (см., например, [14], [15]).

Согласно ей при условии взаимной простоты $(a_1,a_2) = 1$ уравнение разрешимо, если $b > a_1a_2$, и эта граница достижима. Таким образом, если $t_b(a_1,\ldots,a_n) > 0$ для всех $b \geq b_0$ и $t_b(a_1,\ldots,a_n) = 0$ для $b = b_0 - 1$, то значение числа Фробениуса равно b_0.

Для случая булевых переменных формула (4) имеет вид

$$F_{a_1,\ldots,a_n}(z) = \sum_{b=0}^{\infty} z^b t_b(a_1,\ldots,a_n) = \prod_{k=1}^{n}(1+z^{a_k}). \tag{5}$$

Из (5) следует выражение для $t_b\,(a_1,\ldots,a_n)$:

$$t_b(a_1,\ldots,a_n) = \frac{1}{2\pi i}\oint_{|u|=\rho}\frac{(1+u^{a_1})\ldots(1+u^{a_n})}{u^{b+1}}du, \quad \rho < 1.$$

Пусть теперь $x_1 \in A_1,\ldots,\ x_n \in A_n$, где $A_1,\ldots,A_n$ – произвольные подмножества натурального ряда.

Тогда производящая функция $t_b\,(A_1,\ldots,A_n)$ для числа решений уравнения (1) с такой структурой множества аргументов имеет вид:

$$F_{A_1,\ldots,A_n}(z) = \sum_{b=0}^{\infty} z^b t_b\,(A_1,\ldots,A_n). \tag{6}$$

И она может быть представлена следующим образом.

Пусть

$$F_k(z) = \sum_{x\in A_k} z^x \tag{7}$$

есть производящая функция для множества A_k.

Непосредственно из (6) и (7) получаем

Утверждение 2. *Справедлива формула*

$$F_{A_1,\ldots,A_n}(z) = F_1(z)F_2(z)\ldots F_n(z).$$

Пример 3. Пусть $A_k = \{0,1,2\}, k = 1,\ldots,n$. Тогда $F_k(z) = \sum_{x\in A_k} z^x = 1 + z + z^2$.

Отсюда следует $F_{A_1,\ldots,A_n}(z) = \prod_{i=1}^{n}(1+z^{a_i}+z^{2a_i})$.

Определение 1. Среднее число решений уравнения (1) – это величина вида

$$\overline{t_b} = \frac{1}{b^n}\sum_{a_1,\ldots,a_n} t_b(a_1,\ldots,a_n). \tag{8}$$

Перейдем теперь к рассмотрению величины $\overline{t}_b$.

Заметим, что ни один коэффициент в разрешимом уравнении не может превосходить b. Поэтому он может принимать значения от 1 до b. Кроме того, мы предполагаем, что в уравнении ровно n слагаемых. Таким образом, при равномерном распределении значений b величина $\overline{t}_b$ – это среднее число представлений правой части уравнения линейной формой $\sum_{i=1}^{n} a_i x_i$.

Пусть

$$F_n(z) = \sum_{i=1}^{n} \frac{1}{(1 - z^{a_i})}. \tag{9}$$

Лемма 1. *Справедливо соотношение*

$$F_n(z) = n + \sum_{i=1}^{n} A_i z^i, \tag{10}$$

где A_i – это число делителей числа i в множестве $\{a_1, \ldots, a_n\}$.

Доказательство. Из (9) следует

$$F_n(z) = n + \sum_{r_1=1}^{\infty} z^{r_1 a_1} + \sum_{r_2=1}^{\infty} z^{r_2 a_2} + \ldots + \sum_{r_n=1}^{\infty} z^{r_n a_n}. \tag{11}$$

Найдем коэффициент при z^N в (11). Вхождение z^N в s-ю сумму означает, что $N = ra_s$ для какого-то r. Поэтому $N \equiv 0 \pmod{a_s}$, что и доказывает утверждение леммы.

Теорема 2. *Справедлива формула*

$$\overline{t}_b = \frac{1}{b^n} \operatorname*{Coef}_u \left\{ \frac{1}{u^{n+1}} \left(b + \sum_{r=1}^{\infty} \frac{u^r}{(1-u)^r} \right)^n \right\}. \tag{12}$$

Доказательство. По определению и из (8) имеем

$$\overline{t}_b = \frac{1}{b^n} \sum_{\{a_1,\ldots,a_n\}} t_b(a_1, \ldots, a_n) = \frac{1}{b^n} \sum_{\{a_1,\ldots,a_n\}} \operatorname*{Coef}_u \left\{ \frac{1}{u^{b+1}} \prod_{i=1}^{n} \frac{1}{1-u^{a_i}} \right\} = \frac{1}{b^n} \operatorname*{Coef}_u \left\{ \frac{1}{u^{b+1}} \prod_{i=1}^{n} \sum_{a_i=1}^{b} \frac{1}{1-u^{a_i}} \right\}.$$

Теперь используем предыдущую лемму и получаем соотношение

$$\overline{t}_b = \frac{1}{b^n} \operatorname*{Coef}_u \left\{ \frac{1}{u^{b+1}} \left(b + \sum_{r=1}^{\infty} \tau(r) u^r \right)^n \right\}, \tag{13}$$

где $\tau(r)$ – число делителей r.

Производящая функция для числа делителей изучалась во многих работах (см., например, [6]). Для нее известно следующее соотношение:

$$\sum_{r=1}^{\infty} \tau(r) u^r = \sum_{r=1}^{\infty} \frac{u^r}{1-u^r}. \tag{14}$$

Подставляем (14) в (13) и получаем (12).

Это завершает доказательство теоремы.

Следствие 1. Справедливо соотношение

$$\overline{t}_b(a_1, \ldots, a_n) = \frac{1}{2\pi i} \oint_{|u|=\rho} \frac{\left(b + \sum_{r=1}^{\infty} \frac{u^r}{(1-u)^r} \right)^n}{u^{b+1}} du, \quad \rho < 1. \tag{15}$$

3. ЗАДАЧА О РАЗБИЕНИИ

Рассмотрим задачу о разбиении – нахождении числа разбиений P_n для фиксированного n, что эквивалентно нахождению числа решений диофантова уравнения

$$n = \sum_{i=1}^{k} x_i = \sum_{r=1}^{n} r y_r. \tag{16}$$

Если на свойства разбиения нет ограничений, то в качестве вектора **y** может быть выбран любой вектор, удовлетворяющий (14).

Найдем методом коэффициентов выражение производящей функции для числа разбиений:

$$P_n(z) = \sum_{n=1}^{\infty} P_n z^n. \tag{17}$$

Из (16) и (17) следует

$$\begin{gathered} P_n = \sum_{\{y_1\ldots,y_n\}} \operatorname*{Coef}_u \left\{ \frac{u^{\sum_{r=1}^{n} r y_r}}{u^{n+1}} \right\} = \operatorname*{Coef}_u \left\{ \frac{1}{u^{n+1}} \sum_{\{y_1\ldots,y_n\}} u^{\sum_{r=1}^{n} r y_r} \right\} = \\ = \operatorname*{Coef}_u \left\{ \frac{1}{u^{n+1}} \sum_{y_1=0}^{\infty} u^{y_1} \sum_{y_2=0}^{\infty} u^{y_2} \ldots \sum_{y_n=0}^{\infty} u^{y_n} \right\} = \operatorname*{Coef}_u \left\{ \frac{1}{u^{n+1}} \prod_{k=1}^{\infty} \frac{1}{1-u^k} \right\}. \end{gathered} \tag{18}$$

Отсюда и следует выражение для производящей функции

$$P_n(u) = \sum_{n=1}^{\infty} P_n u^n = \prod_{k=1}^{\infty} \frac{1}{1-u^k}. \tag{19}$$

Мы привели классическое выражение для производящей функции числа разбиений.

Если все x_i различны, то на языке $\{y_i\}$ это будет означать, что $y_i \le 1$, $i = 1, \ldots, n$.

Пусть теперь P_n^0 – число разбиений n на различные слагаемые и $P_n^0(z) = \sum_{n=1}^{\infty} P_n^0 z^n$ – производящая функция для него. Тогда аналогично предыдущему рассуждению, имеем

$$P_n^0 = \operatorname*{Coef}_u \left\{ \frac{\prod_{i=1}^{n} \sum_{y_i=0}^{1} u^{i y_i}}{u^{n+1}} \right\} = \operatorname*{Coef}_u \left\{ \frac{\prod_{i=1}^{\infty} (1+u^i)}{u^{n+1}} \right\}.$$

Отсюда следует

$$P_n^0(z) = \sum_{n=1}^{\infty} P_n^0 z^n = \prod_{k=1}^{\infty} (1+z^k).$$

Теорема 3. *Число разбиений на нечетные слагаемые равно числу разбиений на различные слагаемые.*

Доказательство. Если все x_i нечетные, то на языке $\{y_i\}$ это будет означать, что $y_{2r} = 0$, $r = 1, \ldots, [n/2]$.

Пусть теперь P_n^r – число разбиений n на нечетные слагаемые и $P_n^r(z) = \sum_{n=1}^{\infty} P_n^r z^n$ – производящая функция для него.

Для краткости обозначим через $v(z)$ выражение

$$v(z) = \frac{1}{(1-z)(1-z^3)(1-z^5)\ldots} = (1-z^2)(1-z^4)(1-z^6)\ldots = \frac{1}{\prod_{k=1}^{\infty}(1-z^{2k+1})},$$

а через $u(z)$ выражение $\prod_{k=1}^{\infty}(1+z^k)$.

Далее заметим, что

$$u(z)(1-z)(1-z^2)(1-z^3)\ldots = (1-z^2)(1-z^4)(1-z^6)\ldots = \prod_{r=1}^{\infty}(1-z^{2r}).$$

Но это означает, что

$$\frac{\prod_{r=1}^{\infty}(1-z^{2r})}{\prod_{r=1}^{\infty}(1-z^{r})} = \frac{1}{\prod_{r=0}^{\infty}(1-z^{2r+1})} = u(z).$$

Тогда из (18) следует

$$P_n^r(z) = \sum_{n=1}^{\infty} P_n^r z^n = \frac{1}{\prod_{k=1}^{\infty}(1-z^{2k+1})}. \tag{20}$$

Но в этом случае из (20) мы имеем равенство

$$P_n^0(z)\prod_{k=1}^{\infty}(1-z^{k}) = \prod_{k=1}^{\infty}(1-z^{2k}).$$

Но тогда из этого равенства получаем

$$P_n^0(z) = \frac{\prod_{k=1}^{\infty}(1-z^{2k})}{\prod_{k=1}^{\infty}(1-z^{k})} = \frac{1}{\prod_{k=1}^{\infty}(1-z^{2k+1})} = P_n^r(z).$$

Этим завершается доказательство утверждения.

Таким образом, мы здесь привели новое доказательство известного факта из теории разбиений (см., например, [31]).

4. О ПРОБЛЕМЕ ФРОБЕНИУСА

Рассмотрим диофантово уравнение (21) с двумя переменными. Введенные ниже обозначения будут использованы всюду в дальнейшем тексте.

$$ax + by = n. \tag{21}$$

Здесь все числа натуральные. Как известно, эти уравнения разрешимы для всех n, начиная с некоторого – границы разрешимости: $n_0(a, b, c)$ или $n_0(a, b)$. По теореме Сильвестра для взаимно простых a и b:

$$n_0(a,b) = ab - (a+b).$$

В [30] используя метод производящих функций, получена формула для числа решений $t_n(a,b)$ уравнения (21) при условии $(a, b) = 1$, из которой следует не только результат Сильвестра, но и формула для x_0 и y_0, на которых “лежит” граница разрешимости.

Теорема 4. *Если* $(a, b) = 1$, *то*

$$t_n(a,b) = \frac{n}{ab} - \frac{x_0(n)a + y_0(n)b}{ab} + 1, \tag{22}$$

где $x_0(n)$ *и* $y_0(n)$ – *минимальные решения сравнений*: $ax \equiv n (\operatorname{mod} b)$ *и* $by \equiv n (\operatorname{mod} a)$.

Следствие 2. При взаимно простых a и b из неравенства $n > ab - (a + b)$ следует неравенство $t_n(a,b) > 0$.

Доказательство. Так $x_0 \le b-1$ и $y_0 \le a-1$, то

$$t_n(a,b) = \frac{n}{ab} - \frac{x_0(n)a + y_0(n)b}{ab} + 1 \ge \frac{n-(ab-a-b)}{ab}.$$

Но по условию $n > ab-(a+b)$. Отсюда и следует неравенство $t_n(a,b) > 0$.

А это известная формула Сильвестра для двухмерного числа Фробениуса.

Формулу (22) можно записать в более "строгой" форме. Пусть $\varphi(k)$ – функция Эйлера (число чисел, меньших k и взаимно простых с k). Тогда очевидно, что сравнения $ax \equiv n(\mathrm{mod}\, b)$ и $by \equiv n(\mathrm{mod}\, a)$при взаимно простых a и b имеют единственные минимальные решения, представимые в виде:

$$x_0(n) = (na^{\varphi(b)-1})_{\mathrm{mod}\, b} \quad \text{и} \quad y_0(n) = (nb^{\varphi(a)-1})_{\mathrm{mod}\, a}.$$

Отсюда и получается другое выражение для формулы, задающей число решений уравнения (21):

$$t_n(a,b) = \frac{n}{ab} - \frac{(na^{\varphi(b)-1})_{\mathrm{mod}\, b}a + (nb^{\varphi(a)-1})_{\mathrm{mod}\, a}b}{ab} + 1. \tag{23}$$

Пример 4. Рассмотрим уравнение $4x + 7y = n$.

При $n = 12$ из (017) следует $t_{12}(4,7) = \frac{12}{28} - \frac{x_0(n)4 + y_0(n)7}{28} + 1$. Сравнения $4x \equiv 12(\mathrm{mod}\, 7)$ и $7y \equiv 12(\mathrm{mod}\, 4)$ имеют минимальные решения: $x_0(n) = 3$ и $y_0(n) = 0$. Отсюда $t_{12}(4,7) = \frac{12}{28} - \frac{12}{28} + 1 = 1$. Что и соответствует решению (3, 0).

При $n = 13$ из (017) следует $t_{13}(4,7) = \frac{13}{28} - \frac{x_0(n)4 + y_0(n)7}{28} + 1$. Сравнения $4x \equiv 13(\mathrm{mod}\, 7)$ и $7y \equiv 13(\mathrm{mod}\, 4)$ имеют минимальные решения: $x_0(n) = 5$ и $y_0(n) = 3$. Отсюда $t_{13}(4,7) = \frac{13}{28} - \frac{41}{28} + 1 = 0$. Что и соответствует тому факту, что решений у уравнения нет.

При $n = 32$ из (017) следует $t_{32}(4,7) = \frac{32}{28} - \frac{x_0(n)4 + y_0(n)7}{28} + 1$. Сравнения $4x \equiv 32(\mathrm{mod}\, 7)$ и $7y \equiv 32(\mathrm{mod}\, 4)$ имеют минимальные решения: $x_0(n) = 1$ и $y_0(n) = 0$. Отсюда $t_{32}(4,7) = \frac{32}{28} - \frac{4}{28} + 1 = 2$. Что и соответствует двум решениям: (8, 0) и (1, 4).

Следствие 3. Справедлива формула

$$t_{n+ab}(a,b) = t_n(a,b) + 1. \tag{24}$$

Доказательство. Рассмотрим функцию

$$L_n(a,b) = \frac{a(na^{\varphi(b)-1})_{\mathrm{mod}\, b} + b(nb^{\varphi(a)-1})_{\mathrm{mod}\, a}}{ab}.$$

Заметим, что $L_{n+ab}(a,b) = L_n(a,b)$. Имеем периодическую функцию с периодом ab.

Из этого факта и (23) следует (24).

Рассмотрим теперь вопрос о "среднем числе" решений уравнения (21).

Очевидно, что $a, b \le n$. Если параметры a и b равномерно распределены в квадрате nxn, то в качестве "среднего числа" решений можно взять следующую величину:

$$\overline{t_n} = \frac{1}{n^2}\sum_{a,b=0}^{n} t_n(a,b). \tag{25}$$

Величина $\overline{t_n}$ представляет определенный интерес, так как уравнение (21) может вовсе не иметь решений (например, если $(a,b) \not| n$), быть разрешимым для всех n (если $(a,b) = 1$, $n \ge (a-1)(b-1) - 1$ или находиться в любом из "промежуточных" случаев.

Пусть все параметры распределены независимо.

Лемма 2. *Пусть* $V = \{v_1, \ldots, v_k\}$ *— произвольное подмножество натуральных чисел. Тогда имеет место формула*:

$$Q_V(z) = \sum_{i=1}^{n} \frac{1}{1 - z^{v_i}} = \sum_{N=0}^{\infty} B_N z^N, \tag{26}$$

где B_N *— число делителей числа* N, *лежащих в множестве* V *и* $B_0 = k$.

Равенство (26) обосновывается следующим.

Во-первых, по определению справедливо равенство

$$Q_V(z) = \sum_{r=0}^{\infty} z^{rv_1} + \sum_{r=0}^{\infty} z^{rv_2} + \ldots + \sum_{r=0}^{\infty} z^{rv_k}. \tag{27}$$

Во-вторых, B_N равно числу вхождений z^N в (27). В-третьих, вхождение z^N в s-ю сумму в (27) эквивалентно тому, что $N = rv_s$ для некоторого r. Поэтому $N \equiv 0_{\bmod v_s}$. Из этого и вытекает утверждение леммы.

Пример 5. Если $V = \{2, 4\}$, то

$$Q_V(z) = \sum\nolimits_{r=0}^{\infty} z^{r2} + \sum\nolimits_{r=0}^{\infty} z^{r4} = \sum\nolimits_{N=0}^{\infty} B_N z^N,$$

тогда из леммы мы получаем, что

$$B_N = \begin{cases} 0, & \text{если} \quad N \equiv 1_{\bmod 2}, \\ 1, & \text{если} \quad N \equiv 0_{\bmod 2}, \quad N \not\equiv 0_{\bmod 4}. \\ 2, & \text{если} \quad N \equiv 0_{\bmod 4}, \end{cases}$$

В качестве следствия из этой леммы можно получить соотношение

$$B_N = \tau(N)$$

при $N > 0$.

Пусть

$$Q_n(z) = \sum_{i=1}^{n} \frac{1}{1 - z^i} = \sum_{N=0}^{\infty} B_N z^N, \quad J_n = \{1, 2, \ldots n\}, \quad D_N = \{d : d/N\}.$$

Тогда из леммы следует, что

$$B_N = D_N.$$

Но это означает, что

$$B_N = \tau(N) \quad \text{при} \quad N > 0. \tag{28}$$

Из (26) и (28) следует соотношение

$$Q_V(z) = \sum_{N=0}^{\infty} B_N z^N = N + \sum_{N=1}^{\infty} \tau(N) z^N. \tag{29}$$

Теорема 5. *Справедливо соотношение*

$$\overline{t}_n = \frac{2\tau(n)}{n} + \frac{1}{n^2} \sum_{r=1}^{n} \tau(r)\tau(n - r), \tag{30}$$

где $\tau(0) = n$.

Доказательство. Согласно введенным обозначениям, производящая функция для $t_n(a, b)$ выглядит следующим образом:

$$\sum_{n=0}^{\infty} t_n(a, b) z^n = \sum_{x,y} z^{ax+by}.$$

Далее, используя формулу Коши, получаем

$$t_n(a, b) = \frac{1}{2\pi i} \oint_{|u|=\rho} \frac{1}{z^{n+1}(1 - z^a)(1 - z^b)} dz, \quad \rho < 1. \tag{31}$$

Подставляем (31) в (25) и получаем соотношение

$$\overline{t}_n = \frac{1}{n^2}\frac{1}{2\pi i}\oint\limits_{|u|=\rho}\frac{1}{z^{n+1}}\sum_{a,b\le n}\frac{1}{(1-z^a)(1-z^b)}dz, \quad \rho < 1. \tag{32}$$

Далее обращаем внимание на то, что

$$\sum_{a,b\le n}\frac{1}{(1-z^a)(1-z^b)} = \sum_{a=1}^{n}\frac{1}{(1-z^a)}\sum_{b=1}^{n}\frac{1}{(1-z^b)} = Q_n^2(z). \tag{33}$$

Из (28), (31) и (32) следует, что

$$\overline{t}_n = \frac{1}{n^2}\operatorname*{Coef}_z\left\{Q_n^2(z)\right\} = \frac{1}{n^2}\sum_{r=0}^{n}B_rB_{n-r} = \frac{1}{n^2}\operatorname*{Coef}_z\left\{\frac{1}{z^{n+1}}\left(n+\sum_{r=1}^{n}\tau(r)z^r\right)^2\right\}. \tag{34}$$

Но тогда из (26), (29) и (34) вытекает утверждение теоремы.

Следствие 4. Справедливо неравенство $\overline{t}_n \le \log^2 n$.

Доказательство. Применяем неравенство Коши к (34), учитывая (28), и получаем

$$\sum_{r=0}^{n}B_rB_{n-r} \le \left(\sum_{r=0}^{n}B_r^2\right)^{1/2}\left(\sum_{r=0}^{n}B_{n-r}^2\right)^{1/2}. \tag{35}$$

Так как

$$\left(\sum_{r=0}^{n}B_r^2\right)^{1/2} = \left(\sum_{r=0}^{n}B_{n-r}^2\right)^{1/2},$$

то из (28) следует

$$\sum_{r=0}^{n}B_rB_{n-r} \le \sum_{r=0}^{n}B_r^2 \le \sum_{r=0}^{n}\tau^2(r) \le \left(\sum_{r=0}^{n}\tau(r)\right)^2. \tag{36}$$

Известно, что

$$\sum_{r=0}^{n}\tau(r) \le \sum_{r=1}^{n}\left[\frac{n}{r}\right]. \tag{37}$$

Из (36) и (37) следует соотношение

$$\sum_{r=0}^{n}B_rB_{n-r} \le n^2\log^2 n.$$

Из этого соотношения и (34) получаем требуемое неравенство

$$\overline{t}_n \le \log^2 n.$$

Следствие доказано.

5. ЗАКЛЮЧЕНИЕ

В работе были рассмотрены комбинаторные аспекты задач, связанных с разбиением целых чисел на слагаемые. Это хорошо известная и исследованная область. Мы попытались взглянуть на поднятую проблематику сквозь призму получения формул и оценок для числа решений диофантовых уравнений специального типа.

Это позволило не только достаточно естественно вывести уже известные результаты, но и получить новые, что для такой хорошо известной и изученной проблемы представляется полезным и интересным.

В каждом из трех основных разделов работы главные результаты (их четыре) названы теоремами. Прикладные области, где они могут быть использованы, обсуждены частично во введении, где дан краткий экскурс по возможным постановкам задач и полученным результатам.

СПИСОК ЛИТЕРАТУРЫ

1. *Пападимитриу Х., Стайглиц С.* Комбинаторная оптимизация. М.: Мир, 1989.
2. *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
3. *Фомичев В.М., Мельников Д.А.* Криптографические методы защиты информации. М.: Юрайт, 2017.
4. *Erdös P., Graham R.L.* On a linear diophantine problem of Frobenius // Acta Arith. 1972. V. 21. P. 399–408.
5. *Alfonsin J.R.* The Diophantine Frobenius Problem. Oxford: Oxford University Press, 2005.
6. *Харди Х., Райт Э.М.* Введение в теорию чисел. Oxford: Oxford University Press, 1975.
7. *Леонтьев В.К., Гордеев Э.Н.* Производящие функции в задаче о ранце // Докл. АН. 2018. Т. 481. № 5. С. 478–480.
https://doi.org/10.31857/S086956520002139-5
8. *Леонтьев В.К., Гордеев Э.Н.* О некоторых комбинаторных свойствах задачи о рюкзаке // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 8. С. 1439–1447.
https://doi.org/10.1134/S0044466919080076
9. *Егорычев Г.П.* Интегральное представление и вычисление комбинаторных сумм. Новосибирск: Наука, 1977.
10. *Kellerer H., Pferschy U., Pisinger D.* Knapsack problems. Berlin: Springer, 2004.
11. *Леонтьев В.К., Гордеев Э.Н.* Об алгебраической иммунности систем кодирования // Вопросы кибербезопасности. 2019. № 1. С. 59–89.
https://doi.org/10.21681/2311-3456-2019-1-59-68
12. *Гордеев Э.Н., Леонтьев В.К., Медведев Н.В.* О свойствах булевых полиномов, актуальных для криптосистем // Вопросы кибербезопасности. 2017. № 3. С. 63–69.
https://doi.org/10.21681/2311-3456-2017-3-63-69
13. *Мазуров И.Д., Хачай М.Ю.* Комитеты систем линейных неравенств // АиТ. 2004. № 2. С. 43–54.
https://doi.org/10.1023/B:AURC.0000014716.77510.61
14. *Береснев В.Л.* Эффективный алгоритм решения задачи минимизации полиномов от булевых переменных, обладающих свойством связности // Дискретн. анализ и исслед. операций. Сер. 2. 2005. Т. 12. № 1. С. 3–11.
15. *Sylvester J.J.* Problem 7382// The Educational Times, and Journal of the College of Preceptors, New Ser., 36(266) (1883), 177 Solution by W.J. Curran Sharp, ibid., 36(271) (1883), 315 Republished as [15].
16. *Sylvester J.J.* Problem 7382, in: W. J. C. Miller (Ed.), Mathematical questions, with their solutions, from The Educational Times, vol. 41, page 21. London: Francis Hodgson, 1884.
17. *Арнольд В.И.* Экспериментальное наблюдение математических фактов. М: Из-во МЦНМО, 2006.
18. *Arnold V.I.* Arithmetical Turbulence of Selfsimilar Fluctuations Statistics of Large Frobenius numbers of Additive Semigroups of Integer// Moscow Mathematical Journal. 2007. V. 7. № 2. P. 173–193.
19. *Арнольд В.И.* Слабые асимптотики числа решений диофантовых задач // Функциональный анализ и его приложения. 1999. Т. 33. № 4. С. 65–66.
20. *Фомичёв В.А.* Опенка экспонента некоторых графов с помощью чисел Фробениуса для трех аргументов // Прикладная дискретная матем. 2014. Т. 24. № 2. С. 88–96.
21. *Curtis F.* On formulas for the Frobenius number of a numerical semigroup // Math. Scand. 1990. V. 67. P. 190–192.
22. *Amitabha Tripathi*. Formulae for the Frobenius number in three Variables // J. of Number Theory. 2017. V. 170. P. 368–389.
23. *Савельев В.П., Шевченко В.Н.* Задача Фробениуса для трех чисел // Международные научные чтения (памяти Келдыша М.В.): Сб. статей. Международной научно-практической конференции (Москва, 16.11.2019 г.). М.: ЕФИР, 2019. С. 10–15.
24. *Song K.* The Frobenius problem for numerical semigroups generated by the Thabit numbers of the first, second kind base b and the Cunningham numbers // Bull. Korean Math. Soc. 2020. V. 57. P. 623–647.
25. *Rosales J.C., Branco M.B., Torrao D.* The Frobenius problem for Thabit numerical semigroups // J. Number Theory. 2015. V. 155. P. 85–99.
26. *Rosales J.C., Branco M.B., Torrao D.* The Frobenius problem for repunit numerical semigroups // Ramanujan J. 2016. V. 40. P. 323–334.
27. *Rosales J.C., Branco M.B., Torrao D.* The Frobenius problem for Mersenne numerical semigroups // Math. Z. 2017. V. 286. P. 741–749.
28. *Nijenhuis M.* A minimal-path algorithm for the “money changing problem” // The American Mathematical Monthly. 1979. V. 86. P. 832–835.
29. *Фомичёв В.А.* Эквивалентные но Фробениусу примитивные множества чисел // Прикладная дискретная матем. 2014. Т. 23. № 1. С. 20–26.
30. *Леонтьев В.К.* О проблеме Фробениуса // Дискретный анализ и исследование операций. 2022. Т. 29. № 2. С. 24–37.
31. *Эндрюс Г.* Теория разбиений. М.: Наука, 1982.

EDN: AYESAV
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1458–1472

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.642

ПРИМЕНЕНИЕ МОЗАИЧНО-СКЕЛЕТОННЫХ АППРОКСИМАЦИЙ МАТРИЦ В МЕТОДЕ ФИЗИЧЕСКОЙ ОПТИКИ ДЛЯ ЗАДАЧ ЭЛЕКТРОМАГНИТНОГО РАССЕЯНИЯ[1)]

© 2022 г. А. В. Сетуха[1,2,*], С. Л. Ставцев[2,], Р. М. Третьякова[2,***]**

[1] *119991 Москва, Ленинские горы, МГУ им. М.В. Ломоносова, Россия*

[2] *119333 Москва, ул. Губкина, 8, ИВМ РАН, Россия*

**e-mail: setuhaav@rambler.ru*

***e-mail: sstass2000@mail.ru*

****e-mail: r.tretyakova@inm.ras.ru*

Поступила в редакцию 04.03.2022 г.
Переработанный вариант 04.03.2022 г.
Принята к публикации 11.05.2022 г.

В статье для решения задач рассеяния электромагнитных волн рассмотрена модель физической оптики, основанная на приближении Кирхгофа–Макдональда с учетом переотражений. В рамках этой модели используется интегральное представление электромагнитного поля через поверхностные токи. В работе описан итерационный алгоритм, в котором на каждом шаге итераций определение поверхностных токов на ячейках разбиения поверхности выполняется через умножение матрицы влияния на токи, найденные на предыдущей итерации. Для повышения вычислительной эффективности алгоритма матрица влияния сжимается с применением метода мозаично-скелетонных аппроксимаций. При этом учитывается специфика аппроксимируемой матрицы, которая состоит в том, что ее элементы определяются через матрицу дискретного представления интегрального оператора, содержащую матрицу “видимости” ячеек разбиения. Матрица “видимости” указывает, имеет ли отрезок, соединяющий центры двух ячеек, пересечение с облучаемой поверхностью в своих внутренних точках. Проведено тестирование метода на модельных задачах, которое показало применимость предложенного алгоритма к решению задач рассеяния на невыпуклых телах, а также вычислительную эффективность алгоритма. Библ. 18. Фиг. 6.

Ключевые слова: численные методы, малоранговые аппроксимации, дифракция электромагнитных волн, приближение Кирхгофа–Макдональда.

DOI: 10.31857/S0044466922090034

1. ВВЕДЕНИЕ

При численном решении задач рассеяния монохроматических электромагнитных волн на идеально проводящих объектах широкое применение нашли методы, основанные на интегральном представлении электрического и магнитного полей через поверхностные токи [1], [2]. В точной постановке задача сводится к решению интегрального уравнения, записанного на поверхности облучаемых объектов, относительно поверхностных токов. При дискретизации такого уравнения возникает система линейных уравнений с плотной матрицей, причем размерность этой системы определяется числом ячеек разбиения поверхности, а число ячеек разбиения в свою очередь определяется как сложностью геометрии облучаемых объектов, так и требованием малости размера ячеек по отношению к длине волны. Поэтому при применении метода граничных интегральных уравнений вычислительная сложность алгоритма растет с увеличением частоты электромагнитного поля, что ограничивает применение метода для высоких частот.

При численном моделировании рассеяния волн высокой частоты широко применяются давно известные асимптотические методы. Одним из таких методов является метод физической оптики, основанный на приближении Кирхгофа–Макдональда [3]–[5]. В основе этого метода ле-

[1)]Работа выполнена при финансовой поддержке Московского центра фундаментальной и прикладной математики (Соглашение № 075-15-2019-1624 с Минобрнауки РФ).

жит гипотеза о том, что поле, индуцируемое источником излучения, взаимодействует с поверхностью идеально проводящего тела таким образом, что на затененной стороне тела поверхностные токи не возникают, а на каждом освещенном участке тела индуцируются поверхностные токи, такие же, как если бы инициирующее поле воздействовало на касательную плоскость. С математической точки зрения это, по сути, означает, что используется то же самое интегральное представление электромагнитного поля, что и при применении метода граничных интегральных уравнений, однако, само интегральное уравнение не решается, а его решение – поверхностные токи, определяется сразу на основании приближенной формулы, дающей связь между поверхностными токами на ячейках разбиения и инициирующим магнитным полем. В классическом методе физической оптики предполагается, что в качестве этого инициирующего поля выступает первичное поле, в качестве которого может выступать плоская волна, приходящая с бесконечности, или известное поле, возбуждаемое сторонними источниками излучения. Практика показывает, что такой подход применим только при моделировании облучения выпуклых тел.

Более сложная модель возникает при учете переотражений первичного внешнего поля. В этой модели предполагается, что для каждой ячейки, аппроксимирующей поверхность, в качестве инициирующих полей, возбуждающих поверхностный ток на этой ячейке, выступают как первичное внешнее поле, так и поля, индуцируемые поверхностными токами, сформировавшимися на других участках поверхности. В статье дается математическая формулировка такой модели и описывается численный итерационный алгоритм нахождения поверхностных токов. При этом на каждом шаге итераций определение поверхностных токов осуществляется умножением матрицы влияния, описывающей взаимодействие ячеек, на токи, найденные на предыдущей итерации. В результате вычислительная сложность алгоритма при прямом вычислении матрицы влияния становится порядка $\mathcal{O}(n^2)$, где n – число ячеек разбиения.

Существенного ускорения работы такого алгоритма можно добиться за счет применения быстрых матричных алгоритмов. В настоящей работе к решению задачи осуществлено приложение метода мозаично-скелетонных аппроксимаций [6], [7]. Этот метод нашел широкое применение в различных прикладаных задачах [8]–[10]. В данной работе учитывается специфика аппроксимируемой матрицы, которая состоит в том, что ее элементы определяются аналитическим выражением для ядра интегрального оператора с учетом матрицы “видимости” ячеек разбиения (матрица “видимости” указывает, имеет ли отрезок, соединяющий центры двух ячеек, пересечение с облучаемой поверхностью в своих внутренних точках).

Проведено тестирование разработанного численного алгоритма на модельных задачах, в котором поверялись как применимость построенной математической модели к решению задач рассеяния на невыпуклых объектах путем сравнения получаемых диаграмм обратного рассеяния с диаграммами, полученными методом интегральных уравнений и известными из физических экспериментов, так и вычислительная эффективность построенного алгоритма.

2. МЕТОД ФИЗИЧЕСКОЙ ОПТИКИ

Решается задача рассеяния заданного монохроматического первичного электромагнитного поля на системе идеально проводящих объектов (тел и экранов), суммарную поверхность которых обозначим через Σ. Поверхность Σ может состоять из одной или нескольких замкнутых компонент, которые являются поверхностями идеально проводящих тел, и разомкнутых компонент с краем. Каждая такая компонента моделирует идеально проводящий экран.

Напряженности электрического и магнитного полей ищутся в виде $\mathbf{E}(\mathbf{x})e^{-i\omega t}$, $\mathbf{H}(\mathbf{x})e^{-i\omega t}$, где $\mathbf{x} = (x_1, x_2, x_3)$ – точки пространства, t – время, ω – круговая частота колебаний.

Для пространственных составляющих электрического и магнитного полей $\mathbf{E}$ и $\mathbf{H}$ в области вне облучаемых тел справедливы уравнения Максвелла [11, с. 17]:

$$\operatorname{rot}\mathbf{E} = i\mu\mu_0\omega\mathbf{H}; \quad \operatorname{rot}\mathbf{H} = -i\varepsilon\varepsilon_0\omega\mathbf{E}, \tag{1}$$

где ε_0 и μ_0 – электрическая и магнитная постоянные, ε и μ – относительные диэлектрическая и магнитная проницаемости окружающей среды. Далее будем полагать, что среда вне тела имеет относительные диэлектрическую проницаемость $\varepsilon = 1$ и магнитную проницаемость $\mu = 1$.

Предполагается, что источником возбуждения является заданное первичное поле $(\mathbf{E}_0, \mathbf{H}_0)$. В качестве первичного поля может рассматриваться, например, плоская волна, для которой

$$\mathbf{E}_0(x) = \mathbf{E}_i \exp(i\mathbf{kr}), \quad \mathbf{H}_0(x) = \mathbf{k} \times \mathbf{E}_i \frac{\exp(i\mathbf{kr})}{\omega\mu_0}, \tag{2}$$

где $\mathbf{k}$ – заданный волновой вектор, определяющий направление движения волны, причем $|\mathbf{k}| = k = \omega\sqrt{\varepsilon_0\mu_0}$, $\mathbf{r}$ – радиус-вектор точки x, а $\mathbf{E}_i$ – заданный постоянный вектор, определяющий поляризацию волны, причем $(\mathbf{E}_i, \mathbf{k}) = 0$.

Задача состоит в отыскании полных электрического и магнитного полей (total fields)

$$\mathbf{E}_{\text{tot}} = \mathbf{E}_0 + \mathbf{E}, \quad \mathbf{H}_{\text{tot}} = \mathbf{H}_0 + \mathbf{H}, \tag{3}$$

где $(\mathbf{E}, \mathbf{H})$ – неизвестное вторичное поле.

При этом на идеально проводящей поверхности должно выполняться граничное условие для полного электрического поля [11, с. 17]:

$$\mathbf{n} \times \mathbf{E}_{\text{tot}} = 0, \tag{4}$$

где $\mathbf{n}$ – орт вектора нормали к поверхности, а вторичное поле должно удовлетворять условию излучения на бесконечности, которое мы запишем в форме Зоммерфельда [12, с. 110]:

$$\frac{d}{d\boldsymbol{\tau}}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix} - ik\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix} = o\left(\frac{1}{|\mathbf{x}|}\right), \quad \boldsymbol{\tau} = \frac{\mathbf{x}}{|\mathbf{x}|}, \quad |\mathbf{x}| \to \infty. \tag{5}$$

Электрическое и магнитное поля в задаче (1)–(5) будем искать в виде [11, с. 39]:

$$\mathbf{E}(x) = \frac{i}{\omega\varepsilon_0} K[\Sigma, \mathbf{j}](x), \quad \mathbf{H}(x) = R[\Sigma, \mathbf{j}](x),$$

$$K[\Sigma, \mathbf{j}](x) = \int_\Sigma \left(\operatorname{grad}_x \operatorname{div}_x[\mathbf{j}(y)F(x-y)] + k^2\mathbf{j}(y)F(x-y)\right) d\sigma_y; \tag{6}$$

$$R[\Sigma, \mathbf{j}](x) = \int_\Sigma \operatorname{rot}_x[\mathbf{j}(y)F(x-y)]d\sigma_y; \quad F(x-y) = \frac{\exp(ik|x-y|)}{4\pi|x-y|},$$

где $\mathbf{j}$ – неизвестное касательное векторное поле на поверхности Σ. С физической точки зрения $\mathbf{j}$ есть поле поверхностных токов.

В теории дифракции большое значение имеют приближенные методы, позволяющие исследовать дифракцию коротких волн на различных телах. Одним из таких методов, применимых к широкому классу задач, является приближение Кирхгофа–Макдональда, известное в электродинамике как приближение физической оптики [3]–[5].

В этой модели предполагается, что поле, индуцируемое источником излучения, взаимодействует с поверхностью идеально проводящего тела таким образом, что на затененной стороне тела поверхностные токи не возникают, а на каждом освещенном участке тела индуцируются поверхностные токи, такие же, как если бы инициирующее поле воздействовало на касательную плоскость.

В первом приближении мы рассматриваем в качестве инициирующего поля $(\mathbf{E}_{\text{inc}}, \mathbf{H}_{\text{inc}})$ для каждого участка поверхности тела основное первичное поле $(\mathbf{E}_0, \mathbf{H}_0)$ вида (2). В случае, если поле $\mathbf{E}_0$ есть плоская волна, определяемая формулой (2), мы будем называть точку $x \in \Sigma$ освещенной этим полем, если луч, проведенный из точки x в направлении вектора $-\mathbf{k}$, не имеет пересечений с поверхностью Σ, кроме самой точки x. В соответствии с теорией Кирхгоффа–Макдональда в каждой освещенной точке x поверхности тела Σ возбуждается поверхностный ток

$$\mathbf{j}(x) = 2\left[\mathbf{n}(x) \times \mathbf{H}_{\text{inc}}(x)\right]. \tag{7}$$

Такая модель применима для решения задачи дифракции волны на выпуклом теле (когда поверхность Σ ограничивает одну выпуклую область).

В более общем случае предполагаем, что для каждой точки $x \in \Sigma$ в качестве инициирующих полей, возбуждающих поверхностные токи в этой точке, выступают как первичное поле $(\mathbf{E}_0, \mathbf{H}_0)$,

так и поля, индуцируемые поверхностными токами, возникшими на других участках поверхности Σ, которые “видны” из точки x.

Если точка x лежит на разомкнутой компоненте поверхности Σ, то будем предполагать, что в этой точке могут возникать токи $\mathbf{j}^+(x)$ и $\mathbf{j}^-(x)$ на положительной и отрицательной сторонах поверхности соответственно.

Если точка x лежит на замкнутой компоненте поверхности Σ, то считаем, что в такой точке поверхностный ток возникает только на внешней (положительной) стороне поверхности. Этот ток обозначаем как $\mathbf{j}^+(x)$ и при этом считаем, что $\mathbf{j}^-(x) = 0$.

Предполагается, что в каждой точке $x \in \Sigma$ на каждой стороне поверхности Σ возбуждается поверхностный ток в соответствии с формулой (7), где $\mathbf{H}_{\text{inc}}(x)$ есть сумма первичного поля $\mathbf{H}_0$, если точка x освещена этим полем с соответствующей стороны, и магнитного поля, индуцируемого в соответствии с формулой (6) поверхностными токами, размещенными в тех точках поверхности Σ, которые видны из точки x с рассматриваемой стороны (если условие видимости не выполнено, соответствующее слагаемое не прибавляется). Тогда возникает следующее уравнение для поверхностных токов:

$$\mathbf{j}^{\pm}(x) = \pm 2\left[\mathbf{n}(x) \times \mathbf{H}^{\pm}_{\text{inc}}(x)\right], \tag{8}$$

$$\begin{aligned}\mathbf{H}^{\pm}_{\text{inc}}(x) = s^{\pm}(x)\mathbf{H}_0(x) + \int\limits_{\Sigma} a^{\pm}(x,y)a^{+}(y,x)\,\text{grad}_x(F(x-y)) \times \mathbf{j}^{+}(y)dy + \\ + \int\limits_{\Sigma} a^{\pm}(x,y)a^{-}(y,x)\,\text{grad}_x(F(x-y)) \times \mathbf{j}^{-}(y)dy,\end{aligned} \tag{9}$$

где $s^{\pm}(x)$ — параметры освещенности поверхности первичным полем, $a^{\pm}(x,y)$ — функции видимости точки $y \in \Sigma$ наблюдателем, находящимся в точке $x \in \Sigma$ на положительной или отрицательной стороне поверхности соответственно. Мы полагаем, что:

- $s^{+}(x) = 1$, если точка x освещена первичным полем и при этом $(\mathbf{k}, \mathbf{n}(x))/|\mathbf{k}| < -\delta$;
- $s^{-}(x) = 1$, если точка x освещена первичным полем и при этом $(\mathbf{k}, \mathbf{n}(x))/|\mathbf{k}| > \delta$;
- $a^{+}(x,y) = 1$, если отрезок $[x,y]$ не имеет пересечений с поверхностью Σ, кроме как в точках x и y, и при этом $(y-x, \mathbf{n}(x)) > \delta|y-x|$;
- $a^{-}(x,y) = 1$, если отрезок $[x,y]$ не имеет пересечений с поверхностью Σ, кроме как в точках x и y, и при этом $(y-x, \mathbf{n}(x)) < -\delta|y-x|$;
- коэффициенты $s^{\pm}(x)$ и $a^{\pm}(x,y)$ равны 0 во всех остальных случаях,

в исходной постановке метода в данных условиях мы полагаем $\delta = 0$.

Соотношения (8) и (9) представляют собой систему интегральных уравнений, которая решается итерационно. На нулевом шаге полагаем $\mathbf{j}^{+}_0(x) = \mathbf{j}^{-}_0(x) = 0$ для всех $x \in \Sigma$. Далее для каждого $m = 1, 2, \ldots$ находим токи $\mathbf{j}^{+}_m(x)$ и $\mathbf{j}^{-}_m(x)$ по формуле:

$$\mathbf{j}^{\pm}_m(x) = \pm 2\left[\mathbf{n}(x) \times \mathbf{H}^{\pm}_m(x)\right], \tag{10}$$

$$\begin{aligned}\mathbf{H}^{\pm}_m(x) = s^{\pm}(x)\mathbf{H}_0(x) + \int\limits_{\Sigma} a^{\pm}(x,y)a^{+}(y,x)\,\text{grad}_x(F(x-y)) \times \mathbf{j}^{+}_{m-1}(y)dy + \\ + \int\limits_{\Sigma} a^{\pm}(x,y)a^{-}(y,x)\,\text{grad}_x(F(x-y)) \times \mathbf{j}^{-}_{m-1}(y)dy.\end{aligned} \tag{11}$$

Для численной реализации описанного итерационного метода (10), (11) поверхность Σ аппроксимируется системой ячеек σ_i, $i = 1, \ldots, n$, а для токов строится их кусочно-постоянная аппроксимация: предполагается, что на каждой ячейке σ_i токи $\mathbf{j}^{\pm}_m(x)$ принимают постоянные значения $\mathbf{j}^{\pm}_{i,(m)}$, $i = 1, \ldots, n$. При этом на каждой ячейке σ_i выбирается точка коллокации x_i и рассчи-

тывается вектор $\mathbf{n}_i$ – приближение вектора нормали к ячейке. Далее на каждом шаге итерационного процесса дискретные значения токов рассчитываются по формуле:

$$\mathbf{j}^{\pm}_{i,(m)} = \pm 2\mathbf{n}_i \times \mathbf{H}^{\pm}_m(x_i),$$
$$\mathbf{H}^{\pm}_m(x_i) = s^{\pm}_i \mathbf{H}_0(x_i) + \sum_{k=1}^{n} a^{\pm}_{ik} a^{+}_{ki} R[\sigma_k, \mathbf{j}^{+}_{k,(m-1)}](x_i) + \sum_{k=1}^{n} a^{\pm}_{ik} a^{-}_{ki} R[\sigma_k, \mathbf{j}^{-}_{k,(m-1)}](x_i), \quad i = 1,\ldots,n, \tag{12}$$

где коэффициенты $s^{\pm}_i$ и $a^{\pm}_{ik}$ аппроксимируют значения $s^{\pm}(x_i)$ и $a^{\pm}(x_i, x_k)$ соответственно. При этом мы учитываем, что поверхность Σ заменяется при численном решении приближенной поверхностью $\tilde{\Sigma} = \bigcup_{i=1,\ldots,n} \sigma_i$. Коэффициенты $s^{\pm}_i$ и $a^{\pm}_{ik}$ вычисляются как значения $s^{\pm}(x_i)$ и $a^{\pm}(x_i, x_k)$ по описанному выше алгоритму, но с проверкой освещенности и видимости по отношению к поверхности $\tilde{\Sigma}$ и с использованием значения параметра $\delta = 10^{-4}$.

В формуле (12) оператор R вычисляется согласно выражению (6). В основе численного метода вычисления интегралов по ячейкам разбиения лежат квадратурные формулы, описанные в статье [13]. Заметим, что в силу определения функций $a^{\pm}(x, y)$ справедливо равенство $a^{\pm}_{ii} = 0$, поэтому в формуле (12) нет необходимости выделять особенность в операторе $R[\sigma_k, \mathbf{j}^{\pm}_{k,(m-1)}](x_i)$ для $i = k$.

Вычислительная сложность решаемой задачи определяется свойствами поверхности Σ и волновым числом k. Для хорошей аппроксимации поверхностных токов необходимо иметь разбиение, при котором на длину волны приходится несколько ячеек (вычислительные эксперименты показывают, что в методах описываемого типа нужно брать не менее 6 ячеек на длину волны). Это приводит к необходимости увеличивать число ячеек при уменьшении длины волны, что в свою очередь приводит к увеличению вычислительных затрат. Нахождение значений токов $\mathbf{j}^{\pm}_{i,(m)}$ по формулам (12) через значения этих же токов на предыдущем шаге итераций непосредственно на основании записанных формул (12) требует порядка $\mathbb{O}(n^2)$ вычислений значения оператора $R[\sigma_k, \mathbf{j}^{+}_{k,(m-1)}](x_i)$. При этом с вычислительной точки зрения выполнение такого шага итераций можно представить в виде умножения некоторой матрицы на вектор, причем эта матрица является плотной. Поэтому существенного повышения вычислительной эффективности метода можно добиться за счет использования быстрых матричных алгоритмов – это алгоритмы, в которых осуществляется аппроксимация матрицы, позволяющая приближенно выполнять умножение матрицы на вектор с вычислением небольшого числа элементов матрицы и имеющие вычислительную сложность меньше, чем $\mathbb{O}(n^2)$.

Более того, расчет матриц видимости $A^{\pm} = (a^{\pm}_{ik})$ по прямому алгоритму, при котором каждый элемент $a^{\pm}_{ik}$ вычисляется путем проверки пересечения отрезка, соединяющего точки x_i и x_k, с каждой из ячеек разбиения поверхности $\tilde{\Sigma}$, требует $\mathbb{O}(n^3)$ операций.

В настоящей работе для ускорения итерационного процесса, определяемого формулами (12), применяется метод мозаично-скелетных аппроксимаций. Существенная особенность данной задачи с точки зрения применения метода мозаично-скелетных аппроксимаций обусловлена дискретным характером функций видимости $a^{\pm}(x, y)$, а так же необходимостью применения быстрых алгоритмов и к приближенному вычислению самих этих функции. Ниже сначала дадим общее описание метода мозаично-скелетных аппроксимаций матриц, а затем опишем особенности применения этого метода в рассматриваемой задаче.

3. МЕТОД МОЗАИЧНО-СКЕЛЕТНЫХ АППРОКСИМАЦИЙ

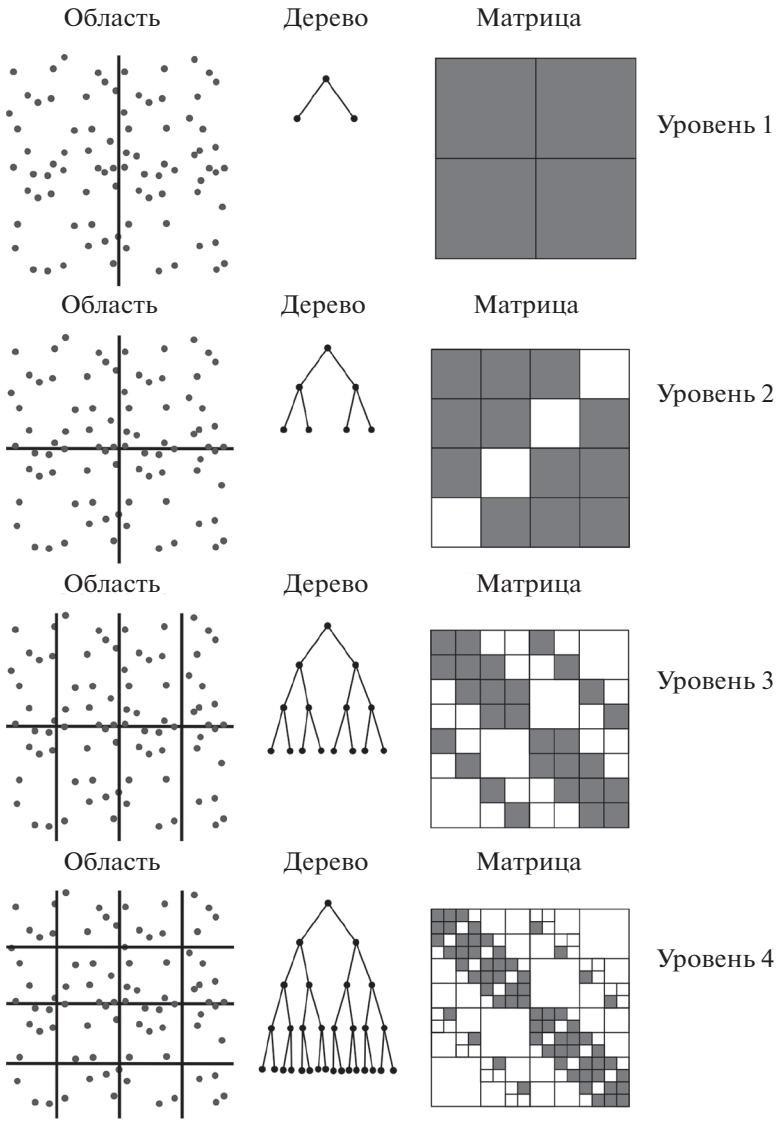
Алгоритм построения аппроксимации $\bar{A}$ для исходной матрицы A на основе метода мозаично-скелетных аппроксимаций состоит из следующих этапов [7].

Шаг 1. Построение дерева кластеров.

Шаг 2. Построение списка блоков матрицы.

Шаг 3. Аппроксимации блоков матрицы.

На этапе построения дерева кластеров используется информация о сетке, на которой проведена дискретизация исходного оператора. Сетка рассматривается как набор точек коллокации



Фиг. 1. Разбиение области и матрицы на части.

x_i, $i = 1,\ldots,n$, используемых в численной схеме (12). Так как для токов на каждой ячейке рассматриваются постоянные значения $\mathbf{j}^{\pm}_{j,(m)}$, $j = 1,\ldots,n$, то каждому столбцу аппроксимируемой матрицы A можно поставить в соответствие произвольную точку y_j в ячейке σ_j. Мы в дальнейшем будем считать, что $y_j = x_j$ (это используется при учете видимости ячеек). Таким образом, квадратная матрица A размера n определяет некоторое взаимодействие "источников", заданных на сетке $\{y_j\}$, $j = 1,\ldots,n$, с точками коллокации x_i, $i = 1,\ldots,n$. Точки каждой из сеток (источников и точек коллокации) группируются по степени геометрической близости в кластеры. Кластеры в свою очередь формируют бинарное дерево, в корне которого лежит кластер, содержащий все узлы сетки, и при переходе на новый уровень дерева текущий кластер разбивается на два дочерних.

Имея два дерева кластеров, отвечающих паре сеток, на которых проведена дискретизация, мы разбиваем матрицу на иерархический набор блоков разного размера, каждый из которых отвечает взаимодействию некоторой группы источников $\{y_j\}$ с некоторой группой точек коллокации $\{x_i\}$. Блоки, отвечающие взаимодействию достаточно удаленных друг от друга кластеров точек, могут быть приближены матрицей малого ранга.

На фиг. 1 представлен пример разбиения матрицы на блоки в мозаично-скелетонном методе для частного случая, когда область разбиения является частью плоскости – квадратом. На этом

рисунке представлено соответствие разбиения области, в которой находятся точки приемники (с источниками), дереву кластера и разбиению матрицы на блоки. Число уровней разбиения задается пользователем. Серым цветом показаны плотные блоки, остальные блоки малоранговые.

Используя алгоритм неполной крестовой аппроксимации [7], каждый малоранговый блок B размера $p \times q$ с заданной точностью ε представляется в виде скелетного разложения $B = UV^{\mathrm{T}}$ (см. [6], [14]), где множители U и V имеют размеры $p \times r$ и $q \times r$ соответственно, а r – ранг B.

Скелетное представление позволяет хранить матрицу в сжатом виде, так как вместо $p \cdot q$ элементов блока хранится только $(p+q)r$ элементов, что дает значительный выигрыш при $r \ll \min(p,q)$. Кроме того, малоранговый блок умножается на вектор за $(p+q)r$ операций вместо $p \cdot q$ операций для плотного блока.

4. БЫСТРЫЕ АЛГОРИТМЫ В ФИЗИЧЕСКОЙ ОПТИКЕ С ИСПОЛЬЗОВАНИЕМ ИЕРАРХИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ДАННЫХ

Перепишем формулу (12) с учетом того, что токи $\mathbf{j}^{\pm}_{i,(m)}$ постоянны на i-й ячейке:

$$\mathbf{H}^{\pm}_m(x_i) = s^{\pm}_i \mathbf{H}_0(x_i) + \sum_{k=1}^{n} R_{\pm+}[\sigma_k](x_i) \times \mathbf{j}^{+}_{k,(m-1)} + \sum_{k=1}^{n} R_{\pm-}[\sigma_k](x_i) \times \mathbf{j}^{-}_{k,(m-1)}, \tag{13}$$

где введены 4 матрицы, для которых необходимо применить мозаично-скелетный алгоритм,

$$\begin{aligned} R_{++}[\sigma_k](x_i) &= a^{+}_{ik} a^{+}_{ki} \tilde{R}[\sigma_k](x_i); \quad & R_{-+}[\sigma_k](x_i) &= a^{-}_{ik} a^{+}_{ki} \tilde{R}[\sigma_k](x_i), \\ R_{+-}[\sigma_k](x_i) &= a^{+}_{ik} a^{-}_{ki} \tilde{R}[\sigma_k](x_i); \quad & R_{--}[\sigma_k](x_i) &= a^{-}_{ik} a^{-}_{ki} \tilde{R}[\sigma_k](x_i), \end{aligned} \tag{14}$$

$$\tilde{R}[\sigma](x_i) = \frac{1}{4\pi} \int_{\sigma} (x_i - y) \Phi(x_i - y) d\sigma_y, \quad \Phi(x_i - y) = (ik|x_i - y| - 1) \frac{\exp(ik|x_i - y|)}{|x_i - y|^3}. \tag{15}$$

В формуле (14) коэффициенты $R_{\pm\pm}[\sigma_k](x_i)$ отличаются друг от друга только коэффициентами $a^{\pm}_{ik}$ и $a^{\pm}_{ki}$. Для удобства записи функцию видимости $a^{+}_{ik} a^{+}_{ki}$, $a^{-}_{ik} a^{+}_{ki}$, $a^{+}_{ik} a^{-}_{ki}$, $a^{-}_{ik} a^{-}_{ki}$ соответствующего оператора R_{++}, R_{-+}, R_{+-}, R_{--} обозначим через $\hat{a}_{ik}$. Тогда формулы (14) запишем в виде одной:

$$\hat{R}[\sigma_k](x_i) = \hat{a}_{ik} \tilde{R}[\sigma_k](x_i), \tag{16}$$

где оператор $\hat{R}$ принимает одно из значений R_{++}, R_{-+}, R_{+-}, R_{--}.

Заметим, что матрицы $(R[\sigma_j](x_i)))$ являются векторными. Задача вычисления векторов $\mathbf{H}^{\pm}_m(x_i)$, $i = 1,\ldots,n$, по формулам (13) сводится к вычислению векторов

$$\widehat{\mathbf{h}}_i = \sum_{k=1}^{n} \hat{R}[\sigma_k](x_i) \times \mathbf{j}_k, \quad i = 1,\ldots,n,$$

где $\mathbf{j}_k$ – один из наборов векторов $\mathbf{j}^{+}_{k,(m-1)}$ или $\mathbf{j}^{-}_{k,(m-1)}$. В свою очередь задача нахождения каждой из координат всех векторов $\widehat{\mathbf{h}}_i$ сводится к умножению матриц, составленных из координат векторов $\hat{R}[\sigma_k](x_i)$, на соответствующие столбцы координат векторов $\mathbf{j}_k$ (нужно выполнить 6 умножений скалярных матриц на векторы). При этом нужно аппроксимировать 3 скалярные матрицы, составленные из координат векторов $\hat{R}[\sigma_k](x_i)$.

Далее заметим, что можно аппроксимировать интегралы (15) так, чтобы задача свелась к выполнению умножений на одну скалярную квадратную матрицу размера n. Для этого запишем приближенное равенство:

$$\tilde{R}[\sigma_k](x_i) = \frac{1}{4\pi}(x_i - x_k) \int_{\sigma_k} \Phi(x_i - y) d\sigma_y$$

(это выражение используем при $i \neq k$, при $i = k$ можем положить $\tilde{R}[\sigma_k](x_i)) = 0$, так как $a_{ik}^{\pm} = 0$ при $i = k$). Тогда

$$\widehat{\mathbf{h}}_i = \sum_{k=1}^{n} [(x_i - x_k) \times \mathbf{j}_k] \hat{R}_{ik}, \quad i = 1,\ldots,n,$$

где

$$\hat{R}_{ik} = \hat{a}_{ik} R_{ik}, \quad R_{ik} = \int_{\sigma_k} \Phi(x_i - y) d\sigma_y, \quad i = 1,\ldots,n, \quad k = 1,\ldots,n, \quad i \neq k, \quad \hat{R}_{ik} = 0 \quad \text{при} \quad i = k. \tag{17}$$

Задача нахождения векторов $\hat{R}[\sigma_k](x_i)$ сводится к 6 умножениям на скалярную квадратную матрицу $\hat{R}_{ik}$ размера n, что видно из записи:

$$\widehat{\mathbf{h}}_i = x_i \times \sum_{k=1}^{n} \mathbf{j}_k \hat{R}_{ik} - \sum_{k=1}^{n} [x_k \times \mathbf{j}_k] \hat{R}_{ik}, \quad i = 1,\ldots,n.$$

Интегралы в выражениях (17) вычислялись по квадратурным формулам, описанным в статье [13].

Перейдем к применению метода мозаично-скелетных аппроксимаций. Для всех четырех матриц $\hat{R}_{ik}$ (14) строится одно и то же дерево кластеров. При построении блочной структуры матрицы $\hat{R} = (\hat{R}_{ij})$ необходимо учитывать дискретный характер функции видимости $\hat{a}_{ik}$. Если в малоранговый блок будут входить как видимые, так и невидимые пары ячеек, то в малоранговом блоке необходимо аппроксимировать нулевые подматрицы, что при заданной точности аппроксимации блока приводит к значительному росту ранга блока. Более того, если известно, что большая подматрица соответствует невидимым друг относительно друга областям, то умножение такой подматрицы на вектор в (13) будет давать нулевую добавку к вектору $\mathbf{H}_m^{\pm}(x_i)$, поэтому не нужно аппроксимировать и хранить такую подматрицу.

При аппроксимации матрицы видимости полагаем, что точки-приемники и точки-источники совпадают ($x_i = y_i$ для всех $i = 1,\ldots,n$).

Согласно мозаично-скелетному методу при проходе по кластерным деревьям от корней к листьям каждой паре узлов дерева соответствуют некоторые области пространства Σ_s и Σ_r, содержащие некоторую группу источников $\{y_j\}$ и некоторую группу точек приемников (точек коллокации)$\{x_i\}$. Этим группам точек соответствует некоторый блок матрицы. Введем следующие типы блоков с точки зрения выполнения условия видимости между их элементами.

1. Назовем блок “нулевым” , если для пары узлов деревьев, определяющих блок, соответствующая функция видимости $\hat{a}_{ij}$ равна нулю для любых i и j.

2. Назовем блок “видимым”, если блок является малоранговым с точки зрения аппроксимации матрицы R_{ik} и при этом для пары узлов деревьев, определяющих блок, соответствующая функция видимости $\hat{a}_{ij}$ равна 1 для любых i и j.

Сначала выполняется мозаичное биение матрицы на исходные малоранговые и плотные блоки с точки зрения аппроксимации матрицы R_{ij} без учета видимости.

Далее все блоки проверяются на предмет отнесения к одному ир типов: “нулевой” или “видимый”. Если блок не принадлежит ни к одному из этих типов, то этот блок разбивается на подблоки следующего уровня и проверяется принадлежность этих подблоков к рассматриваемым типам. Если подблок не принадлежит ни к одному из указанных типов, то он опять разбивается на подблоки следующего уровня и т.д., до перехода на нижний используемый уровень. Блоки нижнего уровня, не относящиеся к рассматриваемым типам, объявляются плотными.

В результате возникло уточненное биение матрицы, которое содержит блоки: “нулевые”, “видимые” малоранговые и “плотные”.

“Нулевые” блоки не вычисляются, не хранятся и не используются при расчете векторов $\mathbf{H}_m^{\pm}(x_i)$ из формулы (13) и токов $\mathbf{j}_{i,(m)}^{\pm}$ из формулы (12). “Видимые” малоранговые блоки (далее будем называть их малоранговыми) аппроксимируются скелетонным представлением. Для плотных блоков вычисляются все элементы матрицы.

В расчетах проверка типа блока матрицы (“нулевой” или “видимый”) осуществляется приближенно. Вычислительные эксперименты показывают, что для определения свойства видимости на поверхности Σ можно воспользоваться следующей процедурой. Для блока, заданного набором точек коллокации $\{x_i\}$, $i = 1,\ldots,p$, и группой источников $\{y_j\}$, $j = 1,\ldots,q$, рассмотрим подмножество точек коллокации $\{x_i^0\}$, $i = 1,\ldots,p_0$ (набор индексов обозначим через $\{I_0\}$) и источников $\{y_j^0\}$, $j = 1,\ldots,q_0$ (набор индексов обозначим через $\{J_0\}$), причем $p_0 \ll p$ и $q_0 \ll q$. Вычисляем функции видимости $\hat{a}_{ij}$ для $i = 1,\ldots p$, $j \in J_0$, и для $i \in I_0$, $j = 1,\ldots q$. Если для всех выбранных индексов функция принимает одно и то же значение (видимости или невидимости), то предполагаем, что весь блок принимает то же самое значение. Наборы точек $\{x_i^0\}$, $i = 1,\ldots,p_0$, и $\{y_j^0\}$, $j = 1,\ldots,q_0$, для блока формировались следующим образом. Каждый блок определяет взаимодействие области Σ_i, в которой находятся точки коллокации $\{x_i\}$, с областью Σ_j, в которой находятся источники $\{y_j\}$. Область Σ_i разбивалась на p_0 частей, в каждой из которой выбиралась точка $\{x_i^0\}$, ближайшая к центральной точке подобласти. Аналогично выбиралась точка $\{y_j^0\}$ из разбиения области Σ_j.

Такая проверка осуществляется только для блоков, начиная с некоторого назначаемого уровня k_0. Если блок имеет низший уровень, то такая приближенная проверка делается для всех его подблоков уровня k_0. Это актуально для тел со сложной геометрией, например, при наличии небольших отверстий и щелей.

В приводимых ниже примерах бралось значение параметра $k_0 = 1$. При этом численные исследования показали, что для всех блоков достаточно выбрать параметры $p_0 = q_0 = 8$.

Теперь рассмотрим алгоритм построения функции видимости для двух точек.

При проверке видимости пары точек сразу проверятся условия:

- если $|(y_k - x_i, \mathbf{n}(x_i))| \le 10^{-4}$ или $|(y_k - x_i, \mathbf{n}(y_k))| \le 10^{-4}$, то $a_{ik}^{\pm} = 0$;
- пусть Σ ограничивает замкнутую (не обязательно выпуклую) область и эта поверхность аппроксимируется замкнутой же поверхностью $\tilde{\Sigma}$ и пусть точка x_i лежит на поверхности $\tilde{\Sigma}$; так как на замкнутой поверхности поверхностные токи размещаются только на внешней стороне, следует положить $a_{ik}^{-} = 0$, $k = 1,\ldots,n$. Также для любой точки y_k, расположенной так, что $(y_k - x_i, \mathbf{n}_i) \le 10^{-4}$, сразу выполнено условие $a_{ik}^{+} = 0$ (точка x_i видна из точки y_k изнутри). Для такой пары точек проверка видимости не проводится.
- если при проверке пересечения отрезка $[x_i, y_k]$ с ячейкой σ_j выяснилось, что точки x_i и y_k не видны, то нет необходимости проверять остальные ячейки поверхности $\tilde{\Sigma}$.

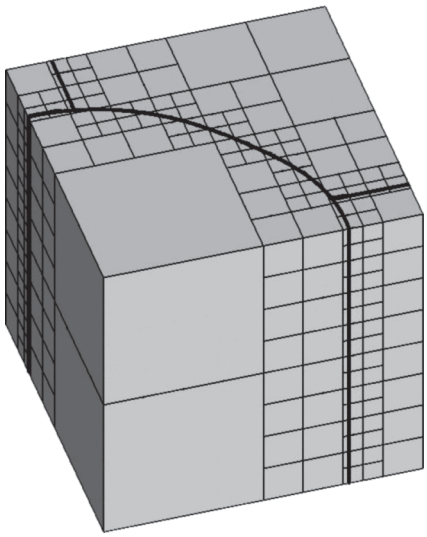
Эти условия, позволяют исключить дальнейшую сложную проверку для многих пар точек, однако, не снижают асимптотическую сложность алгоритма.

Для оставшихся пар точек воспользуемся иерархическими процедурами определения видимости.

Опять осуществим иерархическое разбиение области, содержащей поверхность $\tilde{\Sigma}$, на подобласти. Область, в которой находится поверхность $\tilde{\Sigma}$, помещается в куб. Далее строится восьмеричное дерево, согласно которому кубические области разбиваются на кубические подобласти с уменьшением размеров куба в 2 раза. Если подобласть не содержит ячейки поверхности $\tilde{\Sigma}$, то такая подобласть отмечается и дальше не разбивается. Пример иерархического разбиения области изображен на фиг. 2 (поверхность изображена жирной черной линией). Подготовительный этап выполняется один раз перед расчетом матриц видимости $a_{ik}^{\pm}$.

Для проверки видимости точки $x_i \in \tilde{\Sigma}$ относительно точки $y_k \in \tilde{\Sigma}$ определяются кубические области $\{\Omega_l\}$, которые соответствуют листьям восьмеричного дерева иерархичного разбиения области и которые могут содержать ячейки, пересекаемые отрезком $[x_i, y_k]$. В список таких областей входят:

- области, содержащие точки x_i и y_k;
- области, пересекаемые отрезком $[x_i, y_k]$.



Фиг. 2. Разбиение области согласно восьмеричному дереву.

Набор областей получается при проходе дерева от корня к листьям. Если в результате прохода встретилась область, которую не пересекает отрезок $[x_i, y_k]$, то далее подобласти такой области не проверяются. Если область не содержит ни одной ячейки поверхности $\tilde{\Sigma}$, то такая область в набор областей $\{\Omega_l\}$ не включается.

После формирования списка областей $\{\Omega_l\}$ проверяется пересечение отрезка $[x_i, y_k]$ только с ячейками, лежащими в областях Ω_l.

Решение задачи по формуле (13) с использованием малоранговых аппроксимаций показывает, что на аппроксимацию матрицы тратится гораздо больше времени, чем на умножение малоранговой матрицы на вектор. Поэтому перед итерациями один раз строится блочная малоранговая структура матриц $\hat{R}_{ij}$, а также малоранговая аппроксимация скалярной матрицы (R_{ij}). Матрица в сжатом виде хранится в оперативной памяти, а на каждой итерации пересчет токов выполняется по формулам (12), (13).

Осуществим приближенную оценку сложности построенного вычислительного алгоритма. Сначала оценим сложность алгоритма выделения “нулевых” и “видимых” малоранговых блоков.

Число уровней восьмеричного дерева выбирается таким образом, чтобы с ростом числа ячеек n среднее число ячеек на нижнем уровне оставалось примерно постоянным. Тогда число блоков в мозаичном биении матрицы растет как $\mathbb{O}(n\log n)$ (см. [15]). Значит, для выделения “нулевых” и “видимых” блоков требуется $\mathbb{O}(n\log n)$ вызовов функции видимости для пары расчетных точек x_i и y_j. Сложность алгоритма выделения таких блоков можно оценить как $\mathbb{O}(n\log n)\Psi(n)$, где $\Psi(n)$ – среднее число операций при проверке видимости для пары проверяемых расчетных точек x_i и y_j. Сразу можно сказать, что заведомо выполнено условие $\Psi(n) = \mathbb{O}(n)$.

Более сильную оценку для числа операций при проверке видимости для пары точек, лежащих на поверхности, можно получить, если предположить, что отрезок, соединяющий две любые такие точки, пересекает не более чем L непустых областей одного уровня восьмеричного дерева (областей из списка $\{\Omega_l\}$). Пусть x и y – две точки, лежащие на поверхности $\tilde{\Sigma}$, возникающей при аппроксимации поверхности Σ системой из n_1 ячеек. Предположим, что для проверки функции видимости для этих точек потребовалось $\Psi_{x,y}(n_1)$ проверок пересечения отрезка $[x, y]$ с областями из списка $\{\Omega_l^1\}$. Измельчим сетку на поверхности $\tilde{\Sigma}$, разбив каждую ячейку на 4 части. Тогда на поверхности $\tilde{\Sigma}$ возникнет новая поверхностная сетка с числом ячеек $n_2 = 4n_1$. Увеличим число уровней восьмеричного дерева на 1, разбив каждую область нижнего уровня на 8 кубических областей следующего уровня, и опять выделим семейство областей $\{\Sigma_l^{(2)}\}$, содержащих точки поверхности. При этом в каждой из областей семейства $\{\Sigma_l^{(2)}\}$ опять окажется примерно m ячеек.

В этом случае для каждой области Ω из списка $\{\Omega_l^1\}$, являющейся областью нижнего уровня для первого случая и имеющей пересечение с отрезком $[x, y]$, придется провести не более 8 проверок на пересечение этого же отрезка с областями нижнего уровня для второго дерева, лежащими в области Ω. Поскольку число таких областей Ω не более L, получаем: $\Psi_{x,y}(4n_1) \le \Psi_{x,y}(n_1) + 8L$, откуда следует, что $\Psi_{x,y}(n) = \mathbb{O}(\log n)$.

Если гипотеза об ограниченности числа точек пересечения отрезков с кубами нижнего уровня верна, то сложность алгоритма выделения "нулевых" и "видимых" блоков оценивается в $\mathbb{O}(n\log^2 n)$ операций.

После построения блочной структуры для матриц (14) с помощью алгоритма неполной крестовой аппроксимации необходимо построить скелетные разложения малоранговых блоков. При этом число операций при разложении малоранговых блоков не превосходит число операций при аппроксимации матрицы R_{ij} (без множителей видимости), так как система малоранговых блоков матрицы $\hat{R}_{ij}$ по сути отличается от системы малоранговых блоков матрицы R_{ij} выбрасыванием "нулевых" блоков и отнесением части малоранговых блоков к плотным. Поэтому для алгоритма построения скелетного разложения малоранговых блоков матрицы $\hat{R}_{ij}$ применима оценка $\mathbb{O}(rn\log^2 n)$, где r – мозаичный ранг [15]. Для числа операций при умножении малоранговой части (без учета плотных блоков) матрицы $\hat{R}_{ij}$ на вектор в сжатом формате применима та же оценка, что и для матрицы R_{ij}, и это число операций также оценивается как $\mathbb{O}(rn\log n)$ [15].

Таким образом, в случае если плотные блоки имеют малые размеры, можно надеяться на выполнение оценки $\mathbb{O}(rn\log^2 n)$ как для числа операций по малоранговому разложению матриц $\hat{R}_{ij}$, так и для числа операций при выполнении каждой итерации по формулам (12).

Некоторые данные, подтверждающие справедливость таких оценок при практическом использовании алгоритма, приводятся в следующем разделе.

5. ПРИМЕРЫ РАСЧЕТОВ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В результате решения дискретных уравнений (12) находятся приближенные значения поверхностных токов на ячейках разбиения. Далее по найденным поверхностным токам можно рассчитать распределения электрического и магнитного полей в окружающем пространстве, используя формулы (6).

Для проверки работоспособности разработанной математической модели были проведены тестовые расчеты, в которых для двух тел были получены диаграммы обратного рассеяния, характеризующие поведение в дальней зоне отраженного электрического поля, возникающего при облучении плоской волной (2).

Эффективной площадью рассеяния в направлении единичного вектора $\boldsymbol{\tau}$ называется величина:

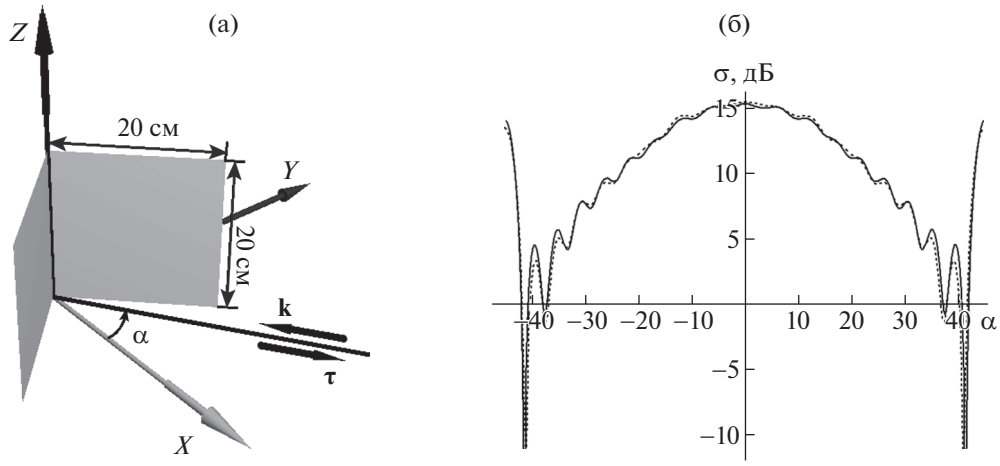
$$\sigma_0(\boldsymbol{\tau}) = \lim_{R\to\infty} 4\pi R^2 \frac{|\mathbf{E}(R\boldsymbol{\tau})|^2}{|\mathbf{E}_0|^2}, \tag{18}$$

где $\mathbf{E}_0$ – напряженность падающего электрического поля, определяемого формулой (2). Величина $\sigma_0(\boldsymbol{\tau})$ приближенно выражается через поверхностные токи по формуле:

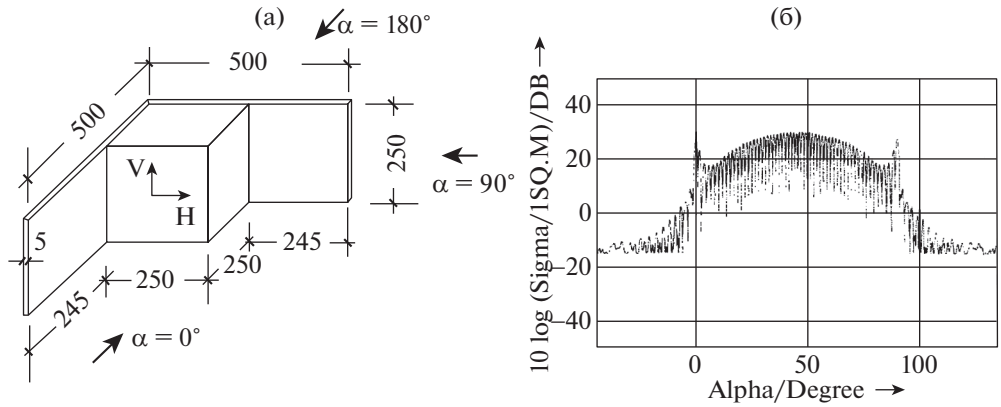
$$\sigma_0(\boldsymbol{\tau}) = \frac{4\pi}{|\mathbf{E}_0|^2}\left\{\frac{k^2}{\omega\varepsilon_0}\right\}^2 \left|\sum_{l=1}^{n} \exp(-ik(\boldsymbol{\tau}, y_l))\big(\mathbf{j}_l - \boldsymbol{\tau}(\mathbf{j}_l, \boldsymbol{\tau})\big)|\sigma_l|\right|^2, \quad \mathbf{j}_l = \mathbf{j}_{l,(m)}^{+} + \mathbf{j}_{l,(m)}^{-}, \tag{19}$$

где $|\sigma_l|$ – площадь ячейки σ_l.

В приводимых ниже примерах положение облучаемых тел рассматривалось в декартовой системе координат $Oxyz$, направление облучения определялось волновым вектором $\mathbf{k}$, причем вектор $\mathbf{k}$ брался параллельным плоскости Oxy в виде $\mathbf{k} = k(-\cos\alpha, -\sin\alpha), 0)$ (α – угол между вектором $-\mathbf{k}$ и осью Ox), k – волновое число (см. схему на фиг. 3а). Были получены диаграммы об-



Фиг. 3. Уголок. Схема объекта (а), диаграмма обратного рассеяния (б).



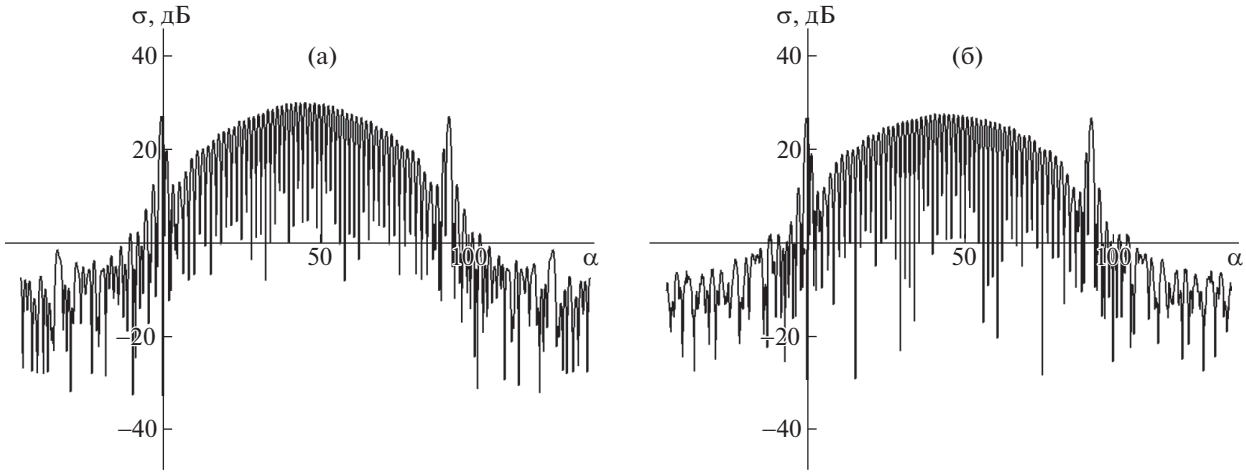
Фиг. 4. Двойной уголок. Схема объекта (а), диаграмма обратного рассеяния, эксперимент (б).

ратного рассеяния, показывающие зависимость эффективной площади рассеяния $\sigma_0(\boldsymbol{\tau})$ в направлении, противоположном направлению облучения ($\boldsymbol{\tau} = -\mathbf{k}/k$) от угла α в децибеллах:

$$\sigma(\alpha) = 10\lg\sigma_0(\boldsymbol{\tau}(\alpha)).$$

В первом примере рассматривалось рассеяние плоской волны на уголковом отражателе. Схема уголкового отражателя показана на фиг. 3а. На фиг. 3б приведены диаграммы обратного рассеяния для частоты $f = 10$ ГГц при Oxy поляризации (в формуле (2) вектор $\mathbf{E}_i$ компланарен плоскости Oxy). В проведенном расчете использовалось разбиение на квадратные ячейки, длина волны составляла $\lambda = 3$ см, число ячеек – 12 800, что соответствовало стороне ячейки 0.25 см (длина волны определяется соотношением $\lambda = c/f$, $c = \sqrt{\varepsilon_0\mu_0}$ – скорость света в вакууме). На графиках пунктирной линией изображена диаграмма, полученная авторами методом граничных интегральных уравнений (использовалась численная схема, описанная в статьях [9], [16], [17]), а сплошной линией – график, полученный методом физической оптики, разработанным в данной статье. Наблюдаемое хорошее согласование приведенных результатов свидетельствует о применимости разработанной итерационной математической модели физической оптики с переотражениями при решении задач рассеяния на невыпуклых телах.

Следующий пример для тела более сложной геометрии иллюстрирует работу алгоритма аппроксимации матриц. Рассматривался двойной уголок, схема которого приведена на фиг. 4 (рисунок из статьи [18]). На фиг. 4, 5 показаны диаграммы обратного рассеяния для частоты 16.66 ГГц (длина волны – 1.8 см). На фиг. 4б показаны данные физического эксперимента, которые были приведены в статье [18]. На фиг. 5 приведены диаграммы обратного рассеяния, по-



Фиг. 5. Диаграмма обратного рассеяния, двойной уголок. Расчет по методу интегральных уравнений (а), расчет по методу физической оптики (б).

лученные авторами методом граничных интегральных уравнений и разработанным методом физической оптики.

В расчетах методом интегральных уравнений и методом физической оптики использовалась поверхностная сетка с числом ячеек 97 240, форма ячеек прямоугольная (близкая к квадратной), максимальная сторона ячейки 0.227 см. Видно хорошее согласование результатов, полученных указанными численными методами и в физическом эксперименте.

Вычислительные затраты при решении задачи дифракции на двойном уголке:

$n = 1600,\ 6400,\ 25600,\ 102\,400;$

$h = 0.866,\ 0.433,\ 0.217,\ 0.108;$

$T_1 = 75.5,\ 512.8,\ 2828.9,\ 23\,459;$

$M_1 = 129,\ 650,\ 3673,\ 26\,730;$

$T_2 = 7.3,\ 19.7,\ 75.9,\ 453.1;$

$M_2 = 89,\ 269,\ 983,\ 3322.$

Здесь представлено:

n – число ячеек;

h – шаг сетки (в сантиметрах);

T_1 – время (в секундах) решения задачи методом интегральных уравнений с применением малоранговых аппроксимаций;

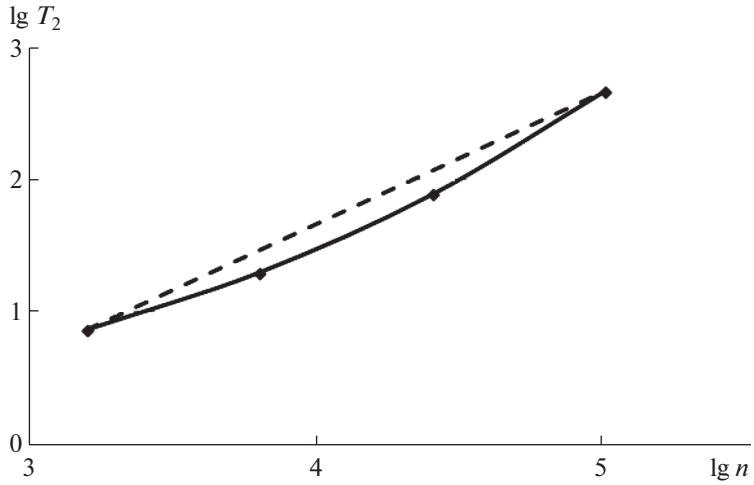
M_1 – память (в мегабайтах), необходимая для решения задачи методом интегральных уравнений с применением малоранговых аппроксимаций;

T_2 – время (в секундах) решения задачи методом физической оптики с применением малоранговых аппроксимаций;

M_2 – память (в мегабайтах), необходимая для решения задачи методом физической оптики с применением малоранговых аппроксимаций.

Рассмотренный пример задачи рассеяния электромагнитной волны на двойном уголке удобен для исследования метода мозаично-скелетных аппроксимаций: с одной стороны, решается задача, которая требует значительных вычислительных ресурсов, а с другой стороны, можно получить решение через интегральные уравнения.

Для оценки вычислительной эффективности разработанного численного алгоритма оценки зависимости числа операций (по порядку) от числа ячеек разбиения была проведена серия расчетов для того же двойного уголока при частоте облучения 4 ГГц (длина волны $h = 7.5$ см). Были использованы 4 расчетные сетки с числом ячеек от 1600 (максимальная сторона ячейки $h = 0.886$ см) до 102 400 (максимальная сторона ячейки 0.108 см). При этом на самой грубой сетке выполня-



Фиг. 6. Зависимость времени решения задачи от числа ячеек.

лось соотношение $h < 8\lambda$, на всех сетках были получены близкие диаграммы обратного рассеяния.

Выще приведены данные, характеризующие вычислительные затраты (время вычислений и затраты оперативной памяти) при применении разработанного алгоритма физической оптики, а также при применении метода граничных интегральных уравнений (использовался алгоритм авторов с применением мозаично-скелетонных аппроксимаций матриц [9], [16], [17]). Рассматривалась задача определения диаграммы обратного диаграммы рассеяния для 180 направлений облучения. Для каждого направления облучения решаются уравнения (13) с одними и теми же аппроксимированными заранее матрицами $\hat{R}[\sigma_k](x_i)$ вида (16) и со своей функцией $\mathbf{H}_0(x)$. В случае применения метода интегральных уравнений многократно решается система линейных уравнений с одной и той же матрицей и различными правыми частями. Расчеты проводились на персональном компьютере с процессором Intel Core-I5 с оперативной памятью 32 Гб.

Из данных видно, что для решения задачи методом физической оптики требуется существенно меньше вычислительных ресурсов, причем выигрыш и по используемой оперативной памяти, и по времени вычислений растет с увеличением числа ячеек.

Для оценки зависимости числа операций (по порядку) от числа ячеек разбиения на фиг. 6 приведен график зависимости $\lg T_2$ от $\lg n$ на основе вышеприведенных данных (сплошная линия). На фиг. 6 штриховой линией показана линейная зависимость, построенная по крайним точкам – эта зависимость имеет вид $\lg T_2 = A \lg n + B$, где $A \approx 0.992$. Тем самым наблюдается близкая к линейной зависимость времени вычислений от числа ячеек разбиения, что подтверждает гипотезу о теоретической оценке сложности алгоритма вида $\mathcal{O}(rn \log^2 n)$.

СПИСОК ЛИТЕРАТУРЫ

1. *Volakis J.L., Sertel K.* Integral equation methods for electromagnetics. SciTech Publishing, 2012. 406 p.
2. *Gibson W.* The method of moments in electromagnetics. Boca Raton: Chapman & Hall/CRC, 2008. 272 p.
3. *Уфимцев П.Я.* Теория дифракционных краевых волн в электродинамике. Введение в физическую теорию дифракции: Пер. с англ. М. : БИНОМ. Лаборатория знаний, 2012. 372 с.
4. *Уфимцев П.Я.* Метод краевых волн в физической теории дифракции. М.: Сов. Радио, 1962. 244 с.
5. *Francisco Saez de Adana* Practical Applications of Asymptotic Techniques in Electromagnetics. Artech House, 2011. 215 p.
6. *Tyrtyshnikov E.* Mosaic-skeleton approximations // Calcolo. 1996. V. 4. № 2. P. 1–3.
7. *Goreinov S.A., Tyrtyshnikov E.E., Zamarashkin N.L.* A theory pseudo-skeleton approximations // Linear Algebra and its Applications. 1997. V. 261. № 1–3. P. 1–21.
8. *Каширин А.А., Смагин С.И., Талтыкина М.Ю.* Применение мозаично-скелетонного метода при численном решении трехмерных задач Дирихле для уравнения Гельмгольца в интегральной форме // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 4. С. 625–638.

9. *Aparinov A., Setukha A., Stavtsev S.* Supercomputer modelling of electromagnetic wave scattering with boundary integral equation method // Communications in Computer and Information Science. 2017. V. 793. P. 325–336.
10. *Stavtsev S.L.* Application of the method of incomplete cross approximation to a nonstationary problem of vortex rings dynamics // RJNAMM. 2012. V. 27. № 3. P. 303–320.
11. *Хёнл Х., Мауэ А., Веспфаль К.* Теория дифракции: Пер. с нем. М.: Мир, 1964. 428 с.
12. *Colton D., KressR.* Integral Equation Methods in Scattering Theory. SIAM, 2013. 270 p.
13. *Захаров Е.В., Сетуха А.В., Безобразова Е.Н.* Метод гиперсингулярных интегральных уравнений в трехмерной задаче дифракции электромагнитных волн на кусочно-однородном диэлектрическом теле // Дифференц. ур-ния. 2015. Т. 51. № 9. С. 1206–1219.
14. *Замарашкин Н.Л., Осинский А.И.* Новые оценки точности псевдоскелетных аппроксимаций матриц // Докл. АН. 2016. Т. 471. № 3. С. 263–266.
15. *Tyrtyshnikov E.* Incomplete cross approximation in the mosaic-skeleton method // Computing. 2000. V. 64. P. 367–380.
16. *Aparinov A.A., Setukha A.V., Stavtsev S.L.* Parallel implementation for some applications of integral equations method // Lobachevskii Journal of Mathematics. 2018. V. 39. № 4. P. 477–485.
17. *Stavtsev S.* Low rank structures in solving electromagnetic problems // Lecture Notes in Computer Science. 2020. V. 11958. P. 165–172.
18. *Klement D., Preissner J., Stein V.* Special problems in applying the physical optics method for backscatter computations of complicated objects // IEEE Transactions on antennas and propagation. 1988. V. 36. № 2. P. 228–237.

EDN: RMYTKJ
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1473–1490

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 519.863

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ИНВЕСТИЦИЯМИ В КОЛЛЕКТИВНОЙ МОДЕЛИ ПЕНСИОННОГО СТРАХОВАНИЯ: ИССЛЕДОВАНИЕ СИНГУЛЯРНЫХ НЕЛИНЕЙНЫХ ЗАДАЧ ДЛЯ ИНТЕГРОДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

© 2022 г. Т. А. Белкина[1,*], Н. Б. Конюхова[2,], С. В. Курочкин[2,***]**

[1] *117418 Москва, Нахимовский пр-т, 47, ЦЭМИ РАН, Россия*

[2] *119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия*

**e-mail: tbel@cemi.rssi.ru*

***e-mail: nadja@ccas.ru*

****e-mail: kuroch@ccas.ru*

Поступила в редакцию 03.03.2022 г.
Переработанный вариант 26.03.2022 г.
Принята к публикации 11.05.2022 г.

Для коллективной модели пенсионного страхования (дуальной модели риска) рассматривается проблема оптимального управления инвестициями с целью максимизации вероятности неразорения страховой компании. Применение динамического программирования для поиска оптимальной стратегии приводит к сингулярным нелинейным краевым задачам для интегродифференциальных уравнений. В случае экспоненциального распределения размера премий даются аналитические исследования этих задач. Приводятся результаты расчетов и дается их сравнение с проведенными ранее расчетами для простых стратегий инвестиций (рисковых и безрисковых) в рассматриваемой модели. Библ. 13. Фиг. 3.

Ключевые слова: коллективная модель пенсионного страхования, вероятность неразорения страховой компании, оптимальное управление инвестициями, уравнение Беллмана, экспоненциальное распределение размера премий, нелинейные интегродифференциальные уравнения, сингулярные краевые задачи.

DOI: 10.31857/S0044466922090058

ВВЕДЕНИЕ

Данная работа относится к проблеме оптимального управления инвестициями в моделях страхования. В то же время, в определенном смысле, она является продолжением работ [1], [2], где рассматриваются стратегии управления определенного вида.

В [1] рассматривается задача о вероятности неразорения (ВНР) для коллективной модели пенсионного страхования (или так называемой модели аннуитета в страховании жизни, см. [3]) в условиях инвестирования всего резерва страховой компании (СК) или фиксированной его доли в рисковый актив, моделируемый геометрическим броуновским движением. Типичный договор страхования в данной модели – это договор пожизненной ренты, предполагающий пожизненное обеспечение страхователя в обмен на передачу права наследования его собственности в пользу СК. Коллективная модель страхового риска рассматривается как дуальная по отношению к классической модели Крамéра–Лундберга (о соотношении этих двух моделей см. [4]). В структуре процесса, описывающего динамику резерва СК, эта двойственность выражается, в частности, в замене на противоположные знаков как у случайной составляющей (составного пуассоновского процесса), так и у детерминированной, отвечающей в данном случае выплатам пенсий с постоянной интенсивностью. Если скачки процесса, имеющие отрицательные знаки в классической модели, означают выплаты по требованиям, то, взятые с положительным знаком, такие скачки (в дальнейшем будем называть их *премиями*) в дуальной модели соответствуют приростам резерва в моменты окончания жизни страхователей. (Другие ссылки на литературу с описанием этой модели и некоторыми результатами ее исследования можно также найти в [1], [2].) Для случая экспоненциального распределения размеров скачков в [1] показано, что ВНР как функция

начального капитала (НК), определенная на неотрицательной вещественной полуоси, является решением сингулярной краевой задачи (КЗ) для интегродифференциального уравнения (ИДУ) с невольтерровым интегральным оператором. Приводится доказательство существования и единственности решения этой задачи, получены асимптотические представления для ВНР при малых и больших значениях НК, предложен эффективный алгоритм численного нахождения решения, проведены расчеты и дана их экономическая интерпретация: показано, что в пенсионном страховании вложение средств в рисковые активы играет важную роль для увеличения платежеспособности компании при малых значениях НК.

В [2] изучается та же коллективная модель пенсионного страхования (пожизненной ренты), но с учетом безрисковой стратегии инвестиций, т.е. в случае, когда весь резерв СК в каждый момент времени инвестируется в безрисковый актив (банковский счет). Дается сравнение этой стратегии с изученными в [1] рисковыми стратегиями инвестиций, при которых, независимо от размера резерва СК, в каждый момент времени постоянную положительную долю этого резерва составляют рисковые активы (акции), а оставшаяся доля инвестируется в банковский счет (такие стратегии называем *простыми рисковыми стратегиями*). Сравнение стратегий основывается на традиционном критерии платежеспособности – ВНР. При экспоненциальном распределении размеров скачков в [2] показано, что в случае безрисковых инвестиций ВНР как функция НК, определенная на неотрицательной вещественной полуоси, является решением сингулярной задачи для ИДУ с невольтерровым интегральным оператором. Получено решение поставленной задачи, проведено аналитическое исследование его свойств, приводятся численные примеры. На примерах дается сравнение влияния рисковых и безрисковых инвестиций на ВНР в данной модели. В частности, показано, что, в то время как при достаточно больших значениях НК (точнее, при значениях, не меньших некоторого числа, определяемого параметрами модели) полное вложение резерва в безрисковый актив делает разорение невозможным, при малых значениях НК применение рисковых инвестиций может увеличивать ВНР в большей мере, чем применение только безрискового актива.

Проведенное в [2] сравнение простых рисковых и безрисковой стратегий инвестиций демонстрирует актуальность проблемы оптимального динамического управления долей рискового актива в инвестиционном портфеле страховщика с целью максимизации ВНР по коллективному портфелю договоров. Аналогичная проблема, рассматриваемая в рамках динамической модели коллективного риска, впервые изучалась в [5] в случае, когда динамика страхового резерва описывается винеровским процессом со сносом (так называемая диффузионная модель риска, являющаяся по сути диффузионной аппроксимацией классической модели Крамéра–Лундберга). Для самóй классической модели оптимальное управление инвестициями было позже исследовано в [6], [7]. Во всех указанных работах рисковый актив моделировался геометрическим броуновским движением. В обоих случаях (и для классической модели, и для ее диффузионной аппроксимации) из полученных результатов следовало, что при малых значениях резерва оптимальная стратегия инвестиций требует заимствования денежных средств при заданной процентной ставке для вложения их в рисковый актив. Это привело к дальнейшим исследованиям, предполагающим ограничения на заимствования в рассматриваемой модели: например, в [8] изучался случай, когда доля рискового актива в портфеле может быть неотрицательной и не превосходящей некоторой величины. Введение же в ограничения возможности принятия отрицательных значений для доли рискового актива, т.е. возможности использования коротких позиций в акциях, привело к парадоксальному на первый взгляд результату, полученному в [9]: при сильном ограничении на заимствования и малых значениях резерва оптимальная стратегия может включать короткие позиции, в то время как в модели без каких-либо ограничений коротких позиций в оптимальной стратегии никогда не возникает.

Целью данной работы является исследование оптимальной стратегии инвестиций в описанной выше коллективной модели пенсионного страхования при отсутствии ограничений на заимствования. Специфика данной модели состоит в том, что, как уже было сказано выше, полное вложение резерва в безрисковый актив при условии, что его значение не меньше некоторой фиксированной величины, определяемой параметрами модели, обеспечивает неразорение с вероятностью единица, определяя, следовательно, оптимальную стратегию при этих значениях резерва, и задача сводится к поиску оптимальной стратегии на некотором конечном отрезке значений резерва. Тем не менее функция Беллмана данной задачи, или оптимальная ВНР как функция НК, от которой зависит оптимальная стратегия, если она существует, должна являться решением некоторого нелинейного ИДУ второго порядка, определенного на всей неотрицательной полуоси (точнее, она должна удовлетворять этому ИДУ всюду, за исключением, быть может, одной

точки, являющейся правой границей указанного конечного отрезка). Исследование сингулярной нелинейной задачи для ИДУ в модели, рассматриваемой в данной работе, позволяет изучить свойства оптимальной стратегии и соответствующей ей ВНР.

Далее, в частности, используются обозначения: $\mathbf{P}(A)$ – вероятность события A; $\mathbf{E}X$ – математическое ожидание случайной величины X. Остальные обозначения и сокращения будут вводиться по мере необходимости.

1. ОПИСАНИЕ МОДЕЛИ И ПОСТАНОВКА ЗАДАЧИ. НЕКОТОРЫЕ ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

1.1. Исходный процесс риска для модели без инвестиций

Приведем здесь краткое описание модели пенсионного страхования, подробно описанной в [1]. Данная модель представляется процессом динамики капитала СК по портфелю договоров пожизненной ренты (процессом риска), имеющим вид

$$R_t = u - ct + \sum_{k=1}^{N(t)} Z_k, \quad t \geq 0. \tag{1.1}$$

Здесь R_t – размер капитала в момент времени t, $t \geq 0$; u – размер НК, $0 < c$ – размер затрат на выплату пенсий в единицу времени, $N(t)$ – однородный пуассоновский процесс с интенсивностью $\lambda > 0$ ($\mathbf{E}N(t) = \lambda t$, $N(0) = 0$), описывающий количество поступивших премий к моменту времени t; Z_k ($k = 1, 2, \ldots$) – размеры премий – независимые одинаково распределенные (невырожденные) неотрицательные случайные величины (СВ), не зависящие от процесса $N(t)$ и имеющие функцию распределения (ФР) $F(z)$ такую, что

$$F(0) = 0, \quad \int_0^\infty z dF(z) = m, \quad 0 < m < \infty. \tag{1.2}$$

Будем также предполагать, что $F(z)$ имеет непрерывную плотность с носителем $(0, \infty)$. В случае экспоненциального распределения размеров премий будет

$$F(z) = 1 - \exp(-z/m), \quad m > 0. \tag{1.3}$$

1.2. Модель, включающая простые стратегии инвестирования. Постановка задачи оптимального управления инвестициями

Прежде чем сформулировать оптимизационную задачу максимизации ВНР с помощью управления инвестициями, опишем так называемые простые стратегии инвестирования. Простые стратегии, включающие два вида активов – рисковый (акции) и безрисковый (банковский счет), устроены таким образом, что с течением времени доля капитала, вложенного в акции (положительная или нулевая), остается постоянной. При меняющихся ценах активов это соответствует постоянной перестройке финансового портфеля. Если доля акций в этом портфеле нулевая, то весь капитал фонда непрерывно инвестируется в безрисковый актив; при постоянной процентной ставке $r > 0$ эволюция этого актива описывается обыкновенным дифференциальным уравнением (ОДУ)

$$dB_t = rB_t dt, \quad t \geq 0, \tag{1.4}$$

где B_t – величина банковского счета в момент времени t. В этом случае динамика капитала компании X_t описывается стохастическим дифференциальным уравнением (СДУ)

$$dX_t = rX_t dt + dR_t, \quad t \geq 0, \tag{1.5}$$

с начальным условием $X_0 = u$.

Процесс (1.5) может интерпретироваться как соответствующий “вырожденному” случаю при рассмотрении простых инвестиционных стратегий, включающих рисковый актив, влияние которых на ВНР для изучаемой здесь модели исследовалось в [1]. Пусть динамика цены рискового актива моделируется процессом геометрического броуновского движения

$$dS_t = \mu S_t dt + \sigma S_t dw_t, \quad t \geq 0,$$

где S_t – цена акции в момент времени t, μ – ожидаемая доходность, $\mu > r$, $0 < \sigma$ – волатильность, w_t – стандартное броуновское движение. При указанных предположениях относительно простых стратегий соответствующий процесс риска удовлетворяет СДУ

$$dX_t = (\mu - r)\alpha X_t dt + rX_t dt + \sigma\alpha X_t dw_t + dR_t, \quad t \geq 0, \tag{1.6}$$

с начальным условием $X_0 = u$, где α – доля рискового актива в портфеле, $0 \leq \alpha \leq 1$.

В случае $\alpha = 0$ СДУ (1.6) приобретает вид (1.5) и соответствует полному вложению капитала в безрисковый актив. *Рисковыми простыми стратегиями* инвестиций (или, для краткости, "рисковыми инвестициями") называем стратегии, соответствующие случаю $0 < \alpha \leq 1$, а при $\alpha = 0$ стратегию называем *безрисковой* (исследование безрисковой стратегии в данной модели и, в частности, ее сравнение с простыми рисковыми стратегиями в смысле их влияния на ВНР проводилось в [2]).

ВНР на бесконечном интервале времени для процесса X_t (как функция от НК u) определяется следующим образом: $\varphi(u) = \mathbf{P}(X_t \geq 0,\ t \geq 0 | X_0 = u)$.

Напомним (см. [1]), что инфинитезимальный оператор, соответствующий однородному марковскому процессу $X_t = X_t^\alpha$, удовлетворяющему СДУ (1.6), имеет вид

$$(\mathscr{A}^\alpha f)(u) = \frac{1}{2}\sigma_\alpha^2 u^2 f''(u) + f'(u)(\mu_\alpha u - c) - \lambda f(u) + \lambda\int_0^\infty f(u+z)dF(z), \quad u > 0, \tag{1.7}$$

где

$$\mu_\alpha = \alpha\mu + (1-\alpha)r > 0, \quad \sigma_\alpha = \alpha\sigma \geq 0. \tag{1.8}$$

При $0 < \alpha \leq 1$ в качестве области определения оператора рассматриваем некоторый класс функций $f(u)$ в пространстве $\mathscr{C}^2(\mathbb{R}_+)$ дважды непрерывно дифференцируемых на $(0,\infty)$ функций.

В [1] для случая экспоненциальной ФР $F(z)$ вида (1.3) показано, что ВНР процесса X_t^α как функция НК при $\alpha > 0$ является дважды непрерывно дифференцируемой на $(0,\infty)$ функцией и удовлетворяет уравнению

$$(\mathscr{A}^\alpha f)(u) = 0, \quad u > 0, \tag{1.9}$$

имеющему (при замене f на φ) вид линейного ИДУ

$$\frac{1}{2}\sigma_\alpha^2 u^2 \varphi''(u) + (\mu_\alpha u - c)\varphi'(u) - \lambda\varphi(u) + \lambda\int_0^\infty \varphi(u+z)dF(z) = 0, \quad u > 0. \tag{1.10}$$

В данной работе вместо постоянной доли α рисковых активов в портфеле в каждый момент времени t рассматривается доля α_t рисковых активов как случайная величина, зависящая от предыдущей информации и определяющая управление в момент времени t. В целом же *допустимым* (неупреждающим) *управлением* в данной модели будем называть случайный процесс $\pi = \{\alpha_t\}_{t\geq 0}$, предсказуемый относительно естественной фильтрации, порожденной парой процессов $\{R_t, w_t\}$. Множество всех допустимых управлений будем обозначать через Π.

Тогда динамика капитала X_t^π при управлении $\pi \in \Pi$ будет определяться СДУ в виде

$$dX_t^\pi = (\mu - r)\alpha_t X_t^\pi dt + rX_t^\pi dt + \sigma\alpha_t X_t^\pi dw_t + dR_t, \quad t \geq 0. \tag{1.11}$$

ВНР процесса X_t^π с начальным условием $X_0^\pi = u$ будем обозначать через $\varphi^\pi(u)$. Ясно, что процесс (1.6) является частным случаем процесса (1.11) при управлении $\pi = \pi_\alpha$, для которого $\alpha_t \equiv \alpha$. Будем рассматривать задачу поиска оптимального управления $\pi^* = \{\alpha_t^*\}_{t\geq 0} \in \Pi$, которое, возможно, может быть задано оптимальной стратегией (функцией от текущего капитала), если она существует. Оптимальность при этом понимается как максимизация ВНР на бесконечном интервале времени:

$$\varphi^*(u) = \sup_{\pi\in\Pi} \varphi^\pi(u) \quad \forall u > 0. \tag{1.12}$$

1.3. Уравнение Гамильтона–Якоби–Беллмана и некоторые дополнительные условия

Применение метода динамического программирования (см., например, [10]), наряду с формулой полной вероятности и формулой Ито, позволяют осуществить эвристический (т.е. без априорного доказательства дифференцируемости) вывод ИДУ для функции Беллмана $V(u)$ рассматриваемой оптимизационной задачи (уравнения Гамильтона–Якоби–Беллмана, которое для краткости далее будем называть *уравнением Беллмана*; по отношению к задачам максимизации ВНР о выводе такого уравнения см., в частности, [6]). В данном случае это уравнение имеет вид

$$\sup_{\alpha\in\mathbb{R}}(\mathscr{A}^{\alpha}V)(u) = 0, \tag{1.13}$$

где $\mathscr{A}^{\alpha}$ определено в (1.7). В дальнейшем будут доказаны утверждения о существовании решения $V(u)$ уравнения (1.13), обладающего следующими свойствами. При $r > 0$ функция $V(u)$ является, по крайней мере, дважды непрерывно дифференцируемой на интервалах $(0, c/r)$, $(c/r, \infty)$ и удовлетворяет уравнению (1.13) всюду, за исключением, быть может, точки c/r, в которой она может не иметь производной. Кроме того, $V(u)$ удовлетворяет условиям

$$V(0) = 0, \quad V(u) = 1, \quad u \geq c/r. \tag{1.14}$$

При $r = 0$ уравнение (1.13) имеет дважды непрерывно дифференцируемое решение $V(u)$ на интервале $[0, \infty)$, удовлетворяющее условиям

$$V(0) = 0, \quad \lim_{u\to\infty} V(u) = 1. \tag{1.15}$$

Окончательный ответ на вопрос о виде и свойствах функции $\varphi^*(u)$ (соответствующей оптимальному управлению в том случае, если оно существует, и тогда супремум в (1.12) достигается) должно дать утверждение о том, что решение $V(u)$ уравнения (1.13), удовлетворяющее вышеописанным условиям, определяет функцию $\varphi^*(u)$ в соответствующей оптимизационной задаче (при $r > 0$ или $r = 0$), т.е. $V(u) \equiv \varphi^*(u)$.

Приведем здесь обоснование сформулированных выше условий на решение $V(u)$, показав, что они вытекают из "физических" свойств функции $\varphi^*(u)$ задачи (1.12). В частности, нулевое начальное условие в обоих случаях возникает в силу наличия отрицательной детерминированной неоднородной составляющей $(-cdt)$ в уравнении (1.11) (напомним, что R_t определено в (1.1)) для процесса X_t^{π}, моментально выводящей его в отрицательную область при нулевом начальном значении и при любом допустимом управлении π (для строгого доказательства в частном случае, а именно, при простых стратегиях, см. лемму 1 в [1]; данное доказательство легко распространяется на случай любого допустимого управления). Следовательно, $\varphi^*(0) = 0$, что объясняет первые условия в (1.14) и (1.15).

Чтобы обосновать вторые условия в (1.14) и (1.15), достаточно заметить, что функция $\varphi^*(u)$ является неубывающей (это нетрудно показать, исходя из ее определения), а также воспользоваться результатами [1], [2] относительно ВНР при использовании простых стратегий. Ясно, что из (1.12), в частности, следует, что для любого $\alpha \geq 0$ и $u > 0$ имеет место неравенство

$$\varphi^*(u) \geq \varphi^{\pi_\alpha}(u), \tag{1.16}$$

где $\varphi^{\pi_\alpha}(u)$ – ВНР для процесса вида (1.6), т.е. процесса (1.11), с начальным состоянием $X_0 = u$, соответствующего управлению π_α. Тогда, так как в случае $\alpha = 0$ и $r > 0$ имеем $\varphi^{\pi_0}(u) = 1$ при $u \geq c/r$ (см. лемму 2 в [2]), то очевидно, что этому же свойству удовлетворяет и функция $\varphi^*(u)$. Рассматривая теперь в случае $r = 0$ положительные α в неравенстве (1.16), выберем такое α, при котором $2\mu_\alpha/\sigma_\alpha^2 > 1$, где μ_α, σ_α^2 определены в (1.8), т.е. $\alpha < 2\mu/\sigma^2$. Для любого положительного α, удовлетворяющего последнему неравенству, инвестиционный портфель, соответствующий простой стратегии, т.е. управлению π_α, становятся "надежным", и из результатов [1] (см. там теоремы 1 и 4) следует, что ВНР $\varphi^{\pi_\alpha}(u)$ процесса (1.11) с начальным состоянием $X_0 = u$ удовлетворяет условию $\lim_{u\to\infty} \varphi^{\pi_\alpha}(u) = 1$, а следовательно, этому условию удовлетворяет и функция $\varphi^*(u)$.

2. АНАЛИЗ И ЧИСЛЕННОЕ РЕШЕНИЕ УРАВНЕНИЯ БЕЛЛМАНА ДЛЯ ЗАДАЧИ ОПТИМИЗАЦИИ ПРИ $r > 0$

Будем исследовать уравнение (1.13) при $u > 0$. С учетом (1.10), (1.8) перепишем его в виде

$$\sup_{\alpha\in\mathbb{R}}\left\{\frac{1}{2}\alpha^2\sigma^2u^2V''(u) + [(\alpha\mu + (1-\alpha)r)u - c]V'(u) - \lambda V(u) + \lambda\int_0^\infty V(u+z)dF(z)\right\} = 0. \tag{2.1}$$

Предположим, что существует решение $V(u)$ уравнения Беллмана (1.13) с условиями (1.14), непрерывно дифференцируемое на $\mathbb{R}_+$, дважды непрерывно дифференцируемое на интервале $(0, c/r)$ и удовлетворяющее требованиям

$$V'(u) > 0, \quad V''(u) < 0, \quad u \in (0, c/r). \tag{2.2}$$

(Заметим, что при неотрицательности второй производной функции $V(u)$ в некоторой точке $u \in (0, c/r)$ супремум в (2.1) не достигается.) Найдем значение α^*, доставляющее экстремум в (2.1) в предположении (2.2) для $u \in (0, c/r)$:

$$\alpha^* = \alpha_V^*(u) := -\frac{(\mu - r)V'(u)}{\sigma^2uV''(u)}. \tag{2.3}$$

Подставляя (2.3) в (2.1), получаем нелинейное ИДУ:

$$(ru - c)V'(u) - \lambda V(u) + \lambda\int_0^\infty V(u+z)dF(z) = \frac{(\mu - r)^2(V'(u))^2}{2\sigma^2V''(u)}, \quad u \in (0, c/r); \tag{2.4}$$

при $u > c/r$, с учетом (1.14), это уравнение также удовлетворяется. Пусть $H(u) = 1 - F(u)$ (дополнительная функция распределения СВ одной премии Z_k (см. (1.1))). Тогда, с учетом условий $H(0) = 1$, $\lim_{u\to\infty} H(u) = 0$ перепишем ИДУ (2.4) в виде:

$$(ru - c)V'(u) + \lambda\int_0^\infty H(z)V'(u+z)dz = \frac{(\mu - r)^2(V'(u))^2}{2\sigma^2V''(u)}, \quad u \in (0, c/r). \tag{2.5}$$

Из условий (1.14), в частности, получаем

$$V'(u) = 0, \quad u > c/r. \tag{2.6}$$

Тогда для $u \in (0, c/r)$ справедливы равенства

$$\int_0^\infty H(z)V'(u+z)dz = \int_0^{c/r-u} H(z)V'(u+z)dz = \int_u^{c/r} H(y-u)V'(y)dy. \tag{2.7}$$

В результате ИДУ (2.5) приобретает вид:

$$(ru - c)V'(u) + \lambda\int_u^{c/r} H(y-u)V'(y)dy = \frac{(\mu - r)^2(V'(u))^2}{2\sigma^2V''(u)}, \quad u \in (0, c/r). \tag{2.8}$$

Обозначив $V'(u) = v(u)$, получаем из (2.8) нелинейное ИДУ для $v(u)$:

$$(ru - c)v(u) + \lambda\int_u^{c/r} H(y-u)v(y)dy = \frac{(\mu - r)^2v^2(u)}{2\sigma^2v'(u)}, \quad u \in (0, c/r). \tag{2.9}$$

Будем искать гладкое на $(0, c/r)$ решение ИДУ (2.9) с условием

$$v(c/r) = 0, \tag{2.10}$$

которое должно выполняться в силу предположения о гладкости функции $V(u)$ на $\mathbb{R}_+$ и условия (2.6).

2.1. Поведение решения $v(u)$ и его производной при $u \uparrow (c/r)$

Пусть $v(u)$ – решение задачи (2.9), (2.10). Изучим поведение $v(u)$ при $u \uparrow (c/r)$, учитывая, что

$$H(z) = 1 + o(1), \quad z \to +0 \tag{2.11}$$

(в предположении, что $F(z)$ непрерывна в нуле).

При $u \uparrow (c/r)$ ищем $v(u)$ в виде:

$$v(u) = \gamma(c/r - u)^{\beta}(1 + o(1)), \quad v'(u) = -\gamma\beta(c/r - u)^{\beta-1}(1 + o(1)), \quad \beta, \gamma > 0. \tag{2.12}$$

Подставляя (2.12) в (2.9), находим, что коэффициент γ не определяется, а коэффициент β – положительный корень квадратного уравнения

$$r\beta^2 + (r - \lambda - (\mu - r)^2/(2\sigma^2))\beta - (\mu - r)^2/(2\sigma^2) = 0, \tag{2.13}$$

т.е.

$$\beta = \left[-\left(r - \lambda - \frac{(\mu - r)^2}{2\sigma^2}\right) + \sqrt{\left(r - \lambda - \frac{(\mu - r)^2}{2\sigma^2}\right)^2 + 2r\frac{(\mu - r)^2}{\sigma^2}}\right]/(2r). \tag{2.14}$$

Замечание 1. Нетрудно проверить, что если в (2.9) положить $H(z) \equiv 1$, то функция $v(u) = \gamma(c/r - u)^{\beta}$ при β вида (2.14) и любом γ окажется точным решением ИДУ (2.9).

Наряду с положительным корнем β уравнения (2.13) будем рассматривать его отрицательный корень, который обозначим через β_-. Тогда по теореме Виета получаем равенства:

$$\beta + \beta_- = (\mu - r)^2/(2r\sigma^2) + \lambda/r - 1, \quad \beta\beta_- = -(\mu - r)^2/(2r\sigma^2). \tag{2.15}$$

Рассмотрим случаи различных соотношений между параметрами в (2.15).

1. Пусть $\lambda \geq 2r$. Тогда из (2.15) получаем, что $\beta + \beta_- > 1$, т.е. $\beta > 1$. В результате из (2.12) следует справедливость предельных равенств

$$\lim_{u \to c/r - 0} v(u) = \lim_{u \to c/r - 0} v'(u) = 0, \tag{2.16}$$

т.е. $V(u)$ (напомним, что функция $v(u)$ является производной функции $V(u)$) дважды непрерывно дифференцируема в точке $u = c/r$.

2. Пусть $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) > 2 - \lambda/r$. Покажем, что $\beta > 1$. Из (2.15) получаем

$$\beta + \beta_- > \lambda/(2r), \quad \beta\beta_- < \lambda/(2r) - 1. \tag{2.17}$$

Теперь предположим противное, т.е. что $\beta \leq 1$. Тогда из первого неравенства в (2.17) получим

$$\beta_- > \lambda/(2r) - 1, \tag{2.18}$$

что несовместимо со вторым неравенством в (2.17).

Следовательно, $\beta > 1$, и снова имеем, что $V(u)$ дважды непрерывно дифференцируема в точке c/r.

3. Пусть $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) = 2 - \lambda/r$. Тогда будет

$$\beta + \beta_- = \lambda/(2r), \quad \beta\beta_- = \lambda/(2r) - 1, \tag{2.19}$$

что влечет соотношение $\beta(\lambda/(2r) - \beta) = \lambda/(2r) - 1$, откуда следует, что $\beta = 1$.

В результате из (2.12) следует справедливость предельных равенств

$$\lim_{u \to c/r - 0} v(u) = 0, \quad \lim_{u \to c/r - 0} v'(u) = -\gamma, \quad \gamma > 0, \tag{2.20}$$

т.е. $V(u)$ непрерывно дифференцируема в точке c/r, но ее вторая производная в этой точке разрывна.

4. Пусть, наконец, $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) < 2 - \lambda/r$. Покажем, что $\beta < 1$. Из (2.15) получаем

$$\beta + \beta_- < \lambda/(2r), \quad \beta\beta_- > \lambda/(2r) - 1. \tag{2.21}$$

Теперь предположим противное, т.е. что $\beta \geq 1$. Тогда из первого неравенства в (2.21) получим

$$\beta_- < \lambda/(2r) - 1, \tag{2.22}$$

что несовместимо со вторым неравенством в (2.21).

Следовательно, $\beta < 1$, и справедливы предельные соотношения

$$\lim_{u \to c/r-0} v(u) = 0, \qquad \lim_{u \to c/r-0} v'(u) = -\infty, \tag{2.23}$$

т.е. вновь, хотя функция $V(u)$ непрерывно дифференцируема в точке c/r, но ее вторая производная терпит разрыв, на этот раз бесконечный.

Будем далее также интересоваться асимптотическим поведением при $u \uparrow (c/r)$ функции

$$A(u) = u\alpha_V^*(u) = -(\mu - r)v(u)/(\sigma^2 v'(u)). \tag{2.24}$$

Учитывая формулы (2.12), получаем

$$A(u) = [(\mu - r)/(\beta\sigma^2)](c/r - u)(1 + o(1)), \quad u \uparrow (c/r). \tag{2.25}$$

2.2. Случай экспоненциального распределения размера премий

Рассмотрим случай экспоненциального распределения размера премий, т.е. при $F(z)$ вида (1.3). Тогда

$$H(z) = 1 - F(z) = \exp(-kz), \quad k = 1/m > 0. \tag{2.26}$$

Полагая $w(u) = v(u)\exp(-ku)$, получаем

$$v(u) = w(u)\exp(ku), \quad v'(u) = \exp(ku)[w'(u) + kw(u)]. \tag{2.27}$$

Подставляя формулы (2.27) в ИДУ (2.9), получаем

$$(ru - c)w(u) + \lambda\int_u^{c/r} w(y)dy = \frac{(\mu - r)^2 w^2(u)}{2\sigma^2[w'(u) + kw(u)]}, \quad u \in (0, c/r). \tag{2.28}$$

Получим теперь из (2.28) уравнение для функции $A(u)$. Учитывая (2.24) и соотношения (2.27), получаем

$$A(u) = -\frac{(\mu - r)w(u)}{\sigma^2[w'(u) + kw(u)]}. \tag{2.29}$$

Тогда ИДУ (2.28) преобразуется к виду

$$(ru - c)w(u) + \lambda\int_u^{c/r} w(y)dy = -[(\mu - r)/2]A(u)w(u), \quad u \in (0, c/r). \tag{2.30}$$

Продифференцируем это соотношение по u:

$$rw(u) + (ru - c)w'(u) - \lambda w(u) = -[(\mu - r)/2][A'(u)w(u) + A(u)w'(u)]. \tag{2.31}$$

Разделив обе части соотношения (2.31) на $w(u)$ и учитывая, что

$$w'(u)/w(u) = -[(\mu - r)/(\sigma^2 A(u)) + k],$$

получаем нелинейное ОДУ для $A(u)$:

$$-(r - \lambda) + (ru - c)\left[\frac{\mu - r}{\sigma^2 A(u)} + k\right] = \frac{\mu - r}{2}\left[A'(u) - kA(u) - \frac{\mu - r}{\sigma^2}\right], \tag{2.32}$$

или, что то же, в виде:

$$A'(u)A(u) = \frac{A^2(u)}{m} + \left[\frac{2(ru - c)}{(\mu - r)m} - \frac{2(r - \lambda)}{\mu - r} + \frac{\mu - r}{\sigma^2}\right]A(u) + \frac{2(ru - c)}{\sigma^2}. \tag{2.33}$$

В результате будем изучать поведение решений ОДУ (2.33) на интервале $[0, c/r]$ с учетом представления (2.25), где β определено в (2.14). В частности, интерес представляют значения $A(+0)$ и

поведение $\alpha_V^*(u)$ вблизи нуля, где $\alpha_V^*(u) = A(u)/u$. Кроме того, для поиска функции $V(u)$ может быть сформулирована следующая

Задача 1. Зная $A(u), u \in [0, c/r]$ ($A(u) = 0, u \geq c/r$), найти функцию $V(u)$ такую, что выполнены уравнение

$$A(u) = -(\mu - r)V'(u)/(\sigma^2 V''(u)), \quad u \in (0, c/r), \tag{2.34}$$

и условия (1.14).

Функция $V(u)$, являющаяся решением задачи 1, как будет показано далее с помощью проверочной теоремы, определяет функцию Беллмана в задаче минимизации ВНР в рассматриваемой модели, при этом соответствующая оптимальная стратегия имеет вид

$$\alpha_V^*(u) = A(u)/u, \quad u > 0, \tag{2.35}$$

а функция $A(u)$ определяет оптимальное количество денежных средств, инвестируемых в рисковый актив при текущем значении резерва u для всех $u > 0$.

2.2.1. Алгоритм численного нахождения решений. Обозначим $\tilde{v}(u) = v(u)/\gamma$, $\tilde{V}(u) = V(u)/\gamma$, $B(u) = A^2(u)$. Тогда, в частности, получим из (2.12), (2.25):

$$\tilde{v}(u) = (c/r - u)^{\beta}(1 + o(1)), \quad u \uparrow (c/r), \tag{2.36}$$

$$B(u) = [(\mu - r)^2/(\beta^2\sigma^4)](c/r - u)^2(1 + o(1)), \quad u \uparrow (c/r), \tag{2.37}$$

где $\beta > 0$ определено в (2.14).

На интервале $[0, c/r]$ из соотношений (2.24), (2.33) и равенства $V'(u) = v(u)$ получаем систему из трех нелинейных ОДУ первого порядка для функций $\tilde{v}(u)$, $\tilde{V}(u)$ и $B(u)$:

$$\tilde{v}'(u)\sqrt{B(u)} = -(\mu - r)\tilde{v}(u)/\sigma^2, \tag{2.38}$$

$$B'(u) = \frac{2B(u)}{m} + 4\left[\frac{ru - c}{(\mu - r)m} - \frac{r - \lambda}{\mu - r} + \frac{\mu - r}{2\sigma^2}\right]\sqrt{B(u)} + \frac{4(ru - c)}{\sigma^2}, \tag{2.39}$$

$$\begin{gathered} \tilde{V}'(u) = \tilde{v}(u), \\ u \in (0, c/r); \end{gathered} \tag{2.40}$$

при этом, очевидно, должны выполняться условия

$$\tilde{v}(c/r) = 0, \quad B(c/r) = 0, \tag{2.41}$$

$$\tilde{V}(0) = 0, \tag{2.42}$$

$$\tilde{V}(c/r) = 1/\gamma, \tag{2.43}$$

где параметр γ пока неизвестен и будет определен в результате расчетов. Для численного нахождения решений систему ОДУ (2.38)–(2.40) будем рассматривать на интервале $[0, \hat{u}]$, где $0 < \hat{u} < c/r$, $c/r - \hat{u}$ мало (по причине вхождения в ОДУ (2.38) и (2.39) величины $\sqrt{B(u)}$, которая вблизи значения $u = c/r$ будет вычисляться неустойчиво). При этом вместо условий (2.41), используя представления (2.36), (2.37), получаем приближенные условия в точке $\hat{u}$:

$$\tilde{v}(\hat{u}) \approx (c/r - \hat{u})^{\beta}, \quad B(\hat{u}) \approx [(\mu - r)^2/(\beta^2\sigma^4)](c/r - \hat{u})^2. \tag{2.44}$$

Вычислительный алгоритм состоит из следующих шагов.

Шаг 1. Прежде всего находим функцию $A(u)$, учитывая ее связь с $B(u)$ и решая справа налево для системы из двух нелинейных ОДУ (2.38), (2.39) задачу Коши (ЗК) от точки $u = \hat{u}$ до точки $u = 0$ с начальными условиями (2.44). Получаем, в частности, значения

$$\tilde{v}(0) = \tilde{v}_0, \quad B(0) = B_0. \tag{2.45}$$

Решение задачи 1 дают следующие 2 шага.

Шаг 2. Для системы из трех ОДУ (2.38)–(2.40) решаем слева направо ЗК от точки $u = 0$ до точки $u = \hat{u}$ с начальными условиями (2.45), (2.42).

Шаг 3. Приближенно продолжаем решение $\tilde{V}(u)$ из точки $u = \hat{u}$ в точку $u = c/r$, применив представление (2.36) с учетом того, что $\tilde{V}(u)$ является первообразной $\tilde{v}(u)$. Тем самым определим значение γ, исходя из условия (2.43). Тогда искомое решение $V(u)$, которое в силу (1.14) должно быть равно единице в точке c/r, определяется равенством

$$V(u) = \tilde{V}(u)/\tilde{V}(c/r), \quad u \in [0, c/r]. \tag{2.46}$$

2.2.2. Утверждение о существовании решения краевой задачи. Напомним, что в общем случае при $u \uparrow (c/r)$ получили для производных $V(u)$ соотношения (см. (2.12)):

$$V'(u) = \gamma(c/r - u)^{\beta}(1 + o(1)), \quad V''(u) = -\gamma\beta(c/r - u)^{\beta-1}(1 + o(1)), \quad \beta, \gamma > 0, \tag{2.47}$$

где β – положительный корень уравнения (2.13), определенный формулой (2.14), а γ – некоторая положительная постоянная, которая в случае экспоненциального распределения размера премий (т.е. при $F(z)$ вида (1.3)) может быть приближенно вычислена в соответствии с шагом 3 описанного выше алгоритма.

Далее заметим, что ИДУ (2.4) при $F(z)$ вида (1.3) приобретает вид

$$(ru - c)V'(u) - \lambda V(u) + \frac{\lambda}{m}\int_0^{\infty} V(u + z)\exp(-z/m)dz = \frac{(\mu - r)^2(V'(u))^2}{2\sigma^2 V''(u)}, \quad u \in (0, c/r), \tag{2.48}$$

а из результатов п. 2.2.1 и подразд. 2.1 следует

Теорема 1. *В случае $r > 0$ справедливы следующие утверждения: существует единственное непрерывно дифференцируемое на $(0, \infty)$ решение $V(u)$ сингулярной нелинейной КЗ* (2.48), (1.14) *для ИДУ, удовлетворяющее условиям* (2.2), *это решение однозначно определяется по алгоритму п.* 2.2.1 *и обладает следующими свойствами*: а) *при выполнении неравенства $\lambda \geq 2r$ или неравенств $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) > 2 - \lambda/r$ функция $V(u)$ дважды непрерывно дифференцируема*; б) *при выполнении условий $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) = 2 - \lambda/r$ вторая производная $V''(u)$ непрерывна на $(0, c/r)$ и $(c/r, \infty)$, но терпит конечный разрыв в точке $u = c/r$, имея ненулевой* (*отрицательный*) *предел при $u \to c/r - 0$*; в) *при выполнении неравенств $\lambda < 2r$ и $(\mu - r)^2/(r\sigma^2) < 2 - \lambda/r$ вторая производная $V''(u)$ также непрерывна на $(0, c/r)$ и $(c/r, \infty)$, при этом она терпит бесконечный разрыв в точке $u = c/r$, уходя в отрицательную бесконечность при $u \to c/r - 0$.*

3. АНАЛИЗ И ЧИСЛЕННОЕ РЕШЕНИЕ УРАВНЕНИЯ БЕЛЛМАНА В ЗАДАЧЕ ОПТИМИЗАЦИИ ПРИ $r = 0$

При $r = 0$ (когда, в частности, интервал $[0, c/r)$ переходит в интервал $[0, \infty)$) уравнение (2.1) принимает вид:

$$\sup_{\alpha \in \mathbb{R}}\left\{\frac{1}{2}\alpha^2\sigma^2 u^2 V''(u) + (\alpha\mu u - c)V'(u) - \lambda V(u) + \lambda\int_0^{\infty} V(u + z)dF(z)\right\} = 0, \quad u > 0. \tag{3.1}$$

Найдем значение α^*, доставляющее экстремум в (3.1) в предположении

$$V'(u) > 0, \quad V''(u) < 0, \quad u \in (0, \infty) \tag{3.2}$$

(в противном случае супремум не достигается):

$$\alpha^* = \alpha_V^*(u) := -\frac{\mu V'(u)}{\sigma^2 u V''(u)}, \quad u \in (0, \infty). \tag{3.3}$$

Подставляя (3.3) в (3.1), получаем нелинейное ИДУ

$$-cV'(u) - \lambda V(u) + \lambda\int_0^{\infty} V(u + z)dF(z) = \frac{\mu^2(V'(u))^2}{2\sigma^2 V''(u)}, \quad u \in (0, \infty), \tag{3.4}$$

с условиями (1.15).

3.1. Случай экспоненциального распределения размера премий

В случае экспоненциального распределения размера премий, т.е. при $F(z)$ вида (1.3), нелинейное ИДУ для $V(u)$ приобретает вид

$$-cV'(u) - \lambda V(u) + \frac{\lambda}{m}\int_0^\infty V(u+z)\exp(-z/m)dz = \frac{\mu^2(V'(u))^2}{2\sigma^2 V''(u)}, \quad u \in (0,\infty). \tag{3.5}$$

Получим уравнение для функции

$$A(u) = u\alpha_V^*(u) = -\mu V'(u)/(\sigma^2 V''(u)) \tag{3.6}$$

(аналогично тому, как это было сделано в случае $r > 0$). Это уравнение имеет вид

$$A'(u)A(u) = \frac{A^2(u)}{m} + \left(\frac{2\lambda}{\mu} + \frac{\mu}{\sigma^2} - \frac{2c}{\mu m}\right)A(u) - \frac{2c}{\sigma^2}, \quad u \in (0,\infty). \tag{3.7}$$

Заметим, что ОДУ (3.7) имеет частное решение $A(u) = A_{par}(u) \equiv a_0$, где a_0 – положительный корень квадратного уравнения

$$\frac{a_0^2}{m} + \left(\frac{2\lambda}{\mu} + \frac{\mu}{\sigma^2} - \frac{2c}{\mu m}\right)a_0 - \frac{2c}{\sigma^2} = 0 \tag{3.8}$$

(второй корень этого уравнения отрицательный).

Найдем функцию $V(u)$, соответствующую решению $A(u) = A_{par}(u) \equiv a_0$.

Из соотношения (3.6) имеем

$$(\ln V'(u))' = -\mu/(\sigma^2 A(u)), \quad u > 0. \tag{3.9}$$

Отсюда при постоянной $A(u) \equiv a_0$ и с учетом условий (1.15) получаем для функции Беллмана аналитическое решение:

$$V(u) = 1 - \exp(-\mu u/(\sigma^2 a_0)), \quad u \ge 0, \tag{3.10}$$

где

$$a_0 = -\left(\frac{2m\lambda}{\mu} + \frac{\mu m}{\sigma^2} - \frac{2c}{\mu}\right)/2 + \sqrt{\left(\frac{2m\lambda}{\mu} + \frac{\mu m}{\sigma^2} - \frac{2c}{\mu}\right)^2/4 + \frac{2cm}{\sigma^2}}. \tag{3.11}$$

Проверим, что функция (3.10) удовлетворяет также нелинейному ИДУ (3.5). Сначала вычислим интегральное слагаемое в (3.5):

$$\begin{aligned}\int_0^\infty V(u+z)\exp(-z/m)/m dz &= \int_0^\infty (1 - \exp(-\mu(u+z)/(\sigma^2 a_0)))\exp(-z/m)/m dz = \\ &= 1 - \exp(-\mu u/(\sigma^2 a_0))/(\mu m/(\sigma^2 a_0) + 1).\end{aligned} \tag{3.12}$$

При подстановке этого выражения и формул для $V(u)$, $V'(u)$, $V''(u)$ в ИДУ (3.5), получаем соотношение

$$-\frac{c\mu}{\sigma^2 a_0} - \frac{\lambda}{(\mu m/(\sigma^2 a_0) + 1)} + \lambda + \frac{\mu^2}{2\sigma^2} = 0, \tag{3.13}$$

которое легко преобразуется к уравнению (3.8) для a_0, т.е. ИДУ (3.5) удовлетворяется.

Таким образом, нами найдено подходящее решение сингулярной нелинейной КЗ (3.5), (1.15) для ИДУ. С использованием проверочных аргументов далее будет показано, что это решение определяет функцию Беллмана в задаче минимизации ВНР в рассматриваемой модели. При этом соответствующая оптимальная стратегия имеет вид

$$\alpha_V^*(u) = a_0/u, \quad u > 0, \tag{3.14}$$

где a_0 – положительный корень уравнения (3.8) (т.е. определяется формулой (3.11)), задающий оптимальное постоянное (не зависящее от размера резерва) количество денежных средств, инвестируемых в рисковый актив.

3.1.1. Сравнение с моделью без инвестиций. Сравним полученное выше решение $V(u)$ (3.10) с ВНР $\varphi(u)$ в модели без инвестиций [2]:

$$\varphi(u) = 1 - \exp(-(\lambda m - c)u/(cm)), \quad u \geq 0, \tag{3.15}$$

где $\lambda m > c$.

Надо показать, что $(\lambda m - c)/(cm) < \mu/(\sigma^2 a_0)$. Предположим противное, т.е. что $(\lambda m - c)/(cm) \geq \mu/(\sigma^2 a_0)$. Тогда из уравнения (3.8) имеем

$$\frac{a_0^2}{m} + \left(\frac{2(\lambda m - c)}{\mu m} + \frac{\mu}{\sigma^2}\right)a_0 - \frac{2c}{\sigma^2} \geq \frac{a_0^2}{m} + \left(\frac{2c}{\sigma^2 a_0} + \frac{\mu}{\sigma^2}\right)a_0 - \frac{2c}{\sigma^2} = \frac{a_0^2}{m} + \frac{\mu}{\sigma^2}a_0, \tag{3.16}$$

откуда получаем неравенство $a_0^2/m + (\mu/\sigma^2)a_0 \leq 0$, т.е. приходим к противоречию.

Для полноты изложения рассмотрим далее вопрос о поведении других решений ОДУ (3.7).

3.1.2. Случай неограниченно возрастающих решений, когда $A(0) > a_0 > 0$. Заметим, что при $A(u) > a_0 > 0$, $u \geq 0$, правая часть ОДУ (3.7) положительна, откуда следует, что $A'(u) > 0$, $u \geq 0$. Тогда траектории решений ЗК для ОДУ (3.7) с условием $A(0) > a_0 > 0$ являются строго возрастающими и неограниченными, так как не могут иметь пределом число, большее a_0. Отсюда, в частности, получаем, что $A(u_0) > A(0)$ $\forall u_0 > 0$.

Остановимся на поведении таких решений подробнее. Рассмотрим вспомогательную ЗК для вспомогательного линейного ОДУ:

$$\tilde{A}'(u) = \frac{\tilde{A}(u)}{m} + \left(\frac{2\lambda}{\mu} + \frac{\mu}{\sigma^2} - \frac{2c}{\mu m}\right) - \frac{2c}{\sigma^2 A(u_0)}, \quad u \geq u_0 > 0, \quad \tilde{A}(u_0) = A(u_0). \tag{3.17}$$

Введем функцию $b(u) = A(u) - \tilde{A}(u)$, $u \geq u_0$, где $b(u_0) = 0$. Тогда $b(u)$ есть решение ЗК:

$$b'(u) = \frac{b(u)}{m} - \frac{2c}{\sigma^2 A(u)} + \frac{2c}{\sigma^2 A(u_0)}, \quad u \geq u_0, \quad b(u_0) = 0. \tag{3.18}$$

Здесь $-2c/(\sigma^2 A(u)) + 2c/(\sigma^2 A(u_0)) \geq 0$ при $u \geq u_0$, так как $A(u) \geq A(u_0)$ при $u \geq u_0$. Тогда $b(u) \geq 0$ и $A(u) - \tilde{A}(u) \geq 0$ при $u \geq u_0$.

Решение линейной ЗК (3.17) для $\tilde{A}(u)$ имеет вид:

$$\begin{gathered}\tilde{A}(u) = A(u_0)\exp((u - u_0)/m)) + C_0(u_0)\int_{u_0}^{u} \exp((u - s)/m)ds = \\ = A(u_0)\exp((u - u_0)/m)) + C_0(u_0)m[\exp((u - u_0)/m)) - 1], \quad u \geq u_0,\end{gathered} \tag{3.19}$$

где $C_0(u_0) = 2\lambda/\mu + \mu/\sigma^2 - 2c/(\mu m) - 2c/(\sigma^2 A(u_0))$.

Заметим, что можно выбрать такое $u_0 > 0$ (быть может, достаточно большое), для которого при любом знаке $C_0(u_0)$ будет $D(u_0) = A(u_0) + mC_0(u_0) > 0$. Тогда существует постоянная $\tilde{D}(u_0) > 0$ такая, что справедливо неравенство

$$\tilde{A}(u) \geq \tilde{D}(u_0)\exp((u - u_0)/m)), \quad u \geq u_0 > 0. \tag{3.20}$$

В самом деле: при $C_0(u_0) \geq 0$ достаточно положить $\tilde{D}(u_0) = A(u_0)$, а при $C_0(u_0) < 0$ будет $\tilde{D}(u_0) = D(u_0)$, т.е. $\tilde{D}(u_0) = \min(A(u_0), D(u_0))$.

Тогда справедливо аналогичное неравенство и для $A(u)$, т.е.

$$A(u) \geq \tilde{D}(u_0)\exp((u - u_0)/m)), \quad u \geq u_0 > 0, \tag{3.21}$$

а из соотношения (3.6) получаем

$$V'(u) \geq K\exp\left(-\widehat{D}(u_0)\int_{u_0}^{u}\exp(-(s-u_0)/m))ds\right) = \\ = K\exp(\widehat{D}(u_0)m(\exp(-(u-u_0)/m)-1)) \geq K_0 > 0, \quad u \geq u_0 > 0, \tag{3.22}$$

откуда следует неравенство

$$V(u) \geq K_0 u + K_1, \quad u \geq u_0 > 0, \tag{3.23}$$

т.е. $V(u)$ неограничена при $u \to \infty$.

3.1.3. Случай решений, не существующих глобально, когда $0 < A(0) < a_0$**.** Рассмотрим теперь решение ОДУ (3.7) при $0 < A(0) < a_0$, когда правая часть ОДУ (3.7) отрицательна, откуда следует, что $A'(0) < 0$. Тогда траектории решений ЗК для ОДУ (3.7) при $0 < A(0) < a_0$ строго убывают с ростом u. При приближении $A(u)$ к нулю сверху при $u \to u^{(0)} - 0$ при некотором $u^{(0)} > 0$ будет $A'(u) \to -\infty$, и решение не существует глобально.

Из результатов подразд. 3.1 следует

Теорема 2. *В случае* $r = 0$ *справедливы следующие утверждения: существует единственное решение* $V(u)$ *сингулярной нелинейной КЗ* (3.5), (1.15) *для ИДУ, оно однозначно определяется аналитической формулой* (3.10), *где* $a_0 > 0$ *определено в* (3.11).

4. ПРОВЕРОЧНЫЕ ТЕОРЕМЫ И ОКОНЧАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ О РЕШЕНИИ ЗАДАЧИ ОПТИМИЗАЦИИ В СЛУЧАЕ ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПРЕМИЙ

Приведем сначала утверждения, устанавливающие связь решений КЗ (2.4), (1.14) и (2.4), (1.15) с ВНР, отвечающей оптимальному управлению в оптимизационной задаче (1.12) для случаев $r > 0$ и $r = 0$ соответственно (так называемые проверочные теоремы).

Теорема 3. *Пусть* $r > 0$ *и существует неубывающее решение* $V(u)$ *КЗ* (2.4), (1.14), *дважды непрерывно дифференцируемое на* $(0,\infty)$ *за исключением, быть может, точки* $u = c/r$. *Тогда оно определяет ВНР, соответствующую оптимальному управлению в задаче* (1.12), *т.е.* $V(u) = \varphi^*(u)$ *для любого* $u \geq 0$. *При этом оптимальное управление имеет вид* $\pi^* = \{\alpha^*(t)\}_{t\geq 0}$, *где* $\alpha^*(t) = \alpha_V^*(X_{t-}^{\pi^*})$, $X_t^{\pi^*}$ *– размер резерва в момент времени* t *при управлении* π^*, *а функция* $\alpha_V^*(u)$ *определена в* (2.3) *при* $u \in (0, c/r)$ *и* $\alpha_V^*(u) = 0$ *при* $u \geq c/r$.

Теорема 4. *Пусть* $r = 0$ *и существует неубывающее дважды непрерывно дифференцируемое решение* $V(u)$ *КЗ* (2.4), (1.15). *Тогда оно определяет ВНР, соответствующую оптимальному управлению в задаче* (1.12), *т.е.* $V(u) = \varphi^*(u)$ *для любого* $u \geq 0$. *При этом оптимальное управление имеет вид* $\pi^* = \{\alpha^*(t)\}_{t\geq 0}$, *где* $\alpha^*(t) = \alpha_V^*(X_{t-}^{\pi^*})$, $X_t^{\pi^*}$ *– размер резерва в момент времени* t *при управлении* π^*, *а функция* $\alpha_V^*(u)$ *определена в* (3.3).

Доказательства теорем 3 и 4 аналогичны доказательству проверочного утверждения в [9] и здесь не приводятся. При этом необходимо отметить, что дополнительная трудность в доказательстве теоремы 3, связанная с применением формулы Ито при отсутствии в общем случае непрерывности второй производной функции $V(u)$ в точке $u = c/r$, преодолевается с помощью аппроксимации $V(u)$ дважды непрерывно дифференцируемыми функциями в малой окрестности данной точки с последующим устремлением радиуса этой окрестности к нулю. Применение такой аппроксимации при использовании проверочных аргументов для обоснования вида ВНР как решения КЗ для ИДУ при инвестировании резерва в безрисковый актив в данной модели демонстрируется в [11].

В случае экспоненциального распределения размеров премий справедливы следующие утверждения.

Теорема 5. *Пусть* $F(z)$ *имеет вид* (1.3) *и* $r > 0$. *Тогда функция* $V(u)$, *определенная в теореме* 1, *задает ВНР, соответствующую оптимальному управлению в задаче* (1.12), *т.е.* $V(u) = \varphi^*(u)$ *для любого* $u \geq 0$. *При этом оптимальное управление имеет вид* $\pi^* = \{\alpha^*(t)\}_{t\geq 0}$, *где* $\alpha^*(t) = A(X_{t-}^{\pi^*})/X_{t-}^{\pi^*}$,

$X_t^{\pi^*}$ – размер резерва в момент времени t при управлении π^*, а функция $A(u)$ при $u \in (0, c/r)$ является решением начальной задачи для ОДУ (2.33) *с условием* $\lim_{u\to c/r-0} A(u) = 0$ (см. (2.25)) *и* $A(u) = 0$ *при* $u \geq c/r$.

Утверждение теоремы 5 непосредственно следует из теорем 1 и 3 с учетом соотношения (2.24) и того факта, что функция $A(u)$ имеет асимптотическое представление (2.25), а в случае экспоненциального распределения размеров премий удовлетворяет также ОДУ (2.33).

Теорема 6. *Пусть $F(z)$ имеет вид* (1.3) *и $r = 0$. Тогда функция $V(u)$, определенная в теореме* 2, *т.е., функция вида* (3.10), *где $a_0 > 0$ определено в* (3.11), *задает ВНР, соответствующую оптимальному управлению в задаче* (1.12), *т.е. $V(u) = \varphi^*(u)$ для любого $u \geq 0$. При этом оптимальное управление имеет вид $\pi^* = \{\alpha^*(t)\}_{t\geq 0}$, где $\alpha^*(t) = a_0/X_{t-}^{\pi^*}$, $X_t^{\pi^*}$ – размер резерва в момент времени t при управлении π^*.*

Утверждение теоремы 6 является очевидным следствием теорем 2 и 4 с учетом соотношения (3.6) и того факта, что в случае экспоненциального распределения размера премий функция $A(u) \equiv a_0$ является решением уравнения (3.7), соответствующим определенной в теореме 2 функции $V(u)$.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ИССЛЕДОВАНИЙ ВНР КАК ФУНКЦИИ НК

По аналогии с [1] и [2], приведем замечание о размерности параметров модели и способе их “обезразмеривания” для последующих вычислений.

Замечание 2. В ИДУ (2.48) (и в ИДУ (3.5)) параметр m – безразмерный (при безразмерных переменных u и V), а каждый из параметров σ^2, μ, r, c и λ имеет размерность $1/[t]$, где $[t]$ – размерность времени. Чтобы перейти к безразмерным величинам, достаточно разделить ИДУ (2.48) (и ИДУ (3.5)) на какую-либо характерную положительную постоянную той же размерности: такое деление приводит задачи (2.48), (1.14) и (3.5), (1.15) к задачам того же вида с новыми параметрами $\tilde{\sigma}^2$, $\tilde{\mu}$, $\tilde{c}$ и $\tilde{\lambda}$, $\tilde{r}$. В частности, удобно в (2.48) и в (3.5) в качестве параметра “обезразмеривания” выбрать λ: если положить $\lambda = 1$, то остальные параметры будут измеряться в долях λ и при необходимости их можно пересчитать в размерном виде простым умножением на величину $\lambda > 0$. В результате, оставляя λ в формулах, как это принято в литературе, полагаем всюду в расчетах $\lambda = 1$.

Для всех примеров расчетов зафиксированы значения $\lambda = 1$, $m = 2$.

Расчеты оптимальной ВНР как функции НК при $r > 0$ осуществлялись по алгоритму п. 2.2.1 данной работы, а при $r = 0$ – по аналитической формуле (3.10), где a_0 определено в (3.11).

Расчеты ВНР как функции НК для простых стратегий инвестиций (с изменением доли α рискового актива в портфеле) осуществлялись при $0 < \alpha \leq 1$ по алгоритму [1, с. 1991], а при $\alpha = 0$ – по алгоритму [2, с. 1688] (см. там теорему 5).

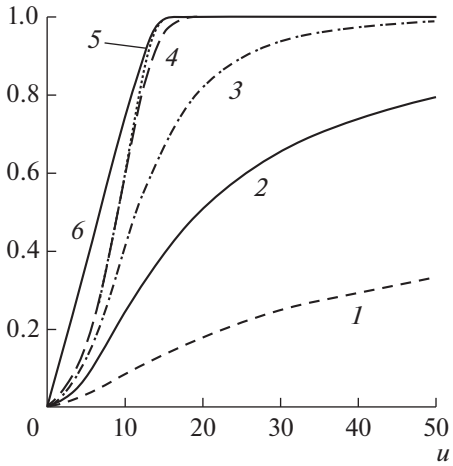
5.1. Случай $r > 0$

Данные к графикам на фиг. 1: $c = 4$; параметры составляющих портфеля: $\mu = 0.25$, $r = 0.24$ ($c/r \approx 16.7$), $\sigma^2 = 0.855$; оптимальной стратегии отвечает график *6*; графикам *1–5* отвечают значения $\alpha = \{2/3;0.5;1/3;0.1;0\}$ соответственно.

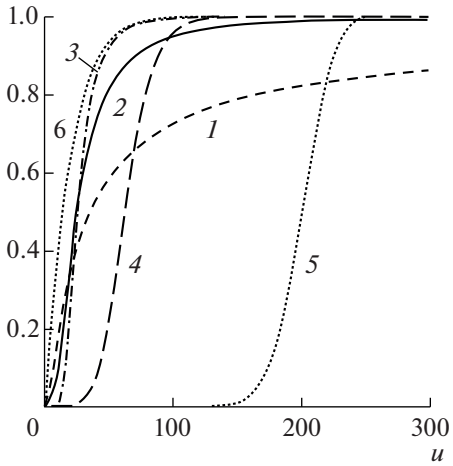
Данные к графикам на фиг. 2: $c = 4$; параметры составляющих портфеля: $\mu = 0.25$, $r = 0.01$ ($c/r = 400$), $\sigma^2 = 0.3$; оптимальной стратегии отвечает график *6*; графикам *1–5* отвечают значения $\alpha = \{1;0.5;1/3;0.1;0\}$ соответственно.

5.2. Случай $r = 0$

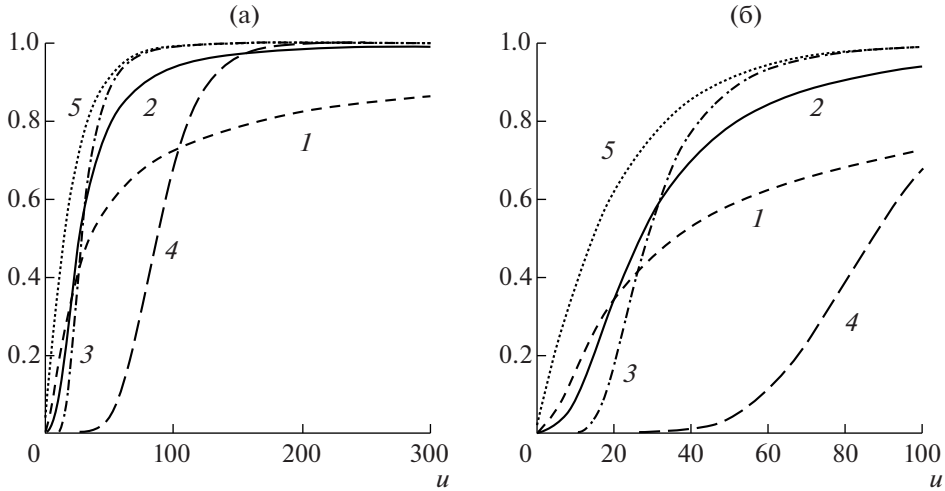
Данные к графикам на фиг. 3а (графики на фиг. 3б те же в другом масштабе): $c = 4$; параметры составляющих портфеля: $\mu = 0.25$, $\sigma^2 = 0.3$; оптимальной стратегии отвечает график *5*; графикам *1–4* отвечают значения $\alpha = \{1;0.5;1/3;0.1\}$ соответственно. Заметим, что значение $\alpha = 0$, соответствующее в случае $r = 0$ отсутствию каких-либо инвестиций (как рисковых, так и безрис-



Фиг. 1. Сравнение оптимальной ВНР с ВНР при простых стратегиях инвестиций с различными долями рискового актива в портфеле, $(\mu - r)/\sigma \approx 0.011$.



Фиг. 2. Сравнение оптимальной ВНР с ВНР при простых стратегиях инвестиций с различными долями рискового актива в портфеле, $(\mu - r)/\sigma \approx 0.438$.



Фиг. 3. Сравнение оптимальной ВНР с ВНР при простых стратегиях инвестиций с различными долями рискового актива в портфеле, $r = 0$, $\mu/\sigma \approx 0.456$.

ковых), приводит здесь к тождественно нулевой функции ВНР (в силу отрицательности нагрузки безопасности в исходном процессе риска, см. [2]).

Во всех представленных примерах видно, что функция ВНР, соответствующая оптимальному управлению, мажорирует функции ВНР, соответствующие постоянным долям рискового актива в инвестиционном портфеле. В описании фиг. 1–3 для сведения приведены значения отношения Шарпа

$$Sh = \frac{\mu - r}{\sigma}, \tag{5.1}$$

соответствующие выбранным в данных примерах параметрам рискового и безрискового активов (об отношении Шарпа см., например, [12, гл. 2]). На фиг. 1 представлен случай с малым значением отношения Шарпа (ожидаемая доходность акций при большой волатильности не на много больше доходности безрискового актива), и при данном наборе выбранных долей акций α в портфеле с постоянной структурой наблюдается (видимая в данном масштабе) монотонность – возрастание ВНР с убыванием α. На самом деле в общем случае это не так. Действительно, нетрудно показать, что функция $A(u)$, определяющая оптимальное количество денежных средств, инвестируемых в акции, при экспоненциальном распределении размеров премий является ограниченной (как решение начальной задачи для уравнения (2.33) на $(0, c/r)$ с условием $A(c/r) = 0$, а при $r = 0$ – постоянной функцией $A \equiv a_0$); следовательно, при достаточно малых значениях резерва оптимальное управление дает значения доли рискового актива, превышающее единицу (что предполагает заимствование денежных средств для инвестирования их в рисковый актив), и даже стремящееся к бесконечности значение этой доли при стремлении резерва к нулю (см. соотношения (2.35), (3.14) для оптимальной стратегии). Таким образом, глобальная монотонность ВНР по α не может ожидаться, и это наглядно видно на фиг. 2 и 3: ВНР, соответствующая полному вложению резерва в рисковый актив, при малых значениях этого резерва мажорирует функции ВНР, соответствующие любой другой постоянной структуре портфеля.

Численные примеры и приведенные выше рассуждения показывают, что вычисление функций ВНР, соответствующих постоянной структуре портфеля, наряду с оптимальной ВНР, позволяют для любого заданного значения начального резерва, ограничившись простыми стратегиями, выбрать ту из них, которая дает значение ВНР, имеющее приемлемую близость к ее оптимальному значению: при разных значениях начального резерва это будут разные стратегии.

Дополнительно приведем следующее

Замечание 3 (об асимптотическом сравнении оптимальной стратегии и стратегий с постоянной структурой инвестиционного портфеля по значениям ожидаемой доходности). Замена оптимальной инвестиционной стратегии некоторой стратегией с постоянной долей рискового актива, для которой при заданном значении начального капитала значение ВНР близко к оптимальному, дает также возможность получить бóльшую ожидаемую доходность в итоговом процессе риска. Действительно, применение оператора (1.7) к функции вида $f(u) = u$ (что соответствует рассмотрению на бесконечно малом интервале времени ожидаемого приращения самого капитала, равного u в начале интервала) приводит к выражению, которое при делении на u представляет собой мгновенную (т.е. в единицу времени) ожидаемую доходность при стратегии, равной α в точке u. Как нетрудно видеть, эта доходность равна

$$\mu_\alpha + (\lambda m - c)/u, \tag{5.2}$$

где $\mu_\alpha = \alpha(\mu - r) + r$, и представляет собой (мгновенную) ожидаемую доходность инвестиционного портфеля, а второе слагаемое в (5.2) – ожидаемая доходность (также в единицу времени) исходного страхового процесса риска. Очевидно тогда, что увеличение доли α рискового актива в портфеле при его постоянной структуре приводит к увеличению общей ожидаемой доходности (в изначально сформулированном предположении $\mu > r$). Однако в то же время при больших значениях u это будет приводить и к увеличению риска, выраженного в том, что ВНР с увеличением α будет уменьшаться, а если эта доля превосходит некоторое значение, зависящее от параметров модели, ВНР обратится в тождественный ноль. Данный вывод основан на асимптотическом представлении ВНР, полученном в [13] (см. также [1]): если $\rho_\alpha := 2\mu_\alpha/\sigma_\alpha^2 > 1$, где μ_α определено выше, $\sigma_\alpha = \alpha\sigma$, то $\varphi(u) = 1 - Ku^{1-\rho_\alpha}(1 + o(1))$, $u \to \infty$, где $0 < K$ – постоянная; при $\rho_\alpha \le 1$ имеем $\varphi \equiv 0$. Отсюда нетрудно видеть, что с ростом α асимптотические характеристики ВНР ухудшаются: ρ_α является убывающей функцией при положительных α, более того, при

$$\alpha \ge \alpha_0^+ := \frac{(\mu - r)}{\sigma^2} + \sqrt{\left[\frac{(\mu - r)}{\sigma^2}\right]^2 + \frac{r}{\sigma^2}} \tag{5.3}$$

величина ρ_α приобретает значения, меньшие или равные единице, что приводит к неминуемому разорению в дальнейшем при сохранении той же стратегии.

При использовании оптимальной стратегии в случае $r > 0$ доля рискового актива равна нулю при значениях u, больших фиксированного числа c/r, и тогда ожидаемая доходность результирующего процесса будет минимальной и равной $r + (\lambda m - c)/u$ (см. (5.2)), но при этом разорения никогда не произойдет (при указанных значениях u). В случае $r = 0$ оптимальная доля стремится к нулю при $u \to \infty$, и тогда ожидаемая доходность стремится к нулю при неограниченном увеличении резерва, но ВНР будет максимально возможной. Однако, заменив оптимальную стратегию простой стратегией с постоянной структурой актива, можно обеспечить значение ожидаемой доходности, имеющее положительный предел, больший безрисковой доходности r, при $u \to \infty$, сохраняя при этом близость ВНР к оптимальному значению при заданном начальном резерве. Указанный предел ожидаемой доходности соответствует значению μ_α (см. (5.2)), которое может быть оценено сверху с помощью неравенства $\mu_\alpha < \mu_{\alpha_0^+} = \mathrm{Sh}^2 + \sqrt{\mathrm{Sh}^4 + \mathrm{Sh}^2 r} + r$, выполненного в ситуации выбора α, исключающего неминуемое разорение (здесь Sh – отношение Шарпа, определенное в (5.1), α_0^+ определено в (5.3)). Такое α может быть выбрано сколь угодно близким, но строго меньшим значения α_0^+.

Заметим также, что сравнение оптимальной и других стратегий одновременно по двум критериям – ВНР и ожидаемой доходности при любых значениях начального капитала – является более общей задачей, решение которой связано с дополнительными исследованиями и численными расчетами.

ЗАКЛЮЧЕНИЕ

Как уже отмечалось в [1] и [2], где исследовался вопрос о влиянии простых рисковых стратегий инвестирования на ВНР в коллективной модели пенсионного страхования (или дуальной модели риска), для более точного ответа на вопрос о том, какую роль рисковые активы могут играть в увеличении ВНР, необходимо изучить задачу оптимального динамического управления портфелем активов (в предположении, что доля рискового актива может меняться со временем). Эта задача, связанная с исследованием сингулярных задач для нелинейных ИДУ, изучена в данной работе. В результате показано, что оптимальное управление порождается некоторой стратегией, т.е. функцией от текущего резерва; эта стратегия в общем случае зависит от решения уравнения Беллмана рассматриваемой оптимизационной задачи, являющегося указанным нелинейным ИДУ. В случае экспоненциального распределения премий доказано существование решения краевой задачи для соответствующего ИДУ и на основании проверочных теорем сделаны выводы о том, что это решение определяет оптимальную ВНР, т.е. ВНР, соответствующую оптимальному управлению. При этом оптимальная стратегия, порождающая оптимальное управление, наиболее простой вид имеет при нулевой процентной ставке ($r = 0$): она соответствует постоянному количеству денежных средств, инвестируемых в рисковый актив; это количество определяется положительным корнем квадратного уравнения с коэффициентами, зависящими от параметров модели. Соответствующая вероятность разорения (разность между единицей и ВНР) имеет явный вид и определяется экспоненциально убывающей функцией начального резерва (напомним, что из результатов [1] следует, что при постоянной доле инвестирования резерва в рисковый актив вероятность разорения асимптотически ведет себя как степенная функция). В случае $r > 0$ оптимальная стратегия соответствует полному вложению резерва в безрисковый актив при значениях резерва, не меньших величины c/r, в противном случае, т.е. при значениях резерва, меньших указанной константы, оптимальное количество денежных средств, инвестируемых в рисковый актив, является решением некоторой начальной задачи для нелинейного ОДУ с предельным нулевым значением в точке c/r. Соответствующая оптимальная ВНР, т.е. функция Беллмана оптимизационной задачи, равна единице при значениях резерва, не меньших c/r, а на интервале $[0, c/r]$ определяется решением КЗ для нелинейного ИДУ. При этом данная функция может терпеть разрыв второй производной в точке c/r, оставаясь при этом непрерывно дифференцируемой.

Проведены численные расчеты, демонстрирующие на графиках соответствующих ВНР преимущество оптимальной стратегии по сравнению с простыми стратегиями, состоящими во вложении постоянной доли резерва в рисковый актив. Расчеты позволяют также при заданном уровне начального резерва выбрать простую стратегию, дающую приемлемое значение ВНР в смысле его близости к оптимальному значению.

СПИСОК ЛИТЕРАТУРЫ

1. *Белкина Т.А., Конюхова Н.Б., Славко Б.В.* Платежеспособность страховой компании в дуальной модели риска с учетом инвестиций: анализ и численные исследования сингулярных краевых задач // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 11. С. 174–198.
2. *Белкина Т.А., Конюхова Н.Б., Славко Б.В.* Безрисковые инвестиции и их сравнение с простыми рисковыми стратегиями в модели пенсионного страхования: решение сингулярных задач для интегродифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 10. С. 1676–1696.
3. *Grandell J.* Aspects of Risk Theory. New York: Springer, 1991. 175 p.
4. *Asmussen S., Albrecher H.* Ruin Probabilities. Advanced series on statistical science and applied probability. V. 14. Singapore: World Scientific, 2010. 602 p.
5. *Browne S.* Optimal investment policies for a firm with a random risk process: exponential utility and minimyzing the probability of ruin // Mathemtics Operations Research. 1995. V. 20. № 4. P. 937–958.
6. *Hipp C., Plum M.* Optimal investment for insurers // Insurance Math. Econom. 2000. V. 27. № 2. P. 215–228.
7. *Hipp C., Plum M.* Optimal investment for investors with state dependent income, and for insurers // Finance Stochast. 2003. V. 7. № 3. P. 299–321.
8. *Azcue P., Muler M.* Optimal investment strategy to minimize the ruin probability of an insurance company under borrowing constraints // Insurance Math. Econom. 2009. V. 44. № 1. P. 26–34.
9. *Belkina T., Hipp C., Luo S., Taksar M.* Optimal constrained investment in the Cramer–Lundberg model // Scandinavian Actuarial Journal. 2014. Issue 5. P. 383–404.
10. *Флеминг У., Ришел Р.* Оптимальное управление детерминированными и стохастическими системами. М.: МИР, 1978. 318 с.
11. *Belkina T.A., Konyukhova N.B.* On sufficient conditions for survival probability in the life annuity insurance model with risk-free investment income// In: “IX Moscow International Conference on Operations Research (ORM2018. Moscow, October 22-27, 2018). Proceedings. In two volumes/ Editor-in-chief F. Ereshko.” Moscow: MAKS Press, 2018. V. 1. P. 213–218.
12. *Шоломицкий А.Г.* Теория риска. Выбор при неопределенности и моделирование риска. М.: Изд. дом ГУ ВШЭ, 2005. 400 с.
13. *Kabanov Yu., Pergamenshchikov S.* In the insurance business risky investments are dangerous: the case of negative risk sums // Finance Stochast. 2016. V. 20. № 2. P. 355–379.

EDN: EFAIHJ
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1491–1521

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.642

ТРЕХМЕРНЫЕ СТАЦИОНАРНЫЕ СФЕРИЧЕСКИ СИММЕТРИЧНЫЕ МОДЕЛИ ЗВЕЗДНОЙ ДИНАМИКИ, ЗАВИСЯЩИЕ ОТ ЛОКАЛЬНОЙ ЭНЕРГИИ[1)]

© 2022 г. Ю. Батт[1,*], Э. Йорн[1], А. Л. Скубачевский[2,]**

[1] *80333 München, Theresienstr. 39, Mathematisches Institut der Universität München, Germany*

[2] *117198 Москва, ул. Миклухо-Маклая, 6, Российский университет дружбы народов, Россия*

**e-mail: batt@mathematik.uni-muenchen.de*

***e-mail: skublector@gmail.com*

Поступила в редакцию 18.02.2022 г.
Переработанный вариант 18.02.2022 г.
Принята к публикации 11.05.2022 г.

Рассматривается система уравнений Власова–Пуассона, описывающая распределение гравитирующих частиц в трехмерном пространстве. Исследуется существование сферически симметричных решений этой системы, которые состоят из трех функций: функции распределения, зависящей от локальной энергии, локальной плотности и ньютоновского потенциала. Изучены две задачи. В первой задаче по заданной положительной, строго убывающей на некотором открытом интервале функции требуется построить сферически симметричное решение системы Власова–Пуассона, в котором локальная плотность совпадает с заданной функцией. Сведение к уравнению Эддингтона позволило получить достаточные условия, при выполнении которых эта задача разрешима. Приведены примеры, когда решение записывается в явном виде. Во второй задаче по заданной положительной функции на открытом интервале требуется построить сферически симметричное решение системы Власова–Пуассона, в котором функция распределения совпадает с заданной функцией. Эта задача сводится к нелинейному интегральному уравнению, которое решается численно. Библ. 12. Фиг. 10. Табл. 5.

Ключевые слова: трехмерная система Власова–Пуассона, стационарные решения, численное приближение.

DOI: 10.31857/S0044466922090083

1. ВВЕДЕНИЕ

Система уравнений Власова–Пуассона (VPS) в трехмерном случае (модель звездной динамики) имеет следующий вид:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} - \frac{\partial}{\partial x}U(t,x)\frac{\partial f}{\partial v} = 0, \tag{V}$$

$$\Delta U(t,x) = 4\pi\rho(t,x) \tag{P_1}$$

$$\text{или} \quad U(t,x) = -\int\frac{\rho(t,y)}{|x-y|}dy, \tag{P_2}$$

$$\rho(t,x) = \int f(t,x,v)dv. \tag{D}$$

Здесь $f = f(t,x,v) \geqslant 0$ – функция распределения гравитирующего вещества, $U = U(t,x) \leqslant 0$ – ньютонов потенциал и $\rho(t,x) \geqslant 0$ – локальная плотность. Эта статья обобщает предыдущую работу [1] по двумерным галактикам до трехмерного случая. Она также открывает путь в дальнейшем к получению явных решений системы уравнений Власова–Пуассона, выходящих за рамки

[1)]Работа третьего автора выполнена при финансовой поддержке Минобрнауки РФ в рамках государственного задания: соглашение № 075-03-2020-223/3 (FSSF-2020-0018).

классических известных примеров, которые содержатся, например, в [2]. Для случая зависящих от времени функций (начальная задача) в работе [3] приводится обзор исследований до 2007 г. Стационарные сферически симметричные функции характеризуются свойством $f(x,v) = f(A_1x, A_2v)$ для любых $A_1, A_2 \in 0(3)$; для краткого описания этого класса решений, имеющих отношение к настоящей работе, см. [1], а также приведенные там ссылки. Стационарные решения краевых задач для системы Власова–Пуассона, в которых функция распределения зависит от локальной энергии, рассматривались в работах [4], [5].

Данная работа посвящена двум задачам. Первая задача известна как "обратная задача": идентифицировать те функции p, определенные на ограниченном интервале $[0, R]$, как локальная плотность стационарной сферически симметричной модели звездной динамики, в которой f зависит от локальной энергии:

$$f(r,u) = q(-E - E_0), \quad \text{где} \quad E_0 > 0 - \text{константа}.$$

Этот вопрос возникает, если мы хотим определить три величины f, ρ, U на основе данных, в качестве которых обычно используется профиль яркости, который может быть превращен в профиль массы. Следующий вопрос заключается в определении потенциала U и функции распределения f (разд. 2–5).

Вторую задачу будем называть *прямой задачей*. Известно, что функция распределения f стационарной сферически симметричной модели звездной динамики является функцией локальной энергии E и углового момента $F := x^2v^2 - (xv)^2$ (это утверждение называется теоремой Джинса) [6]. Прямая задача ставит обратный вопрос, а именно: какие функции q допускают нахождение функций $\rho(r)$ и $U(r)$ наряду с константой $E_0 > 0$ таких, что $f(r,u) = q(-E - E_0)$, ρ и U образуют тройку стационарной сферически симметричной модели звездной динамики (разд. 6–8). Далее мы дадим краткий обзор разделов данной работы.

Разд. 2. Введение потенциального оператора $U = Lp$ на $\mathfrak{D}(L)$ (определение 2.1) с описанием его свойств в лемме 2.1. Каждая неотрицательная строго убывающая функция $p \in \mathfrak{D}(L)$ с носителем $\operatorname{supp} p = [0, R]$ удовлетворяет нелинейному интегральному уравнению $p = FLp$ с соответствующим $F = F[p]$ (лемма 2.2).

Разд. 3. Определение стационарных сферически симметричных решений, зависящих от локальной энергии и доказательство их свойств, лемма об эквивалентности и уравнение Эддингтона (лемма 3.2).

Разд. 4. Обратная задача: формулировка и решение (теорема 4.2).

Разд. 5. Примеры, иллюстрирующие теорему 4.2 и концепцию расширимости.

Разд. 6. Постановка прямой задачи и ее преобразование к эквивалентной задаче решения нелинейного интегрального уравнения вида

$$Lp - E_0 = G_0(p).$$

Разд. 7. Построение аппроксимирующей нелинейной системы (ANS) вида

$$\sum_{k=0}^{n-1} A_{ik}x_k := \sum_{k=0}^{n-1} B_{ik}x_k - C_kx_k = G_0(x_i)$$

и вычисление матрицы (A_{ik}).

Разд. 8. Численный анализ (ANS), описание приближений и сходимости, примеры.

Разд. 9. Содержит работу Тонелли по уравнениям Абеля и Эддингтона с полными доказательствами.

Разд. 10. Состоит из описания двух задач для дальнейшего исследования.

Краткое содержание данной статьи было опубликовано в работе [7].

2. ПОТЕНЦИАЛЬНЫЙ ОПЕРАТОР В СФЕРИЧЕСКИ СИММЕТРИЧНОМ СЛУЧАЕ

Определим потенциальный оператор

$$Lp(x) := \int_{\mathbb{R}^3} \frac{p(y)}{|x-y|}\,dy, \quad x \in \mathbb{R}^3,$$

для сферически симметричных функций p, определенных на $\mathbb{R}^3$. Это значит, что $p(x) = p(r)$, $r := |x|$. Отсюда следует, что функция Lp также сферически симметрична. Действительно, если $A \in O(3)$, тогда, предполагая, что $y = Az$, имеем

$$Lp(Ax) = \int\limits_{\mathbb{R}^3} \frac{p(y)}{|Ax - y|} dy = \int\limits_{\mathbb{R}^3} \frac{p(y)}{|A(x - A^{-1}y)|} dy = \int\limits_{\mathbb{R}^3} \frac{p(Az)}{|x - z|} dz = \int\limits_{\mathbb{R}^3} \frac{p(z)}{|x - z|} dz = Lp(x).$$

Определение 2.1. Обозначим через $\mathcal{D}(L)$ множество функций $p : \mathbb{R}_{0+} \to \mathbb{R}_{0+} \cup \{\infty\}$ со следующими свойствами:

(а) $p \in C(\mathbb{R}_+)$,

(б) для любых $r > 0$ выполняется $\int_0^r p(s)s^2 ds < \infty$, $\int_r^\infty p(s)s ds < \infty$,

(в) существует $\delta > 0$ такое, что $p(r) > 0$ для любых $r \in (0, \delta)$, где $\mathbb{R}_{0+} = \{r \in \mathbb{R} : r \geqslant 0\}$ и $\mathbb{R}_+ = \{r \in \mathbb{R} : r > 0\}$.

Лемма 2.1. *Для $p \in \mathcal{D}(L)$ имеем*

1)

$$Lp(r) = 4\pi\left[\frac{1}{r}\int\limits_0^r p(s)s^2 ds + \int\limits_r^\infty p(s)s ds\right], \quad r > 0, \tag{2.1}$$

2)

$$Lp \in C^2(\mathbb{R}_+) \quad \text{и} \quad (Lp)'(r) = -\frac{4\pi}{r^2}\int\limits_0^r p(s)s^2 ds, \quad r > 0, \tag{2.2}$$

$$(Lp)''(r) = -\frac{2}{r}(Lp)'(r) - 4\pi p(r), \quad r > 0. \tag{2.3}$$

3) *$Lp > 0$ и $(Lp)' < 0$, т.е. Lp строго убывает на $\mathbb{R}_+$. Так как пределы*

$$Lp(0) := \lim_{r\to 0} Lp(r), \quad Lp(\infty) = \lim_{r\to\infty} Lp(r),$$

существуют, то функция Lp имеет строго убывающую обратную

$$(Lp)^{-1} : (Lp(\infty), Lp(0)) \to (0, \infty).$$

Доказательство. 1) Используя сферические координаты

$$x = (r\sin\psi\cos\varphi, r\sin\psi\sin\varphi, r\cos\psi),$$

получаем

$$Lp(x) = \int\limits_0^\pi\int\limits_0^{2\pi}\int\limits_0^\infty \frac{p(s)\sin\psi}{\sqrt{r^2 + s^2 - 2rs\cdot\cos\psi}} s^2 ds d\varphi d\psi = 2\pi\int\limits_0^\infty\int\limits_0^\pi \frac{\sin\psi}{\sqrt{r^2 + s^2 - 2rs\cdot\cos\psi}} d\psi\, p(s)s^2 ds.$$

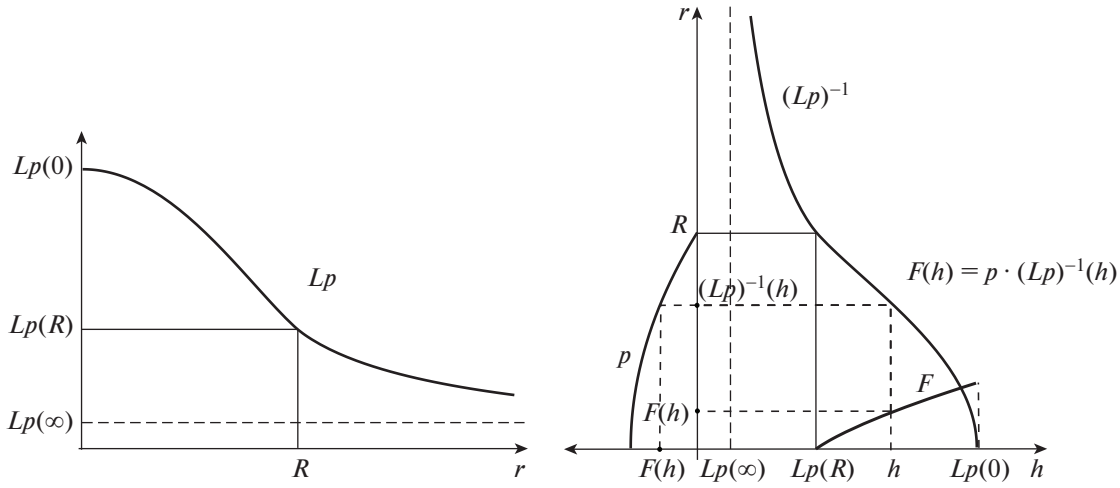
Сделаем подстановку во внутреннем интеграле: $u := \sqrt{r^2 + s^2 - 2rs\cos\psi}$. Тогда

$$\int\limits_0^\pi \frac{\sin\psi}{\sqrt{r^2 + s^2 - 2rs\cdot\cos\psi}} d\psi = \int\limits_{\sqrt{r^2+s^2-2rs}}^{\sqrt{r^2+s^2+2rs}} 1\cdot du\cdot\frac{1}{rs} = \frac{(r+s) - |r-s|}{rs} = \begin{cases} \dfrac{2}{r} & \text{для} \quad s \leqslant r, \\ \dfrac{2}{s} & \text{для} \quad s \geqslant r, \end{cases}$$

откуда следует (2.1).

2) (2.2) является результатом дифференцирования (2.1), а (2.3) следует из дифференцирования (2.2).

3) Неравенство $Lp > 0$ является следствием из определения 2.1 в), а $(Lp)' < 0$ следует из (2.2). Существование пределов и обратной функции является прямым следствием из этих фактов.



Фиг. 1. $F(h) = p \circ (Lp)^{-1}(h)$, $E_0 := Lp(R)$.

Большинство функций $p \in \mathcal{D}(L)$ будут иметь компактный носитель. Определим

$\mathcal{D}_R(L) := \{p \in \mathcal{D}(L) : p > 0 \text{ на } [0, R),\ p = 0 \text{ на } [R, \infty)\}$,

$\mathcal{D}_R^-(L) := \{p \in \mathcal{D}_R(L) : p \text{ строго убывает на } [0, R)\}$.

Функции $p \in \mathcal{D}_R^-(L)$ являются решениями нелинейного интегрального уравнения, что будет показано в следующей лемме.

Лемма 2.2. *Пусть* $p \in \mathcal{D}_R^-(L)$.

Тогда существует единственная строго возрастающая функция

$$F := F[p] : [Lp(R), Lp(0)) \to p((0, R])$$

такая, что

$$p(r) = F \circ Lp(r), \quad r \in (0, R].$$

Доказательство. В лемме 2.1 3) утверждается, что

$$Lp : (0, R] \to [Lp(R), Lp(0))$$

строго убывает и имеет строго убывающую обратную

$$(Lp)^{-1} : [Lp(R), Lp(0)) \to (0, R].$$

Так как p строго убывает на $(0, R]$, то суперпозиция

$$F := p \circ (Lp)^{-1} : [Lp(R), Lp(0)) \to p((0, R])$$

существует и строго возрастает (см. фиг. 1). Тогда

$$F(h) = p \circ (Lp)^{-1}(h), \quad h \in [Lp(R), Lp(0))$$

означает, что

$$F \circ Lp(r) = p(r), \quad r \in (0, R].$$

Если G удовлетворяет уравнению $p = G \circ Lp$, то $G = p \circ (Lp)^{-1} = F$. Таким образом, доказана единственность функции F.

Следствие 2.1. Пусть $F(h) := 0$ для $h \in (0, Lp(R))$. Тогда при выполнении условий леммы 2.2 мы имеем

$$p(r) = F \circ Lp(r) \ \text{ на } \ (0, R] \leftrightarrow p(r) = F \circ Lp(r) \ \text{ на } \ \mathbb{R}_+.$$

Доказательство. Пусть $p(r) = F \circ Lp(r)$ на $(0, R]$. В этом случае, если $R < r$, тогда $p(r) = 0$ и $Lp(R) > Lp(r)$, что означает $F \circ Lp(r) = 0$, т.е. $p(r) = F \circ Lp(r)$ на $\mathbb{R}_+$. Обратное утверждение тривиально.

Далее будем использовать аббревиатуру $P := Lp$.

3. СТАЦИОНАРНЫЕ СФЕРИЧЕСКИ СИММЕТРИЧНЫЕ РЕШЕНИЯ, ЗАВИСЯЩИЕ ОТ ЛОКАЛЬНОЙ ЭНЕРГИИ И ИХ СВОЙСТВА

Определение 3.1. *Тройку* (f, ρ, U) функций $f = f(r,u)$, $\rho = \rho(r)$, $U = U(r)$ будем называть стационарным сферически симметричным, зависящим от E решением системы (VPS), *если* $\rho \in \mathcal{D}_R^-(L)$, и существует функция $q = q(s)$ со следующим свойством:

$$q \in L^1_{\mathrm{loc}}(\mathbb{R}), \quad q(s) = 0 \quad \text{для} \quad s \in (-\infty, 0], \quad q(s) > 0 \quad \text{для} \quad s \in (0, P(0) - E_0) \tag{Q}$$

$$f(r,u) = q(-E(r,u) - E_0), \tag{V}$$

$$E(r,u) := U(r) + \frac{u^2}{2}, \quad u := |v| \quad (\text{мы пишем } f = f_q),$$

$$U(r) = -L\rho(r), \quad L\rho(R) =: E_0, \tag{P}$$

$$\rho(r) = \int_{\mathbb{R}^3} f(r, |v|)dv, \quad r \in \mathbb{R}_+. \tag{D}$$

Отметим, что (Q) описывает свойства q, (V) отсылает к тому, что f является интегралом уравнения Власова (т.е. является постоянной вдоль характеристик), (P) представляет собой уравнение Пуассона (P_2) в проинтегрированном виде, (D) является определением локальной плотности.

В качестве подготовки к следующей важной лемме мы докажем ключевое тождество.

Лемма 3.1. *Пусть* $p \in \mathcal{D}(L)$, $E_0 > 0$, *и* q *удовлетворяет* (Q). *Тогда для* $E_0 \leqslant h < Lp(0)$ *выполняется следующее равенство*:

$$\int_{\mathbb{R}^3} q\left(h - E_0 - \frac{v^2}{2}\right)dv = 4\pi\sqrt{2}\int_0^{h-E_0} q(s)\sqrt{h - E_0 - s}\,ds. \tag{3.1}$$

Доказательство. Так как $q(s) = 0$ для $s \in (-\infty, 0]$, имеем

$$\int_{\mathbb{R}^3} q\left(h - E_0 - \frac{v^2}{2}\right)dv = \int_{|v| < \sqrt{2(h-E_0)}} q\left(h - E_0 - \frac{v^2}{2}\right)dv =$$

$$= \int_0^{\pi}\int_0^{2\pi}\int_0^{\sqrt{2(h-E_0)}} q\left(h - E_0 - \frac{u^2}{2}\right)u^2\,du\,d\varphi\sin\psi\,d\psi = 4\pi\int_0^{\sqrt{2(h-E_0)}} q\left(h - E_0 - \frac{u^2}{2}\right)u^2\,du.$$

Переходя к новой переменной $s := h - E_0 - \dfrac{u^2}{2}$, будем иметь $u = \sqrt{2(h - E_0 - s)}$. Следовательно, получаем (3.1).

Лемма 3.2 (лемма об эквивалентности). (а) *Пусть* (f_q, ρ, U) *— стационарное сферически симметричное, зависящее от* E *решение системы* (*VPS*). *Пусть* $F := F[p]$ (*лемма* 2.2), *где* $p = \rho$. *Тогда*

$$F(h) = 4\pi\sqrt{2}\int_0^{h-E_0} q(s)\sqrt{h - E_0 - s}\,ds \quad \text{для} \quad h \in [E_0, P(0)). \tag{3.2}$$

(б) *Пусть* q *удовлетворяет* (Q), *и пусть*

$$F(h) := 4\pi\sqrt{2}\int_0^{h-E_0} q(s)\sqrt{h - E_0 - s}\,ds \quad \text{для} \quad h \in [E_0, P(0)). \tag{3.3}$$

Предположим, что интегральное уравнение

$$p(r) = F \circ Lp(r) \tag{3.4}$$

имеет решение $p \in \mathcal{D}_R^-(L)$ *на* $\mathbb{R}_+$. *Определим* $\rho := p$, $U(r) := -L\rho(r)$, $E_0 = Lp(R)$ *и* $f_q(r,u) := q(-E(r,u) - E_0)$. *Тогда* (f_q, ρ, u) *является стационарным сферически симметричным, зависящим от* E *решением системы* (*VPS*).

Доказательство. (а) Пусть (f_q, ρ, U) – стационарное сферически симметричное, зависящее от E решение. Затем, в силу леммы 2.2, (D), (V) и (3.1), имеем

$$F \circ L\rho(r) = \rho(r) = \int_{\mathbb{R}^3} f_q(r,|v|)dv = \int_{\mathbb{R}^3} q\left(L\rho(r) - E_0 - \frac{v^2}{2}\right)dv =$$
$$= 4\pi\sqrt{2}\int_0^{L\rho(r)-E_0} q(s)\sqrt{L\rho(r) - E_0 - s}\,ds \quad \text{для} \quad r \in (0, R],$$

откуда следует (3.2).

(б) Наши предположения подразумевают, что (Q), (P), (V) удовлетворены. Кроме того, в силу (3.4), (3.3), (3.1), и (V), имеем

$$\rho(r) = p(r) = F \circ Lp(r) = 4\pi\sqrt{2}\int_0^{Lp(r)-E_0} q(s)\sqrt{Lp(r) - E_0 - s}\,ds =$$
$$= \int_{\mathbb{R}^3} q\left(Lp(r) - E_0 - \frac{v^2}{2}\right)dv = \int_{\mathbb{R}^3} f(r,u)dv, \quad r \in (0, R].$$

Следовательно, (D) также удовлетворено.

В дальнейшем мы будем использовать определение

$$F_0(h) := F(h + E_0) \quad \text{на} \quad [0, P(0) - E_0).$$

Тогда (3.2) примет вид

$$F_0(h) = 4\pi\sqrt{2}\int_0^h q(s)\sqrt{h - s}\,ds, \quad h \in [0, P(0) - E_0). \tag{3.5}$$

Уравнение вида

$$g(x) = \int_0^x f(s)\sqrt{x - s}\,ds$$

называется уравнением Эддингтона. Результаты о его разрешимости основаны на теории для уравнения вида

$$g(x) = \int_0^x \frac{f(s)}{\sqrt{x - s}}ds,$$

которое называется уравнением Абеля. В работе Тонелли [8] доказано существование решений этих уравнений (обзор его работы приведен в [9]).

Для полноты настоящей работы основные результаты и их доказательства приведены в разд. 9.

4. ОБРАТНАЯ ЗАДАЧА

Этот раздел будет посвящен следующему вопросу: при каких условиях заданная функция p, $p \in \mathcal{D}_R^-(L)$, является локальной плотностью стационарного сферически симметричного, зависящего от E решения? В этом случае мы будем говорить, что "p расширима" (на f и U до стационарного сферически симметричного, зависящего от E решения).

Следующее предложение дает первый критерий расширимости для $p \in \mathcal{D}_R^-(L)$.

Теорема 4.1. *Пусть* $p \in \mathcal{D}_R^-(L)$. *Тогда* p *расширима тогда и только тогда, когда уравнение Эддингтона* (3.5) *имеет решение* q *с* (Q) *для* $F := F[p]$ *из леммы* 2.2 *и* $F_0(h) := F(h + E_0)$.

Доказательство. Необходимость. Если p расширима, то существует q с (Q) такая, что

$$f(r,u) = q\left(-U(r) - E_0 + \frac{u^2}{2}\right), \tag{V}$$

где

$$U(r) = -Lp(r), \quad E_0 = Lp(R), \tag{P}$$

$$p(r) = \int_{\mathbb{R}^3} f(r,u)dv. \tag{D}$$

Лемма 3.2, часть (а) показывает, что уравнение Эддингтона (3.5) имеет решение q с (Q).

Достаточность. Если уравнение Эддингтона (3.5) имеет решение q с (Q) для $F := F[p]$, тогда $f := f_q$ удовлетворяет (V) с $U(r) := -Lp(r)$ (P) и $E_0 = Lp(R)$. Следовательно, в силу (V), (3.1) и (3.3) имеем

$$\int_{\mathbb{R}^3} f_q(r,|v|)dv = \int_{\mathbb{R}^3} q\left(Lp(r) - E_0 - \frac{v^2}{2}\right)dv = 4\pi\sqrt{2}\int_0^{Lp(r)-E_0} q(s)\sqrt{Lp(r) - E_0 - s}\,ds = F \circ Lp(r) = p(r).$$

Это совпадает с (D). Тогда p расширима.

Следующая теорема посвящена исследованию разрешимости уравнения Эддингтона вида

$$F_0(h) = 4\pi\sqrt{2}\int_0^h q(s)\sqrt{h-s}\,ds$$

для заданного $F_0(h) := F(h + E_0)$, $F := F[p]$ в более подробной форме.

Эта теорема дает различные условия расширимости для функции $p \in \mathcal{D}_R^-(L)$ в явном виде. Здесь мы используем пространства $L^1_{\mathrm{loc}}[0,T)$ и $AC[0,T)$, которые определены в разд. 9.

Теорема 4.2. *Пусть* $p \in \mathcal{D}_R^-(L)$, $p|_{(0,R]} \in C^2(0,R]$, $F := F[p]$, $E_0 := P(R)$, $F_0(\cdot) = F(\cdot + E_0)$. *Тогда выполняются следующие утверждения.*

1) *Уравнение Эддингтона*

$$F_0(h) = 4\pi\sqrt{2}\int_0^h q(s)\sqrt{h-s}\,ds, \quad 0 \leqslant h < P(0) - E_0, \tag{4.1}$$

имеет единственное вещественнозначное решение $q \in L^1_{\mathrm{loc}}[0, P(0) - E_0)$, *заданное по формуле*

$$q(h) = \frac{1}{4\pi\sqrt{2}}\frac{2}{\pi}\frac{d}{dh}H_{F_0'}(h), \quad 0 \leqslant h < P(0) - E_0,$$

где

$$H_{F_0'}(h) = \int_0^h \frac{F_0'(s)}{\sqrt{h-s}}ds \in AC[0, P(0) - E_0), \quad F_0 \in C^2[0, P(0) - E_0). \tag{4.2}$$

2) p *расширима тогда и только тогда, когда* $q > 0$ *на* $(0, P(0) - E_0)$, *т.е.*

$$\frac{d}{dh}H_{F_0'}(h) = \frac{1}{\sqrt{h}}F_0'(0) + \int_0^h \frac{F_0''(s)}{\sqrt{h-s}}ds > 0 \quad \text{на} \quad (0, P(0) - E_0)$$
$$\left(F_0'(0) = \frac{p'(R)}{P'(R)} \geqslant 0\right). \tag{4.3}$$

3) *Достаточные условия расширимости* p:

(а) $F_0''(s) > 0$ *на* $(0, P(0) - E_0)$,

(б) $X(r) := p'(r) \cdot P''(r) - p''(r)P'(r) > 0$ *на* $(0, R)$,

(в) $\frac{2}{r}p'(r) + p''(r) > 0$ *на* $(0, R)$,

где (а) *и* (б) *эквивалентны и из* (в) *следует* (а) *и* (б).

Доказательство. Из предположений для p и леммы 2.1 следует, что отображение

$$P : (0, R] \to [E_0, P(0))$$

является строго убывающей биекцией в $C^2(0, R]$ со строго убывающим обратным

$$P^{-1} : [E_0, P(0)) \to (0, R]$$

в $C^2[E_0, P(0))$. Суперпозиция с $p \in C^2(0, R]$:

$$F := p \circ P^{-1} : [E_0, P(0)) \to [0, p(0))$$

строго возрастает и $F \in C^2[E_0, P(0))$, $F_0 \in C^2[0, P(0) - E_0)$, $F_0'(\cdot) = F'(\cdot + E_0)$ и $F_0''(\cdot) = F''(\cdot + E_0)$.

1) Чтобы показать, что (4.1) имеет единственное вещественнозначное решение $q \in L^1_{\mathrm{loc}}[0, P(0) - E_0)$, нужно проверить, что для $g := \frac{F_0}{4\pi\sqrt{2}}$ предположения (а), (i) и (ii) леммы А.4 (в приложении) выполняются. Достаточно сделать это для $g := F_0$.

Очевидно, $F_0 \in C^2[0, P(0) - E_0) \subset AC[0, P(0) - E_0)$ и $F_0(0) = F(E_0) = pP^{-1}(E_0) = p(R) = 0$. Для

$$H_{F_0'}(h) := \int_0^h \frac{F_0'(s)}{\sqrt{h - s}} ds$$

(а) (i) означает, что мы должны показать, что $H_{F_0'} \in AC[0, P(0) - E_0)$. Мы имеем $F_0' \in C^1[0, P(0) - E_0)$ и

$$F_0'(0) = F'(E_0) = p'(P^{-1}(E_0)) \cdot (P^{-1})'(E_0) = \frac{p'(R)}{P'(R)} \geqslant 0. \tag{4.4}$$

Интегрируя по частям, получаем

$$H_{F_0'}(h) = -2\sqrt{h - s}F_0'(s)\Big|_0^h + 2\int_0^h F_0''(s)\sqrt{h - s}ds = 2\sqrt{h}F_0'(0) + 2\int_0^h F_0''(s)\sqrt{h - s}ds,$$

и, следовательно, $H_{F_0'} \in AC[0, P(0) - E_0)$, т.е. предположение (i) из леммы А.4 выполняется. Также мы получаем $H_{F_0'}(0) = 0$, что и есть (а) (ii). Из леммы А.4 (а) следует, что (4.1) имеет единственное вещественнозначное решение q, заданное по формуле

$$q(h) := \frac{1}{4\pi\sqrt{2}}\frac{2}{\pi}\frac{d}{dh}H_{F_0'}(h) = \frac{1}{4\pi\sqrt{2}}\frac{2}{\pi}\left[\frac{1}{\sqrt{h}}F_0'(0) + \int_0^h \frac{F_0''(s)}{\sqrt{h - s}}ds\right]. \tag{4.5}$$

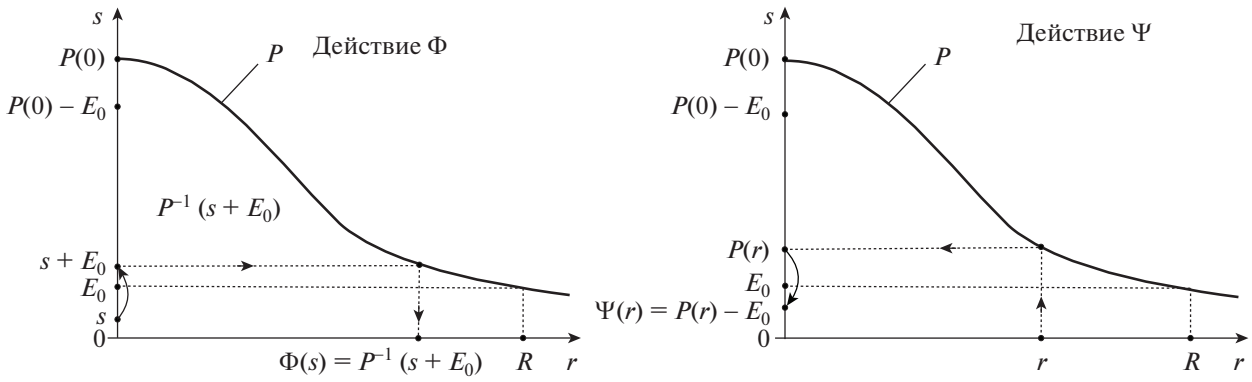
2) Так как $H_{F_0'} \in AC[0, P(0) - E_0)$, то $q \in L^1_{\mathrm{loc}}[0, P(0) - E_0)$. По теореме 1, оно удовлетворяет (Q) и p является расширимым тогда и только тогда, когда $q > 0$ на $(0, P(0) - E_0)$.

3) Доказательство 3) основывается на замене переменных в интеграле

$$\int_0^h \frac{F_0''(s)}{\sqrt{h - s}}ds.$$

Определим C^2-диффеоморфизм Φ

$$\Phi : [0, P(0) - E_0) \to (0, R]$$



Фиг. 2.

как композицию сдвига

$$T : [0, P(0) - E_0) \to [E_0, P(0)), \quad s \mapsto s + E_0,$$

и оператора

$$P^{-1} : [E_0, P(0)) \to (0, R], \quad s \mapsto P^{-1}(s),$$

таким образом, что, $\Phi := P^{-1} \circ T$, $s \mapsto r = P^{-1}(s + E_0)$. Обратное к Φ отображение обозначим $\Psi := T^{-1} \circ P$:

$$\Psi : (0, R] \to [0, P(0) - E_0), \quad r \mapsto s = P(r) - E_0.$$

Действие отображений Φ и Ψ изображено на фиг. 2. Для $0 < s < P(0) - E_0$ имеем $E_0 < s + E_0 < P(0)$. Поэтому $P^{-1}(s + E_0) = \Phi(s) < R$. Для $0 < r < R$ имеем $E_0 < P(r) < P(0)$. Следовательно, $0 < P(r) - E_0 = \Psi(r) < P(0) - E_0$.

Теперь представим $F_0'(s)$, $F_0''(s)$ как функции от $p'(r)$, $p''(r)$, $P'(r)$, $P''(r)$ следующим образом: для $s = \Psi(r)$ имеем

$$F_0(s) = F_0(\Psi(r)) = F(P(r) - E_0 + E_0) = F(P(r)) = pP^{-1}(P(r)) = p(r),$$

$$\frac{d}{dr}F_0(\Psi(r)) = F'(P(r)) \cdot P'(r) = p'(r).$$

Так как $P'(r) < 0$, для $s = \Psi(r)$ получаем

$$F_0'(s) = F_0'(\Psi(r)) = F'(P(r)) = \frac{p'(r)}{P'(r)},$$

$$\frac{d}{dr}F_0'(\Psi(r)) = \frac{d}{dr}F'(P(r) - E_0 + E_0) = \frac{d}{dr}F'(P(r)) = F''(P(r)) \cdot P'(r) = \frac{d}{dr}\frac{p'(r)}{P'(r)} =$$
$$= \frac{P'(r) \cdot p''(r) - p'(r) \cdot P''(r)}{P'(r)^2}.$$

Следовательно, для $s = \Psi(r)$ имеем

$$F_0''(s) = F_0''(\Psi(r)) = F''(P(r)) = \frac{p'(r)P''(r) - p''(r)P'(r)}{|P'(r)|^3}.$$

Заменяя переменные $s \to r$ в интеграле (4.3) по формуле $\Phi(s) := P^{-1} \circ T(s) = P^{-1}(s + E_0) = r$, $\Phi(0) = R$, $\Phi(P(0) - E_0) = 0$, $\Phi(h) = P^{-1} \circ T(h) = P^{-1}(h + E_0)$, $ds = P'(r)dr$, и, учитывая отрицательный знак $P'(r)$, получаем

$$\int_0^h \frac{F_0''(s)}{\sqrt{h-s}}ds = \int_{P^{-1}(h+E_0)}^R \frac{p'(r)P''(r) - p''(r)P'(r)}{|P'(r)|^3\sqrt{h-(P(r)-E_0)}}|P'(r)|dr. \tag{4.6}$$

Из 3) (а) или 3) (б) следует $q > 0$ с учетом (4.3), (4.4).

Пусть 3) (в) выполняется. Тогда по лемме 2.1 имеем

$$p'P'' - p''P' = p'\left(-\frac{2}{r}P' - 4\pi p\right) - p''P' = 4\pi p(-p') + (-P')\left(\frac{2}{r}p' + p''\right) > 0.$$

Откуда следует, что 3) (б) выполняется.

Замечание 4.1. Для численных расчетов, используемых позднее, удобно переписать (4.3) в другом виде. Поскольку $\Phi(h) = P^{-1}(h + E_0)$, то $P(\Phi(h)) = h + E_0$, и поэтому

$$h = P(\Phi(h)) - P(R).$$

Тогда, подставляя (4.4), (4.6) в (4.3), получаем

$$\frac{d}{dh}H_{F_0'}(h) = \frac{1}{\sqrt{P(\Phi(h)) - P(R)}}\left(\frac{p'(R)}{P'(R)} + \int_{\Phi(h)}^R \frac{p'(r)P''(r) - p''(r)P'(r)}{|P'(r)|^2}\sqrt{\frac{P(\Phi(h)) - P(R)}{P(\Phi(h)) - P(r)}}dr\right). \tag{4.7}$$

Подынтегральное выражение имеет сингулярность в точке $r = \Phi(h)$.

Замечание 4.2. Примеры в следующем разделе покажут, что условия 3) (а), (б) и (в) не являются необходимыми для расширимости.

5. ПРИМЕРЫ

Пример 5.1.

$$p(r) := \begin{cases} 1 - \left(\frac{r}{R}\right)^2, & 0 \leqslant r \leqslant R, \\ 0, \quad r > R, & R > 0. \end{cases}$$

Этот пример позволяет явно вычислить другие функции: P, P^{-1}, F, q, $G_0 =: F_0^{-1}$ (которая является правой частью аппроксимирующей нелинейной системы (ANS), возникающей в разд. 7).

Для $0 \leqslant r \leqslant R$

$$p'(r) = -\frac{2}{R^2}r, \quad p''(r) = -\frac{2}{R^2},$$

откуда $\frac{2}{r}p'(r) + p'' = -\frac{6}{R^2} < 0$. Поэтому достаточное условие 3) (в) в теореме 4.2 не выполняется.

Подставив функцию $p(r)$ в (2.1), получаем

$$P(r) = 4\pi\left[\frac{1}{r}\int_0^r\left(1 - \frac{s^2}{R^2}\right)s^2ds + \int_r^R\left(1 - \frac{s^2}{R^2}\right)sds\right] = 4\pi\left[\frac{r^2}{3} - \frac{1}{5}\frac{r^4}{R^2} + \frac{1}{2}(R^2 - r^2) - \frac{1}{4}\frac{1}{R^2}(R^4 - r^4)\right] =$$

$$= 4\pi R^2\left[\frac{1}{20}\left(\frac{r}{R}\right)^4 - \frac{1}{6}\left(\frac{r}{R}\right)^2 + \frac{1}{4}\right] = \pi R^2\left[\frac{1}{5}\left(\frac{r}{R}\right)^4 - \frac{2}{3}\left(\frac{r}{R}\right)^2 + 1\right] := h,$$

$$P'(r) = 4\pi R^2\left[\frac{1}{5}\frac{r^3}{R^4} - \frac{1}{3}\frac{r}{R^2}\right],$$

$$P''(r) = 4\pi R^2\left[\frac{3}{5}\frac{r^2}{R^4} - \frac{1}{3}\frac{1}{R^2}\right],$$

$$P(R) = E_0 = \frac{8}{15}\pi R^2.$$

Следовательно,

$$X(r) = p'P'' - p''P' = -\frac{2}{R^2}r \cdot 4\pi R^2\left[\frac{3}{5}\frac{r^2}{R^4} - \frac{1}{3}\frac{1}{R^2}\right] + \frac{2}{R^2}\cdot 4\pi R^2\left[\frac{1}{5}\frac{r^3}{R^4} - \frac{1}{3}\frac{r}{R^2}\right] = -\frac{16\pi}{5}\frac{r^3}{R^4} < 0.$$

Таким образом, достаточное условие 3) (б) в теореме 4.2 не выполняется.

Так как $P(r) = h$ и $P : [0, R] \to [E_0, P(0)] = \left[\frac{8}{15}\pi R^2, \pi R^2\right]$ является биквадратичной формой по $\frac{r}{R}$, мы можем вычислить P^{-1} следующим образом:

$$\left(\frac{r}{R}\right)^2 = \frac{5}{3} - \sqrt{\frac{25}{9} - 5 + \frac{5}{\pi R^2}h},$$

$$r = P^{-1}(h) = R\sqrt{\frac{5}{3} - \sqrt{\frac{5}{\pi R^2}h - \frac{20}{9}}}.$$

Следовательно,

$$F(h) = p(P^{-1}(h)) = 1 - \left(\frac{P^{-1}(h)}{R}\right)^2 = 1 - \left(\frac{5}{3} - \sqrt{\frac{5}{\pi R^2}h - \frac{20}{9}}\right) =$$

$$= \sqrt{\frac{5}{\pi R^2}h - \frac{20}{3}} - \frac{2}{3}, \quad h \in [E_0, P(0)],$$

$$F_0(h) = F(h + E_0) = \sqrt{\frac{5}{\pi R^2}\left(h + \frac{8}{15}\pi R^2\right) - \frac{20}{9}} - \frac{2}{3} = \sqrt{ah + \frac{4}{9}} - \frac{2}{3}$$

$$\text{при} \quad a := \frac{5}{\pi R^2}, \quad h \in [0, P(0) - E_0].$$

То есть

$$F_0(0) = 0,$$

$$F_0(P(0) - E_0) = F_0\left(\frac{7}{15}\pi R^2\right) = \sqrt{\frac{7}{3} + \frac{4}{9}} - \frac{2}{3} = 1,$$

$$F_0'(h) = \frac{a}{2}\frac{1}{\sqrt{ah + \frac{4}{9}}},$$

$$F_0'(0) = \frac{3}{4}a,$$

$$F_0'(P(0) - E_0) = \frac{3a}{10},$$

$$F_0''(h) = -\frac{a^2}{4}\frac{1}{\left(\sqrt{ah + \frac{4}{9}}\right)^3}.$$

Из [10] следует, что

$$H_{F_0'}(h) = \int_0^h \frac{F_0'(s)}{\sqrt{h - s}}ds = \frac{a}{2}\int_0^h \frac{ds}{\sqrt{as + \frac{4}{9}}\sqrt{h - s}} = \frac{a}{2}\frac{2}{\sqrt{a}}\operatorname{arctg}\sqrt{\frac{(as + 4/9)}{a(h - s)}}\Bigg|_{s=0}^{s=h} = \sqrt{a}\left(\frac{\pi}{2} - \operatorname{arctg}\left(\frac{2}{3}\frac{1}{\sqrt{a}}\frac{1}{\sqrt{h}}\right)\right).$$

Дифференцируя последнее выражение, получаем

$$\frac{d}{dh}H_{F_0'}(h) = -\frac{\sqrt{a}}{1+\frac{4}{9}\frac{1}{a}\frac{1}{h}}\frac{2}{3}\cdot\frac{1}{\sqrt{a}}\left(-\frac{1}{2}\right)\frac{1}{h^{3/2}} = \frac{1}{3}\frac{1}{h+\frac{4}{9}\frac{1}{a}}\cdot\frac{1}{\sqrt{h}} > 0, \quad h \in (0, P(0) - E_0).$$

Таким образом, $H_{F_0'}$ принадлежит $AC[0, P(0))$, $H_{F_0'}(0) = 0$ и $\frac{d}{dh}H_{F_0'}(h) > 0$ на $(0, P(0) - E_0)$. Следовательно, в силу пунктов теоремы 4.2 1) и 4.2 2), функция p расширима и

$$q(h) := \frac{1}{4\pi\sqrt{2}}\frac{2}{\pi}\frac{d}{dh}H_{F_0'}(h) = \frac{\sqrt{2}}{4\pi^2}\frac{1}{3}\frac{1}{h+\frac{4}{9}\frac{1}{a}}\cdot\frac{1}{\sqrt{h}} > 0$$

является единственным решением уравнения Эддингтона (4.1).

Этот пример иллюстрирует как обратную задачу (с известным p), так и прямую (с известным q – в разд. 6). Он будет раскрыт подробнее в разд. 8.

Пример 5.2.

$$p(r) := \begin{cases} \left(1-\frac{r}{R}\right)^2, & 0 \leqslant r \leqslant R, \\ 0, & R < r, \quad R > 0. \end{cases}$$

Это пример позволяет легко решить вопрос о расширимости с помощью теоремы 4.2 3) (б). Для $0 \leqslant r \leqslant R$

$$p'(r) = -\frac{2}{R} + 2\frac{r}{R^2}, \quad p''(r) = \frac{2}{R^2}.$$

Функция $\frac{2}{r}p'(r) + p''(r) = \frac{6}{R^2} - \frac{4}{rR}$ меняет свой знак в точке $\frac{2R}{3} \in (0, R)$. Поэтому теорема 4.2 3) (в) не применима. Из (2.1) получаем

$$P(r) = 4\pi\left[\frac{1}{r}\int_0^r\left(s^2 - \frac{2s^3}{R} + \frac{s^4}{R^2}\right)ds + \int_r^R\left(s - \frac{2s^2}{R} + \frac{s^3}{R^2}\right)ds\right] =$$

$$= 4\pi\left[\frac{1}{3}r^2 - \frac{1}{2}\frac{r^3}{R} + \frac{1}{5}\frac{r^4}{R^2} + \frac{1}{2}(R^2 - r^2) - \frac{2}{3}\frac{1}{R}(R^3 - r^3) + \frac{1}{4}\frac{1}{R^2}(R^4 - r^4)\right] =$$

$$= 4\pi\left[-\frac{1}{6}r^2 + \frac{1}{6}\frac{r^3}{R} - \frac{1}{20}\frac{r^4}{R^2} + \frac{1}{12}R^2\right],$$

$$P'(r) = 4\pi\left[-\frac{1}{3}r + \frac{1}{2}\frac{r^2}{R} - \frac{1}{5}\frac{r^3}{R^2}\right],$$

$$P''(r) = 4\pi\left[-\frac{1}{3} + \frac{r}{R} - \frac{3}{5}\frac{r^2}{R^2}\right] б$$

$$X(r) = (p'(r)P''(r) - p''(r)P'(r)) = 4\pi\left[\left(-\frac{2}{R} + 2\frac{r}{R^2}\right)\left(-\frac{1}{3} + \frac{r}{R} - \frac{3r^2}{5R^2}\right) - \frac{2}{R^2}\left(-\frac{1}{3}r + \frac{1}{2}\frac{r^2}{R} - \frac{1}{5}\frac{r^3}{R^2}\right)\right] =$$

$$= \frac{4\pi}{R}\left(\frac{2}{3} - 2\alpha + \frac{11}{5}\alpha^2 - \frac{4}{5}\alpha^3\right) =: \frac{4\pi}{R}f(\alpha) \quad \text{при} \quad \alpha := \frac{r}{R} \leqslant 1.$$

На $[0,1]$ имеем $f'(\alpha) = -2 + \frac{22}{5}\alpha - \frac{12}{5}\alpha^2$. Легко видеть, что f убывает на $\left[0, \frac{5}{6}\right]$, возрастает на $\left[\frac{5}{6}, 1\right]$ и $f(5/6) = \frac{7}{108}$. Следовательно, $f(\alpha) > 0$ на $[0,1]$ и $X(r) > 0$ на $(0, R)$. Тогда в силу теоремы 4.2 3) (б) p расширима.

В этом случае невозможно вычислить P^{-1} явно (как в примере 5.1), так как $P(r)$ является монотонным неспециальным многочленом степени 4.

Пример 5.3.

$$p(r) := \begin{cases} e^{-r} - e^{-R}, & 0 \leqslant r \leqslant R, \\ 0, \quad R < r, & R > 0. \end{cases}$$

Для $0 < r \leqslant R$ имеем

$$p'(r) = -e^{-r}, \quad p''(r) = e^{-r},$$

$$P(r) := 4\pi\left[\frac{1}{r}\int_0^r (e^{-s} - e^{-R})s^2\,ds + \int_r^R (e^{-s} - e^{-R})s\,ds\right] =$$

$$= 4\pi\left[\frac{1}{r}\left(e^{-s}(-s^2 - 2s - 2) - e^{-R}\frac{s^3}{3}\right)\Bigg|_0^r + \left(e^{-s}(-s-1) - e^{-R}\frac{s^2}{2}\right)\Bigg|_r^R\right] =$$

$$= 4\pi\left[-e^{-r}\left(r + 2 + \frac{2}{r}\right) + \frac{2}{r} - \frac{1}{3}r^2e^{-R} - e^{-R}\left(R + 1 + \frac{R^2}{2}\right) + e^{-r}(r+1) + e^{-R}\frac{r^2}{2}\right] =$$

$$= 4\pi\left[-e^{-r}\left(1 + \frac{2}{r}\right) + \frac{2}{r} - e^{-R}\left(1 + R + \frac{R^2}{2} - \frac{r^2}{6}\right)\right],$$

$$P'(r) = 4\pi\left[e^{-r}\left(1 + \frac{2}{r} + \frac{2}{r^2}\right) - \frac{2}{r^2} + \frac{1}{3}e^{-R}r\right],$$

$$P''(r) = 4\pi\left[-e^{-r}\left(1 + \frac{2}{r} + \frac{4}{r^2} + \frac{4}{r^3}\right) + \frac{4}{r^3} + \frac{1}{3}e^{-R}\right].$$

Вычислим $X(r, R)$ в треугольнике $\{(r, R) : 0 < R,\ r \in (0, R]\}$:

$$X(r, R) = p'(r)P''(r) - p''(r)P'(r) = -e^{-r}(P'(r) + P''(r)) =$$

$$= -4\pi e^{-r}\left[e^{-r}\left(1 + \frac{2}{r} + \frac{2}{r^2}\right) - \frac{2}{r^2} + \frac{1}{3}e^{-R}r - e^{-r}\left(1 + \frac{2}{r} + \frac{4}{r^2} + \frac{4}{r^3}\right) + \frac{4}{r^3} + \frac{1}{3}e^{-R}\right] =$$

$$= 4\pi e^{-r}\left[\left\{e^{-r}\left(\frac{2}{r^2} + \frac{4}{r^3}\right) + \frac{2}{r^2}\right\} - \left\{\frac{4}{r^3} + \frac{e^{-R}}{3}(r+1)\right\}\right] = 4\pi\frac{e^{-r}}{r^3}\left[\{e^{-r}(2r+4) + 2r\} - \left\{4 + \frac{1}{3}e^{-R}(r^4 + r^3)\right\}\right].$$

Поскольку

$$\frac{\partial X(r, R)}{\partial R} = \frac{4\pi}{3}\frac{r+1}{e^{r+R}} > 0 \quad \text{для} \quad 0 < R, \quad r \in (0, R], \tag{5.1}$$

то X строго возрастает в направлении роста значений $R > 0$ для каждого фиксированного $r \in (0, R]$.

На первый взгляд кажется, что X, P, P', P'' имеют сингулярности в точке $r = 0$. Но они могут быть устранены, как показано для X в формулах (5.2) и (5.3). Для P, P', P'' это можно сделать аналогично. Последняя формула для $X(r, R)$ бесполезна при вычислении значений $X(r, R)$ с r вблизи нуля, поскольку X является разностью двух больших положительных членов, имеющих почти одинаковые значения. Поэтому мы сделаем следующее преобразование:

Если $\tilde{X}(r, R) := e^{2r}r^3(4\pi)^{-1} \cdot X(r, R)$, то

$$\tilde{X}(r, R) = \left[(2r+4) - \frac{e^{r-R}}{3}r^3(1+r) + (2r-4)e^r\right] =$$

$$= \left[-\frac{e^{r-R}}{3}r^3(1+r) + (2r+4) + (2r-4)\left(1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} + \frac{r^4}{4!} + \sum_{k=5}^{\infty}\frac{r^k}{k!}\right)\right] =$$

$$= \left[-\frac{e^{r-R}}{3}r^3(1+r) + (2r+4) + \left(2r + 2r^2 + \frac{2r^3}{2!} + \frac{2r^4}{3!} + \frac{2r^5}{4!}\right) + \right.$$

$$\left. + \left(-4 - 4r - \frac{4r^2}{2!} - \frac{4r^3}{3!} - \frac{4r^4}{4!}\right) + (2r-4)\sum_{k=5}^{\infty}\frac{r^k}{k!}\right] =$$

Таблица 1. $X(r,2)$ для $0 \leqslant r \leqslant 2$

r	0	10^{-5}	0.25	0.5	0.75	1	1.25	1.5	1.75	2
$X(r,2)$	3.6220	3.6218	2.33	1.47	0.910	0.541	0.302	0.151	0.056	0

$$= \left[-\frac{e^{r-R}}{3}r^3(1+r) + r^3\left(1-\frac{4}{6}\right) + r^4\left(\frac{2}{6}-\frac{1}{6}\right) + r^5\frac{2}{4!} + (2r-4)\sum_{k=5}^{\infty}\frac{r^k}{k!}\right] =$$

$$= r^3\left[-\frac{e^{r-R}}{3}(1+r) + \frac{1}{3} + \frac{r}{6} + \frac{r^2}{12} + (2r-4)r^2\sum_{k=5}^{\infty}\frac{r^{k-5}}{k!}\right].$$

Разделив $\tilde{X}$ на $e^{2r}r^3(4\pi)^{-1}$, получаем

$$X(r,R) = \frac{4\pi}{e^{2r}}\left[-\frac{e^{r-R}}{3}(1+r) + \frac{1}{3} + \frac{r}{6} + \frac{r^2}{12} + (2r-4)r^2\sum_{k=5}^{\infty}\frac{r^{k-5}}{k!}\right]. \tag{5.2}$$

Отсюда следует, что

$$\lim_{r\to 0} X(r,R) = 4\pi\left[-\frac{1}{3e^R} + \frac{1}{3}\right] \in \left(0, \frac{4\pi}{3}\right] \quad \text{для} \quad R > 0. \tag{5.3}$$

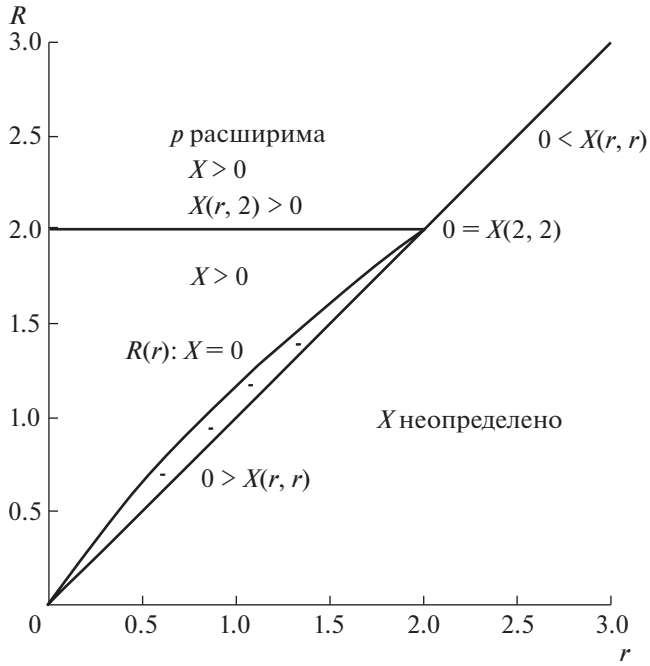
Вначале вычислим $X(r,r)$:

$$X(r,r) = \frac{4\pi}{e^{2r}}\left[-\frac{e^{r-r}}{3}(1+r) + \frac{1}{3} + \frac{r}{6} + \frac{r^2}{12} + (2r-4)r^2\sum_{k=5}^{\infty}\frac{r^{k-5}}{k!}\right] = \frac{4\pi}{e^{2r}}r\left[\frac{r}{12} - \frac{1}{6} + (2r-4)\sum_{k=5}^{\infty}\frac{r^{k-4}}{k!}\right]. \tag{5.4}$$

Из (5.4) заключаем, что $X(r,r) = 0$ для $r = 2$, $X(r,r) > 0$ для $r > 2$ и $X(r,r) < 0$ для $r < 2$.

Применение теоремы 4.2 3) (б) требует определения знака $X(r,R)$ для $0 < r < R$. Мы будем различать два случая

$$(1) \quad R \geqslant 2 \quad \text{и} \quad (2) \quad R < 2.$$



Фиг. 3. Диапазоны положительных и отрицательных значений $X(r,R)$.

В случае (1) для $0 < r < 2$ из (5.1) и табл. 1 (фиг. 3), следует, что

$$X(r,R) > X(r,2) > 0,$$

а для $2 \leqslant r < R$ из (5.1) и (5.4) заключаем, что

$$X(r,R) > X(r,r) \geqslant 0.$$

В случае (1) с $R \geqslant 2$ отсюда следует, что $X(r,R) > 0$ и p расширима по теореме 4.2 3) (б).

В случае (2) с $0 < R < 2$ в силу (5.4) мы имеем $X(r,r) < 0$. С другой стороны, $X(r,2) > 0$ для $r \in [0,2)$. Согласно (5.1) существует единственный нуль функции $X(r,R)$ между $R = r$ и $R = 2$ для каждого фиксированного $r \in (0,2)$. Нули лежат на монотонной кривой, порожденной функцией $X(r,R) = 0$ в некоторой окрестности выше диагонали (см. фиг. 3):

$$R(r) = -\ln\left(6\frac{(r+2)e^{-r} + r - 2}{(r+1)r^3}\right) \quad \text{для} \quad 0 < r \leqslant R < 2.$$

Поэтому расширимость p может быть получена из теоремы 4.2 2) или формулы (4.7) в замечании 4.1. Интеграл нельзя вычислить в явном виде и подынтегральная функция сингулярная на левом конце интервала интегрирования. Но интеграл можно аппроксимировать по формуле Ньютона–Котса, избегая граничных точек (например, по правилу средних точек). Ошибка, возникающая при использовании аппроксимационных формул, должна быть аккуратно оценена относительно предшествующего члена $p'(R)/P'(R)$, чтобы утверждать расширимость p по теореме 4.2 2). Такой случай будет рассмотрен в примере 5.6. Детали мы оставляем читателю.

Пример 5.4.

$$p(r) := \begin{cases} r^{-b} - R^{-b}, & 0 < r \leqslant R, \\ 0, & r > R, \quad R > 0, \quad 0 < b < 3. \end{cases}$$

Отметим сначала, что если $b < 3$, то $p(r) \in \mathfrak{D}_R^-(L)$.

Для $0 < r \leqslant R$

$$p'(r) = -br^{-b-1}, \qquad p''(r) = b(b+1)r^{-b-2}.$$

Для $1 < b < 3$ получаем

$$\frac{2}{r}p'(r) + p''(r) = -2br^{-b-2} + b(b+1)r^{-b-2} = b(b-1)r^{-b-2} > 0,$$

и, следовательно, из теоремы 4.2 3) (в) следует, что p расширима.

Для $0 < b < 3$, $b \neq 2$ и $0 < r < R$ имеем

$$P(r) = 4\pi\left[\frac{1}{r}\int_0^r (s^{-b+2} - R^{-b}s^2)ds + \int_r^R (s^{-b+1} - R^{-b}s)ds\right] =$$

$$= 4\pi\left[\frac{1}{-b+3}r^{-b+2} - \frac{1}{3}R^{-b}r^2 + \left(\frac{1}{-b+2}s^{-b+2} - R^{-b}\frac{s^2}{2}\right)\Bigg|_r^R\right] =$$

$$= 4\pi\left[\frac{r^{-b+2}}{-b+3} - R^{-b}\frac{r^2}{3} + \frac{R^{-b+2}}{-b+2} - \frac{r^{-b+2}}{-b+2} - \frac{R^{-b+2}}{2} + \frac{R^{-b}r^2}{2}\right] =$$

$$= 4\pi\left[\frac{-r^{-b+2}}{(-b+3)(-b+2)} + \frac{1}{6}R^{-b}r^2 + \frac{b}{2(-b+2)}R^{-b+2}\right].$$

Для $0 < r < R$

$$\frac{X(r)}{4\pi} = (-b)r^{-b-1}\left(\frac{b-1}{-b+3}r^{-b} + \frac{1}{3}R^{-b}\right) - b(b+1)r^{-b-2}\left(\frac{-r^{-b+1}}{-b+3} + \frac{1}{3}R^{-b}r\right) =$$
$$= \frac{2b}{-b+3}r^{-2b-1} - \frac{1}{3}(b^2+2b)R^{-b}r^{-b-1} = \frac{R^{-b}r^{-b-1}}{-b+3}\left(2b\left(\frac{R}{r}\right)^b - \frac{b^2}{3}(3-b) - \frac{2b}{3}(3-b)\right) =$$
$$= \frac{R^{-b}r^{-b-1}}{-b+3}\cdot b\cdot\left(2\left(\left(\frac{R}{r}\right)^b - 1\right) + \frac{1}{3}(b^2-b)\right).$$

Для $1 \leqslant b < 3$, $b \neq 2$, мы получаем $X(r) > 0$, и из теоремы 4.2 3) (б) следует, что p расширима. Для $0 < b < 1$ X имеет ноль в точке r_0 на $(0, R)$, являющийся решением уравнения

$$\left(\frac{R}{r}\right)^b - 1 - \frac{1}{6}(b - b^2) = 0.$$

Следовательно,

$$\frac{r_0}{R} = \left(\frac{6}{6 + (b - b^2)}\right)^{1/b}.$$

Таким образом, $\frac{r_0}{R} = 0.9216$ для $b = \frac{1}{2}$, $\frac{r_0}{R} = 1$ для $b = 1$, и $\lim\limits_{b\to 0}\frac{r_0}{R}(b) = e^{-1/6} \approx 0.8465$. Доказательство существования нуля функции X требует применения теоремы 4.2 2). Детали мы опускаем.

Пример 5.5. Данный пример иллюстрирует прямую задачу с заданным q и p вычисленным приближенно в разд. 8.

Пусть $q(s) := c\sqrt{s}$, где $c > 0$. Тогда, осуществив замену переменных $s = h\sin^2 t$, получаем

$$F_0(h) := 4\pi\sqrt{2}\cdot c\int\limits_0^h \sqrt{s}\sqrt{h-s}\,ds = 4\pi\sqrt{2}\cdot c\frac{h^2}{4}\int\limits_0^{\pi/2}(1-\cos 4t)dt = 4\pi\sqrt{2}\cdot c\frac{h^2}{8}\cdot\pi = \frac{\pi^2 c}{\sqrt{2}}h^2.$$

Пример 5.6. Нерасширимая функция p.

Цель этого примера – показать численно, что не все p, удовлетворяющие условиям теоремы 4.2, автоматически являются расширимыми. Другими словами, мы построим функцию p, которая в важных деталях отличается от примеров расширимых функций, приведенных выше: между тем как во всех наших примерах p либо выпуклы, либо вогнуты на $[0, R]$, в этом примере p будет вогнута на подынтервале $[0, w]$ и выпукла на $[w, R]$ для некоторого $0 < w < R$ ($p''(w) = 0$). Важную роль будут играть неравенство (4.3) и формула (4.7).

Построим функцию

$$p(r) = \begin{cases} a_0 + a_1 r + a_2 r^2 + a_3 r^3 + a_4 r^4, & r \in [0, R], \\ 0, & r > R, \end{cases}$$

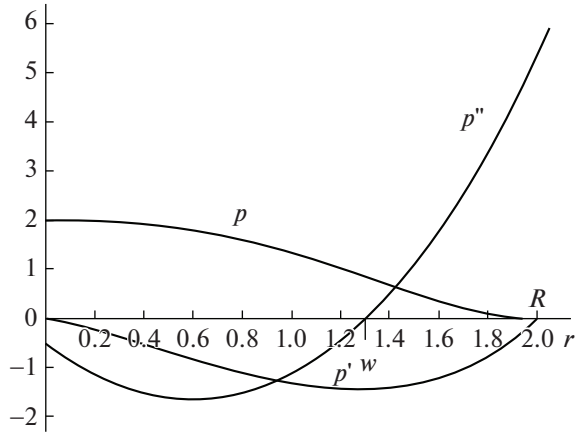
такую, что $p \in \mathcal{D}_R^-(L)$ и $\frac{d}{dh}H_{F_0'}(h) < 0$ для некоторого $h \in (0, P(0) - E_0)$. Тогда в силу теоремы 4.2 2) p будет нерасширимой.

Из условия $p \in \mathcal{D}_R^-(L)$ следует, что

$$p(0) > 0, \tag{5.5}$$
$$p(R) = 0. \tag{5.6}$$

Пусть p удовлетворяет следующим условиям:

$$p'(0) = 0, \tag{5.7}$$
$$p'(R) = 0, \tag{5.8}$$
$$p''(w) = 0 \quad \text{для некоторого} \quad w \in (0, R). \tag{5.9}$$



Фиг. 4. p, p' и p''.

Из равенств (5.8) и (4.4) вытекает, что $F_0'(0) = 0$. Это упрощает формулу (4.7). Очевидно, условия (5.5)–(5.9) выполняются тогда и только тогда, когда $a_0 > 0$, $a_1 = 0$ и

$$a_2R^2 + a_3R^3 + a_4R^4 = -a_0, \tag{5.10}$$

$$2a_2R + 3a_3R^2 + 4a_4R^3 = 0, \tag{5.11}$$

$$2a_2 + 6a_3w + 12a_4w^2 = 0. \tag{5.12}$$

Далее будем предполагать, что

$$R = 2, \quad w = \frac{13}{10}, \quad a_0 = 2. \tag{5.13}$$

Этот выбор R, w, a_0 не может быть произвольным. Проверка различных значений R показывает, что для $w \leqslant R/4$ или $3R/4 \leqslant w$ p никогда не будет строго монотонно убывающей, каким бы не было $a_0 > 0$. Для $R/4 < w \leqslant R/2$ функция p строго убывает, но X нигде на интервале $(0, R)$ не обращается в нуль. Для $R/2 < w < 3R/4$ функция p строго убывает и X имеет нуль на $(0, R)$.

Определитель системы (5.10)–(5.12) относительно a_2, a_3, a_4 не равен нулю. Поэтому существует единственное решение системы (5.10)–(5.12)

$$a_2 = -\frac{39}{146}, \quad a_3 = -\frac{107}{146}, \quad a_4 = \frac{45}{146}.$$

Тогда

$$p(r) = 2 + \frac{1}{146}\left(-39r^2 - 107r^3 + 45r^4\right), \tag{5.14}$$

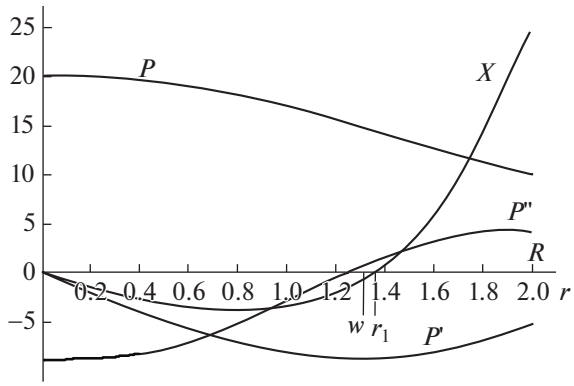
$$p'(r) = \frac{1}{146}\left(-78r - 321r^2 + 180r^3\right), \tag{5.15}$$

$$p''(r) = \frac{1}{146}\left(-78 - 642r + 540r^2\right) \tag{5.16}$$

(см. фиг. 4).

Из (5.15) следует, что $p'(r) < 0$ для $r \in (0, 2)$. Таким образом, $p \in \mathcal{D}_R^-(L)$. Для вычисления $P(r)$ заметим, что

$$L(r^l) = 4\pi\left(-\frac{r^{l+2}}{(l+2)(l+3)} + \frac{R^{l+2}}{l+2}\right), \quad l = 0, 1, 2, \ldots. \tag{5.17}$$



Фиг. 5. X, P, P' и P''.

Отсюда для $l = 0, 2, 3, 4$ непосредственно получим члены в круглых скобках следующей формулы:

$$P(r) = Lp(r) = 4\pi\left[a_0\left(-\frac{r^2}{6}+\frac{R^2}{2}\right)+a_2\left(-\frac{r^4}{20}+\frac{R^4}{4}\right)+a_3\left(-\frac{r^5}{30}+\frac{R^5}{5}\right)+a_4\left(-\frac{r^6}{42}+\frac{R^6}{6}\right)\right] =$$
$$= 4\pi\left[-\frac{1}{3}r^2+\frac{39}{2920}r^4+\frac{107}{4380}r^5-\frac{15}{2044}r^6+\frac{558}{365}\right].$$

Вычисляя $P'(r)$ и $P''(r)$ и подставляя эти производные в равенство $X(r) = p'(r)P''(r) - p''(r)P'(r)$, имеем

$$X(r) = \frac{4\pi}{746\,060}\left[-1093\,540r^2+1183\,812r^3-233\,688r^4-330\,515r^5+329\,025r^6-81\,000r^7\right].$$

Легко видеть, что функция $X = X(r)$ имеет нуль в точке $r_0 = 0$ и $X(1) < 0$, $X(2) > 0$. Поэтому существует нуль $X(r)$ в точке r_1, $1 < r_1 < 2$, с приближенным значением $r_1 \approx w + 0.0575585$ (см. фиг. 5). (Иногда мы работаем с эквивалентными дробями, чтобы избежать ошибок в округлении десятичных представлений.)

Наша цель – доказать существование точки $h \in (0, P(0) - E_0)$ такой, что $\frac{d}{dh}H_{F_0'}(h) < 0$. В силу теоремы 4.2 2) это означает, что p нерасширима. Поскольку $p'(r) = 0$, согласно формуле (4.7) знак $\frac{d}{dh}H_{F_0'}(h)$ равен знаку интеграла в правой части (4.7). Поэтому мы должны исследовать подынтегральную функцию

$$J(r,h) := \frac{X(r)}{|P'(r)|^2}\sqrt{\frac{P(\Phi(h)) - P(R)}{P(\Phi(h)) - P(r)}}.$$

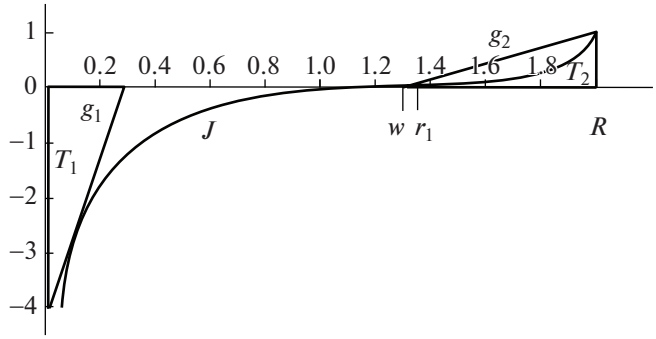
Как было доказано в доказательстве теоремы 4.2,

$$\Phi : [0, P(0) - E_0) \to (0, R], \quad h \mapsto \Phi(h) = P^{-1}(h + E_0)$$

является C^2-диффеоморфизмом. Если $r \in (\Phi(h), R)$, то $P(\Phi(h)) > P(r) > P(R)$. Следовательно, $P(\Phi(h)) - P(r) < P(\Phi(h)) - P(R)$, т.е.

$$\frac{P(\Phi(h)) - P(R)}{P(\Phi(h)) - P(r)} > 1 \quad \text{и} \quad \lim_{r\to\Phi(h)}\frac{P(\Phi(h)) - P(R)}{P(\Phi(h)) - P(r)} = \infty,$$

а поскольку $J(r,h)$ и $X(r)$ всегда имеют одинаковые знаки, то $J(r,h) < 0$ для $\Phi(h) < r < r_1$ и $J(r,h) > 0$ для $r_1 < r < R$. Так как Φ является биекцией $[0, P(0) - E_0)$ на $(0, R]$, для заданного $\tilde{r} := 0{,}01 \in (0, R)$, существует единственное $\tilde{h} \in [0, P(0) - E_0)$ такое, что $\Phi(\tilde{h}) = \tilde{r} = 0.01$.



Фиг. 6. Подынтегральная функция J, гипотенузы g_1, g_2 и треугольники T_1, T_2.

По лемме 2.2 $P(r) \in C^2(\mathbb{R}_+)$. Поэтому, используя формулу Тейлора, мы имеем $P(\Phi(\tilde{h})) - P(r) = P'(r + \theta(\Phi(\tilde{h}) - r))(\Phi(\tilde{h}) - r)$, где $0 < \theta < 1$. Пусть

$$c_0 := \min_{r\in[\tilde{r},R]} |P'(r)|, \quad c_1 := \left|P(\Phi(\tilde{h})) - P(R)\right|, \quad c_2 := \max_{r\in[\tilde{r},R]} |X(r)|.$$

Очевидно, $0 < c_i < \infty$ $(i = 0,1,2)$. Тогда

$$J(r,\tilde{h}) \leqslant \frac{\sqrt{c_1}c_2}{c_0^2}\frac{1}{\sqrt{\Phi(\tilde{h}) - r}}.$$

Поэтому функция $J(r,\tilde{h})$ интегрируема по r на интервале $(\Phi(\tilde{h}), R)$.

Покажем, что $\dfrac{d}{dh}H_{F_0^*}(h)\Big|_{h=\tilde{h}} < 0$, доказав следующее неравенство:

$$\int\limits_{\Phi(\tilde{h})}^{R} J(r,\tilde{h})dr < 0. \tag{5.18}$$

С этой целью введем два прямоугольных треугольника:

$$T_1 := \{(0.01;0),(0.01;-4),(0.31;0)\} \quad \text{и} \quad T_2 := \{(1.3;0),(2;1.21),(2;0)\}$$

с гипотенузами

$$g_1(r) = \frac{4}{0.3}(r - 0.31) \quad \text{при} \quad r \in [0.01, 0.31] \quad \text{и} \quad g_2(r) = \frac{1.21}{0.7}(r - 1.3) \quad \text{при} \quad r \in [1.3, 2]$$

(см. фиг. 6). Численные расчеты показывают, см. табл. 2 и табл. 3, что

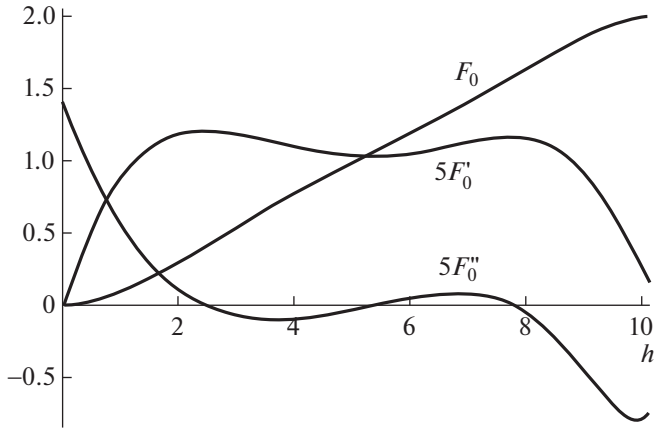
$$0 > g_1(r) > J(r,\tilde{h}) \quad \text{при} \quad r \in [0.01, 0.31] \quad \text{и} \quad 0 < J(r,\tilde{h}) < g_2(r) \quad \text{при} \quad r \in [1.3, 2].$$

Таблица 2. $g_1(r) - J(r,\tilde{h})$

r	0.01	0.1	0.15	0.175	0.2	0.31
$g_1(r) - J(r,\tilde{h})$	∞	0.832	0.182	0.129	0.173	0.937

Таблица 3. $g_2(r) - J(r,\tilde{h})$

r	1.3	1.5	1.7	1.9	2
$g_2(r) - J(r,\tilde{h})$	0.027	0.256	0.369	0.225	0.004



Фиг. 7. F_0, F_0', F_0''.

Площадь треугольника T_1 равна 0.6, а площадь T_2 равна 0.4235. Поэтому, подставляя значения $\Phi(\tilde{h}) = 0,01$ и $R = 2$, имеем

$$\int\limits_{\Phi(\tilde{h})}^{R} J(r,\tilde{h})dr = \int\limits_{0,01}^{0,31} J(r,\tilde{h})dr + \int\limits_{0,31}^{w} J(r,\tilde{h})dr + \int\limits_{w}^{2} J(r,\tilde{h})dr \leqslant \int\limits_{0,01}^{0.31} g_1(r)dr + 0 + \int\limits_{w}^{2} g_2(r)dr =$$
$$= -0.60 + 0.4235 = -0.1765 < 0.$$

Это завершает доказательство того, что функция p нерасширима.

Интересно сравнить при $R = 2$ функции F_0, F_0', F_0'' в примере 5.1 (p расширима) для $h \in [0, 28/15\pi]$ и в примере 5.6 (p нерасширима) для $h \in [0, \approx 10]$. Поскольку в примере 5.1 $F_0' > 0$ и $F_0'' < 0$, то $F_0 > 0$ строго возрастает и F_0' строго убывает. Функция F_0'', как показывает ее формула, строго убывает.

В примере 5.6 невозможно вычислить функцию P^{-1} в явном виде (как в примере 5.1), поскольку $P(r)$ монотонный неэлементарный полином четвертой степени. В этом случае, однако, существует следующая возможность: множество $\{(P(r_k), r_k) : 0 \leqslant r_k \leqslant R;\ k = 1,\ldots,l;\ l > 0\}$ содержится в графике функции P^{-1}. Поэтому можно аппроксимировать P^{-1} полиномом степени m ($l > 2m+1$) по методу наименьших квадратов. С помощью такой аппроксимации мы можем вычислить приближенно F, F_0, F_0', F_0'' (здесь мы опускаем вычисления).

Фигура 7 показывает графики функций F_0 и F_0', F_0'', умноженных на 5, поскольку области значений F_0 и F_0', F_0'' очень отличаются. Мы видим, что $F_0 \geqslant 0$ строго монотонна, $F_0' \geqslant 0$ не является монотонной, а F_0'' не является монотонной и меняет знак.

6. ПРЯМАЯ ЗАДАЧА И ЕЕ ЭКВИВАЛЕНТНАЯ ФОРМУЛИРОВКА

Во введении мы уже упоминали (в связи с теоремой Джинса), что мы понимаем под прямой задачей: при каких условиях существуют функции (f_q, ρ, u), которые задают сферически симметричное, зависящее от E решение системы (VPS), если задана функция $q \in L^1_{\mathrm{loc}}[0,T)$, $q > 0$ на $[0,T)$ для достаточно большого интервала $[0,T)$?

Первый ответ дается в лемме 3.2 б): эти функции существуют, если для функции

$$F(h) := 4\pi\sqrt{2}\int\limits_0^{h-E_0} q(s)\sqrt{h - E_0 - s}\,ds, \quad h \in [E_0, P(0)),$$
$$F = 0 \quad \text{на} \quad (-\infty, E_0),$$

интегральное уравнение

$$p(r) = F \circ Lp(r) \quad \text{на} \quad \mathbb{R}_+ \tag{6.1}$$

имеет решение $p \in \mathcal{D}_R^-(L)$. В следующей лемме приведены свойства задействованных функций F_0, F и p.

Лемма 6.1. *Будем предполагать $T > P(0) - E_0$ и определим*

$$F_0(h) := F(h + E_0) = 4\pi\sqrt{2}\int_0^h q(s)\sqrt{h-s}\,ds, \quad 0 \leqslant h < P(0) - E_0. \tag{6.2}$$

Тогда F_0 строго возрастает на $[0, P(0) - E_0)$, и так же F на $[E_0, P(0))$. Решение $p \in \mathcal{D}(L)$ уравнения (6.1) *такое, что $Lp(R) = E_0$ содержится в $\mathcal{D}_R^-(L)$.*

Доказательство. Из леммы А.5 (б) следует, что $F_0 \in AC[0, P(0) - E_0)$ и F_0 строго возрастает, так как

$$F_0'(h) = \frac{4\pi}{\sqrt{2}}\int_0^h \frac{q(s)}{\sqrt{h-s}}\,ds > 0 \quad \text{п.в. на} \quad [0, P(0) - E_0).$$

Покажем, что $p \in \mathcal{D}_R^-(L)$. Если $0 < r_1 < r_2 < R$, тогда $Lp(r_1) > Lp(r_2)$ (лемма 2.1 3)) и $p(r_1) = F \circ Lp(r_1) > F \circ Lp(r_2) = p(r_2)$, так, что p строго убывает на $(0, R)$. Используя соотношения $F(h) = 0$ для $h \in (-\infty, E_0)$, $Lp(R) = E_0$ и $Lp(r) < Lp(R)$ для $r > R$, имеем $p(r) = F \circ Lp(r) = 0$ $(r > R)$.

Решение прямой задачи тесно связано с решением нелинейного интегрального уравнения

$$p(r) = F \circ Lp(r) \quad \text{на} \quad \mathbb{R}_+,$$

или – что то же самое согласно следствию 2.1 – на $(0, R]$. Чтобы сделать это уравнение более доступным для численных исследований, преобразуем его в несколько иную форму. Так как функция

$$F_0 : [0, P(0) - E_0) \to [0, p(0))$$

является строго возрастающей биекцией (лемма 6.1), она имеет строго возрастающую обратную функцию

$$G_0 : [0, p(0)) \to [0, P(0) - E_0),$$

а функция

$$F : [E_0, P(0)) \to [0, p(0))$$

имеет обратную

$$G : [0, p(0)) \to [E_0, P(0)),$$

которая имеет вид $G(h) = G_0(h) + E_0$. Тогда получаем два эквивалентных утверждения в следующей лемме.

Лемма 6.2. *Пусть $p \in \mathcal{D}_R^-(L)$ и $Lp(R) = E_0$. Тогда*

$$p = F \circ Lp \Leftrightarrow Lp - E_0 = G_0 p \quad \textit{на} \quad [0, R].$$

Доказательство. $\Rightarrow$: Так как $Lp(R) = E_0$, из неравенства $r \leqslant R$ следует, что $Lp(r) \geqslant E_0$. Применим G к обеим частям уравнения $p = F \circ Lp$. Тогда $Gp = Lp$ или $Lp = G_0 p + E_0$.

$\Leftarrow$: Если $Lp - E_0 = G_0 p$, тогда $Lp = G_0 p + E_0 = G(p)$, и применение F дает $F \circ Lp = F \circ Gp = p$.

Теперь пусть функции q и F будут такими же, как в начале этого раздела. Наша цель – решить интегральное уравнение (6.1) или

$$Lp(r) - Lp(R) = G_0 p(r) \quad \text{на} \quad [0, R], \tag{6.3}$$

что эквивалентно в соответствии с леммой 6.2. Первая цель состоит в том, чтобы рассмотреть (6.3) в равноудаленных точках $r = R_k$

$$0 = R_0 < R_1 < R_2 < \ldots < \ldots < R_n := R, \quad R_k := k\frac{R}{n}, \quad k = 0,\ldots,n \quad \text{в} \quad R.$$

Для этого введем пространство кусочно-линейных функций

$$p^{x_0x_1\ldots x_{n-1}x_n}_{R_0R_1\ldots R_{n-1}R_n}(r) := -\frac{r-R_k}{R_k-R_{k-1}}x_{k-1} + \frac{r-R_{k-1}}{R_k-R_{k-1}}x_k \quad \text{на} \quad [R_{k-1},R_k],$$

$k = 1,\ldots,n$, со свойством

$$p^{x_0x_1\ldots x_{n-1}x_n}_{R_0R_1\ldots R_{n-1}R_n}(R_k) = x_k, \quad k = 0,\ldots,n.$$

Базисом в этом пространстве является множество

$$\left\{p^{010}_{R_{k-1}R_kR_{k+1}},\ k = 0,1,\ldots,n-1\right\},$$

где для единообразия обозначений определим

$$p^{010}_{R_{k-1}R_kR_{k+1}}\Big|_{k=0} := p^{10}_{R_0R_1}$$

(R_{-1} не определено). Любая кусочно-линейная функция p с $0 = p(R) = p(R_n) = x_n$ может быть представлена в виде

$$p(r) = \sum_{k=0}^{n-1} x_k p^{010}_{R_{k-1}R_kR_{k+1}}, \tag{6.4}$$

и ее образ под действием L имеет вид

$$Lp(r) = \sum_{k=1}^{n-1} x_k Lp^{010}_{R_{k-1}R_kR_{k+1}}(r).$$

Если $(x_0,x_1,\ldots,x_{n-1})$ является решением системы

$$\sum_{k=0}^{n-1}\left(Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i) - Lp^{010}_{R_{k-1}R_kR_{k+1}}(R)\right)x_k = G_0(x_i), \tag{ANS}$$

$i = 0,1,\ldots,n-1$, p определено по формуле (6.4) и $Lp(R) =: E_0$, то

$$Lp(r) - E_0 = G_0(p(r)) \quad \text{для} \quad r = R_0,R_1,\ldots,R_{n-1}.$$

Система также справедлива при $r = R$, так как $Lp(R) = E_0$ и $G_0(p(R)) = 0$. Будем называть (ANS) *аппроксимирующей нелинейной системой*. Это будет темой следующего раздела.

7. АППРОКСИМИРУЮЩАЯ НЕЛИНЕЙНАЯ СИСТЕМА (ANS)

Аппроксимирующая нелинейная система (ANS) имеет вид

$$Ax = G_0(x),$$

где для вектора $x \in \mathbb{R}^n$ мы определим $G_0(x) := (G_0(x_i))_{i=0}^{n-1}$, A является $n\times n$ матрицей с коэффициентами A_{ik},

$$A_{ik} := Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i) - Lp^{010}_{R_{k-1}R_kR_{k+1}}(R) =: B_{ik} - C_k, \quad i,k = 0,1,\ldots,n-1. \tag{7.1}$$

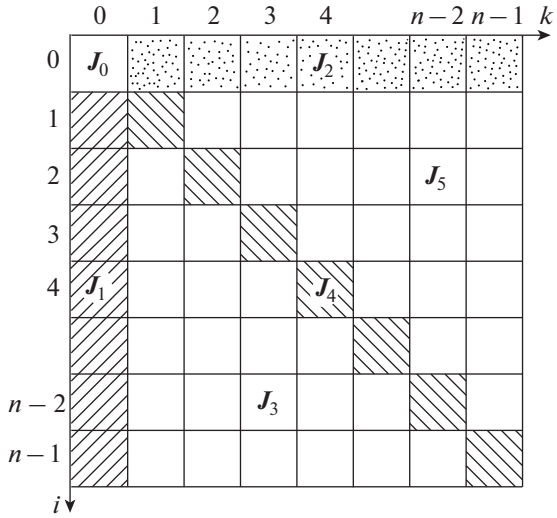
Выражения $Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i)$ состоят из элементов вида

(i) $L_{\mathrm{I}}p^{01}_{QS}(r) := Lp^{01}_{QS}(r)$, $0 \leqslant r \leqslant Q < S$,

(ii) $L_{\mathrm{I}}p^{10}_{ST}(r) := Lp^{10}_{ST}(r)$, $0 \leqslant r \leqslant S < T$,

(iii) $L_{\mathrm{II}}p^{01}_{QS}(r) := Lp^{01}_{QS}(r)$, $0 \leqslant Q < S \leqslant r$,

(iv) $L_{\mathrm{II}}p^{10}_{ST}(r) := Lp^{10}_{ST}(r)$, $0 \leqslant S < T \leqslant r$,



Фиг. 8.

где $p_{QS}^{01}(s) = \dfrac{s-Q}{S-Q}\chi_{[Q,S]}(s)$, $p_{ST}^{10}(s) = \dfrac{s-T}{S-T}\chi_{[S,T]}(s)$. Если $p \in L^{\infty}(R_+)$ и $sp(s) \in L^1(R_+)$, то

$$Lp(0) := \lim_{r\downarrow 0} Lp(r) = 4\pi \cdot \lim_{r\downarrow 0}\left[\frac{1}{r}\int_0^r p(s)s^2\,ds + \int_r^{\infty} p(s)s\,ds\right] 4\pi\int_0^{\infty} p(s)s\,ds.$$

Для $a, b \in \mathbb{R}$ и $n \geqslant 2$ применим формулу

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \ldots ab^{n-2} + b^{n-1}). \tag{7.2}$$

Для вычисления выражений (i)–(iv) требуется следующая лемма.

Лемма 7.1. (i) *Пусть* $0 \leqslant r \leqslant R < S$. *Тогда в силу* (7.2)

$$L_{\mathrm{I}} p_{QS}^{01}(r) = \frac{4\pi}{S-Q}\int_Q^S (s-Q)s\,ds = \frac{4\pi}{S-Q}\left(\frac{s^3}{3} - Q\frac{s^2}{2}\bigg|_Q^S\right) = \frac{4\pi}{S-Q}\left[\frac{1}{3}(S^3-Q^3) - \frac{Q}{2}(S^2-Q^2)\right] =$$

$$= 4\pi\left[\frac{1}{3}(S^2+SQ+Q^2) - \frac{Q}{2}(S+Q)\right] = 4\pi\left[\frac{1}{3}S^2 - \frac{1}{6}SQ - \frac{1}{6}Q^2\right] = \frac{4\pi}{3}\left[S^2 - \frac{1}{2}Q(S+Q)\right].$$

(ii) *Пусть* $0 \leqslant r \leqslant S < T$. *Тогда*

$$L_{\mathrm{I}} p_{ST}^{10}(r) = 4\pi \cdot \frac{(-1)}{T-S}\int_S^T (s-T)s\,ds = -\frac{4\pi}{T-S}\left(\frac{s^3}{3} - T\frac{s^2}{2}\right)\bigg|_S^T = -\frac{4\pi}{T-S}\left[\frac{1}{3}(T^3-S^3) - \frac{1}{2}T(T^2-S^2)\right] =$$

$$= -4\pi\left[\frac{1}{3}(T^2+TS+S^2) - \frac{1}{2}T(T+S)\right] = \frac{4\pi}{3}\left[\frac{1}{2}T(T+S) - S^2\right].$$

(iii) *Пусть* $0 \leqslant Q < S \leqslant r$. *Тогда*

$$L_{\mathrm{II}} p_{QS}^{01}(r) = \frac{4\pi}{r}\frac{1}{S-Q}\int_Q^S (s-Q)s^2\,ds = \frac{4\pi}{r}\frac{1}{S-Q}\left(\frac{s^4}{4} - Q\frac{s^3}{3}\right)\bigg|_Q^S = \frac{4\pi}{r}\frac{1}{S-Q}\left[\frac{1}{4}(S^4-Q^4) - \frac{Q}{3}(S^3-Q^3)\right] =$$

$$= \frac{4\pi}{r}\left[\frac{1}{4}(S^3+S^2Q+SQ^2+Q^3) - \frac{Q}{3}(S^2+QS+Q^2)\right] = \frac{\pi}{r}\left[S^3 - \frac{1}{3}(S^2Q+SQ^2+Q^3)\right].$$

(iv) *Пусть* $0 \leqslant S < T \leqslant r$. *Тогда*

$$L_{\mathrm{II}} p_{ST}^{10}(r) = \frac{-4\pi}{r}\frac{1}{T-S}\int_S^T (s-T)s^2\,ds = -\frac{4\pi}{r}\frac{1}{T-S}\left(\frac{s^4}{4} - \frac{Ts^3}{3}\bigg|_S^T\right) =$$

$$= -\frac{4\pi}{r}\frac{1}{T-S}\left[\frac{1}{4}(T^4-S^4) - \frac{T}{3}(T^3-S^3)\right] =$$

$$= -\frac{4\pi}{r}\left[\frac{1}{4}(T^3+ST^2+S^2T+S^3) - \frac{T}{3}(T^2+ST+S^2)\right] = \frac{\pi}{r}\left[\frac{1}{3}(T^3+T^2S+TS^2) - S^3\right].$$

Чтобы вычислить B_{ik} и C_k, разделим квадрат $\{(i,k) : i,k = 0,1,2,\ldots,n-1\}$ на части $J_0,\ldots,J_5$:

$J_0 := \{(0,0)\}$,

$J_1 := \{(i,0) : i = 1,\ldots,n-1\}$,

$J_2 := \{(0,k) : k = 1,\ldots,n-1\}$,

$J_3 := \{(i,k) : 2 \leqslant i \leqslant n-1, 1 \leqslant k \leqslant i-1\}$,

$J_4 := \{(i,k) : i = k = 1,\ldots,n-1\}$,

$J_5 := \{(i,k) : 1 \leqslant i \leqslant k-1, 2 \leqslant k \leqslant n-1\}$.

Лемма 7.2. *Вычисление* $B_{ik} = Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i)$.

$J_0 = \{0,0\}$. *Согласно лемме* 7.1 (ii)

$$B_{00} = L_{\mathrm{I}}p^{10}_{0R_1}(0) = \frac{4\pi}{3}\frac{1}{2}R_1^2 = \frac{2\pi}{3}\left(\frac{R}{n}\right)^2.$$

$J_1 = \{(i,0) : i = 1,\ldots,n-1\}$. *Согласно лемме* 7.1 (iv)

$$B_{i0} = L_{\mathrm{II}}p^{10}_{0R_1}(R_i) = \frac{\pi}{i\dfrac{R}{n}}\frac{1}{3}\left(\frac{R}{n}\right)^3 = \frac{\pi}{3}\frac{1}{i}\left(\frac{R}{n}\right)^2.$$

$J_2 = \{(0,k) : k = 1,\ldots,n-1\}$. *Согласно лемме* 7.1 (i), (ii)

$$B_{0k} = Lp^{010}_{R_{k-1}R_kR_{k+1}}(0) = L_{\mathrm{I}}p^{01}_{R_{k-1}R_k}(0) + L_{\mathrm{I}}p^{10}_{R_kR_{k+1}}(0) =$$
$$= \frac{4\pi}{3}\left[R_k^2 - \frac{1}{2}R_{k-1}(R_k + R_{k-1})\right] + \frac{4\pi}{3}\left[\frac{1}{2}R_{k+1}(R_{k+1} + R_k) - R_k^2\right] =$$
$$= \frac{4\pi}{3}\left(\frac{R}{n}\right)^2\left[k^2 - \frac{1}{2}(k-1)(2k-1) + \frac{1}{2}(k+1)(2k+1) - k^2\right] =$$
$$= \frac{2\pi}{3}\left(\frac{R}{n}\right)^2\left[-(2k^2 - k - 2k + 1) + (2k^2 + k + 2k + 1)\right] = 4\pi k\left(\frac{R}{n}\right)^2.$$

$J_3 = \{(i,k) : 2 \leqslant i \leqslant n-1, 1 \leqslant k \leqslant i-1\}$. *В силу леммы* 7.1 (iii), (iv)

$$B_{ik} = Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i) = L_{\mathrm{II}}p^{01}_{R_{k-1}R_k}(R_i) + L_{\mathrm{II}}p^{10}_{R_kR_{k+1}}(R_i) =$$
$$= \frac{\pi}{R_i}[R_k^3 - \frac{1}{3}\left(R_k^2R_{k-1} + R_kR_{k-1}^2 + R_{k-1}^3\right) + \frac{1}{3}\left(R_{k+1}^3 + R_{k+1}^2R_k + R_{k+1}R_k^2\right) - R_k^3] =$$
$$= \frac{\pi}{i}\frac{1}{3}\left(\frac{R}{n}\right)^2\left[-\left(k^2(k-1) + k(k-1)^2 + (k-1)^3\right) + (k+1)^3 + (k+1)^2k + (k+1)k^2\right] =$$
$$= \frac{\pi}{i}\frac{1}{3}\left(\frac{R}{n}\right)^2\left[12k^2 + 2\right] = 4\pi\left(\frac{R}{n}\right)^2\frac{1}{i}\left(k^2 + \frac{1}{6}\right).$$

$J_4 = \{(i,k) : i = k = 1,\ldots,n-1\}$. *В силу леммы* 7.1 (iii), (ii)

$$B_{kk} = Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_k) = L_{\mathrm{II}}p^{01}_{R_{k-1}R_k}(R_k) + L_{\mathrm{I}}p^{10}_{R_kR_{k+1}}(R_k) =$$
$$= \frac{\pi}{R_k}\left[R_k^3 - \frac{1}{3}\left(R_k^2R_{k-1} + R_kR_{k-1}^2 + R_{k-1}^3\right)\right] + \frac{4\pi}{3}\left[\frac{1}{2}R_{k+1}(R_{k+1} + R_k) - R_k^2\right] =$$
$$= \frac{\pi}{k}\left(\frac{R}{n}\right)^2\left[k^3 - \frac{1}{3}\left(k^2(k-1) + k(k-1)^2 + (k-1)^3\right)\right] + \frac{4\pi}{3}\left(\frac{R}{n}\right)^2\left[\frac{1}{2}(k+1)\big((k+1) + k\big) - k^2\right] =$$
$$= \frac{\pi}{k}\left(\frac{R}{n}\right)^2\left[2k^2 - \frac{4}{3}k + \frac{1}{3}\right] + 4\pi\left(\frac{R}{n}\right)^2\frac{1}{3}\left[\frac{3}{2}k + \frac{1}{2}\right] = 4\pi\left(\frac{R}{n}\right)^2\left[k - \frac{1}{6} + \frac{1}{12k}\right].$$

$J_5 = \{(i,k) : 1 \leqslant i \leqslant k-1,\ 2 \leqslant k \leqslant n-1\}$. *В силу леммы* 7.1 (i), (ii)

$$B_{ik} = Lp^{010}_{R_{k-1}R_kR_{k+1}}(R_i) = L_1 p^{01}_{R_{k-1}R_k}(R_i) + L_1 p^{10}_{R_kR_{k+1}}(R_i) =$$
$$= \frac{4\pi}{3}\left[R_k^2 - \frac{1}{2}R_{k-1}(R_k + R_{k-1})\right] + \frac{4\pi}{3}\left[\frac{1}{2}(R_{k+1}(R_{k+1} + R_k) - R_k^2\right] =$$
$$= \frac{4\pi}{3}\left(\frac{R}{n}\right)^2\left[k^2 - \frac{1}{2}(k-1)(2k-1) + \frac{1}{2}(k^2 + 2k + 1 + k^2 + k) - k^2\right] =$$
$$= \frac{4\pi}{3}\left(\frac{R}{n}\right)^2\left[k^2 - \frac{1}{2}(2k^2 - k - 2k + 1) + \frac{1}{2}(2k^2 + 3k + 1) - k^2\right] = 4\pi\left(\frac{R}{n}\right)^2 k.$$

Лемма 7.3. *Вычисление* $C_k := Lp^{010}_{R_{k-1}R_kR_{k+1}}(R)$. *Из леммы* 7.1 (iv) *следует, что*

$$C_0 = L_{11} p^{10}_{R_0R_1}(R) = \frac{\pi}{3R}\left(\frac{R}{n}\right)^3.$$

В силу леммы 7 (iii), (iv)

$$C_k = Lp^{010}_{R_{k-1}R_kR_{k+1}}(R) = L_{11} p^{01}_{R_{k-1}R_k}(R) + L_{11} p^{10}_{R_kR_{k+1}}(R) =$$
$$= \frac{\pi}{R}\left[R_k^3 - \frac{1}{3}\left(R_k^2R_{k-1} + R_{k-1}^2R_k + R_{k-1}^3\right) + \frac{1}{3}\left(R_{k+1}^3 + R_{k+1}^2R_k + R_{k+1}R_k^2\right) - R_k^3\right] =$$
$$= \frac{\pi}{R}\left(\frac{R}{n}\right)^3\left[-\frac{1}{3}\left(k^2(k-1) + k(k-1)^2 + (k-1)^3\right) +\right.$$
$$\left. + \frac{1}{3}\left((k+1)^3 + (k+1)^2k + (k+1)k^2\right)\right] = \frac{4\pi}{n}\left(\frac{R}{n}\right)^2\left[k^2 + \frac{1}{6}\right].$$

Следствие 7.1.

$$A_{ik}([0,R]) = R^2 A_{ik}([0,1]), \quad \text{где} \quad A_{ik} = B_{ik} - C_k.$$

8. ЧИСЛЕННЫЙ АНАЛИЗ АППРОКСИМИРУЮЩЕЙ НЕЛИНЕЙНОЙ СИСТЕМЫ (ANS)

Цель этого раздела – показать применение некоторых численных методов для решения системы

$$Ax = G_0(x) \tag{ANS}$$

с матрицей $A = (A_{ik})_{i,k=0}^{n-1}$, $A_{ik} := B_{ik} - C_k$, $i,k = 0,\ldots,n-1$, где $G_0(x) := (G_0(x_i))_{i=0}^{n-1}$, $x \in (\mathbb{R}_+)^n$ для скалярной функции $G_0 : [0, p(0)) \to [0, P(0) - E_0)$. Матрица A зависит от n ($A = A(n)$) и может быть вычислена по формулам из разд. 7. Решение $x = (x_0, x_1, \ldots, x_{n-1})^{\mathrm{T}}$ представляет значения $x_0, x_1, \ldots, x_{n-1}$ многоугольника аппроксимации $p^{x_0x_1\ldots x_{n-1}}_{R_0R_1\ldots R_{n-1}}$ для $R_0 < R_1 < \ldots < R_{n-1}$ с $p^{x_0x_1\ldots x_{n-1}}_{R_0R_1\ldots R_{n-1}}(R) = 0$ и $x_0, x_1, \ldots, x_{n-1}$, $x_n = 0$, $R_k := \frac{k}{n}R$, $k = 0,1,\ldots,n-1$.

Для нахождения решения x системы (ANS) используем метод Ньютона для уравнения $Ax - G_0(x) = 0$. Для сходимости метода Ньютона решающее значение имеет подходящий выбор начальной точки $x^{(0)}$. В последовательности разбиений

$$\pi_{2^i} := \left\{\frac{k}{2^i}R,\ k = 0,\ldots,2^i\right\}, \quad i = 0,1,2,\ldots,$$

используем решение $p_{\pi_{2^i}}$ в качестве начальной точки для следующего разбиения $\pi_{2^{i+1}}$ при n=2,4,8,16,32,64,128. Тогда критерий Канторовича [11] (который во многом зависит от $x^{(0)}$ и $\left\|x^{(0)} - x^{(1)}\right\|$) дает сходимость последовательностей в следующих двух примерах. Для $n = 2^0 = 1$ $Ax - G_0(x) = 0$ представляет собой одномерное нелинейное уравнение. Чтобы найти решение

Таблица 4. Аппроксимации $p_n(r)$ функции $p(r) = 1 - \left(\frac{r}{R}\right)^2$ для $R = 8$ и $n = 2, 4, 8, 16, 32, 64, 128$

r/n	2	4	8	16	32	64	128	$p(r) - p_{128}(r)$
0.0	0.83	0.95	0.989	0.99712	0.99928	0.99982	0.99995	4.5×10^{-5}
0.5	–	–	–	0.99322	0.99537	0.99591	0.99605	4.5×10^{-5}
1.0	–	–	0.973	0.98152	0.98366	0.98420	0.98433	4.5×10^{-5}
1.5	–	–	–	0.96203	0.06414	0.96467	0.96480	4.4×10^{-5}
2.0	–	0.89	0.926	0.93474	0.93681	0.93733	0.93746	4.3×10^{-5}
2.5	–	–	–	0.89965	0.90167	0.90217	0.90230	4.2×10^{-5}
3.0	–	–	0.849	0.85676	0.85872	0.85921	0.85933	4.1×10^{-5}
3.5	–	–	–	0.80608	0.80796	0.80844	0.80855	3.9×10^{-5}
4.0	0.62	0.71	0.740	0.74761	0.74949	0.74985	0.74996	3.7×10^{-5}
4.5	–	–	–	0.68135	0.68303	0.68345	0.68356	3.5×10^{-5}
5.0	–	–	0.601	0.60730	0.60886	0.60925	0.60934	3.2×10^{-5}
5.5	–	–	–	0.52547	0.52688	0.52723	0.52731	2.9×10^{-5}
6.0	–	0.41	0.431	0.43587	0.43709	0.43740	0.43747	2.6×10^{-5}
6.5	–	–	–	0.33850	0.33951	0.33976	0.33982	2.1×10^{-5}
7.0	–	–	0.230	0.23338	0.23413	0.23431	0.23436	1.6×10^{-5}
7.5	–	–	–	0.12053	0.12095	0.12106	0.12108	0.9×10^{-5}
8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\|p - p_n\|$	0.45	0.13	0.032	0.00815	0.00204	0.00051	0.00012	
E_{0n}	79.1	98.9	105.1	106.680	107.094	107.198	107.224	$E_0 = 107.233$
Ошибка, %	26.2	7.73	2.034	0.052	0.013	0.003	0.001	

этого уравнения, можно применить метод вложенных интервалов путем деления пополам (бисекции). Мы использовали Maple [12] для приближенных вычислений и проверки формул.

Пример 5.1 (продолжение). В этом примере вычислим оператор G_0 явно. Действительно, в примере 5.1 было показано, что

$$F : [E_0, P(0)] = \left[\frac{8}{15}\pi R^2, \pi R^2\right] \to [0,1], \quad \text{где} \quad F(h) = \sqrt{ah - \frac{20}{9}} - \frac{2}{3}, \quad a := \frac{5}{\pi R^2}.$$

Также было доказано, что

$$F_0 : [0, P(0) - E_0] = \left[0, \frac{7}{15}\pi R^2\right] \to [0,1], \quad \text{где} \quad F_0(h) = \sqrt{ah + \frac{4}{9}} - \frac{2}{3}.$$

Поэтому

$$G_0 := F_0^{-1} : [0,1] \to \left[0, \frac{7}{15}\pi R^2\right], \quad \text{где} \quad G_0(t) = \frac{\pi R^2}{5}\left(t^2 + \frac{4}{3}t\right).$$

При описанном выборе начальных точек $x^{(0)}$ метод Ньютона сходится для $n = 2, 4$; для $n = 8$, 16, 32, 64, 128 критерий Канторовича гарантирует сходимость. Для каждого n мы выполняли итерации до тех пор, пока норма поправки Ньютона не покажет относительную ошибку по отношению к последней итерации величины 10^{-9} – это было достигнуто менее чем за 5 итераций.

Таблица 4 иллюстрирует сходимость многоугольников p_n как решений (ANS) на отрезке $[0, R]$ для $R = 8$ к точному решению p уравнения $p(r) = FLp(r)$ или $Lp - E_0 = G_0(p)$, которое в данном случае известно как $p(r) = 1 - \left(\frac{r}{R}\right)^2$.

Таблица 5. Аппроксимации $p(r)$ для $G_0(t) = 4\pi\sqrt{1000}\sqrt{t}$ при $R = 8$ и $n = 2, 4, 8, 16, 32, 64, 128$

r/n	2	4	8	16	32	64	128	$\lvert p_{64} - p_{128}\rvert / p_{128}$, %
0.0	55.55	71.64	82.615	86.30120	87.30561	87.56242	87.62698	0.074
0.5	–	–	–	84.22649	85.18691	85.43273	85.49454	0.072
1.0	–	–	75.227	78.31755	79.16157	79.37742	79.43169	0.068
1.5	–	–	–	69.44784	70.12831	70.30214	70.34583	0.062
2.0	–	51.37	56.948	58.80793	59.31085	59.43914	59.47138	0.054
2.5	–	–	–	47.62660	47.96641	48.05298	48.07473	0.045
3.0	–	–	36.160	36.93813	37.14661	37.19968	37.21300	0.036
3.5	–	–	–	27.45442	27.56851	27.59761	27.60492	0.026
4.0	18.69	18.73	19.345	19.54416	19.59812	19.61189	19.61535	0.018
4.5	–	–	–	13.29178	13.31145	13.31652	13.31780	0.010
5.0	–	–	8.578	8.58510	8.58808	8.58891	8.58912	0.02
5.5	–	–	–	5.20257	5.20247	5.20170	5.20152	0.004
6.0	–	3.01	2.917	2.89867	2.89447	2.89345	2.89319	0.009
6.5	–	–	–	1.41754	1.41453	1.41379	1.41361	0.013
7.0	–	–	0.555	0.54872	0.54725	0.54689	0.54680	0.016
7.5	–	–	–	0.11996	0.11958	0.11949	0.11947	0.019
8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000
$\lVert p_n \rVert$	80.19	102.73	117.8	122.826	124.204	124.557	124.646	
Норма–ошибка, %	21.943	12.77	4.117	1.11	0.283	0.071		
E_{0n}	2657	2155	2081	2065.5	2061.21	2060.90	2060.75	
E_0–Ошибка, %	23.300	3.562	0.752	0.208	0.015	0.007		

В центре таблицы перечислены значения $p_n(r_k)$ ($r_k = 0, 0.5, 1, \ldots, 7.5, 8$) многоугольников аппроксимации для $n = 2, 4, \ldots, 128$. Последний столбец содержит разности $p(r_k) - p_{128}(r_k)$, которые меньше, чем 5×10^{-5}, и показывает поточечную сходимость $p_n(r_k) \to p(r_k)$.

Третья с конца строка дает L_2-нормы разностей $p - p_n$, показывая их сходимость к 0 и, в частности, выявляет тот факт, что удвоение n приводит к делению L_2-норм на 4.

В последних двух строках перечислены значения E_{0n} и относительной ошибки $|E_0 - E_{0n}|/|E_0|$ в процентах. Удвоение n также приводит к коэффициенту сходимости $\frac{1}{4}$.

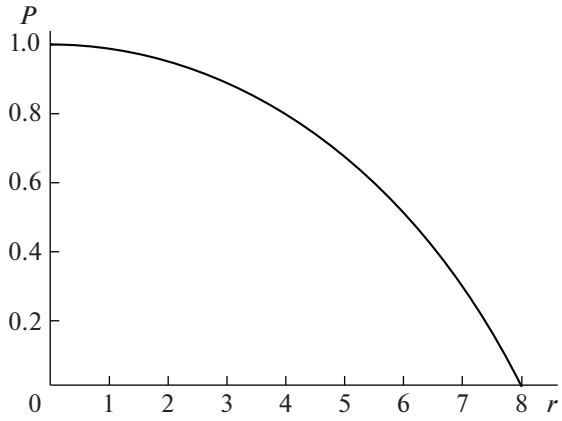
Пример 5.5 (продолжение). Сначала мы должны вычислить $G_0(t) = F_0^{-1}(h)$. Равенство $F_0(h) = \frac{\pi^2 c}{\sqrt{2}}h^2$ подразумевает, что $G_0(t) = \frac{\sqrt[4]{2}}{\pi\sqrt{c}}\sqrt{t}$. Для получения сопоставимых и четко представимых значений x решений в табл. 5 выбираем $c = \sqrt{2}/(4^2\pi^4 1000)$.

В этом примере мы вычисляем решения p_n, как и в первом примере, с помощью метода Ньютона. Критерий Канторовича гарантирует его сходимость для $n = 4, 6, \ldots, 128$. Результаты показаны в табл. 5, которая отличается от табл. 4 следующим.

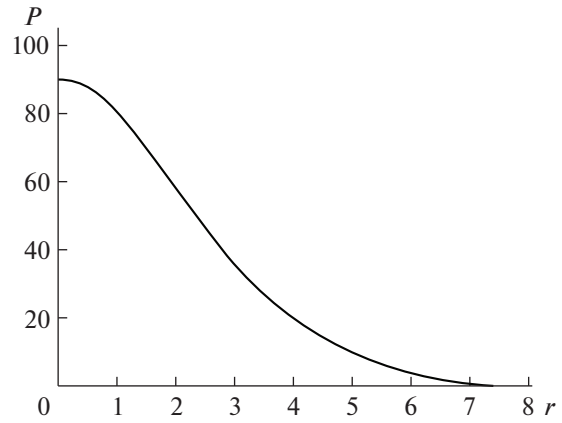
1. В последнем столбце показана относительная ошибка (в процентах) $|p_{128}(r_k) - p_{64}(r_k)|/p_{128}(r_k)\times100$.

2. Третья с конца строка дает ошибку нормы в процентах $= 100\times\left|\lVert p_n\rVert_2 - \lVert p_{2n}\rVert_2\right| / \lVert p_{2n}\rVert_2$.

3. В последней строке E_0 – ошибка в процентах $= 100\times|E_{0,n} - E_{0,2n}|/E_{0,2n}$.



Фиг. 9. Многоугольник p_{128} в табл. 4.



Фиг. 10. Многоугольник p_{128} в табл. 5.

9. УРАВНЕНИЯ АБЕЛЯ И ЭДДИНГТОНА

Цель этого раздела – доказательство существования решения уравнения Эддингтона. Оно основано на доказательстве существования решения уравнения Абеля, которое мы рассматриваем вначале.

Будем использовать следующие обозначения. Для интервала $[0,T)$ положим

$$L^1_{\mathrm{loc}}[0,T) := \left\{p : [0,T) \to \mathbb{R};\ p\big|_{[0,a]} \in L^1[0,a] \text{ для всех } a \in [0,T)\right\},$$

$$AC[0,T) := \{p : [0,T) \to \mathbb{R};\ p\big|_{[0,a]} \text{ абсолютно непрерывна на } [0,a] \text{ для всех } a \in [0,T)\}.$$

Начнем с двух лемм, а затем вернемся к уравнению Абеля.

Лемма A.1. *Пусть* $f \in L^1_{\mathrm{loc}}[0,T)$. *Тогда функция*

$$g(x) := \int_0^x \frac{f(s)ds}{\sqrt{x-s}}ds, \quad x \in [0,T),$$

принадлежит $L^1_{\mathrm{loc}}[0,T)$.

Доказательство. Предположим, что $f \geqslant 0$. Используя равенство

$$\int_a^b \frac{d\sigma}{\sqrt{b-\sigma}\sqrt{\sigma-a}} = \pi, \quad 0 \leqslant a < b,$$

и теорему Фубини, получаем

$$\infty > \int_0^x f(s)ds = \frac{1}{\pi}\int_0^x f(s)\int_s^x \frac{1}{\sqrt{x-\sigma}\sqrt{\sigma-s}}d\sigma ds = \frac{1}{\pi}\int_0^x\left(\int_0^\sigma \frac{f(s)}{\sqrt{\sigma-s}}ds\right)\frac{1}{\sqrt{x-\sigma}}d\sigma = \frac{1}{\pi}\int_0^x g(\sigma)\frac{1}{\sqrt{x-\sigma}}d\sigma.$$

Следовательно, $\left(\sigma \to g(\sigma)\frac{1}{\sqrt{x-\sigma}}\right) \in L^1_{\mathrm{loc}}[0,T)$ и $g \in L^1_{\mathrm{loc}}[0,T)$.

Лемма A.2. *Для любых $f, g \in L^1_{\mathrm{loc}}[0,T)$ равенство*

$$\int_0^x \frac{f(s)}{\sqrt{x-s}}ds = g(x) \quad \text{п.в. на} \quad (0,T) \tag{A.1}$$

справедливо тогда и только тогда, когда

$$\int_0^x f(s)\,ds = \frac{1}{\pi}\int_0^x \frac{g(s)}{\sqrt{x-s}}ds \quad \text{на} \quad (0,T). \tag{A.2}$$

Доказательство. Пусть выполняется (A.1). Тогда

$$\int_0^x \frac{g(s)}{\sqrt{x-s}}ds = \int_0^x\left(\int_0^s \frac{f(\sigma)}{\sqrt{s-\sigma}}d\sigma\right)\frac{1}{\sqrt{x-s}}ds = \int_0^x f(\sigma)\left(\int_\sigma^x \frac{1}{\sqrt{x-s}\sqrt{s-\sigma}}ds\right)d\sigma = \pi\int_0^x f(\sigma)d\sigma,$$

т.е. (A.2) выполняется.

Предположим теперь, что (A.2) выполняется.

Пусть $h(x) := \int_0^x \frac{f(s)}{\sqrt{x-s}}ds - g(x)$. Нам нужно показать, что $h = 0$. Из леммы A.1 следует, что $h \in L^1_{loc}[0,T)$. Аналогично первой части доказательства, используя (A.2), получаем

$$\int_0^x \frac{h(s)}{\sqrt{x-s}}ds = \int_0^x \frac{1}{\sqrt{x-s}}\int_0^s \frac{f(\sigma)}{\sqrt{s-\sigma}}d\sigma ds - \int_0^x \frac{g(s)}{\sqrt{x-s}}ds =$$

$$= \int_0^x\left(\int_\sigma^x \frac{ds}{\sqrt{x-\sigma}\sqrt{\sigma-s}}\right)f(\sigma)d\sigma - \int_0^x \frac{g(x)}{\sqrt{x-s}}ds = \pi\int_0^x f(\sigma)d\sigma - \int_0^x \frac{g(s)}{\sqrt{x-s}}ds = 0.$$

Следовательно,

$$\pi\int_0^x h(s)\,ds = \int_0^x h(s)\int_s^x \frac{1}{\sqrt{x-\sigma}\sqrt{\sigma-s}}d\sigma ds = \int_0^x \frac{1}{\sqrt{x-\sigma}}\left(\int_0^\sigma \frac{h(s)}{\sqrt{\sigma-s}}ds\right)d\sigma = 0, \quad x \in [0,T),$$

откуда $h = 0$.

Теперь рассмотрим разрешимость уравнения Абеля.

Лемма A.3. Существование и единственность решения уравнения Абеля.

(a) *Пусть $g \in L^1_{\mathrm{loc}}[0,T)$ и предположим, что*

(i) $G \in AC[0,T)$, *где* $G(x) := \int_0^x \frac{g(s)}{\sqrt{x-s}}ds$,

(ii) $G(0) = 0$.

Тогда f, определенная как

$$f(x) := \frac{1}{\pi}G'(x),$$

является единственным решением уравнения Абеля

$$g(x) = \int_0^x \frac{f(s)}{\sqrt{x-s}}ds, \quad x \in [0,T). \tag{A.3}$$

(б) *Наоборот, если* $f \in L^1_{\mathrm{loc}}[0,T)$ *и* g *удовлетворяет* (A.3), *тогда* $g \in L^1_{\mathrm{loc}}[0,T)$, (i), (ii) *выполняются и*

$$f(x) = \frac{1}{\pi}G'(x), \quad x \in [0,T).$$

Доказательство. (а) По предположению $f \in L^1_{\mathrm{loc}}[0,T)$, и мы имеем

$$\int_0^x f(s)ds = \frac{1}{\pi}\int_0^x G'(s)ds = \frac{1}{\pi}(G(x)-0) = \frac{1}{\pi}\int_0^x \frac{g(s)}{\sqrt{x-s}}ds, \quad x \in (0,T).$$

Тогда из леммы A.2 вытекает, что f является решением (A.3). Единственность следует из (б).

(б) Из леммы A.1 следует, что $g \in L^1_{\mathrm{loc}}[0,T)$. Пусть

$$G(x) := \int_0^x \frac{g(s)}{\sqrt{x-s}}ds.$$

Так как f удовлетворяет (A.3), из леммы A.2 следует, что

$$\int_0^x f(s)ds = \frac{1}{\pi}G(x), \quad x \in [0,T).$$

т.е., $G \in AC[0,T)$, $G(0) = 0$, и $f(x) = \frac{1}{\pi}G'(x)$, $x \in (0,T)$, что доказывает единственность в (а).

Замечание A.1. Пример $g(s) = \frac{1}{\sqrt{s}}$, $G(x) = \int_0^x \frac{ds}{\sqrt{s}\sqrt{x-s}} = \pi$ показывает, что (ii) не обязательно следует из (i).

Теперь мы рассмотрим разрешимость уравнения Эддингтона.

Лемма A.4 (существование и единственность решения уравнения Эддингтона).

(а) *Пусть* $g \in AC[0,T)$, $g(0) = 0$, *и предположим, что*

(i) $H_{g'} \in AC[0,T)$, *где* $H_{g'}(x) := \int_0^x \frac{g'(s)}{\sqrt{x-s}}ds$,

(ii) $H_{g'}(0) = 0$.

Тогда f, *определенная как*

$$f(x) := \frac{2}{\pi}H'_{g'}(x), \tag{A.4}$$

является единственным решением уравнения Эддингтона

$$g(x) = \int_0^x f(s)\sqrt{x-s}ds. \tag{A.5}$$

(б) *Наоборот, если* $f \in L^1_{\mathrm{loc}}[0,T)$, *и* g *удовлетворяет* (A.5), *тогда* $g \in AC[0,T)$, $g(0) = 0$, (i), (ii) *выполняются, и* f *имеет вид* (A.4). *Кроме того*, $g'(x) = \frac{1}{2}\int_0^x \frac{f(s)}{\sqrt{x-s}}ds$.

Доказательство. (а) Уравнение (A.4), интегрирование по частям, (i), (ii) и теорема Фубини дадут нам следующие равенства:

$$\int_0^x f(s)\sqrt{x-s}ds = \frac{2}{\pi}\int_0^x H'_{g'}(s)\sqrt{x-s}ds = \frac{1}{\pi}\int_0^x \frac{H_{g'}(s)}{\sqrt{x-s}}ds =$$

$$= \frac{1}{\pi}\int_0^x \frac{1}{\sqrt{x-s}}\left(\int_0^s \frac{g'(\sigma)}{\sqrt{s-\sigma}}d\sigma\right)ds = \frac{1}{\pi}\int_0^x\left(\int_\sigma^x \frac{ds}{\sqrt{x-s}\sqrt{s-\sigma}}\right)g'(\sigma)d\sigma = g(x),$$

что совпадает с (A.5). Единственность следует из (б).

(б) Пусть $f \in L^1_{\mathrm{loc}}[0,T)$ и g удовлетворяет (A.5). Покажем, что $g \in AC[0,T)$ и $g(0) = 0$. Определим формально

$$h(x) := \int_0^x \frac{f(s)}{\sqrt{x-s}}\,ds.$$

По лемме A.1 h существует п.в. на $[0,T)$ и $h \in L^1_{\mathrm{loc}}[0,T)$. Из равенства (A.5) и теоремы Фубини следует, что

$$2g(x) = 2\int_0^x f(\sigma)\sqrt{x-\sigma}\,d\sigma = \int_0^x f(\sigma)\left(\int_\sigma^x \frac{ds}{\sqrt{s-\sigma}}\right)d\sigma = \int_0^x\left(\int_0^s \frac{f(\sigma)}{\sqrt{s-\sigma}}\,d\sigma\right)ds = \int_0^x h(s)\,ds.$$

т.е., $g \in AC[0,T)$ и $g(0) = 0$. Кроме того,

$$2g'(x) = h(x) = \int_0^x \frac{f(s)}{\sqrt{x-s}}\,ds,$$

и из леммы A.2 следует, что

$$\int_0^x f(s)\,ds = \frac{2}{\pi}\int_0^x \frac{g'(s)}{\sqrt{x-s}}\,ds = \frac{2}{\pi}H_{g'}(x).$$

Следовательно, $H_{g'} \in AC[0,T)$ и $H_{g'}(0) = 0$. Дифференцируя последнее равенство, получаем (A.4), что доказывает единственность в (а).

10. ПРЕДЛОЖЕНИЯ ДЛЯ ДАЛЬНЕЙШЕЙ РАБОТЫ

В примере 5.6 была построена функция p, удовлетворяющая условиям теоремы 4.2 такая, что в соответствии с численными расчетами функция q может иметь отрицательные значения, т.е. p нерасширима. Было бы интересно дать строгое доказательство этого результата. Еще один вопрос заключается в распространении настоящей работы на случай цилиндрической симметрии.

Авторы благодарны В.В. Козлову за полезные обсуждения работы

СПИСОК ЛИТЕРАТУРЫ

1. *Batt J., Jörn E., Li Y.* Stationary solutions of the flat Vlasov–Poisson System // Arch. Rational Mech. Anal. 2019. V. 231. P. 189–232.
2. *Binney J., Tremaine S.* Galactic Dynamics // Princeton University Press Princeton, 1987.
3. *Rein G.* Collisionless kinetic equations from astrophysics – the Vlasov–Poisson system// Handbook of Differential Equations: Evolutionary Equations. eds. Dafermos C.M. and Feireisl E., Elsiver, Amsterdam, 3 (2007), 383–476.
4. *Веденяпин В.В.* Граничные задачи для стационарного уравнения Власова // Докл. АН СССР. 1986. Т. 290. № 4. С. 777–780.
5. *Архипов Ю.Ю., Веденяпин В.В.* О классификации и устойчивости стационарных решений уравнения Власова на торе и граничной задаче // Тр. МИАН. 1994. Т. 203. С. 13–20.
6. *Batt J., Faltenbacher W., Horst E.* Stationary spherically symmetric models in stellar dynamics // Arch. Rational Mech. Anal. 1986. V. 93. P. 159–183.
7. *Батт Ю., Йорн Э., Скубачевский А.Л.* Стационарные сферически симметричные решения системы Власова–Пуассона в трехмерном случае // Докл. АН. Математика, информатика, процессы управления. 2020. Т. 493. № 1. С. 5–8.
8. *Tonelli L.* Su un problema di Abel // Mathematische Annalen. 1928. V. 99. P. 183–199.
9. *Geigant E.* Inversionsmethoden zur Konstruktion von stationären Lösungen der selbst-konsistenten Problems des Stellardynamik. Diplomarbeit Ludwig-Maximilians-Universität Munich. 1991.
10. *Бронштейн И.Н., Семендяев К.А.* Справочник по математике для инженеров и учащихся втузов: Учебное пособие. СПб: “Лань”, 2010.
11. *Stoer J.* Einführung in die Numerische Mathematik I. Springer-Verlag, 1972.
12. *Kofler M., Bitsch G., Komma M.* Maple. Einführung, Anwendung, Referenz. Addison–Wesley, 2001.

EDN: NKIOCU
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1522–1531

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.588

МЕТОД КОНТИНУАЛЬНЫХ ТЕОРЕМ СЛОЖЕНИЯ И ИНТЕГРАЛЬНЫЕ СООТНОШЕНИЯ МЕЖДУ ФУНКЦИЯМИ КУЛОНА И ФУНКЦИЕЙ АППЕЛЯ F_1

© 2022 г. И. А. Шилин[1,2,*], Дж. Чой[3]

[1] *111250 Москва, ул. Красноказарменная, 14, НИУ Московский энергетический институт, Россия*

[2] *119991 Москва, ул. Малая Пироговская, 1, Московский педагогический государственный университет, Россия*

[3] *Department of Mathematics, Dongguk University, Gyeongju, Republic of Korea*

**e-mail: ilyashilin@li.ru*

Поступила в редакцию 15.06.2021 г.
Переработанный вариант 13.02.2022 г.
Принята к публикации 11.05.2022 г.

Рассматривается введенная авторами функция A, зависящая от одного комплексного, двух действительных переменных и еще одного аргумента, задающего тривиальную или собственную подгруппу трехмерной собственной лоренцевой группы и, таким образом, являющегося действительным числом или парой действительных чисел. Первые три аргумента при этом определяют пространства представления и базисные функции в этих пространствах. Показано, что ее частные значения можно выразить через волновые кулоновские функции или гипергеометрическую функцию Аппеля F_1. Полученная формула преобразования функции A используется для вывода континуальной теоремы сложения для этой функции и вычисления значения одномерного интегрального преобразования типа Фурье–Меллина произведения двух кулоновских функций – его результат выражается через функцию F_1. Библ. 31.

Ключевые слова: волновые кулоновские функции, функция Аппеля F_1, трехмерная собственная лоренцева группа, представление группы, ядро интегрального оператора, интегральное преобразование типа Фурье–Меллина.

DOI: 10.31857/S004446692209006X

1. ВВЕДЕНИЕ

Вычисление матричных элементов неприводимых представлений групп и алгебр Ли (инфинитезимальных операторов) является важной задачей в прикладной теории представлений групп [1]. Помимо применений результатов этой задачи в физике элементарных частиц, атомной физике и квантовой теории поля, важность этой задачи усиливается приложениями полученных результатов в теории специальных функций, возникающих в математической физике.

В настоящей работе рассматривается функция A, зависящая от одного комплексного и двух вещественных аргументов, описывающих соответственно пространства представления трехмерной группы Лоренца G и базисы в этих пространствах, а также еще одного аргумента, являющегося элементом тривиальной или собственной подгруппы в G и, таким образом, зависящей от одного или двух вещественных параметров. В частных случаях значения этой функции удается выразить через хорошо известные гипергеометрические функции: волновые кулоновские функции или функцию Аппеля F_1, зависящую от двух переменных. С другой стороны, через функцию A удается выразить ядра интегральных операторов представления в обычном смысле или в смысле “смешанного” [2] базиса. Таким образом, функции Кулона и Аппеля приобретают теоретико-групповую трактовку. При этом связь между ядрами приводит к интегральным формулам для указанных специальных функций.

2. ОПРЕДЕЛЕНИЕ ФУНКЦИИ A

Пусть σ – фиксированное комплексное число и G – собственная трехмерная лоренцева группа, т.е.

$$G = \{g \in SL(3,\mathbb{R}) \,|\, g^{-1}\mathrm{diag}(1,-1,-1) = \mathrm{diag}(1,-1,-1)g^{\mathrm{T}}\}.$$

Обозначим через $\mathcal{D}$ линейное пространство, состоящее из функций, которые заданы на конусе $C: x_1^2 - x_2^2 - x_3^2 = 0$, $x_1 > 0$, бесконечно дифференцируемы на области определения и при любых $x \in C$ и $\alpha \in \mathbb{C}$ удовлетворяют равенству $f(\alpha x) = |\alpha|^{\sigma} f(x)$. Система функций

$$B_1 = \{f_{p_1}(x) = (x_1 + x_2)^{\sigma} \mathrm{e}^{\mathrm{i}p_1 x_3/(x_1+x_2)} \,|\, p_1 \in \mathbb{R}\} \tag{2.1}$$

является базисом пространства $\mathcal{D}$. Используя обобщенные функции [3]

$$(s)_{\pm}^{\lambda} = \begin{cases} |s|^{\lambda} & \text{при} \quad \pm s > 0, \\ 0 & \text{при} \quad \mp s \geqslant 0, \end{cases}$$

другой базис в $\mathcal{D}$ можно описать формулой

$$B_2 = \{f_{p_2^{\pm}}(x) = (x_2)_{\pm}^{\sigma - \mathrm{i}p_2}(x_1 + x_3)^{\mathrm{i}p_2} \,|\, p_2 \in \mathbb{R}\}. \tag{2.2}$$

Каждый из базисов B_1 и B_2 состоит из собственных функций инфинитезимального оператора (оператора Казимира) для соответствующей однопараметрической подгруппы в G, а их построение подробно описано в [4]. При переходе от σ к другому комплексному числу $\hat{\sigma}$ получается аналогичное линейное пространство $\hat{\mathcal{D}}$, базисы которого, построенные по формулам (2.1) и (2.2), обозначим соответственно $\hat{B}_1$ и $\hat{B}_2$. Далее в работе используется обозначение $\hat{\sigma} = -\sigma - 1$.

Пусть γ – многообразие на конусе C, по одному разу пересекающее все (быть может, за исключением нескольких) его образующие, а H – подгруппа группы G, для которой γ является однородным пространством (например, для окружности $\gamma_r: x_2^2 + x_3^2 = r^2$ подгруппа H_{γ_r} состоит из поворотов в плоскости $x_1 = r$).

Теорема 1 (см. [5]). *При* $\hat{\sigma} = -\sigma - 1$ *билинейные интегральные функционалы*

$$\mathrm{F}_{\gamma}: \mathcal{D} \times \hat{\mathcal{D}} \to \mathbb{C}, \quad (u,v) \mapsto \int_{\gamma} u(x)v(x)\mathrm{d}\gamma,$$

где $\mathrm{d}\gamma$ *– мера на* γ*, инвариантная относительно подгруппы* H*, совпадают.*

Гомоморфизм $T: G \to GL(\mathcal{D})$ в группу невырожденных линейных операторов пространства $\mathcal{D}$, при котором оператор $T(g)$ функции f ставит в соответствие композицию fg^{-1} левого "сдвига" конуса C элементом g^{-1} и функции f, является G-представлением в пространстве $\mathcal{D}$. Группу G можно записать [6] в виде $G = H_{\gamma_r}\tilde{H}H_{\gamma_r}$, где $\tilde{H}$ – подгруппа гиперболических вращений в плоскости Ox_1x_2. Опираясь на это разложение, можно доказать следующее утверждение.

Теорема 2 (см. [7]). *При* $\hat{\sigma} = -\sigma - 1$ *билинейные интегральные функционалы* F_{γ} *инвариантны относительно пары операторов представления* $(T(g), \hat{T}(g))$*, т.е.* $\mathrm{F}_{\gamma}(T(g)f, \hat{T}(g)\hat{f}) = \mathrm{F}_{\gamma}(f, \hat{f})$.

Для дальнейшего изложения определим функцию

$$A: \mathbb{C} \times \mathbb{R}^2 \times G \to \mathbb{C},$$

ставящую набору $(\sigma, p, \hat{p}, g)$ в соответствие число $\mathrm{F}_{\gamma}(T(g)f_p, \hat{f}_{\hat{p}})$. Корректность этого определения следует из теоремы 1. В последующих теоремах четвертый аргумент этой функции будет принадлежать либо тривиальной подгруппе, либо подгруппе H°, состоящей из поворотов

$$h^{\circ}(\xi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\xi & -\sin\xi \\ 0 & \sin\xi & \cos\xi \end{pmatrix}$$

в плоскости Ox_2x_3, но в общем случаем (поскольку $\dim G = 3$) принадлежит одно- или двухпараметрической подгруппе. Хотя параметр ξ, задающий подгруппу H°, принимает все значения из отрезка $[-\pi;\pi]$, ограничимся случаем $0 < \xi < \frac{\pi}{2}$. Так как алгебра Ли группы G порождается кососимметрической матрицей $a = e_{23} - e_{32}$ и двумя симметрическими $b = e_{12} + e_{21}$ и $c = e_{13} + e_{31}$ (где e_{ij} – матрицы канонического базиса пространства матриц размера 3×3), для формы $B(x,y) = \operatorname{tr}\big((\operatorname{ad} x)(\operatorname{ad} y)\big)$ Картана–Киллинга выполняются равенства $B(b,b) = B(c,c) = -B(a,a) = 2$ и $H^\circ = \exp \operatorname{Span}(a)$, то H° является максимальной компактной подгруппой в G.

Теорема 3. *Выполняется формула преобразования*

$$A(\sigma, p, \hat{p}, g) = A(\hat{\sigma}, \hat{p}, p, g^{-1}).$$

Доказательство. Так как T является гомоморфизмом групп, то в силу теоремы 2 имеем

$$A(\sigma, p, \hat{p}, g) = \mathrm{F}_\gamma(T(g)f, \hat{f}) = \mathrm{F}_\gamma(T(g^{-1})T(g)f, \hat{T}(g^{-1})\hat{f}) =$$
$$= \mathrm{F}_\gamma(T(\mathrm{id})f, \hat{T}(g^{-1})\hat{f}) = \mathrm{F}_\gamma(\hat{T}(g^{-1})\hat{f}, f) = A(\hat{\sigma}, \hat{p}, p, g^{-1}).$$

3. ВЫРАЖЕНИЕ ФУНКЦИЙ $A(-\sigma-1, p_2^\pm, p_1, \mathrm{id})$ ЧЕРЕЗ ВОЛНОВЫЕ КУЛОНОВСКИЕ ФУНКЦИИ

При решении линейного дифференциального уравнения $\mathfrak{d}f(p) = 0$, в котором дифференциальный оператор определен формулой

$$\mathfrak{d} = \frac{\mathrm{d}^2}{\mathrm{d}p^2} + 1 - \frac{2\hat{p}}{p} - \frac{\sigma(\sigma+1)}{p^2},$$

важную роль играют кулоновские функции

$$F_\sigma(\hat{p}; p) = C_\sigma(\hat{p}) p^{\sigma+1} \mathrm{e}^{\mathbf{i}p}\, {}_1F_1(1+\sigma+\mathbf{i}\hat{p}; 2\sigma+2; -2\mathbf{i}p)$$

и

$$H_\sigma^\pm(\hat{p}; p) = D_\sigma^\pm(\hat{p}) p^{\sigma+1} \mathrm{e}^{\pm\mathbf{i}p} \times$$
$$\times\left[\frac{\Gamma(-2\sigma-1)}{\Gamma(-\sigma\pm\mathbf{i}p)}\, {}_1F_1(1+\sigma\pm\mathbf{i}\hat{p}; 2\sigma+2; \mp 2\mathbf{i}p) + \frac{\Gamma(2\sigma+1)}{(\mp 2\mathbf{i}\hat{p})^{2\sigma+1}\Gamma(1+\sigma\pm\mathbf{i}\hat{p})}\, {}_1F_1(-\sigma\pm\mathbf{i}\hat{p}; -2\sigma; \mp 2\mathbf{i}p)\right],$$

которые определены при комплексных значениях параметров σ (угловой момент) и $\hat{p} \neq 0$ (параметр Зоммерфельда) и переменной p, и в которых

$$C_\sigma(\hat{p}) = \frac{2^\sigma \sqrt{\Gamma(1+\sigma+\mathbf{i}\hat{p})}\,\Gamma(1+\sigma-\mathbf{i}\hat{p})}{\mathrm{e}^{\hat{p}\pi/2}\Gamma(2\sigma+2)},$$

$$D_\sigma^\pm(\hat{p}) = \frac{(\mp 2\mathbf{i})^{2\sigma+1}\Gamma(1+\sigma\pm\mathbf{i}\hat{p})}{C_\sigma(\hat{p})\Gamma(2\sigma+2)}.$$

Эти функции составляют базисы (см. [8]) $E_1 = \{H_\sigma^+(\hat{p}; p), H_\sigma^-(\hat{p}; p)\}$ и $E_2 = \Big\{F_\sigma(\hat{p}; p), \frac{1}{2}(H_\sigma^+(\hat{p}; p) + H_\sigma^-(\hat{p}; p))\Big\}$ пространства решений $\operatorname{Ker}\mathfrak{d}$.

В работе [9] дано “рекуррентное” определение волновых кулоновских функций, а соответствующие инфинитезимальные операторы изменения индекса рассматриваются как генераторы алгебры Ли. Интерпретация волновых кулоновских функций как матричных элементов перехода между двумя базисами пространства представления дана авторами настоящей статьи в [10]. Покажем, что в частных случаях $(p, \hat{p}, g) = (p_2^\pm, p_1, \mathrm{id})$ значения функции A выражаются через волновые кулоновские функции.

Теорема 4. *При* $\Re(\sigma) > -1$ *и* $p_1 \neq 0$ *имеем*

$$A(-\sigma-1, p_2^+, p_1, \mathrm{id}) = 2p_1^{-\sigma-1}\mathrm{e}^{p_2\pi/2}\sqrt{\Gamma(1+\sigma+\mathbf{i}p_2)\Gamma(1+\sigma-\mathbf{i}p_2)}F_\sigma(p_2; p_1).$$

Доказательство. Так как при гомоморфизме групп $\mathrm{id} \in G$ отображается в тождественный оператор, то, учитывая теорему 1, в качестве многообразия, по которому производится интегрирование, можно выбрать параболу $\omega = \left\{\frac{1}{2}(1+\alpha_2, 1-\alpha^2, 2\alpha) \,|\, \alpha \in \mathbb{R}\right\}$. Поскольку

$$H_\omega = \left\{ h_\omega(\tau) = \frac{1}{2}\begin{pmatrix} 2+\tau^2 & \tau^2 & 2\tau \\ -\tau^2 & 2-\tau^2 & -2\tau \\ 2\tau & 2\tau & 2 \end{pmatrix} \,\Big|\, \tau \in \mathbb{R} \right\},$$

для $x(\alpha) \in \omega$ выполняется включение $h_\omega(\tau)x(\alpha) = x(\alpha+\tau)$ и, следовательно, $\mathrm{d}\omega = \mathrm{d}\alpha$, имеем

$$A(-\sigma-1, p_2^+, p_1, \mathrm{id}) = 2^{-\sigma}\int_{\mathbb{R}}(1-\alpha^2)_+^{\sigma+\mathbf{i}p_2}(1+\alpha)^{-2\mathbf{i}p_2}\mathrm{e}^{\mathbf{i}p_1\alpha}\mathrm{d}\alpha = 2^{-\sigma}\mathrm{e}^{-\mathbf{i}p_1}\int_0^2 t^{\sigma-\mathbf{i}p_2}(2-t)^{\sigma+\mathbf{i}p_2}\mathrm{e}^{\mathbf{i}p_1 t}\mathrm{d}t =$$
$$= 2^{\sigma+1}\mathrm{e}^{-\mathbf{i}p_1}\mathrm{B}(1+\sigma+\mathbf{i}p_2, 1+\sigma-\mathbf{i}p_2)\,{}_1F_1(1+\sigma-\mathbf{i}p_2; 2\sigma+2; 2\mathbf{i}p_1),$$

где для вычисления последнего интеграла использована формула (см. [11, 2.3.6(1)]). Применяя преобразование Куммера $\mathrm{e}^{-z}\,{}_1F_1(a;b;z) = {}_1F_1(b-a;b;-z)$ и определение функции $F_\sigma(p, \hat{p})$, завершаем доказательство.

Отметим, что поскольку алгебра Ли группы G является прямой суммой подалгебры $\mathrm{Span}(a)$ и линейного подпространства $\mathrm{Span}(b,c)$, $\mathrm{Span}(b)$ – максимальная коммутативная подалгебра в $\mathrm{Span}(b,c)$ и спектр оператора $\mathrm{ad}\, b$ имеет вид $\{0,1,-1\}$, то максимальная нильпотентная подалгебра алгебры Ли группы G порождается неподвижной точкой $a+b$ оператора $\mathrm{ad}\, b$. Но $\exp(\tau(a+b)) = H_\omega$, поэтому подгруппа H_ω, транзитивно действующая на параболе ω, является максимальной нильпотентной подгруппой в G. В наших работах [7], [12] показано, что при сужении представления T на эту подгруппу появляются функции Бесселя–Клиффорда, а в классической (четырехмерной) лоренцевой группе – их мультииндексные аналоги (гиперфункции).

Теорема 5. *При* $-1 < \Re(\sigma) < 0$ *и* $p_1 \neq 0$

$$A(-\sigma-1, p_2^-, p_1, \mathrm{id}) = \frac{\mathbf{i}\Gamma(2\sigma+2)C_\sigma(p_2)}{2^{2\sigma}p_1^{\sigma+1}}(\mathrm{e}^{-\mathbf{i}\pi\sigma}H_\sigma^+(p_2; p_1) - \mathrm{e}^{\mathbf{i}\pi\sigma}H_\sigma^-(p_2; p_1)).$$

Доказательство. Интегрируя по параболе ω, имеем

$$A(-\sigma-1, p_2^-, p_1, \mathrm{id}) = 2^{-\sigma}\int_{\mathbb{R}}(1-\alpha^2)_-^{\sigma+\mathbf{i}p_2}(1+\alpha)^{-2\mathbf{i}p_2}\mathrm{e}^{\mathbf{i}p_1\alpha}\mathrm{d}\alpha =$$
$$= 2^{-\sigma}\sum_{k=0}^{1}\mathrm{e}^{(-1)^k\mathbf{i}p_1}\int_0^{+\infty} t^{\sigma+(-1)^k\mathbf{i}p_2}(t+2)^{\sigma-(-1)^k\mathbf{i}p_2}\mathrm{e}^{(-1)^k\mathbf{i}p_1 t}\mathrm{d}t.$$

Применяя формулу [11], 2.3.2(3), получаем

$$A(-\sigma-1, p_2^-, p_1, \mathrm{id}) = 2^{-\sigma}\sum_{k=0}^{1}\mathrm{e}^{(-1)^k\mathbf{i}p_1}\times$$
$$\times\Big[2^{2\sigma+1}\mathrm{B}(1+\sigma+(-1)^k\mathbf{i}p_2, -2\sigma-1)\,{}_1F_1(1+\sigma+(-1)^{k+1}\mathbf{i}p_2;\, 2\sigma+2;\, (-1)^k\mathbf{i}p_1) +$$
$$+ ((-1)^{k+1}\mathbf{i}p_1)^{-2\sigma-1}\Gamma(2\sigma+1)\,{}_1F_1(-\sigma+(-1)^k\mathbf{i}p_2; -\sigma; 2(-1)^{k+1}\mathbf{i}p_1)\Big] =$$
$$= 2p_1^{-\sigma-1}\left(\frac{\Gamma(1+\sigma+\mathbf{i}p_2)H_\sigma^+(p_2; p_1)}{D_\sigma^+(p_2)} + \frac{\Gamma(1+\sigma-\mathbf{i}p_2)H_\sigma^-(p_2; p_1)}{D_\sigma^-(p_2)}\right),$$

откуда и следует утверждение теоремы.

4. ВЫРАЖЕНИЕ ФУНКЦИИ $A(-\sigma-1, p_2^+, -p_1, h(\xi))$ ЧЕРЕЗ ВОЛНОВЫЕ КУЛОНОВСКИЕ ФУНКЦИИ

В этой части статьи докажем утверждение, обобщающее теорему 4 (аналогичное обобщение теоремы 5 опускаем).

Теорема 6. *При* $-1 < \Re(\sigma) < 0$ *имеем*

$$A(-\sigma-1, p_2^+, -p_1, h(\xi)) = \mathrm{B}(1+\sigma-\mathbf{i}p_2, 1+\sigma+\mathbf{i}p_2)\times$$
$$\times\frac{2(\sin\xi+\cos\xi+1)^{\sigma-\mathbf{i}p_2}(\sin\xi-\cos\xi-1)^{\sigma+\mathbf{i}p_2}}{\cos^{2\sigma+2}\xi[p_1\tan\xi]^{\sigma+1}(1+\cos\xi)^{\sigma}\mathrm{e}^{\mathbf{i}p_1\tan\xi}C_\sigma(-p_2)}F_\sigma(-p_2; p_1\tan\xi).$$

Доказательство. Интегрируя вдоль параболы ω, имеем

$$A(-\sigma-1, p_2^+, -p_1, h(\xi)) = \frac{2}{(1-\cos\xi)^{\sigma+1}}\int\limits_{\mathbb{R}}(1-\alpha)_+^{\sigma-\mathbf{i}p_2}(1+\alpha)^{2\mathbf{i}p_2}\left(\alpha+\frac{1+\cos\xi}{\sin\xi}\right)^{-2\sigma-2}\times$$
$$\times\exp\frac{\mathbf{i}p_1\sin\xi\left(\alpha+\dfrac{\cos\xi-1}{\sin\xi}\right)}{(\cos\xi-1)\left(\alpha+\dfrac{\cos\xi+1}{\sin\xi}\right)}\mathrm{d}\alpha = \frac{2\,\mathrm{e}^{\mathbf{i}p_1\sin\xi/(\cos\xi-1)}}{(1-\cos\xi)^{\sigma+1}}\times$$
$$\times\int\limits_{-1}^{1}(1-\alpha)^{\sigma-\mathbf{i}p_2}(1+\alpha)^{\sigma+\mathbf{i}p_2}\left(\alpha+\frac{1+\cos\xi}{\sin\xi}\right)^{-2\sigma-2}\exp\frac{2\mathbf{i}p_1\sin\xi}{(1-\cos\xi)\left(\alpha+\dfrac{\cos\xi+1}{\sin\xi}\right)}\mathrm{d}\alpha.$$

Используя далее подстановку

$$t = \frac{\sin\xi}{1+\cos\xi-\sin\xi} - \frac{\sin\xi}{1+\cos\xi+\alpha\sin\xi},$$

применяем, как и в ходе доказательства теоремы 2, формулу [11, 2.3.6.1], а после выражаем функцию Куммера через функцию Кулона.

5. ВЫРАЖЕНИЕ ФУНКЦИЙ $A(-\sigma-1, p_2^\pm, p_2^\pm, h(\xi))$ И $A(-\sigma-1, p_2^\pm, p_2^\mp, h(\xi))$ ЧЕРЕЗ ФУНКЦИЮ АППЕЛЯ F_1

Функция Аппеля F_1 является частным случаем (при $n = 2$) функции Лауричеллы F_D, которая при $|z_1|,\ldots|z_n| < 1$ определяется рядом

$$F_D\left(\begin{matrix}a; b_1,\ldots,b_n\\ c\end{matrix}\middle| z_1,\ldots,z_n\right) = \sum_{m_1,\ldots,m_n=0}^{\infty}\frac{(a)_{m_1+\ldots+m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c)_{m_1+\ldots+m_n}}\frac{z_1^{m_1}\ldots z_n^{m_n}}{m_1!\ldots m_n!},$$

а вне указанной области является его аналитическим продолжением [13], [14]. Другим частным случаем (при $n = 1$) является гипергеометрическая функция Гаусса. С теоретико-групповой точки зрения функция F_D изучена У. Миллером в работе [15] (а также в [2]), рассмотревшим неприводимые представления алгебры $\mathfrak{sl}(n+3,\mathbb{C})$, а ее частный случай F_1 – в работе [16]. Работам Миллера предшествовала статья Н.Я. Виленкина [17], в которой другие гипергеометрические функции двух переменных, F_A и F_B, были изучены в виде матричных элементов максимально вырожденных представлений группы $SL(n,\mathbb{R})$ (см. также [18]). В работе [19] с помощью теории алгебр Ли для каждой из 34 гипергеометрических функций двух переменных (включая F_1) найдена каноническая система, в которой соответствующая функция возникает в результате разделения переменных. Свойства функции F_1 часто раскрываются в работах, относящихся к квантовой физике (теория рассеивания, теория Редже, фейнмановские интегралы) [20]–[22]. Отметим также связь между F_1 и нормальными (неполными) эллиптическими интегралами [23], [24].

Покажем, что значения функции $A(-\sigma-1, p_2^+, p_2'^+, g)$ в частных случаях выражаются через функцию Аппеля F_1 или линейную комбинацию двух функций F_1.

Теорема 7. *При* $-1 < \Re(\sigma) < 0$ *имеем*

$$A(-\sigma-1, p_2^+, p_2'^+, h((\xi)) = \frac{(1+\sin\xi)^{\mathbf{i}p_2}\,\mathrm{B}(\mathbf{i}p_2-\sigma, 1+\sigma-\mathbf{i}p_2')}{2^{\sigma+\mathbf{i}p_2}\cos^{\sigma+\mathbf{i}(p_2-p_2')-1}(1+\cos\xi-\sin\xi)^{-\sigma-\mathbf{i}p_2'}}\times$$
$$\times F_1\left(\begin{matrix}\mathbf{i}p_2-\sigma; 1+\sigma+\mathbf{i}p_2, -\sigma-\mathbf{i}p_2' \\ 1+\mathbf{i}(p_2-p_2')\end{matrix}\middle|\frac{1+\cos\xi-\sin\xi}{2}, \frac{1-\cos\xi-\sin\xi}{1+\cos\xi-\sin\xi}\right). \tag{5.1}$$

Доказательство. Выбирая в качестве многообразия, по которому происходит интегрирование, параболу ω, получаем

$$A(-\sigma-1, p_2^+, p_2'^+, h((\xi)) = \frac{2(1+\sin\xi)^{\mathbf{i}p_2}}{\cos^{\sigma+1+\mathbf{i}p_2}\xi}\times$$
$$\times\int_{\mathbb{R}}\left(\alpha+\frac{\cos\xi}{1+\sin\xi}\right)^{2\mathbf{i}p_2}(1+\alpha)^{2\mathbf{i}p_2'}(1-\alpha^2)_+^{\sigma}(1+2\alpha\,\mathrm{tg}\,\xi-\alpha^2)_+^{-\sigma-1}\mathrm{d}\alpha.$$

Так как для корней $\alpha_\pm = \dfrac{\sin\xi\pm1}{\cos\xi}$ многочлена $\alpha^2-2\alpha\tan\xi-1$ справедливо неравенство $-1<\alpha_-<1<\alpha_+$, то

$$A(-\sigma-1, p_2^+, p_2'^+, h((\xi)) = \frac{2(1+\sin\xi)^{\mathbf{i}p_2}}{\cos^{\sigma+1+\mathbf{i}p_2}\xi}\int_{\alpha_-}^{1}\frac{(1-\alpha)^{\sigma-\mathbf{i}p_2'}(1+\alpha)^{\sigma+\mathbf{i}p_2'}\,\mathrm{d}\alpha}{\left(\dfrac{\sin\xi+1}{\cos\xi}-\alpha\right)^{\sigma+1+\mathbf{i}p_2}\left(\alpha-\dfrac{\sin\xi-1}{\cos\xi}\right)^{\sigma+1-\mathbf{i}p_2}}.$$

Переходя к новой переменной $t = \dfrac{1+\alpha\cos\xi-\sin\xi}{1+\cos\xi-\sin\xi}$, получаем равенство

$$A(-\sigma-1, p_2, p_2', h((\xi)) = \frac{(1+\sin\xi)^{\mathbf{i}p_2}(1+\cos\xi-\sin\xi)^{\sigma+\mathbf{i}p_2'}}{2^{\sigma+\mathbf{i}p_2}\cos^{\sigma+\mathbf{i}(p_2-p_2')-1}\xi}\times$$
$$\times\int_0^1\frac{t^{-\sigma-1+\mathbf{i}p_2}(1-t)^{\sigma-\mathbf{i}p_2'}\,\mathrm{d}t}{\left(1-\dfrac{1+\cos\xi-\sin\xi}{2}t\right)^{\sigma+1+\mathbf{i}p_2}\left(1-\dfrac{1-\cos\xi-\sin\xi}{1+\cos\xi-\sin\xi}t\right)^{-\sigma-\mathbf{i}p_2'}}.$$

Так как при рассматриваемых значениях параметра ξ выполняются неравенства $1+\cos\xi-\sin\xi>0$ и $\left|\dfrac{1-\cos\xi-\sin\xi}{1+\cos\xi-\sin\xi}\right|<1$, то, используя формулу [25, 2.2.8.5], завершаем доказательство.

С учетом связи [26, 7.2.4.63]

$$F_1\left(\begin{matrix}a,b,c\\ b+c\end{matrix}\middle| z; w\right) = (1-w)^{-a}\,{}_2F_1\left(a,b;b+c;\frac{z-w}{1-w}\right)$$

между гипергеометрическими функциями Аппеля и Гаусса, формулу (5.1) можно переписать с использованием функции Гаусса.

Поскольку

$$A(-\sigma-1, p_2^{\pm}, p_2'^{\mp}, h(\xi)) = \frac{2(1+\sin\xi)^{\mathbf{i}p_2}}{\cos^{\sigma+1+\mathbf{i}p_2}\xi}\times$$
$$\times\int\limits_{\pm1}^{\alpha_{\pm}}\left(\alpha+\frac{\cos\xi}{1+\sin\xi}\right)^{2\mathbf{i}p_2}(1+\alpha)^{2\mathbf{i}p_2'}(\pm\alpha^2\mp1)^{\sigma}(\pm1\pm2\alpha\tan\xi\mp\alpha^2)^{-\sigma-1}\mathrm{d}\alpha,$$

то в случае $(\hat{p}, p) = (p_2^{\pm}, p_2'^{\mp})$ значения функции $A(-\sigma-1, \hat{p}, p, h(\xi))$ тоже выражаются через функции Аппеля F_1, а соответствующие теоремы мы опускаем. В случае $(\hat{p}, p) = (p_2^{-}, p_2'^{-})$ значение функции $A(-\sigma-1, \hat{p}, p, h(\xi))$ можно представить в виде линейной комбинации двух функций Аппеля, а именно имеет место следующая теорема.

Теорема 8. *При* $-1 < \Re(\sigma) < 0$ *имеем*

$$A(-\sigma-1, p_2^{-}, p_2'^{-}, h((\xi)) = \frac{(1+\sin\xi)^{\mathbf{i}\hat{p}_2}}{\cos^{\sigma+\mathbf{i}\hat{p}}\xi(\sin\xi+\cos\xi+1)}\times$$
$$\times\left[\mathrm{B}(1,-\sigma-\mathbf{i}p_2')F_1\left(\begin{matrix}1,\mathbf{i}p_2,1+\sigma-\mathbf{i}p_2'\\1-\sigma-\mathbf{i}p_2'\end{matrix}\middle|\frac{2\cos\xi}{1+\cos\xi+\sin\xi},\frac{1+\cos\xi+\sin\xi}{1+\cos\xi-\sin\xi}\right)-\right.$$
$$\left.-\mathrm{B}(1,1+\sigma-\mathbf{i}p_2)F_1\left(\begin{matrix}1,1+\sigma+\mathbf{i}p_2',1+\sigma-\mathbf{i}p_2'\\-\sigma+\mathbf{i}p_2\end{matrix}\middle|\frac{\cos\xi-\sin\xi-1}{2\cos\xi},\frac{1+\cos\xi-\sin\xi}{2\cos\xi}\right)\right].$$

Доказательство. Так как

$$A(-\sigma-1,\sigma, p_2^{-}, p_2'^{-}, h((\xi)) = \frac{2(1+\sin\xi)^{\mathbf{i}\hat{p}}}{\cos^{\sigma+1+\mathbf{i}\hat{p}}\xi}\int\limits_{\mathbb{R}}\left(\alpha+\frac{\cos\xi}{1+\sin\xi}\right)^{2\mathbf{i}\hat{p}}\times$$
$$\times(1+\alpha)^{2\mathbf{i}p}(1-\alpha^2)_-^{\sigma}(1+2\alpha\tan\xi-\alpha^2)_-^{-\sigma-1}\mathrm{d}\alpha = \frac{2(1+\sin\xi)^{\mathbf{i}\hat{p}}}{\cos^{\sigma+1+\mathbf{i}\hat{p}}\xi}(I_+ - I_-),$$

где

$$I_{\pm} = \int\limits_{\pm\infty}^{\beta_{\pm}}\frac{(\alpha-1)^{\sigma-\mathbf{i}p}(\alpha+1)^{\sigma+\mathbf{i}p}\,\mathrm{d}\alpha}{\left(\alpha-\dfrac{\sin\xi+1}{\cos\xi}\right)^{\sigma+1+\mathbf{i}\hat{p}}\left(\alpha-\dfrac{\sin\xi-1}{\cos\xi}\right)^{\sigma+1-\mathbf{i}\hat{p}}},$$

$\beta_+ = \alpha_+$ и $\beta_- = -1$, то, применяя в интеграле $I_{\pm}$ подстановку $t = 1\pm\alpha$, вычисляем его по формуле [11, 2.2.8.11].

6. КОНТИНУАЛЬНАЯ ТЕОРЕМА СЛОЖЕНИЯ ДЛЯ ФУНКЦИИ A И ВЫЧИСЛЕНИЕ ИНТЕГРАЛА ТИПА ФУРЬЕ–МЕЛЛИНА ОТ ПРОИЗВЕДЕНИЯ ДВУХ КУЛОНОВСКИХ ФУНКЦИЙ

Теорема 9. *При* $-1 < \Re(\sigma) < 0$ *имеем*

$$\int\limits_{\mathbb{R}} p_1^{-2\sigma-2}\mathrm{e}^{\mathbf{i}p_1\tan\xi}F_{\sigma}(p_2;p_1)F_{\sigma}(p_2';-p_1\tan\xi)\mathrm{d}p_1 =$$
$$= \kappa F_1\left(\begin{matrix}\mathbf{i}p_2-\sigma,1+\sigma+\mathbf{i}p_2,\mathbf{i}p_2'-\sigma\\1+\mathbf{i}(p_2+p_2'^{+})\end{matrix}\middle|\frac{1+\cos\xi-\sin\xi}{2},\frac{1-\cos\xi-\sin\xi}{1+\cos\xi-\sin\xi}\right),$$

где

$$\kappa = \frac{2^{-\sigma-\mathbf{i}p_2-1}\pi\cos^{3+\sigma+\mathbf{i}(p_2'-p_2)}\xi(1+\cos\xi)^{\sigma}(1+\sin\xi)^{\mathbf{i}p_2}}{(1+\cos\xi-\sin\xi)^{2\mathbf{i}p_2'}(1-\sin\xi-\cos\xi)^{\sigma-\mathbf{i}p_2}}\times$$
$$\times\frac{\sqrt{\Gamma(1+\sigma+\mathbf{i}p_2)\Gamma(1+\sigma-\mathbf{i}p_2)}\mathrm{B}(\mathbf{i}p_2-\sigma,1+\sigma+\mathbf{i}p_2')C_{\sigma}(p_2')}{\mathrm{B}(1+\sigma+\mathbf{i}p_2',1+\sigma-\mathbf{i}p_2')}.$$

Доказательство. Представим функцию $\hat{T}(g)\hat{f}_{p_2^+}$ в виде суммы двух интегральных операторов

$$\hat{T}(g)\hat{f}_{p_2^+} = \int_{\mathbb{R}} k_+(p_2,p_2')\hat{f}_{p_2'^+}\mathrm{d}p_2 + \int_{\mathbb{R}} k_-(p_2,p_2')\hat{f}_{p_2'^-}\mathrm{d}p_2', \tag{6.1}$$

ядра $k_{\pm}(p_2,p_2')$ которых являются аналогами матричных элементов оператора. Домножив обе части равенства (6.1) на функцию $f_{p_2''^+}\in B_2$ и проинтегрировав по гиперболе $\rho=\{(\cosh\alpha,\pm1,\sinh\alpha)|\alpha\in\mathbb{R}\}$, получаем, что

$$F_{\rho}(\hat{T}(g)\hat{f}_{p_2^+},f_{p_2''^+}) = \int_{\mathbb{R}} k_+(p_2,p_2')\delta(p_2'+p_2'')\mathrm{d}p_2$$

(где δ – дельта-функция), откуда

$$k_+(p_2,p_2') = (2\pi)^{-1}A(-\sigma-1,p_2^+,-p_2'^+,g). \tag{6.2}$$

Таким образом, ядро $k_+(p_2,p_2')$ уже фактически вычислено в теореме 5.

Запишем теперь функции $\hat{f}_{p_2^{\pm}}$ в виде интегрального оператора

$$\hat{f}_{p_2^{\pm}} = \int_{\mathbb{R}} \lambda_{\pm}(p_2,p_1)\hat{f}_{p_1}\mathrm{d}p_1, \tag{6.3}$$

где ядра $\lambda_{\pm}(p_2,p_1)$ является аналогом матричного элемента матрицы перехода между базисами. Домножим обе части равенства (6.3) и проинтегрируем по окружности γ_r. Так как

$$H_{\gamma_r}\simeq H^{\circ}=\left\{h^{\circ}(\tau)=\begin{pmatrix}1&0&0\\0&\cos\tau&-\sin\tau\\0&\sin\tau&\cos\tau\end{pmatrix}\middle|\tau\in\mathbb{R}\right\},$$

$\gamma_r=\{x(\alpha)=(r,r\cos\alpha,r\sin\alpha)|\alpha\in\mathbb{R}\}$, $h^{\circ}(\tau)x(\alpha)=x(\alpha+\tau)$ для $x\in\gamma_r$, $\mathrm{d}\gamma_r=\mathrm{d}\alpha$, то получаем формулу, аналогичную (6.2):

$$\lambda_{\pm}(p_2,p_1) = (2\pi)^{-1}A(-\sigma-1,p_2^{\pm},-p_1,\mathrm{id}),$$

а значит, ядра $\lambda_{\pm}(p_2,p_1)$ уже вычислены в теоремах 4 и 5.

Переписав (6.2) согласно определению функции A в виде

$$k_+(p_2,p_2') = (2\pi)^{-1}F_{\gamma}(\hat{T}(h(\xi)\hat{f}_{p_2},f_{-p_2'}),$$

получим с учетом формулы (6.3), что

$$k_+(p_2,p_2') = (2\pi)^{-1}\int_{\mathbb{R}}\lambda(p_2,p_1)A(-\sigma-1,p_1,-p_2'^+,h(\xi))\mathrm{d}p_1,$$

где A по теореме 4 выражается через функцию Кулона. Так как $h^{-1}(\xi)=h(-\xi)$, то в силу теоремы 2

$$k_+(p_2,p_2') = (2\pi)^{-1}\int_{\mathbb{R}}\lambda_+(p_2,p_1)A(\sigma,-p_2'^+,p_1,h(-\xi))\mathrm{d}p_1,$$

т.е.

$$A(-\sigma-1, p_2^+, -p_2'^+, h(\xi)) = (2\pi)^{-1}\int\limits_{\mathbb{R}} A(-\sigma-1, p_2^{\pm}, -p_1, \mathrm{id})A(\sigma, -p_2'^+, p_1, h(-\xi))\mathrm{d}p_1. \tag{6.4}$$

Равенство (6.4) является континуальной теоремой сложения для функции A.

7. ЗАКЛЮЧЕНИЕ

Кулоновские волновые функции первоначально были определены как решение волнового уравнения Кулона, возникающего после разделения переменных для радиальной функции в задаче о движении электрона в кулоновском поле. В описании состояния электрона часто встречаются произведения двух кулоновских функций [27] или интегралы от таких произведений [28]. Интеграл, вычисленный в теореме 9, можно рассматривать как обобщение интегралов типа Меллина, рассматриваемых в [29]–[31]. Схожие с ним интегралы мы получаем в [10, разд. 8], используя подпредставление трехмерной лоренцевой группы на прямое произведение двух однопараметрических подгрупп.

Теорема сложения (6.4) остается справедливой и для других возможных пар $(\hat{p}, p)$, что приводит к новым интегральным формулам. Например, с помощью теорем 5 и 8 интегралы типа Фурье–Меллина от произведения функций $H_\sigma^+(\hat{p}; p)$ удается выразить через линейные комбинации двух функций F_1.

Учитывая, что функции, принадлежащие базису $\hat{B}_1$, являются собственными функциями оператора $\hat{T}(h_\omega(\tau))$, можно получить континуальную теорему сложения вида

$$\mathrm{e}^{-\mathrm{i}p_1} A(\ldots, -p_1, \mathrm{id}) = \int A(\ldots, -p_2, h_\omega(\tau))A(\ldots, p_2, -p_1, \mathrm{id})\mathrm{d}p_2,$$

которая приводит к представлениям волновых кулоновских функций в виде интеграла от произведения волновой кулоновской функции и функции Аппеля F_1.

Авторы благодарят рецензента за полезные замечания, способствовавшие улучшению статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Климык А.У., Качурик И.И.* Вычислительные методы в теории представлений групп. Киев: Вища школа, 1986.
2. *Миллер У.* Симметрия и разделение переменных. М.: Мир, 1981.
3. *Гельфанд И.М., Шилов Г.Е.* Обобщенные функции и действия над ними. М.: Физматлит, 1959.
4. *Шилин И.А., Чой Дж.* Алгебры Ли и специальные функции, связанные с изотропным конусом. Итоги науки и техники. Серия “Современная математика и ее приложения. Тематические обзоры”. (В печати).
5. *Шилин И.А.* Двойные SO(2,1)-инвариантные интегралы и формулы для функций Уиттекера. Известия вузов. Матем. 2011. № 5. С. 56–66.
6. *Вердиев Й.А.* Инварианты представлений групп Лоренца и их применения в дуальной модели физики частиц. Баку, АН Азербайджана, 1978.
7. *Shilin I.A., Choi J.* Certain relations between Bessel and Whittaker functions related to some diagonal and block-diagonal 3×3-matrices. J. Nonlinear Sci. Appl. 2017. V. 10. P. 560–574.
8. *Gaspard D.* Connection formulas between Coulomb wave functions. J. Math. Phys. 2018. V. 59. 112104.
9. *Chattarji D.* The Coulomb wave function from the viewpoint of the Lie algebra. Il Nuova cimento. 1967. V. 48 (2). P. 524–530.
10. *Shilin I.A., Choi J., Lee J.W.* Some integrals involving Coulomb functions associated with the three-dimensional proper Lorentz group. AIMS Mathematics. 2020. V. 5 (6). P. 5664–5682.
11. *Прудников А.П., Брычков Ю.А., Маричев О.И.* Интегралы и ряды. Элементарные функции. М.: Наука, 1981.
12. *Шилин И.А., Чой Дж.* Некоторые формулы для обычных функций и гиперфункций Бесселя–Клиффорда, связанные с собственной группой Лоренца. Фунд. и прикл. математика. 2019. Т. 22. № 5. С. 195–208.
13. *Bezrodnykh S.I.* Analytic continuation of the Lauricella’s functions $F_A^{(N)}$, $F_B^{(N)}$ and $F_D^{(N)}$. Int. Transforms. Spec. Func. 2020. V. 31 (11). P. 921–940.

14. *Bezrodnykh S.I.* Analytic continuation of the Lauricella function $F_D^{(N)}$ with arbitrary number of variables. Int. Transforms. Spec. Func. 2018. V. 29 (1). P. 21–42.
15. *Miller W.* Lie theory and the Lauricella functions F_D. J. Math. Phys. 1972. V. 13. P. 1393–1399.
16. *Miller W.* Lie theory and the Appell functions F_1. SIAM J. Math. Anal. 1973. V. 4 (4). P. 638–655.
17. *Виленкин Н.Я.* Гипергеометрические функции от нескольких переменных и вырожденные представления группы $SL(n, \mathbb{R})$. Изв. вузов. Матем. 1970. № 4. С. 50–55.
18. *Vilenkin N.Ja., Klimyk A.U.* Representations of Lie groups and special functions. Volume 3. Classical and quantum groups and special functions. Dordrecht, Kluwer, 1992.
19. *Kalnins E.G., Manocha H.L., Miller W.* The Lie theory of two-variable hypergeometric functions. Studies in Appl. Math. 1980. V. 62 (2). P. 143–173.
20. *Kniehl B.A., Tarasov O.V.* Finding new relationships between hypergeometric functions by evaluating Feynman integrals. Nucl. Phys. B. 2012. V. 854. P. 841–852.
21. *Lee J.-C., Yang Y.* The Appell function F_1 and Regge string scattering amplitudes. Phys. Letters. B. 2014. V. 739. P. 370–374.
22. *Shpot M.A.* A massive Feynman integral and some reduction relarions for Appell functions. J. Math. Phys. 2007. V. 48 (12). 123512.
23. *Carlson B.C.* Some series and bounds for incomplete elliptic integrals. J. Math. and. Phys. 1961. V. 40. P. 125–134.
24. *Scarpello G.M., Ritelli D.* π and the hypergeometric functions of complex argument. Journal of Number Theory. 2011. V. 131. P. 1887–1900.
25. *Градштейн И.С., Рыжик И.М.* Таблицы интегралов, рядов и произведений. М.: Наука, 1963.
26. *Прудников А.П., Брычков Ю.А., Маричев О.И.* Интегралы и ряды. Дополнительные главы. М.: Наука, 1986.
27. *Чибисов М.И., Ермолаев А.М., Шеркани М., Бруйар Ф.* Суммы произведений кулоновских волновых функций по вырожденным состояниям. ЖЭТФ. 2020. Т. 117 (2). С. 313–316.
28. *Ancarani L.U., Hervieux P.A.* Analytical formulas for Coulomb integrals involved in scattering problems. Phys. Rev. A. 1998. V. 58 (1). 336.
29. *Arnoldus A.G., George T.F.* Analytical evaluation of elastic Coulomb integrals. J. Math. Phys. 1992. V. 33 (2). P. 578–583.
30. *Nesbet R.K.* Analytical evaluation of integrals over Coulomb wave functions. Comp. Phys. Comm. 1988. V. 52 (1). P. 29–33.
31. *Sil N.C., Crees M.A., Seaton M.J.* Integrals involving products of Coulomb functions and inverse powers of the radial coordinate. J. Phys. B: Atomic and Molecular Physics. 1984. V. 17 (1). P. 1–21.

EDN: GBONRH
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1532–1550

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.6

ТЕСТИРОВАНИЕ БИКОМПАКТНЫХ СХЕМ ДЛЯ ОДНОМЕРНЫХ УРАВНЕНИЙ МАКСВЕЛЛА В СЛОИСТЫХ СРЕДАХ[1)]

© 2022 г. А. А. Белов[1,2,*], Ж. О. Домбровская[1,]**

[1] 119991 Москва, Ленинские горы, 1, стр. 2, МГУ им. М.В. Ломоносова, Россия

[2] 117198 Москва, ул. Миклухо-Маклая, 6, РУДН, Россия

**e-mail: aa.belov@physics.msu.ru*

***e-mail: dombrovskaya@physics.msu.ru*

Поступила в редакцию 26.01.2021 г.
Переработанный вариант 11.02.2022 г.
Принята к публикации 11.03.2022 г.

Проведено тестирование бикомпактных схем для системы одномерных уравнений Максвелла в стационарном и нестационарном случаях. Выполнены расчеты ряда представительных тестовых задач с обобщенными точными решениями. Некоторые из этих задач рассмотрены впервые. Проведено сравнение с наиболее популярными известными подходами: методом конечных разностей во временной области и методом конечных элементов. Эти расчеты убедительно показывают преимущества бикомпактных схем в задачах со слоистыми средами. Библ. 22. Фиг. 18.

Ключевые слова: уравнения Максвелла, бикомпактные схемы, слоистые среды, условия сопряжения, дисперсия вещества.

DOI: 10.31857/S0044466922070031

1. ВВЕДЕНИЕ

1. Задачи для системы уравнений Максвелла встречаются во многих физических и технических приложениях. Примерами являются расчеты элементов радиотехнических цепей и микросхем, распространение излучения по волноводам, оптика диэлектрических и плазмонных наноструктур. Многие реальные задачи рассматриваются в нестационарной постановке. В этом случае электромагнитное поле является не монохроматическим, а содержит набор различных частот.

2. Если задача рассматривается в однородной среде, либо показатель преломления плавно меняется с координатой, то поведение электромагнитного поля определяется уравнениями Максвелла. Если в задаче присутствуют границы раздела, на которых показатель преломления меняется скачком, то дополнительно учитывают условия сопряжения. Они связывают тангенциальные компоненты полей по обе стороны от границы раздела. Решение таких задач является обобщенным: на границе раздела поля испытывают излом или даже скачок. Это представляет серьезную трудность для численного расчета.

Подавляющее большинство конечно-разностных подходов (например, методы конечных разностей, конечных объемов и конечных элементов во временной области, см. [1]–[3]) основано на предположении о достаточной гладкости решения. Поэтому при наличии границ раздела сходимость резко ухудшается либо может вовсе отсутствовать. Фактически задачи в слоистых средах выходят за рамки применимости этого класса методов.

Вторая трудность заключается в том, что показатель преломления может зависеть не только от координат, но и от частоты электромагнитного поля. Такую среду называют диспергирующей. Во временно́м представлении это приводит к тому, что текущий вектор электрической индукции **D** зависит от вектора напряженности **E** не только в текущий момент времени, но и в предыдущие моменты. Система уравнений Максвелла становится интегродифференциальной. Такие задачи намного более трудны, чем чисто дифференциальные.

[1)]Работа выполнена при финансовой поддержке Программы стратегического академического лидерства РУДН.

Указанные трудности удалось преодолеть в предложенных недавно бикомпактных схемах (см. [4]–[6]). На данный момент они построены только для одномерного случая. Обоснование этих методов дано в [7].

3. Одновременно с разработкой новых методов возникает проблема их апробации. Для этого проводят расчеты некоторых тестовых задач. Хороший тест должен удовлетворять следующим требованиям. Во-первых, для него должно быть известно точное решение. Особенно ценны решения, выражающиеся в элементарных функциях, поскольку вычисление специальных функций зачастую представляет отдельную проблему. Во-вторых, решение должно содержать характерные особенности, присущие реальным задачам, – разрывы решения либо его производной на границах раздела сред.

Отметим, что в литературе нередко рассматриваются слишком простые задачи, например, задача о распространении импульса в однородном пространстве. Это не позволяет выявить поведение метода в сложных реальных задачах.

4. Как правило, при апробации нового метода устанавливают факт близости численного решения к точному на фиксированной сетке. Однако такое тестирование нельзя считать исчерпывающим. Необходимо проводить серию расчетов на наборе сеток, в которых от сетки к сетке шаг уменьшается. Такие сетки называют сгущающимися. На каждой сетке следует вычислять погрешность как разность численного и точного решений в некоторой норме и исследовать зависимость погрешности от шага сетки. Это позволяет не только установить сам факт сходимости, но и определить фактический порядок точности. Тест считается успешно пройденным, если фактическая скорость сходимости соответствует теоретической.

Данная работа посвящена апробации бикомпактных схем. Выполнены расчеты ряда представительных тестовых задач с обобщенными точными решениями. Некоторые из этих задач рассмотрены впервые. Проведено сравнение с наиболее популярными известными подходами: методом конечных разностей во временной области (finite-difference time domain, FDTD) и методом конечных элементов (finite element method, FEM). Эти расчеты убедительно показывают преимущества бикомпактных схем в задачах со слоистыми средами.

2. ОДНОМЕРНЫЕ ЗАДАЧИ

Важное место занимают задачи электродинамики плоско-параллельных структур. Примерами являются излучение плоских проводников и их взаимодействие с близко расположенными слоями диэлектриков (см. [8]), проектирование печатных плат (см. [9]), рассеяние падающего излучения набором диэлектрических либо металлических плоскопараллельных пластин (см. [10]) и формирование в них связанных оптических состояний и ряд других задач. В них электромагнитное поле зависит только от одной пространственной координаты. Приведем постановку задач, рассматриваемых в данной работе.

2.1. Стационарный случай

1. Рассмотрим слоистую структуру, состоящую из Q изотропных плоскопараллельных пластин общей толщиной a. Ориентируем координатную ось z перпендикулярно пластинам; оси x и y ориентированы в плоскости пластин. Обозначим координаты границ слоев через $0 = \xi_0 < \xi_1 < \ldots < \xi_Q = a$. При $z < 0$ и $z > a$ расположены полубесконечные диэлектрические среды. Будем считать последние однородными и изотропными.

Пластины могут быть как диэлектриками, так и проводниками и полупроводниками. Обозначим через ε_q диэлектрическую проницаемость, μ_q – магнитную восприимчивость, σ_q – проводимость q-й пластины (для диэлектрических пластин $\sigma_q = 0$).

2. В проводящих пластинах текут внешние токи с объемной плотностью $\mathbf{J}_q = \{J_q, 0, 0\}$, направленные вдоль оси x. Будем считать, что величины J_q зависят только от z. По границам раздела между q-й и $(q-1)$-й пластинами текут поверхностные токи с поверхностными плотностями $\mathbf{J}_q = \{j_q, 0, 0\}$, которые также имеют только x-компоненту.

Указанные токи излучают волны в обоих направлениях оси z, причем в этих волнах электрическое поле $\mathbf{E}$ направлено по оси x, магнитное поле $\mathbf{H}$ – по оси y. Кроме того, пусть на структуру из $z = -\infty$ и $z = +\infty$ нормально падают линейно поляризованные плоские волны с частотой ω и таким же направлением векторов $\mathbf{E}$ и $\mathbf{H}$.

3. Из-за омического нагрева пластины могут становиться оптически неоднородными, т.е. их показатель преломления может зависеть от координаты. Будем считать, что величины ε_q, μ_q, σ_q зависят от z и не зависят от x и y. Тогда и амплитуды полей зависят только от z.

Описанную задачу будем называть *одномерной*.

4. Пусть все объемные и поверхностные токи зависят от времени как $\sim e^{-i\omega t}$, где ω – частота колебаний. Тогда поля E_x и H_y также зависят от времени как $\sim e^{-i\omega t}$. Такую задачу называют *монохроматической* или *стационарной*.

5. Приведем математическую формулировку одномерной стационарной задачи.

Основными уравнениями электродинамики являются уравнения Максвелла. Традиционно в литературе рассматривают их дифференциальную форму. Однако по причинам, указанным далее, мы будем пользоваться интегральной формой этих уравнений. Она имеет следующий вид:

$$\int_\Gamma \mathbf{H}_q d\mathbf{l} = \frac{4\pi}{c}\int_S (\mathbf{J}_q + \sigma_q \mathbf{E}_q) d\mathbf{s} - \frac{i\omega}{c}\int_S \mathbf{D}_q d\mathbf{s}, \quad \mathbf{D}_q = \varepsilon_q \mathbf{E}_q, \quad 1 \le q \le Q; \tag{1}$$

$$\int_\Gamma \mathbf{E}_q d\mathbf{l} = \frac{i\omega}{c}\int_S \mathbf{B}_q d\mathbf{s}, \quad \mathbf{B}_q = \mu_q \mathbf{H}_q, \quad 1 \le q \le Q. \tag{2}$$

Здесь S – произвольная поверхность, ограниченная контуром Г. Используется система единиц СГС.

На внешних границах поставим условия излучения

$$\frac{\partial \mathbf{E}_1}{\partial z} + ik\mathbf{E}_1 = 2ik\mathbf{E}^0, \quad z = 0; \quad \frac{\partial \mathbf{E}_Q}{\partial z} - ik\mathbf{E}_Q = 2ik\mathbf{E}^a e^{-ika}, \quad z = a. \tag{3}$$

Здесь $\mathbf{E}^0$, $\mathbf{E}^a$ – заданные амплитуды волн, падающих справа и слева.

На границах раздела поставим условия сопряжения

$$\begin{gathered} \mathbf{e}_z \times (\mathbf{E}_q - \mathbf{E}_{q-1}) = 0, \quad \mathbf{e}_z(\mathbf{D}_q - \mathbf{D}_{q-1}) = 0, \\ \mathbf{e}_z \times (\mathbf{H}_q - \mathbf{H}_{q-1}) = \frac{4\pi}{c}\mathbf{J}_q, \quad \mathbf{e}_z(\mathbf{B}_q - \mathbf{B}_{q-1}) = 0. \end{gathered} \tag{4}$$

2.2. Нестационарный случай

В этом случае электромагнитное поле содержит не одну частоту, а *набор различных частот*.

В постановке (1)–(4) следует заменить множитель $(-i\omega)$ на производную ∂_t. Пусть объемные $J_q(z,t)$ и поверхностные $j_q(z,t)$ токи являются финитными функциями времени. Слева и справа на рассеиватель падают волновые пакеты $\mathbf{E}^0(t - z/c)\exp(-i\omega^0 t + ik^0 z)$, $\mathbf{E}^a(t + z/c)\exp(-i\omega^0 t - ik^0 z)$ с несущей частотой ω^0 и заданными огибающими $E^{0,a}$. Будем считать огибающие финитными функциями, а поля в начальный момент времени – нулевыми. Таким образом, постановка нестационарной задачи имеет следующий вид:

$$\int_\Gamma \mathbf{H}_q d\mathbf{l} = \frac{4\pi}{c}\int_S \mathbf{J}_q d\mathbf{s} + \frac{1}{c}\frac{\partial}{\partial t}\int_S \mathbf{D}_q d\mathbf{s}, \quad 1 \le q \le Q; \tag{5}$$

$$\int_\Gamma \mathbf{E}_q d\mathbf{l} = -\frac{1}{c}\frac{\partial}{\partial t}\int_S \mathbf{B}_q d\mathbf{s}, \quad 1 \le q \le Q; \tag{6}$$

$$\begin{gathered} \frac{\partial}{\partial z}\mathbf{E}_1 - \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}_1 = 2ik\mathbf{E}^0(t - z/c)\exp(-i\omega^0 t + ik^0 z), \quad z = 0; \\ \frac{\partial}{\partial z}\mathbf{E}_Q + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}_Q = 2ik\mathbf{E}^a(t + z/c)\exp(-i\omega^0 t - ik^0 z), \quad z = a; \end{gathered} \tag{7}$$

$$\begin{gathered} \mathbf{e}_z \times (\mathbf{E}_q - \mathbf{E}_{q-1}) = 0, \quad \mathbf{e}_z(\mathbf{D}_q - \mathbf{D}_{q-1}) = 0, \\ \mathbf{e}_z \times (\mathbf{H}_q - \mathbf{H}_{q-1}) = \frac{4\pi}{c}\mathbf{J}_q, \quad \mathbf{e}_z(\mathbf{B}_q - \mathbf{B}_{q-1}) = 0, \end{gathered} \tag{8}$$

$$\mathbf{E}_q = 0, \quad \mathbf{H}_q = 0, \quad t = 0, \quad 1 \le q \le Q, \quad t = 0. \tag{9}$$

3. БИКОМПАКТНЫЕ СХЕМЫ

1. Задачи в слоистых средах представляют особенную трудность для численного расчета. Из-за наличия изломов и разрывов решения схемы, использующие дифференцирование через границу раздела, страдают от катастрофического падения точности. Поэтому для таких задач требуются двухточечные консервативные схемы. Такие схемы называются *бикомпактными*.

Прообразом бикомпактных схем является классическая двухточечная схема Годунова (см. [11]), разработанная для задачи о распаде разрыва. Калиткин и Корякин впервые сформулировали идею бикомпактности и построили бикомпактную схему для уравнения теплопроводности в слоистой среде в одномерном и двумерном случаях (см. [12], [13]). Позднее Калиткин и Корякин построили бикомпактную схему для уравнения колебаний (см. [14]).

2. Общий подход к построению бикомпактных схем заключается в следующем (см. [14]). Введем сетку так, чтобы границы слоев были узлами. Такие сетки называются *специальными* (см. [15]). Запишем закон сохранения в интегральной форме и аппроксимируем интегралы квадратурами. При этом используем двухточечный шаблон, в котором задействованы 2 соседних узла сетки. Во всех внутренних узлах сетки запишем условия сопряжения для сеточных функций, а в граничных узлах – граничные условия. Подчеркнем, что реализация условий сопряжения и граничных условий должна быть одноточечной либо двухточечной.

В силу аддитивности интегралов и соответствующих квадратур, а также благодаря явному учету условий сопряжения из выполнения законов сохранения в отдельных ячейках сетки следует выполнение тех же законов сохранения во всей расчетной области. Поэтому описанные схемы являются консервативными по Тихонову–Самарскому (см. [16]). Именно консервативность обеспечивает сходимость к правильному обобщенному решению (см. [17]).

3. Для задач электродинамики исходными законами сохранения являются закон Гаусса для электрической индукции, закон Гаусса для магнитной индукции, закон Фарадея, теорема о циркуляции магнитной индукции. Они приводят к уравнениям Максвелла в интегральной форме. Именно поэтому мы ее используем.

Напомним, что условия сопряжения (4) являются следствием интегральных уравнений Максвелла. Чтобы это показать, рассматривают такой контур Г, чтобы часть его находилась по одну сторону от границы раздела, часть – по другую, а сам контур стягивают в точку, лежащую на границе раздела. Таким образом, схемы, в которых эти условия не учитываются, не могут быть консервативными.

4. Бикомпактная схема для уравнений (1)–(4) была предложена в [4]–[6]. Введем специальную сетку $0 = z_0 < z_1 < \ldots < z_N = a$, $z_{n+1} - z_n = \Delta z_n$. Поля могут испытывать разрыв, поэтому в каждом внутреннем узле нужно определить левое и правое предельные значения решения. Поэтому для каждого шага $\Delta z_{n-1/2}, 1 \le n \le N$, введем значения полей E_{2n-2}, H_{2n-2}, относящиеся к левой границе z_{n-1}, и значения полей E_{2n-1}, H_{2n-1}, относящиеся к правой границе z_n.

Разностная схема для задачи (1)–(4) имеет следующий вид:

$$\sqrt{\varepsilon_0}E_0 + \sqrt{\mu_0}H_0 = \sqrt{\varepsilon_0}E^0, \tag{10}$$

$$H_{2n} - H_{2n-2} - \frac{i\omega}{2c}A_{n-1/2}\Delta z_{n-1/2}(E_{2n} + E_{2n-2}) = \frac{4\pi}{c}(J_{n-1/2}\Delta z_{n-1/2} - j_n), \tag{11}$$

$$E_{2n} - E_{2n-2} - \frac{i\omega}{2c}\mu_{n-1/2}\Delta z_{n-1/2}(H_{2n-1} + H_{2n-2}) = \frac{4\pi i\omega}{2c^2}\mu_{n-1/2}\Delta z_{n-1/2}j_n, \tag{12}$$

$$H_{2N-1} - H_{2N-2} - \frac{i\omega}{2c}A_{N-1/2}\Delta z_{N-1/2}(E_{2N-1} + E_{2N-2}) = \frac{4\pi}{c}J_{N-1/2}\Delta z_{N-1/2}, \tag{13}$$

$$E_{2N-1} - E_{2N-2} - \frac{i\omega}{2c}\mu_{N-1/2}\Delta z_{N-1/2}(H_{2N-1} + H_{2N-2}) = 0, \tag{14}$$

$$\sqrt{\varepsilon_N}E_{2N-1} - \sqrt{\mu_N}H_{2N-1} = \sqrt{\varepsilon_N}E^a. \tag{15}$$

Здесь $A = -i\omega/c + 4\pi\sigma/c$. Индекс у ε, μ, σ, J соответствует номеру узла, в котором вычислена соответствующая величина; например, $\varepsilon_{n-1/2} \equiv \varepsilon_q(z_{n-1/2})$, где $z_{n-1/2} = 0.5(z_{n-1} + z_n)$. Поверхностные токи $j_n = j(z_n)$ относятся ко внутренним узлам. В уравнениях (11), (12) индекс n пробегает значения от 1 до $N - 1$. Таким образом, система (10)–(15) содержит $2N + 2$ уравнения.

5. Для нестационарных задач в [4]–[6] был предложен метод спектрального разложения. Кратко напомним его суть.

При распространении волнового пакета в линейной диспергирующей среде для разных спектральных компонент решения реализуются разные значения ε, μ, σ. Разложим пакет на монохроматические компоненты, решим стационарную задачу для каждой из них и просуммируем полученные решения. Спектральное разложение исходного пакета есть прямое преобразование Фурье, суммирование полученных решений – обратное преобразование Фурье. Оба преобразования выполняются с помощью численных квадратур. Описанный алгоритм назовем *нестационарной бикомпактной схемой*.

Таким образом, нестационарная задача сводится к набору стационарных задач. Данный подход имеет простой физический смысл. Он позволяет учитывать произвольный закон дисперсии, включая таблично заданный.

4. СТАЦИОНАРНЫЕ ТЕСТЫ

Для апробации бикомпактных схем мы провели расчеты ряда тестовых задач, у которых точное решение выражается в конечном виде через элементарные функции. Для простоты рассматривались диэлектрические (непроводящие) среды. Наличие ненулевой проводимости не вносит принципиальных трудностей. В приведенных ниже модельных задачах все размерные величины нормированы (т.е. являются относительными).

4.1. Однородная среда

1. Рассмотрим простейшую задачу – распространение плоской монохроматической волны через однородную диэлектрическую среду. Пусть ε, μ постоянны, и объемные и поверхностные токи отсутствуют $J = 0$, $j = 0$. Пусть слева на выделенную область $0 \leqslant z \leqslant a$ падает плоская монохроматическая волна.

2. Решение удобно строить с помощью перехода к волновому уравнению. Запишем одномерную систему уравнений Максвелла в дифференциальной форме

$$\frac{\partial E_x}{\partial z} = \frac{i\mu\omega}{c} H_y, \quad \frac{\partial H_y}{\partial z} = \frac{i\varepsilon\omega}{c} E_z, \quad 0 \leqslant z \leqslant a; \tag{16}$$

$$\frac{\partial E_x}{\partial z} + i\tilde{k}E_x = 2i\tilde{k}E^0, \quad z = 0; \qquad \frac{\partial E_x}{\partial z} - i\tilde{k}E_x = 0, \quad z = a. \tag{17}$$

Здесь $\tilde{k} = \omega\sqrt{\varepsilon\mu}/c$ – волновое число в среде. В записи граничных условий используется именно это значение, поскольку за пределами отрезка $[0, a]$ волна распространяется не в вакууме, а в той же самой среде.

Исключим из (16) H_y. Для этого продифференцируем первое из уравнений по z, выразим из него $\partial H_y/\partial z$ и подставим во второе уравнение. Получим

$$\frac{\partial^2 E_x}{\partial z^2} + \tilde{k}^2 E_x = 0. \tag{18}$$

Общее решение (18) есть

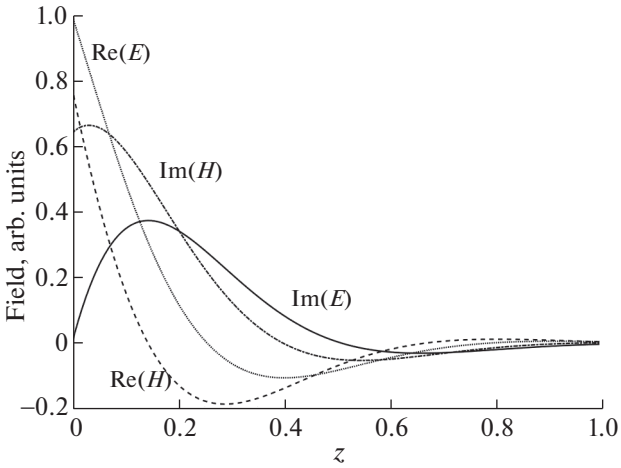
$$E_x = C_1 \exp(i\tilde{k}z) + C_2 \exp(-i\tilde{k}z). \tag{19}$$

С учетом граничных условий (17) имеем $C_1 = E^0$, $C_2 = 0$. Решение H_y найдем с помощью первого из уравнений (16). Таким образом, окончательное решение имеет вид

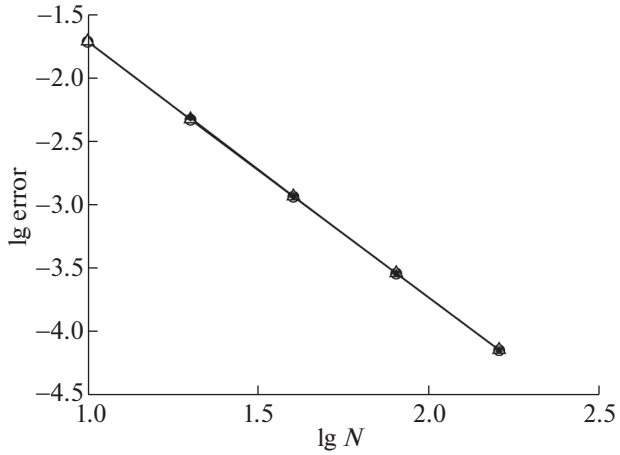
$$E_x = E^0 e^{i\tilde{k}z}, \quad H_y = E^0\sqrt{\frac{\varepsilon}{\mu}} e^{i\tilde{k}z}. \tag{20}$$

Оно известно из учебников.

3. Приведем результаты численных расчетов по предложенной стационарной схеме. Положим $E^0 = 1$, $\omega = \pi/2$, $c = 1$, $a = 1$. Рассмотрим сначала среду с поглощением: $\varepsilon = 1 + 5i$, $\mu = 5$. Такую задачу можно трактовать как жесткую: при $z = a$ амплитуды полей уменьшатся в



Фиг. 1. Решение задачи о среде с поглощением.

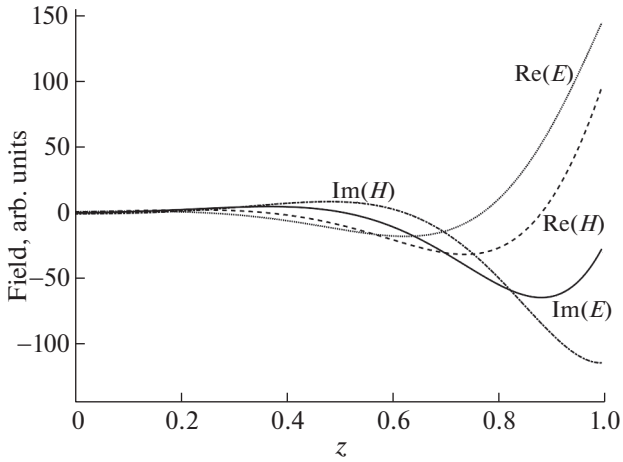


Фиг. 2. Погрешность решения в задаче о среде с поглощением: ○ – E_x, △ – H_y, светлые маркеры – разность численного и точного решений, темные маркеры – оценки по методу Ричардсона.

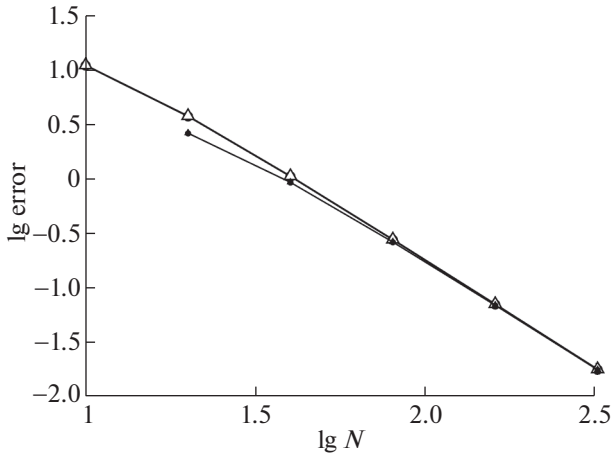
$\exp(i\tilde{k}a) \approx 150$ раз по сравнению с таковыми при $z = 0$. Решение этой задачи представлено на фиг. 1.

Расчеты проводились на наборе сгущающихся сеток. Погрешность расчета определялась двумя способами: во-первых, как разность численного решения и точного, во-вторых, с помощью апостериорной оценки по методу Ричардсона. Нормы L_2 полученных погрешностей для полей E_x и H_y приведены на фиг. 2 в двойном логарифмическом масштабе. Видно, что для обоих полей погрешность убывает. Линии погрешности являются прямыми с наклоном –2. Поэтому фактический порядок точности равен 2, что совпадает с теорией. Видно также, что оценка точности по методу Ричардсона практически совпадает с погрешностью, определенной непосредственно по разности численного решения и точного.

4. Аналогичные расчеты проводились для активной среды: $\varepsilon = 1 - 5i$, $\mu = 5$. В этом случае амплитуды полей при $z = 1$ оказываются в ≈150 раз больше, чем при $z = 0$. Такая задача является плохо обусловленной по левому граничному условию. Вид решения показан на фиг. 3, а полученные погрешности – на фиг. 4. В этом случае справедливы те же выводы, что и в предыдущем. Начальные участки графиков погрешности искривлены, но по мере сгущения сеток погрешность начинает убывать в соответствии с теоретическим вторым порядком точности.



Фиг. 3. Решение задачи об активной среде.



Фиг. 4. Погрешность решения в задаче об активной среде. Обозначения те же, что на фиг. 2.

4.2. Граница раздела диэлектриков

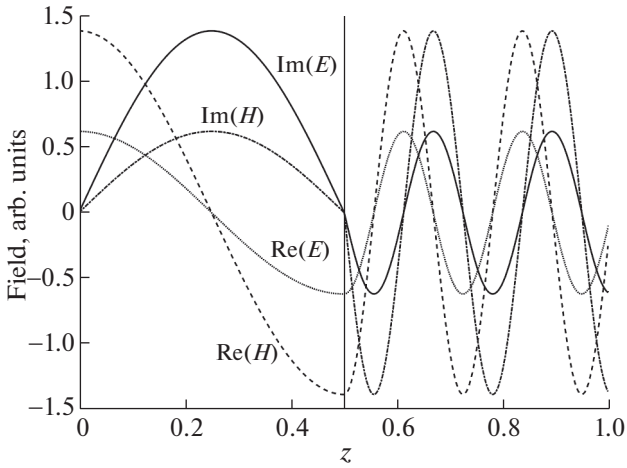
1. Пусть расчетная область состоит из двух однородных недиспергирующих сред с границей в точке $z = b$. Пусть слева от нее материальные параметры равны ε_1, μ_1, а справа – ε_2, μ_2. Пусть волна падает слева $E^0 = 1$, $E^a = 0$, а токи отсутствуют $J_{1,2} = 0$, $j_1 = 0$. Постановка задачи в форме одномерных дифференциальных уравнений Максвелла имеет следующий вид:

$$\frac{\partial E_x}{\partial z} = \frac{i\mu_1\omega}{c}H_y, \quad \frac{\partial H_y}{\partial z} = \frac{i\varepsilon_1\omega}{c}E_x, \quad 0 \leqslant z \leqslant b; \tag{21}$$

$$\frac{\partial E_x}{\partial z} = \frac{i\mu_2\omega}{c}H_y, \quad \frac{\partial H_y}{\partial z} = \frac{i\varepsilon_2\omega}{c}E_x, \quad b \leqslant z \leqslant a; \tag{22}$$

$$\frac{\partial E_x}{\partial z} + \frac{i\omega}{c}E_x = 2i\tilde{k}E^0, \quad z = 0; \quad \frac{\partial E_x}{\partial z} - \frac{i\omega}{c}E_x = 0, \quad z = a. \tag{23}$$

$$E_x\big|_{z=b-0} = E_x\big|_{z=b+0}, \quad H_y\big|_{z=b-0} = H_y\big|_{z=b+0}. \tag{24}$$



Фиг. 5. Решение задачи о диэлектрической границе раздела. Вертикальная прямая – граница раздела.

2. Получим точное решение этой задачи. От (21) и (22) перейдем к уравнениям типа (18) для областей $0 \leqslant z \leqslant b$ и $b \leqslant z \leqslant a$. Общее решение имеет вид

$$E_x = C_1 e^{i\tilde{k}_1 z} + C_2 e^{-i\tilde{k}_1 z}, \quad H_y = \sqrt{\frac{\varepsilon_1}{\mu_1}}(C_1 e^{i\tilde{k}_1 z} - C_2 e^{-i\tilde{k}_1 z}), \quad 0 \leqslant z \leqslant b; \tag{25}$$

$$E_x = C_3 e^{i\tilde{k}_2 z} + C_4 e^{-i\tilde{k}_2 z}, \quad H_y = \sqrt{\frac{\varepsilon_2}{\mu_2}}(C_3 e^{i\tilde{k}_2 z} - C_4 e^{-i\tilde{k}_2 z}), \quad b \leqslant z \leqslant a. \tag{26}$$

Из граничных условий (23) следует, что

$$C_1 = E^0 = 1, \quad C_4 = 0. \tag{27}$$

Подстановка в условия сопряжения дает

$$C_2 = E^0 \frac{\sqrt{\varepsilon_1/\mu_1} - \sqrt{\varepsilon_2/\mu_2}}{\sqrt{\varepsilon_1/\mu_1} + \sqrt{\varepsilon_2/\mu_2}} e^{i\tilde{k}_1 b}, \quad C_3 = E^0 \frac{2\sqrt{\varepsilon_1/\mu_1}}{\sqrt{\varepsilon_1/\mu_1} + \sqrt{\varepsilon_2/\mu_2}} e^{i(\tilde{k}_1 - \tilde{k}_2)b}. \tag{28}$$

Выражения (28) являются частным случаем общеизвестных формул Френеля.

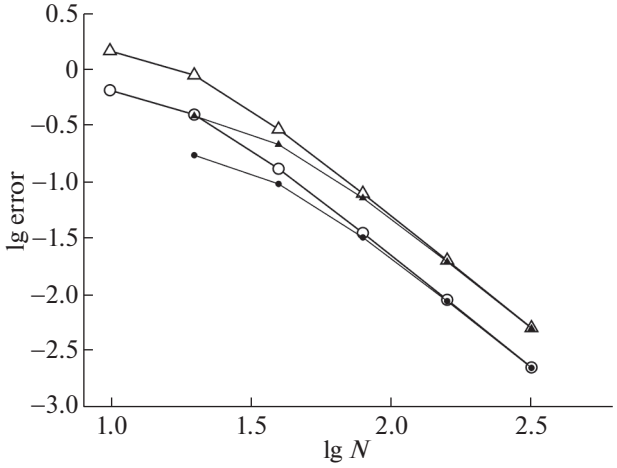
3. Положим $c = 1$, $\omega = \pi$, $a = 1$, $b = 0.5$. Пусть для левой среды $\varepsilon_1 = \mu_1 = 1$, а для правой – $\varepsilon_2 = 10$, $\mu_2 = 2$. На фиг. 5 представлен вид решения: при $z = b$ возникает слабый разрыв у $\operatorname{Im} E$ и $\operatorname{Im} H$, а $\operatorname{Re} E$ и $\operatorname{Re} H$ непрерывны.

Приведем результаты расчета по стационарной бикомпактной схеме. Выберем специальную сетку, у которой узел попадает в точку $z = b$. Очевидно, все последующие сетки, полученные последовательным уменьшением шага вдвое, также являются специальными. На фиг. 6 показана зависимость погрешности в норме L_2 от числа шагов сетки. Масштаб графика двойной логарифмический. Видно, что погрешности, полученные сравнением численного решения с точным, практически совпадают с оценками по методу Ричардсона. Эти кривые стремятся к прямым линиям с наклоном -2, соответствующим теоретическому порядку точности. Напомним, что именно это обеспечивает применимость метода Ричардсона.

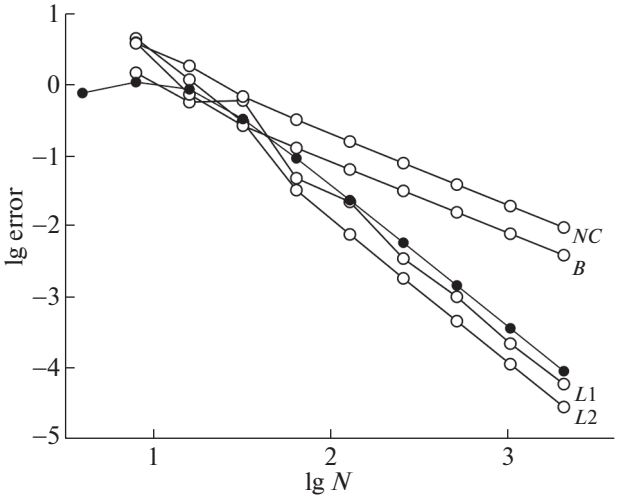
4. Сравним предлагаемую схему с методом конечных элементов на примере пакета FreeFEM++ (см. [18]). Расчеты проводились с использованием следующих элементов: лагранжевы элементы L1 1-го порядка, пузырьковые (B, bubble) элементы 1-го порядка (см. [19]), разрывные (NC, non-conforming) элементы 1-го порядка (см. [20]), лагранжевы элементы L2 2-го порядка. Для простоты мы задавали граничные условия для обоих полей E, H по значению точного решения на левой границе

$$E_x\big|_{z=0} = E_x^{\text{ex}}(0), \quad H_y\big|_{z=0} = H_y^{\text{ex}}(0). \tag{29}$$

Такая постановка соответствует, например, [21], [22].



Фиг. 6. Погрешность решения в задаче о диэлектрической границе раздела. Обозначения те же, что на фиг. 2.



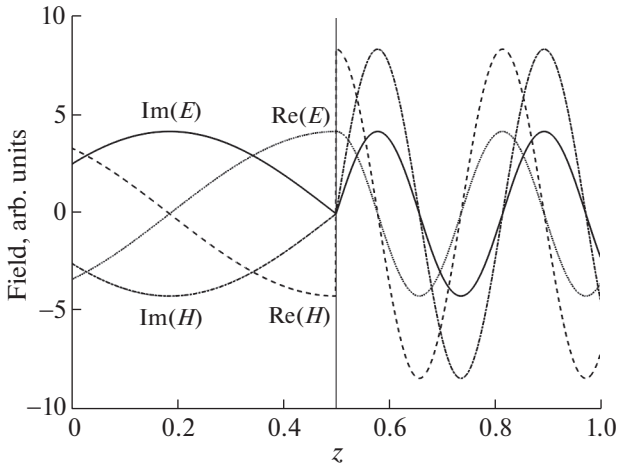
Фиг. 7. Усредненные погрешности (30) в задаче о диэлектрической границе раздела: ● – бикомпактная схема, ○ – метод конечных элементов в пакете FreeFEM++. Типы элементов указаны у кривых.

Расчет проводился на сгущающихся специальных сетках, и на каждой вычислялись погрешности δE, δH относительно точного решения. Чтобы не загромождать график, мы проводили дополнительное усреднение погрешностей по компонентам полей

$$\delta = \frac{1}{2}(\delta E + \delta H). \tag{30}$$

Величины (30) для перечисленных видов конечных элементов приведены на фиг. 7. Масштаб графика двойной логарифмический. Видно, что элементы L2 обеспечивают 2-й порядок точности, причем кривая погрешности является плавной. Для элементов L1 порядок точности в среднем близок ко 2-му, но кривая оказывается дерганой, т.е. отсутствует на ней прямолинейный участок теоретической сходимости. Это затрудняет контроль точности по Ричардсону и ухудшает надежность расчета. Элементы B и NC обеспечивают только первый порядок точности и значительно уступают по точности элементам L1 и L2.

На фиг. 7 приведена также усредненная погрешность бикомпактной схемы. Видно, что эта схема дает несколько худшую точность, чем элементы L2, и практически не уступает в точности элементам L1. При этом трудоемкость бикомпактной схемы такова же, как у элементов L1, и существенно меньше, чем у элементов L2.



Фиг. 8. Решение задачи о границе раздела с поверхностным током. Вертикальная прямая – граница раздела.

5. Таким образом, предложенная схема уверенно справляется с задачами, в которых решение испытывает слабый разрыв. При этом она практически не уступает методу конечных элементов.

4.3. Граница раздела с токами

1. Пусть в предыдущей задаче отсутствуют падающие волны $E^0 = E^a = 0$ и объемные токи $J_{1,2} = 0$, но по границе раздела течет поверхностный ток, излучающий волны в положительном и отрицательном направлении оси z. Такая модель описывает переизлучение тонкого металлического напыления на границе рассеивателя. Эта задача особенно трудна, поскольку у $\operatorname{Re} H$ возникает сильный разрыв, а у $\operatorname{Im} E$ и $\operatorname{Im} H$ – слабый. В литературе такая задача не рассматривалась.

Постановка задачи включает уравнения (21), (22), граничные условия (23); условия сопряжения на границе раздела $z = b$ формулируются следующим образом:

$$E_x\big|_{z=b-0} = E_x\big|_{z=b+0}, \quad H_y\big|_{z=b+0} - H_y\big|_{z=b-0} = \frac{4\pi}{c} j_1. \tag{31}$$

2. Выведем точное решение этой задачи. Перейдем от уравнений Максвелла к уравнениям типа (18) для областей $0 \leqslant z \leqslant b$ и $b \leqslant z \leqslant a$. Общее решение этих уравнений имеет вид (25), (26). Волны, приходящие из бесконечности, отсутствуют, поэтому граничные условия дают

$$C_1 = C_4 = 0. \tag{32}$$

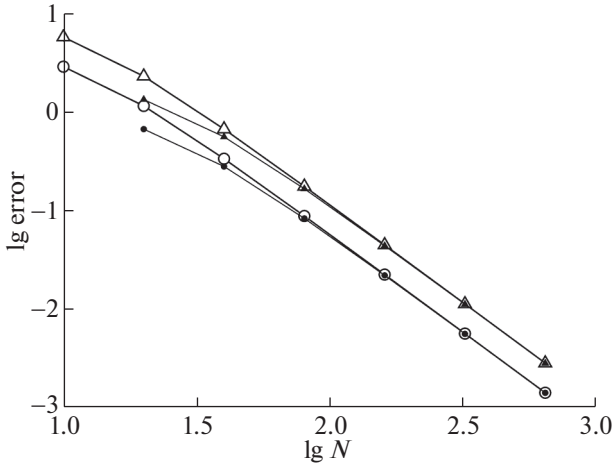
Подставим эти решения в условия сопряжения (31) и найдем C_2, C_3:

$$C_2 = \frac{4\pi}{c} j_1 \frac{e^{-i\tilde{k}_1 b}}{\sqrt{\varepsilon_1/\mu_1} + \sqrt{\varepsilon_2/\mu_2}}, \quad C_3 = \frac{4\pi}{c} j_1 \frac{e^{-i\tilde{k}_2 b}}{\sqrt{\varepsilon_1/\mu_1} + \sqrt{\varepsilon_2/\mu_2}}. \tag{33}$$

Полученное решение приведено на фиг. 8. Оно является новым.

3. В расчетах положим $j = 1$, а все остальные параметры выберем такими же, как в предыдущей задаче. Выберем специальную сетку указанным выше способом. На фиг. 9 показана зависимость погрешности в норме L_2 от числа шагов сетки в двойном логарифмическом масштабе. Видно, что значения погрешности, вычисленные по методу Ричардсона, практически совпадают с погрешностями относительно точного решения. Скорость убывания погрешности соответствует теоретическому второму порядку точности.

4. Был проведен расчет этой задачи в пакете FreeFEM++. Граничные условия задавались аналогично (29). Использовались те же 4 типа конечных элементов.



Фиг. 9. Погрешность решения в задаче о границе раздела с поверхностным током. Обозначения те же, что на фиг. 2.

В пакете FreeFEM++ не предусмотрено задание точечных источников, поэтому для имитации поверхностного тока вводился объемный ток следующего вида:

$$J = \frac{2\pi}{c}\delta(z - a/2) \approx \frac{4\pi}{cd}\begin{cases} \dfrac{x - x_l}{a/2 - x_l}, & x_l < x < a/2; \\ \dfrac{x - l/2}{x_r - a/2}, & a/2 < x < x_r; \\ 0, \quad 0 < x < x_l, & x_r < x < a. \end{cases} \tag{34}$$

Здесь $d \ll a$, $x_l = a/2 - d/2$, $x_r = a/2 + d/2$. Ток (34) является кусочно-линейной функцией координаты. Он отличен от нуля вне отрезка $[x_l, x_r]$, нарастает на отрезке $[x_l, a/2]$ и убывает на отрезке $[a/2, x_r]$. При этом выполняется условие нормировки

$$\int_0^a J dz = \frac{2\pi}{c}. \tag{35}$$

Погрешности, полученные в пакете FreeFEM++ в данной задаче, приведены на фиг. 10 в двойном логарифмическом масштабе. Здесь выбрано $\delta = 10^{-3}$. Видно, что для всех рассмотренных типов конечных элементов погрешность велика: она составляет $\sim 10^1$. При сгущении сеток погрешность не убывает, т.е. сходимость отсутствует. Аналогичные результаты имели место и при других значениях $d = 10^{-1}, 10^{-2}, 10^{-5}$.

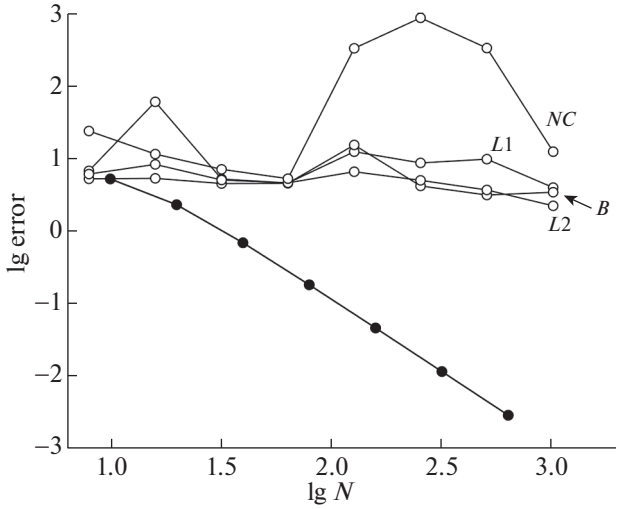
На фиг. 10 приведена также усредненная погрешность бикомпактной схемы. Видно, что она убывает в соответствии с теоретическим 2-м порядком сходимости. Точность бикомпактной схемы кардинально превосходит пакет FreeFEM++.

5. Таким образом, предлагаемая бикомпактная схема на специальных сетках позволяет проводить расчеты решений с сильными разрывами. Это свидетельствует об исключительно высокой надежности данного метода. Пакет FreeFEM++, основанный на методе конечных элементов, не позволил решить эту задачу, что убедительно показывает преимущества предлагаемой схемы.

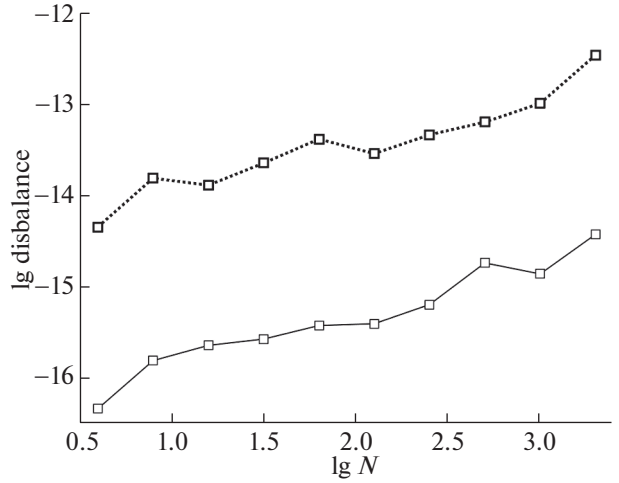
4.4. Плотность энергии электромагнитного поля

Рассмотрим плотность энергии электромагнитного поля

$$\mathscr{E} = \frac{1}{8\pi}(\varepsilon|E_x|^2 + \mu|H_y|^2). \tag{36}$$



Фиг. 10. Усредненные погрешности (30) в задаче о границе раздела с поверхностным током. Обозначения те же, что на фиг. 7.



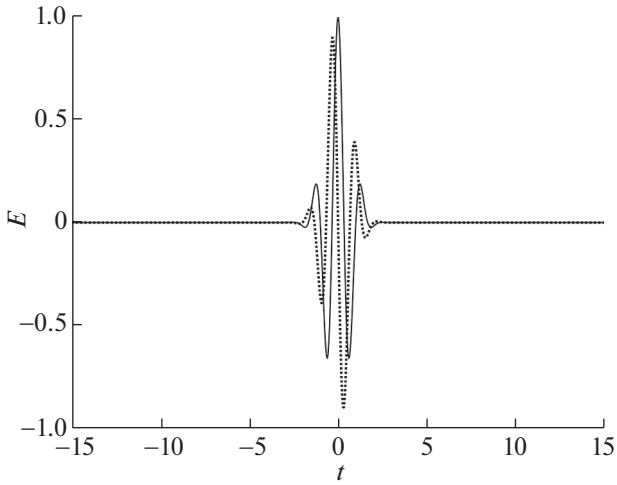
Фиг. 11. Дисбаланс энергии (37) для бикомпактной схемы. Сплошная линия – задача п. 4.2, пунктир – задача п. 4.3.

Если среда прозрачная (т.е. поглощение отсутствует), то энергия сохраняется при распространении волны.

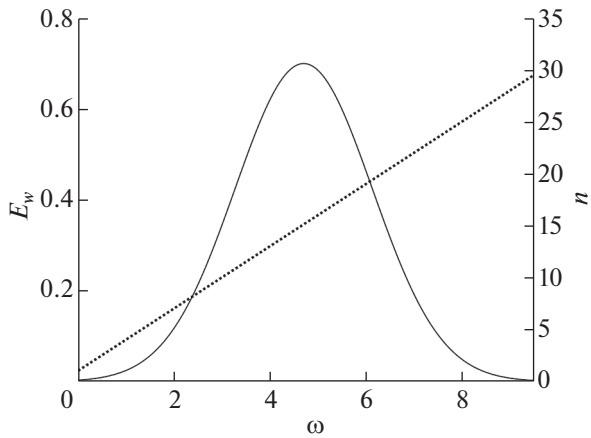
В задачах п. 4.2, 4.3 мы вычисляли $\mathscr{E}$ в каждом узле сетки. Затем вычислялась L_2-норма разности численного $\mathscr{E}_n$ и точного $\mathscr{E}^{\text{ex}}$ значений этой величины

$$\delta\mathscr{E} = \left\| \mathscr{E}^{\text{ex}}(z_n) - \mathscr{E}_n \right\|_{L_2}. \tag{37}$$

Величину $\delta\mathscr{E}$ можно интерпретировать как средний дисбаланс энергии электромагнитного поля. Расчеты велись со сгущением сеток, и средний дисбаланс вычислялся на каждой сетке. На фиг. 11 приведена зависимость $\delta\mathscr{E}$ от числа узлов сетки для обеих задач 4.2, 4.3. Масштаб графика двойной логарифмический. Видно, что в задаче 4.2 дисбаланс составляет $\delta\mathscr{E} \sim 3\times10^{-17}$–$3\times10^{-15}$, а в задаче 4.3 – $\delta\mathscr{E} \sim 3\times10^{-15}$–$3\times10^{-13}$. С увеличением N дисбаланс возрастает, причем средний наклон кривых равен 1/2. Это соответствует закону возрастания $\delta\mathscr{E} \sim N^{1/2}$. Такие значения дисбаланса и характер его роста соответствуют ошибкам компьютерного округления.



Фиг. 12. Временная развертка падающего импульса. Сплошная линия – $\mathrm{Re}\,E$, пунктир – $\mathrm{Im}\,E$.



Фиг. 13. Спектр падающего импульса (сплошная линия) и зависимость показателя преломления от частоты в задаче о диспергирующей среде (пунктир).

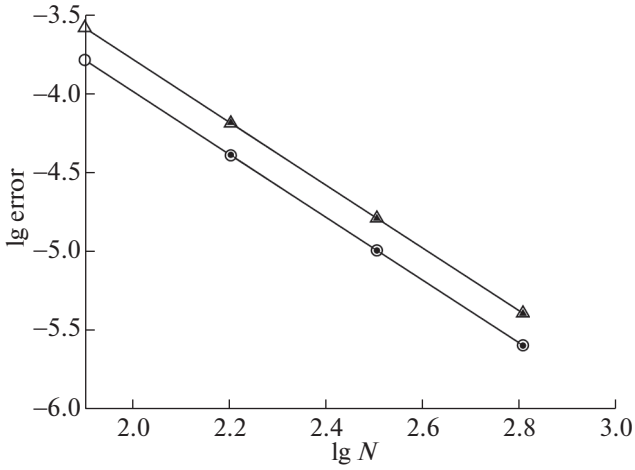
Таким образом, независимо от величины шага сетки предложенная бикомпактная схема передает закон сохранения энергии электромагнитной волны с точностью ошибок округления. Это является дополнительным преимуществом данной схемы.

5. НЕСТАЦИОНАРНЫЕ ТЕСТЫ

В нестационарном случае были проведены расчеты тех же трех тестовых задач: распространение излучения в однородной среде (в том числе диспергирующей), распространение излучения через границу раздела, излучение поверхностных токов). Для определенности в задачах этого раздела падающий импульс считается гауссовым

$$\mathscr{E}(t) = \exp\left(-\frac{t^2}{2W^2} - i\omega^0 t\right), \quad \omega^0 = \frac{3\pi}{2}\pi, \quad W = \frac{\pi}{4}. \tag{38}$$

Напомним, что все величины считаются обезразмеренными. Ширина импульса составляет $\approx\pi$ на половине высоты и $\approx 3\pi$ по основанию. Временная развертка импульса представлена на фиг. 12, а его спектр – на фиг. 13.



Фиг. 14. Погрешность решения в нестационарной задаче о среде с поглощением. Обозначения те же, что на фиг. 2.

5.1. Однородная среда

1. Рассмотрим первую задачу из разд. 4 в нестационарной постановке. Пусть в однородной среде с материальными параметрами $\varepsilon = 1 + 5i$, $\mu = 1$ распространяется импульс (38).

2. Построим точное решение этой задачи. Разложим падающий импульс $\mathscr{E}(t)$ в интеграл Фурье

$$E_\omega = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathscr{E}(t)e^{i\omega t}dt. \tag{39}$$

Решим задачу (1)–(4) для (3): $E^0 = E_\omega$, $E^a = 0$. Это решение имеет вид (20). Выполним обратное преобразование Фурье

$$\begin{gathered} E_x(z,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}E_\omega e^{i\tilde{k}z}e^{-i\omega t}d\omega = \mathscr{E}\left(t - \sqrt{\varepsilon\mu}\,\frac{z}{c}\right), \quad \tilde{k} = \frac{\omega\sqrt{\varepsilon\mu}}{c}, \\ H_y(z,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sqrt{\frac{\varepsilon}{\mu}}E_\omega e^{i\tilde{k}z}e^{-i\omega t}d\omega = \sqrt{\frac{\varepsilon}{\mu}}\mathscr{E}\left(t - \sqrt{\varepsilon\mu}\,\frac{z}{c}\right). \end{gathered} \tag{40}$$

Это решение известно из учебников.

3. Расчет проводился по нестационарной бикомпактной схеме. Полученные погрешности приведены на фиг. 14. Видно, что, во-первых, погрешности, найденные сравнением с точным решением, совпадают с оценками по Ричардсону, и, во-вторых, фактически порядок точности равен 2 и совпадает с теоретическим.

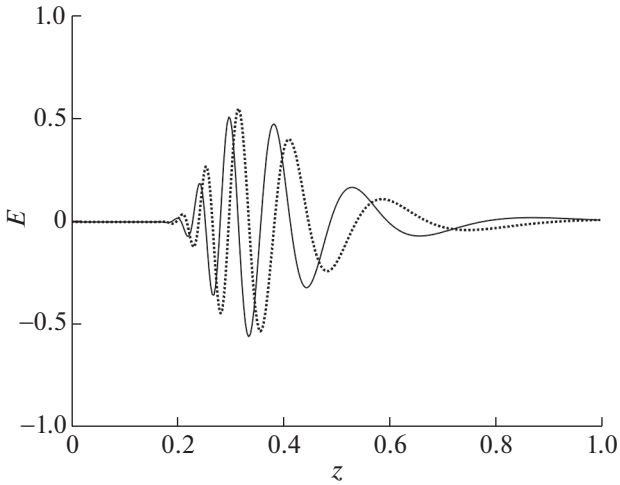
5.2. Диспергирующая среда

1. Рассмотрим предыдущую задачу в усложненной постановке. Пусть теперь среда является диспергирующей, т.е. величины $\varepsilon(\omega)$ и $\mu(\omega)$ зависят от частоты.

2. Точное решение по-прежнему выражается квадратурой (40), однако класс $\varepsilon(\omega)$ и $\mu(\omega)$, при которых ее удается вычислить в элементарных функциях, весьма узок.

Положим $\varepsilon(\omega) = \mu(\omega) = A_1\omega + A_0$. Тогда обе квадратуры в (40) берутся через интеграл Пуассона

$$\begin{gathered} E_x(z,t) = \frac{W}{\sqrt{2a}}\exp\left(\frac{b^2}{4a} - c\right), \quad H_y(z,t) = E_x(z,t), \\ a = \frac{W^2}{2} - i\frac{A_1}{c}z, \quad b = \omega^0 W^2 + i\left(\frac{A_0 z}{c} - t\right), \quad c = \frac{1}{2}(\omega^0 W)^2. \end{gathered} \tag{41}$$



Фиг. 15. Пространственная развертка решения в задаче о диспергирующей среде. Обозначения те же, что на фиг. 12.

Этот случай представляет большую ценность, поскольку обе компоненты решения выражаются явно. По-видимому, это решение является новым.

В конкретном расчете были выбраны значения $A_1 = 3$, $A_0 = 1$. Соответствующая зависимость $n(\omega) = \sqrt{\varepsilon(\omega)\mu(\omega)}$ показателя преломления от частоты приведена на фиг. 13 (правая ось ординат). Видно, что на отрезке частот, содержащихся в импульсе (38), показатель преломления изменяется почти в 30 раз: от $n = 1$ до $n \approx 30$. Поэтому данный тест достаточно представителен.

3. Проанализируем характер решения. На фиг. 15 показана зависимость E_x и H_y от z при фиксированном t. Видно, что высокочастотные компоненты, для которых $n(\omega)$ велико, распространяются медленно, а низкочастотные их заметно обгоняют. В результате импульс расплывается, причем несимметрично.

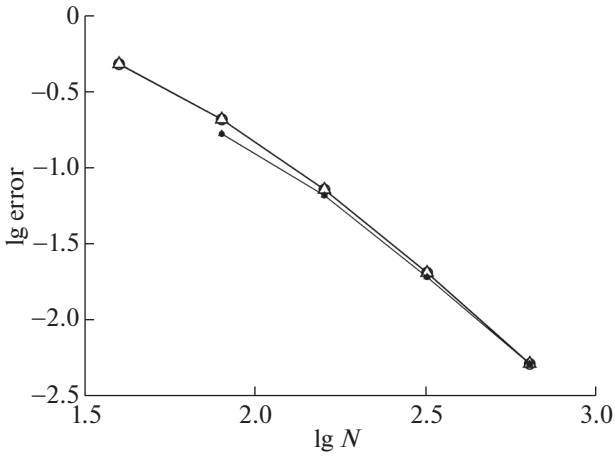
За время T область локализации импульса увеличивается на $cT(1/\min n - 1/\max n)$, это наибольший из пространственных масштабов задачи. Наименьший масштаб есть длина волны самой высокочастотной компоненты $\min\lambda = 2\pi c/\max(\omega n)$. Поэтому отношение наибольшего пространственного масштаба к наименьшему в данной задаче равно

$$\frac{cT(1/\min n - 1/\max n)}{2\pi c/\max(\omega n)} \approx 46T. \tag{42}$$

Видно, что, увеличивая T, разномасштабность (42) можно сделать сколь угодно большой. Поэтому задачу следует относить к жестким. Это является дополнительным преимуществом предложенного точного решения, поскольку оно оказывается хорошим тестом для алгоритмов адаптивного выбора шага по пространству и времени. Особенность этого теста заключается в том, что область, где нужно сгущать пространственную сетку, перемещается с течением времени, и ее положение заранее неизвестно. На данный момент такие алгоритмы еще не построены.

В проведенных расчетах полное время равнялось $T = 30$, и величина (42) достигала ~ 1400. Это достаточно высокая жесткость. Расчет проводился с постоянным шагом по пространству и времени. Поэтому требовались достаточно подробные сетки.

4. На фиг. 16 показаны полученные погрешности. Видно, что при сгущении сеток погрешности убывают в соответствии со вторым порядком точности, и оценки по методу Ричардсона хорошо согласуются с погрешностями, определенными непосредственно сравнением численного решения с точным. Этот тест убедительно подтверждает правильность предложенного метода учета дисперсии.



Фиг. 16. Погрешность решения в задаче о диспергирующей среде. Обозначения те же, что на фиг. 2.

5. Отметим еще два случая, когда хотя бы часть решения удается выразить явно. Во-первых, пусть $\mu = 1$, и $\varepsilon(\omega)$ представимо в виде отношения квадратов некоторых полиномов

$$\begin{gathered}\varepsilon(\omega) = \left(\frac{P_1(\omega)}{P_2(\omega)}\right)^2,\\ P_1(\omega) = A_n\omega^n + A_{n-1}\omega^{n-1} + \ldots + A_0,\\ P_2(\omega) = B_m\omega^m + B_{m-1}\omega^{m-1} + \ldots + B_0.\end{gathered} \tag{43}$$

Тогда интеграл для $H_y(z,t)$ берется через вычеты в нулях полинома P_2. Однако вычислить квадратуру для $E_x(z,t)$ не удается. Такой тест менее информативен, чем описанный выше. Тем не менее диэлектрическую проницаемость многих реальных материалов можно описать равенством (43). Поэтому данный случай также представляет значительную ценность.

Во-вторых, гауссова огибающая является моделью. Вместо нее можно использовать огибающую следующего вида:

$$\mathscr{E}(t) = \frac{A}{B + t^{2k}}, \tag{44}$$

где k – натуральное число. Качественное поведение (44) аналогично (38). В этом случае квадратуры (40) можно вычислить через вычеты при произвольном законе дисперсии $\varepsilon(\omega)$, $\mu(\omega)$.

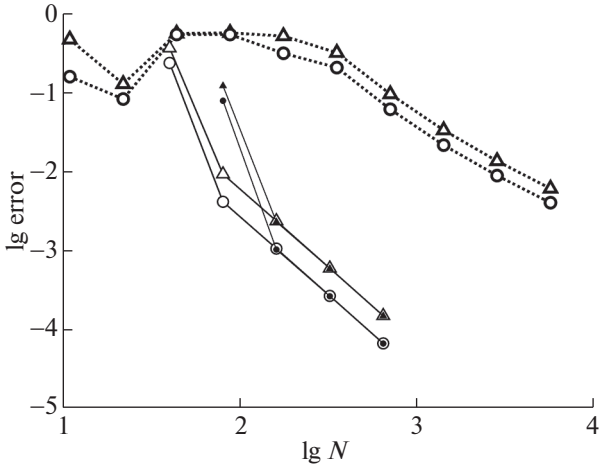
Насколько нам известно, в литературе эти случаи не описаны.

5.3. Граница раздела диэлектриков

1. Рассмотрим эту задачу в нестационарной постановке: пусть слева на границу раздела падает импульс вида (38).

2. Разложим падающий импульс в интеграл Фурье и для каждой спектральной амплитуды E_ω решим стационарную задачу (1)–(4). Точное решение последней имеет вид (25)–(28), где $E^0 = E_\omega$. Обратное преобразование Фурье дает точное решение нестационарной задачи: при $z < b$ решение состоит из падающей и отраженной волн

$$\begin{gathered}E_x = \mathscr{E}\left(\sqrt{\varepsilon_1\mu_1}\,\frac{z}{c} - t\right) + C_2\mathscr{E}\left(\frac{2b - z}{c}\sqrt{\varepsilon_1\mu_1} - t\right),\\ H_y = \sqrt{\frac{\varepsilon_1}{\mu_1}}\mathscr{E}\left(\sqrt{\varepsilon_1\mu_1}\,\frac{z}{c} - t\right) - \sqrt{\frac{\varepsilon_1}{\mu_1}}C_2\mathscr{E}\left(\frac{2b - z}{c}\sqrt{\varepsilon_1\mu_1} - t\right);\end{gathered} \tag{45}$$



Фиг. 17. Погрешность решения в нестационарной задаче о границе раздела диэлектриков. Сплошные линии – расчет по бикомпактной схеме, пунктир – по методу FDTD, другие обозначения те же, что на фиг. 2.

а при $z > b$ – из прошедшей

$$E_x = C_3\mathscr{E}\left(\frac{b}{c} + \sqrt{\varepsilon_2\mu_2}\,\frac{z-b}{c} - t\right),$$
$$H_y = C_3\sqrt{\frac{\varepsilon_2}{\mu_2}}\mathscr{E}\left(\frac{b}{c} + \sqrt{\varepsilon_2\mu_2}\,\frac{z-b}{c} - t\right). \tag{46}$$

Найти это решение в литературе не удалось. По-видимому, оно является новым.

3. Выберем такие же материальные параметры, как в п. 4: $\varepsilon_1 = \mu_1$, $\varepsilon_2 = 10$, $\mu_2 = 2$. Расчет проводился по нестационарной схеме, соответствующие погрешности приведены на фиг. 17. Видно, что, во-первых, погрешности, найденные сравнением с точным решением, совпадают с оценками по Ричардсону, и, во-вторых, фактический порядок точности равен 2 и совпадает с теоретическим.

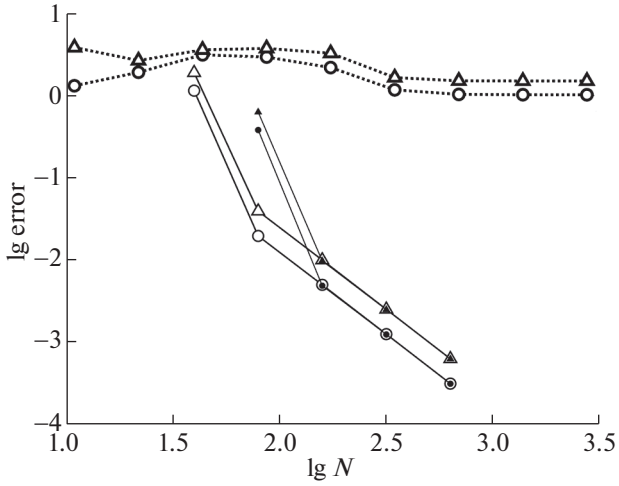
4. Для сравнения проводился расчет нестационарной задачи по схеме “с перешагиванием” метода FDTD. Граница раздела расположена в целом узле (напомним, что в этом методе к целым узлам относится поле E_x, а к полуцелым – поле H_y). Соответствующие погрешности приведены на фиг. 17. Видно, что они существенно хуже, чем у бикомпактной схемы, а порядок точности только 1-й (вместо теоретического 2-го).

Причина снижения порядка точности у метода FDTD заключается в том, что схема “с перешагиванием” некомпактна: шаблон занимает 1.5 шага по пространству и по времени. Поэтому, как бы ни была расположена граница раздела (в целом узле или в полуцелом), обеспечить достаточную гладкость решения внутри шаблона не удается. Это приводит к большой немонотонности решения (т.е. появлению сильных нефизичных осцилляций). Величина этих осцилляций составляет $O(h + \tau)$. Это и приводит к падению точности вблизи границ раздела.

5.4. Граница раздела с токами

1. В нестационарном случае зададим поверхностный ток на границе раздела в виде (38).

2. Построим точное решение этой задачи. Разложим поверхностный ток в интеграл Фурье и обозначим фурье-образ j_1 через j_ω. Решим стационарную задачу (1)–(4) при $E^0 = E^a = 0$ и $j_1 = j_\omega$. Ее точное решение имеет вид (25), (26), (32), (33).



Фиг. 18. Погрешность решения в нестационарной задаче о границе раздела с поверхностными токами. Обозначения те же, что на фиг. 17.

Выполним обратное преобразование Фурье и получим точное решение нестационарной задачи. При $z < b$ оно имеет вид

$$E_x(z,t) = \frac{4\pi}{c}\frac{1}{\sqrt{\varepsilon_1/\mu_1}+\sqrt{\varepsilon_2/\mu_2}}\, j_1\left(\frac{b-z}{c}\sqrt{\varepsilon_1\mu_1}-t\right),$$
$$H_y(z,t) = -\sqrt{\frac{\varepsilon_2}{\mu_2}}E_x(z,t). \tag{47}$$

При $z > b$ нужно поменять местами индексы 1 и 2 и изменить знак разности $(b - z) \to (z - b)$:

$$E_x(z,t) = \frac{4\pi}{c}\frac{1}{\sqrt{\varepsilon_1/\mu_1}+\sqrt{\varepsilon_2/\mu_2}}\, j_1([z-b]\sqrt{\varepsilon_2\mu_2}/c-t),$$
$$H(z,t) = -\sqrt{\frac{\varepsilon_2}{\mu_2}}E_x(z,t). \tag{48}$$

Решение (47), (48) является новым.

3. Погрешности расчета по нестационарной бикомпактной схеме приведены на фиг. 18. Видно, что, начиная со второй сетки ($N \approx 80$), погрешность убывает в соответствии со вторым порядком точности. Величины погрешностей, вычисленные по методу Ричардсона, практически совпадают с погрешностями, вычисленными по разности численного и точного решений.

4. Для сравнения был проведен расчет нестационарной задачи по явной схеме “с перешагиванием” метода FDTD. При этом расчет велся по однородной схеме, т.е. без явного выделения особенности. Соответствующие погрешности даны на фиг. 18. Видно, что этот метод вообще не может обеспечить сходимость: при уменьшении шага погрешность не убывает.

5. Таким образом, приведенные примеры расчетов убедительно показывают преимущества предложенного метода по сравнению с известными подходами.

Авторы искренне благодарны Н.Н. Калиткину за ценные замечания и обсуждения и А.Н. Боголюбову за внимание к работе.

СПИСОК ЛИТЕРАТУРЫ

1. *Inan U.S., Marshall R.A.* Numerical electromagnetics. The FDTD method. Cambridge University Press, Cambridge, 2011.
2. *Sullivan D.M.* Electromagnetic simulation using the FDTD method. IEEE Press. 2000.
3. *Taflove A., Johnson S.G., Oskooi A.* Advances in FDTD Computational Electromagnetics: Photonics and Nanotechnology. Artech House, London. 2013.

4. *Белов А.А., Домбровская Ж.О.* Бикомпактная разностная схема для уравнений Максвелла в слоистых средах // Докл. АН. 2020. Т. 492. С. 15–19.
5. *Белов А.А., Домбровская Ж.О.* Прецизионные методы расчета нестационарных задач интегральной фотоники // Изв. РАН. Сер. физ. 2022. Т. 86. № 2. С. 259–265.
6. *Belov A.A., Dombrovskaya Zh.O., Bogolyubov A.N.* A bicompact scheme and spectral decomposition method for difference solution of maxwell's equations in layered media // Comput. and Math. with Appl. 2021. T. 96C. C. 178–187.
7. *Белов А.А., Домбровская Ж.О.* Прецизионные методы решения одномерных уравнений Максвелла в слоистых средах // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 1. С. 51–65.
8. *Харвей А.Ф.* Техника сверхвысоких частот. М.: Сов. радио. 1965.
9. *Ушаков Н.Н.* Технология элементов вычислительных машин. М.: Высш. школа, 1976.
10. *Егоров А.А., Ловецкий К.П., Севастьянов Л.А., Хохлов А.А.* Многослойные оптические покрытия. М.: Изд. РУДН, 2014.
11. *Годунов С.К.* Разностный метод численного расчета разрывных решений уравнений гидродинамики // Матем. сб. 1959. Т. 47 (89). № 3. С. 271–306.
12. *Калиткин Н.Н., Корякин П.В.* Бикомпактные схемы и слоистые среды // Докл. АН. 2008. Т. 419. № 6. С. 744–748.
13. *Калиткин Н.Н., Корякин П.В.* Одномерные и двумерные бикомпактные схемы в слоистых средах // Матем. моделирование. 2009. Т. 21. № 8. С. 44–62.
14. *Калиткин Н.Н., Корякин П.В.* Численные методы. Т. 2. Методы математической физики. М.: Академия, 2013.
15. *Толстых А.И.* Компактные разностные схемы и их применение в задачах аэрогидродинамики. М.: Наука, 1990.
16. *Тихонов А.Н., Самарский А.А.* О сходимости разностных схем в классе разрывных коэффициентов // Докл. АН. 1959. Т. 8. № 3. С. 529–532.
17. *Самарский А.А.* Теория разностных схем. М.: Наука, 1989.
18. *Hecht F.* New development in freefem++ // J. of Numeric. Math. 2012. V. 20. № 3–4. P. 251–266.
19. *Yvonnet J., Villon P., Chinesta F.* Bubble and hermite natural element approximations // Lect. Not. on Comput. Sci. and Engineer. 2007. C. 283–298.
20. *Droniou J., Eymard R., Thierry G., Guichard C., Herbin R.* The gradient discretisation method // Chapter Nonconforming finite element methods. C. 285–305. 07 2018.
21. *Тихонравов А.В.* Амплитудно-фазовые свойства спектральных коэффициентов слоистых сред // Ж. вычисл. матем. и матем. физ. 1985. Т. 25. № 3. С. 442–450.
22. *Furman Sh.A., Tikhonravov A.V.* Basics of optics of multilayer systems. Gifsur-Yvette. 1992.

EDN: AVIENM
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1551–1562

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 517.9

МАССОВЫЙ ПОТОК И РАСПРЕДЕЛЕНИЕ ДАВЛЕНИЯ ГАЗА В ДЛИННОМ КОНЦЕНТРИЧЕСКОМ КОЛЬЦЕВОМ КАНАЛЕ ПРИ НЕПОЛНОЙ АККОМОДАЦИИ МОЛЕКУЛ ГАЗА

© 2022 г. О. В. Гермидер[1], В. Н. Попов[1,*]

[1] *163002 Архангельск, набережная Северной Двины, 17 , Северный (Арктический) федеральный ун-т, Россия*
**e-mail: v.popov@narfu.ru*

Поступила в редакцию 12.10.2021 г.
Переработанный вариант 22.01.2022 г.
Принята к публикации 11.05.2022 г.

В зависимости от значений коэффициента аккомодации тангенциального импульса молекул газа найдено решение линеаризованной задачи об изотермическом течении разреженного газа в длинном концентрическом кольцевом канале. В качестве основного уравнения используется модельное кинетическое уравнение Бхатнагара–Гросса–Крука, а в качестве граничного условия – модель зеркально-диффузного отражения молекул газа от стенок канала. Получены значения массового потока и распределение давления по длине канала при произвольных значениях параметра разрежения. Проведено их исследование в свободномолекулярном и гидродинамическом режимах течения газа в канале. Библ. 25. Фиг. 7.

Ключевые слова: модельные кинетические уравнения, течение газа в каналах, коэффициент аккомодации тангенциального импульса молекул газа.

DOI: 10.31857/S0044466922090113

ВВЕДЕНИЕ

Исследование течений газа в микро- и наноканалах имеет важное значение при разработке технических систем [1]. Характер этих течений определяется соответствующими процессами переноса, экспериментальное изучение которых затруднено [2]. При использовании кинетического подхода, основанного на решении уравнения Больцмана, появляется возможность восстановления интегральных характеристик течения с последующим их анализом в широком диапазоне изменения значений параметра разрежения без проведения сложно реализуемого физического эксперимента [3], [4]. Для получения численного решения уравнения Больцмана в режимах течения газа, отличных от свободномолекулярного, необходимы значительные вычислительные ресурсы, сократить которые позволяет применение модельных кинетических уравнений [5]–[11]. Наиболее простым и широко используемым для описания течения газа является модельное кинетическое уравнение Бхатнагара–Гросса–Крука (БГК) [12]. Сравнительный анализ [13] показал, что уравнение БГК позволяет получить вполне удовлетворительные результаты при описании изотермических течений. Однако известные к настоящему моменту времени методы его решения являются ресурсозатратными [11]. В связи с этим продолжает оставаться актуальной задача поиска более эффективных и менее ресурсозатратных вычислительных методов и алгоритмов решения модельных кинетических уравнений для описания процессов переноса в каналах.

В представленной работе исследуется один из аспектов общей проблемы изотермического течения газа в канале, связанный с восстановлением массового потока и давления газа в длинном концентрическом кольцевом канале. В постановочной части данная работа близка к [6] и представляет интерес как с прикладной, так и с теоретической точек зрения. С одной стороны, каналы с такой конфигурацией сечения имеют место в практическом приложении [7]. С другой стороны, особая геометрия сечения позволяет проследить изменение свойств потока разреженного газа, вызванное его торможением у внутренней цилиндрической поверхности. В качестве характерного размера поперечного сечения канала в [6] выбран гидродинамический (гидравлический) диаметр [14]. Численное решение уравнения БГК в [6] получено методом дискретных скоростей, анализ результатов выполнен в широком диапазоне разреженности газа с получением и гидродинамического предела. В данной работе развивается подход, предложенный в [10] и [15].

Оптимизация построения численного решения БГК достигается с применением свойств сумм полиномов Чебышёва [19]. В качестве характерного размера поперечного сечения канала, как и в [20], используется радиус внутреннего цилиндра. Значения массового потока и распределение давления находятся с использованием системы компьютерной алгебры Maple 18. Проводится сравнительный анализ с результатами из [6] и с аналитическими решениями в свободномолекулярном режиме и режиме течения со скольжением, которые выписываются в явном виде. Исследуется влияние коэффициента аккомодации тангенциального импульса молекул газа на величину массового потока газа и давления газа в канале. Предлагаемый в работе метод может быть применен к широкому классу задач, связанных с решением линеаризованных модельных кинетических уравнений, в частности, к расчету течений газа в каналах при различных поперечных сечениях канала. Данный метод без использования свойства конечных сумм был применен в [16] для решения эллипсоидально-статистического модельного уравнения в рамках задачи течения газа между двумя параллельными пластинами. Следует отметить также возрастание точности при увеличении гладкости решения и вычислительную устойчивость предлагаемого метода [17] и [18].

1. ПОСТАНОВКА ЗАДАЧИ. КИНЕТИЧЕСКОЕ УРАВНЕНИЕ

Рассмотрим изотермическое течение разреженного газа в длинном концентрическом кольцевом канале, образованном двумя цилиндрами с радиусами R_1' и R_2' ($R_1' < R_2'$), под действием заданного градиента давления, направленного вдоль оси канала z'. Считаем, что цилиндры поддерживаются при постоянной температуре и коэффициенты аккомодации тангенциального импульса молекул газа совпадают и равны α. Канал соединяет два резервуара, давления в которых обозначим через p_1' и p_2' соответственно. Полагаем, что $p_2' > p_1'$ и длина канала $L' \gg R_2'$. Состояние газа в точке $\mathbf{r}'$ определяем функцией распределения молекул газа $f'(\mathbf{r}', \mathbf{v})$, где $\mathbf{v}$ – молекулярная скорость газа. В качестве масштабов длины, скорости, концентрации, температуры, функции распределения выберем соответственно величины [8]: R_2', $\beta^{-1/2}$, n_0', T_0', $n_0'\beta^{3/2}$, где $\beta = m'/(2k_Б T_0')$, $k_Б$ – постоянная Больцмана, m' – масса молекул газа, n_0', T_0' – концентрация, температура газа в некоторой точке, принятой за начало координат; $p' = n' k_Б T'$. Тогда для безразмерных величин имеем следующие соотношения:

$$\mathbf{r} = \frac{\mathbf{r}'}{R_2'}, \quad R_1 = \frac{R_1'}{R_2'}, \quad R_2 = 1, \quad f = \frac{f'}{n_0'\beta^{3/2}},$$

$$\mathbf{C} = \beta^{1/2}\mathbf{v}, \quad \mathbf{u} = \beta^{1/2}\mathbf{u}', \quad n = \frac{n'}{n_0'}, \quad T = 1.$$

Полагаем, что модуль безразмерного градиента давления G_p является малым по величине. Тогда при малом перепаде давления на концах длинного канала ($L' \gg R_2'$) распределение давления вдоль канала можно считать линейным [13]:

$$p_0' = p_{av}', \quad p'(z) = p_{av}'(1 + G_p z), \quad G_p = \frac{p_2' - p_1'}{L p_{av}'}, \tag{1.1}$$

где $p_{av}' = (p_2' + p_1')/2$.

Если отношение $p_2^* = p_2'/p_1'$ является большим, то распределение давления перестает быть линейным. Рассматривая течение газа в средней части канала и вводя обозначения $p^* = p'/p_1'$ и $z^* = z'/L'$, находим $p^*(z^*)$ из условия постоянства полного потока массы J_M' (уравнения неразрывности) [13]:

$$J_M^* dz^* = J_{M,p} dp^*, \tag{1.2}$$

с граничными условиями $p^*(-1/2) = 1$ и $p^*(1/2) = p_2^*$. Здесь J_M^* – неизвестный параметр, представляющий собой нормированный поток массы

$$J_M^* = \frac{L}{\pi(R_2'^2 - R_1'^2)p_1'} J_M' \sqrt{\frac{2k_Б T_1'}{m'}}, \tag{1.3}$$

а величина приведенного потока массы $J_{M,p}$:

$$J_{M,p}G_p = \frac{1}{\pi(R_2'^2 - R_1'^2)p_0'} J_M' \sqrt{\frac{2k_Б T_0'}{m'}}, \tag{1.4}$$

определяется с использованием линеаризованного модельного кинетического уравнения БГК. Зависимость $J_{M,p}$ от z^* проявляется неявно, через параметр разреженности δ. Для модели жестких сфер имеем следующее соотношение [10]

$$\delta = \frac{R_2' p'}{\eta_g'(T')} \sqrt{\frac{m'}{2k_Б T'}}, \tag{1.5}$$

где $\eta_g'(T')$ – динамическая вязкость газа.

Учитывая осесимметричный характер течения газа в канале, введем цилиндрические координаты $\mathbf{r} = (\rho, r_\varphi, r_z)$ в конфигурационном пространстве и $\mathbf{C} = (C_\perp, C_\psi, C_z)$ в пространстве скоростей. Пусть $\zeta = \cos\psi$, в линейном приближении получим

$$\begin{gathered} f(\mathbf{r},\mathbf{C}) = f_0(C)\big(1 + G_p(z + h(\rho,\mathbf{C}))\big), \quad f_0(C) = \pi^{-3/2}\exp(-C^2), \\ J_{M,p} = \frac{4}{(1-R_1^2)}\int_{R_1}^{1} U_z(\rho)\rho d\rho, \end{gathered} \tag{1.6}$$

$$\begin{gathered} u_z = G_p U_z, U_z(\rho) = \frac{2}{\pi}\int_0^{+\infty} \exp(-C_\perp^2) C_\perp \int_{-1}^{1} \frac{1}{\sqrt{1-\zeta^2}} Z(\rho,\zeta,C_\perp) d\zeta dC_\perp, \\ Z(\rho,\zeta,C_\perp) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} \exp(-C_z^2) C_z h(\rho,\mathbf{C}) dC_z. \end{gathered} \tag{1.7}$$

Линеаризованное уравнение БГК запишем в виде [15]

$$\left(\frac{\partial Z}{\partial \rho}\zeta + \frac{\partial Z}{\partial \zeta}\frac{(1-\zeta^2)}{\rho}\right) C_\perp + \delta Z(\rho,\zeta,C_\perp) + \frac{1}{2} = \delta U_z(\rho). \tag{1.8}$$

В качестве граничного условия на цилиндрах используем модель зеркально-диффузного отражения Максвелла [13]. В этом случае имеем

$$Z(1,\zeta,C_\perp) = (1-\alpha)Z(1,C_\perp,-\zeta), \quad \zeta < 0, \tag{1.9}$$

$$Z(R_1,\zeta,C_\perp) = (1-\alpha)Z(R_1,C_\perp,-\zeta), \quad \zeta > 0. \tag{1.10}$$

2. РЕШЕНИЕ КРАЕВОЙ ЗАДАЧИ

Неизвестную функцию $Z(\rho,\zeta,C_\perp)$, где $\rho \in [R_1,1]$, $\zeta \in [-1,1]$ и $C_\perp \in [0,+\infty)$, раскладываем в ряд по полиномам Чебышёва I рода T_{k_i} [19] и, ограничиваясь в этом ряду членами с номерами $k_i \le n_i$ $(i = \overline{1,3})$, получаем

$$Z(\rho,\zeta,C_\perp) = \mathbf{T}_1(x_1) \otimes \mathbf{T}_2(x_2) \otimes \mathbf{T}_3(x_3)\mathbf{A}, \tag{2.1}$$

где $x_1 = (2\rho - 1 - R_1)/(1 - R_1)$, $x_2 = \zeta$, $x_3 = (C_\perp - 1)/(C_\perp + 1)$, $\mathbf{T}_i$ – матрица размера $1 \times n_i'$ $(n_i' = n_i + 1$, $i = \overline{1,3})$:

$$\mathbf{T}_i(x_i) = \big(T_0(x_i)T_1(x_i)\ldots T_{n_i-1}(x_i)T_{n_i}(x_i)\big),$$

$\mathbf{A}$ – матрица размера $n_1'n_2'n_3' \times 1$:

$$\mathbf{A} = \big(a_{000}a_{001}\ldots a_{n_1n_2n_3-1}a_{n_1n_2n_3}\big)^{\mathrm{T}},$$

через $\mathbf{T}_1(x_1) \otimes \mathbf{T}_2(x_2)$ обозначено произведение Кронекера двух матриц [21].

Подставляя (2.1) в (1.8)–(1.10) и учитывая (1.7), получаем

$$\mathbf{B}(x_1,x_2,x_3)\mathbf{A} = -\frac{1}{2}, \tag{2.2}$$

$$\mathbf{T}_1(\pm 1) \otimes (\mathbf{T}_2(x_2) - (1-\alpha)\mathbf{T}_2(-x_2)) \otimes \mathbf{T}_3(x_3)\mathbf{A} = 0, \quad \pm x_2 < 0, \tag{2.3}$$

где

$$\mathbf{B}(x_1,x_2,x_3)=\frac{2}{1-R_1}\left(\frac{d\mathbf{T}_1(x_1)}{dx_1}\otimes\mathbf{T}_2(x_2)x_2+\mathbf{T}_1(x_1)\otimes\frac{d\mathbf{T}_2(x_2)}{dx_2}\frac{(1-x_2^2)}{x_1+(1+R_1)/(1-R_1)}\right)\otimes$$

$$\otimes\,\mathbf{T}_3(x_3)\frac{1+x_3}{1-x_3}+\delta\mathbf{T}_1(x_1)\otimes\left(\mathbf{T}_2(x_2)\otimes\mathbf{T}_3(x_3)-\mathbf{P}_2\otimes\mathbf{P}_3\right),$$

$$\mathbf{P}_2=\frac{2}{\pi}\int_{-1}^{1}\frac{\mathbf{T}_2(x_2)}{\sqrt{1-x_2^2}}dx_2=(20\ldots00),$$

$$\mathbf{P}_3=2\int_{-1}^{1}\frac{1+x_3}{(1-x_3)^3}\mathbf{T}_3(x_3)\exp\left(-\frac{(1+x_3)^2}{(1-x_3)^2}\right)dx_3.$$

В качестве точек коллокации в (2.2) для x_1 выберем точки экстремумов многочлена $T_{n_1}(x_1)$ на отрезке $[-1;1]$, для x_2 и для x_3 – нули $T_{n_2'}(x_2)$ и $T_{n_3'}(x_3)$ на этом отрезке [19]:

$$x_{1,k_1}=\cos\left(\frac{\pi(n_1-k_1)}{n_1}\right),\quad k_1=\overline{0,n_1},\tag{2.4}$$

$$x_{i,k_i}=\cos\left(\frac{\pi(2n_i-2k_i+1)}{2(n_i+1)}\right),\quad k_i=\overline{0,n_i},\quad i=2,3.\tag{2.5}$$

Для нахождения значений полиномов Чебышёва в точках (2.4) и (2.5) воспользуемся геометрическим определением $T_{j_i}(x_i)=\cos(j_i\arccos x_i)$, где $x_i\in[-1,1]$ [19]. Тогда

$$T_{j_1}(x_{1,k_1})=\cos\left(\frac{\pi j_1(n_1-k_1)}{n_1}\right),\quad j_1,k_1=\overline{0,n_1},$$

$$T_{j_i}(x_{i,k_i})=\cos\left(\frac{\pi j_i(2n_i-2k_i+1)}{2(n_i+1)}\right),\quad j_i,k_i=\overline{0,n_i},\quad i=2,3.$$

Подставляя (2.4) и (2.5) в (2.2), приходим к системе линейных $n_1'n_2'n_3'$-уравнений, в которой заменяем уравнения с $x_{1,0}$, x_{2,k_2} ($k_2=\overline{n_2'/2,n_2}$) на уравнения, вытекающие из граничного условия (2.3) для $x_2>0$:

$$\mathbf{T}_1(-1)\otimes(\mathbf{T}_2(x_{2,k_2})-(1-\alpha)\mathbf{T}_2(x_{2,n_2-k_2}))\otimes\mathbf{T}_3(x_{3,k_3})\mathbf{A}=0,\quad k_3=\overline{0,n_3},$$

с x_{1,n_1}, x_{2,k_2} ($k_2=\overline{0,n_2'/2-1}$) на уравнения (2.3) для $x_2<0$:

$$\mathbf{T}_1(1)\otimes(\mathbf{T}_2(x_{2,k_2})-(1-\alpha)\mathbf{T}_2(x_{2,n_2-k_2}))\otimes\mathbf{T}_3(x_{3,k_3})\mathbf{A}=0,\quad k_3=\overline{0,n_3}.$$

Здесь и ниже считаем, что n_2 – нечетное число. Для минимизации влияние ошибок округления при вычислении значений искомой функции, удовлетворяющей решению линеаризованного кинетического уравнения и граничного условия в точках коллокации в качестве неизвестных в системе линейных $n_1'n_2'n_3'$-уравнений будем использовать значения функции $Z(\rho,\zeta,C_\perp)$ в точках (2.4) и (2.5). Пусть

$$\mathbf{Z}=\left(Z_{000},Z_{001}\ldots Z_{n_1n_2n_3-1}Z_{n_1n_2n_3}\right)^{\mathrm{T}},\quad Z_{k_1k_2k_3}=Z(\rho_{k_1},\zeta_{k_2},C_{\perp,k_3}),$$

получим

$$\begin{gathered}\mathbf{LZ}=\mathbf{F},\\ \mathbf{L}=\frac{2}{1-R_1}\mathbf{B}_1+\delta(\mathbf{I}-\mathbf{B}_2),\\ \mathbf{B}_1=\mathbf{E}_{\mathbf{JH}}\circ\left(\frac{2}{n_1}\mathbf{J}_2\mathbf{J}_1\otimes\mathbf{I}_{\mathbf{d}_2}+\frac{2}{n_2'}\mathbf{I}_{\mathbf{d}_1}\otimes(\mathbf{H}_2\mathbf{H}_1)\right)\otimes\mathbf{I}_{\mathbf{d}_3},\\ \mathbf{B}_2=\frac{4}{n_2'n_3'}\mathbf{E}_{\mathbf{JH}}\circ\mathbf{I}_{\mathbf{b}}\otimes\mathbf{E}_{\mathbf{G}}.\end{gathered}\tag{2.6}$$

Здесь $\mathbf{F}$ – матрица-столбец размером $n_1'n_2'n_3'\times1$, элементы которой равны $-1/2$ за исключением нулевых элементов с индексами l из промежутков $\overline{n_2'n_3'/2-1,n_2'n_3'-1}$ и $\overline{(n_1'n_2'-n_2'/2-2)n_3'-1,n_1'n_2'n_3'-1}$; $\mathbf{I}$ – матрица размера $n_1'n_2'n_3'\times n_1'n_2'n_3'$, на главной диагонали которой стоят единицы, а остальные ненулевые ее элементы принимают следующие значения:

$$I_{(n_2-i_2)n_3'+i_3,i_2n_3'+i_3} = \alpha - 1, \quad I_{n_1n_2'n_3'+i_2n_3'+i_3,n_1n_2'n_3'+(n_2-i_2)n_3'+i_3} = \alpha - 1,$$

где $i_2 = \overline{0, n_2'/2 - 1}$ и $i_3 = \overline{0, n_3}$; $\mathbf{E_{JH}}$ – матрица размером $n_1'n_2' \times n_1'n_2'$ с нулевыми строками l ($l = \overline{n_2'/2, n_2}$ или $l = \overline{n_1n_2' + 1, n_1n_2' + (n_2 - 1)/2}$), а все остальные строки содержат единицы; через $\circ$ обозначено произведение Адамара двух матриц [21]; матрицы $\mathbf{J}_2$ $\mathbf{H}_2$ имеют размеры $n_k' \times n_k'$ ($k = 1, 2$) соответственно и элементы

$$J_{2,i_1,j_1} = \frac{dT_{1,j_1}(x_{1,i_1})}{dx_1} \quad (i_1, j_1 = \overline{0, n_1}),$$

$$H_{2,i_2,j_2} = \frac{dT_{2,j_2}(x_{2,i_2})}{dx_2}(1 - x_{2,i_2}^2) \quad (i_2, j_2 = \overline{0, n_2});$$

матрицы $\mathbf{J}_1$, $\mathbf{H}_1$ и $\mathbf{G}_1$ имеют соответственно размеры $n_k' \times n_k'$ ($k = \overline{1,3}$) и содержат элементы $J_{1,i_1,j_1} = \gamma_1\gamma_2 T_{i_1}(x_{1,j_1})$, $H_{1,i_2,j_2} = \gamma_1 T_{i_2}(x_{2,j_2})$ и $G_{1,i_3,j_3} = \gamma_1 T_{i_3}(x_{3,j_3})$ ($i_k, j_k = \overline{0, n_k}$), где коэффициент γ_1 равен 1 для всех элементов матриц $\mathbf{J}_1$, $\mathbf{H}_1$ и $\mathbf{G}_1$ за исключением элементов первых их строк, для которых $\gamma_1 = 1/2$, а коэффициент γ_2 равен 1/4 для первого и последнего элемента последней строки матрицы $\mathbf{J}_1$, для всех других элементов этой строки, как и элементов первого и последнего столбца γ_2 равен 1/2, для остальных элементов $\gamma_2 = 1$; $\mathbf{I_b}$ – блочно-диагональная матрица ($n_1'n_2' \times n_1'n_2'$), состоящая из одинаковых блоков размера $n_2' \times n_2'$ с единичными элементами; $\mathbf{I_{d_k}}$ – диагональная $n_k' \times n_k'$-матрица ($k = \overline{1,3}$):

$$I_{d_1,i_1,i_1} = \frac{1}{x_{1,i_1} + (1 + R_1)/(1 - R_1)}, \quad I_{d_2,i_2,i_2} = x_{2,i_2}, \quad I_{d_3,i_3,i_3} = \frac{1 + x_{3,i_3}}{1 - x_{3,i_3}}, \quad i_k = \overline{0, n_k};$$

матрица $\mathbf{E_G}$ размера $n_3' \times n_3'$ имеет равные строки $\mathbf{P}_3\mathbf{G}_1$.

Решение уравнения (2.6) находим LU-методом. На основе полученных элементов матрицы $\mathbf{Z}$ восстанавливаем $U_z(\rho)$:

$$U_z(\rho) = \varrho \mathbf{T}_1\left(\frac{2\rho - 1 - R_1}{1 - R_1}\right)\mathbf{J}_1 \otimes \mathbf{E_{HG,r}}\mathbf{Z}, \quad \varrho = \frac{8}{n_1n_2'n_3'}, \tag{2.7}$$

где $\mathbf{E_{HG,r}}$ – блочная $1 \times n_2'n_3'$-матрица, состоящая из одинаковых блоков $\mathbf{E_{G,r}}$ размера $1 \times n_3'$, представляющих собой первую строку матрицы $\mathbf{E_G}$.

Подставляя (2.7) в (1.6), получаем

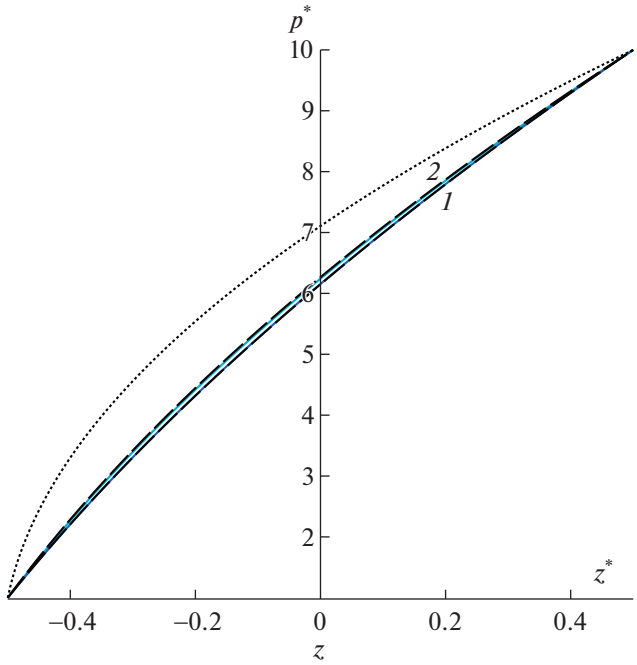
$$J_{M,p} = \frac{\varrho}{1 + R_1}\mathbf{P}_1\mathbf{J}_1 \otimes \mathbf{E_{HG,r}}\mathbf{Z}, \quad \mathbf{P}_1 = \int_{-1}^{1} \mathbf{T}_1(x_1)\big(x_1(1 - R_1) + 1 + R_1\big)\,dx_1. \tag{2.8}$$

Неизвестный параметр J_M^* в дифференциальном уравнении (1.2) с граничными условиями $p^*(-1/2) = 1$ и $p^*(1/2) = p_2^*$ находим, интегрируя левую и правую части уравнения (1.2). В уравнении (1.2) зависимость $J_{M,p}$ от $p^*(z^*)$ проявляется неявно через δ. Учитывая независимость J_M^* от z^* и равенство (1.5), из которого вытекает $\delta(z^*) = \delta_1 p^*(z^*)$, имеем

$$J_M^* = \int_1^{p_2^*} J_{M,p}\,dp^* = \frac{1}{\delta_1}\int_{\delta_1}^{\delta_2} J_{M,p}\,d\delta. \tag{2.9}$$

Используя полученное значение параметра J_M^*, распределение давления $p^*(z^*)$ по длине канала восстанавливаем путем решения уравнения (1.2) с граничным условием $p^*(-1/2) = 1$ методом Рунге–Кутты.

Распределение давления $p^*(z^*)$ в режиме течения со скольжением находим как и в случае гидродинамического режима. Ввиду громозкости полученного выражения для $p^*(z^*)$ явный вид его приводить не будем. Результаты вычислений $p^*(z^*)$ для $R_1 = 0.1$ и $p_2^* = 10$ представлены графически штриховыми линиями *1* и *2* на фиг. 1 и соответствуют значениям коэффициента аккомодации тангенциального импульса молекул газа $\alpha = 0.85$ и $\alpha = 1$ из промежутка изменения α для реальных газов [25]. Пунктирная кривая на фиг. 1 соответствует гидродинамическому пределу (3.8). Распределения давления $p^*(z^*)$, найденные с использованием кинетического подхода (1.2) и (2.6)–(2.9) при тех же.



Фиг. 1. Распределение $p^*(z^*)$ при $\delta_1 = 1$ и $p_2^* = 10$ для $R_1 = 0.1$.

3. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

В свободномолекулярном режиме течения ($\delta = 0$) уравнение (1.8) с граничными условиями (1.9) и (1.10) переходит в уравнение Больцмана для бесстолкновительного газа и может быть получено аналитически [22]. В этом случае вид функции $Z(\rho, \zeta, C_\perp)$ определяется исходя из следующих условий:

для молекул, которые отразились от внешнего цилиндра и летят к нему

$$Z(\rho,\zeta,C_\perp) = -\frac{\alpha\rho\zeta + (2-\alpha)\sqrt{1-\rho^2(1-\zeta^2)}}{2\alpha C_\perp}, \quad -\frac{\sqrt{R_1^2-\rho^2}}{\rho} \le \zeta \le \frac{\sqrt{R_1^2-\rho^2}}{\rho}, \tag{3.1}$$

для молекул, которые отразились от внешнего цилиндра и летят к внутреннему цилиндру

$$Z(\rho,\zeta,C_\perp) = -\frac{\alpha\rho\zeta + (1-(1-\alpha)R_1)\sqrt{1-\rho^2(1-\zeta^2)}}{2\alpha C_\perp}, \quad -1 \le \zeta \le -\frac{\sqrt{R_1^2-\rho^2}}{\rho}, \tag{3.2}$$

для молекул, которые отразились от внутреннего цилиндра и летят к внешнему цилиндру

$$Z(\rho,\zeta,C_\perp) = -\frac{\alpha\rho\zeta - (1-(1-\alpha)R_1^{-1})\sqrt{R_1^2-\rho^2(1-\zeta^2)}}{2\alpha C_\perp}, \quad \frac{\sqrt{R_1^2-\rho^2}}{\rho} < \zeta \le 1. \tag{3.3}$$

Подставляя (3.1)–(3.3) в (1.7), приходим к следующему выражению для $U_z(\rho)$:

$$U_z(\rho) = \frac{1}{2\sqrt{\pi}}\left(\frac{1-(1-\alpha)R_1^{-1}}{\alpha}\int\limits_{\frac{\sqrt{R_1^2-\rho^2}}{\rho}}^{1}\frac{\sqrt{R_1^2-\rho^2(1-\zeta^2)}}{\sqrt{1-\zeta^2}}d\zeta - \right.$$
$$\left. -\frac{1-(1-\alpha)R_1}{\alpha}\int\limits_{-1}^{-\frac{\sqrt{R_1^2-\rho^2}}{\rho}}\frac{\sqrt{1-\rho^2(1-\zeta^2)}}{\sqrt{1-\zeta^2}}d\zeta - \frac{2-\alpha}{\alpha}\int\limits_{-\frac{\sqrt{R_1^2-\rho^2}}{\rho}}^{\frac{\sqrt{R_1^2-\rho^2}}{\rho}}\frac{\sqrt{1-\rho^2(1-\zeta^2)}}{\sqrt{1-\zeta^2}}d\zeta\right). \tag{3.4}$$

В этом случае выражение (1.6) для приведенного потока массы $J_{M,p}$ не зависит от δ и распределение давления $p^*(z^*)$ вдоль канала является линейной функцией $p^*(z^*) = (1 + p_2^*)/2 + (p_2^* - 1)z^*$. При этом величина J_M^* равна $J_{M,p}(p_2^* - 1)$.

В гидродинамическом режиме течения ($\delta^{-1} \ll 1$) решение уравнения (1.8), соответствующее уравнению Навье–Стокса для одноатомного газа с граничным условием прилипания на цилиндрах [23]

$$U_z(\rho) = \frac{\delta(\rho^2 - 1)}{4}\left(1 - \frac{\ln\rho}{\ln R_1}\right), \tag{3.5}$$

не зависит от $C_\perp$ и ζ и имеет вид

$$Z(\rho, \zeta, C_\perp) = \frac{\delta(\rho^2 - 1)}{4}\left(1 - \frac{\ln\rho}{\ln R_1}\right). \tag{3.6}$$

В этом случае, подставляя (3.5) в (1.6), получаем

$$J_{M,p} = \frac{\delta(R_1^2(1 - \ln R_1) - 1 - \ln R_1)}{4\ln R_1}. \tag{3.7}$$

Тогда из уравнения (1.2) с граничными условиями $p^*(-1/2) = 1$ и $p^*(-1/2) = p_2^*$ находим J_M^* и $p^*(z^*)$:

$$p^*(z^*) = \sqrt{\frac{2J_M^*\delta z^*}{J_{M,p}} + \frac{p_2^{*2} + 1}{2}}, \quad J_M^*\delta = \frac{p_2^{*2} - 1}{2}J_{M,p}. \tag{3.8}$$

В режиме течения со скольжением тангенциальная массовая скорость газа пропорциональна ее нормальному градиенту вблизи стенок канала [13]. Следуя [6] и [24], запишем граничные условия скольжения в виде

$$U_z(R_i) = (-1)^{i+1}\frac{\sigma_p}{\delta}\frac{dU_z}{d\rho}(R_i), \quad i = 1, 2, \tag{3.9}$$

где σ_p – безразмерный коэффициент вязкого скольжения. Для диффузного рассеяния ($\alpha = 1$) в рамках модели БГК коэффициент σ_p равен 1.016 [24], для зеркально-диффузного отражения Максвелла коэффициент σ_p может быть получен по формуле (см. [6] и [13])

$$\sigma_p(\alpha) = \frac{2 - \alpha}{\alpha}\big(\sigma_p(1) - 0.1211(1 - \alpha)\big).$$

Решение уравнения Навье–Стокса

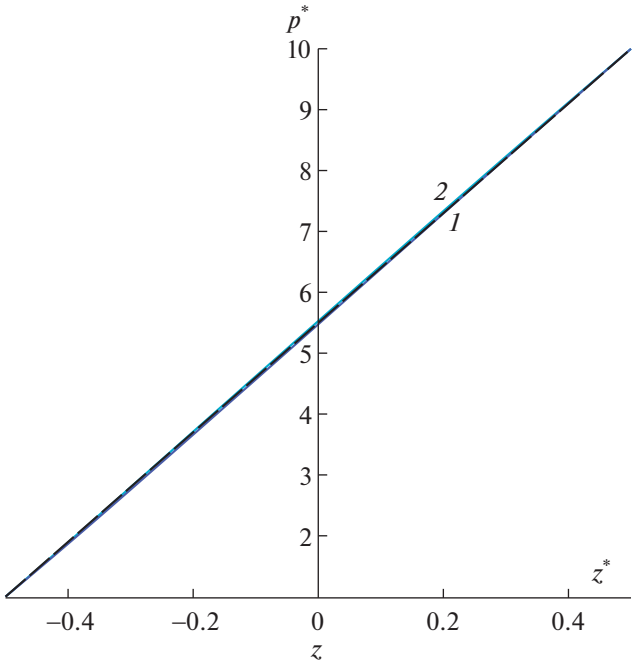
$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{dU_z(\rho)}{d\rho}\right) = \delta \tag{3.10}$$

с граничными условиями скольжения (3.9) находим аналитически

$$U_z(\rho) = \frac{1}{4}\left(\delta\rho^2 - \frac{\delta R_1(R_1 + 1)(\delta R_1 - \delta - 2\sigma)\ln\rho}{\delta R_1\ln R_1 - \sigma(1 + R_1)} - 2\sigma - \frac{\delta\big(\delta R_1\ln R_1 - \sigma(1 + R_1^3)\big) + 4R_1^2\sigma^2}{\delta R_1\ln R_1 - \sigma(1 + R_1)}\right). \tag{3.11}$$

Подставляя (3.11) в (1.6), получаем

$$\begin{aligned} J_{M,p} = \frac{1}{4(1 - R_1)(\delta R_1\ln R_1 - \sigma(1 + R_1))}\big(&((1 - R_1 + R_1^2 - R_1^3)(1 - \ln R_1) + 2(1 - R_1))R_1\delta^2 + \\ &+ (1 - 4R_1 + 4R_1^3 - R_1^4 - 4R_1(1 - R_1 + R_1^2)\ln R_1)\delta\sigma + 4(1 - R_1 - R_1^2 + R_1^3)\sigma^2\big). \end{aligned} \tag{3.12}$$



Фиг. 2. Распределение $p^*(z^*)$ при $\delta_1 = 0.1$ и $p_2^* = 10$ для $R_1 = 0.1$.

Представляя в виде ряда по малому параметру δ^{-1} выражение для массовой скорости газа (3.9) и ограничиваясь в этом разложении линейными членами этого ряда относительно δ^{-1}, имеем

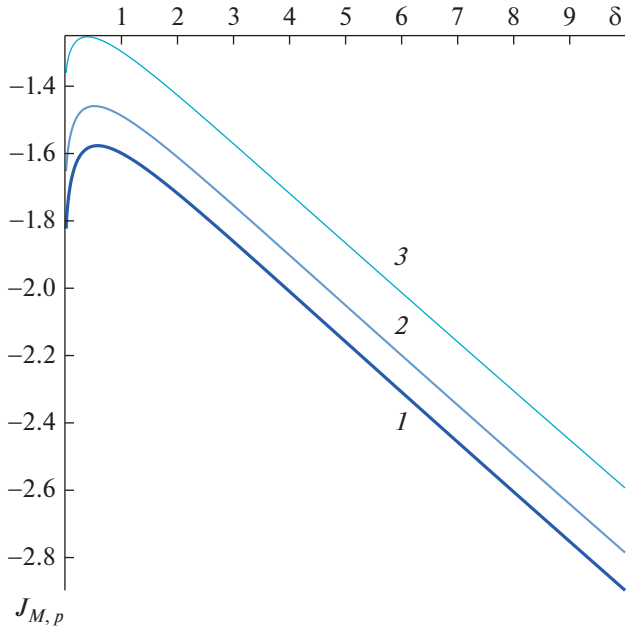
$$U_z(\rho) = U_{H,z}(\rho) + \sigma_p U_{s,z}(\rho), \tag{3.13}$$

$$U_{s,z}(\rho) = \frac{(1 + 2R_1^2 \ln R_1 - R_1^3 + 2R_1 \ln R_1 - R_1^2 + R_1)\ln\rho}{4R_1 \ln^2 R_1} + \frac{1}{4}\frac{R_1^2 - 2\ln R_1 - 1}{\ln R_1}. \tag{3.14}$$

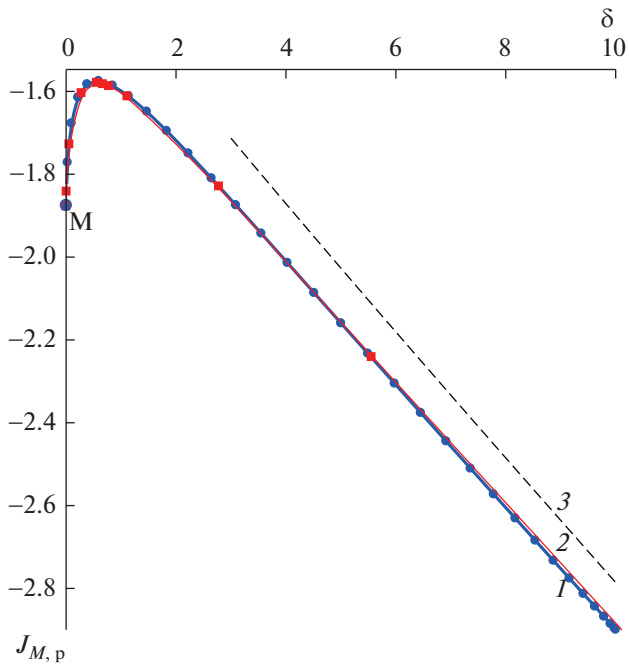
Здесь $U_{H,z}(\rho)$ – гидродинамическое решение с граничными условиями прилипания (3.5). Видим, что учет скольжения вносит поправки в (3.5) порядка, не превышающего δ^0, что соответствует поиску решения краевой задачи (3.9) и (3.10) в виде (3.13). Подставляя (3.13) в (1.6), получаем

$$\begin{gathered} J_{M,p} = J_{M,H} + \sigma_p J_{M,s} = \frac{\delta(R_1^2(1 - \ln R_1) - 1 - \ln R_1)}{4\ln R_1} + \\ + \sigma_p \frac{4R_1^3 \ln^2 R_1 - 4R_1^2 \ln^2 R_1 - 4R_1^3 \ln R_1 + R_1^4 + 4R_1 \ln^2 R_1 + 4R_1 \ln R_1 - 2R_1^2 + 1}{4R_1(R_1 - 1)\ln^2 R_1}. \end{gathered} \tag{3.15}$$

Значения α для случая $\delta_1 = 1$ показаны сплошными кривыми, которые визуально совпадают со штриховыми кривыми *1* и *2*. Вычисленные при этом значения параметра J_M^* равны соответственно –20.1147 и –17.4534 и приближаются к значениям $J_M^* = -18.8847$ и $J_M^* = -16.0763$, полученным аналитически с использованием уравнений гидродинамики со скольжением. Гидродинамический предел J_M^* согласно (3.8) равен –7.1781. Результаты вычислений $p^*(z^*)$ для $R_1 = 0.1$, $\delta_1 = 0.1$ и $p_2^* = 10$ представлены на фиг. 2. В этом случае распределения $p^*(z^*)$ при $\alpha = 0.85$ (кривая *1*) и $\alpha = 1$ (кривая *2*) приближаются к отрезку прямой $p^*(z^*) = (1 + p_2^*)/2 + (p_2^* - 1)z^*$ ($z^* \in [-1/2;1/2]$), который определяет распределение давления в свободномолекулярном режиме течения газа в канале. Вычисленные при этом значения параметра J_M^* равны соответственно –14.3225 и –11.3970 и приближаются к значениям $J_M^* = -16.8732$ и $J_M^* = -12.5010$, полученным по формуле $J_M^* = (p_2^* - 1)J_{M,p}$, справедливой для



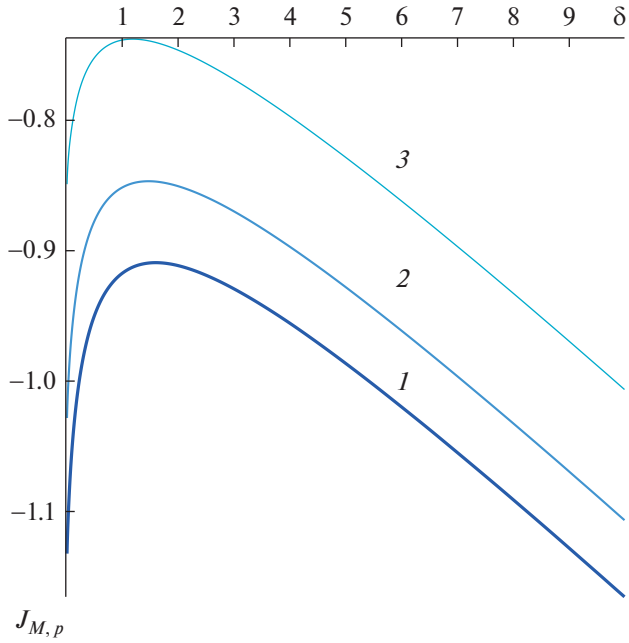
Фиг. 3. Зависимость приведенного массового потока газа $J_{M,p}$ от δ для $R_1 = 0.1$: *1*, *2* и *3* – $\alpha = 0.85, 0.9, 1$.



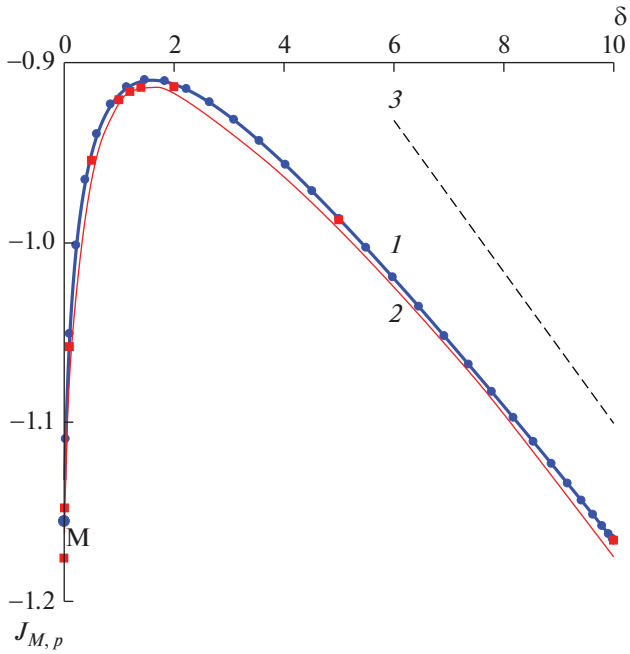
Фиг. 4. Результаты вычисления $J_{M,p}$ в сравнении с [6] для $R_1 = 0.1$ и $\alpha = 0.85$.

свободномолекулярного режима течения. Для $R_1 = 0.5$ в поведении функции $p^*(z^*)$ качественных изменений не выявлено.

Результаты вычислений приведенного массового потока $J_{M,p}$ по формуле (2.8) в зависимости от δ представлены на фиг. 3 и 4 для $R_1 = 0.1$ и фиг. 5 и 6 для $R_1 = 0.5$. Аппроксимация выполнена B-спайнами третьей степени. На фиг. 3 и 5 кривыми *1–3* показаны изменения величины $J_{M,p}$ при

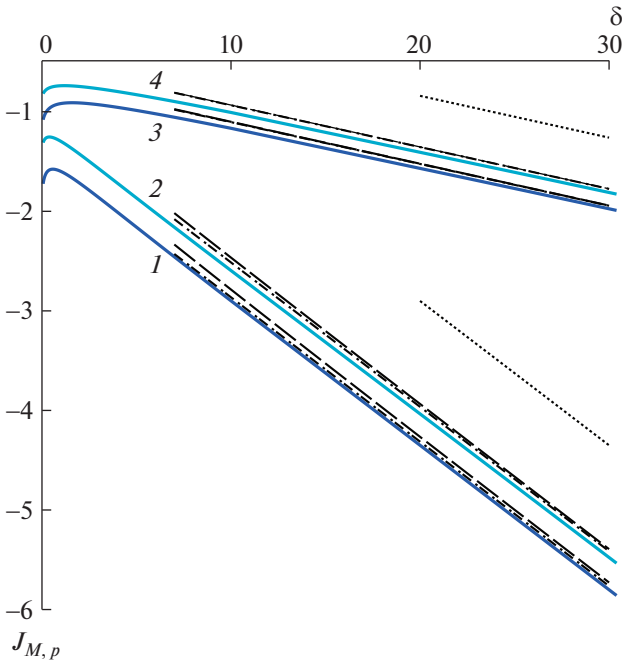


Фиг. 5. Зависимость приведенного массового потока газа $J_{M,p}$ от δ для $R_1 = 0.5$: *1*, *2* и *3* – $\alpha = 0.85, 0.9, 1$.



Фиг. 6. Результаты вычисления $J_{M,p}$ в сравнении с [6] для $R_1 = 0.5$ и $\alpha = 0.85$.

значениях коэффициента аккомодации тангенциального импульса молекул газа $\alpha = 0.85$, 0.9 и 1 соответственно. Точками M на фиг. 4 и 6 показаны свободномолекулярныые пределы, найденные согласно (1.6) с использованием (3.4). Из анализа фиг. 3 и 5 следует, что величина приведенного массового потока увеличивается с уменьшением значений α для $0 \le \delta \le 10$, при этом положение точки максимума $J_{M,p}$ определяется в большей степени отношением $R_1 = R_1'/R_2'$. Кривыми *2* на фиг. 4 и 6 показаны результаты вычислений $J_{M,p}$ из [6] с учетом умножения параметра разрежения на $1/2 + R_1/(2(1 - R_1))$ и величины приведенного потока – на $2(R_2 - R_1)$, что обусловлено



Фиг. 7. График функции $J_{M,p}(\delta)$ в зависимости от R_1 и α.

введением в [6] в качестве характерного линейного размера гидродинамического диаметра канала $2(R_2' - R_1')$; штриховыми линиями (кривые *3*) нанесены результаты вычислений $J_{M,p}$ по формуле (3.12). Как видно из фиг. 4 и 6 , наблюдается хорошее согласие полученных результатов в настоящей работе с [6]. Фигура 7 иллюстрирует поведение кривых потоков массы газа в режиме течения со скольжением ($\delta > 10$). Кривые *1–4* соответствуют: *1* – $R_1 = 0.1$, $\alpha = 0.85$, *2* – $R_1 = 0.1$, $\alpha = 1$ и *3* – $R_1 = 0.5$, $\alpha = 0.85$, *4* – $R_1 = 0.5$, $\alpha = 1$. Штриховыми и штрихпунктирными линиями на фиг. 5 нанесены результаты вычислений $J_{M,p}$ по формулам (3.12) и (3.15) при тех же значениях R_1 и α, что и кривые *1–4*. Пунктирные кривые соответствуют гидродинамическому пределу (3.7). На фиг. 7 видно, что при $\delta \geq 20$ величина $J_{M,p}$, полученная по формуле (2.8), приближается к линейной функции с углом наклона, который увеличивается с уменьшением отношения $R_1 = R_1'/R_2'$.

ЗАКЛЮЧЕНИЕ

Методом коллокации Чебышёва решена задача о вычислении потока массы и давления газа с использованием линеаризованного БГК уравнения при изотермическом течении разреженного газа в длинном концентрическом кольцевом канале. Реализация метода коллокации выполнена с использованием свойств сумм многочленов Чебышёва и матричных произведений Адамара и Кронекера с целью минимизизации влияния ошибок округления при вычислении значений искомой функции, удовлетворяющей решению линеаризованного кинетического уравнения и граничного условия в точках коллокации. Вычислены значения массового потока газа для широкого диапазона изменения значений параметра разрежения. Показано, что применение многочленов Чебышёва при исследовании течений газа дает возможность эффективно находить интегральные характеристики этих течений с последующим их анализом. Предложенный метод построения численного решения линеаризованного кинетического уравнения может быть применен для исследования свойств процессов переноса в длинных осесимметричных каналах при наличии малого градиента давления и режимах изотермического течения газа от свободномолекулярного до гидродинамического.

СПИСОК ЛИТЕРАТУРЫ

1. *Sharipov F.* Rarefied gas dynamics: Fundamentals for research and practice. Berlin: Wiley-VCH, 2016.

2. *Рудяк В.Я., Лежнев Е.В.* Вязкость газов в наноканалах // Письма в ЖТФ. 2020. Т. 46. № 20. С. 51.

3. *Ambrus V., Sharipov F., Sofonea V.* Comparison of the Shakhov and ellipsoidal models for the Boltzmann equation and DSMC for ab initio-based particle interactions // Computers and Fluids. 2020. V. 211. 104637.

4. *Lin C., Luo K.H., Fei L., Succi S.* A multi-component discrete Boltzmann model for nonequilibrium reactive flows // Scientific reports. 2017. V. 7. 14580.

5. *Graur I., Ho M.T.* Rarefied gas flow through a long rectangular channel of variable cross section // Vacuum. 2014. V. 101. P. 328.

6. *Breyiannis G., Varoutis S., Valougeorgis D.* Rarefied gas flow in concentric annular tube: Estimation of the Poiseuille number and the exact hydraulic diameter // European Journal of Mechanics B/Fluids. 2008. V. 27. P. 609.

7. *Taheri P., Struchtrup H.* Poiseuille flow of moderately rarefied gases in annular channels // International Journal of Heat and Mass Transfer. 2012. V. 55. P. 1291.

8. *Titarev V.* Implicit high-order method for calculating rarefied gas flow in a planar microchannel // Journal of Computational Physics. 2012. V. 231. P. 109.

9. *Гермидер О.В., Попов В.Н.* Течение разреженного газа между двумя коаксиальными цилиндрами под действием градиента температуры в рамках зеркально-диффузной модели отражения // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 8. С. 1401.

10. *Гермидер О.В., Попов В.Н.* Неизотермическое течение разреженного газа в длинном цилиндрическом канале при произвольных перепадах давления и температуры // Изв. РАН. МЖГ. 2020. № 3. С. 125.

11. *Valougeorgis D., Vasileiadis N., Titarev V.* Validity range of linear kinetic modeling in rarefied pressure driven single gas flows through circular capillaries // European Journal of Mechanics - B/Fluids. 2017. V. 64. P. 2.

12. *Bhatnagar P.L., Gross E.P., Krook M.* A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems // Phys. Rev. 1954. V. 94. P. 511.

13. *Шарипов Ф.М., Селезнев В.Д.* Движение разреженных газов в каналах и микроканалах. Екатеринбург: УрО РАН, 2008.

14. *Буляница А.Л.* Математическое моделирование в микрофлюидике: основные положения // Научное приборостроение. 2005. Т. 15. № 2. С. 51.

15. *Гермидер О.В., Попов В.Н.* Применение полиномов Чебышёва для вычисления потоков разреженного газа в каналах с цилиндрической геометрией // Сибирские электронные математические известия. 2019. Т. 16. С. 1947.

16. *Гермидер О.В., Попов В.Н.* Метод коллокации и его применение для решения линеаризованного уравнения Холвея // Матем. моделирование. 2020. Т. 32. № 9. С. 3.

17. *Шильков А.В.* Разложение оператора рассеяния уравнения переноса частиц в ряд по сферическим тензорам // Препринты ИПМ им. М.В. Келдыша. 2018. Т. 249.

18. *Baseri A., Abbasbandy S., Babolian E.* A collocation method for fractional diffusion equation in a long time with Chebyshev functions // Applied Mathematics and Computation. 2018. V. 322. P. 55–65.

19. *Mason J., Handscomb D.* Chebyshev polynomials. Florida: CRC Press, 2003.

20. *Шахов Е.М.* Течение разреженного газа между коаксиальными цилиндрами под действием градиента давления // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 7. С. 1107.

21. *Liu S., Trenkler G.* Hadamard, Khatri-Rao, Kronecker and Other Matrix Products // International Journal of Information and Systems Sciences. 2008. V. 4. № 1. P. 160.

22. *Germider O.V., Popov V.N., Yushkanov A.A.* Calculation of gas flow between two coaxial cylinders in a free-molecular regime under specular-diffuse boundary conditions // Journal of Engineering Physics and Thermophysics. 2019. V. 92. № 1. P. 280.

23. *Landau L.D., Lifshitz E.M.* Fluid Mechanics. New York: Pergamon, 1989.

24. *Graur I., Sharipov F.* Gas flow through an elliptical tube over the whole range of the gas rarefaction // European Journal of Mechanics B/ Fluids. 2008. V. 27. P. 335.

25. *Nacer M.H., Graur I., Perrier P., Meolans G., Wuest M.* Gas flow through microtubes with different internal surface coatings // Journal of Vacuum Science and Technology A. 2014. V. 32. 21601.

EDN: MELYJV
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1563

МАТЕМАТИЧЕСКАЯ ФИЗИКА

A FICTITIOUS TIME LIE-GROUP INTEGRATOR FOR THE BRINKMAN–FORCHHEIMER MOMENTUM EQUATION MODELING FLOW OF FULLY DEVELOPED FORCED CONVECTION[1)]

© 2022 г. S. Abbasbandy[1,*], M. Hajiketabi[1,]**

[1] *Department of Applied Mathematics, Faculty of Science, Imam Khomeini, International University, Ghazvin, 34149-16818, Iran*

**e-mail: abbasbandy@sci.ikiu.ac.ir*

***e-mail: mohammadhajiketabi@gmail.com*

Поступила в редакцию 02.02.2022 г.
Переработанный вариант 02.02.2022 г.
Принята к публикации 11.05.2022 г.

Численная схема решения уравнения Бринкмана–Форхгеймера для моделирования стационарного течения в насыщенном пористом канале. Исследуется численная схема решения уравнения Бринкмана–Форхгеймера для моделирования стационарного течения в насыщенном пористом канале. С помощью введения дополнительной новой переменной (фиктивного времени) и после дискретизации уравнения по двум пространственным переменным получена система обыкновенных дифференциальных уравнений по этому фиктивному времени. Результирующая система уравнений решается с помощью метода геометрического численного интегрирования. Это позволяет сохранять групповые свойства исходной системы, такие как асимптотическое поведение решения и отсутствие искусственных ложных решений. Такой метод позволил построить численное решение с малой погрешностью, что показало сравнение с известными аналитическими и численными результатами. Проведен анализ устойчивости и сходимости метода и выявлено влияние различных физических параметров на поведение решения.

Ключевые слова: уравнение импульса Бринкмана–Форхгеймера, насыщенный пористый канал, сложный метод интегрирования по времени, схема сохранения группы.

DOI: 10.31857/S0044466922090125

[1)]Полный текст статьи печатается в английской версии журнала.

EDN: PPOANY
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 9, с. 1564–1584

ИНФОРМАТИКА

УДК 531.1

РЕГУЛЯРНЫЕ АППРОКСИМАЦИИ НАИБЫСТРЕЙШЕГО ДВИЖЕНИЯ МОБИЛЬНОГО РОБОТА ПРИ ОГРАНИЧЕННЫХ ФАЗОВЫХ КООРДИНАТАХ[1)]

© 2022 г. А. Н. Дарьина[1,*], А. И. Дивеев[1,], Д. Ю. Карамзин[1,***], Ф. Л. Перейра[2,****], Е. А. Софронова[1,*****], Р. А. Чертовских[2,******]**

[1] *119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия*

[2] *Исследовательский центр систем и технологий (SYSTEC), Университет г. Порто, Порто, Португалия*

**e-mail: daryina@ccas.ru*

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

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

*****e-mail: flp@fe.up.pt*

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

*******e-mail: roman.home@gmail.com*

Поступила в редакцию 01.05.2021 г.
Переработанный вариант 05.04.2022 г.
Принята к публикации 11.05.2022 г.

Исследована задача быстродействия для мобильного управляемого робота при фазовых ограничениях. Предложен численный метод решения, основанный на принципе максимума Понтрягина. Известно, что задача управления мобильным роботом, как и любое движение, согласно унициклической модели, принадлежит классу существенно нерегулярных задач относительно фазовых ограничений. Решение такой задачи с помощью принципа максимума затруднено тем, что нет формулы для меры-множителя Лагранжа. Неясно, как выразить эту меру через другие экстремальные значения, и тем самым непонятно, как свести условия принципа максимума к соответствующей краевой задаче. Чтобы обойти указанную трудность, в работе предложен один метод регуляризации задачи, основанный на ε-возмущении. Приведены результаты численного эксперимента. Результаты эксперимента демонстрируют непрерывность меры-множителя. Библ. 21. Фиг. 5.

Ключевые слова: оптимальное управление, фазовые ограничения, принцип максимума Понтрягина.

DOI: 10.31857/S0044466922090095

1. ВВЕДЕНИЕ

В работе изучается движение мобильного робота при ограничениях, накладываемых на фазовые координаты. В качестве критерия оптимальности рассматривается быстродействие, т.е. время, затрачиваемое на движение из заданной точки A в точку B. Задача о наибыстрейшем движении мобильного робота при фазовых ограничениях, или задача о наибыстрейшем обходе роботом препятствия, формулируется следующим образом:

$$\begin{gathered} J(p,u,T) := T \to \min, \quad \dot{x}_1 = p\cos\alpha, \quad \dot{x}_2 = p\sin\alpha, \quad \dot{\alpha} = u, \\ x(0) = A, \quad x(T) = B, \quad \alpha(0) = \alpha_0, \quad p \in [-1,1], \quad u \in [-1,1], \quad x_1^2 + x_2^2 \geq 1. \end{gathered} \tag{1}$$

Здесь $x = (x_1, x_2) \in \mathbb{R}^2$ – основная фазовая переменная, $p, u \in \mathbb{R}$ – скалярные параметры управления, $A = (a_1, a_2)$, $B = (b_1, b_2)$ – заданные точки на плоскости, лежащие вне единичного круга. Задача рассматривается на отрезке времени $[0, T]$. Конечный момент времени T не фиксирован, и,

[1)]Работа первого и третьего авторов поддержана Минобрнауки РФ, проект 075-15-2020-799. Четвертый и шестой авторы выражают признательность Фонду науки и технологии при Министерстве образования и науки Португальской республики (FCT/MCTES).

более того, необходимо найти наименьшее возможное такое время T, за которое можно, начиная движение из точки A, попасть в точку B с помощью каких-либо допустимых управлений $p(t)$, $u(t)$. При этом робот не должен заезжать на единичный круг на плоскости, который считается препятствием. Переменная α – это дополнительная фазовая переменная, отвечающая углу поворота мобильного робота. Скорость поворота робота управляется функцией $u(t)$, а продольная скорость – функцией $p(t)$. Обе функции считаются измеримыми. Робот оснащен обратным ходом, что соответствует отрицательным значениям p. Начальное значение угла поворота α_0 задано.

В задаче (1) реализуется так называемая простейшая унициклическая модель движения, которая дополняется возможностью разворота робота на месте, т.е. движением при $p = 0$. Она отвечает, например, упрощенной модели трехколесного робота с передним приводом. Отметим, что продольная и угловая скорости в этой модели являются независимыми, что не является характерным для некоторых других мобильных роботов (например, для гусеничного робота). Тем не менее предлагаемый в этой работе подход применим и к другим моделям мобильных роботов. Общий метод исследования обсуждается в разд. 4.

Важно отметить существование решения задачи (1). Действительно, очевидно, что в этой задаче всегда найдется допустимый процесс, и, значит, существование решения в классе измеримых управлений p, u следует из известной теоремы Филиппова (см. [1]) (поскольку правая часть динамической системы линейна по управлению, а множество значений управления – выпуклый компакт). Обозначим решение этой задачи (которых может быть много) через $(x^*, \alpha^*, p^*, u^*, T^*)$.

Итак, задача состоит в оптимальном (по быстродействию) обходе роботом препятствия, которое задано в виде единичного круга на плоскости. В нашей работе предлагается некоторый численный метод нахождения решения $(x^*, \alpha^*, p^*, u^*, T^*)$, основанный на принципе максимума Понтрягина (см. [2]).

Отметим, что решение задачи (1) с помощью принципа максимума осложнено рядом трудностей. Во-первых, любая допустимая траектория, соприкасающаяся с границей единичного круга, не будет регулярной по Р.В. Гамкрелидзе (см. [3], [2, гл. 6]). Действительно, в любой точке границы имеет место

$$x_1 \cos\alpha + x_2 \sin\alpha = 0,$$

а значит, градиент функции $\Gamma(x, p, u) := -2p(x_1 \cos\alpha + x_2 \sin\alpha)$ – скалярного произведения правой части динамической системы на градиент фазоограничения – по (p, u) равен нулю. Более того, по этой же причине в задаче (1) заведомо не будут выполняться даже и более слабые, чем условия регулярности, известные условия управляемости относительно фазовых ограничений (см. [4]). (Такая ситуация весьма распространена в технических системах, где обычно количество фазовых переменных превышает количество переменных управления.) Поэтому мера-множитель Лагранжа, отвечающая за фазовое ограничение, может иметь сингулярную составляющую. В связи с этим, т.е. в связи с отсутствием регулярности и наличием сингулярной составляющей меры, не представляется возможным свести условия принципа максимума к соответствующей краевой задаче, решение которой привело бы к решению задачи управления. Здесь, в частности, возникает проблема нахождения точек выхода на границу и схода с границы фазового ограничения.

Кроме того, применение принципа максимума затруднено наличием особого режима по угловой скорости u. Робот не всегда будет следовать с максимальной по модулю угловой скоростью, т.е. при $|u| = 1$. Однако на участках, где $|u| < 1$, условие максимума, очевидно, не информативно. Возникает проблема нахождения точек выхода на особый режим управления и схода с него.

Оба явления, т.е. отсутствие регулярности относительно фазового ограничения и особый режим управления по угловой скорости, имеют совершенно разную природу. Тем не менее одним из возможных решений и в первом, и во втором случаях является подходящая регуляризация исходной задачи. В работе предлагается с помощью малого параметра $\varepsilon > 0$ определенным образом возмутить исходную задачу (1), так чтобы возмущенная задача стала регулярной в упомянутом выше смысле. Для этого будут использованы дополнительные переменные управления.

2. ОБЩАЯ ТЕОРИЯ

В этом разделе приведем некоторые факты из теории принципа максимума для задач быстродействия с фазовыми ограничениями. Рассмотрим такую задачу в общем виде:

$$\begin{gathered} T \to \min, \quad \dot{x} = f(x,u), \quad x(0) = A, \quad x(T) = B, \\ u(t) \in U \quad \text{для п.в.} \ \ t \in [0,T], \\ g(x(t)) \le 0 \quad \forall t \in [0,T]. \end{gathered} \tag{2}$$

Здесь, $t \in [0,T]$ – время, x – фазовая переменная в $\mathbb{R}^n$, u – переменная управления в $\mathbb{R}^m$. Управляющая функция $u(\cdot)$ измерима. Множество U имеет вид

$$U := \{u \in \mathbb{R}^m : \varphi(u) \le 0\},$$

где φ – заданная вектор-функция. Точки A и B определяют начальную и конечную позиции (краевые условия). Допустимая траектория $x(\cdot)$ является абсолютно непрерывной функцией, удовлетворяет дифференциальному ограничению $\dot{x}(t) = f(x(t),u(t))$ для п.в. $t \in [0,T]$, начальному и конечному краевому условию, а также неравенству $g(x(t)) \le 0$ $\forall t \in [0,T]$, которое задает фазовые ограничения. Число T определяет конечный момент времени движения, оно не фиксировано, но подлежит минимизации.

Отображения

$$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \quad \varphi : \mathbb{R}^m \to \mathbb{R}^r, \quad \text{и} \quad g : \mathbb{R}^n \to \mathbb{R}$$

предполагаются достаточно гладкими (точнее, $f, \varphi \in C^1$, $g \in C^2$).

Пусть $u(\cdot)$ – управление, а $x(\cdot)$ – отвечающая ему траектория на отрезке времени $[0,T]$. Тройка (x,u,T) называется допустимым процессом, если выполнены концевые и фазовые ограничения. Допустимый процесс (x^*,u^*,T^*) называется оптимальным, если конечное время T в (2) принимает наименьшее возможное значение на множестве всех допустимых процессов (так называемый глобальный минимум). Ниже предположим, что в задаче (2) существует по крайней мере один оптимальный процесс (x^*,u^*,T^*).

Для формулировки принципа максимума Понтрягина потребуются некоторые определения и обозначения. Назовем точку $u \in U$ регулярной, если градиенты $\nabla\varphi^i(u)$, $i \in I(u)$, линейно независимы. Здесь $I(u) := \{i : \varphi^i(u) = 0\}$ – так называемое множество активных индексов (т.е. множество индексов активных ограничений); символ ∇ означает классический градиент скалярной функции.

Пусть задана измеримая функция $u(\cdot) : \mathbb{R} \to \mathbb{R}$.

Определение 1. Замыканием по мере функции $u(\cdot)$ в точке τ называется множество векторов $v \in \mathbb{R}^m$ таких, что

$$\ell(\{t \in [\tau-\varepsilon, \tau+\varepsilon] : |u(t) - v| \le \varepsilon\}) > 0 \quad \forall \varepsilon > 0,$$

где ℓ – мера Лебега на $\mathbb{R}$.

Многозначное отображение, ставящее в соответствие точке на прямой замыкание по мере функции в этой точке, обозначим как $\operatorname{Clos} u$. В контексте изучения задач оптимального управления понятие замыкания по мере измеримой функции было предложено в [5]. В [6] элементы множества $\operatorname{Clos} u(\tau)$ называются существенными значениями функции u в точке τ. Некоторые свойства замыкания по мере функции описаны также в [7].

Рассмотрим функцию

$$\Gamma(x,u) := \langle \nabla g(x), f(x,u) \rangle.$$

Сформулируем условие регулярности.

Определение 2. Процесс (x^*,u^*,T^*) называется регулярным относительно фазовых ограничений, если для всех $t \in [0,T^*]$, и всех $u \in \operatorname{Clos} u^*(t)$ таких, что $g(x^*(t)) = \Gamma(x^*(t),u) = 0$, точка u регулярна, и выполнено условие

$$\Gamma'_u(x^*(t),u) \notin \operatorname{lin}\{\nabla\varphi^i(u), i \in I(u)\}.$$

Здесь Γ'_u означает классическую частную производную функции Γ по u.

Рассмотрим расширенную функцию Гамильтона–Понтрягина

$$\bar{H}(x,u,\psi,\mu) := \langle \psi, f(x,u)\rangle - \mu\Gamma(x,u),$$

где ψ, μ – переменные из $\mathbb{R}^n$ и $\mathbb{R}$ соответственно.

Теорема 1. *Предположим, что регулярный относительно фазовых ограничений процесс* (x^*,u^*,T^*) *является оптимальным в задаче* (2). *Тогда существуют множители Лагранжа: число* $\lambda \in [0,1]$, *абсолютно непрерывная векторная функция* $\psi \in \mathbb{W}_{1,\infty}([0,T^*];\mathbb{R}^n)$, *и непрерывная скалярная функция* $\mu \in \mathbb{C}([0,T^*];\mathbb{R})$ *такие, что*

$$\lambda + \left|\psi(t) - \mu(t)\nabla g(x^*(t))\right| > 0 \quad \forall t \in [0,T^*];$$

$$\dot{\psi}(t) = -\bar{H}'_x(x^*(t),u^*(t),\psi(t),\mu(t)) \quad \textit{для п.в.} \ \ t \in [0,T^*];$$

$$\max_{u\in U} \bar{H}(x^*(t),u,\psi(t),\mu(t)) = \bar{H}(x^*(t),u^*(t),\psi(t),\mu(t)) \quad \textit{для п.в.} \ \ t \in [0,T^*],$$

$$\max_{u\in U} \bar{H}(x^*(T^*),u,\psi(T^*),\mu(T^*)) = \lambda.$$

Кроме того, функция $\mu(\cdot)$ *не убывает, и*

$$\int_0^{T^*} g(x^*(t))d\mu(t) = 0.$$

Теорема 1 была получена в [7] при некоторых дополнительных предположениях гладкости, а в [8] и [9] была доказана в полной общности. Краткая версия доказательства приведена в разд. 6. Отметим также, что теорема 1 является уточнением соответствующего результата из [2, гл. 6]. Уточнение касается непрерывности меры-множителя $\mu(\cdot)$.

Приведем несколько замечаний, которые будут использованы далее.

Замечание 1. Из условий принципа максимума вытекает известный закон сохранения вдоль экстремали (см. [2], [10])

$$\max_{u\in U} \bar{H}(x^*(t),u,\psi(t),\mu(t)) = \lambda \quad \forall t \in [0,T^*].$$

Замечание 2. Из условий принципа максимума следует, что $\mu(\cdot)$ постоянна на любом отрезке времени $[a,b]$, на котором оптимальная траектория лежит во внутренности фазовых ограничений, т.е. когда $g(x^*(t)) < 0 \ \forall t \in [a,b]$.

Замечание 3. Ввиду правила множителей Лагранжа, условие максимума влечет

$$\bar{H}'_u(x^*(t),u^*(t),\psi(t),\mu(t)) \in N_U(u^*(t)) \quad \text{для п.в.} \ \ t \in [0,T^*],$$

где N_U означает предельный нормальный конус ко множеству (см. [11]). Предполагая, что все точки множества U регулярны, можно утверждать существование дополнительного множителя Лагранжа – векторной функции $\xi : [0,T^*] \to \mathbb{R}^r$ с неотрицательными компонентами такой, что $\langle \xi(t), \varphi(u^*(t))\rangle = 0$,

$$\bar{H}'_u(x^*(t),u^*(t),\psi(t),\mu(t)) = \xi(t)\varphi'(u^*(t)) \quad \text{для п.в.} \ \ t \in [0,T^*].$$

Замечание 4. Пусть набор множителей (λ,ψ,μ) удовлетворяет принципу максимума, и $a \in \mathbb{R}$ – произвольное число. Тогда набор множителей

$$(\lambda, \psi(t) + a\nabla g(x^*(t)), \mu(t) + a)$$

снова удовлетворяет всем условиям принципа максимума (см. [2], [12]). Поэтому можно считать, что $\mu(T^*) = 0$ и тогда $\mu(t) \geq 0$, или, что эквивалентно, $\mu(0) = 0$ и тогда $\mu(t) \leq 0$.

Принцип максимума из теоремы 1 предоставляет так называемую полную систему уравнений для нахождения оптимального процесса, в которой количество уравнений формально равно количеству неизвестных. Действительно, в условиях регулярности на границе фазового ограничения замечание 3 позволяет выразить μ через другие экстремальные значения. Вне границы фазового ограничения μ постоянна, согласно замечанию 2. Поэтому, используя свойство непрерывности, меру-множитель μ можно выразить через x^*, ψ, u^*. Если одновременно с этим с помощью условия максимума получается выразить и экстремальное управление, то условия

принципа максимума сводятся к решению соответствующей краевой задачи, которая может быть решена стандартными методами. Опираясь на эти соображения, можно предложить алгоритм вычисления экстремалей. Такой алгоритм будет описан ниже в вычислительной части работы (разд. 5) и применен для нахождения приближенного решения задачи о наибыстрейшем обходе препятствия мобильным роботом (1).

Отметим, что условие регулярности, сформулированное в определении 2, нельзя назвать удобным для предварительного анализа задачи, поскольку оно оперирует с заданным оптимальным процессом, который, вообще говоря, заранее неизвестен (его нам и следует вычислить). В связи с этим представляется более практичным иметь некоторое априорное условие регулярности в терминах функций f, φ и g, которое бы гарантировало регулярность любого допустимого процесса, а значит, и заведомую применимость метода, основанного на теореме 1. Такое априорное условие регулярности несложно привести.

Условие P). *Для всех $x \in \mathbb{R}^n$, $u \in U$ таких, что $g(x) = \Gamma(x,u) = 0$, имеет место, что точка и регулярна, и*

$$\Gamma'_u(x,u) \notin \operatorname{lin}\{\nabla\varphi^i(u), i \in I(u)\}.$$

Это априорное условие легко проверить для конкретной задачи. В контексте задач быстродействия, его естественно дополнить следующим замечанием.

Замечание 5. Пусть $0 \in \operatorname{int} f(x,U)\ \forall x$. Тогда условие P) достаточно проверять на множестве

$$\{x \in \mathbb{R}^n, u \in U : g(x) = \Gamma(x,u) = 0, f(x,u) \neq 0\}.$$

Это следует из леммы 1 в разд. 6. Тогда условие P) гарантирует регулярность любого оптимального процесса задачи. Это модифицированное условие P) используется ниже при анализе движения мобильного робота.

3. РЕГУЛЯРИЗАЦИЯ ЗАДАЧИ

Любой допустимый процесс задачи (1) нерегулярен относительно фазовых ограничений, как только траектория касается границы допустимой области. Действительно, в любой точке границы области имеем $x_1\cos\alpha + x_2\sin\alpha = 0$, и, значит, градиент функции $\Gamma(x,p,u) = -2p(x_1\cos\alpha + x_2\sin\alpha)$ по (p,u) равен нулю. Поэтому теорему 1 не удается применить для численного решения поставленной задачи, как это было сделано, например, в [9]. Однако предлагается рассмотреть некоторое возмущение исходной задачи такое, что в возмущенной задаче выполняется условие P). Теорему 1 будем применять к возмущенной задаче.

Итак, зафиксируем малое число $\varepsilon > 0$ и рассмотрим ε-задачу

$$\begin{gathered} J(p,u,v,w,T) := T \to \min, \quad \dot{x}_1 = p\cos\alpha + \varepsilon w_1 + \varepsilon v_1, \quad \dot{x}_2 = p\sin\alpha + \varepsilon w_2 + \varepsilon v_2, \\ \dot{\alpha} = u, \quad x(0) = A, \quad x(T) = B, \quad \alpha(0) = \alpha_0, \\ p^2 + w_1^2 + w_2^2 \le 1, \quad u^2 + v_1^2 + v_2^2 \le 1, \quad x_1^2 + x_2^2 \ge 1. \end{gathered} \tag{3}$$

Здесь $w = (w_1, w_2) \in \mathbb{R}^2$ и $v = (v_1, v_2) \in \mathbb{R}^2$ – дополнительные переменные управления, отвечающие за регуляризацию задачи. Ясно, что задача (3) является релаксацией задачи (1), так как любой допустимый процесс задачи (1) будет допустимым и в задаче (3), если положить $v = w = 0$. Кроме того, решение в задаче (3) существует. Обозначим ее решение (какое-либо) через $\wp_\varepsilon := (x_\varepsilon, \alpha_\varepsilon, p_\varepsilon, u_\varepsilon, v_\varepsilon, w_\varepsilon, T_\varepsilon)$.

Покажем, что ε-задача регулярна. Имеет место следующее утверждение.

Предложение 1. *Любое решение $\wp_\varepsilon$ задачи* (3) *является регулярным относительно фазовых ограничений.*

Доказательство. Имеем

$$\Gamma(x,\alpha,p,u,v,w) = -2x_1(p\cos\alpha + \varepsilon w_1 + \varepsilon v_1) - 2x_2(p\sin\alpha + \varepsilon w_2 + \varepsilon v_2).$$

В силу замечания 5 условие регулярности будет заведомо выполнено для любого оптимального процесса, если на множестве

$$x,\alpha,p,u,v,w: \ x_1^2 + x_2^2 = 1, \quad \Gamma(x,\alpha,p,u,v,w) = 0, \quad f(\alpha,p,u,v,w) \neq 0$$

градиент функции Γ по (p, u, v, w) и градиенты активных ограничений на управление линейно независимы, т.е. линейно независима следующая система из трех векторов (множитель 2 сокращен):

$$a_1 := \begin{bmatrix} -x_1\cos\alpha - x_2\sin\alpha \\ 0 \\ -\varepsilon x_1 \\ -\varepsilon x_2 \\ -\varepsilon x_1 \\ -\varepsilon x_2 \end{bmatrix}, \quad a_2 := \begin{bmatrix} p \\ 0 \\ 0 \\ 0 \\ w_1 \\ w_2 \end{bmatrix}, \quad a_3 := \begin{bmatrix} 0 \\ u \\ v_1 \\ v_2 \\ 0 \\ 0 \end{bmatrix}.$$

Здесь функция f – правая часть динамической системы в (3), т.е.

$$f := \begin{bmatrix} p\cos\alpha + \varepsilon w_1 + \varepsilon v_1 \\ p\sin\alpha + \varepsilon w_2 + \varepsilon v_2 \\ u \end{bmatrix}.$$

Векторы a_1, a_2, a_3, очевидно, попарно линейно независимы при активных ограничениях на управление, так как $|x| = 1$. Поэтому, как легко видеть, если три вектора линейно зависимы, то первый вектор (с точностью до знака) является линейной комбинацией двух других: $a_1 = \alpha a_2 + \beta a_3$, причем $\alpha \neq 0$ и $\beta \neq 0$. Тем самым $u = 0$, и

$$\begin{gathered} -x_1\cos\alpha - x_2\sin\alpha = \alpha p, \\ -\varepsilon x_1 = \beta v_1, \quad -\varepsilon x_2 = \beta v_2, \\ -\varepsilon x_1 = \alpha w_1, \quad -\varepsilon x_2 = \alpha w_2. \end{gathered}$$

Тогда, подставляя полученное в выражения для Γ, имеем

$$\Gamma(x, \alpha, p, u, v, w)/2 = \alpha(p^2 + |w|^2) + \beta|v|^2 = \alpha + \beta = 0.$$

Таким образом, поскольку рассматривается случай, когда ограничения на управление активны, $\alpha = -\beta$. Учитывая, что $|x| = 1$, имеем $|v| = 1 \Rightarrow |\alpha| = |\beta| = \varepsilon \Rightarrow |w| = 1 \Rightarrow p = 0$. Следовательно, получили $p = u = 0$ и $w = -v \Rightarrow f = 0$. Это означает, что условие P) выполнено с учетом замечания 5. Тем самым установлено, что ε-задача регулярна. Предложение 1 доказано.

Несложно проверяется, что решения ε-задачи аппроксимируют решение задачи (1), а именно, имеет место следующее утверждение.

Лемма 1. *Для любой числовой последовательности* $\varepsilon_k \to 0$ *существует решение* x^*, α^*, T^* *задачи* (1) *и подпоследовательность* ε_{k_i} *такие, что* $x_{\varepsilon_{k_i}} \rightrightarrows x^*$, $\alpha_{\varepsilon_{k_i}} \rightrightarrows \alpha^*$, *и* $T_{\varepsilon_{k_i}} \to T^*$. *Если решение задачи* (1) *единственно, то* $\varepsilon_{k_i} = \varepsilon_k$.

Утверждение является простым следствием слабой секвенциальной компактности шара в гильбертовом пространстве и того, что ε-задача является релаксацией линейно-выпуклой задачи (1).

Таким образом, установлено, что если требуется найти численно хотя бы одно (какое-либо) решение задачи (1), то при малом $\varepsilon > 0$ таким решением может быть решение ε-задачи (3), которое, в свою очередь, можно найти приближенно с помощью решения условий принципа максимума Понтрягина.

Перейдем к условиям принципа максимума для задачи (3). Фазовое ограничение $|x|^2 \leq 1$ заменим на эквивалентное ему $|x|^2/2 \leq 1/2$, что, очевидно, не отразится на свойстве регулярности ε-задачи. Имеем

$$\begin{gathered} \bar{H}(x, \alpha, p, u, v, w, \psi, \mu) := \psi_1(p\cos\alpha + \varepsilon w_1 + \varepsilon v_1) + \psi_2(p\sin\alpha + \varepsilon w_2 + \varepsilon v_2) + \psi_3 u - \\ - \mu\Gamma(x, \alpha, p, u, v, w) = (\psi_1 + \mu x_1)(p\cos\alpha + \varepsilon w_1 + \varepsilon v_1) + (\psi_2 + \mu x_2)(p\sin\alpha + \varepsilon w_2 + \varepsilon v_2) + \psi_3 u, \end{gathered}$$

где $\psi = (\psi_1, \psi_2, \psi_3) \in \mathbb{R}^3$, $\mu \in \mathbb{R}$ – сопряженные переменные.

Согласно принципу максимума существуют (зависящие от ε) число $\lambda \geq 0$, абсолютно непрерывная вектор-функция $\psi(t)$ и непрерывная убывающая скалярная функция $\mu(t)$ такие, что

$$\lambda + |\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)| + |\psi_2(t) + \mu(t)x_{\varepsilon 2}(t)| + |\psi_3(t)| > 0 \quad \forall t \in [0, T_\varepsilon]. \tag{4}$$

Уравнение для сопряженной переменной принимает вид

$$\begin{gathered}\dot{\psi}_1(t) = -\mu(t)(p_\varepsilon(t)\cos\alpha_\varepsilon(t) + \varepsilon w_{\varepsilon 1}(t) + \varepsilon v_{\varepsilon 1}(t)),\\ \dot{\psi}_2(t) = -\mu(t)(p_\varepsilon(t)\sin\alpha_\varepsilon(t) + \varepsilon w_{\varepsilon 2}(t) + \varepsilon v_{\varepsilon 2}(t)),\\ \dot{\psi}_3(t) = p_\varepsilon(t)\sin\alpha_\varepsilon(t)(\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)) - p_\varepsilon(t)\cos\alpha_\varepsilon(t)(\psi_2(t) + \mu(t)x_{\varepsilon 2}(t)),\end{gathered} \tag{5}$$

причем $\psi_3(T_\varepsilon) = 0$.

Условие максимума распадается на два независимых условия

$$\begin{gathered}\max_{p^2+|w|^2\leq 1}((\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))(p\cos\alpha_\varepsilon(t) + \varepsilon w_1) + (\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))(p\sin\alpha_\varepsilon(t) + \varepsilon w_2)) =\\ = (\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))(p_\varepsilon(t)\cos\alpha_\varepsilon(t) + \varepsilon w_{\varepsilon 1}(t)) + (\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))(p_\varepsilon(t)\sin\alpha_\varepsilon(t) + \varepsilon w_{\varepsilon 2}(t)),\end{gathered} \tag{6}$$

$$\begin{gathered}\max_{u^2+|v|^2\leq 1}((\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))\varepsilon v_1 + (\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))\varepsilon v_2 + \psi_3(t)u) =\\ = (\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))\varepsilon v_{\varepsilon 1}(t) + (\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))\varepsilon v_{\varepsilon 2}(t) + \psi_3(t)u_\varepsilon(t)\end{gathered} \tag{7}$$

для п.в. $t \in [0, T_\varepsilon]$.

Закон сохранения и условие трансверсальности по времени дают

$$\max_{p^2+|w|^2\leq 1, u^2+|v|^2\leq 1} \bar{H}(x_\varepsilon(t), \alpha_\varepsilon(t), p, u, v, w, \psi(t), \mu(t)) = \lambda \quad \forall t \in [0, T_\varepsilon]. \tag{8}$$

Кроме того, функция $\mu(t)$ постоянна на тех отрезках времени, где оптимальная траектория целиком лежит внутри фазового ограничения $|x_\varepsilon(t)| > 1$. Также, не ограничивая общности, согласно замечанию 4 можно полагать, что $\mu(T_\varepsilon) = 0$. Таким образом, $\mu(t) \geq 0$.

Утверждение 1. *Имеет место более сильное, чем* (4), *условие нетривиальности*

$$|\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)| + |\psi_2(t) + \mu(t)x_{\varepsilon 2}(t)| + |\psi_3(t)| > 0 \quad \forall t \in [0, T_\varepsilon]. \tag{9}$$

Действительно, если (9) нарушается, то из (8) имеем $\lambda = 0$, а следовательно, нарушается и (4).

Таким образом, множитель λ далее не играет существенной роли, поскольку в выражениях ниже он не участвует, и можно считать после нормировки, что

$$(\psi_1(T_\varepsilon))^2 + (\psi_2(T_\varepsilon))^2 = 1.$$

Теперь найдем выражения для вычисления управлений p_ε, u_ε, v_ε, w_ε и меры-множителя Лагранжа μ.

Из условия максимума (7) имеем

$$u_\varepsilon(t) = \frac{\psi_3(t)}{\sqrt{A_\varepsilon}}, \quad v_{\varepsilon 1}(t) = \frac{\varepsilon(\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))}{\sqrt{A_\varepsilon}}, \quad v_{\varepsilon 2}(t) = \frac{\varepsilon(\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))}{\sqrt{A_\varepsilon}},$$

где

$$A_\varepsilon := \varepsilon^2(\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))^2 + \varepsilon^2(\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))^2 + (\psi_3(t))^2 > 0.$$

Утверждение 2. *Имеет место более сильное, чем условие* (9), *условие нетривиальности*

$$|\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)| + |\psi_2(t) + \mu(t)x_{\varepsilon 2}(t)| > 0 \quad \forall t \in [0, T_\varepsilon]. \tag{10}$$

Действительно, вблизи границы области, используя замечание 3, из условия P) можно вывести оценку

$$|\mu(t)| \leq \mathrm{const}(|\psi_1(t)| + |\psi_2(t)|),$$

где const – некоторая постоянная (зависящая от ε, но не от t). Поэтому, если (10) нарушается в какой-то точке $\tau \in [0, T_\varepsilon]$, то используя неравенство Гронуолла, считая при этом, что $\mu(\tau) = \psi_1(\tau) = \psi_2(\tau) = 0$ (что возможно в силу замечания 4), из сопряженной системы выводим, что $\mu(t) = \psi_1(t) = \psi_2(t) = 0$ $\forall t \in [0, T_\varepsilon]$. Это однако противоречит (9) при $t = T_\varepsilon$.

Из условия максимума (6) имеем

$$p_\varepsilon(t) = \frac{1}{\sqrt{B_\varepsilon}}((\psi_1(t) + \mu(t)x_{\varepsilon1}(t))\cos\alpha_\varepsilon(t) + (\psi_2(t) + \mu(t)x_{\varepsilon2}(t))\sin\alpha_\varepsilon(t)),$$

$$w_{\varepsilon1}(t) = \frac{1}{\sqrt{B_\varepsilon}}\varepsilon(\psi_1(t) + \mu(t)x_{\varepsilon1}(t)), \quad w_{\varepsilon2}(t) = \frac{1}{\sqrt{B_\varepsilon}}\varepsilon(\psi_2(t) + \mu(t)x_{\varepsilon2}(t)),$$

где

$$B_\varepsilon = ((\psi_1(t) + \mu(t)x_{\varepsilon1}(t))\cos\alpha_\varepsilon(t) + (\psi_2(t) + \mu(t)x_{\varepsilon2}(t))\sin\alpha_\varepsilon(t))^2 +$$
$$+ \varepsilon^2(\psi_1(t) + \mu(t)x_{\varepsilon1}(t))^2 + \varepsilon^2(\psi_2(t) + \mu(t)x_{\varepsilon2}(t))^2 > 0.$$

Вычислим μ. На границе фазового ограничения $\Gamma(x_\varepsilon, \alpha_\varepsilon, p_\varepsilon, u_\varepsilon, v_\varepsilon, w_\varepsilon) = 0$, т.е. (зависимость от времени в дальнейшем для сокращения записи опускаем)

$$x_{\varepsilon1}(p_\varepsilon\cos\alpha_\varepsilon + \varepsilon w_{\varepsilon1} + \varepsilon v_{\varepsilon1}) + x_{\varepsilon2}(p_\varepsilon\sin\alpha_\varepsilon + \varepsilon w_{\varepsilon2} + \varepsilon v_{\varepsilon2}) = 0.$$

Подставляя сюда полученные выше выражения для p_ε, w_ε, v_ε, получаем

$$\frac{(b\mu + a)b + \varepsilon^2\mu + c}{\sqrt{(b\mu + a)^2 + \varepsilon^2\mu^2 + 2c\mu + d}} = -\frac{\varepsilon^2\mu + c}{\sqrt{\varepsilon^2\mu^2 + 2c\mu + d + \psi_3^2}},$$

где

$$a = \psi_1\cos\alpha_\varepsilon + \psi_2\sin\alpha_\varepsilon, \quad b = x_{\varepsilon1}\cos\alpha_\varepsilon + x_{\varepsilon2}\sin\alpha_\varepsilon,$$
$$c = \varepsilon^2(x_{\varepsilon1}\psi_1 + x_{\varepsilon2}\psi_2), \quad d = \varepsilon^2(\psi_1^2 + \psi_2^2).$$

Возводим в квадрат и получаем уравнение 4-й степени на μ:

$$r_4\mu^4 + r_3\mu^3 + r_2\mu^2 + r_1\mu + r_0 = 0,$$

где

$$r_4 = b^4\varepsilon^2 + b^2\varepsilon^4, \quad r_3 = 2b^2(b^2c + ab\varepsilon^2 + 2c\varepsilon^2),$$
$$r_2 = a^2\varepsilon^2(b^2 - \varepsilon^2) + 2abc(2b^2 + \varepsilon^2) + b^2(3c^2 + b^2d + 2d\varepsilon^2) + (b^2 + \varepsilon^2)^2\psi_3^2,$$
$$r_1 = 2[b^2cd + a^2c(b^2 - \varepsilon^2) + ab(c^2 + d(b^2 + \varepsilon^2)) + (ab + c)(b^2 + \varepsilon^2)\psi_3^2],$$
$$r_0 = a(-ac^2 + ab^2d + 2bcd) + (ab + c)^2\psi_3^2.$$

Решая его, после исключения приобретенных корней, получаем формулу

$$\mu = -\frac{r_3}{4r_4} \pm \frac{C_7}{2} + \frac{1}{2}\sqrt{2C_5 - C_6 - \frac{4r_2r_3/r_4^2 - r_3^3/r_4^3 - 8r_1/r_4}{4C_7}}, \tag{11}$$

где

$$C_7 := \sqrt{C_5 + C_6}, \quad C_6 := C_4 + \frac{1}{32^{1/3}r_4}C_3, \quad C_5 := \frac{r_3^2}{4r_4^2} - \frac{2r_2}{3r_4}, \quad C_4 := \frac{2^{1/3}C_1}{3r_4C_3},$$
$$C_3 := (C_2 + \sqrt{-4C_1^3 + C_2^2})^{1/3}, \quad C_2 := 2r_2^3 - 9r_1r_2r_3 + 27r_0r_3^2 + 27r_1^2r_4 - 72r_0r_2r_4,$$
$$C_1 := r_2^2 - 3r_1r_3 + 12r_0r_4.$$

Формула (11) используется для расчета $\mu(t)$ только на границе фазового ограничения, т.е. на тех участках траектории робота, где $x_{\varepsilon1}^2(t) + x_{\varepsilon2}^2(t) = 1$. Вне границы μ не изменяется со временем. Кроме того, μ непрерывна, что позволяет найти точки стыка (т.е. точки, в которых экстремаль выходит на границу или сходит с нее). Полученное значение для μ подставляется в выражения для p_ε, v_ε, w_ε и в сопряженное уравнение. Таким образом, условия принципа максимума сводятся к решению соответствующей краевой задачи.

Итак, построена регулярная аппроксимация исходной задачи (1) и приведены нужные формулы для сведения к краевой задаче. При этом возмущенная задача (3), как и исходная, также является задачей быстродействия, что мы специально отмечаем. В общем случае, т.е. для задачи

управления (2), регуляризация быстродействия быстродействием представляется нетривиальной задачей. По всей видимости, такой тип регуляризации осуществим лишь при некоторых дополнительных предположениях. Тем не менее, если пренебречь свойством быстродействия аппроксимирующей задачи и, тем самым, возмущать помимо динамики еще и минимизирующий функционал, то оказывается возможным предложить достаточно общую схему регуляризации задачи управления. Более того, можно добиться даже более сильного условия регулярности относительно фазового ограничения. Такую общую схему регуляризации обсуждаем в следующем разделе.

4. РЕГУЛЯРИЗАЦИЯ В ОБЩЕМ СЛУЧАЕ

Обратимся к регуляризации общей задачи быстродействия с фазовыми ограничениями. Существует множество различных подходов к регуляризации. Ниже мы рассмотрим лишь один из таких подходов.

Рассмотрим возмущенную задачу

$$\begin{gathered} J(u,v,T) := T + \int_0^T |v|^2 dt \to \min, \quad \dot{x} = f(x,u) + \varepsilon v, \\ x(0) = A, \quad x(T) = B, \quad \varphi(u) \le 0, \quad g(x) \le 0. \end{gathered} \tag{12}$$

Здесь $v = (v_1,\ldots,v_n)$ – дополнительное n-мерное управление.

Напомним, что фазовое ограничение называется регулярным, если $\nabla g(x) \ne 0 \ \forall x : g(x) = 0$. Ограничения на управление $\varphi(u) \le 0$ регулярны, если регулярны все точки множества U. Эти предположения (которые естественны для ряда прикладных задач) ниже считаются a priori выполненными. В этих предположениях проверим для данных задачи (12) выполнение условия регулярности в смысле определения 2.

Для задачи (12) имеем

$$\Gamma(x,u,v) = \langle \nabla g(x), f(x,u) \rangle + \varepsilon \langle \nabla g(x), v \rangle.$$

Градиент Γ и градиенты активных ограничений на управление по (u,v) равны:

$$(\nabla g(x) f_u'(x,u), \varepsilon \nabla g(x)), \quad (\nabla \varphi^i(u), 0),\ i \in I(u).$$

Значит, в силу регулярности фазовых ограничений и регулярности ограничений на управление в задаче (12) *любой* допустимый процесс регулярен относительно фазовых ограничений. Более того, как видно, выполняется даже более сильное условие регулярности, чем сформулированное в определении 2 (см., например, [7, определение 3.5]). Таким образом, получаем следующее утверждение.

Предложение 1. *Пусть фазовые ограничения регулярны, и регулярны все точки $u \in U$. Тогда любой допустимый процесс задачи* (12) *регулярен относительно фазовых ограничений.*

Исследуем вопрос о существовании решения в возмущенной задаче. Заметим, что никаких геометрических ограничений на управление v не имеется, и, следовательно, исходя из вида функционала, можно считать, что управление v ограничено лишь по норме $\|\cdot\|_{L_2}$. Однако при известных предположениях можно гарантировать, что решение задачи (12) принадлежит пространству $\mathbb{W}_{1,2}([0,T])$.

Теорема 2. *Предположим, что*

(a) *множество U компактно*;

(b) *множество $f(x,U)$ выпукло при всех x*;

(c) *имеет место оценка $|f(x,u)| \le \mathrm{const}(1+|x|)\ \forall u \in U$*;

(d) *в исходной задаче* (2) *существует допустимый процесс.*

Тогда существует решение $(x_\varepsilon, u_\varepsilon, v_\varepsilon, T_\varepsilon)$ задачи (12) *в классе траекторий $x \in \mathbb{W}_{1,2}([0,T])$ и управлений $u \in \mathbb{L}_\infty([0,T])$, $v \in \mathbb{L}_2([0,T])$ такое, что $x_\varepsilon \rightrightarrows x_0$, $\|v_\varepsilon\|_{\mathbb{L}_2} \to 0$ при $\varepsilon \to 0$, где x_0 – одно из решений исходной задачи* (2).

Доказательство. Прежде всего отметим существование решения в исходной задаче (2). Действительно, в силу сделанных предположений (a)–(d) это следует из теоремы Филиппова (см. [1]). Обозначим это решение через (x^*, u^*, T^*).

Возьмем натуральное число N и рассмотрим ε, N-задачу, которая получается из задачи (12) путем добавления одного ограничения на управления: $|v| \le N$. Процесс $(x^*, u^*, 0, T^*)$ является допустимым в ε, N-задаче, и, значит, в силу той же теоремы у нее существует решение, которое мы обозначим через $(x_{\varepsilon,N}, u_{\varepsilon,N}, v_{\varepsilon,N}, T_{\varepsilon,N})$. Устремим $N \to \infty$ и найдем искомый процесс $(x_\varepsilon, u_\varepsilon, v_\varepsilon, T_\varepsilon)$.

Имеем

$$T_{\varepsilon,N} + \|v_{\varepsilon,N}\|_{L_2}^2 = J(x_{\varepsilon,N}, u_{\varepsilon,N}, v_{\varepsilon,N}, T_{\varepsilon,N}) \le J(x^*, u^*, 0, T^*) = T^*. \tag{13}$$

Отсюда, поскольку $T_{\varepsilon,N} > 0$, выводим, что последовательность $\{T_{\varepsilon,N}\}$ ограничена. Переходя к подпоследовательности, $T_{\varepsilon,N} \to T_\varepsilon$ при $N \to \infty$. Аналогично, переходя к подпоследовательности, $v_{\varepsilon,N} \xrightarrow{w} v_\varepsilon$ слабо в $\mathbb{L}_2([0, T_\varepsilon])$.

Из условия (c) и из (13) с помощью стандартных оценок, включающих неравенство Гронуола, получаем, что последовательность функций $x_{\varepsilon,N}$ ограничена в равномерной метрике $\|x_{\varepsilon,N}\|_C \le \mathrm{const}$. Поэтому семейство функций

$$y_{\varepsilon,N}(t) := A + \int_0^t f(x_{\varepsilon,N}(s), u_{\varepsilon,N}(s))ds$$

равномерно ограничено и равностепенно непрерывно. Поэтому, переходя к подпоследовательности, можно считать, что $y_{\varepsilon,N} \rightrightarrows y_\varepsilon$ равномерно на $[0, T_\varepsilon]$. При этом $\dot{y}_{\varepsilon,N} \xrightarrow{w} \dot{y}_\varepsilon$ слабо в L_2. Семейство функций

$$z_{\varepsilon,N}(t) := \int_0^t v_{\varepsilon,N}(s)ds$$

также равномерно ограничено и равностепенно непрерывно. Это очевидно следует из оценки (13), так как по неравенству Коши–Буняковского

$$\left|\int_{t_1}^{t_2} v_{\varepsilon,N}^j(t)dt\right| \le \sqrt{|t_1 - t_2|}\,\|v_{\varepsilon,N}^j\|_{L_2}, \quad j = 1, \ldots, n.$$

Поэтому, переходя к подпоследовательности, $z_{\varepsilon,N} \rightrightarrows z_\varepsilon$, и, значит, имеем $x_{\varepsilon,N} \rightrightarrows x_\varepsilon := y_\varepsilon + z_\varepsilon$. Из условий (a), (b) в силу равномерной сходимости $x_{\varepsilon,N}$ из леммы Мазура следует, что $\dot{y}_\varepsilon(t) \in f(x_\varepsilon(t), U)$ для п.в. t. По лемме об измеримом селекторе существует управление $u_\varepsilon : [0, T_\varepsilon] \to U$ такое, что $\dot{y}_\varepsilon(t) = f(x_\varepsilon(t), u_\varepsilon(t))$ для п.в. t. Таким образом,

$$x_\varepsilon(t) = A + \int_0^t f(x_\varepsilon(s), u_\varepsilon(s))ds + \varepsilon\int_0^t v_\varepsilon(s)ds,$$

и x_ε удовлетворяет фазовым и граничным условиям. Значит, $(x_\varepsilon, u_\varepsilon, v_\varepsilon, T_\varepsilon)$ – допустимый процесс в задаче (12). Из оценки (13) в силу слабой полунепрерывности снизу функционала J имеем

$$T_\varepsilon + \|v_\varepsilon\|_{L_2}^2 \le T^*.$$

Перейдем к пределу при $\varepsilon \to 0$. Рассуждая таким же образом, переходя к подпоследовательности, $T_\varepsilon \to T_0$, $x_\varepsilon \rightrightarrows x_0 \in \mathbb{W}_{1,2}([0, T_0])$, $v_\varepsilon \xrightarrow{w} v_0 \in \mathbb{L}_2([0, T_0])$, $\varepsilon \to 0$. Аналогично, найдется управление u_0 такое, что процесс (x_0, u_0, T_0) является допустимым в задаче (2). Однако

$$T_0 + \|v_0\|_{L_2}^2 \le T^* \Rightarrow T_0 = T^*, \quad v_0 = 0,$$

и $v_\varepsilon \to 0$ сильно в силу оптимальности. Следовательно, процесс (x_0, u_0, T_0) также является решением задачи (2). Теорема 2 доказана.

В классе траекторий из $\mathbb{W}_{1,2}([0,T])$ также справедлив принцип максимума. Его несложно вывести по аналогии с доказательством теоремы 2, применив к ε, N-задаче принцип максимума для траекторий класса $\mathbb{W}_{1,\infty}([0,T])$ и перейдя в полученных условиях к пределу при $N \to \infty$. Для простоты изложения предположим, что $g(A) + g(B) < 0$, т.е. что концы траектории лежат строго во внутренности фазового ограничения. С одной стороны, такое предположение позволяет избежать в пределе эффекта вырождающегося принципа максимума (см. [4], [12]) – случая, который требует отдельного рассмотрения. С другой стороны, оно не является существенным для задачи об обходе препятствия мобильным роботом. Тогда, переходя стандартным образом к пределу при $N \to \infty$ в условиях принципа максимума, можно утверждать существование нетривиального набора множителей Лагранжа $(\lambda_\varepsilon, \psi_\varepsilon, \mu_\varepsilon)$, удовлетворяющего сопряженной системе, условию максимума и условию трансверсальности по времени. (Условие максимума выводится с помощью интегральной формы и слабой сходимости.) Условие нетривиальности

$$\lambda_\varepsilon + \left|\psi_\varepsilon(t) - \mu_\varepsilon(t)\nabla g(x_\varepsilon(t))\right| > 0 \quad \forall t \in [0, T_\varepsilon] \tag{14}$$

выводится с помощью условия регулярности, которое выполняется в ε-задаче, и рассуждений, полностью аналогичных приведенным в [13] при доказательстве теоремы 3.1 (отличие заключается лишь в использовании расширенной версии неравенства Гронуолла, см., например, упражнение 3.4 в [14]).

Рассмотрим ε-условие максимума по v. Имеем

$$v_\varepsilon(t) \in \operatorname*{argmax}_{v \in \mathbb{R}^n} (\varepsilon < v, \psi_\varepsilon(t) - \mu_\varepsilon(t)\nabla g(x_\varepsilon(t)) > -\lambda_\varepsilon |v|^2).$$

Отсюда ввиду (14) получаем, что $\lambda_\varepsilon > 0$, следовательно,

$$v_\varepsilon(t) = \frac{\varepsilon}{2\lambda_\varepsilon}(\psi_\varepsilon(t) - \mu_\varepsilon(t)\nabla g(x_\varepsilon(t))). \tag{15}$$

Таким образом, показано, что функция v_ε ограничена, и, значит, поскольку выполнены условия регулярности, можно применить теорему 1. Поэтому функция μ_ε непрерывна, и, следовательно, v_ε также непрерывна. Отметим, что сопряженная система задачи (12) отличается от исходной сопряженной системы задачи (2) наличием малого регуляризирующего члена $\varepsilon\mu_\varepsilon(t)g''(x_\varepsilon(t))v_\varepsilon(t)$. Здесь функция v_ε определяется выражением (15). При каждом $\varepsilon > 0$ эта функция ограничена и даже, как было отмечено, непрерывна. Однако неограниченные значения не исключаются при $\varepsilon \to 0$.

Заметим, что, добавив в функционал J малый член $\varepsilon\|u\|_{L_2}$, наряду с регулярностью относительно фазовых ограничений, можно добиться и выполнения усиленного условия Лежандра. Тогда можно утверждать (см. [7]), что множитель μ_ε является даже липшицевой функцией (однако с константой Липшица, зависящей от ε).

Теперь применим описанную выше схему возмущения к задаче о движении мобильного робота (1). При этом понятно, что задача (12) представляет собой некоторую общую схему возмущения относительно фазового ограничения, но не учитывает возможных особых режимов. Однако, очевидно, что не все время в пути робот будет разворачиваться с максимальной по модулю угловой скоростью, и поэтому анализ особого режима по угловой скорости является критически важным для модели (1). Проблему особых режимов можно решить (по аналогии с разд. 3) с помощью дополнительной регуляризации задачи и введения дополнительных переменных управления. Другим подходом может служить возмущение минимизируемого функционала с помощью малой интегральной добавки. Ниже рассмотрим именно второй подход как альтернативу первому подходу, рассмотренному в разд. 3.

Рассмотрим возмущенную задачу

$$J(p,u,v,T) := T + \varepsilon\int_0^T |p|^2 dt + \varepsilon\int_0^T |u|^2 dt + \int_0^T |v|^2 dt \to \min, \tag{16a}$$

$$\dot{x}_1 = p\cos\alpha + \varepsilon v_1, \quad \dot{x}_2 = p\sin\alpha + \varepsilon v_2, \quad \dot{\alpha} = u, \tag{16b}$$

$$x(0) = A, \quad x(T) = B, \quad \alpha(0) = \alpha_0, \tag{16c}$$

$$p \in [-1,1], \quad u \in [-1,1], \tag{16d}$$

$$\frac{1}{2}(x_1^2 + x_2^2) \geq \frac{1}{2}. \tag{16e}$$

Здесь $v = (v_1, v_2)$ – дополнительная двумерная переменная управления. (Поясним, что рассмотрение переменной v_3 излишне, поскольку градиент фазового ограничения по α тождественно равен нулю.)

Имеем

$$\Gamma(x,\alpha,p,u,v) = -x_1(p\cos\alpha + \varepsilon v_1) - x_2(p\sin\alpha + \varepsilon v_2),$$
$$\bar{H}(x,\alpha,p,u,v,\psi,\mu,\lambda) = (\psi_1 + \mu x_1)(p\cos\alpha + \varepsilon v_1) + (\psi_2 + \mu x_2)(p\sin\alpha + \varepsilon v_2) +$$
$$+ \psi_3 u - \varepsilon\lambda|p|^2 - \varepsilon\lambda|u|^2 - \lambda|v|^2.$$

Рассмотрим зависящие от параметра ε число $\lambda > 0$ (случай $\lambda = 0$, как несложно проверить, исключен), абсолютно непрерывную вектор-функцию $\psi(t)$ и непрерывную и убывающую скалярную функцию $\mu(t)$, которые удовлетворяют принципу максимума. Имеем

$$\begin{gathered}\dot{\psi}_1(t) = -\mu(t)(p_\varepsilon(t)\cos\alpha_\varepsilon(t) + \varepsilon v_{\varepsilon 1}(t)),\\ \dot{\psi}_2(t) = -\mu(t)(p_\varepsilon(t)\sin\alpha_\varepsilon(t) + \varepsilon v_{\varepsilon 2}(t)),\\ \dot{\psi}_3(t) = p_\varepsilon(t)\left[\sin\alpha_\varepsilon(t)(\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)) - \cos\alpha_\varepsilon(t)(\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))\right],\end{gathered} \tag{17}$$

а также

$$\psi_3(T_\varepsilon) = 0. \tag{18}$$

Анализ условия максимума по переменным v_1, v_2 дает

$$v_{\varepsilon 1}(t) = \frac{\varepsilon}{2\lambda}(\psi_1(t) + \mu(t)x_{\varepsilon 1}(t)), \quad v_{\varepsilon 2}(t) = \frac{\varepsilon}{2\lambda}(\psi_2(t) + \mu(t)x_{\varepsilon 2}(t)). \tag{19}$$

Условие максимума по переменной p влечет

$$p_\varepsilon(t) = \operatorname{sgn} a_\varepsilon(t)\min\left\{\frac{|a_\varepsilon(t)|}{2\varepsilon\lambda}, 1\right\}, \tag{20}$$

где

$$a_\varepsilon(t) = (\psi_1(t) + \mu(t)x_{\varepsilon 1}(t))\cos\alpha_\varepsilon(t) + (\psi_2(t) + \mu(t)x_{\varepsilon 2}(t))\sin\alpha_\varepsilon(t). \tag{21}$$

Из условия максимума по переменной u следует, что

$$u_\varepsilon(t) = \operatorname{sgn}\psi_3(t)\min\left\{\frac{|\psi_3(t)|}{2\varepsilon\lambda}, 1\right\}. \tag{22}$$

Заметим, что $|\psi_1(T_\varepsilon)| + |\psi_2(T_\varepsilon)| > 0$. Действительно, иначе, учитывая, что $\mu(T_\varepsilon) = 0$ и $\psi_3(T_\varepsilon) = 0$ ввиду условия трансверсальности по времени получаем $\lambda = 0$. Таким образом, после нормировки можно считать, что

$$(\psi_1(T_\varepsilon))^2 + (\psi_2(T_\varepsilon))^2 = 1. \tag{23}$$

Найдем формулы для множителей λ и μ. Используя условие трансверсальности по времени, учитывая, что $\mu(T_\varepsilon) = 0$, $\psi_3(T_\varepsilon) = 0$ и (23), решая соответствующее квадратное уравнение и отбрасывая лишние корни, приходим к следующим выражениям для λ:

Случай 1. Пусть $a_\varepsilon(T_\varepsilon) \leq -\varepsilon^2/\sqrt{1-\varepsilon}$. Тогда

$$\lambda = \lambda_1 = \frac{-a_\varepsilon(T_\varepsilon) + \sqrt{a_\varepsilon^2(T_\varepsilon) + \varepsilon^2(\varepsilon+1)}}{2(\varepsilon+1)}. \tag{24}$$

Случай 2. Пусть $a_\varepsilon(T_\varepsilon) \in (-\varepsilon^2/\sqrt{1-\varepsilon}, \varepsilon^2/\sqrt{1-\varepsilon})$. Тогда

$$\lambda = \lambda_2 = \sqrt{\frac{a_\varepsilon^2(T_\varepsilon) + \varepsilon^3}{4\varepsilon}}. \tag{25}$$

Случай 3. Пусть $a_\varepsilon(T_\varepsilon) \geq \varepsilon^2/\sqrt{1-\varepsilon}$. Тогда

$$\lambda = \lambda_3 = \frac{a_\varepsilon(T_\varepsilon) + \sqrt{a_\varepsilon^2(T_\varepsilon) + \varepsilon^2(\varepsilon+1)}}{2(\varepsilon+1)}. \tag{26}$$

На границе фазового ограничения имеем

$$\Gamma(x_\varepsilon, \alpha_\varepsilon, p_\varepsilon, u_\varepsilon, v_\varepsilon) = 0,$$

т.е.

$$x_{\varepsilon 1}(p_\varepsilon \cos\alpha_\varepsilon + \varepsilon v_{\varepsilon 1}) + x_{\varepsilon 2}(p_\varepsilon \sin\alpha_\varepsilon + \varepsilon v_{\varepsilon 2}) = 0.$$

(Зависимость от времени в дальнейшем для удобства записи опущена.) Подставляя в это равенство полученные выше выражения для p_ε и v_ε, при $a_\varepsilon \leq -2\varepsilon\lambda$, $a_\varepsilon \in (-2\varepsilon\lambda, 2\varepsilon\lambda)$ и $a_\varepsilon \geq 2\varepsilon\lambda$ получаем следующие выражения для вычисления μ:

Случай А. Пусть $a_\varepsilon(t) \leq -2\varepsilon\lambda$. Тогда

$$\mu = \frac{2\lambda}{\varepsilon^2}(x_{\varepsilon 1}\cos\alpha_\varepsilon + x_{\varepsilon 2}\sin\alpha_\varepsilon) - (\psi_1 x_{\varepsilon 1} + \psi_2 x_{\varepsilon 2}). \tag{27}$$

Случай Б. Пусть $a_\varepsilon(t) \in (-2\varepsilon\lambda, 2\varepsilon\lambda)$. Тогда

$$\mu = -\frac{(\psi_1\cos\alpha_\varepsilon + \psi_2\sin\alpha_\varepsilon)(x_{\varepsilon 1}\cos\alpha_\varepsilon + x_{\varepsilon 2}\sin\alpha_\varepsilon) + \varepsilon^3(\psi_1 x_{\varepsilon 1} + \psi_2 x_{\varepsilon 2})}{(x_{\varepsilon 1}\cos\alpha_\varepsilon + x_{\varepsilon 2}\sin\alpha_\varepsilon)^2 + \varepsilon^3}. \tag{28}$$

Случай В. Пусть $a_\varepsilon(t) \geq 2\varepsilon\lambda$. Тогда

$$\mu = -\frac{2\lambda}{\varepsilon^2}(x_{\varepsilon 1}\cos\alpha_\varepsilon + x_{\varepsilon 2}\sin\alpha_\varepsilon) - (\psi_1 x_{\varepsilon 1} + \psi_2 x_{\varepsilon 2}). \tag{29}$$

Заметим, что каждого λ_1, λ_2 и λ_3 реализуются свои случаи А, Б и В (итак, всего имеется 9 разных случаев).

Вне границы фазового ограничения μ не изменяется со временем. Кроме того, μ непрерывна, что позволяет найти точки стыка (т.е. точки, в которых экстремаль выходит на границу или сходит с нее). Полученное значение для μ подставляется в выражения для p_ε, v_ε и в сопряженное уравнение. Таким образом, условия принципа максимума сводятся к решению соответствующей краевой задачи.

Теперь, считая выполненным (23), $\mu(T_\varepsilon) = 0$ и $\psi_3(T_\varepsilon) = 0$, получим необходимые оценки сверху на $|\psi|$ и μ в зависимости от ε. Поскольку из формул для λ в общем случае имеем $\lambda = O(\varepsilon^{-1})$, то из формул для v_ε находим

$$\mu(t) \leq \mathrm{const}(\varepsilon^{-2}|v_\varepsilon(t)| + |\psi_1(t)| + |\psi_2(t)|).$$

Отсюда, используя сопряженное уравнение (17) и неравенство Гронуолла (напомним, что v_ε непрерывна), стандартными рассуждениями получаем, что

$$|\psi(t)| \leq \|v_\varepsilon\|_{L_2} O(\varepsilon^{-2}) \Rightarrow \mu(0) = \|v_\varepsilon\|_{L_2} O(\varepsilon^{-2}).$$

Поскольку $\|v_\varepsilon\|_{L_2} \to 0$ при $\varepsilon \to 0$, имеем $\mu(0) = o(\varepsilon^{-2})$. Однако знанием о скорости сходимости $\|v_\varepsilon\|_{L_2}$ к нулю мы не обладаем. Таким образом, можно гарантировать лишь следующее. Пусть

$$|\psi_1(T_\varepsilon)| + |\psi_2(T_\varepsilon)| = \varepsilon^2.$$

Тогда ψ и μ ограничены на всем отрезке движения одной константой, которая не зависит от ε.

В завершение теоретической части отметим, что радиус окружности, задающей фазовое ограничение, может быть выбран произвольным образом, т.е. не обязательно быть единичным. В таком случае немного меняется формула для μ, где появляется радиус окружности r. Тем не менее, с точки зрения вычислений случай именно единичной окружности интересен тем, что он является пограничным в следующем смысле: ясно, что робот может двигаться одновременно с максимальными угловой и продольной скоростями только по единичной окружности. Если как-ли-

бо изменить радиус, то это уже не так: при $r > 1$ угловая скорость не максимальна, а при $r < 1$ – продольная.

5. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Рассмотрим следующую схему приближенного решения исходной задачи (1). Фиксируем некоторое $\varepsilon \in (0,1)$ и переходим к возмущенной задаче (16), которая регуляризована. К возмущенной задаче применяем численный метод, предложенный в [9], [15]. Используем решение этой задачи для заданного ε в качестве начального приближения к решению задачи (16) с меньшим значением ε. Решение этой задачи для наименьшего из рассмотренных значений ε считаем приближением к решению исходной задачи (1).

5.1. Алгоритм решения задачи (16)

Принцип максимума для этой задачи сводится к краевой задаче с помощью выражений, приведенных в предыдущем разделе. Краевая задача принципа максимума решается методом стрельбы в обратном времени (т.е. перебор значений параметров, определяющих решение краевой задачи, проводится при $t = T$).

Итак, рассматривается краевая задача для системы обыкновенных дифференциальных уравнений (16b) и (17) и краевых условий (16c) и (18). При $t = T$ имеем следующие условия на решение: мобильный робот находится в точке B (16c) и третья компонента сопряженной переменной равна нулю (18). Используя (23), считаем, что $\psi_1(T) = \cos(\theta)$ и $\psi_2(T) = \sin(\theta)$, где $\theta \in [0, 2\pi)$. Также положим $\alpha(T) = \beta$, где $\beta \in [0, 2\pi)$. Если значения θ и β заданы, то при $t = T$ заданы все функции, входящие в рассматриваемую систему дифференциальных уравнений. Суть метода стрельбы состоит в том, чтобы численно найти значения параметров θ и β, чтобы решение $(x(t), \psi(t))$ системы дифференциальных уравнений (16b) и (17), удовлетворяющее указанным краевым условиям при $t = T$, также удовлетворяло бы (в заданной точностью) краевым условиям при $t = 0$ (значения всех параметров, определяющих точность численного решения задачи, приведены в следующем пункте.). Такая траектория является решением краевой задачи, полученной из принципа максимума Понтрягина, и, следовательно, является экстремалью рассматриваемой задачи оптимального управления.

При заданном наборе значений величин θ и β система уравнений (16b) и (17) численно интегрируется в обратном времени классическим методом Рунге–Кутты четвертого порядка с постоянным шагом. Перед началом интегрирования вычисляется значение λ по формулам (24)–(26). Считая, что исходная точка B не принадлежит границе, полагаем $\mu(t) = 0$, пока траектория не достигнет границы фазового ограничения (16e). При вычислении правых частей дифференциальных уравнений функции управления $p(t)$, $u(t)$, $v_1(t)$ и $v_2(t)$ вычисляются по формулам (20), (22) и (19) соответственно.

Правые части интегрируемых дифференциальных уравнений, вообще говоря, негладкие, ввиду возможных переключений режимов управления $p(t)$ и $u(t)$, т.е. при t, когда выполнено $|a_\varepsilon(t)| = 2\varepsilon\lambda$ (переключение $p(t)$, $a_\varepsilon(t)$ задано выражением (21)) и $|\psi_3(t)| = 2\varepsilon\lambda$ (переключение $u(t)$). В эти моменты времени соответствующее управление становится или перестает быть максимальным (равным 1 по абсолютной величине). Чтобы избежать потерю точности в окрестности таких точек, при расчете правых частей в методе Рунге–Кутты считаем шаг неуспешным, если при его вычислении происходит такое переключение для одного из управлений. В таком случае уменьшаем шаг интегрирования вдвое и повторяем попытку выполнения шага. Уменьшение шага прекращается после семикратного уменьшения начального шага (что отвечает уменьшению начального шага примерно на два порядка), после чего продолжаем интегрирование с начальным шагом по времени с новым режимом по рассмотренному управлению.

Если траектория пересекает границу фазового ограничения (16e), то в точке пересечения вычисляем значение меры-множителя Лагранжа μ, используя выражения (27)–(29). Если его абсолютное значение не превосходит заданной (малой) величины, считаем, что в этой точке $\mu(t)$ непрерывно, в этот момент времени происходит заход на границу фазового ограничения, дальше траектория следует по этой границе, расчет меры-множителя на каждом шаге по времени происходит по формулам (27)–(29). Если же его значение больше заданной величины, то, ввиду раз-

рыва $\mu(t)$ в этой точке, построенная траектория не является частью экстремали и, следовательно, не представляет интереса.

Для траекторий, проходящих по границе фазового ограничения, осуществляем следующую процедуру. На каждом шаге по времени фиксируем μ и продолжаем интегрирование системы дифференциальных уравнений с этим фиксированным значением меры-множителя Лагранжа. Такая траектория сходит с границы и может являться частью экстремали, соединяющей границу с точкой A или снова заходящей на границу.

На каждом шаге по времени измеряется расстояние от текущей позиции робота до краевого условия при $t = 0$ в (16c), и если оно меньше заданного значения, то построенная траектория считается решением краевой задачи. Если таких траекторий найдено несколько, то выбираем ту из них, которая отвечает меньшему значению функционала (16a).

Были использованы следующие условия, по которым траектория отвергалась из кандидатов в экстремали рассмотренной задачи:

– время движения робота больше заданного (достаточно большого) значения,

– при следовании по границе нарушение монотонности μ,

– при сходе с границы нарушение фазового ограничения.

Для наибольшего значения ε начальный выбор значений параметров θ и β проводили на равномерной сетке в квадрате $[0, 2\pi)^2$, локальные минимумы соответствующих невязок уточнялись методом деления пополам. Для бóльших значений ε в качестве начального значения указанных параметров брали их “оптимальное” значение, найденное для меньшего значения ε.

5.2. Результаты расчетов

Программой, реализующей указанный алгоритм, была численно решена задача оптимального управления (0) для следующих исходных данных:

$$x(0) = A = (-1.5, 0.3),$$
$$x(T) = B = (1.5, -0.5),$$
$$\alpha(0) = \alpha_0 = -\pi/4.$$

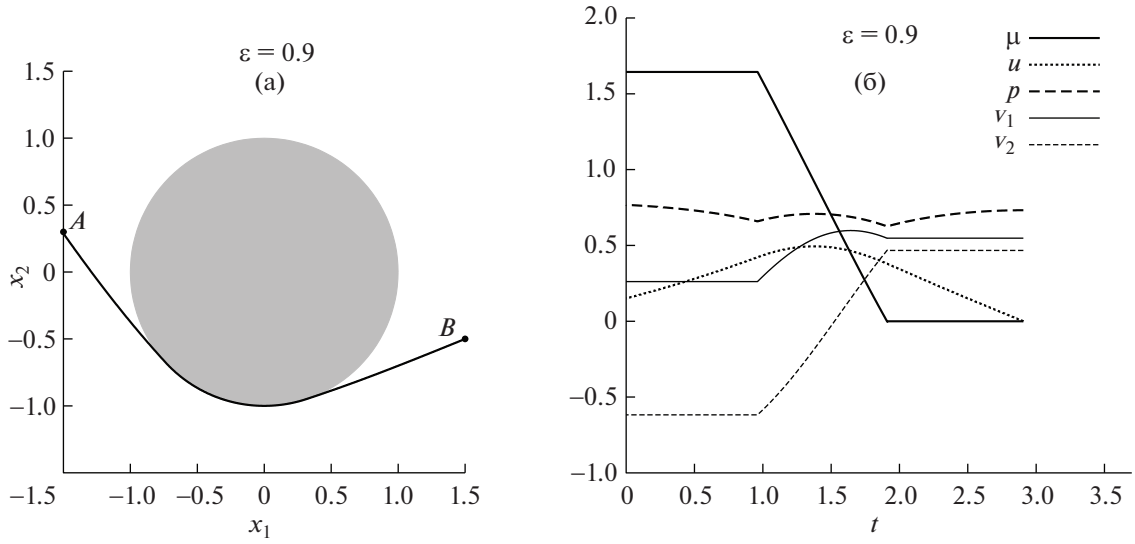
Расположение начальной и конечной точек маршрута робота, а также препятствие показаны на фиг. 1a.

Шаг интегрирования системы дифференциальных уравнений брали равным 10^{-3}. Если траектория пересекала границу фазового ограничения (16e), точку пересечения находили линейной интерполяцией, в этой точке вычисляли μ и считали ее непрерывной (т.е. в этой точке происходит заход на границу), если абсолютная величина этого значения не превосходила 10^{-2}. Решение краевой задачи считалось построенным, если расстояние от траектории до начального условия (т.е. в (16c) при $t = 0$) не превосходило 10^{-2}. Построение траектории останавливалось, если время следования по траектории превосходило 20 единиц времени.

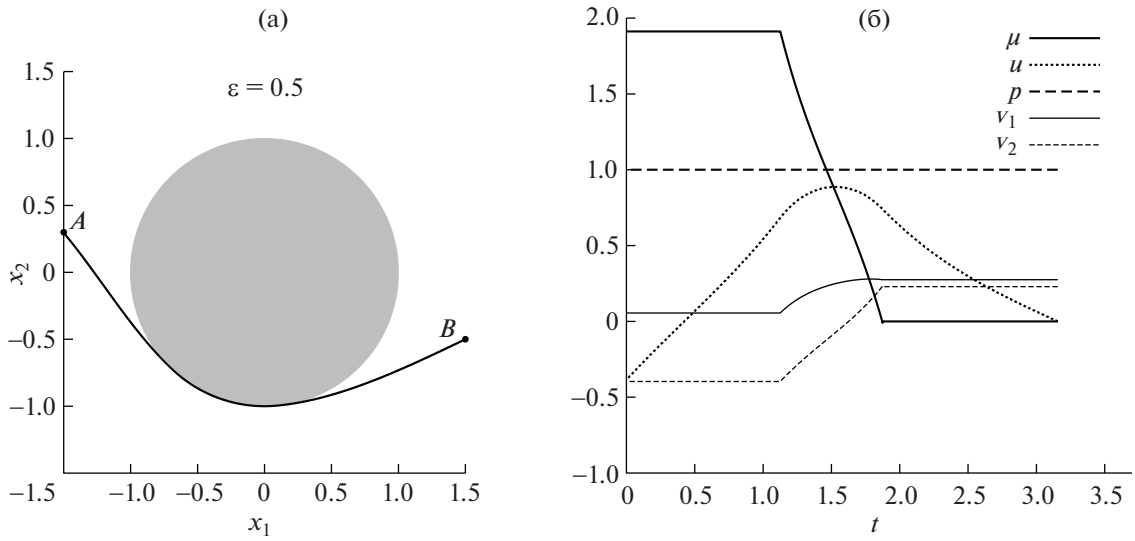
Хотя, как описано выше, решение задачи строилось в обратном времени, т.е. от $\tilde{t} = 0$ до $\tilde{t} = -T$, для построения фигур все функции преобразованы и показаны в прямом времени: $t = \tilde{t} + T$. Чтобы упростить сравнение, для всех значений ε на вертикальных и горизонтальных осях всех фигур, приведенных ниже, показаны одинаковые интервалы значений.

Были проведены расчеты для $\varepsilon = 0.9$, 0.5, 0.2, 0.1 и 0.05, вычисленные оптимальные траектории и соответствующие управления показаны на фиг. 1–5 соответственно.

На первый взгляд, траектории робота, отвечающие решениям задачи оптимального управления при разных значениях ε, отличаются друг от друга мало (ср. фиг. 1a–5a). Каждая из этих траекторий состоит из трех частей – части заходящей на границу фазового ограничения, движению по границе и части, сходящей с границы. Однако режимы управления, обеспечивающие такое движение, различаются существенно. Так, при $\varepsilon = 0.9$ ни скорость поворота робота $u(t)$, ни его продольная скорость $p(t)$, не являются максимальными ни в какой точке траектории (фиг. 1б). При $\varepsilon = 0.5$ (фиг. 2б) скорость поворота снова нигде не достигает максимального значения, но его продольная скорость становится максимальной на всем участке движения, т.е. $p(t) = 1$ $\forall t \in [0, T]$. При $\varepsilon = 0.2$ (фиг. 3б) продольная скорость остается всегда максимальной, а скорость поворота достигает максимального значения на некотором участке движения вдоль границы.



Фиг. 1. Вычисленное решение регуляризованной задачи (0) при $\varepsilon = 0.9$: (а) – оптимальная траектория робота, (б) – эволюция вдоль этой траектории меры-множителя Лагранжа $\mu(t)$ и управлений $u(t)$, $p(t)$, $v_1(t)$ и $v_2(t)$. Препятствие показано на (а) серым цветом.

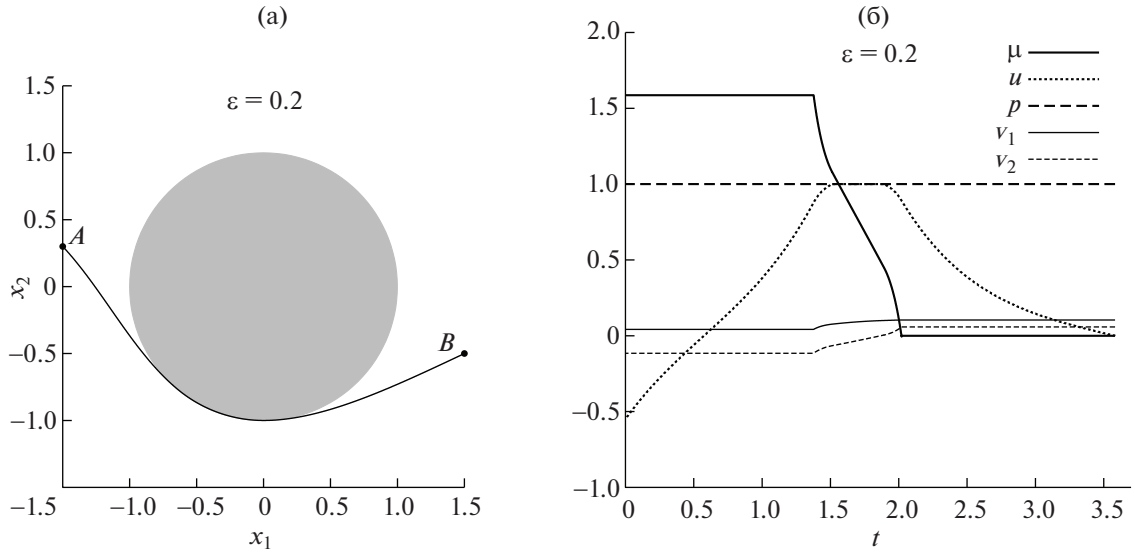


Фиг. 2. То же, что на фиг. 1, но при $\varepsilon = 0.5$.

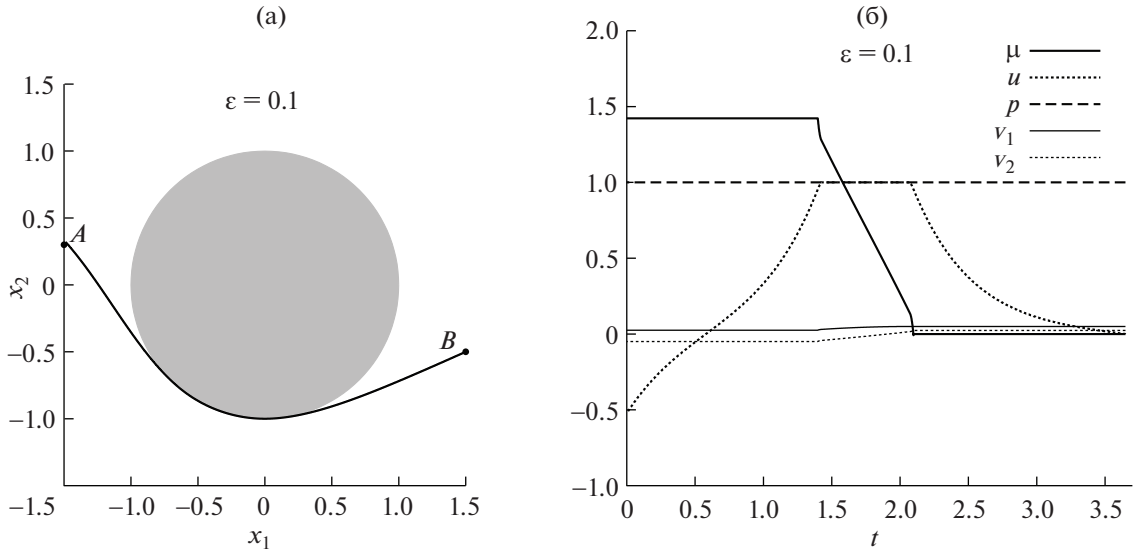
При $\varepsilon = 0.1$ и 0.05 качественного изменения в сравнении с предыдущим случаем нет, но скорость поворота достигает максимального значения на большем участке границы (фиг. 4б и 5б), чем для траектории при $\varepsilon = 0.2$.

Время следования по оптимальным траекториям T_ε растет с уменьшением ε: $T_{0.9} = 2.91$, $T_{0.5} = 3.16$, $T_{0.2} = 3.58$, $T_{0.1} = 3.64$ и $T_{0.05} = 3.65$. Мы не рассматривали меньших значений ε, поскольку разница между двумя последними, $T_{0.05}$ и $T_{0.1}$, мала – порядка 10^{-2}. Кроме того, отметим поведение регуляризаторов, т.е. функций v_1 и v_2 (тонкие линии на фиг. 1б–5б). Видно, что они стремятся к нулю в равномерной метрике при $\varepsilon \to 0$, причем с линейной относительно ε скоростью, что также свидетельствует о сходимости метода.

Следует отметить характер изменения μ при уменьшении ε (при движении по границе, так как при движении вне границы величина меры-множителя Лагранжа не меняется). При $\varepsilon = 0.9$

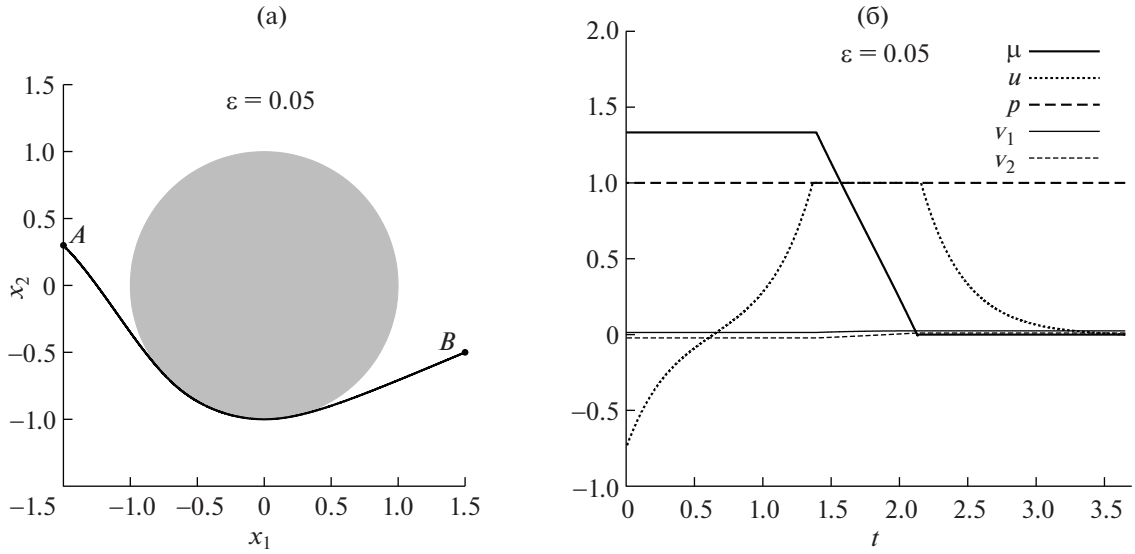


Фиг. 3. То же, что на фиг. 1, но при $\varepsilon = 0.2$.



Фиг. 4. То же, что на фиг. 1, но при $\varepsilon = 0.1$.

фиг. 1б между точкой входа и схода с границы $\mu(t)$ – практически линейная функция. При $\varepsilon = 0.5$ и 0.2 (фиг. 2б и 3б) заметно характерное нелинейное поведение – вблизи точки захода и схода с границы функция быстро меняется, оставаясь близкой к линейной при удалении от этих точек. Такая нелинейность (рост градиента) особенно заметна при $\varepsilon = 0.1$ (фиг. 4б). Однако при $\varepsilon = 0.05$ линейность уже полностью восстанавливается. Это означает, что предельная мера-множитель Лагранжа, которая отвечает принципу максимума в исходной задаче (1), по всей вероятности непрерывна. В этой связи отметим, что в известной литературе существует совсем немного примеров, которые демонстрируют наличие сингулярной составляющей меры и, в частности, атомов (см., например, [16], [17]). Представляется любопытным, что в такой существенно нерегулярной задаче, как движение мобильного робота относительно препятствия, мера-множитель все же сохраняет свойство непрерывности. Заметим, что такое свойство не гарантируется никакой существующей теорией и установлено пока лишь эмпирически.



Фиг. 5. То же, что на фиг. 1, но при $\varepsilon = 0.05$.

6. ПРИЛОЖЕНИЕ

Приведем доказательство теоремы 1. Напомним следующее понятие (см. [4]).

Определение 3. В концевых точках выполняются условия управляемости относительно фазовых ограничений, если найдутся векторы $u_A, u_B \in U$ такие, что

$$\langle \nabla g(A), f(A, u_A) \rangle < 0 \quad \text{при} \quad g(A) = 0,$$

$$\langle \nabla g(B), f(B, u_B) \rangle > 0 \quad \text{при} \quad g(B) = 0.$$

Относительно этих условий и условия регулярности имеет место следующее простое утверждение.

Предложение 2. *Пусть допустимый процесс* (x, u, T) *регулярен относительно фазовых ограничений. Тогда в концевых точках выполнены условия управляемости относительно фазовых ограничений.*

Доказательство очевидно следует из определений. Перейдем к доказательству теоремы.

Доказательство теоремы 1. Рассмотрим заданный оптимальный процесс управления (x^*, u^*, T^*), регулярный относительно фазовых ограничений. В силу предложения 2 в концевых точках выполнены условия управляемости относительно фазовых ограничений. Тогда ввиду результатов, полученных в [4], [7], [13], а также уточнений, сделанных в [8, лемма 4.1], существует невырождающий набор множителей Лагранжа (λ, ψ, μ) такой, что функция $\mu(t)$ непрерывна на $(0, T^*)$, и который удовлетворяет всем условиям принципа максимума, кроме условия нетривиальности, которое слабее, чем в теореме:

$$\lambda + \ell\left(\left\{t \in [0, T^*] : \psi(t) - \mu(t)\nabla g(x^*(t)) \neq 0\right\}\right) > 0. \tag{30}$$

Ясно, что ввиду (30) и замечания 1 можно считать, что функция μ непрерывна на всем отрезке времени $[0, T^*]$.

Необходимо показать, что в условиях регулярности, сформулированных в этой работе, условие (30) влечет

$$\lambda + \left|\psi(t) - \mu(t)\nabla g(x^*(t))\right| > 0 \quad \forall t \in [0, T^*]. \tag{31}$$

Для простоты изложения предположим, что $u^*(t) \in \operatorname{int} U$ для п.в. t. (Общий случай рассматривается в полной аналогии.) При этом предположении из условия максимума следует, что

$$H'_u(x^*(t), u, \psi(t), \lambda) = \mu(t)\Gamma'_u(x^*(t), u) \quad \forall u \in \mathcal{U}(t), \quad t \in [0, T^*], \tag{32}$$

где H означает обычную функцию Гамильтона–Понтрягина

$$H(x,u,\psi,\lambda) := \langle \psi, f(x,u)\rangle - \lambda f_0(x,u).$$

Рассмотрим множество точек

$$\mathcal{T} := \{t \in [0,T^*] : g(x^*(t)) = 0,\ \exists u_0 \in \mathcal{U}(t) : \Gamma(x^*(t),u_0) = 0\}.$$

Отметим, что множество $\mathcal{T}$ замкнуто.

Рассмотрим произвольную точку $t \in \mathcal{T}$. В силу регулярности $\Gamma'_u(x^*(t),u_0) \neq 0$ для некоторого $u_0 = u_0(t) \in \mathcal{U}(t)$. Поэтому, взяв скалярное произведение (32) с $\Gamma'_u(x^*(t),u_0)$, учитывая, что $|\Gamma'_u(x^*(t),u_0)|^2 > 0$, получаем, что

$$\mu(t) = \frac{\left\langle H'_u(x^*(t),u_0(t),\psi(t),\lambda), \Gamma'_u(x^*(t),u_0(t))\right\rangle}{|\Gamma'_u(x^*(t),u_0)|^2} \quad \forall t \in \mathcal{T}. \tag{33}$$

Используя стандартные рассуждения, основанные на соображении компактности, и условие регулярности, можно заключить, что $|\Gamma'_u(x^*(t),u_0)| \geq \varepsilon\ \forall t \in T$ для некоторого $\varepsilon > 0$. Тогда из (33) следует оценка

$$\mu(t) \leq \mathrm{const}(\lambda + |\psi(t)|) \quad \forall t \in \mathcal{T}. \tag{34}$$

Здесь константа const не зависит от t.

Предположим, что условие (31) нарушено. Тогда $\lambda = 0$, и существует точка $t_0 \in [0,T^*]$ такая, что $\psi(t_0) = \mu(t_0)\nabla g(x^*(t_0))$. Для $t \in [0,T^*]$ положим

$$\tilde{\mu}(t) := \mu(t) - \mu(t_0), \quad \tilde{\psi}(t) := \psi(t) - \mu(t_0)\nabla g(x^*(t)).$$

В силу замечания 4 новый набор множителей Лагранжа $(\lambda,\tilde{\psi},\tilde{\mu})$ снова удовлетворяет принципу максимума и (30), причем $\tilde{\mu}(t_0) = 0$, $\tilde{\psi}(t_0) = 0$. Для удобства переобозначим $\tilde{\mu}$, $\tilde{\psi}$ снова через μ и ψ соответственно.

Сначала рассмотрим случай $t_0 \in \mathcal{T}$. После переобозначений оценка (34) принимает следующий вид

$$|\mu(t)| \leq C_1|\psi(t)| \quad \forall t \in \mathcal{T}. \tag{35}$$

Здесь $C_1 > 0$ – некоторая константа.

В силу свойства Липшиц-непрерывности сопряженной функции ψ имеем

$$|\psi(t)| \leq K|t - t_0| \quad \forall t \in [0,T^*], \tag{36}$$

при некотором $K > 0$.

Докажем оценку

$$|\mu(t)| \leq C_1K|t - t_0| \quad \forall t \in [0,T^*]. \tag{37}$$

Действительно, при $t \in \mathcal{T}$ эта оценка вытекает из (35) и (36). Рассмотрим точку $t \in [0,T^*]$, $t > t_0$, такую, что $t \notin \mathcal{T}$. Поскольку множество $\mathcal{T}$ замкнуто, а $t_0 \in \mathcal{T}$, существует точка $t_* \in \mathcal{T}$, $t_0 \leq t_* < t$, такая, что $\mathcal{T} \cap (t_*,t] = \varnothing$. Для точки t_* оценка (37) уже доказана. В то же время из [7, предложение 4.8] вытекает постоянство функции μ на $[t_*,t]$. Эти рассуждения доказывают (37) справа от t_0. Точно такие же рассуждения проводятся и слева от точки t_0. Оценка (37) установлена.

Используя полученную оценку и (36), из сопряженного уравнения имеем

$$|\psi(t)| \leq KC_2(C_1 + 1)\int_{t_0}^{t}(s - t_0)ds, \quad t \in [0,T^*],$$

где $C_2 > 0$.

Отсюда

$$|\psi(t)| \le KC_2(C_1+1)\frac{(t-t_0)^2}{2} \quad \forall t \in [0,T^*]. \tag{38}$$

Применяя аналогичную (37) оценку, но используя (38) вместо (36), имеем

$$|\mu(t)| \le KC_2(C_1+1)C_1\frac{(t-t_0)^2}{2} \quad \forall t \in [0,T^*].$$

Из этой оценки и из (38), используя сопряженное уравнение, получаем

$$|\psi(t)| \le KC_2^2(C_1+1)^2\frac{|t-t_0|^3}{3!} \quad \forall t \in [0,T^*].$$

Продолжая такой итеративный процесс, на n-м шаге имеем

$$|\psi(t)| \le \frac{KC_2^n(C_1+1)^n}{(n+1)!}|t-t_0|^{n+1} \quad \forall t \in [0,T^*].$$

При этом

$$|\mu(t)| \le \frac{KC_2^n(C_1+1)^nC_1}{(n+1)!}|t-t_0|^{n+1} \quad \forall t \in [0,T^*].$$

Переход к пределу при $n \to \infty$ показывает, что $\psi = 0, \mu = 0$. Значит, все множители равны нулю, что противоречит условию нетривиальности (30).

Случай $t_0 \notin T$ легко свести к уже рассмотренному случаю из [7, предложение 4.8]. Условие (30) доказано.

Лемма 2. *Пусть* (x^*,u^*,T^*) *— оптимальный процесс в задаче* (2). *Предположим, что выполнено условие:* $0 \in \operatorname{int} f(x,U)$ $\forall x \in \mathbb{R}^n$. *Тогда* $\exists\delta > 0$:

$$\ell(t \in [0,T^*] : |f(x^*(t),u^*(t))| \le \delta) = 0.$$

Доказательство проводится стандартными рассуждениями от противного.

7. ЗАКЛЮЧЕНИЕ

В работе изучена задача управления мобильным роботом при наличии фазовых ограничений. В качестве критерия оптимальности рассмотрено быстродействие. На основе принципа максимума Понтрягина предложен алгоритм для вычисления решения. Однако прямое приложение принципа максимума не приводит к нужному результату, что обусловлено существенной нерегулярностью задачи относительно фазовых ограничений. Поэтому был предложен метод регуляризации задачи, который позволяет решить задачу приближенно с помощью ε-аппроксимаций. Этот метод применен к тестовой задаче о безынерционном движении трехколесного мобильного робота с передним приводом. Приведены результаты численного эксперимента. Настоящее исследование является завершением цикла работ [18]–[21] авторского коллектива по тематике задач быстродействия для мобильного робота при фазовых ограничениях.

Авторы выражают признательность рецензенту за сделанные замечания и интересный комментарий.

СПИСОК ЛИТЕРАТУРЫ

1. *Filippov A.F.* On certain problems of optimal regulation // Bull. Moscow State Univer., ser. Math. and Mech. 1959. V. 2. P. 25–38.
2. *Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V., Mishchenko E.F.* The Mathematical Theory of Optimal Processes, Interscience, New York, 1962.
3. *Gamkrelidze R.V.* Optimal control processes with restricted phase coordinates // Izv. Akad. Nauk SSSR Ser. Mat. 1960. V. 24, P. 315–356.
4. *Arutyunov A.V.* Optimality conditions: Abnormal and Degenerate Problems. Mathematics and Its Application. Kluwer Acad. Publ., 2000.

5. *Дубовицкий А.Я., Милютин А.А.* Необходимые условия слабого экстремума в задачах оптимального управления со смешанными ограничениями типа неравенства // Ж. вычисл. матем. и матем. физ. 1968. Т. 8. № 4. С. 725–779.
6. *Clarke F.H., Vinter R.B.* Optimal multiprocesses // SIAM J. Control Optim. 1989. V. 27. № 5. P. 1072–1091.
7. *Arutyunov A.V., Karamzin D.Yu.* On some continuity properties of the measure lagrange multiplier from the maximum principle for state constrained problems // SIAM J. Control Optim. 2015. V. 53. № 4. P. 2514–2540.
8. *Karamzin D., Pereira F.L.* On a few questions regarding the study of state-constrained problems in optimal control // J. Optim. Theor. Appl. 2019. V. 180. № 1.
9. *Chertovskih R., Karamzin D., Khalil N.T., Pereira F.L.* An indirect method for regular state-constrained optimal control problems in flow fields // IEEE Transact. Automat. Control. 2021. V. 66. № 2. P. 787–793.
10. *Арутюнов А.В., Карамзин Д.Ю.* Свойства экстремалей в задачах оптимального управления с фазовыми ограничениями // Дифференц. ур-ния. 2016. Т. 52. № 11. С. 1464–1476.
11. *Mordukhovich B.S.* Variational Analysis and Generalized Differentiation. Vol. I. Basic Theory. Springer-Verlag, 2006, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Berlin.
12. *Arutyunov A., Karamzin D.* A survey on regularity conditions for state-constrained optimal control problems and the non-degenerate maximum principle // J. Optim. Theor. Appl. 2020. V. 184. № 3. P. 697–723.
13. *Arutyunov A.V., Karamzin D.Yu., Pereira F.L.* The maximum principle for optimal control problems with state constraints by R.V. Gamkrelidze: Revisited // J. Optim. Theory Appl. 2011. V. 149.
14. *Arutyunov A.V., Karamzin D.Yu., Pereira F.L.* Optimal Impulsive Control. The Extension Approach. Springer, 2011.
15. *Chertovskih R., Karamzin D., Khalil N.T., Lobo Pereira F.* Regular path-constrained time-optimal control problems in three-dimensional flow fields // Europ. J. Control. 2020. V. 56. P. 98–106.
16. *Афанасьев А.П., Дикусар В.В., Милютин А.А., Чуканов С.А.* Необходимое условие в оптимальном управлении. М.: Наука, 1990. 318 с.
17. *Захаров Е.В., Карамзин Д.Ю.* К исследованию условий непрерывности меры-множителя Лагранжа в задачах с фазовыми ограничениями // Дифференц. ур-ния. 2015. Т. 51. № 3. С. 395–401.
18. *Дарьина А.Н., Дивеев А.И., Карамзин Д.Ю., Перейра Ф.Л., Софронова Е.А., Чертовских Р.А.* Исследование метода возмущений для решения нерегулярных задач оптимального управления с фазовыми ограничениями // Вопросы теории безопасности и устойчивости систем. 2020. № 22. С. 25–51.
19. *Дарьина А.Н., Дивеев А.И., Карамзин Д.Ю., Перейра Ф.Л., Софронова Е.А., Чертовских Р.А.* Регулярные возмущения оптимального движения трехколесного мобильного робота с передним приводом при ограниченных фазовых переменных // Вопросы теории безопасности и устойчивости систем. 2020. № 22. С. 52–77.
20. *Chertovskih R., Daryina A., Diveev A., Karamzin D., Pereira F.L., Sofronova E.* Regular perturbations to the motion of a three-wheeled mobile robot with the front-wheel drive under restricted state variables // Europ. Control Conf. (ECC 2020), St-Petersburg, May 2020. P. 1210–1215. article 9143809.
21. *Chertovskih R., Daryina A., Diveev A., Karamzin D., Pereira F.L., Sofronova E.* Investigation of a perturbation method to solve essentially non-regular time-optimal control problems with state constraints // Europ. Control Conf. (ECC 2020), St-Petersburg, May 2020. P. 849–854. article 9143703.