

Том 64, Номер 7

ISSN 0044-4669

Июль 2024



ФИЦ ИУ РАН

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ



НАУКА

— 1727 —

СОДЕРЖАНИЕ

Том 64, номер 7, 2024

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

- Операторно-разностные аппроксимации на нестандартных прямоугольных сетках
П. Н. Вабищевич 1097
- Регуляризация решения вырожденных систем алгебраических уравнений на примере идентификации ви-
риального уравнения состояния реального газа
А. Г. Видулов 1112
- MDM-алгоритм и задача Сильвестра
В. Н. Малозёмов, Н. А. Соловьева, Г. Ш. Тамасян 1128
- О новых классах решений задачи о σ -коммутировании ($\sigma \neq 0, \pm 1$) тёплицевой и ганкелевой матриц в рам-
ках унифицированного подхода
В. Н. Чугунов, Х. Д. Икрамов 1145

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

- Аппроксимация задач оптимального управления для полулинейных эллиптических уравнений
конвекции-диффузии с граничным наблюдением конормальной производной, с управлениями в коэф-
фициентах оператора конвективного переноса и нелинейном члене уравнения
Ф. В. Лубышев, М. Э. Файрузов 1163

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

- Применение универсальной быстрой тригонометрической интерполяции и экстраполяции для опреде-
ления координат точки запуска летательного аппарата
А. Д. Чернышов, О. Ю. Никифорова, В. В. Горяйнов, И. Г. Рукин 1183

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

- Дуализм в теории солитонных решений
Л. А. Бекларян, А. Л. Бекларян 1196
- Обобщение метода матриц рассеяния на задачи в нелинейных диспергирующих средах
А. А. Белов, Ж. О. Домбровская 1217

О равномерной сходимости аппроксимаций касательной и нормальной производных потенциала простого слоя вблизи границы двумерной области <i>Д. Ю. Иванов</i>	1233
Об однозначности определения сеточного фундаментального решения уравнения Лапласа в теории дискретного потенциала <i>И. Э. Степанова, И. И. Колотов, А. Г. Ягола, А. Н. Левашов</i>	1253
О численном решении уравнений Власова—Ампера <i>Е. В. Чижонков</i>	1268

МАТЕМАТИЧЕСКАЯ ФИЗИКА

KP_1 — схема ускорения внешних итераций по области термализации нейтронов и по источнику деления при решении подкритической краевой задачи <i>А. М. Волощенко</i>	1281
Сравнительный анализ влияния поверхностных квантовых эффектов на оптические характеристики наночастиц щелочных и благородных металлов <i>Ю. А. Еремин, В. В. Лопушенко</i>	1305
Численная диагностика разрушения решения одной тепло-электрической модели полупроводника <i>М. О. Корпусов, Р. С. Шафир, А. К. Матвеева</i>	1314
Использование весовых схем Монте-Карло для слабоионизованных разреженных течений газа <i>А. А. Шевырин, Е. А. Бондарь</i>	1323

УДК 519.63

ОПЕРАТОРНО-РАЗНОСТНЫЕ АППРОКСИМАЦИИ НА НЕСТАНДАРТНЫХ ПРЯМОУГОЛЬНЫХ СЕТКАХ¹⁾

© 2024 г. П. Н. Вабищевич^{1,2,*}

¹119991 Москва, Ленинские горы, МГУ им. М. В. Ломоносова, Россия

²677007 Якутск, ул. Кулаковского, 42, Северо-Восточный федеральный университет им. М. К. Аммосова, Россия

*e-mail: vabishchevich@gmail.com

Поступила в редакцию 07.02.2024 г.

Переработанный вариант 03.03.2024 г.

Принята к публикации 05.03.2024 г.

При приближенном решении краевых задач для уравнений с частными производными широко используются разностные методы. Наиболее просто строятся сеточные аппроксимации при разбиении расчетной области на прямоугольные ячейки. Обычно узлы сетки совпадают с вершинами ячеек. Помимо таких узловых аппроксимаций применяются также сетки с узлами в центрах ячеек. Краевые задачи удобно формулировать в терминах инвариантных операторов векторного (тензорного) анализа, которым сопоставляются соответствующие сеточные аналоги. В работе строятся аналоги операторов градиента и дивергенции на нестандартных прямоугольных сетках, узлы которых состоят как из вершин расчетных ячеек, так и их центров. Предложенный подход иллюстрируется аппроксимациями краевой задачи для стационарного двумерного уравнения конвекции-диффузии. Отмечены ключевые особенности построения аппроксимаций для векторных задач при ориентации на прикладные задачи механики твердого тела. Библ. 20. Фиг. 6.

Ключевые слова: прямоугольная расчетная сетка, операторы векторного анализа, сеточные операторы, операторно-разностные схемы, уравнение конвекции-диффузии.

DOI: 10.31857/S0044466924070016, EDN: xjilpp

ВВЕДЕНИЕ

Традиционная технология численного решения краевых задач для уравнений с частными производными связана с использованием сеточных (разностных) методов [1, 2]. В методах этого класса ищется приближенное решение в узлах расчетной сетки. Переход от непрерывной задачи к дискретной проводится различными способами. Простейший вариант (непосредственная аппроксимация) связан с прямой заменой дифференциальных выражений в уравнениях и граничных условиях на соответствующие сеточные выражения. Более полно специфика рассматриваемых задач учитывается при применении метода баланса (интегроинтерполяционного метода) [3, 4], когда сеточная задача рассматривается как интегральное следствие (закон сохранения) для отдельных частей расчетной области. Такие аппроксимации особенно уместны при использовании непрямоугольных сеток. В настоящее время в англоязычной литературе такой подход носит название метод конечных объемов (см., например, [5], [6]) и часто рассматривается как самостоятельный класс методов, наряду с методом конечных элементов. К разностным при этом относят методы с использованием сеточных аппроксимаций на прямоугольных сетках.

Приближенное решение удобно строить на основе формулировки краевой задачи с применением операторов векторного (тензорного) анализа, прежде всего, операторов градиента и дивергенции. Такие операторы инвариантны относительно выбора системы координат, что позволяет использовать единый подход для построения дискретных задач на произвольных расчетных сетках. Сама сеточная краевая задача строится с привлечением общих построенных дискретных аналогов операторов

¹⁾ Работа выполнена при финансовой поддержке РНФ (проект № 24-11-00058)

векторного анализа. Сеточные операторы векторного анализа должны наследовать основные свойства дифференциальных операторов. При выборе соответствующих гильбертовых пространств сеточных функций основное внимание уделяется условиям сопряженности операторов, выполнению основных интегральных соотношений с операторами векторного анализа. Такие согласованные аппроксимации являются основой MD (Mimetic Discretization) технологии [7, 8], которая развивается в различных направлениях.

Согласованные аппроксимации операторов векторного анализа относительно легко строятся на прямоугольных сетках. В этой связи отметим пионерскую работу [9], более поздние результаты отражены в книге [10]. При использовании произвольных сеток успех достигается при применении метода опорных операторов [11, 12]. Наиболее просто сеточные аналоги операторов векторного анализа на нерегулярных расчетных сетках получаются при использовании триангуляций Делоне и разбиений Вороного [13].

В настоящей работе мы будем рассматривать базовые двумерные эллиптические краевые задачи в прямоугольнике, когда уравнение включает операторы градиента и дивергенции. В расчетной области, которая разбивается на прямоугольные ячейки, сетку можно ввести различными способами. Стандартный вариант связан с выбором вершин ячеек в качестве узлов расчетной сетки. Второй хорошо известный вариант базируется на задании узлов сетки как центров ячеек. Построение разностных аппроксимаций при одновременном использовании этих двух расчетных сеток предложено в работах [14, 15]. В других работах Фрязинова И. В. (см., например, [16, 17]) такая двухсеточная технология применена при рассмотрении различных краевых задач для эллиптических уравнений второго порядка. Сеточная задача рассматривается на прямой сумме соответствующих гильбертовых пространств сеточных функций, определенных на двух сетках.

Мы строим операторно-разностные аппроксимации на объединенной прямоугольной сетке. Основное внимание уделяется согласованным аппроксимациям операторов градиента и дивергенции. Основные особенности использования сеточных операторов векторного анализа проиллюстрированы на краевой задаче для уравнения конвекции-диффузии со смешанными производными. Отмечены также ключевые моменты построения разностных схем на объединенной сетке для векторных эллиптических уравнений второго порядка.

1. СЕТКИ И СЕТОЧНЫЕ ФУНКЦИИ

Аппроксимации по пространству при использовании прямоугольных расчетных сеток будем рассматривать на примере двумерных краевых задач, когда расчетная область Ω — прямоугольник с границей $\partial\Omega$:

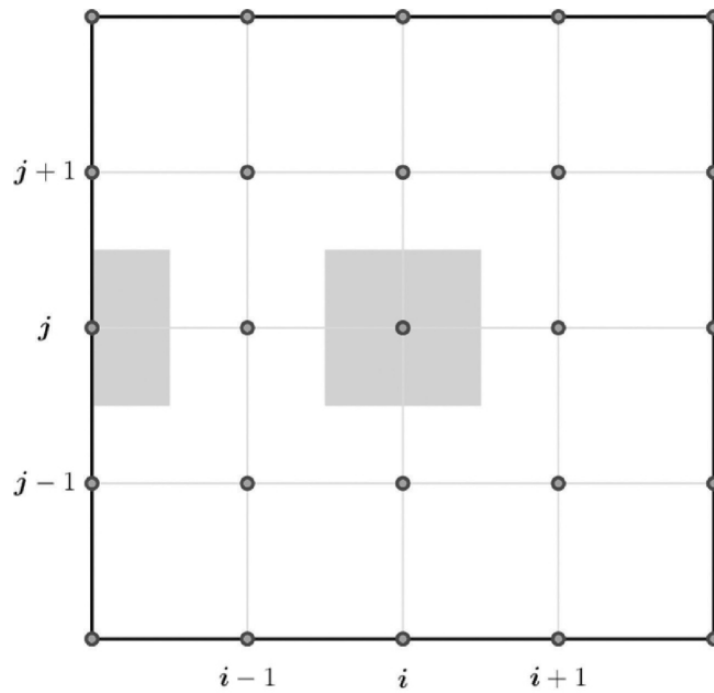
$$\Omega = \left\{ \mathbf{x} \mid \mathbf{x} = \left(x^{(1)}, x^{(2)} \right), 0 < x^{(\alpha)} < l_{\alpha}, \alpha = 1, 2 \right\}.$$

Будем рассматривать неравномерные прямоугольные сетки, когда Ω разбивается на ячейки линиями $x^{(\alpha)} = \text{const}$, $\alpha = 1, 2$:

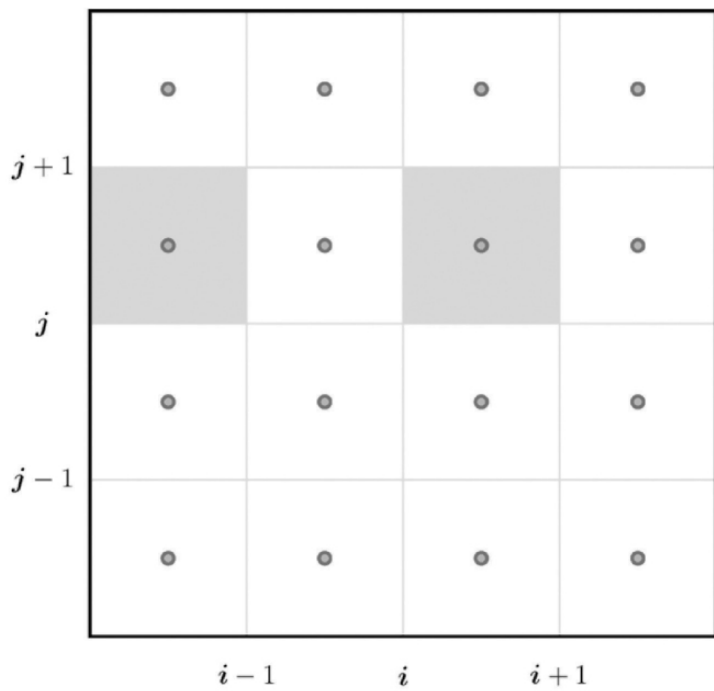
$$\begin{aligned} x^{(1)} &= x_i^{(1)} = x_{i-1}^{(1)} + h_i^{(1)}, \quad i = 1, 2, \dots, N_1, \quad x_0^{(1)} = 0, \quad x_{N_1}^{(1)} = l_1, \\ x^{(2)} &= x_j^{(2)} = x_{j-1}^{(2)} + h_j^{(2)}, \quad j = 1, 2, \dots, N_2, \quad x_0^{(2)} = 0, \quad x_{N_2}^{(2)} = l_2. \end{aligned}$$

Узлы сетки связываются с ячейками. Стандартные разностные аппроксимации связаны с совпадающими с вершинами ячеек узлами сетки — узловая (node-center) аппроксимация (фиг. 1). Второй используемый класс расчетных сеток характеризуется тем, что узлами сетки являются центры ячеек — центроидная (cell-center) аппроксимация (фиг. 2).

Будем использовать объединенную сетку, узлы которой состоят как из вершин ячеек, так и их центров (фиг. 3). Подобные сетки введены в вычислительную практику в обзорной работе Самарского А. А. и Фрязинова И. В. [15]. Более подробное рассмотрение вопросов аппроксимации краевых задач для скалярных эллиптических и параболических уравнений второго порядка выполнено, например, в [16, 17]. При выделении приближенного решения на каждом отдельных (исходной узловой и сопряженной центроидной) сетках задача рассматривалась на гильбертовом пространстве сеточных векторных функций — прямой сумме сеточных гильбертовых пространств скалярных функций.

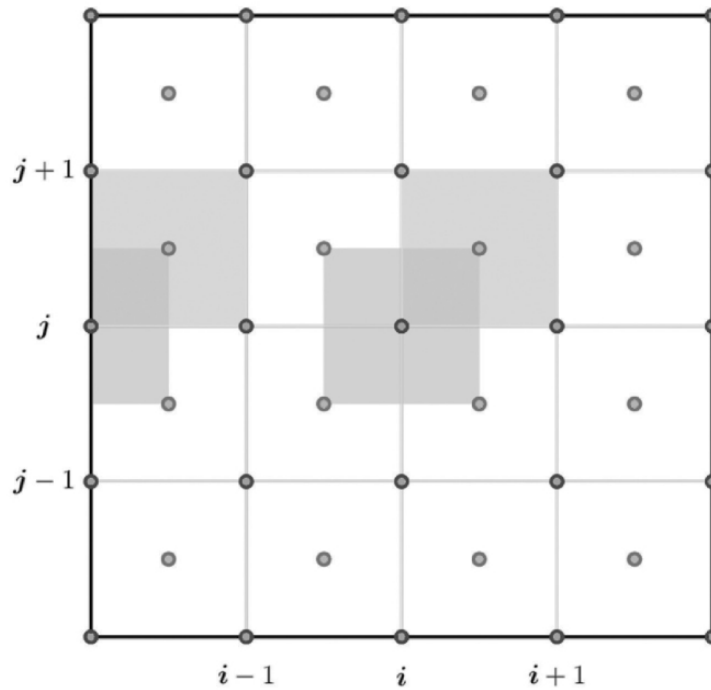


Фиг. 1. Узлы сетки и контрольные объемы при узловой аппроксимации.



Фиг. 2. Узлы сетки и контрольные объемы при центроидной аппроксимации.

Мы не будем усложнять дискретную задачу и будем работать с сеточными скалярными функциями на объединенной сетке.



Фиг. 3. Узлы и контрольные объемы для объединенной сетки.

Для функций $u(\mathbf{x})$, $\mathbf{x} \in \Omega$, зададим гильбертово пространство $\mathcal{H} = L_2(\Omega)$ со скалярным произведением и нормой

$$(u, v) = \int_{\Omega} u(\mathbf{x})v(\mathbf{x}) d\mathbf{x}, \quad \|u\| = (u, u)^{1/2}.$$

Аналогично вводятся гильбертовы пространства сеточных функций на введенных сетках.

При узловых аппроксимациях (фиг. 1) сеточная функции y_{ij} задается в узлах сетки ω^n :

$$\omega^n = \left\{ \mathbf{x} = \mathbf{x}_{ij} \mid \mathbf{x}_{ij} = \left(x_i^{(1)}, x_j^{(2)} \right), i = 0, 1, \dots, N_1, j = 0, 1, \dots, N_2 \right\}.$$

С отдельным узлом $\mathbf{x}_{ij}$ мы связываем часть расчетной области Ω_{ij} . В частности, для внутренних узлов $i = 1, 2, \dots, N_1 - 1$, $j = 1, 2, \dots, N_2 - 1$ имеем

$$\Omega_{ij} = \left\{ \mathbf{x} \mid \mathbf{x} = \left(x^{(1)}, x^{(2)} \right), x_{i-1/2}^{(1)} < x^{(1)} < x_{i+1/2}^{(1)}, x_{j-1/2}^{(2)} < x^{(2)} < x_{j+1/2}^{(2)} \right\},$$

при использовании обозначений

$$x_{i+1/2}^{(1)} = \frac{1}{2} \left(x_{i+1}^{(1)} + x_i^{(1)} \right), i = 0, 1, \dots, N_1 - 1, \quad x_{j+1/2}^{(2)} = \frac{1}{2} \left(x_{j+1}^{(2)} + x_j^{(2)} \right), j = 0, 1, \dots, N_2 - 1.$$

Для узлов на границе расчетной области (см. фиг. 1) части расчетной области выделяются аналогично. Для сеточных функций $y(\mathbf{x})$, $\mathbf{x} \in \omega^n$ пространству $\mathcal{H} = L_2(\Omega)$ сопоставим конечномерное гильбертово пространство $H(\omega^n)$, в котором

$$(y, v)_n = \sum_{\mathbf{x} \in \omega^n} y(\mathbf{x}) v(\mathbf{x}) V^n(\mathbf{x}), \quad \|y\|_n = (y, y)_n^{1/2},$$

где

$$V^n(\mathbf{x}) = \text{meas}(\Omega_{ij}), \quad \mathbf{x} = \mathbf{x}_{ij} \in \omega^n.$$

При центроидных аппроксимациях (фиг. 2) сеточная функции $y_{i+1/2,j+1/2}$ задается в узлах сетки ω^c :

$$\omega^c = \left\{ \mathbf{x} = \mathbf{x}_{i+1/2,j+1/2} \mid \mathbf{x}_{i+1/2,j+1/2} = \left(x_{i+1/2}^{(1)}, x_{j+1/2}^{(2)} \right), i = 0, 1, \dots, N_1 - 1, j = 0, 1, \dots, N_2 - 1 \right\}.$$

С узлом $\mathbf{x}_{i+1/2,j+1/2}$ связывается ячейка расчетной области $\Omega_{i+1/2,j+1/2}$:

$$\Omega_{i+1/2,j+1/2} = \left\{ \mathbf{x} \mid \mathbf{x} = \left(x^{(1)}, x^{(2)} \right), x_i^{(1)} < x^{(1)} < x_{i+1}^{(1)}, x_j^{(2)} < x^{(2)} < x_{j+1}^{(2)} \right\},$$

$$V^c(\mathbf{x}) = \text{meas}(\Omega_{i+1/2,j+1/2}), \quad \mathbf{x} = \mathbf{x}_{i+1/2,j+1/2} \in \omega^c.$$

Для сеточных функций $y(\mathbf{x})$, $\mathbf{x} \in \omega^c$, зададим конечномерное гильбертово пространство $H(\omega^c)$, в котором

$$(y, v)_c = \sum_{\mathbf{x} \in \omega^c} y(\mathbf{x}) v(\mathbf{x}) V^c(\mathbf{x}), \quad \|y\|_c = (y, y)_c^{1/2}.$$

Для объединенной сетки имеем $\omega = \omega^n \cup \omega^c$. Для узлов этой сетки будем использовать обозначения $\mathbf{x}_k$, $k = 0, 1, \dots, N_1 + N_2 + 2N_1N_2$. С отдельным узлом связывается часть области

$$\Omega_k = \frac{1}{2} \begin{cases} \Omega_{i,j}, & \mathbf{x}_k \in \omega^n, \\ \Omega_{i+1/2,j+1/2}, & \mathbf{x}_k \in \omega^c, \end{cases} \quad V(\mathbf{x}) = \frac{1}{2} \begin{cases} V^n(\mathbf{x}), & \mathbf{x} \in \omega^n, \\ V^c(\mathbf{x}), & \mathbf{x} \in \omega^c. \end{cases}$$

Определим пространство $H(\omega)$ со скалярным произведением и нормой

$$(y, v) = \sum_{\mathbf{x} \in \omega} y(\mathbf{x}) v(\mathbf{x}) V(\mathbf{x}), \quad \|y\| = (y, y)^{1/2},$$

для сеточных функций $y(\mathbf{x})$, $\mathbf{x} \in \omega$, на объединенной сетке.

2. СЕТОЧНЫЕ ОПЕРАТОРЫ

Разностные аппроксимации краевой задачи (1), (4) будем проводить на основе метода баланса (интегроинтерполяционного метода) [1], метода конечного объема [5, 6]. Мы приходим к дискретной задаче для сеточных функций при интегрировании исходного уравнения по соответствующему конечному объему, при записи закона сохранения (баланса) на дискретном уровне. Для узлов $\mathbf{x} = \mathbf{x}_{ij}$ сетки ω^n в качестве контрольных объемов выступают Ω_{ij} , а для $\mathbf{x} = \mathbf{x}_{i+1/2,j+1/2} \in \omega^c$ — $\Omega_{i+1/2,j+1/2}$.

Мы ориентируемся на формулировку исследуемых проблем с использованием инвариантных (не зависящих от выбора системы координат) операторов векторного (тензорного) анализа. В технологии VAGO (Vector Analysis Grid Operator) [13] мы строим дискретные аналоги таких операторов. При применении MD (Mimetic Discretization) подхода [7, 8] мы отслеживаем основные свойства дискретных операторов векторного анализа. Это особенно важно при аппроксимациях на общих нерегулярных расчетных сетках, когда покоординатное представление не имеет большого значения.

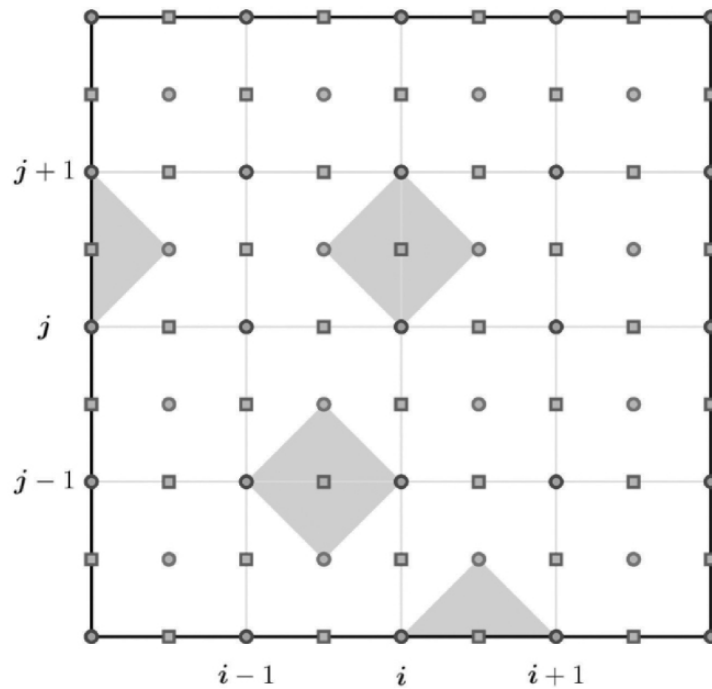
Помимо скалярных сеточных функций $y(\mathbf{x})$, $\mathbf{x} \in \omega$, будем рассматривать векторные функции на своей сетке ω^* с узлами $\mathbf{x}_m \in \omega^*$, $m = 0, 1, \dots, 2N_1N_2 + N_1 + N_2 - 1$. Узлы этой сетки расположены на пересечении граней контрольных объемов для узлов сеток ω^n и ω^c :

$$\omega^* = \left\{ \mathbf{x} = \mathbf{x}_m \mid \mathbf{x}_m = (x_i^{(1)}, x_{j+1/2}^{(2)}), i = 0, 1, \dots, N_1, j = 0, 1, \dots, N_2 - 1, \right. \\ \left. (x_{i+1/2}^{(1)}, x_j^{(2)}), i = 0, 1, \dots, N_1 - 1, j = 0, 1, \dots, N_2 \right\}.$$

Контрольные объемы (фиг. 4) есть четырехугольники с ортогональными диагоналями. Такой ортогональный четырехугольник, связанный с отдельным узлом $\mathbf{x}_m \in \omega^*$, обозначим через Ω_m^* , $m = 0, 1, \dots, 2N_1N_2 + N_1 + N_2 - 1$.

Для векторных сеточных функций $\mathbf{y}(\mathbf{x}) \in \omega^*$ определим гильбертово пространство $\mathbf{H}(\omega^*)$ со скалярным произведением и нормой

$$(\mathbf{y}, \mathbf{v}) = \sum_{\mathbf{x} \in \omega^*} \mathbf{y}(\mathbf{x}) \mathbf{v}(\mathbf{x}) V^*(\mathbf{x}), \quad \|\mathbf{y}\| = (\mathbf{y}, \mathbf{y})^{1/2},$$



Фиг. 4. Контрольные объемы для векторных величин и узлы сеток ω и ω^* .

где

$$V^*(\mathbf{x}) = \text{meas}(\Omega_m^*), \quad \mathbf{x} = \mathbf{x}^* \in \omega^*.$$

Рассмотрение сеточных операторов векторного анализа мы начнем с оператора градиента. Для приближенного вычисления $\mathbf{v} = \text{grad } u$ для заданной функции $u(\mathbf{x})$, $\mathbf{x} \in \Omega$, в узлах сетки ω^* положим

$$\tilde{\mathbf{v}}(\mathbf{x}^*) = \text{grad}_h y(\mathbf{x}), \quad \mathbf{x} \in \omega, \quad \mathbf{x}^* \in \omega^*, \quad (1)$$

где сеточный оператор градиента $\text{grad}_h : H(\omega) \rightarrow \mathbf{H}(\omega^*)$. Для внутренних узлов ($\mathbf{x}_m^* \in \Omega$) сетки ω^* градиент $\tilde{\mathbf{v}}(\mathbf{x}) = \{\tilde{v}_1(\mathbf{x}), \tilde{v}_2(\mathbf{x})\}$ вычисляется по правилу

$$\tilde{v}_1(\mathbf{x}_m^*) = \begin{cases} \frac{2}{h_{i+1}^{(1)} + h_i^{(1)}} (y(\mathbf{x}_{i+1/2, j+1/2}) - y(\mathbf{x}_{i-1/2, j+1/2})), & \mathbf{x}_m^* = (x_i^{(1)}, x_{j+1/2}^{(2)}), \\ \frac{1}{h_i^{(1)}} (y(\mathbf{x}_{i+1, j}) - y(\mathbf{x}_{i, j})), & \mathbf{x}_m^* = (x_{i+1/2}^{(1)}, x_j^{(2)}), \end{cases} \quad (2)$$

$$\tilde{v}_2(\mathbf{x}_m^*) = \begin{cases} \frac{1}{h_j^{(2)}} (y(\mathbf{x}_{i, j+1}) - y(\mathbf{x}_{i, j})), & \mathbf{x}_m^* = (x_i^{(1)}, x_{j+1/2}^{(2)}), \\ \frac{2}{h_{j+1}^{(2)} + h_j^{(2)}} (y(\mathbf{x}_{i+1/2, j+1/2}) - y(\mathbf{x}_{i+1/2, j-1/2})), & \mathbf{x}_m^* = (x_{i+1/2}^{(1)}, x_j^{(2)}). \end{cases}$$

Для сеточного оператора градиента в узлах сетки ω^* на границе $\partial\Omega$ можно использовать расчетные формулы (2), считая, что

$$\mathbf{x}_{-1/2, j+1/2} = (0, x_{j+1/2}^{(2)}), \quad \mathbf{x}_{N_1+1/2, j+1/2} = (l_1, x_{j+1/2}^{(2)}), \quad i = 0, 1, \dots, N_1 - 1,$$

$$\mathbf{x}_{i+1/2, -1/2} = (x_{i+1/2}^{(1)}, 0), \quad \mathbf{x}_{i+1/2, N_2+1/2} = (x_{i+1/2}^{(1)}, l_2), \quad j = 0, 1, \dots, N_2 - 1.$$

Тем самым нормальная компонента градиента аппроксимируется по значениям в приграничных узлах сетки ω^c и на границе области.

Аппроксимации (1), (2) можно получить, отталкиваясь от интегрального определения градиента. Для контрольного объема Ω_m^* имеем

$$\int_{\Omega_m^*} \text{grad } u(\mathbf{x}) d\mathbf{x} = \int_{\partial\Omega_m^*} u(\mathbf{x}) \mathbf{n} d\mathbf{x}, \quad (3)$$

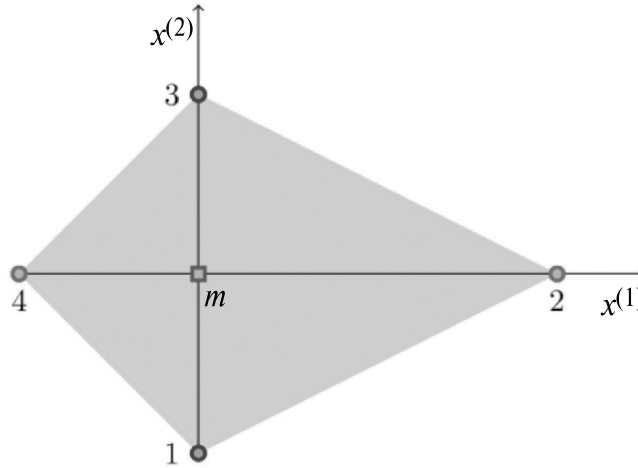
где $\mathbf{n}$ — внешняя нормаль к $\partial\Omega_m^*$. С учетом этого сеточный оператор градиента в узле $\mathbf{x}_m \in \omega^*$ определяется согласно

$$(\text{grad}_h y)(\mathbf{x}_m) = \frac{1}{V^*(\mathbf{x})} \int_{\partial\Omega_m^*} \tilde{y}(\mathbf{x}) \mathbf{n} d\mathbf{x}, \quad (4)$$

где $\tilde{y}(\mathbf{x})$ — кусочно-линейная интерполяция на частях границы контрольного объема $\partial\Omega_m^*$ по значениям сеточной функции $y(\mathbf{x})$, $\mathbf{x} \in \omega$. Тем самым для интегрирования по контуру в правой части (3) мы используем квадратурную формулу трапеций. При использовании обозначений фиг. 5 из (4) получим

$$(\text{grad}_h y)(\mathbf{x}_m) = \left\{ \frac{y(\mathbf{x}_2) - y(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|}, \frac{y(\mathbf{x}_3) - y(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|} \right\}. \quad (5)$$

Для $\mathbf{x}_m \in \partial\Omega$ аппроксимация градиента выполняется, когда контрольный объем Ω_m^* вырождается в треугольник.



Фиг. 5. К аппроксимации градиента.

Аналогичный подход будем использовать при построении сеточного оператора дивергенции. По заданному векторному полю $\mathbf{v}(\mathbf{x})$ в узлах $\mathbf{x} \in \omega^*$ приближенно вычисляется $y(\mathbf{x}) = \text{div } \mathbf{v}$ в узлах $\mathbf{x} \in \omega$.

Применение теоремы о дивергенции к контрольному объему Ω_k для узла $\mathbf{x}_k \in \omega$ дает

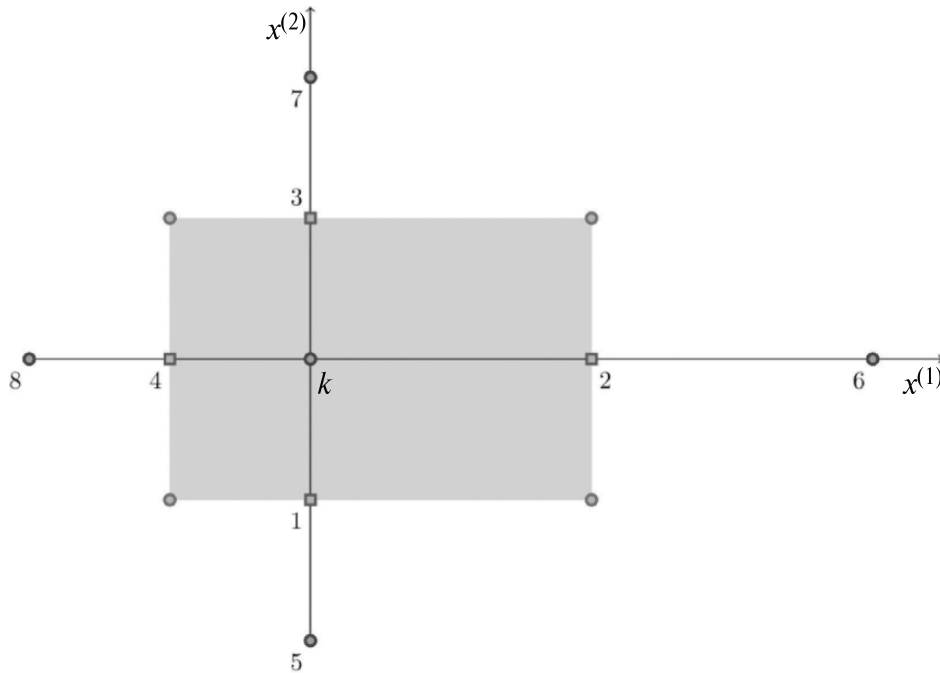
$$\int_{\Omega_k} \text{div } \mathbf{v}(\mathbf{x}) d\mathbf{x} = \int_{\partial\Omega_k} \mathbf{v} \cdot \mathbf{n} d\mathbf{x}. \quad (6)$$

Сеточный оператор дивергенции $\text{div}_h : \mathbf{H}(\omega^*) \rightarrow H(\omega)$ с учетом (6) определим согласно

$$(\text{div}_h \mathbf{w})(\mathbf{x}_k) = \frac{1}{2V(\mathbf{x})} \int_{\partial\Omega_k} \tilde{\mathbf{w}} \cdot \mathbf{n} d\mathbf{x}, \quad (7)$$

где $\tilde{\mathbf{w}}(\mathbf{x})$ — кусочно-постоянная интерполяция сеточной векторной функции $\mathbf{w}(\mathbf{x})$, $\mathbf{x} \in \omega^*$, на частях границы контрольного объема $\partial\Omega_k$. Конкретное представление сеточного оператора дивергенции в обозначениях фиг. 6 запишется в виде

$$(\text{div}_h \mathbf{w})(\mathbf{x}_k) = \frac{w_1(\mathbf{x}_2) - w_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{w_2(\mathbf{x}_3) - w_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}. \quad (8)$$



Фиг. 6. К аппроксимации дивергенции.

Тем самым мы имеем стандартную аппроксимацию для

$$\operatorname{div} \mathbf{w} = \frac{\partial w_1}{\partial x^{(1)}} + \frac{\partial w_2}{\partial x^{(2)}}$$

на неравномерной прямоугольной сетке. Аналогично (2) можно ориентироваться на запись

$$(\operatorname{div}_h \mathbf{w})(\mathbf{x}_k) = \begin{cases} 2 \frac{w_1(\mathbf{x}_{i+1/2,j}) - w_1(\mathbf{x}_{i-1/2,j})}{h_{i+1}^{(1)} + h_i^{(1)}} + 2 \frac{w_2(\mathbf{x}_{i,j+1/2}) - w_2(\mathbf{x}_{i,j-1/2})}{h_{j+1}^{(2)} + h_j^{(2)}}, \\ \mathbf{x}_k = (x_i^{(1)}, x_j^{(2)}), \\ \frac{w_1(\mathbf{x}_{i+1,j+1/2}) - w_1(\mathbf{x}_{i,j+1/2})}{h_{i+1}^{(1)}} + \frac{w_2(\mathbf{x}_{i+1/2,j+1}) - w_2(\mathbf{x}_{i+1/2,j})}{h_{j+1}^{(2)}}, \\ \mathbf{x}_k = (x_{i+1/2}^{(1)}, x_{j+1/2}^{(2)}) \end{cases} \quad (9)$$

при сформулированных ранее предположениях об узлах вне расчетной области Ω .

МД технологии связываются с согласованностью аппроксимаций градиента и дивергенции. Для достаточно гладких функций непрерывного аргумента имеет место ключевое равенство

$$\operatorname{div}(u\mathbf{v}) = \mathbf{v} \cdot \operatorname{grad} u + u \operatorname{div} \mathbf{v}.$$

Из него следует, что для функций $u(\mathbf{x}) = 0$, $\mathbf{x} \in \partial\Omega$

$$\int_{\Omega} \mathbf{v} \cdot \operatorname{grad} u \, d\mathbf{x} = - \int_{\Omega} u \operatorname{div} \mathbf{v} \, d\mathbf{x}.$$

При выполнении такого соотношения на сеточном уровне

$$(\mathbf{w}, \operatorname{grad}_h y) = -(y, \operatorname{div}_h \mathbf{w}) \quad (10)$$

мы говорим о согласованной аппроксимации градиента и дивергенции. В условиях (10) для $y(\mathbf{x}) \in H_0(\omega)$, $\mathbf{w}(\mathbf{x}) \in \mathbf{H}(\omega^*)$ мы имеем

$$\operatorname{grad}_h^* = -\operatorname{div}_h, \quad (11)$$

где

$$H_0(\omega) = \{y(\mathbf{x}) \mid y(\mathbf{x}) \in H(\omega), \quad y(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega\}.$$

Равенство (10) доказывается прямыми вычислениями при, например, использовании представлений (2) и (9). Из (11) сразу же следует, что сеточный оператор Лапласа $\text{div}_h \text{grad}_h : H_0(\omega) \rightarrow H_0(\omega)$ является самосопряженным и отрицательным.

3. КРАЕВАЯ ЗАДАЧА ДЛЯ УРАВНЕНИЯ КОНВЕКЦИИ-ДИФфуЗИИ

Вопросы построения приближенных решений на объединенной сетке мы проиллюстрируем на примере краевой задачи для эллиптического уравнения со смешанными производными. Искомое решение $u(\mathbf{x})$, $\mathbf{x} \in \Omega$, является решением уравнения конвекции-диффузии

$$-\sum_{\alpha,\beta=1}^2 \frac{\partial}{\partial x_\alpha} \left(k_{\alpha\beta}(\mathbf{x}) \frac{\partial u}{\partial x_\beta} \right) + \sum_{\alpha=1}^2 c_\alpha(\mathbf{x}) \frac{\partial u}{\partial x_\alpha} = f(\mathbf{x}), \quad \mathbf{x} \in \Omega. \quad (12)$$

Будем считать, что для коэффициентов диффузии выполнены условия симметричности: $k_{\alpha\beta} = k_{\beta\alpha}$, $\alpha, \beta = 1, 2$. Уравнение (12) рассматривается в предположении равномерной эллиптичности, когда

$$\kappa_1 \sum_{\alpha=1}^2 \xi_\alpha^2 \leq \sum_{\alpha,\beta=1}^2 k_{\alpha\beta}(\mathbf{x}) \xi_\alpha \xi_\beta \leq \kappa_2 \sum_{\alpha=1}^2 \xi_\alpha^2, \quad \kappa_1 > 0, \quad (13)$$

для произвольных ξ_α , $\alpha = 1, 2$. Конвективный перенос в уравнении (12) определяется вектором $\mathbf{c}(\mathbf{x}) = \{c_1(\mathbf{x}), c_2(\mathbf{x})\}$. Будем считать поле скорости бездивергентным, так что

$$\frac{\partial c_1}{\partial x_1} + \frac{\partial c_2}{\partial x_2} = 0. \quad (14)$$

Для простоты, граничные условия возьмем в виде

$$u(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega. \quad (15)$$

На множестве функций, удовлетворяющих (15), определим дифференциальный оператор $\mathcal{D}$ выражением

$$\mathcal{D}u = - \sum_{\alpha,\beta=1}^2 \frac{\partial}{\partial x_\alpha} \left(k_{\alpha\beta}(\mathbf{x}) \frac{\partial u}{\partial x_\beta} \right), \quad x \in \Omega.$$

Удобно записать оператор диффузии $\mathcal{D}$ с использованием операторов векторного анализа:

$$\mathcal{D}u = - \text{div} (K(\mathbf{x}) \text{grad} u), \quad (16)$$

где $K(\mathbf{x})$ — симметричная матрица с элементами $k_{\alpha\beta}(\mathbf{x})$, $\alpha, \beta = 1, 2$. В $\mathcal{H}$ оператор $\mathcal{D}$ является самосопряженным и положительно-определенным:

$$\mathcal{D} = \mathcal{D}^* \geq \delta \mathcal{I}, \quad \delta = \pi^2 \kappa_1 (l_1^{-2} + l_2^{-2}), \quad (17)$$

где κ_1 — постоянная в (13), а $\mathcal{I}$ — единичный оператор.

Оператор конвективного переноса в уравнении (12) имеет недивергентную (характеристическую) форму [18]

$$\mathcal{C}_1 u = \sum_{\alpha=1}^2 c_\alpha(\mathbf{x}) \frac{\partial u}{\partial x_\alpha}, \quad x \in \Omega,$$

тем самым

$$\mathcal{C}_1 u = \mathbf{c}(\mathbf{x}) \cdot \text{grad} u. \quad (18)$$

Принимая во внимание условие (14), оператор конвективного переноса можно записать в дивергентной (консервативной) форме

$$\mathcal{C}_2 u = \text{div}(\mathbf{c}(\mathbf{x})u). \quad (19)$$

На множестве функций (15) для операторов $\mathcal{C}_\alpha$, $\alpha = 1, 2$, справедливо равенство

$$\mathcal{C}_1^* = -\mathcal{C}_2. \quad (20)$$

Для оператора конвективного переноса в симметричной форме имеем

$$\mathcal{C}_0 u = \frac{1}{2} (\mathbf{c}(\mathbf{x}) \cdot \text{grad } u + \text{div}(\mathbf{c}(\mathbf{x})u)). \quad (21)$$

С учетом (21) получим

$$\mathcal{C}_0 = \frac{1}{2} (\mathcal{C}_1 + \mathcal{C}_2), \quad \mathcal{C}_0 = -\mathcal{C}_0^*. \quad (22)$$

Отметим, что ключевое свойство кососимметричности оператора $\mathcal{C}_0$ справедливо и для векторов $\mathbf{c}(x)$, для которых условие (14) не выполнено.

Принимая во внимание введенные обозначения, краевую задачу (12), (15) запишем в виде операторного уравнения

$$\mathcal{D}u + \mathcal{C}u = f, \quad (23)$$

где $\mathcal{C} = \mathcal{C}_0$ для всех $\mathbf{c}(\mathbf{x})$ и $\mathcal{C} = \mathcal{C}_1, \mathcal{C}_2$ — при выполнении (14). Для решения уравнения (23) с учетом (17), (22) имеем априорную оценку

$$\|u\| \leq \frac{1}{\delta} \|f\|. \quad (24)$$

При переходе от (12)–(15) к дискретной задаче мы хотим наследовать подобные свойства непрерывной задачи.

При аппроксимации краевой задачи для эллиптического уравнения (12)–(15) дифференциальным оператором диффузии $\mathcal{D}$ и конвекции $\mathcal{C}$ сопоставляются сеточные операторы D и C . Ищется приближенное решение $y(\mathbf{x}) \in H_0$ из уравнения

$$Dy + Cy = \varphi, \quad (25)$$

где, в простейшем случае $f(\mathbf{x}) = 0$, $\mathbf{x} \in \partial\Omega$, положим $\varphi(\mathbf{x}) = f(\mathbf{x})$, $\mathbf{x} \in \omega$.

С учетом (16) для сеточного оператора диффузии используется представление

$$(Dy)(\mathbf{x}_k) = -\text{div}_h (K(\mathbf{x}) \text{grad}_h y), \quad \mathbf{x}_k \in \omega \cap \Omega. \quad (26)$$

Принимая во внимание (11), для $y(\mathbf{x}), v(\mathbf{x}) \in H_0(\omega)$ имеем

$$(Dy, v) = (K(\mathbf{x}) \text{grad}_h y, \text{grad}_h v).$$

Это равенство и симметричность $K(\mathbf{x})$ обеспечивает самосопряженность сеточного оператора диффузии. При ограничениях (13) получим

$$(Dy, y) \geq \kappa_1 \|\text{grad}_h y\|^2. \quad (27)$$

Положительная определенность оператора D следует из сеточного аналога неравенства Фридрихса

$$\|\text{grad}_h y\|^2 \geq M \|y\|^2 \quad (28)$$

для $y(\mathbf{x}) \in H_0(\omega)$ с постоянной M , которая не зависит от сетки. Это неравенство устанавливается также, как это делается для обычных прямоугольных неравномерных сеток (см., например, [19]). Из (27), (28) следует сеточный аналог (17)

$$D = D^* \geq \delta_h I, \quad \delta_h = \kappa_1 M. \quad (29)$$

Построение сеточных аналогов операторов конвективного переноса удобно начать с аппроксимации оператора $\mathcal{C}_2$. В соответствии с определением (19) при построении сеточного оператора дивергенции мы должны определить аппроксимации нормальной компоненты вектора $\mathbf{c}(\mathbf{x})u(\mathbf{x})$ к соответствующей грани контрольного объема для узла $\mathbf{x} \in \omega$. Приближение сеточной функции $y(\mathbf{x})$, $\mathbf{x} \in \omega$,

в узлах сетки ω^* можно выполнить различными способами. В контрольном объеме Ω_m^* для узла $\mathbf{x} \in \omega^*$ выделим две компоненты вектора $\mathbf{c}(\mathbf{x})$ вдоль диагоналей Ω_m^* . Имеем $\mathbf{c}(\mathbf{x}) = \{c_n(\mathbf{x}), c_c(\mathbf{x})\}$, причем $c_n(\mathbf{x})$ соответствует диагонали по узлам сетки ω^n , а $c_c(\mathbf{x})$ — диагонали по узлам сетки ω^c . Если при аппроксимации оператора дивергенции на границе контрольного объема задействована $c_n(\mathbf{x})$, то будем аппроксимировать $y(\mathbf{x})$, $\mathbf{x} \in \omega^*$, по узлам сетки ω^n , а при задействовании $c_c(\mathbf{x})$ проведем аппроксимацию по узлам сетки ω^c . По такому правилу строятся аппроксимации $\bar{y}(\mathbf{x})$, $\mathbf{x} \in \omega^*$, в

$$(C_2 y)(\mathbf{x}_k) = \operatorname{div}_h (\mathbf{c}(\mathbf{x}) \bar{y}(\mathbf{x})), \quad \mathbf{x}_k \in \omega. \quad (30)$$

Формула (30) при использовании обозначений на фиг. 6 с учетом (8) соответствует заданию сеточного оператора конвективного переноса в дивергентной форме

$$(C_2 y)(\mathbf{x}_k) = \frac{c_1(\mathbf{x}_2) \bar{y}_1(\mathbf{x}_2) - c_1(\mathbf{x}_4) \bar{y}_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{c_2(\mathbf{x}_3) \bar{y}_2(\mathbf{x}_3) - c_2(\mathbf{x}_1) \bar{y}_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}, \quad (31)$$

где

$$\begin{aligned} \bar{y}_2(\mathbf{x}_1) &= \frac{1}{2}(y(\mathbf{x}_5) + y(\mathbf{x}_k)), & \bar{y}_1(\mathbf{x}_2) &= \frac{1}{2}(y(\mathbf{x}_6) + y(\mathbf{x}_k)), \\ \bar{y}_2(\mathbf{x}_3) &= \frac{1}{2}(y(\mathbf{x}_7) + y(\mathbf{x}_k)), & \bar{y}_1(\mathbf{x}_4) &= \frac{1}{2}(y(\mathbf{x}_8) + y(\mathbf{x}_k)). \end{aligned}$$

Тем самым аппроксимация выполняется на обычном пятиточечном шаблоне *крест*.

Для операторов конвективного переноса имеет место соотношение

$$C_1 u = C_2 u - u \operatorname{div} \mathbf{c}.$$

Согласованные аппроксимации соответствуют тому, что подобное равенство выполнено и на сеточном уровне, когда

$$C_1 y = C_2 y - y \operatorname{div}_h \mathbf{c}. \quad (32)$$

В методе опорных операторов [11, 12] соотношение типа (32) используется для построения одного сеточного оператора при выбранной аппроксимации другого сеточного оператора. При (31), (32) для сеточного оператора конвективного переноса в недивергентной форме C_1 получим

$$(C_1 y)(\mathbf{x}_k) = \frac{c_1(\mathbf{x}_2) w_1(\mathbf{x}_2) + c_1(\mathbf{x}_4) w_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{c_2(\mathbf{x}_3) w_2(\mathbf{x}_3) + c_2(\mathbf{x}_1) w_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}, \quad (33)$$

где

$$\begin{aligned} w_2(\mathbf{x}_1) &= \frac{1}{2}(y(\mathbf{x}_k) - y(\mathbf{x}_5)), & w_1(\mathbf{x}_2) &= \frac{1}{2}(y(\mathbf{x}_6) - y(\mathbf{x}_k)), \\ w_2(\mathbf{x}_3) &= \frac{1}{2}(y(\mathbf{x}_7) - y(\mathbf{x}_k)), & w_1(\mathbf{x}_4) &= \frac{1}{2}(y(\mathbf{x}_k) - y(\mathbf{x}_8)). \end{aligned}$$

Для оператора конвективного переноса в кососимметричной форме C_0 в соответствии с (21) используется представление

$$C_0 = \frac{1}{2}(C_1 + C_2). \quad (34)$$

С учетом (31) и (33) из (34) следует

$$(C_0 y)(\mathbf{x}_k) = \frac{c_1(\mathbf{x}_2) y(\mathbf{x}_6) + c_1(\mathbf{x}_4) y(\mathbf{x}_8)}{2|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{c_2(\mathbf{x}_3) y(\mathbf{x}_7) + c_2(\mathbf{x}_1) y(\mathbf{x}_5)}{2|\mathbf{x}_3 - \mathbf{x}_1|}. \quad (35)$$

Ключевая особенность аппроксимаций (31), (33) и (35) связана с тем, что мы можем работать с узлами только одной сетки. Например, для $y(\mathbf{x}) \in H(\omega^n)$ имеем $(C_0 y)(\mathbf{x}) \in H(\omega^n)$, а для $y(\mathbf{x}) \in H(\omega^c)$ имеем $(C_0 y)(\mathbf{x}) \in H(\omega^c)$. Аналогичные свойства имеет и сеточный оператор диффузии D , определяемый согласно (26), в случае диагональной матрицы $K(\mathbf{x})$ ($k_{\alpha\beta}(\mathbf{x}) = 0$, $\alpha \neq \beta = 1, 2$).

Для сеточных операторов конвективно переноса устанавливаются аналоги равенств (20), (22) в пространстве $H_0(\omega)$:

$$C_1^* = -C_2, \quad C_0^* = -C_0. \quad (36)$$

В силу (29) и (36) для приближенного решения $y(\mathbf{x}) \in H_0(\omega)$ сеточной задачи конвекции-диффузии (25) получим априорную оценку

$$\|y\| \leq \frac{1}{\delta_h} \|\varphi\|,$$

которая является сеточным аналогом оценки (24) для решения дифференциальной задачи.

4. ВЕКТОРНЫЕ ЗАДАЧИ

Операторно-разностные аппроксимации на объединенных сетках можно применять при приближенном решении краевых задач для векторных уравнений. В качестве типичного примера будем рассматривать уравнение

$$-\operatorname{div}(a(\mathbf{x})\varepsilon(\mathbf{u})) - \operatorname{grad}(b(\mathbf{x}) \operatorname{div} \mathbf{u}) = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (37)$$

для искомого вектора $\mathbf{u} = \{u_1(\mathbf{x}), u_2(\mathbf{x})\}$. В уравнении (37) через $\varepsilon(\mathbf{u})$ обозначена симметричная часть градиента векторной функции $\mathbf{u}$:

$$\varepsilon(\mathbf{u}) = \frac{1}{2}(\operatorname{grad} \mathbf{u} + (\operatorname{grad} \mathbf{u})^t).$$

Дополним это уравнение однородными граничными условиями Дирихле

$$\mathbf{u}(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega. \quad (38)$$

На достаточно гладкие коэффициенты уравнения (37) накладываются ограничения $a(\mathbf{x}) \geq \kappa > 0$, $b(\mathbf{x}) \geq 0$. Примером уравнения (37) выступают уравнение Ламе в линейной теории упругости.

В пространстве векторных функций $\mathbf{L}_2(\Omega)$ для скалярного произведения и нормы положим

$$(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \mathbf{u}(\mathbf{x}) \mathbf{v}(\mathbf{x}) d\mathbf{x}, \quad \|\mathbf{u}\| = (\mathbf{u}, \mathbf{u})^{1/2}.$$

На множестве функций $\mathbf{u}(\mathbf{x})$, удовлетворяющих (38), определим операторы

$$\mathcal{A}\mathbf{u} = -\operatorname{div}(a(\mathbf{x})\varepsilon(\mathbf{u})), \quad \mathcal{B}\mathbf{u} = -\operatorname{grad}(b(\mathbf{x}) \operatorname{div} \mathbf{u}), \quad (39)$$

для которых

$$(\mathcal{A}\mathbf{u}, \mathbf{v}) = \int_{\Omega} a(\mathbf{x})\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{v}) d\mathbf{x}, \quad (\mathcal{B}\mathbf{u}, \mathbf{v}) = \int_{\Omega} b(\mathbf{x}) \operatorname{div} \mathbf{u} \operatorname{div} \mathbf{v} d\mathbf{x}.$$

В силу этого операторы $\mathcal{A}, \mathcal{B}$ являются симметричными и неотрицательными при неотрицательных коэффициентах $a(\mathbf{x}), b(\mathbf{x})$:

$$\mathcal{A} = \mathcal{A}^* \geq 0, \quad \mathcal{B} = \mathcal{B}^* \geq 0. \quad (40)$$

При отмеченных ограничениях $a(\mathbf{x}) \geq \kappa > 0$ неравенство Корна [20] для векторных функций обеспечивает положительную определенность оператора $\mathcal{A}$. При разностной аппроксимации краевой задачи (37), (38) мы наследуем ключевые свойства (40) сеточных аналогов операторов $\mathcal{A}, \mathcal{B}$.

При приближенном решении двумерной задачи (37), (38) ищется сеточная векторная функция $\mathbf{y}(\mathbf{x})$, $\mathbf{x} \in \omega$, с компонентами $y_\alpha(\mathbf{x})$, $\alpha = 1, 2$. При построении сеточного аналога оператора $\mathcal{A}$ сначала определяется тензор второго ранга (матрица), который связывается с градиентом векторной функции. Принимая во внимание (5) и работая с отдельными компонентами вектора, для $\operatorname{grad}_h \mathbf{y}$ получим

$$(\operatorname{grad}_h \mathbf{y})(\mathbf{x}_m) = \begin{pmatrix} \frac{y_1(\mathbf{x}_2) - y_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} & \frac{y_1(\mathbf{x}_3) - y_1(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|} \\ \frac{y_2(\mathbf{x}_2) - y_2(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} & \frac{y_2(\mathbf{x}_3) - y_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|} \end{pmatrix}, \quad \mathbf{x}_m \in \omega^*, \quad (41)$$

при использовании обозначений фиг. 5. Для элементов тензора

$$\varepsilon_h(\mathbf{y}) = \{(\varepsilon_h(\mathbf{y}))_{\alpha\beta}\}, \quad \alpha, \beta = 1, 2,$$

это дает

$$\begin{aligned} (\varepsilon_h(\mathbf{y}))_{11}(\mathbf{x}_m) &= \frac{y_1(\mathbf{x}_2) - y_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|}, \quad (\varepsilon_h(\mathbf{y}))_{22}(\mathbf{x}_m) = \frac{y_2(\mathbf{x}_3) - y_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}, \\ (\varepsilon_h(\mathbf{y}))_{12}(\mathbf{x}_m) &= (\varepsilon_h(\mathbf{y}))_{21}(\mathbf{x}_m) = \frac{y_1(\mathbf{x}_3) - y_1(\mathbf{x}_1)}{2|\mathbf{x}_3 - \mathbf{x}_1|} + \frac{y_2(\mathbf{x}_2) - y_2(\mathbf{x}_4)}{2|\mathbf{x}_2 - \mathbf{x}_4|}, \quad \mathbf{x}_m \in \omega^*. \end{aligned} \quad (42)$$

Аналогично (8) (см. фиг. 6) определяется вектор

$$\begin{aligned} (\operatorname{div}_h \sigma)(\mathbf{x}_k) &= \left\{ \frac{\sigma_{11}(\mathbf{x}_2) - \sigma_{11}(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{\sigma_{12}(\mathbf{x}_3) - \sigma_{12}(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}, \right. \\ &\quad \left. \frac{\sigma_{21}(\mathbf{x}_2) - \sigma_{21}(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{\sigma_{22}(\mathbf{x}_3) - \sigma_{22}(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|} \right\}, \quad \mathbf{x}_k \in \omega, \end{aligned} \quad (43)$$

для тензора второго ранга $\sigma(\mathbf{x})$, заданного в узлах $\mathbf{x}_m \in \omega^*$. В нашем случае $\sigma(\mathbf{x}) = a(\mathbf{x})(\varepsilon_h(\mathbf{y}))(\mathbf{x})$, $\mathbf{x}_m \in \omega^*$. Расчетные формулы (42), (43) позволяют определить сеточный аналог оператора $\mathcal{A}$ следующим образом:

$$(\mathcal{A}\mathbf{y})(\mathbf{x}_k) = -\operatorname{div}_h (a(\mathbf{x}) \varepsilon_h(\mathbf{y}))(\mathbf{x}_k) \quad \mathbf{x}_k \in \omega \cap \Omega. \quad (44)$$

На подпространстве

$$\mathbf{H}_0(\omega) = \{\mathbf{y}(\mathbf{x}) \mid \mathbf{y}(\mathbf{x}) \in \mathbf{H}(\omega), \quad \mathbf{y}(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega\}$$

для сеточного оператора A получим

$$(\mathcal{A}\mathbf{y}, \mathbf{v}) = \sum_{\mathbf{x} \in \omega} a(\mathbf{x}) \varepsilon_h(\mathbf{y}) : \varepsilon_h(\mathbf{v}) V(\mathbf{x}).$$

При $a(\mathbf{x}) \geq 0$ это дает

$$A = A^* \geq 0.$$

Для установления положительной определенности A при $a(\mathbf{x}) \geq \kappa > 0$ необходимо провести более детальное рассмотрение, следуя, например, [12].

Отметим ключевые моменты в построении сеточного аналога оператора $\mathcal{B}$. Для векторных сеточных функций $\mathbf{y}(\mathbf{x})$, $\mathbf{x} \in \omega$, сначала рассчитывается приближение для дивергенции в узлах сетки ω^* . При использовании обозначений на фиг. 5 положим

$$(\widetilde{\operatorname{div}}_h \mathbf{y})(\mathbf{x}_m) = \frac{y_1(\mathbf{x}_2) - y_1(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|} + \frac{y_2(\mathbf{x}_3) - y_2(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|}, \quad \mathbf{x}_m \in \omega^*. \quad (45)$$

Сеточный оператор градиента в узлах сетки ω (см. фиг. 6) определим для сеточной скалярной функции $w(\mathbf{x})$, $\mathbf{x} \in \omega^*$:

$$(\widetilde{\operatorname{grad}}_h w)(\mathbf{x}_k) = \left\{ \frac{w(\mathbf{x}_2) - w(\mathbf{x}_4)}{|\mathbf{x}_2 - \mathbf{x}_4|}, \frac{w(\mathbf{x}_3) - w(\mathbf{x}_1)}{|\mathbf{x}_3 - \mathbf{x}_1|} \right\}, \quad \mathbf{x}_k \in \omega. \quad (46)$$

Введенные аппроксимации градиента (46) $(\widetilde{\operatorname{grad}}_h : H(\omega^*) \rightarrow \mathbf{H}(\omega))$ и дивергенции (45) $(\widetilde{\operatorname{div}}_h : \mathbf{H}(\omega) \rightarrow H(\omega^*))$ являются согласованными в смысле выполнения на $\mathbf{H}_0(\omega)$ равенства

$$\widetilde{\operatorname{grad}}_h^* = -\widetilde{\operatorname{div}}_h. \quad (47)$$

Сеточный аналог оператора $\mathcal{B}$ определим как

$$(\mathcal{B}\mathbf{y})(\mathbf{x}_k) = -\widetilde{\operatorname{grad}}_h (b(\mathbf{x}) \widetilde{\operatorname{div}}_h \mathbf{y}), \quad \mathbf{x}_k \in \omega \cap \Omega. \quad (48)$$

С учетом (47) получим

$$(\mathcal{B}\mathbf{y}, \mathbf{v}) = (b(\mathbf{x}) \widetilde{\operatorname{div}}_h \mathbf{y}, \widetilde{\operatorname{div}}_h \mathbf{v}).$$

Для $b(\mathbf{x}) \geq 0$ это дает

$$B = B^* \geq 0.$$

Тем самым при аппроксимациях (44), (48) наследуются отмеченные выше (смотри (40)) основные свойства операторов $\mathcal{A}, \mathcal{B}$.

СПИСОК ЛИТЕРАТУРЫ

1. Самарский А. А. Теория разностных схем. М.: Наука, 1989.
2. Strikwerda J. C. Finite Difference Schemes and Partial Differential. Philadelphia: Society for Industrial Mathematics, 2004.
3. Самарский А. А. Уравнения параболического типа с разрывными коэффициентами и разностные методы их решения // Тр. Всес. совещания по дифференциальным уравнениям. (Ереван, ноябрь 1958 г.). Ереван: Изд-во АН ЛрмССР, 1960. С. 148–160.
4. Тихонов А. Н., Самарский А. А. Однородные разностные схемы // Ж. вычисл. матем. и матем. физ. 1961. Т. 1. № 1. С. 4–63.
5. Eymard R., Gallouet T., Herbin R. Finite volume methods // Handbook of Numerical Analysis. V. 7. Amsterdam: North Holland, 2000. P. 713–1020.
6. Li R., Chen Z., Wu W. Generalized Difference Methods for Differential Equations: Numerical Analysis of Finite Volume Methods. New York: Marcel Dekker, 2000.
7. Shashkov M. Conservative Finite-Difference Methods on General Grids. Boca Raton: CRC press, 1996.
8. da Veiga L. B., Lipnikov K., Manzini G. The Mimetic Finite Difference Method for Elliptic Problems. Berlin: Springer, 2014.
9. Лебедев В. И. Разностные аналоги ортогональных разложений, основных дифференциальных операторов и некоторых краевых задач математической физики. I // Ж. вычисл. матем. и матем. физ. 1964. Т. 4. № 3. С. 449–465.
10. Castillo J. E., Miranda G. F. Mimetic Discretization Methods. Boca Raton: CRC Press, 2013.
11. Самарский А. А., Тишкин В. Ф., Фаворский А. П., Шашков М. Ю. Операторные разностные схемы // Дифференц. уравнения. 1981. Т. 17. № 7. С. 1317–1327.
12. Самарский А. А., Колдоба А. В., Повещенко Ю. А. и др. Разностные схемы на нерегулярных сетках. Минск: Изд. Критерий, 1996.
13. Vabishchevich P. N. Finite-difference approximation of mathematical physics problems on irregular grids // Computational Methods in Applied Mathematics. 2005. V. 5. № 3. P. 294–330.
14. Фрязинов И. В. Об одной аппроксимации смешанных производных // Ж. вычисл. матем. и матем. физ. 1975. Т. 15. № 3. С. 644–660.
15. Самарский А. А., Фрязинов И. В. О разностных методах аппроксимации задач математической физики // Успехи матем. наук. 1976. Т. 31. № 6(192). С. 167–197.
16. Фрязинов И. В. Об одной разностной аппроксимации задач для эллиптического уравнения // Ж. вычисл. матем. и матем. физ. 1976. Т. 16. № 1. С. 102–118.
17. Фрязинов И. В. Аппроксимация двумерных эллиптических и параболических уравнений на паре согласованных сеток // Матем. моделирование. 1994. Т. 6. № 4. С. 53–64.
18. Самарский А. А., Вабищевич П. Н. Численные методы решения задач конвекции–диффузии. М.: URSS, 1999.
19. Самарский А. А., Андреев В. Б. Разностные методы для эллиптических уравнений. М.: Наука, 1976.
20. Duvaut G., Lions J. L. Inequalities in Mechanics and Physics. Berlin: Springer-Verlag, 1976.

OPERATOR-DIFFERENCE APPROXIMATIONS ON NON-STANDARD RECTANGULAR GRIDS

P. N. Vabishchevich^{a,b,*}

^a 119991 Moscow, Lenin Hills, 1, Lomonosov Moscow State University

^b 677007 Yakutsk, Kulakovsky Str., 42, Northeastern Federal University named after M. K. Ammosov

*e-mail: vabishchevich@gmail.com

Received February 7, 2024

Revised March 3, 2024

Accepted March 5, 2024

Abstract. In the approximate solution of boundary value problems for partial differential equations, difference methods are widely used. Grid approximations are most simply constructed when dividing the calculated area into rectangular cells. Usually the grid nodes coincide with the vertices of the cells. In addition to such nodal approximations, grids with nodes in the centers of cells are also used. It is convenient to formulate boundary value problems in terms of invariant operators of vector (tensor) analysis, which are compared with the corresponding grid analogues. The paper builds analogues of gradient and divergence operators on non-standard rectangular grids, the nodes of which consist of both the vertices of the calculated cells and their centers. The proposed approach is illustrated by approximations of the boundary value problem for the stationary two-dimensional convection-diffusion equation. The key features of the construction of approximations for vector problems with orientation to applied problems of solid mechanics are noted.

Keywords: rectangular computational grid, vector analysis operators, grid operators, operator-difference schemes, convection-diffusion equation.

РЕГУЛЯРИЗАЦИЯ РЕШЕНИЯ ВЫРОЖДЕННЫХ СИСТЕМ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ НА ПРИМЕРЕ ИДЕНТИФИКАЦИИ ВИРИАЛЬНОГО УРАВНЕНИЯ СОСТОЯНИЯ РЕАЛЬНОГО ГАЗА¹⁾

© 2024 г. А. Г. Викулов^{1,*}

¹ 125993 Москва, Волоколамское шоссе, 4, Московский авиационный ин-т, Россия

*e-mail: vikulovag81@gmail.com

Поступила в редакцию 06.06.2023 г.

Переработанный вариант 05.03.2024 г.

Принята к публикации 05.04.2024 г.

Для проведения термодинамического расчета цикла в двухфазной области необходимо уравнение состояния рабочего тела, в качестве которого используется вириальное уравнение с неизвестными функциями температуры. Построена вырожденная система алгебраических уравнений относительно неизвестных коэффициентов, которые являются значениями вириальных функций на дискретной температурной сетке. На основе метода регуляризации разработан вариационно-итерационный алгоритм решения вырожденной системы уравнений. Проведен вычислительный эксперимент, подтверждающий работоспособность метода. Библ. 12. Фиг. 8.

Ключевые слова: система алгебраических уравнений, вырожденная матрица, метод регуляризации, уравнение состояния, реальный газ, вириалькоэффициенты.

DOI: 10.31857/S0044466924070024, EDN: xjbxff

ВВЕДЕНИЕ

Задачи математической физики формулируются в виде дифференциальных уравнений в частных производных. При их решении вариационными методами получаются системы линейных алгебраических уравнений (СЛАУ):

$$Au = f \quad (1)$$

с квадратной матрицей A порядка n в левой, вектором f в правой частях и искомым вектором u . Кроме того, постановки многих физических задач имеют вид СЛАУ. Особую трудность представляют собой СЛАУ с большим количеством неизвестных. Во-первых, появляются вычислительные ошибки округления, во-вторых, постановка задачи может приводить к СЛАУ с матрицей в левой части, близкой к вырожденной.

Из-за ошибок округления в левой части уравнения (1) вместо матрицы A появляется матрица $A + \Delta$, в правой части вместо вектора f — вектор $f + \delta$. Тогда вместо уравнения (1) получается уравнение (см. [1])

$$(A + \Delta)u = f + \delta \Leftrightarrow A_\delta u_\delta = f_\delta, \quad (2)$$

где $A_\delta = A + \delta$ — приближенная матрица в левой части, $f_\delta = f + \delta$ — приближенный вектор в правой части, u_δ — приближенное решение. Ошибки округления тем меньше, чем меньше число обусловленности матрицы A . Если определитель матрицы A очень мал, то матрица близка к вырожденной. В этом случае возникает задача нахождения единственного решения, которая относится к существенно некорректным и требует регуляризации [2]. Если же определитель матрицы A заметно больше нуля, то решение системы (2) единственное, но для приближенно заданного вектора в правой части

¹⁾ Работа выполнена при финансовой поддержке РФФИ (код проекта 23-49-10060).

и, возможно, матрицы в левой части задача нахождения приближенного решения может быть плохо обусловленной — решение неустойчиво — и также требовать регуляризации.

В [3] разработана методика регуляризации решения плохо обусловленных и вырожденных СЛАУ на основе интервальных оценок. Наличие погрешностей в исходных данных и ошибок округления приводит к тому, что точное решение систем уравнений с неточными правой и левой частями может не иметь смысла. Интервальные оценки решения позволяют избежать этой трудности, так как гарантируют принадлежность решения определенному множеству. Задача решается методом регуляризации путем минимизации сглаживающего функционала на заданных интервалах.

В [4], [5] сформулированы условия регуляризуемости сдвигом СЛАУ (1) с комплексной квадратной матрицей порядка n в левой, комплексным вектором $f \in C^n$ в правой частях и искомым вектором $u \in C^n$. В [6], [7] эти условия дополнены новыми, справедливость которых доказана в виде теоремы со вспомогательными леммами. Семейство СЛАУ [6]

$$(A + \lambda B)u = f, \quad (3)$$

где B — произвольная квадратная матрица порядка n с комплексными элементами, $\lambda \in C$ — комплексный параметр, является регуляризацией сдвигом системы (1), если СЛАУ (2) имеет единственное решение u_λ при достаточно малых $|\lambda| > 0$ и решение u_λ сходится при $\lambda \rightarrow 0$ к некоторому решению СЛАУ (1) в случае ее разрешимости. Если матрица A обратима, то и матрица $A + \lambda B$ обратима для любой матрицы B и достаточно малых $|\lambda| > 0$, и для точного решения u СЛАУ (1) справедлива оценка [6]

$$\|u_\lambda - u\| = O(\lambda). \quad (4)$$

Если матрица A необратима, то не при всех $f \in C^n$ СЛАУ (1) разрешима, и в случае разрешимости СЛАУ (1) и однозначной разрешимости СЛАУ (3) может не иметь места оценка (4). Поэтому в [3]— [6] выделяются пары матриц A , B , для которых семейство СЛАУ (2) является регуляризацией сдвигом системы (1).

На основе метода регуляризации [2] разработаны прикладные методики нахождения приближенного решения плохо обусловленных и вырожденных СЛАУ, возникающих при формулировке задач геофизики [8, 9].

В настоящей работе на основе метода регуляризации разработана итерационно-вариационная методика решения вырожденных систем линейных и нелинейных алгебраических уравнений. Она применена для идентификации вириального уравнения состояния реального газа, коэффициенты которого неизвестны.

Для вычисления работы и изменения энтропии в термодинамическом процессе двухфазной смеси из жидкости и влажного пара необходимо уравнение состояния, связывающее между собой давление, удельный объем и температуру, или экспериментальные данные, аппроксимируемые приближенной зависимостью между параметрами состояния на ограниченных интервалах. Зная изменение энтропии в термодинамическом процессе, по h - s -диаграмме влажного пара данного вещества можно найти изменение энтальпии в этом процессе.

Существование критической точки, в которой равны удельные объемы жидкой v' и газовой v'' фаз двухфазной смеси на нижней и верхней пограничных кривых [10],

$$v = v'' = v_{cr},$$

говорит о непрерывности газообразного и жидкого состояний вещества в отличие от твердой фазы, имеющей принципиальные различия с газом и жидкостью. Математическое выражение этого факта заключается в возможности описания жидкой и газовой фаз единым уравнением состояния.

Диаграммы состояния реального вещества в газообразной и жидкой фазах заметно отличаются от диаграмм состояния идеального газа, поэтому не описываются уравнением состояния идеального газа. Окончательного ответа на вопрос об уравнении состояния реального газа до сих пор нет. Однако существуют приближения, с той или иной точностью описывающие термодинамические процессы реального вещества.

Промежуточным приближением между уравнениями состояния идеального и реального газов является уравнение Ван-дер-Ваальса. Более точное приближение к уравнению состояния реального газа получено методами статистической физики и носит название вириального уравнения [10]:

$$pv = RT \left(1 - \sum_{n=1}^{\infty} \frac{n}{n+1} \frac{\beta_n}{v^n} \right), \quad (5)$$

где $\beta_n = \beta_n(T)$ — вириальные коэффициенты, являющиеся функциями только температуры. Чем больше удельный объем, тем меньшее количество коэффициентов необходимо учитывать в вириальном уравнении. При $v \rightarrow \infty$ уравнение (5) совпадает с уравнением состояния идеального газа, и вириальные коэффициенты не требуются.

Вириальные функции $v_n(T)$ определяются экспериментально, а задача их вычисления имеет характер обратной. Так как функции $v_n(T)$ зависят только от температуры, то при постоянном давлении $p = p_c = \text{const}$ удельный объем является функцией только температуры. Возьмем при постоянном давлении p_c частичную сумму коэффициентов, ограниченную N первыми слагаемыми вириального уравнения (5):

$$p_c v(p_c, T) = RT \left(1 - \sum_{n=1}^N \frac{n}{n+1} \frac{\beta_n(T)}{v^n(T)} \right). \quad (6)$$

Для определения N функций $\beta_n(T)$ необходимы N линейно независимых режимов $v_c(T) = v(p_c, T)$, где $n = 1, 2, \dots, N$; $c = 1, 2, \dots, N$. Линейная независимость режимов проверяется по значению определителя Грама, который должен быть отличен от нуля [1]:

$$\det \begin{pmatrix} (v_1, v_1) & (v_1, v_2) & \dots & (v_1, v_N) \\ (v_2, v_1) & (v_2, v_2) & \dots & (v_2, v_N) \\ \dots & \dots & \dots & \dots \\ (v_N, v_1) & (v_N, v_2) & \dots & (v_N, v_N) \end{pmatrix} \neq 0, \quad (7)$$

где скалярные произведения функций $v_c(T)$ вычисляются в гильбертовом пространстве квадратично интегрируемых функций $L_2(T_{\min}, T_{\max})$:

$$(v_c, v_n)_{L_2} = \int_{T_{\min}}^{T_{\max}} v_c(T) v_n(T) dT.$$

Перепишем вириальное уравнение (6) в виде системы алгебраических уравнений относительно функций $\beta_n(T)$:

$$\sum_{n=1}^N \frac{n}{n+1} \frac{\beta_n(T)}{v^n(T)} = 1 - \frac{p_c v_c(T)}{RT}, \quad c = 1, 2, \dots, N.$$

В течение термодинамического процесса температура меняется во времени. Введем временную сетку, и определим значения температуры от времени в точках этой сетки:

$$T_j = T(\tau_j), \quad j = 0, 1, 2, \dots, M. \quad (8)$$

Для фиксированного значения температуры удельный объем постоянен:

$$v_{cj} = v_c(T_j) = \text{const}.$$

Поэтому для каждого значения температуры (8) получается система линейных алгебраических уравнений относительно коэффициентов $\beta_{nj} = \beta_n(T_j)$:

$$\sum_{n=1}^N \frac{n}{n+1} \frac{\beta_n(T_j)}{v^n(T_j)} = 1 - \frac{p_c v_c(T_j)}{RT_j}, \quad c = 1, 2, \dots, N, \quad j = 1, 2, \dots, M,$$

или

$$\sum_{n=1}^N \frac{n}{n+1} \frac{\beta_{nj}}{v_j^n} = 1 - \frac{p_c v_{cj}}{RT_j}, \quad c = 1, 2, \dots, N, \quad j = 1, 2, \dots, M. \quad (9)$$

Поскольку все коэффициенты при искомым β_{nj} имеют не нулевое значение, ранг основной матрицы в системе (9) равен рангу расширенной матрицы, и, по теореме Кронекера—Капелли, эта система совместна. Ранг матрицы равен числу неизвестных N , следовательно, система имеет единственное решение [11]. Однако в силу специфики задач идентификации, связанной с использованием экспериментальной информации в правой — а в данном случае и в левой — частях, определитель основной матрицы может быть очень мал, что при наличии ошибок вычисления может затруднить решение системы уравнений стандартными методами.

Для насыщенного пара давление однозначно определяется температурой $p = p(T)$. Так как функции $\beta_n(T)$ зависят только от температуры, то, если исключить степень сухости, удельный объем будет только функцией температуры: $v = v[p(T), T] = v(T)$. Этого можно добиться, рассматривая состояние жидкости или пара на линии насыщения. Поэтому будем считать удельный объем равным удельному объему сухого насыщенного пара на линии насыщения: $v(T) = v''(T)$. Уравнение состояния с частичной суммой приобретает вид:

$$p(T)v[p(T), T] = RT \left(1 - \sum_{n=1}^N \frac{n}{n+1} \frac{\beta_n(T)}{v^n(T)} \right).$$

На введенной временной сетке для фиксированного значения температуры удельный объем постоянен:

$$v_j = v[p(T_j), T_j] = v(T_j) = v'_j + v''_j = \text{const}, \quad j = 1, 2, \dots, M.$$

а совокупность коэффициентов ВЕТА_n не изменяется в зависимости от температуры — каждый коэффициент однозначно определен единственным значением температуры.

Отсюда следует, что в области насыщенного пара система линейных алгебраических уравнений является общей для всех температур, и коэффициенты β_n не зависят от температуры:

$$\sum_{n=1}^M \frac{n}{n+1} \frac{\beta_n}{v^n[p(T_j), T_j]} = 1 - \frac{p(T_j)v[p(T_j), T_j]}{RT_j}, \quad j = 1, 2, \dots, M,$$

или

$$\sum_{n=1}^M \frac{n}{n+1} \frac{\beta_n}{v_j^n} = 1 - \frac{p_j v_j}{RT_j}, \quad j = 1, 2, \dots, M. \quad (10)$$

В векторной форме система (10) принимает вид уравнения (1), если положить

$$A = \{A_{jn}\} = \left\{ \frac{n}{n+1} \frac{\beta_n}{v_j^n} \right\}, \quad u = \{\beta_j\}, \quad f = \left\{ 1 - \frac{p_j v_j}{RT_j} \right\}, \quad j = 1, 2, \dots, M, \quad n = 1, 2, \dots, M.$$

Для определения M коэффициентов β_j необходимы M линейно независимых измерений $v_j = v[p(T_j)]$, где $j = 1, 2, \dots, M$. Линейная независимость измерений проверяется по значению определителя Грама (7), который должен быть отличен от нуля [1]. Скалярные произведения функций $v(T_j)$ вычисляются в гильбертовом пространстве квадратично интегрируемых функций $L_2(T_{\min}, T_{\max})$:

$$(v_j, v_n)_{L_2} = \int_{T_{\min}}^{T_{\max}} v_j v_n dT = v_j v_n (T_{\max} - T_{\min}) = v(T_j) v(T_n) (T_{\max} - T_{\min});$$

$$j = 1, 2, \dots, M; \quad n = 1, 2, \dots, M.$$

1. РЕГУЛЯРИЗАЦИЯ РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ

В практических задачах в силу погрешности исходных данных вместо точного значения правой части f уравнения (1), где $f \in F$, $u \in U$; $F, U \in R$ ($F = R^n$, $U = R^m$, m не обязательно равно n) — метрические пространства, — известно ее приближенное значение f_δ , которое может принадлежать множеству $R_A = AD_A$, не являющемуся компактом. В этих случаях нельзя строить приближенное решение уравнения (7) по формуле $u_\delta = A^{-1}f_\delta$, так как обратный оператор $A^{-1}f_\delta$ может быть не определен, а решение u_δ — не единственным и не принадлежать заданному компакту D_A . По этой причине решение u_δ не является корректным по Тихонову, и вводится понятие квазирешения $u_\delta \in U$, а метод подбора при условии компактности множества U позволяет найти квазирешение (псевдо-решение), минимизирующее норму разности левой и правой частей [2]:

$$\rho_F = (Au_\delta, f) = \inf_{u \in U} \rho_F(Au, f).$$

Если U — компакт, то квазирешение существует для любого $f \in F$ и если, кроме того, $f \in AU$, то квазирешение $u_\delta = u$. Квазирешение может быть и не одно — в этом случае под квазирешением понимается любой элемент из множества квазирешений.

Уравнение

$$Au = f_\delta, \quad (11)$$

с приближенно заданной правой частью f_δ , принадлежащей области значений R_A оператора A , не являющейся компактом, называется существенно некорректным (существенно некорректной задачей), а его квазирешение — условно корректным по Тихонову. В существенно некорректных задачах исключается последнее из четырех условий применимости метода наименьших квадратов — условие единственности решения [1], — но остается возможность минимизации нормы разности левой и правой частей уравнения (11) и, следовательно, квадратичного функционала

$$Fu = \|Au - f_\delta\|^2.$$

Поскольку исходная система (1) может не иметь точного решения, то можно найти лишь квазирешение u_μ , принадлежащее множеству векторов Q , удовлетворяющих условию

$$\inf_{u \in U} \|Au - f\| = \mu \geq 0, \Rightarrow \|Au - f\| \leq \mu. \quad (12)$$

Для уравнения (11) с приближенно заданной правой частью квазирешение u_δ принадлежит множеству векторов Q_δ , удовлетворяющих условию

$$\inf_{u \in U} \|Au - f_\delta\| = \mu_\delta \geq 0, \Rightarrow \|Au - f_\delta\| \leq \mu_\delta. \quad (13)$$

Из неравенств треугольника

$$\begin{aligned} \|Au - f_\delta\| &\leq \|Au - f\| + \|f - f_\delta\|, \\ \|Au - f\| &\leq \|Au - f_\delta\| + \|f_\delta - f\|, \end{aligned}$$

с учетом (12), (13) следуют неравенства

$$\begin{aligned} \|Au - f_\delta\| &\leq \mu_\delta \leq \mu + \delta, \\ \|Au - f\| &\leq \mu \leq \mu_\delta + \delta, \end{aligned}$$

из которых

$$\begin{aligned} |\mu - \mu_\delta| &\leq \delta, \\ \|Au - f_\delta\| &\leq \mu_\delta + 2\delta. \end{aligned} \quad (14)$$

Пусть D_A — совокупность всех квазирешений системы (1). Нормальным относительно элемента u_μ решением (нормальным решением) системы (1) называется квазирешение u_0 с минимальной нормой $\|u^\sim u_\mu\|$ [2]:

$$\|u_0 - u_\mu\| = \inf_{u \in D_A} \|u - u_\mu\|.$$

Для любой системы вида (1) нормальное решение существует и единственно. Приближение $\tilde{u}_\delta$ к нормальному решению u_0 принадлежит множеству векторов $\tilde{Q}_\delta$, удовлетворяющих условию (14).

Произвольный вектор $\tilde{u}_\delta$ из множества $\tilde{Q}_\delta$ неустойчив к малым изменениям правой части уравнения (11), поэтому необходимо ограничить множество $\tilde{Q}_\delta$ дополнительным условием в виде минимума стабилизирующего квадратичного функционала (стабилизатора) [2]

$$\Omega_0 = \inf_{u \in \tilde{Q}_\delta} \Omega[u].$$

Поскольку нормальное решение является квазирешением, последовательность элементов, сходящаяся к нормальному решению в пространстве U ($D_A \subset U$), одновременно сходится и в пространстве F (как, например, в методе наименьших квадратов для линейного оператора A):

$$u \rightarrow u_0, \quad Au \rightarrow Au_0 = f_0 \approx f.$$

Поэтому стабилизатор Ω выбирается в виде квадрата расстояния между элементами $u, u_0 \in D_A \subset U$:

$$\Omega(u) = \rho_U^2(u, u_0) = \|u - u_0\|^2. \quad (15)$$

Тогда по аналогии с методом наименьших квадратов вариационный метод определения приближения $\tilde{u}_\delta$ к нормальному квазирешению u_0 будет основываться на минимизации квадратичного функционала, состоящего из двух слагаемых: квадрата нормы разности левой и правой частей уравнения (11) — сходимости в пространстве F — и стабилизатора Ω в виде квадрата нормы разности приближенного и нормального квазирешений, умноженного на нормирующий множитель γ , — сходимости в пространстве U [2]:

$$Mu = \|Au - f_\delta\|^2 + \gamma\Omega(u) = \rho_F^2(Au, f_\delta) + \gamma\Omega(u), \quad (16)$$

где $\gamma > 0$ — параметр регуляризации. Параметр регуляризации однозначно определяется по уравнению, следующему из нестрогого неравенства (14) [2]:

$$\|Au(\gamma) - f_\delta\|^2 = \mu_\delta + 2\delta. \quad (17)$$

С учетом (15) квадратичный функционал (16) запишется в виде

$$\begin{aligned} Mu &= \rho_F^2(Au, f_\delta) + \gamma\rho_U^2(u, u_0) = \|Au - f_\delta\|^2 + \gamma\|u - u_0\|^2 = \\ &= (Au - f_\delta, Au - f_\delta) + \gamma(u - u_0, u - u_0) = \\ &= (Au, Au) - 2(Au, f_\delta) + (f_\delta, f_\delta) + \gamma(u, u) - 2\gamma(u, u_0) + \gamma(u_0, u_0). \end{aligned} \quad (18)$$

Пусть функционал (18) принимает минимальное значение в $\tilde{Q}_\delta \subset D_A$ для элемента u_0 . Для любого другого элемента $u \in \tilde{Q}_\delta$

$$u = u_0 + tv, \quad (19)$$

где t — произвольное действительное число, $v \in \tilde{Q}_\delta$ — произвольный элемент из множества $\tilde{Q}_\delta$, значение функционала больше:

$$Mu = M(u_0 + tv) \geq Mu_0, \quad (20)$$

причем, равенство выполняется для элемента $u = u_0$. Вычислим значение функционала (18) с учетом выражения (19):

$$\begin{aligned} M(u_0 + tv) &= (A(u_0 + tv), A(u_0 + tv)) - 2(A(u_0 + tv), f_\delta) + (f_\delta, f_\delta) + \\ &+ \gamma((u_0 + tv), (u_0 + tv)) - 2\gamma((u_0 + tv), u_0) + \gamma(u_0, u_0) = \\ &= (A(u_0 + tv), A(u_0 + tv)) - 2(A(u_0 + tv), f_\delta) + (f_\delta, f_\delta) + \\ &+ \gamma(u_0, u_0) + 2\gamma t(u_0, v) + \gamma t^2(v, v) - 2\gamma(u_0, u_0) - 2\gamma t(v, u_0) + \gamma(u_0, u_0) = \\ &= (A(u_0 + tv), A(u_0 + tv)) - 2(A(u_0 + tv), f_\delta) + (f_\delta, f_\delta) + \gamma t^2(v, v). \end{aligned} \quad (21)$$

Для линейного оператора A

$$\begin{aligned} M(u_0 + tv) &= (Au_0 + tAv, Au_0 + tAv) - 2(Au_0 + tAv, f_\delta) + (f_\delta, f_\delta) + \gamma t^2(v, v) = \\ &= (Au_0, Au_0) + t(Au_0, Av) + t(Av, Au_0) + t^2(Av, Av) - 2(Au_0, f_\delta) - 2t(Av, f_\delta) + \\ &+ (f_\delta, f_\delta) + \gamma t^2(v, v) = (Au_0, Au_0) + 2t(Au_0, Av) + t^2(Av, Av) - 2(Au_0, f_\delta) - \\ &- 2t(Av, f_\delta) + (f_\delta, f_\delta) + \gamma t^2(v, v). \end{aligned} \quad (22)$$

Из (20) следует, что функционал Mu приобретает минимальное значение при $t = 0$:

$$\begin{aligned} \frac{d}{dt} M(u_0 + tv) &= [2(Au_0, Av) + 2t(Av, Av) - 2(Av, f_\delta) + 2\gamma t(v, v)]_{t=0} = \\ &= 2(Au_0, Av) - 2(Av, f_\delta) = 2(Au_0 - f_\delta, Av) = 0. \end{aligned} \quad (23)$$

Если R_A плотно в F , то для линейного оператора A по теореме о плотности [1] следует, что единственным элементом, ортогональным к произвольному элементу $Av \in F$, является нулевой элемент:

$$Au_0 - f_\delta = 0.$$

Таким образом, минимум квадратичного функционала (18) соответствует нормальному решению уравнения (11).

Для произвольного элемента $\tilde{u}_\delta \neq u_0$, являющегося приближением к нормальному решению, параметр $t \neq 0$ и определяется из уравнения (23):

$$\begin{aligned} 2(A\tilde{u}_\delta, Av) + 2t(Av, Av) - 2(Av, f_\delta) + 2\gamma t(v, v) &= 0, \\ (A\tilde{u}_\delta, Av) + t(Av, Av) - (Av, f_\delta) + \gamma t(v, v) &= 0, \\ (A\tilde{u}_\delta - f_\delta, Av) &= -t[(Av, Av) + \gamma(v, v)], \\ (A\tilde{u}_\delta - f_\delta, Av) &= -t(\|Av\|^2 + \gamma\|v\|^2), \\ t &= -\frac{(A\tilde{u}_\delta - f_\delta, Av)}{\|Av\|^2 + \gamma\|v\|^2}. \end{aligned} \quad (24)$$

Поскольку $v \in \tilde{Q}_\delta$, которое согласно (17) формируется из элементов вида $Au - f_\delta$, можно считать

$$v = A\tilde{u}_\delta - f_\delta. \quad (25)$$

Тогда из (24), (25) имеем

$$t = -\frac{(Av, v)}{\|Av\|^2 + \gamma\|v\|^2}.$$

При итерационном поиске нормального решения

$$u_{n+1} = u_n + t_n v_n, \quad n = 1, 2, \dots, \quad (26)$$

где

$$v_n = Au_n - f_\delta, \quad (27)$$

$$t_n = -\frac{(Av_n, v_n)}{\|Av_n\|^2 + \gamma\|v_n\|^2}, \quad (28)$$

а параметр регуляризации согласно (17) находится из условия

$$\|Au_{n+1}(\gamma) - f_\delta\|^2 = \mu_\delta + 2\delta$$

или с учетом (26), (28)

$$\begin{aligned}
 \left\| A \left(u_n - \frac{(Av_n, v_n)}{\|Av\|^2 + \gamma_n \|v\|^2} v_n \right) - f_\delta \right\|^2 &= \mu_\delta + 2\delta, \\
 \left\| Au_n - \frac{(Av_n, v_n)}{\|Av\|^2 + \gamma_n \|v\|^2} Av_n - f_\delta \right\|^2 &= \mu_\delta + 2\delta, \\
 \left\| v_n - \frac{(Av_n, v_n)}{\|Av\|^2 + \gamma_n \|v\|^2} Av_n \right\|^2 &= \mu_\delta + 2\delta, \\
 \left\| \left((\|Av_n\|^2 + \gamma_n \|v_n\|^2) v_n - (Av_n, v_n) Av_n \right) \right\|^2 &= (\mu_\delta + 2\delta) (\|Av_n\|^2 + \gamma_n \|v_n\|^2).
 \end{aligned} \tag{29}$$

В преобразовании (21) используется только симметрия скалярного произведения, а в преобразовании (22) — симметрия скалярного произведения и линейность оператора A , поэтому в отличие от теоремы о минимуме квадратичного функционала [1], на которой основывается стандартная группа прямых вариационных методов, не требуется положительность и, тем более, положительная определенность оператора A . Поскольку симметрия является свойством скалярного произведения, то методика (26)–(28) вариационно-итерационного метода регуляризации применима к любому линейному оператору A и, в частности, к линейному матричному оператору.

Если оператор A нелинеен, то производная функционала (21) равна

$$\begin{aligned}
 \frac{d}{dt} M(u_0 + tv) &= \frac{d}{dt} \left[\|A(u_0 + tv)\|^2 - 2(A(u_0 + tv), f_\delta) + (f_\delta, f_\delta) + \gamma t^2(v, v) \right]_{t=0} = \\
 &= \left[\frac{d}{dt} \|A(u_0 + tv)\|^2 - 2 \frac{d}{dt} (A(u_0 + tv), f_\delta) + 2\gamma t(v, v) \right]_{t=0} = \\
 &= \frac{d}{dt} (A(u_0 + tv), A(u_0 + tv))_{t=0} - 2 \frac{d}{dt} (A(u_0 + tv), f_\delta)_{t=0} = \\
 &= \left(\frac{dA(u_0 + tv)}{dt} v, A(u_0 + tv) \right)_{t=0} + \left(A(u_0 + tv), \frac{dA(u_0 + tv)}{dt} v \right)_{t=0} - 2 \left(\frac{dA(u_0 + tv)}{dt} v, f_\delta \right)_{t=0} = \\
 &= 2 \left(\frac{dA(u_0 + tv)}{dt} v, A(u_0 + tv) \right)_{t=0} - 2 \left(\frac{dA(u_0 + tv)}{dt} v, f_\delta \right)_{t=0} = \\
 &= 2 \left(A(u_0 + tv) - f_\delta, \frac{dA(u_0 + tv)}{dt} v \right)_{t=0} = 2 \left(Au - f_\delta, \frac{dAu}{dt} v \right)_{t=0} = 0.
 \end{aligned} \tag{30}$$

Если R_A плотно и имеет непрерывные производные dAu/dt в F , то для нелинейного оператора A по теореме о плотности следует, что единственным элементом, ортогональным к произвольному элементу $(dAu/dt)v \in F$, является нулевой элемент

$$Au_0 - f_\delta = 0.$$

Следовательно, минимум квадратичного функционала (18) соответствует нормальному решению уравнения (11).

Для произвольного элемента $\tilde{u}_\delta \neq 0$, являющегося приближением к нормальному решению, параметр $t \neq 0$ и определяется из уравнения (30):

$$\begin{aligned}
 \frac{d}{dt} M(\tilde{u}_\delta + tv) &= 2 \left(Au - f_\delta, \frac{dAu}{dt} v \right) + 2\gamma t(v, v) = 0, \\
 t &= - \frac{\left(Au - f_\delta, \frac{dAu}{dt} v \right)}{\gamma(v, v)} = - \frac{\left(\frac{dA(u_0 + tv)}{dt} v, v \right)}{\gamma \|v\|^2}.
 \end{aligned}$$

При итерационном поиске нормального решения

$$u_{n+1} = u_n + t_n v_n, \quad n = 1, 2, \dots, \tag{31}$$

где

$$v_n = Au_n - f_\delta,$$

параметр t определяется из уравнения (30):

$$t_n = - \frac{\left(\frac{dA(u_n + t_n v_n)}{dt} v_n, v_n \right)}{\gamma_n \|v_n\|^2},$$

а параметр регуляризации согласно (17) находится из условия

$$\|A(u_n + t_n(\gamma_n) v_n) - f_\delta\|^2 = \mu_\delta + 2\delta$$

или с учетом (28)

$$\|A(u_n + t_n(\gamma_n) v_n) - f_\delta\|^2 = \mu_\delta + 2\delta. \quad (32)$$

Если оператор $A = A_\delta$ в левой части уравнения (11) задан приближенно, как и правая часть $f = f_\delta$:

$$A_\delta u = f_\delta, -$$

то полученные соотношения полностью сохраняются как для линейного, так и для нелинейного оператора A в уравнении.

Во избежание комплексных значений параметра регуляризации расстояние между функциями $u, u_0 \in D_A \subset U$ в выражении (15) стабилизатора Ω должно быть мнимым:

$$\Omega(u) = \rho_U^2(u, u_0) = i^2 \|u - u_0\|^2.$$

Тогда квадратичный функционал (13) принимает вид

$$Mu = \|Au - f_\delta\|^2 + \gamma \Omega(u) = \rho_F^2(Au, f_\delta) + i^2 \gamma \|u - u_0\|^2,$$

а в уравнениях (29), (30) параметр регуляризации γ_n умножается на i^2 . После определения действительного параметра регуляризации на каждой итерации n используется исходная форма функционала со стабилизатором (15), и все полученные уравнения сохраняются:

$$\begin{cases} \rho_U(u, u_0) = \|u - u_0\| \\ \rho_U(u, u_0) = i \|u - u_0\| \end{cases} \Rightarrow \begin{cases} \Omega(U) = \rho_U^2(u, u_0) = \|u - u_0\|^2 \\ \Omega(U) = \rho_U^2(u, u_0) = i^2 \|u - u_0\|^2 \end{cases},$$

где стабилизатор в мнимом метрическом пространстве U используется для вычисления действительного параметра регуляризации γ_n , а стабилизатор в действительном пространстве U — для вычисления действительного параметра итерационного процесса t_n .

2. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

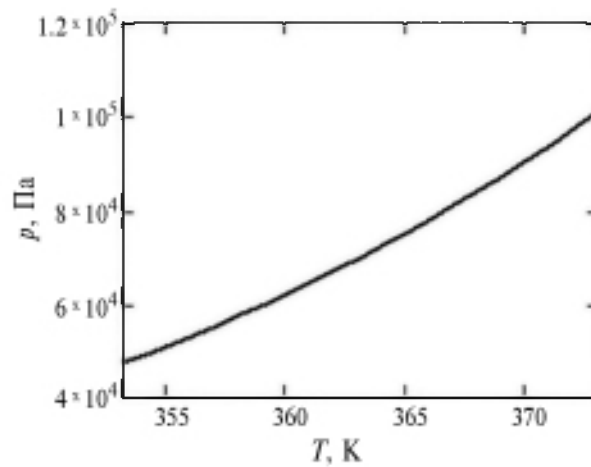
Для вычислительного эксперимента возьмем воду и водяной пар в интервале температуры от 80 до 100°C.

Давление влажного пара однозначно определяется его температурой (см. фиг. 1), а удельный объем связан со степенью сухости влажного пара соотношением [10]

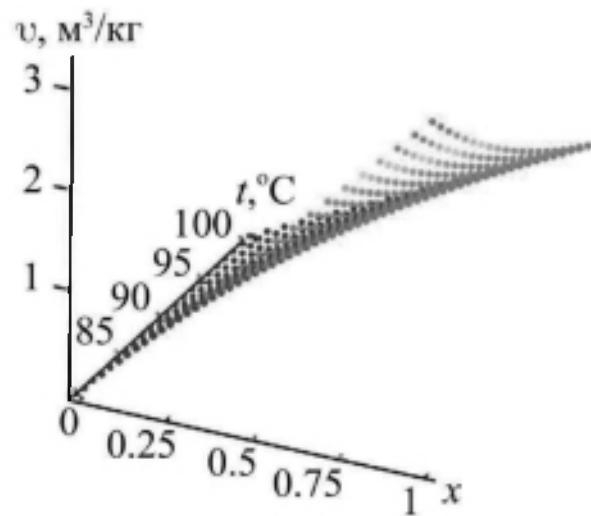
$$v = v' + x(v'' - v').$$

Поскольку удельные объемы сухого насыщенного пара и жидкости на линии насыщения также зависят от температуры, то удельный объем влажного пара является функцией температуры и степени сухости (см. фиг. 2).

Чтобы исключить степень сухости из уравнения состояния для вычисления вириальных функций, можно использовать одно из двух состояний: жидкость на линии насыщения левее критической точки при $x = 0$ (см. фиг. 3), сухой насыщенный пар на линии насыщения правее критической точки при



Фиг. 1. Распределение давления влажного пара от температуры.



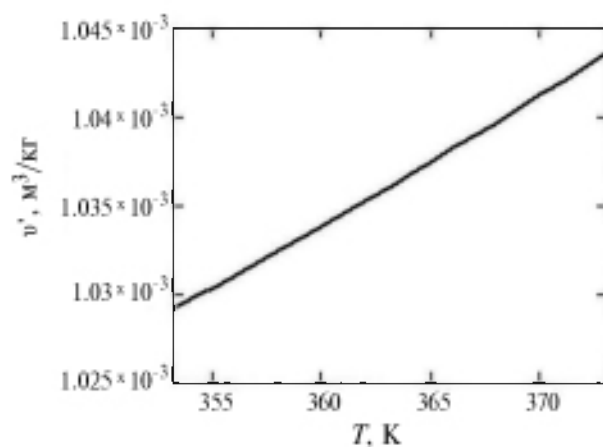
Фиг. 2. Распределение удельного объема влажного пара от температуры и степени сухости.

$x = 1$ (см. фиг. 4). Так как объем жидкости мал, удобнее производить расчеты для сухого насыщенного пара. Найденные вириальные коэффициенты будут справедливы и в области влажного пара, где изобара совпадает с изотермой.

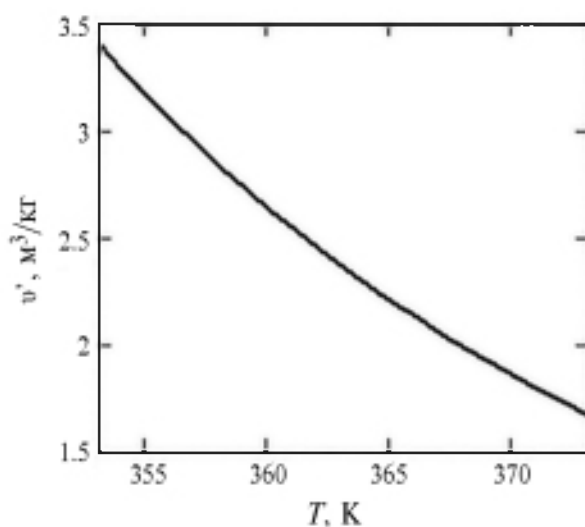
Непосредственное вычисление определителя для системы алгебраических уравнений (10) показывает, что его значение практически равно нулю, поэтому решение системы нуждается в регуляризации. Параметр регуляризации и функция итерационного процесса в зависимости от номера итерации показаны на фиг. 5, 6, а сравнение табличного распределения давления от температуры и расчетного, найденного по уравнению состояния (5), см. фиг. 7. Удовлетворительное визуальное совпадение достигается уже на третьей итерации, но с ростом числа итераций уменьшается значение невязки между расчетным и табличным распределениями.

Для вычислительного эксперимента возьмем воду и водяной пар в интервале температуры от 80 до 100°C [12]:

$$t = \begin{pmatrix} 80 & 81 & 82 & 83 & 84 & 85 & 86 & 87 & 88 & 89 & 90 \\ 91 & 92 & 93 & 94 & 95 & 96 & 97 & 98 & 99 & 100 \end{pmatrix}^T \text{ } ^\circ\text{C}.$$



Фиг. 3. Распределение удельного объема жидкости на линии насыщения от температуры.



Фиг. 4. Распределение удельного объема сухого насыщенного пара на линии насыщения от температуры.

Количество расчетных точек равно

$$M = 21.$$

Рассчитаем абсолютные температуры

$$j = 0, 1, \dots, M-1, \quad T_j = t_j + 273.15 \text{ K}.$$

Давление влажного пара однозначно определяется его температурой (фиг. 1) [12]:

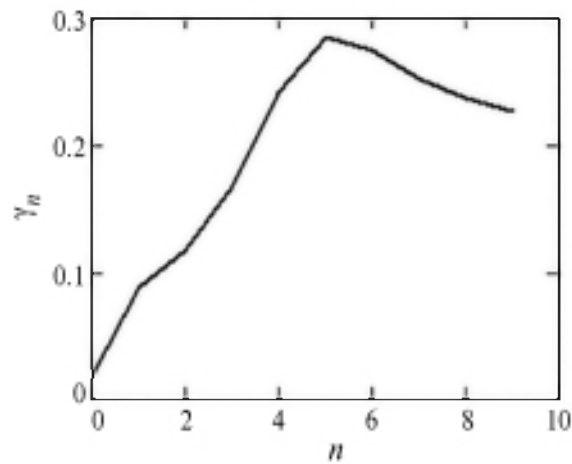
$$p = (4.7359 \ 4.9310 \ 5.1328 \ 5.3415 \ 5.5572 \ 5.7803 \ 6.0107 \ 6.2488 \ 6.4947 \ 6.7486 \ 7.0108 \\ 7.2814 \ 7.5607 \ 7.8488 \ 8.1460 \ 8.4525 \ 8.7685 \ 9.0943 \ 9.4301 \ 9.7760 \ 10.1325)^T \cdot 10^4 \text{ Па}.$$

Удельный объем связан со степенью сухости влажного пара соотношением [10]

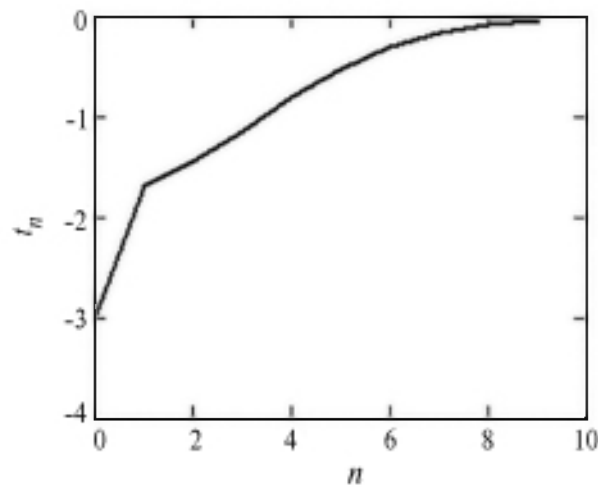
$$v = v' + x(v'' - v'),$$

где удельный объем жидкости

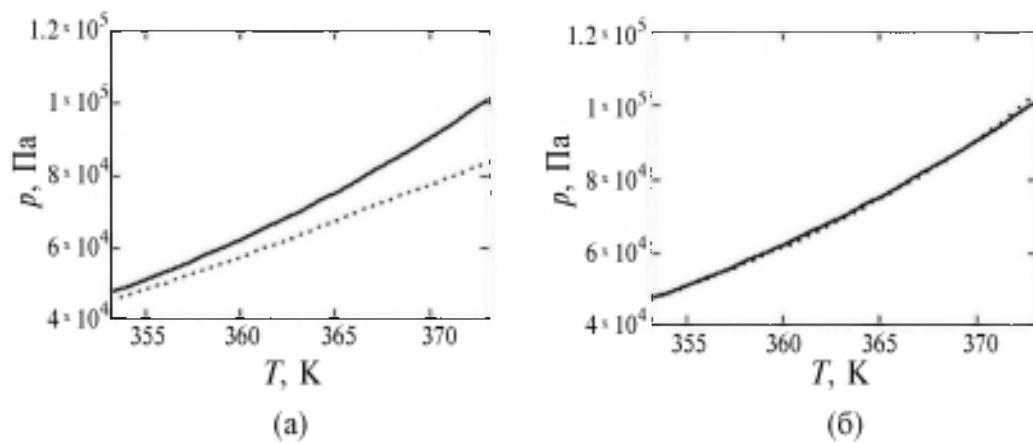
$$v' = (1.0292 \ 1.0299 \ 1.0305 \ 1.0312 \ 1.0319 \ 1.0326 \ 1.0333 \ 1.0340 \ 1.0347 \ 1.0354 \ 1.0361 \\ 1.0369 \ 1.0376 \ 1.0384 \ 1.0391 \ 1.0398 \ 1.0406 \ 1.0414 \ 1.0421 \ 1.0429 \ 1.0437)^T \cdot 10^{-4} \text{ м}^3/\text{кг}$$



Фиг. 5. Распределение параметра регуляризации $\gamma_n = \gamma(n)$ от номера итерации n .



Фиг. 6. Распределение функции итерационного процесса $t_n = t(\gamma_n) = t[\gamma(n)]$ от номера итерации n .



Фиг. 7. Распределение табличного (сплошная линия) и расчетного (пунктирная линия) распределений давления влажного пара в зависимости от температуры на итерации n : (а) $n = 1$; (б) $n = 10$.

и удельный объем сухого насыщенного пара

$$v'' = \begin{pmatrix} 3.4104 & 3.2839 & 3.1629 & 3.0471 & 2.9362 & 2.8300 & 2.7284 & 2.6309 & 2.5376 & 2.4482 & 2.3624 \\ 2.2801 & 2.2012 & 2.1256 & 2.0529 & 1.9832 & 1.9163 & 1.8520 & 1.7902 & 1.7309 & 1.6738 \end{pmatrix} \text{ м}^3/\text{кг}.$$

Поскольку удельные объемы сухого насыщенного пара и жидкости на линии насыщения также зависят от температуры (давления), то удельный объем влажного пара является функцией температуры (давления) и степени сухости (фиг. 2).

Чтобы исключить степень сухости из уравнения состояния для вычисления вириальных функций, можно использовать одно из двух состояний: жидкость на линии насыщения левее критической точки при $x = 0$ (фиг. 3), сухой насыщенный пар на линии насыщения правее критической точки при $x = 1$ (фиг. 4). Так как объем жидкости мал, удобнее производить расчеты для сухого насыщенного пара. Найденные вириальные функции будут справедливы и в области влажного пара, где изобара совпадает с изотермой.

Удельная газовая постоянная для воды

$$R = \frac{R_\mu}{\mu} = 461.30 \frac{\text{Дж}}{\text{кг} \cdot \text{К}},$$

где молярная масса воды $\mu = 18.015 \cdot 10^{-3}$ кг/моль, универсальная газовая постоянная $R_\mu = 8.31$ Дж/(моль·К). Вычислим вектор правых частей системы алгебраических уравнений (4.12):

$$j = 0, 1, \dots, M-1,$$

$$f_j = 1 - \frac{p_j v_j}{RT_j},$$

и матрицу левых частей

$$j = 0, 1, \dots, M-1, \quad n = 0, 1, \dots, M-1,$$

$$A_{j,n} = \frac{n}{n+1} \frac{1}{(v_j)^n}.$$

Непосредственное вычисление определителя для системы алгебраических уравнений (10) показывает, что его значение практически равно нулю, поэтому решение системы нуждается в регуляризации. Зададим начальное приближение для искомых коэффициентов:

$$j = 0, 1, \dots, M-1,$$

$$u_{0,j} = 1.$$

На каждой итерации i вычислим оператор Au и невязку v :

$$i = 1, \dots, N, \quad j = 0, 1, \dots, M-1,$$

$$Au_{i-1,j} = \sum_{n=0}^{M-1} \frac{n}{n+1} \frac{v_{i-1,n}}{(v_j)^n},$$

$$v_{i-1,j} = Au_{i-1,j} - f_j.$$

Запишем уравнение (29) относительно неизвестного параметра регуляризации в пространстве $L_2 [T_0, T_{M-1}]$, рассчитывая интегральные суммы вместо определенных интегралов:

$$\sum_{k=1}^{M-1} \left\{ \left[\left(\sum_{j=1}^{M-1} (Av_{i-1,j})^2 (T_j - T_{j-1}) + i^2 \gamma \sum_{j=1}^{M-1} (v_{i-1,j})^2 (T_j - T_{j-1}) \right) v_{i-1,k} - \left(\sum_{j=1}^{M-1} Av_{i-1,j} v_{i-1,j} (T_j - T_{j-1}) \right) Av_{i-1,k} \right] (T_k - T_{k-1}) \right\} =$$

$$= (\mu_\delta + 2\delta) \left[\sum_{j=1}^{M-1} (Av_{i-1,j})^2 (T_j - T_{j-1}) + i^2 \gamma_{i-1} \sum_{j=1}^{M-1} (v_{i-1,j})^2 (T_j - T_{j-1}) \right]. \quad (33)$$

Зададим ошибки определения исходных данных $\delta = 0.1$, разрешимости уравнения $\mu_\delta = 0$, количество итераций $N = 10$, начальное приближение параметра регуляризации на каждой итерации $\gamma_{i-1} = 1$ и решим уравнение (33) численно.

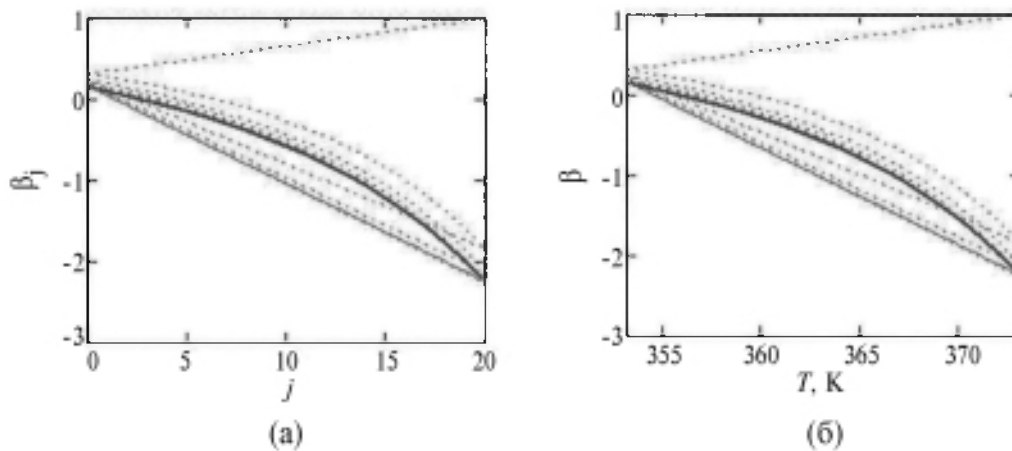
Затем по найденному параметру регуляризации вычислим функцию итерационного процесса (28):

$$t_{i-1} = \frac{\sum_{j=1}^{M-1} A v_{i-1,j} v_{i-1,j} (T_j - T_{j-1})}{\sum_{j=1}^{M-1} (A v_{i-1,j})^2 (T_j - T_{j-1}) + \gamma_{i-1} \sum_{j=1}^{M-1} (v_{i-1,j})^2 (T_j - T_{j-1})}. \quad (34)$$

Тогда новое приближение к искомым коэффициентам равно

$$u_{i,j} = u_{i-1,j} + t_{i-1} v_{i-1,j}. \quad (35)$$

Найденное распределение вириальных функций показано на фиг. 8.



Фиг. 8. Распределение вириальных функций: (а) — в зависимости от номера точки на равномерной температурной сетке; (б) — в зависимости от температуры.

3. ЗАКЛЮЧЕНИЕ

1. Для термодинамического расчета цикла двухфазной смеси необходимо уравнение состояния, в качестве которого может быть использовано вириальное уравнение реального газа с неизвестными функциями температуры. Эти функции могут быть идентифицированы по табличным распределениям давления и удельного объема в зависимости от температуры. В двухфазной области изобара совпадает с изотермой, поэтому давление однозначно определяется температурой. Удельный объем жидкости зависит от температуры и степени сухости. Чтобы исключить последнюю, удобно пользоваться степенью сухости, равной 0 — жидкость — или 1 — пар. Поскольку объем жидкости мал, удобнее пользоваться данными для сухого насыщенного пара. В результате получается система алгебраических уравнений с неизвестными вириальными коэффициентами, которые являются значениями вириальных функций на дискретной температурной шкале. Определитель системы очень мал, поэтому ее решение нуждается в регуляризации.
2. Минимизируя сглаживающий функционал метода регуляризации А. Н. Тихонова, как это делается в методе наискорейшего спуска для стандартного квадратичного функционала, можно выразить функцию итерационного процесса через параметр регуляризации, который определяется по задаваемому значению невязки левой и правой части. В результате, получается итерационно-вариационный метод, который может применяться как для линейного, так и нелинейного матричного оператора A в системе алгебраических уравнений $Au = f$.

3. Проведенный вычислительный эксперимент для воды в диапазоне температуры от 80 до 100°C и степени сухости $x = 1$ подтвердил регуляризующие свойства метода, позволившего определить вириальные функции температуры в указанном интервале из системы алгебраических уравнений с нулевым определителем. Поскольку вириальные функции зависят только от температуры, они являются общими для всей двухфазной области ниже линии насыщения.
4. Рассмотренная задача относится к существенно некорректным, и имеет решение, условно корректное по Тихонову. Регуляризующим условием отбора решения является требование одновременной сходимости правой части и квазирешения в соответствующих гильбертовых пространствах к правой части исходного уравнения и его нормальному квазирешению. Стабилизатор сглаживающего функционала, отвечающий за сходимость квазирешения, рассчитывается в мнимом метрическом пространстве для определения на каждой итерации действительного параметра регуляризации, и в действительном пространстве для вычисления функции итерационного процесса, зависящей от параметра регуляризации.
5. Разработанный метод отличается тем, что не требует обратимости оператора в левой части и допускает неразрешимость и вырожденность уравнения $Au = f$ с приближенно заданными правой и левой частями, позволяя находить приближение к его нормальному решению итерационно.

СПИСОК ЛИТЕРАТУРЫ

1. Ректорис К. Вариационные методы в математической физике и технике: Пер. с англ. М.: Мир, 1985. 590 с.
2. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1979. 285 с.
3. Джураев Х. Ш., Хасидов Я. Об одном регуляризующем алгоритме получения приближений к нормальному решению вырожденных систем линейных алгебраических уравнений // Молодой ученый. 2021. № 14. 6 с.
4. Морозов В. А., Назимов А. Б. О необходимых и достаточных условиях регуляризуемости вырожденных систем линейных алгебраических уравнений методом сдвига // Ж. вычисл. матем. и матем. физ. 1986. Т. 26. № 9. 1283 с.
5. Морозов В. А., Назимов А. Б. К проблеме регуляризации систем линейных алгебраических уравнений // Докл. АН СССР. 1986. Т. 286. № 3. 286 с.
6. Морозов В. А., Мухамадиев Э. М., Назимов А. Б. О проблеме регуляризации сдвигом вырожденных систем линейных алгебраических уравнений // Ж. вычисл. матем. и матем. физ. 2007. Т. 47. № 12. 1971 с.
7. Морозов В. А., Мухамадиев Э. М., Назимов А. Б. К проблеме регуляризации сдвигом вырожденных систем линейных алгебраических уравнений // Докл. АН. 2008. Т. 419. № 4. 454 с.
8. Барлиани А. Г. Модифицированный алгоритм Тихонова для решения вырожденных систем уравнений // Гео-Сибирь. 2010. Т. 1. № 1. 120 с.
9. Колотов И. И., Лукьяненко Д. В., Степанова И. Э., Ягола А. Г. О единственности решения систем линейных алгебраических уравнений, к которым редуцируются обратные линейные задачи гравиметрии и магнитометрии: локальный случай // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 8. 1317 с.
10. Кириллин В. А., Сычев В. В., Шейдлин А. Е. Техническая термодинамика. М.: Наука, 1979. 512 с.
11. Калиткин Н. Н. Численные методы. М.: Наука, 1978. 512 с.
12. Ривкин С. Л., Александров А. А. Термодинамические свойства воды и водяного пара: Справочник. 2-е изд. М.: Энергоатомиздат, 1984, 80 с.

REGULARIZATION OF THE SOLUTION OF DEGENERATE SYSTEMS OF ALGEBRAIC EQUATIONS BY THE EXAMPLE OF IDENTIFICATION OF THE VIRIAL EQUATION OF STATE OF A REAL GAS

A. G. Vikulov^{a,*}

^a 125993 Moscow, Volokolamskoye Highway 4, Moscow Aviation Institute, Russia

*e-mail: vikulovag81@gmail.com

Received June 6, 2023

Revised March 5, 2024

Accepted April 5, 2024

Abstract. To carry out the thermodynamic calculation of the cycle in the two-phase region, an equation of state of the working fluid is necessary, as which a virial equation with unknown temperature functions is used. A degenerate system of algebraic equations is constructed with respect to unknown coefficients, which are the values of virial functions on a discrete temperature grid. Based on the regularization method, a variational iterative algorithm for solving a degenerate system of equations has been developed. A computational experiment has been conducted to confirm the efficiency of the method.

Keywords: system of algebraic equations, degenerate matrix, regularization method, equation of state, real gas, virial coefficients.

MDM-АЛГОРИТМ И ЗАДАЧА СИЛЬВЕСТРА¹⁾© 2024 г. В. Н. Малоземов^{1,*}, Н. А. Соловьева^{2,**}, Г. Ш. Тамасян^{3,4,***}¹ 199034 С.-Петербург, Университетская наб., 7/9, С.-Пб гос. ун-т, Россия² 191023 С.-Петербург, наб. к-ла Грибоедова, 30/32, С.-Пб гос. экон. ун-т, Россия³ 197198 С.-Петербург, ул. Ждановская, 13, ВКА им. А. Ф. Можайского, Россия⁴ 198178 С.-Петербург, Большой пр-т. ВО, 61, ИПМ РАН, Россия

*e-mail: v.malozemov@spbu.ru

**e-mail: 4vinyo@gmail.com

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

Поступила в редакцию 06.02.2024 г.

Переработанный вариант 10.03.2024 г.

Принята к публикации 05.04.2024 г.

При разработке численных методов решения нелинейных минимаксных задач возникла следующая вспомогательная задача: в выпуклой оболочке некоторого конечного множества в евклидовом пространстве найти точку, имеющую наименьшую норму. В 1971 г. Б. Митчелл, В. Демьянов и В. Малоземов предложили нестандартный алгоритм решения этой задачи, который в дальнейшем получил название MDM-алгоритма (по главным буквам фамилий авторов). В данной статье рассматривается конкретная минимаксная задача: найти шар наименьшего объема, содержащий заданное конечное множество точек. Она называется задачей Сильвестра и является частным случаем задачи о чебышевском центре множества. Задаче Сильвестра сопоставляется выпуклая задача квадратичного программирования с симплексными ограничениями. Для решения этой задачи в статье предлагается использовать вариант MDM-алгоритма. С его помощью строится минимизирующая последовательность планов, такая, что у соседних планов различаются только две компоненты. Номера этих компонент выбираются на основе некоторых условий оптимальности. Доказывается слабая сходимость полученной последовательности планов, из которой следует сходимость по норме соответствующей последовательности векторов к единственному решению задачи Сильвестра. Приводятся четыре характерных примера на плоскости. Библ. 10. Фиг. 23.

Ключевые слова: задача Сильвестра, квадратичное программирование, MDM-алгоритм.**DOI:** 10.31857/S0044466924070038, **EDN:** xiwvrk

ВВЕДЕНИЕ

В пространстве $\mathbb{R}^n$ с евклидовой нормой заданы m попарно различных точек $a_1, \dots, a_m$. Рассмотрим задачу Сильвестра [1, 2]: найти шар наименьшего объема, содержащий все эти точки.

Задача Сильвестра формализуется следующим образом:

$$\max_{i \in 1:m} \left\{ \frac{1}{2} \|a_i - x\|^2 \right\} \rightarrow \min_{x \in \mathbb{R}^n}. \quad (1)$$

Здесь x — центр шара. Задача (1) имеет решение x^* и оно единственно [2, с. 116].

Обозначим через A матрицу со столбцами $a_1, \dots, a_m$ и через b вектор из $\mathbb{R}^m$ с компонентами $b[i] = \frac{1}{2} \|a_i\|^2$. Введем вспомогательную задачу квадратичного программирования

$$\begin{aligned} Q(u) &:= \frac{1}{2} \langle A^T A u, u \rangle - \langle b, u \rangle \rightarrow \min, \\ \sum_{i=1}^m u[i] &= 1, \quad u[i] \geq 0 \text{ при всех } i \in 1:m. \end{aligned} \quad (2)$$

Множество ее планов (векторов u , удовлетворяющих ограничениям задачи) обозначим через U . Очевидно, что у задачи (2) существует оптимальный план.

¹⁾ Результаты разд. 5 получены в Институте проблем машиноведения РАН при финансовой поддержке РФФИ (проект № 23-41-00060).

Справедливо следующее утверждение (см., например, [2, с. 117–119] и [3]).

Предложение. Если u^* — оптимальный план задачи (2), то вектор $x^* = Au^*$ будет решением задачи (1). При этом

$$\max_{i \in 1:m} \left\{ \frac{1}{2} \|a_i - x^*\|^2 \right\} = -Q(u^*). \quad (3)$$

Отметим, что целевая функция $Q(u)$ задачи (2) на множестве планов U допускает представление

$$Q(u) = -\frac{1}{2} \sum_{i=1}^m u[i] \cdot \|a_i - x\|^2, \quad (4)$$

где

$$x = Au = \sum_{i=1}^m u[i] a_i.$$

Действительно, в силу определения $b[i]$ имеем

$$\begin{aligned} \sum_{i=1}^m u[i] \cdot \|a_i - x\|^2 &= 2\langle b, u \rangle - 2\langle Au, x \rangle + \|x\|^2 = \\ &= -2\left(\frac{1}{2} \|x\|^2 - \langle b, u \rangle\right) = -2Q(u). \end{aligned}$$

Из (4) следует, что $Q(u) \leq 0$ при всех $u \in U$.

На самом деле, справедлив более точный результат. Чтобы записать его, введем обозначения. Пусть e_i , $i \in 1 : m$, — орты в пространстве $\mathbb{R}^m$ (они являются угловыми точками симплекса U) и

$$I^+(u) = \{i \in 1 : m \mid u[i] > 0\} \text{ — носитель плана } u.$$

Лемма 1. Для произвольного орта e_i справедливо равенство $Q(e_i) = 0$. Для остальных планов $u \in U$ задачи (2) выполняется строгое неравенство $Q(u) < 0$.

Доказательство. Рассмотрим сначала случай, когда $u = e_j$ при некотором $j \in 1 : m$. Тогда $I^+(u) = \{j\}$ и $x = Au = Ae_j = a_j$. Согласно (4) получим

$$Q(e_j) = -\frac{1}{2} \sum_{i=1}^m e_j[i] \cdot \|a_i - a_j\|^2 = -\frac{1}{2} \|a_j - a_j\|^2 = 0.$$

Теперь возьмем план $u \in U$, не совпадающий ни с одним из ортов e_i . Тогда носитель $I^+(u)$ содержит как минимум два индекса. Обозначим их p и q . На основании (4) запишем

$$\begin{aligned} Q(u) &= -\frac{1}{2} u[p] \cdot \|a_p - x\|^2 - \frac{1}{2} u[q] \cdot \|a_q - x\|^2 - \\ &\quad - \frac{1}{2} \sum_{i \neq p, i \neq q} u[i] \cdot \|a_i - x\|^2. \end{aligned}$$

Первые два слагаемых в правой части этого равенства не равны нулю одновременно, так как точки a_i попарно различны, а следующая за ними сумма неотрицательна. Значит, $Q(u) < 0$. Лемма доказана.

Вернемся к предложению. Оно показывает, что решение задачи Сильвестра (1) сводится к решению задачи (2). В данной статье для решения задачи (2) предложен простой алгоритм, который, по существу, является вариантом MDM-алгоритма [4–7]. С помощью этого алгоритма строится последовательность $u_0, u_1, \dots, u_i, \dots$ планов задачи (2), слабо сходящаяся к некоторому решению u^* задачи (2). Последовательность $\{u_i\}$ имеет важную особенность: при переходе от u_i к u_{i+1} у плана u_i пересчитываются только две компоненты. Доказано, что соответствующая последовательность векторов $v_0, v_1, \dots, v_i, \dots$, где $v_i = Au_i$, сходится по норме к решению x^* задачи Сильвестра.

В статье дано подробное изложение результатов, представленных в тезисах [8].

1. КРИТЕРИЙ ОПТИМАЛЬНОСТИ

Получим нестандартный вариант критерия оптимальности для задачи (2).

Теорема 1. План u задачи (2) будет оптимальным тогда и только тогда, когда для вектора $x = Au$ при некотором вещественном μ выполняются соотношения

$$\langle a_i, x \rangle - b[i] = \mu, \quad \text{если } u[i] > 0, \quad (5)$$

$$\langle a_i, x \rangle - b[i] \geq \mu, \quad \text{если } u[i] = 0. \quad (6)$$

Доказательство. Напомним стандартный критерий оптимальности, использующий условия Куна—Таккера: для того, чтобы план u задачи (2) был оптимальным, необходимо и достаточно, чтобы нашлись вещественные числа $\mu, \lambda_1, \dots, \lambda_m$, такие, что

$$A^T Au - b = \mu e + \sum_{i=1}^m \lambda_i e_i, \quad (7)$$

$$\lambda_i u[i] = 0 \quad \text{и} \quad \lambda_i \geq 0 \quad \text{при всех } i \in 1:m. \quad (8)$$

Здесь e — вектор из $\mathbb{R}^m$, все компоненты которого равны единице, и e_i — i -й орт в $\mathbb{R}^m$. Покажем, что условия (5), (6) и (7), (8) эквивалентны.

Предположим, что выполняются условия (7), (8). Обозначим $x = Au$ и распишем равенство (7) покомпонентно:

$$\langle a_i, x \rangle - b[i] = \mu + \lambda_i, \quad i \in 1:m. \quad (9)$$

Если $u[i] > 0$, то, согласно (8), $\lambda_i = 0$. В этом случае из (9) следует (5). При $u[i] = 0$ воспользуемся тем, что $\lambda_i \geq 0$. Тогда из (9) будет следовать (6).

Проверим обратное утверждение. Пусть выполнены условия (5), (6). Обозначим

$$\lambda_i = \langle a_i, x \rangle - b[i] - \mu, \quad i \in 1:m. \quad (10)$$

Формула (10) равносильна формуле (9), которая в свою очередь представляет собой покомпонентную запись равенства (7). Далее, по определению, $\lambda_i = 0$ при $u[i] > 0$ и $\lambda_i \geq 0$ при $u[i] = 0$. Это гарантирует выполнение условий (8).

Теорема доказана.

Замечание 1. Для константы μ можно получить явное представление

$$\mu = \|x\|^2 - \langle b, u \rangle.$$

Покажем это. Согласно (5), (6) и определению $b[i]$ имеем

$$\|x\|^2 - \|a_i - x\|^2 = 2(\langle a_i, x \rangle - b[i]) = 2\mu \quad \text{при } u[i] > 0,$$

$$\|x\|^2 - \|a_i - x\|^2 \geq 2\mu \quad \text{при } u[i] = 0.$$

Отсюда следует, что

$$\|a_i - x\|^2 = \|x\|^2 - 2\mu \quad \text{при } u[i] > 0,$$

$$\|a_i - x\|^2 \leq \|x\|^2 - 2\mu \quad \text{при } u[i] = 0.$$

Значит,

$$R^2 := \max_{i \in 1:m} \|a_i - x\|^2 = \|x\|^2 - 2\mu.$$

Вместе с тем, по формуле (3), $R^2 = -2Q(u)$, так что

$$\mu = \frac{1}{2} \|x\|^2 + Q(u) = \|x\|^2 - \langle b, u \rangle.$$

Критерий оптимальности (5), (6) можно еще упростить. Для этого введем оценку плана u — величину

$$\Delta(u) = \max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} - \min_{i \in 1:m} \{ \langle a_i, x \rangle - b[i] \}. \quad (11)$$

Здесь $x = Au$ и $I^+(u) = \{i \in 1 : m \mid u[i] > 0\}$ — носитель плана u . Величина $\Delta(u)$ при всех $u \in U$ неотрицательна. Действительно,

$$\Delta(u) \geq \max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} - \min_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} \geq 0. \quad (12)$$

Теорема 2. План u задачи (2) будет оптимальным тогда и только тогда, когда $\Delta(u) = 0$.

Доказательство. Для оптимального плана u равенство $\Delta(u) = 0$ непосредственно следует из условий (5), (6). Обратно, пусть $\Delta(u) = 0$. Согласно (12) имеем

$$\max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} = \min_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \}.$$

Значит,

$$\langle a_i, x \rangle - b[i] =: \mu \quad \text{при } i \in I^+(u). \quad (13)$$

Еще раз воспользуемся условием $\Delta(u) = 0$. Получим

$$\mu = \max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} = \min_{i \in 1:m} \{ \langle a_i, x \rangle - b[i] \},$$

так что

$$\langle a_i, x \rangle - b[i] \geq \mu \quad \text{при } i \in 1 : m. \quad (14)$$

Условия (13) и (14) гарантируют оптимальность плана u .

Теорема доказана.

С помощью $\Delta(u)$ можно оценить близость вектора $x = Au$ к решению x^* задачи (1).

Теорема 3. Пусть u — план задачи (2) и $x = Au$. Тогда

$$\|x - x^*\|^2 \leq \Delta(u). \quad (15)$$

Доказательство. По предложению справедливо равенство $x^* = Au^*$, где u^* — оптимальный план задачи (2). Согласно теореме 2, $\Delta(u^*) = 0$.

Имеем

$$\|x - x^*\|^2 = \langle x, x - x^* \rangle - \langle x^*, x - x^* \rangle.$$

Воспользуемся ограничениями на план задачи (2). Получим

$$\begin{aligned} \langle x, x - x^* \rangle &= \langle u, A^T(x - x^*) \rangle = \\ &= \sum_{i=1}^m u[i] \times \left((\langle a_i, x \rangle - b[i]) - (\langle a_i, x^* \rangle - b[i]) \right) \leq \\ &\leq \max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \} - \min_{i \in 1:m} \{ \langle a_i, x^* \rangle - b[i] \}. \end{aligned} \quad (16)$$

Аналогично

$$\begin{aligned} \langle x^*, x - x^* \rangle &= \sum_{i=1}^m u^*[i] \times \left((\langle a_i, x \rangle - b[i]) - (\langle a_i, x^* \rangle - b[i]) \right) \geq \\ &\geq \min_{i \in 1:m} \{ \langle a_i, x \rangle - b[i] \} - \max_{i \in I^+(u^*)} \{ \langle a_i, x^* \rangle - b[i] \}. \end{aligned} \quad (17)$$

Вычтем неравенство (17) из неравенства (16). Так как $\Delta(u^*) = 0$, то придем к требуемому неравенству (15).

Теорема доказана.

Оценке $\Delta(u)$ при $u \in U$ можно дать эквивалентное определение в терминах исходной задачи (1). А именно,

$$\Delta(u) = \frac{1}{2} \left(\max_{i \in 1:m} \|a_i - x\|^2 - \min_{i \in I^+(u)} \|a_i - x\|^2 \right), \quad (18)$$

где $x = Au$. Действительно, так как $b[i] = \frac{1}{2} \|a_i\|^2$, то

$$\langle a_i, x \rangle - b[i] = \frac{1}{2} (\|x\|^2 - \|a_i - x\|^2).$$

Согласно определению (11) имеем

$$\Delta(u) = \frac{1}{2} \left(\max_{i \in I^+(u)} \{ \|x\|^2 - \|a_i - x\|^2 \} - \min_{i \in 1:m} \{ \|x\|^2 - \|a_i - x\|^2 \} \right).$$

Воспользуемся следующими свойствами функций максимума и минимума:

$$\begin{aligned} \max_{i \in 1:m} \{ \alpha_i + c \} &= c + \max_{i \in 1:m} \{ \alpha_i \}, & \min_{i \in 1:m} \{ \alpha_i + c \} &= c + \min_{i \in 1:m} \{ \alpha_i \}, \\ \max_{i \in 1:m} \{ -\alpha_i \} &= - \min_{i \in 1:m} \{ \alpha_i \}, & \min_{i \in 1:m} \{ -\alpha_i \} &= - \max_{i \in 1:m} \{ \alpha_i \}. \end{aligned}$$

Получим

$$\begin{aligned} \Delta(u) &= \frac{1}{2} \left(\max_{i \in I^+(u)} \{ -\|a_i - x\|^2 \} - \min_{i \in 1:m} \{ -\|a_i - x\|^2 \} \right) = \\ &= \frac{1}{2} \left(\max_{i \in 1:m} \|a_i - x\|^2 - \min_{i \in I^+(u)} \|a_i - x\|^2 \right), \end{aligned}$$

что соответствует формуле (18).

Теорема 4. План u задачи (2) будет оптимальным тогда и только тогда, когда для вектора $x = Au$ выполняются равенства

$$\|a_s - x\| = \max_{i \in 1:m} \|a_i - x\| \text{ при всех } s \in I^+(u). \quad (19)$$

Доказательство. Согласно (18) критерий оптимальности $\Delta(u) = 0$ равносильно соотношению

$$\max_{i \in 1:m} \|a_i - x\| = \min_{i \in I^+(u)} \|a_i - x\|. \quad (20)$$

Если u — оптимальный план, то согласно (20) при всех $s \in I^+(u)$ имеем

$$\|a_s - x\| \geq \min_{i \in I^+(u)} \|a_i - x\| = \max_{i \in 1:m} \|a_i - x\|.$$

Обратное неравенство очевидно. Значит, справедливы равенства (19).

Наоборот, если выполнено условие (19), то справедливо соотношение (20). Оно эквивалентно равенству $\Delta(u) = 0$, гарантирующему оптимальность плана u .

Теорема доказана.

Теорему 4 можно переформулировать в геометрических терминах:

план u задачи (2) будет оптимальным тогда и только тогда, когда точки a_s при всех $s \in I^+(u)$ лежат на поверхности шара с центром в точке $x = Au$ и радиусом R , равным $\max_{i \in 1:m} \|a_i - x\|$.

Такие точки a_s будем называть опорными.

2. ВАРИАЦИЯ ПЛАНА ЗАДАЧИ (2)

Зафиксируем план u задачи (2), отличный от оптимального. Согласно теореме 2, его оценка $\Delta(u)$ положительна. Рассмотрим вариацию $\tilde{u}$ плана u следующего вида:

$$\tilde{u} = u + \delta(e_{i'} - e_{i''}), \quad (21)$$

где e_i — орты в пространстве $\mathbb{R}^m$. Распишем равенство векторов (21) по компонентам:

$$\tilde{u}[i] = \begin{cases} u[i'] + \delta & \text{при } i = i', \\ u[i''] - \delta & \text{при } i = i'', \\ u[i] & \text{при остальных } i \in 1 : m. \end{cases}$$

Будем считать, что выполнено условие

$$0 < \delta \leq u[i'']. \quad (22)$$

Представление (21) и условие (22) гарантируют принадлежность вектора $\tilde{u}$ множеству планов U задачи (2). Выбор параметра δ и индексов i' , i'' из множества $1 : m$ должен обеспечить максимально возможное уменьшение значения $Q(\tilde{u})$ целевой функции по сравнению со значением $Q(u)$.

Вычислим $Q(\tilde{u})$. Пусть $x = Au$. Отметим, что $Ae_i = a_i$. На основании (21) имеем

$$\begin{aligned} \frac{1}{2} \langle A^T A \tilde{u}, \tilde{u} \rangle &= \frac{1}{2} \|A \tilde{u}\|^2 = \frac{1}{2} \|x + \delta(a_{i'} - a_{i''})\|^2 = \\ &= \frac{1}{2} \|x\|^2 - \delta \langle a_{i''} - a_{i'}, x \rangle + \frac{1}{2} \delta^2 \|a_{i'} - a_{i''}\|^2, \\ \langle b, \tilde{u} \rangle &= \langle b, u \rangle - \delta(b[i''] - b[i']). \end{aligned}$$

Вычитая второе равенство из первого, получаем

$$Q(\tilde{u}) = Q(u) - \delta[(\langle a_{i''}, x \rangle - b[i'']) - (\langle a_{i'}, x \rangle - b[i'])] + \frac{1}{2} \delta^2 \|a_{i'} - a_{i''}\|^2. \quad (23)$$

Теперь выберем индексы i' и i'' из следующих соображений: $i' \in 1 : m$, $i'' \in I^+(u)$ и

$$\begin{aligned} \langle a_{i''}, x \rangle - b[i''] &= \max_{i \in I^+(u)} \{ \langle a_i, x \rangle - b[i] \}, \\ \langle a_{i'}, x \rangle - b[i'] &= \min_{i \in 1:m} \{ \langle a_i, x \rangle - b[i] \}. \end{aligned} \quad (24)$$

Тогда выражение в квадратных скобках из правой части формулы (23) будет равно $\Delta(u)$ (см. определение (11)). Сама формула (23) примет вид

$$Q(\tilde{u}) = Q(u) - \delta \Delta(u) + \frac{1}{2} \delta^2 \|a_{i'} - a_{i''}\|^2. \quad (25)$$

Условие $i'' \in I^+(u)$ означает, что $u[i''] > 0$. Это важное для дальнейшего неравенство. Оно, в частности, согласуется с условием (22).

Выражение $Q(\tilde{u})$ как функция от δ является квадратным трехчленом. Его минимум достигается при $\delta = \tilde{\delta}$, где

$$\tilde{\delta} = \frac{\Delta(u)}{\|a_{i'} - a_{i''}\|^2}. \quad (26)$$

Чтобы удовлетворить условию (22), придется подкорректировать δ . Положим

$$\delta = \begin{cases} \tilde{\delta}, & \text{если } \tilde{\delta} \leq u[i''], \\ u[i''], & \text{если } \tilde{\delta} > u[i'']. \end{cases} \quad (27)$$

Ясно, что $0 < \delta \leq \tilde{\delta}$.

Все параметры в формуле (21) для $\tilde{u}$ определены. Индексы i' и i'' находятся по формуле (24), а параметр δ — по формуле (27). При этом значение $Q(\tilde{u})$ вычисляется по формуле (25).

Отметим, что при $\delta = \tilde{\delta}$ равенство (25) упрощается:

$$Q(\tilde{u}) = Q(u) - \frac{1}{2} \tilde{\delta} \Delta(u).$$

Отсюда в силу неравенства $\delta \leq \tilde{\delta}$ следует, что в общем случае

$$Q(\tilde{u}) \leq Q(u) - \frac{1}{2} \delta \Delta(u). \quad (28)$$

3. MDM-АЛГОРИТМ

Опишем, как применяется MDM-алгоритм к решению задачи Сильвестра.

В качестве начального приближения возьмем произвольный план u_0 задачи (2). Ему соответствует вектор $x_0 = Au_0$.

Пусть уже имеется k -е приближение $u_k \in U$. Ему соответствует вектор $x_k = Au_k$. Проверим план u_k на оптимальность. Для этого выберем индексы $i'_k \in 1:m$ и $i''_k \in I^+(u_k)$ из условий

$$\begin{aligned}\langle a_{i'_k}, x_k \rangle - b[i'_k] &= \min_{i \in 1:m} \{ \langle a_i, x_k \rangle - b[i] \}, \\ \langle a_{i''_k}, x_k \rangle - b[i''_k] &= \max_{i \in I^+(u_k)} \{ \langle a_i, x_k \rangle - b[i] \}.\end{aligned}$$

Вычислим оценку $\Delta(u_k)$ плана u_k :

$$\Delta(u_k) = \langle a_{i''_k} - a_{i'_k}, x_k \rangle - (b[i''_k] - b[i'_k]).$$

Если $\Delta(u_k) = 0$, то u_k — решение задачи (2), а x_k — решение задачи (1). Процесс закончен.

Пусть $\Delta(u_k) > 0$. В этом случае будем говорить, что план u_k порождает k -ю итерацию. Вычислим

$$\tilde{\delta}_k = \frac{\Delta(u_k)}{\|a_{i'_k} - a_{i''_k}\|^2} \quad (29)$$

и положим

$$\delta_k = \begin{cases} \tilde{\delta}_k, & \text{если } \tilde{\delta}_k \leq u_k[i''_k], \\ u_k[i''_k], & \text{если } \tilde{\delta}_k > u_k[i''_k]. \end{cases} \quad (30)$$

Очевидно, что $0 < \delta_k \leq \tilde{\delta}_k$. Если $\delta_k = \tilde{\delta}_k$, то k -ю итерацию будем называть *неусеченной*. При $\delta_k < \tilde{\delta}_k$ получаем *усеченную итерацию*. Формулы

$$\begin{aligned}u_{k+1} &= u_k + \delta_k (e_{i'_k} - e_{i''_k}), \\ x_{k+1} &= x_k + \delta_k (a_{i'_k} - a_{i''_k})\end{aligned} \quad (31)$$

определяют очередное $(k+1)$ -е приближение. Описание алгоритма завершено.

С помощью предложенного алгоритма строятся две последовательности

$$\begin{aligned}u_0, u_1, \dots, u_k, \dots, \\ x_0, x_1, \dots, x_k, \dots\end{aligned} \quad (32)$$

Они конечны, если при некотором k выполнится равенство $\Delta(u_k) = 0$. Они бесконечны, когда $\Delta(u_k) > 0$ при всех $k = 0, 1, \dots$. В следующем разделе изучается вопрос о сходимости этих последовательностей.

4. СХОДИМОСТЬ MDM-АЛГОРИТМА

Итак, считаем, что

$$\Delta(u_k) > 0 \quad \text{при всех } k = 0, 1, \dots \quad (33)$$

Согласно (28) при всех $k = 0, 1, \dots$ выполняется неравенство

$$Q(u_k) - Q(u_{k+1}) \geq \frac{1}{2} \delta_k \Delta(u_k). \quad (34)$$

Так как δ_k и $\Delta(u_k)$ положительны, то $Q(u_k) > Q(u_{k+1})$. Получили, что последовательность $\{Q(u_k)\}$ строго убывает. Кроме того, она ограничена снизу наименьшим значением $Q(u)$ на множестве планов U . Значит, последовательность $\{Q(u_k)\}$ имеет предел. Обозначим этот предел через α и запишем

$$\lim_{k \rightarrow \infty} Q(u_k) = \alpha. \quad (35)$$

Напомним, что MDM-алгоритм состоит из итераций двух типов: неусеченных (при $\delta_k = \tilde{\delta}_k$) и усеченных (при $\delta_k < \tilde{\delta}_k$).

Лемма 2. *Количество идущих подряд усеченных итераций конечно.*

Доказательство существенно опирается на тот факт, что в последовательности $\{u_k\}$ все планы попарно различны в силу строгого убывания последовательности $\{Q(u_k)\}$.

Пусть итерация с номером k_0 усеченная и далее следуют также усеченные итерации. Согласно (30), при $k \geq k_0$ справедливо равенство $\delta_k = u_k[i_k'']$. Правило перехода от плана u_k к плану u_{k+1} принимает вид

$$u_{k+1}[i] = \begin{cases} u_k[i_k'] + u_k[i_k''] & \text{при } i = i_k', \\ 0 & \text{при } i = i_k'', \\ u_k[i] & \text{при остальных } i \in 1 : m. \end{cases} \quad (36)$$

По определению, $u_k[i_k''] > 0$. Если $u_k[i_k'] > 0$, то у плана u_{k+1} появляется дополнительная нулевая компонента. При $u_k[i_k'] = 0$ формула (36) упрощается

$$u_{k+1}[i] = \begin{cases} u_k[i_k''] & \text{при } i = i_k', \\ 0 & \text{при } i = i_k'', \\ u_k[i] & \text{при остальных } i \in 1 : m. \end{cases}$$

Ненулевая компонента $u_k[i_k'']$ переместилась на позицию $i = i_k'$, на которой раньше стоял ноль. Количество нулевых компонент не изменилось. Вместе с тем, количество последовательных преобразований, перемещающих положительную компоненту текущего плана на место нулевой компоненты, конечно. Значит, через конечное число усеченных итераций с $u_k[i_k'] = 0$ мы получим план, у которого аналогичная компонента положительна. У следующего плана появится дополнительная нулевая компонента.

Описанный процесс не может быть бесконечным, так как все планы u_k ненулевые.

Лемма доказана.

Таким образом, MDM-алгоритм может содержать конечные блоки подряд идущих усеченных итераций.

Из последовательности $\{u_k\}$ исключим все планы, порождающие усеченные итерации. Останется бесконечная подпоследовательность планов $\{u_{k_j}\}$, порождающих неусеченные итерации.

Теорема 5. *Для векторов $x_{k_j} = Au_{k_j}$ справедливо предельное соотношение*

$$\lim_{j \rightarrow \infty} x_{k_j} = x^*,$$

где x^* — решение задачи (1).

Доказательство. Воспользуемся определением неусеченной итерации и формулой (29). Запишем

$$\delta_{k_j} = \tilde{\delta}_{k_j} = \frac{\Delta(u_{k_j})}{\|a_{i_{k_j}'} - a_{i_{k_j}''}\|^2} \geq \frac{\Delta(u_{k_j})}{d^2},$$

где $d = \max_{i \in 1:m, j \in 1:m} \|a_i - a_j\|$. Подставив это неравенство в (34) при $k = k_j$, получим

$$Q(u_{k_j}) - Q(u_{k_j+1}) \geq \frac{[\Delta(u_{k_j})]^2}{2d^2}. \quad (37)$$

Поскольку последовательность $\{Q(u_k)\}$ сходится, то левая часть неравенства (37) стремится к нулю при $j \rightarrow \infty$. Отсюда следует, что

$$\lim_{j \rightarrow \infty} \Delta(u_{k_j}) = 0. \quad (38)$$

На основании (15) приходим к заключению теоремы.

Вернемся к предельному соотношению (35) и выясним, чему равна константа α .

Теорема 6. *Справедливо предельное соотношение*

$$\lim_{k \rightarrow \infty} Q(u_k) = Q(u^*),$$

где u^* — некоторое решение задачи (2).

Так что $\alpha = Q(u^*)$.

Доказательство. По определению множества планов U задачи (2) последовательность $\{u_{k_j}\}$ ограничена. Из нее можно выделить сходящуюся подпоследовательность. Не умаляя общности, будем считать, что вся последовательность $\{u_{k_j}\}$ сходится к вектору $\check{u}$,

$$\lim_{j \rightarrow \infty} u_{k_j} = \check{u}. \quad (39)$$

Очевидно, что $\check{u}$ — план задачи (2). Покажем, что это оптимальный план.

Достаточно проверить, что $\Delta(\check{u}) = 0$. Допустим противное $\Delta(\check{u}) > 0$. Выберем столь большое J , чтобы при $j \geq J$ выполнялись неравенства

$$u_{k_j}[i] > 0 \quad \text{при } i \in I^+(\check{u}), \quad (40)$$

$$\left| (\langle a_i, x_{k_j} \rangle - b[i]) - (\langle a_i, x^* \rangle - b[i]) \right| < \varepsilon \quad \text{при } i \in 1:m. \quad (41)$$

Здесь выбрано $\varepsilon = \frac{1}{4}\Delta(\check{u})$, x^* — решение задачи (1). В силу (40) имеем $I^+(\check{u}) \subset I^+(u_{k_j})$. Далее

$$\begin{aligned} \Delta(u_{k_j}) &= \max_{i \in I^+(u_{k_j})} \{ \langle a_i, x_{k_j} \rangle - b[i] \} - \min_{i \in 1:m} \{ \langle a_i, x_{k_j} \rangle - b[i] \} \geq \\ &\geq \max_{i \in I^+(\check{u})} \{ \langle a_i, x_{k_j} \rangle - b[i] \} - \min_{i \in 1:m} \{ \langle a_i, x_{k_j} \rangle - b[i] \}. \end{aligned}$$

Теперь воспользуемся неравенством (41). Получим

$$\begin{aligned} \Delta(u_{k_j}) &\geq \max_{i \in I^+(\check{u})} \{ \langle a_i, x^* \rangle - b[i] - \varepsilon \} - \min_{i \in 1:m} \{ \langle a_i, x^* \rangle - b[i] + \varepsilon \} = \\ &= \max_{i \in I^+(\check{u})} \{ \langle a_i, x^* \rangle - b[i] \} - \min_{i \in 1:m} \{ \langle a_i, x^* \rangle - b[i] \} - 2\varepsilon. \end{aligned} \quad (42)$$

Из (39) следует, что

$$\lim_{j \rightarrow \infty} Au_{k_j} = A\check{u},$$

так что, по теореме 5, $x^* = A\check{u} =: \check{x}$. Заменим в (42) x^* на $\check{x}$. Придем к неравенству

$$\Delta(u_{k_j}) \geq \Delta(\check{u}) - 2\varepsilon = \frac{1}{2}\Delta(\check{u}) > 0.$$

Это противоречит соотношению (38).

Установлено, что $\check{u}$ оптимальный план задачи (2). Будем писать u^* вместо $\check{u}$. На основании (39) по непрерывности получаем

$$\lim_{j \rightarrow \infty} Q(u_{k_j}) = Q(u^*).$$

Строгое убывание последовательности $\{Q(u_k)\}$ гарантирует справедливость заключения теоремы.

В теореме 6, по существу, установлена слабая сходимость последовательности $\{u_k\}$.

Теорема 7 (о сильной сходимости последовательности $\{x_k\}$). *Справедливо предельное соотношение*

$$\lim_{k \rightarrow \infty} \|x_k - x^*\| = 0,$$

где x^* — решение задачи (1).

Доказательство. Согласно теореме 1 имеем

$$\|x^*\|^2 = \langle x^*, Ax^* \rangle = \langle A^T x^*, u^* \rangle = \sum_{i=1}^m u^*[i] (\langle a_i, x^* \rangle - b[i]) + b[i] =$$

$$= \langle b, u^* \rangle + \sum_{i \in I^+(u^*)} u^*[i] \mu^* = \langle b, u^* \rangle + \mu^*.$$

Отсюда следует, что

$$\mu^* - \frac{1}{2} \|x^*\|^2 = \frac{1}{2} \|x^*\|^2 - \langle b, u^* \rangle = Q(u^*). \quad (43)$$

Аналогично

$$\begin{aligned} \langle x^*, x_k \rangle &= \langle A^T x^*, u_k \rangle = \sum_{i=1}^m u_k[i] (\langle a_i, x^* \rangle - b[i]) + b[i] \geq \\ &\geq \langle b, u_k \rangle + \min_{i \in 1:m} \{ \langle a_i, x^* \rangle - b[i] \} = \langle b, u_k \rangle + \mu^*. \end{aligned}$$

Теперь запишем

$$\begin{aligned} \|x_k - x^*\|^2 &= \|x_k\|^2 - 2\langle x^*, x_k \rangle + \|x^*\|^2 \leq \\ &\leq \|x_k\|^2 - 2(\langle b, u_k \rangle + \mu^*) + \|x^*\|^2 = \\ &= 2\left(\frac{1}{2} \|x_k\|^2 - \langle b, u_k \rangle\right) - 2(\mu^* - \frac{1}{2} \|x^*\|^2). \end{aligned}$$

Согласно (43) получаем

$$\|x_k - x^*\|^2 = 2(Q(u_k) - Q(u^*)).$$

Остается сослаться на теорему 6. Теорема доказана.

5. ПРИМЕРЫ

Рассмотрим четыре характерных примера на применение MDM-алгоритма к решению задачи Сильвестра на плоскости. Вычисления заканчивались, когда на очередном приближении u_k выполнялось одно из условий:

$$\sqrt{\Delta(u_k)} < 5 \cdot 10^{-3}$$

или

$$\|x_k - x_{k+1}\| < 10^{-5}.$$

При описании результатов используется обозначение $R_k = \sqrt{-2Q(u_k)}$ для приближенного значения радиуса минимального круга.

Пример 1. Возьмем три точки, лежащих на окружности, с центром в начале координат и радиусом 2 (см. точки синего цвета на фиг. 1):

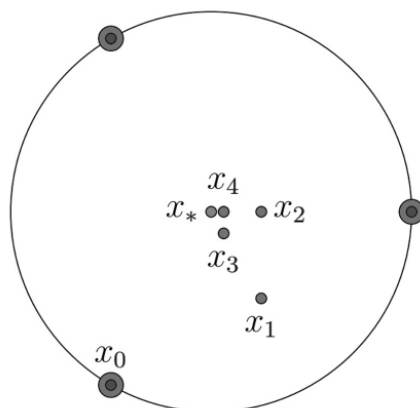
$$a_1 = \begin{pmatrix} -1 \\ -\sqrt{3} \end{pmatrix}, \quad a_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}.$$

В качестве начального приближения возьмем вектор $u_0 = (1, 0, 0)$, что соответствует $x_0 = a_1$. На фиг. 1 зеленым цветом отмечен искомый центр минимальной окружности. Точки на окружности, выделенные красным цветом, являются опорными.

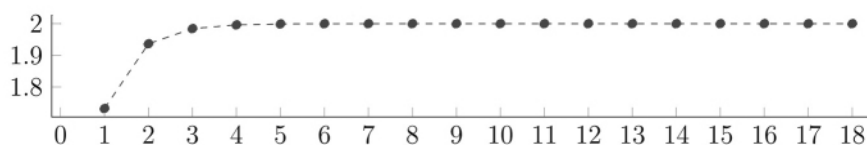
Ниже в виде графиков приведены результаты вычислений. Всюду по оси абсцисс указывается номер итерации.

В качестве начального приближения можно взять центр тяжести заданной системы $a_1, \dots, a_m$, а именно вектор u_0 с компонентами $u_0[i] = \frac{1}{m}$, $i \in 1 : m$. Можно поступить иначе: искусственно к исходному набору точек a_k добавить еще одну точку $a_{m+1} := \frac{1}{m} \sum_{i=1}^m a_i$. При этом минимальный круг сохранится. В качестве начального приближения возьмем план u_0 с компонентами $u_0[i] = 0$, $i \in 1 : m$, $u_0[m+1] = 1$.

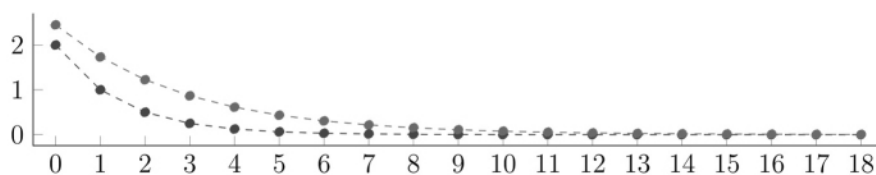
Пусть в примере 1 начальное приближение строится с привлечением дополнительной точки $a_4 = 0$. Положим $u_0 = (0, 0, 0, 1)$, что соответствует $x_0 = a_4 = 0$. На фиг. 6–10 приведены результаты расчетов.



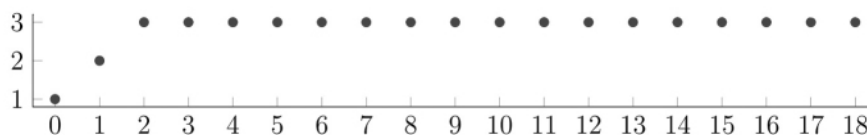
Фиг. 1. Ход решения при $u_0 = (1, 0, 0)$.



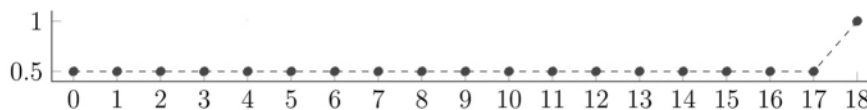
Фиг. 2. График последовательности R_k .



Фиг. 3. Графики оценок $\bullet \|x_k - x_*\|$ и $\bullet \sqrt{\Delta(u_k)}$.

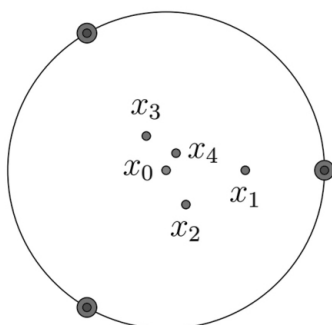
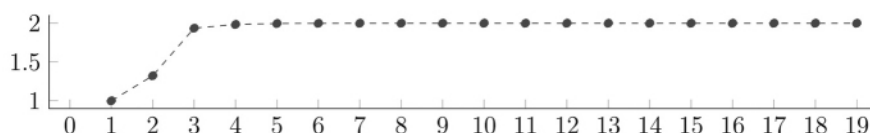
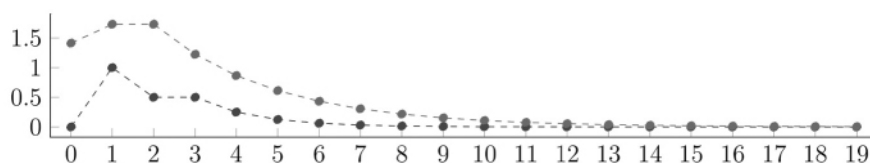
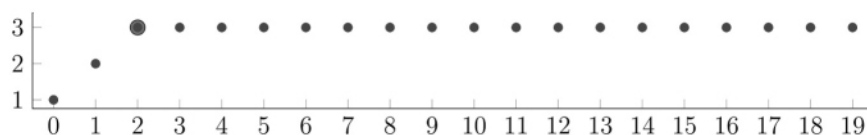
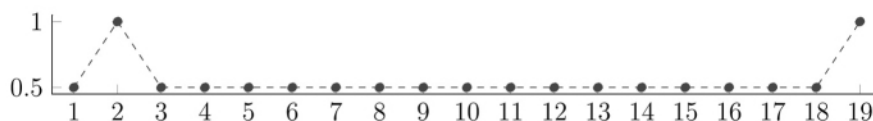


Фиг. 4. График последовательности $|I^+(u_k)|$.



Фиг. 5. График отношения $\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|}$.

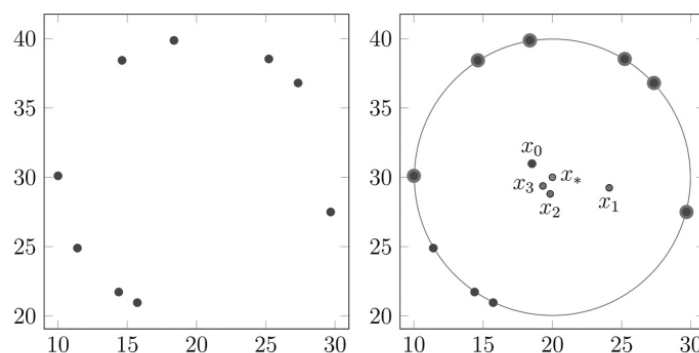
На фиг. 8 видно, что план u_0 не оптимальный, $\sqrt{\Delta(u_0)} > 0$. Вместе с тем, соответствующая ему точка $x_0 = 0$ является решением исходной задачи (1). Оптимальным планом задачи (2) будет вектор $(1/3, 1/3, 1/3, 0)$.

Фиг. 6. Ход решения при $u_0 = (0, 0, 0, 1)$.Фиг. 7. График последовательности R_k .Фиг. 8. Графики оценок $\bullet \|x_k - x_*$ и $\bullet \sqrt{\Delta(u_k)}$.Фиг. 9. График последовательности $|I^+(u_k)|$.Фиг. 10. График отношения $\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|}$.

На фиг. 9 красным цветом выделена усеченная итерация. В обоих случаях вычисления закончились по условию $\sqrt{\Delta(u_k)} < 0.005$.

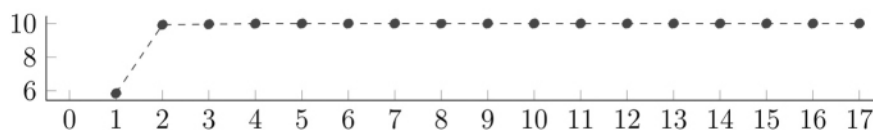
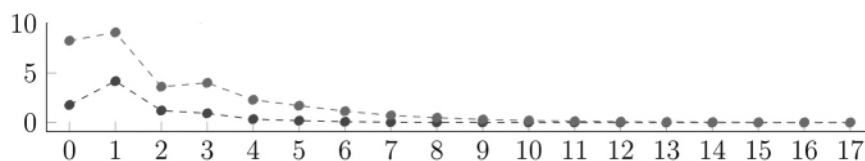
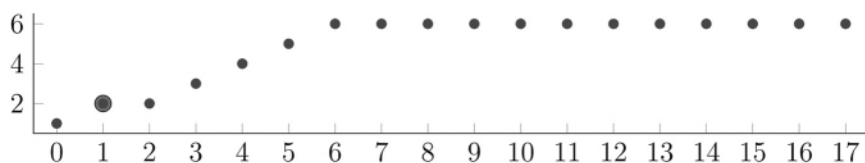
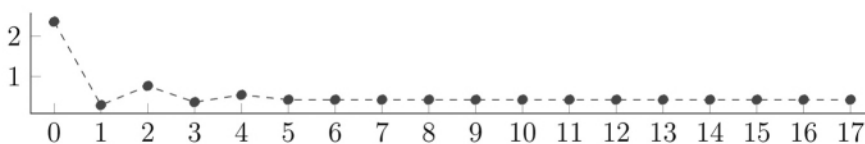
Пример 2. Предположим, что нужно выяснить, лежит ли данный набор точек на некоторой окружности. Это можно сделать, решая соответствующую задачу Сильвестра.

В примере генерировалось 9 случайных точек, лежащих на границе круга $(x - 20)^2 + (y - 30)^2 \leq 10$ (см. фиг. 11).

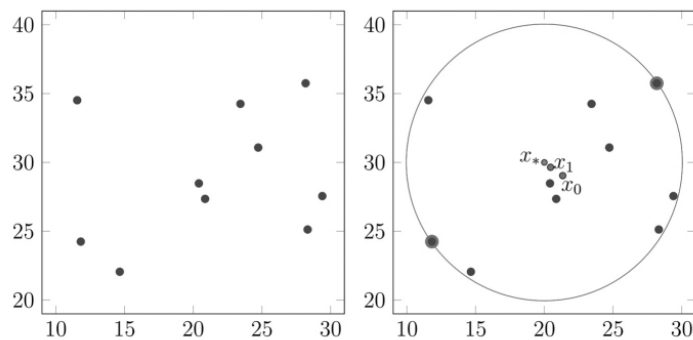


Фиг. 11. Исходные данные и ход решения задачи.

Начальное приближение строилось с привлечением дополнительной точки. На графиках 12–15 представлены соответствующие результаты расчетов. Вычисления закончились на 17-й итерации по достижению условия $\|x_{16} - x_{17}\| < 10^{-5}$.

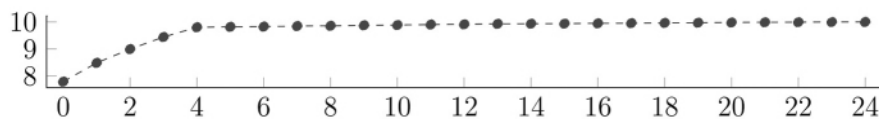
Фиг. 12. График последовательности R_k .Фиг. 13. Графики оценок $\|x_k - x_*$ и $\sqrt{\Delta(u_k)}$.Фиг. 14. График последовательности $|I^+(u_k)|$.Фиг. 15. График отношения $\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|}$.

Пример 3. В этом примере генерировалось 10 случайных точек, лежащих в круге $(x - 20)^2 + (y - 30)^2 \leq 10$ (см. фиг. 16).

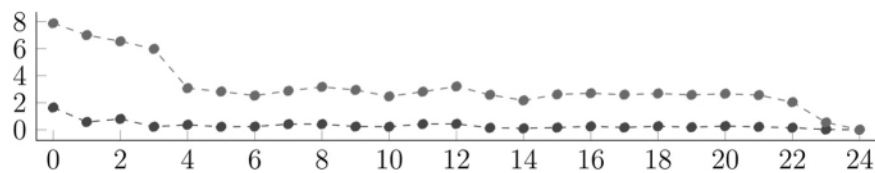


Фиг. 16. Исходные данные и ход решения задачи.

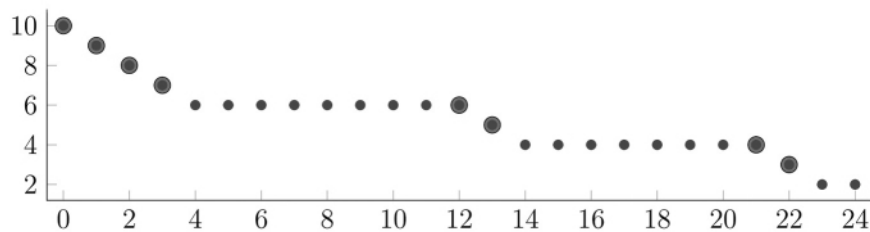
На фиг. 16–20 представлены результаты решения задачи Сильвестра, когда в качестве начального приближения выбран центр тяжести исходных точек. В этом случае мощность носителя $|I^+(u_0)|$ максимальна и равна 10. На фиг. 19 отчетливо демонстрируется нецелесообразность выбора начального приближения, у которого мощность носителя сильно превышает размерность пространства. Вычисления закончились на 24-й итерации при выполнении правила останова $\sqrt{\Delta(u_{24})} < 0.005$.



Фиг. 17. График последовательности R_k .

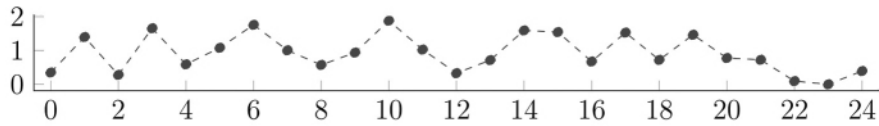


Фиг. 18. Графики оценок $\|x_k - x_*\|$ и $\sqrt{\Delta(u_k)}$.

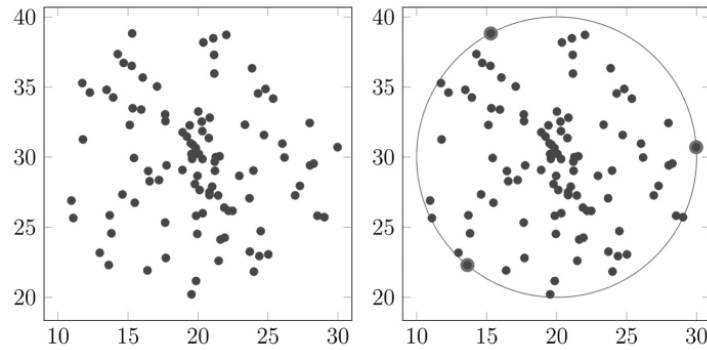


Фиг. 19. График последовательности $|I^+(u_k)|$.

Пример 4. Решим задачу Сильвестра со 100 случайными точками лежащими в круге $(x - 20)^2 + (y - 30)^2 \leq 10$ (см. фиг. 21).

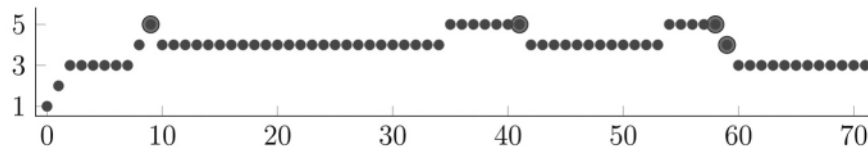


Фиг. 20. График отношения $\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|}$.

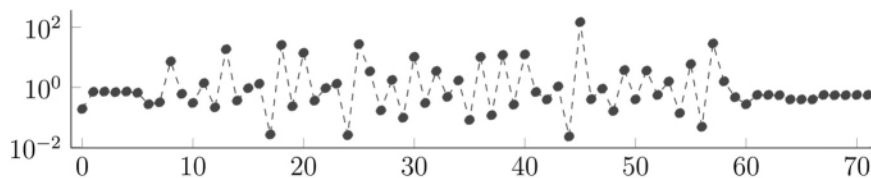


Фиг. 21. Исходные данные и решение задачи.

На фиг. 22, 23 представлены результаты расчетов, когда в качестве начального приближения выбрана одна из исходных точек. Вычисления закончились на 71-й итерации при выполнении условия $\|x_{70} - x_{71}\| < 10^{-5}$.



Фиг. 22. График последовательности $|I^+(u_k)|$.



Фиг. 23. График отношения $\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|}$.

Замечание 2. Эффективность MDM-алгоритма решения задачи Сильвестра зависит от величины $|I^+(u^*)|$. Наиболее благоприятный случай, когда $|I^+(u^*)| = n + 1$.

Замечание 3. Если план u_k порождает усеченную итерацию, то при переходе от u_k к u_{k+1} мощность $|I^+(u_k)|$ уменьшается на единицу при $u_k[i'_k] > 0$ и остается прежней при $u_k[i'_k] = 0$ (см. доказательство леммы 2).

6. ЗАКЛЮЧЕНИЕ

Известно, что задача Сильвестра о шаре наименьшего объема, содержащего m заданных точек в n -мерном евклидовом пространстве, сводится к выпуклой задаче квадратичного программирования [2, с. 117, 118]. Двойственная к ней задача после перемены знака у целевой функции принимает

вид (2). Обычно решается именно задача (2). По её оптимальному плану легко восстанавливается единственное решение задачи Сильвестра. Для решения задачи (2) в квадратичном программировании имеются конечные методы (например, метод Данцига [9]), которые эффективно работают при относительно небольших значениях n и m . На каждом шаге таких методов нужно решать систему линейных уравнений и проверять некоторые условия. При больших n и m погрешность вычислений становится критической, что часто приводит к неверному результату. В связи с этим разрабатываются простые итеративные методы, бесконечные в общем случае, с гарантированной сходимостью. В данной статье предлагается такой метод для решения задачи (2). Он аналогичен MDM-алгоритму для нахождения точки выпуклой оболочки конечного множества с наименьшей нормой [4]. Приводятся примеры решения задачи Сильвестра на плоскости. Они преследуют скромную цель — пояснить, как выбирать начальное приближение, и продемонстрировать характер сходимости MDM-алгоритма с учетом усеченных и неусеченных итераций.

В статье используется принципиальная схема MDM-алгоритма, ориентированная на доказательство его сходимости. Однако существует и рабочая схема, которая применяется при решении реальных задач. Вопрос о рабочей схеме и связанный с ним вопрос о сравнении MDM-алгоритма с другими методами решения задачи Сильвестра, в данной работе не рассматриваются. Авторы планируют посвятить этой теме отдельное исследование, в котором будет учтена информация из обзорной статьи [10].

СПИСОК ЛИТЕРАТУРЫ

1. Зуховицкий С. И. Алгоритм для отыскания точки, наименее уклоняющейся (в смысле П. Л. Чебышева) от данной системы m точек // ДАН УССР, 1951, № 6. С. 404–407.
2. Гавурин М. К., Малоземов В. Н. Экстремальные задачи с линейными ограничениями. Л.: Изд-во ЛГУ, 1984. 176 с.
3. Малоземов В. Н., Плоткин А. В. Двойственность в квадратичном программировании. Задача Сильвестра // Семинар “CNSA & NDO”. Избранные доклады. 8 декабря 2021 г. (дата обращения: 31.01.2024).
4. Митчелл Б. Ф., Демьянов В. Ф., Малоземов В. Н. Нахождение ближайшей к началу координат точки многогранника // Вестник ЛГУ. 1971. № 19. С. 38–45.
5. Малоземов В. Н. МДМ-методу — 50 лет // Семинар “CNSA & NDO”. Избранные доклады. 10 ноября 2021 г. (дата обращения: 31.01.2024).
6. Lopez J., Barbero A., Dorronsoro J. R. On the equivalence of the SMO and MDM algorithms for SVM training / Springer-Verlag Berlin Heidelberg. W. Daelemans et al. (Eds.): ECML PKDD 2008, Part I, LNAI 5211, pp. 288–300.
7. Малозёмов В. Н., Соловьева Н. А. МДМ-метод для решения общей квадратичной задачи математической диагностики // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2023. 10(3). С. 516–529.
8. Малоземов В. Н., Соловьева Н. А., Тамасян Г. Ш. MDM-алгоритм и задача Сильвестра // Математические методы распознавания образов: Тезисы докладов 21-й Всероссийской конф. с международным участием, Москва, 12–15 декабря 2023 года. М.: РАН, 2023. С. 87–89.
9. Даугавет В. А. Численные методы квадратичного программирования. СПб.: Изд-во СПбГУ, 2004. 128 с.
10. E. Alper Yildirim. Two algorithms for the minimum enclosing ball problem // SIAM J. OPTIM. Vol. 19. N 3. 2008. P. 1368–1391.

MDM ALGORITHM AND SYLVESTER'S PROBLEM

V. N. Malozemov^{a,*}, N. A. Solovyova^{b,**}, G. Sh. Tamasyan^{c,d,***}^a 199034, St. Petersburg, Universitetskaya Emb., 7/9, S.-Pb State University, Russia^b 191023, St. Petersburg, Griboyedova Channel Emb., 30/32, S.-Pb State Economy. Univ., Russia^c 197198, St. Petersburg, Zhdanovskaya Str., 13, A. F. Mozhaisky MCA^d 198178, St. Petersburg, Bolshoy Ave. VO, 61, IAM RAS, Russia

*e-mail: v.malozemov@spbu.ru

**e-mail: 4vinyo@gmail.com

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

Received February 6, 2024

Revised March 1, 2024

Accepted April 5, 2024

Abstract. When developing numerical methods for solving nonlinear minimax problems, the following auxiliary problem arose: in the convex hull of some finite set in Euclidean space, find the point with the lowest norm. In 1971, B. Mitchell, V. Demyanov and V. Malozemov proposed a non-standard algorithm for solving this problem, which later became known as the MDM algorithm (based on the capital letters of the authors' surnames). This article discusses a specific minimax problem: to find a ball of the smallest volume containing a given finite set of points. It is called the Sylvester problem and is a special case of the Chebyshev center of the set problem. The convex quadratic programming problem with simplex constraints is compared to the Sylvester problem. To solve this problem, the article suggests using a variant of the MDM algorithm. With its help, a minimizing sequence of plans is built, such that only two components differ from neighboring plans. The numbers of these components are selected based on certain optimality conditions. The weak convergence of the obtained sequence of plans is proved, from which follows the norm convergence of the corresponding sequence of vectors to the only solution of the Sylvester problem. Four typical examples on the plane are given.

Keywords: Sylvester's problem, quadratic programming, MDM algorithm.

О НОВЫХ КЛАССАХ РЕШЕНИЙ ЗАДАЧ И О σ -КОММУТИРОВАНИИ ($\sigma \neq 0, \pm 1$) ТЁПЛИЦЕВОЙ И ГАНКЕЛЕВОЙ МАТРИЦ В РАМКАХ УНИФИЦИРОВАННОГО ПОДХОДА¹⁾

© 2024 г. В. Н. Чугунов^{1,*}, Х. Д. Икрамов^{2,**}

¹119333 Москва, ул. Губкина, 8, ИВМ РАН, Россия

²119992 Москва, Ленинские горы, МГУ, ВМК, Россия

*e-mail: chugunov.vadim@gmail.com

**e-mail: ikramov@cs.msu.su

Поступила в редакцию 03.04.2024 г.

Переработанный вариант 03.04.2024 г.

Принята к публикации 03.04.2024 г.

В предыдущей работе авторов был предложен унифицированный подход к конструированию пар матриц (T, H) , решающих задачу о σ -коммутировании тёплицевой и ганкелевой матриц. В настоящей статье этот подход применяется к выводу новых классов решений. Библ. 8.

Ключевые слова: тёплицева матрица, ганкелева матрица, σ -коммутирование.

DOI: 10.31857/S0044466924070043, EDN: xiuhsq

1. ВВЕДЕНИЕ

Понятие σ -коммутирования матриц является относительно новым и еще недостаточно изученным в линейной алгебре.

Матрицы K и M σ -коммутируют (или квази-коммутируют), если найдется такое число σ , что $KM = \sigma MK$ (см. [1]). В этой же работе [1] отмечается, что квази-коммутативность является важным соотношением в квантовой физике (см. [2], [3]), а также в теории представлений аффинных алгебр Гекке (Неске) (см. [4]).

Задача о σ -коммутировании тёплицевой и ганкелевой матриц заключается в описании пар ненулевых матриц (T, H) таких, что T — тёплицева, H — ганкелева и выполняется соотношение

$$TH = \sigma HT. \quad (1)$$

Необходимо сказать несколько слов о параметре σ . Произвольный выбор σ возможен лишь в случае, когда хотя бы одна из матриц T или H вырождена. Если же обе матрицы невырождены, то σ является одним из корней n -й степени из единицы. В настоящей работе от параметра σ мы требуем, чтобы $\sigma \neq 0, \pm 1$.

Заметим, что так как след произведения двух матриц не меняется при перестановке сомножителей и $\sigma \neq 1$, то матрицы TH и HT имеют нулевой след.

Хотя задача о σ -коммутировании тёплицевой и ганкелевой матриц пока не имеет полного решения, в [5] были описаны некоторые частные множества требуемых пар матриц (T, H) . Пусть $\mathcal{P}_n$ — $n \times n$ матрица-перестановка

¹⁾Работа первого автора поддержана Московским центром фундаментальной и прикладной математики в ИВМ РАН (Соглашение № 075-15-2022-286 с Минобрнауки РФ).

$$\mathcal{P}_n = \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & \cdots & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix}, \quad (2)$$

называемая иногда перьединичной матрицей.

Теорема 1. Ненулевые тѐплицева матрица T и ганкелева матрица H σ -коммутируют ($\sigma \neq 0, \pm 1$), если T и H входят хотя бы в один из описываемых ниже классов:

Класс 1'. Матрица T является циркулянтном

$$T = F_n^* D_1 F_n,$$

а H — ганкелевым циркулянтном

$$H = F_n^* D_2 F_n \mathcal{P}_n.$$

Здесь F_n — (нормированная) матрица дискретного преобразования Фурье

$$F_n = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \varepsilon & \varepsilon^2 & \cdots & \varepsilon^{n-1} \\ 1 & \varepsilon^2 & \varepsilon^4 & \cdots & \varepsilon^{2(n-1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \varepsilon^{n-1} & \varepsilon^{2(n-1)} & \cdots & \varepsilon^{(n-1)^2} \end{pmatrix},$$

$\varepsilon = \exp\left(\frac{2\pi i}{n}\right)$ — первообразный корень n -й степени из единицы; $D_1 = \text{diag}\left(d_1^{(1)}, d_2^{(1)}, \dots, d_n^{(1)}\right)$ и $D_2 = \text{diag}\left(d_1^{(2)}, d_2^{(2)}, \dots, d_n^{(2)}\right)$ — диагональные матрицы; при этом

$$d_1^{(2)} d_1^{(1)} = 0,$$

$$d_j^{(2)} \left(d_j^{(1)} - \sigma d_{n+2-j}^{(1)}\right) = 0, \quad j = 2, 3, \dots, n.$$

Класс 2'. Матрица T является косым циркулянтном

$$T = G_{-1} F_n^* D_1 F_n G_{-1}^*,$$

а H — ганкелевым косым циркулянтном

$$H = G_{-1} F_n^* D_2 F_n G_{-1}^* \mathcal{P}_n,$$

где

$$G_{-1} = \text{diag}(1, \psi, \psi^2, \dots, \psi^{n-1}),$$

$\psi = e^{i\pi/n}$ есть корень n -й степени из (-1) . Диагональные матрицы $D_1 = \text{diag}\left(d_1^{(1)}, d_2^{(1)}, \dots, d_n^{(1)}\right)$ и $D_2 = \text{diag}\left(d_1^{(2)}, d_2^{(2)}, \dots, d_n^{(2)}\right)$ должны удовлетворять соотношениям

$$\begin{aligned} d_1^{(2)} \left(d_1^{(1)} - \sigma d_2^{(1)}\right) &= 0, & d_2^{(2)} \left(d_2^{(1)} - \sigma d_1^{(1)}\right) &= 0, \\ d_j^{(2)} \left(d_j^{(1)} - \sigma d_{n+3-j}^{(1)}\right) &= 0, & j &= 3, 4, \dots, n. \end{aligned}$$

Класс 3'. Пусть $n = 2r$, матрицы T и H имеют вид

$$T = \alpha_1 \begin{pmatrix} 0_{r,r} & I_r \\ 0_{r,r} & 0_{r,r} \end{pmatrix}, \quad H = \beta_1 \begin{pmatrix} \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \sigma \mathcal{P}_r \end{pmatrix},$$

α_1, β_1 — произвольные числа.

Класс 4'. Пусть $n = 2r$, матрицы T и H имеют вид

$$T = \alpha_1 \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ I_r & 0_{r,r} \end{pmatrix}, \quad H = \beta_1 \begin{pmatrix} \sigma \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \mathcal{P}_r \end{pmatrix},$$

α_1, β_1 — произвольные числа.

Замечание 1. Классы 1 и 2 были выделены в [5] во множествах пар матриц (T, H) , в которых T — тёплицев, а H — ганкелев циркулянт или же обе матрицы суть косые циркулянты.

Данные множества получены из различных соображений. Позднее, в работе [6] авторами был предложен унифицированный подход к получению полного набора требуемых матриц. Сущность его состоит в сужении множества всех пар матриц (T, H) до множеств, объединению которых принадлежат все решения рассматриваемой задачи, после чего задача о σ -коммутировании исследуется на каждой конкретной более узкой комбинации наборов (T, H) . Основную роль в этом процессе играет следующее утверждение.

Лемма 1 (см. [6]). *Всякую пару (T, H) , решающую задачу о σ -коммутировании тёплицевой и ганкелевой матриц, можно представить в виде*

$$\begin{aligned} T &= \alpha_2 (A - \sigma A^\top), \\ H &= \beta_2 B \mathcal{P}_n, \end{aligned} \quad (3)$$

где A и B — нескаларные тёплицевы матрицы, удовлетворяющие условиям

$$\begin{aligned} AB &= BA, \\ A^\top B + BA^\top &= \mu AB, \\ \mu &= \frac{1 + \sigma^2}{\sigma}, \\ \mu &\neq \pm 2, \end{aligned} \quad (4)$$

α_2, β_2 — произвольные числа.

Сформулированная лемма позволяет считать основным уравнением соотношение

$$A^\top B + BA^\top = \mu AB. \quad (5)$$

Заметим, что в данном уравнении матрицы A и B определены с точностью до скалярного множителя.

Для упрощения уравнения (5) воспользуемся следующим утверждением.

Лемма 2 (см. [7]). *Две нескаларные тёплицевы матрицы $\check{T}_1$ и $\check{T}_2$ коммутируют тогда и только тогда, когда $\check{T}_1$ и $\check{T}_2$ принадлежат хотя бы одному из следующих классов:*

Класс КТМ_1. Обе матрицы $\check{T}_1$ и $\check{T}_2$ верхнетреугольные или же обе нижнетреугольные.

Класс КТМ_2. Матрицы $\check{T}_1$ и $\check{T}_2$ суть φ -циркулянты для одного и того же числа $\varphi \neq 0$.

Класс КТМ_3. Одна из матриц $\check{T}_1$ или $\check{T}_2$ является линейной функцией от другой.

Согласно приведенной лемме, для коммутирующих тёплицевых матриц A и B возможны лишь следующие четыре случая: 1) обе матрицы A и B являются верхними треугольными; 2) обе матрицы A и B — нижние треугольные; 3) обе матрицы A и B суть φ -циркулянты для одного и того же числа $\varphi \neq 0$; 4) $B = \delta A + \gamma I_n$.

Целью настоящей работы является применение унифицированного подхода к конструированию пар матриц (T, H) , решающих задачу о σ -коммутировании тёплицевой и ганкелевой матриц, для получения новых классов решений. А именно, будет полностью исследован случай, когда $B = \delta A + \gamma I_n$ и $\mu \neq 0$.

Введем вспомогательную операцию. Пару пар матриц $(\widetilde{T}, \widetilde{H})$ назовем $\mathcal{P}_n$ -преобразованием пары (T, H) , если

$$\widetilde{T} = \mathcal{P}_n T \mathcal{P}_n, \quad \widetilde{H} = \mathcal{P}_n H \mathcal{P}_n.$$

Учитывая персимметричность тёплицевых матриц, имеем

$$\widetilde{T} = T^\top, \quad \widetilde{H} = \mathcal{P}_n H \mathcal{P}_n.$$

Очевидно, что операция $\mathcal{P}_n$ -преобразования пары (T, H) сохраняет тёплицевость матрицы T . Ниже мы покажем, что эта операция сохраняет и ганкелевость матрицы H , а также свойство σ -коммутирования в случае σ -коммутирования самих матриц T и H .

Структура статьи следующая. В разд. 2 формулируется теорема, являющаяся главным результатом статьи и содержащая описание новых классов пар σ -коммутирующих матриц (T, H) с точностью до $\mathcal{P}_n$ -преобразования. Доказательство теоремы проводится в разд. 3.

Для дальнейшего нам понадобятся перечисляемые ниже факты, первый из которых мы доказываем, а остальные формулируем, указывая источник.

Лемма 3. *Если произведение нескаларных симметричных или кососимметричных тёплицевых матриц T_1 и T_2 является тёплицевой матрицей, то матрицы T_1 и T_2 будут обе циркулянтами или обе косыми циркулянтами.*

Доказательство леммы 3. Известен следующий факт, принадлежащий своего рода тёплицеву фольклору: произведение произвольных нескаларных тёплицевых матриц T_1 и T_2 (необязательно симметричных или кососимметричных) является тёплицевой матрицей тогда и только тогда, когда T_1 и T_2 либо верхне- (нижне-) треугольные матрицы либо φ -циркулянты для одного и того же числа φ . Первый случай невозможен в силу нескаларности и одной из симметрий. Значит, T_1 и T_2 суть φ -циркулянты.

Рассмотрим матрицу T_1 с элементами первой строки $t_0, t_1, \dots, t_{n-1}$ и первого столбца $t_0, t_{-1}, \dots, t_{-(n-1)}$. Пусть она является нескаларным симметричным φ -циркулянт, тогда для ненулевого внедиагонального элемента t_j имеем

$$\begin{aligned} t_j &= \{\text{матрица } T_1 \text{ симметрична}\} = t_{-j} = \\ &= \{\text{матрица } T_1 - \varphi\text{-циркулянт}\} = \varphi t_{n-j} = \\ &= \{\text{матрица } T_1 \text{ симметрична}\} = \varphi t_{-(n-j)} = \\ &= \{\text{матрица } T_1 - \varphi\text{-циркулянт}\} = \varphi^2 t_j \end{aligned}$$

или

$$(1 - \varphi^2) t_j = 0.$$

Так как $t_{j_0} \neq 0$ для некоторого j_0 в силу нескаларности T_1 , то $\varphi^2 = 1$, т.е. T_1 и T_2 — суть циркулянты или косые циркулянты.

Если же матрица T_1 является нескаларным кососимметричным φ -циркулянт, то для ненулевого элемента t_j имеем

$$\begin{aligned} t_j &= \{\text{матрица } T_1 \text{ кососимметрична}\} = -t_{-j} = \\ &= \{\text{матрица } T_1 - \varphi\text{-циркулянт}\} = -\varphi t_{n-j} = \\ &= \{\text{матрица } T_1 \text{ кососимметрична}\} = \varphi t_{-(n-j)} = \\ &= \{\text{матрица } T_1 - \varphi\text{-циркулянт}\} = \varphi^2 t_j \end{aligned}$$

или

$$(1 - \varphi^2) t_j = 0.$$

Поскольку $t_{j_0} \neq 0$ для некоторого $j_0 \neq 0$ в силу нескаларности T_1 , то $\varphi^2 = 1$, т.е. снова T_1 и T_2 суть циркулянты или косые циркулянты. Лемма 3 доказана.

Лемма 4 (см. [8]). *Ненулевые симметричная тёплицева матрица T_1 и кососимметричная тёплицева матрица T_2 антикоммутируют, т.е.*

$$T_1 T_2 + T_2 T_1 = 0, \quad (6)$$

тогда и только тогда, когда они принадлежат одному из следующих классов:

Класс АСТКТМ_1. Матрицы T_1 и T_2 — циркулянты, делящие нуль, т.е. $T_1 T_2 = 0$.

Класс АСТКТМ_2. Матрицы T_1 и T_2 суть косые циркулянты, делящие нуль.

Класс АСТКТМ_3. Матрицы T_1 и T_2 имеют вид

$$T_1 = \alpha_3 (L + L^\top), \quad T_2 = \beta_3 (L - L^\top),$$

где L — ненулевая строго нижнетреугольная тѐплицева матрица, для которой $L^2 = 0$; α_3, β_3 — произвольные числа.

Класс АСТКТМ_4. Матрицы T_1 и T_2 задаются формулами

$$T_1 = \alpha_3 C, \quad T_2 = \beta_3 (S - CSC).$$

Здесь C — симметричный инволютивный циркулянт, S — косо циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы C ; α_3, β_3 — произвольные числа.

Класс АСТКТМ_5. Матрицы T_1 и T_2 задаются формулами

$$T_1 = \alpha_3 S, \quad T_2 = \beta_3 (C - SCS).$$

Здесь S — симметричный инволютивный косо циркулянт, C — циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы S ; α_3, β_3 — произвольные числа.

Класс АСТКТМ_6. Порядок n — четное число. Матрицы T_1 и T_2 задаются формулами

$$T_1 = \alpha_3 (C - SCS), \quad T_2 = \beta_3 S,$$

где S — кососимметричный инволютивный косо циркулянт, а C — циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы S ; α_3, β_3 — произвольные числа.

Класс АСТКТМ_7. Матрицы T_1 и T_2 задаются формулами

$$T_1 = \alpha_3 (Z + Z^\top), \quad T_2 = \beta_3 (Z - Z^\top),$$

где Z — инволютивный φ -циркулянт, причем $|\varphi| = 1$, $\varphi \neq \pm 1$; α_3, β_3 — произвольные числа.

Лемма 5 (см. [8]). Пусть S — ненулевой кососимметричный косо циркулянт, C — нескаларный циркулянт. Матрица $SC - CS^\top$ тогда и только тогда является косым циркулянтом, когда выполняются следующие условия: n — четное число, векторы, составленные из поддиагональных элементов первых столбцов в матрицах C и S , пропорциональны, и матрица S с точностью до скалярного множителя есть инволюция.

Лемма 6 (см. [8]). Пусть C — циркулянт, а S — косо циркулянт, причем обе матрицы нескаларные и имеют один и тот же порядок. Матрица $SC - CS^\top$ тогда и только тогда является тѐплицевой, когда векторы, составленные из поддиагональных элементов первых столбцов в C и S , коллинеарны.

Лемма 7 (см. [8]). Пусть C — симметричный и нескаларный циркулянт, а S — кососимметричный и нескаларный косо циркулянт. Матрица $SC - CS^\top$ тогда и только тогда является циркулянтом, когда векторы, составленные из поддиагональных элементов первых столбцов в матрицах C и S , пропорциональны, а циркулянт C с точностью до скалярного множителя есть инволюция.

Лемма 8 (см. [8]). Пусть C — циркулянт, а S — косо циркулянт, причем обе матрицы нескаларные и имеют один и тот же порядок. Матрица $CS - SC^\top$ тогда и только тогда является тѐплицевой, когда векторы, составленные из поддиагональных элементов первых столбцов в C и S , коллинеарны.

Лемма 9 (см. [8]). Пусть S — симметричный и нескаларный косо циркулянт, а C — кососимметричный и нескаларный циркулянт. Матрица $CS - SC^\top$ тогда и только тогда является косым циркулянтом, когда векторы, составленные из поддиагональных элементов первых столбцов в матрицах C и S , пропорциональны, а косо циркулянт S с точностью до скалярного множителя есть инволюция.

Лемма 10 (см. [8]). Пусть T — тѐплицева $n \times n$ -матрица вида $T = t_0 I_n + L + L^\top$, где L — нижняя треугольная тѐплицева матрица с первым столбцом $(l_0 = 0, l_1, \dots, l_{n-1})^\top$ такая, что $L^2 = 0$. Если T является нескаларным циркулянтом или косым циркулянтом, то n — четное число $n = 2r$, а T — циркулянт вида

$$T = \begin{pmatrix} t_0 I_r & l_r I_r \\ l_r I_r & t_0 I_r \end{pmatrix}. \quad (7)$$

Доказательство леммы 10. Рассмотрим отдельно случаи нечетного и четного порядка матрицы T .

Пусть сначала $n = 2r + 1$, тогда из условия $L^2 = 0$ следует, что $l_j = 0$ для $j = 1, 2, \dots, r$. Из вида матрицы T имеем соотношения $\{T\}_{1, n+1-j} = l_{n-j}$ при $j = 1, 2, \dots, r$. Так как T — циркулянт или

косой циркулянт, то $l_{n-j} = \{T\}_{1,n+1-j} = \pm \{T\}_{j+1,1} = \pm l_j = 0$ для $j = 1, 2, \dots, r$. Получаем, что $L = 0$ и T — скалярная матрица.

Если же n — четное число $n = 2r$, то из условия $L^2 = 0$ следует, что $l_j = 0$ для $j = 1, 2, \dots, r-1$. Аналогично, из вида матрицы T имеем соотношения $\{T\}_{1,n+1-j} = l_{n-j}$ при $j = 1, 2, \dots, r-1$. Так как T — циркулянт или косой циркулянт, то $l_{n-j} = \{T\}_{1,n+1-j} = \pm \{T\}_{j+1,1} = \pm l_j = 0$ для $j = 1, 2, \dots, r-1$. Приходим к тому, что $l_j = 0$ для $j = 1, 2, \dots, r-1, r+1, \dots, n-1$. Значит, в силу нескаларности матрица T имеет вид (7) и является циркулянтом. Лемма 10 доказана.

2. ГЛАВНЫЙ РЕЗУЛЬТАТ

Прежде, чем сформулировать главный результат, остановимся на некоторых особенностях $\mathcal{P}_n$ -преобразования пары (T, H) в пару $(\tilde{T}, \tilde{H})$.

Если положить $H = \check{T}\mathcal{P}_n$ для некоторой тёплицевой матрицы $\check{T}$, то

$$\tilde{H} = \mathcal{P}_n H \mathcal{P}_n = \mathcal{P}_n \check{T} \mathcal{P}_n \mathcal{P}_n = \mathcal{P}_n \check{T} = \check{T}^\top \mathcal{P}_n.$$

Таким образом, $\mathcal{P}_n$ -преобразование сохраняет ганкелевость матрицы $\tilde{H}$.

В случае, когда матрицы T и H σ -коммутируют, можем записать

$$\tilde{T} \tilde{H} = \mathcal{P}_n T \mathcal{P}_n \mathcal{P}_n H \mathcal{P}_n = \mathcal{P}_n T H \mathcal{P}_n = \sigma \mathcal{P}_n H T \mathcal{P}_n = \sigma \mathcal{P}_n H \mathcal{P}_n \mathcal{P}_n T \mathcal{P}_n = \sigma \tilde{H} \tilde{T}.$$

Тем самым $\mathcal{P}_n$ -преобразование сохраняет σ -коммутирование тёплицевой и ганкелевой матриц.

Теорема 2. Пусть $\hat{L}$ — нижняя треугольная тёплицева матрица порядка r , где $r = \lfloor n/2 \rfloor$. Ненулевые тёплицева матрица T и ганкелева матрица H σ -коммутируют, если эти матрицы (с точностью до $\mathcal{P}_n$ -преобразования) входят хотя бы в один из следующих классов:

Класс 1. Если порядок матриц является четным числом $n = 2r$, то сами матрицы T и H имеют вид

$$T = \alpha \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \hat{L} & 0_{r,r} \end{pmatrix}, \quad H = \beta \begin{pmatrix} \sigma \hat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \hat{L} \mathcal{P}_r \end{pmatrix},$$

α, β — произвольные числа.

Класс 2. Если порядок матриц является нечетным числом $n = 2r + 1$, то сами матрицы T и H имеют вид

$$T = \alpha \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \hat{L} & 0_r & 0_{r,r} \end{pmatrix}, \quad H = \beta \begin{pmatrix} \sigma \hat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r}^\top \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \hat{L} \mathcal{P}_r \end{pmatrix},$$

α, β — произвольные числа.

3. ОБОСНОВАНИЕ ГЛАВНОГО РЕЗУЛЬТАТА

Рассмотрим уравнение (5) при условии $B = \delta A + \gamma I_n$.

Начнем исследование со специального случая, когда $\delta = 0$. Поскольку A и B определены с точностью до скалярных множителей, то можно считать, что $B = I_n$ и уравнение (5) имеет вид

$$2A^\top = \mu A.$$

Отсюда путем несложных выкладок выводим соотношение

$$\left(1 - \frac{\mu^2}{4}\right) A = 0,$$

которое не может быть выполнено, так как $A \neq 0$ и $\mu \neq \pm 2$. Следовательно, $\delta \neq 0$. Так как матрицы A и B определены с точностью до скалярных множителей, то δ можно положить равным единице, и тогда $B = A + \gamma I_n$. Решаемое уравнение приобретает вид

$$A^\top A + A A^\top = \mu A^2 + \gamma \mu A - 2\gamma A^\top. \quad (8)$$

В дальнейшем будем считать, что $\mu \neq 0$. Исследуем несколько взаимоисключающих случаев.

1. Если матрица A симметрична, то приходим к уравнению

$$2A^2 = \mu A^2 + \gamma \mu A - 2\gamma A,$$

которое запишем в виде

$$(\mu - 2)(A^2 + \gamma A) = 0.$$

Поскольку $\mu \neq 2$, получаем условие

$$A^2 + \gamma A = 0$$

или

$$A(A + \gamma I_n) = 0.$$

Произведение тёплицевых матриц в левой части дает (нулевую) тёплицеву матрицу. Так как матрица A симметрична, то, по лемме 3, A и, поэтому, B суть циркулянты или косые циркулянты. В таком случае исходные матрицы T и H являются тёплицевыми и ганкелевыми циркулянтами либо косыми циркулянтами. Эти ситуации уже были рассмотрены согласно замечанию 1.

2. Матрица A является линейным многочленом от кососимметричной матрицы K , т.е. $A = a_0 I_n + K$. В этом случае решаемое уравнение

$$\begin{aligned} (a_0 I_n - K)(a_0 I_n + K) + (a_0 I_n + K)(a_0 I_n - K) = \\ = \mu(a_0 I_n + K)^2 + \gamma \mu(a_0 I_n + K) - 2\gamma(a_0 I_n - K) \end{aligned}$$

можно упростить до вида

$$2a_0^2 I_n - 2K^2 = \mu a_0^2 I_n + 2\mu a_0 K + \mu K^2 + (\mu - 2)\gamma a_0 I_n + (\mu + 2)\gamma K,$$

допускающего запись

$$(\mu - 2)a_0^2 I_n + (\mu - 2)\gamma a_0 I_n + (\mu + 2)K^2 = -(2\mu a_0 + (\mu + 2)\gamma)K.$$

В левой части стоит симметричная матрица, в правой части — кососимметричная. Приходим к системе

$$\begin{aligned} (\mu - 2)a_0^2 I_n + (\mu - 2)\gamma a_0 I_n + (\mu + 2)K^2 &= 0, \\ (2\mu a_0 + (\mu + 2)\gamma)K &= 0. \end{aligned}$$

Перепишем первое уравнение системы в виде

$$(\mu + 2)K^2 = -(\mu - 2)a_0^2 I_n - (\mu - 2)\gamma a_0 I_n.$$

Так как $\mu \neq -2$, то K^2 является тёплицевой матрицей. В силу косой симметрии матрицы K по лемме 3 сама матрица K и, поэтому, A и B суть циркулянты или косые циркулянты. В таком случае исходные матрицы T и H являются тёплицевыми и ганкелевыми циркулянтами либо косыми циркулянтами, что было уже рассмотрено согласно замечанию 1.

3. Матрица A не является ни симметричной, ни линейным многочленом от кососимметричной матрицы. В этом случае введем вспомогательные тёплицевы матрицы — симметричную матрицу X и кососимметричную матрицу Y — формулами

$$X = \frac{A + A^\top}{2}, \quad Y = \frac{A - A^\top}{2},$$

тогда

$$A = X + Y, \quad A^\top = X - Y. \quad (9)$$

В силу условий рассматриваемого случая матрицы X и Y не скалярны. Подставляя эти выражения в (8), получаем соотношение

$$(X - Y)(X + Y) + (X + Y)(X - Y) = \mu(X + Y)^2 + \gamma \mu(X + Y) - 2\gamma(X - Y),$$

которое можно преобразовать к виду

$$2X^2 - 2Y^2 = \mu X^2 + \mu(XY + YX) + \mu Y^2 + \gamma(\mu - 2)X + \gamma(\mu + 2)Y$$

или к виду

$$(2 - \mu)X^2 - (2 + \mu)Y^2 + \gamma(2 - \mu)X = \mu(XY + YX) + \gamma(\mu + 2)Y.$$

В левой части стоит симметричная матрица, в правой — кососимметричная. Приходим к системе

$$\begin{aligned} (2 - \mu)X^2 - (2 + \mu)Y^2 + \gamma(2 - \mu)X &= 0, \\ \mu(XY + YX) + \gamma(\mu + 2)Y &= 0 \end{aligned}$$

или

$$\begin{aligned} (\mu - 2)X^2 + (\mu + 2)Y^2 + \gamma(\mu - 2)X &= 0, \\ \mu(XY + YX) + \gamma(\mu + 2)Y &= 0. \end{aligned} \quad (10)$$

Учитывая, что $\sigma \neq \pm 1$, от искоемых матриц X и Y перейдем к симметричной матрице X_0 и кососимметричной матрице Y_0 . Эти тёплицевы матрицы определяются формулами

$$X = X_0, \quad Y = \frac{\sigma - 1}{\sigma + 1} Y_0. \quad (11)$$

Вспомяная, что $\mu = \frac{1+\sigma^2}{\sigma}$, можем записать

$$\begin{aligned} X^2 &= X_0^2, \\ Y^2 &= \frac{(\sigma - 1)^2}{(\sigma + 1)^2} Y_0^2 = \frac{\sigma^2 + 1 - 2\sigma}{\sigma^2 + 1 + 2\sigma} Y_0^2 = \frac{\frac{1+\sigma^2}{\sigma} - 2}{\frac{1+\sigma^2}{\sigma} + 2} Y_0^2 = \frac{\mu - 2}{\mu + 2} Y_0^2. \end{aligned}$$

После подстановки этих соотношений в систему (10) и почленного деления первого уравнения на ненулевое число $(\mu - 2)$, а второго на $\frac{\sigma-1}{\sigma+1}$, получаем

$$\begin{aligned} X_0^2 + Y_0^2 + \gamma X_0 &= 0, \\ \mu(X_0 Y_0 + Y_0 X_0) + \gamma(\mu + 2)Y_0 &= 0. \end{aligned} \quad (12)$$

Положим

$$X_0 = X_1 - \frac{\gamma}{2} I_n, \quad Y_0 = Y_1 \quad (13)$$

и подставим эти выражения в (12):

$$\begin{aligned} X_1^2 + Y_1^2 &= \frac{\gamma^2}{4} I_n, \\ \mu(X_1 Y_1 + Y_1 X_1) + 2\gamma Y_1 &= 0. \end{aligned} \quad (14)$$

Так как $\mu \neq 0$, то можем переписать эту систему в виде

$$\begin{aligned} X_1^2 + Y_1^2 &= \frac{\gamma^2}{4} I_n, \\ \left(X_1 + \frac{\gamma}{\mu} I_n\right) Y_1 + Y_1 \left(X_1 + \frac{\gamma}{\mu} I_n\right) &= 0. \end{aligned} \quad (15)$$

Введем тёплицевы симметричную матрицу Q и кососимметричную матрицу V такие, что

$$X_1 = Q - \frac{\gamma}{\mu} I_n, \quad Y_1 = V.$$

Из (9), (11), (13) получаем

$$A = X + Y = X_0 + \frac{\sigma - 1}{\sigma + 1} Y_0 = X_1 - \frac{\gamma}{2} I_n + \frac{\sigma - 1}{\sigma + 1} Y_1 = Q - \left(\frac{\gamma}{\mu} + \frac{\gamma}{2}\right) I_n + \frac{\sigma - 1}{\sigma + 1} V, \quad (16)$$

$$B = A + \gamma I_n,$$

где матрицы Q и V таковы, что

$$\begin{aligned} Q^2 - \frac{2\gamma}{\mu} Q + V^2 &= \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n, \\ QV + VQ &= 0. \end{aligned} \quad (17)$$

Таким образом, тёплицевы симметричная матрица Q и кососимметричная матрица V должны антикоммутировать. При этом ни Q , ни V не должны быть скалярными матрицами, так как в противном случае A является скалярной или симметричной матрицей либо линейным многочленом от кососимметричной матрицы, а эти случаи уже были разобраны. Согласно лемме 4, Q и V принадлежат хотя бы одному из семи классов АСТКТМ.

3.1. Если Q и V — циркулянты, то A и B тоже циркулянты, и тогда исходные матрицы T и H будут соответственно циркулянтном и ганкелевым циркулянтном. Этот случай был разобран (см. замечание 1).

3.2. Если Q и V — косые циркулянты, то A и B тоже косые циркулянты, и тогда исходные матрицы T и H будут соответственно косым циркулянтном и ганкелевым косым циркулянтном. Этот случай также был разобран.

3.3. Пусть теперь

$$Q = \alpha_3 (L + L^\top), \quad V = \beta_3 (L - L^\top),$$

где L — ненулевая строго нижнетреугольная тёплицева матрица, для которой $L^2 = 0$. Подставляя эти выражения в первое уравнение системы (17), получаем

$$\alpha_3^2 (LL^\top + L^\top L) - \frac{2\gamma\alpha_3}{\mu} (L + L^\top) - \beta_3^2 (LL^\top + L^\top L) = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n$$

или

$$(\alpha_3^2 - \beta_3^2) (LL^\top + L^\top L) - \frac{2\gamma\alpha_3}{\mu} (L + L^\top) = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n. \quad (18)$$

Так как $L^2 = 0$, то, вводя параметр $r = \lfloor n/2 \rfloor$ и нижнюю треугольную тёплицеву матрицу $\hat{L}$ порядка r , можем записать, что

$$L = \begin{pmatrix} 0_{n-r,r} & 0_{n-r,n-r} \\ \hat{L} & 0_{r,n-r} \end{pmatrix}.$$

Рассмотрим сначала случай, когда $\beta_3 = \pm\alpha_3$. Так как Q — не скалярная матрица, то $\alpha_3 \neq 0$ и $L \neq 0$. Из соотношения (18) выводим, что $\gamma = 0$ и, значит, $B = A$.

Введем параметр $\vartheta = \pm 1$, тогда $\beta_3 = \vartheta\alpha_3$ и из (16) имеем

$$\begin{aligned} A &= Q + \frac{\sigma - 1}{\sigma + 1} V = \alpha_3 (L + L^\top) + \beta_3 \frac{\sigma - 1}{\sigma + 1} (L - L^\top) = \\ &= \frac{\alpha_3}{\sigma + 1} \left\{ (\sigma + 1) (L + L^\top) + \vartheta(\sigma - 1) (L - L^\top) \right\} = \\ &= \frac{\alpha_3}{\sigma + 1} \left\{ [\sigma(\vartheta + 1) + (1 - \vartheta)] L + [\sigma(1 - \vartheta) + (1 + \vartheta)] L^\top \right\}. \end{aligned}$$

Дальнейший анализ разобьем на два подслучая $\vartheta = -1$ и $\vartheta = 1$. Пусть сначала $\vartheta = -1$. В этом случае

$$A = B = \frac{2\alpha_3}{\sigma + 1} (L + \sigma L^\top).$$

Для четного $n = 2r$ можем написать

$$A = B = \frac{2\alpha_3}{\sigma + 1} \begin{pmatrix} 0_{r,r} & \sigma \hat{L}^\top \\ \hat{L} & 0_{r,r} \end{pmatrix},$$

и тогда

$$\begin{aligned} T &= \alpha_2 (A - \sigma A^\top) = \frac{2\alpha_2\alpha_3}{\sigma + 1} \left[\begin{pmatrix} 0_{r,r} & \sigma \hat{L}^\top \\ \hat{L} & 0_{r,r} \end{pmatrix} - \sigma \begin{pmatrix} 0_{r,r} & \hat{L}^\top \\ \sigma \hat{L} & 0_{r,r} \end{pmatrix} \right] = \\ &= \frac{2\alpha_2\alpha_3(1 - \sigma^2)}{\sigma + 1} \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \hat{L} & 0_{r,r} \end{pmatrix} = \alpha \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \hat{L} & 0_{r,r} \end{pmatrix}, \end{aligned}$$

$$H = \beta_2 B \mathcal{P}_n = \frac{2\alpha_3 \beta_2}{\sigma + 1} \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \widehat{L} \mathcal{P}_r \end{pmatrix} = \beta \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \widehat{L} \mathcal{P}_r \end{pmatrix}.$$

Приходим к паре матриц класса 1 с $\alpha = 2\alpha_2\alpha_3(1 - \sigma)$ и $\beta = \frac{2\alpha_3\beta_2}{\sigma+1}$.

Если же $n = 2r + 1$, то

$$A = B = \frac{2\alpha_3}{\sigma + 1} \begin{pmatrix} 0_{r,r} & 0_r & \sigma \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix},$$

и тогда

$$\begin{aligned} T = \alpha_2 (A - \sigma A^\top) &= \frac{2\alpha_2\alpha_3}{\sigma + 1} \left[\begin{pmatrix} 0_{r,r} & 0_r & \sigma \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} - \sigma \begin{pmatrix} 0_{r,r} & 0_r & \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \sigma \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} \right] = \\ &= \frac{2\alpha_2\alpha_3(1 - \sigma^2)}{\sigma + 1} \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} = \alpha \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix}, \\ H = \beta_2 B \mathcal{P}_n &= \frac{2\alpha_3\beta_2}{\sigma + 1} \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \widehat{L} \mathcal{P}_r \end{pmatrix} = \beta \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \widehat{L} \mathcal{P}_r \end{pmatrix}. \end{aligned}$$

Получаем пару матриц из класса 2 с $\alpha = 2\alpha_2\alpha_3(1 - \sigma)$ и $\beta = \frac{2\alpha_3\beta_2}{\sigma+1}$.

Если $\vartheta = 1$, то имеем

$$A = B = \frac{2\alpha_3}{\sigma + 1} (\sigma L + L^\top).$$

В случае четного $n = 2r$ можем написать

$$A = B = \frac{2\alpha_3}{\sigma + 1} \begin{pmatrix} 0_{r,r} & \widehat{L}^\top \\ \sigma \widehat{L} & 0_{r,r} \end{pmatrix},$$

и тогда

$$\begin{aligned} T = \alpha_2 (A - \sigma A^\top) &= \frac{2\alpha_2\alpha_3}{\sigma + 1} \left[\begin{pmatrix} 0_{r,r} & \widehat{L}^\top \\ \sigma \widehat{L} & 0_{r,r} \end{pmatrix} - \sigma \begin{pmatrix} 0_{r,r} & \sigma \widehat{L}^\top \\ \widehat{L} & 0_{r,r} \end{pmatrix} \right] = \\ &= \frac{2\alpha_2\alpha_3(1 - \sigma^2)}{\sigma + 1} \begin{pmatrix} 0_{r,r} & \widehat{L}^\top \\ 0_{r,r} & 0_{r,r} \end{pmatrix} = \\ &= 2\alpha_2\alpha_3(1 - \sigma) \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \widehat{L} & 0_{r,r} \end{pmatrix}^\top = \alpha \mathcal{P}_n \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \widehat{L} & 0_{r,r} \end{pmatrix} \mathcal{P}_n, \\ H = \beta_2 B \mathcal{P}_n &= \frac{2\alpha_3\beta_2}{\sigma + 1} \begin{pmatrix} \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \sigma \widehat{L} \mathcal{P}_r \end{pmatrix} = \beta \mathcal{P}_n \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \widehat{L} \mathcal{P}_r \end{pmatrix} \mathcal{P}_n. \end{aligned}$$

Эта пара матриц представляет собой $\mathcal{P}_n$ -преобразование пары матриц класса 1 с $\alpha = 2\alpha_2\alpha_3(1 - \sigma)$ и $\beta = \frac{2\alpha_3\beta_2}{\sigma+1}$.

Если же $n = 2r + 1$, то

$$A = B = \frac{2\alpha_3}{\sigma + 1} \begin{pmatrix} 0_{r,r} & 0_r & \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \sigma \widehat{L} & 0_r & 0_{r,r} \end{pmatrix},$$

и

$$T = \alpha_2 (A - \sigma A^\top) = \frac{2\alpha_2\alpha_3}{\sigma + 1} \left[\begin{pmatrix} 0_{r,r} & 0_r & \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \sigma \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} - \sigma \begin{pmatrix} 0_{r,r} & 0_r & \sigma \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} \right] =$$

$$\begin{aligned}
&= \frac{2\alpha_2\alpha_3(1-\sigma^2)}{\sigma+1} \begin{pmatrix} 0_{r,r} & 0_r & \widehat{L}^\top \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & 0_{r,r} \end{pmatrix} = \\
&= 2\alpha_2\alpha_3(1-\sigma) \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix}^\top = \alpha\mathcal{P}_n \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} \mathcal{P}_n, \\
H &= \beta_2 B\mathcal{P}_n = \frac{2\alpha_3\beta_2}{\sigma+1} \begin{pmatrix} \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \sigma\widehat{L}\mathcal{P}_r \end{pmatrix} = \\
&= \beta \begin{pmatrix} \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \sigma\widehat{L}\mathcal{P}_r \end{pmatrix} = \beta\mathcal{P}_n \begin{pmatrix} \sigma\widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \widehat{L}\mathcal{P}_r \end{pmatrix} \mathcal{P}_n.
\end{aligned}$$

Эта пара матриц представляет собой $\mathcal{P}_n$ -преобразование пары матриц класса 2 с $\alpha = 2\alpha_2\alpha_3(1-\sigma)$ и $\beta = \frac{2\alpha_3\beta_2}{\sigma+1}$.

Далее считаем, что $\alpha_3^2 - \beta_3^2 \neq 0$. Перепишем уравнение (18) в виде

$$(L + L^\top)^2 - \frac{2\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)}(L + L^\top) = \frac{1}{(\alpha_3^2 - \beta_3^2)}\left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right)I_n$$

или

$$\left(L + L^\top - \frac{\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)}I_n\right)^2 = \left[\frac{1}{(\alpha_3^2 - \beta_3^2)}\left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) + \frac{\gamma^2\alpha_3^2}{\mu^2(\alpha_3^2 - \beta_3^2)^2}\right]I_n.$$

В силу леммы 3 матрица $U = L + L^\top - \frac{\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)}I_n$, квадрат которой есть скалярная матрица, должна быть симметричным циркулянтном или косым циркулянтном. Из леммы 10 следует, что $n = 2r$ — четное число и U — циркулянт вида (7). Тогда Q и V с точностью до скалярных множителей можно записать как

$$Q = \alpha_3 \begin{pmatrix} 0_{r,r} & I_r \\ I_r & 0_{r,r} \end{pmatrix}, \quad V = \beta_3 \begin{pmatrix} 0_{r,r} & -I_r \\ I_r & 0_{r,r} \end{pmatrix}.$$

В этом случае $Q^2 = \alpha_3^2 I_n$ и $V^2 = -\beta_3^2 I_n$. Подстановка в первое уравнение системы (17) дает условие

$$\frac{2\gamma}{\mu}Q = \alpha_3^2 I_n - \beta_3^2 I_n - \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right)I_n.$$

Поскольку Q — не скалярная матрица, отсюда следует, что $\gamma = 0$, а тогда $\beta_3^2 = \alpha_3^2$, вопреки предположению.

3.4. Пусть теперь тёплицевы симметричная матрица Q и кососимметричная матрица V задаются формулами

$$Q = \alpha_3 C, \quad V = \beta_3 (S - CSC),$$

где C — симметричный инволютивный не скалярный циркулянт, а S — косой циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы C .

Подстановка матрицы Q в первое уравнение системы (17) дает условие

$$\alpha_3^2 I_n - \frac{2\gamma\alpha_3}{\mu}C + V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right)I_n,$$

из которого выводим

$$V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} - \alpha_3^2\right)I_n + \frac{2\gamma\alpha_3}{\mu}C.$$

Отсюда следует, что V^2 является циркулянтном. По лемме 3 сама матрица V должна быть циркулянтном или косым циркулянтном. Но если V — косой циркулянт, то C — скалярная матрица, вопреки предположению. Приходим к уже разобранному случаю, когда Q и V — циркулянты.

3.5. Рассмотрим класс $АСТКТМ_5$, в котором матрицы Q и V задаются формулами

$$Q = \alpha_3 S, \quad V = \beta_3 (C - SCS),$$

где S — симметричный инволютивный нескаларный косой циркулянт, а C — циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы S .

Подстановка матрицы Q в первое уравнение системы (17) дает условие

$$\alpha_3^2 I_n - \frac{2\gamma\alpha_3}{\mu} S + V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n,$$

из которого выводим

$$V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} - \alpha_3^2 \right) I_n + \frac{2\gamma\alpha_3}{\mu} S.$$

Отсюда следует, что V^2 является косым циркулянтном. По лемме 3 сама матрица V должна быть циркулянтном или косым циркулянтном. Но если V — циркулянт, то S — скалярная матрица, вопреки предположению. Приходим к уже разобранному случаю, когда Q и V — косые циркулянты.

3.6. Исследуем теперь симметричную и кососимметричную тёплицевы матрицы из класса $АСТКТМ_6$. В этом случае

$$Q = \alpha_3 (C - SCS), \quad V = \beta_3 S,$$

где S — кососимметричный инволютивный косой циркулянт, а C — циркулянт с поддиагональными элементами, совпадающими с поддиагональными элементами матрицы S ; n — четное число.

Подставим V в первое уравнение системы (17)

$$\left(Q - \frac{\gamma}{\mu} I_n \right)^2 = \left(\frac{\gamma^2}{4} - \beta_3^2 \right) I_n.$$

Поскольку матрица Q симметрична, то, согласно лемме 3, она должна быть циркулянтном или косым циркулянтном.

В данном случае разложение Q в сумму циркулянта и косого циркулянта имеет вид

$$\begin{aligned} Q = \alpha_3 (C - SCS) &= 2\alpha_3 \frac{1}{2} (C - SCS) = 2\alpha_3 \left(C - \frac{1}{2} (C + SCS) \right) = \\ &= 2\alpha_3 \left(C - \frac{1}{2} S (SC - CS^\top) \right), \end{aligned}$$

так как по лемме 5 матрица $SC - CS^\top$ является косым циркулянтном.

Предположим сначала, что Q есть косой циркулянт. Тогда C — скалярная матрица и, поскольку поддиагональные элементы в S и C совпадают, то скалярными матрицами будут и S , и сама Q .

Если Q — симметричный циркулянт, то

$$C + SCS = \zeta I_n$$

для некоторого числа ζ . Введем вспомогательный циркулянт C_1 формулой

$$C = C_1 + \frac{\zeta}{2} I_n,$$

тогда

$$C_1 + SC_1 S = 0$$

или

$$SC_1 - C_1 S^\top = 0.$$

В силу леммы 7 матрица $SC_1 - C_1S^\top$ является циркулянтном тогда и только тогда, когда циркулянт C_1 есть скалярное кратное симметричной инволюции. При этом у матриц S и C_1 совпадают поддиагональные элементы. Представим эти матрицы в виде

$$S = L - L^\top, \quad C_1 = c_0 I_n + L + L^\top,$$

где L — ненулевая строго нижняя треугольная тёплицева матрица. Условия инволютивности S и того факта, что C_1 — скалярное кратное симметричной инволюции, запишем в виде системы

$$\begin{aligned} (L - L^\top)^2 &= I_n, \\ (c_0 I_n + L + L^\top)^2 &= \lambda I_n, \end{aligned}$$

которую после раскрытия скобок можно переписать как

$$\begin{aligned} L^2 + (L^\top)^2 - LL^\top - L^\top L &= I_n, \\ c_0^2 I_n + 2c_0 L + 2c_0 L^\top + L^2 + (L^\top)^2 + LL^\top + L^\top L &= \lambda I_n. \end{aligned}$$

Складывая уравнения системы, получаем соотношение

$$c_0^2 I_n + 2c_0 L + 2c_0 L^\top + 2L^2 + 2(L^\top)^2 = (1 + \lambda) I_n,$$

приводимое к виду

$$L(c_0 I_n + L) + L^\top(c_0 I_n + L^\top) = \frac{1 + \lambda - c_0^2}{2} I_n.$$

Поскольку L — тёплицева строго нижняя треугольная матрица, то из этого равенства следует, что

$$L(c_0 I_n + L) = 0.$$

В силу условия $L \neq 0$ заключаем, что $c_0 = 0$ и $L^2 = 0$. Теперь из леммы 10 выводим, что $n = 2r$ — четное число и C_1 — циркулянт вида (7). Сами матрицы Q и V с точностью до скалярных множителей можно записать как

$$Q = v_1 \begin{pmatrix} 0_{r,r} & I_r \\ I_r & 0_{r,r} \end{pmatrix}, \quad V = v_2 \begin{pmatrix} 0_{r,r} & I_r \\ -I_r & 0_{r,r} \end{pmatrix}.$$

Приходим к матрицам Q и V из п. 3.3.

3.7. И, наконец, исследуем ситуацию, когда нескаларные матрицы Q и V задаются формулами

$$Q = \alpha_3 (Z + Z^\top), \quad V = \beta_3 (Z - Z^\top),$$

где Z — инволютивный φ -циркулянт, причем $|\varphi| = 1$, $\varphi \neq \pm 1$.

Подставляя выражения для Q и V в первое уравнение системы (17), получаем

$$\begin{aligned} 2\alpha_3^2 I_n + \alpha_3^2 (ZZ^\top + Z^\top Z) - \frac{2\gamma\alpha_3}{\mu} (Z + Z^\top) + 2\beta_3^2 I_n - \\ - \beta_3^2 (ZZ^\top + Z^\top Z) = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n. \end{aligned}$$

Вычитая из обеих частей $4\beta_3^2 I_n$, имеем

$$(\alpha_3^2 - \beta_3^2) (2I_n + ZZ^\top + Z^\top Z) - \frac{2\gamma\alpha_3}{\mu} (Z + Z^\top) = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n - 4\beta_3^2 I_n$$

или

$$(\alpha_3^2 - \beta_3^2) (Z + Z^\top)^2 - \frac{2\gamma\alpha_3}{\mu} (Z + Z^\top) = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n - 4\beta_3^2 I_n.$$

Рассмотрим сначала случай $\alpha_3 = \pm\beta_3$. Из нескаларности Q выводим, что $\gamma = 0$, а тогда $\beta_3 = 0$ и $\alpha_3 = 0$. В результате приходим к равенству $T = H = 0$, которое противоречит условию, что T и H — ненулевые матрицы.

Считая в дальнейшем, что $\alpha_3^2 - \beta_3^2 \neq 0$, разделим последнее соотношение на $\alpha_3^2 - \beta_3^2$:

$$\left(Z + Z^\top\right)^2 - \frac{2\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)}(Z + Z^\top) = \frac{1}{\alpha_3^2 - \beta_3^2} \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n - \frac{4\beta_3^2}{\alpha_3^2 - \beta_3^2} I_n.$$

Добавляя к обеим частям матрицу $\frac{\gamma^2\alpha_3^2}{\mu^2(\alpha_3^2 - \beta_3^2)^2} I_n$, приходим к равенству

$$\begin{aligned} \left(Z + Z^\top\right)^2 - \frac{2\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)}(Z + Z^\top) + \frac{\gamma^2\alpha_3^2}{\mu^2(\alpha_3^2 - \beta_3^2)^2} I_n = \\ = \frac{1}{\alpha_3^2 - \beta_3^2} \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n - \frac{4\beta_3^2}{\alpha_3^2 - \beta_3^2} I_n + \frac{\gamma^2\alpha_3^2}{\mu^2(\alpha_3^2 - \beta_3^2)^2} I_n, \end{aligned}$$

допускающему запись в виде

$$\begin{aligned} \left(Z + Z^\top - \frac{\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)} I_n\right)^2 = \\ = \frac{1}{\alpha_3^2 - \beta_3^2} \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n - \frac{4\beta_3^2}{\alpha_3^2 - \beta_3^2} I_n + \frac{\gamma^2\alpha_3^2}{\mu^2(\alpha_3^2 - \beta_3^2)^2} I_n. \end{aligned}$$

Получаем, что квадрат симметричной матрицы $Z + Z^\top - \frac{\gamma\alpha_3}{\mu(\alpha_3^2 - \beta_3^2)} I_n$ является скалярной матрицей. В силу леммы 3, сама эта матрица, как и матрица $Z + Z^\top$, должна быть либо циркулянт, либо косым циркулянт.

Пусть сначала матрица $Z + Z^\top$ — циркулянт. Положим $Z = C + S$, где C — циркулянт, а S — косой циркулянт; при этом S имеет нулевую диагональ. Из условия, что $Z + Z^\top$ — циркулянт, получаем соотношение $S^\top = -S$. Тогда $Z + Z^\top = C + C^\top$ и $Q = \alpha_3(C + C^\top)$.

Запишем условия инволютивности матрицы $Z = C + S$ и транспонированной матрицы $Z^\top$:

$$\begin{aligned} C^2 + CS + SC + S^2 &= I_n, \\ (C^\top)^2 - C^\top S - SC^\top + S^2 &= I_n. \end{aligned}$$

Разность этих соотношений дает равенство

$$S(C + C^\top) - (C + C^\top)S^\top = (C^\top)^2 - C^2. \quad (19)$$

Согласно лемме 7 матрица $C + C^\top$ является скалярным кратным инволютивной матрицы. То же самое можно сказать о матрице Q . Положим $Q^2 = \lambda I_n$ и подставим это равенство в первое уравнение (17):

$$\lambda I_n - \frac{2\gamma}{\mu} Q + V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n.$$

Отсюда находим

$$V^2 = \frac{2\gamma}{\mu} Q + \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2}\right) I_n - \lambda I_n.$$

Квадрат тѐплицевой кососимметричной матрицы V является циркулянт.

Если $\gamma \neq 0$, то V^2 — нескаларный циркулянт, и тогда в силу леммы 3 матрица V — циркулянт. Отсюда следует, что A и B — циркулянты, а сами матрицы T и H будут тѐплицевым и ганкелевым циркулянтами. Этот случай уже был разобран. Аналогично обосновывается случай $\gamma = 0$ и V — циркулянт.

Рассмотрим теперь ситуацию, когда $\gamma = 0$ и V — скалярное кратное инволютивного косого циркулянта. Из соотношений

$$V = \beta_3 (Z - Z^\top) = \beta_3 (C + S - C^\top - S^\top) = \beta_3 (C + 2S - C^\top)$$

следует, что $C^\top = C$, $V = 2\beta_3 S$. Подстановка равенства $C^\top = C$ в (19) дает условие

$$SC - CS^\top = 0.$$

Согласно лемме 6, C и S обладают тем свойством, что векторы, составленные из поддиагональных элементов их первых столбцов, пропорциональны. В силу леммы 7 матрица C с точностью до скалярного кратного является инволюцией. То же верно для матрицы S , коллинеарной матрице V . Запишем эти три факта следующим образом.

Пусть L — строго нижняя треугольная тёплицева матрица. С учетом условия, что S имеет нулевую диагональ, положим

$$C = c_0 I_n + L + L^\top, \quad S = \gamma(L - L^\top), \\ C^2 = \lambda_1 I_n, \quad S^2 = \lambda_2 I_n.$$

Из этих соотношений выводим систему

$$(c_0 I_n + L + L^\top)^2 = \lambda_1 I_n, \\ (L - L^\top)^2 = \frac{\lambda_2}{\gamma^2} I_n,$$

которая после раскрытия скобок приобретает вид

$$c_0^2 I_n + 2c_0 L + 2c_0 L^\top + L^2 + (L^\top)^2 + LL^\top + L^\top L = \lambda_1 I_n, \\ L^2 + (L^\top)^2 - LL^\top - L^\top L = \frac{\lambda_2}{\gamma^2} I_n.$$

Суммируя эти уравнения, получаем

$$c_0^2 I_n + 2c_0 L + 2c_0 L^\top + 2L^2 + 2(L^\top)^2 = \left(\lambda_1 + \frac{\lambda_2}{\gamma^2}\right) I_n.$$

Поскольку L — строго нижняя треугольная тёплицева матрица, приходим к выводу, что

$$L(c_0 I_n + L) = 0.$$

Отсюда следует, что $c_0 = 0$, а тогда $L^2 = 0$. На основании леммы 10 заключаем, что $n = 2r$ — четное число, а матрицы C и S , как и матрицы Q и V , имеют вид

$$C = \nu_1 \begin{pmatrix} 0_{r,r} & I_r \\ I_r & 0_{r,r} \end{pmatrix}, \quad S = \nu_2 \begin{pmatrix} 0_{r,r} & -I_r \\ I_r & 0_{r,r} \end{pmatrix}.$$

Подобный случай был разобран при исследовании третьего класса АСТКТМ.

Теперь исследуем ситуацию, когда матрица $Z + Z^\top$ — косой циркулянт. Положим $Z = C + S$, где C — циркулянт, а S — косой циркулянт; при этом C имеет нулевую диагональ. Из условия, что $Z + Z^\top$ — косой циркулянт, получаем соотношение $C^\top = -C$. Тогда $Z + Z^\top = S + S^\top$ и $Q = \alpha_3 (S + S^\top)$.

Запишем условия инволютивности матрицы $Z = C + S$ и транспонированной матрицы $Z^\top$:

$$C^2 + CS + SC + S^2 = I_n, \\ C^2 - S^\top C - CS^\top + (S^\top)^2 = I_n.$$

Разность этих соотношений дает равенство

$$C(S + S^\top) - (S + S^\top)C^\top = (S^\top)^2 - S^2. \quad (20)$$

Согласно лемме 9 матрица $S + S^\top$ является скалярным кратным инволютивной матрицы. То же самое можно сказать о матрице Q . Положим $Q^2 = \lambda I_n$ и подставим это равенство в первое уравнение (17):

$$\lambda I_n - \frac{2\gamma}{\mu} Q + V^2 = \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n$$

или

$$V^2 = \frac{2\gamma}{\mu} Q + \left(\frac{\gamma^2}{4} - \frac{\gamma^2}{\mu^2} \right) I_n - \lambda I_n.$$

Квадрат тёплицевой кососимметричной матрицы является косым циркулянтном.

Если $\gamma \neq 0$, то V^2 — нескаларный косой циркулянт, и тогда в силу леммы 3 матрица V — косой циркулянт. Отсюда следует, что A и B — косые циркулянты, а сами матрицы T и H будут тёплицевым и ганкелевым косыми циркулянтами. Этот случай уже был разобран. Аналогично обосновывается случай $\gamma = 0$ и V — косой циркулянт.

Пусть теперь $\gamma = 0$. Неисследованной остается ситуация, когда V — скалярное кратное инволютивного циркулянта. Напомним, что V — кососимметричная матрица. Из соотношений

$$V = \beta_3 (Z - Z^\top) = \beta_3 (C + S - C^\top - S^\top) = \beta_3 (2C + S - S^\top)$$

следует, что $S^\top = S$, $V = 2\beta_3 C$. Подстановка равенства $S^\top = S$ в (20) дает условие

$$CS - SC^\top = 0.$$

Согласно лемме 8, C и S обладают тем свойством, что векторы, составленные из поддиагональных элементов их первых столбцов, пропорциональны. В силу леммы 9 матрица S с точностью до скалярного кратного является инволюцией. То же верно для матрицы C , коллинеарной матрице V . Запишем эти три факта следующим образом.

Пусть L — строго нижняя треугольная тёплицева матрица, тогда

$$C = L - L^\top, \quad S = \gamma(s_0 I_n + L + L^\top),$$

$$C^2 = \lambda_1 I_n, \quad S^2 = \lambda_2 I_n.$$

Из этих соотношений выводим систему

$$\begin{aligned} (L - L^\top)^2 &= \lambda_1 I_n, \\ (s_0 I_n + L + L^\top)^2 &= \frac{\lambda_2}{\gamma^2} I_n, \end{aligned}$$

которую после раскрытия скобок можно записать как

$$\begin{aligned} L^2 + (L^\top)^2 - LL^\top - L^\top L &= \lambda_1 I_n, \\ s_0^2 I_n + 2s_0 L + 2s_0 L^\top + L^2 + (L^\top)^2 + LL^\top + L^\top L &= \frac{\lambda_2}{\gamma^2} I_n. \end{aligned}$$

Суммируя эти уравнения, получаем

$$s_0^2 I_n + 2s_0 L + 2s_0 L^\top + 2L^2 + 2(L^\top)^2 = \left(\lambda_1 + \frac{\lambda_2}{\gamma^2} \right) I_n.$$

Поскольку L — строго нижняя треугольная тёплицева матрица, приходим к выводу, что

$$L(s_0 I_n + L) = 0.$$

Отсюда следует, что $s_0 = 0$, а тогда $L^2 = 0$.

Матрица S — косой циркулянт. По лемме 10 матрица S должна в то же время быть циркулянтом и, значит, скалярной матрицей. Отсюда следует, что скалярными будут матрицы Q и V и матрицы A и B , что противоречит условию рассматриваемого случая.

Теорема 2 доказана. Она дает полное решение системы (4) при условии $\mu \neq 0$.

Проведем небольшой анализ решений, учитывая свойство персимметричности тёплицевых матриц. Если пара (T, H) принадлежит классу 1, то

$$T = \alpha \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \widehat{L} & 0_{r,r} \end{pmatrix}, \quad H = \beta \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \\ 0_{r,r} & \widehat{L} \mathcal{P}_r \end{pmatrix},$$

$$TH = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \sigma \widehat{L} \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \end{pmatrix},$$

$$HT = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \widehat{L} \mathcal{P}_r \widehat{L} & 0_{r,r} \end{pmatrix} = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_{r,r} \\ \widehat{L} \widehat{L}^\top \mathcal{P}_r & 0_{r,r} \end{pmatrix}.$$

Для пары из класса 2 получаем

$$T = \alpha \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} & 0_r & 0_{r,r} \end{pmatrix}, \quad H = \beta \begin{pmatrix} \sigma \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ 0_{r,r} & 0_r & \widehat{L} \mathcal{P}_r \end{pmatrix},$$

$$TH = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \sigma \widehat{L} \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \end{pmatrix},$$

$$HT = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} \mathcal{P}_r \widehat{L} & 0_r & 0_{r,r} \end{pmatrix} = \alpha \beta \begin{pmatrix} 0_{r,r} & 0_r & 0_{r,r} \\ 0_r^\top & 0 & 0_r^\top \\ \widehat{L} \widehat{L}^\top \mathcal{P}_r & 0_r & 0_{r,r} \end{pmatrix}.$$

Таким образом, для каждого класса $TH = \sigma HT$.

Важно отметить, что множества пар матриц (T, H) , составляющих классы 4' и 3' теоремы 1, являются подмножествами класса 1 теоремы 2 и его $\mathcal{P}_n$ -преобразования. Данное наблюдение получается, если в качестве нижней треугольной матрицы $\widehat{L}$ взять единичную матрицу.

СПИСОК ЛИТЕРАТУРЫ

1. *Guterman A. E., Markova O. V., Mehrmann V.* Length realizability for pairs of quasi-commuting matrices // *Linear Algebra and Appl.* 2019. V. 568. P. 135–154.
2. *Kassel C.* Quantum Groups, Grad. Texts in Math. V. 155. Springer-Verlag. New York, 1995.
3. *Manin Yu. I.* Quantum Groups and Non-commutative Geometry. CRM. Montreal, 1988.
4. *Chriss N., Ginzburg V.* Representation Theory and Complex Geometry. Birkhauser. Boston. Basel. Berlin, 1997.
5. *Чугунов В. Н.* О некоторых множествах пар σ -коммутирующих ($\sigma \neq 0, \pm 1$) тёплицевой и ганкелевой матриц // Численные методы и вопросы организации вычислений. XXXII. Зап. научн. сем. ПОМИ. Т. 482. СПб. 2019. С. 288–294.
6. *Чугунов В. Н., Икрамов Х. Д.* Об одном частном решении задачи о σ -коммутировании ($\sigma \neq 0, \pm 1$) тёплицевой и ганкелевой матриц // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. N 11. С. 1817–1828.
7. *Гельфгат В. И.* Условия коммутирования тёплицевых матриц // Ж. вычисл. матем. и матем. физ. 1998. Т. 38. N 1. С. 11–14.
8. *Икрамов Х. Д., Чугунов В. Н.* Об описании пар антикоммутирующих тёплицевой и ганкелевой матриц // Зап. научн. сем. ПОМИ. Т. 463. СПб. 2017. С. 160–223.

ON NEW CLASSES OF SOLUTIONS TO THE PROBLEM OF σ -COMMUTATION ($\sigma \neq 0, \pm 1$) OF THE TOEPLITZ AND HANKEL MATRICES WITHIN THE FRAMEWORK OF A UNIFIED APPROACH

V. N. Chugunov^{a,*}, Kh. D. Ikramov^{b,**}

^a 119333 Moscow, Gubkina str., 8, ICM RAS, Russia

^b 119992 Moscow, Leninskie Gory, Moscow State University, CMC Faculty, Russia

*e-mail: chugunov.vadim@gmail.com

**e-mail: ikramov@cs.msu.su

Received April 3, 2024

Revised April 3, 2024

Accepted April 3, 2024

Abstract. In the previous work by the authors, a unified approach to the design of pairs of matrices (T, H) solving the problem of σ -commutation of Toeplitz and Hankel matrices was proposed. In this paper, this approach is applied to the derivation of new classes of solutions.

Keywords: Teplitz matrix, Hankel matrix, σ -commutation.

АППРОКСИМАЦИЯ ЗАДАЧ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ДЛЯ ПОЛУЛИНЕЙНЫХ ЭЛЛИПТИЧЕСКИХ УРАВНЕНИЙ КОНВЕКЦИИ-ДИФФУЗИИ С ГРАНИЧНЫМ НАБЛЮДЕНИЕМ КОНОРМАЛЬНОЙ ПРОИЗВОДНОЙ, С УПРАВЛЕНИЯМИ В КОЭФФИЦИЕНТАХ ОПЕРАТОРА КОНВЕКТИВНОГО ПЕРЕНОСА И НЕЛИНЕЙНОМ ЧЛЕНЕ УРАВНЕНИЯ

© 2024 г. Ф. В. Лубышев¹, М. Э. Файрузов^{1,*}

¹450076 Уфа, ул. Заки Валиди, 32, Уфимский университет науки и технологий, Россия

*e-mail: fairuzovme@mail.ru

Поступила в редакцию 22.05.2023 г.

Переработанный вариант 25.02.2024 г.

Принята к публикации 05.04.2024 г.

В работе изучаются разностные аппроксимации задачи оптимального управления с граничным наблюдением конормальной производной состояния, описываемой задачей Дирихле для полулинейных эллиптических уравнений с управлениями в коэффициентах оператора конвективного переноса и нелинейном члене уравнения. Рассматриваются вопросы корректности постановки задачи оптимального управления. Строятся разностные аппроксимации задачи оптимального управления. Изучаются вопросы сходимости аппроксимаций по функционалу и управлению. Проводится регуляризация аппроксимаций. Библ. 23.

Ключевые слова: задача оптимального управления, полулинейные эллиптические уравнения, краевая задача Дирихле, разностные аппроксимации, сходимость аппроксимаций, регуляризация аппроксимаций.

DOI: 10.31857/S0044466924070055, EDN: xisfoz

1. ВВЕДЕНИЕ

Характер конкретных постановок задач оптимального управления для распределенных систем существенно зависит от того, куда входят управления (в свободные члены уравнений состояния или в коэффициенты уравнений), линейными или нелинейными дифференциальными уравнениями математической физики (УМФ) описываются состояния систем управления (см. [1]–[11]), а также зависит от того, какой гладкостью обладают входные данные систем управления. В настоящее время наиболее полно исследован случай таких систем управления, когда управления входят в линейные уравнения состояния и линейные предельные условия, а также когда входные данные и функции состояния систем являются достаточно гладкими. Характерной особенностью задач оптимального управления нелинейного типа является то, что отображение из множества допустимых управлений в пространство состояний является нелинейным. Нелинейные задачи оптимального управления для УМФ относятся к наиболее сложному классу задач оптимизации. Проблема численного решения задач оптимального управления приводит к необходимости их аппроксимации задачами более простой природы — “конечномерными задачами оптимизации” (см. [1]). Обзор работ, посвященных основам общей теории и методам устойчивости и аппроксимации задач оптимального управления, а также результатам в этой области, представлен, например, в [1]–[11]. Одним из наиболее удобных, универсальных и широко распространенных методов конечномерных аппроксимаций задач оптимального управления является метод сеток (см. [12]–[17]). Центральными в проблеме аппроксимации являются вопросы “конструирования” аппроксимаций, сходимости аппроксимаций по состоянию, функционалу, управлению, регуляризации аппроксимаций (см., например, [1]–[11]). Для систем управления, описываемых УМФ, построения и исследования аппроксимаций проводились в основном для

линейных задач оптимального управления, причем с достаточно гладкими коэффициентами УМФ и состояниями.

При исследовании многих процессов в движущихся средах в качестве основных можно выделить диффузионный перенос той или иной субстанции и перенос, обусловленный движением среды, т.е. конвективный перенос (см. [12]). Задачи конвекции-диффузии являются типичными для математических моделей механики жидкости и газа. Так, распределение тепла, примесей может происходить не только за счет диффузии, но и быть обусловленным движением среды. Принципиальные особенности физико-химических процессов в механике жидкости и газа могут быть порождены именно учетом движения сред под действием тех или иных сил. Конвективный-диффузионный процесс может играть определяющую роль при моделировании самых разнообразных процессов. В частности, важное значение приобретают экологические проблемы, связанные с описанием процессов распределения примесей в атмосфере и водоемах, с моделированием загрязнения грунтовых вод. В газо- и гидродинамике в качестве базовых моделей многих процессов выступают краевые задачи как для стационарных, так и нестационарных уравнений конвекции-диффузии — эллиптические или параболические уравнения второго порядка с младшими членами. В настоящее время в теории численных методов решения задач УМФ и задач оптимального управления наиболее глубокие результаты получены при рассмотрении процессов с самосопряженными операторами. Это относится как к методам, базирующимся на конечно-разностных аппроксимациях состояния, так и к методам на основе конечно-элементных аппроксимаций.

Данная работа посвящена исследованию численных методов решения наиболее важных для теории и практики задач оптимального управления для эллиптических уравнений второго порядка с несамосопряженными операторами — задач конвекции-диффузии. Процесс управления описывается полулинейным уравнением конвекции-диффузии с управлениями в коэффициентах оператора конвективного переноса и нелинейном члене уравнения. На управления накладываются локальные ограничения. Наличие несамосопряженного оператора вызывает определенные трудности при построении и изучении аппроксимаций дифференциальных уравнений, в частности при доказательстве корректности разностных аппроксимаций, и при исследовании связи между исходной задачей оптимального управления и аппроксимирующей сеточной задачей. Исследованы вопросы сходимости аппроксимаций: установлены оценки точности аппроксимаций по состоянию и функционалу и сходимость аппроксимаций по управлению. На основе метода А. Н. Тихонова и полученных результатов проведена регуляризация предложенных аппроксимаций, позволяющая строить минимизирующие последовательности для целевых функционалов нелинейных задач оптимального управления, сильно сходящиеся в пространствах управлений к множествам точек минимумов функционалов исходных постановок.

2. ПОСТАНОВКА ЗАДАЧИ

Пусть $\Omega = \{x = (x_1, x_2) \in \mathbb{R}^2 : 0 < x_\alpha < l_\alpha, \alpha = 1, 2\}$ — прямоугольник с границей Γ . Пусть управляемый процесс описывается в Ω краевой задачей Дирихле для полулинейного уравнения эллиптического типа с переменными коэффициентами:

$$L(g)u = - \sum_{\alpha=1}^2 \frac{\partial}{\partial x_\alpha} \left(k_\alpha(x) \frac{\partial u}{\partial x_\alpha} \right) + \sum_{\alpha=1}^2 b_\alpha(x) \frac{\partial u}{\partial x_\alpha} + d(x)q(u) = f(x), \quad x \in \Omega, \quad (1)$$

$$u(x) = 0, \quad x \in \Gamma. \quad (2)$$

Рассмотрим задачу минимизации функционала

$$J(g) = \int_{\Gamma_*} \left| \frac{\partial u(x, g)}{\partial N} - \psi(x) \right|^2 d\Gamma_*, \quad (3)$$

на решениях $u(x) = u(x; g)$ краевой задачи (1)–(2) при следующих ограничениях на управления $g(x) = (g_1(x), g_2(x), g_3(x)) = (d(x), b_1(x), b_2(x))$:

$$g \in U = \prod_{k=1}^3 U_k \subset B = W_\infty^1(\Omega) \times W_\infty^1(\Omega) \times W_\infty^1(\Omega) = (W_\infty^1(\Omega))^3, \quad (4)$$

где $\Gamma_* \subseteq \Gamma$, $\text{mes} \Gamma_* > 0$, $\psi(x) \in W_2^1(\Gamma_*)$ — заданная функция, а $\xi_\alpha, \bar{\xi}_\alpha, \hat{R}_\beta^{(\alpha)}$, $\alpha = 1, 2, 3$, $\beta = 1, 2$ — заданные числа, п.в. — почти всюду. В данном случае мы имеем, так называемую, задачу оптимального управления с граничным наблюдением конормальной производной состояния. Функционал цели (3) в данной экстремальной задаче характеризует меру отклонения в L_2 — норме конормальной производной состояния на границе Γ области Ω от заданной функции $\psi(x)$. В дальнейшем, для определенности будем считать, что

$$\Gamma_* = \Gamma_{+1} = \{x_1 = l_1, 0 < x_2 < l_2\}.$$

Предполагается выполнение следующих условий: $k_\alpha(x)$, $\alpha = 1, 2$, $q(u)$, $f(x)$ — заданные функции, причем

$$k_\alpha(x) \in C^1(\bar{\Omega}), \quad f(x) \in C(\bar{\Omega}), \quad k_\alpha(x) \geq \nu > 0, \quad \alpha = 1, 2, \quad x \in \Omega, \\ \text{и } \frac{\partial^2 k_\alpha}{\partial x_\alpha \partial x_\beta}, \quad \frac{\partial f}{\partial x_\beta}, \quad \alpha, \beta = 1, 2 \text{ — ограниченные на } \Omega \text{ кусочно-гладкие функции;} \quad (5)$$

функция $q(s)$ определена на $\mathbb{R}$ со значениями в $\mathbb{R}$ и удовлетворяет условиям

$$q(0) = 0, \quad 0 < q_0 \leq [q(s_1) - q(s_2)] / (s_1 - s_2) \leq L_q < \infty, \quad \forall s_1, s_2 \in \mathbb{R}, s_1 \neq s_2; \quad (6)$$

$$\xi_1 > 0, \quad -m \leq \xi_2 \leq \bar{\xi}_2 \leq m, \quad -p \leq \xi_3 \leq \bar{\xi}_3 \leq p, \quad m, p = \text{const} > 0, \\ \delta_* = \max_{\substack{\varepsilon_1, \varepsilon_2 > 0 \\ \varepsilon_1 + \varepsilon_2 \leq \nu}} \left\{ \frac{\nu - (\varepsilon_1 + \varepsilon_2)}{C_\Omega^2} + \xi_1 q_0 - \frac{m^2}{4\varepsilon_1} - \frac{p^2}{4\varepsilon_2} \right\} > 0, \quad C_\Omega^2 = \left(\frac{8}{l_1^2} + \frac{8}{l_2^2} \right)^{-1}. \quad (7)$$

Используя метод монотонных операторов [18, с. 78], [19, с. 202] можно показать, что существует единственное обобщенное решение задачи (1)–(2) в классе $\dot{W}_2^1(\Omega)$. Более того, это решение $u(x) \in W_2^2(\Omega) \cap \dot{W}_2^1(\Omega) = W_{2,0}^2(\Omega)$, причем

$$\sup_{g \in U} \|u(x, g)\|_{W_2^2(\Omega)} \leq C_1 \|f(x)\|_{L_2(\Omega)}. \quad (8)$$

Будем считать, что задача (1)–(2) принадлежит также пространству $W_2^3(\Omega)$.

Замечание 2.1. В краевой задаче для состояния (1)–(2) выделим случай: $q(u) = u$, важный для приложений [12], [21]. Тогда источниковый член уравнения (1) будет содержать слагаемое $d(x)u(x)$ пропорциональное функции состояния $u = u(x)$. В этом случае для существования решения $u(x)$ задачи (1)–(2) из пространства $W_2^3(\Omega)$ достаточно (см. [21, с. 31]), чтобы выполнялись, сформулированные выше условия на исходные данные и управления.

Замечание 2.2. Рассмотрим условие (7): $\delta_* > 0$. Оно выполняется, если, например, постоянная C_Ω “достаточно мала” или если положительное число $\xi_1 \cdot q_0$ “достаточно большое”. Выражения “достаточно малое” и “достаточно большое” уточняются условием $\delta_* > 0$, где δ_* определено в (7).

Здесь

$$C_\Omega^2 = \left(\frac{8}{l_1^2} + \frac{8}{l_2^2} \right)^{-1}$$

есть константа из неравенства Пуанкаре–Фридрихса

$$\int_\Omega u^2 d\Omega \leq C_\Omega^2 \int_\Omega \sum_{\alpha=1}^2 \left(\frac{\partial u}{\partial x_\alpha} \right)^2 d\Omega,$$

с константой C_Ω , зависящей лишь от этой области Ω .

По поводу условия (7) смотри также монографию [23, с. 94].

Условия (6) описывают сильную монотонность и липшиц-непрерывность нелинейной функции $q(u)$.

Напомним следующее

Определение 2.1 (см. [1, с. 49]). Последовательность $\{u_n\} \subset B$ сходится к множеству U слабо в B , где B — банахово пространство, если $\{u_n\}$ имеет хотя бы одну слабо сходящуюся подпоследовательность, причем все точки v , являющиеся слабым пределом какой-либо подпоследовательности $\{u_n\}$ принадлежат U .

Пусть: $H = W_2^1(\Omega) \times W_2^1(\Omega) \times W_2^1(\Omega)$, $J_* = \inf\{J(g) : g \in U\}$, $U_* = \{g_* \in U : J(g_*) = J_*\}$.

Корректность постановки задачи (1)–(4) устанавливает

Теорема 2.1. Пусть выполнены условия, принятые при постановке задачи (1)–(4). Тогда

$$J_* = \inf\{J(g) : g \in U\} > -\infty,$$

$$U_* = \{g_* \in U : J(g_*) = J_*\} \neq \emptyset,$$

где U_* — слабо компактно в H и любая минимизирующая последовательность $\{g_n\} \subset U$ для функционала $J(g)$ слабо в H сходится к множеству U_* .

Доказательство. Используя оценку (8), можно показать, что функционал $J(g)$ слабо непрерывен на U в слабой топологии пространства H . Кроме того, множество допустимых управлений U является выпуклым ограниченным и замкнутым множеством в рефлексивном пространстве H . Тогда справедливость теоремы 2.1 следует из [1, с. 49, гл. 1, п. 3, определение 6, теорема 2, с. 51, теорема 4. Теорема доказана.

3. РАЗНОСТНАЯ АППРОКСИМАЦИЯ ЗАДАЧ УПРАВЛЕНИЯ. АПРИОРНАЯ ОЦЕНКА ФУНКЦИИ СОСТОЯНИЯ СЕТОЧНОЙ ЗАДАЧИ В НОРМЕ СЕТОЧНОГО ПРОСТРАНСТВА $W_{2,0}^2(\bar{\omega})$

Для аппроксимации задачи (3), (4), (1), (2) введем в $\bar{\Omega}$ сетки [13]: $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$, $\omega = \omega_1 \times \omega_2$, $\gamma = \bar{\omega} \setminus \omega$, $\omega^{(\pm 1)} = \omega_1^\pm \times \omega_2$, $\omega^{(\pm 2)} = \omega_1 \times \omega_2^\pm$. Пусть V — множество сеточных функций, заданных на сетке $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$, а $\overset{\circ}{V}$ — его подмножество, состоящее из сеточных функций, обращающихся в нуль на $\gamma = \bar{\omega} \setminus \omega$. Для сеточных функций из множества $\overset{\circ}{V}$ введем следующие нормы:

$$\|y\|_{W_2^1(\bar{\omega})}^2 = |y|_{W_2^1(\bar{\omega})}^2 + \|y\|_{L_2(\omega)}^2 = |y|_{W_2^1(\bar{\omega})}^2 + \sum_{x \in \omega} h_1 h_2 y^2,$$

где

$$|y|_{W_2^1(\bar{\omega})}^2 = \|y\|_{W_2^1(\bar{\omega})}^2 = \sum_{\alpha=1}^2 \sum_{x \in \omega^{(+\alpha)}} h_1 h_2 y_{\bar{x}_\alpha}^2 = \|\nabla y\|_{L_2(\omega)}^2,$$

$\omega^{(+1)} = \omega_1^+ \times \omega_2$, $\omega^{(+2)} = \omega_1 \times \omega_2^+$, которая превращает множество $\overset{\circ}{V}$ в нормированное пространство, обозначаемое через $\overset{\circ}{W}_{2,0}^1(\bar{\omega})$, и

$$\|y\|_{W_2^2(\bar{\omega})}^2 = |y|_{W_2^2(\bar{\omega})}^2 + \|y\|_{W_2^1(\bar{\omega})}^2,$$

где

$$|y|_{W_2^2(\bar{\omega})}^2 \equiv \|y\|_{W_{2,0}^2(\bar{\omega})}^2 = \sum_{\alpha=1}^2 \sum_{x \in \omega} h_1 h_2 y_{\bar{x}_\alpha x_\alpha}^2 + 2 \sum_{\omega_1^- \times \omega_2^+} h_1 h_2 y_{\bar{x}_1 x_2}^2,$$

превращающее его в нормированное пространство, обозначаемое через $W_{2,0}^2(\bar{\omega})$.

Здесь, как обычно, через $L_2(\omega)$ обозначено пространство сеточных функций, заданных на сетке ω со скалярным произведением и нормой

$$(y, v)_\omega = (y, v)_{L_2(\omega)} = \sum_{x \in \omega} h_1 h_2 y(x) v(x), \quad \|y\|_\omega^2 = (y, y)_\omega = (y, y)_{L_2(\omega)}.$$

Далее, так как для любой сеточной функции $y(x) \in \overset{\circ}{V}$ справедливо неравенство (разностный аналог неравенства Фридрихса)

$$\|y\|_{L_2(\omega)}^2 \leq C_\Omega^2 \|y\|_{\overset{\circ}{W}_{\frac{1}{2}}(\bar{\omega})}^2, \quad C_\Omega^2 = \left(\frac{8}{l_1^2} + \frac{8}{l_2^2} \right)^{-1}, \quad (9)$$

то как следствие из этого неравенства, получим, что в пространстве $\overset{\circ}{V}$ полунорма $|y|_{W_{\frac{1}{2}}(\bar{\omega})} \equiv \|y\|_{\overset{\circ}{W}_{\frac{1}{2}}(\bar{\omega})}$ определяет эквивалентную $\|\cdot\|_{W_{\frac{1}{2}}(\bar{\omega})}$ норму. Действительно,

$$(1 + C_\Omega^2)^{-1} \|y\|_{W_{\frac{1}{2}}(\bar{\omega})}^2 \leq \|y\|_{\overset{\circ}{W}_{\frac{1}{2}}(\bar{\omega})}^2 \leq \|y\|_{W_{\frac{1}{2}}(\bar{\omega})}^2. \quad (10)$$

Далее, в пространстве $\overset{\circ}{V}$ сеточных функций, обращающихся в нуль на границе γ , полунорма $|y|_{W_{\frac{1}{2}}(\bar{\omega})} \equiv \|y\|_{W_{\frac{1}{2},0}(\bar{\omega})}$ эквивалентна норме $\|y\|_{W_{\frac{1}{2}}(\bar{\omega})}$. Действительно, имеет место оценка

$$\|y\|_{\overset{\circ}{W}_{\frac{1}{2}}(\bar{\omega})}^2 = |y|_{W_{\frac{1}{2}}(\bar{\omega})}^2 \leq \frac{l_1^2 + l_2^2}{8} \|y_{\bar{x}_1 \bar{x}_2}\|_{L_2(\omega_1^+ \times \omega_2^+)}^2. \quad (11)$$

Применяя неравенство (9) к этой оценки, получим

$$C_\Omega^{-2} \|y\|_{L_2(\omega)}^2 \leq \frac{l_1^2 + l_2^2}{8} \|y_{\bar{x}_1 \bar{x}_2}\|_{L_2(\omega_1^+ \times \omega_2^+)}^2, \quad (12)$$

откуда вытекает оценка

$$\frac{8}{16 + (1 + C_\Omega^2)(l_1^2 + l_2^2)} \|y\|_{W_{\frac{1}{2}}(\bar{\omega})}^2 \leq \|y\|_{W_{\frac{1}{2},0}(\bar{\omega})}^2 \leq \|y\|_{W_{\frac{1}{2}}(\bar{\omega})}^2. \quad (13)$$

Лемма 3.1. Для любой сеточной функции $y(x)$, заданной на равномерной сетке $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$ и обращающейся в нуль на границе γ , справедливо неравенство

$$\|y\|_{L_2(\omega)} \leq M \|y\|_{W_{\frac{1}{2},0}(\bar{\omega})}, \quad M = \frac{\sqrt{l_1 l_2}}{32} \left[\pi(l_1^2 + l_2^2) + \frac{2l_1^3 l_2^3}{(l_1^2 + l_2^2)^2} \right]. \quad (14)$$

Доказательство. Для функций $y(x)$, заданных на равномерной сетке $\bar{\omega}$ и обращающихся в нуль на γ , справедлива оценка

$$\|y\|_{C(\bar{\omega})} \leq M_0 \|y\|_{W_{\frac{1}{2},0}(\bar{\omega})}, \quad M_0 = \frac{1}{32} \left[\pi(l_1^2 + l_2^2) + \frac{2l_1^3 l_2^3}{(l_1^2 + l_2^2)^2} \right]. \quad (15)$$

Следовательно,

$$\|y\|_{L_2(\omega)}^2 = \sum_{\omega} h_1 h_2 y^2(x) \leq l_1 l_2 M_0^2 \|y\|_{W_{\frac{1}{2},0}(\bar{\omega})}^2, \quad (16)$$

откуда получаем оценку (14).

Лемма 3.2. Для любой сеточной функции $v(x)$, заданной на равномерной сетке $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$, справедливо неравенство

$$\sum_{\bar{\omega}_2} \bar{h}_2 v_{\bar{x}_1}^2(l_1, x_2) \leq \varepsilon \sum_{\bar{\omega}_1 \times \bar{\omega}_2} v_{\bar{x}_1 x_1}(x_1, x_2) h_1 \bar{h}_2 + M \left(\frac{1}{l_1} + \frac{1}{\varepsilon^3} \right) \|v\|_{L_2(\bar{\omega})}^2, \quad (17)$$

где $\varepsilon > 0$ — произвольное число, а M — некоторая положительная постоянная, не зависящая ни от сетки $\bar{\omega}$, ни от функции $v(x)$.

Доказательство. При произвольном фиксированном x_2 для функции $v(x) = v(x_1, x_2)$ справедлива оценка (см. [13], с.294, теорема 3)

$$v_{\bar{x}_1}^2(l_1, x_2) \leq \varepsilon \sum_{\bar{\omega}_1} v_{\bar{x}_1 x_1}^2(x_1, x_2) h_1 + M \left(\frac{1}{l_1} + \frac{1}{\varepsilon^3} \right) \sum_{\bar{\omega}_1} v^2(x_1, x_2) \bar{h}_1.$$

Умножая это неравенство на h_2 и суммируя полученное соотношение по $\bar{\omega}_2$, получаем оценку (17).

Лемма 3.3. Для любой сеточной функции $y(x)$, заданной на равномерной сетке $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$ и обращающейся в нуль на границе γ , справедливо неравенство

$$\sum_{\bar{\omega}_2} \hbar_2 y_{\bar{x}_1}^2(l_1, x_2) \leq C \|y\|_{W_{2,0}^2(\bar{\omega})}^2. \quad (18)$$

Доказательство следует из оценок (14) и (17).

Перейдем теперь к аппроксимации задачи оптимального управления.

Задаче оптимального управления (3), (4), (1)–(2) (с входными данными $k_\alpha(x)$, $\alpha = 1, 2$, $q(u)$, $f(x)$) поставим в соответствие следующие разностные аппроксимации: минимизировать сеточный функционал

$$J_h(\Phi_h) = \sum_{x_2 \in \bar{\omega}_2} \hbar_2 |a_1(l_1, x_2) y_{\bar{x}_1}(l_1, x_2) - \psi^{h_2}(x_2)|^2, \quad (19)$$

при условиях, что сеточная функция $y(x) = y(x, \Phi_h)$, $x \in \bar{\omega}_h$, является решением следующей сеточной задачи Дирихле:

$$\begin{aligned} \Lambda(\Phi_h)y(x) &= - \sum_{\alpha=1}^2 (a_\alpha(x) y_{\bar{x}_\alpha})_{x_\alpha} + \sum_{\alpha=1}^2 \Phi_{\alpha+1,h}(x) y_{y_\alpha}^\circ(x) + \Phi_{1h}(x) q(y) = \\ &= S^{x_1} S^{x_2} f(x) = f^h(x), \quad x \in \omega, \end{aligned} \quad (20)$$

$$y(x) = 0, \quad x \in \gamma_h, \quad (21)$$

а сеточные управления Φ_h таковы, что

$$\Phi_h(x) = (\Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in \prod_{\alpha=1}^3 U_{\alpha h} = U_h \subset B_h = (W_\infty^1(\omega))^3, \quad (22)$$

$$\begin{aligned} U_{\alpha h} &= \{ \Phi_{\alpha h}(x) \in W_\infty^1(\omega) : \xi_\alpha \leq \Phi_{\alpha h}(x) \leq \bar{\xi}_\alpha, x \in \omega, |\Phi_{\alpha h x_1}(x)| \leq R_1^{(\alpha)}, \\ &x \in (\omega_1 \setminus \{x_1 - l_1 - h_1\}) \times \omega_2, |\Phi_{\alpha h x_2}(x)| \leq R_2^{(\alpha)}, x \in \omega_1 \times (\omega_2 \setminus \{x_2 = l_2 - h_2\}) \}, \end{aligned} \quad (23)$$

$\alpha = 1, 2, 3$. Здесь $\psi^{h_2}(x_2) = S^{x_2} \psi$ — сеточная аппроксимация функции $\psi(x)$ с помощью операции усреднения по Стеклову (см. [17, с. 56], [12, с. 158]):

$$a_\alpha(x) = k_\alpha^{(-0.5_\alpha)}(x), \quad x \in \omega, \quad \alpha = 1, 2.$$

Разностную схему (20), (21) в дальнейшем будем рассматривать как операторное уравнение

$$\bar{A}_h y = \bar{F}_h, \quad (24)$$

с оператором $\bar{A}_h$, действующим из $W_{2,0}^2(\bar{\omega})$ в $L_2(\omega)$ и определяемым соотношением

$$\bar{A}_h y(x) = \Lambda y(x), \quad \bar{F}_h = f^h(x), \quad x \in \omega. \quad (25)$$

Лемма 3.4. Для оператора $\bar{A}_h : W_{2,0}^2(\bar{\omega}) \rightarrow L_2(\omega)$ справедливы следующие оценки:

$$(\bar{A}_h v_1 - \bar{A}_h v_2, v_1 - v_2)_{L_2(\omega)} \geq \delta_* \|v_1 - v_2\|_{W_{2,0}^1(\bar{\omega})}^2, \quad (26)$$

$$C_0 \|v_1 - v_2\|_{W_{2,0}^2(\bar{\omega})} - M_1 \|v_1 - v_2\|_{W_{2,0}^1(\bar{\omega})} \leq \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\bar{\omega})} \leq M_2 \|v_1 - v_2\|_{W_{2,0}^2(\bar{\omega})}, \quad (27)$$

где

$$C_0 = \frac{\nu^2}{2\bar{\nu}}, \quad M_1 = C_\Omega L_q \xi_2^* + 2^{1/2}(R + \theta), \quad \bar{\nu} = \max_\alpha \{ \|k_\alpha(x)\|_{L_2(\Omega)} \},$$

$$M_2 = 2^{1/2}\bar{v} + \frac{M_1 l_1 l_2}{16C_\Omega}, \quad R = \max_\alpha \left\{ \left\| \frac{\partial k_\alpha(x)}{\partial x_\alpha} \right\|_{L_\infty(\Omega)} \right\}, \quad \theta = \max\{m, p\}, \quad \xi_2^* = \max\{|\xi_2|, |\bar{\xi}_2|\},$$

константы C_Ω , L_q , v , m , p , ξ_2 , $\bar{\xi}_2$, δ_* определены в условиях (4)–(7).

Доказательство. Ограничимся доказательством оценки (27). Используя формулы разностного дифференцирования произведения [13], и обозначая $v(x) = v_1(x) - v_2(x)$, из (20), (21) получим

$$\begin{aligned} \bar{A}_h v_1 - \bar{A}_h v_2 &= - \sum_{\alpha=1}^2 a_\alpha(x) v_{\bar{x}_\alpha x_\alpha} - \sum_{\alpha=1}^2 (a_\alpha(x))_{x_\alpha} \cdot v_{x_\alpha} + \\ &+ \sum_{\alpha=1}^2 \Phi_{\alpha+1,h}(x) v_{x_\alpha}^\circ + \Phi_{1h}(x) [q(v_1) - q(v_2)], \quad x \in \omega. \end{aligned} \quad (28)$$

Откуда найдем

$$\sum_{\alpha=1}^2 a_\alpha(x) v_{\bar{x}_\alpha x_\alpha} = D(x, v_1, v_2), \quad x \in \omega, \quad (29)$$

$$\begin{aligned} D(x, v_1, v_2) &= \Phi_{1h}(x) [q(v_1) - q(v_2)] - [\bar{A}_h v_1 - \bar{A}_h v_2] - \\ &- \sum_{\alpha=1}^2 a_{\alpha x_\alpha} v_{x_\alpha} + \sum_{\alpha=1}^2 \Phi_{\alpha+1,h}(x) v_{x_\alpha}^\circ, \quad x \in \omega. \end{aligned} \quad (30)$$

Умножим обе части (29) на сеточную функцию $\frac{v_{\bar{x}_1 x_1}}{a_2(x)}$ и приведем полученное соотношение к виду

$$\frac{1}{a_2(x)} \left\{ a_1(x) v_{\bar{x}_1 x_1}^2 + a_2(x) \left(\frac{v_{\bar{x}_1 x_1} + v_{\bar{x}_2 x_2}}{2} \right)^2 \right\} = \frac{v_{\bar{x}_1 x_1}}{a_2(x)} D(x, v_1, v_2) + I(v),$$

где

$$I(v) = \left(\frac{v_{\bar{x}_1 x_1} + v_{\bar{x}_2 x_2}}{2} \right)^2 - v_{\bar{x}_1 x_1} \cdot v_{\bar{x}_2 x_2}. \quad (31)$$

Из (31) установим оценку

$$\frac{v}{\bar{v}} \leq \frac{1}{v} |D| \cdot |v_{\bar{x}_1 x_1}| + I(v) \leq \frac{1}{v} \left[\frac{1}{4\epsilon} D^2(x, v_1, v_2) + \epsilon |v_{\bar{x}_1 x_1}|^2 \right] + I(v), \quad \epsilon > 0. \quad (32)$$

Полагая $\epsilon = \frac{v^2}{2\bar{v}}$ в неравенстве (32), получим оценку

$$\frac{v}{\bar{v}} v_{\bar{x}_1 x_1}^2 \leq \frac{\bar{v}}{\sqrt{3}} D^2(x, v_1, v_2) + I(v). \quad (33)$$

Умножим (33) на $h_1 h_2$ и просуммируем по сетке ω . В результате получим оценку

$$\|v_{\bar{x}_1 x_1}\|_{L_2(\omega)}^2 \leq \frac{\bar{v}^2}{v^4} \|D(x, v_1, v_2)\|_{L_2(\omega)}^2 + \frac{2\bar{v}}{v} \sum_{\omega} I(v) h_1 h_2. \quad (34)$$

Покажем, что

$$\sum_{\omega} I(v) h_1 h_2 \leq 0 \quad \forall v \in \overset{\circ}{V}. \quad (35)$$

Действительно, имеем

$$\sum_{\omega} I(v) h_1 h_2 \leq \frac{1}{2} \sum_{\omega} (v_{\bar{x}_1 x_2}^2 + v_{x_1 \bar{x}_2}^2) h_1 h_2 - \sum_{\omega} v_{\bar{x}_1 x_1} \cdot v_{\bar{x}_2 x_2} h_1 h_2. \quad (36)$$

Далее

$$\sum_{\omega} (v_{\bar{x}_1 x_2}^2 + v_{x_1 \bar{x}_2}^2) h_1 h_2 = \sum_{x_1 \in \omega_1} \sum_{x_2=2h_2}^{N_2 h_2} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2 + \sum_{x_1=2h_1}^{N_1 h_1} \sum_{x_2 \in \omega_2} v_{x_1 \bar{x}_2}^2 \leq 2 \sum_{x \in \omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2. \quad (37)$$

Используя формулу суммирования по частям (дважды) с учетом соотношений

$$v_{\bar{x}_2 x_2} \Big|_{\substack{x_1=0 \\ x_1=l_1}} = 0, \quad v_{\bar{x}_1 x_1} \Big|_{\substack{x_2=0 \\ x_2=l_2}} = 0,$$

получим

$$\begin{aligned} \sum_{\omega} v_{\bar{x}_1 x_1} \cdot v_{\bar{x}_2 x_2} h_1 h_2 &= - \sum_{\omega_2} h_2 \sum_{\omega_1^+} h_1 v_{\bar{x}_1} \cdot (v_{\bar{x}_2 x_2})_{\bar{x}_1} = \\ &= \sum_{\omega_1^+} h_1 \sum_{\omega_2} h_2 v_{\bar{x}_1 \bar{x}_2} \cdot (v_{\bar{x}_1})_{\bar{x}_2} = \sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2. \end{aligned} \quad (38)$$

Теперь с учетом (37), (38) из (36) получим желаемую оценку (35).

Вернемся теперь к неравенству (34). С учетом (35) из неравенства (34) получим

$$\|v_{\bar{x}_1 x_1}\|_{L_2(\omega)}^2 \leq \frac{\bar{v}^2}{v^4} \|D(x, v_1, v_2)\|_{L_2(\omega)}^2. \quad (39)$$

Аналогично устанавливается оценка

$$\|v_{\bar{x}_2 x_2}\|_{L_2(\omega)}^2 \leq \frac{\bar{v}^2}{v^4} \|D(x, v_1, v_2)\|_{L_2(\omega)}^2. \quad (40)$$

Следовательно,

$$\sum_{\alpha=1}^2 \|v_{\bar{x}_\alpha x_\alpha}\|_{L_2(\omega)}^2 \leq \frac{2\bar{v}^2}{v^4} \|D(x, v_1, v_2)\|_{L_2(\omega)}^2. \quad (41)$$

Далее из (38) имеем

$$\sum_{x \in \omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2 \leq \frac{1}{2} \sum_{\omega} h_1 h_2 (v_{\bar{x}_1 x_2}^2 + v_{x_1 \bar{x}_2}^2). \quad (42)$$

Принимая во внимание оценку (42) и определение сеточной нормы $\|v\|_{W_{2,0}^2(\bar{\omega})}$, получим неравенство

$$\|v\|_{W_{2,0}^2(\bar{\omega})}^2 = \sum_{\omega} (v_{\bar{x}_1 x_1}^2 + v_{\bar{x}_2 x_2}^2) h_1 h_2 + 2 \sum_{\omega_1^+ \times \omega_2^+} h_1 h_2 v_{\bar{x}_1 \bar{x}_2}^2 \leq 2 \sum_{\omega} h_1 h_2 (v_{\bar{x}_1 x_1}^2 + v_{\bar{x}_2 x_2}^2). \quad (43)$$

Теперь из неравенств (43), (41) получаем оценку

$$\|v\|_{W_{2,0}^2(\bar{\omega})} \leq \frac{2\bar{v}}{v^2} \|D(x, v_1, v_2)\|_{L_2(\omega)}. \quad (44)$$

Перейдем теперь к оценке правой части неравенств (44). Имеем

$$\begin{aligned} \|D(x, v_1, v_2)\|_{L_2(\omega)} &\leq \|\Phi_{1h}(x)[q(v_1) - q(v_2)]\|_{L_2(\omega)} + \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)} + \\ &+ \sum_{\alpha=1}^2 \left[\|a_{\alpha x_\alpha}\|_{L_2(\omega)} + \|\Phi_{\alpha+1,h}(x) v_{\bar{x}_\alpha}^\circ\|_{L_2(\omega)} \right]. \end{aligned} \quad (45)$$

Принимая во внимание очевидные оценки

$$|a_{\alpha x_\alpha}| < R, \quad R = \max_{\alpha} \left\| \frac{\partial k_{\alpha}(x)}{\partial x_{\alpha}} \right\|_{L_{\infty}(\Omega)}, \quad \alpha = 1, 2; \quad \|v_{\bar{x}_\alpha}^\circ\|_{L_2(\omega)} \leq \|v_{\bar{x}_\alpha}\|_{L_2(\omega^{+\alpha})}, \quad \alpha = 1, 2,$$

неравенство (9) и ограничения (4), (5), (7), получим из (45)

$$\|D(x, v_1, v_2)\|_{L_2(\omega)} \leq [C_{\Omega} L_q \xi_2^* + 2^{1/2}(R + \theta)] \|v\|_{W_{\frac{1}{2}}^1(\bar{\omega})} + \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)}, \quad (46)$$

где

$$\theta = \max\{m, p\}, \quad \xi_2^* = \max\{|\xi_2|, |\bar{\xi}_2|\}.$$

Из неравенств (44), (46) получаем

$$\frac{v^2}{2\bar{v}} \|v\|_{W_{2,0}^2(\omega)} - [C_\Omega L_q \xi_2^* + 2^{1/2}(R + \theta)] \|v\|_{\dot{W}_{\frac{1}{2}}^1(\bar{\omega})} \leq \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)}, \quad (47)$$

откуда следует левая часть неравенства (27). Установим теперь правую часть неравенства (27). Из (28) нетрудно получить оценку

$$\begin{aligned} \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)} &\leq \sum_{\alpha=1}^2 \left(\|a_\alpha(x) v_{\bar{x}_\alpha x_\alpha}\|_{L_2(\omega)} + \|a_{\alpha x_\alpha}(x) v_{x_\alpha}\|_{L_2(\omega)} + \|\Phi_{\alpha+2,h}(x) v_{x_\alpha}^\circ\|_{L_2(\omega)} \right) + \\ &+ \|\Phi_{1h}(x)[q(v_1) - q(v_2)]\|_{L_2(\omega)} \leq 2^{1/2}\bar{v} \|v\|_{W_{2,0}^2(\omega)} + [C_\Omega L_q \xi_2^* + 2^{1/2}(R + \theta)] \|v\|_{\dot{W}_{\frac{1}{2}}^1(\bar{\omega})}. \end{aligned} \quad (48)$$

Далее, так как сеточная функция $v_{\bar{x}_1}(x_1, x_2)$ обращается в нуль при $x_2 = 0$, $x_2 = l_2$, то (см. [13])

$$\sum_{\omega_2} v_{\bar{x}_1}^2(x) h_2 \leq \frac{l_2^2}{8} \sum_{\omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_2. \quad (49)$$

Умножим эти неравенства на h_1 и просуммируем по $x_1 \in \omega_1^+$, тогда будем иметь

$$\sum_{\omega_1^+ \times \omega_2} v_{\bar{x}_1}^2(x) h_1 h_2 \leq \frac{l_2^2}{8} \sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2. \quad (50)$$

Аналогично можно получить оценку

$$\sum_{\omega_1 \times \omega_2^+} v_{\bar{x}_2}^2(x) h_1 h_2 \leq \frac{l_1^2}{8} \sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2 h_1 h_2. \quad (51)$$

Следовательно,

$$\|v\|_{\dot{W}_{\frac{1}{2}}^1(\bar{\omega})}^2 \leq \frac{l_1^2 + l_2^2}{8} \|v_{\bar{x}_1 \bar{x}_2}\|_{L_2(\omega_1^+ \times \omega_2^+)}^2. \quad (52)$$

Далее, так как (см. (38))

$$\sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 \bar{x}_2}^2(x) h_1 h_2 = \sum_{\omega} v_{\bar{x}_1 x_1} \cdot v_{\bar{x}_2 x_2} h_1 h_2,$$

то

$$2 \sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 x_2}(x) h_1 h_2 \leq \sum_{\omega} (v_{\bar{x}_1 x_2}^2 + v_{x_1 \bar{x}_2}^2) h_1 h_2. \quad (53)$$

Прибавляя к обеим частям неравенства (53) выражение

$$2 \sum_{\omega_1^+ \times \omega_2^+} v_{\bar{x}_1 x_2}^2(x) h_1 h_2,$$

получим оценку

$$\|v_{\bar{x}_1 \bar{x}_2}\|_{L_2(\omega_1^+ \times \omega_2^+)}^2 \leq \frac{1}{4} \|v\|_{W_{2,0}^2(\bar{\omega})}^2. \quad (54)$$

Теперь на основе оценок (52), (54) получим неравенство

$$\|v\|_{\dot{W}_{\frac{1}{2}}^1(\bar{\omega})} \leq \frac{l_1^2 + l_2^2}{32} \|v\|_{W_{2,0}^2(\bar{\omega})}. \quad (55)$$

Вернемся к неравенству (48). Для оценки правой части (48) воспользуемся неравенством (55). В результате установим оценку

$$\|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)} \leq \left\{ 2^{1/2} \bar{v} + [2^{1/2}(R + \theta) + C_\Omega L_q \xi_2^*] \frac{l_1 l_2}{16 C_\Omega} \right\} \|v\|_{W_{2,0}^2(\bar{\omega})}.$$

Лемма доказана.

Следствие 3.1. Для всяких сеточных функций $v_1(x)$, $v_2(x)$, обращающихся в нуль на границе γ сетки $\bar{\omega}$ имеет место оценка

$$\|v_1 - v_2\|_{W_{2,0}^2(\bar{\omega})} \leq \frac{1}{C_0} \left(1 + \frac{M_1 C_\Omega}{\delta_*} \right) \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)}, \quad (56)$$

где константы C_0 , M_1 , C_Ω , δ_* определены в лемме 3.4.

Доказательство. В силу оценки (26) и сеточного аналога неравенства Фридрихса (9) имеем

$$\begin{aligned} \|v\|_{W_{2,0}^2(\bar{\omega})}^2 &\leq \frac{1}{\delta_*} (\bar{A}_h v_1 - \bar{A}_h v_2, v_1 - v_2)_{L_2(\omega)} \leq \\ &\leq \frac{1}{\delta_*} \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)} \|v\|_{L_2(\omega)} \leq \frac{C_\Omega}{\delta_*} \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)} \|v\|_{W_{2,0}^2(\bar{\omega})}. \end{aligned} \quad (57)$$

Теперь из (27) и (57) следует

$$C_0 \|v_1 - v_2\|_{W_{2,0}^2(\bar{\omega})} \leq \left(1 + \frac{M_1 C_\Omega}{\delta_*} \right) \|\bar{A}_h v_1 - \bar{A}_h v_2\|_{L_2(\omega)},$$

т.е. справедливость неравенства (56) доказана.

Вернемся к разностной схеме (24), (25).

Следствие 3.2. Из неравенства (56) и $A_h 0 = 0$ следует, что для решения $y(x)$ разностного уравнения (24), (25) справедлива оценка

$$\|y(x, \Phi_h)\|_{W_{2,0}^2(\bar{\omega})} \leq C \|f^h(x)\|_{L_2(\bar{\omega})}, \quad (58)$$

а также оценки

$$\begin{aligned} \|y(x, \Phi_h)\|_{C(\bar{\omega})} &\leq \hat{M}_1 \|f^h(x)\|_{L_2(\bar{\omega})}, \\ \|y(x, \Phi_h)\|_{L_2(\bar{\omega})} &\leq \hat{M}_2 \|f^h(x)\|_{L_2(\bar{\omega})}, \\ \|y(x, \Phi_h)\|_{W_2^1(\bar{\omega})} &\leq \hat{M}_3 \|f^h(x)\|_{L_2(\bar{\omega})}, \end{aligned} \quad (59)$$

здесь $\hat{M}_1, \hat{M}_2, \hat{M}_3 = \text{const} > 0$.

Теорема 3.1. Пусть разностная аппроксимация задача оптимального управления (1)–(4) определяется соотношениями (19)–(23) и выполнена оценка (58). Тогда

$$\begin{aligned} J_{h*} &= \inf \{ J_h(\Phi_h) : \Phi_h \in U_h \} > -\infty, \\ U_{h*} &= \{ \Phi_{h*} \in U_h : J_h(\Phi_{h*}) = J_{h*} \} \neq \emptyset. \end{aligned}$$

Доказательство теоремы аналогично доказательству теоремы 2.1 и проводится на основе методики [1], [22].

4. АПРИОРНАЯ ОЦЕНКА ПОГРЕШНОСТИ МЕТОДА ПО СОСТОЯНИЮ В НОРМЕ СЕТОЧНОГО ПРОСТРАНСТВА $W_{2,0}^2(\bar{\omega})$

Установим теперь априорную оценку погрешности метода по состоянию в норме сеточного пространства $W_{2,0}^2(\bar{\omega})$. Через $\bar{\psi}_h(x)$ обозначим невязку разностного уравнения (24), (25) (погрешности аппроксимации разностного уравнения (24), (25) на решении исходной задачи для состояния (1), (2))

$$\bar{\psi}_h(x) = \bar{F}_h(x) - \bar{A}_h u. \quad (60)$$

Для определения погрешности $z = y - u$ схемы (24), (25) получим уравнение

$$\bar{A}_h y - \bar{A}_h u = \bar{\Psi}_h(x). \quad (61)$$

В силу неравенства корректности (56) имеем

$$\|y(x, \Phi_h) - u(x, g)\|_{W_{2,0}^2(\bar{\omega})} \leq M \|\bar{\Psi}_h(x)\|_{L_2(\omega)}. \quad (62)$$

Таким образом, оценка погрешности разностной схемы (24), (25) в сеточной норме $W_{2,0}^2(\bar{\omega})$ сводится к оценке $L_2(\omega)$ — нормы погрешности аппроксимации $\bar{\Psi}_h(x)$ разностной схемы (24), (25).

Лемма 4.1. Погрешность аппроксимации разностной схемы по состоянию (24), (25) представима в виде

$$\bar{\Psi}_h(x) = \sum_{\alpha=1}^2 \left[(\eta_{\alpha}^{(1)})_{x_{\alpha}} - \frac{1}{2} \left(B_{\alpha}^{-}(x) + B_{\alpha}^{+}(x) + \hat{B}_{\alpha}^{-}(x) + \hat{B}_{\alpha}^{+}(x) \right) - \eta_3^{(\alpha)}(x) \right], \quad x \in \omega, \quad (63)$$

где сеточные функции $\eta_{\alpha}^{(1)}$, $B_{\alpha}^{\pm}(x)$, $\hat{B}_{\alpha}^{\pm}(x)$, $\eta_3^{(\alpha)}(x)$ имеют вид:

$$\begin{aligned} \eta_{\alpha}^{(1)}(x) &= \left[S^{x_{\beta}} k_{\alpha}^{(-0.5_{\alpha})}(x) \right] u_{\bar{x}_{\alpha}}(x) - S^{x_{\beta}} \left(k_{\alpha} \frac{\partial u}{\partial \xi_{\alpha}} \right)^{(-0.5_{\alpha})}(x), \quad x \in \omega^{+\alpha}, \\ B_{\alpha}^{-}(x) &= \left[S^x B_{\alpha}(x) \right] u_{x_{\alpha}}(x) - S^x \left(b_{\alpha} \frac{\partial u}{\partial \xi_{\alpha}} \right)(x), \\ B_{\alpha}^{+}(x) &= \left[S^x B_{\alpha}(x) \right] u_{\bar{x}_{\alpha}}(x) - S^x \left(b_{\alpha} \frac{\partial u}{\partial \xi_{\alpha}} \right)(x), \\ \hat{B}_{\alpha}^{-}(x) &= (\Phi_{\alpha+1,h}(x) - S^x b_{\alpha+1}) u_{x_{\alpha}}(x), \\ \hat{B}_{\alpha}^{+}(x) &= (\Phi_{\alpha+1,h}(x) - S^x b_{\alpha+1}) u_{\bar{x}_{\alpha}}(x), \\ \eta_3^{(1)}(x) &= [S^x d(x)] q(u(x)) - S^x [d(\xi) q(u(\xi))](x), \\ \eta_3^{(2)}(x) &= (\Phi_{1h}(x) - S^x d(x)) q(u(x)). \end{aligned} \quad (64)$$

Доказательство. Имеем

$$\begin{aligned} \bar{\Psi}_h(x) &= \bar{F}_h(x) - \bar{A}_h u(x) = \\ &= f^h(x) - \left\{ - \sum_{\alpha=1}^2 (a_{\alpha}(x) u_x)_{x_{\alpha}} + \sum_{\alpha=1}^2 \Phi_{\alpha+1,h}(x) u_{x_{\alpha}}^{\circ} + \Phi_{1h}(x) q(u) \right\}. \end{aligned} \quad (65)$$

Для представления погрешности аппроксимации в подходящем виде применим к обеим частям дифференциального уравнения (1) усредняющий оператор $S^x = S^{x_1} S^{x_2}$ с центром в точке $x \in \omega$ (иначе говоря, проинтегрируем уравнение (1) по площадке $e(x)$ и разделим на $h_1 h_2$). Тогда после некоторых преобразований получим соотношение (обобщенное уравнение баланса):

$$\begin{aligned} &- \sum_{\alpha=1}^2 \left(S^{x_{\beta}} \left(k_{\alpha}(x) \frac{\partial u}{\partial \xi_{\alpha}} \right)^{(-0.5_{\alpha})} \right)_{x_{\alpha}}(x) + \sum_{\alpha=1}^2 S^{x_1} S^{x_2} \left(b_{\alpha}(x) \frac{\partial u}{\partial \xi_{\alpha}} \right)(x) + \\ &+ S^{x_1} S^{x_2} (d(\xi) q(u(\xi)))(x) = S^{x_1} S^{x_2} f(x), \quad x \in \omega, \end{aligned} \quad (66)$$

где $S^{x_{\alpha}}$ — усредняющие операторы Стеклова по направлению x_{α} , $\alpha = 1, 2$, ξ — переменная, по которой производится интегрирование в $S^{x_{\alpha}}$, x — точка сетки ω .

Используя соотношение (66), приведем погрешность аппроксимации (65) к виду (63), где сеточные функции — составляющие погрешности аппроксимации имеют вид (64). Лемма 4.1 доказана.

Перейдем к оценке составляющих $(\eta_{\alpha}^{(1)}(x))_{x_{\alpha}}$, $\alpha = 1, 2$, погрешности аппроксимации $\bar{\Psi}_h(x)$.

Лемма 4.2. Пусть $k_{\alpha}(\xi) \in W_{\infty}^1(\Omega)$, $\alpha = 1, 2$, $u(\xi) \in W_2^3(\Omega)$. Тогда для выражений $(\eta_{\alpha}^{(1)}(x))_{x_{\alpha}}$ справедливы оценки:

$$\|(\eta_{\alpha}^{(1)}(x))_{x_{\alpha}}\|_{L_2(\omega)} \leq 4\sqrt{3}|h| \|k_{\alpha}(x)\|_{W_{\infty}^1(\Omega)} [u|_{W_2^2(\Omega)} + |u|_{W_2^3(\Omega)}], \quad \alpha = 1, 2. \quad (67)$$

Доказательство. Нетрудно видеть

$$(\eta_1^{(1)}(x))_{x_1} = \frac{1}{h_2} \int_{e_2(x_2)} \left(k_1(x_1 - 0.5h_1, \xi_2) \left[u_{\bar{x}_1} - \frac{\partial u(x_1 - 0.5h_1, \xi_2)}{\partial \xi_1} \right] \right) d\xi_2. \quad (68)$$

Используя разностные формулы дифференцирования произведения сеточных функций, из (68) получим представление

$$(\eta_1^{(1)}(x))_{x_1} = \overset{\circ}{B}(x) + B(x), \quad (69)$$

где

$$\overset{\circ}{B}(x) = \frac{1}{h_2} \int_{e_2(x_2)} (k_1(x_1 - 0.5h_1, \xi_2))_{x_1} \left[u_{\bar{x}_1} - \frac{\partial u(x_1 - 0.5h_1, \xi_2)}{\partial \xi_1} \right] d\xi_2, \quad (70)$$

$$B(x) = \frac{1}{h_2} \int_{e_2(x_2)} k_1(x_1 + 0.5h_1, \xi_2) \left(u_{\bar{x}_1} - \frac{\partial u(x_1 - 0.5h_1, \xi_2)}{\partial \xi_1} \right)_{x_1} d\xi_2. \quad (71)$$

Оценим величины $\overset{\circ}{B}(x)$, $B(x)$. Нетрудно видеть

$$\overset{\circ}{B}(x) = \frac{1}{h_1 h_2} \int_{e_2(x_2)} \left[\int_{e_1(x_1)} \frac{\partial k_1(\xi_1, \xi_2)}{\partial \xi_1} d\xi_1 \right] \cdot \left[u_{\bar{x}_1}(x) - \frac{\partial u(x_1 - 0.5h_1, \xi_2)}{\partial \xi_1} \right] d\xi_2. \quad (72)$$

Откуда

$$|\overset{\circ}{B}(x)| \leq \frac{1}{h_2} \left\| \frac{\partial k_1(\xi_1, \xi_2)}{\partial \xi_1} \right\|_{L_2(\Omega)} \int_{e_2(x_2)} \left| u_{\bar{x}_1}(x) - \frac{\partial u(x_1 - 0.5h_1, \xi_2)}{\partial \xi_1} \right| d\xi_2. \quad (73)$$

В результате оценки интеграла в правой части неравенства (73) получим

$$|\overset{\circ}{B}(x)| \leq \left\| \frac{\partial k_1(\xi_1, \xi_2)}{\partial \xi_1} \right\|_{L_2(\Omega)} (h_1 h_2)^{-1/2} (h_1 + h_2) |u|_{W_2^2(e^1(x))}, \quad (74)$$

где $e^1(x) = \{\xi = (\xi_1, \xi_2) : x_1 - h_1 \leq \xi_1 \leq x_1, x_2 - 0.5h_2 \leq \xi_2 \leq x_2 + 0.5h_2\}$.

Оценим теперь величину $B(x)$. Можно показать, что справедлива цепочка представлений

$$\begin{aligned} B(x) &= \frac{1}{h_1 h_2} \int_{e_1(x_1)} \int_{e_2(x_2)} k_1(x_1 + 0.5h_1, \xi_2) \left[u_{\bar{x}_1 x_1}(x) - \frac{\partial^2 u(\xi)}{\partial \xi_1^2} \right] d\xi_1 d\xi_2 = \\ &= \frac{1}{h_1^3 h_2} \iint_{e(x)} k_1(x_1 + 0.5h_1, \xi_2) \left\{ \int_{x_1}^{x_1+h_1} \left[\int_{\theta_1-h_1}^{\theta_1} \left(\int_{\xi_2}^{x_2} \frac{\partial^3 u(r_1, m_2)}{\partial r_1^2 \partial m_2} dm_2 + \right. \right. \right. \\ &\quad \left. \left. \left. + \int_{\xi_1}^{r_1} \frac{\partial^3 u(m_1, \xi_2)}{\partial m_1^3} dm_1 \right) dr_1 \right] d\theta_1 \right\} d\xi_1 d\xi_2. \end{aligned} \quad (75)$$

Откуда найдем

$$\begin{aligned} |B(x)| &\leq \frac{1}{h_1^3 h_2} \|k_1(\xi)\|_{L_2(\Omega)} \left[h_1^2 h_2 \int_{x_1-h_1}^{x_1+h_1} \int_{e_2(x_2)} \left| \frac{\partial^3 u(\xi_1, \xi_2)}{\partial \xi_1^2 \partial \xi_2} \right| d\xi_1 d\xi_2 + \right. \\ &\quad \left. + 2h_1^3 \int_{x_1-h_1}^{x_1+h_1} \int_{e_2(x_2)} \left| \frac{\partial^3 u(\xi_1, \xi_2)}{\partial \xi_1^3} \right| d\xi_1 d\xi_2 \right] \leq \|k_1(\xi)\|_{L_2(\Omega)} \left[h_1^{-1} (h_1 h_2)^{1/2} |u|_{W_2^3(e_1^- \times e_2)} + \right. \\ &\quad \left. + 2h_2^{-1} (h_1 h_2)^{1/2} |u|_{W_2^3(e_1^- \times e_2)} + h_1^{-1} (h_1 h_2)^{1/2} |u|_{W_2^3(e_1^+ \times e_2)} + 2h_2^{-1} (h_1 h_2)^{1/2} |u|_{W_2^3(e_1^+ \times e_2)} \right]. \end{aligned}$$

В результате получим

$$|B(x)| \leq 2\|k_1(\xi)\|_{L_2(\Omega)}(h_1 h_2)^{-1/2}(h_1 + h_2)[|u|_{W_2^3(e_1^- \times e_2)} + |u|_{W_2^3(e_1^+ \times e_2)}], \quad x \in \omega, \quad (76)$$

где $e_1^+ = \{x_1 \leq \xi_1 \leq x_1 + h_1\}$, $e_1^- = \{x_1 - h_1 \leq \xi_1 \leq x_1\}$. Из соотношений (70), (74), (76) получаем оценку величины $(\eta_1^{(1)}(x))_{x_1}$:

$$\begin{aligned} |(\eta_1^{(1)}(x))_{x_1}| &\leq 2\|k_1(\xi)\|_{W_\infty^1(\Omega)}(h_1 h_2)^{-1/2}(h_1 + h_2) \times \\ &\times \left[|u|_{W_2^2(e_1^- \times e_2)} |u|_{W_2^3(e_1^- \times e_2)} + |u|_{W_2^3(e_1^+ \times e_2)} \right], \quad x \in \omega. \end{aligned} \quad (77)$$

Аналогично установим оценку

$$\begin{aligned} |(\eta_2^{(1)}(x))_{x_2}| &\leq 2\|k_2(\xi)\|_{W_\infty^1(\Omega)}(h_1 h_2)^{-1/2}(h_1 + h_2) \times \\ &\times \left[|u|_{W_2^2(e_1 \times e_2^-)} |u|_{W_2^3(e_1 \times e_2^-)} + |u|_{W_2^3(e_1 \times e_2^+)} \right], \quad x \in \omega. \end{aligned} \quad (78)$$

Оценка (67) является следствием оценок (77), (78). Лемма доказана.

Лемма 4.3. Пусть решение задачи (1), (2) для состояния $u(x) = u(x, g)$ принадлежит классу $W_{2,0}^2(\Omega)$. Тогда справедливы оценки:

$$\begin{aligned} \|B_\alpha^\pm\|_{L_2(\omega)} &\leq Mh\|b_\alpha\|_{L_\infty(\Omega)}\|u\|_{W_{2,0}^2(\Omega)}, \quad \alpha = 1, 2, \\ \|\hat{B}_\alpha^\pm\|_{L_2(\omega)} &\leq M\|\Phi_{\alpha+1,h}(x) - S^x b_\alpha(x)\|_{L_\infty(\omega)}\|u\|_{W_{2,0}^2(\Omega)}, \quad \alpha = 1, 2, \\ \|\eta_3^{(1)}\|_{L_2(\omega)} &\leq ML_q h \|d(x)\|_{L_\infty(\Omega)}\|u\|_{W_{2,0}^2(\Omega)}, \\ \|\eta_3^{(2)}\|_{L_2(\omega)} &\leq ML_q \|\Phi_{1h}(x) - S^x d(x)\|_{L_\infty(\omega)}\|u\|_{W_{2,0}^2(\Omega)}. \end{aligned} \quad (79)$$

Доказательство оценок (79) не вызывает труда, они довольно громоздки, и мы их опускаем.

Теорема 4.1. Пусть входные данные задачи (1), (2) $k_\alpha(x)$, $f(x)$ таковы, что ее решение $u(\xi) = u(\xi, g)$ (из $W_{2,0}^2(\Omega)$) для любых управлений $g \in U$, где множество U имеет вид (4), принадлежит классу $W_2^3(\Omega)$. Пусть $y(x) = y(x, \Phi_h)$ — решение сеточной задачи (20), (21). Тогда справедлива следующая оценка погрешности метода по состоянию в норме сеточного пространства $W_{2,0}^2(\bar{\omega})$:

$$\begin{aligned} \|y(x, \Phi_h) - u(x, g)\|_{W_{2,0}^2(\bar{\omega})} &\leq \left\{ |h| \sum_{\alpha=1}^2 \|k_\alpha\|_{W_\infty^1(\Omega)} \|u(x, g)\|_{W_2^3(\Omega)} + \right. \\ &+ \left[|h| \left(\sum_{\alpha=1}^2 \|b_\alpha\|_{L_\infty(\Omega)} + L_q \|d\|_{L_\infty(\Omega)} \right) + \sum_{\alpha=1}^2 \|\Phi_{\alpha+1,h}(x) - S^x b_\alpha(x)\|_{L_\infty(\omega)} + \right. \\ &\left. \left. + L_q \|\Phi_{1h}(x) - S^x d(x)\|_{L_\infty(\omega)} \right] \|u(x, g)\|_{W_{2,0}^2(\Omega)} \right\}. \end{aligned} \quad (80)$$

Доказательство теоремы следует из оценки (62), представления погрешности $\bar{\psi}_h$ в виде (63), (64) и оценок норм составляющих погрешности $\bar{\psi}_h$ из (67) и (79).

5. ОЦЕНКИ ПОГРЕШНОСТИ СЕТОЧНОГО ФУНКЦИОНАЛА, СКОРОСТИ СХОДИМОСТИ АППРОКСИМАЦИЙ ПО ФУНКЦИОНАЛУ, СХОДИМОСТЬ ПО УПРАВЛЕНИЮ. РЕГУЛЯРИЗАЦИЯ АППРОКСИМАЦИЙ

Определим некоторые кусочно-постоянные восполнения сеточных функций. Пусть $v(x_\alpha)$ — сеточная функция, заданная на одномерной сетке $\bar{\omega}_\alpha \subset [0, l_\alpha]$, $\alpha = 1, 2$. Кусочно-постоянное восполнение $v(x_\alpha)$ на $[0, l_\alpha]$, $\alpha = 1, 2$, определим по формуле

$$\bar{v}(\xi_\alpha) = \sum_{i_\alpha=0}^{N_\alpha} v(x_\alpha)^{(i_\alpha)} \varphi_{i_\alpha}(\xi), \quad \varphi_{i_\alpha}(\xi) = \begin{cases} 1, & \xi_\alpha \in e_\alpha(x_\alpha^{(i_\alpha)}), \\ 0, & \xi_\alpha \notin e_\alpha(x_\alpha^{(i_\alpha)}). \end{cases} \quad (81)$$

Справедлива следующая лемма.

Лемма 5.1. Пусть $\overline{\psi^{h_2}}(\xi_2)$, $\overline{u_{\bar{x}_1}}(l_1, \xi_2)$, $\overline{a_1}(l_1, \xi_2)$ — кусочно-постоянные восполнения на $[0, l_2]$ сеточных функций $\psi^{h_2}(\xi_2)$, $u_{\bar{x}_1}(l_1, \xi_2)$, $a_1(l_1, \xi_2)$, $x_2 \in \bar{\omega}_2 \in [0, l_2]$ по формулам, аналогичным (81). Тогда справедливы оценки

$$\|\psi^{h_2}(x_2)\|_{L_2(\omega_2)} \leq \|\Psi(\xi_2)\|_{L_2(0, l_2)}, \quad (82)$$

$$\|\overline{\psi^{h_2}}(x_2)\|_{L_2(0, l_2)} \leq \|\Psi(\xi_2)\|_{L_2(0, l_2)}, \quad (83)$$

$$\|\psi^{h_2}(\xi_2) - \overline{\psi^{h_2}}(x_2)\|_{L_2(0, l_2)} \leq |h| \|\Psi(\xi_2)\|_{W_2^1(0, l_2)}, \quad (84)$$

$$\|\overline{u_{\bar{x}_1}}(l_1, x_2)\|_{L_2(0, l_2)}^2 = \|u_{\bar{x}_1}(l_1, x_2)\|_{L_2(\bar{\omega}_2)}^2 \leq C \left[|h| \|u\|_{W_2^1(\Omega)}^2 + \|u\|_{W_2^2(\Omega)}^2 \right] \leq M \|u\|_{W_2^2(\Omega)}^2, \quad (85)$$

$$\left\| k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \overline{a_1}(l_1, \xi_2) \overline{u_{\bar{x}_1}}(l_1, \xi_2) \right\|_{L_2(0, l_2)} \leq M |h| \|k_1\|_{W_\infty^1(\Omega)} \|u\|_{W_2^3(\Omega)}. \quad (86)$$

Доказательство. Установим оценку (85). Нетрудно показать, что справедливо представление

$$u_{\bar{x}_1}(x) = u_{x_1}(x_1 - h_1, x_2) = \tilde{h}_2(h_1 \tilde{h}_2)^{-1/2} |u(\xi)|_{W_2^2(e^1(x_1 - h_1, x_2))} + \\ + (h_1 \tilde{h}_2)^{-1/2} |u(\xi)|_{W_2^1(e^1(x_1 - h_1, x_2))}, \quad x \in \omega_1^+ \times \bar{\omega}_2, \quad (87)$$

где $e^1(x) = \bar{e}_1(x_1) \times e_2(x_2) = (x_1, x_1 + h_1) \times e_2(x_2)$.

Из (87) при $x_1 = l_1$ получаем

$$u_{\bar{x}_1}(l_1, x_2) = \tilde{h}_2(h_1 \tilde{h}_2)^{-1/2} |u(\xi)|_{W_2^2((l_1 - h_1, l_1) \times (0, l_2))} + (h_1 \tilde{h}_2)^{-1/2} |u(\xi)|_{W_2^1((l_1 - h_1, l_1) \times (0, l_2))}. \quad (88)$$

Принимая во внимание оценку (88), установим

$$\|u_{\bar{x}_1}(l_1, x_2)\|_{L_2(\bar{\omega}_2)}^2 \leq 2 \left[h_1^{-1} \tilde{h}_2^2 |u(\xi)|_{W_2^2((l_1 - h_1) \times (0, l_2))}^2 + h_1^{-1} |u(\xi)|_{W_2^1((l_1 - h_1) \times (0, l_2))}^2 \right]. \quad (89)$$

Используя неравенство [21] имеем

$$|u(\xi)|_{W_2^1(\Omega_h)}^2 \leq M_0 h_1 \|u(\xi)\|_{W_2^2(\Omega)}^2, \quad (90)$$

где $\Omega_h = (l_1 - h_1, l_1) \times (0, l_2)$, из (89) получим оценку (85).

Докажем теперь оценку (86). Нетрудно видеть, что

$$\left\| k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \overline{a_1}(l_1, \xi_2) \overline{u_{\bar{x}_1}}(l_1, \xi_2) \right\|_{L_2(0, l_2)}^2 = \sum_{\bar{\omega}_2} \int_{e_2(x_2)} \left[k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \right. \\ \left. - a_1(l_1, x_2) u_{\bar{x}_1}(l_1, x_2) \right]^2 d\xi_2 = \sum_{\bar{\omega}_2} \int_{e_2(x_2)} \left[\left(\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - u_{\bar{x}_1}(l_1, x_2) \right) a_1(l_1, x_2) + \right. \\ \left. + (k_1(l_1, \xi_2) - a_1(l_1, x_2)) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right]^2 d\xi_2 \leq 2 \sum_{\bar{\omega}_2} \int_{e_2(x_2)} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - u_{\bar{x}_1}(l_1, x_2) \right]^2 \times \\ \times |a_1(l_1, x_2)|^2 d\xi_2 + 2 \sum_{\bar{\omega}_2} \int_{e_2(x_2)} [k_1(l_1, \xi_2) - a_1(l_1, x_2)]^2 \left| \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right|^2 d\xi_2. \quad (91)$$

Оценим первое слагаемое в правой части (91). Имеем

$$|a_1(l_1, x_2)| = \left| \frac{1}{\tilde{h}_2} \int_{e_2(x_2)} k_1(l_1 - 0.5h_1, r) dr \right| \leq \|k_1(\xi)\|_{L_\infty(\Omega)} \quad \forall x_2 \in \bar{\omega}_2. \quad (92)$$

Далее

$$\begin{aligned}
& \int_{e_2(x_2)} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - u_{\bar{x}_1}(l_1, x_2) \right]^2 d\xi_2 = \int_{e_2(x_2)} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \frac{1}{h_1} \int_{l_1-h_1}^{l_1} \frac{\partial u(\xi_1, x_2)}{\partial \xi_1} d\xi_1 \right]^2 d\xi_2 = \\
& = \int_{e_2(x_2)} \left\{ \frac{1}{h_1} \int_{l_1-h_1}^{l_1} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \frac{\partial u(\xi_1, \xi_2)}{\partial \xi_1} + \frac{\partial u(\xi_1, \xi_2)}{\partial \xi_1} - \frac{\partial u(\xi_1, x_2)}{\partial \xi_1} \right] d\xi_1 \right\}^2 d\xi_2 = \\
& = \frac{1}{h_1^2} \int_{e_2(x_2)} \left\{ \int_{l_1-h_1}^{l_1} \left[\int_{\xi_1}^{l_1} \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} dm_1 + \int_{x_2}^{\xi_2} \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} dm_2 \right] d\xi_1 \right\}^2 d\xi_2 \leq \\
& \leq \frac{1}{h_1^2} \int_{e_2(x_2)} \left\{ \int_{l_1-h_1}^{l_1} \left[\int_{\xi_1}^{l_1} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right| dm_1 + \int_{x_2}^{\xi_2} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right| dm_2 \right] d\xi_1 \right\}^2 d\xi_2 = \quad (93) \\
& = \frac{1}{h_1^2} \int_{e_2(x_2)} \left\{ h_1 \int_{l_1-h_1}^{l_1} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right| dm_1 + \int_{l_1-h_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right| dm_2 d\xi_1 \right\}^2 d\xi_2 \leq \\
& \leq \frac{2}{h_1^2} \int_{e_2(x_2)} \left\{ h_1^2 \left(\int_{\xi_1}^{l_1} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right| dm_1 \right)^2 + \left(\int_{l_1-h_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right| dm_2 d\xi_1 \right)^2 \right\} d\xi_2 = \\
& = 2 \int_{e_2(x_2)} \left(\int_{l_1-h_1}^{l_1} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right| dm_1 \right)^2 d\xi_2 + \frac{2h_2}{h_1^2} \left(\int_{l_1-h_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right| dm_2 d\xi_1 \right)^2 \leq \\
& \leq 2 \int_{e_2(x_2)} h_1 \left(\int_{\xi_1}^{l_1} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right| dm_1 \right)^2 d\xi_2 + \frac{2h_2}{h_1^2} h_1 h_2 \int_{\xi_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right| dm_2 d\xi_1 = \\
& = 2h_1 \int_{\xi_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(m_1, \xi_2)}{\partial m_1^2} \right|^2 dm_1 d\xi_2 + \frac{2h_2^2}{h_1} \int_{\xi_1}^{l_1} \int_{e_2(x_2)} \left| \frac{\partial^2 u(\xi_1, m_2)}{\partial \xi_1 \partial m_2} \right|^2 dm_2 d\xi_1.
\end{aligned}$$

На основе оценок (92), (93) получаем оценку первого слагаемого в правой части неравенства (91):

$$\sum_{\bar{\omega}_2} \int_{e_2(x_2)} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - u_{\bar{x}_1}(l_1, x_2) \right]^2 |a_1(l_1, x_2)|^2 d\xi_2 \leq M|h| \|k_1(\xi)\|_{L_\infty(\Omega)} |u(\xi)|_{W_2^2(\Omega_h)}^2. \quad (94)$$

Используя неравенство (см. [21, с. 24])

$$|u(\xi)|_{W_2^2(\Omega_h)}^2 \leq M|h| \|u\|_{W_2^3(\Omega)}^2, \quad (95)$$

из (94) установим оценку

$$\sum_{\bar{\omega}_2} \int_{e_2(x_2)} \left[\frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - u_{\bar{x}_1}(l_1, x_2) \right]^2 |a_1(l_1, x_2)|^2 d\xi_2 \leq M|h|^2 \|k_1(\xi)\|_{L_\infty(\Omega)} \|u\|_{W_2^3(\Omega)}^2. \quad (96)$$

Перейдем теперь к оценке второго слагаемого в правой части неравенства (90). Имеем

$$\begin{aligned}
 & \sum_{\omega_2} \int_{e_2(x_2)} [k_1(l_1, \xi_2) - a_1(l_1, x_2)]^2 \left| \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right|^2 d\xi_2 \leq \\
 & \leq \left(\sum_{\omega_2} \int_{e_2(x_2)} [k_1(l_1, \xi_2) - a_1(l_1, x_2)]^4 d\xi_2 \right)^{1/2} \left(\sum_{\omega_2} \int_{e_2(x_2)} \left| \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right|^4 d\xi_2 \right)^{1/2} = \\
 & = \left(\sum_{\omega_2} \int_{e_2(x_2)} \left| \frac{1}{\bar{h}_2} \int_{e_2(x_2)} [k_1(l_1, \xi_2) - k_1(l_1 - 0.5h_1, r_2)] dr_2 \right|^4 d\xi_2 \right)^{1/2} \times \\
 & \times \left(\sum_{\omega_2} \int_{e_2(x_2)} \left| \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right|^4 d\xi_2 \right)^{1/2} \leq M \|u(\xi)\|_{W_2^2(\Omega)}^2 \times \\
 & \times \left(\sum_{\omega_2} \int_{e_2(x_2)} \left| \frac{1}{\bar{h}_2} \int_{e_2(x_2)} \left[\int_{l_1-0.5h_1}^{l_1} \frac{\partial k_1(m_1, \xi_2)}{\partial m_1} dm_1 + \int_{r_2}^{\xi_2} \frac{\partial k_1(l_1 - 0.5h_1, m_2)}{\partial m_2} dm_2 \right] dr_2 \right|^4 d\xi_2 \right)^{1/2} \leq \\
 & \leq M_1 |h|^2 \left(\left\| \frac{\partial k_1(\xi)}{\partial \xi_1} \right\|_{L_\infty(\Omega)} + \left\| \frac{\partial k_1(\xi)}{\partial \xi_2} \right\|_{L_\infty(\Omega)} \right)^2 \|u(\xi)\|_{W_2^2(\Omega)}^2.
 \end{aligned} \tag{97}$$

Принимая во внимание оценки (96), (97) и неравенство (91), получаем оценку (86). Лемма доказана.

Теорема 5.1. Пусть выполнены условия теоремы 4.1. Для любых управлений $g \in U$ и $\Phi_h \in U_h$ и любых $h > 0$ для погрешности сеточного функционала (19) справедлива оценка

$$|J(g) - J_h(\Phi_h)| \leq M \left[|h| + \|S^x d(x) - \Phi_{1h}(x)\|_{L_\infty(\omega)} + \sum_{\alpha=1}^2 \|S^x b_\alpha(x) - \Phi_{\alpha+1,h}(x)\|_{L_\infty(\omega)} \right]. \tag{98}$$

Доказательство. Принимая во внимание (3), (19) приведем погрешность аппроксимации функционала $\epsilon_h(g, \Phi) = J(g) - J_h(\Phi_h)$ к специальному виду

$$\epsilon_h(g, \Phi_h) = \sum_{k=1}^2 D_h^{(k)}, \tag{99}$$

где

$$D_h^{(1)} = \left\| k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \psi(\xi_2) \right\|_{L_2(0, l_2)}^2 - \left\| \bar{a}_1(l_1, \xi_2) \bar{u}_{\bar{x}_1}(l_1, \xi_2) - \bar{\psi}^{h_2}(\xi_2) \right\|_{L_2(0, l_2)}^2, \tag{100}$$

$$D_h^{(2)} = \left\| \bar{a}_1(l_1, \xi_2) \bar{u}_{\bar{x}_1}(l_1, \xi_2) - \bar{\psi}^{h_2}(\xi_2) \right\|_{L_2(0, l_2)}^2 - \left\| a_1(l_1, x_2) y_{\bar{x}_1}(l_1, x_2) - \psi^{h_2}(x_2) \right\|_{L_2(\omega_2)}^2. \tag{101}$$

Оценим величины $D_h^{(k)}$, $k = 1, 2$. Нетрудно убедиться, что справедливы оценки

$$\begin{aligned}
 |D_h^{(1)}| \leq & \left[\left\| k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} - \bar{a}_1(l_1, \xi_2) \bar{u}_{\bar{x}_1}(l_1, \xi_2) \right\| + \right. \\
 & \left. + \left\| \psi(\xi_2) - \bar{\psi}^{h_2}(\xi_2) \right\|_{L_2(0, l_2)} \right] \cdot \left[\left\| k_1(l_1, \xi_2) \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right\|_{L_2(0, l_2)} + \right. \\
 & \left. + \left\| \bar{a}_1(l_1, \xi_2) \bar{u}_{\bar{x}_1}(l_1, \xi_2) \right\|_{L_2(0, l_2)} + \left\| \psi(\xi_2) \right\|_{L_2(0, l_2)} + \left\| \bar{\psi}^{h_2}(\xi_2) \right\|_{L_2(0, l_2)} \right],
 \end{aligned} \tag{102}$$

$$|D_h^{(2)}| \leq \|a_1(l_1, x_2)[u_{\bar{x}_1}(l_1, x_2) - y_{\bar{x}_1}(l_1, x_2)]\|_{L_2(\bar{\omega}_2)} \times \\ \times [\|a_1(l_1, x_2)u_{\bar{x}_1}(l_1, x_2)\|_{L_2(\bar{\omega}_2)} + \|a_1(l_1, x_2)y_{\bar{x}_1}(l_1, x_2) + 2\psi^{h_2}(x_2)\|_{L_2(\bar{\omega}_2)}]. \quad (103)$$

Наконец проведем оценки правых частей неравенств (102), (103). Для этого воспользуемся оценками (82)–(86), неравенствами

$$\left\| \frac{\partial u(l_1, \xi_2)}{\partial \xi_1} \right\|_{L_2(0, l_2)} \leq M_1 \|u(\xi)\|_{W_2^2(\Omega)}, \quad (104)$$

$$\|y_{\bar{x}_1}(l_1, x_2)\|_{L_2(\bar{\omega}_2)} \leq M_2 \|y(x)\|_{W_{2,0}^2(\bar{\omega})}, \quad (105)$$

$$\|y_{\bar{x}_1}(l_1, x_2) - u_{\bar{x}_1}(l_1, x_2)\|_{L_2(\bar{\omega}_2)} \leq M_2 \|y - u\|_{W_{2,0}^2(\bar{\omega})}, \quad (106)$$

вытекающими из теорем вложения Соболева и их разностных аналогов, а также оценками (8), (10), (20). В результате, для величины $\varepsilon_h(g, \Phi_h)$ установим оценку (98). Теорема доказана.

Будем допускать, что вычисления функционалов $J_h(\Phi_h)$, определяемых соотношением (19), ведутся приближенно, и получающийся приближенный функционал $J_{h\delta_h}(\Phi_h)$ связан с $J_h(\Phi_h)$ соотношениями

$$J_{h\delta_h}(\Phi_h) = J_h(\Phi_h) + \theta_{\delta_h}(\Phi_h), \quad |\theta_{\delta_h}(\Phi_h)| \leq \delta_h, \\ \forall \Phi_h \in U_h, \quad \delta_h \rightarrow +0 \text{ при } |h| \rightarrow 0. \quad (107)$$

Предположим, что при каждом $h = (h_1, h_2)$ и соответствующей сетки $\bar{\omega} = \bar{\omega}_h$ с помощью какого-либо метода минимизации получено сеточное управление $\hat{\Phi}_h = \Phi_{h\delta_h\varepsilon_h} \in U_h$, дающее решение задачи (19)–(23), (107) в смысле

$$J_{h\delta_h*} \leq J_{h\delta_h}(\hat{\Phi}_h) \leq J_{h\delta_h*} + \varepsilon_h, \quad \hat{\Phi}_h \in U_h, \quad (108)$$

где последовательность $\{\varepsilon_h\}$ такова, что $\varepsilon_h > 0$ и $\varepsilon_h \rightarrow 0$ при $h \rightarrow 0$.

Определим отображения

$$R_h : H \rightarrow H_h \text{ и } N_h : H_h \rightarrow H$$

по правилу: $R_h g = \Phi_h$, где

$$g = (g_1(t), g_2(t), g_3(t)) = (d(t), b_1(t), b_2(t)), \quad \Phi_h(x) = (\bar{R}_h g_1(x), \bar{R}_h g_2(x), \bar{R}_h g_3(x)),$$

а $\bar{R}_h g_k(x)$, $k = 1, 2, 3$ — дискретизации на сетке $\omega = \omega_1 \times \omega_2$ управляющих функций $g_k(t)$, $k = 1, 2, 3$, по формуле $\bar{R}_h g_k(x) = S^x g_k$, $k = 1, 2, 3$; $N_h \Phi_h = g$, где $\Phi_h = (\Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x))$, $g = (Q_h \Phi_{1h}(t), Q_h \Phi_{2h}(t), Q_h \Phi_{3h}(t))$, $Q_h \Phi_{kh}(t)$, $k = 1, 2, 3$ — кусочно-линейные восполнения на Ω сеточных управлений $\Phi_{kh}(x)$, $k = 1, 2, 3$, $x \in \omega$, такие что $R_h g \in U_h$, $N_h \Phi_h \in U$.

Справедлива

Теорема 5.2. Пусть выполнены условия теоремы 5.1. Тогда последовательность разностных задач минимизации (19), (20)–(21), (22)–(23) при $|h| \rightarrow 0$ аппроксимирует исходную экстремальную задачу (3), (4), (1)–(2) по функционалу, т.е. $\lim J_{h\delta_h*} = J_*$ при $|h| \rightarrow 0$, причем справедлива оценка скорости сходимости

$$|J_{h\delta_h*} - J_*| \leq |J_{h*} - J_*| + \delta_h \leq C|h| + \delta_h; \quad (109)$$

если последовательность сеточных управлений $\{\hat{\Phi}_h\} \subset U_h$ определена из условий (108), то последовательность управлений $\{N_h \hat{\Phi}_h\}$ является минимизирующей для задачи (3), (4), (1)–(2) и справедлива оценка скорости сходимости

$$0 \leq J(N_h \hat{\Phi}_h) - J_* \leq C[|h| + \varepsilon_h + \delta_h];$$

последовательность $\{N_h \hat{\Phi}_h\}$ слабо в H сходится к множеству $U_* \neq \emptyset$ оптимальных управлений задачи (3), (4), (1)–(2).

Доказательство теоремы 5.2 проводится на основе методики [1], [22] и опирается на полученные выше результаты.

Согласно теореме 2.1 задача (3), (1), (2), (4) является корректно поставленной в слабой топологии H . В сильной топологии пространства H (в метрике пространства H) задача, вообще говоря, некорректна, т.е. могут существовать минимизирующие последовательности, не сходящиеся по норме H к множеству U_* . Поэтому для построения минимизирующей последовательности, сходящейся в метрике пространства H к множеству U_* воспользуемся методом регуляризации А. Н. Тихонова [1], [22]. Введем стабилизатор $\Omega(g) = \|g\|_H^2$ и его разностный аналог $\Omega_h(\Phi_h) = \|\Phi_h\|_{H_h}^2$, где $H_h = (W_2^1(\omega))^3$. При каждом $h > 0$ рассмотрим задачу минимизации дискретного функционала Тихонова

$$T_{h\delta_h\alpha_h}(\Phi_h) = J_{h\delta_h}(\Phi_h) + \alpha_h\Omega_h(\Phi_h), \quad \Phi_h \in U_h, \quad (110)$$

на множестве U_h , где $\{\alpha_h\}$ — произвольная последовательность положительных чисел, сходящихся к нулю при $|h| \rightarrow 0$.

Определим последовательность $\check{\Phi}_h = \{\Phi_{h\delta_h\alpha_h v_h}\}$ из условий

$$T_{h\delta_h\alpha_h*} \equiv \inf\{T_{h\delta_h\alpha_h} : \Phi_h \in U_h\} \leq T_{h\delta_h\alpha_h}(\check{\Phi}) \leq T_{h\delta_h\alpha_h*} + v_h, \quad \check{\Phi}_h \in U_h, \quad (111)$$

где $\{v_h\}$ — положительная последовательность, сходящаяся к нулю при $h \rightarrow 0$.

Теорема 5.3. Пусть выполнены условия теоремы 5.1, последовательность $\check{\Phi}_h = \{(\check{\Phi}_{1h}(x), \check{\Phi}_{2h}(x), \check{\Phi}_{3h}(x))\}$ определена из условий (111). Тогда последовательность управлений $\{N_h\check{\Phi}_h\}$ является минимизирующей для задачи (3), (1), (2), (4) и справедлива оценка скорости сходимости

$$0 \leq J(N_h\check{\Phi}_h) - J_* \leq C[|h| + \delta_h + v_h + \alpha_h];$$

если, кроме того, параметры α_h, δ_h, v_h согласованы так, что $(|h| + v_h + \delta_h)/\alpha_h \rightarrow 0$ при $|h| \rightarrow 0$, то последовательность $\{N_h\check{\Phi}_h\} = \{Q_h\check{\Phi}_{1h}(t), Q_h\check{\Phi}_{2h}(t), Q_h\check{\Phi}_{3h}(t)\}$ сильно в H сходится к множеству $U_{**} = \{g_{**} \in U_* : \Omega(g_{**} = \|g_{**}\|_H^2) = \inf\{\|g_*\|_H^2 : g_* \in U_*\} = \Omega_*\}$ — нормальных решений задачи (3), (1), (2), (4), т.е.

$$\lim_{|h| \rightarrow 0} \inf\{\|N_h\check{\Phi}_h - g_{**}\|_H : g_{**} \in U_{**}\} = 0 \text{ и } \lim \Omega(N_h\check{\Phi}_h) = \Omega_* \text{ при } |h| \rightarrow 0.$$

Доказательство теоремы проводится на основе методики [1], [22] и опирается на полученные выше результаты.

СПИСОК ЛИТЕРАТУРЫ

1. Васильев Ф. П. Методы решения экстремальных задач. М.: Наука, 1981. 400 с.
2. Лионс Ж.-Л. Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972. 414 с.
3. Лурье К. А. Оптимальное управление в задачах математической физики. М.: Наука, 1975. 480 с.
4. Литвинов В. Г. Оптимизация в эллиптических граничных задачах с приложениями к механике. М.: Наука, 1987. 368 с.
5. Райтум У. Ё. Задачи оптимального управления для эллиптических уравнений. Рига: Зинатне, 1989. 276 с.
6. Егоров А. И. Оптимальное управление тепловыми и диффузионными процессами. М.: Наука, 1978.
7. Ишмухаметов А. З. Вопросы устойчивости и аппроксимации задач оптимального управления. М.: ВЦ РАН, 1999.
8. Ишмухаметов А. З. Вопросы устойчивости и аппроксимации задач оптимального управления системами с распределенными параметрами. М.: ВЦ РАН, 2001.
9. Потапов М. М. Аппроксимация экстремальных задач в математической физике (гиперболические уравнения). М.: Изд-во МГУ, 1985.

10. *Лубышев Ф. В.* Разностные аппроксимации задач оптимального управления системами, описываемыми уравнениями в частных производных. Уфа: БашГУ, 1999. 244 с.
11. *Лубышев Ф. В., Манапова А. Р.* Аппроксимации задач оптимального управления старшими коэффициентами эллиптических уравнений в недивергентной форме с неограниченной нелинейностью в коэффициентах // Дифференц. ур-ния. 2021. Т. 57. № 6. С. 796–820.
12. *Самарский А. А., Вабищевич П. Н.* Вычислительная теплопередача. Изд. 2-е. М.: Книжный дом “ЛИБРОКОМ”, 2009. 784 с.
13. *Самарский А. А., Андреев В. Б.* Разностные схемы для эллиптических уравнений. М.: Наука, 1976.
14. *Самарский А. А.* Теория разностных схем. М.: Наука, 1989.
15. *Самарский А. А., Николаев Е. С.* Методы решения сеточных уравнений. М.: Наука, 1978. 589 с.
16. *Самарский А. А., Гулин А. В.* Устойчивость разностных схем. М.: Наука, 1973. 416 с.
17. *Самарский А. А., Лазаров Р. Д., Макаров В. Л.* Разностные схемы для дифференциальных уравнений с обобщенными решениями. М.: Высшая школа, 1987.
18. *Гаевский Х., Грёгер К., Захариас К.* Нелинейные операторные уравнения и операторные дифференциальные уравнения. М.: Мир, 1978. 336 с.
19. *Куфнер А., Фучик С.* Нелинейные дифференциальные уравнения. М.: Наука, 1988. 304 с.
20. *Соболев С. Л.* Некоторые применения функционального анализа в математической физике. М.: Наука, 1988. 334 с.
21. *Оганесян Л. А., Руховец Л. А.* Вариационно-разностные методы решения эллиптических уравнений. Ереван: Из-во АН Армянской ССР, 1979. 334 с.
22. *Тихонов А. Н., Арсенин В. Я.* Методы решения некорректных задач. М.: Наука, 1986. 287 с.
23. *Ладженская О. А.* Краевые задачи математической физики. М.: Наука, 1973. 408 с.

APPROXIMATION OF OPTIMAL CONTROL PROBLEMS FOR SEMILINEAR ELLIPTIC CONVECTION-DIFFUSION EQUATIONS WITH BOUNDARY OBSERVATION OF THE CONORMAL DERIVATIVE, WITH CONTROLS IN THE COEFFICIENTS OF THE CONVECTIVE TRANSPORT OPERATOR AND THE NONLINEAR TERM OF THE EQUATION

F. V. Lubyshev^a, M. E. Fairuzov^{a,*}

^a450076, Ufa, Zaki Validi Str. 32, Ufa Science and Technology University

*e-mail: fairuzovme@mail.ru

Received May 22, 2023

Revised February 25, 2024

Accepted April 5, 2024

Abstract. The paper studies difference approximations of the optimal control problem with boundary observation of the conormal derivative of the state described by the Dirichlet problem for semi-linear elliptic equations with controls in the coefficients of the convective transport operator and the nonlinear term of the equation. The issues of the correctness of the formulation of the optimal control problem were considered. Differential approximations of the optimal control problem are constructed. The issues of convergence of approximations in terms of functionality and control are studied. The regularization of approximations is carried out.

Keywords: optimal control problem, semi-linear elliptic equations, Dirichlet boundary value problem, difference approximations, convergence of approximations, regularization of approximations.

ПРИМЕНЕНИЕ УНИВЕРСАЛЬНОЙ БЫСТРОЙ ТРИГОНОМЕТРИЧЕСКОЙ ИНТЕРПОЛЯЦИИ И ЭКСТРАПОЛЯЦИИ ДЛЯ ОПРЕДЕЛЕНИЯ КООРДИНАТ ТОЧКИ ЗАПУСКА ЛЕТАТЕЛЬНОГО АППАРАТА

© 2024 г. А. Д. Чернышов^{1,*}, О. Ю. Никифорова^{1,**}, В. В. Горяйнов^{2,***}, И. Г. Рукин^{1,****}

¹394000 Воронеж, пр-т Революции, 19, Воронежский государственный университет инженерных технологий, Россия

²394006 Воронеж, ул. 20-летия Октября, 84, Воронежский государственный технический университет, Россия

*e-mail: chernyshovad@mail.ru

**e-mail: niki22@mail.ru

***e-mail: gorvit77@mail.ru

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

Поступила в редакцию 30.10.2023 г.

Переработанный вариант 28.02.2024 г.

Принята к публикации 05.04.2024 г.

Приведены основы быстрой универсальной тригонометрической интерполяции для неперiodических функций, позволяющей получить приближенное решение с высокой точностью. Получены формулы для вычисления координат точки запуска летательного аппарата (ЛА) с высокой точностью за счет одновременного применения метода универсальных быстрых тригонометрических интерполяций и метода экстраполяции на концах некоторого заданного отрезка. Численные эксперименты показывают, что после запуска первого ЛА координаты пускового механизма могут быть определены за 7 сек с точностью 10^{-17} м. Библ. 17. Фиг. 3. Табл. 3.

Ключевые слова: быстрые разложения, быстрая тригонометрическая интерполяция, экстраполяция, высокая точность, летательный аппарат, координаты, точка запуска.

DOI: 10.31857/S0044466924070063, EDN: xirkgm

ВВЕДЕНИЕ

Классические ряды Фурье медленно сходятся и в общем случае не допускают почленное дифференцирование [1]. При применении быстрых разложений с граничной функцией M_p p -го порядка коэффициенты Фурье с увеличением порядкового номера m в ряде быстро уменьшаются по закону m^{-p-2} , что допускает возможность p -кратного дифференцирования ряда и обеспечивает высокую точность и скорость его сходимости [2]. Увеличению скорости сходимости рядов Фурье для неперiodической функции посвящены многочисленные исследования, например, [3–5]. Наиболее эффективными в данном направлении являются быстрые разложения [2, 6]. В работе [7] показано, что быстрые разложения не только увеличивают скорость сходимости ряда в сотни раз по сравнению со сходимостью классических рядов Фурье, но и позволяют их почленно дифференцировать p раз. В этой связи в дальнейших рассмотрениях будем использовать универсальные быстрые разложения. Это увеличивает точность определения координат механизма, из которого происходят запуски ЛА. Под ЛА будем понимать ракету или снаряд. При многократном запуске ЛА из одного механизма задача определения его координат упрощается.

В данной статье для определения координат точки запуска ЛА все рассмотрения будут проводиться в прямоугольной декартовой геоцентрической системе координат, связанной с Землей, с учетом реактивной силы предполагаемой ракеты, силы тяжести, зависящей от расстояния до центра Земли, нелинейной силы сопротивления атмосферы и силы Кориолиса. Вследствие нелинейности полученной системы с дробными степенями относительно неизвестных для определения коэффициентов

Фурье будем использовать метод универсальных быстрых тригонометрических интерполяций. Поточечное определение коэффициентов ряда применяют в методе коллокаций [8–10] и в методе тригонометрической интерполяции [11]. Ранее метод быстрых тригонометрических интерполяций с успехом использовался для вычисления определенных интегралов с переменным верхним пределом от сложных или неявно заданных функций [12] и решения ряда прикладных задач, например, [13–15]. Приведем основы универсальных быстрых интерполяционных разложений.

1. БЫСТРЫЕ УНИВЕРСАЛЬНЫЕ ТРИГОНОМЕТРИЧЕСКИЕ ИНТЕРПОЛЯЦИИ

Пусть некоторая регистрирующая установка (РУ) определяет положение подвижного ЛА в пространстве. Промежуток интерполяции $2a$ определяется разностью между концом $\tilde{t}_2$ и началом $\tilde{t}_1$ времени работы РУ

$$\tilde{t}_2 - \tilde{t}_1 = 2a. \quad (1)$$

Для математических вычислений удобнее ввести вспомогательную временную переменную t , связанную с реальным временем $\tilde{t}$ формулой

$$t = \tilde{t} - \frac{\tilde{t}_2 + \tilde{t}_1}{2}. \quad (2)$$

При построении быстрых разложений будем использовать следующий полный функциональный базис Фурье:

$$\left\{ 1, \cos \frac{m\pi t}{a}, \sin \frac{m\pi t}{a} \right\}, \quad t \in [-a, a], \quad m = 1, 2, \dots \quad (3)$$

Пусть некоторая непериодическая функция $f(t) \in L_2^a[-a \leq t \leq a]$ из пространства Гильберта рассматривается на временном отрезке $[-a \leq t \leq a]$.

Быстрым разложением $f(t)$, $[-a \leq t \leq a]$ назовем сумму некоторой граничной функции $M_p(t)$ и ряда Фурье для разности $f(t) - M_p(t)$, т.е.

$$\begin{aligned} f(t) &= M_p(t) + a_0 + \sum_{m=1}^{\infty} a_m \cos m\pi \frac{t}{a} + b_m \sin m\pi \frac{t}{a}, \\ a_0 &= \frac{1}{2a} \int_{-a}^a (f(t) - M_p(t)) dt, \quad a_m = \frac{1}{a} \int_{-a}^a (f(t) - M_p(t)) \cos m\pi \frac{t}{a} dt, \\ b_m &= \frac{1}{a} \int_{-a}^a (f(t) - M_p(t)) \sin m\pi \frac{t}{a} dt, \quad t \in [-a, a]. \end{aligned} \quad (4)$$

Граничная функция $M_p(t)$ в (4) введена с целью увеличения точности и скорости сходимости ряда Фурье для разности $(f(t) - M_p(t))$ и возможности его многократного дифференцирования, где $M_p(t)$ определим следующей суммой из $p + 1$ слагаемых:

$$M_p(t) = \sum_{q=0}^p A_q P_q(t), \quad (5)$$

где A_q — некоторые постоянные, а $P_q(t)$ — быстрые полиномы.

Коэффициенты A_q определяются по формуле ([6])

$$A_q = f^{(q)}(a) - f^{(q)}(-a). \quad (6)$$

Полиномы четной и нечетной степеней $P_q(t)$ запишем рекуррентными формулами через определенные интегралы

$$\begin{aligned} P_0(t) &= \frac{t}{2a}, \quad P_{2q-1}(t) = \int_0^t P_{2q-2}(t) dt, \\ P_{2q}(t) &= \int_0^t P_{2q-1}(t) dt - \frac{t}{a} \int_0^a P_{2q-1}(t) dt, \quad q = 1, 2, \dots \end{aligned} \quad (7)$$

Полином первой степени $P_0(t)$ является начальным, остальные полиномы получаются по рекуррентным формулам (7). Например, при $p = 6$, что соответствует граничной функции 6-го порядка $M_6(t)$, получим

$$\begin{aligned} P_0(t) &= \frac{t}{2a}, \quad P_1(t) = \frac{t^2}{4a}, \quad P_2(t) = \frac{t^3}{12a} - \frac{at}{12}, \quad P_3(t) = \frac{t^4}{48a} - \frac{at^2}{24}, \\ P_4(t) &= \frac{t^5}{240a} - \frac{at^3}{72} + \frac{7a^3t}{720}, \quad P_5(t) = \frac{t^6}{1440a} - \frac{at^4}{288} + \frac{7a^3t^2}{1440}, \\ P_6(t) &= \frac{t^7}{7 \cdot 1440a} - \frac{at^5}{1440} + \frac{7a^3t^3}{3 \cdot 1440} - \frac{31ta^5}{21 \cdot 1440}. \end{aligned} \quad (8)$$

Поскольку зависимость $f(t)$ предполагается неизвестной, известны только дискретные значения $f(t_j)$, то воспользоваться интегральными формулами Фурье для нахождения a_0, a_m, b_m не представляется возможным. Поэтому для определения коэффициентов a_0, a_m, b_m используем дискретную систему, полученную в виде частичной суммы ряда (4) при $t = t_j$:

$$f(t_j) = M_p(t_j) + a_0 + \sum_{m=1}^{N-1} a_m \cos m\pi \frac{t_j}{a} + b_m \sin m\pi \frac{t_j}{a}, \quad m = 1 \div N-1, \quad j = -N \div N-1. \quad (9)$$

где $2N$ число внутренних точек деления, нанесенных на отрезок $[-a, a]$.

Из базиса (3) с непрерывной переменной $t \in [-a, a]$ построим базис с дискретной переменной $t_j = \frac{ja}{N}$:

$$\left\{ 1, \cos m\pi \frac{t_j}{a}, \sin m\pi \frac{t_j}{a} \right\}, \quad t_j = \frac{ja}{N}, \quad t_j \in [-a, a], \quad j = -N \div N-1. \quad (10)$$

Здесь t_j — координаты интерполяционных точек на отрезке $[-a, a]$. Базис (10) будем называть дискретным. При использовании универсальных быстрых интерполяционных разложений по дискретному базису необходимо выполнить три условия:

- 1) отрезок $[-a, a]$ следует разбить на равные части;
- 2) количество внутренних точек разбиения должно быть четным $2N$.

Нечетное количество внутренних точек разбиения ($2N+1$) строго не допускается, иначе дискретная система (9) не будет ортогональной на отрезке $[-a, a]$, что приведет к увеличению погрешности. В случае, если в системе (9) каждый коэффициент ряда Фурье вычислять с некоторой погрешностью, тогда погрешность будет накапливаться. Если же система (9) ортогональная, то точность существенно увеличивается;

3) в число расчетных точек должна входить только совокупность внутренних точек равномерного разбиения и левая граница отрезка $[-a, a]$, т.е. $t = -a$.

Доказательство дискретной ортогональности базиса (10) приведено в [16]. Также в [16] показано, что система (14) имеет решение относительно a_0, a_m, b_m в конечном виде с высокой точностью:

$$a_0 = \frac{1}{2N} \sum_{j=1}^{N-1} (f(t_j) - M_p(t_j)), \quad (11)$$

$$a_m = \frac{1}{N} \sum_{j=-N}^{N-1} (f(t_j) - M_p(t_j)) \cos m\pi \frac{t_j}{a}, \quad m = 1 \div N-1, \quad (12)$$

$$b_m = \frac{1}{N} \sum_{j=1-N}^{N-1} (f(t_j) - M_p(t_j)) \sin m\pi \frac{t_j}{a}, \quad m = 1 \div N-1. \quad (13)$$

Подставляя a_0 , a_m , b_m из (11), (12), (13) в (9), где возьмем $M_p(t) = M_6(t)$, получим формулу для полной тригонометрической интерполяции $f(t)$ на отрезке $[-a, a]$ в явном виде при задании $f(t)$ дискретными значениями $f(t_j)$ с граничной функцией 6-го порядка:

$$\begin{aligned} f(t) = & \frac{1}{2N} \sum_{j=-N}^{N-1} (f(t_j) - M_6(t_j)) + \frac{1}{N} \sum_{n=1}^{N-1} \sum_{j=-N}^{N-1} (f(t_j) - M_6(t_j)) \cos n\pi \frac{t_j}{a} \cos n\pi \frac{t}{a} + \\ & + \frac{1}{N} \sum_{n=1}^{N-1} \sum_{j=-N}^{N-1} (f(t_j) - M_6(t_j)) \sin n\pi \frac{t_j}{a} \sin n\pi \frac{t}{a} + M_6(t), \quad f(t) \in L_2^{5+2}(t \in [-a, a]). \end{aligned} \quad (14)$$

При $t = t_j$ из (14) будем иметь точное равенство, т.е. интерполяционная кривая (14) точно проходит через экспериментальные точки $t = t_j$. Кроме этого свойства решение допускает еще и дифференцирование заданное число раз, что не рассматривается у других авторов.

В [16] получены оценки для коэффициентов a_m и b_m :

$$|a_m| \leq \frac{2a^7}{m^7 \pi^7} |f^{(7)}(t)| \sim (m\pi)^{-7}, \quad (15)$$

$$|b_m| \leq \frac{2a^8}{m^8 \pi^8} \max |f^{(8)}(t)| \sim (m\pi)^{-8}. \quad (16)$$

Из данных оценок следует, что даже вторые члены полного ряда Фурье будут иметь порядок $|a_2| \leq (2\pi)^{-7}$ и $|b_2| \leq (2\pi)^{-8}$, что является чрезвычайно малой величиной. Таким образом, при использовании универсального быстрого разложения из полного ряда Фурье достаточно учитывать всего два члена — один первый член ряда по синусам и один первый член ряда по косинусам.

Также из оценок (15) и (16) следует, что данный здесь метод по скорости сходимости ряда во много раз превосходит все известные методы и обладает следующими замечательными свойствами.

1. Данный здесь приближенный метод является аналитическим.
2. Интерполяционная кривая (14) с коэффициентами a_0 , a_m , b_m , вычисленными по формулам (11), (12), (13), точно проходит через заданные интерполяционные точки $t = t_j$.
3. Вследствие чрезвычайно высокой скорости сходимости ряда Фурье и высокой точности в ряде достаточно учитывать только несколько первых слагаемых с синусами и с косинусами. Это приводит к весьма экономной расчетной алгебраической системе и позволяет быстро определять координаты пусковой установки.
4. Данный полный ряд Фурье с граничной функцией $M_6(t)$ допускает шестикратное почленное дифференцирование, что является недоступным в других методах и позволяет проводить исследования.
5. Погрешность предлагаемого приближенного метода значительно меньше погрешности входных данных рассматриваемых инженерных задач (силы тяжести, коэффициентов сопротивления атмосферы и т.д.). Поэтому вполне можно принимать подобное приближенное решение в качестве точного на отрезке $t \in [-a, a]$.

2. ОПРЕДЕЛЕНИЕ КООРДИНАТ ПУСКОВОГО МЕХАНИЗМА МЕТОДОМ УНИВЕРСАЛЬНЫХ БЫСТРЫХ ТРИГОНОМЕТРИЧЕСКИХ ИНТЕРПОЛЯЦИЙ

2.1. Постановка задачи

Пусть с поверхности Земли из некоторого пускового механизма запускается ЛА, который может выходить за пределы земной атмосферы, при этом ЛА движется по заранее неизвестной траектории, но координаты его определяются при $t \geq \tilde{t}_1$, где $\tilde{t}_1$ — начало работы регистрирующей установки (РУ). Введем неподвижную $(x = x_1, y = x_2, z = x_3)$ систему координат, жестко связанную с Землей. Начало

системы координат находится в центре Земли. Ось $z = x_3$ неподвижной системы направим по оси вращения Земли к Северному полюсу.

Для удобства вычислений введем вспомогательную временную переменную t . Координаты ЛА в зависимости от t регистрируются с помощью РУ. Момент начала работы РУ будем принимать за значение $t = \tilde{t}_1 = -a$, конец учитываемых точек работы РУ принимается за значение $t = \tilde{t}_2 = a$, весь промежуток работы РУ равен $\tilde{t}_2 - \tilde{t}_1 = 2a$. Переменная t изменяется в пределах $t \in [-a, a]$, в середине отрезка $[\tilde{t}_1, \tilde{t}_2]$ переменная t принимает значение, равное нулю, при $t \in [-a, 0)$ переменная t принимает отрицательные значения $t < 0$, при $t \in (0, a]$ переменная t принимает положительные значения $t > 0$. Математическая переменная t связана с реальным временем $\tilde{t}$ зависимостью (2). Все рассмотрение проведем в наиболее приближенном виде к реальной обстановке.

Пусть масса ЛА вследствие равномерного сгорания топлива уменьшается со временем по линейному закону $m(t) = m_0 (1 - \gamma (t - \tilde{t}_1))$. Начальную массу ЛА m_0 и скорость сгорания топлива γ для данного класса ЛА считаем известными. Если ЛА — это артиллерийский снаряд, движущийся по инерции за счет начальной скорости, то следует считать $\gamma = 0$.

Абсолютное ускорение ЛА складывается из суммы относительного ускорения $(\ddot{x}_1, \ddot{x}_2, \ddot{x}_3)$ и вектора ускорения Кориолиса со следующими проекциями $2\omega_0 \times \mathbf{v}_r = 2(-\omega_0 \dot{x}_2, \omega_0 \dot{x}_1, 0)$.

Запишем уравнения движения центра масс ЛА в системе (x_1, x_2, x_3) в виде трех дифференциальных уравнений

$$\begin{aligned} D_2(t) &= \begin{cases} m(t) (\ddot{x}_1 - 2\omega_0 \dot{x}_2 + G(t)x_1) + F_f(t)\dot{x}_1 - H_1 = 0, \\ m(t) (\ddot{x}_2 + 2\omega_0 \dot{x}_1 + G(t)x_2) + F_f(t)\dot{x}_2 - H_2 = 0, \\ m(t) (\ddot{x}_3 + G(t)x_3) + F_f(t)\dot{x}_3 - H_3 = 0, \end{cases} \\ F_f(t) &= \rho \left(\beta_1 + \beta_2 \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2} + \beta_3 (\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) \right), \\ G(t) &= \frac{gR_3^2}{(x_1^2 + x_2^2 + x_3^2)^{3/2}}, \quad m(t) = m_0 (1 - \gamma (t - \tilde{t}_1)). \end{aligned} \quad (17)$$

Здесь величина $G(t)$ учитывает переменное притяжение к Земле за счет изменения расстояния до центра Земли, g — ускорение Земного притяжения. Индекс 2 внизу оператора $D_2(t)$ означает, что каждое из трех уравнений — второго порядка. В системе (17) в (H_1, H_2, H_3) может включаться реактивная сила, а также некоторые внешние силы. В записи системы (17) учтено также ускорение Кориолиса, переменное притяжение к Земле, атмосферное сопротивление с учетом до третьей степени скорости ЛА, так как нелинейный коэффициент силы атмосферного сопротивления F_f умножается на вектор скорости $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$, $\beta_1, \beta_2, \beta_3$ — коэффициенты сопротивления, γ — коэффициент сгорания топлива, m_0 — масса ЛА в момент начала работы РУ при $t = \tilde{t}_1$. В системе (17) с таким же успехом можно было бы учесть силы притяжения к Солнцу, Луне и т.д. Усложнения были бы не принципиальными. Для атмосферной плотности ρ примем экспоненциальную модель [17]

$$\rho = \rho_0 \cdot e^{-(x_3(t) - R_3)/7170}, \quad (18)$$

где ρ_0 — атмосферная плотность на поверхности Земли.

К системе (17) следует добавить начальные условия, которые записываются обычно заданием начального положения и начальной скорости

$$x_1(T_0) = 0, \quad x_2(T_0) = 0, \quad x_3(T_0) = x_{3,0} = R_3 + h_0, \quad (19)$$

$$\dot{x}_k(T_0) = \dot{x}_{k,0}, \quad k = 1, 2, 3, \quad (20)$$

где T_0 — неизвестный момент запуска ЛА, h_0 — высота над поверхностью Земли, откуда стартует ЛА, которую считаем известной для данного класса ЛА.

2.2. Решение задачи

Будем полагать, что начальные данные ЛА (19) и (20) хотя и существуют, но заранее момент T_0 запуска ЛА, а также ее начальные координаты (19) и начальная скорость (20) неизвестны, их надо определить. Известно только множество значений координат ЛА, полученные из измерительных данных:

$$t_{-N} = \tilde{t}_1, \quad t_N = \tilde{t}_2, \quad x_k(t_j) = x_j^{(k)}, \quad t_j = j \frac{a}{N}, \quad j = -N \div N - 1, \quad k = 1, 2, 3. \quad (21)$$

Из (21) при $j = -N$, $t_j = t_{-N} = -a$ получаем первую известную начальную точку из системы данных (21) для промежутка $t \in [-a, a]$

$$x_k(-a) = x_{-N}^{(k)}, \quad x_k(a) = x_N^{(k)}. \quad (22)$$

Здесь $(x_{-N}^1, x_{-N}^2, x_{-N}^3)$ — первая точка из данной системы точек. Значения проекций начальной скорости при $j = -N$

$$\dot{x}_1(t_{-N}) = \dot{x}_{1,0}, \quad \dot{x}_2(t_{-N}) = \dot{x}_{2,0}, \quad \dot{x}_3(t_{-N}) = \dot{x}_{3,0} \quad (23)$$

полагаем заранее неизвестными. Решение задачи в данной постановке никаким известным методом невозможно, кроме предлагаемого метода быстрых разложений [2, 6]. Левые части трех уравнений (17) примем за некоторые три функции от переменной $t \in [-a, a]$, к каждой из которых применим быстрые разложения с граничной функцией шестого порядка $M_6(t)$ из (5) по аналогии с разложением (4) с использованием частичных сумм, т.е.

$$\begin{aligned} x_k(t) &= M_6^k(t) + a_0^k + \sum_{m=1}^{N-1} a_m^k \cos m\pi \frac{t}{a} + b_m^k \sin m\pi \frac{t}{a}, \quad k = 1, 2, 3, \quad t \in [-a, a], \\ M_6^k(t) &= A_0^k \frac{t}{2a} + A_1^k \frac{t^2}{4a} + A_2^k \left(\frac{t^3}{12a} - \frac{at}{12} \right) + A_3^k \left(\frac{t^4}{48a} - \frac{at^2}{24} \right) + \\ &+ A_4^k \left(\frac{t^5}{240a} - \frac{at^3}{72} + \frac{7a^3t}{720} \right) + A_5^k \left(\frac{t^6}{1440a} - \frac{at^4}{288} + \frac{7a^3t^2}{1440} \right) + \\ &+ A_6^k \left(\frac{t^7}{7 \cdot 1440a} - \frac{at^5}{1440} + \frac{7a^3t^3}{3 \cdot 1440} - \frac{31a^5t}{21 \cdot 1440} \right). \end{aligned} \quad (24)$$

Коэффициенты Фурье в (24) вычисляются по формулам (11), (12), (13):

$$\begin{aligned} a_0^k &= \frac{1}{2N} \sum_{j=-N}^{N-1} (x_k(t_j) - M_6^k(t_j)), \quad a_m^k = \frac{1}{N} \sum_{j=-N}^{N-1} (x_k(t_j) - M_6^k(t_j)) \cos m\pi \frac{t_j}{a}, \\ b_m^k &= \frac{1}{N} \sum_{j=-N}^{N-1} (x_k(t_j) - M_6^k(t_j)) \sin m\pi \frac{t_j}{a}, \quad k = 1, 2, 3, \quad m = 1 \div N-1. \end{aligned} \quad (25)$$

В силу свойств рядов Фурье быстрое разложение (24) справедливо только на замкнутом отрезке $t \in [-a, a]$, включая и его концы, вне отрезка решение будем строить при помощи экстраполяций.

Вначале определим коэффициенты A_0^k по формуле

$$A_0^k = x_k(a) - x_k(-a), \quad x_k(-a) = x_{-N}^{(k)}, \quad x_k(a) = x_N^{(k)}, \quad k = 1, 2, 3. \quad (26)$$

Поскольку $x_k(a)$ и $x_k(-a)$ известны из данных (21), то и A_0^k определяется из (26). Остальные $A_1^k \div A_6^k$ будем находить из системы 18-ти уравнений. Для нахождения A_1^k в первом уравнении (24) положим $t = a$, что пока отдельно нигде не использовалось, и потому такое уравнение является независимым от всех ранее записанных уравнений:

$$x_k(t=a) = x_N = \frac{1}{2}A_0^k + \frac{a}{4}A_1^k - \frac{a^3}{48}A_3^k + \frac{a^5}{480}A_5^k + a_0^k + \sum_{m=1}^{N-1} a_m^k (-1)^m, \quad k = 1, 2, 3. \quad (27)$$

Уравнение (27) содержит неизвестные a_0^k, a_m^k , исключить которые можно с помощью формул (25). Для нахождения A_2^k в системе (17) вначале положим $t = a$, затем $t = -a$ и составим разностные уравнения

$$D_2(t=a) - D_2(t=-a) = 0. \quad (28)$$

После дифференцирования левой и правой частей уравнений (17) последовательно один, два, три, четыре раза [6], что допустимо в силу свойств быстрых разложений и, положив $t = a$ и $t = -a$ в каждом из них, составим еще четыре разностных уравнения

$$\begin{aligned} D_2'(t=a) - D_2'(t=-a) &= 0, & D_2''(t=a) - D_2''(t=-a) &= 0, \\ D_2'''(t=a) - D_2'''(t=-a) &= 0, & D_2^{(4)}(t=a) - D_2^{(4)}(t=-a) &= 0. \end{aligned} \quad (29)$$

Заметим, что уравнения (28) и (29) записаны в символическом виде, но представить их явно при помощи систем компьютерной алгебры не представляет особых трудностей.

Из уравнений (28) и (29), также как и из уравнения (27) исключены коэффициенты a_0^k, a_m^k, b_m^k с помощью формул (25). Система 18-ти алгебраических уравнений (27)–(29) относительно $A_1^k \div A_6^k$ является замкнутой. После подстановки найденных $A_1^k \div A_6^k$ в (25) определим a_0^k, a_m^k, b_m^k . Затем подставив найденные $A_1^k \div A_6^k$ и a_0^k, a_m^k, b_m^k в выражение (24) для $x_k(t)$, будем иметь аналитическое решение для координат ЛА (или артиллерийского снаряда), справедливое на замкнутом временном отрезке $t \in [-a, a]$. Таким образом, универсальная быстрая тригонометрическая интерполяция позволяет построить траекторию полета ЛА на отрезке $t \in [-a, a]$. Для определения координат пускового механизма необходимо получить траекторию полета ЛА при $t \leq -a$. Поэтому предложенный выше метод интерполяций можно использовать для организации экстраполяций при $t \leq -a$, где интерполяционное разложение (24) по построению уже неприменимо. Однако из (24) можно вычислить производные $x_k^p(t)$, $k = 1, 2, 3$, $p = 0 \div 6$, до шестого порядка включительно при $t = -a$, с помощью которых организуем ряд Тейлора, являющийся искомой экстраполяцией справедливой при $t \leq -a$, а также и при $t \geq -a$:

$$\begin{aligned} x_k(t) &= \sum_{p=0}^6 x_k^{(p)}(-a) \frac{(t+a)^p}{p!}, \\ x_k'(-a) &= A_0^k \frac{1}{2a} - A_1^k \frac{1}{2} + A_2^k \frac{a}{6} - A_4^k \frac{a^3}{90} + A_6^k \frac{a^5}{945} + \sum_{m=1}^{\infty} m \frac{\pi}{a} (-1)^m, \\ x_k''(-a) &= A_1^k \frac{1}{2a} - A_2^k \frac{1}{2} + A_3^k \frac{a}{6} - A_5^k \frac{a^3}{90} - \sum_{m=1}^{\infty} \left(m \frac{\pi}{a}\right)^2 (-1)^m a_m^k, \\ x_k'''(-a) &= A_2^k \frac{1}{2a} - A_3^k \frac{1}{2} + A_4^k \frac{a}{6} - A_6^k \frac{a^3}{90} - \sum_{m=1}^{\infty} \left(m \frac{\pi}{a}\right)^3 (-1)^m b_m^k, \\ x_k^{(4)}(-a) &= A_3^k \frac{1}{2a} - A_4^k \frac{1}{2} + A_5^k \frac{a}{6} + \sum_{m=1}^{\infty} \left(m \frac{\pi}{a}\right)^4 a_m^k (-1)^m, \\ x_k^{(5)}(-a) &= A_4^k \frac{1}{2a} - A_5^k \frac{1}{2} + A_6^k \frac{a}{6} + \sum_{m=1}^{\infty} \left(m \frac{\pi}{a}\right)^5 b_m^k (-1)^m, \\ x_k^{(6)}(-a) &= A_5^k \frac{1}{2a} - A_6^k \frac{1}{2} - \sum_{m=1}^{\infty} \left(m \frac{\pi}{a}\right)^6 a_m^k (-1)^m. \end{aligned} \quad (30)$$

Радиус сходимости ряда Тейлора неизвестен. Поэтому в примере, демонстрирующем применение метода быстрых универсальных тригонометрических интерполяций, для определения координат пускового механизма будет произведено сравнение значений $x_k(t)$ при $t \geq -a$, вычисленных по формулам быстрых разложений (24) и рядом Тейлора (30), с целью экспериментального установления радиуса сходимости ряда Тейлора (30).

АЛГОРИТМ ПРИМЕНЕНИЯ МЕТОДА БЫСТРЫХ УНИВЕРСАЛЬНЫХ ТРИГОНОМЕТРИЧЕСКИХ ИНТЕРПОЛЯЦИЙ ДЛЯ ОБНАРУЖЕНИЯ КООРДИНАТ ПУСКОВОГО МЕХАНИЗМА

1. Расположить РУ в удобном месте для регистрации координат ЛА.
2. Включить передатчики П с записями ложных телефонных сигналов и включить РУ для определения координат ЛА.
3. Определить моменты времени начала $\tilde{t}_1$ и конца $\tilde{t}_2$ работы РУ при регистрации координат ЛА.
4. По формуле (1) вычислить величину a и определить количество N .
5. Данные РУ представить в виде (21).
6. По формуле (26) найти коэффициенты A_0^k .
7. По формулам (25) выразить коэффициенты a_0^k, b_m^k, b_m^k через коэффициенты $A_1^k \div A_6^k$ граничной функции.
8. Записать замкнутую нелинейную алгебраическую систему уравнений (27)–(29) для определения $A_1^k \div A_6^k$.
9. Найденные a_0^k, b_m^k, b_m^k и $A_1^k \div A_6^k$ подставить в (24) и получить искомую быструю интерполяцию для координат $x_k(t)$, справедливую на замкнутом отрезке $t \in [-a, a]$ и допускающую дифференцирование на всем отрезке и на его концах 6 раз.
10. Вычислить значения $x_k(t)$ и их первых шести производных в точке $t = -a$ и составить для $x_k(t)$ ряд Тейлора в виде (30).
11. Из условия $x_3(T_0) = R_3$ определить значение T_0 — время запуска ЛА и вычислить в точке $t = T_0$ координаты $x_k(t)$ из (30). Тем самым получим координаты пускового механизма.

3. ПРИМЕР ОПРЕДЕЛЕНИЯ КООРДИНАТ ПУСКОВОГО МЕХАНИЗМА

Пусть в качестве тестового примера в начальные моменты времени запуска ЛА движется по винтовой линии

$$x_1^* = R_0 \cos \omega_1 t + \delta_1, \quad x_2^* = R_0 \sin \omega_1 t + \delta_2, \quad x_3^* = R_3 + K + w(t + a) + \delta_3, \quad (31)$$

где R_0, K, ω_1, w — некоторые заданные постоянные, $\delta_1, \delta_2, \delta_3$ — постоянные, определяющие ошибки РУ нахождения координат траектории ЛА из-за воздействия шумов. И пусть функции $H_1 \div H_3, F_f, G$ в правых частях системы (17) заданы выражениями

$$H_1 = m_0 (1 - \gamma(t - \tilde{t}_1)) (-R_0 \omega_1^2 \cos \omega_1 t - 2\omega_0 R_0 \omega_1 \cos \omega_1 t + G(R_0 \cos \omega_1 t + \delta_1)) - F_f R_0 \omega_1 \sin \omega_1 t,$$

$$H_2 = m_0 (1 - \gamma(t - \tilde{t}_1)) (-R_0 \omega_1^2 \sin \omega_1 t - 2\omega_0 R_0 \omega_1 \sin \omega_1 t + G(R_0 \cos \omega_1 t + \delta_2)) + F_f R_0 \omega_1 \cos \omega_1 t,$$

$$H_3 = m_0 (1 - \gamma(t - \tilde{t}_1)) G(R_3 + K + w(t + a) + \delta_3) + F_f w,$$

$$F_f = \rho_0 \cdot e^{-(K+w(t+a)+\delta_3)/7170} \left(\beta_1 + \beta_2 \sqrt{R_0^2 \omega_1^2 + w^2} + \beta_3 (R_0^2 \omega_1^2 + w^2) \right),$$

$$G = \frac{gR_3^2}{\left((R_0 \cos \omega_1 t + \delta_1)^2 + (R_0 \sin \omega_1 t + \delta_2)^2 + (R_3 + K + w(t + a) + \delta_3)^2 \right)^{3/2}}.$$

Граничные условия для точного решения (31) имеют вид

$$x_1^*(-a) = R_0 \cos \omega_1 a + \delta_1, \quad x_2^*(-a) = -R_0 \sin \omega_1 a + \delta_2, \quad x_3^*(-a) = R_3 + K + \delta_3.$$

Тестовая траектория (31) здесь выбрана одной из простейших, типа винтовой линии, что не является принципиальным. Подобные траектории используются часто, чтобы затруднить определение положения пускового механизма. При использовании метода быстрых разложений с таким же успехом могла бы быть использована и любая другая тестовая траектория ЛА, если она будет предложена.

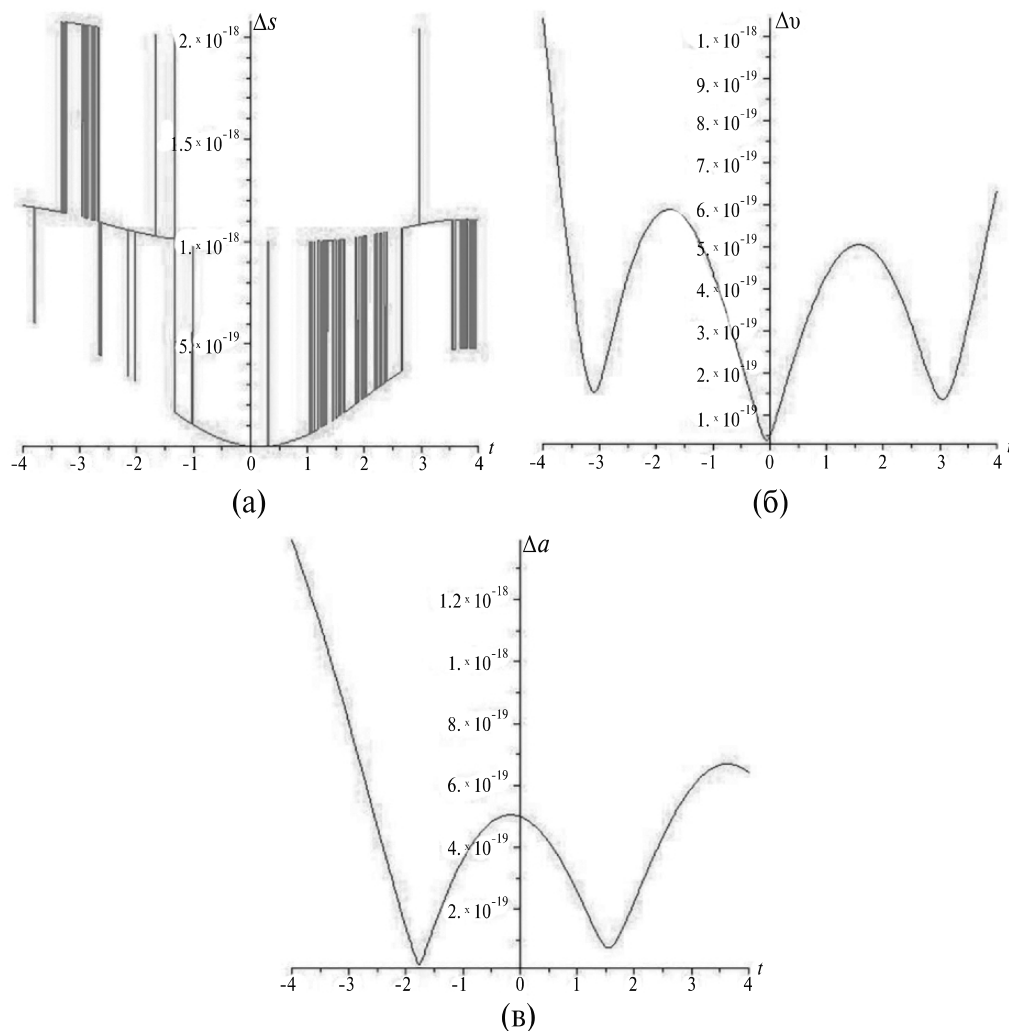
Имея тестовое точное решение (31), можно путем сравнения с приближенным аналитическим решением (24) вычислить абсолютную погрешность в определении положения ЛА, его скорости и ускорения.

Значения параметров выберем следующими: $m_0 = 80$ кг, $\omega_1 = 2\pi/7200$ с⁻¹, $w = 5$ м/с, $R_0 = 100$ м, $K = 50$ м, $g = 9.80665$ м/с², $R_3 = 6372000$ м, $\omega_0 = 1.995 \cdot 10^{-7}$ с⁻¹, $\rho = 1.2754$ кг/м³, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $\gamma = 1/80$, $\delta_1 = \delta_2 = \delta_3 = 0$. Значения указанных величин задавались случайным образом. Все расчеты выполнены в системе компьютерной алгебры Maple.

Абсолютные погрешности траектории ЛА Δs , его скорости Δv и ускорения Δa вычислялись по формулам

$$\begin{aligned}\Delta s &= \sqrt{(x_1^* - x_1)^2 + (x_2^* - x_2)^2 + (x_3^* - x_3)^2}, \\ \Delta v &= \sqrt{(\dot{x}_1^* - \dot{x}_1)^2 + (\dot{x}_2^* - \dot{x}_2)^2 + (\dot{x}_3^* - \dot{x}_3)^2}, \\ \Delta a &= \sqrt{(\ddot{x}_1^* - \ddot{x}_1)^2 + (\ddot{x}_2^* - \ddot{x}_2)^2 + (\ddot{x}_3^* - \ddot{x}_3)^2}.\end{aligned}\quad (32)$$

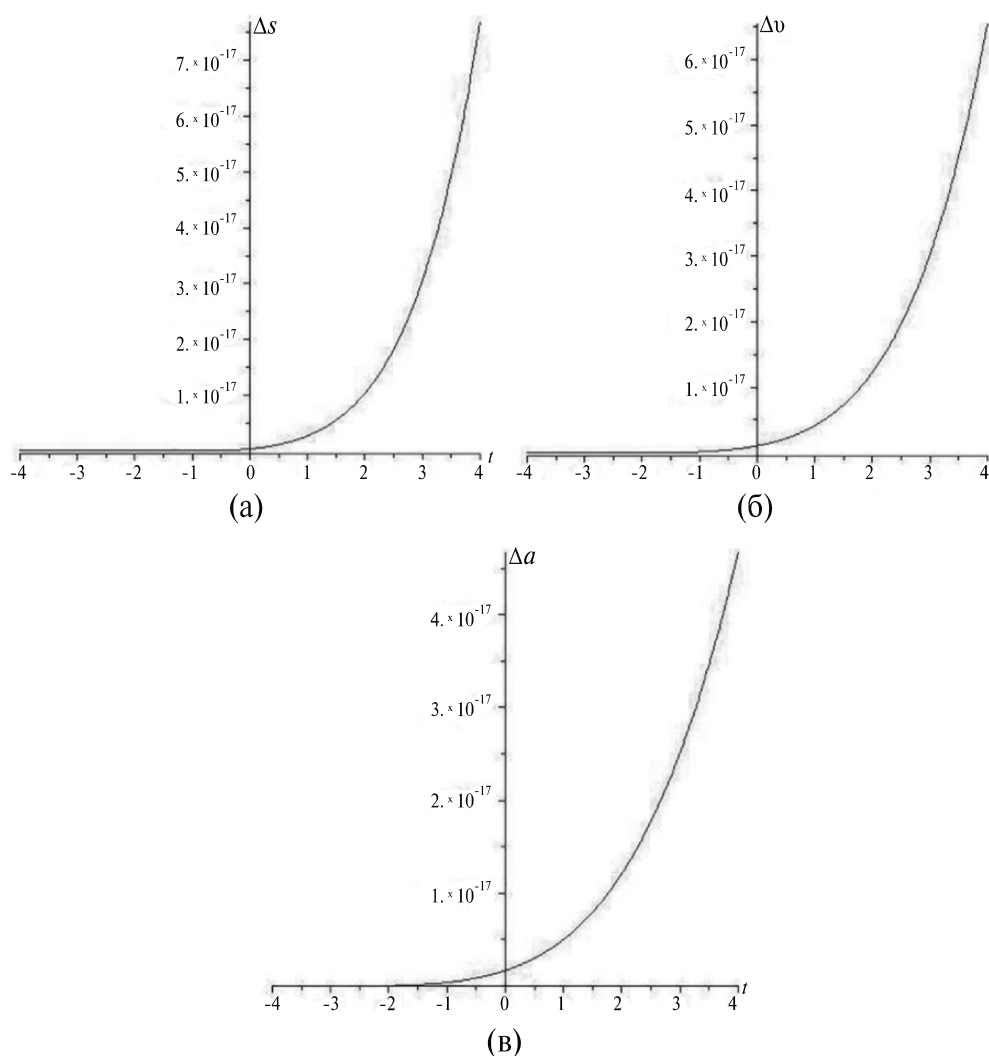
На фиг. 1 показана абсолютная погрешность приближенного аналитического решения (24) по сравнению с точным (31) при значениях $N = 2$ и времени $a = 4$. Из фиг. 1 видно, что порядок погрешностей Δs равен 10^{-18} м, $\Delta v = 10^{-18}$ м/с, а для Δa составляет 10^{-18} м/с². Поэтому, приближенное аналитическое решение (24), справедливое на замкнутом отрезке $t \in [-a, a]$, можно принять за точное.



Фиг. 1. Абсолютная погрешность решения (24) по сравнению с (31): (а) — траектории полета ЛА, (б) — скорости ЛА, (в) — ускорения ЛА.

На фиг. 2 представлена абсолютная погрешность ряда Тейлора (30) по сравнению с аналитическим решением (24). Данная погрешность также вычислялась по формулам (32), где за x_1^* , x_2^* , x_3^* принима-

лось решение (24). Порядки погрешностей Δs , Δv и Δa , которые соответственно составляют 10^{-17} м, 10^{-17} м/с, 10^{-17} м/с² позволяют сделать вывод о том, что радиус сходимости ряда Тейлора (30) больше a . Следовательно, ряд Тейлора (30) можно использовать для экстраполяции траектории полета ЛА на отрезок $t \in [T_0, -a]$ с целью определения координат пускового механизма установки.



Фиг. 2. Абсолютная погрешность ряда Тейлора (30) по сравнению с решением (24): (а) — траектории полета ЛА, (б) — скорости ЛА, (в) — ускорения ЛА.

Также, как и точность определения координат пускового механизма, важным параметром является время обнаружения этого механизма. Уменьшить значение этого критерия можно, сократив время поиска координат пускового механизма. Так, например, время расчета программы по обнаружению пускового механизма, выполненном на персональном компьютере с процессором Intel Core i3-4160 и ОЗУ 8 ГБ при значениях $N = 2$ и времени $a = 4$ составляет 6.5 с. Если же при $N = 2$ время a увеличилось до $a = 8$, то время расчета программы будет 7.3 с. Следовательно, на определение координат пускового механизма можно отводить не более 8 с.

Представляет интерес исследование устойчивости решения. Максимальные значения абсолютных погрешностей Δs , Δv и Δa при различных значениях коэффициентов сопротивления β_1 , β_2 , β_3 приведены в табл. 1, из которой следует, что при малых возмущениях исходных данных ($\beta_1, \beta_2, \beta_3$) абсолютная погрешность изменяется незначительно, т.е. решение устойчиво по отношению к возмущениям коэффициентов сопротивления.

Таблица 1. Устойчивость решения по отношению к возмущениям коэффициентов сопротивления

$\beta_1 = \beta_2 = \beta_3$	$\Delta S_{\max}, \text{м}$	$\Delta v_{\max}, \text{м/с}$	$\Delta a_{\max}, \text{м/с}^2$
0.99	$2.031 \cdot 10^{-18}$	$1.039 \cdot 10^{-18}$	$1.383 \cdot 10^{-18}$
1.00	$2.052 \cdot 10^{-18}$	$1.042 \cdot 10^{-18}$	$1.392 \cdot 10^{-18}$
1.01	$2.151 \cdot 10^{-18}$	$1.044 \cdot 10^{-18}$	$1.400 \cdot 10^{-18}$

В табл. 2 приведены данные расчетов, показывающих влияние коэффициентов сопротивления $\beta_1, \beta_2, \beta_3$ на точность результатов. В обоих случаях полученные абсолютные погрешности ничтожно малы.

Таблица 2. Влияние коэффициентов сопротивления на точность результатов

$\beta_1 = \beta_2 = \beta_3$	$\Delta S_{\max}, \text{м}$	$\Delta v_{\max}, \text{м/с}$	$\Delta a_{\max}, \text{м/с}^2$
0.0	$2.000 \cdot 10^{-18}$	$8.059 \cdot 10^{-19}$	$8.502 \cdot 10^{-19}$
1.0	$2.052 \cdot 10^{-18}$	$1.042 \cdot 10^{-18}$	$1.392 \cdot 10^{-18}$
1.2	$2.061 \cdot 10^{-18}$	$1.112 \cdot 10^{-18}$	$1.557 \cdot 10^{-18}$

Предположим, что РУ выдает координаты траектории ЛА с некоторыми ошибками $\delta_1, \delta_2, \delta_3$ из-за воздействия шумов. Максимальные значения абсолютных погрешностей $\Delta s, \Delta v$ и Δa при различных величинах ошибок $\delta_1, \delta_2, \delta_3$ приведены в табл. 3, из которой видно, что ошибки измерений незначительно влияют на точность вычислений, т.е. порядок абсолютных погрешностей $\Delta s, \Delta v$ и Δa ничтожно малый.

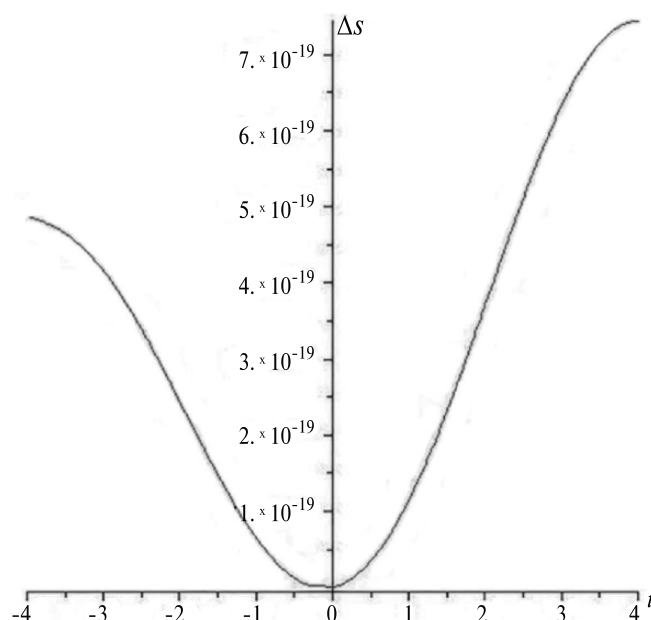
Таблица 3. Влияние ошибок измерений на точность расчетов

$\delta_1 = \delta_2 = \delta_3, \text{м}$	$\Delta S_{\max}, \text{м}$	$\Delta v_{\max}, \text{м/с}$	$\Delta a_{\max}, \text{м/с}^2$
1.0	$3.043 \cdot 10^{-18}$	$1.040 \cdot 10^{-18}$	$1.391 \cdot 10^{-18}$
2.0	$3.394 \cdot 10^{-18}$	$1.048 \cdot 10^{-18}$	$1.397 \cdot 10^{-18}$
5.0	$3.402 \cdot 10^{-18}$	$1.050 \cdot 10^{-18}$	$1.399 \cdot 10^{-18}$

Для анализа влияния погрешности в определении массы ЛА на точность определения траектории его полета проведено сравнение траекторий ЛА с различными массами. Одно значение массы ЛА примем за эталонное и будем считать его заложенным в программу. Другую массу ЛА считаем реальной массой ЛА, траекторию которого нужно определить. Расчеты проведены с использованием зашумленности данных ($\delta_1 = \delta_2 = \delta_3 = 1 \text{ м}$). Пусть в программу внесена масса ЛА 80 кг. Допустим, что реальная масса ЛА 100 кг. Абсолютная погрешность траектории ΔS движения ЛА массой 100 кг по сравнению с движением ЛА с массой 80 кг представлена на фиг. 3. Как видно из фиг. 3 точность определения траектории ЛА остается очень высокая даже при не точном определении массы ЛА на 20%.

ЗАКЛЮЧЕНИЕ

Методом универсальных быстрых тригонометрических интерполяций получено приближенное аналитическое решение задачи о быстром определении координат пускового механизма. Учитывая в рядах Фурье универсальных быстрых разложений только по одному коэффициенту у косинусов и у синусов, абсолютная погрешность расчетов для траектории ЛА составила 10^{-18} м . Абсолютная погрешность определения скорости и ускорения ЛА составляет 10^{-18} м/с и 10^{-18} м/с^2 соответственно. В связи с этим, приближенное аналитическое решение (24) можно считать точным решением. Время, затрачиваемое на определение координат пускового механизма, составляет не более 8 сек.



Фиг. 3. Абсолютная погрешность траектории ΔS движения ЛА с массой 100 кг по сравнению с движением ЛА массой 80 кг.

Подобная высокая точность и, одновременно с этим, высокая скорость нахождения координат пускового механизма позволят более эффективно проводить его поиск.

СПИСОК ЛИТЕРАТУРЫ

1. Толстов Г. П. Ряды Фурье. М.: Наука, 1980. 384 с.
2. Чернышов А. Д. Метод быстрых разложений для решения нелинейных дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 2024. Т. 54. № 1. С. 13–24.
3. Бейкер Дж., Грейвс-Моррис П. Аппроксимации Паде. М.: Мир, 1986. 502 с.
4. Нерсисян А., Погосян А., Бархударян Р. Ускоренная сходимость рядов Фурье // Известия НАН Армении. Математика. 2006. Т. 41. № 2. С. 43–56.
5. Балакирев Ю. Г. Об улучшении сходимости модальных разложений в задачах синтеза динамических характеристик составных конструкций // Изв. РАН. МТТ. 2002. № 6. С. 158–170.
6. Chernyshov A. D., Saiko D. S., Kovaleva E. N. Universal fast expansion for solving nonlinear problems // Journal of Physics: Conference Series. 2020. С. 012147.
7. Чернышов А. Д., Горайнов В. В., Лешонков О. В., Соболева Е. А., Никифорова О. Ю. Сравнение скорости сходимости быстрых разложений с разложениями в классический ряд Фурье // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2019. № 1. С. 27–34.
8. Самохин А. Б., Тыртышников Е. Е. Численный метод решения объемных интегральных уравнений на неравномерной сетке // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 5. С. 878–884.
9. Исаев В. И., Шапеев В. П. Варианты метода коллокаций и наименьших квадратов повышенной точности для численного решения уравнений Навье—Стокса // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 10. С. 1758–1770.

10. Беляев В. А., Брындин Л. С., Голушко С. К., Семисалов Б. В., Шанеев В. П. Н-, Р- и НР-варианты метода коллокации и наименьших квадратов для решения краевых задач для бигармонического уравнения в нерегулярных областях и их приложения // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 4. С. 531–552.
11. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М. Наука, 1987. 800 с.
12. Лешонков О. В., Соболева Е. А., Чернышов А. Д. Метод быстрых разложений для вычисления определенных интегралов с переменным верхним пределом от сложных или неявно заданных функций // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 6. С. 926–935.
13. Чернышов А. Д., Горяйнов В. В., Чернышов О. А. Применение метода быстрых разложений для расчета траекторий космических кораблей // Известия вузов. Авиационная техника. 2015. № 2. С. 41–47.
14. Chernyshov A. D., Goryainov V. V., Danshin A. A. Analysis of the stress field in a wedge using the fast expansions with pointwise determined coefficients // IOP Conf. Series: Journal of Physics: Conf. Series. 973. 2018. 012002.
15. Чернышов А. Д., Горяйнов В. В., Попов М. И. Исследование погрешности быстрой тригонометрической интерполяции при решении задачи о напряжениях в бруске // Известия Российской академии наук. Механика твердого тела. 2023. № 1. С. 115–128.
16. Чернышов А. Д., Никифорова О. Ю., Горяйнов В. В., Кузнецов С. Ф., Рукин И. Г. Универсальные быстрые тригонометрические интерполяции для интегро-дифференциальных задач различного порядка // Вестник Чувашиского государственного педагогического университета им. И. Я. Яковлева. Серия: Механика предельного состояния. 2022. № 4 (54). С. 57–70.
17. Дубровина А. А. Применение методов условной многомерной минимизации к задаче расчета траектории баллистической ракеты // Вестник Концерна ВКО «Алмаз – Антей». 2017. № 3. С. 59–62.

THE USE OF UNIVERSAL FAST TRIGONOMETRIC INTERPOLATION AND EXTRAPOLATION TO DETERMINE THE COORDINATES OF THE LAUNCH POINT OF THE AIRCRAFT

A. D. Chernyshov^{a,*}, O. Yu. Nikiforova^{a,**}, V. V. Goryainov^{b,***}, I. G. Rukin^{a,****}

^a 394000 Voronezh, Revolution Ave., 19, Voronezh State Engineering Technologies University, Russia

^b 394006 Voronezh, 20th Anniversary of October Str. 84, Voronezh State Technical University, Russia

*e-mail: chernyshovad@mail.ru

**e-mail: niki22@mail.ru

***e-mail: gorvit77@mail.ru

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

Received 30 October, 2023

Revised 28 February, 2024

Accepted 5 April, 2024

Abstract. The basics of fast trigonometric interpolation are given for nonperiodic functions, making it possible to obtain an approximate solution with high accuracy. Formulas are obtained to calculate the coordinates of the launch point of an aircraft with high accuracy due to the simultaneous application of the method for universal fast trigonometric interpolations and the method for extrapolations at the ends of a certain specified segment. Numerical experiments show that after the launch of the first aircraft the coordinates of the launching mechanism can be determined in 7 seconds with an accuracy up to 10^{-17} m.

Keywords: fast decompositions, fast trigonometric interpolation, extrapolation, high accuracy, aircraft, coordinates, launch point.

ДУАЛИЗМ В ТЕОРИИ СОЛИТОННЫХ РЕШЕНИЙ¹⁾

© 2024 г. Л. А. Бекларян^{1,*}, А. Л. Бекларян^{2,**}

¹117418 Москва, Нахимовский пр-т, 47, Центральный Экономико-Математический Институт РАН, Россия

²101000 Москва, ул. Мясницкая, 20, Национальный исследовательский университет “Высшая школа экономики”, Россия

*e-mail: lbeklaryan@outlook.com

**e-mail: abeklaryan@hse.ru

Поступила в редакцию 12.09.2023 г.

Переработанный вариант 23.02.2024 г.

Принята к публикации 05.03.2024 г.

Статья посвящена дуализму теорий солитонных решений и решений функционально-дифференциальных уравнений точечного типа. Представлены основы формализма такого дуализма, центральным элементом которого является понятие солитонного букета, а также дуальная пара “функция-оператор”. В рамках такого подхода удастся описать все пространство солитонных решений с заданной характеристикой, а также их асимптотику как по пространству, так и по времени. На примере модели транспортного потока на манхэттенской решетке описано все семейство ограниченных солитонных решений. Библ. 28. Фиг. 2.

Ключевые слова: манхэттенская решетка, солитонные решения, функционально-дифференциальное уравнение.

DOI: 10.31857/S0044466924070072, EDN: xiovu0

ВВЕДЕНИЕ

Изучение солитонных решений (решений типа бегущей волны), является одним из активно развивающихся направлений как в разработке теории, так и в приложениях. Центральный вопрос касается как существования таких решений, а так и их многообразия. Имеется большое число работ, посвященных этой тематике. Приведем некоторые из них, как в виде статей, монографий, так и в виде обзора [1–4]. Методы исследования в данной проблематике также многообразны. Одним из универсальных методов исследования таких систем является конструктивное построение решений, использующее явный вид правой части, а также наличие различных симметрий, либо аналитических свойств в виде периодичности, бесконечной дифференцируемости, аналитичности. Далее, методами теории возмущений солитонные решения устанавливаются и для близких правых частей. Обзор по работам такого направления для бесконечномерных (однородных) систем с потенциалами Френкеля—Конторовой и Ферми—Паста—Улама приведен в работе [4]. Следует отметить, что такой подход не позволяет описать пространство всех солитонных решений, а также асимптотику их возможного роста.

В работе представлены некоторые фрагменты формализма теории солитонных решений для бесконечномерных систем, разрабатываемые авторами. В частности, к таким системам относятся конечно-разностные аналоги уравнения математической физики. В рамках такого формализма солитонные решения с заданной характеристикой находятся во взаимно однозначном соответствии с решениями индуцированного функционально-дифференциального уравнения точечного типа с параметром в виде характеристики семейства солитонных решений. На этом пути удастся описать все пространство солитонных решений с характеристиками в выделенных диапазонах. Описывается возможный рост таких решений, где рост также определяется характеристикой солитонного решения.

¹⁾Работа выполнена при финансовой поддержке РНФ (проект № 23-11-00080).

В разрабатываемом формализме правая часть (оператор) бесконечномерного дифференциального уравнения и правая часть индуцированного функционально-дифференциального уравнения образуют дуальную пару “функция-оператор”. Подобный дуализм оказывается весьма полезным. В ряде случаев оказывается возможным использование свойств конечномерности фазового пространства решений индуцированного функционально-дифференциального уравнения. В иных случаях, для солитонных решений удастся воспользоваться спектральными свойствами правой части бесконечномерного дифференциального уравнения. Для дуальной пары “функция-оператор”, возникает естественный вопрос о существовании дуальной пары с таким же индуцированным функционально-дифференциальным уравнением, но более простым оператором. Такой вопрос получает свое положительное решение.

Подход в изучении солитонных решений с заданной характеристикой для бесконечномерной динамической системы, основанный на *соответствии* таких решений решениям индуцированного функционально-дифференциального уравнения использовался в ряде работ [5, 6, 8–11, 20–24].

За пределами обсуждения мы оставляем такую важную проблему в самом общем виде, как разрешимость функционально-дифференциальных уравнений. По этому вопросу можно обратиться к работам [11, 25–28] и ряду других работ.

Рассматривается пример в виде модели движения в манхэттенской решетке. В такой модели с большим числом узлов решетки полагаем, что она является полной по всем направлениям и однородной. Ранее, в работе [17] рассматривалась агрегированная модель манхэттенской решетки, в которой в каждом узле решетки имеется накопитель большой емкости для проведения маневра “разворот”, а общее число транспортного потока представлено без детализации по направлениям движения. В данной работе рассматривается агрегированная модель с ограниченной емкостью накопителей.

Установившиеся потоки в такой решетке описываются солитонными решениями для бесконечномерного обыкновенного дифференциального уравнения в форме конечно-разностного аналога параболического уравнения. Параметром такого семейства уравнений выступает характеристика солитонного решения. В рамках предложенного формализма удастся описать все пространство солитонных решений с характеристиками в выделенных диапазонах, а также получено описание всего семейства ограниченных солитонных решений.

1. СОЛИТОННЫЕ БУКЕТЫ И ДУАЛЬНЫЕ ПАРЫ “ФУНКЦИЯ-ОПЕРАТОР”

Для определения солитонного букета приведем хорошо известное определение условий Каратеодори.

Определение 1. Функция $f : \mathbb{B}_R \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$, где $\mathbb{B}_R$ — интервал на числовой прямой, удовлетворяет условиям Каратеодори, если:

- отображение $f : \mathbb{B}_R \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ определено при всех $z_1, \dots, z_s \in \mathbb{R}^n$ и почти всех $t \in \mathbb{B}_R$;
- для каждой точки $\bar{t} \in \mathbb{B}_R$ и каждого набора $\bar{z}_1, \dots, \bar{z}_s \in \mathbb{R}^n$ существует конечный замкнутый интервал $\Delta \subseteq \mathbb{B}_R$, для которого $\bar{t} \in \Delta$, и $r > 0$ такие, что функция f при почти каждом $t \in \Delta$ непрерывна по совокупности переменных $z_1, \dots, z_s \in \mathbb{R}^n$, $z_j \in B_r(\bar{z}_j)$, $B_r(\bar{z}_j) = \{z_j : \|\bar{z}_j - z_j\|_{\mathbb{R}^n} \leq r\}$, $j = 1, \dots, s$;
- $f(t, z_1, \dots, z_s)$ измерима по переменной $t \in \Delta$ при всех $z_1, \dots, z_s \in \mathbb{R}^n$, $z_j \in B_r(\bar{z}_j)$, $j = 1, \dots, s$;
- существует интегрируемая функция $\zeta_{\Delta r}(t)$, $t \in \Delta$ такая что $\|f(t, z_1, \dots, z_s)\|_{\mathbb{R}^n} \leq \zeta_{\Delta r}(t)$ при почти всех $t \in \Delta$ и всех $z_1, \dots, z_s \in \mathbb{R}^n$, $z_j \in B_r(\bar{z}_j)$, $j = 1, \dots, s$.

Из условий Каратеодори следует, что композиция измеримых функций $z_1(t), \dots, z_s(t)$, $t \in \Delta' \subseteq \Delta$ и функции $f(t, z_1, \dots, z_s)$ является измеримой функцией на Δ' , а также интегрируемой.

Определение 2. Совокупность конструкций $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ называется солитонным букетом и однозначно определяется набором $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, где:

- (1) Υ — конечно порожденная группа с образующими $\{\check{\gamma}_1, \dots, \check{\gamma}_d\}$ и выделенными элементами $\{\gamma_1, \dots, \gamma_s\}$, а также соответствующее пространство

$$\mathcal{K}_\Upsilon^n = \prod_{\gamma \in \Upsilon} \overline{R_\gamma}, \quad R_\gamma = \mathbb{R}^n,$$

бесконечных последовательностей $\varkappa = \{x_\gamma\}_{\gamma \in \Upsilon}$, $x_\gamma \in \mathbb{R}^n$, $x_\gamma = (x_{\gamma 1}, \dots, x_{\gamma n})'$, со стандартной топологией полного прямого произведения;

(2) $\mathbb{T}_\Upsilon = \{T_{\hat{\gamma}} : \hat{\gamma} \in \Upsilon\}$ — конечно порожденная группа сдвигов, действующая в пространстве $\mathcal{K}_\Upsilon^n$ по следующему правилу:

$$T_{\hat{\gamma}}\{x_\gamma\}_{\gamma \in \Upsilon} = \{x_{\gamma\hat{\gamma}}\}_{\gamma \in \Upsilon}, \quad \hat{\gamma} \in \Upsilon, \quad \{x_\gamma\}_{\gamma \in \Upsilon} \in \mathcal{K}_\Upsilon^n, \quad T_{\hat{\gamma}} \in \mathbb{T}_\Upsilon;$$

(3) $\eta : \Upsilon \rightarrow Q$ — эпиморфизм, где Q — группа диффеоморфизмов прямой, сохраняющих ориентацию, т.е. $Q \subset \text{Diff}_+^{(1)}(\mathbb{R})$, и, соответственно, Q — конечно порожденная группа с образующими $\check{q}_j = \eta(\check{\gamma}_j)$, $j = 1, \dots, d$, и выделенными элементами $q_j = \eta(\gamma_j)$, $j = 1, \dots, s$;

(4) $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ — функция, удовлетворяющая условиям Каратеодори;

(5) G_Γ — оператор, такой что

$$G_\Gamma : \mathbb{R} \times \mathcal{K}_\Upsilon^n \rightarrow \mathcal{K}_\Upsilon^n, \quad \Gamma = (\Upsilon, d, s, \eta, Q, g), \quad (1)$$

и что координата $(G_\Gamma(t, \varkappa))_e$ бесконечномерной вектор-функции $G_\Gamma(t, \varkappa)$, соответствующая единичному элементу e группы Υ , зависит только лишь от конечного числа координат и равна

$$(G_\Gamma(t, \varkappa))_e = g(t, x_{\gamma_1}, \dots, x_{\gamma_s}); \quad (2)$$

(6) при почти каждом $t \in \mathbb{R}$ выполняются “почти перестановочные” соотношения

$$T_{\hat{\gamma}} G_\Gamma(t, \varkappa) = \frac{d}{dt} \eta(\hat{\gamma})(t) \cdot G_\Gamma(\eta(\hat{\gamma})(t), T_{\hat{\gamma}} \varkappa), \quad \forall \varkappa \in \mathcal{K}_\Upsilon^n, \quad \forall \hat{\gamma} \in \Upsilon. \quad (3)$$

Для солитонного букета $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ в фазовом пространстве $\mathcal{K}_\Upsilon^n$ с фазовой переменной $\varkappa \in \mathcal{K}_\Upsilon^n$ определим систему

$$\dot{\varkappa}(t) = G_\Gamma(t, \varkappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (4)$$

$$\varkappa(\eta(\hat{\gamma})(t)) = T_{\hat{\gamma}} \varkappa(t) \quad \forall t \in \mathbb{R}, \quad \forall \hat{\gamma} \in \Upsilon. \quad (5)$$

Производная в бесконечномерном ОДУ (4) понимается как *производная по Гато*. Вектор-функция $\varkappa : [a, b] \rightarrow \mathcal{K}_\Upsilon^n$, $\varkappa(\cdot) = \{x_\gamma(\cdot)\}_{\gamma \in \Upsilon}$ называется *слабо дифференцируемой* в точке $t \in \mathbb{R}$, если для любого $\gamma \in \Upsilon$ функция $x_\gamma(\cdot)$ дифференцируема в точке $t \in \mathbb{R}$. Очевидно, что такая вектор-функция слабо дифференцируема в точке $t \in \mathbb{R}$ тогда и только тогда, когда она дифференцируема по Гато. Вектор-функция $\varkappa : [a, b] \rightarrow \mathcal{K}_\Upsilon^n$, $\varkappa(\cdot) = \{x_\gamma(\cdot)\}_{\gamma \in \Upsilon}$ называется *слабо абсолютно непрерывной (слабо измеримой)*, если для любого $\gamma \in \Upsilon$ функция $x_\gamma(\cdot)$ абсолютно непрерывна (измерима) на интервале $[a, b]$. *Решением системы* (4), (5) называется всякая слабо абсолютно непрерывная вектор-функция $\varkappa : \mathbb{R} \rightarrow \mathcal{K}_\Upsilon^n$, $\varkappa(\cdot) = \{x_\gamma(\cdot)\}_{\gamma \in \Upsilon}$, почти всюду удовлетворяющая уравнению (4) и при любом $t \in \mathbb{R}$ ограничению (5). Нелокальные ограничения (5) означают, что для решений системы *сдвиг по пространству равен сдвигу по времени*. Решения такой системы называются *решениями типа бегущей волны (солитонные решения)*, а группа Q называется *характеристикой бегущей волны*.

В паре с системой (4), (5) рассматривается индуцированное функционально-дифференциальное уравнение точечного типа

$$\dot{x}(t) = g(t, x(q_1(t)), \dots, x(q_s(t))), \quad t \in \mathbb{R}. \quad (6)$$

Каждый солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ определяет дуальную пару $(G_\Gamma|Q, g)$ функция-оператор. Для каждого солитонного букета (дуальной пары) существует канонический солитонный букет (каноническая дуальная пара), описание которой приводится ниже.

Лемма 1. *Каждый солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ определяет бесконечномерную систему (4), (5) и функционально-дифференциальное уравнение точечного типа (6). Каждое функционально-дифференциальное уравнение точечного типа (6) с функцией g , удовлетворяющей условиям Каратеодори, и конечно-порожденная группа $Q \subseteq \text{Diff}_+^{(1)}(\mathbb{R})$ с образующими $\check{q}_j \in Q$, $j = 1, \dots, d$, и выделенной системой элементов $q_j \in Q$, $j = 1, \dots, s$, индуцируют канонический солитонный букет*

$(Q, d, s, \mathcal{I}, G_\Gamma|Q, g)$ с $\Gamma = (Q, d, s, \mathcal{I}, Q, g)$ и каноническую дуальную пару $(G_\Gamma|Q, g)$, где $\mathcal{I}$ — тождественный автоморфизм группы Q .

Доказательство. Первое утверждение очевидно. Докажем второе утверждение.

Для этого нам следует для набора $(Q, d, s, \mathcal{I}, G_\Gamma|Q, g)$ задать оператор G_Γ в координатной форме

$$(G_\Gamma(t, \kappa))_q = \frac{d}{dt} \mathcal{I}(q)(t) \cdot g(\mathcal{I}(q)(t), x_{q_1q}, \dots, x_{q_sq}), \quad q \in Q. \quad (7)$$

Такой набор будет определять солитонный букет. Аналог системы (4), (5), определенной в фазовом пространстве $\mathcal{K}_Q^n$ с фазовой переменной $\kappa \in \mathcal{K}_Q^n$, $\kappa = \{x_q\}_{q \in Q}$, задающей решения типа бегущей волны, будет иметь вид

$$\dot{\kappa}(t) = G_\Gamma(t, \kappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (8)$$

$$\kappa(\hat{q}(t)) = T_{\hat{q}} \kappa(t) \quad \forall t \in \mathbb{R}, \quad \forall \hat{q} \in Q. \quad (9)$$

Аналог дифференциального уравнения (6), индуцированный солитонным букетом $(Q, d, s, \mathcal{I}, G_\Gamma|Q, g)$ с $\Gamma = (Q, d, s, \mathcal{I}, g)$, совпадает с этим же уравнением.

Для заданных Q, g канонический солитонный букет выделяется наиболее простой структурой набора $\Gamma = (Q, d, s, \mathcal{I}, Q, g)$ и, соответственно, оператора G_Γ .

Для солитонного букета справедливо следующее важное утверждение о дуализме солитонных решений и решений индуцированного функционально-дифференциального уравнения точечного типа.

Теорема 1. Пусть $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ является солитонным букетом. Тогда каждому решению $x(t), t \in \mathbb{R}$, функционально-дифференциального уравнения точечного типа (6) соответствует решение $\kappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}, t \in \mathbb{R}$ бесконечномерной системы (4)–(5) (солитонное решение) и наоборот. Такие решения связаны соотношениями $x_\gamma(t) = x_e(\eta(\gamma)(t)), x_e(t) = x(t), \gamma \in \Upsilon, t \in \mathbb{R}$.

Доказательство. Если $\kappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}, t \in \mathbb{R}$, решение бесконечномерной системы (4), (5), то функция $x_e(\cdot)$ будет удовлетворять функционально-дифференциальному уравнению точечного типа (6).

В обратную сторону. Пусть $x(\cdot)$ является решением функционально-дифференциального уравнения точечного типа (6). Определим бесконечномерную вектор-функцию $\kappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}, t \in \mathbb{R}$, где $x_\gamma(t) = x(\eta(\gamma)(t)), \gamma \in \Upsilon$. Тогда, в силу уравнения (6), будет выполняться система равенств

$$\begin{aligned} \dot{x}_\gamma(t) &= \frac{d}{dt} \eta(\gamma)(t) \cdot g(\eta(\gamma)(t), x(\eta(\gamma_1\gamma)(t)), \dots, x(\eta(\gamma_s\gamma)(t))), \\ \forall \gamma \in \Upsilon \quad &\text{для п.в. } t \in \mathbb{R} \end{aligned} \quad (10)$$

и вектор функция $\kappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}, t \in \mathbb{R}$, будет решением бесконечномерной системы (4), (5).

2. ВОПРОСЫ ЛОКАЛИЗАЦИИ ПРОСТРАНСТВА СОЛИТОННЫХ РЕШЕНИЙ И РЕШЕНИЙ ФУНКЦИОНАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Пусть $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ является солитонным букетом.

Так как пространство $\mathcal{K}_\Upsilon^n$ только лишь метризуемо, то исследование пространства солитонных решений (пространства решений системы (4), (5)) весьма затруднительно. Более продвинутое изучение требует локализации пространства солитонных решений, основанной на тех или иных представлениях. Одним из подходов по локализации пространства солитонных решений может быть требование их гладкости, или аналитичности [4]. В рамках такого способа локализации пространства солитонных решений, как правило, в пространстве бесконечно дифференцируемых функций удастся установить существование солитонных решений, а в пространстве аналитических функций удастся установить единственность солитонных решений.

Для более продвинутого исследования пространства солитонных решений предлагается иной подход локализации пространства солитонных решений. Такой подход основан на учете асимптотики решений по пространству как траекторий в бесконечномерном пространстве последовательностей $\mathcal{K}_\Upsilon^n$. Для этого в пространстве $\mathcal{K}_\Upsilon^n$ выбирается семейство банаховых подпространств

$\mathcal{K}_{\Upsilon\infty\mu h}^n$, $\mu \in (0, 1)$, $h \in \mathbb{R}_{\geq 0}$, $\mathbb{R}_{\geq 0} = [0, +\infty)$, непрерывно вложенных друг в друга так, что для любых $\mu' < \mu$, $\mathcal{K}_{\Upsilon\infty\mu h}^n \subset \mathcal{K}_{\Upsilon\infty\mu' h}^n$. При этом пространства $\mathcal{K}_{\Upsilon\infty\mu h}^n$ должны быть таковыми, чтобы отображение $G_{\Gamma p\mu h} = G_{\Gamma}|_{\mathbb{R} \times \mathcal{K}_{\Upsilon p\mu h}^n}$ действовало в пространство $\mathcal{K}_{\Upsilon p\mu h}^n$, т.е. отображение $G_{\Gamma p\mu h} : \mathbb{R} \times \mathcal{K}_{\Upsilon p\mu h}^n \rightarrow \mathcal{K}_{\Upsilon p\mu h}^n$ было корректно определено и удовлетворяло ряду хороших свойств, стандартных для дифференциальных уравнений с такой правой частью.

Определение 3. Вектор-функция $\varkappa : \mathbb{B}_R \rightarrow \mathbb{X}$, где $\mathbb{X}$ — банахово пространство и $\mathbb{B}_R$ — интервал на числовой прямой, называется сильно измеримой на $\mathbb{B}_R$, если она является на этом интервале пределом почти везде сходящейся последовательности счетнозначных вектор-функций $\varkappa_k(\cdot)$, $k \in \mathbb{Z}_+$. Если функция $\|\varkappa(t)\|_{\mathbb{X}}$ к тому же интегрируема, то вектор-функцию $\varkappa(\cdot)$ называют интегрируемой по Бохнеру (сильно интегрируемой).

Определение 4. Пусть $\mathbb{X}$ — банахово пространство и задано отображение (оператор) $\Phi : \mathbb{B}_R \times \mathbb{X} \rightarrow \mathbb{X}$, где $\mathbb{B}_R$ — интервал на числовой прямой. Тогда Φ удовлетворяет условиям Каратеодори в сильной форме, если:

- Φ определено при почти каждом $t \in \mathbb{B}_R$ и каждом $\varkappa \in \mathbb{X}$;
- для каждой точки $\bar{t} \in \mathbb{B}_R$ и каждом векторе $\bar{\varkappa} \in \mathbb{X}$ существуют конечный замкнутый интервал $\Delta \subseteq \mathbb{B}_R$, для которого $\bar{t} \in \Delta$, и $\sigma > 0$ такие, что вектор-функция $\Phi(t, \varkappa)$ при почти каждом $t \in \Delta$ непрерывна по переменной $\varkappa$ в шаре $B_{\sigma}(\bar{\varkappa}) = \{\varkappa : \|\bar{\varkappa} - \varkappa\|_{\mathbb{X}} \leq \sigma\}$;
- вектор-функция $\Phi(t, \varkappa)$ сильно измерима по переменной $t \in \Delta$ при всяком векторе $\varkappa$ из шара $B_{\sigma}(\bar{\varkappa}) = \{\varkappa : \|\bar{\varkappa} - \varkappa\|_{\mathbb{X}} \leq \sigma\}$;
- существует интегрируемая функция $\zeta_{\Delta\sigma}(t)$, $t \in \Delta$, такая, что при почти всех $t \in \Delta$ и всех $\varkappa$ из шара $B_{\sigma}(\bar{\varkappa}) = \{\varkappa : \|\bar{\varkappa} - \varkappa\|_{\mathbb{X}} \leq \sigma\}$ выполняется оценка $\|\Phi(t, \varkappa)\|_{\mathbb{X}} \leq \zeta_{\Delta\sigma}(t)$.

В случае конечномерности пространства $\mathbb{X}$ измеримость, дифференцируемость, абсолютная непрерывность в сильной и слабой форме и, соответственно, условия Каратеодори и условия Каратеодоре в сильной форме совпадают. Из условий Каратеодори в сильной форме следует, что композиция сильно измеримой вектор-функции $\varkappa(t)$, $t \in \Delta' \subseteq \Delta$, с образом в шаре $B_{\sigma}(\bar{\varkappa})$ и вектор-функции $\Phi(t, \varkappa)$ является сильно измеримой вектор-функцией, а также сильно интегрируемой (интегрируемой по Бохнеру) на интервале Δ' .

Определим семейство подпространств $\mathcal{K}_{\Upsilon\infty\mu h}^n$, $\mu \in (0, 1)$, $h \in \mathbb{R}_{\geq 0}$, следующим образом:

$$\mathcal{K}_{\Upsilon\infty\mu h}^n = \left\{ \varkappa : \varkappa \in \mathcal{K}_{\Upsilon}^n; \sup_{\gamma \in \Upsilon} \|x_{\gamma}\|_{\mathbb{R}^n} \mu^{l(\gamma)h} < +\infty \right\}, \quad (11)$$

$$\|\varkappa\|_{\Upsilon\infty\mu h} = \sup_{\gamma \in \Upsilon} \|x_{\gamma}\|_{\mathbb{R}^n} \mu^{l(\gamma)h},$$

где $l(\gamma)$ обозначает длину слова $\gamma \in \Upsilon$ в конечно-порожденной группе Υ ([19]). Параметр h в таких банаховых пространствах определяется по группе Q из рассматриваемого солитонного букета следующим образом:

$$h \geq \check{h}, \quad (12)$$

$$\check{h} = \max_{j \in \{1, \dots, d\}} h_{\check{q}_j}, \quad h_{\check{q}_j} = \sup_{t \in \mathbb{R}} |t - \check{q}_j(t)|, \quad j = 1, \dots, d,$$

где $\check{q}_j$, $j = 1, \dots, d$, образующие группы Q .

Сформулируем дополнительные условия на функцию g в виде условий роста, не выше линейного, а также условия для группы диффеоморфизмов Q из рассматриваемого солитонного букета:

- функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям Каратеодори;
- для любых $z_j, \bar{z}_j$, $j = 1, \dots, s$, и почти всех t имеет место рост не выше линейного

$$\|g(t, z_1, \dots, z_s)\|_{\mathbb{R}^n} \leq M_0(t) + M_1 \sum_{j=1}^s \|z_j\|_{\mathbb{R}^n}$$

и условие Липшица

$$\|g(t, z_1, \dots, z_s) - g(t, \bar{z}_1, \dots, \bar{z}_s)\|_{\mathbb{R}^n} \leq M_2 \sum_{j=1}^s \|z_j - \bar{z}_j\|_{\mathbb{R}^n};$$

(в) существует $\mu^* \in (0, 1]$ такое, что

$$M_0(\cdot) \in \mathcal{L}_{\mu^*}^1 L_\infty(\mathbb{R}) \quad \left(\mathcal{L}_{\mu^*}^1 L_\infty(\mathbb{R}) = \left\{ x(\cdot) : \sup_{t \in \mathbb{R}} \|x(t)\mu^{|t|}\|_{\mathbb{R}} < +\infty \right\} \right)$$

(г) задана группа диффеоморфизмов Q с системой образующих $\check{q}_1, \dots, \check{q}_d$ и выделенными элементами $q_1, \dots, q_s \in Q$, для любого $q \in Q$ величины

$$h_q = \sup_{t \in \mathbb{R}} |t - q(t)|, \quad h_{\dot{q}} = \sup_{t \in \mathbb{R}} \dot{q}(t)$$

конечны, семейство величин $h_{\dot{q}}, q \in Q$, равномерно ограничены и поэтому величины

$$\check{h} = \max_{j \in \{1, \dots, d\}} \sup_{t \in \mathbb{R}} |t - \check{q}_j(t)|, \quad h_{\dot{Q}} = \sup_{q \in Q} h_{\dot{q}}$$

конечны.

Все перечисленные условия для функции g , как правой части дифференциального уравнения, являются стандартными в теории дифференциальных уравнений.

Условие (г) мы всегда можем удовлетворить, сделав замену времени, но при этом может нарушиться условие Липшица (поточечно условие Липшица выполняется, но константы Липшица не являются равномерно ограниченными) и условие (в) (рост по переменной t может быть выше экспоненциального, что является нарушением стандартных условий на правую часть уравнения даже и для обыкновенного дифференциального уравнения).

Важное замечание состоит в том, что для конечно-разностных аналогов уравнений математической физики, в рамках дуальной пары “функция-оператор”, условия (а)–(г) на функцию g , как правую часть индуцированного функционально-дифференциального уравнения точечного типа (6), выполняются. При этом, группа Q изоморфна коммутативной группе $\mathbb{Z}^n$, где n — размерность фазового пространства функционально-дифференциального уравнения точечного типа (6).

Правая часть $g(\cdot)$ функционально-дифференциального уравнения точечного типа (6) является элементом банахового пространства функций $V_{\mu^*}(\mathbb{R} \times \mathbb{R}^{ns}, \mathbb{R}^n)$ с липшицевой нормой

$$\begin{aligned} V_{\mu^*}(\mathbb{R} \times \mathbb{R}^{ns}, \mathbb{R}^n) &= \{g(\cdot) : g(\cdot) \text{ удовлетворяет условиям (а)–(г)}\}, \\ \|g(\cdot)\|_{\text{Lip}} &= \sup_{t \in \mathbb{R}} \|g(t, 0, \dots, 0)(\mu^*)^{|t|}\|_{\mathbb{R}^n} + \\ &+ \sup_{(t, z_1, \dots, z_s, \bar{z}_1, \dots, \bar{z}_s) \in \mathbb{R}^{1+2ns}} \frac{\|g(t, z_1, \dots, z_s) - g(t, \bar{z}_1, \dots, \bar{z}_s)\|_{\mathbb{R}^n}}{\sum_{j=1}^s \|z_j - \bar{z}_j\|_{\mathbb{R}^n}}, \end{aligned} \quad (13)$$

где параметр $\mu \in \mathbb{R}_+$ совпадает с соответствующей константой из условия (в). Очевидно, что для функции $g(\cdot) \in V_{\mu^*}(\mathbb{R} \times \mathbb{R}^{ns}, \mathbb{R}^n)$ наименьшее значение константы M_2 из условия Липшица (условие (2)) совпадает со значением второго слагаемого в определении нормы $\|g(\cdot)\|_{\text{Lip}}$.

Опишем свойства оператора $G_{\Gamma \infty \mu h} = G|_{\mathbb{R} \times \mathcal{K}_{\Gamma \infty \mu h}^n}$. Для Υ из рассматриваемого солитонного букета и любых $\mu \in (0, 1)$, $h \in [0, +\infty)$, $t \in \mathbb{R}$, положим

$$\begin{aligned} \mathcal{P}_{\Upsilon \mu h} &= \sum_{j=1}^s \mu^{-l(\gamma_j)h}, \quad \mathcal{M}_{0\Upsilon \infty \mu h}(t) = h_{\dot{Q}} \sup_{\gamma \in \Upsilon} M_0(\eta(\gamma)(t)) \mu^{l(\gamma)h}, \\ \mathcal{M}_{1\Upsilon \infty \mu h} &= h_{\dot{Q}} M_1 \mathcal{P}_{\Upsilon \mu h}, \quad \mathcal{M}_{2\Upsilon \infty \mu h} = h_{\dot{Q}} M_2 \mathcal{P}_{\Upsilon \mu h}. \end{aligned} \quad (14)$$

Лемма 2. Пусть задан солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, функция $g(\cdot)$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Тогда для любых $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$, имеет место условие

$$\mathcal{M}_{0\Upsilon \infty \mu h}(\cdot) \in \mathcal{L}_{\mu^*}^1 L_\infty(\mathbb{R}). \quad (15)$$

Доказательство. Заметим, что для любых $t \in \mathbb{R}$, $\gamma \in \Gamma$ имеет место неравенство

$$|\eta(\gamma)(t)| \leq |t| + l(\gamma)\check{h}. \quad (16)$$

Используя неравенство (16), для любого $t \in \mathbb{R}$ получим цепочку оценок

$$\begin{aligned} \mathcal{M}_{0\Upsilon\infty\mu h}(t) &= h_{\check{Q}} \sup_{\gamma \in \Upsilon} M_0(\eta(\gamma)(t)) \mu^{l(\gamma)h} = \\ &= h_{\check{Q}} \sup_{\gamma \in \Upsilon} M_0(\eta(\gamma)(t)) (\mu^*)^{|\eta(\gamma)(t)|} (\mu^*)^{-|\eta(\gamma)(t)|} \mu^{l(\gamma)h} \leq \\ &\leq h_{\check{Q}} \sup_{\gamma \in \Upsilon} M_0(\eta(\gamma)(t)) (\mu^*)^{|\eta(\gamma)(t)|} (\mu^*)^{-|t|} \left(\frac{\mu}{\mu^*}\right)^{l(\gamma)h} \leq \\ &\leq h_{\check{Q}} \|M_0(\cdot)\|_{\mu^*} (\mu^*)^{-|t|} \sup_{\gamma \in \Upsilon} \left(\frac{\mu}{\mu^*}\right)^{l(\gamma)h} \end{aligned} \quad (17)$$

откуда и следует, что $\mathcal{M}_{0\Upsilon\infty\mu h}(\cdot) \in \mathcal{L}_{\mu^*}^1 L_\infty(\mathbb{R})$.

Лемма 3. Пусть задан солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, в котором функция $g(\cdot)$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Тогда для любых $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$, отображение $G_{\Gamma\infty\mu h}$, являющееся ограничением отображения G_Γ на подпространство $\mathbb{R} \times \mathcal{K}_{\Gamma\infty\mu h}^n$, при почти каждом $t \in \mathbb{R}$ и каждом $\varkappa \in \mathcal{K}_{\Gamma\infty\mu h}^n$ действует из $\mathbb{R} \times \mathcal{K}_{\Gamma\infty\mu h}^n$ в $\mathcal{K}_{\Gamma\infty\mu h}^n$, для него справедливы оценки

$$\|G_{\Gamma\infty\mu h}(t, \varkappa)\|_{\Upsilon\infty\mu h} \leq \mathcal{M}_{0\Upsilon\infty\mu h}(t) + \mathcal{M}_{1\Upsilon\infty\mu h} \|\varkappa\|_{\Upsilon\infty\mu h}, \quad (18)$$

$$\|G_{\Gamma\infty\mu h}(t, \varkappa) - G_{\Gamma\infty\mu h}(t, \bar{\varkappa})\|_{\Upsilon\infty\mu h} \leq \mathcal{M}_{2\Upsilon\infty\mu h} \|\varkappa - \bar{\varkappa}\|_{\Upsilon\infty\mu h} \quad (19)$$

и вектор-функция $G_{\Gamma\infty\mu h}(t, \varkappa)$ удовлетворяет условию Каратеодори в сильной форме.

Доказательство. Покажем, что отображение $G_{\Gamma\infty\mu h}$ действует из пространства $\mathbb{R} \times \mathcal{K}_{\Gamma\infty\mu h}^n$ в пространство $\mathcal{K}_{\Gamma\infty\mu h}^n$. Для любых $\varkappa \in \mathcal{K}_{\Gamma\infty\mu h}^n$ и почти каждом $t \in \mathbb{R}$, используя первую оценку из условия (б), получим

$$\begin{aligned} \|G_{\Gamma\infty\mu h}(t, \varkappa)\|_{\Upsilon\infty\mu h} &\leq \sup_{\gamma \in \Upsilon} \left\| \frac{d}{dt} \eta(\gamma)(t) g(\eta(\gamma)(t), x_{\gamma_1\gamma}, \dots, x_{\gamma_s\gamma}) \right\|_{R^n} \mu^{l(\gamma)h} \leq \\ &\leq h_{\check{Q}} \sup_{\gamma \in \Upsilon} \left[M_0(\eta(\gamma)(t)) + h_{\check{Q}} M_1 \sum_{j=1}^s \|x_{\gamma_j\gamma}\|_{R^n} \right] \mu^{l(\gamma)h} \leq \\ &\leq h_{\check{Q}} \sup_{\gamma \in \Upsilon} M_0(\eta(\gamma)(t)) \mu^{l(\gamma)h} + h_{\check{Q}} M_1 \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \|x_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma)h} = \\ &= \mathcal{M}_{0\Upsilon\infty\mu h}(t) + h_{\check{Q}} M_1 \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \|x_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma_j\gamma)h} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h} \leq \\ &\leq \mathcal{M}_{0\Upsilon\infty\mu h}(t) + h_{\check{Q}} M_1 \|\varkappa\|_{\Upsilon\infty\mu h} \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h}. \end{aligned} \quad (20)$$

Проведем оценку выражения, стоящего под знаком суммы в (20)

$$\sum_{j=1}^s \sup_{\gamma \in \Upsilon} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h} \leq \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \mu^{-(l(\gamma_j)+l(\gamma))h} \mu^{l(\gamma)h} = \sum_{j=1}^s \mu^{-l(\gamma_j)h}. \quad (21)$$

Если для последней суммы в неравенстве (20) использовать оценки (21) и обозначения (14), то получим оценку (18). В частности, отсюда будет следовать, что отображение $G_{\Gamma\infty\mu h}$ действует из пространства $\mathbb{R} \times \mathcal{K}_{\Gamma\infty\mu h}^n$ в пространство $\mathcal{K}_{\Gamma\infty\mu h}^n$.

Используя условие Липшица для функции $g(\cdot)$, для любых $\varkappa, \bar{\varkappa} \in K_{\Upsilon\infty\mu h}^n$ и почти каждом $t \in \mathbb{R}$ получим цепочку неравенств:

$$\begin{aligned}
 & \|G_{\Gamma\infty\mu h}(t, \varkappa) - G_{\Gamma\infty\mu h}(t, \bar{\varkappa})\|_{\Upsilon\infty\mu h} = \\
 & = \sup_{\gamma \in \Upsilon} \left\| \frac{d}{dt} \eta(\gamma)(t) [g(\eta(\gamma)(t), x_{\gamma_1\gamma}, \dots, x_{\gamma_s\gamma}) - g(\eta(\gamma)(t), \bar{x}_{\gamma_1\gamma}, \dots, \bar{x}_{\gamma_s\gamma})] \right\|_{R^n} \mu^{l(\gamma)} \leq \\
 & \leq h_{\dot{Q}} \sup_{\gamma \in \Upsilon} M_2 \sum_{j=1}^s \|x_{\gamma_j\gamma} - \bar{x}_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma)h} \leq h_{\dot{Q}} M_2 \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \|x_{\gamma_j\gamma} - \bar{x}_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma)h} = \\
 & = h_{\dot{Q}} M_2 \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \|x_{\gamma_j\gamma} - \bar{x}_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma_j\gamma)h} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h} \leq \\
 & \leq h_{\dot{Q}} M_2 \|\varkappa - \bar{\varkappa}\|_{\Upsilon\infty\mu h} \sum_{j=1}^s \sup_{\gamma \in \Upsilon} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h}. \tag{22}
 \end{aligned}$$

В силу оценки (21) для суммы из неравенства (22) и обозначений (14), непосредственно будет следовать оценка (19).

Остается показать, что отображение $G_{\Gamma\infty\mu h}$ удовлетворяет условию Каратеодори в сильной форме. Фиксируем произвольную точку $\bar{t} \in \mathbb{R}$ и вектор $\bar{\varkappa} \in K_{\Upsilon\infty\mu h}^n$. Выберем замкнутый интервал Δ , $\bar{t} \in \Delta$ и $\sigma > 0$. Тогда для почти каждого $t \in \Delta$ и каждого $\varkappa$ из шара $B_{\Upsilon\sigma}(\bar{\varkappa}) = \{\varkappa : \|\bar{\varkappa} - \varkappa\|_{\Upsilon\infty\mu h} \leq \sigma\}$ в силу неравенства (18) выполняется оценка $\|G_{\Gamma\infty\mu h}(t, \varkappa)\|_{\Upsilon\infty\mu h} \leq \zeta_{\Delta\sigma}(t)$, где $\zeta_{\Delta\sigma}(t) = \mathcal{M}_{0\Upsilon\infty\mu h}(t) + \mathcal{M}_{1\Upsilon\infty\mu h}[\|\bar{\varkappa}\|_{\Upsilon\infty\mu h} + \sigma]$. Так как функция $\zeta_{\Delta\sigma}(t)$ непрерывная по $t \in \Delta$, то она является интегрируемой на интервале Δ .

Непрерывность отображения $G_{\Gamma\infty\mu h}$ при почти каждом $t \in \Delta$ по переменной $\varkappa$ из шара $B_{\Upsilon\sigma}(\bar{\varkappa}) = \{\varkappa : \|\bar{\varkappa} - \varkappa\|_{\Upsilon\infty\mu h} \leq \sigma\}$ следует из оценки (19).

Остается показать сильную измеримость по переменной $t \in \Delta$ вектор-функции $G_{\Gamma\infty\mu h}(t, \varkappa)$ при каждом фиксированном векторе $\varkappa$ из шара $B_{\Upsilon\sigma}(\bar{\varkappa})$. Пусть Υ^r , $r = 0, 1, \dots$, множество слов в группе Υ длины не более чем r [19]. Покажем, что при каждом фиксированном векторе $\varkappa$ из шара $B_{\Upsilon\sigma}(\bar{\varkappa})$ и почти каждом $t \in \Delta$ имеет место сходимость

$$\lim_{r \rightarrow +\infty} \|P_{\Upsilon^r} G_{\Gamma\infty\mu h}(t, \varkappa) - G_{\Gamma\infty\mu h}(t, \varkappa)\|_{\Upsilon\infty\mu h} = 0, \tag{23}$$

равномерная по переменной $t \in \Delta$. Следуя преобразованиям из (17), (20), (21), получим

$$\begin{aligned}
 & \|P_{\Upsilon^r} G_{\Gamma\infty\mu h}(t, \varkappa) - G_{\Gamma\infty\mu h}(t, \varkappa)\|_{\Gamma\infty\mu h} = \\
 & = \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \left\| \frac{d}{dt} \eta(\gamma)(t) g(\eta(\gamma)(t), x_{\gamma_1\gamma}, \dots, x_{\gamma_s\gamma}) \right\|_{R^n} \mu^{l(\gamma)h} \leq \\
 & \leq h_{\dot{Q}} \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} [M_0(\eta(\gamma)(t)) + M_1 \sum_{j=1}^s \|x_{\gamma_j\gamma}\|_{R^n}] \mu^{l(\gamma)h} \leq \\
 & \leq h_{\dot{Q}} \left[\sup_{\gamma \in \Upsilon \setminus \Upsilon^r} M_0(\eta(\gamma)(t)) \mu^{l(\gamma)h} + \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} M_1 \sum_{j=1}^s \|x_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma)h} \right] \leq \\
 & \leq h_{\dot{Q}} \left[\sup_{\gamma \in \Upsilon \setminus \Upsilon^r} M_0(\eta(\gamma)(t)) \mu^{l(\gamma)h} + M_1 \sum_{j=1}^s \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \|x_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma)h} \right] \leq \\
 & \leq h_{\dot{Q}} \|M_0(\cdot)\|_{\mu^*} (\mu^*)^{-|t|} \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \left(\frac{\mu}{\mu^*} \right)^{l(\gamma)h} + \\
 & + h_{\dot{Q}} M_1 \sum_{j=1}^s \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \|x_{\gamma_j\gamma}\|_{R^n} \mu^{l(\gamma_j\gamma)h} \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \mu^{-l(\gamma_j\gamma)h} \mu^{l(\gamma)h} \leq
 \end{aligned}$$

$$\begin{aligned} &\leq h_{\dot{Q}} \|M_0(\cdot)\|_{\mu^*} (\mu^*)^{-|t|} \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \left(\frac{\mu}{\mu^*}\right)^{l(\gamma)h} + \\ &+ h_{\dot{Q}} M_1 \|P_{\Upsilon \setminus \Upsilon^{r-i} \mathcal{K}}\|_{\Gamma_{\infty \mu h}} \sum_{j=1}^s \mu^{-l(\gamma_j)h}. \end{aligned} \quad (24)$$

Тогда для $\varkappa = 0$ получим

$$\|P_{\Upsilon^r} G_{\Gamma_{\infty \mu h}}(t, 0) - G_{\Gamma_{\infty \mu h}}(t, 0)\|_{\Gamma_{\infty \mu h}} \leq h_{\dot{Q}} \|M_0(\cdot)\|_{\mu^*} (\mu^*)^{-|t|} \sup_{\gamma \in \Upsilon \setminus \Upsilon^r} \left(\frac{\mu}{\mu^*}\right)^{l(\gamma)h}. \quad (25)$$

При $r \rightarrow +\infty$ последний сомножитель в правой части неравенства (25) стремится к нулю. Поэтому, вся правая часть неравенства (25) равномерно по $t \in \Delta$ стремится к нулю. Вектор-функция $P_{\Upsilon^r} G_{\Gamma_{\infty \mu h}}(t, 0)$ имеет конечное число координат, не равных нулю. Поэтому она будет не только слабо измеримой, но также и сильно измеримой. А равномерный предел сильно измеримых вектор-функций также является сильно измеримой. Следовательно, вектор-функция $G_{\Gamma_{\infty \mu h}}(t, 0)$ также будет сильно измеримой. Используя сдвиг по переменной $\varkappa$ в банаховом пространстве $\mathcal{K}_{\Gamma_{\infty \mu h}}^n$, получим сильную измеримость вектор функции $G_{\Gamma_{\infty \mu h}}(t, \varkappa)$ при произвольном $\varkappa$.

В рамках дуализма между солитонными решениями и решениями индуцированного функционально-дифференциального точечного типа (6) мы будем изучать решения последних в параметрическом семействе банаховых подпространств (пространства функций с весами) с параметром $\mu \in (0, 1)$

$$\mathcal{L}_{\mu}^n C^{(0)}(\mathbb{R}) = \left\{ x(\cdot) : x(\cdot) \in C^{(0)}(\mathbb{R}, \mathbb{R}^n), \sup_{t \in \mathbb{R}} \|x(t)\mu^{|t|}\|_{\mathbb{R}^n} < +\infty \right\} \quad (26)$$

и нормой

$$\|x(\cdot)\|_{\mu}^{(0)} = \sup_{t \in \mathbb{R}} \|x(t)\mu^{|t|}\|_{\mathbb{R}^n}. \quad (27)$$

3. УТОЧНЕНИЕ СООТВЕТСТВИЙ В РАМКАХ ИЗУЧАЕМОГО ДУАЛИЗМА

Для заданных $\mu \in (0, 1)$, $h \in \mathbb{R}_{\geq 0}$, наряду с системой (4), (5) в исходном фазовом пространстве $\mathcal{K}_{\Upsilon}^n$, в подпространстве $\mathcal{K}_{\Gamma_{\infty \mu h}}^n$ рассматривается система

$$\frac{d\varkappa(t)}{dt} = G_{\Gamma_{\infty \mu h}}(t, \varkappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (28)$$

$$\varkappa(\eta(\hat{\gamma})(t)) = T_{\hat{\gamma}} \varkappa(t) \quad \forall \hat{\gamma}, t, \quad \hat{\gamma} \in \Upsilon, \quad t \in \mathbb{R}, \quad (29)$$

в виде дифференциального уравнения (28) с нелокальными ограничениями (29). Здесь дифференциальное уравнение (28) понимается в сильной форме (производная по Фреше), а не в слабой форме (производная по Гато) как в (4). По лемме 3 оператор $G_{\Gamma_{\infty \mu h}}$ удовлетворяет условиям Каратеодоре в сильном смысле и условию Липшица по переменной $\varkappa$. Для любой сильно измеримой (тем более непрерывной) вектор-функции $\varkappa : \mathbb{R} \rightarrow \mathcal{K}_{\Gamma_{\infty \mu h}}^n$ суперпозиция $G_{\Gamma_{\infty \mu h}}(t, \varkappa(t))$, $t \in \mathbb{R}$, на каждом конечном интервале является сильно измеримой и интегрируемой по Бохнеру вектор-функцией.

Решением такого дифференциального уравнения с фазовым пространством $\mathcal{K}_{\Gamma_{\infty \mu h}}^n$ называется всякая сильно абсолютно непрерывная вектор-функция $\varkappa(t)$, $t \in \mathbb{R}$ (непрерывная вектор-функция, почти всюду дифференцируемая по Фреше, производная $\dot{\varkappa}(\cdot)$ интегрируема по Бохнеру и вектор-функция $\varkappa(\cdot)$ однозначно восстанавливается по производной $\dot{\varkappa}(\cdot)$) почти всюду, удовлетворяющая этому уравнению. Для такого уравнения (28) справедлива теорема существования и единственности задачи Коши. Решения бесконечномерного дифференциального уравнения (28) имеет асимптотику по пространству, которая определяется параметром μ . Важной задачей является определение асимптотики такого решения системы (28), (29) (солитонное решение) и по времени.

Теорема 1 устанавливает соответствие между всем пространством солитонных решений (решений системы (4), (5)) и решениями функционально-дифференциального уравнения точечного типа (6).

Следующая задача состоит в том, чтобы уточнить это соответствие для подсемейства солитонных решений, как решений системы (28), (29) (солитонные решения), с решениями функционально-дифференциального уравнения точечного типа (6) в терминах асимптотики по пространству, а также по времени для первых, и асимптотики по времени для вторых.

Лемма 4. Пусть функция $g(\cdot)$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Если функция $x(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$, $\mu \in (0, \mu^*)$, является решением функционально-дифференциального уравнения точечного типа (6), то $\dot{x}(\cdot) \in \mathcal{L}_\mu^n L_\infty(R)$.

Доказательство. Пусть $x(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$, $\mu \in (0, \mu^*)$ и $x(\cdot)$ является решением исходного функционально-дифференциального уравнения точечного типа (6). Используя оценку

$$|q(t)| \leq |t| + l(q)\check{h} \quad \forall q \in Q, \quad (30)$$

для почти всякого $t \in \mathbb{R}$ получим

$$\begin{aligned} \|\dot{x}(t)\mu^{|t|}\|_{R^n} &= \|g(t, x(q_s(t)), \dots, x(q_s(t)))\mu^{|t|}\|_{R^n} \leq M_0(t)\mu^{|t|} + \\ &+ M_1 \sum_{j=1}^s \|x(q_j(t))\|_{R^n} \mu^{|t|} = M_0(t)\mu^{|t|} + M_2 \sum_{j=1}^s \|x(q_j(t))\|_{R^n} \mu^{|q_j(t)|} \mu^{-|q_j(t)|} \mu^{|t|} \leq \\ &\leq M_0(t)\mu^{|t|} + M_1 \sum_{j=1}^s \sup_{t' \in R} [\|x(q_j(t'))\|_{R^n} \mu^{|q_j(t')|}] \mu^{-|q_j(t)|} \mu^{|t|} = \\ &= M_0(t)\mu^{|t|} + M_1 \|x(\cdot)\|_\mu \sum_{j=1}^s \mu^{-|q_j(t)|} \mu^{|t|} \leq M_0(t)\mu^{|t|} + \\ &+ M_1 \|x(\cdot)\|_\mu \sum_{j=1}^s \mu^{-[|t|+l(q_j)\check{h}]} \mu^{|t|} = M_0(t)\mu^{|t|} + M_1 \|x(\cdot)\|_\mu \sum_{j=1}^s \mu^{-l(q_j)\check{h}}. \end{aligned}$$

Второе слагаемое является константой, а равномерная ограниченность первого слагаемого по $t \in \mathbb{R}$ следует из условия $\mu \in (0, \mu^*)$ и ограничений на рост функции $M_0(\cdot)$. Тогда для почти всякого $t \in \mathbb{R}$ величина $\|x(t)\mu^{|t|}\|_{R^n}$ также равномерно ограничена, что и доказывает лемму.

В рамках отмеченного дуализма сформулируем теорему о взаимосоответствии между решениями системы (28), (29) (солитонными решениями) и решениями функционально-дифференциального уравнения точечного типа (6).

Теорема 2. Пусть задан солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, в котором функция $g: \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Если функция $x(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$, $\bar{\mu} \in (0, \mu^*)$, является решением функционально-дифференциального уравнения точечного типа (6), тогда для произвольных заданных $\mu \in (0, \bar{\mu}]$, $h \in [\check{h}, +\infty)$ ей соответствует вектор-функция $\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}$, $\varkappa(t) \in \mathcal{K}_{\Upsilon \infty \mu h}^n$, $t \in \mathbb{R}$, такая, что $x_\gamma(t) = x_e(\eta(\gamma)(t))$, $x_e(t) = x(t)$, $\gamma \in \Upsilon$, $t \in \mathbb{R}$, и вектор-функция $\varkappa(\cdot)$:

а) является сильно абсолютно непрерывной;

б) является решением бесконечномерного обыкновенного дифференциального уравнения (28) и удовлетворяет ограничениям (29),

в) для функции $\rho(t) = \|\varkappa(t)\|_{\Upsilon \infty \mu h}$, $t \in \mathbb{R}$, справедливо вложение $\rho(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$.

Если для произвольных заданных $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$ вектор-функция $\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}$, $\varkappa(t) \in \mathcal{K}_{\Upsilon \infty \mu h}^n$, $t \in \mathbb{R}$, сильно абсолютно непрерывная, является решением бесконечномерной системы (28)–(29) (солитонное решение) и для функции $\rho(t) = \|\varkappa(t)\|_{\Upsilon \infty \mu h}$, $t \in \mathbb{R}$, справедливо вложение $\rho(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$, то ей соответствует решение $x(t)$, $t \in \mathbb{R}$, функционально-дифференциального уравнения точечного типа (6). Такие решения связаны соотношениями $x_\gamma(t) = x_e(\eta(\gamma)(t))$, $x_e(t) = x(t)$, $\gamma \in \Upsilon$, $t \in \mathbb{R}$, и выполняется вложение $x(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$.

Доказательство. Проведем доказательство первой части теоремы. Пусть $x(\cdot) \in \mathcal{L}_\mu^n C^{(0)}(R)$, $\bar{\mu} \in (0, \mu^*)$ и $x(\cdot)$ является решением функционально-дифференциального уравнения (6). Тогда по лемме 4 имеет место условие $\dot{x}(\cdot) \in \mathcal{L}_\mu^n L_\infty(R)$. Определим вектор-функцию $\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}$,

$x_\gamma(t) = x_e(\eta(\gamma)(t))$, $x_e(t) = x(t)$, $\gamma \in \Upsilon$, $t \in \mathbb{R}$. Покажем, что для произвольного $\mu \leq \bar{\mu}$ при каждом $t \in \mathbb{R}$ имеет место вложение $\varkappa(t) \in \mathcal{K}_{\Upsilon\infty\mu h}^n$.

Для произвольного $\mu \leq \bar{\mu}$ при каждом $t \in \mathbb{R}$, используя оценку $|\eta(\gamma)(t)| \leq |t| + l(\gamma)\check{h}$, $\forall \gamma \in \Upsilon$, получим

$$\begin{aligned} \|\varkappa(t)\|_{\Upsilon\infty\mu h} &= \sup_{\gamma \in \Upsilon} \|x(\eta(\gamma)(t))\|_{R^n} \mu^{l(\gamma)h} = \sup_{\gamma \in \Upsilon} \|x(\eta(\gamma)(t))\|_{R^n} \mu^{\lfloor |t| + l(\gamma)h \rfloor} \mu^{-|t|} = \\ &= \mu^{-|t|} \sup_{\gamma \in \Upsilon} \|x(\eta(\gamma)(t))\|_{R^n} \bar{\mu}^{\lfloor |t| + l(\gamma)h \rfloor} \left(\frac{\mu}{\bar{\mu}}\right)^{\lfloor |t| + l(\gamma)h \rfloor} \leq \\ &\leq \mu^{-|t|} \sup_{\gamma \in \Upsilon} \|x(\eta(\gamma)(t))\|_{R^n} \bar{\mu}^{\eta(\gamma)(t)} \left(\frac{\mu}{\bar{\mu}}\right)^{\lfloor |t| + l(\gamma)h \rfloor} \leq \\ &\leq \mu^{-|t|} \|x(\cdot)\|_{\bar{\mu}} \sup_{\gamma \in \Upsilon} \left(\frac{\mu}{\bar{\mu}}\right)^{\lfloor |t| + l(\gamma)h \rfloor} = \bar{\mu}^{-|t|} \|x(\cdot)\|_{\bar{\mu}} \sup_{\gamma \in \Upsilon} \left(\frac{\mu}{\bar{\mu}}\right)^{l(\gamma)h} = \bar{\mu}^{-|t|} \|x(\cdot)\|_{\bar{\mu}}, \end{aligned} \quad (31)$$

что и доказывает справедливость вложения $\varkappa(t) \in \mathcal{K}_{\Upsilon\infty\mu h}^n$ при каждом $t \in \mathbb{R}$. Очевидно, что таким образом определенная вектор-функция $\varkappa(t)$, $t \in \mathbb{R}$, удовлетворяет нелокальным ограничениям (29).

По лемме 4 имеют место условия $x(\cdot)$, $\dot{x}(\cdot) \in \mathcal{L}_{\bar{\mu}}^n C^{(0)}(R)$. Тогда для семейств функций $\bar{\mu}^{l(\gamma)h} x_\gamma(\cdot)$, $\bar{\mu}^{l(\gamma)h} \frac{d}{dt} x_\gamma(t)$, $\gamma \in \Upsilon$, при каждом $t \in \mathbb{R}$ и, соответственно, при почти каждом $t \in \mathbb{R}$ имеют место оценки

$$\begin{aligned} \|\bar{\mu}^{l(\gamma)h} x_\gamma(t)\|_{R^n} &= \|\bar{\mu}^{l(\gamma)h} x(\eta(\gamma)(t))\|_{R^n} = \bar{\mu}^{-|t|} \|\bar{\mu}^{\lfloor |t| + l(\gamma)h \rfloor} x(\eta(\gamma)(t))\|_{R^n} \leq \\ &\leq \bar{\mu}^{-|t|} \|\bar{\mu}^{\eta(\gamma)(t)} x(\eta(\gamma)(t))\|_{R^n} = \bar{\mu}^{-|t|} \|x(\cdot)\|_{\bar{\mu}}, \end{aligned} \quad (32)$$

$$\begin{aligned} \|\bar{\mu}^{l(\gamma)h} \frac{d}{dt} x_\gamma(t)\|_{R^n} &= \|\bar{\mu}^{l(\gamma)h} \frac{d}{dt} x(\eta(\gamma)(t))\|_{R^n} = \|\bar{\mu}^{l(\gamma)h} \dot{x}(\eta(\gamma)(t))\|_{R^n} \frac{d}{dt} \eta(\gamma)(t) \leq \\ &\leq h_{\dot{Q}} \|\bar{\mu}^{l(\gamma)h} \dot{x}(\eta(\gamma)(t))\|_{R^n} = h_{\dot{Q}} \bar{\mu}^{-|t|} \|\bar{\mu}^{\lfloor |t| + l(\gamma)h \rfloor} \dot{x}(\eta(\gamma)(t))\|_{R^n} \leq \\ &\leq h_{\dot{Q}} \bar{\mu}^{-|t|} \|\bar{\mu}^{\eta(\gamma)(t)} \dot{x}(\eta(\gamma)(t))\|_{R^n} \leq h_{\dot{Q}} \bar{\mu}^{-|t|} \|\dot{x}(\cdot)\|_{\bar{\mu}}. \end{aligned} \quad (33)$$

Следовательно, семейства функций $\bar{\mu}^{l(\gamma)h} x_\gamma(\cdot)$, $\bar{\mu}^{l(\gamma)h} \frac{d}{dt} x_\gamma(t)$, $\gamma \in \Upsilon$, на любом конечном интервале $[a, b]$ равномерно ограничены. В таком случае, семейство функций $\bar{\mu}^{l(\gamma)h} x_\gamma(\cdot)$, $\gamma \in \Upsilon$, будет также и равномерно непрерывным. Тогда вектор-функция $\varkappa(t)$, $t \in \mathbb{R}$, будет сильно абсолютно непрерывной. Так как $\bar{\mu} \in (0, \mu^*)$, то из условия $\mu \leq \bar{\mu}$ следует условие $\mu < \mu^*$. Тогда из леммы 3 следует, что вектор-функция $G_{\Upsilon\infty\mu h}(t, \varkappa)$ удовлетворяет условию Каратеодори в сильной форме, а также вектор-функция $\varkappa(\cdot)$ является решением обыкновенного дифференциального уравнения (28). Вложение $\rho(\cdot) \in \mathcal{L}_{\bar{\mu}}^n C^{(0)}(R)$ непосредственно следует из оценки (31). Первая часть теоремы доказана.

Доказательство второй части теоремы. В силу леммы 3 для произвольных $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$ вектор-функция $G_{\Upsilon\infty\mu h}(t, \varkappa)$ корректно определена и удовлетворяет условию Каратеодори в сильной форме. Для произвольных $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$ вектор-функция $\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \Upsilon}$, $\varkappa(t) \in \mathcal{K}_{\Upsilon\infty\mu h}^n$, $t \in \mathbb{R}$, сильно абсолютно непрерывна и является решением бесконечномерной системы (28), (29) (решение типа бегущей волны). Из условий (29) следуют соотношения $x_\gamma(t) = x_e(\eta(\gamma)(t))$, $\gamma \in \Upsilon$, $t \in \mathbb{R}$. Тогда функция $x(t) = x_e(t)$, $t \in \mathbb{R}$, будет абсолютно непрерывной, а из вида координаты с индексом e в бесконечномерном дифференциальном уравнении (28) следует, что она удовлетворяет функционально-дифференциальному уравнению (6). Остается установить асимптотику возможного роста решения $x(\cdot)$. Для любых $t \in \mathbb{R}$ справедлива оценка $\|x(t)\|_{R^n} \leq \|\varkappa(t)\|_{\Upsilon\infty\mu h}$. Так как для функции $\rho(t) = \|\varkappa(t)\|_{\Upsilon\infty\mu h}$, $t \in \mathbb{R}$, справедливо вложение $\rho(\cdot) \in \mathcal{L}_{\bar{\mu}}^n C^{(0)}(R)$, то отсюда будет следовать вложение $x(\cdot) \in \mathcal{L}_{\bar{\mu}}^n C^{(0)}(R)$. Теорема доказана.

В приведенной теореме вложение $\rho(\cdot) \in \mathcal{L}_{\bar{\mu}}^n C^{(0)}(R)$ для функции $\rho(t) = \|\varkappa(t)\|_{\Upsilon\infty\mu h}$, $t \in \mathbb{R}$, является важным. Поэтому опишем асимптотику по времени решений бесконечномерного дифференциального уравнения (28).

Предложение 1. Пусть задан солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, в котором функция g удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Тогда для

произвольных $\mu \in (0, \mu^*)$, $h \in [\tilde{h}, +\infty)$ в задаче Коши

$$\frac{d\kappa(t)}{dt} = G_{\Gamma\infty\mu h}(t, \kappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (34)$$

$$\kappa(\bar{t}) = \bar{\kappa}, \quad \bar{\kappa} \in \mathcal{K}_{\Gamma\infty\mu h}^n, \quad \bar{t} \in \mathbb{R}, \quad (35)$$

при любых заданных начальном условии $\bar{\kappa} \in \mathcal{K}_{\Gamma\infty\mu h}^n$ и начальном времени $\bar{t} \in \mathbb{R}$ существует, причем единственное, решение $\kappa(t)$, $t \in \mathbb{R}$. Для такого решения вектор-функция $\kappa(\cdot)$ является сильно абсолютно непрерывной, для функции $\rho(t) = \|\kappa(t)\|_{\Gamma\infty\mu h}$, $t \in \mathbb{R}$, справедливо вложение

$$\rho(\cdot) \in \mathcal{L}_{\hat{\mu}}^n C^{(0)}(R), \quad \hat{\mu} = \min\{\mu^*, e^{-\mathcal{M}_{2\Gamma\infty\mu h}}\} \quad (36)$$

и функция $\rho(t)$, $t \in \mathbb{R}$, как элемент пространства $\mathcal{L}_{\hat{\mu}}^n C^{(0)}(R)$, непрерывно зависит от начального данного $\bar{\kappa}$.

Доказательство. Из леммы 3 следует, что вектор-функция $G_{\Gamma\infty\mu h}(t, \kappa)$ корректно определена и удовлетворяет условию Каратеодори в сильной форме. Так как вектор-функция $G_{\Gamma\infty\mu h}(t, \kappa)$ по переменной κ удовлетворяет условию Липшица с константой $\mathcal{M}_{2\Gamma\infty\mu h}$, то для задачи Коши (34), (35) существует, причем единственное, решение $\kappa(t)$, $t \in \mathbb{R}$, такое, что вектор-функция $\kappa(\cdot)$ является сильно абсолютно непрерывной. Остается установить асимптотику роста такого решения. Решение $\kappa(t)$, $t \in \mathbb{R}$, имеет представление

$$\kappa(t) = \bar{\kappa} + \int_{\bar{t}}^t G_{\Gamma\infty\mu h}(\tau, \kappa(\tau)) d\tau$$

откуда и следует оценка

$$\frac{d}{dt} \|\kappa(t)\|_{\Gamma\infty\mu h} \leq \|G_{\Gamma\infty\mu h}(t, \kappa(t))\| \leq \mathcal{M}_{0\Gamma\infty\mu h}(t) + \mathcal{M}_{2\Gamma\infty\mu h} \|\kappa(t)\|_{\Gamma\infty\mu h},$$

а из последнего неравенства и вложения (15) из леммы 2 для функции $\mathcal{M}_{0\Gamma\infty\mu h}(\cdot)$ будет следовать вложение $\rho(\cdot) \in \mathcal{L}_{\hat{\mu}}^n C^{(0)}(R)$, $\hat{\mu} = \min\{\mu^*, e^{-\mathcal{M}_{2\Gamma\infty\mu h}}\}$.

Непрерывная зависимость решения дифференциального уравнения (34) от начального данного $\bar{\kappa}$ является хорошо известным фактом теории дифференциальных уравнений.

4. ОСНОВНАЯ НАЧАЛЬНО-КРАЕВАЯ ЗАДАЧА

В силу теоремы 2, изучение пространства солитонных решений эквивалентно изучению пространства решений системы (28), (29). Вопрос разрешимости такой системы переформулируем в форме разрешимости краевой задачи

$$\frac{d\kappa(t)}{dt} = G_{\Gamma\infty\mu h}(t, \kappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (37)$$

$$\kappa(\eta(\hat{\gamma})(\bar{t})) = T_{\hat{\gamma}}\kappa(\bar{t}) \quad \forall \hat{\gamma} \in \Upsilon, \quad (38)$$

при заданном начальном моменте $\bar{t} \in \mathbb{R}$.

Теорема 3. Пусть $(\Upsilon, d, s, \eta, G_{\Gamma}|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ является солитонным букетом, в котором функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Тогда при заданном начальном моменте $\bar{t} \in \mathbb{R}$ любое решение задачи (28), (29) (солитонное решение) является решением краевой задачи (37), (38) и наоборот.

Доказательство. В одну сторону доказательство очевидно и для этого не нужны никакие дополнительные условия. Пусть функция $\kappa(\cdot)$ является решением краевой задачи (37), (38). Уравнение (37) перепишем в эквивалентной интегральной форме: для любых $t \in \mathbb{R}$ справедливо равенство

$$\kappa(t) = \kappa(\bar{t}) + \int_{\bar{t}}^t G_{\Gamma\infty\mu h}(\tau, \kappa(\tau)) d\tau, \quad (39)$$

где интеграл понимается в смысле интеграла Бохнера. Из равенства (39) следует справедливость следующего равенства: для заданного фиксированного $\tilde{\gamma} \in \Upsilon$ и произвольного $t \in \mathbb{R}$

$$T_{\tilde{\gamma}}\kappa(t) = T_{\tilde{\gamma}}\kappa(\bar{t}) + \int_{\bar{t}}^t T_{\tilde{\gamma}}G_{\Gamma\infty\mu h}(\tau, \kappa(\tau)) d\tau.$$

В силу коммутационного соотношения (3) получим равенство

$$T_{\tilde{\gamma}}\kappa(t) = T_{\tilde{\gamma}}\kappa(\bar{t}) + \int_{\bar{t}}^t \frac{d}{dt} \eta(\tilde{\gamma})(t) G_{\Gamma\infty\mu h}(\eta(\tilde{\gamma})(\tau), T_{\tilde{\gamma}}\kappa(\tau)) d\tau, \quad t \in \mathbb{R}, \quad (40)$$

справедливое при любых фиксированных $\tilde{\gamma} \in \Upsilon$. Используя очевидное тождество $\eta(\tilde{\gamma}^{-1})(\eta(\tilde{\gamma})(t)) \equiv t$ и, сделав замену времени под знаком интеграла, придем к уравнению

$$T_{\tilde{\gamma}}\kappa(\eta(\tilde{\gamma}^{-1})(\eta(\tilde{\gamma})(t))) = T_{\tilde{\gamma}}\kappa(\bar{t}) + \int_{\eta(\tilde{\gamma})(\bar{t})}^{\eta(\tilde{\gamma})(t)} G_{\Gamma\infty\mu h}(\tau, T_{\tilde{\gamma}}\kappa(\eta(\tilde{\gamma}^{-1})(\tau))) d\tau,$$

также справедливое при всех $t \in \mathbb{R}$. Полученное уравнение эквивалентно тому, что вектор-функция $u(t) = T_{\tilde{\gamma}}\kappa(\eta(\tilde{\gamma}^{-1})(t))$ является решением задачи Коши для уравнения (37) с начальным значением вида $u(\eta(\tilde{\gamma})(\bar{t})) = T_{\tilde{\gamma}}\tilde{\kappa}$, где $\tilde{\kappa} = \kappa(\bar{t})$, и начальным моментом $\eta(\tilde{\gamma})(\bar{t})$.

С другой стороны, используя ограничение (38) для рассматриваемого решения $\kappa(\cdot)$, соотношение (40) преобразовывается к виду

$$\kappa(\eta(\tilde{\gamma})(t)) = T_{\tilde{\gamma}}\kappa(\bar{t}) + \int_{\bar{t}}^t \frac{d}{dt} \eta(\tilde{\gamma})(t) G_{\Gamma\infty\mu h}(\eta(\tilde{\gamma})(\tau), \kappa(\eta(\tilde{\gamma})(\tau))) d\tau, \quad t \in \mathbb{R}.$$

Проведя замену времени окончательно получим

$$\kappa(\eta(\tilde{\gamma})(t)) = T_{\tilde{\gamma}}\kappa(\bar{t}) + \int_{\eta(\tilde{\gamma})(\bar{t})}^{\eta(\tilde{\gamma})(t)} G_{\Gamma\infty\mu h}(\tau, \kappa(\tau)) d\tau, \quad t \in \mathbb{R}. \quad (41)$$

Полученное уравнение эквивалентно тому, что вектор-функция $\kappa(t)$ является решением задачи Коши для уравнения (37) с начальным значением вида $\kappa(\eta(\tilde{\gamma})(\bar{t})) = T_{\tilde{\gamma}}\tilde{\kappa}$, где $\tilde{\kappa} = \kappa(\bar{t})$, и начальным моментом $\eta(\tilde{\gamma})(\bar{t})$. В силу единственности решения задачи Коши для уравнения (37) имеет место тождество $u(t) \equiv \kappa(t)$.

Учитывая вид функции $u(t) = T_{\tilde{\gamma}}\kappa(\eta(\tilde{\gamma}^{-1})(t))$ получим, что для любых $t \in \mathbb{R}$ справедливы равенства

$$T_{\tilde{\gamma}}\kappa(\eta(\tilde{\gamma}^{-1})(t)) = \kappa(t).$$

Сделав подстановку $t = \eta(\tilde{\gamma})(\tau)$, получим, что для любых $\tau \in \mathbb{R}$ будут выполняться равенства

$$T_{\tilde{\gamma}}\kappa(\tau) = \kappa(\eta(\tilde{\gamma})(\tau)).$$

Так как это справедливо для каждого заданного $\tilde{\gamma} \in \Upsilon$, то они совпадают с нелокальными ограничениями (29), что и доказывает теорему.

Сформулируем важное следствие.

Следствие 1. Пусть $(\Upsilon, d, s, \eta, G_{\Gamma}|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ является солитонным букетом, в котором функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Фиксируем элемент $\tilde{\gamma} \in \Upsilon$ и начальный момент $\bar{t} \in \mathbb{R}$. Тогда при любых $\mu \in (0, \mu^*)$, $h \in [\tilde{h}, +\infty)$ всякое решение системы

$$\dot{\kappa}(t) = G_{\Gamma\infty\mu h}(t, \kappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (42)$$

$$\kappa(\eta(\tilde{\gamma})(t)) = T_{\tilde{\gamma}}\kappa(t) \quad \forall t \in \mathbb{R}, \quad (43)$$

является решением краевой задачи

$$\dot{\kappa}(t) = G_{\Gamma\infty\mu h}(t, \kappa) \quad \text{для п.в. } t \in \mathbb{R},$$

$$\kappa(\eta(\tilde{\gamma})(\bar{t})) = T_{\tilde{\gamma}}\kappa(\bar{t})$$

и наоборот.

Доказательство. Доказательство леммы полностью проведено в ходе доказательства теоремы 3.

На основе теоремы 3 и следствия 1 мы можем сформулировать краевую задачу, описывающую все солитонные решения (эквивалентную системе (28), (29)).

Теорема 4. Пусть $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$ является солитонным букетом, в котором функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Фиксируем начальный момент $\bar{t} \in \mathbb{R}$. Тогда при любых $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$ всякое решение системы (28), (29) (солитонное решение) является решением краевой задачи

$$\dot{\varkappa}(t) = G_{\Gamma\mu h}(t, \varkappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (44)$$

$$\varkappa(\eta(\check{\gamma}_k)(\bar{t})) = T_{\check{\gamma}_k} \varkappa(\bar{t}), \quad k = 1, \dots, d, \quad (45)$$

и наоборот.

Доказательство. Утверждение непосредственно следует из теоремы 3, следствия 1 и того, что множество элементов $\check{\gamma}_k$, $k = 1, \dots, d$, являются образующими группы Υ .

При заданных начальном моменте $\bar{t} \in \mathbb{R}$, начальном значении $\bar{x} \in \mathbb{R}^n$ и координате (элементе) $\bar{\gamma} \in \Upsilon$ сформулируем основную начально-краевую задачу

$$\dot{\varkappa}(t) = G_{\Gamma\infty\mu h}(t, \varkappa) \quad \text{для п.в. } t \in \mathbb{R}, \quad (46)$$

$$\varkappa(\eta(\check{\gamma}_k)(\bar{t})) = T_{\check{\gamma}_k} \varkappa(\bar{t}), \quad k = 1, \dots, d, \quad (47)$$

$$(\varkappa(\bar{t}))_{\bar{\gamma}} = \bar{x}, \quad \bar{\gamma} \in \Upsilon, \quad \bar{x} \in \mathbb{R}^n, \quad \bar{t} \in \mathbb{R}. \quad (48)$$

Теорема 5 (гипотеза). Пусть задан солитонный букет $(\Upsilon, d, s, \eta, G_\Gamma|Q, g)$ с $\Gamma = (\Upsilon, d, s, \eta, Q, g)$, в котором функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Если при $\mu \in (0, \mu^*)$, $h \in [\check{h}, +\infty)$ справедливо неравенство

$$\check{h} \dot{M}_2 \sum_{j=1}^s \mu^{-l(q_j)h} < h \ln \mu^{-1}, \quad (49)$$

то при любых заданных начальном моменте $\bar{t} \in \mathbb{R}$, начальном значении $\bar{x} \in \mathbb{R}^n$, координате $\bar{\gamma} \in \Upsilon$ для индуцированной основной начально-краевой задачи (46)–(48) существует решение $\varkappa(t) \in \mathcal{K}_{\Gamma\infty\mu h}^n$, $t \in \mathbb{R}$. Такое решение является единственным, а также непрерывно зависит от начального значения $\bar{x} \in \mathbb{R}^n$ и функции $g(\cdot) \in V_{\mu^*}(\mathbb{R} \times \mathbb{R}^{ns}, \mathbb{R}^n)$.

Доказательство гипотезы для случая, когда группа диффеоморфизма Q является циклической группой сдвигов и $h = \check{h}$, приведена в [11]. Пример манхэттенской решетки, который будет приведен ниже, относится к этому случаю. Полное доказательство гипотезы будет приведено в последующих публикациях.

Замечание 1. Пусть выполняется условие (49) и $\check{h} \neq 0$. Тогда для функции $\rho(t) = \|\varkappa(t)\|_{\Gamma\infty\mu h}$, $t \in \mathbb{R}$, где $\varkappa(\cdot)$ — решение уравнения (46), справедливо вложение (справедлива асимптотика по времени)

$$\rho(\cdot) \in \mathcal{L}_{\hat{\mu}}^n C^{(0)}(R), \quad \hat{\mu} = \mu^{\frac{h}{\check{h}}}. \quad (50)$$

Доказательство. Неравенство (49) имеет вид $\check{h} M_{2\Gamma\mu h} < h \ln \mu^{-1}$. Так как $\mu < \mu^*$, то, пользуясь оценкой (49) и уточнив значение $\hat{\mu}$ в (36) из предложения 1, получим вложение $\rho(\cdot) \in \mathcal{L}_{\hat{\mu}}^n C^{(0)}(R)$.

Из замечания 1 следует, что в случае $h = \check{h}$, $\check{h} \neq 0$ имеет место вложение $\rho(\cdot) \in \mathcal{L}_{\mu}^n C^{(0)}(R)$ и асимптотика решений уравнения (46) как по пространству, так и по времени задается параметром μ .

Параметрами в неравенстве (49) выступают группа Q и h . Все решения такого неравенства образуют некоторый максимальный связный интервал $(\mu_1(Q, h), \mu_2(Q, h)) \subseteq (0, 1)$. При малых $\varepsilon > 0$ пространство $\mathcal{K}_{\Gamma\infty(\mu_2(Q, h)-\varepsilon)h}^n$ уже чем пространство $\mathcal{K}_{\Gamma\infty(\mu_1(Q, h)+\varepsilon)h}^n$. Справедливость гипотезы будет гарантировать существование решение основной начально-краевой задачи в узком пространстве $\mathcal{K}_{\Gamma\infty(\mu_2(Q, h)-\varepsilon)h}^n$ и единственность в широком пространстве $\mathcal{K}_{\Gamma\infty(\mu_1(Q, h)+\varepsilon)h}^n$.

5. ФУНКЦИОНАЛЬНО-ДИФФЕРЕНЦИАЛЬНОЕ УРАВНЕНИЕ ТОЧЕЧНОГО ТИПА

Сформулируем задачу Коши

$$\dot{x}(t) = g(t, x(q_1(t)), \dots, x(q_s(t))), \quad t \in \mathbb{R}, \quad (51)$$

$$x(\hat{t}) = \bar{x}, \quad \bar{x} \in \mathbb{R}^n, \quad t \in \mathbb{R}. \quad (52)$$

В действительности, соответствия, которые описаны в теореме 2, продолжаются до соответствий между решениями основной начально-краевой задачи (46)–(48) и решениями задачи Коши (51), (52). В них начальное значение $\bar{x}$ одно и то же, а начальные моменты связаны соотношением $\hat{t} = \bar{\gamma}(\hat{t})$.

Пользуясь дуализмом и, соответственно, теоремой 2 о взаимосоответствиях, мы можем переформулировать гипотезу из предыдущего раздела в отношении к функционально-дифференциальному уравнению точечного типа (6).

Теорема 6. Пусть функция $g : \mathbb{R} \times \mathbb{R}^{ns} \rightarrow \mathbb{R}^n$ удовлетворяет условиям (а)–(в), а группа диффеоморфизмов Q — условиям (г). Если при $\mu \in (0, \mu^*)$ справедливо неравенство

$$h_{\dot{Q}} M_2 \sum_{j=1}^s \mu^{-l(q_j)\tilde{h}} < \ln \mu^{-1}, \quad (53)$$

то для задачи Коши (51), (52) существует решение $x(\cdot) \in \mathcal{L}_{\mu}^n C^{(0)}(R)$. Такое решение является единственным, а также как элемент пространства $\mathcal{L}_{\mu}^n C^{(0)}(R)$ непрерывно зависит от начального значения $\bar{x} \in \mathbb{R}^n$ и функции $g(\cdot) \in V_{\mu^*}(\mathbb{R} \times \mathbb{R}^{ns}, \mathbb{R}^n)$.

Доказательство. По лемме 1 такие функция g и группа диффеоморфизмов Q индуцируют канонический солитонный букет $(Q, d, s, \mathcal{I}, G_{\Gamma}|Q, g)$ с $\Gamma = (Q, d, s, \mathcal{I}, Q, g)$ и каноническую дуальную пару $(G_{\Gamma}|Q, g)$, где $\mathcal{I}$ — тождественный автоморфизм группы Q . Для такого солитонного букета рассмотрим основную начально-краевую задачу (46)–(48), в которой положено $h = \tilde{h}$. В силу предложения 1, для решений бесконечномерного уравнения (46) имеет место асимптотика по времени $\rho(\cdot) \in \mathcal{L}_{\mu}^n C^{(0)}(R)$, $\rho(t) = \|\varkappa(t)\|_{Y_{p\mu\tilde{h}}}$, $t \in \mathbb{R}$, $\hat{\mu} = \min\{\mu^*, e^{-M_{2\Upsilon_{p\mu\tilde{h}}}}\}$. Неравенство (53) имеет вид $M_{2\Upsilon_{p\mu\tilde{h}}} < \ln \mu^{-1}$. Так как $\mu < \mu^*$, то мы получим вложение $\rho(\cdot) \in \mathcal{L}_{\mu}^n C^{(0)}(R)$. Далее доказательство следует из теоремы о соответствиях 2 между решениями основной начально-краевой задачи и решениями индуцированного функционально-дифференциального уравнения точечного типа и справедливости гипотезы из предыдущего раздела.

Доказательство гипотезы для случая, когда группа диффеоморфизмов Q является циклической группой сдвигов, приведена в [11].

6. ПРИМЕР МАНХЭТТЕНСКОЙ РЕШЕТКИ

Схема движения в манхэттенской решетке (группа $\mathbb{Z}^2$) такова, что из каждой точки узла решетки движение возможно в трех направлениях по основным магистралям (вперед, влево, вправо) и движение в обратном направлении через накопители. Рассмотрим агрегированную модель такого движения, при котором в каждом узле решетки наблюдаемой величиной является только лишь общий объем трафика без разделения его по направлениям движения [18].

Для группы $\Gamma = \mathbb{Z}^2$ через $\{\gamma_1, \gamma_2\}$ обозначим систему ее образующих, где $\gamma_1 = (1, 0)$, $\gamma_2 = (0, 1)$. Движение описывается одномерным конечно-разностным аналогом параболического уравнения [17]

$$\begin{aligned} \dot{x}_{\gamma}(t) &= \sum_{l=1}^2 \alpha_{\gamma_l} x_{\gamma_l \gamma} - \sum_{l=1}^2 (\alpha_{\gamma_l} + \beta_{\gamma_l}) x_{\gamma} + \sum_{l=1}^2 \beta_{\gamma_l} x_{\gamma_l^{-1} \gamma} + \varphi(x_{\gamma}), \\ x_{\gamma} &\in \mathbb{R} \quad \forall \gamma \in \mathbb{Z}^2 \quad \text{п.в. } t \in \mathbb{R}. \end{aligned} \quad (54)$$

Для каждого $\gamma \in \mathbb{Z}^2$ (для каждого узла двумерной решетки) через x_{γ} обозначается величина стока транспортного потока в нем. Величины $\alpha_{\gamma_l} [x_{\gamma_l \gamma} - x_{\gamma}]$, $\beta_{\gamma_l} [x_{\gamma_l^{-1} \gamma} - x_{\gamma}]$, $l = 1, 2$, задают интенсивность изменения потока в узле γ в зависимости от разности значений величины потоков в узлах γ , $\gamma_l \gamma$, $\gamma_l^{-1} \gamma$, $l = 1, 2$, соответственно. При отрицательных значениях таких разностей происходит переток потока из узла γ в узлы $\gamma_l \gamma$, $\gamma_l^{-1} \gamma$, $l = 1, 2$, соответственно. При положительных значениях таких разностей происходит переток потока из накопителей в узел γ . Потенциал φ задается непрерывной функцией

и имеет вид

$$\varphi(t) = \begin{cases} 0, & t \in (-\infty, 0] \cup [\Delta, +\infty), \\ \varphi(t) > 0, & t \in (0, \Delta). \end{cases} \quad (55)$$

Уровень потока, равный Δ , является критическим для узла и определяется его техническими возможностями. При значениях величины потока, меньших критического значения Δ , происходит переток потока из накопителей в узел γ . При значениях величины потока, больших критического значения Δ , он остается в этом же узле. Это соответствует случаю, когда объемы накопителей для разгрузки узла малы. Ранее, в работе [17] рассматривался случай, когда объемы накопителей для разгрузки узла достаточны и происходит переток потока из узла γ в накопители, чему соответствовал отрицательный аффинный потенциал на полупрямой $(\Delta, +\infty)$.

Определение 5. Будем говорить, что решение $\{x_\gamma(\cdot)\}_{\gamma \in \mathbb{Z}^2}$ системы (54), определенное для всех $t \in \mathbb{R}$, имеет тип бегущей волны (солитонное решение), если существует $\tau \geq 0$, не зависящая от t и γ , что при всех $\gamma \in \mathbb{Z}^2$ и $t \in \mathbb{R}$ выполнено равенство

$$x_\gamma(t + \tau) = x_{\gamma_l \gamma}(t), \quad l = 1, 2. \quad (56)$$

Константу τ будем называть *характеристикой бегущей волны*.

Задача заключается в том, чтобы установить в каких диапазонах характеристики τ солитонное решение существуют, какова асимптотика таких решений и их зависимость от параметров системы. Среди солитонных решений выделяется важный класс ограниченных солитонных решений, как один из наиболее информативных классов решений.

Определим солитонный букет, порождающий рассматриваемую систему и представим систему (54), (56) в операторной форме. Через $\mathcal{K}_{\mathbb{Z}^2}^1$ обозначим пространство последовательностей с элементами $\varkappa = \{x_\gamma\}_{\gamma \in \mathbb{Z}^2}$, $x_\gamma \in \mathbb{R}$, наделенной тихоновской топологией. Определим линейный оператор $\mathbb{A}$, нелинейный оператор $\mathbb{F}$ и группу операторов сдвига $\mathbb{T}_{\mathbb{Z}^2} = \{T_{\hat{\gamma}} : \hat{\gamma} \in \mathbb{Z}^2\}$, действующие непрерывно из пространства $\mathcal{K}_{\mathbb{Z}^2}^1$ в себя по следующему правилу: для любых $\gamma, \hat{\gamma} \in \mathbb{Z}^2$, $\varkappa \in \mathcal{K}_{\mathbb{Z}^2}^1$, $\varkappa = \{x_\gamma\}_{\gamma \in \mathbb{Z}^2}$

$$\begin{aligned} (\mathbb{A}\varkappa)_\gamma &= \sum_{l=1}^2 \alpha_{\gamma_l} x_{\gamma_l \gamma} - \sum_{l=1}^2 (\alpha_{\gamma_l} + \beta_{\gamma_l}) x_\gamma + \sum_{l=1}^2 \beta_{\gamma_l} x_{\gamma_l^{-1} \gamma}, \quad (\mathbb{F}(\varkappa))_\gamma = \varphi(x_\gamma), \\ T_{\hat{\gamma}} \{x_\gamma\}_{\gamma \in \mathbb{Z}^2} &= \{x_{\gamma \hat{\gamma}}\}_{\gamma \in \mathbb{Z}^2}. \end{aligned} \quad (57)$$

Определим циклическую группу $Q = \langle \check{q} \rangle$, $\check{q}(t) = t + \tau$, эпиморфизм $\eta : \mathbb{Z}^2 \rightarrow Q$, где $\eta(\gamma_l) = \check{q}$, $l = 1, 2$, для образующих γ_l , $l = 1, 2$, группы $\mathbb{Z}^2$, а также отображение $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ вида

$$\begin{aligned} g(z_1, z_0, z_{-1}) &= \alpha z_1 - (\alpha + \beta) z_0 + \beta z_{-1} + \varphi(z_0), \\ \alpha &= \alpha_{\gamma_1} + \alpha_{\gamma_2}, \quad \beta = \beta_{\gamma_1} + \beta_{\gamma_2}. \end{aligned}$$

Коэффициенты α_{γ_1} , α_{γ_2} , β_{γ_1} , β_{γ_2} являются неотрицательными.

Зададим солитонный букет $(\mathbb{Z}^2, 2, 1, \eta, G_\Gamma | Q, g)$ с $\Gamma = (\mathbb{Z}^2, 2, 1, \eta, Q, g)$, в котором оператор G_Γ имеет вид $G_\Gamma = \mathbb{A} + \mathbb{F}$. Система (54), (56) имеет следующее эквивалентное операторное представление

$$\dot{\varkappa}(t) = G_\Gamma(\varkappa) \quad \text{п.в. } t \in \mathbb{R}. \quad (58)$$

$$\varkappa(\eta(\hat{\gamma})(t)) = T_{\hat{\gamma}} \varkappa(t) \quad \forall \hat{\gamma} \in \mathbb{Z}^2, \quad \forall t \in \mathbb{R}. \quad (59)$$

Рассмотрим индуцированное функционально-дифференциальное уравнение точечного типа

$$\dot{x}(t) = \alpha x(t + \tau) - (\alpha + \beta) x(t) + \beta x(t - \tau) + \varphi(x(t)) \quad \text{п.в. } t \in \mathbb{R}. \quad (60)$$

Наряду с системой (58), (59) будем рассматривать ее ограничение на банахово подпространство $\mathcal{K}_{\mathbb{Z}^2 \infty \mu \tau}^1$, $\mu \in (0, 1)$, $\tau \in [0, +\infty)$

$$\mathcal{K}_{\mathbb{Z}^2 \infty \mu \tau}^1 = \{\varkappa : \varkappa \in \mathcal{K}_{\mathbb{Z}^2}^1; \sup_{\gamma \in \mathbb{Z}^2} |x_\gamma| \mu^{l(\gamma)\tau} < +\infty\}, \quad l(\gamma) = |(i, j)| = |i| + |j| \quad (61)$$

с нормой

$$\|\mathcal{X}\|_{\mathbb{Z}^2 \times \mu\tau} = \sup_{\gamma \in \mathbb{Z}^2} |x_\gamma| \mu^{l(\gamma)\tau}.$$

Такая система имеет вид

$$\dot{\mathcal{X}}(t) = G_{\Gamma \infty \mu \tau}(\mathcal{X}) \quad \text{п.в. } t \in \mathbb{R}, \quad (62)$$

$$\mathcal{X}(\eta(\hat{\gamma})(t)) = T_{\hat{\gamma}} \mathcal{X}(t) \quad \forall \hat{\gamma} \in \mathbb{Z}^2, \quad \forall t \in \mathbb{R}. \quad (63)$$

Преобразуем функционально-дифференциальное уравнение точечного типа (60) из рассматриваемой дуальной пары. Проведем замену времени с тем, чтобы отклонения аргумента стали целочисленными, а характеристика τ рассматривалась как параметр

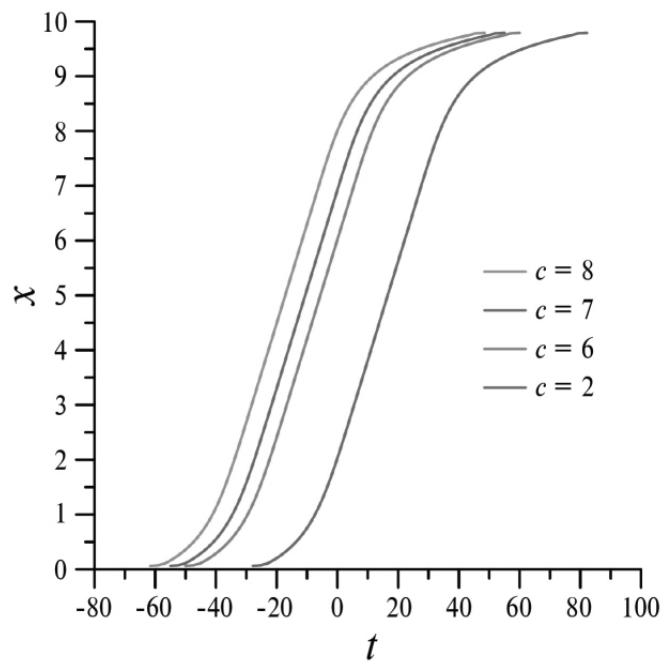
$$\dot{\bar{x}}(t) = \tau[\alpha \bar{x}(t+1) - (\alpha + \beta) \bar{x}(t) + \beta \bar{x}(t-1) + \varphi(\bar{x}(t))] \quad \text{п.в. } t \in \mathbb{R}.$$

Такое уравнение было исследовано в монографии [11] при минимальных ограничениях на потенциал $\varphi(\cdot)$ в виде наличия условия Липшица (квазилинейные потенциалы).

Пусть потенциал φ удовлетворяет условию Липшица с константой L_φ . Рассмотрим трансцендентное уравнение относительно двух переменных $\tau \in (0, +\infty)$ и $\mu \in (0, 1)$

$$2C_\varphi \tau (\mu^{-1} + 1) = \ln \mu^{-1}, \quad C_\varphi = \max \{ \alpha + \beta; L_\varphi \}. \quad (64)$$

Множество решений уравнения (64) описывается функциями $\mu_1(\tau)$, $\mu_2(\tau)$, заданными на фиг. 1.



Фиг. 1. Графики функций $\mu_1(\tau)$, $\mu_2(\tau)$.

Сформулируем теорему существования и единственности решения для индуцированного функционально-дифференциального уравнения (60).

Теорема 7 (см. [11]). Пусть потенциал φ удовлетворяет условию Липшица с константой L_φ . Тогда при любых начальных данных $a \in \mathbb{R}$, начальном моменте $\hat{t} \in \mathbb{R}$ и характеристиках $\tau > 0$, удовлетворяющих условию

$$0 < \tau < \hat{\tau},$$

в пространстве $\mathcal{L}_\mu^1 C^{(0)}(\mathbb{R}) \forall \mu, \mu^\tau \in (\mu_1(\tau), \mu_2(\tau))$, для функционально-дифференциального уравнения (60) существует и причем единственное решение $x(t)$, $t \in \mathbb{R}$, такое, что оно удовлетворяет начальному условию $x(\hat{t}) = a$. Такое решение, как элемент пространства $\mathcal{L}_\mu^1 C^{(0)}(\mathbb{R})$, непрерывно зависит от начального данного $a \in \mathbb{R}$, характеристики τ и потенциала $\varphi(\cdot)$.

Ранее мы определили максимальный связный интервал $(\mu_1(Q, h), \mu_2(Q, h))$, для значений μ из которого справедливо основное условие (49). В рассматриваемом примере для параметров выполняются условия $h = \check{h} = \tau$ и имеют место соотношения $\mu_1(Q, \tau) = \sqrt[3]{\mu_1(\tau)}$, $\mu_2(Q, \tau) = \sqrt[3]{\mu_2(\tau)}$.

Теорема 8. Пусть потенциал φ удовлетворяет условию Липшица с константой L_φ . Тогда при любых начальных данных $\bar{\gamma} \in \mathbb{Z}^2$, $a \in \mathbb{R}$, $\bar{t} \in \mathbb{R}$, и характеристиках $\tau > 0$, удовлетворяющих условию

$$0 < \tau < \hat{\tau},$$

для исходной бесконечномерной системы дифференциальных уравнений (54) существует единственное солитонное решение $\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \mathbb{Z}^2}$, $t \in \mathbb{R}$ (решением системы (54), (56)), с характеристикой τ такое, что оно удовлетворяет начальному условию $x_{\bar{\gamma}}(\bar{t}) = a$, для любого параметра μ , $\mu^\tau \in (\mu_1(\tau), \mu_2(\tau))$, значения вектор-функции

$$\varkappa(t) = \{x_\gamma(t)\}_{\gamma \in \mathbb{Z}^2}$$

при любом $t \in \mathbb{R}$ принадлежат пространству $\mathcal{K}_{\mathbb{Z}^2 \infty \mu \tau}^1$, а функция

$$\rho(t) = \|\varkappa(t)\|_{\mathbb{Z}^2 \infty \mu \tau}, \quad t \in \mathbb{R},$$

принадлежит пространству $\mathcal{L}_\mu^1 C^{(0)}(\mathbb{R})$. Такое решение непрерывно зависит от начального данного $a \in \mathbb{R}$, характеристики τ и потенциала $\varphi(\cdot)$.

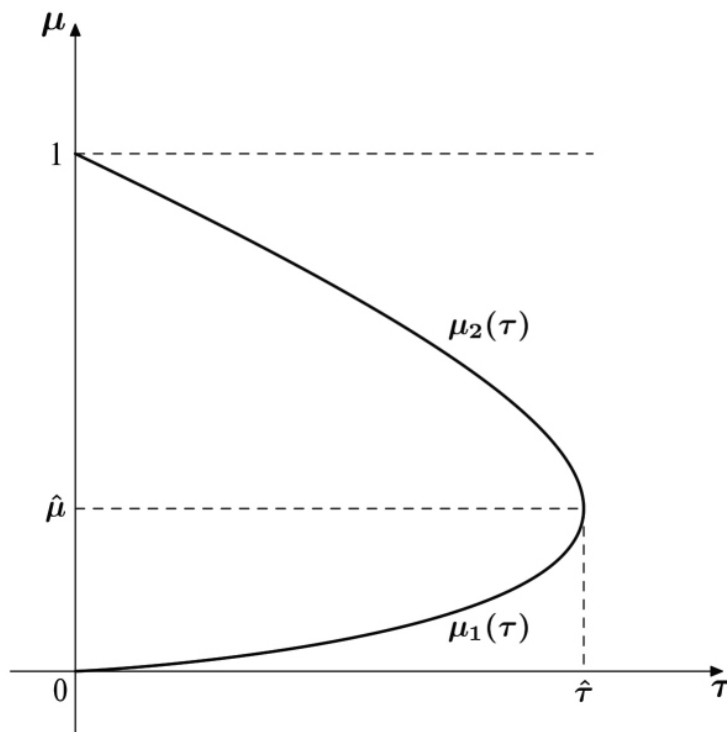
Доказательство. Данное утверждение непосредственно следует из теорем 2 и 7.

6.1. Ограниченные солитонные решения

Среди солитонных решений особое место занимают ограниченные солитонные решения. В силу теоремы 2, ограниченными солитонными решениями соответствуют ограниченные решения индуцированного функционально-дифференциального уравнения и они принадлежат каждому из семейства пространств $\mathcal{L}_\mu^1 C^{(0)}(\mathbb{R})$, $\mu \in (0, 1)$, а по теореме 7 в пространствах функций $\mathcal{L}_\mu^1 C^{(0)}(\mathbb{R})$, $\mu^\tau \in (\mu_1(\tau), \mu_2(\tau))$, $0 < \tau < \hat{\tau}$, справедливо утверждение о единственности для ограниченных решений.

Для начала опишем стационарные решения:

$$x(t) \equiv \xi, \quad t \in \mathbb{R}, \quad \xi \in (-\infty, 0] \cup [\Delta, +\infty). \quad (65)$$



Фиг. 2. График ограниченных решений для различных c .

Рассматриваемое уравнение обладает рядом симметрий. В силу автономности уравнения и теоремы существования и единственности решения (теорема 7), пространство решений инвариантно относительно сдвига решений по времени. Более того, в силу одномерности фазового пространства для функционально-дифференциального уравнения (60), всякое периодическое решение является стационарным. Поэтому не стационарные решения являются строго монотонными и, в силу условий (65), они являются ограниченными.

Дальнейшее усложнение рассматриваемой модели связано с учетом также и направления движения. Это приводит к модели для той же двумерной решетки, где в каждом узле решетки имеется многомерное пространство переменных (16-мерное пространство переменных) с более сложной системой их взаимовлияния.

СПИСОК ЛИТЕРАТУРЫ

1. Френкель Я. И., Конторова Т. А. О теории пластической деформации и двойственности // Ж. экспериментальной и теор. физ. 1938. Т. 8. С. 89–97.
2. Toda М. Теория нелинейных решеток. М.: Мир, 1984. 262 с.
3. Мива Т., Джимбо М., Датэ Э. Солитоны: дифференциальные уравнения, симметрии и бесконечные алгебры. М.: МЦНМО, 2005. 112 с.
4. Пустыльников Л. Д. Бесконечномерные нелинейные обыкновенные дифференциальные уравнения и теория КАМ // УМН. 1997. Т. 52. № 3. С. 105–160.
5. Бекларян Л. А. Краевая задача для дифференциального уравнения с отклоняющимся аргументом // Докл. АН СССР. 1986. Т. 291. № 1. С. 19–22.
6. Бекларян Л. А. Дифференциальное уравнение с отклоняющимся аргументом как бесконечномерная динамическая система // ВЦ АН СССР. Сообщ. по приклад. матем. 1989. 18 с.
7. Бекларян Л. А. Об одном методе регуляризации краевых задач для дифференциальных уравнений с отклоняющимся аргументом // Докл. АН СССР. 1991. Т. 317. № 5. С. 1033–1037.
8. Бекларян Л. А. Групповые особенности дифференциальных уравнений с отклоняющимся аргументом и связанные с ними метрические инварианты // Итоги науки и техн. Сер. Современ. мат. и ее прил. Темат. обз. 1999. Т. 67. С. 161–182.
9. Beklaryan L. A. Equations of Advanced-Retarded Type and Solutions of Traveling-Wave Type for Infinite-Dimensional Dynamic Systems // J. of Math. Sci. 2004. V. 124. N. 4. P. 5098–5109.
10. Beklaryan L. A., Khachatryan N. K. Traveling wave type solutions in dynamic transport models // Functional differential equations. 2006. V. 13. N. 2. P. 125–155.
11. Бекларян Л. А. Введение в теорию функционально-дифференциальных уравнений. Групповой подход. М.: Факториал Пресс, 2007. 286 с.
12. Бекларян Л. А. О квазибегущих волнах // Матем. сб. 2010. Т. 201. № 12. С. 21–68.
13. Бекларян Л. А., Хачатрян Н. К. Об одном классе динамических моделей грузоперевозок // Ж. вычисл. матем. и матем. физ. 2013. Т. 53. № 10. С. 1649–1667.
14. Бекларян Л. А. Квазибегущие волны как естественное расширение класса бегущих волн // Вестн. Тамбовского гос. ун-та. 2014. Т. 19. № 2. С. 331–340.
15. Бекларян Л. А., Бекларян А. Л. Вопрос существования ограниченных солитонных решений в задаче о продольных колебаниях упругого бесконечного стержня в поле с сильно нелинейным потенциалом // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 12. С. 2024–2039.

16. Бекларян А. Л., Бекларян Л. А. Вопрос существования ограниченных солитонных решений в задаче о продольных колебаниях упругого бесконечного стержня в поле с нелинейным потенциалом общего вида // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 6. С. 933–950.
17. Beklaryan L. A., Beklaryan A. L., Akopov A. S. Soliton Solutions for the Manhattan Lattice // Inter. J. of Appl. Math. 2023. V. 36. N. 4. P. 569–589.
18. Akopov A. S., Beklaryan L. A. Traffic Improvement in Manhattan Road Networks With the Use of Parallel Hybrid Biobjective Genetic Algorithm // IEEE Access. 2024. V. 12. P. 19532–19552.
19. Григорчук Р. И., Курчанов П. Ф. Некоторые вопросы теории групп, связанные с геометрией // Итоги науки и техн. Совр. пробл. матем. Фундам. напр. 1990. Т. 58. С. 191–256.
20. Keener J. P. Propagation and Its Failure in Coupled Systems of Discrete Excitable Cells // SIAM J. of Appl. Math. 1987. V. 47. N. 3. P. 556–572.
21. Zinner B. Existence of Traveling Wavefront Solutions for the Discrete Nagumo Equation // J. of Different. Equat. 1992. V. 96. N. 1. P. 1–27
22. Cahn J. W., Mallet-Paret J., Van Vleck E. S. Traveling Wave Solutions for Systems of ODEs on Two-Dimensional Spatial Lattice // SIAM J. on Appl. Math. 1998. V. 59. N. 2. P. 455–493.
23. Mallet-Paret J. The Fredholm Alternative for Functional-Differential Equations Mixed Type // J. of Dynamic. and Different. Equat. 1999. V. 11. N. 1. P. 1–47.
24. Mallet-Paret J. The Global Structure of Traveling Waves in Spatially Discrete Dynamical Systems // J. of Dynamic. and Different. Equat. 1999. V. 11. N. 1. P. 49–127.
25. Эльсгольц Л. Э., Норкин С. Б. Введение в теорию дифференциальных уравнений с отклоняющимся аргументом. М.: Наука, 1971. 296 с.
26. Hale J. K. Theory of Functional Differential Equations. New York, Springer, 1977. 366 p.
27. Hale J. K., Verduyn L. S. Introduction to Functional-Differential Equations. New York: Springer, 1993. 450 p.
28. Azbelev N. V., Maksimov V. P., Rakhmatullina L. F. Introduction to the Theory of Functional Differential Equations: Methods and Applications. NY, Hindawi, 2007.

DUALISM IN THE SOLITON SOLUTION THEORY

L. A. Beklaryan^{a,*}, A. L. Beklaryan^{b,**}

^a 117418, Moscow, Nakhimovsky Ave, 47, Central Economics and Mathematics Institute, RAS

^b 101000, Moscow, Myasnitskaya Str., 20, National Research University Higher School of Economics

*e-mail: lbeklaryan@outlook.com

**e-mail: abeklaryan@hse.ru

Received 12 September, 2023

Revised 23 February, 2024

Accepted 05 March, 2024

Abstract. The article is dedicated to dualism of the soliton solution theories and solutions of functional differential point form equations. The basics of the formalism of such dualism are presented with the central element being the idea of soliton bunch and the dual pair: function-operator. Within such an approach, it is possible to describe the whole space of soliton solutions with a set characteristic as well as their asymptotics both in space and time. An example of traffic flow on the Manhattan grid shows the whole family of limited soliton solutions.

Keywords: Manhattan grid, soliton solutions, functional differential equations.

ОБОБЩЕНИЕ МЕТОДА МАТРИЦ РАССЕЯНИЯ НА ЗАДАЧИ В НЕЛИНЕЙНЫХ ДИСПЕРГИРУЮЩИХ СРЕДАХ¹⁾

© 2024 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 07.02.2023 г.

Переработанный вариант 07.03.2024 г.

Принята к публикации 03.04.2024 г.

В последние годы большое внимание уделяется устройствам интегральной фотоники на основе нелинейных сред. Предложено обобщение метода матриц переноса на задачи в плоскопараллельных слоистых средах с квадратичной и кубической нелинейностью. Падающее излучение может быть как монохроматической волной, так и некогерентным импульсом. Ранее такие задачи удавалось решать только сеточными методами. Предлагаемые подходы существенно расширяют область применимости матричных методов и кардинально превосходят по экономичности известные сеточные методы. Библ. 16. Фиг. 4.

Ключевые слова: уравнения Максвелла, слоистые среды, нелинейные среды, условия сопряжения, дисперсия вещества.

DOI: 10.31857/S0044466924070081, EDN: xilgen

1. ВВЕДЕНИЕ

1.1. Приложения

Интегральной фотоникой называется направление оптики, посвященное разработке наноразмерных устройств, позволяющих управлять излучением в ближнем ИК- и видимом диапазоне. Многие такие устройства сводятся к набору плоскопараллельных либо клиновидных пластин, на которые нормально либо наклонно падает электромагнитное излучение. Важной задачей является нахождение таких толщин слоев, которые обеспечивают заданный спектральный состав отраженного либо прошедшего излучения (см. [1], [2]).

В последние годы большое внимание уделяется устройствам интегральной фотоники на основе нелинейных сред. Напомним (см. [3]), что нелинейной называют среду, у которой диэлектрическая восприимчивость $\chi(\mathbf{E})$ зависит от величины электрического поля $\mathbf{E}$. Традиционно рассматривают квадратичные и кубические среды. Простейшие материальные уравнения для них имеют вид

$$\mathbf{P} = \chi^I \mathbf{E} + \chi^{II} \mathbf{E}^2, \quad (1)$$

и

$$\mathbf{P} = \chi^I \mathbf{E} + \chi^{III} \mathbf{E}^3 \quad (2)$$

соответственно. Здесь $\mathbf{P}$ — модуль вектора поляризации вещества.

Нелинейная интегральная фотоника вызывает большой интерес, поскольку такие устройства представляются перспективной базой для создания различных элементов оптической вычислительной техники (см. [4], [5]). Среди них, например, переключатель с гистерезисом (элемент памяти), логический элемент “и” или “или”, трансфазорный усилитель (оптический транзистор), ограничитель оптической мощности.

¹⁾ Работа выполнена при финансовой поддержке РНФ (код проекта № 22-71-00070).

1.2. Вычислительные методы

1. Приведенные приложения описываются начально-краевой задачей для нестационарных уравнений Максвелла. Отметим, что нестационарность является принципиально важной. Так, при прохождении монохроматической волны через нелинейную среду генерируется излучение с кратными (удвоенной либо утроенной) частотами. При прохождении немонохроматического импульса картина еще усложняется. При этом становится значимой *частотная дисперсия* среды, т.е. зависимость материальных параметров от частоты. Это представляет значительную трудность для численного расчета. Второй трудностью является наличие *границ раздела сред*, на которых материальные параметры изменяются скачком. На этих границах записывают условия сопряжения (в англоязычной литературе они называются *interface conditions*). В результате решение оказывается обобщенным, т.е. имеет ограниченную гладкость.

2. Подробный обзор известных подходов приведен в [6]. Многие известные алгоритмы для решения уравнений Максвелла (матричные методы, методы дискретных источников, модальные методы, псевдоспектральные методы, методы мультипольной аппроксимации, методы конечных разностей и конечных элементов в частотной области) дают хорошие результаты для достаточно узких классов задач. Так, матричные методы применимы, если рассеиватель является плоскопараллельным, среды — кусочно-однородными, а излучение — монохроматическим.

Ранее нелинейные задачи удавалось решать только сеточными методами — методами конечных разностей (FD- time domain), конечных элементов (FE- time domain) и конечных объемов (FV- time domain) во временной области. Практически все они сводятся к схеме типа “крест”, в которой электрическое и магнитное поля относятся к целым и полуцелым пространственным ячейкам и временным слоям.

Алгоритмы FDTD, FETD, FVTD имеют следующие принципиальные недостатки.

1. Ни в одном из них не реализованы физически корректные условия сопряжения на границах раздела сред. Поэтому методы данного класса не являются полностью консервативными. Это приводит к резкому ухудшению точности и появлению нефизичных пилообразных осцилляций численного решения вблизи границ раздела (см. [7]).

2. Известные способы учета частотной дисперсии вносят существенную неустранимую погрешность (см. [8], [9]).

3. Эти недостатки удалось преодолеть в предложенных нами бикомпактных схемах и методе спектрального разложения (см. [10]). Это двухточечные полностью консервативные схемы, основанные на сеточной аппроксимации интегральных уравнений Максвелла и условий сопряжения на границах раздела сред. Доказано, что бикомпактные схемы сходятся как на гладких, так и на обобщенных решениях. В методе спектрального разложения немонохроматическая задача сводится к набору монохроматических задач с помощью сеточного преобразования Фурье. Это позволяет учитывать произвольный закон частотной дисперсии. Расчеты тестовых и прикладных задач (см. [11]) убедительно показали преимущества бикомпактных схем и метода спектрального разложения по сравнению с другими известными подходами.

4. В настоящей статье предложено обобщение метода спектрального разложения и метода матриц переноса на задачи в слоистых средах с квадратичной и кубической нелинейностью. Предлагаемые подходы существенно расширяют область применимости матричных методов. Ранее данные задачи удавалось решать только сеточными методами. Для финитных импульсов квадратуры в методе спектрального разложения сходятся по экспоненциальному закону. Поэтому предлагаемые методы являются сверхбыстрым. Их трудоемкость кардинально меньше таковой у известных сеточных методов.

2. КВАДРАТИЧНЫЕ СРЕДЫ

2.1. Монохроматическая задача

1. Рассмотрим слоистую структуру, состоящую из Q плоскопараллельных пластин общей толщиной a . Ориентируем координатную ось z перпендикулярно пластинам; оси x и y расположены в плоскости пластин. Обозначим координаты границ слоев через $0 = \xi_0 < \xi_1 < \dots < \xi_Q = a$. При $z < 0$

и $z > a$ расположены полубесконечные диэлектрические среды. Их диэлектрические проницаемости таковы же, как в крайних слоях $0 < z < \xi_1$ и $\xi_{Q-1} < z < a$ соответственно.

2. Пусть пластины состоят из немагнитных диэлектриков. Часть из них является линейными, часть — квадратичными средами. Будем считать все среды изотропными. Тогда вектор поляризации $\mathbf{P}$ параллелен вектору электрического поля $\mathbf{E}$. Материальные уравнения квадратичных слоев имеют вид (1). Для линейных слоев полагаем $\chi^{II} = 0$. Значения $\chi^{I,II}$ на границах раздела меняются скачком. Материальные параметры зависят от частоты, т.е. среды являются частотно диспергирующими.

3. Пусть на описанную структуру из $z = -\infty$ нормально падает плоская электромагнитная волна; а волна, падающая из $z = +\infty$, отсутствует. Приближение плоской волны хорошо работает на оси оптического пучка. При этом мы пренебрегаем краевыми эффектами на границах пучка (дифракционная расходимость, сужение пучка из-за самофокусировки и т.д.)

Пусть падающая волна содержит только одну частоту ω . Такой случай назовем *монохроматическим*. При прохождении такого излучения через квадратичную среду формируется волна с удвоенной частотой 2ω . Обе эти волны испытывают переотражения на границах раздела сред. Будем считать, что показатели преломления для частот ω и 2ω различаются. В этом случае говорят, что условие волнового синхронизма не выполнено.

В монохроматической задаче мы рассматриваем решение, соответствующее установившимся колебаниям. При этом информация о начальных условиях теряется. Поэтому мы не включили их в постановку задачи.

4. В каждом q -м слое электрическое поле удовлетворяет скалярному волновому уравнению

$$\frac{\partial^2 E_q}{\partial z^2} - \frac{1}{c^2} (1 + 4\pi\chi_q^I) \frac{\partial^2 E_q}{\partial t^2} = \frac{4\pi}{c^2} \frac{\partial^2}{\partial t^2} (\chi_q^{II} E_q^2), \quad 1 \leq q \leq Q. \quad (3)$$

Используется система единиц СГС. Подчеркнем, что уравнение (3) является нестационарным, даже если падающее излучение содержит только одну частоту.

Для каждой спектральной компоненты магнитное поле выражается через электрическое по формуле

$$H_q = \frac{c}{i\omega} \frac{\partial E_q}{\partial z}. \quad (4)$$

Для волны с удвоенной частотой нужно заменить $\omega \rightarrow 2\omega$.

На границах раздела сред выполняются условия сопряжения

$$E_q|_{z=\xi_q} = E_{q+1}|_{z=\xi_q}, \quad H_q|_{z=\xi_q} = H_{q+1}|_{z=\xi_q}, \quad 1 \leq q \leq Q-1. \quad (5)$$

Эти равенства имеют место для каждой спектральной компоненты.

На границах расчетной области поставим условия излучения (см. [12]) для каждой спектральной компоненты. Для исходной гармоники они принимают вид

$$\frac{\partial E_0}{\partial z} + ik_0^\omega E_0 = 2ik_0\mathcal{E}, \quad \frac{\partial E_Q}{\partial z} - ik_Q^\omega E_Q = 0. \quad (6)$$

Здесь

$$k_q^\omega = n_q^\omega \omega / c \quad (7)$$

есть волновое число в q -м слое на частоте ω . Линейный показатель преломления равен $n_q^\omega = \sqrt{1 + 4\pi\chi^I(\omega)}$. Для удвоенной частоты имеем

$$\frac{\partial E_0}{\partial z} + ik_0^{2\omega} E_0 = 0, \quad \frac{\partial E_Q}{\partial z} - ik_Q^{2\omega} E_Q = 0. \quad (8)$$

Волновое число для удвоенной частоты $k_q^{2\omega}$ определяется равенством (7) с заменой $\omega \rightarrow 2\omega$.

2.2. Нулевое приближение

1. В реальных средах вклад нелинейности в поляризацию вещества является малым: $\chi^{II} E^2 / \chi^I E \sim E / E_{\text{ат}}$ для квадратичных сред и $\chi^{III} E^2 / \chi^I E^2 \sim (E / E_{\text{ат}})^2$ для кубических. Здесь $E_{\text{ат}} \approx 10^7$ СГСЭ —

характерная величина внутриатомного электрического поля (см. [3]). Для типичных источников излучения $E/E_{\text{ат}} \sim 10^{-4}$. Поэтому к описанной нелинейной задаче можно применить метод последовательных приближений.

2. Нулевое приближение соответствует пространственно однородной среде, причем $\chi^{II} = 0$. Тогда правая часть уравнения (3) равна нулю. Согласно граничному условию (6), в падающем излучении присутствует только частота ω . Поэтому нулевое приближение будет содержать только эту частоту $E^0(z, t) = E^0(z)e^{-i\omega t}$. Тогда в (3) можно заменить $\partial^2/\partial t^2 \rightarrow -\omega^2$ и получить уравнение для комплексной амплитуды поля $E^0(z)$

$$\frac{\partial^2 E_q^0}{\partial z^2} + (k_q^\omega)^2 E_q^0 = 0, \quad 1 \leq q \leq Q. \quad (9)$$

3. Решение этой задачи хорошо известно. В каждом слое оно представляется суперпозицией плоских монохроматических волн, распространяющихся в обоих направлениях оси z . Запишем это решение в q -слое:

$$E_q^0(z) = A_q \exp(ik_q^\omega z) + B_q \exp(-ik_q^\omega z), \quad 1 \leq q \leq Q. \quad (10)$$

Далее вычислим магнитное поле согласно (4):

$$H_q^0(z) = n_q^\omega [A_q \exp(ik_q^\omega z) - B_q \exp(-ik_q^\omega z)], \quad 1 \leq q \leq Q. \quad (11)$$

4. Из условий излучения (6) следует, что

$$A_0 = \mathcal{E}, \quad B_Q = 0. \quad (12)$$

Подставим (10), (11) в условия сопряжения (5)

$$A_q \exp(ik_q^\omega \xi_q) + B_q \exp(-ik_q^\omega \xi_q) = A_{q+1} \exp(ik_{q+1}^\omega \xi_q) - B_{q+1} \exp(-ik_{q+1}^\omega \xi_q), \quad (13)$$

$$n_q^\omega [A_q \exp(ik_q^\omega \xi_q) - B_q \exp(-ik_q^\omega \xi_q)] = n_{q+1}^\omega [A_{q+1} \exp(ik_{q+1}^\omega \xi_q) - B_{q+1} \exp(-ik_{q+1}^\omega \xi_q)]. \quad (14)$$

Здесь $1 \leq q \leq Q - 1$. Равенства (12)–(14) есть система линейных алгебраических уравнений относительно коэффициентов решения (10), (11). Она содержит $2Q - 2$ уравнения относительно $2Q - 2$ неизвестных. Ее можно решить, например, методом Гаусса для ленточных матриц.

Таким образом, решение в нулевом приближении построено.

5. Полученное решение является реализацией хорошо известного метода матриц переноса. Отметим, что классическая формулировка этого метода (см. [13]) основана на матрицах, выражающих набег фазы полей в q -м слое:

$$M_q = \begin{pmatrix} \cos k_q^\omega d_q & i(n_q^\omega)^{-1} \sin k_q^\omega d_q \\ in_q^\omega \sin k_q^\omega d_q & \cos k_q^\omega d_q \end{pmatrix}. \quad (15)$$

Через эти матрицы выражаются амплитудные коэффициенты отражения и преломления и значения поля E в границах раздела сред. Последние позволяют без труда найти поля внутри слоев.

Приведенная нами форма метода матриц рассеяния более предпочтительна в контексте настоящей работы, поскольку она непосредственно обобщается на первое и последующие приближения.

2.3. Первое приближение

1. Рассмотрим q -й слой. Пусть $\chi_q^{II} \neq 0$. Найденное выше решение (10) есть комплексная амплитуда электрического поля. Восстановим временную зависимость, домножая на $\exp(-i\omega t)$. Подставим это выражение в правую часть уравнения (3) и получим уравнение для поправки первого приближения

$$\begin{aligned} & \frac{\partial^2}{\partial z^2} \delta E_q^1 - \frac{1}{c^2} (1 + 4\pi\chi_q^I) \frac{\partial^2}{\partial t^2} \delta E_q^1 = \\ & = \frac{4\pi}{c^2} \chi_q^{II} [A_q \exp(ik_q^\omega z) + B_q \exp(-ik_q^\omega z)]^2 \exp(-2i\omega t). \end{aligned} \quad (16)$$

Здесь значения восприимчивости $\chi^{I,II}$ относятся к частоте 2ω .

Поправка δE_q^1 считается малой по сравнению с нулевым приближением. Поэтому среду для нее можно считать линейной. Влияние нелинейности является эффектом второго порядка малости, которым можно пренебречь. Правая часть уравнения (16) соответствует “перекачке” энергии из основной гармоники в кратную.

2. Видно, что в правой части (16) имеется только удвоенная частота 2ω . Поэтому решение может зависеть от времени только как $\exp(-2i\omega t)$, т.е. $\delta E_q^1(z, t) = \delta E_q^1(z) \exp(-2i\omega t)$. Заменяя $\partial^2/\partial t^2 \rightarrow -4\omega^2$, получим уравнение для комплексной амплитуды первой поправки

$$\frac{\partial^2}{\partial z^2} \delta E_q^1 + (k_q^{2\omega})^2 \delta E_q^1 = -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} [A_q^2 \exp(2ik_q^\omega z) + 2A_q B_q + B_q^2 \exp(-2ik_q^\omega z)]. \quad (17)$$

Здесь коэффициенты A_q, B_q известны из нулевого приближения.

3. Построим общее решение уравнения (17). Оно состоит из общего решения однородного уравнения и частного решения неоднородного уравнения. Первое имеет вид

$$C_q \exp(ik_q^{2\omega} z) + D_q \exp(-ik_q^{2\omega} z), \quad (18)$$

где C_q, D_q — произвольные постоянные.

Частное решение будем искать в виде

$$\alpha_q \exp(2ik_q^\omega z) + \beta_q \exp(-2ik_q^\omega z) + \gamma_q, \quad (19)$$

где $\alpha_q, \beta_q, \gamma_q$ — некоторые коэффициенты. Чтобы найти их, подставим (19) в (17). После несложных выкладок получим

$$\begin{aligned} \alpha_q &= -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{A_q^2}{(k_q^{2\omega})^2 - 4(k_q^\omega)^2}, \\ \beta_q &= -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{B_q^2}{(k_q^{2\omega})^2 - 4(k_q^\omega)^2}, \\ \gamma_q &= -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{2A_q B_q}{(k_q^{2\omega})^2}. \end{aligned} \quad (20)$$

Поскольку условие волнового синхронизма не выполнено, то $k_q^{2\omega} \neq 2k_q^\omega$ ни в одном слое. Поэтому знаменатели в (20) не обращаются в нуль.

Таким образом, амплитуда поправки первого порядка малости для электрического поля равна

$$\delta E_q^1 = C_q \exp(ik_q^{2\omega} z) + D_q \exp(-ik_q^{2\omega} z) + \alpha_q \exp(2ik_q^\omega z) + \beta_q \exp(-2ik_q^\omega z) + \gamma_q. \quad (21)$$

Вычислим поправку первого порядка для магнитного поля

$$\delta H_q^1 = n_q^{2\omega} [C_q \exp(ik_q^{2\omega} z) - D_q \exp(-ik_q^{2\omega} z)] + n_q^\omega [\alpha_q \exp(2ik_q^\omega z) - \beta_q \exp(-2ik_q^\omega z)]. \quad (22)$$

4. Определим коэффициенты C_q, D_q . Из условий излучения (8) следует, что

$$C_1 = 0, \quad D_Q = 0. \quad (23)$$

Из условий сопряжения при $z = \xi_q, 1 \leq q \leq Q-1$, получаем

$$\begin{aligned} C_q \exp(ik_q^{2\omega} \xi_q) + D_q \exp(-ik_q^{2\omega} \xi_q) - \\ - C_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) - D_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q) = \\ = \alpha_{q+1} \exp(2ik_{q+1}^\omega \xi_q) + \beta_{q+1} \exp(-2ik_{q+1}^\omega \xi_q) + \gamma_{q+1} - \\ - \alpha_q \exp(2ik_q^\omega \xi_q) - \beta_q \exp(-2ik_q^\omega \xi_q) - \gamma_q, \end{aligned} \quad (24)$$

$$n_q^{2\omega} [C_q \exp(ik_q^{2\omega} \xi_q) - D_q \exp(-ik_q^{2\omega} \xi_q)] -$$

$$\begin{aligned}
& -n_{q+1}^{2\omega} [C_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) - D_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q)] = \\
& = n_{q+1}^{\omega} [\alpha_{q+1} \exp(2ik_{q+1}^{\omega} \xi_q) - \beta_{q+1} \exp(-2ik_{q+1}^{\omega} \xi_q)] - \\
& \quad - n_q^{\omega} [\alpha_q \exp(2ik_q^{\omega} \xi_q) - \beta_q \exp(-2ik_q^{\omega} \xi_q)]. \quad (25)
\end{aligned}$$

Здесь предполагается, что среды по обе границы раздела являются квадратичными без волнового синхронизма. В частности, если одна из сред (например, $(q+1)$ -я) является линейной, то для нее $\chi^{II} = 0$, и коэффициенты $\alpha_{q+1} = \beta_{q+1} = 0$.

Матрица линейной системы (23)–(25) аналогична матрице линейной системы (12)–(14) для коэффициентов нулевого приближения. Она также легко обращается методом Гаусса для ленточных матриц.

5. Таким образом, коэффициенты в выражении для поправки первого порядка вычислены. В изотропном случае поля нулевого приближения и поправки первого порядка имеют одинаковую поляризацию. Поэтому окончательно выражения для полей в первом приближении имеют вид

$$E_q^1(z, t) = E_q^0(z) e^{-i\omega t} + \delta E_q^1(z) e^{-2i\omega t}, \quad H_q^1(z, t) = H_q^0(z) e^{-i\omega t} + \delta H_q^1(z) e^{-2i\omega t}. \quad (26)$$

Полученное решение является новым.

2.4. Волновой синхронизм

1. Наибольший интерес для практики представляет случай, когда фазовые скорости распространения света на основной и на кратной частотах совпадают. В этом случае говорят, что выполняется условие *волнового синхронизма*. При этом происходит конструктивная интерференция этих волн, и амплитуда кратной гармоники нарастает по мере ее распространения в среде.

2. Реализовать это условие можно с помощью двойного лучепреломления в анизотропном кристалле (см. [3]). В отрицательном одноосном кристалле скорости волн основной частоты и второй гармоники могут быть одинаковы, если они имеют разные поляризации: основная гармоника является обыкновенной волной, а кратная — необыкновенной.

Формально задачи в анизотропных средах выходят далеко за пределы рассмотренных простейших постановок. Однако в описанном выше частном случае направления поляризации основной гармоники (т.е. нулевого приближения) и удвоенной гармоники (т.е. поправки первого порядка) соответствуют собственным состояниям поляризации кристалла, поэтому их поляризации не изменяются при распространении волны. Значит, амплитуды этих волн описываются скалярными уравнениями, и к ним можно применить предложенный подход.

3. Пусть q -й слой рассеивателя выполнен из отрицательного одноосного анизотропного кристалла. Обозначим главные показатели преломления через n_o (для обыкновенной волны) и n_e (для необыкновенной). Пусть частотная дисперсия является нормальной: с ростом частоты показатель преломления увеличивается.

Пусть угол между волновым вектором $\mathbf{k}$ и главной осью кристалла таков, что выполняется условие волнового синхронизма, т.е. $n_o^{\omega} = n_e^{2\omega}$. При этом падение излучения на рассеиватель может оказаться как нормальным, так и наклонным. Это зависит от взаимной ориентации главной оси кристалла и плоскостей, образующих границы слоев. Ниже приведены формулы для случая нормального падения. Обобщение на наклонное падение описано далее в разд. 5.

4. Амплитуда поправки первого порядка описывается уравнением (17), в котором нужно положить $k_q^{2\omega} = 2k_q^{\omega}$. Общее решение однородного уравнения по-прежнему выражается формулой (18). Частное решение неоднородного уравнения будем искать в следующем виде:

$$\alpha z \exp(2ik_q^{\omega} z) + \beta z \exp(-2ik_q^{\omega} z) + r. \quad (27)$$

Коэффициенты α , β , γ определяются подстановкой этого решения в уравнение (17). Они имеют вид

$$\alpha_q = -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{A_q^2}{4ik_q^{\omega}}, \quad \beta_q = -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{B_q^2}{4ik_q^{\omega}}, \quad \gamma_q = -\frac{16\pi}{c^2} \omega^2 \chi_q^{II} \frac{A_q B_q}{2ik_q^{\omega}}. \quad (28)$$

Приведем окончательные формулы поправок первого порядка малости для амплитуд электрического и магнитного полей

$$\delta E_q^1 = (C_q + \alpha_q z) \exp(ik_q^{2\omega} z) + (D_q + \beta_q z) \exp(-ik_q^{2\omega} z) + \gamma_q, \quad (29)$$

$$\begin{aligned} \delta H_q^1 = & (n_e)_q^{2\omega} [C_q \exp(ik_q^{2\omega} z) - D_q \exp(-ik_q^{2\omega} z)] + \\ & + \frac{c}{2i\omega} [\alpha_q \exp(ik_q^{2\omega} z) + \beta_q \exp(-ik_q^{2\omega} z)] + \\ & + (n_e)_q^{2\omega} z [\alpha_q \exp(ik_q^{2\omega} z) - \beta_q \exp(-ik_q^{2\omega} z)]. \end{aligned} \quad (30)$$

Напомним, что временная зависимость этих поправок выражается множителем $\exp(-2i\omega t)$.

5. Остается определить коэффициенты C_q, D_q . Найдем их из граничных условий и условий сопряжения. Имеем

$$C_0 = 0, \quad D_Q = 0, \quad (31)$$

$$\begin{aligned} C_q \exp(ik_q^{2\omega} \xi_q) + D_q \exp(-ik_q^{2\omega} \xi_q) - \\ - C_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) + D_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q) = \\ = \alpha_{q+1} \xi_q \exp(ik_{q+1}^{2\omega} \xi_q) + \beta_{q+1} \xi_q \exp(-ik_{q+1}^{2\omega} \xi_q) + \gamma_{q+1} - \\ - \alpha_q \xi_q \exp(ik_q^{2\omega} \xi_q) - \beta_q \xi_q \exp(-ik_q^{2\omega} \xi_q) - \gamma_q, \end{aligned} \quad (32)$$

$$\begin{aligned} (n_e)_q^{2\omega} [C_q \exp(ik_q^{2\omega} \xi_q) - D_q \exp(-ik_q^{2\omega} \xi_q)] - \\ - (n_e)_{q+1}^{2\omega} [C_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) - D_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q)] = \\ = \frac{ic}{2\omega} [\alpha_q \exp(ik_q^{2\omega} \xi_q) + \beta_q \exp(-ik_q^{2\omega} \xi_q) - \\ - \alpha_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) - \beta_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q)] + \\ + (n_e)_{q+1}^{2\omega} \xi_q [\alpha_{q+1} \exp(ik_{q+1}^{2\omega} \xi_q) - \beta_{q+1} \exp(-ik_{q+1}^{2\omega} \xi_q)] - \\ - (n_e)_q^{2\omega} \xi_q [\alpha_q \exp(ik_q^{2\omega} \xi_q) - \beta_q \exp(-ik_q^{2\omega} \xi_q)]. \end{aligned} \quad (33)$$

Если одна из сред линейна, то для нее коэффициенты α, β, γ обращаются в нуль.

Таким образом, решение полностью построено. Оно ранее было неизвестно.

3. КУБИЧЕСКИЕ СРЕДЫ

3.1. Постановка задачи

1. Пусть среда является изотропной, и ее материальное уравнение описывается равенством (2). Постановка *монохроматической* задачи в слоистой среде с кубической нелинейностью практически дословно повторяет п. 2.1 с одним отличием. В правой части уравнения (3) нужно заменить $\chi_q^{II} E^2 \rightarrow \chi_q^{III} E^3$.

2. При прохождении излучения в такой среде генерируется волна с утроенной частотой. Будем считать, что из-за дисперсии показатели преломления для частот ω и 3ω различаются, т.е. условие волнового синхронизма не выполнено. Для утроенной гармоники отсутствуют волны, падающие из $z = \pm\infty$.

3.2. Последовательные приближения

1. Нулевое приближение соответствует случаю $\chi_q^{III} = 0$. Поэтому его вывод полностью повторяет п. 2.2, а поля даются формулами (10), (11).

2. Построим поправку первого порядка малости. Пусть в q -м слое $\chi_q^{III} \neq 0$. Подставим решение для нулевого приближения в правую часть волнового уравнения. Его правая часть зависит от времени

как $\exp(-3i\omega t)$. Предполагая такую же временную зависимость у поправки первого приближения, получим уравнение для ее амплитуды $\delta E_q^1(z)$

$$\frac{\partial^2}{\partial z^2} \delta E_q^1 + (k_q^{3\omega})^2 \delta E_q^1 = -\frac{36\pi}{c^2} \omega^3 \chi_q^{III} [A_q \exp(ik_q^\omega z) + B_q \exp(-ik_q^\omega z)]^3. \quad (34)$$

Представим общее решение как сумму общего решения однородного уравнения

$$C_q \exp(ik_q^{3\omega} z) + D_q \exp(-ik_q^{3\omega} z) \quad (35)$$

и частного решения неоднородного уравнения

$$\alpha_q \exp(3ik_q^\omega z) + \beta_q \exp(ik_q^\omega z) + \gamma_q \exp(-ik_q^\omega z) + \delta_q \exp(-3ik_q^\omega z). \quad (36)$$

Коэффициенты $\alpha_q, \beta_q, \gamma_q, \delta_q$ имеют вид

$$\begin{aligned} \alpha_q &= -\frac{36\pi}{c^2} \omega^3 \chi_q^{III} \frac{A_q^3}{(k_q^{3\omega})^2 - 9(k_q^\omega)^2}, \quad \beta_q = -\frac{36\pi}{c^2} \omega^3 \chi_q^{III} \frac{3A_q^2 B_q}{(k_q^{3\omega})^2 - (k_q^\omega)^2}, \\ \gamma_q &= -\frac{36\pi}{c^2} \omega^3 \chi_q^{III} \frac{3A_q B_q^2}{(k_q^{3\omega})^2 - (k_q^\omega)^2}, \quad \delta_q = -\frac{36\pi}{c^2} \omega^3 \chi_q^{III} \frac{B_q^3}{(k_q^{3\omega})^2 - 9(k_q^\omega)^2}. \end{aligned} \quad (37)$$

Таким образом, поправки первого порядка малости для электрического и магнитного полей равны

$$\begin{aligned} \delta E_q^1 &= C_q \exp(ik_q^{3\omega} z) + D_q \exp(-ik_q^{3\omega} z) + \\ &+ \alpha_q \exp(3ik_q^\omega z) + \beta_q \exp(ik_q^\omega z) + \gamma_q \exp(-ik_q^\omega z) + \delta_q \exp(-3ik_q^\omega z), \end{aligned} \quad (38)$$

$$\begin{aligned} \delta H_q^1 &= n_q^{3\omega} [C_q \exp(ik_q^{3\omega} z) - D_q \exp(-ik_q^{3\omega} z)] + \\ &+ n_q^\omega [\alpha_q \exp(3ik_q^\omega z) - \delta_q \exp(-3ik_q^\omega z)] + \\ &+ \frac{n_q^\omega}{3} [\beta_q \exp(ik_q^\omega z) - \gamma_q \exp(-ik_q^\omega z)]. \end{aligned} \quad (39)$$

3. Коэффициенты C_q, D_q определяются подстановкой (38), (39) в условия излучения и сопряжения. Они удовлетворяют системе уравнений

$$C_1 = 0, \quad D_Q = 0, \quad (40)$$

$$\begin{aligned} C_q \exp(ik_q^{3\omega} \xi_q) + D_q \exp(-ik_q^{3\omega} \xi_q) - \\ - C_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) + D_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q) = \\ = \alpha_{q+1} \exp(3ik_{q+1}^\omega \xi_q) + \beta_{q+1} \exp(ik_{q+1}^\omega \xi_q) + \\ + \gamma_{q+1} \exp(-ik_{q+1}^\omega \xi_q) + \delta_{q+1} \exp(-3ik_{q+1}^\omega \xi_q) - \\ - \alpha_q \exp(3ik_q^\omega \xi_q) - \beta_q \exp(ik_q^\omega \xi_q) - \\ - \gamma_q \exp(-ik_q^\omega \xi_q) - \delta_q \exp(-3ik_q^\omega \xi_q), \end{aligned} \quad (41)$$

$$\begin{aligned} n_q^{3\omega} [C_q \exp(ik_q^{3\omega} \xi_q) - D_q \exp(-ik_q^{3\omega} \xi_q)] - \\ - n_{q+1}^{3\omega} [C_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) - D_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q)] = \\ = n_{q+1}^\omega [\alpha_{q+1} \exp(3ik_{q+1}^\omega \xi_q) - \delta_{q+1} \exp(-3ik_{q+1}^\omega \xi_q)] - \\ - n_q^\omega [\alpha_q \exp(3ik_q^\omega \xi_q) - \delta_q \exp(-3ik_q^\omega \xi_q)] + \\ + \frac{n_{q+1}^\omega}{3} [\beta_{q+1} \exp(ik_{q+1}^\omega \xi_q) - \gamma_{q+1} \exp(-ik_{q+1}^\omega \xi_q)] - \\ - \frac{n_q^\omega}{3} [\beta_q \exp(ik_q^\omega \xi_q) - \gamma_q \exp(-ik_q^\omega \xi_q)]. \end{aligned} \quad (42)$$

Как и ранее, если одна из сред линейна, то для нее коэффициенты $\alpha, \beta, \gamma, \delta$ равны нулю.

4. Окончательно выражения для полей в первом приближении имеют вид

$$E_q^1 = E_q^0 e^{-i\omega t} + \delta E_q^1 e^{-3i\omega t}, \quad H_q^1 = H_q^0 e^{-i\omega t} + \delta H_q^1 e^{-3i\omega t}. \quad (43)$$

Напомним, что в изотропной среде векторы нулевого приближения и поправки первого порядка малости параллельны. Полученное решение является новым.

3.3. Волновой синхронизм

В этом случае $(n_e)_q^{3\omega} = (n_o)_q^\omega$, $k_q^{3\omega} = 3k_q^\omega$. Построение решения проводится аналогично п. 2.4. Опуская несложные, но громоздкие выкладки, приведем окончательный результат.

Поправка первого порядка малости для электрического и магнитного полей имеет следующий вид:

$$\delta E_q^1 = (C_q + \alpha_q z) \exp(ik_q^{3\omega} z) + (D_q + \delta_q z) \exp(-ik_q^{3\omega} z) + \beta_q \exp(ik_q^\omega z) + \gamma_q \exp(-ik_q^\omega z), \quad (44)$$

$$\begin{aligned} \delta H_q^1 = & (n_e)_q^{3\omega} [C_q \exp(ik_q^{3\omega} z) - D_q \exp(-ik_q^{3\omega} z)] + \\ & + \frac{c}{3i\omega} [\alpha_q \exp(ik_q^{3\omega} z) + \delta_q \exp(-ik_q^{3\omega} z)] + \\ & + (n_e)_q^{3\omega} z [\alpha_q \exp(ik_q^{3\omega} z) - \delta_q \exp(-ik_q^{3\omega} z)] + \\ & + \frac{(n_e)_q^{3\omega}}{3} [\beta_q \exp(ik_q^\omega z) - \gamma_q \exp(-ik_q^\omega z)]. \end{aligned} \quad (45)$$

Коэффициенты C_q , D_q определяются из системы уравнений

$$\begin{aligned} C_q \exp(ik_q^{3\omega} \xi_q) + D_q \exp(-ik_q^{3\omega} \xi_q) - \\ - C_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) + D_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q) = \\ = \alpha_{q+1} \xi_q \exp(ik_{q+1}^{3\omega} \xi_q) + \beta_{q+1} \exp(ik_{q+1}^\omega \xi_q) + \\ + \gamma_{q+1} \exp(-ik_{q+1}^\omega \xi_q) + \delta_{q+1} \xi_q \exp(-ik_{q+1}^{3\omega} \xi_q) - \\ - \alpha_q \xi_q \exp(ik_q^{3\omega} \xi_q) - \beta_q \exp(ik_q^\omega \xi_q) - \\ - \gamma_q \exp(-ik_q^\omega \xi_q) - \delta_q \xi_q \exp(-ik_q^{3\omega} \xi_q), \end{aligned} \quad (46)$$

$$\begin{aligned} (n_e)_q^{3\omega} [C_q \exp(ik_q^{3\omega} \xi_q) - D_q \exp(-ik_q^{3\omega} \xi_q)] - \\ - (n_e)_{q+1}^{3\omega} [C_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) - D_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q)] = \\ = \frac{c}{3i\omega} [\alpha_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) + \delta_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q)] + \\ + (n_e)_{q+1}^{3\omega} \xi_q [\alpha_{q+1} \exp(ik_{q+1}^{3\omega} \xi_q) - \delta_{q+1} \exp(-ik_{q+1}^{3\omega} \xi_q)] + \\ + \frac{(n_e)_{q+1}^{3\omega}}{3} [\beta_{q+1} \exp(ik_{q+1}^\omega \xi_q) - \gamma_{q+1} \exp(-ik_{q+1}^\omega \xi_q)] - \\ - \frac{c}{3i\omega} [\alpha_q \exp(ik_q^{3\omega} \xi_q) + \delta_q \exp(-ik_q^{3\omega} \xi_q)] - \\ - (n_e)_q^{3\omega} \xi_q [\alpha_q \exp(ik_q^{3\omega} \xi_q) - \delta_q \exp(-ik_q^{3\omega} \xi_q)] - \\ - \frac{(n_e)_q^{3\omega}}{3} [\beta_q \exp(ik_q^\omega \xi_q) - \gamma_q \exp(-ik_q^\omega \xi_q)]. \end{aligned} \quad (47)$$

Как и в случае квадратичных сред, векторы нулевого приближения и первой поправки ортогональны друг другу.

4. ВЕРИФИКАЦИЯ

1. Построенные поправки первого порядка малости удовлетворяют известным *качественными закономерностям* (см. [3]). Во-первых, для квадратичных сред поправка пропорциональна квадрату поля нулевого приближения, а для кубических сред — кубу. Во-вторых, в случае волнового синхронизма амплитуда кратной гармоники линейно растет по мере распространения волны в среде. Это непосредственно видно из формул (29) и (44).

2. Вычислим *невязку* на исходном волновом уравнении. Пусть q -й слой является квадратичной средой. Подставим решение (26) в уравнение (3). В нелинейных слоях последнее не будет выполнено тождественно, поскольку решение является приближенным. Найдем разность левой и правой частей этого уравнения. Нулевое приближение E_0 удовлетворяет уравнению (9), а поправка δE_1 — уравнению (17). Поэтому часть слагаемых в невязке сократится. Также вычислим вторую производную по t , учитывая явную зависимость нулевого приближения и поправки от времени. Окончательно невязка примет вид

$$\Psi_q^1(z, t) = \frac{4\pi}{c^2} \chi_q^{II} \omega^2 \left[9E_q^0(z) \delta E_q^1(z) e^{-3i\omega t} + 16 (\delta E_q^1(z))^2 e^{-4i\omega t} \right]. \quad (48)$$

3. Видно, что невязка содержит размерные величины и, тем самым, зависит от выбранной системы единиц. Это неудобно для количественных оценок. Введем *относительную невязку*, нормированную на левую часть уравнения. Рассмотрим

$$\tilde{\Psi}_q^1 = \frac{\Psi_q}{\|(E_q^1)_{zz}\| + n_q^2/c^2 \|(E_q^1)_{tt}\|} \approx \frac{c^2}{n_q^2 \omega^2} \frac{\Psi_q}{2\|E_q^0 + 4\delta E_q^1\|}. \quad (49)$$

Здесь учтена характерная зависимость решения от времени и координаты $(E_q^0)_{tt} = -\omega^2 E_q^0$, $(E_q^0)_{zz} \sim -k_q^2 E_q^0$, $(\delta E_q^1)_{tt} = -4\omega^2 \delta E_q^1$, $(\delta E_q^1)_{zz} \sim -4k_q^2 \delta E_q^1$. Подставляя (48) в (49), получим

$$|\tilde{\Psi}_q^1| \leq \frac{2\pi \chi_q^{II}}{n_q^2} \frac{9|E_q^0(z) \delta E_q^1(z)| + 16|(\delta E_q^1(z))^2|}{\|E_q^0 + 4\delta E_q^1\|}. \quad (50)$$

4. Для сравнения вычислим невязку на нулевом приближении E_q^0 . Поскольку оно удовлетворяет уравнению (9), то невязка равна правой части, т.е. нелинейности в (3)

$$|\tilde{\Psi}_q^0| \leq \frac{2\pi \chi_q^{II}}{n_q^2} |E_q^0(z)|. \quad (51)$$

Поправка δE_q^1 предполагается малой по сравнению с нулевым приближением. Видно, что отношение $\|\tilde{\Psi}_q^1/\tilde{\Psi}_q^0\|$ по порядку величины составляет $\|\delta E_q^1/E_q^0\|$, т.е. тоже является величиной *первого порядка малости*. Иными словами, первое приближение точнее удовлетворяет исходному уравнению, чем нулевое. Это верифицирует правильность построенного решения. Примеры конкретных значений $|\tilde{\Psi}|$ для реальных сред приведены далее.

5. Аналогично рассматривается случай кубических сред. Нормированная невязка для первого приближения имеет вид

$$|\tilde{\Psi}_q^1| \leq 4\pi \chi_q^{III} \frac{75|E_q^0(z)^2 \delta E_q^1(z)| + 147|E_q^0(z)(\delta E_q^1(z))^2| + 81|(\delta E_q^1(z))^3|}{\|E_q^0 + 9\delta E_q^1\|}. \quad (52)$$

Нормированная невязка нулевого приближения равна

$$|\tilde{\Psi}_q^0| = \frac{2\pi \chi_q^{II}}{n_q^2} |E_q^0(z)|^2. \quad (53)$$

Нетрудно видеть, что отношение $\|\tilde{\Psi}_q^1/\tilde{\Psi}_q^0\|$ имеет порядок величины $\|\delta E_q^1/E_q^0\|^2$.

5. НАКЛОННОЕ ПАДЕНИЕ

Решение, построенное в предыдущих пунктах, можно обобщить на случай наклонного падения излучения на рассеиватель. Это обобщение основано на том, что для изотропных сред поляризация волн основной и удвоенной частот (т.е. нулевого приближения и первой поправки соответственно) по-прежнему одинакова.

При наклонном падении в матричных методах рассматривают не амплитуды полей, а их *тангенциальные проекции* на границу раздела. Кроме того, по сравнению со случаем нормального падения меняется *набег фазы* при прохождении волны через каждый слой рассеивателя.

Пусть угол падения, т.е. угол между волновым вектором $\mathbf{k}$ и нормалью к плоскостям рассеивателя, равен θ . Для учета наклонного падения в приведенных выше формулах нужно заменить $k \rightarrow k \cos \theta$ для всех спектральных компонент. При s -поляризации поле E параллельно границе раздела. Поэтому формулы для E остаются в силе. Поле H имеет как тангенциальную, так и нормальную компоненты. Поэтому в приведенных выше формулах нужно заменить $H \rightarrow H / \cos \theta$. Аналогично для p -поляризации необходимо заменить $E \rightarrow E / \cos \theta$, оставляя без изменений H .

6. НЕМОНОХРОМАТИЧЕСКИЙ СЛУЧАЙ

1. Пусть теперь на структуру, описанную в п. 2.1, падает не монохроматическая волна, а волновой пакет, содержащий спектр частот. Такой случай будем называть *немонохроматическим*. К постановке задачи нужно добавить начальное условие. Будем считать его нулевым $E|_{t=0} = 0$.

2. Ранее для линейных гиперболических задач мы предложили *метод спектрального разложения*. Напомним его суть. Выполним численное преобразование Фурье падающих волновых пакетов по формуле трапеций, используя одинаковые сетки по времени и наборы частот $\{\omega_m\}$, $\omega_{m+1} - \omega_m = \Delta\omega_m$:

$$\tilde{f}^{0,a}(\omega_m) = \sum_{k=0}^K f^{0,a}(\zeta_k) e^{-i\omega t_k} g_k \Delta\zeta_k. \quad (54)$$

Здесь $g_0 = g_K = 0.5$, $g_k = 1$, $k \neq 0, K$ — веса квадратурной формулы трапеций.

Тем или иным методом решим монохроматические задачи, соответствующие различным частотам ω_m . Соответствующие решения обозначим $\tilde{E}(z_n, \omega_m)$, $\tilde{H}(z_n, \omega_m)$. Напомним, что в каждой из этих задач реализуются свои значения $\varepsilon(z_n, \omega_m)$. Поэтому предлагаемый подход позволяет учитывать произвольный гладкий закон частотной дисперсии.

Просуммируем полученные спектральные амплитуды решения по всем частотам ω_m :

$$\begin{aligned} E(t_k, z_n) &= \sum_{m=0}^M \tilde{E}(\omega_m, z_n) \exp(i\omega t_k) g_m \Delta\omega_m, \\ H(t_k, z_n) &= \sum_{m=0}^M \tilde{H}(\omega_m, z_n) \exp(i\omega t_k) g_m \Delta\omega_m. \end{aligned} \quad (55)$$

Это дает решение $E(z, t)$, $H(z, t)$ исходной нестационарной задачи.

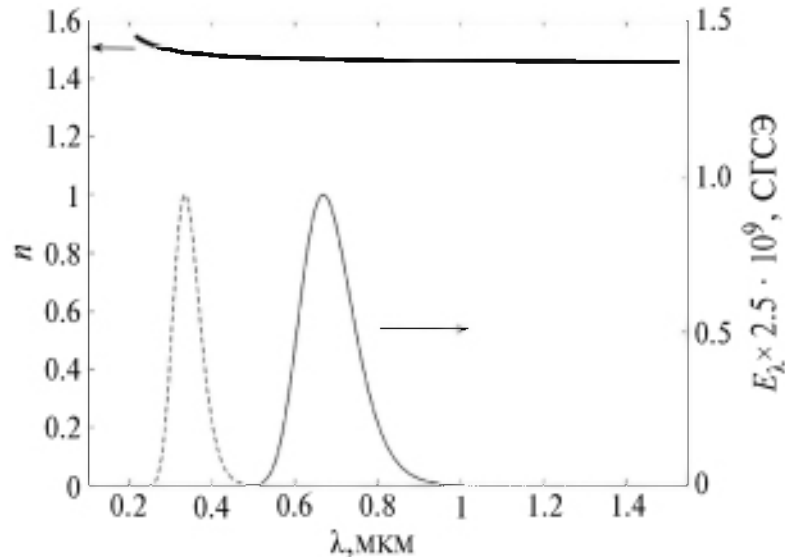
3. Для решения линейных монохроматических задач в работе [14] предложено использовать метод матриц рассеяния. Такой алгоритм был назван нестационарным матричным методом.

В настоящей работе предложено обобщение этого подхода на случай нелинейных сред. Разложим падающий импульс на монохроматические компоненты. Решим каждую монохроматическую задачу с помощью полученных выше обобщений метода матриц переноса. Затем численно выполним обратное преобразование Фурье.

4. Обоснование предлагаемого подхода заключается в следующем. Как отмечено выше, в реальных задачах вклад нелинейности в поляризацию мал по сравнению с линейным слагаемым. Поэтому возможно применение метода последовательных приближений. Приближения определяются из линейных задач. Метод спектрального разложения применяется к этим линейным задачам. Сеточная погрешность квадратур для интегралов Фурье может быть уменьшена вплоть до ошибок компьютерного округления. Поэтому точность метода спектрального разложения соответствует точности самих последовательных приближений.

7. ПРИМЕР РАСЧЕТА

1. В качестве примера рассмотрим рассеиватель, состоящий из одного слоя калийдигидрофосфата (KDP) толщиной 0.5 мкм, расположенного в воздухе. Пусть излучение распространяется перпендикулярно оптической оси. Соответствующий линейный показатель преломления указанного материала возьмем согласно [15, 16]. Зависимость показателя преломления от длины волны приведена на фиг. 1. Этот материал обладает квадратичной нелинейностью. Соответствующая восприимчивость равна $\chi^{II} = 3 \cdot 10^{-9}$ СГСЭ (см. [3]). Вклад частотной дисперсии χ^{II} является величиной второго порядка малости, поэтому данным эффектом пренебрежем.



Фиг. 1. Зависимость линейного показателя преломления от длины волны для кристалла KDP (жирная линия). Спектр падающего импульса (тонкая сплошная линия) и спектр импульса, соответствующего кратным гармоникам (штриховая линия).

2. Рассмотрим задачу в монохроматической постановке. Пусть на рассеиватель падает излучение с длиной волны $\lambda = 0.25$ мкм, амплитуда составляет $\mathcal{E} = 10^6$ СГСЭ. Этот случай можно рассматривать как изотропный согласно п. 2.2, 2.3. Пространственная развертка нормированной невязки (49) приведена на фиг. 2. В линейных средах она по построению равна нулю. В нелинейном материале невязка составляет не более 1%. Таким образом, исходное волновое уравнение выполняется с хорошей точностью.

На фиг. 3 приведены амплитуды нулевого приближения и поправки первого порядка малости для электрического поля. Видно, что электрическое поле непрерывно. Это подтверждает выполнение условий сопряжения. Условия излучения выполнены по построению. Поэтому данное решение удовлетворяет всем условиям задачи, и метод можно считать верифицированным.

Видно также, что амплитуда поправки увеличивается по мере распространения волны в среде. Поэтому наибольшее влияние нелинейность оказывает после прохождения рассеивателя. Здесь наибольшая амплитуда поправки достигает $\sim 20\%$ от амплитуды нулевого приближения.

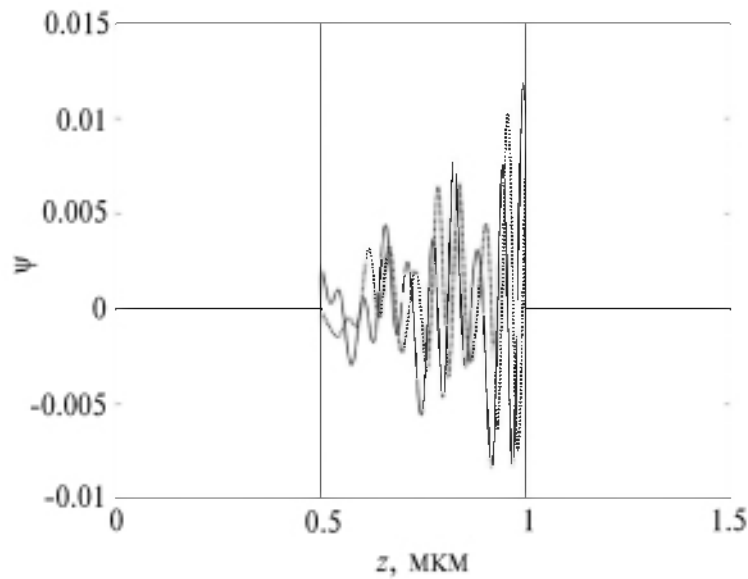
3. Пусть теперь падающее излучение является гауссовым импульсом

$$\mathcal{E}(t) = \mathcal{E}_0 \exp(-t^2/t_0^2). \quad (56)$$

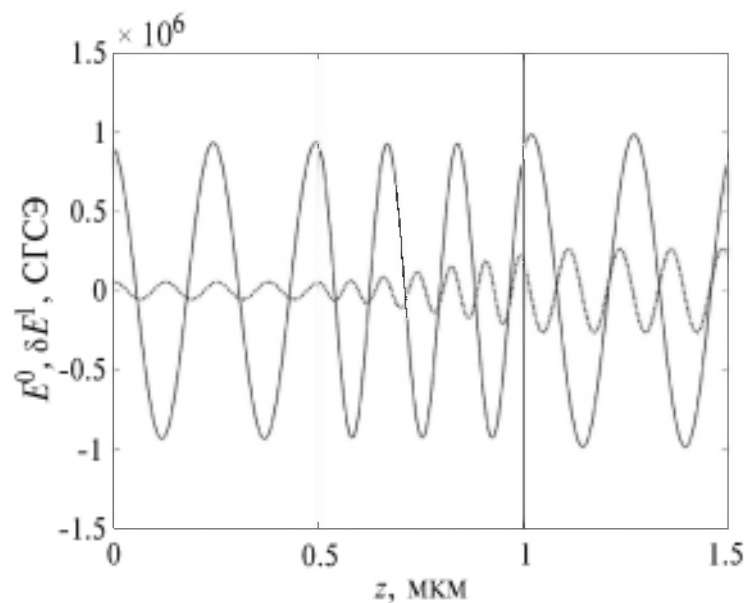
Его спектр также имеет гауссов профиль, показанный на фиг. 1. Вместе с ним приведен спектр импульса, в котором все гармоники удвоены. Увеличим толщину рассеивателя до 5 мкм, чтобы пространственная развертка импульса умещалась в нем целиком.

4. В нелинейной среде электрическое поле влияет на показатель преломления. Для квадратичной среды он равен

$$n_q(E) = \sqrt{n_q + 4\pi\chi_q^{II}E_q} \approx n_q + 2\pi\chi_q^{II}E_q. \quad (57)$$



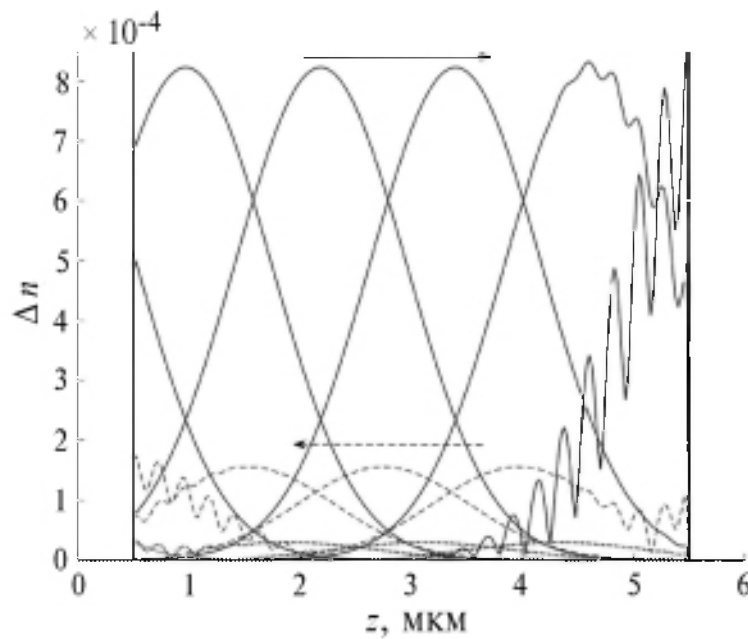
Фиг. 2. Нормированная невязка (49). Сплошная линия — $\text{Re } \bar{\psi}$, пунктир — $\text{Im } \bar{\psi}$. Вертикальные линии — границы раздела сред.



Фиг. 3. Амплитуды нулевого приближения (сплошная линия) и поправки первого порядка малости (пунктир). Вертикальные линии — границы раздела сред.

На фиг. 4 показано отклонение показателя преломления от линейного $\Delta n_q = n_q(E) - n_q$ в нескольких последовательных моментах времени. Для иллюстрации значение линейного показателя преломления взят в середине частотного диапазона. Видно, что эволюция показателя преломления имеет вид импульса, который распространяется в среде. Он проникает в среду и движется слева направо (сплошные линии), доходит до правой границы, частично отражается от нее и распространяется в обратную сторону (штриховые линии). Затем он частично отражается от левой границы, распространяется слева направо (пунктир) и т.д.

Этот эффект представляет интерес для оптических вычислений. Пусть имеется два импульса. Один из них (сильный) играет роль накачки (в англоязычной литературе он называется pump). Этот



Фиг. 4. Пространственно-временная модуляция показателя преломления немонохроматическим импульсом. Стрелками указано направление распространения импульса показателя преломления. Остальные обозначения — см. текст.

импульс модулирует показатель преломления. Одновременно запускают второй (слабый) импульс, его часто называют probe. Он рассеивается на среде с модулированным в пространстве и времени показателе преломления. Такая схема перспективна для оптического вычисления свертки либо произведения этих двух импульсов.

8. ЗАКЛЮЧЕНИЕ

В настоящей работе построены обобщения метода матриц переноса на ряд актуальных задач целинейной оптики в слоистых средах. Рассмотрены среды с квадратичной и кубической нелинейностью, при наличии и при отсутствии волнового синхронизма, как в монохроматическом, так и в немонохроматическом случае. Предложенные подходы имеют ряд преимуществ по сравнению с традиционными. Во-первых, они существенно расширяют область применимости матричных методов. Ранее такие задачи удавалось решать только сеточными методами. Во-вторых, для финитных импульсов квадратуры в методе спектрального разложения сходятся по экспоненциальному закону, т.е. при уменьшении шага сеток по частоте и времени вдвое число верных знаков примерно удваивается (см. [14]). Такая сходимость напоминает ньютоновскую. Поэтому предлагаемые методы являются сверхбыстрыми. Их трудоемкость кардинально меньше таковой у известных сеточных методов.

Авторы искренне благодарны Л. А. Севастьянову за ценные замечания и обсуждения и А. Н. Боголюбову за внимание к работе.

СПИСОК ЛИТЕРАТУРЫ

1. Гусев Е. Л. Математические методы синтеза слоистых структур. Новосибирск: Наука, 1993.
2. Гусев Е. Л. Качественные закономерности взаимосвязи параметров в оптимальных структурах в задачах оптимального синтеза неоднородных структур из дискретного набора материалов при волновых воздействиях // Докл. АН. 1996. Т. 346. № 3. С. 324–326.
3. Ахманов С. А., Никитин С. Ю. Физическая оптика. М.: Изд-во МГУ; Наука, 2004.
4. MacGillivray W. R., Smith S. D., MacKenzie H. A., Tooley F. A. P. Photonic logic // Optica Acta: Inter. J. of Optic. 1985. V. 32. N 5. P. 511–524.

5. Фишер Р., Мюллер Р. Нелинейные оптические устройства — основные элементы будущих цифровых оптических компьютеров? // Квантовая электроника. 1989. Т. 16. № 8. С. 1723–1730.
6. 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.
7. Домбровская Ж. О., Боголюбов А. Н. Немонотонность схемы fdtd при моделировании границ раздела между диэлектриками // Ученые записки физического факультета Московского Университета. 2017. № 4. С. 1740302.
8. Petropoulos P. G. Stability and phase error analysis of fdtd in dispersive dielectrics // IEEE Trans. Antennas Propagat. 1994. V. 42. N 1. P. 62–69.
9. Young J. L., Kittichartphayak A., Kwok Y. M., Sullivan D. On the dispersion errors related to (fd)²/td type schemes // IEEE Trans. Microwave Theory Techn. 1995. V. 43. N 8. P. 1902–1910.
10. Белов А. А., Домбровская Ж. О. Бикомпактная разностная схема для уравнений Максвелла в слоистых средах // Докл. АН. 2020. Т. 492. С. 15–19.
11. Белов А. А., Домбровская Ж. О. Тестирование бикомпактных схем для одномерных уравнений Максвелла в слоистых средах // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 9. С. 1532–1550.
12. Свешников А. Г. Принципы излучения // Докл. АН. 1950. Т. 3. № 5. С. 517–520.
13. Macleod H. A. Thin-film optical filters. Fourth edition. Taylor and Francis. Boca Raton London New York, 2010.
14. Белов А. А., Боголюбов А. Н., Домбровская Ж. О., Жбанников С. О. Сверх- быстрый метод расчёта одномерных задач фотоники // Физические основы приборостроения. 2020. Т. 9. № 2. С. 2–9.
15. Zernike F. Refractive indices of ammonium dihydrogen phosphate and potassium dihydrogen phosphate between 2000 Å and 1.5 μ // J. Opt. Soc. Am. 1964. T. 54. N 10. C. 1215–1220.
16. Polyanskiy M. N. Refractive index database. <https://refractiveindex.info>. Accessed on 2022-02-13.

GENERALIZATION OF THE SCATTERING MATRIX METHOD TO PROBLEMS IN NONLINEAR DISPERSING MEDIA

A. A. Belov^{a,b*}, Zh. O. Dombrovskaya^{a,**}

^a 119991 Moscow, Lenin Hills, 1, p. 2, Lomonosov Moscow State University, Russia

^b 117198 Moscow, Miklukho-Maklaya Str., 6, RUDN, Russia

*e-mail: aa.belov@physics.msu.ru

**e-mail: dombrovskaya@physics.msu.ru

Received 7 February, 2023

Revised 7 March, 2024

Accepted 3 April, 2024

Abstract. In recent years, much attention has been paid to integrated photonics devices based on nonlinear media. A generalization of the transfer matrix method to problems in plane-parallel layered media with quadratic and cubic nonlinearity is proposed. Incident radiation can be either a monochromatic wave or a non-monochromatic pulse. Previously, such problems could only be solved using grid methods. The proposed approaches significantly expand the field of applicability of matrix methods and radically exceed the efficiency of the known grid methods.

Keywords: Maxwell's equations, layered media, nonlinear media, conjugation conditions, dispersion of matter.

О РАВНОМЕРНОЙ СХОДИМОСТИ АППРОКСИМАЦИЙ КАСАТЕЛЬНОЙ И НОРМАЛЬНОЙ ПРОИЗВОДНЫХ ПОТЕНЦИАЛА ПРОСТОГО СЛОЯ ВБЛИЗИ ГРАНИЦЫ ДВУМЕРНОЙ ОБЛАСТИ

© 2024 г. Д. Ю. Иванов^{1,*}

¹ 127994 Москва, ул. Образцова, 9, Российский университет транспорта (РУТ (МИИТ)),
Академия базовой подготовки, Россия

*e-mail: ivanovdyu@yandex.ru

Поступила в редакцию 06.09.2023 г.

Переработанный вариант 06.09.2023 г.

Принята к публикации 03.04.2024 г.

Предлагаются полуаналитические аппроксимации касательной производной (КП) и нормальной производной (НП) потенциала простого слоя (ППС) вблизи границы двумерной области, выполненные в рамках коллокационного метода граничных элементов и не требующие аппроксимации координатных функций границы. Для получения аппроксимаций используются аналитическое интегрирование по гладкой компоненте функции расстояния и специальный аддитивно-мультипликативный способ выделения особенностей. Доказано, что такие аппроксимации обладают более равномерной сходимостью вблизи границы области по сравнению с аналогичными аппроксимациями КП и НП ППС на основе простого мультипликативного способа выделения особенностей. Установлена одна из причин сильно неравномерной сходимости традиционных аппроксимаций КП и НП ППС на основе квадратурных формул Гаусса. Библ. 26. Табл. 2.

Ключевые слова: метод граничных элементов, почти сингулярный интеграл, аналитическое интегрирование.

DOI: 10.31857/S0044466924070094, EDN: xikdvh

1. ВВЕДЕНИЕ

Метод граничных элементов (МГЭ) наряду с методом конечных элементов (МКЭ) и методом конечных разностей (МКР) является одним из основных методов решения краевых задач математической физики (см. [1, п. 1.8]). В рамках МГЭ дискретизации подвергается только граница области, в отличие от двух других методов, где дискретизируется вся область. В любой точке x , принадлежащей области, решения получают в виде потенциалов, имеющих вид криволинейных интегралов по длине дуги или поверхностных интегралов по площади поверхности в двумерном и трехмерном случаях соответственно. В коллокационном МГЭ (КМГЭ) (см. [1, п. 2.5]) для приближенного вычисления потенциалов граница разбивается на участки — так называемые граничные элементы, на каждом из которых осуществляется полиномиальная интерполяция функции плотности по длине дуги. Возникающие после этого интегралы на граничных элементах, вообще говоря, не могут быть вычислены точно, и в двумерном случае для их вычисления традиционно используются формулы Гаусса с постоянной (единичной) весовой функцией, которые обычно называют просто формулами Гаусса (см. [2, гл. 3, § 5, п. 3]). Если наблюдаемая точка x расположена достаточно далеко от точек x' , принадлежащих граничному элементу, то интеграл вычисляется с помощью формул Гаусса с удовлетворительной точностью. Если точка x находится на граничном элементе, то подынтегральная функция в общем случае имеет особенность при $x' = x$, и интеграл не может быть удовлетворительно вычислен с помощью формул Гаусса. Для вычисления таких сингулярных интегралов (СИ) применяются специальные методы, в том числе полуаналитические (см. [3]–[5]). Если же точка x не принадлежит граничному элементу, но находится от него достаточно близко, то подынтегральная функция не имеет особенности, но является быстро изменяющейся вблизи точки x_0 — проекции точки x на границу, поэтому

интеграл также не может быть вычислен с помощью формул Гаусса с приемлемой точностью. Такие интегралы принято называть почти сингулярными интегралами (ПСИ) (см. [6]–[10]), и для их вычисления также используются полуаналитические методы (см. [4]–[13]).

Необходимость вычисления потенциалов и их производных вблизи границы области возникает при решении задач в тонкостенных и многослойных конструкциях, тонких покрытиях, пленках, на концах трещин (см. [6], [10], [14]). В этих случаях МКЭ и МКР оказываются неэффективны, так как для их реализации требуется избыточная дискретизация пограничного слоя, и проявляется преимущество МГЭ, не требующего дискретизации области. Но при вычислениях вблизи границы большое значение имеет высокоточная аппроксимация самой границы, и линейная аппроксимация граничных элементов считается неудовлетворительной (см. [6], [14]). Дело в том, что из-за грубой аппроксимации границы может получиться даже так, что точка наблюдения, расположенная вблизи границы, может лежать по одну сторону от точной кривой, но по другую сторону от приближенной. Поэтому некоторые специальные методы вычисления ПСИ, реализованные вначале для линейных граничных элементов (см. [8]), были осуществлены впоследствии для квадратичных граничных элементов (см. [9]). Заметим, что линейными и квадратичными граничными элементами можно считать все те, у которых координатные функции граничной кривой заменены их линейными и квадратичными аппроксимациями соответственно. В последнее время отмечена необходимость еще более точной аппроксимации границы, например, с помощью сплайнов, и такая аппроксимация была реализована в рамках метода адаптивного деления почти сингулярных граничных элементов (см. [15], [16]) и методов нелинейного преобразования переменной интегрирования (см. [14], [17]).

Необходимость аппроксимации границы возникает по двум причинам. Первая причина имеет общий характер: на практике часто известно лишь конечное число граничных точек, на основе которых приходится делать интерполяцию всей границы. Вторая причина заключается в невозможности реализации метода для граничных элементов сложной формы. Вторая проблема характерна для полуаналитических методов, основанных на аналитическом интегрировании по длине дуги или площади поверхности, которое осуществимо лишь в случае простых координатных функций, так как степени функции расстояния находятся под знаком логарифма или в знаменателе подынтегрального выражения. По этой причине в работах [8], [4], [5], [11]–[13] для аппроксимации функции расстояния используются линейные приближения координатных функций, а в [10] используются линейные граничные элементы. В работе [9] применяются квадратичные граничные элементы, но для того, чтобы аналитическое интегрирование стало возможным, квадрат функции расстояния обрезается до квадратичной функции. Очевидно, что при таком подходе нет смысла использовать граничные элементы более высоких степеней, чем вторая.

Тем не менее в двумерном случае в рамках КМГЭ существует полуаналитический метод аппроксимации ПСИ, который может быть реализован для любой аналитически заданной границы без приближения координатных функций. Метод основан на переходе к новой переменной интегрирования $\rho \equiv (r^2 - d^2)^{1/2}$, где d и r — расстояния от наблюдаемой точки x до ее проекции x_0 и до граничной точки интегрирования x' соответственно. Для того чтобы стало возможным аналитическое интегрирование по ρ , применяется мультипликативный способ выделения особенностей: подынтегральная функция, порожденная фундаментальным решением, представляется в виде произведения двух функций переменной ρ , одна из которых при малых значениях d является быстро изменяющейся вблизи $\rho = 0$ и берется в качестве весовой, а другая, включающая якобиан, является медленно изменяющейся и аппроксимируется с помощью полиномиальной интерполяции по переменной ρ . Сложность аналитического интегрирования не зависит от формы границы, так как интегралы по ρ зависят от граничной кривой только параметрически. Заметим, что аппроксимация границы в данном случае осуществляется опосредованно — через интерполяцию медленно изменяющейся функции. Но поскольку интерполянт входит в виде множителя в числитель подынтегрального выражения, сложность интеграла при увеличении степени интерполянта принципиально не возрастает, и теоретически порядок аппроксимации может быть повышен до любого конечного значения. Недостатком данного метода является его реализация в настоящее время только для двумерной области с достаточно гладкой границей.

По существу этот метод изложен в работах [6], [7], [3] (в более ранней работе [3] рассматривается аппроксимация СИ, когда $d = 0$ и $\rho = r$), но формулировка и обоснование метода в этих работах выполнены только для квадратичных граничных элементов. В работе автора [18] доказано, что точное интегрирование по ρ может быть использовано для любой достаточно гладкой аналитически заданной границы без ее аппроксимации. В работах [6], [7] метод применялся для вычисления ПСИ, характерных для потенциала простого слоя (ППС) и его производных в случае уравнения Лапласа. В работах автора на основе аналитического интегрирования по ρ были получены аппроксимации теплового ППС (см. [18]) и его нормальной производной (НП) (см. [19]), а также потенциала двойного слоя (ПДС): теплового (см. [20]) и гармонического (см. [21]). Кроме случая теплового ППС, где также используется простой мультипликативный способ выделения особенностей, ПСИ в работах [19]–[21] специальным образом представляются в виде суммы двух ПСИ, каждый из которых аппроксимируется указанным мультипликативным способом. Такая аддитивно-мультипликативная аппроксимация (АМА) обеспечивает более равномерную сходимость аппроксимаций ПДС и НП ППС в приграничной области, чем простая мультипликативная аппроксимация (ПМА), описанная в [6], [7]. В настоящей работе с помощью АМА и ПМА получены аппроксимации касательной производной (КП) и НП гармонического ППС в любой точке двумерной приграничной области, не включающей саму границу. Для аппроксимации медленно изменяющихся функций используется кусочно-квадратичная интерполяция (ККИ). Как и в работах [18]–[21], точное интегрирование по ρ осуществляется не только для вычисления ПСИ, но и для вычисления интегралов на всех граничных элементах, принадлежащих некоторой фиксированной по ширине дуге граничной кривой вблизи точки наблюдения. На остальной части границы интегралы на граничных элементах, возникающие после интерполяции функции плотности, вычисляются с помощью формул Гаусса с порядком аппроксимации не меньше пятого, что вместе с использованием ККИ обеспечивает кубическую поточечную сходимость аппроксимаций КП и НП ППС. Такие аппроксимации КП и НП ППС, где использование АМА и ПМА сочетается с использованием формул Гаусса, условимся называть АМА КП ППС, АМА НП ППС, ПМА КП ППС, ПМА НП ППС. Кроме того, в настоящей работе исследуются аппроксимации КП и НП ППС, в которых интегралы на граничных элементах, возникающие после ККИ функции плотности, вычисляются исключительно с помощью формул Гаусса. Такие аппроксимации мы считаем традиционными и называем традиционными аппроксимациями КП и НП ППС на основе формул Гаусса.

В разд. 3 настоящей работы для границы класса C^5 доказано, что АМА КП ППС и ПМА КП ППС устойчиво сходятся в приграничной области с порядками аппроксимации не менее, чем $O(h^3 \ln |d|)$ и $O(h^3 d^{-1})$ соответственно (h — длина граничного элемента), а АМА НП ППС и ПМА НП ППС — $O(h^3)$ и $O(h^3 d^{-1})$ соответственно. В разд. 4 и для КП ППС, и для НП ППС доказано, что если расстояние от наблюдаемой точки до граничного элемента уменьшается пропорционально h , то АМА ПСИ и ПМА ПСИ являются аппроксимациями третьего и второго порядков соответственно, а на основе формул Гаусса вообще не могут быть получены аппроксимации ПСИ при фиксированном числе узлов квадратурной формулы. В заключительном разд. 5 приведены результаты вычисления КП и НП ППС вблизи границы единичного круга, которые подтверждают теоретические выводы.

Доказанное здесь утверждение относительно аппроксимаций ПСИ на основе формул Гаусса можно рассматривать как одну из причин сильно неравномерной сходимости традиционных аппроксимаций НП и КП ППС на основе формул Гаусса вблизи границы области. Но существуют и другие объяснения этого явления. Как правило, указывается, что при приближении точки наблюдения x к узлам формулы Гаусса x_μ аппроксимации потенциалов и их производных неограниченно возрастают, и точность катастрофически падает (см., например, [4], [12]). Но точность существенно уменьшается и тогда, когда точка x при пересечении границы проходит мимо узлов x_μ , и аппроксимации остаются конечными. В случае ПДС и НП ППС это можно объяснить тем, что их точные значения при переходе через границу терпят конечный разрыв, а их традиционные аппроксимации на основе формул Гаусса непрерывны и сходятся к точным значениям во всех точках замкнутой области, кроме x_μ , если граница достаточно гладкая. Такая сходимость не может быть равномерной, поскольку в достаточной близости от границы всегда сохраняется абсолютная погрешность, близкая к величине разрыва.

2. ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Пусть Ω_+ — двумерная открытая ограниченная односвязная область с границей $\partial\Omega$. В декартовых координатах (x_1, x_2) зададим параметрические уравнения кривой $\partial\Omega$: $x_1 = \tilde{x}_1(s)$, $x_2 = \tilde{x}_2(s)$. Параметр s по модулю равен длине дуги, откладываемой от некоторой фиксированной точки и заканчивающейся в точке $\tilde{x}(s) \equiv (\tilde{x}_1(s), \tilde{x}_2(s))$, и увеличивается, когда область Ω_+ при обходе границы $\partial\Omega$ остается слева. Функции $\tilde{x}_1(s)$, $\tilde{x}_2(s)$ ($s \in \mathbb{R}$), периодические с периодом $2S$ (S — половина длины $\partial\Omega$), осуществляют взаимнооднозначное отображение множества $I_S \equiv [-S, S]$ на множество $\partial\Omega$. Условимся далее писать $\partial\Omega \in C^n$, если существуют непрерывные на замкнутом множестве $\overline{I_S}$ производные $\tilde{x}_i^{(l)}(s)$ ($l = \overline{0, n}$, $i = 1, 2$), причем $\tilde{x}_i^{(l)}(-S+0) = \tilde{x}_i^{(l)}(S-0)$. Будем считать, что $\partial\Omega \in C^2$, если не оговорено особо.

Обозначим через $\mathbf{e}(s)$ единичный вектор, направленный по касательной к кривой $\partial\Omega$ в точке $\tilde{x}(s)$ в сторону увеличения параметра s , а через $\mathbf{n}(s)$ — единичную нормаль к кривой $\partial\Omega$, проходящую через точку $\tilde{x}(s)$ и направленную внутрь области Ω_+ . Векторы $\mathbf{e}(s)$, $\mathbf{n}(s)$ образуют правую систему, и их координаты (x_1, x_2) вычисляются с помощью формул $\mathbf{e}(s) \equiv (\tilde{x}'_1(s), \tilde{x}'_2(s))$, $\mathbf{n}(s) \equiv (-\tilde{x}'_2(s), \tilde{x}'_1(s))$.

Через $C(\partial\Omega)$ обозначим банахово пространство периодических с периодом $2S$ и непрерывных на всей числовой оси $\mathbb{R}$ вещественных функций $f(s)$, с нормой $\|f\|_{C(\partial\Omega)} = \sup_{s \in I_S} |f(s)|$. Через $C^n(\partial\Omega)$ ($n \in \mathbb{Z}_+$) обозначим банаховы пространства функций $f \in C(\partial\Omega)$, имеющих непрерывные производные

$f^{(l)}(s)$ ($s \in \mathbb{R}$, $l = \overline{1, n}$), с нормой $\|f\|_{C^n(\partial\Omega)} = \sum_{l=0}^n \sup_{s \in I_S} |f^{(l)}(s)|$ ($C^0(\partial\Omega) = C(\partial\Omega)$).

Обозначим через D треть радиуса круга Ляпунова кривой $\partial\Omega$ (см. [22, п. 94]). Введем в рассмотрение местные системы декартовых координат (ξ_s, η_s) с началами в точках $\tilde{x}(s)$ и осями ординат, направленными с соответствующими векторами $\mathbf{n}(s)$. Точки $\tilde{x}_d(s)$ ($s \in I_S$) с местными координатами $(\xi_s, \eta_s) = (0, d)$ при фиксированном $d \in I_D \equiv [-D, 0] \cup (0, D]$ образуют замкнутую линию $\partial\Omega_d \in C^1$, при этом соответствие между точками $\tilde{x}_d(s)$ и $\tilde{x}(s)$ взаимно однозначное ($\tilde{x}_0(s) \equiv \tilde{x}(s)$), а нормали $\tilde{x}(s)\tilde{x}_d(s)$ к кривой $\partial\Omega$ являются и нормальными к кривой $\partial\Omega_d$ (см. [22, п. 102]).

Обозначим через Ω_D множество, образованное точками $x \equiv (x_1, x_2) = \tilde{x}_d(s)$ ($d \in I_D$, $s \in I_S$). На множестве Ω_D зададим функции $u(x) \equiv (2\pi)^{-1}G(x)v$ и $w(x) \equiv (2\pi)^{-1}Q(x)v$, где

$$G(x)v \equiv \int_{I_S} g(x, s')v(s')ds', \quad g(x, s') \equiv \partial_{\mathbf{e}(s)} \ln r^{-1} = -2^{-1}(\partial_s r^2)/r^2, \quad (2.1a)$$

$$Q(x)v \equiv \int_{I_S} q(x, s')v(s')ds', \quad q(x, s') \equiv \partial_{\mathbf{n}(s)} \ln r^{-1} = -2^{-1}(\partial_d r^2)/r^2; \quad (2.1b)$$

$v \in C(\partial\Omega)$, $r(x, s') = |\mathbf{r}|$, $\mathbf{r}(x, s') \equiv \overrightarrow{x\tilde{x}(s')}$; дифференцирования $\partial_{\mathbf{e}(s)}$ и $\partial_{\mathbf{n}(s)}$ осуществляются по переменной $x = \tilde{x}_d(s)$ в направлениях $\mathbf{e}(s)$ и $\mathbf{n}(s)$ соответственно. Функции $u(x)$ и $w(x)$ — КП и НП двумерного гармонического ППС с плотностью v . Условимся, что мы можем иногда для краткости не писать аргументы функции, если они такие же, какие используются при определении функции.

Условимся линейный оператор $\mathbf{A}$, отображающий банахово пространство X в банахово пространство Y , обозначать как $\mathbf{A} [X \rightarrow Y]$. При каждом фиксированном $x \in \Omega_D$ оператор $G(x) [C(\partial\Omega) \rightarrow \mathbb{R}]$ — линейный функционал, отображающий пространство $C(\partial\Omega)$ в банахово пространство вещественных чисел $\mathbb{R}$.

Получим представления функций $g(x, s')$ и $q(x, s')$, которые будут использованы для полуаналитических аппроксимаций функций $u(x)$ и $w(x)$ в области Ω_D . Пусть $\mathbf{r}_0(s, s') \equiv \overrightarrow{\tilde{x}(s)\tilde{x}(s')}$, $r_0(s, s') \equiv |\mathbf{r}_0|$. Зададим на множестве $\Theta \equiv \{(s, s') : s \in \overline{I_S}, s' - s \in \overline{I_S}\}$ функции $\psi_i(s, s')$ ($i = \overline{0, 3}$): при $s' \neq s$ равенствами $\psi_i \equiv \varphi_i/(s' - s)^2$ ($i = 0, 2$), $\psi_i \equiv \varphi_i/(s' - s)$ ($i = 1, 3$), где

$$\varphi_0(s, s') \equiv r_0^2 = [\tilde{x}_1(s') - \tilde{x}_1(s)]^2 + [\tilde{x}_2(s') - \tilde{x}_2(s)]^2, \quad (2.2)$$

$$\varphi_1(s, s') \equiv -(\mathbf{e}(s), \mathbf{r}_0)_{\mathbb{R}^2} = -\tilde{x}'_1(s) [\tilde{x}_1(s') - \tilde{x}_1(s)] - \tilde{x}'_2(s) [\tilde{x}_2(s') - \tilde{x}_2(s)] = \partial_s \varphi_0, \quad (2.3)$$

$$\varphi_2(s, s') \equiv -(\mathbf{n}(s), \mathbf{r}_0)_{\mathbb{R}^2} = \tilde{x}'_2(s) [\tilde{x}_1(s') - \tilde{x}_1(s)] - \tilde{x}'_1(s) [\tilde{x}_2(s') - \tilde{x}_2(s)], \quad (2.4)$$

$$\varphi_3(s, s') \equiv \partial_s \varphi_2 = \tilde{x}_2''(s) [\tilde{x}_1(s') - \tilde{x}_1(s)] - \tilde{x}_1''(s) [\tilde{x}_2(s') - \tilde{x}_2(s)] \quad (2.5)$$

(здесь $(\cdot, \cdot)_{\mathbb{R}^2}$ — скалярное произведение в евклидовом пространстве $\mathbb{R}^2$), а при $s' = s$ равенствами $\psi_0 = -\psi_1 \equiv 1$, $\psi_2 = -\psi_3 \equiv 2^{-1} [\tilde{x}_2'(s) \tilde{x}_1''(s) - \tilde{x}_1'(s) \tilde{x}_2''(s)]$. В силу [23, лемма] при условии $\partial\Omega \in C^{n+2}$ ($n \in \mathbb{Z}_+$) существуют непрерывные на множестве Θ производные $\partial_s^k \partial_{s'}^l \psi_i$ ($k = 0, n-l$, $l = 0, n$, $i = \overline{0, 3}$).

Мы считаем, что значение параметра s соответствует точке наблюдения $x = \tilde{x}_d(s)$, а значение s' — точке интегрирования $\tilde{x}(s')$ в выражениях (2.1) для КП и НП ППС. Местные координаты (ξ_s, η_s) точек $\tilde{x}_d(s)$ и $\tilde{x}(s')$ равны $(0, d)$ и $((e(s), r_0)_{\mathbb{R}^2}, (n(s), r_0)_{\mathbb{R}^2})$ соответственно, поэтому с учетом равенств (2.2)–(2.4) имеем

$$r^2 = \left| \overrightarrow{\tilde{x}_d(s) \tilde{x}(s')} \right|^2 = r_0^2 - 2d (n(s), r_0)_{\mathbb{R}^2} + d^2 = \varphi_0 + 2d \varphi_2 + d^2. \quad (2.6)$$

Введем в рассмотрение функцию $\varphi(d, s, s')$ на множестве $\Upsilon \equiv \overline{I_D} \times \Theta$: $\varphi \equiv r^2 - d^2 = \varphi_0 + 2d \varphi_2$. Кривая $\partial\Omega$ и окружность радиуса $d \in I_D$ с центром $\tilde{x}_d(s)$ имеют только одну общую точку $\tilde{x}(s)$, поэтому если $s' \neq s$, то $2d \cos \alpha < r_0$, где α — угол между лучами $\tilde{x}(s) \tilde{x}(s')$ и $\tilde{x}(s) \tilde{x}_d(s)$. Тогда из формулы (2.6) следует, что $\varphi \geq 0$ при $(d, s, s') \in \Upsilon$ ($\varphi > 0$ при $s \neq s'$, $\varphi = 0$ при $s = s'$). На множестве Υ зададим непрерывные функции $\rho(d, s, s')$, $\psi(d, s, s')$: $\rho = \sqrt{\varphi}$, если $s' \geq s$; $\rho = -\sqrt{\varphi}$, если $s' < s$; $\psi \equiv \psi_0 + 2d \psi_2$.

Так как контур $\partial\Omega$ не имеет точек самопересечения, то $c_r \equiv \inf_{(s, s') \in \Theta} \psi_0 > 0$ ($c_r \leq 1$). Если $(s, s') \in \Theta$ и ϑ — острый угол между нормальными, проходящими через точки $\tilde{x}(s)$ и $\tilde{x}(s')$, то $\vartheta \leq c_K |s' - s| \leq c_K c_r^{-1/2} r_0$, при этом $c_K \equiv \sup_{s \in I_S} K(s)$, $K(s) = 2 |\psi_2(s, s)|$ — кривизна кривой $\partial\Omega$ в точке $\tilde{x}(s)$. Отсюда

следует, что величина $3D$, где $D \equiv (3c_K)^{-1} c_r^{1/2}$, может быть взята в качестве радиуса круга Ляпунова (см. [22, п. 94]). Так как $\psi_0(s, s) = 1$, $|\psi_2(s, s)| = 2^{-1} K(s)$ и $D \leq 1/(3c_K)$, то при $(d, s) \in \overline{I_D} \times \overline{I_S}$ имеем оценку $\psi(d, s, s) \geq 2/3$. Кроме того, $\psi = \varphi/(s' - s)^2 > 0$ при $(d, s, s') \in \Upsilon$, $s' \neq s$, поэтому $\psi > 0$ на множестве Υ . Следовательно, производные $\partial_{s'}^j \rho$ ($j = \overline{0, n+1}$) непрерывны на Υ при условии $\partial\Omega \in C^{n+2}$ ($n \in \mathbb{Z}_+$), так как $\partial_{s'} \rho = \psi^{-1/2} [-\psi_1(s', s) - d \psi_3(s', s)]$.

В силу формул (2.1), (2.3), (2.5), (2.6) имеем при $d \in I_D$, $(s, s') \in \Theta$ равенства $2^{-1} \partial_s r^2 = \varphi_1 + d \varphi_3$, $2^{-1} \partial_{d'} r^2 = \varphi_2 + d$, и $g(\tilde{x}_d(s), s') = -(\varphi_1 + d \varphi_3) / (\varphi + d^2)$, $q(\tilde{x}_d(s), s') = -(\varphi_2 + d) / (\varphi + d^2)$. Учитывая также, что $\rho / (s' - s) = \psi^{1/2}$, получаем по два представления функций g и q при $x = \tilde{x}_d(s)$, $(d, s, s') \in \Upsilon$, $d \neq 0$:

$$g(x, s') = a_1(d, \rho) \delta_1(d, s, s') + a_2(d, \rho) \delta_2(d, s, s'), \quad (2.7a)$$

$$q(x, s') = a_3(d, \rho) \delta_3(d, s, s') + a_4(d, \rho) \delta_4(d, s, s'), \quad (2.7b)$$

$$g(x, s') = a_5(d, \rho) \delta_5(d, s, s'), \quad (2.8a)$$

$$q(x, s') = a_6(d, \rho) \delta_6(d, s, s'), \quad (2.8b)$$

где $a_1(d, \rho) \equiv \rho / (\rho^2 + d^2)$, $a_3(d, \rho) \equiv \rho^2 / (\rho^2 + d^2)$, $a_2(d, \rho) = a_4(d, \rho) \equiv d / (\rho^2 + d^2)$, $a_5(d, \rho) = a_6(d, \rho) \equiv 1 / (\rho^2 + d^2)$, $\delta_1 \equiv -\psi_1 / \psi^{1/2}$, $\delta_2 \equiv -\varphi_3$, $\delta_3 \equiv -\psi_2 / \psi$, $\delta_4 \equiv 1$, $\delta_5 \equiv -\varphi_1 - d \varphi_3$, $\delta_6 \equiv -\varphi_2 - d$. Так как $\psi > 0$ на множестве Υ , то при условии $\partial\Omega \in C^{n+2}$ ($n \in \mathbb{Z}_+$) существуют непрерывные на множестве Υ производные $\partial_{s'}^j \delta_i$ ($j = \overline{0, n}$, $i = \overline{1, 6}$).

Значения $s' - s$, соответствующие при фиксированном $s \in I_S$ точкам $\tilde{x}(s')$ с местными координатами $\xi_s = -D$ и $\xi_s = D$, обозначим через Σ'_s, Σ''_s соответственно ($\Sigma'_s < 0 < \Sigma''_s$).

Теорема 1 (см. [18, теорема 5]). Пусть $\partial\Omega \in C^{n+2}$ ($n \in \mathbb{Z}_+$). Тогда на множестве $\Upsilon' \equiv \{(d, s, s') : d \in \overline{I_D}, s \in \overline{I_S}, s' - s \in \Xi_s \equiv [\Sigma'_s, \Sigma''_s]\}$ существуют положительная, ограниченная сверху функция $\delta_0(d, s, s') \equiv (\partial_{s'} \rho)^{-1}$ и непрерывные производные $\partial_{s'}^j \delta_0$ ($j = \overline{0, n}$).

Следствие 1. Пусть $\partial\Omega \in C^{n+2}$ ($n \in \mathbb{Z}_+$). Тогда функции $\rho_{d,s}(\sigma) \equiv \rho(d, s, s + \sigma)$ при любых фиксированных $s \in I_S$, $d \in \overline{I_D}$ диффеоморфно с гладкостью C^{n+1} отображают множества Ξ_s на соответствующие множества $\rho_{d,s}(\Xi_s)$. Функции $\delta_0(d, s, \rho) \equiv \delta_0(d, s, s + \sigma_{d,s}(\rho))$, $\tilde{\delta}_i(d, s, \rho) \equiv \delta_i(d, s, s + \sigma_{d,s}(\rho)) \delta_0$ ($i = \overline{1, 6}$; $\sigma_{d,s}(\rho)$ — функция, обратная к функции $\rho_{d,s}(\sigma)$) имеют непрерывные на множестве $\tilde{\Upsilon}' \equiv \{(d, s, \rho) : d \in \overline{I_D}, s \in \overline{I_S}, \rho \in \rho_{d,s}(\Xi_s)\}$ производные $\partial_\rho^j \tilde{\delta}_i$ ($j = \overline{0, n}$, $i = \overline{0, 6}$).

В заключение раздела рассмотрим квадратичные аппроксимации вещественной функции $f(z) \in C([z_0, z_1])$ ($z_0 < z_1$). Обозначим через $\hat{f}(z)$ интерполяционный многочлен Лагранжа с узлами $z_0, z_1, z_2 \equiv 2^{-1}(z_0 + z_1)$ (см. [2, гл. 2, § 5, п. 2]):

$$\hat{f}(z) \equiv f(z_0) + (z - z_0)f_1(z_0, z_1) + (z - z_0)(z - z_1)f_2(z_0, z_1, z_2).$$

Здесь f_1, f_2 — разделенные разности 1-го и 2-го порядков:

$$f_1(z_0, z_1) \equiv (f(z_1) - f(z_0)) / (z_1 - z_0), \quad f_2(z_0, z_1, z_2) \equiv (f(z_1, z_2) - f(z_0, z_1)) / (z_2 - z_0)$$

(см. [2, гл. 2, § 5, п. 1]). При $z \in (z_0, z_2) \cup (z_2, z_1)$ имеют место две формулы при $f \in C([z_0, z_1])$ и $f \in C^1([z_0, z_1])$ соответственно:

$$f(z) = \hat{f}(z) + p_3(z, z_0, z_1)f_3(z, z_0, z_1, z_2), \quad (2.9)$$

$$f(z) = \hat{f}(z) + p_3(z, z_0, z_1)f_3(z_0, z_1, z_2, z_2) + p_4(z, z_0, z_1)f_4(z, z_0, z_1, z_2, z_2) \quad (2.10)$$

(см. [2, гл. 2, § 5, п. 3] и [2, гл. 2, § 11, п. 5]). Здесь $p_3(z, z_0, z_1) \equiv (z - z_0)(z - z_1)(z - z_2)$, $p_4(z, z_0, z_1) \equiv (z - z_0)(z - z_1)(z - z_2)^2$; f_3, f_4 — разделенные разности 3-го и 4-го порядков:

$$f_3(z, z_0, z_1, z_2) \equiv (f_2(z_0, z_1, z_2) - f_2(z, z_0, z_1)) / (z_2 - z),$$

$$f_3(z_0, z_1, z_2, z_2) \equiv \lim_{z_\epsilon \rightarrow z_2} f_3(z_0, z_1, z_2, z_\epsilon) \quad (z_\epsilon \in (z_0, z_2) \cup (z_2, z_1)),$$

$$f_4(z, z_0, z_1, z_2, z_2) \equiv (f_3(z_0, z_1, z_2, z_2) - f_3(z, z_0, z_1, z_2)) / (z_2 - z)$$

(см. [2, гл. 2, § 5, п. 1] и [2, гл. 2, § 11, п. 4]). Для любого $z \in (z_0, z_2) \cup (z_2, z_1)$ существуют $\zeta_1, \zeta_2 \in [z_0, z_1]$, такие что

$$f_3(z, z_0, z_1, z_2) = 6^{-1}f^{(3)}(\zeta_1) \quad (f \in C^3([z_0, z_1])), \quad (2.11)$$

$$f_4(z, z_0, z_1, z_2, z_2) = 24^{-1}f^{(4)}(\zeta_2) \quad (f \in C^4([z_0, z_1])) \quad (2.12)$$

(см. [2, гл. 2, § 5, п. 3] и [2, гл. 2, § 11, п. 5]). В силу формул (2.9), (2.11) при $z \in [z_0, z_1]$, $f \in C^3([z_0, z_1])$ имеет место мажорантная оценка погрешности интерполяции

$$|\hat{f}(z) - f(z)| \leq c_\lambda \sup_{z \in [z_0, z_1]} |f^{(3)}(z)| 8^{-1} (z_1 - z_0)^3 \quad (c_\lambda \equiv \sqrt{3}/27). \quad (2.13)$$

Также интерполяционный многочлен $\hat{f}(z)$ может быть записан в виде $\hat{f}(z) = \sum_{m=0}^2 f(z_m) \Lambda_m^{[z_0, z_1]}(z)$ (см. [2, гл. 2, § 2, п. 1]), где $\Lambda_m^{[z_0, z_1]}(z)$ ($m = \overline{0, 2}$) — квадратичные функции, такие что $\Lambda_m^{[z_0, z_1]}(z_m) = 1$ и $\Lambda_m^{[z_0, z_1]}(z_j) = 0$, если $j \neq m$. При $z \in [z_0, z_1]$ имеем неравенства $|\Lambda_m^{[z_0, z_1]}(z)| \leq 1$ ($m = \overline{0, 2}$) и, как следствие, оценки

$$|\hat{f}(z)| \leq c_{\Lambda, 0} \max_{m=\overline{0, 2}} |f(z_m)| \quad (f \in C([z_0, z_1]), \quad c_{\Lambda, 0} \equiv 3). \quad (2.14)$$

При $z \in [z_0, z_1]$ для производных $\hat{f}^{(j)}(z)$ ($j = 1, 2$) справедливы формулы

$$\hat{f}^{(1)}(z) = \frac{2z - (z_2 + z_1)}{(z_0 - z_2)(z_0 - z_1)} \int_{z_2}^{z_0} f^{(1)}(\zeta) d\zeta + \frac{2z - (z_2 + z_0)}{(z_1 - z_2)(z_1 - z_0)} \int_{z_2}^{z_1} f^{(1)}(\zeta) d\zeta,$$

$$\hat{f}^{(2)}(z) = \frac{\int_{z_2}^{z_0} f^{(2)}(\zeta) (z_0 - \zeta) d\zeta}{(z_0 - z_2)(z_0 - z_1)} + \frac{\int_{z_2}^{z_1} f^{(2)}(\zeta) (z_1 - \zeta) d\zeta}{(z_1 - z_2)(z_1 - z_0)},$$

полученные с помощью формулы Тейлора с дополнительным членом в виде определенного интеграла (см. [24, п. 318]). В силу этих формул при $z \in [z_0, z_1]$ имеют место оценки

$$|\hat{f}^{(j)}(z)| \leq c_{\Lambda, j} \sup_{z \in [z_0, z_1]} |f^{(j)}(z)| \quad (f \in C^j([z_0, z_1]), \quad j = 1, 2; \quad c_{\Lambda, 1} \equiv 3, \quad c_{\Lambda, 2} \equiv 2^{-1}). \quad (2.15)$$

3. ПОЛУАНАЛИТИЧЕСКИЕ АППРОКСИМАЦИИ КП И НП ППС

Рассмотрим представления КП и НП ППС, которые будут использованы для построения их полуаналитических аппроксимаций. Пусть $x = \tilde{x}_d(s)$, $d \in I_D$, $s \in I_S$. В силу равенств (2.7), (2.8) функционалы $G(x)$, $Q(x)$ ($x \in \Omega_D$) могут быть представлены в виде сумм двумя способами:

$$G(x) = A_1(x) + A_2(x) + A_7(x), \quad Q(x) = A_3(x) + A_4(x) + A_8(x), \quad (3.1)$$

$$G(x) = A_5(x) + A_7(x), \quad Q(x) = A_6(x) + A_8(x); \quad (3.2)$$

$$A_i(x)f \equiv \int_{\rho_{d,s}(\Xi_s)} a_i(d, \rho) B_i(d, s, \rho) f \, d\rho, \quad B_i(d, s, \rho) f \equiv \tilde{\delta}_i(d, s, \rho) f(s + \sigma_{d,s}(\rho)) \quad (i = \overline{1, 6});$$

$$A_j(x)f \equiv \int_{\overline{I_S \setminus \Xi_s}} B_j(d, s, s + \sigma) f \, d\sigma \quad (j = 7, 8),$$

$$B_7(d, s, s')f \equiv g(\tilde{x}_d(s), s') f(s'), \quad B_8(d, s, s')f \equiv q(\tilde{x}_d(s), s') f(s').$$

Здесь $f \in C(\partial\Omega)$. В силу следствия 1 и неравенства $r \geq D$, имеющего место, если $(d, s, s') \in \overline{\Upsilon \setminus \Upsilon'}$, для норм функционалов $A_i(x)$ [$C(\partial\Omega) \rightarrow \mathbb{R}$] ($x \in \Omega_D$) справедливы оценки

$$\|A_i(x)\| \leq \chi_i(d) \tilde{c}_i \hat{c}_{i,0} \quad (i = \overline{1, 8}), \quad (3.3)$$

$$\chi_1(d) \equiv |\ln |d||, \quad \chi_2(d) = \chi_3(d) = \chi_4(d) = \chi_7(d) = \chi_8(d) \equiv 1, \quad \chi_5(d) = \chi_6(d) \equiv |d|^{-1},$$

$$\tilde{c}_i \equiv \sup_{d \in I_D} \chi_i^{-1}(d) \int_{I_S} |a_i(d, \rho)| \, d\rho, \quad \hat{c}_{i,0} \equiv \sup_{(d,s,\rho) \in \Upsilon'} |\tilde{\delta}_i(d, s, \rho)| \quad (i = \overline{1, 6}),$$

$$\tilde{c}_7 = \tilde{c}_8 \equiv 2S, \quad \hat{c}_{7,0} \equiv \sup_{(d,s,s') \in \overline{\Upsilon \setminus \Upsilon'}} |g(\tilde{x}_d(s), s')|, \quad \hat{c}_{8,0} \equiv \sup_{(d,s,s') \in \overline{\Upsilon \setminus \Upsilon'}} |q(\tilde{x}_d(s), s')|.$$

Заметим, что $A_1(x) + A_2(x) = A_5(x)$, $A_3(x) + A_4(x) = A_6(x)$. Тем не менее представления (3.1) позволяют получить более сильные оценки норм функционалов $G(x)$ и $Q(x)$, чем (3.2). А именно, пусть $D \equiv \min \left\{ (3c_K)^{-1} c_r^{1/2}, e^{-1} \right\}$. Тогда в силу (3.2), (3.3) для норм функционалов $G(x)$, $Q(x)$ [$C(\partial\Omega) \rightarrow \mathbb{R}$] ($x \in \Omega_D$) справедливы оценки $\|G(x)\| \leq |d|^{-1} (\tilde{c}_5 \hat{c}_{5,0} + \tilde{c}_7 \hat{c}_{7,0})$, $\|Q(x)\| \leq |d|^{-1} (\tilde{c}_6 \hat{c}_{6,0} + \tilde{c}_8 \hat{c}_{8,0})$, а в силу (3.1), (3.3) имеем

$$\|G(x)\| \leq |\ln |d|| c_G, \quad c_G \equiv \tilde{c}_1 \hat{c}_{1,0} + \tilde{c}_2 \hat{c}_{2,0} + \tilde{c}_7 \hat{c}_{7,0}; \quad (3.4a)$$

$$\|Q(x)\| \leq c_Q, \quad c_Q \equiv \tilde{c}_3 \hat{c}_{3,0} + \tilde{c}_4 \hat{c}_{4,0} + \tilde{c}_8 \hat{c}_{8,0}. \quad (3.4b)$$

В соответствии с КМГЭ для осуществления дискретизации разобьем границу $\partial\Omega$ на граничные элементы и осуществим ККИ функции плотности. А именно, пусть $L/2 \in \mathbb{N}$, $h \equiv S/(L+1)$, $s_l \equiv lh$, $l \in \mathbb{Z}$. Тогда $\tilde{x}(s_{l+2L+2}) = \tilde{x}(s_l)$. Введем в рассмотрение пространства H_L вещественных сеточных функций f со значениями f_l , заданными в точках коллокации s_l ($f_{l+2L+2} = f_l$), с нормой $\|f\|_{H_L} = \max_{-L-1 \leq l \leq L} |f_l|$. Зададим проецирующие операторы $\mathbf{P}_L$ [$C(\partial\Omega) \rightarrow H_L$]:

$$(\mathbf{P}_L f)_l \equiv f(s_l), \quad f(s) \in C(\partial\Omega) \quad (3.5)$$

($\|\mathbf{P}_L\| \leq 1$), и интерполирующие операторы $\ddot{\mathbf{P}}_L$ [$H_L \rightarrow C(\partial\Omega)$]:

$$(\ddot{\mathbf{P}}_L f)(s) \equiv \sum_{m=0}^2 f_{2l-1+m} \Lambda_m^{[s_{2l-1}, s_{2l+1}]}(s) \quad (s \in [s_{2l-1}, s_{2l+1}], \quad l = \overline{-L/2, L/2}).$$

В силу оценки (2.14) операторы $\ddot{\mathbf{P}}_L$ [$H_L \rightarrow C(\partial\Omega)$] ограничены в совокупности: $\|\ddot{\mathbf{P}}_L\| \leq c_{\Lambda,0}$. На основании неравенства (2.13) имеем оценки

$$\|\ddot{\mathbf{P}}_L \mathbf{P}_L f - f\|_{C(\partial\Omega)} \leq c_\kappa \|f^{(3)}\|_{C(\partial\Omega)} h^3 \quad (f \in C^3(\partial\Omega)). \quad (3.6)$$

С помощью равенств

$$\ddot{G}(x)f \equiv G(x)\ddot{\mathbf{P}}_L f, \quad \ddot{Q}(x)f \equiv Q(x)\ddot{\mathbf{P}}_L f \quad (f \in H_L)$$

зададим сеточные функционалы $\ddot{G}(x), \ddot{Q}(x) [H_L \rightarrow \mathbb{R}]$ ($x \in \Omega_D$). Используя оценки (3.4), (3.6), получаем оценки аппроксимации функционалов $G(x), Q(x)$ ($x \in \Omega_D$) функционалами $\ddot{G}(x)\mathbf{P}_L, \ddot{Q}(x)\mathbf{P}_L$:

$$\left| \ddot{G}(x)\mathbf{P}_L f - G(x)f \right| \leq |\ln |d|| c_G c_\lambda \|f^{(3)}\|_{C(\partial\Omega)} h^3 \quad (f \in C^3(\partial\Omega)), \quad (3.7a)$$

$$\left| \ddot{Q}(x)\mathbf{P}_L f - Q(x)f \right| \leq c_Q c_\lambda \|f^{(3)}\|_{C(\partial\Omega)} h^3 \quad (f \in C^3(\partial\Omega)). \quad (3.7b)$$

Функционалы $\ddot{G}(x), \ddot{Q}(x)$ имеют вид матриц-строк длиной $2L + 2$:

$$\ddot{G}(x)f = \sum_{l=-L-1}^L \ddot{g}_l(x)f_l \quad (f \in H_L), \quad \ddot{g}_{2l}(x) \equiv \int_{s_{2l-1}}^{s_{2l+1}} g_{1,l}(x, s') ds', \quad (3.8)$$

$$\ddot{Q}(x)f = \sum_{l=-L-1}^L \ddot{q}_l(x)f_l \quad (f \in H_L), \quad \ddot{q}_{2l}(x) \equiv \int_{s_{2l-1}}^{s_{2l+1}} q_{1,l}(x, s') ds',$$

где при $l = \overline{-L/2, L/2}$

$$\ddot{g}_{2l-1}(x) \equiv \int_{s_{2l-3}}^{s_{2l-1}} g_{2,l}(x, s') ds' + \int_{s_{2l-1}}^{s_{2l+1}} g_{0,l}(x, s') ds' \quad (l = \overline{-L/2, L/2}),$$

$$g_{m,l}(x, s') \equiv g(x, s') \Lambda_m^{[s_{2l-1}, s_{2l+1}]}(s') \quad (m = 0, 1); \quad g_{2,l}(x, s') \equiv g(x, s') \Lambda_m^{[s_{2l-3}, s_{2l-1}]}(s'),$$

$$\ddot{q}_{2l-1}(x) \equiv \int_{s_{2l-3}}^{s_{2l-1}} q_{2,l}(x, s') ds' + \int_{s_{2l-1}}^{s_{2l+1}} q_{0,l}(x, s') ds' \quad (l = \overline{-L/2, L/2}),$$

$$q_{m,l}(x, s') \equiv q(x, s') \Lambda_m^{[s_{2l-1}, s_{2l+1}]}(s') \quad (m = 0, 1), \quad q_{2,l}(x, s') \equiv q(x, s') \Lambda_2^{[s_{2l-3}, s_{2l-1}]}(s').$$

Интегралы $\ddot{g}_l(x), \ddot{q}_l(x)$ в общем случае не могут быть вычислены точно, поэтому требуется их аппроксимация.

В соответствии с равенствами (3.1), (3.2) функционалы $\ddot{G}(x), \ddot{Q}(x)$ могут быть представлены в виде сумм двумя способами: $\ddot{G}(x) = \ddot{A}_1(x) + \ddot{A}_2(x) + \ddot{A}_7(x)$, $\ddot{Q}(x) = \ddot{A}_3(x) + \ddot{A}_4(x) + \ddot{A}_8(x)$ и $\ddot{G}(x) = \ddot{A}_5(x) + \ddot{A}_7(x)$, $\ddot{Q}(x) = \ddot{A}_6(x) + \ddot{A}_8(x)$, при этом $\ddot{A}_i(x) \equiv A_i(x)\ddot{\mathbf{P}}_L$ ($i = \overline{1, 8}$). Введем в рассмотрение функционалы $\hat{A}_i(x)$, аппроксимирующие $\ddot{A}_i(x)$ ($i = \overline{1, 6}$). Для этого заменим функции $B_i(d, s, \rho)\ddot{\mathbf{P}}_L f$ ($f \in H_L$) их ККИ $\hat{B}_i(d, s, \rho)f$ по переменной ρ :

$$\hat{A}_i(x)f \equiv \int_{\rho_{d,s}(\Xi_s)} a_i(d, \rho) \hat{B}_i(d, s, \rho) f d\rho \quad (i = \overline{1, 6}),$$

$$\hat{B}_i(d, s, \rho)f \equiv \hat{B}_{i,l}(d, s, \rho)f \quad (\rho \in [\rho_{d,s,l}, \rho_{d,s,l+1}], \quad l = \overline{-L-1, L}),$$

$$\hat{B}_{i,l}(d, s, \rho)f \equiv \begin{cases} \sum_{m=0}^2 B_i(d, s, \rho_{d,s,l,m})\ddot{\mathbf{P}}_L f \Lambda_m^{[\rho_{d,s,l}, \rho_{d,s,l+1}]}(\rho) & (\rho_{d,s,l} < \rho_{d,s,l+1}) \\ B_i(d, s, \rho_{d,s,l})\ddot{\mathbf{P}}_L f & (\rho_{d,s,l} = \rho_{d,s,l+1}) \end{cases};$$

$$\rho_{d,s,l,0} \equiv \rho_{d,s,l}, \quad \rho_{d,s,l,1} \equiv \rho_{d,s,l+1}, \quad \rho_{d,s,l,2} \equiv 2^{-1}(\rho_{d,s,l} + \rho_{d,s,l+1}).$$

Здесь $\rho_{d,s,l} \equiv \rho_{d,s}(\alpha_{s,l})$; $\alpha_{s,l} \equiv \min \{s_l - s, \Sigma_s''\}$, если $s_l \geq s$, и $\alpha_{s,l} \equiv \max \{s_l - s, \Sigma_s'\}$, если $s_l < s$. По определению $\bigcup_{l=-L-1}^L [\alpha_{s,l}, \alpha_{s,l+1}] = \Xi_s$ ($s \in \overline{I_S}$). Промежутки $[\alpha_{s,l}, \alpha_{s,l+1}]$ можно рассматривать как пересечения промежутков $[s_l, s_{l+1}]$ с множеством Ξ_s в том смысле, что выполняются равенства $(\alpha_{s,l}, \alpha_{s,l+1}) = (s_l, s_{l+1}) \cap \Xi_s$.

Аппроксимации $\hat{A}_7(x)$, $\hat{A}_8(x)$ функционалов $\ddot{A}_7(x)$, $\ddot{A}_8(x)$ построим с помощью формулы Гаусса с γ узлами:

$$\hat{A}_i(x) f \equiv 2^{-1} \sum_{l=-L-1}^L (\beta_{s,l+1} - \beta_{s,l}) \hat{B}_{i,l}(d, s) f \quad (i = 7, 8),$$

$$\hat{B}_{i,l}(d, s) f \equiv \sum_{j=1}^{\gamma} \omega_j B_i(d, s, \beta_{s,l,j}) \ddot{P}_L f, \quad \beta_{s,l,j} \equiv 2^{-1}(\beta_{s,l} + \beta_{s,l+1}) + 2^{-1}(\beta_{s,l+1} - \beta_{s,l}) z_j,$$

где $\beta_{s,l} \equiv \max \{s_l, s + \Sigma_s''\}$, если $s_l \geq s$, и $\beta_{s,l} \equiv \min \{s_l, s + \Sigma_s'\}$, если $s_l < s$; z_j — корни многочлена $(d'/dz^\gamma)(z^2 - 1)^\gamma$ на интервале $(-1, 1)$ (см. [2, гл. 3, § 5, п. 2]); для весовых коэффициентов ω_j выполняются условия $\sum_{j=1}^{\gamma} \omega_j = 2$, $\omega_j > 0$ (см. [2, гл. 3, § 5, п. 1]). По определению $\bigcup_{l=-L-1}^L [\beta_{s,l}, \beta_{s,l+1}] = \overline{I_S} \setminus \overline{\Xi_s}$ ($s \in \overline{I_S}$). Промежутки $[\beta_{s,l}, \beta_{s,l+1}]$ можно рассматривать как пересечения промежутков $[s_l, s_{l+1}]$ с множеством $\overline{I_S} \setminus \overline{\Xi_s}$ в том смысле, что выполняются равенства $(\beta_{s,l}, \beta_{s,l+1}) = (s_l, s_{l+1}) \cap \overline{I_S} \setminus \overline{\Xi_s}$.

Зададим функционалы $\hat{G}_I(x)$, $\hat{G}_{II}(x)$, $\hat{Q}_I(x)$, $\hat{Q}_{II}(x)$ [$H_L \rightarrow \mathbb{R}$] ($x \in \Omega_D$):

$$\hat{G}_I(x) \equiv \hat{A}_1(x) + \hat{A}_2(x) + \hat{A}_7(x), \quad \hat{G}_{II}(x) \equiv \hat{A}_5(x) + \hat{A}_7(x),$$

$$\hat{Q}_I(x) \equiv \hat{A}_3(x) + \hat{A}_4(x) + \hat{A}_8(x), \quad \hat{Q}_{II}(x) \equiv \hat{A}_6(x) + \hat{A}_8(x).$$

В силу оценки (2.14) функционалы $\hat{B}_i(d, s, \rho)$ [$H_L \rightarrow \mathbb{R}$] ($i = \overline{1, 6}$) равномерно ограничены по $(d, s, \rho) \in \tilde{\Upsilon}'$: $\|\hat{B}_i(d, s, \rho)\| \leq c_{\Lambda,0}^2 \hat{c}_{i,0}$, а функционалы $\hat{B}_{j,l}(d, s)$ ($j = 7, 8$) [$H_L \rightarrow \mathbb{R}$] равномерно ограничены по $(d, s) \in \overline{I_D} \times \overline{I_S}$, $l \in \mathbb{Z}$: $\|\hat{B}_{j,l}(d, s)\| \leq 2c_{\Lambda,0} \hat{c}_{j,0}$. Поэтому при $x \in \Omega_D$ справедливы неравенства $\|\hat{A}_i(x)\| \leq \chi_i(d) \tilde{c}_i c_{\Lambda,0}^2 \hat{c}_{i,0}$ ($i = \overline{1, 6}$), $\|\hat{G}_j(x)\| \leq \tilde{c}_j c_{\Lambda,0} \hat{c}_{j,0}$ ($j = 7, 8$), на основании которых получаем следующее утверждение.

Теорема 2. Пусть $\partial\Omega \in C^2$, $\gamma \in \mathbb{N}$. Тогда функционалы $\ln^{-1} |d| \hat{G}_I(x)$, $d \hat{G}_{II}(x)$, $\hat{Q}_I(x)$, $d \hat{Q}_{II}(x)$ [$H_L \rightarrow \mathbb{R}$] ($L/2 \in \mathbb{N}$, $x \in \Omega_D$) ограничены в совокупности.

В силу следствия 1 и неравенства $r \geq D$, имеющего место, если $(d, s, s') \in \overline{\Upsilon} \setminus \overline{\Upsilon}'$, при указанных гладкостях кривой $\partial\Omega$ и $j = \overline{0, n}$, $n \in \mathbb{Z}_+$ могут быть определены константы

$$\hat{c}_{i,j} \equiv \sup_{(d,s,\rho) \in \tilde{\Upsilon}'} \left| \partial_\rho^j \tilde{\delta}_i \right| \quad (i = \overline{0, 6}, \partial\Omega \in C^{m+2}),$$

$$\hat{c}_{7,j} \equiv \sup_{(d,s,s') \in \overline{\Upsilon} \setminus \overline{\Upsilon}'} \left| \partial_{s'}^j g(\tilde{x}_d(s), s') \right|, \quad \hat{c}_{8,j} \equiv \sup_{(d,s,s') \in \overline{\Upsilon} \setminus \overline{\Upsilon}'} \left| \partial_{s'}^j q(\tilde{x}_d(s), s') \right| \quad (\partial\Omega \in C^{m+1}).$$

Используя неравенства (2.13)–(2.15) и $\rho_{d,s,l+1} - \rho_{d,s,l} \leq c_h h$ ($c_h \equiv \sup_{(d,s,s') \in \Upsilon'} \partial_{s'} \rho$), при $f \in C^2(\partial\Omega)$, $\ddot{f} \equiv \ddot{P}_L \mathbf{P}_L f$, $x \in \Omega_D$ и указанных гладкостях кривой $\partial\Omega$ получаем оценки

$$\begin{aligned} \left| \hat{A}_i(x) \mathbf{P}_L f - \ddot{A}_i(x) \mathbf{P}_L f \right| &\leq \chi_i(d) \tilde{c}_i c_\Lambda 8^{-1} c_h^3 h^3 \operatorname{ess\,sup}_{(d,s,\rho) \in \tilde{\Upsilon}'} \left| \partial_\rho^3 B_i(d, s, \rho) \ddot{f} \right| \leq \\ &\leq 8^{-1} \chi_i(d) \tilde{c}_i c_h^3 c_\Lambda c_i \|f\|_{C^2(\partial\Omega)} h^3 \quad (\partial\Omega \in C^5), \end{aligned} \quad (3.9)$$

$$c_i \equiv \hat{c}_{i,3} c_{\Lambda,0} + (3\hat{c}_{i,2} \hat{c}_{0,0} + 3\hat{c}_{i,1} \hat{c}_{0,1} + \hat{c}_{i,0} \hat{c}_{0,2}) c_{\Lambda,1} + 3(\hat{c}_{i,1} \hat{c}_{0,0}^2 + \hat{c}_{i,0} \hat{c}_{0,1} \hat{c}_{0,0}) c_{\Lambda,2} \quad (i = \overline{1, 6});$$

$$\begin{aligned} \left| \hat{A}_i(x) \mathbf{P}_L f - \ddot{A}_i(x) \mathbf{P}_L f \right| &\leq \tilde{c}_i h^{2\gamma} \operatorname{ess\,sup}_{(d,s,s') \in \overline{\Upsilon} \setminus \Upsilon'} \left| \partial_{s'}^{2\gamma} B_j(d, s, s') \ddot{f} \right| \leq \\ &\leq 2S\tilde{c}_i \|f\|_{C^2(\partial\Omega)} h^{2\gamma} \quad (\partial\Omega \in C^{2\gamma+1}), \end{aligned} \quad (3.10)$$

$$c_i \equiv (\gamma!)^4 [(2\gamma)!]^{-3} (2\gamma+1)^{-1} [\hat{c}_{i,2\gamma} c_{\Lambda,0} + 2\gamma \hat{c}_{i,2\gamma-1} c_{\Lambda,1} + \gamma(2\gamma-1) \hat{c}_{i,2\gamma-2} c_{\Lambda,2}] \quad (i = 7, 8).$$

Здесь $\operatorname{ess\,sup}$ — существенный супремум (см. [25, гл. III, п. 1, пп. 11]). При получении оценки (3.10) используется оценка остаточного члена формулы Гаусса (см. [2, гл. 3, § 5, п. 2]).

На основании оценок (3.7), (3.9), (3.10) при $f \in C^3(\partial\Omega)$, $x \in \Omega_D$, $h \leq 1$, $\partial\Omega \in C^{2\gamma+1}$, $\gamma \geq 2$, получаем оценки

$$\left| \hat{G}_I(x) \mathbf{P}_L f - G(x) f \right| \leq |\ln |d|| c_{G,I} \|f\|_{C^3(\partial\Omega)} h^3, \quad c_{G,I} \equiv 8^{-1} c_h^3 c_\lambda (\tilde{c}_1 c_1 + \tilde{c}_2 c_2) + \tilde{c}_7 c_7 + c_G c_\lambda,$$

$$\left| \hat{G}_{II}(x) \mathbf{P}_L f - G(x) f \right| \leq |d|^{-1} c_{G,II} \|f\|_{C^3(\partial\Omega)} h^3, \quad c_{G,II} \equiv 8^{-1} c_h^3 c_\lambda \tilde{c}_5 c_5 + \tilde{c}_7 c_7 + c_G c_\lambda,$$

$$\left| \hat{Q}_I(x) \mathbf{P}_L f - Q(x) f \right| \leq c_{Q,I} \|f\|_{C^3(\partial\Omega)} h^3, \quad c_{Q,I} \equiv 8^{-1} c_h^3 c_\lambda (\tilde{c}_3 c_3 + \tilde{c}_4 c_4) + \tilde{c}_8 c_8 + c_Q c_\lambda,$$

$$\left| \hat{Q}_{II}(x) \mathbf{P}_L f - Q(x) f \right| \leq |d|^{-1} c_{Q,II} \|f\|_{C^3(\partial\Omega)} h^3, \quad c_{Q,II} \equiv 8^{-1} c_h^3 c_\lambda \tilde{c}_6 c_6 + \tilde{c}_8 c_8 + c_Q c_\lambda,$$

из которых вытекает следующее утверждение.

Теорема 3. Пусть $\partial\Omega \in C^{2\gamma+1}$, $\gamma \geq 2$, $\gamma, L/2 \in \mathbb{N}$. Тогда функционалы $\ln^{-1} |d| \hat{G}_I(x) \mathbf{P}_L$, $d \hat{G}_{II}(x) \mathbf{P}_L$ [$C^3(\partial\Omega) \rightarrow \mathbb{R}$], $\hat{Q}_I(x) \mathbf{P}_L$, $d \hat{Q}_{II}(x) \mathbf{P}_L$ [$C^3(\partial\Omega) \rightarrow \mathbb{R}$] сходятся при $L \rightarrow \infty$ по операторной норме к соответствующим функционалам $\ln^{-1} |d| G(x)$, $d G(x)$, $Q(x)$, $d Q(x)$ [$C^3(\partial\Omega) \rightarrow \mathbb{R}$] равномерно относительно $x \in \Omega_D$ со скоростью не меньше, чем кубическая.

Определения и основные сведения, касающиеся сходимости последовательностей операторов в различных топологиях см. в [25, гл. VI, п. 1, пп. 1–3].

С учетом формул (3.8) функционалы $\hat{G}_I(x)$, $\hat{G}_{II}(x)$, $\hat{Q}_I(x)$, $\hat{Q}_{II}(x)$ имеют вид матриц-строк длиной $2L+2$:

$$\hat{G}_I(x) f = \sum_{l=-L-1}^L [b_{1,l} + b_{2,l} + b_{7,l}] f_l, \quad \hat{G}_{II}(x) f = \sum_{l=-L-1}^L [b_{5,l} + b_{7,l}] f_l \quad (f \in H_L), \quad (3.11a)$$

$$\hat{Q}_I(x) f = \sum_{l=-L-1}^L [b_{3,l} + b_{4,l} + b_{8,l}] f_l, \quad \hat{Q}_{II}(x) f = \sum_{l=-L-1}^L [b_{6,l} + b_{8,l}] f_l \quad (f \in H_L), \quad (3.11b)$$

где при $l = \overline{-L/2, L/2}$

$$\begin{aligned} b_{i,2l}(x) &\equiv \sum_{k=-1}^0 \int_{\rho_{d,s,2l+k}}^{\rho_{d,s,2l+k+1}} a_i \hat{\delta}_{i,2l+k,1} d\rho \quad (i = \overline{1,6}), \\ b_{i,2l-1}(x) &\equiv \sum_{k=-3}^{-2} \int_{\rho_{d,s,2l+k}}^{\rho_{d,s,2l+k+1}} a_i \hat{\delta}_{i,2l+k,2} d\rho + \sum_{k=-1}^0 \int_{\rho_{d,s,2l+k}}^{\rho_{d,s,2l+k+1}} a_i \hat{\delta}_{i,2l+k,0} d\rho \quad (i = \overline{1,6}), \\ \hat{\delta}_{i,2l+k,m}(d, s, \rho) &\equiv \\ &\equiv \sum_{m'=0}^2 \tilde{\delta}_i(d, s, \rho_{d,s,2l+k,m'}) \Lambda_m^{[s_{2l-1}, s_{2l+1}]}(s + \sigma_{d,s}(\rho_{d,s,2l+k,m'})) \Lambda_{m'}^{[\rho_{d,s,2l+k}, \rho_{d,s,2l+k+1}]}(\rho) \quad (m = 0, 1), \\ \hat{\delta}_{i,2l+k,2}(d, s, \rho) &\equiv \\ &\equiv \sum_{m'=0}^2 \tilde{\delta}_i(d, s, \rho_{d,s,2l+k,m'}) \Lambda_2^{[s_{2l-3}, s_{2l+1}]}(s + \sigma_{d,s}(\rho_{d,s,2l+k,m'})) \Lambda_{m'}^{[\rho_{d,s,2l+k}, \rho_{d,s,2l+k+1}]}(\rho), \end{aligned}$$

$$\begin{aligned}
b_{7,2l}(x) &\equiv 2^{-1} \sum_{k=-1}^0 (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j g_{1,l}(x, \beta_{s,2l+k,j}), \\
b_{7,2l-1}(x) &\equiv 2^{-1} \sum_{k=-3}^{-2} (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j g_{2,l}(x, \beta_{s,2l+k,j}) + \\
&\quad + 2^{-1} \sum_{k=-1}^0 (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j g_{0,l}(x, \beta_{s,2l+k,j}), \\
b_{8,2l}(x) &\equiv 2^{-1} \sum_{k=-1}^0 (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j q_{1,l}(x, \beta_{s,2l+k,j}), \\
b_{8,2l-1}(x) &\equiv 2^{-1} \sum_{k=-3}^{-2} (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j q_{2,l}(x, \beta_{s,2l+k,j}) + \\
&\quad + 2^{-1} \sum_{k=-1}^0 (\beta_{s,2l+k+1} - \beta_{s,2l+k}) \sum_{j=1}^{\gamma} \omega_j q_{0,l}(x, \beta_{s,2l+k,j}).
\end{aligned}$$

Интегралы $b_{i,l}(x)$ ($l = \overline{-L-1, L}$, $i = \overline{1, 6}$) вычисляются аналитически.

Теорема 3 позволяет получить по две аппроксимации КП ППС $u(x)$ и НП ППС $w(x)$ при $x \in \Omega_D$:

$$\hat{u}_I(x) \equiv (2\pi)^{-1} \hat{G}_I(x) \mathbf{P}_L v, \quad \hat{u}_{II}(x) \equiv (2\pi)^{-1} \hat{G}_{II}(x) \mathbf{P}_L v, \quad (3.12a)$$

$$\hat{w}_I(x) \equiv (2\pi)^{-1} \hat{Q}_I(x) \mathbf{P}_L v, \quad \hat{w}_{II}(x) \equiv (2\pi)^{-1} \hat{Q}_{II}(x) \mathbf{P}_L v. \quad (3.12b)$$

Функционалы $\hat{G}_I(x)$, $\hat{G}_{II}(x)$, $\hat{Q}_I(x)$, $\hat{Q}_{II}(x)$ могут быть вычислены с помощью формул (3.11), операторы $\mathbf{P}_L$ определяются формулой (3.5). АМА КП ППС $\hat{u}_I(x)$ и АМА НП ППС $\hat{w}_I(x)$ предлагаются в настоящей работе, ПМА КП ППС $\hat{u}_{II}(x)$ и ПМА НП ППС $\hat{w}_{II}(x)$ получены на основе работ [6], [7].

В силу теоремы 2 и неравенства $\|\mathbf{P}_L\| \leq 1$ функционалы $\ln^{-1}|d| \hat{G}_I(x) \mathbf{P}_L$, $d \hat{G}_{II}(x) \mathbf{P}_L$, $\hat{Q}_I(x) \mathbf{P}_L$, $d \hat{Q}_{II}(x) \mathbf{P}_L$ [$C(\partial\Omega) \rightarrow \mathbb{R}$] ($L/2 \in \mathbb{N}$, $x \in \Omega_D$) ограничены в совокупности, поэтому аппроксимации $\ln^{-1}|d| \hat{u}_I(x)$, $d \hat{u}_{II}(x)$, $\hat{w}_I(x)$, $d \hat{w}_{II}(x)$ равномерно устойчивы вблизи границы $\partial\Omega$ к возмущениям функции плотности v в норме $C(\partial\Omega)$. Сформулируем основной результат настоящей работы.

Следствие 2. Пусть $\partial\Omega \in C^{2\gamma+1}$, $\gamma \geq 2$, $\gamma, L/2 \in \mathbb{N}$, $R > 0$, $t \in (0, D]$. Тогда функции $\ln^{-1}|d| \hat{u}_I(x)$, $d \hat{u}_{II}(x)$, $\hat{w}_I(x)$, $d \hat{w}_{II}(x)$ сходятся при $L \rightarrow \infty$ со скоростью не меньше, чем кубическая, к соответствующим функциям $\ln^{-1}|d| u(x)$, $du(x)$, $w(x)$, $dw(x)$ равномерно относительно $x \in \Omega_D$ и функций $v \in C^3(\partial\Omega)$, удовлетворяющих неравенству $\|v\|_{C^3(\partial\Omega)} \leq R$. В частности, функции $\hat{u}_I(\tilde{x}_d(s))$, $\hat{u}_{II}(\tilde{x}_d(s))$, $\hat{w}_{II}(\tilde{x}_d(s))$ сходятся при $L \rightarrow \infty$ к соответствующим функциям $u(\tilde{x}_d(s))$, $u(\tilde{x}_d(s))$, $w(\tilde{x}_d(s))$ равномерно относительно $|d| \in [t/L, D]$, $s \in I_S$ с порядками аппроксимации не ниже, чем $O(L^{-3} \ln L)$, $O(L^{-2})$, $O(L^{-2})$ соответственно. Кроме того, функции $\ln^{-1}|d| \hat{u}_I^\delta(x)$, $d \hat{u}_{II}^\delta(x)$, $\hat{w}_I^\delta(x)$, $d \hat{w}_{II}^\delta(x)$, где $\hat{u}_I^\delta(x) \equiv (2\pi)^{-1} \hat{G}_I(x) \mathbf{P}_L v_\delta$, $\hat{u}_{II}^\delta(x) \equiv (2\pi)^{-1} \hat{G}_{II}(x) \mathbf{P}_L v_\delta$, $\hat{w}_I^\delta(x) \equiv (2\pi)^{-1} \hat{Q}_I(x) \mathbf{P}_L v_\delta$, $\hat{w}_{II}^\delta(x) \equiv (2\pi)^{-1} \hat{Q}_{II}(x) \mathbf{P}_L v_\delta$, сходятся при $L \rightarrow \infty$, $\delta \rightarrow +0$ к соответствующим функциям $\ln^{-1}|d| u(x)$, $du(x)$, $w(x)$, $dw(x)$ равномерно относительно $x \in \Omega_D$ и функций $v \in C^3(\partial\Omega)$, $v_\delta \in C(\partial\Omega)$, удовлетворяющих неравенствам $\|v\|_{C^3(\partial\Omega)} \leq R$, $\|v_\delta - v\|_{C(\partial\Omega)} \leq \delta$.

В заключение раздела рассмотрим традиционные аппроксимации КП и НП ППС на основе формул Гаусса. Для этого построим аппроксимации $\hat{A}_9(x)$, $\hat{A}_{10}(x)$ соответствующих функционалов $\hat{A}_5(x)$, $\hat{A}_6(x)$ на основе формулы Гаусса с μ узлами:

$$\hat{A}_i(x) f \equiv 2^{-1} \sum_{l=-L-1}^L \sum_{2n=0}^1 (\alpha_{s,l+n+1/2} - \alpha_{s,l+n}) \hat{B}_{i,l+n}(d, s) f \quad (i = 9, 10), \quad (3.13)$$

$$\hat{B}_{i,l+n}(d, s)f \equiv \sum_{j=1}^{\mu} \omega_j B_{i-2}(d, s, \alpha_{s,l+n,j}) \ddot{P}_L f,$$

$$\alpha_{s,l+n,j} \equiv 2^{-1}(\alpha_{s,l+n} + \alpha_{s,l+n+1/2}) + 2^{-1}(\alpha_{s,l+n+1/2} - \alpha_{s,l+n}) z_j.$$

Здесь вводится дополнительная точка $\alpha_{s,l+1/2} \equiv 0$ при $s \in [s_l, s_{l+1})$, делящая промежуток численного интегрирования $[\alpha_{s,l}, \alpha_{s,l+1}]$ на два: $[\alpha_{s,l}, \alpha_{s,l+1/2}]$ и $[\alpha_{s,l+1/2}, \alpha_{s,l+1}]$. Это делается для того, чтобы точка s никогда не совпадала с узлами формул Гаусса. В связи с этим для удобства записи считаем, что $\alpha_{s,l+1/2} \equiv \alpha_{s,l}$, если $s \notin [s_l, s_{l+1})$, и в формуле (3.13) делаем двойное суммирование, где во внутренней сумме второе ненулевое слагаемое появляется только при таких l , при которых $s \in (s_l, s_{l+1})$.

Аппроксимации КП ППС $u(x)$ и НП ППС $w(x)$ на множестве Ω_D зададим с помощью формул

$$\begin{aligned} \tilde{u}(x) &\equiv (2\pi)^{-1} \tilde{G}(x) \mathbf{P}_L v, & \tilde{w}(x) &\equiv (2\pi)^{-1} \tilde{Q}(x) \mathbf{P}_L v, \\ \tilde{G}(x) &\equiv \hat{A}_9(x) + \hat{A}_7(x), & \tilde{Q}(x) &\equiv \hat{A}_{10}(x) + \hat{A}_8(x). \end{aligned} \quad (3.14)$$

Операторы $\mathbf{P}_L$ определяются формулой (3.5). Функционалы $\tilde{G}(x)$, $\tilde{Q}(x)$ имеют вид матриц-строк длиной $2L + 2$:

$$\tilde{G}(x)f = \sum_{l=-L-1}^L [b_{9,l} + b_{7,l}] f_l, \quad \tilde{Q}(x)f = \sum_{l=-L-1}^L [b_{10,l} + b_{8,l}] f_l,$$

где при $l = \overline{-L/2, L/2}$

$$\begin{aligned} b_{9,2l}(x) &\equiv 2^{-1} \sum_{k=-1}^0 \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j g_{1,l}(x, \alpha_{s,2l+k+n,j}), \\ b_{9,2l-1}(x) &\equiv 2^{-1} \sum_{k=-3}^{-2} \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j g_{2,l}(x, \alpha_{s,2l+k+n,j}) + \\ &+ 2^{-1} \sum_{k=-1}^0 \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j g_{0,l}(x, \alpha_{s,2l+k+n,j}), \\ b_{10,2l}(x) &\equiv 2^{-1} \sum_{k=-1}^0 \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j q_{1,l}(x, \alpha_{s,2l+k+n,j}), \\ b_{10,2l-1}(x) &\equiv 2^{-1} \sum_{k=-3}^{-2} \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j q_{2,l}(x, \alpha_{s,2l+k+n,j}) + \\ &+ 2^{-1} \sum_{k=-1}^0 \sum_{2n=0}^1 (\alpha_{s,2l+k+n+1/2} - \alpha_{s,2l+k+n}) \sum_{j=1}^{\gamma} \omega_j q_{0,l}(x, \alpha_{s,2l+k+n,j}). \end{aligned}$$

Вычисление функций $b_{7,l}$, $b_{8,l}$ см. формулы (3.11). Аппроксимации $\tilde{u}(x)$, $\tilde{w}(x)$ можно считать традиционными аппроксимациями КП ППС $u(x)$ и НП ППС $w(x)$, так как все интегралы $\tilde{g}_l(x)$, $\tilde{q}_l(x)$, в том числе и на промежутках $[\alpha_{s,l}, \alpha_{s,l+1}]$ ($\alpha_{s,l} < \alpha_{s,l+1}$), вычисляются с помощью формул Гаусса.

При условии $\partial\Omega \in C^{n+1}$ ($n \in \mathbb{Z}_+$) для любой замкнутой области $\Omega'_D \subset \Omega_D$ можно определить константы $\tilde{c}'_{9,j} \equiv \sup_{x \in \Omega'_D, s' \in I_S} |\partial_{s'}^j g(x, s')|$, $\tilde{c}'_{10,j} \equiv \sup_{x \in \Omega'_D, s' \in I_S} |\partial_{s'}^j q(x, s')|$ ($j = \overline{0, n}$) и получить для функционалов $\hat{A}_9(x)$, $\hat{A}_{10}(x)$ оценки, аналогичные (3.10). Тогда по аналогии со следствием 2 получаем следующее утверждение.

Следствие 3. Пусть $\partial\Omega \in C^{2\mu+1}$, $\mu \geq \gamma \geq 2$, $\mu, \gamma, L/2 \in \mathbb{N}$, $R > 0$. Тогда функции $\tilde{u}(x)$, $\tilde{w}(x)$ сходятся при $L \rightarrow \infty$ со скоростью не меньше, чем кубическая, к соответствующим функциям $u(x)$, $w(x)$ равномерно относительно $x \in \Omega'_D$ и функций $v \in C^3(\partial\Omega)$, удовлетворяющих неравенству $\|v\|_{C^3(\partial\Omega)} \leq R$. Функции $\tilde{u}_\delta(x) \equiv (2\pi)^{-1} \tilde{G}(x) \mathbf{P}_L v_\delta$, $\tilde{w}_\delta(x) \equiv (2\pi)^{-1} \tilde{Q}(x) \mathbf{P}_L v_\delta$ сходятся при $L \rightarrow \infty$, $\delta \rightarrow +0$ к соответствующим функциям $u(x)$, $w(x)$ равномерно относительно $x \in \Omega'_D$ и функций $v \in C^3(\partial\Omega)$, $v_\delta \in C(\partial\Omega)$, удовлетворяющих неравенствам $\|v\|_{C^3(\partial\Omega)} \leq R$, $\|v_\delta - v\|_{C(\partial\Omega)} \leq \delta$.

4. АППРОКСИМАЦИИ ПСИ

Интегралы на граничных элементах, возникающие при вычислении аппроксимаций $\hat{A}_i(x)\mathbf{P}_L f$ и точных функций $A_i(x)f$ ($i = \overline{1,6}$), при $f = 1$ имеют вид

$$\hat{J}_i \equiv \int_{\rho_0}^{\rho_1} a_i(d, \rho) \hat{\delta}_i(d, s, \rho) d\rho, \quad J_i \equiv \int_{\rho_0}^{\rho_1} a_i(d, \rho) \tilde{\delta}_i(d, s, \rho) d\rho \quad (d \in I_D, s \in I_S),$$

$$\hat{\delta}_i(d, s, \rho) \equiv \sum_{m=0}^2 \tilde{\delta}_i(d, s, \rho_m) \Lambda_m(\rho, \rho_0, \rho_1);$$

$$\rho_0 \equiv \rho_{d,s}(\alpha h), \quad \rho_1 \equiv \rho_{d,s}((\alpha + 1)h), \quad \rho_2 \equiv 2^{-1}(\rho_0 + \rho_1).$$

Если $d = th$ и при этом величины $\alpha, t \in \mathbb{R}$ ограничены и выполняется хотя бы одно из условий: $\alpha(\alpha + 1) > 0$ или $t \neq 0$, то интегралы J_i можно считать ПСИ, так как точка наблюдения $\tilde{x}_d(s)$ не находится на граничном элементе $[\tilde{x}(\alpha h), \tilde{x}((\alpha + 1)h)] \subset \partial\Omega$, но расстояние между ними неограниченно уменьшается при $h \rightarrow +0$. Заметим, что $J_1 + J_2 = J_5 \equiv J_G$, $J_3 + J_4 = J_6 \equiv J_Q$.

Определение. Вещественные функции $f_1(t, h), f_2(t, h)$ эквивалентны при $h \rightarrow +0, t \in [t_1, t_2]$ с порядком аппроксимации $k > 0$, если существуют постоянные $c_1, c_2, h_0 > 0$, такие что при любых $t \in [t_1, t_2], h \in (0, h_0]$ выполняются неравенства $c_1 h^k \leq |f_1(t, h) - f_2(t, h)| \leq c_2 h^k$. Функции $f_1(t, h), f_2(t, h)$ не эквивалентны при $h \rightarrow +0, t \in [t_1, t_2]$, если существуют постоянные $c_1, h_0 > 0$, такие что при любых $t \in [t_1, t_2], h \in (0, h_0]$ выполняется неравенство $|f_1(t, h) - f_2(t, h)| \geq c_1$.

Теорема 4. Пусть $\partial\Omega \in C^6$, значения $s \in I_S, \alpha \in [-2, -1) \cup (0, 1]$ фиксированы, $(\partial_\rho^3 \tilde{\delta}_i)(0, s, 0) \neq 0$ ($i = 1, 4, 5, 6$). Тогда

(i) существует достаточно малое число $t_1 > 0$, такое что при $h \rightarrow +0, d = th, t \in [-t_1, t_1]$ функции $\hat{J}_{G,I}(t, h) \equiv \hat{J}_1 + \hat{J}_2$ и $J_G(t, h)$ эквивалентны с третьим порядком аппроксимации, а функции $\hat{J}_{G,II}(t, h) \equiv \hat{J}_5, \hat{J}_{Q,II}(t, h) \equiv \hat{J}_6$ эквивалентны соответствующим функциям $J_G(t, h), J_Q(t, h)$ со вторым порядком аппроксимации;

(ii) существует достаточно малое число $t_0 > 0$, такое что для любого $t_2 \in (0, t_0]$ функции $\hat{J}_{Q,I}(t, h) \equiv \hat{J}_3 + \hat{J}_4$ и $J_Q(t, h)$ при $h \rightarrow +0, d = th, |t| \in [t_2, t_0]$ эквивалентны с третьим порядком аппроксимации.

Доказательство. Введем в рассмотрение функции $z_m(d, h) \equiv \rho_m/h$ ($m = \overline{0,2}$) при $d \in \overline{I_D}, h \in (0, \Sigma_s/2]$ ($\Sigma_s \equiv \min\{-\Sigma'_s, \Sigma''_s\}$). Доопределим функции z_0, z_1, z_2 при $h = 0, d \in \overline{I_D}$: $z_0(d, 0) \equiv \alpha \rho'_{d,s}(0), z_1(d, 0) \equiv (\alpha + 1) \rho'_{d,s}(0), z_2(d, 0) \equiv 2^{-1}(z_0 + z_1)$, тогда в силу теоремы 1 функции z_m при $d \in \overline{I_D}, h \in [0, \Sigma_s/2]$ непрерывны и справедливы оценки $z_1 - z_0 \geq z'_0, \alpha z'_0 \leq z_0 < z_1 \leq (\alpha + 1)z'_1$, где $z'_0 \equiv \inf_{s'-s \in \Xi_s, d \in \overline{I_D}} \partial_{s'} \rho > 0, z'_1 \equiv \sup_{s'-s \in \Xi_s, d \in \overline{I_D}} \partial_{s'} \rho$.

После замены переменной $\rho = zh$ в интегралах $\hat{J}_i, J_i$ ($i = \overline{1,6}$) остаточные члены $\Delta \hat{J}_i \equiv J_i - \hat{J}_i$ при $d = th, |t| \leq 2D/\Sigma_s, h \in (0, \Sigma_s/2]$ могут быть записаны в следующем виде:

$$\Delta \hat{J}_i(t, h) = \eta_i(h) \int_{z_0(th, h)}^{z_1(th, h)} a_i(t, z) [\tilde{\delta}_i(th, s, zh) - \hat{\delta}_i(th, s, zh)] dz,$$

где $\eta_1(h) = \eta_2(h) = \eta_4(h) \equiv 1, \eta_3(h) \equiv h, \eta_5(h) = \eta_6(h) \equiv h^{-1}$. В силу формулы (2.10) при $\rho = zh, z \in (z_0, z_2) \cup (z_2, z_1), d \in \overline{I_D}, h \in (0, \Sigma_s/2]$ имеем равенства

$$\Delta \hat{\delta}_i = h^3 p_3(z, z_0, z_1) \tilde{\delta}_{i,3}(d, s, z_0 h, z_1 h, z_2 h, z_2 h) + h^4 p_4(z, z_0, z_1) \tilde{\delta}_{i,4}(d, s, zh, z_0 h, z_1 h, z_2 h, z_2 h),$$

где $\Delta \hat{\delta}_i(d, \rho) \equiv \tilde{\delta}_i - \hat{\delta}_i, i = \overline{1,6}; \tilde{\delta}_{i,3}, \tilde{\delta}_{i,4}$ — соответствующие разделенные разности 3-го и 4-го порядков. Заметим, что $\tilde{\delta}_{i,3}$ не зависят от z , а функции $a_i(t, z) p_4(z, z_0, z_1)$ в силу условия $z_0 z_1 > 0$ не меняют знака при $z \in [z_0, z_1]$ и фиксированных t . С учетом формулы (2.12) и обобщенной теоремы о среднем (см. [24, п. 304]) имеем при $|t| \leq 2D/\Sigma_s, h \in (0, \Sigma_s/2]$

$$\Delta \hat{J}_i = \eta_i(h) h^3 \tilde{\delta}_{i,3}(t, \zeta_{i,1}, h) j_{i,3}(t, h) + \eta_i(h) h^4 \tilde{\delta}_{i,4}(t, \zeta_{i,2}, h) j_{i,4}(t, h) \quad (i = \overline{1,6}); \quad (4.1)$$

$$j_{i,3}(t, h) \equiv \int_{\tilde{z}_0(t, h)}^{\tilde{z}_1(t, h)} a_i(t, z) p_3(z, \tilde{z}_0, \tilde{z}_1) dz, \quad j_{i,4}(t, h) \equiv \int_{\tilde{z}_0(t, h)}^{\tilde{z}_1(t, h)} a_i(t, z) p_4(z, \tilde{z}_0, \tilde{z}_1) dz;$$

$\tilde{z}_m(t, h) \equiv z_m(th, h)$ ($m = \overline{0, 2}$); $\tilde{\delta}_{i,3}(t, z, h) \equiv 6^{-1} \left(\partial_\rho^3 \tilde{\delta}_i \right) (th, s, zh)$, $\tilde{\delta}_{i,4}(t, z, h) \equiv 24^{-1} \left(\partial_\rho^4 \tilde{\delta}_i \right) (th, s, zh)$; $\zeta_{i,1}$, $\zeta_{i,2}$ — некоторые точки на промежутке $[\alpha z'_0, (\alpha + 1)z'_1]$, точное положение которых зависит от t, h .

В условиях теоремы в силу следствия 1 существуют константы $h_0 \in (0, \Sigma_s/2]$, $t_0 \in (0, D/h_0]$, такие что функции $\tilde{\delta}_{i,3}$, $\tilde{\delta}_{i,4}$ ($i = \overline{1, 6}$) непрерывны и $\tilde{\delta}_{i,3} \neq 0$ ($i = 1, 4, 5, 6$) при $t \in [-t_0, t_0]$, $z \in [\alpha z'_0, (\alpha + 1)z'_1]$, $h \in [0, h_0]$. Так как при $|t| \leq 2D/\Sigma_s$, $h \in [0, \Sigma_s/2]$ функции $\tilde{z}_0$, $\tilde{z}_1$ непрерывны и $\tilde{z}_0 \tilde{z}_1 > 0$, то функции $j_{i,3}$, $j_{i,4}$ ($i = \overline{1, 6}$) при этом также непрерывны.

При $t = 0$ и фиксированных $h \in [0, \Sigma_s/2]$ для функций $f_i(z) \equiv a_i p_3$ ($i = 1, 5, 6$) в силу условий $\tilde{z}_1 - \tilde{z}_0 > 0$, $\tilde{z}_0 \tilde{z}_1 > 0$ имеют место неравенства $f_i(\tilde{z}_0 + \alpha) > -f_i(\tilde{z}_1 - \alpha)$, где $0 < \alpha < 2^{-1}(\tilde{z}_1 - \tilde{z}_0)$. Поэтому при $h \in [0, \Sigma_s/2]$ выполняются неравенства $j_{i,3}(0, h) > 0$ ($i = 1, 5, 6$). Также при $h \in [0, \Sigma_s/2]$ выполняется равенство $j_{2,3}(0, h) = 0$. Следовательно, существуют постоянные $t_1 \in (0, t_0]$, $c_1, c_2 > 0$, такие что при $t \in [-t_1, t_1]$, $h \in (0, h_0]$ справедливы оценки

$$\begin{aligned} c_1 &\leq \left| \tilde{\delta}_{1,3}(t, \zeta_{1,1}, h) j_{1,3}(t, h) + \tilde{\delta}_{2,3}(t, \zeta_{2,1}, h) j_{2,3}(t, h) \right| \leq c_2, \\ c_1 &\leq \left| \tilde{\delta}_{k,3}(t, \zeta_{k,1}, h) j_{k,3}(t, h) \right| \leq c_2 \quad (k = 5, 6), \\ \left| \tilde{\delta}_{i,4}(t, \zeta_{i,2}, h) j_{i,4}(t, h) \right| &\leq c_2 \quad (i = 1, 2, 5, 6). \end{aligned} \quad (4.2)$$

С учетом формул (4.1) и оценок (4.2) утверждение (i) доказано.

При $0 \neq |t| \leq 2D/\Sigma_s$, $h \in [0, \Sigma_s/2]$ выполняется неравенство $j_{4,3}(t, h) > 0$. Пусть $t_2 \in (0, t_0]$. Тогда существуют постоянные $c_3, c_4 > 0$, такие что при $|t| \in [t_2, t_0]$, $h \in (0, h_0]$ справедливы оценки

$$\begin{aligned} c_3 &\leq \left| \tilde{\delta}_{4,3}(t, \zeta_{4,1}, h) j_{4,3}(t, h) \right| \leq c_4, \\ \left| \tilde{\delta}_{3,3}(t, \zeta_{3,1}, h) j_{3,3}(t, h) \right| &\leq c_4, \quad \left| \tilde{\delta}_{i,4}(t, \zeta_{i,2}, h) j_{i,4}(t, h) \right| \leq c_4 \quad (i = 3, 4). \end{aligned} \quad (4.3)$$

С учетом формул (4.1) и оценок (4.3) утверждение (ii) также доказано. Теорема 4 полностью доказана.

Обозначим через $\tilde{J}_{G,\mu}$, $\tilde{J}_{Q,\mu}$ аппроксимации соответствующих интегралов $J_G = \int_{ah}^{(\alpha+1)h} g(\tilde{x}_d(s), s + \sigma) d\sigma$, $J_Q = \int_{ah}^{(\alpha+1)h} q(\tilde{x}_d(s), s + \sigma) d\sigma$, полученные с помощью формулы Гаусса с μ узлами. Они возникают при построении приближенных функций $\hat{A}_9(x) \mathbf{P}_L f$, $\hat{A}_{10}(x) \mathbf{P}_L f$ при $f = 1$.

Теорема 5. Пусть $\partial\Omega \in C^{2\mu+1}$, значения $\mu \in \mathbb{N}$, $s \in I_S$, $\alpha \in [-2, -1) \cup (0, 1]$ фиксированы. Тогда

(i) существует достаточно малое число $t_0 > 0$, такое что при $h \rightarrow +0$, $d = th$, $t \in [-t_0, t_0]$ функции $\tilde{J}_{G,\mu}(t, h)$ и $J_G(t, h)$ не эквивалентны;

(ii) существует достаточно малое число $t_1 > 0$, такое что для любого $t_2 \in (0, t_1]$ функции $\tilde{J}_{Q,\mu}(t, h)$ и $J_Q(t, h)$ при $h \rightarrow +0$, $d = th$, $|t| \in [t_2, t_1]$ не эквивалентны.

Доказательство. Остаточные члены $\Delta \tilde{J}_{G,\mu} \equiv J_G - \tilde{J}_{G,\mu}$, $\Delta \tilde{J}_{Q,\mu} \equiv J_Q - \tilde{J}_{Q,\mu}$ имеют вид $\Delta \tilde{J}_{G,\mu} = r_G(d, \zeta_G, h)$, $\Delta \tilde{J}_{Q,\mu} = r_Q(d, \zeta_Q, h)$ ($d \in \overline{I_D}$, $h \in (0, S/2]$), где

$$r_G(d, \sigma, h) \equiv c_\mu h^{2\mu+1} \partial_\sigma^{2\mu} g(\tilde{x}_d(s), s + \sigma), \quad r_Q(d, \sigma, h) \equiv c_\mu h^{2\mu+1} \partial_\sigma^{2\mu} q(\tilde{x}_d(s), s + \sigma); \quad (4.4)$$

$c_\mu \equiv (\mu!)^4 [(2\mu)!]^{-3} (2\mu + 1)^{-1}$; ζ_G, ζ_Q — некоторые точки на промежутке $[ah, (\alpha + 1)h]$, точное положение которых зависит от d, h (см. [2, гл. 3, § 5, п. 2]). Согласно равенствам (2.8) при $d \in \overline{I_D}$, $\sigma \in [ah, (\alpha + 1)h]$, $h \in (0, S/2]$ имеем формулы

$$g(\tilde{x}_d(s), s + \sigma) \equiv (-\sigma \tilde{\psi}_1 - d \sigma \tilde{\psi}_3) / (\sigma^2 \tilde{\psi} + d^2), \quad (4.5a)$$

$$q(\tilde{x}_d(s), s + \sigma) \equiv (-\sigma^2 \tilde{\psi}_2 - d) / (\sigma^2 \tilde{\psi} + d^2), \quad (4.5b)$$

где $\tilde{\psi}_i(\sigma) \equiv \psi_i(s, s + \sigma)$ ($i = \overline{1, 3}$), $\tilde{\psi}(d, \sigma) \equiv \psi(d, s, s + \sigma)$. В условиях теоремы производные $\tilde{\psi}_i^{(j)}$ ($i = \overline{1, 3}$), $\partial_0^j \tilde{\psi}$ ($j = \overline{0, 2\mu}$) непрерывны и $\tilde{\psi} \neq 0$ при $\sigma \in \overline{I_S}$, $d \in \overline{I_D}$, поэтому в силу формул (4.5) и условия $\alpha(\alpha + 1) > 0$ функции r_G , r_Q при $d = th$, $\sigma = yh$, $y \in [\alpha, (\alpha + 1)]$, $|t| \leq 2D/S$, $h \in (0, S/2]$ могут быть представлены в виде

$$r_G(th, yh, h) = -c_\mu r'_G(t, y, h) + o_G(t, y, h), \quad (4.6a)$$

$$r_Q(th, yh, h) = -c_\mu t r'_Q(t, y, h) + o_Q(t, y, h); \quad (4.6b)$$

$$r'_G \equiv \partial_y^{2\mu} \left[y \tilde{\psi}_1(yh) (y^2 \tilde{\psi}(th, yh) + t^2)^{-1} \right], \quad r'_Q \equiv \partial_y^{2\mu} \left[(y^2 \tilde{\psi}(th, yh) + t^2)^{-1} \right];$$

$o_G, o_Q \rightarrow 0$ при $h \rightarrow +0$ равномерно по $y \in [\alpha, (\alpha + 1)]$, $|t| \leq 2D/S$. При $t = 0$, $y \in [\alpha, (\alpha + 1)]$, $h \in (0, S/2]$ имеем равенства

$$r'_G(0, y, h) = \partial_y^{2\mu} \left[\tilde{\psi}_1(yh) (y \tilde{\psi}(0, yh))^{-1} \right] = (y^{-1})^{(2\mu)} \tilde{\psi}_1(yh) / \tilde{\psi}(0, yh) + o'_G(y, h), \quad (4.7a)$$

$$r'_Q(0, y, h) = \partial_y^{2\mu} \left[(y^2 \tilde{\psi}(0, yh))^{-1} \right] = (y^{-2})^{(2\mu)} / \tilde{\psi}(0, yh) + o'_Q(y, h), \quad (4.7b)$$

где $o'_G, o'_Q \rightarrow 0$ при $h \rightarrow +0$ равномерно по $y \in [\alpha, (\alpha + 1)]$.

Так как $\tilde{\psi}_1(0) = -1$, то в силу формул (4.6a), (4.7a) и непрерывности функции r'_G при $|t| \leq 2D/S$, $y \in [\alpha, (\alpha + 1)]$, $h \in (0, S/2]$ существуют постоянные $h_0 \in (0, S/2]$, $t_0 \in (0, D/h_0]$, $c_0 > 0$, такие что $|r_G(d, \sigma, h)| \geq c_0$ при $d = th$, $t \in [-t_0, t_0]$, $\sigma \in [\alpha h, (\alpha + 1)h]$, $h \in (0, h_0]$. Утверждение (i) доказано.

В силу формулы (4.7b) и непрерывности функции r'_Q при $|t| \leq 2D/S$, $y \in [\alpha, (\alpha + 1)]$, $h \in (0, S/2]$ существуют постоянные $h_1 \in (0, S/2]$, $t_1 \in (0, D/h_1]$, $c_1 > 0$, такие что $|r'_Q(t, y, h)| \geq c_1$ при $t \in [-t_1, t_1]$, $y \in [\alpha, \alpha + 1]$, $h \in (0, h_1]$. Поэтому в силу формулы (4.6b) для любого $t_2 \in (0, t_1]$ существуют постоянные $h_2 \in (0, h_1]$, $c_2 \in (0, t_2 c_1]$, такие что $|r_Q(d, \sigma, h)| \geq c_2$ при $d = th$, $|t| \in [t_2, t_1]$, $\sigma \in [\alpha h, (\alpha + 1)h]$, $h \in (0, h_2]$. Утверждение (ii) доказано. Доказательство теоремы 5 завершено.

Заметим, что остаточные члены аппроксимаций интегралов J на основе квадратурных формул Ньютона—Котеса (КФНК) также допускают представление типа (4.4) (см. [2, гл. 3, § 4, п. 2]), где порядок производной на единицу меньше степени длины промежутка интегрирования. Поэтому теорема 5 справедлива и для аппроксимаций на основе КФНК.

5. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Рассмотрим вычисление КП и НП ППС с плотностью $v(s) = \cos s$ во внешности единичного круга. Пусть $x = \tilde{x}_d(s)$, $d \in I_D$, $s \in I_S$. Точные значения КП ППС $\bar{u}(x)$ и НП ППС $\bar{w}(x)$ вычисляются по формулам

$$\bar{u}(x) = -2^{-1} (1 - d)^{-1} \sin s, \quad \bar{w}(x) = 2^{-1} (1 - d)^{-2} \cos s.$$

Полуаналитические аппроксимации $\hat{u}_I(x)$, $\hat{u}_{II}(x)$, $\hat{w}_I(x)$, $\hat{w}_{II}(x)$ вычисляются с помощью формул (3.12), при этом для вычисления функционалов $\hat{A}_7(x)$, $\hat{A}_8(x)$ используется формула Гаусса с $\gamma = 2$ узлами. Традиционные аппроксимации $\tilde{u}_i(x)$, $\tilde{w}_i(x)$ вычисляются с помощью формул (3.14), при этом для вычисления функционалов $\hat{A}_9(x)$, $\hat{A}_{10}(x)$ используется формула Гаусса с $\mu_i = 12 \cdot 2^{i-1}$ узлами ($i = \overline{1, 3}$), а для вычисления функционалов $\hat{A}_7(x)$, $\hat{A}_8(x)$ по прежнему используется формула Гаусса с $\gamma = 2$ узлами (веса и абсциссы формул Гаусса с большим числом узлов см., например, [26, гл. 17, табл. 1]). Для данной геометрии $D = 2/(3\pi)$, $\Sigma''_s = -\Sigma'_s = \arcsin(2/(3\pi))$. Вычисления осуществляем с двойной точностью.

Для экспериментального определения точности аппроксимаций вычисляем $\Delta \hat{u}_I$, $\Delta \hat{u}_{II}$, $\Delta \tilde{u}_i$ и $\Delta \hat{w}_I$, $\Delta \hat{w}_{II}$, $\Delta \tilde{w}_i$ — максимумы модулей абсолютных погрешностей приближенных решений $\hat{u}_I(x)$, $\hat{u}_{II}(x)$, $\tilde{u}_i(x)$ и $\hat{w}_I(x)$, $\hat{w}_{II}(x)$, $\tilde{w}_i(x)$ на множествах точек вида $\{x_{l,j} \equiv \tilde{x}_{d_m}(s_{l,j}), j = \overline{0, 7}, l = \overline{-2^{k+1} + 1, 2^{k+1} - 2}\}$, где $s_{l,j}$ принимает значения $(l \pm \alpha_j) h_k$, d_m принимает значения $t_m h_k$ при фиксированных $k \in \{\overline{1, 6}\}$, $m \in \{-2, -3, -4, -5, -14\}$, при этом

$\alpha_0 \equiv 0$, $\alpha_j \equiv 10^{-(6-j)}$ при $j = \overline{1, 5}$, $\alpha_6 \equiv 0.25$, $\alpha_7 \equiv 0.5$; $h_k \equiv \pi/(2^{k+1} - 1)$, $t_m \equiv -10^m$. Например, $\Delta \hat{u}_I \equiv \max_{l,j} |\hat{u}_I(x_{l,j}) - \bar{u}(x_{l,j})|$. В табл. 1 в каждой основной ячейке представлены значения $\Delta \hat{u}_I$, $\Delta \hat{u}_{II}$, $\Delta \tilde{u}_1$, $\Delta \tilde{u}_2$, $\Delta \tilde{u}_3$, а в табл. 2 — значения $\Delta \hat{w}_I$, $\Delta \hat{w}_{II}$, $\Delta \tilde{w}_1$, $\Delta \tilde{w}_2$, $\Delta \tilde{w}_3$ в соответствующем порядке сверху вниз.

Вычисляя на основании табл. 1 и табл. 2 степени ν скоростей сходимости аппроксимаций КП и НП ППС (например, для $\hat{u}_I$ степень скорости сходимости $\nu = \ln[\Delta \hat{u}_I(h_k)/\Delta \hat{u}_I(h_{k+1})]/\ln(h_k/h_{k+1})$ при переходе от h_k к h_{k+1}), можно заметить, что при $h \rightarrow +0$, $d = th$ и фиксированных $t \in I_D$ АМА КП ППС $\hat{u}_I$ и АМА НП ППС $\hat{w}_I$, предлагаемые в настоящей работе, обладают скоростями сходимости соответственно чуть меньше кубической и кубической, что хорошо согласуется с утверждениями следствия 2 и теоремы 4. ПМА КП ППС $\hat{u}_{II}$ и ПМА НП ППС $\hat{w}_{II}$, полученные на основе работ [6], [7], обладают квадратичными скоростями сходимости при $h \rightarrow +0$, $d = th$ и фиксированных $t \in I_D$, что также согласуется со следствием 2 и теоремой 4. Таким образом, вычислительные эксперименты подтверждают теоретические выводы, и АМА КП ППС, АМА НП ППС обладают более равномерной сходимостью в приграничной области, чем ПМА КП ППС, ПМА НП ППС соответственно.

Традиционные аппроксимации КП ППС $\tilde{u}_i$ и НП ППС $\tilde{w}_i$ на основе формул Гаусса имеют нулевые скорости сходимости при $h \rightarrow +0$, $d = th$ и фиксированных $t \in I_D$, что согласуется с теоремой 5. Здесь вычислительные эксперименты подтверждают известный факт, что традиционные аппроксимации КП и НП ППС на основе формул Гаусса обладают сильно неравномерной сходимостью в приграничной области, и для достижения удовлетворительной точности требуется избыточная дискретизация или увеличение порядка аппроксимации формул Гаусса.

При увеличении $|t|$ при фиксированных h скорости сходимости аппроксимаций $\tilde{u}_i$ и $\tilde{w}_i$ возрастают до кубических, что согласуется со следствием 3, так как при этом происходит увеличение расстояния $|d| = |t|h$ от точки наблюдения до границы, и на достаточном удалении от границы сходимость аппроксимаций КП и НП ППС становится равномерной.

Заметим, что значения $\Delta \tilde{w}_i$ при $t \rightarrow 0$, $d = th$ и фиксированных h , μ приближаются к 0.5 — наибольшей абсолютной величине разрыва НП ППС с плотностью $v = \cos \varphi$ на границе $\partial\Omega$. Это объясняется непрерывностью аппроксимаций традиционных НП ППС на основе формул Гаусса в граничных точках, не совпадающих с узлами формулы Гаусса, и конечным разрывом точных значений НП ППС на границе. Поэтому сходимость традиционных аппроксимаций НП ППС на основе формул Гаусса не может быть равномерной, поскольку в достаточной близости от границы всегда сохраняется абсолютная погрешность, близкая к величине разрыва.

6. ЗАКЛЮЧЕНИЕ

В настоящей работе исследованы полуаналитические аппроксимации КП и НП ППС: АМА и ПМА, позволяющие вычислить КП и НП ППС в любой точке двумерной приграничной области, не включающей саму границу, с конечной полушириной, зависящей от кривизны граничной кривой. В этих аппроксимациях для вычисления интегралов, возникающих после ККИ функции плотности, используются точное интегрирование по переменной ρ и ККИ медленно изменяющихся подынтегральных функций, если граничные элементы расположены вблизи точки наблюдения, и используются формулы Гаусса с порядком аппроксимации не ниже пятого, если граничные элементы расположены достаточно далеко от точки наблюдения. При этом установлено, что использование АМА, предлагаемой в настоящей работе, предпочтительнее ПМА, описанной в работах [6], [7], а именно, скорости равномерной сходимости АМА КП ППС и АМА НП ППС вблизи границы выше, чем соответствующие скорости равномерной сходимости ПМА КП ППС и ПМА НП ППС. В частности, если расстояние от наблюдаемой точки до граничного элемента уменьшается пропорционально h , то скорости сходимости АМА КП ППС и АМА НП ППС почти кубическая и кубическая соответственно, а скорости сходимости ПМА КП ППС и ПМА НП ППС квадратичные. Также установлено, что использование исключительно формул Гаусса для вычисления интегралов, возникающих после ККИ функции плотности, невозможно вблизи границы вследствие серьезного нарушения равномерной сходимости. В частности, если расстояние от наблюдаемой точки до граничного элемента уменьшается пропорционально h , то сходимость традиционных аппроксимаций КП и НП ППС на основе формул Гаусса отсутствует.

Таблица 1. Максимумы модулей абсолютных погрешностей аппроксимаций КП ППС

$-t_m$	$h_1 = \pi/3$	$h_1 = \pi/7$	$h_1 = \pi/15$	$h_1 = \pi/31$	$h_1 = \pi/63$	$h_1 = \pi/127$
10^{-2}	$3.32 \cdot 10^{-2}$	$2.82 \cdot 10^{-3}$	$3.16 \cdot 10^{-4}$	$3.69 \cdot 10^{-5}$	$4.44 \cdot 10^{-6}$	$5.48 \cdot 10^{-7}$
	$3.40 \cdot 10^{-2}$	$5.58 \cdot 10^{-3}$	$3.17 \cdot 10^{-3}$	$7.35 \cdot 10^{-4}$	$1.77 \cdot 10^{-4}$	$4.35 \cdot 10^{-5}$
	$3.32 \cdot 10^{-2}$	$3.63 \cdot 10^{-3}$	$2.34 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$	$2.34 \cdot 10^{-2}$	$2.33 \cdot 10^{-2}$
	$3.32 \cdot 10^{-2}$	$2.82 \cdot 10^{-3}$	$8.03 \cdot 10^{-4}$	$7.33 \cdot 10^{-4}$	$7.32 \cdot 10^{-4}$	$7.31 \cdot 10^{-4}$
	$3.32 \cdot 10^{-2}$	$2.82 \cdot 10^{-3}$	$3.21 \cdot 10^{-4}$	$3.77 \cdot 10^{-5}$	$4.57 \cdot 10^{-6}$	$6.76 \cdot 10^{-7}$
10^{-3}	$3.39 \cdot 10^{-2}$	$2.94 \cdot 10^{-3}$	$3.25 \cdot 10^{-4}$	$3.81 \cdot 10^{-5}$	$4.60 \cdot 10^{-6}$	$5.73 \cdot 10^{-7}$
	$3.63 \cdot 10^{-2}$	$8.84 \cdot 10^{-3}$	$6.30 \cdot 10^{-3}$	$1.47 \cdot 10^{-3}$	$3.56 \cdot 10^{-4}$	$8.75 \cdot 10^{-5}$
	$5.88 \cdot 10^{-2}$	$5.68 \cdot 10^{-2}$	$1.27 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$
	$3.45 \cdot 10^{-2}$	$3.95 \cdot 10^{-2}$	$4.67 \cdot 10^{-2}$	$4.68 \cdot 10^{-2}$	$4.68 \cdot 10^{-2}$	$4.68 \cdot 10^{-2}$
	$3.39 \cdot 10^{-2}$	$3.51 \cdot 10^{-3}$	$4.70 \cdot 10^{-3}$	$4.96 \cdot 10^{-3}$	$4.96 \cdot 10^{-3}$	$4.96 \cdot 10^{-3}$
10^{-4}	$3.40 \cdot 10^{-2}$	$2.95 \cdot 10^{-3}$	$3.26 \cdot 10^{-4}$	$3.83 \cdot 10^{-5}$	$4.61 \cdot 10^{-6}$	$5.66 \cdot 10^{-7}$
	$3.83 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	$9.49 \cdot 10^{-3}$	$2.22 \cdot 10^{-3}$	$5.37 \cdot 10^{-4}$	$1.32 \cdot 10^{-4}$
	$2.54 \cdot 10^{-1}$	$3.86 \cdot 10^{-1}$	$4.97 \cdot 10^{-1}$	$4.97 \cdot 10^{-1}$	$4.97 \cdot 10^{-1}$	$4.97 \cdot 10^{-1}$
	$7.37 \cdot 10^{-2}$	$1.90 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$
	$5.53 \cdot 10^{-2}$	$4.14 \cdot 10^{-2}$	$6.89 \cdot 10^{-2}$	$6.89 \cdot 10^{-2}$	$6.89 \cdot 10^{-2}$	$6.89 \cdot 10^{-2}$
10^{-5}	$3.40 \cdot 10^{-2}$	$2.95 \cdot 10^{-3}$	$3.26 \cdot 10^{-4}$	$3.83 \cdot 10^{-5}$	$4.62 \cdot 10^{-6}$	$5.75 \cdot 10^{-7}$
	$4.10 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	$3.09 \cdot 10^{-3}$	$7.45 \cdot 10^{-4}$	$1.83 \cdot 10^{-4}$
	$5.93 \cdot 10^{-1}$	$6.72 \cdot 10^{-1}$	$7.21 \cdot 10^{-1}$	$7.21 \cdot 10^{-1}$	$7.21 \cdot 10^{-1}$	$7.21 \cdot 10^{-1}$
	$4.20 \cdot 10^{-1}$	$5.54 \cdot 10^{-1}$	$6.70 \cdot 10^{-1}$	$6.70 \cdot 10^{-1}$	$6.70 \cdot 10^{-1}$	$6.70 \cdot 10^{-1}$
	$1.70 \cdot 10^{-1}$	$3.01 \cdot 10^{-1}$	$4.17 \cdot 10^{-1}$	$4.17 \cdot 10^{-1}$	$4.17 \cdot 10^{-1}$	$4.17 \cdot 10^{-1}$
10^{-14}	$3.40 \cdot 10^{-2}$	$2.95 \cdot 10^{-3}$	$3.27 \cdot 10^{-4}$	$3.83 \cdot 10^{-5}$	$4.62 \cdot 10^{-6}$	$5.67 \cdot 10^{-7}$
	$9.05 \cdot 10^{-2}$	$5.21 \cdot 10^{-2}$	$4.87 \cdot 10^{-2}$	$3.33 \cdot 10^{-2}$	$2.78 \cdot 10^{-3}$	$6.56 \cdot 10^{-4}$
	$9.85 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.83 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.64 \cdot 10^{-1}$	$1.05 \cdot 10^0$
	$1.17 \cdot 10^0$	$1.16 \cdot 10^0$	$1.20 \cdot 10^0$	$1.14 \cdot 10^0$	$1.06 \cdot 10^0$	$1.06 \cdot 10^0$
	$1.19 \cdot 10^0$	$1.36 \cdot 10^0$	$1.25 \cdot 10^0$	$1.11 \cdot 10^0$	$1.12 \cdot 10^0$	$1.11 \cdot 10^0$

В заключение отметим, что предлагаемый метод практически и теоретически осуществим для любой аналитически заданной граничной кривой класса гладкости C^5 . Это накладывает ограничение для его практического использования, поскольку часто бывает известно положение лишь конечного числа граничных точек, а аппроксимация границы в рамках описанного метода не предусмотрена. Тем не менее можно предположить, что аппроксимация границы с помощью аналитически заданной гладкой кривой является самостоятельной проблемой. Тогда при условии ее удовлетворительного решения метод может быть всегда реализован, независимо от сложности аппроксимирующих функций. Например, граничные элементы можно аппроксимировать с помощью сплайнов высоких порядков, обеспечивающих непрерывность производных достаточно высоких порядков на стыках граничных элементов и высокоточную аппроксимацию границы, необходимую при вычислениях вблизи границы. При этом другие полуаналитические методы, основанные на точном интегрировании по длине дуги, не могут быть использованы, так как для их реализации, т.е. возможности аналитического ин-

Таблица 2. Максимумы модулей абсолютных погрешностей аппроксимаций НП ППС

$-t_m$	$h_1 = \pi/3$	$h_1 = \pi/7$	$h_1 = \pi/15$	$h_1 = \pi/31$	$h_1 = \pi/63$	$h_1 = \pi/127$
10^{-2}	$3.01 \cdot 10^{-2}$	$2.66 \cdot 10^{-3}$	$2.77 \cdot 10^{-4}$	$3.15 \cdot 10^{-5}$	$3.76 \cdot 10^{-6}$	$4.59 \cdot 10^{-7}$
	$3.04 \cdot 10^{-2}$	$4.58 \cdot 10^{-3}$	$3.22 \cdot 10^{-3}$	$7.39 \cdot 10^{-4}$	$1.78 \cdot 10^{-4}$	$4.36 \cdot 10^{-5}$
	$3.00 \cdot 10^{-2}$	$1.06 \cdot 10^{-3}$	$1.55 \cdot 10^{-2}$	$1.56 \cdot 10^{-2}$	$1.56 \cdot 10^{-2}$	$1.56 \cdot 10^{-2}$
	$3.01 \cdot 10^{-2}$	$2.66 \cdot 10^{-3}$	$1.06 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$
	$3.01 \cdot 10^{-2}$	$2.66 \cdot 10^{-3}$	$2.76 \cdot 10^{-4}$	$3.15 \cdot 10^{-5}$	$3.76 \cdot 10^{-6}$	$1.90 \cdot 10^{-6}$
10^{-3}	$3.11 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.85 \cdot 10^{-4}$	$3.24 \cdot 10^{-5}$	$3.86 \cdot 10^{-6}$	$5.73 \cdot 10^{-7}$
	$3.20 \cdot 10^{-2}$	$7.92 \cdot 10^{-3}$	$6.41 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$	$3.57 \cdot 10^{-4}$	$8.75 \cdot 10^{-5}$
	$1.21 \cdot 10^{-1}$	$3.00 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$
	$3.21 \cdot 10^{-2}$	$7.81 \cdot 10^{-3}$	$1.68 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$
	$3.11 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.26 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$
10^{-4}	$3.12 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.85 \cdot 10^{-4}$	$3.25 \cdot 10^{-5}$	$3.87 \cdot 10^{-6}$	$4.73 \cdot 10^{-7}$
	$3.27 \cdot 10^{-2}$	$1.19 \cdot 10^{-2}$	$9.65 \cdot 10^{-3}$	$2.23 \cdot 10^{-3}$	$5.39 \cdot 10^{-4}$	$1.32 \cdot 10^{-4}$
	$4.51 \cdot 10^{-1}$	$4.79 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$
	$3.19 \cdot 10^{-1}$	$4.20 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$
	$4.36 \cdot 10^{-2}$	$2.15 \cdot 10^{-1}$	$3.54 \cdot 10^{-1}$	$3.54 \cdot 10^{-1}$	$3.54 \cdot 10^{-1}$	$3.54 \cdot 10^{-2}$
10^{-5}	$3.12 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.86 \cdot 10^{-4}$	$3.25 \cdot 10^{-5}$	$3.87 \cdot 10^{-6}$	$4.73 \cdot 10^{-7}$
	$3.33 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	$3.09 \cdot 10^{-3}$	$7.45 \cdot 10^{-4}$	$1.83 \cdot 10^{-4}$
	$5.95 \cdot 10^{-1}$	$4.98 \cdot 10^{-1}$	$4.99 \cdot 10^{-1}$	$4.99 \cdot 10^{-1}$	$4.99 \cdot 10^{-1}$	$4.99 \cdot 10^{-1}$
	$4.81 \cdot 10^{-1}$	$4.92 \cdot 10^{-1}$	$4.96 \cdot 10^{-1}$	$4.96 \cdot 10^{-1}$	$4.96 \cdot 10^{-1}$	$4.96 \cdot 10^{-1}$
	$4.27 \cdot 10^{-1}$	$4.69 \cdot 10^{-1}$	$4.85 \cdot 10^{-1}$	$4.85 \cdot 10^{-1}$	$4.85 \cdot 10^{-1}$	$4.85 \cdot 10^{-1}$
10^{-14}	$3.12 \cdot 10^{-2}$	$2.75 \cdot 10^{-3}$	$2.86 \cdot 10^{-4}$	$3.25 \cdot 10^{-5}$	$3.87 \cdot 10^{-6}$	$4.73 \cdot 10^{-7}$
	$5.49 \cdot 10^{-2}$	$5.18 \cdot 10^{-2}$	$4.92 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$	$2.78 \cdot 10^{-3}$	$6.65 \cdot 10^{-4}$
	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$
	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$
	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$

тегрирования, требуется более простая аппроксимация координатных функций: линейная или квадратичная.

СПИСОК ЛИТЕРАТУРЫ

1. Бреббия К., Теллес Ж., Вроубел Л. Методы граничных элементов. М.: Мир, 1987.
2. Березин И. С., Жидков Н. П. Методы вычислений. Т. 1. М.: Физматгиз, 1962.
3. Gao X. W., Yang K., Wang J. An adaptive element subdivision technique for evaluation of various 2D singular boundary integrals // Eng. Anal. Bound. Elem. 2008. V. 32. Iss. 8. P. 692–696.
4. Крутицкий П. А., Федотова А. Д., Колыбасова В. В. Квадратурная формула для потенциала простого слоя // Дифференц. ур-ния. 2019. Т. 55. № 9. С. 1269–1284.

5. Крутицкий П. А., Резниченко И. О. Улучшенная квадратурная формула для потенциала двойного слоя // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 2. С. 230–244.
6. Zhang Y.-M., Gu Y., Chen J.-T. Stress analysis for multilayered coating systems using semi-analytical BEM with geometric non-linearities // Comput. Mech. 2011. V. 47. Iss. 5. P. 493–504.
7. Gu Y., Chen W., Zhang B., Qu W. Two general algorithms for nearly singular integrals in two dimensional anisotropic boundary element method // Comput. Mech. 2014. V. 53. Iss. 6. P. 1223–1234.
8. Niu Z., Cheng C., Zhou H., Hu Z. Analytic formulations for calculating nearly singular integrals in two-dimensional BEM // Eng. Anal. Bound. Elem. 2007. V. 31. Iss. 12. P. 949–964.
9. Niu Z., Hu Z., Cheng C., Zhou H. A novel semi-analytical algorithm of nearly singular integrals on higher order elements in two dimensional BEM // Eng. Anal. Bound. Elem. 2015. V. 61. P. 42–51.
10. Cheng C., Pan D., Han Z., Wu M., Niu Z. A state space boundary element method with analytical formulas for nearly singular integrals // Acta Mech. Solida Sin. 2018. V. 31. № 4. P. 433–444.
11. Крутицкий П. А., Колыбасова В. В. Численный метод решения интегральных уравнений в задаче с наклонной производной для уравнения Лапласа вне разомкнутых кривых // Дифференц. уравнения. 2016. Т. 52. № 9. С. 1262–1276.
12. Крутицкий П. А., Резниченко И. О. Квадратурная формула для гармонического потенциала двойного слоя // Дифференц. уравнения. 2021. Т. 57. № 7. С. 932–950.
13. Крутицкий П. А., Резниченко И. О. Квадратурная формула для потенциала двойного слоя в случае уравнения Гельмгольца // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 3. С. 421–436.
14. Gong Y. P., Dong C. Y., Bai Y. Evaluation of nearly singular integrals in isogeometric boundary element method // Eng. Anal. Bound. Elem. 2017. V. 75. P. 21–35.
15. Gao X.-W., Zhang J.-B., Zheng B.-J., Zhang C. Element-subdivision method for evaluation of singular integrals over narrow strip boundary elements of super thin and slender structures // Eng. Anal. Bound. Elem. 2016. V. 66. P. 145–154.
16. Zhang J., Wang P., Lu C., Dong Y. A spherical element subdivision method for the numerical evaluation of nearly singular integrals in 3D BEM // Eng. Comput. 2017. V. 34. Iss. 6. P. 2074–2087.
17. Gong Y., Dong C., Qin F., Hattori G., Trevelyan J. Hybrid nearly singular integration for isogeometric boundary element analysis of coatings and other thin 2D structures // Comput. Methods Appl. Mech. Eng. 2020. V. 367. 113099.
18. Иванов Д. Ю. Уточнение коллокационного метода граничных элементов вблизи границы области в случае двумерных задач нестационарной теплопроводности с граничными условиями второго и третьего рода // Вестн. Томск. гос. ун-та. Матем. и мех. 2019. № 57. С. 5–25.
19. Иванов Д. Ю. Об аппроксимации нормальной производной теплового потенциала простого слоя вблизи границы двумерной области // Вестн. Томск. гос. ун-та. Матем. и мех. 2023. № 83. С. 31–51.
20. Иванов Д. Ю. Уточнение коллокационного метода граничных элементов вблизи границы двумерной области с помощью полуаналитической аппроксимации теплового потенциала двойного слоя // Вестн. Томск. гос. ун-та. Матем. и мех. 2020. № 65. С. 30–52.
21. Иванов Д. Ю. О равномерной сходимости аппроксимаций потенциала двойного слоя вблизи границы двумерной области // Вестн. Удмурт. ун-та. Матем. Мех. Комп. науки. 2022. Т. 32. Вып. 1. С. 26–43.
22. Смирнов В. И. Курс высшей математики. Т. 4. Ч. 2. М.: Наука, 1981.

23. *Иванов Д. Ю.* Устойчивая разрешимость в пространствах дифференцируемых функций некоторых двумерных интегральных уравнений теплопроводности с операторно-полугрупповым ядром // Вестн. Томск. гос. ун-та. Матем. и мех. 2015. № 38. С. 33–45.
24. *Фихтенгольц Г. М.* Курс дифференциального и интегрального исчисления. Т. 2. М.: Физматлит, 2003.
25. *Данфорд Н., Шварц Дж. Т.* Линейные операторы. Общая теория. М.: Изд-во иностр. лит., 1962.
26. *Крылов В. И., Шульгина Л. Т.* Справочная книга по численному интегрированию. М.: Наука, 1966.

ON THE UNIFORM CONVERGENCE OF APPROXIMATIONS OF THE TANGENT AND NORMAL DERIVATIVES OF THE POTENTIAL OF A SIMPLE LAYER NEAR THE BOUNDARY OF A TWO-DIMENSIONAL DOMAIN

D. Yu. Ivanov^{a,*}

^a 127994 Moscow, Obraztsova str., 9, Russian University of Transport, (RUT (MIIT)), Academy of Basic Training, Russia

*e-mail: ivanovdyu@yandex.ru

Received 6 September, 2023

Revised 6 September, 2023

Accepted 3 April, 2024

Abstract. Semi-analytical approximations of the tangent derivative (TD) and the normal derivative (ND) of the potential of a simple layer (PSL) near the boundary of a two-dimensional region are proposed, performed within the framework of the collocation method of boundary elements and not requiring approximation of the coordinate functions of the boundary. To obtain approximations, analytical integration over the smooth component of the distance function and a special additive-multiplicative method of distinguishing features are used. It is proved that such approximations have a more uniform convergence near the boundary of the domain compared with similar approximations of TD and PSL ND based on a simple multiplicative method of distinguishing features. One of the reasons for the highly uneven convergence of traditional approximations of the TD and PSL ND based on Gauss's quadrature formulas has been established.

Keywords: boundary element method, almost singular integral, analytical integration.

ОБ ОДНОЗНАЧНОСТИ ОПРЕДЕЛЕНИЯ СЕТОЧНОГО ФУНДАМЕНТАЛЬНОГО РЕШЕНИЯ УРАВНЕНИЯ ЛАПЛАСА В ТЕОРИИ ДИСКРЕТНОГО ПОТЕНЦИАЛА¹⁾

© 2024 г. И. Э. Степанова^{1,*}, И. И. Колотов², А. Г. Ягола², А. Н. Левашов²

¹ 123242 Москва, ул. Б. Грузинская, 10, стр. 1, Институт физики Земли им. О. Ю. Шмидта, Россия

² 119992 Москва, Ленинские горы, МГУ, Россия

*e-mail: tet@ifz.ru

Поступила в редакцию 13.03.2024 г.

Переработанный вариант 13.03.2024 г.

Принята к публикации 14.03.2024 г.

В работе рассматривается проблема однозначного определения фундаментального решения сеточного аналога уравнения Лапласа в рамках теории дискретного гравитационного потенциала. Сеточное фундаментальное решение конечно-разностного аналога уравнения Лапласа играет ключевую роль при восстановлении непрерывно распределенного источника гравитационного или магнитного поля по разнородным и разноточным данным, полученным в точках некоторого сеточного множества. Библ. 15.

Ключевые слова: однозначное определение, фундаментальное решение, дискретный гравитационный потенциал.

DOI: 10.31857/S0044466924070108, EDN: xifmwy

ВВЕДЕНИЕ

В настоящее время весьма остро стоит проблема интерпретации разнородной и разноточной геофизической и геодезической информации, получаемой благодаря спутниковому зондированию. Можно констатировать, что существующие методы обработки больших и сверхбольших массивов постоянно меняющихся данных нуждаются в кардинальном совершенствовании. Как отмечалось в работах [1–4], аппроксимационный подход к решению обратных линейных и нелинейных задач геофизики, геодезии и геоморфологии позволяет редуцировать практически все постановки по определению параметров геологической среды к системам линейных (в некоторых случаях и нелинейных) алгебраических уравнений.

При цифровой обработке сигналов критически важным становится учет дискретности получаемых данных. Даже спутниковые системы высокого разрешения не могут предоставить исследователям всю информацию о непрерывно меняющихся физических полях и изучаемых природных объектах, равно как и об объектах антропогенного происхождения. Поэтому основные принципы теории дискретного гравитационного и магнитного потенциалов, предложенные В. Н. Страховым в начале 1990-х годов и развитые далее в работах [1, 2], могут, на наш взгляд, применяться при решении сложных и масштабных задач по распознаванию изображений, восстановлению источников гравитационного и магнитного полей и т.п. Если одновременно с теорией дискретного потенциала будет использоваться метод линейных интегральных представлений [3–6], то такой подход позволит достичь более высокой степени соответствия математической модели геофизическим реалиям.

1. ОСНОВЫ ТЕОРИИ ДИСКРЕТНОГО ПОТЕНЦИАЛА

В рамках теории дискретного гравитационного потенциала [1] вместо вектора с координатами в декартовом пространстве R^n : $x = (x_1, x_2, \dots, x_n) \in R^n$, $-\infty < x_k < +\infty$, $k = 1, \dots, n$, рассматривается его сеточный аналог. Общее число сеток равно 2^n . Из этих сеток одна считается основной,

¹⁾ Работа выполнена в рамках госзадания ИФЗ РАН.

и в узлах этой сетки определены значения сеточного гравитационного потенциала $V_S(x^{(S)})$, где $x^{(S)}$ имеет следующий вид:

$$x^{(S)} = (x_1^{(S)}, x_2^{(S)}, \dots, x_n^{(S)})^T, \quad x_k^{(S)} = ph_k; \quad -\infty < p < +\infty, \quad p = \pm 0, \pm 1, \pm 2, \dots;$$

$$h_k = \text{const}, \quad h = (h_1, h_2, \dots, h_n)^T.$$

“Неосновные” сетки служат для вычисления значений высших производных потенциала. Если e_k — единичный орт вдоль k -й оси, то первая производная гравитационного потенциала запишется в виде

$$\frac{\partial V_S(x^{(S)})}{\partial x_k^{(S)}} = \frac{V_S(x^{(S)} + h_k e_k) - V_S(x^{(S)})}{h_k}. \quad (1)$$

Значения производной по k -й оси относятся к узлам $x^{(S,k)}$, $k = 1, 2, \dots, n$, вспомогательной сетки:

$$x^{(S,k)} = x^{(S)} + \frac{h_k}{2} e_k, \quad k = 1, 2, \dots, n.$$

Вторые производные сеточного потенциала находятся из выражения

$$\frac{\partial^2 V_S(x^{(S)})}{\partial x_p^{(S)} \partial x_q^{(S)}} = \frac{1}{h_q} \left[\frac{\partial V_S(x^{(S)} + h_p e_p)}{\partial x_p^{(S)}} - \frac{\partial V_S(x^{(S)})}{\partial x_p^{(S)}} \right]. \quad (2)$$

Значения $\frac{\partial^2 V_S(x^{(S)})}{\partial x_p^{(S)} \partial x_q^{(S)}}$ при $p \neq q$ относятся к узлам третьей вспомогательной сетки: $x^{(S,3)} = x^{(S)} + \frac{h_p e_p}{2} + \frac{h_q e_q}{2}$; если же $p = q$, то разностные аналоги вторых производных вычисляются в узлах основной сети.

Аналоги операторов Лапласа и Пуассона в сеточном пространстве могут быть записаны в виде

$$\Delta \{V_S(x^{(S)})\} = 0;$$

$$\Delta \{V_S(x^{(S)})\} = -\frac{GC_n m_S(x^{(S)})}{H^{n-2}}, \quad H = \left(\prod_{k=1}^n h_k \right)^{\frac{1}{n}}, \quad m_S(x^{(S)}) = H^n \rho(x^{(S)}). \quad (3)$$

В (3) Δ — выбранный конечно-разностный аналог оператора Лапласа, а $m_S(x^{(S)}) = H^n \rho(x^{(S)})$ — значения сеточных масс, порождающих сеточное гравитационное поле.

В теории дискретного потенциала важнейшей методологической установкой считается формулировка вариационной задачи, в которой наряду с условиями (3) рассматривается дополнительное требование:

$$\left| \Delta_2 \{V_S(x^{(S)})\} \right|_{E(D_S)}^2 + \alpha \Phi(V_S(x^{(S)}), g_S(x^{(S)})) = \min_{V_S(x^{(S)}), x^{(S)} \in \text{Int } D_S}, \quad (4)$$

где $\alpha \geq 0$ — некоторый параметр (можно сказать, параметр регуляризации [7]); D_S — заданная сеточная область; $x^{(S)} \in D_S$; $E(D_S)$ — евклидова норма в сеточной области D_S ; $\Phi(t, z)$ — неотрицательный функционал на векторах одинаковой размерности t и z , Δ_2 — это аппроксимация оператора Лапласа на другом шаблоне, отличная от $\Delta = \Delta_1$. Вариационную задачу (3), (4) можно решать одним из методов, описанных, например, в [8]. Но в наших предыдущих работах (см. [3–6]) мы показали, что при решении больших плохо обусловленных систем линейных алгебраических уравнений (СЛАУ) целесообразно применять специальные методы, которые позволяют эффективно устранять влияние помехи в правой части и матрице.

Система уравнений [3] не является, очевидно, замкнутой: она не позволяет определять сеточные потенциалы однозначно во всем пространстве (даже в некоторой ограниченной трехмерной сеточной области ее решение будет неединственным). Вопросы однозначной разрешимости СЛАУ имеют важное значение для адекватной реальности интерпретации геофизических данных [9–11]. Чем больше произвол в выборе эквивалентного по внешнему полю распределения масс, тем больше неопределенность в истинности того или иного утверждения о геологическом строении среды. В отсутствие

априорной информации об объекте, порождающем поле, вообще ничего нельзя сказать о локализации и геометрии источника сигнала.

Поэтому, прежде чем переходить к формулировке краевых задач для выделения единственного сеточного аналога гравитационного или магнитного потенциала, необходимо рассмотреть принципиально важный для дальнейшего вопрос нахождения фундаментального решения разностного аналога оператора Лапласа.

Дискретными аналогами потенциала и фундаментального решения уравнения Лапласа служат (ср. с формулами (3)):

$$V_S \left(x^{(S)} \right) = G \sum_{\xi^{(S)} \in J_S} m_S \left(\xi^{(S)} \right) \Omega_n^{(S)} \left(\xi^{(S)} - x^{(S)} \right), \quad (5)$$

$$\Delta \left(\Omega_n^{(S)} \left(x^{(S)} \right) \right) = - \frac{C_n}{H^{n-2}} e \left(x^{(S)} \right), \quad (6)$$

$$e \left(x^{(S)} \right) = \begin{cases} 1, & x^{(S)} = 0, \\ 0, & |x^{(S)}| > 0. \end{cases} \quad (7)$$

В (5) есть совокупность векторов координат $\xi^{(S)}$ тех узлов сетки, в которых имеются ненулевые сеточные массы; иначе говоря, J_S есть сеточный носитель источников поля. Функция $\Omega_n^{(S)} \left(x^{(S)} \right)$ есть сеточный аналог фундаментального решения $\Omega_n(x)$ уравнения Лапласа в континуальной теории, Δ — выбранный, как и в (3), конечно-разностный аналог оператора Лапласа. Ясно, что $\Omega_n^{(S)} \left(x^{(S)} \right)$ играет роль сеточного фундаментального решения. Условиями (5)–(7) сеточное фундаментальное решение уравнения Лапласа ни в одной из размерностей однозначно не определяется. В [1] В. Н. Страхов выдвинул гипотезу о том, что имеет место единственность решения задачи по определению сеточного фундаментального решения в R^3 , близкого по евклидовой норме к потенциалу призмы, заданной в области $|\bar{r}| \leq \frac{h}{2}$, $\bar{r} = (x, y, z)^T$, $|\bar{r}| \equiv \sqrt{x^2 + y^2 + z^2}$.

Заметим, что уравнениям (6), (7) удовлетворяет, например, следующая сеточная функция трех переменных:

$$\Omega_3^{(0)} \left(x^{(S)} \right) = \frac{1}{6} \cdot \left\{ \left(x_1^{(S)} \right)^2 + \left(x_2^{(S)} \right)^2 + \left(x_3^{(S)} \right)^2 \right\} + Ax_1^{(S)} + Bx_2^{(S)} + Cx_3^{(S)} + D,$$

где A, B, C, D — произвольные вещественные постоянные. Вместо линейной функции сеточных переменных можно взять любой гармонический полином, который удовлетворяет сеточному аналогу уравнения Лапласа (к примеру, $2 \left(x_1^{(S)} \right)^2 + 2 \left(x_2^{(S)} \right)^2 - 4 \left(x_3^{(S)} \right)^2$). Таким образом, об однозначной разрешимости задачи по определению сеточного фундаментального уравнения Лапласа речь может идти только при наложении дополнительных ограничений на функцию $\Omega_3^{(S)} \left(x^{(S)} \right)$. Далее, сеточное фундаментальное решение мы будем искать в некоторой ограниченной области сеточного трехмерного пространства. В качестве такой области может выступать как прямоугольный параллелепипед (в случае декартовой системы координат), так и шар (если рассматривать сеточные аналоги уравнений Лапласа и Пуассона в сферической системе координат).

2. СЕТОЧНОЕ ФУНДАМЕНТАЛЬНОЕ РЕШЕНИЕ УРАВНЕНИЯ ЛАПЛАСА В ДЕКАРТОВОЙ СИСТЕМЕ КООРДИНАТ

Оператор Лапласа в (6) будем считать заданным на шаблоне “крест” [12]. Тогда для трехмерного сеточного пространства имеем

$$\Lambda^+ \left(\Omega_3^{(S)} \left(x^{(S)} \right) \right) = - \frac{4\pi}{H} e_S \left(x^{(S)} \right);$$

$$\Lambda^+ \left(\Omega_3^{(S)}(x^{(S)}) \right) = -3 \left(\sum_{p=1}^3 C_p \right) \Omega_3^{(S)}(x^{(S)}) + \sum_{p=1}^3 \left\{ \Omega_3^{(S)}(x^{(S)} - h_p e_p) + \Omega_3^{(S)}(x^{(S)} + h_p e_p) \right\}, \quad (8)$$

$$C_p = \frac{H^3}{h_1 h_2 h_3}, \quad H = \sqrt[3]{h_1 h_2 h_3}, \quad x^{(S)} = (x_1^{(S)}, x_2^{(S)}, x_3^{(S)}), \quad x_p^{(S)} = k h_p,$$

$$-K \leq k \leq +K; \quad h_p = \text{const}, \quad p = 1, 2, 3; \quad |x_p^{(S)}| \leq d_p = K h_p, \quad p = 1, 2, 3.$$

Как утверждается в [1], введение условия $\lim_{|x^{(S)}| \rightarrow \infty} |\Omega_n^{(S)}(x^{(S)})| = 0$ при $n \geq 3$ не снимает неединственности. В самом деле, соображения симметрии показывают, что имеют место равенства

$$\Omega_2^{(S)}(x^{(S)}) = \Omega_3^{(S)}(x_1^{(S)}, x_2^{(S)}, x_3^{(S)}) = \Omega_3^{(S)}(|x_1^{(S)}|, |x_2^{(S)}|, |x_3^{(S)}|) = \Omega_3^{(S)}(|x_{i_1}^{(S)}|, |x_{i_2}^{(S)}|, |x_{i_3}^{(S)}|),$$

$$(i_1, i_2, i_3) \in S_3,$$

где S_3 — группа перестановок трех индексов. Следовательно, достаточно найти значения $\Omega_3^{(S)}(x^{(S)})$ в области $S_{+++}^3 = \{x^{(S)} : x_1^{(S)} \geq 0, x_2^{(S)} \geq 0, x_3^{(S)} \geq 0, x_1^{(S)} \leq x_2^{(S)} \leq x_3^{(S)}\}$, т.е. в части первого координатного октанта.

При определении значений сеточного фундаментального решения в узлах области S_{+++}^3 можно поставить условие минимальности отклонения сеточного решения от его непрерывного аналога:

$$F \left[\Omega_3^{(S)}(x^{(S)}) \right] = \sum_{x^{(S)} \in S_{+++}^3} \left(\frac{\Omega_3^{(S)}(x^{(S)}) - \Omega_{3,c}(x^{(S)})}{\Omega_{3,c}(x^{(S)})} \right) = \min, \quad (9)$$

где $\Omega_{3,c}(x^{(S)})$ — значения непрерывного фундаментального решения уравнения Лапласа в точках сеточной области.

Таким образом, одним из вариантов “устранения” неопределенности при нахождении сеточного фундаментального решения уравнения Лапласа является минимизация функционала (9) при условиях (8); такого рода вариационные постановки исследовались нами неоднократно при применении метода линейных интегральных представлений к решению обратных линейных задач геофизики [3–6].

В настоящей работе мы попытаемся заменить трехмерное сеточное пространство (т.е. неограниченное множество точек $x^{(S)} = (x_1^{(S)}, x_2^{(S)}, \dots, x_n^{(S)})^T$, $x_k^{(S)} = p h_k$; $-\infty < p < +\infty$, $p = \pm 0, \pm 1, \pm 2, \dots$; $h_k = \text{const}$, ансамблем, или семейством, расширяющихся компактов, представляющих собой прямоугольные параллелепипеды $|x_p^{(S)}| \leq d_p = K_n h_p$, $p = 1, 2, 3$; здесь d_p , K_n и h_p , $p = 1, 2, 3$, — некоторые положительные константы, причем $K_n \rightarrow \infty$ при $n \rightarrow \infty$.

Рассмотрим уравнения Лапласа и Пуассона в некотором прямоугольном параллелепипеде. Для того чтобы корректно поставить задачу по определению $\Omega(x^{(S)})$ в указанном параллелепипеде нужно задать граничные условия. Будем считать, что на гранях параллелепипеда K_{n_q} с номером n_q , принадлежащего описанному выше семейству расширяющихся компактов, значения фундаментального решения равны нулю (непрерывным аналогом дискретной постановки такого рода является задача Дирихле для уравнения Пуассона):

$$\Omega(x^{(S)}) = 0. \quad (10)$$

Решение задачи (8)–(10) приводит к необходимости применения метода прогонки. В указанном методе матрицы систем линейных алгебраических уравнений имеют трехдиагональный или блочно-

трехдиагональный вид (если постановка рассматривается в трехмерном сеточном пространстве):

$$\begin{aligned}
 A_{NN} &= \begin{bmatrix} A_{11} & B_{12} & \dots & 0 \\ B_{21} & \dots & \dots & \\ 0 & \dots & A_{NN} & \end{bmatrix}, \quad A_{kk} = \begin{bmatrix} a & b \dots & 0 \\ b & \dots & 0 \\ 0 & \dots & a \end{bmatrix}, \quad \mathbf{Y}_k = \begin{pmatrix} \mathbf{y}_1^{(k)} \\ \mathbf{y}_2^{(k)} \\ \dots \\ \mathbf{y}_{N_2}^{(k)} \end{pmatrix}, \\
 a &= 2(1 + \alpha_1 + \alpha_2), \quad \alpha_1 = \frac{h_3^2}{h_1^2}, \quad \alpha_2 = \frac{h_3^2}{h_2^2}, \quad b = -\alpha_1; \\
 c &= -\alpha_2; \quad \mathbf{y}_j^{(k)} = \left(y_{1,j}^{(k)}, y_{2,j}^{(k)}, \dots, y_{N_1,j}^{(k)} \right)^T, \quad 1 \leq k \leq N_3 - 1; \\
 B_{12} &= B_{21}, \quad B_{ij} = \begin{bmatrix} c & 0 \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & c \end{bmatrix}, \quad j = i \pm 1; \quad i = 1, \dots, N_1 - 1, \quad j = 1, \dots, N_2 - 1.
 \end{aligned} \tag{11}$$

Необходимо сделать важное замечание: соотношение (8) — это конечно-разностный аналог уравнения Пуассона, заданный в неограниченной сеточной области, фактически во всем сеточном пространстве. Для однозначной разрешимости такого рода уравнений необходимо задать граничные условия, причем в тех точках, где это возможно сделать. Поэтому нужно определиться со свойствами сеточных областей, в которых мы в дальнейшем будем рассматривать краевые задачи для конечно-разностных аналогов уравнений Лапласа и Пуассона. Введем следующие обозначения:

$$\begin{aligned}
 \omega_3 &= (ih_1, jh_2, kh_3), \quad 1 \leq i \leq N_1 - 1, \quad 1 \leq j \leq N_2 - 1, \quad 1 \leq k \leq N_3 - 1; \\
 \omega &= (ih_1, jh_2), \quad 1 \leq i \leq N_1 - 1, \quad 1 \leq j \leq N_2 - 1; \\
 \gamma_3 &= (0, jh_2, kh_3) \cup (N_1, jh_2, kh_3) \cup (jh_1, 0, kh_3) \cup (jh_1, N_2, kh_3) \cup (jh_1, jh_2, 0) \cup (ih_1, jh_2, N_3), \\
 &1 \leq i \leq N_1 - 1, \quad 1 \leq j \leq N_2 - 1, \quad 1 \leq k \leq N_3 - 1; \\
 \gamma &= (0, jh_2) \cup (N_1, jh_2) \cup (jh_1, 0) \cup (jh_1, N_2), \quad 1 \leq i \leq N_1 - 1, \quad 1 \leq j \leq N_2 - 1.
 \end{aligned} \tag{12}$$

Таким образом, мы задали равномерные сетки в трехмерном и двумерном пространствах (ω_3, ω) , а также обозначили границы соответствующих областей.

Дифференциально-разностному оператору в (6) соответствует при решении краевых задач схема матричной прогонки [12] следующего вида:

$$\begin{aligned}
 -\mathbf{Y}_{k-1} + \mathbf{C}_k \mathbf{Y}_k + \mathbf{Y}_{k+1} &= \mathbf{F}_k, \quad 1 \leq k \leq N_3 - 1, \\
 \mathbf{C} \mathbf{Y} &= (\Lambda \mathbf{y}_1, \Lambda \mathbf{y}_2, \dots, \Lambda \mathbf{y}_{N_2-1}), \quad \mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_2-1}), \\
 \mathbf{y}_j &= (y_{1,j}, y_{2,j}, \dots, y_{N_1-1,j}), \quad 1 \leq j \leq N_2 - 1, \\
 \Lambda \mathbf{y}_j &= (\Lambda_1 y_{1,j}, \Lambda_1 y_{2,j}, \dots, \Lambda_1 y_{N_1-1,j}), \\
 \Lambda_1 y_{i,j} &= 2y_{i,j} - C_3^2 y_{\bar{x}_1 x_1; i, j} - C_3^2 y_{\bar{x}_2 x_2; i, j}, \quad 1 \leq i \leq N_1 - 1; \quad 1 \leq j \leq N_2 - 1.
 \end{aligned} \tag{13}$$

В (13) через $\mathbf{Y}_k$ обозначен блочный вектор значений сеточной функции в слое, соответствующем значению третьей координаты $x_3^{(S)} = kh_3$, $1 \leq k \leq N_3 - 1$. Правые части $\mathbf{F}_k$ при изменении индекса $1 \leq k \leq N_3 - 1$ — это значения сеточной функции $\frac{4\pi}{H} e_S(x^{(S)})$ в соответствующих точках, константа C_3^2 определяется в (8) (точнее говоря, это — квадрат C_3). К уравнениям (13) необходимо добавить граничные условия с тем, чтобы возникла постановка краевой задачи. Мы рассмотрим граничные условия I рода:

$$\mathbf{Y}_0 = \mathbf{F}_0, \quad \mathbf{Y}_{N_3} = \mathbf{F}_{N_3}. \tag{14}$$

Постановка (13), (14) — это частный случай более общей задачи: найти векторы $\mathbf{Y}_k$, $0 \leq k \leq N_3$, удовлетворяющие следующей системе:

$$\begin{aligned}
 -\mathbf{A}_k \mathbf{Y}_{k-1} + \mathbf{C}_k \mathbf{Y}_k + \mathbf{B}_k \mathbf{Y}_{k+1} &= \mathbf{F}_k, \quad 1 \leq k \leq N_3 - 1, \\
 C_0 \mathbf{Y}_0 - \mathbf{B}_0 \mathbf{Y}_1 &= \mathbf{F}_0, \quad k = 0, \\
 -\mathbf{A}_{N_3} \mathbf{Y}_{N_3-1} + \mathbf{C}_{N_3} \mathbf{Y}_{N_3} &= \mathbf{F}_{N_3}, \quad k = N_3,
 \end{aligned} \tag{15}$$

где $\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k$ — блочные матрицы; $\mathbf{C}_k$ — блочная квадратная матрица $N_2 \times N_2$ каждый блок которых имеет размер $N_1 \times N_1$; $\mathbf{A}_k$ — блочная прямоугольная матрица размера $N_2 - 1 \times N_2$, блоки которой имеют размер $N_1 \times N_1$; $\mathbf{B}_k$ — блочная прямоугольная матрица $N_2 + 1 \times N_2$ с блоками размера $N_1 \times N_1$; $\mathbf{Y}_k$ — блочный вектор размера N_2 с блоками длины N_1 .

Формулы матричной прогонки имеют вид:

$$\begin{aligned} \alpha_{k+1} &= (\mathbf{C}_k - \mathbf{A}_k \alpha_{k-1})^{-1} \mathbf{B}_k, \quad k = 1, 2, \dots, N_3 - 1, \quad \alpha_1 = \mathbf{C}_0^{-1} \mathbf{B}_0; \\ \beta_{k+1} &= (\mathbf{C}_k - \mathbf{A}_k \alpha_{k-1})^{-1} (\mathbf{F}_k + \mathbf{A}_k \beta_k), \quad k = 1, 2, \dots, N_3, \quad \beta_1 = \mathbf{C}_0^{-1} \mathbf{F}_0; \\ \mathbf{Y}_k &= \alpha_{k+1} \mathbf{Y}_{k+1} + \beta_{k+1}, \quad k = N_1 - 1, \quad N_3 - 2, \dots, 0, \quad \mathbf{Y}_{N_3} = \beta_{N_3+1}, \end{aligned} \quad (16)$$

причем алгоритм (16) корректен, если матрицы $(\mathbf{C}_k - \mathbf{A}_k \alpha_{k-1})^{-1}$ не вырождены для $k = 1, 2, \dots, N_3$.

В (12) показано, что при $\mathbf{C}_k = \mathbf{C}$, $A_{N_3} = B_0 = 0$, $C_0 = C_{N_3} = E$, $C_0 = C_{N_3} = E$, $B_k = A_k = E$, $1 \leq k \leq N_3$, а квадратная матрица $\mathbf{C}_k$ задана в (16), условия корректности алгоритма (16) принимают вид

$$\|\mathbf{C}^{-1}\| = \max_l |\lambda_l(\mathbf{C}^{-1})| = \frac{1}{\min_l |\lambda_l \mathbf{C}|} \leq 0.5. \quad (17)$$

В (17) максимум и минимум берется по собственным значениям матрицы $\mathbf{C}^{-1}$ и $\mathbf{C}$ соответственно. Из определения матрицы $\mathbf{C}$ можно получить, что при замене $\lambda_l = 2 + C_3^2 \lambda_{l_1}^{(1)} + C_3^2 \lambda_{l_2}^{(2)}$, $l = (l_1, l_2)$, задача по поиску собственных значений λ_1 для оператора

$$\begin{aligned} \Lambda_1 y_{i,j} &= 2y_{i,j} - C_3^2 y_{x_1 x_1; i,j} - C_3^2 y_{x_2 x_2; i,j} = \lambda_l y_{i,j}, \quad 1 \leq i \leq N_1 - 1; \quad 1 \leq N_2 \leq N_2 - 1, \\ y_{k,0} &= y_{k,N_2} = y_{0,j} = y_{N_1,j} = 0. \end{aligned} \quad (18)$$

сводится к следующей задаче на собственные значения (12):

$$\begin{aligned} \Lambda y(x) + \lambda_l y(x) &= 0, \quad x = (x_1, x_2) \in \omega, \quad \Lambda = \Lambda_x + \Lambda_y, \\ y(x) &= 0, \quad x = (x_1, x_2) \in \gamma; \\ y_{ij} &= y(x_{1,i}, x_{2,j}); \quad x_{1,i} = ih_1, \quad x_{2,j} = jh_2; \\ \Lambda_x \mu_{l_1}^{(1)} + \lambda_{l_1}^{(1)} \mu_{l_1}^{(1)} &= 0, \quad 1 \leq i \leq N_1 - 1, \\ \Lambda_y \mu_{l_2}^{(2)} + \lambda_{l_2}^{(2)} \mu_{l_2}^{(2)} &= 0, \quad 1 \leq j \leq N_2 - 1, \\ \mu_{l_1}^{(1)}(0) = \mu_{l_1}^{(1)}(N_1) = 0 &= \mu_{l_2}^{(2)}(0) = \mu_{l_2}^{(2)}(N_2), \\ \lambda_{l_\alpha}^{(\alpha)} &= \frac{4}{h_\alpha^2} \sin^2 \left(\frac{l_\alpha \pi}{2N_\alpha} \right) = \frac{4}{h_\alpha^2} \sin^2 \left(\frac{l_\alpha \pi h_\alpha}{2d_\alpha} \right), \quad l_\alpha = 1, 2, \dots, N_\alpha - 1, \\ \mu_{l_1}^{(1)}(i) &= \sqrt{\frac{2}{d_1}} \sin \left(\frac{l_1 \pi i}{N_1} \right), \quad l_1 = 1, \dots, N_1 - 1; \\ \mu_{l_2}^{(2)}(j) &= \sqrt{\frac{2}{d_2}} \sin \left(\frac{l_2 \pi j}{N_2} \right), \quad l_2 = 1, \dots, N_2 - 1; \\ y_{ij} &= \mu_l(i, j) = \mu_{l_1}^{(1)}(i) \cdot \mu_{l_2}^{(2)}(j), \\ \Lambda_x y_{ij} &= \frac{1}{h_1^2} [y_{i+1,j} - 2y_{ij} + y_{i-1,j}], \quad \Lambda_y y_{ij} = \frac{1}{h_2^2} [y_{i,j+1} - 2y_{ij} + y_{i,j-1}]. \end{aligned} \quad (19)$$

Само же собственное значение равно

$$\lambda_l = 2 + C_3^2 \cdot \sum_{\alpha=1}^2 \lambda_{l_\alpha}^{(\alpha)}. \quad (20)$$

В (17) через $y(x)$ обозначена сеточная функция двух аргументов, определенная на двумерной сетке $x = (x_1, x_2) \in \omega$ (см. (12)). При этом в каждом слое рассматривается «своя» задача на собственные значения конечно-разностного оператора вида (18). Из (20) становится ясно, что $\lambda_1 > 2$, поэтому

алгоритм матричной прогонки является для задачи (8)–(10) корректным, и мы можем утверждать, что справедлива следующая

Теорема 1. *Фундаментальное решение сеточного аналога уравнения Лапласа в трехмерном случае определяется условиями (8)–(10) однозначно.*

Подчеркнем, что корректность алгоритма матричной прогонки означает, что решение задачи (8)–(10) не только существует, но и единственно.

3. СЕТОЧНОЕ ФУНДАМЕНТАЛЬНОЕ РЕШЕНИЕ В СФЕРИЧЕСКОЙ СИСТЕМЕ КООРДИНАТ

Наряду с сеточными областями в трехмерном евклидовом пространстве мы будем рассматривать их сферические аналоги:

$$\begin{aligned} \mathbf{r}^{(S)} &= (r^{(S)}, \varphi^{(S)}, \vartheta^{(S)})^T, \quad r_k^{(S)} = r_{k-1}^{(S)} + h_{r,k}; \quad k = 1, \dots; \\ \varphi_m^{(S)} &= \varphi_{m-1}^{(S)} + mh_\varphi, \quad m = 1, \dots, N_2, \quad h_\varphi = \text{const}; \\ \vartheta_l^{(S)} &= \vartheta_{l-1}^{(S)} + h_{\vartheta,l}, \quad l = 1, \dots, N_3. \end{aligned} \quad (21)$$

В (21) по радиусу и углу ϑ введены сетки с произвольными переменными шагами $h_{r,k}$, $h_{\vartheta,l}$, и индексы k и l принимают значения из счетного и конечного множеств соответственно (ср. со случаем декартовой системы координат, когда значения координат сеточных масс $\mathbf{x}_{k,j,m}^{(S)} = x_k^{(S)}, y_l^{(S)}, z_m^{(S)}$ принадлежат прямому произведению $\square \times \square \times \square$ — трех экземпляров группы целых чисел $\square$). Если сеточная область ограничена по радиусу и двум углам, будем полагать, что $k = 0, \dots, N_1; l = 0, \dots, N_2; m = 0, \dots, N_3$. Можно считать, что заданы три сеточные функции целочисленного аргумента $h_1(k) = h_{r,k}$, $h_2(m) = h_\varphi$, $h_3(l) = h_{\vartheta,l}$ значения которых в целых точках вещественной прямой равны шагам по соответствующим координатам. Определим также средние шаги по каждой из сферических координат:

$$\begin{aligned} \bar{h}_i(m) &= \begin{cases} \frac{1}{2} h_i(1), & m = 0, \\ \frac{1}{2} \{h_i(m+1) + h_i(m)\}, & 1 \leq m \leq N_i - 1, \\ \frac{1}{2} h_i(N_i), & m = N_i, \end{cases} \quad i = 1, 2, 3; \\ h_1(k) &= h_{r,k}, \quad h_2(m) = h_\varphi, \quad h_3(l) = h_{\vartheta,l}. \end{aligned} \quad (22)$$

В сферической системе координат оператор Лапласа имеет вид

$$\Delta_{r,\varphi,\vartheta} f(x^{(i)}) \equiv \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \vartheta} \frac{\partial^2 f}{\partial \varphi^2} + \frac{1}{r^2} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial f}{\partial \vartheta} \right), \quad (23)$$

$$x^{(i)} = (r^{(i)}, \varphi^{(i)}, \vartheta^{(i)}).$$

Уравнению (23) соответствует в декартовых координатах следующий дифференциальный оператор эллиптического типа (в общем случае):

$$A_k f \equiv \text{div} (k(x) \nabla f), \quad x = (x_1, x_2, x_3) \in R^3, \quad k(x) \geq k_0 > 0. \quad (24)$$

Для описания сеточного фундаментального решения в криволинейной системе координат необходимо знать, как выглядят в ней дискретные аналоги операторов Лапласа и Пуассона. Для этого введем сеточные функции $\rho(i)$ и $\tau(l)$ [12]:

$$\begin{aligned} \rho(k) &= r_k, \quad 1 \leq k \leq N_1; \quad \rho = \begin{cases} \frac{1}{4} h_{r,1}, & r_0 = 0, \\ 0, & r_0 > 0, \end{cases} \\ \tau(l) &= \sin \vartheta_l^{(S)}, \quad 1 \leq l \leq N_2; \\ \tau(0) &= \begin{cases} \frac{1}{4} \sin h_{\vartheta,1}, & \vartheta_0 = 0, \\ \sin \vartheta_0, & \vartheta_0 > 0, \end{cases} \quad \tau(N_2) = \begin{cases} \frac{1}{4} \sin h_{\vartheta,N_2}, & \vartheta_{N_2} = \pi, \\ \sin \vartheta_{N_2}, & \vartheta_{N_2} < \pi. \end{cases} \end{aligned} \quad (25)$$

Для функций $\rho(k)$, $\tau(l)$ мы ниже будем использовать также обозначения $\rho_k = \rho(k)$ и $\tau(\vartheta_l^{(S)} = \tau(l))$.

Конечно-разностные аналоги операторов дифференцирования по сферическим координатам запишем в следующем виде:

$$\begin{aligned}\partial_{\bar{r}} f &\equiv \frac{f(r_k, \vartheta, \varphi) - f(r_{k-1}, \vartheta, \varphi)}{h_{r,k}}, \quad \partial_r f \equiv \frac{f(r_{k+1}, \vartheta, \varphi) - f(r_k, \vartheta, \varphi)}{h_{r,k+1}}, \\ \partial_{\bar{\vartheta}} f &\equiv \frac{f(r, \vartheta_l, \varphi) - f(r, \vartheta_{l-1}, \varphi)}{h_{\vartheta,l}}, \quad \partial_r f \equiv \frac{f(r, \vartheta_{l+1}, \varphi) - f(r, \vartheta_l, \varphi)}{h_{\vartheta,l+1}}, \\ \partial_{\bar{\varphi}} f &\equiv \frac{f(r, \vartheta, \varphi_m) - f(r, \vartheta, \varphi_{m-1})}{h_{\varphi}}, \quad \partial_{\varphi} f \equiv \frac{f(r, \vartheta, \varphi_{m+1}) - f(r, \vartheta, \varphi_m)}{h_{\varphi}}.\end{aligned}\quad (26)$$

Тогда для сеточного фундаментального решения в сферических координатах можно записать следующую систему конечно-разностных уравнений, аналогичную (8) для декартовых координат:

$$\begin{aligned}\Delta_{\mathbf{h}(r,\vartheta,\varphi)} \Omega_3^{(S)}(\mathbf{x}_{k,l,m}^{(S)}) &\equiv \\ &\equiv \frac{1}{\rho_k^2} \frac{\rho_{k+1}^2 \Omega_3^{(S)}(r_k^{(S)} + h_{r,k}, \varphi_m^{(S)}, \vartheta_l^{(S)}) - 2\rho_k^2 \Omega_3^{(S)}(r_k^{(S)} + h_{r,k}, \varphi_m^{(S)}, \vartheta_l^{(S)}) + \rho_{k-1}^2 \Omega_3^{(S)}(r_k^{(S)} - h_{r,k-1}, \varphi_m^{(S)}, \vartheta_l^{(S)})}{\bar{h}_r} + \\ &+ \frac{1}{\rho_k^2} \frac{\tau(\vartheta_{l+1}^{(S)}) \Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)} + h_{\vartheta,l}) - 2\tau(\vartheta_l^{(S)}) \Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)}) + \tau(\vartheta_{l-1}^{(S)}) \Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)} - h_{\vartheta,l-1})}{\bar{h}_{\vartheta}} + \\ &+ \frac{\Omega_3^{(S)}(r_k^{(S)}, \varphi_{m+1}^{(S)}, \vartheta_l^{(S)}) - 2\Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)}) + \Omega_3^{(S)}(r_k^{(S)}, \varphi_{m-1}^{(S)}, \vartheta_l^{(S)})}{\rho_k^2 [\tau(\vartheta_l^{(S)})]^2 h_{\varphi}^2} \equiv [\Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)})]_{r\bar{r}} + \\ &+ \frac{1}{\rho_k^2 \cdot \bar{h}_3(l)} [\Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)})]_{\vartheta, \bar{\vartheta}} + \frac{1}{\rho_k^2 [\tau(\vartheta_l^{(S)})]^2 h_{\varphi}^2} [\Omega_3^{(S)}(r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)})]_{\varphi \bar{\varphi}} = -\frac{4\pi}{H} e_S(\mathbf{x}_{k,l,m}^{(S)}), \\ k = 1, \dots, N_1 - 1; \quad l = 1, \dots, N_2 - 1; \quad m = 1, \dots, N_3 - 1; \quad \mathbf{x}_{k,l,m}^{(S)} = (r_k^{(S)}, \varphi_m^{(S)}, \vartheta_l^{(S)}).\end{aligned}\quad (27)$$

В (27) через $e_S(\mathbf{x}_{k,l,m}^{(S)})$ обозначена функция, определенная в (7); H — это среднее геометрическое из $\bar{h}_r$, h_{φ} , $\bar{h}_{\vartheta}$: $H = \sqrt[3]{\bar{h}_r h_{\varphi} \bar{h}_{\vartheta}}$, $\bar{h}_r = \frac{1}{N_1} \sum_{k=1}^{N_1} h_{r,k}$; $\bar{h}_{\vartheta} = \frac{1}{N_2} \sum_{l=1}^{N_2} h_{\vartheta,l}$.

Заметим, что в (27) по радиусу и углу ϑ вводится равномерная сетка, но для общего случая мы приведем систему конечно-разностных уравнений ниже.

Для однозначности определения сеточной функции $\Omega_3^{(S)}(\mathbf{x}_{(l,k,m)}^{(S)})$ необходимо, как и при формулировке сеточной краевой задачи (13), (14), указать, каким граничным условиям удовлетворяет указанная функция. Будем (для простоты) полагать, что сеточная функция $\Omega_3^{(S)}(\mathbf{x}_{(l,k,m)}^{(S)})$ задана в шаровом слое:

$$\begin{aligned}DS_R &= \{x^{(S)} : \rho(0) \leq |x^{(S)}| \leq R, \arcsin(4\tau(0)) \leq \vartheta \leq \pi - \arcsin(4\pi(N_2)), 0 \leq \varphi \leq 2\pi\}, \\ \rho(0) &= \begin{cases} \frac{1}{4} h_{r,l}, & r_0 = 0, \\ 0, & r_0 > 0, \end{cases} \\ \tau(0) &= \begin{cases} \frac{1}{4} \sin h_{\vartheta,l}, & \vartheta_0 = 0, \\ \sin \vartheta_0, & \vartheta_0 > 0, \end{cases} \quad \tau(N_2) = \begin{cases} \frac{1}{4} \sin h_{\vartheta,N_2}, & \vartheta_{N_2} = \pi, \\ \sin \vartheta_{N_2}, & \vartheta_{N_2} < \pi. \end{cases}\end{aligned}\quad (28)$$

В (28) r_0 и ϑ_0 , ϑ_{N_2-1} — некоторые константы, $h_{r,1}$ и $h_{\vartheta,1}$ — первые шаги дискретной сети по радиусу и углу ϑ соответственно, h_{ϑ,N_2-1} — последний шаг по углу ϑ . Для краткости обозначим $\mathbf{x}_{(l,k,m)}^{(S)}$ через $x^{(S)}$.

Тогда вместе с условиями

$$\Omega_3^{(S)} \Big|_{x^{(S)} \in \Gamma(DS_{R_n})} = 0, \quad n = 1, 2, \dots, \quad (29)$$

получим краевую задачу в сеточном пространстве, аналогичную (13), (14). В (29) через $\Gamma(DS_{R_n})$ обозначена граница n -го шарового слоя (вместо n -го параллелепипеда в декартовом случае).

Матрица оператора Λ_1 является блочной трехдиагональной, и он имеет следующий вид:

$$\Lambda_1 \Omega_3^{(S)}(x^{(S)}) = 2\Omega_n(x^{(S)}) - \rho_k^2 \left[\tau(\vartheta_l^{(S)}) \right]^2 \cdot h_\varphi^2 \Omega_3^{(S)}(x^{(S)})_{\bar{r}r} - \rho_k^2 \left[\tau(\vartheta_l^{(S)}) \right]^2 \cdot h_\varphi^2 \Omega_3^{(S)}(x^{(S)})_{\bar{\vartheta}\vartheta}, \quad (30)$$

$$1 \leq k \leq N_1 - 1; \quad 1 \leq l \leq N_2.$$

В (30) фактически рассматриваются значения сеточной функции в m -м слое сетки, соответствующем значению угла $\varphi_m^{(S)} = \varphi_{m+1}^{(S)} + mh_\varphi$, $m = 1, \dots, N_2$, $h_\varphi = \text{const}$, где $\varphi_m^{(S)}$ — дискретный аналог переменной φ в “континуальной” сферической системе координат (см. (21)).

Как и в рассмотренном в разд. 2 случае декартовой системы координат, для решения задачи (28)–(30) можно применить алгоритм матричной прогонки. Матрица системы линейных алгебраических уравнений будет блочно-трехдиагональной, элементы ее имеют вид (11), но вместо $\alpha_1 = \frac{h_3^2}{h_1^2}$; $\alpha_2 = \frac{h_3^2}{h_2^2}$ нужно будет записать следующие выражения:

$$\alpha_1 = \alpha_1(r_k^{(S)}, \vartheta_l^{(S)}) = h_\varphi^2 (r_k^{(S)})^2 \sin^2(\vartheta_l^{(S)}) \cdot \left(\frac{1}{h_{r,k} \cdot h_{r,k+1}} + \frac{(r_{k-1}^{(S)})^2}{(r_k^{(S)})^2 \cdot (h_{r,k})^2} \right), \quad (31)$$

$$\alpha_2 = \alpha_2(r_k^{(S)}, \vartheta_l^{(S)}) = h_\varphi^2 (r_k^{(S)})^2 \sin^2(\vartheta_l^{(S)}) \cdot \left(\frac{1}{h_{\vartheta,l} \cdot h_{\vartheta,l+1}} + \frac{\sin \vartheta_{l-1}^{(S)}}{\sin \vartheta_l^{(S)} \cdot (h_{\vartheta,l})^2} \right).$$

В (31) вместо функции $\tau(\vartheta_l^{(S)})$ фигурирует $\sin \vartheta_l^{(S)}$, что приводит к одним и тем же выражениям для элементов СЛАУ в случае, когда $\vartheta_0 > 0$ (см. определение вспомогательных сеточных функций (25)). Алгоритм матричной прогонки применим, если выполняется условие (17). Соотношения (31) легко получаются, если вспомнить, как записываются конечно-разностные аналоги вторых производных сеточного фундаментального решения по радиусу и углу ϑ :

$$\Delta_r^{(S)} \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) \equiv \frac{1}{(r_k^{(S)})^2} \left[(r_k^{(S)})^2 \left\{ \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) \right\}_r \right]_{\bar{r}} = \frac{\Omega_3^{(S)}(r_{k+1}^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)})}{h_{r,k+1} \cdot h_{r,k}} -$$

$$- \left(\frac{1}{h_{r,k} \cdot h_{r,k+1}} + \frac{(r_{k-1}^{(S)})^2}{(r_k^{(S)})^2 \cdot (h_{r,k})^2} \right) \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) + \frac{(r_{k-1}^{(S)})^2 \cdot \Omega_3^{(S)}(r_{k-1}^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)})}{(r_k^{(S)})^2 \cdot (h_{r,k})^2};$$

$$\Delta_\vartheta^{(S)} \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) \equiv \frac{1}{\sin \vartheta_l^{(S)}} \left[\sin \vartheta_l^{(S)} \left\{ \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) \right\}_\vartheta \right]_{\bar{\vartheta}} = \frac{\Omega_3^{(S)}(r_k^{(S)}, \vartheta_{l+1}^{(S)}, \varphi_m^{(S)})}{h_{\vartheta,l+1} \cdot h_{\vartheta,l}} -$$

$$- \left(\frac{1}{h_{\vartheta,l} \cdot h_{\vartheta,l+1}} + \frac{\sin \vartheta_{l-1}^{(S)}}{\sin \vartheta_l^{(S)} \cdot (h_{\vartheta,l})^2} \right) \Omega_3^{(S)}(r_k^{(S)}, \vartheta_l^{(S)}, \varphi_m^{(S)}) + \frac{\sin \vartheta_{l-1}^{(S)} \cdot \Omega_3^{(S)}(r_k^{(S)}, \vartheta_{l-1}^{(S)}, \varphi_m^{(S)})}{\sin \vartheta_l^{(S)} \cdot (h_{\vartheta,l})^2}. \quad (32)$$

Собственные значения оператора в Λ_1 в (30) можно оценить следующим образом.

Введем в пространстве сеточных функций от сферических координат $r_k^{(S)}$, $\vartheta_l^{(S)}$ скалярное произведение по формуле (таким образом, значение координаты $\varphi_m^{(S)}$ фиксировано):

$$(u, v) = \sum_{k=0}^{N_1} \sum_{l=0}^{N_2} u(k, l, m) v(k, l, m) \bar{h}_1(k) \bar{h}_2(l). \quad (33)$$

Согласно общей теории линейных операторов [12], собственные значения матрицы оператора Λ_1 определяются из соотношения Рэлея:

$$\lambda = \sup_{\|x^{(S)}\|_E=1} \frac{\|Ax^{(S)}\|_E}{\|x^{(S)}\|_E}. \quad (34)$$

В (34) точная верхняя грань берется по векторам сеточного пространства, имеющим евклидову норму, равную 1. Более подробно:

$$\|x^{(S)}\|_E = \sqrt{(x^{(S)}, x^{(S)})}. \quad (35)$$

В (35) через $(x^{(S)}, y^{(S)})$ обозначено скалярное произведение векторов сеточного пространства, определенное ранее в (33). В [12] получены оценки максимального и минимального собственных значений разностного эллиптического оператора в цилиндрических координатах.

Если для самосопряженного оператора A имеют место неравенства [12]

$$C_1 \delta(u, u) \leq (Au, u) \leq C_2 \Delta(u, u), \quad (36)$$

то собственные значения этого оператора лежат в диапазоне

$$C_1 \delta \leq \lambda_A \leq C_2 \Delta. \quad (37)$$

В случае сферических координат можно сделать аналогичные выводы относительно собственных значений указанного оператора, если ввести скалярное произведение (33) для сеточных функций, являющихся решениями задачи Дирихле. Для того чтобы убедиться в этом, запишем соотношение (20) следующим образом:

$$\begin{aligned} \Lambda_1 \Omega_3^{(S)}(x^{(S)}) &= 2\Omega_3^{(S)}(x^{(S)}) - \rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 \Omega_3^{(S)}(x^{(S)})_{\bar{r}r} - \rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 \Omega_3^{(S)}(x^{(S)})_{\bar{\vartheta}\vartheta} = \\ &= \left(2E - \rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 A_1 - \rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 A_2 \right) \Omega_3^{(S)}(x^{(S)}), \\ &1 \leq k \leq N_1 - 1; \quad 1 \leq l \leq N_2. \end{aligned} \quad (38)$$

Таким образом, собственные значения оператора Λ_1 можно оценить, если известны собственные значения операторов $\rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 A_1$ и $\rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot h_\varphi^2 A_2$. Поскольку шаг по углу φ является константой ($h_\varphi = \text{const}$), то нам нужно исследовать свойства операторов A_1 и A_2 , точнее говоря, их суммы, умноженной на $\rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2$. Но $\rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot [A_1 + A_2]$ является самосопряженным оператором при определении скалярного произведения согласно (33).

В самом деле,

$$\begin{aligned} \rho_k^2 \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot [A_1 + A_2] \Omega_3^{(S)}(x^{(S)}) &= \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot \left\{ \rho_k^2 \left(\Omega_3^{(S)}(x^{(S)}) \right) \right\}_{\bar{r}} + \\ &+ \tau \left(\vartheta_l^{(S)} \right) \left\{ \tau \left(\vartheta_l^{(S)} \right) \cdot \left(\Omega_3^{(S)}(x^{(S)}) \right) \right\}_{\bar{\vartheta}}. \end{aligned} \quad (39)$$

В (39) первое слагаемое обозначим через $B_1 \Omega_3^{(S)}$, а второе слагаемое — через $B_2 \Omega_3^{(S)}$. Если скалярно умножить $B_1 \Omega_3^{(S)} + B_2 \Omega_3^{(S)}$ на некоторую сеточную функцию $u(x^{(S)})$, удовлетворяющую условиям Дирихле на границе области задания $\Omega_3^{(S)}$ (скалярное произведение при этом определено по (33)), то мы получим

$$\begin{aligned} (B_1 \Omega_3^{(S)} + B_2 \Omega_3^{(S)}, u) &= \left[\tau \left(\vartheta_l^{(S)} \right) \right]^2 \cdot \left\{ \rho_k^2 (u(x^{(S)})) \right\}_{\bar{r}} \cdot \Omega^{(S)} + \\ &+ \tau \left(\vartheta_l^{(S)} \right) \left\{ \tau \left(\vartheta_l^{(S)} \right) \cdot (u(x^{(S)})) \right\}_{\bar{\vartheta}} \cdot \Omega_3^{(S)} = \left(\Omega_3^{(S)}, B_1 u + B_2 u \right). \end{aligned} \quad (40)$$

Таким образом, мы показали, что оператор $B_1 + B_2$ является самосопряженным в евклидовом пространстве сеточных функций со скалярным произведением (33).

Собственные значения B_1 могут быть оценены через собственные значения одномерного оператора Лапласа для задачи Дирихле на отрезке по r от $\rho(0)$ до R :

$$\begin{aligned} \beta_{1\min} \delta_1 &\leq \lambda_1^{(l_1)} \leq \beta_{1\max} \Delta_1, \quad l_1 = 1, \dots, N_1 - 1; \\ \beta_{1\min} &= \min \{ \tau^2(l_2), l_2 = 0, \dots, N_2 \}, \quad \beta_{1\max} = \max \{ \tau^2(l_2), l_2 = 0, \dots, N_2 \}; \\ \delta_1 &= \min \left[\frac{4\pi^2(1)}{h_{1,r}^2}, \max_{1 \leq k \leq N_1-1} \frac{2}{\bar{h}_1(k)} \cdot \left(\frac{\rho^2(k)}{h_{k,r}} + \frac{\rho^2(k+1)}{h_{k+1}} \right) \right], \\ \Delta_1 &= \min \left[\frac{4\pi^2(1)}{h_{1,r}^2}, \max_{1 \leq k \leq N_1-1} \frac{2}{\bar{h}_1(k)} \cdot \left(\frac{\rho^2(k)}{h_{k,r}} + \frac{\rho^2(k+1)}{h_{k+1}} \right) \right], \\ \bar{h}_1(k) &= \frac{1}{2} [h_{k,r} + h_{k+1,r}]. \end{aligned} \quad (41)$$

В (41) мы сохранили обозначения для шагов по координате r и функции $\rho(k)$, значения которой в точках сетки равны радиусу, за исключением начала координат (см. (28)).

Для оценки собственных значений оператора B_2 введем новую переменную

$$\eta = 0.5 \ln \left(\frac{1 - \cos \vartheta}{1 + \cos \vartheta} \right). \quad (42)$$

В (42) учтено, что $\tau(\vartheta_l^{(S)})$ — это $\sin \vartheta_l^{(S)}$, за исключением концов отрезка по углу ϑ (в случае, когда $0 \leq \vartheta \leq \pi$ или выполняется хотя бы одно из неравенств, входящих в это двойное неравенство).

По переменной η мы получим вместо B_2 оператор Лапласа и краевую задачу Дирихле на отрезке:

$$\begin{aligned} \eta &\in [\beta_{2\min}; \beta_{2\max}], \\ \beta_{2\min} &= \min \left\{ 0.5 \ln \left(\frac{1 - \cos(4 \arcsin \tau(0))}{1 + \cos(4 \arcsin \tau(0))} \right), 0.5 \ln \left(\frac{1 - \cos(4 \arcsin \tau(N_2 - 1))}{1 + \cos(4 \arcsin \tau(N_2 - 1))} \right) \right\}, \\ \beta_{2\max} &= \max \left\{ 0.5 \ln \left(\frac{1 - \cos(4 \arcsin \tau(0))}{1 + \cos(4 \arcsin \tau(0))} \right), 0.5 \ln \left(\frac{1 - \cos(4 \arcsin \tau(N_2 - 1))}{1 + \cos(4 \arcsin \tau(N_2 - 1))} \right) \right\}. \end{aligned} \quad (43)$$

Собственные значения оператора Лапласа на указанном отрезке при условии равенства всех шагов по координате η лежат в диапазоне:

$$\begin{aligned} \delta_2 &\leq \lambda_2^{(l_2)} \leq \Delta_2, \quad \lambda_2^{(l_2)} = \frac{4}{\bar{h}_2^2} \sin^2 \frac{\pi l_2}{2N_2}, \quad l_2 = 1, \dots, N_2 - 1; \\ \delta_2 &= \frac{4}{\bar{h}_2^2} \sin^2 \frac{\pi}{2N_2} \geq \frac{8}{\tilde{d}_2^2}, \quad \tilde{d}_2 = \beta_{2\max} - \beta_{2\min}, \\ \tilde{h}_2 N_2 &= \tilde{d}_2; \quad \Delta_2 = \frac{4}{\bar{h}_2^2} \cos^2 \frac{\pi}{2N_2} \leq \frac{4}{\tilde{h}_2^2}. \end{aligned} \quad (44)$$

Окончательно, будем иметь следующие оценки для собственных значений оператора Δ_1 на каждом из слоев дискретной сети, соответствующем некоторому значению угла φ :

$$\begin{aligned} \lambda_{\Delta_1}^{(l_1, l_2)} &= 2 + h_\varphi^2 \left(\lambda_1^{(l_1)} + \lambda_2^{(l_2)} \right), \\ 2 + h_\varphi^2 (\beta_{1\min} \delta_1 + \delta_2) \beta_{1\min} \delta_1 + \delta_2 &\leq \lambda_{\Delta_1}^{(l_1, l_2)} \leq 2 + h_\varphi^2 (\beta_{1\max} \Delta_1 + \Delta_2), \\ 1 &\leq l_1 \leq N_1 - 1; \quad 1 \leq l_2 \leq N_2 - 1. \end{aligned} \quad (45)$$

Если учесть, что вторые производные по радиусу вычисляются по формулам (32) (первое равенство), то можно сделать вывод о том, что для сеточного фундаментального решения, зависящего только от радиуса, справедлива

Теорема 2. *Сферически симметричное сеточное фундаментальное решение определяется из (27)–(29) в каждом шаровом слое DS_R однозначно: выполнены условия корректности метода монотонной прогонки по радиальному направлению для задачи Дирихле.*

Снова заметим, что корректность алгоритма прогонки гарантирует также существование решения (27)–(29), зависящего только от радиуса.

Оценка (45) для собственных значений конечно-разностного оператора Δ_1 в общем случае позволяет утверждать, что система (27)–(29) будет однозначно разрешима и при зависимости правой части от всех трех сферических координат, что особенно важно для приложений.

Таким образом, справедлива

Теорема 3. Система (27)–(29) может быть решена с помощью метода матричной прогонки при любой зависимости правой части от сферических координат: для матриц $C_k = C$, $A_{N_3} = B_0 = 0$, $C_0 = C_{N_3} = E$, $B_m = A_m = E$, $1 \leq m \leq N_3$, в каждом слое, соответствующем $\varphi_m^{(S)}$, $m = 1, \dots, N_3$ выполнены условия корректности алгоритма (17).

4. ВАРИАЦИОННАЯ ПОСТАНОВКА ПО ОПРЕДЕЛЕНИЮ СЕТОЧНОГО ФУНДАМЕНТАЛЬНОГО РЕШЕНИЯ НА СЕМЕЙСТВЕ РАСШИРЯЮЩИХСЯ КОМПАКТОВ

Из свойств конечно-разностного аналога оператора Лапласа следует, что функционал (9) — выпуклый и непрерывный. Мы видим также, что дискретный аналог оператора Лапласа (8) — линейный оператор, действующий на пространстве сеточных функций в декартовой системе координат.

В [14] доказывается утверждение о том, что для выпуклого, замкнутого и ограниченного множества G множество $G^* = \{\varphi^* \in G : J(\varphi^*) = \mu\} \neq \emptyset$ выпукло, замкнуто и ограничено. Здесь $J(\varphi^*)$ — выпуклый и непрерывный функционал на некотором гильбертовом пространстве, φ^* — элемент этого пространства. В качестве гильбертова пространства мы выберем пространство сеточных функций Ξ со скалярным произведением, определенным по формуле:

$$\begin{aligned} (u, v) = & \sum_{k=1}^{N_1-1} \sum_{l=1}^{N_2-1} \sum_{m=1}^{N_3-1} u(k, l, m) v(k, l, m) h_1 h_2 h_3 + \\ & + 0.5 \left[\sum_{k=1}^{N_1-1} \sum_{l=1}^{N_2-1} \{u(k, l, 0) v(k, l, 0) + u(k, l, N_3) v(k, l, N_3)\} \right] + \\ & + 0.5 \left[\sum_{k=1}^{N_1-1} \sum_{m=1}^{N_3-1} \{u(k, 0, m) v(k, 0, m) + u(k, N_2, m) v(k, N_2, m)\} \right] + \\ & + 0.5 \left[\sum_{m=1}^{N_3-1} \sum_{l=1}^{N_2-1} \{u(0, l, m) v(0, l, m) + u(N_1, l, m) v(N_1, l, m)\} \right]. \end{aligned} \quad (46)$$

Таким образом, сеточные функции $u(k, l, m), v(k, l, m) \in \Xi$.

Замечание. Введенное в (46) скалярное произведение соответствует постановкам краевых задач с условиями Дирихле.

Введем множества

$$G_R = G \cap K_R, \quad F_R = \{f \in \Phi, \quad f = \Lambda^+ \varphi, \quad \varphi \in G_R\},$$

$$K_R = \{\varphi \in \Phi : \|\varphi\| \leq R\}, \quad S_R = \{\varphi \in \Phi : \|\varphi\| = R\}, \quad \varphi \in \Xi. \quad (47)$$

В (47) через Λ^+ обозначен оператор Лапласа на шаблоне “крест” (см. (8)). Из того, что $K_R \subseteq K_{R'}$ при $\bar{R} \leq R' \leq R'' \leq \bar{R}$ следует справедливость включения $G_R \subseteq G_{R'}$, $F_R \subseteq F_{R'}$.

Согласно теореме 2, доказанной в [14], при $\forall R$ F_R выпукло, слабо замкнуто и ограничено. Это значит, что при $\forall R \geq \bar{R}$ существует единственный $f_R^* \in F_R$ такой, что

$$\|f_R^* - f\|_{\Xi'} = \inf_{f \in F_R} \|f - \bar{f}\|_{\Xi'}. \quad (48)$$

Через Ξ' мы обозначили сеточное пространство функций, определенных во внутренних узлах исходной сетки (поскольку шаблон “крест” предполагает вычисление значений функций именно в таких

узлах). Назовем f^* и f_R^* метрическими проекциями f соответственно на множества F и F_R . Очевидно,

$$\begin{aligned}\mu &= \inf_{\varphi \in G} J(\varphi) = \inf_{f \in F} \|f - \bar{f}\|_{\Xi'}^2 = \|f^* - \bar{f}\|_{\Xi'}^2, \\ \mu(R) &= \inf_{\varphi \in G_R} J(\varphi) = \inf_{f \in F_R} \|f - \bar{f}\|_{\Xi'}^2 = \|f_R^* - \bar{f}\|_{\Xi'}^2.\end{aligned}\quad (49)$$

Поскольку $J(\varphi)$ — непрерывный и выпуклый функционал, G_R — выпукло, замкнуто и ограничено, F_R — выпукло, слабо замкнуто и ограничено, G и F — подмножества гильбертовых пространств, $G_{R'} \subseteq G_{R''}$, $F_{R'} \subseteq F_{R''}$ при $\bar{R} \leq R' \leq R'' \leq \bar{\bar{R}}$, как было показано в [14], то $\mu(R)$ — непрерывная функция при $R \in [\bar{R}, \bar{\bar{R}}]$.

Можно показать (см. [14]), что $J(\varphi)$ достигает своей точной нижней грани на G_R в единственной точке, такой что $\|\varphi_R^*\| = R$.

Через R мы обозначили расстояние от начала координат до точки в пространстве сеточных фундаментальных решений. Как было уже отмечено, норма пространства Ξ связана со скалярным произведением, введенным в (46).

Если под $J(\varphi)$ мы будем понимать функционал качества (9), отражающий близость сеточного фундаментального решения к классическому, континуальному, то можно утверждать, что справедлива

Теорема 4. *На семействе расширяющихся компактов K_{R_n} , $R_N \rightarrow \infty$, существует единственное решение задачи (8)–(9).*

При построении решения задачи (8), (9) на семействе расширяющихся компактов следует использовать регуляризирующие алгоритмы, обеспечивающие высокую точность искомого сеточного фундаментального решения [15].

ЗАКЛЮЧЕНИЕ

1. В статье рассмотрены условия однозначной разрешимости системы линейных алгебраических уравнений для определения сеточного фундаментального решения в рамках теории дискретного гравитационного потенциала в декартовых и сферических координатах в некоторой ограниченной сеточной области. Приводятся различные варианты постановок условно-вариационных задач по поиску сеточного фундаментального решения.
2. В работе описывается алгоритм построения сеточных фундаментальных решений на семействе расширяющихся компактов, который позволяет находить значения дискретного потенциала и его производных в любой точке трехмерного сеточного пространства, что важно для геофизических приложений [15].

Авторы выражают глубокую благодарность А. С. Леонову за полезные замечания и внимание к работе.

СПИСОК ЛИТЕРАТУРЫ

1. Страхов В. Н., Степанова И. Э., Гричук Л. В. Теория дискретного гравитационного потенциала и ее использование в гравиметрии // В сб.: “Вопросы теории и практики геологической интерпретации гравитационных, магнитных и электрических полей”, Труды международной конференции. Воронеж. Воронежский государственный университет. 1996. С. 49–71.
2. Арсанукаев З. З. Вычисление пространственных элементов аномальных полей с использованием методов теории дискретных гравитационных полей // Физ. Земли. 2004. № 11. С. 47–69.
3. Страхов В. Н., Степанова И. Э. Метод S-аппроксимаций и его использование при решении задач гравиметрии (локальный вариант) // Физ. Земли. 2002. № 2. С. 3–19.
4. Страхов В. Н., Степанова И. Э. Метод S-аппроксимаций и его использование при решении задач гравиметрии (региональный вариант) // Физ. Земли. 2002. № 7. С. 3–12.

5. *Stepanova I. E., Kerimov I. A., Yagola A. G.* Approximation approach in various modifications of the method of linear integral representations // *Izvestiya. Physics of the Solid Earth*. 2019. V. 55. N. 2. P. 218–231.
6. *Раевский Д. Н., Степанова И. Э.* О решении обратных задач гравиметрии с помощью модифицированного метода S-аппроксимаций // *Физ. Земли*. 2015. № 2. С. 44–54.
7. *Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г.* Численные методы решения некорректных задач. М.: Наука, 1990. 230 с.
8. *Лаврентьев М. А., Люстерник Л. А.* Курс вариационного исчисления. М.Л.: Гостоптехиздат, 1950. 296 с.
9. *Kolotov I. I., Lukyanenko D. V., Stepanova I. E., Shchepetilov A. V., Yagola A. G.* On the uniqueness of solution to systems of linear algebraic equations to which the inverse problems of gravimetry and magnetometry are reduced: a regional variant // *Comput. Math. and Math. Phys.* 2023. V. 63. № 9. P. 1588–1599.
10. *Kolotov I. I., Lukyanenko D. V., Stepanova I. E., Yagola A. G.* On the uniqueness of solutions to systems of linear algebraic equations resulting from the reduction of linear inverse problems of gravimetry and magnetometry: a local case // *Comput. Math. and Math. Phys.* 2023. V. 63. № 8. P. 1452–1465.
11. *Леонов А. С.* Метод минимальной псевдообратной матрицы // *Ж. вычисл. матем. и матем. физ.* 1987. Т. 27. № 8. С. 1123–1138.
12. *Самарский А. А., Николаев Е. С.* Методы решения сеточных уравнений. М.: Наука, 1978. 592 с.
13. *Михайлов В. П.* Дифференциальные уравнения в частных производных. М.: Наука, 1976. 391 с.
14. *Гавурин М. К., Фабровская Ю. Б.* Об одном итеративном методе разыскания суммы квадратов // *Ж. вычисл. матем. и матем. физ.* 1966. Т. 6. № 6. С. 1094–1097.
15. *Leonov A. S.* Extraoptimal A Posteriori Estimates of the Solution Accuracy in the Ill-Posed Problems of the Continuation of Potential Geophysical Fields // *Izvestiya, Physics of the Solid Earth*, 2011. V. 47. N. 6. P. 531–540.

ON THE UNAMBIGUITY OF DETERMINING THE GRID FUNDAMENTAL SOLUTION OF THE LAPLACE EQUATION IN THE THEORY OF DISCRETE POTENTIAL

I. E. Stepanova^{a,*}, I. I. Kolotov^b, A. G. Yagoda^b, A. N. Levashov^b

^a 123242 Moscow, B. Gruzinskaya str., 10, Bldg. 1, O. Yu. Schmidt Institute of Physics of the Earth, Russia

^b 119992 Moscow, Lenin Hills, Moscow State University, Russia

*e-mail: tet@ifz.ru

Received 13 March, 2024

Revised 13 March, 2024

Accepted 14 March, 2024

Abstract. The paper considers the problem of unambiguously determining the fundamental solution of the grid analogue of the Laplace equation within the framework of the theory of discrete gravitational potential. The grid fundamental solution of the finite-difference analogue of the Laplace equation plays a key role in reconstructing a continuously distributed source of a gravitational or magnetic field from heterogeneous and unequally accurate data obtained at points of a certain grid set.

Keywords: unambiguous definition, fundamental solution, discrete gravitational potential.

О ЧИСЛЕННОМ РЕШЕНИИ УРАВНЕНИЙ ВЛАСОВА—АМПЕРА

© 2024 г. Е. В. Чижонков^{1,*}¹ 119899 Москва, Ленинские горы, МГУ им. М. В. Ломоносова, Россия

*e-mail: chizhonk@mech.math.msu.su

Поступила в редакцию 02.11.2023 г.

Переработанный вариант 10.03.2024 г.

Принята к публикации 03.04.2024 г.

Для кинетической модели плазмы, базирующейся на уравнениях Власова—Ампера, построена неявная схема типа Мак-Кормака. По сравнению с явной схемой она имеет более слабое ограничение на устойчивость, но сохраняет прежнюю вычислительную эффективность, т.е. не использует внутренние итерации. При этом погрешность полной энергии соответствует второму порядку точности алгоритма, а полный заряд (число частиц) сохраняется на сеточном уровне. В качестве моделируемого физического процесса рассматривается формирование плазменных волн, возбуждаемых коротким мощным лазерным импульсом. Библ. 31. Фиг. 8.

Ключевые слова: кинетическая модель плазмы, уравнения Власова—Ампера, плазменные колебания и волны, численное моделирование, неявная схема Мак-Кормака.

DOI: 10.31857/S0044466924070116, EDN: xidnpk

ВВЕДЕНИЕ

Плазма имеет решающее значение для различных приложений. Понимание сложного поведения плазмы привело к важным достижениям в области космической физики, термоядерной энергетики и создания мощных микроволновых излучений и крупномасштабных ускорителей частиц (см. [1]). Однако изучение поведения плазмы с использованием традиционных экспериментальных методов является сложной задачей из-за ее высокой динамичности, что побуждает к разработке различных методов моделирования. Теоретическим фундаментом современного моделирования плазменных процессов, безусловно, является уравнение Власова (см. [2]). Однако, как и большинство кинетических уравнений, оно требует не только нетривиальных алгоритмов для получения приближенного решения, но и весьма значительных объемов вычислений для их реализации.

Для численного решения уравнения Власова, как правило, используются два основных подхода: на основе частиц и на основе сеток. Подход на основе сеток часто называют *эйлеровым*. Метод *частиц в ячейке* — хорошо известный способ отслеживания движения конечного числа макрочастиц (см. [3]). Хотя такой подход эффективен в вычислительном отношении и подходит для задач большой размерности, он может создавать собственный численный шум, особенно для потоков плазмы с небольшими возмущениями. Поэтому в последнее время предпочтение отдается *эйлерову* моделированию на основе решения кинетического уравнения непосредственно в фазовом пространстве. Хорошие обзоры сеточных *эйлеровых* методов можно найти в работах [4]–[6].

В настоящей работе основное внимание уделяется системе Власова—Ампера. Эту систему можно рассматривать как нулевой магнитный предел общей системы Власова—Максвелла. Кроме того, она эквивалентна наиболее популярной системе Власова—Пуассона, если решение гладкое и отсутствуют внешние силы. Однако в отличие от системы Власова—Пуассона здесь электрическое поле не находится из уравнения Пуассона, а смещается во времени за счет тока, генерируемого движущимися зарядами. Для решения системы Власова—Ампера известны численные алгоритмы, основанные на различных идеях: метод “частиц в ячейке” (см. [7]), метод конечных разностей типа Мак-Кормака (см. [8]), полулагранжев метод (см. [9]), схема конечного объема второго порядка (см. [10]), разрывный метод Галеркина (см. [11]) и пр. Следует отметить, что сейчас основным направлением в разработке подобных алгоритмов является контроль за сохранением энергии при снятии ограничений на

устойчивость, присущих явным расчетным схемам (см. [12], [13]). Это связано с тем, что при моделировании на длительных временных диапазонах паразитная энергия, создаваемая или уничтожаемая численными методами, может накапливаться и приводить к нефизическим результатам, таким как самонагрев или охлаждение плазмы (см. [14]). Проблема становится более заметной, если используется сетка с недостаточным разрешением или большие временные шаги.

Работа посвящена развитию эйлерова сеточного метода типа Мак-Кормака (см. [8]). Новизна заключается в сохранении эффективности явной схемы (т.е. без использования итераций) при наличии устойчивости, как у неявной схемы (см. [15]). При этом погрешность полной энергии соответствует второму порядку точности алгоритма, а полный заряд (число частиц) сохраняется на сеточном уровне. В качестве моделируемого физического процесса рассматривается динамика плазменных (ленгмюровских) колебаний и волн, возбуждаемых коротким мощным лазерным импульсом (см. [16]).

Статья имеет следующую структуру. В разд. 1 приведена простейшая кинетическая постановка задачи Коши на основе уравнений Власова—Ампера, описывающая нерелятивистские колебания и волны плазмы. В разд. 2 описана оригинальная неявная схема Мак-Кормака для модельного скалярного уравнения переноса, а также отмечены наиболее важные вычислительные свойства этой схемы. В разд. 3 построена неявная схема Мак-Кормака для системы нелинейных уравнений из первого раздела. Наконец, в разд. 4 описаны результаты вычислительных экспериментов, иллюстрирующих эффективность предложенного подхода. В заключении систематизированы результаты проведенных исследований.

1. ПОСТАНОВКА ЗАДАЧИ

Будем считать, что одномерное движение разреженной плазмы описывается в фазовой плоскости (x, v) следующим безразмерным уравнением Власова:

$$\frac{\partial F}{\partial t} + v \frac{\partial F}{\partial x} - E(x, t) \frac{\partial F}{\partial v} = 0, \quad (1)$$

где $F = F(x, v, t)$ — функция распределения электронов, $E = E(x, t)$ — напряженность электрического поля. При отсутствии внешних (заданных) полей функции E и F связаны соотношением

$$\frac{\partial}{\partial x} E(x, t) = 1 - N(x, t), \quad (2)$$

где функция $N(x, t) = \int_{-\infty}^{\infty} F dv$ имеет смысл плотности электронов. Как правило, вместо функции электрического поля в уравнениях (1), (2) используется его потенциал, удовлетворяющий в векторном случае соотношению $\text{grad } \varphi = \mathbf{E}$, тогда эту систему двух уравнений называют уравнениями Власова—Пуассона. Формальной причиной для этого является преобразование уравнения (2) относительно φ в известное уравнение Пуассона.

В уравнении (2) символ “1” означает, что ионы неподвижны, а их плотность (безразмерная) постоянна. Поэтому согласно одному из уравнений Максвелла, которое часто называют уравнением Ампера, дополнительно имеем

$$\frac{\partial E}{\partial t} = \int_{-\infty}^{\infty} v F dv. \quad (3)$$

Если для замыкания системы к уравнению (1) добавляют уравнение (3), то такую систему двух уравнений называют уравнениями Власова—Ампера.

Система уравнений (1)–(3) давно и хорошо известна (см., например, [17]). При моделировании плазмы в каждом конкретном случае применяют либо подсистему (1), (2), либо — (1), (3). Учитывая форму уравнения (1), для обозначения любой из подсистем используют термин “1D1V-система”.

Рассмотрим возбуждение ленгмюровских колебаний и волн с помощью короткого мощного лазерного импульса. Согласно кинетической модели распространения лазерных импульсов в плазме

(см. [18], [19]) вынужденное изменение электрического поля можно описать, используя градиент квадрата модуля медленно изменяющейся комплексной амплитуды высокочастотного лазерного поля (так называемой огибающей). При этом, если импульс имеет гауссово пространственное распределение, формулу возмущающего электрического поля удобно задать в виде

$$E_0(x) = \alpha x \exp \left\{ -2 \frac{x^2}{\rho_*^2} \right\}, \quad \alpha = \left(\frac{a_*}{\rho_*} \right)^2, \quad (4)$$

где параметры ρ_* и a_* характеризуют масштаб области локализации и максимальную величину $E_{\max} = \frac{a_*^2}{\rho_* 2\sqrt{e}} \approx 0.3 \frac{a_*^2}{\rho_*}$ электрического поля. Здесь $e = 2.71828 \dots$ — основание натурального логарифма. Вид функции (4) выбран из соображений, что подобное возмущение электрического поля может возбуждаться в разреженной плазме:

$$\omega_l \gg \omega_p = \left(\frac{4\pi e^2 n_0}{m} \right)^{1/2},$$

(n_0 — значение невозмущенной плотности электронов, e, m — заряд и масса электрона) лазерным импульсом с частотой ω_l при его фокусировке в линию (этого можно добиться при использовании цилиндрической линзы, см. детали в [20]).

Формула (4) порождает однозначное возмущение равновесного значения электронной плотности на основании уравнения (2), т.е.

$$N_0(x) = 1 - \frac{\partial}{\partial x} E_0(x). \quad (5)$$

При этом параметры ρ_* и a_* в (4) должны обеспечивать положительность начальной электронной плотности, т.е. выполнение условия $\alpha < 1$. В свою очередь, соотношение (5) дает возможность задать начальное значение функции распределения, необходимое для решения уравнения Власова (1),

$$F(x, v, t = 0) = N_0(x) f_M(v, T_0), \quad (6)$$

где функция

$$f_M(v, T_0) = \frac{1}{\sqrt{2\pi T_0}} \exp \left\{ -\frac{v^2}{2T_0} \right\} \quad (7)$$

определяет максвелловское распределение скоростей при фиксированной температуре T_0 .

Для рассматриваемой постановки (1)–(7) справедлива теорема существования и единственности (см. [21]): “Если непрерывная начальная функция $F(x, v, t = 0) > 0$ такова, что

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} F(x, v, t = 0) dv - 1 \right] dx = 0, \\ F(x, v, t = 0) < G(v), \quad \int_{-\infty}^{\infty} v^2 G(v) dv < \infty,$$

где $G(v)$ — монотонно убывающая функция $|v|$, то решение задачи Коши для системы (1)–(2) существует и единственно в классе непрерывных функций $E(x, t)$, имеющих ограниченную производную $\partial E / \partial x$ на всей оси x и обращающихся в нуль при $x \rightarrow \pm \infty$ ”.

Заметим, что в теореме используется по существу ограничение на параметр α ($\alpha < 1$), так как требуется обеспечить положительность начальной плотности электронов.

Ниже нас будет интересовать в полупространстве $t > 0$ решение задачи Коши для уравнений (1), (3) с начальными условиями (4), (6). Обратим внимание, что в постановке принципиально отсутствует периодичность решения по пространству, т.е. рассматривается решение с конечной энергией.

2. НЕЯВНАЯ СХЕМА МАК-КОРМАКА ДЛЯ МОДЕЛЬНОЙ ЗАДАЧИ

Следуя оригинальной идее [15], напомним основные формулы неявной схемы Мак-Кормака для простейшего уравнения переноса с постоянным коэффициентом a :

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0, \quad x \in \mathbb{R},$$

снабженного начальным условием $u(x, 0) = u_0(x)$ таким, что $u_0(x) = 0$ при $|x| \rightarrow \infty$.

Определим дискретизацию независимых переменных с помощью постоянных параметров τ и h так, что

$$t^n = n\tau, \quad n \geq 0, \quad x_i = ih, \quad i = 0, \pm 1, \pm 2, \dots,$$

и будем обозначать зависящую переменную $u(x, t)$ в узле сетки (x_i, t^n) как u_i^n .

Запишем неявную схему Мак-Кормака в традиционной форме “предиктор–корректор”, считая известными величины u_i^n :

1) предиктор с результатом u_i^p —

$$\begin{aligned} \Delta u_i^n &= -\frac{a\tau}{h} (u_{i+1}^n - u_i^n), \\ \left(1 + \frac{\lambda\tau}{h}\right) \delta u_i^p &= \Delta u_i^n + \frac{\lambda\tau}{h} \delta u_{i+1}^p, \quad u_i^p = u_i^n + \delta u_i^p, \end{aligned} \quad (8)$$

где вычисления проводятся в сторону уменьшения индекса $i = \dots, k+1, k, k-1, \dots$;

2) корректор с результатом u_i^c —

$$\begin{aligned} \Delta u_i^p &= -\frac{a\tau}{h} (u_i^p - u_{i-1}^p), \\ \left(1 + \frac{\lambda\tau}{h}\right) \delta u_i^c &= \Delta u_i^p + \frac{\lambda\tau}{h} \delta u_{i-1}^c, \quad u_i^c = u_i^p + \delta u_i^c, \end{aligned} \quad (9)$$

где вычисления проводятся в сторону увеличения индекса $i = \dots, k-1, k, k+1, \dots$.

В формулах (8), (9) верхний индекс p (или c) обозначает шаг предиктор (или корректор) или n — временной слой t^n , λ — постоянный параметр схемы, который будет определен ниже.

Окончательная формула, формирующая решение на следующем временном слое с номером $(n+1)$, имеет вид

$$u_i^{n+1} = (u_i^p + u_i^c) / 2.$$

Отметим, что если положить $\lambda = 0$, то неявная схема трансформируется в обычную явную схему Мак-Кормака (см. [22]) с условием устойчивости $\omega = |a|\tau/h \leq 1$ (условие Куранта) и первым дифференциальным приближением (см. [23]) вида

$$u_t + au_x = \frac{ah^2}{6} (\omega^2 - 1) u_{xxx} + O(\tau^3 + h^3).$$

Если же выбирать λ из условия

$$\lambda \geq \frac{1}{2} \max \left\{ |a| - \frac{h}{\tau}, 0 \right\},$$

то неявная схема (при $\lambda > 0$) будет безусловно устойчивой, а в правую часть первого дифференциального приближения добавится слагаемое (см. [24])

$$-\tau^2 a \lambda \left(\lambda + \frac{h}{\tau} \right) u_{xxx}.$$

Иными словами, на гладких решениях погрешность аппроксимации неявной схемы асимптотически не хуже, чем погрешность явной схемы. При этом замена явной схемы на неявную требует незначительного увеличения вычислительной работы, а сама неявная процедура носит безытерационный характер. Вышесказанное серьезно увеличивает привлекательность неявной схемы Мак-Кормака.

3. НЕЯВНАЯ СХЕМА ДЛЯ УРАВНЕНИЙ ВЛАСОВА–АМПЕРА

При численном решении задачи Коши для уравнений (1), (3) с начальными условиями (4), (6) искомые функции рассматриваются в ограниченной области

$$\Omega = \{|x| \leq x_{\max}, |v| \leq v_{\max}, 0 \leq t \leq t_{\max}\}.$$

Дискретизация независимых переменных проводится с помощью постоянных параметров τ , h_x , h_v так, что

$$t^n = n\tau, \quad n \geq 0, \quad x_i = ih_x - x_{\max}, \quad 0 \leq i \leq M_x, \\ v_j = \left(j - \frac{1}{2}\right)h_v - v_{\max}, \quad 1 \leq j \leq M_v,$$

где $h_x = 2x_{\max}/M_x$, $h_v = 2v_{\max}/M_v$, M_v — четное. Искомую переменную $F(x, v, t)$ в узле сетки (x_i, v_j, t^n) будем обозначать как $F_{i,j}^n$, аналогично $E(x, t)$ в узле сетки $(x_i, t^n) = E_i^n$.

Введем полезные обозначения для операторов пространственных разностей “вперед” D^+ и “назад” D^- :

$$D_x^+ G_{i,j} = G_{i+1,j} - G_{i,j}, \quad D_x^- G_{i,j} = G_{i,j} - G_{i-1,j}, \\ D_v^+ G_{i,j} = G_{i,j+1} - G_{i,j}, \quad D_v^- G_{i,j} = G_{i,j} - G_{i,j-1},$$

и запишем неявную схему Мак-Кормака (см. [15]) для уравнения (1) в традиционной форме предиктор—корректор, считая известными величины $F_{i,j}^n$, $E_i^{n+1/2}$:

1) шаг предиктор с результатом $F_{i,j}^p$ —

$$\Delta F_{i,j}^n = -\tau \left(\frac{v_j}{h_x} D_x^+ F_{i,j}^n - \frac{E_i^{n+1/2}}{h_v} D_v^+ F_{i,j}^n \right), \\ \left(I - \lambda \frac{\tau}{h_x} |v_j| D_x^+ \right) \left(I - \lambda \frac{\tau}{h_v} |E_i^{n+1/2}| D_v^+ \right) \delta F_{i,j}^p = \Delta F_{i,j}^n, \\ F_{i,j}^p = F_{i,j}^n + \delta F_{i,j}^p, \quad (10)$$

2) шаг корректор с результатом $F_{i,j}^c$ —

$$\Delta F_{i,j}^p = -\tau \left(\frac{v_j}{h_x} D_x^- F_{i,j}^p - \frac{E_i^{n+1/2}}{h_v} D_v^- F_{i,j}^p \right), \\ \left(I + \lambda \frac{\tau}{h_x} |v_j| D_x^- \right) \left(I + \lambda \frac{\tau}{h_v} |E_i^{n+1/2}| D_v^- \right) \delta F_{i,j}^c = \Delta F_{i,j}^p, \\ F_{i,j}^c = F_{i,j}^p + \delta F_{i,j}^c. \quad (11)$$

В формулах (10), (11) верхний индекс p (или c) обозначает шаг предиктор (или корректор) или n — временной слой t^n , λ — постоянный параметр схемы, который будет определен ниже.

Окончательные формулы, формирующие решение на следующем временном слое с номером $(n+1)$, имеют вид

$$F_{i,j}^{n+1} = [F_{i,j}^p + F_{i,j}^c] / 2. \quad (12)$$

Обратим внимание, что величины $x_{\max}$ и $v_{\max}$ следует выбирать достаточно большими, чтобы в процессе численного решения задачи Коши начальные возмущения не успели распространиться до границы расчетной области. Это дает возможность использовать краевые условия первого рода для функции $F(x, v, t)$ и, в свою очередь, для дискретных вариаций этой функции, т.е. величин $\delta F_{i,j}^p$ и $\delta F_{i,j}^c$, — однородные краевые условия.

Для численного решения уравнения Ампера воспользуемся схемой второго порядка точности по времени аналогично [8]:

$$\frac{E_i^{n+1} - E_i^{n-1}}{2\tau} = \sum_{j=1}^{M_v} v_j F_{i,j}^n h_v \quad \left(\approx \int_{-\infty}^{\infty} v F dv \right), \\ E_i^{n+1/2} = [E_i^n + E_i^{n+1}] / 2. \quad (13)$$

Заметим, что начальное условие (6) дает $E_t(x, t = 0) = 0$, что приводит к справедливости $E(x, t = \tau) = E_0(x) + O(\tau^2)$. Поэтому для инициализации вычислений без потери точности можно положить $E_i^1 = E_i^0$, $0 \leq i \leq M_x$.

Для вычисления приближенного значения электронной плотности $N(x, t)$ используется простейшая составная формула прямоугольников (см. [25]):

$$N_i^n = \sum_{j=1}^{M_v} F_{i,j}^n h_v.$$

При этом, следуя [8], суммирование по индексу j (аналог интегрирования по переменной v) всюду в схеме проводится “в парах”.

Сделаем несколько замечаний о предложенной схеме. В формулах (10), (11) используются известные или предварительно вычисленные коэффициенты переноса v_j и $E_i^{n+1/2}$, что на каждом из этапов (предиктор или корректор) приводит к необходимости решать фактически линейное уравнение Власова. По этой причине в предыдущем разделе приведены аргументы в пользу безусловной устойчивости неявной схемы Мак-Кормака для модельного линейного уравнения с постоянным коэффициентом переноса. При этом, с точки зрения применения теоретических конструкций типа “замороженных коэффициентов” (см. [26]), разница между переменными и постоянными коэффициентами не играет принципиального значения.

Отметим, что при использовании в работе [8] явной схемы Мак-Кормака традиционное ограничение на устойчивость (типа двумерного условия Куранта) возникало именно на этапе решения такого же линейного уравнения Власова, что соответствует частному случаю $\lambda = 0$ в (10), (11). Иными словами, ослабление жесткого ограничения происходит за счет обобщения исходного алгоритма (с явной схемы до неявной) только в части решения уравнения Власова. Уравнение Ампера решается таким же способом, как в [8].

Для предлагаемого алгоритма совместного решения уравнений Власова—Ампера в настоящее время отсутствует теоретическое обоснование безусловной устойчивости, поэтому пока можно говорить только о более слабом ограничении на устойчивость для решаемой нелинейной задачи. Следует отметить, что при численном решении уравнений, описывающих динамику вязкого сжимаемого газа, в исходной работе [15] также имелась в виду не безусловная устойчивость неявной схемы для нелинейной задачи, а только ослабление ограничений по сравнению с явной схемой.

Важный аспект построенной схемы — отсутствие вспомогательных итераций при решении нелинейной задачи. В первую очередь, это связано со вторым порядком точности метода, который часто соответствует схемам типа предиктор—корректор. Хорошо известен подобный пример для численного решения ОДУ, когда решение обладает достаточной гладкостью: вместо дополнительных итераций можно использовать формулы явного пересчета (см. [26], с. 368).

4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Для однозначного задания начальных условий и сохранения преемственности с предыдущими расчетами (см., например, [27], [28]) зафиксируем параметры в (6): $\rho_* = 0.5$, $\alpha = 0.3$, $T_0 = 1.6 \times 10^{-3}$ и заметим, что на больших расстояниях от прямой $x = 0$ функция распределения электронов не возмущается:

$$F(x \rightarrow \pm\infty, v, t) = f_M(v, T_0).$$

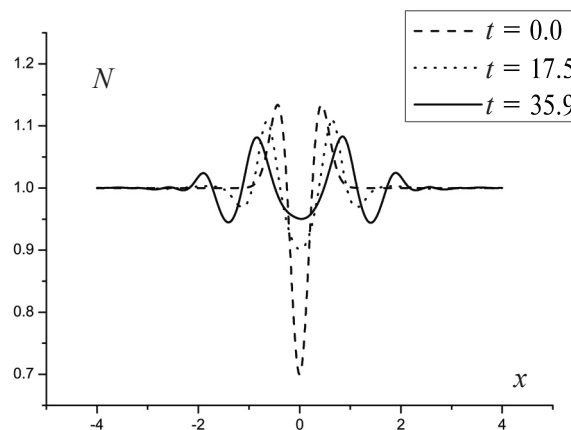
Однако в целях численного моделирования расчетную область необходимо ограничить. При гидродинамическом моделировании ленгмюровских колебаний холодной плазмы, независимо от времени наблюдения за ними, вполне достаточно положить $x_{\max} = 4.5\rho_*$ (см. [27]); здесь же (в кинетике) влияние температуры будет трансформировать колебания в волны (см. [28], [29]), пространственная динамика которых существенно зависит от времени наблюдения за ними. Иными словами, величина $x_{\max}$ должна быть согласована со значением $t_{\max}$, которое, в свою очередь, зависит от скорости распространения волны и неявно определяется температурой T_0 плазмы. В данном случае плазма предполагается “теплой”, т.е. приводящей к небольшой скорости распространения волны (в линейном

случае $\approx \sqrt{3T_0}$), однако время наблюдения может быть значительным. Отметим, что исходные уравнения обезразмерены так, что плазменная частота ω_p равна единице, поэтому характерное значение $t_{\max}$ составляет 5–6 периодов, т.е. порядка 40 безразмерных единиц времени. В силу вышесказанного, величина $x_{\max}$ бралась равной 10, т.е. $x_{\max} = 20\rho_*$. В свою очередь, ограничение расчетной области по переменной v задавалось как $v_{\max} = 0.34$, что при выбранном значении T_0 позволяет использовать однородные условия

$$F(x, v \rightarrow \pm\infty, t) \approx 0.$$

Обратим внимание на вибрирующий процесс распространения волн при выбранных значениях параметров. Начальное пространственное распределение электронной плотности $N_0(x)$ порождает избыток положительного заряда в окрестности начала координат (т.е. при $x = 0$). По этой причине электроны начинают движение в направлении центра области, что примерно через полпериода порождает распределение плотности с глобальным максимумом также при $x = 0$. В свою очередь, сформировавшийся избыток электронов в центре области приводит к их движению в обратном направлении. Если бы плазма была холодной ($T_0 = 0$), то после очередной половины периода пространственное распределение электронной плотности возвратилось бы к начальному (как при $t = 0$), и в результате наблюдалось бы 2π -периодическое по времени решение в соответствии с аналитическими результатами [30] (подробные доказательства приведены в [31]).

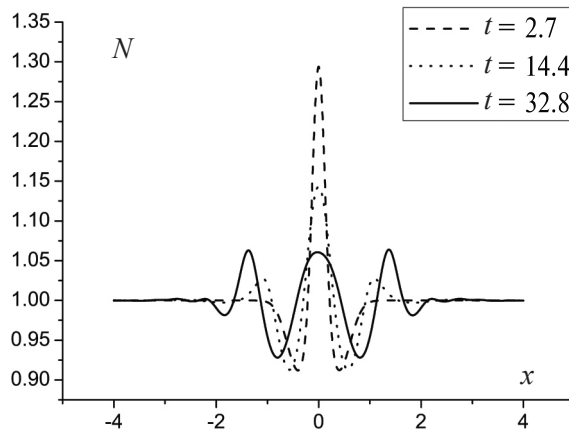
Однако кинетическая модель учитывает влияние температуры плазмы (даже весьма небольшой!), поэтому при движении частиц от прямой $x = 0$ наблюдается типичный волновой эффект, т.е. перемещение энергии. При этом, как и ранее, формируется избыток положительного заряда в центре области, приводящий к соответствующему движению электронов (подходящий образ — “прилив”). Вслед за “приливом” опять следует полпериода “отлива” волны, т.е. переноса энергии и т.д. Хорошей иллюстрацией к описанной динамике волны являются фиг. 1 и 2. На них в специально выбранные моменты времени наблюдаются пространственные распределения плотности электронов N , обеспечивающие вибрирующее поведение волны. Например, на фиг. 1 приведены распределения плотности, аналогичные начальному, т.е. с минимумами при $x = 0$. Интервалы между выбранными моментами времени составляют порядка трех периодов, что делает наглядным перенос энергии в моменты “отливов”. В свою очередь, на фиг. 2 приведены распределения плотности с максимумами при $x = 0$. Интервалы между выбранными моментами времени также составляют порядка трех периодов, что делает наглядным изменение формы волны в моменты “приливов”.



Фиг. 1. Пространственные распределения функции плотности N в моменты времени $t = 0, 17.5, 35.9$ при $\alpha = 0.3$.

Отметим также, что несмотря на гладкость представленных пространственных распределений, значения сеточных параметров невелики:

$$h_x = 2.5 \times 10^{-3}, \quad h_v = 8.5 \times 10^{-4}, \quad \tau = 5 \times 10^{-2},$$



Фиг. 2. Пространственные распределения функции плотности N в моменты времени $t = 2.7, 14.4, 32.8$ при $\alpha = 0.3$.

что было обусловлено необходимостью качественной аппроксимации экспоненциально убывающих начальных функций $E_0(x)$ и $F(x, v, t = 0)$. Если использовать явную схему (при $\lambda = 0$), то необходимое условие устойчивости (см. [8]) для взятых параметров

$$\tau \leq \frac{1}{v_{\max}/h_x + E_{\max}/h_v}$$

приводит к ограничению $\tau \leq 5 \times 10^{-3}$. При этом попытка увеличить в 2 раза шаг по времени в явной схеме сразу приводит к разрушению решения из-за неустойчивости по отношению к ошибкам округлений. Приведенные же иллюстрации получены с помощью неявной схемы (при $\lambda = 1$), когда значение шага по времени τ в 10 раз превышало ограничение для явной схемы, поэтому важно учитывать законы сохранения.

На фиг. 3 приведена динамика относительной погрешности полной энергии, вычисляемой по формуле

$$\text{Total}_{\text{en}}(t) = \frac{1}{2} \left(\int_{-v_{\max}}^{v_{\max}} \int_{-x_{\max}}^{x_{\max}} v^2 F dv dx + \int_{-x_{\max}}^{x_{\max}} E^2 dx \right).$$

Приближенное интегрирование по v проводилось, как описано выше, а приближенное интегрирование по x было реализовано с помощью составной формулы трапеций (см. [25]). При уменьшении всех сеточных параметров в 2 раза относительная погрешность полной энергии в полном соответствии с теоретическими свойствами схемы Мак-Кормака уменьшалась в 4 раза. В свою очередь, полный заряд плазмы (количество частиц)

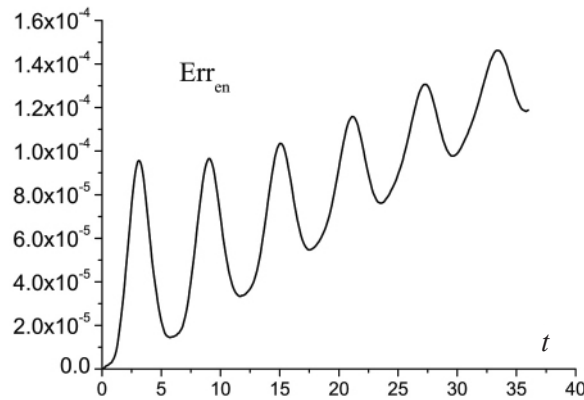
$$\text{Total}_{\text{part}}(t) = \int_{-v_{\max}}^{v_{\max}} \int_{-x_{\max}}^{x_{\max}} F dv dx$$

сохраняется на сеточном уровне (точно!), поэтому соответствующие графики не приводятся.

Следует отметить, что при расчетах уравнения Власова традиционной иллюстрацией является динамика максимального по области значения электрического поля, т.е. его зависимость от времени:

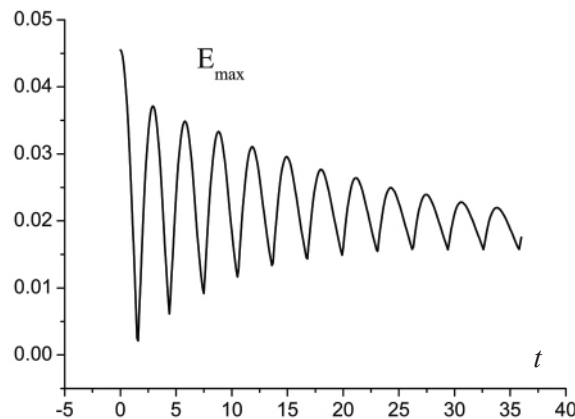
$$E_{\max}(t) = \max_{|x| \leq x_{\max}} E(x, t).$$

На фиг. 4 представлены типичные осцилляции этой величины. Так как температура плазмы достаточно мала, то период колебаний близок к 2π , поэтому, как в холодной плазме (см. [30]), у графика



Фиг. 3. Динамика относительной погрешности полной энергии при $\alpha = 0.3$.

наблюдается примерно два максимума на период. С другой стороны, даже малая температура (в расчетах — $T_0 = 1.6 \times 10^{-3}$) приводит к монотонному убыванию амплитуды электрического поля, что и наблюдается на протяжении примерно шести периодов. Затухание амплитуды связано с распротранением волны при сохранении полной энергии и хорошо согласуется с графиками на фиг. 1–3.

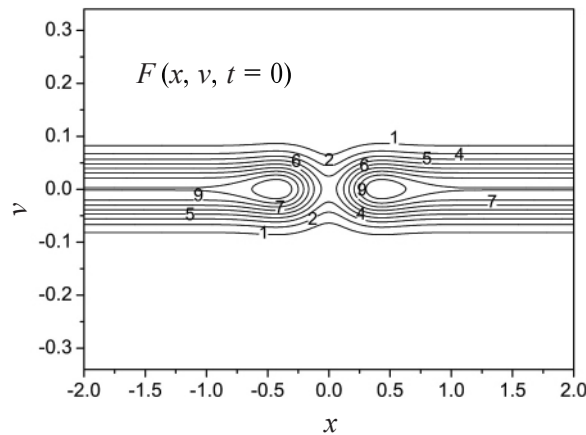


Фиг. 4. Динамика максимального значения электрического поля при $\alpha = 0.3$.

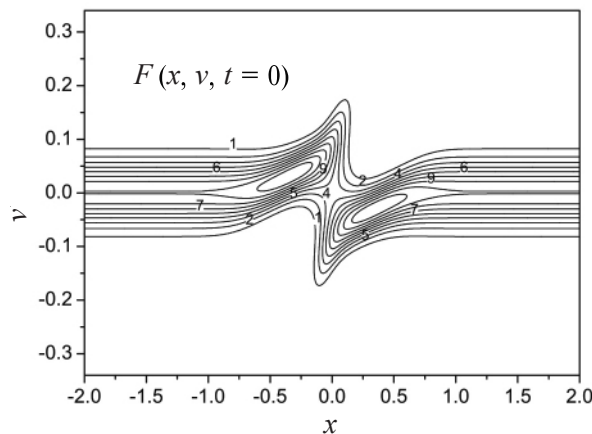
Приведем дополнительно результаты вычислительного эксперимента при $\alpha = 0.55$, т.е. пример волны, возбуждаемой более мощным лазерным импульсом. Остальные параметры брались без изменений. Цель расчетов — иллюстрация быстрой потери гладкости функции распределения F , следовательно, необходимость измельчения сеточных параметров, что является весомым аргументом в пользу применения неявной схемы.

Рассмотрим сначала фиг. 5, где изображены линии уровня функции распределения в начальный момент времени. Возмущение электронной плотности $N_0(x)$ заметно только в небольшой окрестности начала координат (при $|x| \leq 1$), вне которой наглядно максвелловское распределение по скоростям. Напомним, что фиг. 5 соответствует состоянию “отлива” для рассматриваемой волны. Далее на фиг. 6 приведены линии уровня функции распределения в момент времени $t = 2.5$, соответствующий первому “приливу”, т.е. когда при $x = 0$ наблюдается максимум электронной плотности. Наблюдаемые искривления и сгущения линий уровня иллюстрируют большие градиенты решения в окрестности прямой $x = 0$, что, в свою очередь, порождает сильное увеличение плотности электронов — избыток отрицательного заряда. Наконец, на фиг. 7 приведены линии уровня функции распределения

в момент времени $t = 5.5$, соответствующий следующему “приливу”, т.е. когда при $x = 0$ опять наблюдается максимум электронной плотности. Легко видеть, что функция распределения имеет гладкость значительно худшую, чем на фиг. 5 и фиг. 6. При этом наблюдаемая картина — свойство решения, а не недостаток сеточного разрешения, так как иллюстрация полностью идентична полученной при двукратном измельчении всех параметров сетки. Это означает, что дальнейшее продолжение моделирования волны при взятых параметрах практически невозможно: требуется значительное сгущение сетки в подобласти $|x| \leq 1$. Иными словами, вычислительные возможности предложенной неявной схемы здесь уже исчерпаны, т.е. для дальнейшего моделирования требуется применение суперкомпьютеров, параллельных алгоритмов и т.п.

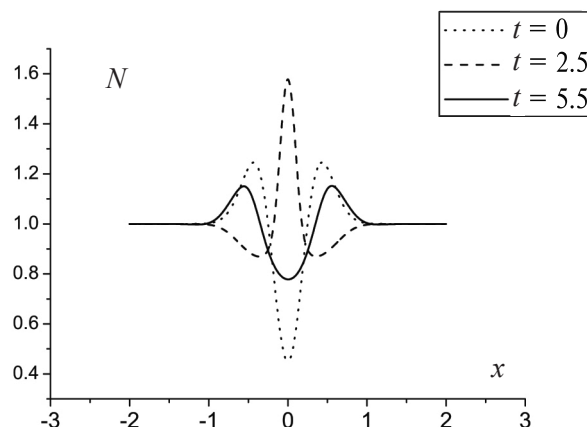


Фиг. 5. Линии уровня функции распределения электронов F в момент времени $t = 0$ при $\alpha = 0.55$.



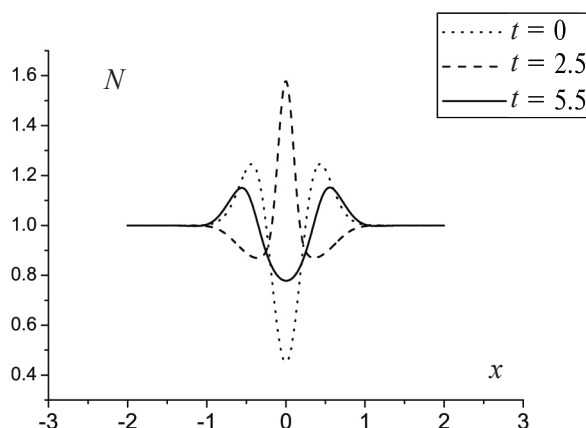
Фиг. 6. Линии уровня функции распределения электронов F в момент времени $t = 2.5$ при $\alpha = 0.55$.

Однако представляется уместным замечание о том, что вариант с $\alpha = 0.55$, как и ранее рассмотренный с $\alpha = 0.3$, полностью удовлетворяет приведенной в разд. 1 теореме существования. Иными словами, вычислительные проблемы — это следствие больших градиентов решения непрерывной задачи. Отметим также, что когда вместо кинетической модели рассматривается гидродинамическая модель холодной плазмы, значение $\alpha = 0.55$ недоступно в принципе: в гидродинамике при $\alpha = 0.55$ плотность электронов становится сингулярной уже при $t_{\text{бр}} < \pi$ (см. [30], [31]). В свою очередь, пространственные распределения электронной плотности в указанные выше моменты времени $t = 0, 2.5, 5.5$, которые приведены на фиг. 8, не вызывают никаких замечаний о гладкости. Это означает, что наблю-



Фиг. 7. Линии уровня функции распределения электронов F в момент времени $t = 5.5$ при $\alpha = 0.55$.

дение только за моментами функции распределения F не всегда позволяет контролировать вычислительный процесс полностью.



Фиг. 8. Пространственные распределения функции плотности N в моменты времени $t = 0, 2.5, 5.5$ при $\alpha = 0.55$.

ЗАКЛЮЧЕНИЕ

В работе для кинетической модели плазмы, базирующейся на уравнениях Власова—Ампера, построена неявная схема типа Мак-Кормака. По сравнению с явной схемой она имеет более слабое ограничение на устойчивость, но сохраняет прежнюю вычислительную эффективность, т.е. не использует внутренние итерации. При этом погрешность полной энергии соответствует второму порядку точности алгоритма, а полный заряд (число частиц) сохраняется на сеточном уровне.

В качестве моделируемого физического процесса рассматривается динамика плазменных (ленгмюровских) колебаний и волн, возбуждаемых коротким мощным лазерным импульсом. Показано, что перенос энергии носит вибрирующий характер типа “приливов” и “отливов”. Также приведен пример, не имеющий аналога в гидродинамике холодной плазмы и иллюстрирующий потерю гладкости функции распределения при сохранении гладкости функции плотности электронов, т.е. момента нулевого порядка.

Как правило, уравнение Власова используется для моделирования свойств уже сформировавшейся волны, имеющих пространственно-периодический характер. Поэтому определенный интерес представляет процесс именно формирования бегущих волн, распространяющихся из области первоначально возмущенной плотности по причине теплового (хаотического) движения электронов.

Результаты работы могут быть обобщены на постановки кинетических задач большей размерности и учет дополнительных физических факторов в модели плазмы.

СПИСОК ЛИТЕРАТУРЫ

1. *Chen F. F.* Introduction to plasma physics and controlled fusion. 3rd ed. New York: Springer, 2016. P. 355–411.
2. *Власов А. А.* О вибрационных свойствах электронного газа // Успехи физ. наук. 1967. Т. 93. № 3. 444 с.
3. *Григорьев Ю. Н., Вишивков В. А., Федорук М. П.* Численное моделирование методами частиц-в-ячейках. Новосибирск: Изд-во СО РАН, 2004. С. 190–255.
4. *Arber T., Vann R.* A critical comparison of Eulerian-grid-based Vlasov solvers // J. Comput. Phys. 2002. V. 180. P. 339.
5. *Filbet F., Sonnendrucker E.* Comparison of Eulerian Vlasov solvers // Comput. Phys. Commun. 2003. V. 150. P. 247.
6. *Dimarco G., Pareschi L.* Numerical methods for kinetic equations // Acta Numerica. 2014. V. 23. P. 369.
7. *G.Chen G., Chacon L., Barnes D.* An energy- and charge-conserving, implicit, electrostatic particle-in-cell algorithm // J. Comput. Phys. 2011. V. 230. № 18. P. 7018.
8. *Horne R. B., Freeman M. P.* A new code for electrostatic simulation by numerical integration of the Vlasov and Ampere equations using MacCormack's method // J. Comput. Phys. 2001. V. 171. P. 182.
9. *Crouseilles N., Respaud T.* A charge preserving scheme for the numerical resolution of the Vlasov–Ampere equations // Commun. Comput. Phys. 2011. V. 10. P. 1001.
10. *Elkina N., Buchner J.* A new conservative unsplit method for the solution of the Vlasov equation // J. Comput. Phys. 2006. V. 213. N 2. P. 862.
11. *Cheng Y., Christlieb A. J., Zhong X.* Energy-conserving discontinuous Galerkin methods for the Vlasov–Ampere system // J. Comput. Phys. 2014. V. 256. P. 630.
12. *Anderson S. E., Taitano W. T., Chacon L., Simakov A. N.* An efficient, conservative, time-implicit solver for the fully kinetic arbitrary-species 1D-2V Vlasov-Ampere system // J. Comput. Phys. 2020. V. 419. P. 109686.
13. *Liu H., Cai X., Cao Y., Lapenta G.* An efficient energy conserving semi-Lagrangian kinetic scheme for the Vlasov-Ampere system // J. Comput. Phys. 2023. V. 492. N 1. P. 112412.
14. *Cohen B., Langdon A., Hewett D., Proccassini R.* Performance and optimization of direct implicit particle simulation // J. Comput. Phys. 1989. V. 81. N 1. P. 151.
15. *MacCormack R. W.* A numerical method for solving the equations of compressible viscous flow // AIAA J. 1982. V. 20. N 9. P. 1275.
16. *Горбунов Л. М.* Зачем нужны сверхмощные лазерные импульсы? // Природа. 2007. Т. 21. № 4. С. 11.
17. *Гельфанд И. М., Зуева Н. М., Имшенник В. С., Локуцевский О. В., Рябенский В. С., Хазин Л. Г.* К теории нелинейных колебаний электронной плазмы // Ж. вычисл. матем. и матем. физ. 1967. Т. 7. № 2. С. 322.
18. *Mora P. A., Antonsen T. M.* Kinetic modeling of intense, short pulses propagating in tenuous plasmas // Phys. Plasm. 1997. V. 4. N 4. P. 217.

19. Андреев Н. Е., Горбунов Л. М., Рамазашвили Р. Р. К теории трехмерной кильватерной волны, возбуждаемой мощным лазерным импульсом в разреженной плазме // *Физика плазмы*. 1997. Т. 23. № 4. С. 303.
20. Sheppard C. J. R. Cylindrical lenses — focusing and imaging: a review [Invited] // *Appl. Optic*. 2013. V. 52. № 4. P. 538.
21. Иорданский С. В. О задаче Коши для кинетического уравнения плазмы // *Тр. МИАН*. 1961. Т. 60. С. 81.
22. MacCormack R. W. The effect of viscosity in hypervelocity impact cratering // *J. Spacecr. Rocket*. 2003. V. 40. № 5. P. 757.
23. Шокин Ю. И., Яненко Н. Н. Метод дифференциального приближения. Применение к газовой динамике. Новосибирск: Наука, 1985. С. 17–30.
24. Furst J., Furmanek P. An implicit MacCormack scheme for unsteady flow calculations // *Comput. Fluid*. 2011. V. 46. P. 231.
25. Бахвалов Н. С., Корнев А. А., Чижонков Е. В. Численные методы. Решения задач и упражнения. Серия Классический университетский учебник. Учеб. пособие для вузов. 2-е изд., испр. и дополн. М.: Лаборатория знаний, 2016. С. 94–101.
26. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. Серия Классический университетский учебник. Учеб. пособие для вузов. 7-е изд. М.: БИНОМ. Лаборатория знаний, 2013. С. 524–527.
27. Чижонков Е. В. Математические аспекты моделирования колебаний и кильватерных волн в плазме. М.: Физматлит, 2018. С. 12–240.
28. Chizhonkov E. V., Frolov A. A. Effect of electron temperature on formation of travelling waves in plasma: Kinetic and hydrodynamic models // *Russ. J. Numer. Anal. Math. Model*. 2023. V. 38. № 2. P. 63.
29. Фролов А. А., Чижонков Е. В. Численное моделирование плазменных колебаний с учетом теплового движения электронов // *Вычисл. методы и программ.* 2018. Т. 19. С. 194.
30. Розанова О. С., Чижонков Е. В. О существовании глобального решения одной гиперболической задачи // *Докл. РАН. Математика, информатика, процессы управления*. 2020. Т. 492. № 1. С. 97.
31. Rozanova O. S., Chizhonkov E. V. On the conditions for the breaking of oscillations in a cold plasma // *Z. Angew. Math. Phys*. 2021. V. 72. № 13. P. 1.

ON THE NUMERICAL SOLUTION OF THE VLASOV-AMPERE EQUATIONS

E. V. Chijonkov^{a,*}

^a 119899 Moscow, Lenin Hills, Lomonosov Moscow State University, Russia

*e-mail: chizhonk@mech.math.msu.su

Received 2 November, 2023

Revised 10 March, 2024

Accepted 3 April, 2024

Abstract. An implicit McCormack-type scheme is constructed for the kinetic plasma model based on the Vlasov-Ampere equations. Compared to the explicit scheme, it has a weaker stability constraint, but retains the same computational efficiency, i.e. it does not use internal iterations. In this case, the error of the total energy corresponds to the second order of accuracy of the algorithm, and the total charge (number of particles) is stored at the grid level. The formation of plasma waves excited by a short powerful laser pulse is considered as a simulated physical process.

Keywords: kinetic plasma model, Vlasov-Ampere equations, plasma oscillations and waves, numerical modeling, implicit McCormack scheme.

УДК 519.635

KP_1 -СХЕМА УСКОРЕНИЯ ВНЕШНИХ ИТЕРАЦИЙ ПО ОБЛАСТИ ТЕРМАЛИЗАЦИИ НЕЙТРОНОВ И ПО ИСТОЧНИКУ ДЕЛЕНИЯ ПРИ РЕШЕНИИ ПОДКРИТИЧЕСКОЙ КРАЕВОЙ ЗАДАЧИ

© 2024 г. А. М. Волощенко^{1,*}¹125047 Москва, Миусская пл., 4, Институт прикладной математики РАН, Россия

*e-mail: volosch@kiam.ru

Поступила в редакцию 07.02.2024 г.

Переработанный вариант 22.02.2024 г.

Принята к публикации 03.04.2024 г.

Для уравнения переноса в трехмерной r, ϑ, z геометрии построена KP_1 -схема ускорения сходимости внешних итераций по области термализации нейтронов и по источнику деления при решении подкритической краевой задачи, согласованная с Weighted Diamond Differencing (WDD) схемой, а также рассмотрено ее обобщение на случай нодальных Linear Discontinues (LD) и Linear Best (LB)-схем 3-го и 4-го порядка точности по пространственным переменным. Для решения P_1 -системы для ускоряющих поправок использован алгоритм, основанный на использовании циклического метода расщепления, аналогичный использованному ранее при построении KP_1 -схемы ускорения сходимости внутренних итераций. Рассмотрен алгоритм определения энергетической зависимости для ускоряющих поправок KP_1 -схемы ускорения сходимости внешних итераций. Рассмотрен вопрос выбора критерия сходимости внешних итераций, предложен интегральный по рассеянным вверх по энергии тепловым нейтронам критерий сходимости внешних итераций по области термализации нейтронов. Рассмотрена модификация алгоритма на случай трехмерной x, y, z геометрии. Приведены численные примеры использования KP_1 -схемы ускорения сходимости внешних итераций для решения характерных задач переноса нейтронов в трехмерной геометрии. Библ. 39. Фиг. 2. Табл. 7.

Ключевые слова: KP_1 -схема ускорения внешних итераций, выбор спектральной функции и критерия сходимости внешних итераций, уравнение переноса.

DOI: 10.31857/S0044466924070127, EDN: xibspv

ВВЕДЕНИЕ

Ускорение внешних итераций, которые возникают, при наличии в матрице межгрупповых переходов вверх по группам, является существенным элементом численной методики решения многогруппового уравнения переноса нейтронов. Существуют достаточно важные и представительные классы задач, при решении которых для уменьшения вычислительных затрат необходимо уменьшить число внешних итераций по области термализации нейтронов (thermal up-scattering) и (или) по источнику деления (fission up-scattering).

Число внешних итераций по области термализации при наличии, например, в защитной композиции воды, разогретой до температуры порядка 600 K⁰ (реально реализуемая ситуация для водородных реакторов), может быть достаточно велико (порядка 70 при точности сходимости внешних итераций $\varepsilon_{upsc} = 10^{-3}$ и использовании пяти групп нейтронов с термализацией (для которых присутствуют переходы из нижерасположенных групп) константных систем¹ BUGLE96 [1] или BGL1000/440 [2]). При увеличении числа групп нейтронов в тепловой области, что имеет место при использовании мало- и мультигрупповых систем констант (V7-27N19G, V7-200N47G из SCALE-6.1 [3]; v7.1-200n47g, v7.0-238n, v7.1-252n из SCALE-6.2.4 [4]; мультигрупповые версии БНАБ-93 [5] и БНАБ-РФ [6], имеющие соответственно 13, 34, 36, 90, 97 и 73 групп с термализацией),

¹Под константной системой понимается библиотека многогрупповых сечений для нейтронов и фотонов для заданного набора нуклидов, полученная из файлов оцененных данных путем усреднения детального хода сечений на заданную энергетическую сетку.

число внешних итераций по области термализации также возрастает и, в зависимости от расчетной композиции, может достигать нескольких тысяч, что делает проблематичным проведение соответствующих расчетов без ускорения внешних итераций.

Другая область приложений, в которых алгоритм ускорения внешних итераций также очень важен — это ускорение внешних итераций по источнику деления при решении подкритических задач с заданным внешним источником. ADS (accelerator driven systems) системы, в которых поле нейтронов в подкритической системе поддерживается источником spallation нейтронов, т.е. вторичных нейтронов, возникающих при взаимодействии с мишенью пучка протонов высокой энергии, являются практически важным примером таких систем. Как показывают теоретические оценки и практические расчеты, число внешних итераций по источнику деления при расчете таких систем может быть достаточно велико (порядка 70 при $k_{eff} = 0.94$), и их число неограниченно возрастает при $k_{eff} \rightarrow 1$. Здесь и ниже k_{eff} — эффективный коэффициент размножения системы с делящимися нуклидами (см. [7], [8]). Таким образом, расчет систем, близких к критическим, в отсутствие эффективного алгоритма ускорения внешних итераций по источнику деления, представляет серьезную вычислительную проблему.

В настоящей работе мы рассмотрим обобщение KP_1 -схемы (см. [7], [9–17]) применительно к задаче ускорения внешних итераций по области термализации нейтронов и по источнику деления при решении подкритической задачи. Эффективность и устойчивость данного алгоритма существенно зависят от выбора формы спектра для ускоряющих поправок. В [18–20] эффективность данного алгоритма была исследована на примере 1D геометрий. В [16], развитием которой является настоящая работа, были представлены дополнительные численные результаты по использованию алгоритма при решении практических задач в 1D, 2D и 3D геометриях в зависимости от выбора формы спектра для ускоряющих поправок и сформулирован вариант алгоритма ускорения внешних итераций, который, как показал опыт расчетов, обеспечивает необходимую эффективность при решении широкого класса практических задач. Данный алгоритм реализован в 1D, 2D и 3D S_n программах РОЗ-6, КАСКАД-С и КАТРИН, включенных в систему программ CNCSN-2009 (см. [21]), и оптимизирован в последних версиях этих программ. В [22] предложен Transport Two Grid (TTG) алгоритм ускорения внешних итераций по области термализации нейтронов, являющийся обобщением Two Grids (TG) алгоритма, предложенного ранее в [18], а также приведены примеры его использования в x, y, z геометрии.

Последовательность изложения в настоящей работе следующая. В разд. 1–4 для случая r, ϑ, z геометрии будет построена KP_1 -схема для ускорения внешних итераций по области термализации и по источнику деления при решении подкритической задачи, согласованная с WDD схемой, а также рассмотрено обобщение алгоритма на случай x, y, z геометрии и нодальных схем². Разделы 5 и 6 посвящены описанию алгоритма определения вида спектра для корректирующих поправок KP_1 -схемы для гомогенной и гетерогенной сред. В разд. 7 приведены численные результаты по использованию KP_1 -схемы ускорения внешних итераций. Обсуждению полученных результатов посвящен разд. 8.

1. WDD СХЕМА ДЛЯ УРАВНЕНИЯ ПЕРЕНОСА В ТРЕХМЕРНОЙ R, Θ, Z ГЕОМЕТРИИ

В r, ϑ, z геометрии уравнение переноса имеет вид (см. [8])

$$\eta \frac{\partial \psi}{\partial \vartheta} + \mu r \frac{\partial \psi}{\partial z} + \xi \frac{\partial}{\partial r}(r\psi) - \frac{\partial}{\partial \varphi}(\mu\psi) + \sigma r\psi(r, \vartheta, z; \mu, \varphi) = rS(r, \vartheta, z; \mu, \varphi), \quad (1)$$

где ψ , σ и S — поток нейтронов (фотонов), полное сечение и источник в группе q (в разд. 1 и 2 индекс номера группы опускается); ξ , η и μ — направляющие косинусы единичного вектора Ω направления скорости частицы:

$$\xi = (\Omega \mathbf{n}_r) = \sqrt{1 - \mu^2} \cos \phi, \quad \eta = (\Omega \mathbf{n}_\vartheta) = \sqrt{1 - \mu^2} \sin \phi, \quad \mu = (\Omega \mathbf{n}_z) = \cos \theta, \quad (2)$$

который изменяется в пределах единичной сферы (восьми октантов): $-1 \leq \xi, \eta \leq 1$, φ — азимутальный угол: $0 \leq \varphi \leq 2\pi$; переменные r, ϑ и z изменяются в пределах: $0 \leq r_{\text{int}} \leq r \leq r_{\text{ext}}$, $0 \leq \vartheta_0 \leq \vartheta \leq \vartheta_{\text{end}} \leq 2\pi$, $0 \leq z_{\text{bot}} \leq z \leq z_{\text{top}} = H$.

²В нодальных схемах для повышения порядка аппроксимации разностной схемы используются дополнительные пространственные моменты решения в разностной ячейке, а также дополнительные балансные уравнения.

Для граничных направлений $\varphi = \pi$ и $\varphi = 2\pi$, используемых для организации расчета ячеек в r , ϑ , z геометрии, уравнение (1) имеет вид

$$\mu r \frac{\partial \Psi}{\partial z} + \xi \left[\frac{\partial}{\partial r} (r\Psi) - \Psi \right] + \sigma r \Psi(r, z, \mu, \varphi) = rS(r, z, \mu, \varphi). \quad (3)$$

К уравнению (1) следует также добавить граничные условия при $r = r_{\text{int}} \geq 0$, $r = r_{\text{ext}}$, $\vartheta = \vartheta_0$, $\vartheta = \vartheta_{\text{end}}$, $z = z_{\text{bot}}$, $z = z_{\text{top}}$.

Для аппроксимации интеграла рассеяния введем квадратуру на единичной сфере угловых направлений (8 октантов): $-1 \leq \xi, \eta, \mu \leq 1$, $0 \leq \theta \leq \pi$, $0 \leq \varphi \leq 2\pi$, следующего вида:

$$\begin{aligned} \int_{-1}^1 d\mu \int_0^{2\pi} d\varphi f(\mu, \varphi) &= \int_{-1}^1 d\mu \int_0^{\pi} d\varphi f(\mu, \varphi) + \int_{-1}^1 d\mu \int_{\pi}^{2\pi} d\varphi f(\mu, \varphi) = \\ &= \sum_{l=1}^{L_1} \sum_{m=1}^{M_1^{(1)}} w_{l,m}^{(1)} f(\mu_l^{(1)}, \varphi_{l,m}^{(1)}) + \sum_{l=1}^{L_2} \sum_{m=1}^{M_1^{(2)}} w_{l,m}^{(2)} f(\mu_l^{(2)}, \varphi_{l,m}^{(2)}). \end{aligned} \quad (4)$$

Таким образом, для передней $0 < \varphi < \pi$ ($\mu_{l,m} > 0$) и задней $\pi < \varphi < 2\pi$ ($\mu_{l,m} < 0$) полусфер, вообще говоря, может использоваться различная квадратура. Квадратура, например, для передней полусферы (I, II, V и VI октанты) разбивает ее на L_1 полос по переменной μ веса $w_l^{(1)}$, а каждая полоса l , $l = 1, \dots, L_1$, дополнительно еще разбивается на $M_l^{(1)}$ секторов веса $w_{l,m}^{(1)}$, где индекс m на l -м слое пробегает значения $m = 1, \dots, M_l^{(1)}$. Нумерация узлов по φ на l -м слое соответствует возрастанию $\xi_{l,m}^{(1)}$ (убыванию $\varphi_{l,m}^{(1)}$ при $0 \leq \varphi \leq \pi$). Аналогично устроена квадратура и для задней полусферы (III, IV, VII и VIII октанты). Нумерация узлов по φ на l -м слое для нее соответствует возрастанию $\xi_{l,m}^{(2)}$ (и $\varphi_{l,m}^{(2)}$) при $\pi \leq \varphi \leq 2\pi$. Для обычно используемых симметричных квадратур квадратура для задней полусферы получается из квадратуры для передней полусферы путем зеркального отражения относительно плоскости $\varphi = \pi$. Ниже мы будем, как правило, опускать индекс полусферы.

Уравнение баланса нулевого порядка получается путем интегрирования уравнения (1) по разностной ячейке $(r_{i-1/2}, r_{i+1/2}) \cdot (\vartheta_{j-1/2}, \vartheta_{j+1/2}) \cdot (z_{k-1/2}, z_{k+1/2}) \cdot (\varphi_{l,m+1/2}, \varphi_{l,m-1/2}) \cdot (\mu_{l-1/2}, \mu_{l+1/2})$, $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$ (ниже мы опускаем целые индексы, в том числе, номер группы, в тех случаях, когда это не вызывает недоразумений):

$$\begin{aligned} &|\eta| \Delta r v_z (\Psi_P - \Psi_F) + |\mu| v_r v_\vartheta (\Psi_T - \Psi_B) + v_z v_\vartheta \times \\ &\times \left[|\xi| (A^+ \Psi_R - A^- \Psi_L) + \frac{C}{w} (\alpha_{m+1/2} \Psi_{m+1/2} - \alpha_{m-1/2} \Psi_{m-1/2}) \right] + \sigma V \Psi = VS, \end{aligned} \quad (5)$$

где

$$\begin{aligned} \alpha_{l,m+1/2} - \alpha_{l,m-1/2} &= w_{l,m} \xi_{l,m}, \quad \alpha_{l,1/2} = \alpha_{l,M_{l+1/2}} = 0, \quad m = 1, \dots, M_l, \quad l = 1, \dots, L, \\ A_{i\pm 1/2} &= r_{i\pm 1/2}, \quad v_{r,i} = \frac{1}{2} (r_{i+1/2}^2 - r_{i-1/2}^2), \quad C_i = r_{i+1/2} - r_{i-1/2} = \Delta r_i, \\ \Delta z_k &= v_{z,k} = z_{k+1/2} - z_{k-1/2}, \\ \Delta \vartheta_j &= v_{\vartheta,j} = \vartheta_{j+1/2} - \vartheta_{j-1/2}, \quad V_{i,j,k} = \Delta z_k v_{r,i} v_{\vartheta,j}, \\ \Delta \vartheta_j &= v_{\vartheta,j} = \vartheta_{j+1/2} - \vartheta_{j-1/2}, \quad V_{i,j,k} = \Delta z_k v_{r,i} v_{\vartheta,j}, \\ \Psi_{R(L)} \equiv \Psi_r^\pm &= \begin{cases} \Psi_{i\pm 1/2,j,k,l,m}, & \xi_{l,m} > 0, \\ \Psi_{i\mp 1/2,j,k,l,m}, & \xi_{l,m} < 0, \end{cases} \quad \Psi_{T(B)} \equiv \Psi_r^\pm = \begin{cases} \Psi_{i,j,k\pm 1/2,l,m}, & \mu_l > 0, \\ \Psi_{i,j,k\mp 1/2,l,m}, & \mu_l < 0, \end{cases} \\ A^\pm &= \begin{cases} A_{i\pm 1/2}, & \xi_{l,m} > 0, \\ A_{i\mp 1/2}, & \xi_{l,m} < 0, \end{cases} \quad \Psi_{P(F)} \equiv \Psi_\vartheta^\pm = \begin{cases} \Psi_{i,j\pm 1/2,k,l,m}, & \eta_{l,m} > 0, \\ \Psi_{i,j\mp 1/2,k,l,m}, & \eta_{l,m} < 0. \end{cases} \end{aligned} \quad (6)$$

Здесь $\Psi_{R(L)}$, $\Psi_{P(F)}$ и $\Psi_{T(B)}$ — нулевые пространственные моменты потока на пространственных гранях ячейки.

Для получения WDD схемы к уравнению баланса (5) следует добавить четыре дополнительных уравнения:

$$\begin{aligned}\psi_r^+ &= (1 + P_r) \psi - P_r \psi_r^-, \quad \psi_\vartheta^+ = (1 + P_\vartheta) \psi - P_\vartheta \psi_\vartheta^-, \quad \psi_z^+ = (1 + P_z) \psi - P_z \psi_z^-, \\ \psi_{m+1/2} &= (1 + P_\xi) \psi - P_\xi \psi_{m-1/2}, \quad 0 \leq P_r, P_\vartheta, P_z, P_\xi \leq 1,\end{aligned}\quad (7)$$

где P_r, P_ϑ, P_z и P_ξ — весовые коэффициенты WDD схемы. Для расчета весовых коэффициентов в программе используется адаптивная WDD (AWDD) схема (см. [23]–[26]), которая определяет эти веса в результате оценки градиентов решения по каждой из переменных. AWDD схема обеспечивает положительность и приемлемый (регулируемый) уровень монотонности разностного решения с сохранением достаточно высокой точности расчета интегральных величин, не препятствует сходимости внутренних итераций.

2. ПОСТРОЕНИЕ KP_1 -СХЕМЫ УСКОРЕНИЯ ВНУТРЕННИХ ИТЕРАЦИЙ В R, Θ, Z ГЕОМЕТРИИ

Перепишем уравнение баланса (5) и систему дополнительных уравнений WDD схемы (7) в r, ϑ, z геометрии на $(n + 1)$ -й внутренней итерации в более удобном для последующего изложения виде:

$$\begin{aligned}\eta_{l,m} \Delta r_i \Delta z_k \left(\psi_{i,j+1/2,k,l,m}^{n+1/2} - \psi_{i,j-1/2,k,l,m}^{n+1/2} \right) + \mu_l v_i \Delta \vartheta_j \left(\psi_{i,j,k+1/2,l,m}^{n+1/2} - \psi_{i,j,k-1/2,l,m}^{n+1/2} \right) + \Delta z_k \Delta \vartheta_j \times \\ \times \left[\xi_{l,m} \left(A_{i+1/2} \psi_{i+1/2,j,k,l,m}^{n+1/2} - A_{i-1/2} \psi_{i-1/2,j,k,l,m}^{n+1/2} \right) + \right. \\ \left. + \frac{C_i}{w_{l,m}} \left(\alpha_{l,m+1/2} \psi_{i,j,k,l,m+1/2}^{n+1/2} - \alpha_{l,m-1/2} \psi_{i,j,k,l,m-1/2}^{n+1/2} \right) \right] +\end{aligned}\quad (8)$$

$$+ \sigma_l V_{i,j,k} \psi_{i,j,k,l,m}^{n+1/2} = V_{i,j,k} \sum_{\kappa=0}^K \frac{2\kappa+1}{4\pi} \sigma_{s,\kappa} \sum_{\nu=-\kappa}^{\kappa} Y_{\kappa}^{\nu}(\mu_l, \varphi_{l,m}) \Phi_{\kappa,i,j,k}^{\nu,n} + V_{i,j,k} F_{i,j,k,l,m},$$

$$\psi_{i,j,k,l,m}^{n+1/2} = a_{i,j,k,l,m}^{n+1/2} \psi_{i+1/2,j,k,l,m}^{n+1/2} + \left(1 - a_{i,j,k,l,m}^{n+1/2} \right) \psi_{i-1/2,j,k,l,m}^{n+1/2}, \quad a = \begin{cases} \frac{1}{1 + P_r}, & \xi > 0, \\ \frac{1}{1 + P_r}, & \xi < 0, \end{cases} \quad 0 \leq P_r \leq 1, \quad (9)$$

$$\psi_{i,j,k,l,m}^{n+1/2} = b_{i,j,k,l,m}^{n+1/2} \psi_{i,j+1/2,k,l,m}^{n+1/2} + \left(1 - b_{i,j,k,l,m}^{n+1/2} \right) \psi_{i,j-1/2,k,l,m}^{n+1/2}, \quad b = \begin{cases} \frac{1}{1 + P_\vartheta}, & \eta > 0, \\ \frac{1}{1 + P_\vartheta}, & \eta < 0, \end{cases} \quad 0 \leq P_\vartheta \leq 1, \quad (10)$$

$$\psi_{i,j,k,l,m}^{n+1/2} = c_{i,j,k,l,m}^{n+1/2} \psi_{i,j,k+1/2,l,m}^{n+1/2} + \left(1 - c_{i,j,k,l,m}^{n+1/2} \right) \psi_{i,j,k-1/2,l,m}^{n+1/2}, \quad c = \begin{cases} \frac{1}{1 + P_z}, & \mu > 0, \\ \frac{1}{1 + P_z}, & \mu < 0, \end{cases} \quad 0 \leq P_z \leq 1, \quad (11)$$

$$\psi_{i,j,k,l,m+1/2}^{n+1/2} = (1 + P_\xi) \psi_{i,j,k,l,m}^{n+1/2} - P_\xi \psi_{i,j,k,l,m-1/2}^{n+1/2} = \sum_{m'=1/2,1,\dots,m} d_{m'}^{l,m} \psi_{i,j,k,l,m'}^{n+1/2}, \quad 0 \leq P_\xi \leq 1. \quad (12)$$

В уравнении (8) $\sigma_{s,\kappa}$ — это κ -я гармоника заданной в P_K приближении индикатрисы рассеяния в группе, $F_{i,j,k,l,m}$ — заданный многогрупповой источник в группе; сферические гармоники $Y_{\kappa}^{\nu}(\mu, \varphi)$ и угловые моменты потока $\Phi_{\kappa,i,j,k}^{\nu,n}$ определены следующим образом:

$$Y_{\kappa}^{\nu}(\mu, \varphi) = \left[(2 - \delta_{0,\nu}) \frac{(\kappa - |\nu|)!}{(\kappa + |\nu|)!} \right]^{1/2} P_{\kappa}^{|\nu|}(\mu) \begin{cases} \sin |\nu| \varphi, & \nu = -\kappa, -\kappa + 1, \dots, -1, \\ \cos \nu \varphi, & \nu = 0, 1, \dots, \kappa, \end{cases} \quad \kappa = 0, 1, \dots, |\nu| \leq \kappa, \quad (13)$$

$$\Phi_{\kappa,i,j,k}^{\nu,n} = \sum_{l,m} w_{l,m} Y_{\kappa}^{\nu}(\mu_l, \varphi_{l,m}) \psi_{i,j,k,l,m}^n, \quad (14)$$

где $P_K^\nu(\nu)$ — это присоединенные полиномы Лежандра. Отметим здесь, что $Y_0^0 = 1$, $Y_1^{-1} = \eta$, $Y_1^0 = \mu$, $Y_1^1 = \xi$.

К системе уравнений (8)–(12) следует также добавить граничные условия при $r = r_{1/2} = r_{\text{int}} \geq 0$, $r = r_{I+1/2} = r_{\text{ext}}$, $\vartheta = \vartheta_{1/2} = \vartheta_0$, $\vartheta = \vartheta_{J+1/2} = \vartheta_{\text{end}}$, $z = z_{1/2} = z_{\text{bot}}$, $z = z_{K+1/2} = z_{\text{top}}$:

$$\begin{aligned} \Psi_{1/2,j,k,l,m}^{n+1/2} &= F_{j,k,l,m}^{\text{int}} + \sum_{\xi_{l',m'} < 0} w_{l',m'} R_{lm,l'm'}^{\text{int}} \Psi_{1/2,j,k,l',m'}^{n+1/2}, \quad \xi_{l,m} > 0, \\ \Psi_{I+1/2,j,k,l,m}^{n+1/2} &= F_{j,k,l,m}^{\text{ext}} + \sum_{\xi_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{\text{ext}} \Psi_{I+1/2,j,k,l',m'}^n, \quad \xi_{l,m} < 0, \end{aligned} \quad (15)$$

$$\begin{aligned} \Psi_{i,1/2,k,l,m}^{n+1/2} &= F_{i,k,l,m}^0 + \sum_{\eta_{l',m'} < 0} w_{l',m'} R_{lm,l'm'}^0 \Psi_{i,1/2,k,l',m'}^{n+1/2}, \quad \eta_{l,m} > 0, \\ \Psi_{i,J+1/2,k,l,m}^{n+1/2} &= F_{i,k,l,m}^{\text{end}} + \sum_{\eta_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{\text{end}} \Psi_{i,J+1/2,k,l',m'}^n, \quad \eta_{l,m} < 0, \end{aligned} \quad (16)$$

$$\begin{aligned} \Psi_{i,j,1/2,l,m}^{n+1/2} &= F_{i,j,l,m}^{\text{bot}} + \sum_{\mu_{l'} < 0} w_{l',m'} R_{lm,l'm'}^{\text{bot}} \Psi_{i,j,1/2,l',m'}^{n+1/2}, \quad \mu_l > 0, \\ \Psi_{i,j,K+1/2,l,m}^{n+1/2} &= F_{i,j,l,m}^{\text{top}} + \sum_{\mu_{l'} > 0} w_{l',m'} R_{lm,l'm'}^{\text{top}} \Psi_{i,j,K+1/2,l',m'}^n, \quad \mu_l < 0. \end{aligned} \quad (17)$$

При использовании периодических граничных условий, например, по переменной ϑ , что имеет место при наличии поворотной симметрии для сектора симметрии $(\vartheta_0, \vartheta_{\text{end}})$, граничные условия (16) заменяются следующими:

$$\begin{aligned} \Psi_{i,1/2,k,l,m}^{n+1/2} &= \Psi_{i,J+1/2,k,l,m}^n, \quad \eta_{l,m} > 0, \\ \Psi_{i,J+1/2,k,l,m}^{n+1/2} &= \Psi_{i,1/2,k,l,m}^n, \quad \eta_{l,m} < 0. \end{aligned} \quad (18)$$

Для ускорения сходимости внутренних итераций в КР₁-схеме используются линейные поправки к нулевому и первым угловым моментам решения следующего вида:

$$\begin{aligned} \Psi_{i,j,k,l,m}^{n+1} &= \Psi_{i,j,k,l,m}^{n+1/2} + \frac{1}{4\pi} (f_{i,j,k}^0 + 3\xi_{l,m} f_{i,j,k}^r + 3\eta_{l,m} f_{i,j,k}^\vartheta + 3\mu_l f_{i,j,k}^z), \\ \Psi_{i\pm 1/2,j,k,l,m}^{n+1} &= \Psi_{i\pm 1/2,j,k,l,m}^{n+1/2} + \frac{1}{4\pi} (f_{i\pm 1/2,j,k}^0 + 3\xi_{l,m} f_{i\pm 1/2,j,k}^r), \\ \Psi_{i,j\pm 1/2,k,l,m}^{n+1} &= \Psi_{i,j\pm 1/2,k,l,m}^{n+1/2} + \frac{1}{4\pi} (f_{i,j\pm 1/2,k}^0 + 3\eta_{l,m} f_{i,j\pm 1/2,k}^\vartheta), \\ \Psi_{i,j,k\pm 1/2,l,m}^{n+1} &= \Psi_{i,j,k\pm 1/2,l,m}^{n+1/2} + \frac{1}{4\pi} (f_{i,j,k\pm 1/2}^0 + 3\mu_l f_{i,j,k\pm 1/2}^z). \end{aligned} \quad (19)$$

Для получения системы уравнений для определения ускоряющих поправок f^0 , f^r , f^ϑ и f^z воспользуемся следующей процедурой, близкой к “4-step” процедуре Ларсена (см. [27], [12], а также [13]). Прежде всего, подействуем на уравнение баланса (8) операторами проектирования $\hat{L}_0$, $\hat{L}_r$, $\hat{L}_\vartheta$ и $\hat{L}_z$:

$$\begin{aligned} L_0 \Psi &= \sum_{l,m} w_{l,m} \Psi_{i,j,k,l,m}, \quad L_r \Psi = \sum_{l,m} w_{l,m} \xi_{l,m} \Psi_{i,j,k,l,m}, \\ L_\vartheta \Psi &= \sum_{l,m} w_{l,m} \eta_{l,m} \Psi_{i,j,k,l,m}, \quad L_z \Psi = \sum_{l,m} w_{l,m} \mu_l \Psi_{i,j,k,l,m}, \end{aligned} \quad (20)$$

где суммирование производится по восьми октантам. В предположении, что используемая квадратура достаточно точно интегрирует соответствующие сферические гармоники, получим

$$\begin{aligned} \Delta r_i \Delta z_k &\left(f_{i,j+1/2,k}^\vartheta + f_{i,j-1/2,k}^\vartheta \right) + v_i \Delta \vartheta_j \left(f_{i,j,k+1/2}^z - f_{i,j,k-1/2}^z \right) + \\ &+ \Delta z_k \Delta \vartheta_j \left(A_{i+1/2} f_{i+1/2,j,k}^r - A_{i-1/2} f_{i-1/2,j,k}^r \right) + \sigma^0 V_{i,j,k} f_{i,j,k}^0 = V_{i,j,k} Q_{i,j,k}^0, \end{aligned}$$

$$\begin{aligned}
& \frac{1}{3} \Delta z_k \Delta \vartheta_j \left(A_{i+1/2} f_{i+1/2,j,k}^0 - A_{i-1/2} f_{i-1/2,j,k}^0 \right) - \frac{1}{3} C_i \Delta z_k \Delta \vartheta_j f_{i,j,k}^0 + \\
& \quad + (\sigma^{11} V_{i,j,k} + M_r C_i \Delta z_k \Delta \vartheta_j) f_{i,j,k}^r = V_{i,j,k} Q_{i,j,k}^r, \\
& \frac{1}{3} \Delta r_i \Delta z_k \left(f_{i,j+1/2,k}^0 - f_{i,j-1/2,k}^0 \right) + (\sigma^{11} V_{i,j,k} + M_\vartheta C_i \Delta z_k \Delta \vartheta_j) f_{i,j,k}^\vartheta = V_{i,j,k} Q_{i,j,k}^\vartheta, \\
& \frac{1}{3} v_i \Delta \vartheta_j \left(f_{i,j,k+1/2}^0 - f_{i,j,k-1/2}^0 \right) + \sigma^{11} V_{i,j,k} f_{i,j,k}^z = V_{i,j,k} Q_{i,j,k}^z,
\end{aligned} \tag{21}$$

где

$$\begin{aligned}
Q_{i,j,k}^0 &= \sigma_{s,0} \left(\Phi_{i,j,k}^{0,n+1/2} - \Phi_{i,j,k}^{0,n} \right), \quad Q_{i,j,k}^r = \sigma_{s,1} \left(\Phi_{i,j,k}^{r,n+1/2} - \Phi_{i,j,k}^{r,n} \right), \quad Q_{i,j,k}^\vartheta = \sigma_{s,1} \left(\Phi_{i,j,k}^{\vartheta,n+1/2} - \Phi_{i,j,k}^{\vartheta,n} \right), \\
Q_{i,j,k}^z &= \sigma_{s,1} \left(\Phi_{i,j,k}^{z,n+1/2} - \Phi_{i,j,k}^{z,n} \right), \quad \sigma^{00} = \sigma - \sigma_{s,0}, \quad \sigma^{11} = \sigma - \sigma_{s,1}, \\
\Phi_{i,j,k}^0 &\equiv \Phi_{0,i,j,k}^0, \quad \Phi_{i,j,k}^r \equiv \Phi_{1,i,j,k}^1, \quad \Phi_{i,j,k}^\vartheta \equiv \Phi_{1,i,j,k}^{-1}, \quad \Phi_{i,j,k}^z \equiv \Phi_{1,i,j,k}^0,
\end{aligned} \tag{22}$$

$$\begin{aligned}
M_r &= \frac{3}{4\pi} \sum_{l,m} \xi_{l,m} \left(\alpha_{l,m+1/2} \pi_{l,m+1/2}^r - \alpha_{l,m-1/2} \pi_{l,m-1/2}^r \right), \\
M_\vartheta &= \frac{3}{4\pi} \sum_{l,m} \eta_{l,m} \left(\alpha_{l,m+1/2} \pi_{l,m+1/2}^\vartheta - \alpha_{l,m-1/2} \pi_{l,m-1/2}^\vartheta \right).
\end{aligned} \tag{23}$$

Величины $\pi_{l,m\pm 1/2}^r$, $\pi_{l,m\pm 1/2}^\vartheta$ в (23) могут быть вычислены с помощью следующих рекуррентных соотношений:

$$\begin{aligned}
\pi_{l,m\pm 1/2}^r &= \left(1 + P_{l,m}^\xi \right) \xi_{l,m} - P_{l,m}^\xi \pi_{l,m-1/2}^r, \quad \pi_{l,1/2}^r = \xi_{l,1/2}, \\
\pi_{l,m\pm 1/2}^\vartheta &= \left(1 + P_{l,m}^\xi \right) \eta_{l,m} - P_{l,m}^\xi \pi_{l,m-1/2}^\vartheta, \quad \pi_{l,1/2}^\vartheta = \eta_{l,1/2} = 0.
\end{aligned} \tag{24}$$

В систему из четырех уравнений (21) входит десять неизвестных величин: $f_{i,j,k}^0$, $f_{i,j,k}^r$, $f_{i,j,k}^\vartheta$, $f_{i,j,k}^z$, $f_{i\pm 1/2,j,k}^0$, $f_{i\pm 1/2,j,k}^r$, $f_{i,j\pm 1/2,k}^0$, $f_{i,j\pm 1/2,k}^\vartheta$, $f_{i,j,k\pm 1/2}^0$ и $f_{i,j,k\pm 1/2}^z$. Для замыкания системы (21) к ней необходимо добавить шесть дополнительных уравнения, а также граничные условия. Искомые дополнительные уравнения можно получить из естественного требования (см. [13]) инвариантности системы (21) при заменах в уравнении баланса (1.11.3):

$$\psi \rightarrow a\psi_{i+1/2} + (1-a)\psi_{i-1/2}, \quad \psi \rightarrow b\psi_{j+1/2} + (1-b)\psi_{j-1/2}, \quad \psi \rightarrow c\psi_{k+1/2} + (1-c)\psi_{k-1/2}. \tag{25}$$

Если используемая квадратура достаточно точна, то полученные таким образом соотношения совпадают с теми, которые получаются применением операторов $\hat{L}_0$ и $\hat{L}_r$ к уравнению (9), операторов $\hat{L}_0$ и $\hat{L}_\vartheta$ к уравнению (10) и операторов $\hat{L}_0$ и $\hat{L}_z$ к уравнению (11) (как это рекомендуется в “4-step” процедуре Ларсена, [27], [12]):

$$\begin{aligned}
f_{i,j,k}^0 &= A_{i,j,k}^0 f_{i+1/2,j,k}^0 + \left(1 - A_{i,j,k}^0 \right) f_{i-1/2,j,k}^0 + A_{i,j,k}^1 \left(f_{i+1/2,j,k}^r - f_{i-1/2,j,k}^r \right), \\
f_{i,j,k}^r &= A_{i,j,k}^2 f_{i+1/2,j,k}^r + \left(1 - A_{i,j,k}^2 \right) f_{i-1/2,j,k}^r + A_{i,j,k}^1 \left(f_{i+1/2,j,k}^0 - f_{i-1/2,j,k}^0 \right) / 3,
\end{aligned} \tag{26}$$

$$\begin{aligned}
f_{i,j,k}^0 &= B_{i,j,k}^0 f_{i,j+1/2,k}^0 + \left(1 - B_{i,j,k}^0 \right) f_{i,j-1/2,k}^0 + B_{i,j,k}^1 \left(f_{i,j+1/2,k}^\vartheta - f_{i,j-1/2,k}^\vartheta \right), \\
f_{i,j,k}^\vartheta &= B_{i,j,k}^2 f_{i,j+1/2,k}^\vartheta + \left(1 - B_{i,j,k}^2 \right) f_{i,j-1/2,k}^\vartheta + B_{i,j,k}^1 \left(f_{i,j+1/2,k}^0 - f_{i,j-1/2,k}^0 \right) / 3,
\end{aligned} \tag{27}$$

$$\begin{aligned}
f_{i,j,k}^0 &= C_{i,j,k}^0 f_{i,j,k+1/2}^0 + \left(1 - C_{i,j,k}^0 \right) f_{i,j,k-1/2}^0 + C_{i,j,k}^1 \left(f_{i,j,k+1/2}^z - f_{i,j,k-1/2}^z \right), \\
f_{i,j,k}^z &= C_{i,j,k}^2 f_{i,j,k+1/2}^z + \left(1 - C_{i,j,k}^2 \right) f_{i,j,k-1/2}^z + C_{i,j,k}^1 \left(f_{i,j,k+1/2}^0 - f_{i,j,k-1/2}^0 \right) / 3,
\end{aligned} \tag{28}$$

где

$$\begin{aligned} A_{i,j,k}^0 &= \frac{1}{4\pi} \sum_{l,m} w_{l,m} a_{i,j,k,l,m}, & A_{i,j,k}^1 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \xi_{l,m} a_{i,j,k,l,m}, & A_{i,j,k}^2 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \xi_{l,m}^2 a_{i,j,k,l,m}, \\ B_{i,j,k}^0 &= \frac{1}{4\pi} \sum_{l,m} w_{l,m} b_{i,j,k,l,m}, & B_{i,j,k}^1 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \eta_{l,m} b_{i,j,k,l,m}, & B_{i,j,k}^2 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \eta_{l,m}^2 b_{i,j,k,l,m}, \\ C_{i,j,k}^0 &= \frac{1}{4\pi} \sum_{l,m} w_{l,m} c_{i,j,k,l,m}, & C_{i,j,k}^1 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \mu_{l,m} c_{i,j,k,l,m}, & C_{i,j,k}^2 &= \frac{3}{4\pi} \sum_{l,m} w_{l,m} \mu_{l,m}^2 c_{i,j,k,l,m}. \end{aligned} \quad (29)$$

Для получения граничных условий для P_1 системы для ускоряющих поправок воспользуемся следующей процедурой, обобщающей использованную ранее для случая одномерных (см. [19], [28]) и двумерных (см. [15]) геометрий. Домножим уравнения (15), определяющие закон отражения при $r = r_{\text{int}}$ и $r = r_{\text{ext}}$, на $w_{l,m} \xi_{l,m}$ и просуммируем их по направлениям $\xi_{l,m} > 0$ и $\xi_{l,m} < 0$ соответственно. С учетом вида поправок (19), получим

$$l_0^r f_{1/2,j,k}^0 + 3l_1^r f_{1/2,j,k}^r = 0, \quad m_0^r f_{I+1/2,j,k}^0 + 3m_1^r f_{I+1/2,j,k}^r = \lambda_{j,k}^r, \quad (30)$$

где

$$\begin{aligned} l_p^r &= \sum_{\xi_{l,m} > 0} w_{l,m} \xi_{l,m} \left(\xi_{l,m}^p - \sum_{\xi_{l',m'} < 0} w_{l',m'} \xi_{l',m'}^p R_{lm,l'm'}^{\text{int}} \right), \\ m_p^r &= \sum_{\xi_{l,m} < 0} w_{l,m} \xi_{l,m} \left(\xi_{l,m}^p - \sum_{\xi_{l',m'} > 0} w_{l',m'} \xi_{l',m'}^p R_{lm,l'm'}^{\text{ext}} \right), \\ p = 0, 1, \quad \lambda_{j,k}^r &= 4\pi \sum_{\xi_{l,m} < 0} w_{l,m} \xi_{l,m} \sum_{\xi_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{\text{ext}} \left(\Psi_{I+1/2,j,k,l',m'}^{n+1/2} - \Psi_{I+1/2,j,k,l',m'}^n \right). \end{aligned} \quad (31)$$

Для определения граничных условий для поправок на границах $\vartheta = \vartheta_0$ и $\vartheta = \vartheta_{\text{end}}$ домножим уравнения (16), определяющие закон отражения на этих границах, на $w_{l,m} \eta_{l,m}$ и просуммируем их по направлениям $\eta_{l,m} > 0$ и $\eta_{l,m} < 0$ соответственно. С учетом вида поправок (19), получим

$$l_0^\vartheta f_{i,1/2,k}^0 + 3l_1^\vartheta f_{i,1/2,k}^\vartheta = 0, \quad m_0^\vartheta f_{i,J+1/2,k}^0 + 3m_1^\vartheta f_{i,J+1/2,k}^\vartheta = \lambda_{i,k}^\vartheta, \quad (32)$$

где

$$\begin{aligned} l_p^\vartheta &= \sum_{\eta_{l,m} > 0} w_{l,m} \eta_{l,m} \left(\eta_{l,m}^p - \sum_{\eta_{l',m'} < 0} w_{l',m'} \eta_{l',m'}^p R_{lm,l'm'}^0 \right), \\ m_p^\vartheta &= \sum_{\eta_{l,m} < 0} w_{l,m} \eta_{l,m} \left(\eta_{l,m}^p - \sum_{\eta_{l',m'} > 0} w_{l',m'} \eta_{l',m'}^p R_{lm,l'm'}^{\text{end}} \right), \\ p = 0, 1, \quad \lambda_{i,k}^\vartheta &= 4\pi \sum_{\eta_{l,m} < 0} w_{l,m} \eta_{l,m} \sum_{\eta_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{\text{end}} \left(\Psi_{i,J+1/2,k,l',m'}^{n+1/2} - \Psi_{i,J+1/2,k,l',m'}^n \right). \end{aligned} \quad (33)$$

При использовании по переменной ϑ периодических граничных условий, применяя аналогичную процедуру к уравнениям (18), получим

$$\begin{aligned} l_0^\vartheta \left(f_{i,1/2,k}^0 - f_{i,J+1/2,k}^0 \right) + 3l_1^\vartheta \left(f_{i,1/2,k}^\vartheta - f_{i,J+1/2,k}^\vartheta \right) &= \lambda_{i,k}^{\vartheta,\text{end}}, \\ m_0^\vartheta \left(f_{i,J+1/2,k}^0 - f_{i,1/2,k}^0 \right) + 3m_1^\vartheta \left(f_{i,J+1/2,k}^\vartheta - f_{i,1/2,k}^\vartheta \right) &= \lambda_{i,k}^{\vartheta,0}, \end{aligned} \quad (34)$$

где

$$l_p^\vartheta = \sum_{\eta_{l,m} > 0} w_{l,m} \eta_{l,m} \eta_{l,m}^p, \quad m_p^\vartheta = \sum_{\eta_{l,m} < 0} w_{l,m} \eta_{l,m} \eta_{l,m}^p, \quad p = 0, 1,$$

$$\begin{aligned}\lambda_{i,k}^{\vartheta,0} &= 4\pi \sum_{\eta_{l,m} < 0} w_{l,m} \eta_{l,m} \left(\Psi_{i,1/2,k,l',m'}^{n+1/2} - \Psi_{i,1/2,k,l',m'}^n \right), \\ \lambda_{i,k}^{\vartheta,end} &= 4\pi \sum_{\eta_{l,m} > 0} w_{l,m} \eta_{l,m} \left(\Psi_{i,J+1/2,k,l',m'}^{n+1/2} - \Psi_{i,J+1/2,k,l',m'}^n \right).\end{aligned}\quad (35)$$

Для определения граничных условий для поправок на границах $z = z_{bot}$ и $z = z_{top}$ домножим уравнения (18), определяющие закон отражения на этих границах, на $w_{l,m}\mu_l$ и просуммируем их по направлениям $\mu_l > 0$ и $\mu_l < 0$ соответственно. С учетом вида поправок (19), получим

$$l_0^z f_{i,j,1/2}^0 + 3l_1^z f_{i,j,1/2}^z = 0, \quad m_0^z f_{i,j,K+1/2}^z + 3m_1^z f_{i,j,K+1/2}^z = \lambda_{i,j}^z, \quad (36)$$

где

$$\begin{aligned}l_p^z &= \sum_{\mu_l > 0} w_{l,m} \mu_l \left(\mu_l^p - \sum_{\mu_{l'},m' < 0} w_{l',m'} \mu_{l'}^p R_{lm,l'm'}^{bot} \right), \\ m_p^z &= \sum_{\mu_l < 0} w_{l,m} \mu_l \left(\mu_l^p - \sum_{\mu_{l'},m' > 0} w_{l',m'} \mu_{l'}^p R_{lm,l'm'}^{top} \right), \\ p = 0, 1, \quad \lambda_{i,k}^z &= 4\pi \sum_{\mu_l < 0} w_{l,m} \mu_l \sum_{\mu_{l'},m' > 0} w_{l',m'} R_{lm,l'm'}^{top} \left(\Psi_{i,j,K+1/2,l',m'}^{n+1/2} - \Psi_{i,j,K+1/2,l',m'}^n \right).\end{aligned}\quad (37)$$

Следует отметить, что в уравнениях (30), (32), (34) и (36) величины $\lambda_{j,k}^r$, $\lambda_{i,k}^{\vartheta}$, $\lambda_{i,k}^{\vartheta,0}$, $\lambda_{i,k}^{\vartheta,end}$ и $\lambda_{i,j}^z$ стремятся к нулю по мере сходимости итераций. По этой причине возможно использование также и упрощенных граничных условий на границах $r = r_{ext}$, $\vartheta = \vartheta_0$, $\vartheta = \vartheta_{end}$ и $z = z_{top}$, отличающихся от вышеприведенных тем, что правые части в них полагаются равными нулю:

$$\lambda_{j,k}^r = \lambda_{i,k}^{\vartheta} = \lambda_{i,j}^z = 0, \quad \lambda_{i,k}^{\vartheta,0} = \lambda_{i,k}^{\vartheta,end} = 0. \quad (38)$$

В частности, условия поворотной симметрии (34) по переменной ϑ приобретают вид

$$f_{i,1/2,k}^0 = f_{i,J+1/2,k}^0, \quad f_{i,1/2,k}^{\vartheta} = f_{i,J+1/2,k}^{\vartheta}. \quad (39)$$

Использование упрощенных граничных условий, в свою очередь, позволяет упростить постановку граничных условий в методе расщепления (МР) (см. [30], [16], [17]), используемому для нахождения решения KP_1 -системы для ускоряющих поправок.

Учитывая, что в 2D и 3D геометриях в задачах с существенной ролью анизотропии рассеяния, когда выполняется неравенство

$$\lambda = \frac{\mu_0 c}{1 - \mu_0 c} > 1, \quad \text{где} \quad \mu_0 = \frac{\sigma_{s1}}{\sigma_{s0}}, \quad c = \frac{\sigma_{s0}}{\sigma_t}, \quad (40)$$

сходимость KP_1 -схемы нарушается (см. [7], [29]), для обеспечения сходимости KP_1 -схемы в задачах переноса нейтронов и фотонов нами используется следующий простой прием (см. [14], [15]): на четных внутренних итерациях ускоряющая коррекция в соответствии с (19) производится для четырех угловых моментов решения, включая нулевой, на нечетных — только для скалярного потока. Ясно, что такое решение не является оптимальным, но оно обеспечивает сходимость KP_1 -схемы в неоднородных геометриях для широкого класса практических задач.

P_1 система для ускоряющих поправок для KP_1 -схемы в r , ϑ , z геометрии состоит из четырех балансных уравнений (21), шести дополнительных уравнений (26)–(29) и граничных условий (31)–(37). Эта система может быть решена итерационно с использованием, например, метода расщепления (см. [30], [16], [17]).

Односкоростное уравнение переноса в x , y , z геометрии имеет вид

$$\begin{aligned}\xi \frac{\partial \Psi}{\partial x} + \eta \frac{\partial \Psi}{\partial y} + \mu \frac{\partial \Psi}{\partial z} + \sigma \Psi(x, y, z; \theta, \varphi) &= S(x, y, z; \theta, \varphi), \\ x_{left} \leq x \leq x_{right}, \quad y_{front} \leq y \leq x_{back}, \quad z_{bot} \leq z \leq z_{top},\end{aligned}\quad (41)$$

где ξ , η и μ — направляющие косинусы единичного вектора Ω направления скорости частицы:

$$\xi = (\Omega \mathbf{n}_x) = \sqrt{1 - \mu^2} \cos \phi, \quad \eta = (\Omega \mathbf{n}_y) = \sqrt{1 - \mu^2} \sin \phi, \quad \mu = (\Omega \mathbf{n}_z) = \sqrt{1 - \mu^2} \cos \theta, \quad (42)$$

которые изменяются в тех же пределах, что и в случае r , ϑ , z геометрии: $-1 \leq \xi, \eta, \mu \leq 1$. Все расчетные формулы КР₁-схемы для x , y , z геометрии получаются из соответствующих формул для r , ϑ , z геометрии путем замены:

$$r \rightarrow x, \quad \vartheta \rightarrow y, \quad A_{i\pm 1/2} \rightarrow 1, \quad C_1 \rightarrow 0, \quad v_i \rightarrow \Delta x_i, \quad M_r \rightarrow 0, \quad M_\vartheta \rightarrow 0. \quad (43)$$

3. КР₁-СХЕМА ДЛЯ УСКОРЕНИЯ ВНЕШНИХ ИТЕРАЦИЙ ПО ОБЛАСТИ ТЕРМАЛИЗАЦИИ НЕЙТРОНОВ

Обозначим через $\mathfrak{T}$ интервал групп, в котором имеются переходы вверх по группам за счет термализации нейтронов. Группа $q \in \mathfrak{T}$, если в эту группу имеются переходы из нижерасположенных групп. Мы будем исходить из системы балансных соотношений для интервала групп $\mathfrak{T}$ в r , ϑ , z геометрии:

$$\begin{aligned} & \eta_{l,m} \Delta r_i \Delta z_k \left(\Psi_{i,j+1/2,k,l,m}^{q,n+1/2} - \Psi_{i,j-1/2,k,l,m}^{q,n+1/2} \right) + \mu_l v_i \Delta \vartheta_j \left(\Psi_{i,j,k+1/2,l,m}^{q,n+1/2} - \Psi_{i,j,k-1/2,l,m}^{q,n+1/2} \right) + \Delta z_k \Delta \vartheta_j \times \\ & \quad \times \left[\xi_{l,m} \left(A_{i+1/2} \Psi_{i+1/2,j,k,l,m}^{q,n+1/2} - A_{i-1/2} \Psi_{i-1/2,j,k,l,m}^{q,n+1/2} \right) + \right. \\ & \quad \left. + \frac{C_i}{w_{l,m}} \left(\alpha_{l,m+1/2} \Psi_{i,j,k,l,m+1/2}^{q,n+1/2} - \alpha_{l,m-1/2} \Psi_{i,j,k,l,m-1/2}^{q,n+1/2} \right) \right] + \\ & \quad + \sigma_t^q V_{i,j,k} \Psi_{i,j,k,l,m}^{q,n+1/2} - V_{i,j,k} \sum_{p \leq q} \sum_{\kappa=0}^K \frac{2\kappa+1}{4\pi} \sigma_{s,\kappa}^{p \rightarrow q} \sum_{\nu=-\kappa}^{\kappa} Y_{\kappa}^{\nu}(\mu_l, \varphi_{l,m}) \Phi_{\kappa,i,j,k}^{p,\nu,n+1/2} = \\ & = V_{i,j,k} \sum_{p > q} \sum_{\kappa} \frac{2\kappa+1}{4\pi} \sigma_{s,\kappa}^{p \rightarrow q} \sum_{\nu=-\kappa}^{\kappa} Y_{\kappa}^{\nu}(\mu_l, \varphi_{l,m}) \Phi_{\kappa,i,j,k}^{p,\nu,n} + V_{i,j,k} F_{i,j,k,l,m}^q, \quad q, p \in \mathfrak{T}, \end{aligned} \quad (44)$$

дополнительных уравнений

$$\Psi_{i,j,k,l,m}^{q,n+1/2} = a_{i,j,k,l,m}^{q,n+1/2} \Psi_{i+1/2,j,k,l,m}^{q,n+1/2} + \left(1 - a_{i,j,k,l,m}^{q,n+1/2} \right) \Psi_{i-1/2,j,k,l,m}^{q,n+1/2}, \quad a = \begin{cases} \frac{1}{1 + P_r}, & \xi > 0, \\ \frac{1}{1 + P_r}, & \xi < 0, \end{cases} \quad 0 \leq P_r \leq 1, \quad (45)$$

$$\Psi_{i,j,k,l,m}^{q,n+1/2} = b_{i,j,k,l,m}^{q,n+1/2} \Psi_{i,j+1/2,k,l,m}^{q,n+1/2} + \left(1 - b_{i,j,k,l,m}^{q,n+1/2} \right) \Psi_{i,j-1/2,k,l,m}^{q,n+1/2}, \quad b = \begin{cases} \frac{1}{1 + P_\vartheta}, & \eta > 0, \\ \frac{1}{1 + P_\vartheta}, & \eta < 0, \end{cases} \quad 0 \leq P_\vartheta \leq 1, \quad (46)$$

$$\Psi_{i,j,k,l,m}^{q,n+1/2} = c_{i,j,k,l,m}^{q,n+1/2} \Psi_{i,j,k+1/2,l,m}^{q,n+1/2} + \left(1 - c_{i,j,k,l,m}^{q,n+1/2} \right) \Psi_{i,j,k-1/2,l,m}^{q,n+1/2}, \quad c = \begin{cases} \frac{1}{1 + P_z}, & \mu > 0, \\ \frac{1}{1 + P_z}, & \mu < 0, \end{cases} \quad 0 \leq P_z \leq 1, \quad (47)$$

$$\Psi_{i,j,k,l,m+1/2}^{q,n+1/2} = \left(1 + P_{\xi,i,j,k,l,m}^q \right) \Psi_{i,j,k,l,m}^q \Psi_{i,j,k,l,m-1/2}^{q,n+1/2} = \sum_{m'=1/2,1,\dots,m} d_{m'}^{q,l,m} \Psi_{i,j,k,l,m'}^{q,n+1/2}, \quad 0 \leq P_\xi \leq 1, \quad (48)$$

и граничных условий при $r = r_{1/2} = r_{\text{int}} \geq 0$, $r = r_{I+1/2} = r_{\text{ext}}$, $\vartheta = \vartheta_{1/2} = \vartheta_0$, $\vartheta = \vartheta_{J+1/2} = \vartheta_{\text{end}}$, $z = z_{1/2} = z_{\text{bot}}$, $z = z_{K+1/2} = z_{\text{top}}$:

$$\Psi_{1/2,j,k,l,m}^{q,n+1/2} = F_{j,k,l,m}^{q,\text{int}} + \sum_{\xi_{l',m'} < 0} w_{l',m'} R_{lm,l'm'}^{q,\text{int}} \Psi_{1/2,j,k,l',m'}^{q,n+1/2}, \quad \xi_{l,m} > 0,$$

$$\Psi_{I+1/2,j,k,l,m}^{q,n+1/2} = F_{j,k,l,m}^{q,ext} + \sum_{\xi_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{q,ext} \Psi_{I+1/2,j,k,l',m'}^{q,n+1/2}, \quad \xi_{l,m} < 0, \quad (49)$$

$$\Psi_{i,1/2,k,l,m}^{q,n+1/2} = F_{i,k,l,m}^{q,0} + \sum_{\eta_{l',m'} < 0} w_{l',m'} R_{lm,l'm'}^{q,0} \Psi_{i,1/2,k,l',m'}^{q,n+1/2}, \quad \eta_{l,m} > 0,$$

$$\Psi_{i,J+1/2,k,l,m}^{q,n+1/2} = F_{i,k,l,m}^{q,end} + \sum_{\eta_{l',m'} > 0} w_{l',m'} R_{lm,l'm'}^{q,end} \Psi_{i,J+1/2,k,l',m'}^{q,n+1/2}, \quad \eta_{l,m} < 0, \quad (50)$$

$$\Psi_{i,j,1/2,l,m}^{q,n+1/2} = F_{i,k,l,m}^{q,bot} + \sum_{\mu_{l'} < 0} w_{l',m'} R_{lm,l'm'}^{q,bot} \Psi_{i,j,1/2,l',m'}^{q,n+1/2}, \quad \mu_l > 0,$$

$$\Psi_{i,j,K+1/2,l,m}^{q,n+1/2} = F_{i,k,l,m}^{q,top} + \sum_{\mu_{l'} > 0} w_{l',m'} R_{lm,l'm'}^{q,top} \Psi_{i,j,K+1/2,l',m'}^{q,n+1/2}, \quad \mu_l < 0. \quad (51)$$

В уравнениях (44)–(51) индекс $n + 1/2$ обозначает результат решения многогрупповой системы кинетических уравнений (44)–(51) с правой частью, полученной на n -й внешней итерации.

В KP_1 -схеме ускорения внешних итераций ищутся поправки к нулевому и первому угловым моментам решения в виде

$$\begin{aligned} \Psi_{i,j,k,l,m}^{q,n+1} &= \Psi_{i,j,k,l,m}^{q,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} (f_{i,j,k}^0 + 3\xi_{l,m} f_{i,j,k}^r + 3\eta_{l,m} f_{i,j,k}^{\theta} + 3\mu_l f_{i,j,k}^z), \\ \Psi_{i\pm 1/2,j,k,l,m}^{q,n+1} &= \Psi_{i\pm 1/2,j,k,l,m}^{q,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} (f_{i\pm 1/2,j,k}^0 + 3\xi_{l,m} f_{i\pm 1/2,j,k}^r), \\ \Psi_{i,j\pm 1/2,k,l,m}^{q,n+1} &= \Psi_{i,j\pm 1/2,k,l,m}^{q,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} (f_{i,j\pm 1/2,k}^0 + 3\eta_{l,m} f_{i,j\pm 1/2,k}^{\theta}), \\ \Psi_{i,j,k\pm 1/2,l,m}^{q,n+1} &= \Psi_{i,j,k\pm 1/2,l,m}^{q,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} (f_{i,j,k\pm 1/2}^0 + 3\mu_l f_{i,j,k\pm 1/2}^z). \end{aligned} \quad (52)$$

В уравнения (52) входит функция ε_{β}^q , которая определяет энергетическую зависимость ускоряющих поправок для материала $\beta = 1, \dots, NK$, где NK — число материалов в защитной композиции. Предполагается, что функция ε_{β}^q , которую мы также будем называть спектральной функцией, нормирована на единицу:

$$\sum_{q \in \mathfrak{I}} \varepsilon_{\beta}^q = 1. \quad (53)$$

Пространственная зависимость спектральной функции ε_{β}^q имеет место только на границах различных материалов. Однако, как показывает численный эксперимент [18], этой зависимостью можно пренебречь.

Возможен также выбор спектральной функции, которая не зависит от номера материала β , а выбирается в зависимости от свойств гомогенизированной защитной композиции (см. [16]). Однако такой выбор, как показал численный эксперимент, несколько снижает эффективность алгоритма ускорения внешних итераций.

Форма ускоряющих поправок с факторизованной энергетической зависимостью позволяет построить арифметически простую схему ускорения внешних итераций, согласованную с используемой разностной аппроксимацией уравнения переноса (см. [19]). Эффективность и устойчивость этой схемы существенно зависит от выбора спектральной функции ε^q . В [19] была использована простейшая плоская (т.е. независимая от энергии) форма спектральной функции $\varepsilon^q = 1/G$. Однако этот выбор ε^q не является оптимальным. Использование в качестве ε^q распределения Максвелла с характерной температурой среды является более адекватным выбором в данной задаче. В этом случае величины ε^q вычисляются путем усреднения спектра Максвелла по энергетическому интервалу q -й группы $\{E_{q+1/2}, E_{q-1/2}\}$:

$$\begin{aligned} \varepsilon^q &: \frac{1}{\Delta E_q} \int_{E_{q+1/2}}^{E_{q-1/2}} E \exp\left(-\frac{E}{kT}\right) dE, \quad \Delta E_q = E_{q-1/2} - E_{q+1/2}, \quad k = 8.617385 \cdot 10^{-11} \text{ МэВ/К}, \\ &\sum_{q \in \mathfrak{I}} \varepsilon^q = 1. \end{aligned} \quad (54)$$

Более общий алгоритм определения формы спектра для ускоряющих поправок ϵ^q основан на оценке энергетической зависимости собственной функции, отвечающей наиболее медленно убывающей фурье-гармонике ошибки итерационного метода Гаусса—Зейделя (т.е. чистых, неускоренных внешних итераций) для рассматриваемой задачи (см. [18]). Данный алгоритм определения ϵ^q рассмотрен ниже в разд. 5 и 6.

Для нахождения ускоряющих поправок в уравнении (52) подействуем на балансное уравнение (44) и (аналогично “4-step” процедуре [27]) на дополнительные соотношения (44)–(46) операторами проектирования $\hat{\mathfrak{L}}_0$, $\hat{\mathfrak{L}}_r$, $\hat{\mathfrak{L}}_\vartheta$ и $\hat{\mathfrak{L}}_z$:

$$\begin{aligned}\hat{\mathfrak{L}}_0\Psi &= \sum_{q \in \mathfrak{T}} \sum_{l,m} w_{l,m} \Psi_{i,j,k,l,m}^q, & \hat{\mathfrak{L}}_r\Psi &= \sum_{q \in \mathfrak{T}} \sum_{l,m} w_{l,m} \xi_{i,m} \Psi_{i,j,k,l,m}^q, \\ \hat{\mathfrak{L}}_\vartheta\Psi &= \sum_{q \in \mathfrak{T}} \sum_{l,m} w_{l,m} \eta_{l,m} \Psi_{i,j,k,l,m}^q, & \hat{\mathfrak{L}}_z\Psi &= \sum_{q \in \mathfrak{T}} \sum_{l,m} w_{l,m} \mu_l \Psi_{i,j,k,l,m}^q,\end{aligned}\quad (55)$$

где суммирование проводится по энергетическому диапазону $q \in \mathfrak{T}$ и по восьми октантам единичной сферы угловых направлений. В итоге получим

$$\begin{aligned}& \Delta r_i \Delta z_k \left(f_{i,j+1/2,k}^\vartheta - f_{i,j-1/2,k}^\vartheta \right) + v_i \Delta \vartheta_j \left(f_{i,j,k+1/2}^z - f_{i,j,k-1/2}^z \right) + \\& + \Delta z_k \Delta \vartheta_j \left(A_{i+1/2} f_{i+1/2,j,k}^r - A_{i-1/2} f_{i-1/2,j,k}^r \right) + \sum^{00} V_{i,j,k} f_{i,j,k}^0 = V_{i,j,k} Q_{i,j,k}^0, \\& \frac{1}{3} \Delta z_k \Delta \vartheta_j \left(A_{i+1/2} f_{i+1/2,j,k}^0 - A_{i-1/2} f_{i-1/2,j,k}^0 \right) - \frac{1}{3} C_i \Delta z_k \Delta \vartheta_j f_{i,j,k}^0 + \\& + \left(\sum^{11} V_{i,j,k} + \mathfrak{M}_r C - i \Delta z_k \Delta \vartheta_j \right) f_{i,j,k}^r = V_{i,j,k} Q_{i,j,k}^r, \\& \frac{1}{3} \Delta r_i \Delta z_k \left(f_{i,j+1/2,k}^0 - f_{i,j-1/2,k}^0 \right) + \left(\sum^{11} V_{i,j,k} + \mathfrak{M}_\vartheta C_i \Delta z_k \Delta \vartheta_j \right) f_{i,j,k}^\vartheta = V_{i,j,k} Q_{i,j,k}^\vartheta, \\& \frac{1}{3} v_i \Delta \vartheta_j \left(f_{i,j,k+1/2}^0 - f_{i,j,k-1/2}^0 \right) + \sum^{11} V_{i,j,k} f_{i,j,k}^0 Q_{i,j,k}^z,\end{aligned}\quad (56)$$

где

$$\begin{aligned}\sum^{00} &= \sum_{q \in \mathfrak{T}} \left\{ \epsilon_\beta^q \sigma^q - \sum_{p \in \mathfrak{T}} \epsilon_\beta^p \sigma_{s,0}^{p \rightarrow q} \right\}, & \sum^{11} &= \sum_{q \in \mathfrak{T}} \left\{ \epsilon_\beta^q \sigma^q - \sum_{p \in \mathfrak{T}} \epsilon_\beta^p \sigma_{s,1}^{p \rightarrow q} \right\}, \\ Q_{i,j,k}^0 &= \sum_{q \in \mathfrak{T}} \sum_{p > q} \sigma_{s,0}^{p \rightarrow q} \left(\Phi_{i,j,k}^{p,0,n+1/2} - \Phi_{i,j,k}^{p,0,n} \right), & Q_{i,j,k}^\alpha &= \sum_{q \in \mathfrak{T}} \sum_{p > q} \sigma_{s,1}^{p \rightarrow q} \left(\Phi_{i,j,k}^{p,\alpha,n+1/2} - \Phi_{i,j,k}^{p,\alpha,n} \right), \quad \alpha = r, \vartheta, z, \\ \Phi_{i,j,k}^{q,0} &\equiv \Phi_{0,i,j,k}^{q,0}, & \Phi_{i,j,k}^{q,r} &\equiv \Phi_{1,i,j,k}^{q,1}, \quad \Phi_{i,j,k}^{q,\vartheta} \equiv \Phi_{1,i,j,k}^{q,-1}, \quad \Phi_{i,j,k}^{q,z} \equiv \Phi_{1,i,j,k}^{q,0},\end{aligned}\quad (57)$$

а дополнительные уравнения имеют вид

$$\begin{aligned}f_{i,j,k}^0 &= \mathfrak{V}_{i,j,k}^0 f_{i+1/2,j,k}^0 + \left(1 - \mathfrak{V}_{i,j,k}^0 \right) f_{i-1/2,j,k}^0 + \mathfrak{V}_{i,j,k}^1 \left(f_{i+1/2,j,k}^r - f_{i-1/2,j,k}^r \right), \\ f_{i,j,k}^r &= \mathfrak{V}_{i,j,k}^2 f_{i+1/2,j,k}^r + \left(1 - \mathfrak{V}_{i,j,k}^2 \right) f_{i-1/2,j,k}^r + \frac{\mathfrak{V}_{i,j,k}^1 \left(f_{i+1/2,j,k}^0 - f_{i-1/2,j,k}^0 \right)}{3},\end{aligned}\quad (58)$$

$$\begin{aligned}f_{i,j,k}^0 &= \mathfrak{B}_{i,j,k}^0 f_{i,j+1/2,k}^0 + \left(1 - \mathfrak{B}_{i,j,k}^0 \right) f_{i,j-1/2,k}^0 + \mathfrak{B}_{i,j,k}^1 \left(f_{i,j+1/2,k}^\vartheta - f_{i,j-1/2,k}^\vartheta \right), \\ f_{i,j,k}^\vartheta &= \mathfrak{B}_{i,j,k}^2 f_{i,j+1/2,k}^\vartheta + \left(1 - \mathfrak{B}_{i,j,k}^2 \right) f_{i,j-1/2,k}^\vartheta + \frac{\mathfrak{B}_{i,j,k}^1 \left(f_{i,j+1/2,k}^0 - f_{i,j-1/2,k}^0 \right)}{3},\end{aligned}\quad (59)$$

$$\begin{aligned}f_{i,j,k}^0 &= \mathfrak{C}_{i,j,k}^0 f_{i,j,k+1/2}^0 + \left(1 - \mathfrak{C}_{i,j,k}^0 \right) f_{i,j,k-1/2}^0 + \mathfrak{C}_{i,j,k}^1 \left(f_{i,j,k+1/2}^z - f_{i,j,k-1/2}^z \right), \\ f_{i,j,k}^z &= \mathfrak{C}_{i,j,k}^2 f_{i,j,k+1/2}^z + \left(1 - \mathfrak{C}_{i,j,k}^2 \right) f_{i,j,k-1/2}^z + \frac{\mathfrak{C}_{i,j,k}^1 \left(f_{i,j,k+1/2}^0 - f_{i,j,k-1/2}^0 \right)}{3}.\end{aligned}\quad (60)$$

Здесь величины $\mathbb{M}_r$, $\mathbb{M}_\vartheta$, $\mathbb{A}_x$, $\mathbb{B}_x$ и $\mathbb{C}_x$, $x = 0, 1, 2$, представляют собой усредненные по группам $q \in \mathfrak{T}$ с весом ε_β^q значения соответствующих групповых величин, определенных уравнениями (23) и (29):

$$\begin{aligned} \mathbb{M}_r &= \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q M_r^q, \quad \mathbb{M}_\vartheta = \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q M_\vartheta^q, \quad \mathbb{A}_{i,j,k}^x = \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q A_{i,j,k}^{x,q}, \quad \mathbb{B}_{i,j,k}^x = \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q B_{i,j,k}^{x,q}, \\ \mathbb{C}_{i,j,k}^x &= \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q C_{i,j,k}^{x,q}, \quad x = 0, 1, 2. \end{aligned} \quad (61)$$

Для получения граничных условий для системы (56)–(60) повторим процедуру получения граничных условий (30)–(37) для алгоритма ускорения внутренних итераций, дополнив ее усреднением по группам $q \in \mathfrak{T}$ с весом ε_β^q граничных условий (49)–(51). С учетом вида используемых поправок (52), получим

$$\begin{aligned} l_0^r f_{1/2,j,k}^0 + 3l_1^r f_{1/2,j,k}^r &= 0, \quad m_0^r f_{I+1/2,j,k}^0 + 3m_1^r f_{I+1/2,j,k}^r = 0, \\ l_0^\vartheta f_{i,1/2,k}^0 + 3l_1^\vartheta f_{i,1/2,k}^\vartheta &= 0, \quad m_0^\vartheta f_{i,J+1/2,k}^0 + 3m_1^\vartheta f_{i,J+1/2,k}^\vartheta = 0, \\ l_0^z f_{i,j,1/2}^0 + 3l_1^z f_{i,j,1/2}^z &= 0, \quad m_0^z f_{i,j,K+1/2}^0 + 3m_1^z f_{i,j,K+1/2}^z = 0, \end{aligned} \quad (62)$$

где

$$l_p^\alpha = \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q l_p^{q,\alpha}, \quad w_p^\alpha = \sum_{q \in \mathfrak{T}} \varepsilon_\beta^q w_p^{q,\alpha}, \quad p = 0, 1, \quad \alpha = r, \vartheta, z. \quad (63)$$

Из сравнения P_1 систем (21)–(37) и (56)–(63) следует, что KP_1 -схема ускорения внешних итераций использует свертки по группам с весом ε_β^q величин $\mathbb{M}_r$, $\mathbb{M}_\vartheta$, $\mathbb{A}_x$, $\mathbb{B}_x$ и $\mathbb{C}_x$, $x = 0, 1, 2$; l_p^α , m_p^α , $p = 0, 1$; $\alpha = r, \vartheta, z$ рассчитываемых в KP_1 -схеме ускорения внутренних итераций. Для решения системы (56)–(63) может быть использован метод расщепления (см. [30], [16], [17]).

Поточечно по групповой критерий сходимости внешних итераций по области термализации нейтронов

$$\max_{i,j,k,q \in \mathfrak{T}} \left| \frac{\Phi_{i,j,k}^{q,(n)} - \Phi_{i,j,k}^{q,(n-1)}}{\Phi_{i,j,k}^{q,(n)}} \right| < \varepsilon_{upsc}^{flux}, \quad (64)$$

где $\Phi_{i,j,k}^{q,(n)}$ — скалярный поток в (i, j, k) -й пространственной ячейке q -й группы на n -й внешней итерации, а $\varepsilon_{upsc}^{flux}$ — заданная точность сходимости внешних итераций по скалярному потоку, как показал численный эксперимент, является чрезмерно сильным. Из уравнения (57) можно сделать вывод, что в качестве более слабого критерия сходимости внешних итераций по области термализации нейтронов может быть выбран следующий поточечный, интегральный по рассеянным вверх по энергии тепловым нейтронам, критерий сходимости:

$$\max_{i,j,k} \left| \frac{U_{upsc,i,j,k}^{(n)} - U_{upsc,i,j,k}^{(n-1)}}{U_{upsc,i,j,k}^{(n)}} \right| < \varepsilon_{upsc}^{dens}, \quad U_{upsc,i,j,k}^{(n)} = \sum_{q \in \mathfrak{T}} \sum_{p > q} \sigma_{s,0}^{p \rightarrow q} \Phi_{i,j,k}^{p,n}, \quad (65)$$

где $U_{upsc,i,j,k}^{(n)}$ — плотность источника рассеиваемых вверх по энергии нейтронов для (i, j, k) -й пространственной ячейки, а $\varepsilon_{upsc}^{dens}$ — заданная точность сходимости внешних итераций по критерию (65). Имеющийся численный эксперимент подтверждает достаточность выбора критерия (65) для решения практических задач, а также уменьшение числа внешних итераций для достижения заданной точности сходимости.

Отметим, что критерий (65) вполне аналогичен критерию сходимости по плотности нейтронов деления, используемому в задачах с внешними итерациями по источнику деления.

Для уменьшения объема вычислительной работы в 3D S_n программе КАТРИН была принята стратегия последовательного увеличения точности сходимости внутренних итераций за счет уменьшения $\varepsilon_{in}^{(n)}$ по мере уменьшения достигнутой ошибки сходимости внешних итераций ε^* :

$$\varepsilon_{in}^{(n)} = \begin{cases} 10\varepsilon_{upsc}^{dens}, & n = 1, 2, \\ \max\{\varepsilon_{in}, \varepsilon_{upsc}^{dens}\}, & \varepsilon^* > 10, \\ \varepsilon_{in}, & 1 < \varepsilon^* < 10, \\ \min\{\varepsilon_{in}, 10^{-2}\varepsilon_{upsc}^{dens}\}, & \varepsilon^* < 1, \end{cases} \quad (66)$$

где ε_{in} и $\varepsilon_{upsc}^{dens}$ — заданные точности сходимости внутренних и внешних итераций по скалярному потоку. Кроме того, значение начального веса $P_{\vartheta}^{cut} = 0.95$ по переменной ϑ , используемое в алгоритме коррекции AWDD схемы, для 1-й внешней итерации по области термализации (для которой имеют место максимальные градиенты потоков тепловых нейтронов) уменьшено до значения 0.93.

Рассмотренный КР₁ метод ускорения внешних итераций по области термализации нейтронов, согласованный с WDD схемой, обобщается и на случай семейства нодальных WLD—WLB схем 3-го и 4-го порядков точности. В этом случае поправки к нулевому и первому угловым моментам решения, аналогично (52), ищутся в виде

$$\Psi_{i,j,k,l,m}^{q,\alpha,n+1} = \Psi_{i,j,k,l,m}^{q,\alpha,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} \left(f_{i,j,k,l,m}^{\alpha,0} + 3\xi_{l,m} f_{i,j,k,l,m}^{\alpha,r} + 3\eta_{l,m} f_{i,j,k,l,m}^{\alpha,\vartheta} + 3\mu_l f_{i,j,k,l,m}^{\alpha,z} \right), \quad \alpha = 0, r, \vartheta, z, \quad (67)$$

$$\Psi_{i\pm 1/2,j,k,l,m}^{q,\gamma,n+1} = \Psi_{i\pm 1/2,j,k,l,m}^{q,\gamma,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} \left(f_{i\pm 1/2,j,k}^{\gamma,0} + 3\xi_{l,m} f_{i\pm 1/2,j,k}^{\gamma,r} \right), \quad \gamma = 0, \vartheta, z,$$

$$\Psi_{i,j\pm 1/2,k,l,m}^{q,\gamma,n+1} = \Psi_{i,j\pm 1/2,k,l,m}^{q,\gamma,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} \left(f_{i,j\pm 1/2,k}^{\gamma,0} + 3\eta_{l,m} f_{i,j\pm 1/2,k}^{\gamma,\vartheta} \right), \quad \gamma = 0, r, z,$$

$$\Psi_{i,j,k\pm 1/2,l,m}^{q,\gamma,n+1} = \Psi_{i,j,k\pm 1/2,l,m}^{q,\gamma,n+1/2} + \varepsilon_{\beta}^q \frac{1}{4\pi} \left(f_{i,j,k\pm 1/2}^{\gamma,0} + 3\mu_l f_{i,j,k\pm 1/2}^{\gamma,z} \right), \quad \gamma = 0, r, \vartheta. \quad (68)$$

В уравнениях (67) и (68) α и γ — индексы пространственных моментов. Уравнение (67) содержит 16 поправок, уравнения (68) — 18 поправок. Дальнейшее построение КР₁-схемы ускорения внешних итераций по области термализации нейтронов, согласованной с WLB—WLD схемой, аналогично вышеприведенному для случая WDD схемы и опирается на результаты построения КР₁-схемы ускорения внутренних итераций, согласованной с WLB—WLD схемой (см. [31], [32]). Используемая в 3D S_n программе КАТРИН стратегия последовательного увеличения точности сходимости внутренних итераций за счет уменьшения $\varepsilon_{in}^{(n)}$ по мере уменьшения достигнутой ошибки сходимости внешних итераций ε^* для случая WLB—WLD схемы следующая:

$$\varepsilon_{in}^{(n)} = \begin{cases} 10\varepsilon_{upsc}^{dens}, & n = 1, 2, \\ \varepsilon_{in}, & \varepsilon^* > 1, \\ \min\{\varepsilon_{in}, 10^{-3}\varepsilon_{upsc}^{dens}\}, & \varepsilon^* < 1, \end{cases} \quad (69)$$

где ε_{in} и $\varepsilon_{upsc}^{dens}$ — заданные точности сходимости внутренних и внешних итераций по скалярному потоку.

4. КР₁-СХЕМА ДЛЯ УСКОРЕНИЯ ВНЕШНИХ ИТЕРАЦИЙ ПО ИСТОЧНИКУ ДЕЛЕНИЯ ПРИ РЕШЕНИИ ПОДКРИТИЧЕСКИХ ЗАДАЧ

В данном разделе мы рассмотрим, как модифицируется вышеописанная КР₁-схема ускорения внешних итераций в задаче об ускорении внешних итераций по источнику деления при решении подкритической краевой задачи. Обозначим через $\mathfrak{F}$ интервал групп, в котором имеются переходы вверх по группам за счет источника деления. Группа $q \in \mathfrak{F}$, если либо спектр деления χ^q , либо сечение деления σ_j^q в этой группе отличны от нуля. Мы будем исходить из системы балансных соотношений для

интервала групп $\mathfrak{F}$ в r, ϑ, z геометрии:

$$\begin{aligned} & \eta_{l,m} \Delta r_i \Delta z_k \left(\Psi_{i,j+1/2,k,l,m}^{q,n+1/2} - \Psi_{i,j-1/2,k,l,m}^{q,n+1/2} \right) + \mu_l v_i \Delta \vartheta_j \left(\Psi_{i,j,k+1/2,l,m}^{q,n+1/2} - \Psi_{i,j,k-1/2,l,m}^{q,n+1/2} \right) + \Delta z_k \Delta \vartheta_j \times \\ & \times \left[\xi_{l,m} \left(A_{i+1/2} \Psi_{i+1/2,j,k,l,m}^{q,n+1/2} - A_{i-1/2} \Psi_{i-1/2,j,k,l,m}^{q,n+1/2} \right) + \frac{C_i}{w_{i,m}} \left(\alpha_{l,m+1/2} \Psi_{i,j,k,l,m+1/2}^{q,n+1/2} - \right. \right. \\ & \left. \left. - \alpha_{l,m-1/2} \Psi_{i,j,k,l,m-1/2}^{q,n+1/2} \right) \right] + \sigma_{\ell}^q V_{i,j,k} \Psi_{i,j,k,l,m}^{q,n+1/2} - V_{i,j,k} \sum_{p \in \mathfrak{F}} \sum_{\kappa=0}^K \frac{2\kappa+1}{4\pi} \sigma_{s,\kappa}^{p \rightarrow q} \sum_{\nu=-\kappa}^{\kappa} Y_{\kappa}^{\nu}(\mu_l, \varphi_{l,m}) \Phi_{\kappa,i,j,k}^{p,\nu,n+1/2} = \\ & = V_{i,j,k} \frac{\chi^q}{4\pi} \sum_{p \in \mathfrak{F}} \nu \sigma_f^p \Phi_{i,j,k}^{p,0,n} + V_{i,j,k} F_{i,j,k,l,m}^q, \quad q, p \in \mathfrak{F}. \end{aligned} \quad (70)$$

Осуществляя процедуру, аналогичную описанной в предыдущем разделе, можно прийти к выводу, что в данном случае

$$\begin{aligned} \Sigma^{00} &= \sum_{q \in \mathfrak{F}} \left\{ \epsilon_{\beta}^q \sigma^q - \sum_{q \in \mathfrak{F}} \epsilon_{\beta}^p \sigma_{s,0}^{p \rightarrow q} - \chi^q \sum_{q \in \mathfrak{F}} \epsilon_{\beta}^p \nu \sigma_f^p \right\}, \quad \Sigma^{11} = \sum_{q \in \mathfrak{F}} \left\{ \epsilon_{\beta}^q \sigma^q - \sum_{q \in \mathfrak{F}} \epsilon_{\beta}^p \sigma_{s,1}^{p \rightarrow q} \right\}, \\ Q_{i,j,k}^0 &= \sum_{q \in \mathfrak{F}} \chi^q \sum_{p \in \mathfrak{F}} \nu \sigma_f^p \left(\Phi_{i,j,k}^{p,0,n+1/2} - \Phi_{i,j,k}^{p,0,n} \right), \quad Q_{i,j,k}^{\alpha} = 0, \quad \alpha = r, \vartheta, z. \end{aligned} \quad (71)$$

Все остальные соотношения п. 5.1 с заменой $\mathfrak{T} \rightarrow \mathfrak{F}$ справедливы и в этом случае. Следует, однако, отметить, что в случае подкритической задачи без фурье-анализа основной гармоники ошибки не удастся определить вид спектральной функции ϵ_{β}^q , близкий к оптимальному.

В качестве критерия сходимости внешних итераций по источнику деления естественно воспользоваться поточечным критерием сходимости для плотности нейтронов деления:

$$\max_{i,j,k} \left| \frac{N_{fis,i,j,k}^{neut,(n)} - N_{fis,i,j,k}^{neut,(n-1)}}{N_{fis,i,j,k}^{neut,(n)}} \right| < \epsilon_{fis}^{dens}, \quad N_{fis,i,j,k}^{neut,(n)} = \sum_{q=1}^Q \nu \sigma_f^q \Phi_{i,j,k}^{q,(n)}, \quad (72)$$

где ϵ_{fis}^{dens} — заданная точность сходимости внешних итераций по источнику деления.

5. ОЦЕНКА ФОРМЫ СПЕКТРА ДЛЯ УСКОРЯЮЩИХ ПОПРАВКОВ ДЛЯ ГОМОГЕННОЙ СРЕДЫ

Следуя [18], [19], получим оценку спектрального радиуса итерационного метода Гаусса—Зейделя (т.е. чистых неускоренных внешних итераций) решения системы групповых уравнений при наличии up-scattering на примере 1D плоской геометрии. В этом случае уравнение переноса на $n+1/2$ -й внешней итерации для интервала групп с up-scattering $q \in \mathfrak{T}$ может быть записано в виде

$$\mu \frac{\partial \Psi^{q,n+1/2}}{\partial x} + \sigma \Psi^{q,n+1/2}(x, \mu) = \sum_{l=0}^L \frac{2l+1}{4\pi} P_l(\mu) \left[\sum_{p \leq q} \sigma_{s,l}^{p \rightarrow q} \Phi_l^{p,n+1/2}(x) + \sum_{p > q} \sigma_{s,l}^{p \rightarrow q} \Phi_l^{p,n}(x) + F_l^q(x) \right], \quad q \in \mathfrak{T}, \quad (73)$$

где угловые моменты решения Φ_l^q определены следующим образом:

$$\Phi_l^q(x) = 2\pi \int_{-1}^1 \Psi^q(\mu', x) P_l(\mu') d\mu'.$$

Вычитая из исходного уравнения переноса уравнение (73) для $n+1/2$ -й внешней итерации, получим уравнение для поправок:

$$\mu \frac{\partial f^{q,n+1/2}}{\partial x} + \sigma f^{q,n+1/2}(x, \mu) = \sum_{l=0}^L \frac{2l+1}{4\pi} P_l(\mu) \left[\sum_{p \leq q} \sigma_{s,l}^{p \rightarrow q} f_l^{p,n+1/2}(x) + \sum_{p > q} \sigma_{s,l}^{p \rightarrow q} f_l^{p,n}(x) \right], \quad q \in \mathfrak{T}, \quad (74)$$

где

$$f^{q,n}(\mu, x) = \psi^q(\mu, x) - \psi^{q,n}(\mu, x), \quad f_l^{q,n}(x) = \Phi_l^q(x) - \Phi_l^{q,n}(x). \quad (75)$$

Представим пространственную зависимость поправок в виде разложения в ряд Фурье:

$$f^{q,n+1/2}(x, \mu) = \int_{-\infty}^{+\infty} f^{q,n+1/2}(\lambda, \mu) e^{i\lambda x} d\lambda, \quad \Phi_l^{q,n+1/2}(x, \mu) = \int_{-\infty}^{+\infty} \Phi_l^{q,n+1/2}(\lambda, \mu) e^{i\lambda x} d\lambda. \quad (76)$$

Используя дополнительно предположение о гомогенности рассматриваемой среды, из уравнения (74) получим

$$i\mu\lambda f^{q,n+1/2}(\lambda, \mu) + \sigma_t^q f^{q,n+1/2}(\lambda, \mu) = \sum_{l=0}^L \frac{2l+1}{4\pi} P_l(\mu) \left[\sum_{q \leq l} \sigma_{s,l}^{p \rightarrow q} f_l^{p,n+1/2}(\lambda) + \sum_{p > q} \sigma_{s,l}^{p \rightarrow q} f_l^{p,n}(\lambda) \right], \quad q \in \mathfrak{T}. \quad (77)$$

В матричной форме уравнение (77) можно переписать в следующем виде:

$$\mathbf{f}^{n+1/2}(\lambda, \mu) = \left(i\mu\lambda \hat{I} + \hat{T} \right)^{-1} \sum_{l=0}^L \frac{2l+1}{4\pi} P_l(\mu) \left[\hat{D}_l \mathbf{f}_l^{n+1/2}(\lambda) + \hat{U}_l \mathbf{f}_l^n(\lambda) \right], \quad (78)$$

где $\hat{I}$ — единичная матрица, $\hat{T}$ — диагональная матрица полных сечений, $\mathbf{f}^{n+1/2}$ — зависящий от угла вектор ошибки, $\hat{D}_l$ и $\hat{U}_l$ — матрицы, составленные из l -х гармоник матриц сечений переходов, описывающих down-scattering и up-scattering соответственно для диапазона групп $q \in \mathfrak{T}$.

В случае изотропного рассеяния ($L = 0$), умножая уравнение (78) на $P_0(\mu) = 1$ и интегрируя его по всем углам с учетом соотношения

$$\int_0^1 \frac{d\mu}{1 + \lambda^2 \mu^2} = \frac{1}{\lambda} \operatorname{arctg} \lambda, \quad (79)$$

получим

$$\mathbf{f}_0^{n+1/2}(\lambda) = \hat{A}(\lambda) \mathbf{f}_0^n(\lambda) = \left(1 - \hat{C} \hat{D}_0 \right)^{-1} \hat{C} \hat{U}_0 \mathbf{f}_0^n(\lambda), \quad \hat{C} = \frac{1}{\lambda} \operatorname{arctg} \left(\frac{\lambda}{\hat{T}} \right), \quad (80)$$

где $\hat{C}$ — диагональная матрица с элементами

$$C_{q,q} = \frac{1}{\lambda} \operatorname{arctg} \left(\frac{\lambda}{\sigma_t^q} \right). \quad (81)$$

Матрица $\hat{A}$ в уравнении (80) есть итерационная матрица метода Гаусса—Зейделя, и ее собственные значения $\omega(\lambda)$ определяют скорость сходимости итерационного процесса для гармоники Фурье λ . Эти собственные значения и соответствующие им собственные вектора $\omega^q(\lambda)$, $q = 1, \dots, Q$, где q — число групп с up-scattering, могут быть определены, используя стандартную библиотеку расчета собственных значений. Спектральный радиус итерационного процесса определяется как максимальное значение собственного значения оператора $\hat{A}$ по всем гармоникам λ :

$$\rho = \max_{\lambda} \{ \rho_M(\lambda) \}, \quad \rho_M(\lambda) = \max_q \{ \omega^q(\lambda) \}, \quad q = 1, \dots, Q. \quad (82)$$

Этот максимум достигается при значении $\lambda = 0$ (см. [18], [19]). Используя разложение

$$\operatorname{arctg}(x) \approx x - \frac{x^3}{3} + \dots, \quad |x| = 1,$$

получим, что для случая изотропного рассеяния

$$\mathbf{f}_0^{n+1/2}(0) = \hat{A}(0)\mathbf{f}_0^n(0) = (\hat{T} - \hat{D}_0)^{-1} \hat{U}_0 \mathbf{f}_0^n(0). \quad (83)$$

Собственный вектор оператора $\hat{A}(0)$, отвечающий максимальному собственному значению этого оператора (главный собственный вектор), описывает энергетическое распределение наиболее медленно убывающей гармоники ошибки. Для ее эффективного подавления именно энергетическое распределение этой гармоники (нормированное на 1) естественно использовать в качестве формы спектра для ускоряющих поправок ϵ^q (52) для достижения максимальной эффективности KP_1 -схемы ускорения внешних итераций (см. [18]).

В задаче термализации нейтронов при отсутствии поглощения нейтронов энергетическое распределение главного собственного вектора оператора $\hat{A}(0)$ совпадает с распределением Максвелла соответствующей температуры (см. [18]). Поэтому использование в качестве ϵ^q распределения Максвелла, усредненного по групповым интервалам (54), во многих случаях обеспечивает эффективность KP_1 -схемы ускорения внешних итераций сопоставимую с использованием главного собственного вектора оператора $\hat{A}(0)$.

Приведем здесь полученное в [19] выражение для итерационной матрицы KP_1 -схемы ускорения внешних итераций для случая изотропного рассеяния и гомогенной среды:

$$\hat{A}_{KP_1} = (1 - k\hat{\Xi}\hat{U})\hat{A} - k\hat{\Xi}\hat{U}, \quad k = a \left[a(a - b) + \frac{\lambda^2}{3} \right]^{-1}, \quad a = \sum_q \epsilon^q \sigma_t^q, \quad = \sum_q \sum_p \epsilon^p \sigma_{s,0}^{p \rightarrow q}, \quad (84)$$

где $\hat{A}$ — итерационная матрица метода Гаусса—Зейделя (80), $\hat{\Xi}$ — матрица, имеющая ранг 1, с элементами

$$\Xi_{q,p} = \epsilon^q, \quad q = 1, \dots, Q. \quad (85)$$

Как показали результаты численного эксперимента (см. [18]), спектральный радиус оператора определяется вторым собственным значением оператора $\hat{A}_{KP_1}$ (см. [18]). Для легких нуклидов, для которых эффект термализации существен, второе собственное значение оператора $\hat{A}(0)$ значительно меньше первого, и, следовательно, эффект ускорения весьма значителен.

6. ОЦЕНКА ФОРМЫ СПЕКТРА ДЛЯ УСКОРЯЮЩИХ ПОПРАВКОВ ДЛЯ ГЕТЕРОГЕННОЙ СРЕДЫ

Оценка главной собственной функции метода Гаусса—Зейделя для реальной неоднородной гетерогенной задачи является достаточно сложной задачей. В [18] предлагается использовать в алгоритме ускорения внешних итераций несколько функций ϵ^q по числу имеющихся материалов среды. Хотя такой подход не учитывает влияния на спектр нейтронов в зоне с некоторым материалом спектра нейтронов в соседних зонах с другими материалами, численный эксперимент, как уже отмечалось в разд. 3, показал его эффективность.

Однако не для всех материалов среды могут присутствовать переходы вверх по группам, обусловленные термализацией и (или) делением, которые используются при определении спектральной функции для материала. Для таких материалов можно воспользоваться спектральной функцией ϵ_{over}^q , полученной для однородной гомогенизированной среды. Для случая 3D геометрии гомогенизированные сечения среды могут быть рассчитаны по формулам

$$\bar{\sigma}_t^q = \frac{\sum_{i,j,k} V_{i,j,k} \sigma_{t,i,j,k}^q}{\sum_{i,j,k} V_{i,j,k}}, \quad \bar{\sigma}_{s,0}^{p \rightarrow q} = \frac{\sum_{i,j,k} V_{i,j,k} \sigma_{s,0,i,j,k}^{p \rightarrow q}}{\sum_{i,j,k} V_{i,j,k}}, \quad p, q = 1, \dots, Q. \quad (86)$$

7. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ ИСПОЛЬЗОВАНИЯ KP_1 -СХЕМЫ УСКОРЕНИЯ ВНЕШНИХ ИТЕРАЦИЙ

В данном разделе мы представим численные результаты, иллюстрирующие эффективность согласованной KP_1 -схемы ускорения внешних итераций по области термализации нейтронов для 1D, 2D

и 3D геометрий в зависимости от выбора разностной схемы (AWDD схемы 2-го порядка точности и AWLD схемы 3-го порядка точности), числа групп и критерия сходимости итераций по области термализации нейтронов (64) или (65). AWLD схема отличается от LD схемы (см. [33], [34]) наличием адаптивного алгоритма коррекции отрицательных потоков (в одномерных геометриях и в r, z и r, ϑ, z геометриях (в последних двух геометриях коррекция использовалась только по угловой переменной); в x, z и x, y, z геометриях коррекция не применялась). Выбор спектральной функции ϵ_{β}^q в уравнениях (52) и (67), (68) проводился на основе определения спектра собственных значений оператора $\hat{A}_0$ (83) для каждого из материалов задачи. Расчетные времена для тестовых задач в 1D, 2D и 3D x, y, z геометриях приведены для ПК Intel Core i7 3930K, а для задач в 3D r, ϑ, z геометрии — для ПК Intel Core i7-6950X.

Используемый ниже в таблицах термин $P_L S_N$ -приближение означает, что для описания анизотропии рассеяния в задаче используется индикатриса рассеяния в P_L приближении, содержащем $L + 1$ гармоник индикатрисы рассеяния, включая нулевую, а в качестве квадратуры используется S_N квадратура порядка N .

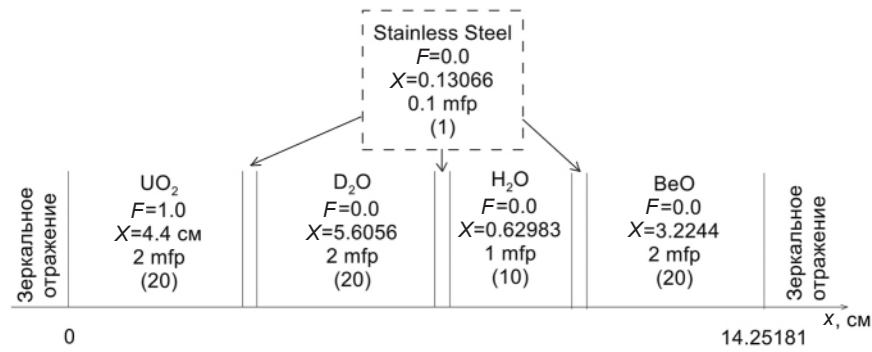
В табл. 1 и 2 приведено число внешних итераций по области термализации нейтронов при решении двух задач в 1D плоской и цилиндрической геометриях. В качестве S_n квадратуры при проведении расчетов для табл. 2–7 использовалась LCT_n триангулярная квадратура Лежандра—Чебышёва (см. [35], [36]).

В табл. 3 и 4 представлены результаты использования КР₁-схемы ускорения внешних итераций при решении задач радиационной защиты в 2D и 3D геометриях.

В табл. 5 представлены расчетные результаты для трех тестовых задач в x, y, z геометрии из [22], представляющих собой пластины с постоянным изотропным источником в 1-й группе и условием зеркального отражения на внешних границах. Пластина 1 состоит из графита и имеет размеры $63.594 \times 2.1198 \times 2.1198$ см ($30 \times 1 \times 1$ длин свободного пробега (mfp)). Пластина 2 состоит из тяжелой воды (D₂O) с размерами $84.084 \times 2.8028 \times 2.8028$ см ($30 \times 1 \times 1$ mfp). Пластина 3 имеет размеры $62.252 \times 2.8028 \times 2.8028$ см. По переменной x она состоит из трех слоев: железо (Fe), тяжелая вода (D₂O) и графит, размерами: 13.026 (10 mfp), 28.028 (10 mfp) и 21.198 см (10 mfp) соответственно. Для пластин 1–3 использовалась пространственная сетка из $30 \times 2 \times 2$ интервалов. Для пластины 3

Таблица 1. Число внешних итераций по области термализации нейтронов в зависимости от выбора критерия сходимости итераций при расчете в $P_5 S_{16}$ приближении топливной ячейки в плоской геометрии из [20] (фиг. 1) на пространственной сетке из 73 ячеек при точности сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-3}$, $\epsilon_{upsc}^{flux} = 10^{-3}$ соответственно. В скобках приведено число внешних итераций и расчетное время без ускорения внешних итераций

Программа	Библиотека сечений	Число групп в области термализации	Разностная схема	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
PO3-6.7	SCALE-6.1.2, v7.0-238n	90	AWDD	8 (43)	2.03E-2 (5.76E-2)	5	1.69E-2
			AWLD	8 (43)	3.80E-2 (1.31E-1)	5	2.99E-2
PO3-6.7	SCALE-6.1.2, V7-200N	34	AWDD	8 (41)	7.55E-3 (9.63E-3)	6	7.03E-3
			AWLD	8 (41)	1.46E-2 (3.26E-2)	6	1.38E-2
PO3-6.7	SCALE-6.1.2, V7-27N	13	AWDD	9 (28)	1.30E-3 (2.34E-3)	7	1.30E-3
			AWLD	9 (28)	2.86E-3 (4.69E-3)	7	1.56E-3



Фиг. 1. Топливная ячейка в плоской геометрии из [20]. Здесь F — внутренний изотропный источник; X — размер зоны в см, а также в длинах свободного пробега (mfp); в скобках приведено число пространственных интервалов в зоне.

Таблица 2. Число внешних итераций по области термализации нейтронов в зависимости от выбора критерия сходимости итераций при расчете в P_5S_{16} приближении радиационной защиты РУ ВВЭР-1000 в 1D цилиндрической геометрии на пространственной сетке из 213 ячеек при точности сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-3}$, $\epsilon_{upsc}^{flux} = 10^{-3}$ соответственно. В скобках приведено число внешних итераций и расчетное время без ускорения внешних итераций

Программа	Библиотека сечений	Число групп в области термализации	Разностная схема	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
PO3-6.7	BGL1000, 47n+20g	5	AWDD	5 (54)	2.553E-2 (1.14E-1)	5	2.52E-2
			AWLD	5 (54)	4.22E-2 (1.13E-1)	5	4.19E-2
PO3-6.7	SCALE-6.1.2, V7-200N47G	34	AWDD	16 (139)	6.73E-1 (4.91)	11	5.32E-1
			AWLD	16 (139)	6.17E-1 (2.78)	11	5.31E-1
PO3-6.7	SCALE-6.1.2, V7-27N19G	13	AWDD	9 (105)	3.49E-2 (3.32E-1)	7	3.07E-3
			AWLD	9 (105)	5.65E-2 (4.78E-1)	7	4.87E-2

для каждой из зон по переменной использовалась сетка из 10 интервалов; постоянный изотропный источник был помещен в зоне с железом.

Приведенное в табл. 5 число внешних итераций при использовании библиотеки V7-27N совпадает с числом итераций из [22]. Данные тестовые задачи могут быть решены также и в x, z геометрии. Результаты для этой геометрии приведены в табл. 6.

Для сравнения, в табл. 1, 2, 3, 5, 6 в скобках приведено также число внешних итераций и расчетное время без ускорения внешних итераций. Из-за медленной сходимости внешних итераций для варианта с тяжелой водой (табл. 5 и 6) в этом случае использовался более слабый критерий сходимости $\epsilon_{upsc}^{flux} = 10^{-3}$.

В табл. 7 представлены результаты использования KP_1 -схемы ускорения внешних итераций при решении подкритической задачи (ADS бенчмарк [37]), изображенной на фиг. 1, в зависимости от значения k_{eff} для решаемой задачи. Расчет проводился с заданным внутренним источником со спектром в области мишени, характерным для вторичных нейтронов, порождаемых пучком протонов с энер-

Таблица 3. Число внешних итераций и расчетные времена при расчете радиационной защиты РУ ВВЭР-1000 в 2D r, z геометрии на пространственной сетке 190×245 в P_3S_{16} приближении для нейтронов с $E > 3$ МэВ и P_3S_8 приближении для нейтронов с $E < 3$ МэВ и фотонов с точностью сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-3}$, $\epsilon_{upsc}^{flux} = 10^{-3}$. В скобках приведено число внешних итераций и расчетное время без ускорения внешних итераций

Программа	Библиотека сечений	Число групп в области термализации	Разностная схема	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
КАСКАД-С-4.0	BGL1000, 47n+20g	5	AWDD	11 (92)	2.55 (7.36)	10	2.45
			AWLD	8 (93)	4.54 (17.0)	7	4.38
КАСКАД-С-4.0	SCALE-6.1.2, V7-200N47G	34	AWDD	20 (237)	11.6 (68.6)	14	10.0
			AWLD	26 (239)	27.7 (144.4)	12	22.4
КАСКАД-С-4.0	SCALE-6.1.2, V7-27N19G	13	AWDD	11 (181)	2.41 (20.5)	10	2.12
			AWLD	11 (182)	5.76 (45.2)	10	5.29

Таблица 4. Число внешних итераций и расчетные времена при расчете радиационной защиты РУ ВВЭР-1000 в 3D r, θ, z геометрии на пространственной сетке $190 \times 120 \times 192$ по программе КАТРИН в P_3S_{16} приближении для нейтронов с $E > 3$ МэВ и P_3S_8 приближении для нейтронов с $E < 3$ МэВ и фотонов с точностью сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-3}$, $\epsilon_{upsc}^{flux} = 10^{-3}$

Программа	Библиотека сечений	Число групп в области термализации	Разностная схема	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
КАТРИН-3.5	BGL1000, 47n+20g	5	AWDD	12	9 час. 42 мин	12	9 час. 42 мин
			AWLD	16	17 час. 41 мин	15	18 час. 18 мин
КАТРИН-3.5	SCALE-6.1.2, V7-200N47G	34	AWDD	30	16 час. 48 мин	17	11 час. 12 мин
			AWLD	29	34 час.	18	28 час. 54 мин
КАТРИН-3.5	SCALE-6.1.2, V7-27N19G	13	AWDD	13	9 час. 37 мин	12	9 час. 23 мин
			AWLD	16	19 час. 10 мин	15	18 час. 39 мин

гией 1 ГэВ, падающим перпендикулярно на нижний торец мишени из свинца, с использованием 28 групповых констант, подготовленных программой CONSYST (см. [38]) без учета термализации нейтронов. Следует отметить, что так как в рассчитываемой композиции отсутствует замедлитель, число итераций без ускорения в этой задаче весьма велико для большинства групп. Данная задача решалась с $\epsilon^q = \epsilon_{aver}^q$, так как использование $\epsilon^q = \epsilon_{\beta}^q$ в данной задаче приводит к неустойчивому алго-

Таблица 5. Число внешних итераций по области термализации нейтронов в зависимости от выбора критерия сходимости итераций при расчете с использованием LD схемы в P_3S_8 приближении пластины в x, y, z геометрии из [22] на пространственной сетке из $30 \times 2 \times 2$ ячеек при точности сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-4}$, $\epsilon_{upsc}^{flux} = 10^{-4}$ соответственно. В скобках приведено число внешних итераций и расчетное время без ускорения внешних итераций (из-за медленной сходимости внешних итераций для варианта с тяжелой водой использовался критерий сходимости $\epsilon_{upsc}^{flux} = 10^{-3}$)

Программа	Библиотека сечений	Число групп в области термализации	Задача	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
КАТРИН-3.5	SCALE-6.1.2, v7.0-238n	90	Графит	12 (421)	8.08E-1 (18.4)	7	6.25E-1
			Тяжелая вода (D ₂ O)	7 (904)	4.86E-1 (27.6)	5	4.31E-1
			Fe + D ₂ O + графит	24 (410)	1.38 (18.2)	15	1.11
КАТРИН-3.5	SCALE-6.1.2, V7-200N	34	Графит	10 (381)	3.63E-1 (6.46)	7	3.24E-1
			D ₂ O	7 (900)	2.74E-1 (11.0)	6	2.72E-1
			Fe + D2O + графит	19 (388)	5.38E-1 (6.26)	13	4.69E-1
КАТРИН-3.5	SCALE-6.1.2, V7-27N	13	Графит	6 (277)	5.78E-2 (1.53)	6	5.71E-2
			D ₂ O	6 (848)	5.64E-2 (3.78)	6	6.30E-2
			Fe + D2O + графит	13 (262)	1.13E-1 (1.47)	11	1.07E-1

ритму в силу существенных граничных эффектов между зоной замедлителя и зонами с делящимися материалами.

При использовании для данной задачи 28 групповых констант программы CONSYST, учитывающих термализацию нейтронов, допустимо использовать $\epsilon^q = \epsilon_\beta^q$. В этом случае внешние итерации по источнику деления (группы 1–28) чередуются с внешними итерациями по области термализации нейтронов (группы 24–28).

8. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Как следует из приведенных численных результатов, а также из опыта решения практических задач, KP_1 -схема ускорения внешних итераций, согласованная с разностной аппроксимацией уравнения переноса, а также использующая спектральные функции ϵ_β^q , определяемые на основе оценки спектра собственных значений оператора $\hat{A}(0)$ (83) для материалов задачи, является устойчивой и обеспечивает существенный вычислительный выигрыш. В частности, время расчета радиационной защиты РУ ВВЭР-1000 в 3D r, ϑ, z геометрии с библиотекой V7-27N19G (табл. 4) при переходе от ϵ_{aver}^q к ϵ_β^q уменьшилось на 20%.

Использование своей спектральной функции ϵ_β^q для каждого материала задачи целесообразно при наличии протяженных геометрических зон с различными материалами, когда можно пренебречь взаимным влиянием спектров нейтронов в соседних зонах. В задачах радиационной защиты это требование обычно соблюдается. Если же задача состоит из плотно упакованных зон с различными ма-

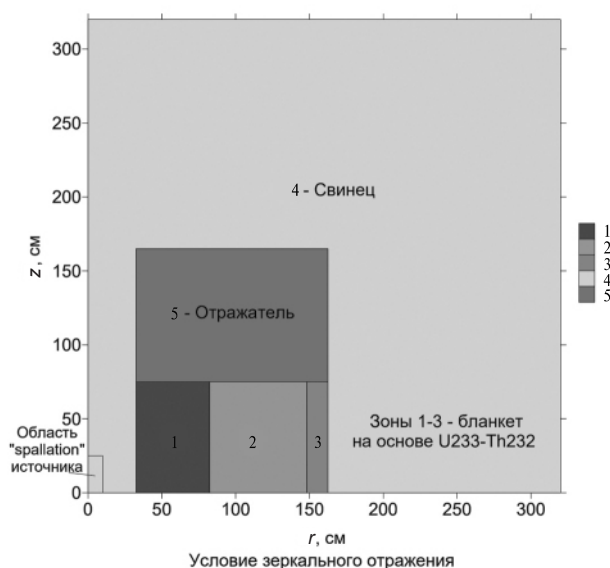
Таблица 6. Число внешних итераций по области термализации нейтронов в зависимости от выбора критерия сходимости итераций при расчете пластины в x, z геометрии из [22] с использованием LD схемы в P_3S_8 приближении на пространственной сетке из 30×2 ячеек при точности сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{upsc}^{dens} = 10^{-4}$, $\epsilon_{upsc}^{flux} = 10^{-4}$ соответственно. В скобках приведено число внешних итераций и расчетное время без ускорения внешних итераций (из-за медленной сходимости внешних итераций для варианта с тяжелой водой использовался критерий сходимости $\epsilon_{upsc}^{flux} = 10^{-3}$)

Программа	Библиотека сечений	Число групп в области термализации	Задача	Критерий сходимости итераций $\epsilon_{upsc}^{flux} = 10^{-3}$		Критерий сходимости итераций $\epsilon_{upsc}^{dens} = 10^{-3}$	
				Число итераций	Расчетное время, мин	Число итераций	Расчетное время, мин
КАСКАД-4.0	SCALE-6.1.2, v7.0-238n	90	Графит	12 (421)	8.08E-1 (6.47)	7	6.09E-1
			Тяжелая вода (D ₂ O)	7 (905)	4.91E-1 (9.96)	5	4.37E-1
			Fe + D ₂ O + графит	23 (481)	4.43E-1 (6.16)	15	3.51E-1
КАСКАД-4.0	SCALE-6.1.2, V7-200N	34	Графит	10 (381)	4.00E-1 (2.19)	7	3.48E-1
			D ₂ O	5 (900)	2.59E-3 (3.85)	5	2.61E-1
			Fe + D ₂ O + графит	19 (381)	1.70E-1 (2.10)	13	1.48E-1
КАСКАД-4.0	SCALE-6.1.2, V7-27N	13	Графит	6 (277)	7.35E-2 (5.14E-1)	6	7.13E-2
			D ₂ O	7 (848)	7.04E-2 (1.39)	6	6.95E-2
			Fe + D ₂ O + графит	11 (256)	3.60E-2 (4.70E-1)	8	3.20E-2

Таблица 7. Число внешних итераций и расчетное время (в круглых скобках) при расчете подкритического ADS бенчмарка (фиг. 2) в r, z геометрии по программе КАСКАД-С-4.0 в P_5S_8 приближении с пространственной r, z сеткой 51×46 , поточечным критерием сходимости внутренних и внешних итераций $\epsilon_{in} = 10^{-4}$ и $\epsilon_{fis}^{dens} = 10^{-3}$ соответственно

k_{eff} задачи	Разностная схема	Число внешних итераций и расчетное время, мин	
		Без ускорения внешних итераций	КР ₁
0.945	AWDD	66 (1.44)	5 (0.27)
	AWLD	67 (3.05)	6 (0.57)
0.965	AWDD	89 (1.90)	5 (0.29)
	AWLD	89 (3.90)	6 (0.57)
0.985	AWDD	146 (2.97)	9 (0.46)
	AWLD	147 (6.53)	9 (0.86)

териалами, когда длина свободного пробега нейтрона в зоне существенно больше ее размеров, целесообразно использовать спектральную функцию ϵ_{over}^q , усредненную по материалам задачи. К примерам таких задач можно отнести расчет активной зоны с отражателем на k_{eff} , в частности, расчет 2D C5G7MOX (см. [39]) бенчмарка и его 3D обобщений. Поэтому в расчетных программах РОЗ-6, КАСКАД-С и КАТРИН введен дополнительный признак, позволяющий пользователю вместо стандартно используемого набора спектральных функций ϵ_{β}^q выбрать для решаемой задачи использование ϵ_{aver}^q .



Фиг. 2. Геометрия ADS бенчмарка в r, z геометрии.

Использование интегрального по рассеянным вверх по энергии тепловым нейтронам, поточечного критерия сходимости (65) обеспечивает дополнительное уменьшение числа внешних итераций по области термализации нейтронов и уменьшение расчетных времен по сравнению с поточечным по групповым критерием сходимости (64) при сохранении достаточной точности определения потока тепловых нейтронов.

СПИСОК ЛИТЕРАТУРЫ

1. White J. E., Ingersoll D. T., Slater C. O., Roussin R. W. BUGLE-96: A Revised Multigroup Cross-Section Library derived from ENDF/B-VI for LWR and Pressure Vessel Dosimetry Applications // DLC-185, Radiation Safety Information Computational Center. 1996.
2. Bucholz J., Antonov S., Belousov S. BGL440 and BGL1000 Broad Group Neutron/Photon Cross Section Libraries Derived from ENDF/B-VI Nuclear Data // IAEA, INDC(BUL)-15. 1996.
3. SCALE 6.1: A Comprehensive Modeling and Simulation Suite for Nuclear Safety Analysis and Design // ORNL/TM-2005/39, Version 6.1, RSICC code package C00785. 2011.
4. Wieselquist W. A., Lefebvre R. A. and Jessee M. A., Eds. SCALE Code System // ORNL/TM-2005/39, Version 6.2.4, Oak Ridge National Laboratory, Oak Ridge, Tennessee. 2020.
5. Manturov G. N., Nikolaev M. N. and Tsiboulia A. M. BNAB-93 Group Data Library, Part1: Nuclear Data for the Calculations of Neutron and Photon Radiation Fields // Vienna, IAEA, INDC(CCP)-409. 1997.
6. Кощеев В. Н., Мантуров Г. Н., Николаев М. Н., Цибуля А. М. Библиотека групповых констант БНАБ-РФ для расчетов реакторов и защиты // Изв. вузов. Ядерная энергетика. 2014. № 3. С. 93–101.
7. Марчук Г. И., Лебедев В. И. Численные методы в теории переноса нейтронов. М.: Атомиздат, 1981.
8. Белл Д., Глестон С. Теория ядерных реакторов. пер. с англ., М.: Атомиздат, 1974.
9. Лебедев В. И. О КР-методе ускорения сходимости итераций при решении кинетического уравнения // Числ. методы решения задач матем. физ. М., 1966. С. 154–176.

10. *Kopp H. J.* Synthetic method solution of the transport equation // Nucl. Sci. and Eng. 1963. V. 17. N 1. P. 65–74.
11. *Alcouffe R. E.* Diffusion synthetic acceleration methods for the diamond-differenced discrete-ordinates equations // Nucl. Sci. and Eng. 1977. V. 64. N 2. P. 344–355.
12. *Adams M. L., Larsen E. W.* Fast iterative methods for discrete-ordinates particle transport calculations // Progress in Nuclear Energy. 2002. V. 40. Iss. 1. P. 3–159.
13. *Voloschenko A. M.* Completely consistent P_1 synthetic acceleration scheme for charged-particle transport calculations // Proc. 1996 Top. Meet. Radiation Protection and Shielding, April 21–25, 1996, No. Falmouth, USA. V. 1. P. 408–415.
14. *Voloschenko A. M.* Experience in the use of the consistent P_1 synthetic acceleration scheme for transport equation in 2D geometry // Proc. of Inter. Conf. on Math. and Comput., Reactor Phys., and Environmental Anal. in Nucl. Appl., 27–30 September, 1999, Madrid, Spain, V. 1. P. 104–113.
15. *Волощенко А. М.* KP_1 -схема ускорения внутренних итераций для уравнения переноса в двумерной геометрии, согласованная со взвешенной алмазной схемой // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. № 9. С. 1379.
16. *Voloschenko A. M.* Consistent P_1 synthetic acceleration scheme for transport equation in 3D geometries // Proc. of Intern. Conf. on Math. and Comput., Supercomputing, Reactor Phys., and Nucl. and Biological Applications. Avignon, France, September 12–15, 2005, paper 070.
17. *Волощенко А. М.* KP_1 -схема ускорения внутренних итераций для уравнения переноса в трехмерной геометрии, согласованная со взвешенной алмазной схемой // Ж. вычисл. матем. и матем. физ. 2009. Т. 49. № 2. С. 1–30.
18. *Adams B. T., Morel J. E.* A Two-Grid Acceleration Scheme for the Multigroup S_N Equations with Neutron Upscattering // Nucl. Sci. Eng. 1993. V. 115. P. 253.
19. *Averin A. V., Voloschenko A. M.* Consistent P_1 synthetic acceleration method for outer iterations // Transp. Theory and Stat. Phys. 1994. V. 23. N 5. P. 701–730.
20. *Adams B. T., Morel J. E.* An Acceleration Scheme for the Multigroup S_N Equations with Fission and Thermal Upscattering // Proc. of Joint Inter. Conf. on Math. Meth. and Supercomput. for Nucl. Appl., Saratoga Springs, New York, USA, 5–9 October, 1997. V. 1. P. 343.
21. CNCSN 2009: One, Two- and Three-Dimensional Coupled Neutral and Charged Particle Discrete Ordinates Parallel Multi-Threaded Code System // RSICC code package CCC-726, 2009.
22. *Evans T. M., Clarno K. T., Morel J. E.* A Transport Accelerating Scheme for Multigroup Discrete Ordinates with Upscattering // Nucl. Sci. Eng. 2010. V. 165. P. 292–304.
23. *Басс Л. П., Волощенко А. М. и Гермогенова Т. А.* Методы дискретных ординат в задачах о переносе излучения. М., ИПМ им. М. В. Келдыша АН СССР, 1986.
24. *Carlson B. G.* A method of characteristics and other improvements in solution methods for the transport equation // Nucl. Sci. Eng. 1976. V. 61. P. 408.
25. *Voloschenko A. M., Germogenova T. A.* Numerical Solution of the Time-Dependent Transport Equation with Pulsed Sources // Transp. Theory and Stat. Phys. 1994. V. 23. N 6. P. 845.
26. *Alcouffe R. E.* An Adaptive Weighted Diamond-Differencing Method for Three-Dimensional XYZ Geometry // TANS. 1993. V. 68A. P. 206.
27. *Larsen E. W.* Diffusion-synthetic acceleration methods for discrete-ordinates problems // Transp. Theory and Stat. Phys. 1984. V. 13. N 1&2. P. 107–126.
28. *Волощенко А. М., Воронков А. В., Сычугова Е. П.* Согласованная P_1SA схема ускорения внутренних и внешних итераций для уравнения переноса нейтронов и фотонов в одномерных геометриях в пакете РЕАКТОР // Препринт ИПМ им. М. В. Келдыша РАН. 1996. № 2.
29. *Adams M. L., Wareing T. A.* Diffusion-synthetic acceleration given anisotropic scattering, general quadratures, and multidimensions // Trans. Am. Nucl. Soc. 1993. V. 68A. P. 203–204.
30. *Марчук Г. И.* Методы расщепления. М.: Наука, 1988.
31. *Волощенко А. М.* KP_1 -схема ускорения внутренних итераций для уравнения переноса в трехмерной геометрии, согласованная с нодальными схемами. Основные уравнения и численные результаты // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 3. С. 441–464.

32. Волощенко А. М. KP_1 -схема ускорения внутренних итераций для уравнения переноса в трехмерной геометрии, согласованная с нодальными схемами II. Метод расщепления для решения P_1 -системы для ускоряющих поправок // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 5. С. 57–82.
33. Walters W. F. and O'Dell R. D. A Comparison of Linear Nodal, Linear Discontinuous and Diamond Schemes for Solving the Transport Equation in (x, y) Geometry // TANS. 1981. V. 39. P. 465.
34. Alcouffe R. E. A Robust Linear Discontinuous Method for the RZ S_N Transport Equation // Trans. Am. Nucl. Soc. 2003. V. 89. P. 363–366.
35. Lathrop K. D., Carlson B. G. Discrete ordinates angular quadrature of the neutron transport equation // LANL Report LA-3186. 1965. P. 1–48.
36. Волощенко А. М., Руссков А. А. Квадратурные формулы типа Гаусса для сферы с узлами, обладающими симметрией правильной призмы // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 7. С. 1095–1110.
37. Slessarev I., Tchistiakov A. IAEA-ADS Benchmark (Stage 1), Results and Preliminary Analysis // RCM-Meeting, Bologna, March 24–26, 1997.
38. Мантуров Г. Н., Николаев М. Н., Цибуля А. М. Программа подготовки констант CONSYST. Описание применения // Препринт ФЭИ-2828. Обнинск. 2000. С. 1–41.
39. Lewis E. E., Smith M. A., Tsoulfanidis N., Palmiotti G., Taiwo T. A., Blomquist R. N. Benchmark specification for Deterministic 2-D/3-D MOX fuel assembly transport calculations without spatial 20omogenization (C5G7 MOX), NEA/NSC/DOC(2001)4.

KP_1 -SCHEME FOR ACCELERATION OF UPSCATTER ITERATIONS OVER NEUTRON THERMALIZATION REGION AND IN THE FISSION SOURCE WHEN SOLVING A SUBCRITICAL BOUNDARY VALUE PROBLEM

A. M. Voloshchenko^{a,*}

^a 1125047, Moscow, Miusskaya pl., 4, Keldysh Institute of Applied Mathematics RAS, Russia

*e-mail: volosch@kiam.ru

Received 7 February, 2024

Revised 22 February, 2024

Accepted 3 April, 2024

Abstract. For the transport equation in three-dimensional r, ϑ, z geometry, a KP_1 -scheme for accelerating the convergence of upscatter iterations over the neutron thermalization region and the fission source when solving a subcritical boundary value problem is constructed, consistent with the Weighted Diamond Differentiation (WDD) scheme, and its generalization to the case of nodal Linear Discontinues (LD) and Linear Best (LB) schemes of the 3rd and 4th order of accuracy in spatial variables is considered. To solve the P_1 system for accelerating corrections, an algorithm based on the use of a cyclic splitting method was used, similar to the one used earlier in the construction of the KP_1 -scheme for accelerating the convergence of internal iterations. An algorithm for determining the energy dependence for accelerating corrections of the KP_1 -scheme for accelerating the convergence of upscatter iterations is studied. The question of choosing a criterion for the convergence of upscatter iterations is considered, and an integral criterion for the convergence of upscatter iterations in the field of neutron thermalization is proposed for thermal neutrons scattered upward in energy. A modification of the algorithm for the case of three-dimensional x, y, z geometry is considered. Numerical examples of using KP_1 -schemes for accelerating the convergence of upscatter iterations to solve typical neutron transport problems in three-dimensional geometry are given.

Keywords: KP_1 -scheme for accelerating upscatter iterations, choice of the spectral function and the criterion of convergence of upscatter iterations, transport equation.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЛИЯНИЯ ПОВЕРХНОСТНЫХ КВАНТОВЫХ ЭФФЕКТОВ НА ОПТИЧЕСКИЕ ХАРАКТЕРИСТИКИ НАНОЧАСТИЦ ЩЕЛОЧНЫХ И БЛАГОРОДНЫХ МЕТАЛЛОВ¹⁾

© 2024 г. Ю. А. Еремин^{1,*}, В. В. Лопушенко^{1,**}

¹119991 Москва, Ленинские горы, 1, стр. 52, МГУ имени М. В. Ломоносова, ВМК, Россия

*e-mail: eremin@cs.msu.ru

**e-mail: lopushnk@cs.msu.ru

Поступила в редакцию 21.03.2024 г.

Переработанный вариант 21.03.2024 г.

Принята к публикации 05.04.2024 г.

На основе метода дискретных источников построена математическая модель, позволяющая проводить сравнительный анализ влияния объемных и поверхностных квантовых эффектов на оптические свойства наночастиц щелочных и благородных металлов, располагающихся в плотной внешней среде. Установлено существенное отличие в проявлениях объемных и поверхностных квантовых эффектов в частицах щелочного металла. В частности, в таких частицах сдвиг плазмонного резонанса в случае объемного квантового эффекта происходит в область коротких волн (blue shift), в то время как поверхностный эффект приводит к сдвигу в длинноволновую область (red shift). Показано, что этот сдвиг существенно зависит от плотности окружающей среды и может достигать 50 нм в спектральной области. Библ. 26. Фиг. 4.

Ключевые слова: метод дискретных источников, математические модели, квантовая наноплазмоника, уравнения Максвелла, мезоскопические граничные условия, функции поверхностного отклика, щелочные и благородные металлы.

DOI: 10.31857/S0044466924070131, EDN: xiacgq

1. ВВЕДЕНИЕ

Нанолазмонные структуры используются в широком круге практических приложений, включая, например, лазерную фототермию для разрушения раковых опухолей посредством магнито-плазмонных частиц, нанолазмонные солнечные жидкости для улавливания, концентрации и трансформации солнечной энергии, нанобиосенсоры для ускоренного обнаружения вирусов в растворах и резонаторы плазмонного нанолазера. Все эти инновационные технологии предполагают наличие плазмонных наноструктур, в которых размеры металлических элементов могут оказаться порядка 10 нм [1]. В этом случае в металле начинают проявляться такие квантовые эффекты, как пространственная нелокальность [2], выброс электронов за поверхность металла [3], затухание Ландау и туннельный эффект [4]. Учет подобных квантовых эффектов необходим для правильного понимания принципов функционирования подобных структур.

С целью описания объемного эффекта пространственной нелокальности была разработана гидродинамическая теория Друде и ее обобщение — теория обобщенного нелокального оптического отклика (GNOR) [5]. Она позволяет анализировать поведение плазмонных наноструктур и объяснять возникающие оптические эффекты, которые могут существенно искажать картину, предсказанную классической теорией Максвелла. Например, с ее помощью описываются такие эффекты, как снижение амплитуды плазмонного резонанса (ПР) в благородных металлах и его сдвиг в область коротких длин волн (blue shift) [6, 7]. Теория GNOR хорошо описывает оптические свойства частиц благородных металлов (Au, Ag, Pt) [8] в то время как ее результаты, относящиеся к щелочным металлам

¹⁾Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках реализации программы Московского центра фундаментальной и прикладной математики по соглашению № 075-15-2022-284.

(Na, Al), вызывают скептическое отношение [9]. Дело в том, что дополнительное граничное условие, которое необходимо в рамках теории GNOR для однозначной разрешимости граничной задачи рассеяния, является слишком строгим, не допускающим возможность выхода электронов за пределы металла [10]. Это обстоятельство в значительной степени стимулировало поиск новой теории, описывающей квантовые эффекты, оставаясь в рамках классической теории Максвелла.

Значительного прогресса удалось добиться в рамках мезоскопической теории (МТ), появившейся в самое последнее время [9, 11]. Эта теория для рассмотрения поверхностных квантовых эффектов использует функции поверхностного отклика — параметры Фейбельмана [12]. МТ позволяет учитывать такие эффекты, как выброс электронов за пределы металла (spill out) и затухание Ландау [11]. МТ предполагает, что внутри плазмонного металла выполняются классические уравнения Максвелла, а на поверхности формулируются мезоскопические граничные условия, описывающие поведение индуцированных внешним полем свободных зарядов и токов [11]. Считается, что МТ является мостиком между чисто квантовым и классическим описанием происходящих явлений в наноплазмонике [13].

В данной работе мы используем Метод дискретных источников (МДИ) [14, 15]. Он представляет собой численно-аналитический поверхностно ориентированный метод. Приближенное решение в рамках МДИ строится на основе полей дискретных источников, локализованных внутри рассеивателя. Оно удовлетворяет всем условиям граничной задачи дифракции: уравнениям Максвелла вне и внутри объекта и условиям излучения на бесконечности, а амплитуды дискретных источников (ДИ) определяются из граничных условий, поставленных на поверхностях раздела сред с различными характеристиками. Таким образом, представление для полей является аналитической функцией в окрестности поверхности рассеивателя. Это обстоятельство дает возможность проводить апостериорную оценку погрешности полученного приближенного решения путем вычисления невязки полей на поверхности локального рассеивателя. Как было неоднократно показано, критерий невязки служит надежным инструментом оценки погрешности полученного приближенного решения [16].

В случае моделирования осесимметричных рассеивателей численная схема МДИ существенно упрощается [14]. При этом ДИ располагаются на оси симметрии и приближенное решение представляется в виде конечной линейной комбинации азимутальных гармоник с использованием так называемых распределенных мультиполей низшего порядка по отношению к азимутальной переменной φ . Разлагая внешнее возбуждение в ряд Фурье по φ , задача удовлетворения граничных условий на поверхности рассеивателя сводится к последовательному сшиванию гармоник Фурье на образующей поверхности вращения.

В данной работе на основе метода Дискретных источников проводится сравнительный анализ влияния объемных и поверхностных квантовых эффектов в рамках GNOR и МТ на оптические характеристики золотых и натриевых наночастиц. Установлены существенные отличия во влиянии объемных и поверхностных квантовых эффектов на оптические характеристики частиц щелочного металла. Отмечается существенное влияние параметров окружающей среды на оптические характеристики частиц щелочного металла при учете поверхностного квантового эффекта. В частности, в плотной среде сдвиг плазмонного резонанса в длинноволновую область (red shift) может достигать 50 нм.

2. ПОСТАНОВКА ГРАНИЧНОЙ ЗАДАЧИ ДИФРАКЦИИ С МЕЗОСКОПИЧЕСКИМИ ГРАНИЧНЫМИ УСЛОВИЯМИ

Рассмотрим математическую постановку граничной задачи дифракции поля электромагнитной плоской волны $\mathbf{E}_0, \mathbf{H}_0$ на однородной осесимметричной плазмонной частице. Пусть частица занимает область D_i с гладкой замкнутой поверхностью $\partial D_i \in C^{(1,\gamma)}$ (пространство Гёльдера), а внешнюю область обозначим через D_e . Будем предполагать, что все среды внутри и вне являются немагнитными, а зависимость от времени выбрана в виде $\exp(j\omega t)$.

Математическая постановка задачи дифракции поля плоской волны включает в себя классическую систему уравнений Максвелла для полного поля внутри $D_i \{\mathbf{E}_i, \mathbf{H}_i\}$ и рассеянного поля во внешней области $D_e \{\mathbf{E}_e, \mathbf{H}_e\}$

$$\operatorname{rot} \mathbf{H}_{i,e} = jk\varepsilon_{i,e} \mathbf{E}_{i,e}, \quad \operatorname{rot} \mathbf{E}_{i,e} = -jk\mathbf{H}_{i,e} \quad \text{в } D_{e,i}, \quad (1a)$$

мезоскопические граничные условия сопряжения для полных полей на поверхности ∂D_i

$$\mathbf{n}_i \times (\mathbf{E}_i(P) - \mathbf{E}_e(P) - \mathbf{E}_0(P)) = -d_\perp \mathbf{n}_i \times \nabla \{ \mathbf{n}_i \cdot (\mathbf{E}_i(P) - \mathbf{E}_e(P) - \mathbf{E}_0(P)) \}, \quad P \in \partial D_i,$$

$$\mathbf{n}_i \times (\mathbf{H}_i(P) - \mathbf{H}_e(P) - \mathbf{H}_0(P)) = -j\omega d_\parallel \{ \mathbf{n}_i \times [\mathbf{D}_i(P) - \mathbf{D}_e(P) - \mathbf{D}_0(P)] \} \times \mathbf{n}_i \quad (16)$$

и условия излучения Сильвера—Мюллера [17] на бесконечности

$$\lim_{r \rightarrow \infty} r \cdot \left(\sqrt{\epsilon_e} \mathbf{E}_e \times \frac{\mathbf{r}}{r} - \mathbf{H}_e \right) = 0; \quad r = |M|. \quad (1b)$$

Здесь $\epsilon_{i,e}$ — диэлектрические проницаемости сред в областях $D_{i,e}$, при этом $\text{Im } \epsilon_e = 0$, $\text{Im } \epsilon_i \leq 0$, $k = \frac{\omega}{c}$, $d_\perp$, $d_\parallel$ — параметры Фейбельмана, $\mathbf{n}_i$ — единичная нормаль к поверхности ∂D_i . Параметры Фейбельмана для плоского интерфейса $y = \text{const}$ определяются как

$$d_\perp(\omega) = \frac{\int \rho(x, \omega) x dx}{\int \rho(x, \omega) dx}; \quad d_\parallel(\omega) = \frac{\int \partial_x J_x(x, \omega) x dx}{\int \partial_x J_x(x, \omega) dx},$$

здесь $\rho(x, \omega)$ — плотность индуцированных поверхностных зарядов, а $J_y(x, \omega)$ — плотность поверхностных токов. Уместно отметить, что для случая границы раздела сред металл-диэлектрик параметр Фейбельмана $d_\parallel = 0$ [11], что является следствием условия непротекания нормальной компоненты тока проводимости через интерфейс металл-диэлектрик [18]. Параметр $d_\perp$ представляет собой комплексную функцию, зависящую от длины падающей волны. Его действительная часть описывает положение центра индукции зарядов по отношению к границе металл-диэлектрик, а мнимая отвечает за поглощение энергии вблизи поверхности (затухание Ландау) [19].

Будем полагать, что граничная задача в постановке (1) имеет единственное решение. Для случая однородной сферы аналитическое решение было получено в [20].

3. МЕТОД ДИСКРЕТНЫХ ИСТОЧНИКОВ

Как уже отмечалось ранее, в рамках МДИ представление для полей строится как конечная линейная комбинация мультиполей низшего порядка, распределенных вдоль оси симметрии OZ . Система уравнений Максвелла (1a) везде вне разрывов среды и условия излучения Сильвера—Мюллера (1b) на бесконечности удовлетворяются при этом в явном аналитическом виде. Соответствующие амплитуды ДИ определяются из мезоскопических граничных условий (16). Следует подчеркнуть, что МДИ весьма удобен для решения задачи рассеяния с мезоскопическими граничными условиями, так как поля вблизи поверхности частицы представляют собой аналитические функции и, следовательно, можно вычислять производные любого типа и порядка на ее поверхности.

При построении приближенного решения граничной задачи (1) ограничимся случаем p -поляризации падающей плоской волны, падающей наклонно по отношению к оси OZ . Именно в этом случае возникает наиболее заметный ПР [21]. Будем полагать, что электрический вектор $\mathbf{E}_0$ лежит в плоскости падения XZ , а угол между волновым вектором и осью вращения составляет $\pi - \theta_0$. В этом случае поле внешнего возбуждения принимает вид

$$\mathbf{E}_0^P = (\mathbf{e}_x \cos \theta_0 + \mathbf{e}_z \sin \theta_0) \cdot \chi(x, z), \quad \mathbf{H}_0^P = -\sqrt{\epsilon_e} \mathbf{e}_y \cdot \chi(x, z), \quad (2)$$

где

$$\xi(x, z) = \exp \{ -jk_e (x \sin \theta_0 - z \cos \theta_0) \},$$

$k_e = k\sqrt{\epsilon_e}$, а $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ — декартов базис.

Будем строить поля в областях $D_{i,e}$ на основе векторных потенциалов, индуцированных источниками, распределенными вдоль оси симметрии рассеивателя. Тогда векторные потенциалы в цилиндрической системе координат могут быть записаны как [14]

$$\mathbf{A}_{mn}^{1,\alpha} = Y_m^\alpha(\zeta, z_n^\alpha) \{ \mathbf{e}_\varphi \cos[(m+1)\varphi] - \mathbf{e}_\varphi \sin[(m+1)\varphi] \}, \quad \alpha = i, e,$$

$$\mathbf{A}_{mn}^{2,\alpha} = Y_m^\alpha(\zeta, z_n^\alpha) \{ \mathbf{e}_\varphi \sin[(m+1)\varphi] - \mathbf{e}_\varphi \cos[(m+1)\varphi] \}, \quad \mathbf{A}_n^{3,\alpha} = Y_0^\alpha(\zeta, z_n^\alpha) \mathbf{e}_z. \quad (3)$$

Здесь соответствующие функции имеют вид

$$Y_m^e(\zeta, z_n^e) = h_m^{(2)}(k_{e,s} R_{z_n^e}) \left(\frac{\rho}{R_{z_n^e}} \right)^m, \quad Y_m^i(\zeta, z_n^i) = j_m(k_i R_{z_n^i}) \left(\frac{\rho}{R_{z_n^i}} \right)^m,$$

где $h_m^{(2)}$ — сферические функции Ханкеля, удовлетворяющие условиям излучения, j_m — сферические функции Бесселя, $k_i = k\sqrt{\epsilon_i}$, $\zeta = (\zeta, z)$, $\rho^2 = x^2 + y^2$, $R_{z_n}^2 = \rho^2 + (z - z_n)^2$, $\{z_n^a\}_{n=1}^{N_a^m}$ — координаты ДИ, распределенных вдоль оси вращения OZ , число которых может различаться в зависимости от номера гармоники по φ .

Итак, представления для полей $\{\mathbf{E}_{i,e}, \mathbf{H}_{i,e}\}$ в случае p -поляризации приобретают вид

$$\mathbf{E}_\alpha^N = \sum_{m=0}^M \sum_{n=1}^{N_\alpha^m} \left\{ p_{mn}^\alpha \frac{j}{k\epsilon_\alpha} \text{rotrot} \mathbf{A}_{mn}^{1,\alpha} + q_{mn}^\alpha \frac{1}{\epsilon_\alpha} \text{rot} \mathbf{A}_{mn}^{2,\alpha} \right\} + \sum_{n=1}^{N_\alpha^0} r_n^\alpha \frac{j}{k\epsilon_\alpha} \text{rotrot} \mathbf{A}_n^{3,\alpha}, \quad \mathbf{H}_\alpha^N = \frac{j}{k} \text{rot} \mathbf{E}_\alpha^N, \quad \alpha = i, e. \quad (4)$$

Легко убедиться, что построенные поля (4) удовлетворяют всем условиям граничной задачи (1) за исключением мезоскопических граничных условий (16). Неизвестные амплитуды ДИ $\mathbf{p}_m^N = \{p_{mn}^{e,i}, q_{mn}^{e,i}, r_n^{e,i}\}$ определяются из этих граничных условий. При этом сведение поверхностной аппроксимации полей в условии (1в) к последовательности одномерных аппроксимаций на образующей поверхности вращения осуществляется разложением плоской волны $\chi(x, z)$ в (2) в ряд Фурье вида

$$\exp\{-jk_e \sin \theta_0\} = \sum_{m=0}^{-\infty} (2 - \delta_{0m}) (-j)^m j_m(k_e \rho \sin \theta_0) \cos(m\varphi). \quad (5)$$

Здесь δ_{0m} — символ Кронекера, j_m — цилиндрическая функция Бесселя. Из (5) видно, что сходимость ряда определяется значениями максимального радиуса рассеивателя ρ и угла падения θ_0 . После определения амплитуд ДИ легко вычислить все компоненты поля (4) вне и на поверхности рассеивателя.

Важной характеристикой в теории дифракции является диаграмма направленности рассеянного поля [17]. Используя асимптотику рассеянного поля (4) на бесконечности, можно записать компоненты θ, φ диаграммы направленности для p -поляризации как [14]

$$\begin{aligned} F_\varphi^p(\theta, \varphi) &= j \sum_{m=0}^M (j \sin \theta)^m \cos(m+1)\varphi \sum_{n=1}^{N_e^m} \{p_{mn}^e \cos \theta + q_{mn}^e\} \exp\{jk_e z_n^e \cos \theta\} - \\ &\quad - j \sin \theta \sum_{n=1}^{N_e^0} r_n^e \exp\{jk_e z_n^e \cos \theta\}, \\ F_\varphi^p(\theta, \varphi) &= -j \sum_{m=0}^M (j \sin \theta)^m \sin(m+1)\varphi \sum_{n=1}^{N_e^m} \{p_{mn}^e + q_{mn}^e \cos \theta\} \exp\{jk_e z_n^e \cos \theta\}. \end{aligned} \quad (6)$$

Основной наш интерес будет связан с исследованием поведения сечения экстинкции в спектральной области. В случае p -поляризации сечение экстинкции вычисляется через компоненты диаграммы в направлении распространения плоской волны и принимает следующий вид:

$$\sigma_{\text{ext}}^p = -\frac{4\pi}{k_e} \text{Im} F_\theta^p(\pi - \theta_0, \pi). \quad (7)$$

Сечение экстинкции (7) показывает, какую часть энергии плоской волны забирает рассеиватель на поглощение и рассеяние.

4. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Отметим, что наш основной интерес будет сосредоточен на анализе различий в оптических свойствах щелочных и благородных металлов натрия (Na) от (Au) [22, 23]. Несмотря на кажущуюся “экзотичность” натрия, исследованию его оптических свойств посвящена значительная часть публикаций в области квантовой наноплазмоники [20, 23, 24]. Для описания плазмонных свойств Na и Au мы будем использовать следующие квантовые параметры:

$$\begin{aligned} \text{Au: } \hbar\omega_p &= 9.02\text{eV}, \quad \hbar\gamma = 0.071\text{eV}, \quad v_F = 1.39 \cdot 10^{12} \mu\text{m/sec}, \quad D = 8.62 \cdot 10^8 \mu\text{m}^2/\text{sec}. \\ \text{Na: } \hbar\omega_p &= 5.89\text{eV}, \quad \hbar\gamma = 0.10\text{eV}, \quad v_F = 1.06 \cdot 10^{12} \mu\text{m/sec}, \quad D = 2.67 \cdot 10^8 \mu\text{m}^2/\text{sec}. \end{aligned}$$

Кроме того, зависящие от длины волны коэффициенты преломления для золота брались из [25], а для натрия — из [26].

Существенное отличие представленной в данной статье модели от подхода [15] состоит в том, что вместо экспериментальных значений функций поверхностного отклика (SRF) — параметров Фейбелмана, измеренных для соответствующих металлов в вакууме [10], нами использовались аналитические формулы, которые позволяют получать параметры плазмонных наночастиц, расположенных в различных внешних средах: спирт, пластик, стекло и т.д. [11, 22, 23]. Для металлов с малой работой выхода электронов за поверхность металла (jelly metals) таких как Na, K, Al, параметр Фейбелмана может быть вычислен, по формуле [11]

$$d_{\perp}(w) = j \frac{\xi}{\sqrt{\frac{\epsilon_i}{\epsilon_e}}}. \quad (8)$$

Здесь ξ — длина корреляции нелокальности в модели GNOR, $\xi^2(\omega) = \epsilon_b \frac{\beta^2 + D(\gamma + j\omega)}{\omega^2 - j\gamma\omega}$, $\epsilon_b = \epsilon_i + \frac{\omega_p^2}{\omega^2 - j\gamma\omega}$, ω_p — плазменная частота металла, $\beta^2 = \left(\frac{3}{5}\right) v_F^2$, v_F — скорость Ферми, γ — скорость затухания Друде, D — коэффициент диффузии электронов в металле. Заметим, что для благородных металлов с большой работой выхода электронов (Au, Ag, Pt) соответствующая формула имеет вид [22]

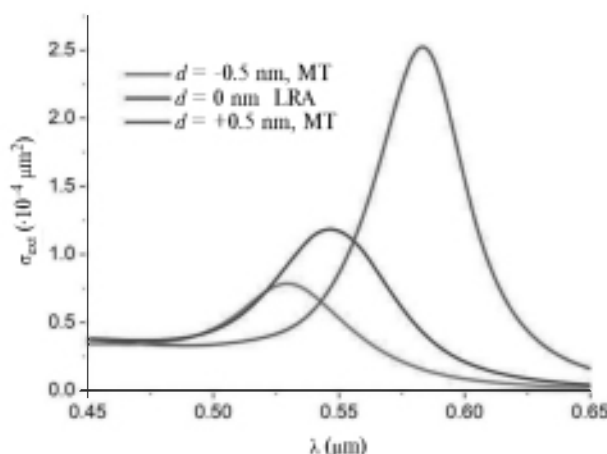
$$d_{\perp}(w) = -j \frac{\epsilon_i \epsilon_e}{\epsilon_i - \epsilon_e} \cdot \frac{\beta}{\omega_p \sqrt{\epsilon_b}} \left(\frac{\epsilon_b}{\epsilon_i} - 1 \right)^{\frac{3}{2}}. \quad (9)$$

Как показано в [22] (фиг. 1б), полученное соотношение хорошо аппроксимирует экспериментальные результаты для случая $\epsilon_e = 1$. Мы будем использовать формулы (8), (9) в наших непосредственных расчетах.

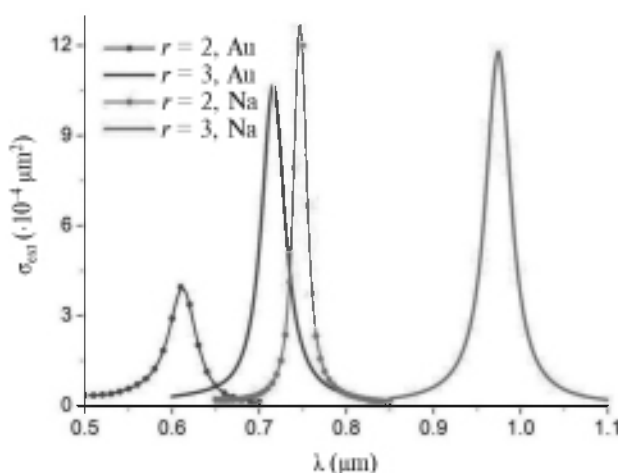
На фиг. 1 приведены результаты расчетов для случая золотой сферы диаметром 10 нм, расположенной в оптическом стекле SF5 ($\sqrt{\epsilon_i} = 1.67$). Представлены графики для локального случая (LRA) ($d_{\perp} = 0$) и двух значений $d_{\perp} = \pm 0.5$ нм. Видно, что при $d_{\perp} = -0.5$ нм, когда происходит “вдавливание” (spill in) электронного облака внутрь наночастицы, ПР сдвигается в область коротких волн (blue shift) и уменьшается по амплитуде (damping). Вдавливание электронного облака внутрь как бы снижает эффективный объем плазмонного облака, взаимодействующего с фотонами. Подобное поведение обычно характерно для объемного квантового эффекта (GNOR) [2, 15]. В случае же $d_{\perp} = -0.5$ нм, ПР сдвигается в область длинных волн (red shift) и увеличивается в амплитуде, что объясняется увеличением объема плазмонного облака.

Все дальнейшие рассуждения будут проходить на примере вытянутых эквиобъемных сфероидов (нанорисинков) диаметром 10 нм, расположенных в кристалле SiO₂ ($\sqrt{\epsilon_i} = 1.46$), при угле падения плоской волны $\theta_0 = 90$. Следует отметить, что подобные структуры часто используются в практических приложениях [24]. Фиг. 2 соответствует локальному случаю и демонстрирует сечения экстинкции золотых и натриевых сфероидов с соотношением осей $r = 2, 3$. Существенное отличие графиков для натриевых и золотых частиц состоит в величине сдвига ПР в инфракрасную область и противоположном изменении амплитуд. Это обстоятельство обусловлено поведением показателей преломления золота и натрия в частотной области [25, 26].

На фиг. 3а приведены результаты для золотых сфероидальных частиц с соотношением осей $r = 2, 3$ в трех различных вариантах: классическом (LRA), с учетом объемного эффекта на основе теории



Фиг. 1. Сечение экстинкции золотой сферы диаметром 10 нм, расположенной в стекле SF5, полученное в рамках моделей LRA и MT с параметрами Фейбельмана $d_{\pm} = \pm 0.5$ нм.

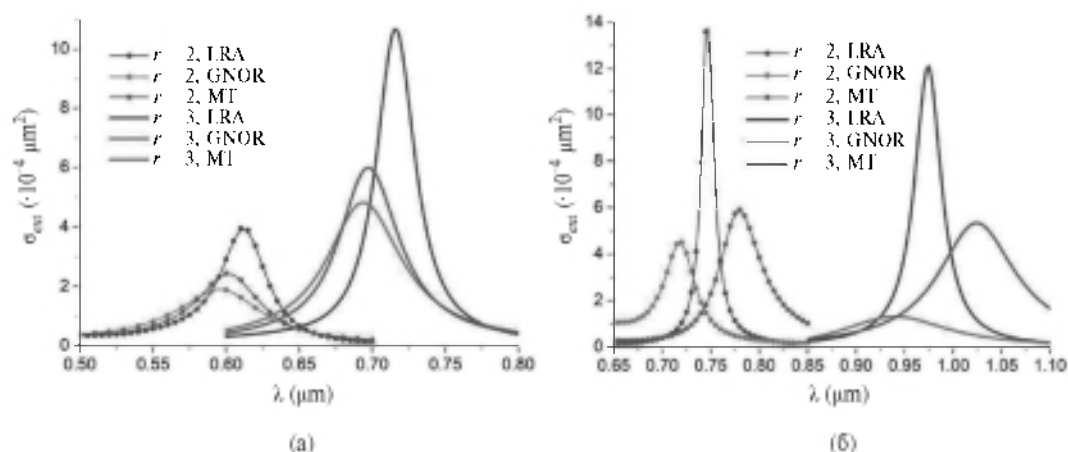


Фиг. 2. Сечение экстинкции золотых и натриевых сфероидов с различным соотношением осей в локальном случае (LRA). Эквивалентный диаметр $D = 10$ нм, угол падения $\theta_0 = 90^\circ$, внешняя среда — SiO_2 .

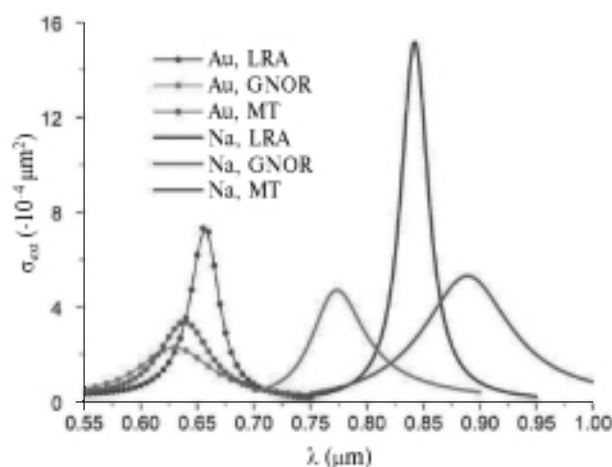
GNOR и с учетом поверхностного эффекта в рамках подхода MT. Отметим, что подробное описание теории GNOR можно найти в [15]. Видно, что в обоих случаях учета квантового эффекта как на основе GNOR, так и MT сдвиг ПР осуществляется в область коротких волн с одновременным уменьшением его амплитуды по сравнению с локальным случаем.

Однако при рассмотрении тех же конфигураций для натриевых частиц ситуация кардинально отличается (фиг. 3б). Так, при учете объемного эффекта (GNOR) ПР сдвинут в область коротких волн, а при учете поверхностного эффекта (MT) — в область длинных волн. Данное обстоятельство существенно отличает проявления поверхностного и объемного квантовых эффектов в щелочных металлах.

На фиг. 4 приведены результаты, соответствующие вытянутым сфероидам ($r = 2$) из различных металлов Au и Na, расположенным во внешней среде SF5 с более высоким показателем преломления. Из рисунка видно, что для золотых частиц результаты качественно мало отличаются от изображенных на фиг. 3а за исключением амплитуды, но для натриевых частиц сдвиги при объемном и поверхностном эффектах характеризуются гораздо большими значениями (см. фиг. 3б). Так, для поверхностного эффекта сдвиг достигает почти 50 нм.



Фиг. 3. Сравнение результатов, полученных в рамках моделей LRA, GNOR и MT для (а) золотых и (б) натриевых сфероидов с соотношением осей $r_2/r_3 = 2, 3$. Эквивалентный диаметр $D = 10$ нм, угол падения $\theta_i = 90^\circ$, внешняя среда — SiO_2 .



Фиг. 4. Значения сечения экстинкции, полученные в рамках моделей LRA, GNOR и MT для золотых и натриевых сфероидов с соотношением осей $r_2/r_3 = 2$, эквивалентным диаметром $D = 10$ нм, в стекле SF5. Угол падения $\theta_i = 90^\circ$.

5. ЗАКЛЮЧЕНИЕ

Сформулируем основные результаты статьи.

1. На основе метода дискретных источников построена математическая модель задачи дифракции для системы уравнений Максвелла с мезоскопическими граничными условиями, позволяющая проводить анализ плазмонных структур, состоящих из благородных и щелочных металлов и заключенных в различные среды.
2. Проведено исследование оптических свойств нанорисинков в частотной области. Установлено, что учет объемного эффекта на основе теории GNOR для щелочного металла приводит к сдвигу плазмонного резонанса в область коротких волн, в то время как поверхностный эффект сдвигает ПР в длинноволновую область.
3. Определено, что сдвиг ПР для натриевых частиц существенно зависит от плотности внешней среды и может достигать 50 нм в направлении длинных волн.

СПИСОК ЛИТЕРАТУРЫ

1. *Shi H., Zhu X., Zhang S., et al.* Plasmonic metal nanostructures with extremely small features: New effects, fabrication and applications // *Nanoscale Adv.* 2021. V. 3. 4349.
2. *David C., Garca de Abajo F.* Spatial Nonlocality in the Optical Response of Metal Nanoparticles // *J. Phys. Chem. C.* 2011. V. 15. P. 19470–19475.
3. *David C., Garca de Abajo F.* Surface plasmon dependence on the electron density profile at metal surfaces // *ACS Nano.* 2014. V. 8. N 9. 9558.
4. *Savage K.J., Hawkeye M.M., Esteban R., et al.* Revealing the Quantum Regime in Tunnelling Plasmonics // *Nature.* 2012. V. 491. P. 574–577.
5. *Mortensen N.A., Raza S., Wubs M., et al.* A generalized non-local optical response theory for plasmonic nanostructures // *Nat. Commun.* 2014. V. 5. 3809.
6. *Toscano G., Straubel J., Kwiatkowski A., et al.* Resonance shifts and spill-out effects in self-consistent hydrodynamic nanoplasmonics // *Nat. Commun.* 2015. V. 6. 7132.
7. *Tserkezis C., Yan W., Hsieh W., et al.* On the Origin of Nonlocal Damping in Plasmonic Monomers and Dimers // *Int. J. Mod. Phys. B.* 2017. V. 31. 1740005.
8. *Kupresak M., Zheng X., Vandenbosch G.A.E., Moshchalkov V.V.* Appropriate Nonlocal Hydrodynamic Models for the Characterization of Deep-Nanometer Scale Plasmonic Scatterers // *Adv. Theory Simul.* 2020. V. 3. 1900172.
9. *Mortensen N.A.* Mesoscopic electrodynamics at metal surfaces (Review) // *Nanophotonics* 2021. V. 10. N 10. P. 2563–2616.
10. *Echarri R.A., Goncalves P.A.D., Tserkezis C., et al.* Optical response of noble metal nanostructures: Quantum surface effects in crystallographic facets // *Optica.* 2021. V. 8. N 5. P. 710–721.
11. *Stamatopoulou P.E., Tserkezis C.* Finite-size and quantum effects in plasmonics: manifestations and theoretical modelling [Invited] // *Optical Materials Express.* 2022. V. 12. N. 5. P. 1869–1893.
12. *Feibelman P.J.* Surface electromagnetic fields // *Prog. Surf. Sci.* 1982. V. 12. P. 287–407.
13. *Yang F., Ding K.* Transformation optics approach to mesoscopic plasmonics // *Phys. Rev. B.* 2022. V. 105. L121410.
14. *Еремин Ю.А., Свешников А.Г.* Квазиклассические модели квантовой наноплазмоники на основе метода Дискретных источников (обзор) // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 4. С. 34–62.
15. *Еремин Ю.А., Лопушенко В.В.* Анализ влияния квантовых эффектов на оптические характеристики плазмонных наночастиц методом дискретных источников // *Ж. вычисл. матем. и матем. физ.* 2023. Т. 63. № 11. С. 1911–1921.
16. *Eremin Yu.A., Tsitsas N.L., Kouroublakis M., Fikioris G.* New scheme of the discrete sources method for two-dimensional scattering problems by penetrable obstacles // *J. Computat. Appl. Math.* 2023. V. 417. 114556.
17. *Колтон Д., Кресс Р.* Методы интегральных уравнений в теории рассеяния. М.: Мир, 1987.
18. *Raza S., Bozhevolnyi S.I., Wubs M., Mortensen N.A.* Nonlocal optical response in metallic nanostructures. Topical Review // *J. Phys. Condens. Matter.* 2015. V. 27. N183204.
19. *Mortensen N.A., Goncalves P.A.D., Shuklin F.A., et al.* Surface-response functions obtained from equilibrium electron-density profiles // *Nanophotonics.* 2021. V. 10. N 14. P. 3647–3657.

20. *Goncalves P. A. D., Christensen T., Rivera N., et al.* Plasmon–emitter interactions at the nanoscale // *Nat. Commun.* 2020. V. 11. P. 366.
21. *Еремин Ю. А., Свешников А. Г.* Математические модели задач нанооптики и биофотоники на основе метода дискретных источников // *Ж. вычисл. матем. и матем. физ.* 2007. Т. 47. № 2. С. 266.
22. *Eriksen M. H., Tserkezis C., Mortensen N. A., Cox J. D.* Nonlocal effects in atom-plasmon interactions // *ArXiv*. 2023. 2308.09134v.
23. *Bundgaard I. J., Hansen C. N., Stamatopoulou P. E., Tserkezis C.* Quantum-informed plasmonics for strong coupling: the role of electron spill-out // *ArXiv*. 2023. 2311.06030v.
24. *Zhang H., Huang C.* Optical response and spill-out effects of metal nanostructures with arbitrary shape // *J. Opt. Soc. Am. B*. 2021. V. 38. N 11. P. 3285–3291.
25. *Johnson P. B., Christy R. W.* Optical constants of the noble metals // *Phys. Rev. B*. 1972. V. 6. P. 4370–4379.
26. *Smith N. V.* Optical constants of sodium and potassium from 0.5 to 4.0 eV by split-beam ellipsometry // *Phs. Rev.* 1969. V. 183. P. 634–644.

COMPARATIVE ANALYSIS OF THE INFLUENCE OF SURFACE QUANTUM EFFECTS ON THE OPTICAL CHARACTERISTICS OF NANOPARTICLES OF ALKALI AND NOBLE METALS

Yu. A. Eremin^{a,*}, V. V. Lopushenko^{a,**}

^a 119991 Moscow, Lenin Hills 1, Bldg. 52, Lomonosov MSU, Faculty of computational mathematics and cybernetics, Россия

*e-mail: eremin@cs.msu.ru

**e-mail: lopushnk@cs.msu.ru

Received 21 March, 2024

Revised 21 March, 2024

Accepted 5 April, 2024

Abstract. Based on the discrete element method, a mathematical model has been built making it possible to carry out a comparative analysis of the influence of volume and surface quantum effects on the optical properties of alkali and noble metal nanoparticles located in a dense external environment. A significant difference in the manifestations of volume and surface quantum effects in alkali metal nanoparticles has been detected. In particular, in such particles plasmon resonance in the case of volume quantum effect shifts to the shortwave region (blue shift) while the surface effect leads to a shift to the longwave region (red shift). It is shown that this shift significantly depends on the density of the environment and can reach 50 nm in the spectral region.

Keywords: discrete source method, mathematical models, quantum nanoplasmonics, Maxwell's equations, mesoscopic boundary conditions, surface response functions, alkali and noble metals.

ЧИСЛЕННАЯ ДИАГНОСТИКА РАЗРУШЕНИЯ РЕШЕНИЯ ОДНОЙ ТЕПЛО-ЭЛЕКТРИЧЕСКОЙ МОДЕЛИ ПОЛУПРОВОДНИКА¹⁾

© 2024 г. М. О. Корпусов^{1,2,*}, Р. С. Шафир^{2,**}, А. К. Матвеева^{1,3,***}

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

² 117198 Москва, ул. Миклухо-Маклая, 6, РУДН, Россия

³ 115409 Москва, Каширское шоссе, 31, НИЯУ МИФИ, Россия

*e-mail: korpusov@gmail.com

**e-mail: romanshafir@mail.ru

***e-mail: matveeva2778@yandex.ru

Поступила в редакцию 10.03.2024 г.

Переработанный вариант 10.03.2024 г.

Принята к публикации 05.04.2024 г.

Предложена система уравнений с нелинейностью относительно потенциала электрического поля и температуры, описывающая процесс нагрева полупроводниковых элементов электрической платы, причем с течением времени возможно возникновение теплового и электрического “пробоев”. В работе рассматривается метод численной диагностики разрушения решения. В процессе численного исследования этой задачи использовался подход, основанный на сведении исходной системы уравнений в частных производных к дифференциально-алгебраической системе с последующим решением этой системы с помощью одностадийной схемы Розенброка с комплексным коэффициентом. Численная диагностика разрушения точного решения указанной системы уравнений основывалась на методике вычисления апостериорной асимптотически точной оценки погрешности, получаемой при вычислении приближенного решения на последовательно сгущающихся сетках. Получены численные оценки момента времени разрушения для различных начальных условий. Библ. 34. Фиг. 3.

Ключевые слова: нелинейные уравнения соболевского типа, разрушение, blow-up, локальная разрешимость, численная диагностика разрушения решения, оценки времени разрушения.

DOI: 10.31857/S0044466924070146, EDN: xhyctn

1. ВВЕДЕНИЕ

В настоящее время в микроэлектронике широко используются полупроводники, что обусловлено их уникальными физическими свойствами. В этой связи возникает насущная необходимость в исследовании как вопросов устойчивости полупроводников в электрическом поле, так и вопросов их неустойчивости в электрическом поле. Физический смысл разрушения — пробой полупроводника, что делает работу актуальной.

Настоящая работа посвящена численной диагностике разрушения решения. Рассмотрена следующая начально-краевая задача:

$$\begin{aligned} \frac{\varepsilon}{4\pi} \frac{\partial}{\partial t} \frac{\partial^2 \phi}{\partial x^2} + \sigma \frac{\partial^2 \phi}{\partial x^2} + \gamma \frac{\partial^2 \psi}{\partial x^2} &= 0, \quad x \in (a, b), t \in (0, T], \\ \varepsilon_0 \frac{\partial \psi}{\partial t} &= \frac{\partial^2 \psi}{\partial x^2} + q_0 \left| \frac{\partial \phi}{\partial x} \right|^p, \quad x \in (a, b), t \in (0, T], \\ \phi(x, 0) &= \phi_0(x), \quad \psi(x, 0) = \psi_0(x), \quad x \in (a, b), \\ \phi(a, t) &= \phi_a(t), \quad \phi(b, t) = \phi_b(t), \quad t \in [0, T], \\ \psi(a, t) &= \psi_a(t), \quad \psi(b, t) = \psi_b(t), \quad t \in [0, T]. \end{aligned} \quad (1)$$

¹⁾ Работа выполнена при финансовой поддержке РНФ (проект № 23-11-00056) и при поддержке Фонда развития теоретической физики и математики БАЗИС (проект № 22-2-2-17-1).

Система принадлежит теплоэлектрической теории. Функции $\phi = \phi(x, t)$ и $\psi = \psi(x, t)$, где $x \in \mathbb{R}^1$, $t \in [0, T]$, описывают распределение потенциала внутри полупроводника и его температуры соответственно.

Отметим, что система уравнений (1) относится к уравнениям соболевского типа (см. [1]). Исследованию уравнений соболевского типа посвящено большое количество работ. Так, в работах Г. А. Свиридюка, С. А. Загребинной, А. А. Замышляевой [3]– [5] были рассмотрены в общем виде и в виде примеров начально–краевые задачи для большого многообразия классов линейных и нелинейных уравнений соболевского типа.

С развитием компьютерной техники стали актуальны вопросы численного исследования задач теории разрушения решений. Численные методы предоставляют мощный инструмент для решения сложных математических задач, которые не могут быть решены аналитически. В данной работе мы используем метод, который основан на идеях апостериорной оценки точности. Метод подробно описан в работах [20–22]. Также отметим работу [23].

Работа продолжает исследования, начатые в [13]– [19].

2. СВЕДЕНИЕ СИСТЕМЫ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ К СИСТЕМЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

В данной работе мы используем численный подход к исследованию начально краевой задачи. Для численного решения дифференциально-алгебраической системы мы используем жесткий метод прямых, известный как SMOL. Этот метод был подробно описан в работах [24, 25]. Он помогает свести исходную систему с уравнением в частных производных к более простой задаче, требующей решения системы обыкновенных дифференциальных уравнений. Эта упрощенная система может быть эффективно решена с помощью одностадийной схемы Розенброка с комплексными коэффициентами CROS1 [26].

Рассмотрим следующую начально-краевую задачу (1). Сначала введем сетку X_N по пространственной переменной x с шагом $h = \frac{b-a}{N}$, где N — число интервалов сетки. Тогда $X_N = \{x_n, 1 \leq n \leq N+1 : x_n = a + (n-1)h\}$. После конечно-разностной аппроксимации производных по переменной x со вторым порядком точности получим следующую систему уравнений:

$$\begin{aligned} & \frac{\varepsilon}{4\pi} \frac{d}{dt} \left(\frac{\phi_{n+1} - 2\phi_n + \phi_{n-1}}{h^2} \right) + \sigma \left(\frac{\phi_{n+1} - 2\phi_n + \phi_{n-1}}{h^2} \right) + \\ & + \gamma \left(\frac{\psi_{n+1} - 2\psi_n + \psi_{n-1}}{h^2} \right) = 0, \quad t \in (0, T], n = \overline{2, N}, \\ \varepsilon_0 \frac{d\psi_n}{dt} &= \frac{\psi_{n+1} - 2\psi_n + \psi_{n-1}}{h^2} + q_0 \left| \frac{\phi_{n+1} - \phi_{n-1}}{2h} \right|^p, \quad t \in (0, T], n = \overline{2, N}, \\ \phi_1(t) &= \phi_a(t), \quad \phi_{N+1}(t) = \phi_b(t), \quad t \in [0, T], \\ \psi_1(t) &= \psi_a(t), \quad \psi_{N+1}(t) = \psi_b(t), \quad t \in [0, T], \\ \phi_n(0) &= \phi_0(x_n), \quad \psi_n(0) = \psi_0(x_n), \quad n = \overline{2, N}, \end{aligned}$$

где $\phi_n \equiv \phi_n(t) \equiv \phi(x_n, t)$, $\psi_n \equiv \psi_n(t) \equiv \psi(x_n, t)$ — неизвестные функции при $n = \overline{2, N}$.

Исключим из данной системы функции $\phi_1(t)$, $\phi_{N+1}(t)$, $\psi_1(t)$, $\psi_{N+1}(t)$, воспользовавшись граничными условиями. Тогда придем к следующей системе обыкновенных дифференциальных уравнений:

$$\begin{aligned} & \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (\phi_3 - 2\phi_2) = -\frac{\varepsilon}{4\pi h^2} \phi'_a(t) - \frac{\sigma}{h^2} (\phi_3 - 2\phi_2 + \phi_a(t)) - \\ & - \frac{\gamma}{h^2} (\psi_3 - 2\psi_2 + \psi_a(t)), \quad t \in (0, T], \\ & \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (\phi_{n+1} - 2\phi_n + \phi_{n-1}) = -\frac{\sigma}{h^2} (\phi_{n+1} - 2\phi_n + \phi_{n-1}) - \\ & - \frac{\gamma}{h^2} (\psi_{n+1} - 2\psi_n + \psi_{n-1}), \quad t \in (0, T], n = \overline{3, N-1}, \end{aligned} \tag{2}$$

$$\begin{aligned} \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (-2\phi_N + \phi_{N-1}) = -\frac{\varepsilon}{4\pi h^2} \phi'_b(t) - \frac{\sigma}{h^2} (\phi_b(t) - 2\phi_N + \phi_{N-1}) - \\ - \frac{\gamma}{h^2} (\psi_b(t) - 2\psi_N + \psi_{N-1}), \quad t \in (0, T], \end{aligned}$$

$$\varepsilon_0 \frac{d\psi_2}{dt} = \frac{1}{h^2} (\psi_3 - 2\psi_2 + \psi_a(t)) + q_0 \left| \frac{\phi_3 - \phi_a(t)}{2h} \right|^p, \quad t \in (0, T],$$

$$\begin{aligned} \varepsilon_0 \frac{d\psi_n}{dt} = \frac{1}{h^2} (\psi_{n+1} - 2\psi_n + \psi_{n-1}) + \\ + q_0 \left| \frac{\phi_{n+1} - \phi_{n-1}}{2h} \right|^p, \quad t \in (0, T], n = \overline{3, N-1}, \end{aligned}$$

$$\varepsilon_0 \frac{d\psi_N}{dt} = \frac{1}{h^2} (\psi_b(t) - 2\psi_N + \psi_{N-1}) + q_0 \left| \frac{\phi_b(t) - \phi_{N-1}}{2h} \right|^p, \quad t \in (0, T],$$

$$\phi_n(0) = \phi_0(x_n), \quad \psi_n(0) = \psi_0(x_n), \quad n = \overline{2, N}.$$

Введем вектор неизвестных $\mathbf{y} = (\phi_2, \dots, \phi_N, \psi_2, \dots, \psi_N)$. Тогда $\phi_n = y_{n-1}$, $\psi_n = y_{(N-1)+(n-1)}$ при $n = \overline{2, N}$.

Перепишем систему (2), заменив функции ϕ_n , ψ_n через компоненты вектора $\mathbf{y}$:

$$\begin{aligned} \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (y_2 - 2y_1) = -\frac{\varepsilon}{4\pi h^2} \phi'_a(t) - \frac{\sigma}{h^2} (y_2 - 2y_1 + \phi_a(t)) - \\ - \frac{\gamma}{h^2} (y_{N+1} - 2y_N + \psi_a(t)), \quad t \in (0, T], \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (y_n - 2y_{n-1} + y_{n-2}) = -\frac{\sigma}{h^2} (y_n - 2y_{n-1} + y_{n-2}) - \\ - \frac{\gamma}{h^2} (y_{(N-1)+n} - 2y_{(N-1)+(n-1)} + y_{(N-1)+(n-2)}), \quad t \in (0, T], n = \overline{3, N-1}, \end{aligned}$$

$$\begin{aligned} \frac{\varepsilon}{4\pi h^2} \frac{d}{dt} (-2y_{N-1} + y_{N-2}) = -\frac{\varepsilon}{4\pi h^2} \phi'_b(t) - \frac{\sigma}{h^2} (\phi_b(t) - 2y_{N-1} + y_{N-2}) - \\ - \frac{\gamma}{h^2} (\psi_b(t) - 2y_{2(N-1)} + y_{2(N-1)-1}), \quad t \in (0, T], \end{aligned}$$

$$\varepsilon_0 \frac{dy_N}{dt} = \frac{1}{h^2} (y_{N+1} - 2y_N + \psi_a(t)) + q_0 \left| \frac{y_2 - \phi_a(t)}{2h} \right|^p, \quad t \in (0, T],$$

$$\begin{aligned} \varepsilon_0 \frac{dy_{(N-1)+(n-1)}}{dt} = \frac{1}{h^2} (y_{(N-1)+n} - 2y_{(N-1)+(n-1)} + y_{(N-1)+(n-2)}) + \\ + q_0 \left| \frac{y_n - y_{n-2}}{2h} \right|^p, \quad t \in (0, T], n = \overline{3, N-1}, \end{aligned}$$

$$\varepsilon_0 \frac{dy_{2(N-1)}}{dt} = \frac{1}{h^2} (\psi_b(t) - 2y_{2(N-1)} + y_{2(N-1)-1}) + q_0 \left| \frac{\phi_b(t) - y_{N-2}}{2h} \right|^p, \quad t \in (0, T],$$

$$y_n(0) = \phi_0(x_n) \quad \text{при } n = \overline{1, N-1}, \quad y_n(0) = \psi_0(x_n) \quad \text{при } n = \overline{N, 2(N-1)}.$$

Систему (3) можно записать в векторном виде:

$$\begin{aligned} \mathbf{D} \frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}), \\ \mathbf{y}(0) = \mathbf{y}_0. \end{aligned} \quad (4)$$

Матричная функция $\mathbf{D}$ имеет следующую структуру :

$$\begin{aligned} D_{k,k-1} &= \frac{\varepsilon}{4\pi h^2}, \quad \text{если } k = \overline{2, N-1}, \\ D_{k,k} &= \begin{cases} -2\frac{\varepsilon}{4\pi h^2}, & \text{если } k = \overline{1, N-1}, \\ \varepsilon_0, & \text{если } k = \overline{N, 2(N-1)}, \end{cases} \\ D_{k,k+1} &= \frac{\varepsilon}{4\pi h^2}, \quad \text{если } k = \overline{1, N-2}, \end{aligned}$$

остальные элементы матрицы $\mathbf{D}$ равны нулю.

Далее выпишем компоненты вектора $\mathbf{f} = (f_1 \ f_2 \ f_3 \ \dots \ f_{2N-2})^T$:

$$\begin{aligned} f_1 &= -\frac{\varepsilon}{4\pi h^2} \phi'_a(t) - \frac{\sigma}{h^2} (y_2 - 2y_1 + \phi_a(t)) - \frac{\gamma}{h^2} (y_{N+1} - 2y_N + \psi_a(t)), \\ f_k &= -\frac{\sigma}{h^2} (y_{k+1} - 2y_k + y_{k-1}) - \\ &\quad - \frac{\gamma}{h^2} (y_{(N-1)+(k+1)} - 2y_{(N-1)+k} + y_{(N-1)+(k-1)}), \quad k = \overline{2, N-2}, \\ f_{N-1} &= -\frac{\varepsilon}{4\pi h^2} \phi'_b(t) - \frac{\sigma}{h^2} (\phi_b(t) - 2y_{N-1} + y_{N-2}) - \\ &\quad - \frac{\gamma}{h^2} (\psi_b(t) - 2y_{2(N-1)} + y_{2(N-1)-1}), \\ f_N &= \frac{1}{h^2} (y_{N+1} - 2y_N + \psi_a(t)) + q_0 \left| \frac{y_2 - \phi_a(t)}{2h} \right|^p, \\ f_k &= \frac{1}{h^2} (y_{k+1} - 2y_k + y_{k-1}) + q_0 \left| \frac{y_{k-(N-2)} - y_{k-(N-2)-2}}{2h} \right|^p, \quad k = \overline{N+1, 2N-3}, \\ f_{2(N-1)} &= \frac{1}{h^2} (\psi_b(t) - 2y_{2(N-1)} + y_{2(N-1)-1}) + q_0 \left| \frac{\phi_b(t) - y_{N-2}}{2h} \right|^p. \end{aligned}$$

Для численного решения системы (4) применим одностадийную схему Розенброка с комплексным коэффициентом (CROS1). Введем равномерную сетку T_M по переменной t с шагом $\tau = \frac{T-0}{M}$, где M — число интервалов. Тогда $T_M = \{t_k, 1 \leq k \leq M+1 : t_k = 0 + (k-1)\tau\}$. Теперь можно применить схему CROS1 для решения системы (4):

$$\begin{aligned} \mathbf{y}(t_{k+1}) &= \mathbf{y}(t_k) + \tau \operatorname{Re} \mathbf{w}, \\ \left[\mathbf{D} - \frac{1+i}{2} \mathbf{f}_y(\mathbf{y}(t_k)) \tau \right] \mathbf{w} &= \mathbf{f}(\mathbf{y}(t_k)). \end{aligned} \quad (5)$$

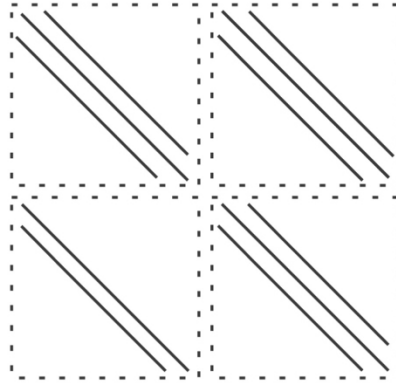
Также нам понадобится матрица Якоби $\mathbf{f}_y$, компоненты которой вычисляются по формуле: $(f_y)_{n,m} = \frac{\partial f_n}{\partial y_m}$. Выпишем ненулевые элементы матрицы Якоби:

$$\begin{aligned} (f_y)_{k,k-1} &= \begin{cases} -\frac{\sigma}{h^2}, & \text{если } k = \overline{2, N-1}, \\ \frac{1}{h^2}, & \text{если } k = \overline{N, 2N-2}, \end{cases} \\ (f_y)_{k,k} &= \begin{cases} \frac{2\sigma}{h^2}, & \text{если } k = \overline{1, N-1}, \\ -\frac{2}{h^2}, & \text{если } k = \overline{N, 2N-2}, \end{cases} \\ (f_y)_{k,k+1} &= \begin{cases} -\frac{\sigma}{h^2}, & \text{если } k = \overline{2, N-1}, \\ \frac{1}{h^2}, & \text{если } k = \overline{N, 2N-3}, \end{cases} \end{aligned}$$

$$\begin{aligned}
(f_y)_{k,N-1+(k-1)} &= -\frac{\gamma}{h^2}, \quad \text{если } k = \overline{2, N-2}, \\
(f_y)_{k,N-1+k} &= \frac{2\gamma}{h^2}, \quad \text{если } k = \overline{1, N-1}, \\
(f_y)_{k,N-1+(k+1)} &= -\frac{\gamma}{h^2}, \quad \text{если } k = \overline{2, N-2}, \\
(f_y)_{N,2} &= \frac{q_0 p}{(2h)^p} (y_2 - \phi_a(t)) |y_2 - \phi_a(t)|^{p-2}, \\
(f_y)_{k,k-(N-2)} &= \frac{q_0 p}{(2h)^p} (y_{k-(N-2)} - y_{k-(N-2)-2}) |y_{k-(N-2)} - y_{k-(N-2)-2}|^{p-2}, \\
&\quad \text{если } k = \overline{N+1, 2N-3}, \\
(f_y)_{k,k-(N-2)-2} &= -\frac{q_0 p}{(2h)^p} (y_{k-(N-2)} - y_{k-(N-2)-2}) |y_{k-(N-2)} - y_{k-(N-2)-2}|^{p-2}, \\
&\quad \text{если } k = \overline{N+1, 2N-3}, \\
(f_y)_{2(N-1),N-2} &= -\frac{q_0 p}{(2h)^p} (\phi_b(t) - y_{N-2}) |\phi_b(t) - y_{N-2}|^{p-2}.
\end{aligned}$$

Другие компоненты матрицы $\mathbf{f}_u$ в случае рассматриваемого уравнения равны нулю.

Таким образом, матрица системы (5) состоит из четырех блоков размерности $(N-1) \times (N-1)$ (структура матрицы представлена на фиг. 1). Этот факт дает возможность применить алгоритм решения СЛАУ, который найдет решение системы (5) за $O(N)$ операций.



Фиг. 1. Структура матрицы СЛАУ (5).

Отметим, что схема CROS1 имеет точность $O(\tau^2)$. А поскольку при аппроксимации пространственных производных были использованы формулы точности $O(h^2)$, предложенный алгоритм решения системы (4) имеет точность $O(\tau^2 + h^2)$.

Для оценки точности численного решения, а также диагностики разрушения решения можно эффективно применить метод сгущающихся сеток. Пусть базовая сетка по переменным (x, t) имеет следующий вид: $X_{N^{(0)}} \times T_{M^{(0)}} = \{(x_n, t_k), 1 \leq n \leq N^{(0)} + 1, 1 \leq k \leq M^{(0)} + 1\}$. Сгущать сетку мы будем в целое число раз r_t по переменной t и в r_x раз по переменной x . Выберем $r_t = r_x = 2$. Тогда каждая следующая сетка $X_{N_s} \times T_{M_s}$, где $N_s = r_x^{s-1} N^{(0)}$, $M_s = r_t^{s-1} M^{(0)}$, (s — номер сетки) будет иметь с базовой сеткой общие узлы. В узлах базовой сетки можно оценить эффективный порядок точности по формуле:

$$p_s^{\text{eff}} = \log_{r_t} \left(\frac{|u_{s-2}(x, t) - u_{s-3}(x, t)|}{|u_{s-1}(x, t) - u_{s-2}(x, t)|} \right), \quad (6)$$

где $u_s(x, t)$ — численное решение на сетке с номером s .

В точках (x, t) , в которых решение исходной задачи является достаточно гладким, имеет место сходимость:

$$p_s^{\text{eff}} \rightarrow p^{\text{theor}} = 2 \quad \text{при } s \rightarrow \infty.$$

Нарушение же данной сходимости будет соответствовать нарушению гладкости решения, что в нашем случае позволит локализовать по времени и пространству разрушение решения.

3. ЧИСЛЕННОЕ РЕШЕНИЕ

3.1. Численный эксперимент 1

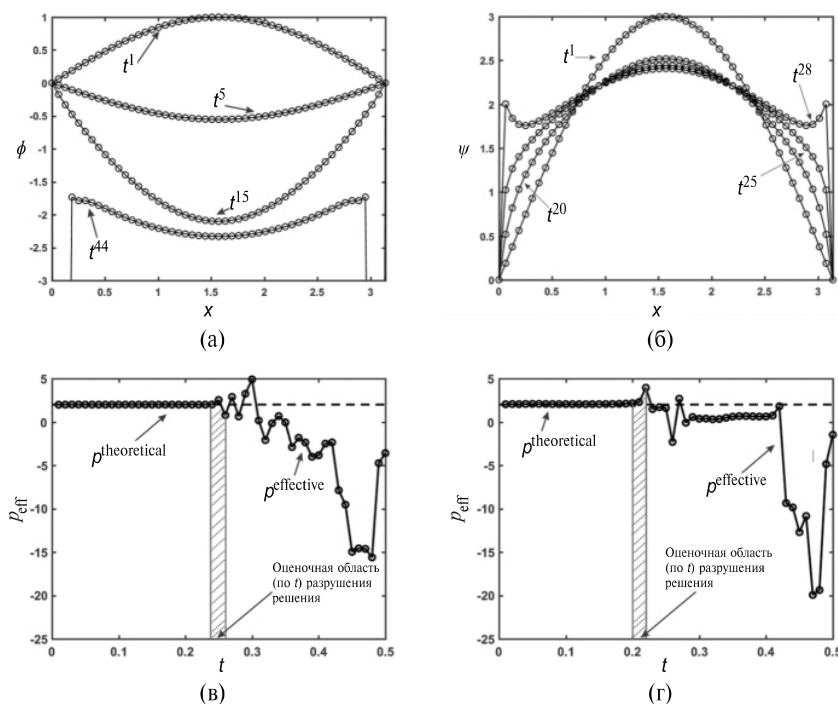
Возьмем для численного решения системы (1) следующий набор параметров: $\varepsilon = 1$, $\sigma = 1$, $\gamma = 1$, $\varepsilon_0 = 1$, $q_0 = 1$, $p = 3$.

Задачу будем решать на следующем промежутке: $a = 0$, $b = \pi$, $T = 0.5$.

Начальные и граничные условия выберем в виде: $\phi_0 = \sin x$, $\psi_0 = 3 \sin x$, $\phi_a = 0$, $\phi_b = 0$, $\psi_a = 0$, $\psi_b = 0$.

Число интервалов базовой сетки: $N^{(0)} = 50$, $M^{(0)} = 50$. Число последовательно используемых для вычислений сеток, включая начальную: $K = 6$. Получив приближенное численное решение на разных сетках, мы можем проверить сходимость эффективного порядка точности к теоретическому для каждого временного слоя по формуле (6).

Из графиков фиг. 2а,б видно, что классическое решение существует лишь до определенного момента времени: t_{20} , что соответствует $t = 0.20$. Также из анализа графика ясно, что потеря гладкости у решения происходит сначала вблизи границы. Поэтому расчет эффективных порядков точности в первую очередь необходимо проводить вблизи граничных точек. Результаты расчетов эффективных порядков точности представлены на графиках фиг. 2в,г. Анализ данных графиков позволяет оценить время разрушения: $0.20 \leq T_{\text{blow-up}} \leq 0.25$.



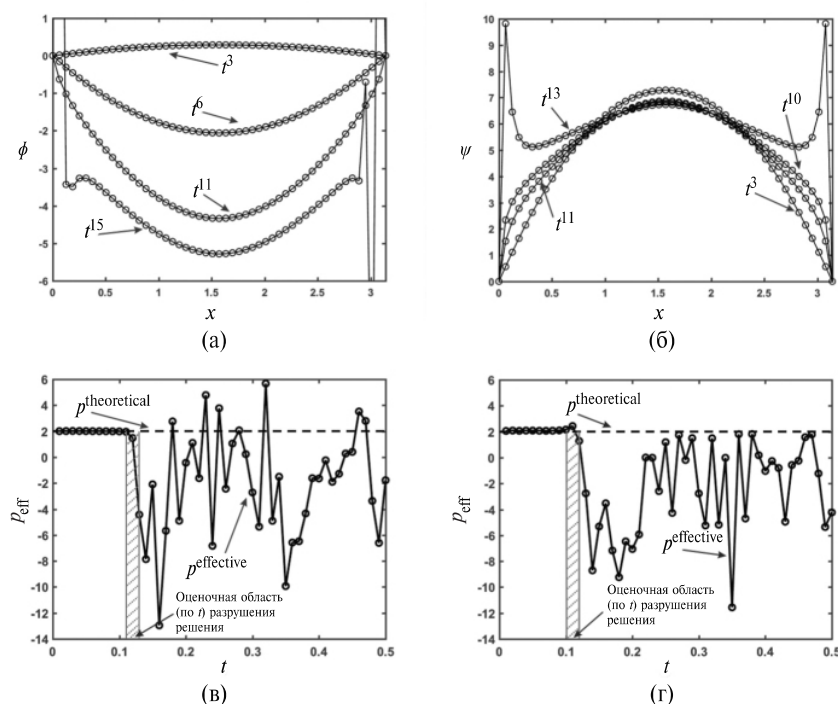
Фиг. 2. Результаты численного расчета для эксперимента 1: а — зависимость потенциала от координаты в различные моменты времени; б — зависимость температуры от координаты в различные моменты времени; в — зависимость эффективного порядка точности для потенциала от времени (вблизи границы); г — зависимость эффективного порядка точности для температуры от времени (вблизи границы).

3.2. Численный эксперимент 2

Выберем другие начальные условия для численного эксперимента: $\phi_0 = x(\pi - x)$, $\psi_0 = 3x(\pi - x)$.

Число интервалов базовой сетки и число сгущений (параметр K) оставим прежними: $N^{(0)} = 50$, $M^{(0)} = 50$, $K = 6$.

Из графиков фиг. 3а,б видно, что классическое решение существует лишь до определенного момента времени: t_{10} , что соответствует $t = 0.10$. Результаты расчетов эффективных порядков точности представлены на графиках фиг. 3в,г. Анализ данных графиков позволяет оценить время разрушения: $0.10 \leq T_{\text{blow-up}} \leq 0.11$.



Фиг. 3. Результаты численного расчета для эксперимента 2: а — зависимость потенциала от координаты в различные моменты времени; б — зависимость температуры от координаты в различные моменты времени; в — зависимость эффективного порядка точности для потенциала от времени (вблизи границы); г — зависимость эффективного порядка точности для температуры от времени (вблизи границы).

СПИСОК ЛИТЕРАТУРЫ

1. Al'shin A. B., Korpusov M. O., Sveshnikov A. G. Blow-up in nonlinear Sobolev type equations. De Gruyter Series in Nonlinear Analysis and Applications. 2011. V. 15. P. 648.
2. Гинзбург В. Л., Рухадзе А. А. Волны в магнитоактивной плазме. Современные проблемы физики. М.: Наука, 1970.
3. Свиридюк Г. А. К общей теории полугрупп операторов // Успехи матем. наук. 1994. Т. 49. N 4. С. 47–74.
4. Загребина С. А. Начально-конечная задача для уравнений соболевского типа с сильно (L,p)–радиальным оператором // Матем. заметки ЯГУ. 2012. Т. 19. N 2. С. 39–48.
5. Zamyshlyayeva A. A., Sviridyuk G. A. Nonclassical equations of mathematical physics. Linear Sobolev type equations of higher order // Вестн. Южно-Ур. ун-та. Сер. Матем. Механ. Физ. 2016. V. 8. N 4. P. 5–16.
6. Капитонов Б. В. Теория потенциала для уравнения малых колебаний вращающейся жидкости // Матем. сб. 1979. Т. 109(151). N 4(8). С. 607–628.

7. *Габов С. А., Свешников А. Г.* Линейные задачи теории нестационарных внутренних волн. М.: Наука, 1990.
8. *Габов С. А.* Новые задачи математической теории волн. М.: Физматлит, 1998.
9. *Плетнер Ю. Д.* Фундаментальные решения операторов типа Соболева и некоторые начально-краевые задачи // Ж. вычисл. матем. и матем. физ. 1992. Т. 32. N 12. С. 1885–1899.
10. *Похожаев С. И., Митидиери Э.* Априорные оценки и отсутствие решений нелинейных уравнений и неравенств в частных производных // Тр. МИАН. 2001. Т. 234. С. 3–383.
11. *Galakhov E. I.* Some nonexistence results for quasilinear elliptic problems // J. Math. Anal. Appl. 2000. V. 252. N 1. P. 256–277.
12. *Галахов Е. И., Салиева О. А.* Об отсутствии неотрицательных монотонных решений для некоторых коэрцитивных неравенств в полупространстве // СМФН. 2017. Т. 63. N 4. С. 573–585.
13. *Корпусов М. О.* Критические показатели мгновенного разрушения или локальной разрешимости нелинейных уравнений соболевского типа // Изв. РАН. Сер. матем. 2015. Т. 79. N 5. С. 103–162.
14. *Корпусов М. О.* О разрушении решений нелинейных уравнений типа уравнения Хохлова—Заболотской // ТМФ. 2018. Т. 194. N 3. С. 403–417.
15. *Korpusov M. O., Ovchinnikov A. V., Panin A. A.* Instantaneous blow-up versus local solvability of solutions to the Cauchy problem for the equation of a semiconductor in a magnetic field // Math. Methods Appl. Sci. 2018. V. 41. N 17. P. 8070–8099.
16. *Шафир Р. С.* Разрешимость и разрушение слабых решений задач Коши для $3 + 1$ -мерных уравнений дрейфовых волн в плазме // Матем. заметки. 2021. Т. 111. N 3. P. 459–475.
17. *Корпусов М. О., Шафир Р. С.* О разрушении решений задач Коши для нелинейных уравнений теории сегнетоэлектричества // ТМФ. 2022. Т. 212. N 3. P. 327–339.
18. *Корпусов М. О., Шафир Р. С.* О задачах Коши для нелинейных соболевских уравнений теории сегнетоэлектричества // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. N 1. P. 123–144.
19. *Корпусов М. О., Перлов А. Ю., Тимошенко А. В., Шафир Р. С.* О глобальной во времени разрешимости одной нелинейной системы уравнений тепло-электрической модели с квадратичной нелинейностью // ТМФ. 2023. Т. 217. N 2. P. 378–390.
20. *Альшина Е. А., Калиткин Н. Н., Корякин П. В.* Диагностика особенностей точного решения при расчетах с контролем точности // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. N 10. С. 1837–1847.
21. *Калиткин Н. Н., Альшин А. Б., Альшина Е. А., Рогов Б. В.* Вычисления на квазиравномерных сетках. М.: Физматлит, 2005.
22. *Al'shin A. B., Al'shina E. A.* Numerical diagnosis of blow-up of solutions of pseudoparabolic equations // J. of Math. Scie. 2008. T 148. N 1. С. 143–162.
23. *Лукьяненко Д. В., Панин А. А.* Разрушение решения уравнения стратификации объемного заряда в полупроводниках: численный анализ при сведении исходного уравнения к дифференциально-алгебраической системе // Выч. мет. программирование. 2016. Т. 17. N 4. P. 437–446.
24. *E. Hairer and G. Wanner* Solving of Ordinary Differential Equations // Stiff and Differential-Algebraic Problems. Springer, 2002.
25. *Калиткин Н. Н.* Численные методы решения жестких систем // Матем. моделирование. 1995. Т. 7. N 6. С. 111–117.
26. *Rosenbrock H. H.* Some general implicit processes for the numerical solution of differential equations // Computer Journal. 1963. Т. 5. N 4. С. 329–330.

27. Альшин А. Б., Альшина Е. А., Калиткин Н. Н., Корягина А. Б. Схемы Розенброка с комплексными коэффициентами для жестких и дифференциально-алгебраических систем // Ж. вычисл. матем. и матем. физ. 2006. Т. 46. N 8. С. 1392–1414.
28. Корпусов М. О., Плетнер Ю. Д., Свешников А. Г. О нестационарных волнах в средах с анизотропной дисперсией // Ж. вычисл. матем. и матем. физ. 1999. Т. 39. N 6. С. 1006–1022.
29. Кудашев В. Р., Михайловский А. Б., Шарапов С. Е. К нелинейной теории дрейфовой моды, индуцированной тороидальностью // Физ. плазмы. 1987. Т. 13. N 4. С. 417–421.
30. Каменец Ф. Ф., Лахин В. П., Михайловский А. Б. Нелинейные электронные градиентные волны // Физ. плазмы. 1987. Т. 13. N 4. С. 412–416.
31. Ситенко А. П., Сосенко П. П. О коротковолновой конвективной турбулентности и аномальной электронной теплопроводности плазмы // Физ. плазмы. 1987. Т. 13. N 4. С. 456–462.
32. Владимиров В. С. Уравнения математической физики. М.: Наука, 1988.
33. Панин А. А. О локальной разрешимости и разрушении решения абстрактного нелинейного интегрального уравнения Вольтерра // Матем. заметки. 2015. Т. 97. N 6. С. 884–903.
34. Демидович Б. П. Лекции по математической теории устойчивости. М.: Наука, 1967.

NUMERICAL DIAGNOSTICS OF THE DESTRUCTION OF A SOLUTION OF ONE THERMOELECTRIC SEMICONDUCTOR MODEL

M. O. Korpusov^{a,b,*}, R. S. Shafir^{b,**}, A. K. Matveyeva^{a,c,***}

^a 119991 Moscow, Lenin Hills, 1, Lomonosov MSU, Russia

^b 117198 Moscow, Miklukho-Maklay Str., 6, RUDN University, Russia

^c 115409 Moscow, Kashirskoye Highway, 31, National Research Nuclear University MEPhI, Russia

*e-mail: korpusov@gmail.com

**e-mail: romanshafir@mail.ru

***e-mail: matveeva2778@yandex.ru

Received 10 March, 2024

Revised 10 March, 2024

Accepted 5 April, 2024

Abstract. A system of equations with nonlinearity with respect to the potential of the electric field and temperature is proposed, describing the heating process of semiconductor elements of an electric board, and over time, thermal and electrical “breakdowns” may occur. The paper considers a method for the numerical diagnosis of solution destruction. In the process of numerical investigation of this problem, an approach was used based on the reduction of the initial system of partial differential equations to a differential algebraic system, followed by the solution of this system using a one-stage Rosenbrock scheme with a complex coefficient. Numerical diagnostics of the destruction of the exact solution of the specified system of equations was based on the method for calculating a posteriori asymptotically accurate error estimate obtained when calculating an approximate solution on successively thickening grids. Numerical estimates of the moment of destruction are obtained for various initial conditions.

Keywords: Sobolev type nonlinear equations, failure, blow-up, local solvability, numerical diagnostics of solution failure, numerical diagnostics of solution destruction, destruction time estimates.

ИСПОЛЬЗОВАНИЕ ВЕСОВЫХ СХЕМ МОНТЕ–КАРЛО ДЛЯ СЛАБОИОНИЗОВАННЫХ РАЗРЕЖЕННЫХ ТЕЧЕНИЙ ГАЗА¹⁾

© 2024 г. А. А. Шевырин^{1,*}, Е. А. Бондарь¹

¹630090 Новосибирск, ул. Институтская, 4/1, Институт теоретической и прикладной механики им. С. А. Христиановича СО РАН, Россия

*e-mail: shevr@itam.nsc.ru

Поступила в редакцию 23.11.2023 г.

Переработанный вариант 18.03.2024 г.

Принята к публикации 03.04.2024 г.

Представлено описание процедур метода прямого статистического моделирования Монте–Карло слабоионизованных течений, возникающих при обтекании возвращаемых космических аппаратов. Для реакций ионизации и рекомбинации приводятся выражения для модельной зависимости вероятности реакций от скоростей и энергий реагентов. Представлен алгоритм диссоциативной рекомбинации, вычислительная эффективность реализации которого достигается обходом моделирования взаимодействия электронов и тяжелых частиц. Изложен подход к построению весовой схемы для упругих столкновений и химических реакций, которая существенно повышает вычислительную эффективность расчетов. Представлен пример использования описанных численных моделей и процедур для исследования слабоионизованного течения около возвращаемой капсулы в типичных условиях входа орбитальных космических аппаратов. Результаты расчетов сравниваются с данными измерений параметров плазмы в летном эксперименте. Библ. 32. Фиг. 4. Табл. 1.

Ключевые слова: весовые схемы, высотная аэротермодинамика, ионизация, метод прямого статистического моделирования Монте–Карло, неравновесное течение газа, низкотемпературная плазма, рекомбинация заряда.

DOI: 10.31857/S0044466924070157, EDN: xhxkek

ВВЕДЕНИЕ

При разработке и эксплуатации пилотируемых возвращаемых космических аппаратов необходимо исследование их аэротермодинамики, а также предсказание участков траектории спуска с возможным экранированием радиосвязи. Такие комплексные численные исследования включают в себя расчет обтекания аппаратов. Верхний участок траектории характеризуется влиянием разреженности и неравновесности течения, поэтому необходимо применение кинетического подхода к описанию газа. Основным численным методом, реализующим такой подход, является метод прямого статистического моделирования Монте–Карло (ПСМ) (см. [1]).

В программном комплексе SMILE++ (см. [2], [3]), разработанном в ИТПМ СО РАН, имплементирован метод ПСМ с учетом основных физико-химических процессов: возбуждение внутренних степеней свободы молекул газа (см. [4]), химические реакции в газовой фазе (см. [5], [6]) и на поверхности аппарата (см. [7], [8]), ионизация (см. [9]) и тепловое излучение газа (см. [10]).

Решение, получаемое методом ПСМ, должно удовлетворять уравнениям сохранения. Уравнения сохранения массы, импульса и энергии ионизованного газа в наиболее общем виде приводятся в работе [11]. Анализ этих уравнений для течений около возвращаемых космических аппаратов проводился неоднократно (см., например, [12], [13]). Применительно к реализации моделей метода ПСМ вопрос рассмотрен в работе [14].

Для расчета электрически нейтральных ионизованных течений методом ПСМ для заряженных частиц используется упрощенный подход (см. [15]), в котором электроны не моделируются как частицы, а концентрация электронов оценивается как сумма концентраций ионов. Дальнейшее развитие этого подхода состоит в учете различных процессов, связанных с появлением, исчезновением

¹⁾Работа выполнена в рамках государственного задания ИТПМ СО РАН с использованием оборудования ЦКП “Механика” (ИТПМ СО РАН).

заряженных частиц (см. [16]), а также учетом электронного транспорта (см. [17]). Адекватное моделирование переноса энергии электронами может быть важно не только для определения характеристик плазмы около обтекаемого аппарата, но также и при расчете теплового излучения газа (см. [18]).

Ранее нами в [14] был предложен подход, состоящий в сопряженном моделировании нейтральной плазмы методом ПСМ и решении уравнения на энергию свободных электронов в континуальном приближении. Получены условия применимости этого подхода, и показано их выполнение для течений около спускаемых космических аппаратов при параметрах обтекания, когда ионизация становится заметной. В настоящей работе основной акцент сделан на описании используемых процедур. При этом анализ физических аспектов применения приближенных моделей метода ПСМ представлен в предыдущих работах и здесь детально рассматриваться не будет.

Целью работы является описание численных процедур, имплементированных в программном комплексе SMILE++, связанных с моделированием появления и исчезновения ионов в слабоионизованных течениях: ассоциативной ионизации и диссоциативной рекомбинации. Реализация обеих реакций основана на идеях модели полной столкновительной энергии (Total collision energy, TCE [1]) и температурной зависимости скорости реакции с коэффициентом в форме обобщенной аррениусовской зависимости $k(T) = AT^B \exp\left(\frac{-E_a}{k_B T}\right)$. Наиболее сложным аспектом в реализации являлась необходимость повышения числа моделирующих частиц для заряженных компонентов, использования весовых схем метода ПСМ.

Применение представленных численных моделей для расчета плотности плазмы около спускаемых аппаратов продемонстрировано на примере обтекания капсулы RAM C-II. Полученные результаты демонстрируют необходимость применения весовых схем в расчетах методом ПСМ и соответствие расчетных данных результатам летного эксперимента.

1. МОДЕЛИРОВАНИЕ ИОНИЗОВАННЫХ ТЕЧЕНИЙ МЕТОДОМ ПСМ

1.1. Модель ассоциативной ионизации

В ударном слое перед обтекаемым аппаратом вследствие диссоциации молекул образуются атомы азота и кислорода. При столкновении атомов могут образовываться соответствующие молекулярные ионы.

В модели TCE предполагается, что если распределение энергии реагентов равновесное, то константа скорости реакции $k(T)$ будет иметь температурную зависимость в форме Аррениуса с заданными параметрами. Константы скорости рассматриваемых реакций взяты из [19] и приведены в табл. 1.

Таблица 1. Параметры константы скоростей реакций $k(T) = AT^B \exp\left(\frac{-E_a}{k_B T}\right)$ по [19]

Реакция	Ионизация			Рекомбинация		
	$A, \text{ м}^3/\text{с}$	B	$E_a, \text{ Дж}$	$A, \text{ м}^3/\text{с}$	B	$E_a, \text{ Дж}$
$\text{N} + \text{O} \leftrightarrow \text{NO}^+ + e$	8.766×10^{18}	0	4.416×10^{-19}	1.321×10^{-9}	-1.187	0
$\text{N} + \text{N} \leftrightarrow \text{N}_2^+ + e$	3.387×10^{17}	0	9.346×10^{-19}	7.274×10^{-12}	-0.65	0

Для реакций ассоциативной ионизации $A + B \rightarrow AB^+ + e$ вероятность $P_r(E_c)$ для сталкивающихся частиц-реагентов как функция энергии E_c сталкивающейся пары в системе ее центра масс удовлетворяет выражению для константы скорости реакции

$$k(T) = \int_0^\infty P_r(E_c) f_c^{\text{eq}}(T, E_c) dE_c,$$
 (1)

где $f_c^{\text{eq}}(T, E_c)$ — плотность равновесного распределения $f_c(E_c)$, нормированная на единицу. Индекс c показывает, что распределение соответствует *сталкивающимся* частицам. Такая функция распределения зависит от сечения столкновения σ , которое для модели переменных твердых сфер (VHS) и обратно-степенного потенциала взаимодействия пропорционально степени модуля относительной скорости $g = |\mathbf{v}_B - \mathbf{v}_A|$: $\sigma \sim g^{-2\alpha}$, α — справочный параметр, принятый равным 0.25. В настоящей

работе рассматриваются только ассоциативная ионизация атомов, поэтому число внутренних степеней свободы частиц А и В равно нулю. Равновесная функция распределения может быть выражена как (см. [1])

$$f_c(E_c) = \left(\frac{E_c}{k_B T} \right)^{1-\alpha} \frac{\exp\left(\frac{-E_c}{k_B T}\right)}{\Gamma(2-\alpha)k_B T}. \quad (2)$$

В модели ТСЕ предполагается, что

$$P_r(E_c) = C_1(E_c - E_a)^{C_2} \left(1 - \frac{E_a}{E_c} \right)^{1-\alpha}, \quad (3)$$

где константы C_1 и C_2 могут быть выражены через параметры аррениусовской зависимости из решения уравнения (1) с функцией распределения (2):

$$C_1 = \frac{\varepsilon A}{\sigma_{\text{ref}} k_B^{B-0.5+\alpha} T_{\text{ref}}^\alpha} \sqrt{\frac{m_r \pi}{8k_B}} \frac{\Gamma(2-\alpha)}{\Gamma(B+\frac{3}{2})}, \quad C_2 = B + \alpha - \frac{1}{2}. \quad (4)$$

Для данных табл. 1 параметр $B = 0$ для всех трех реакций и сохранен в формулах для общности. В выражении (4), $\varepsilon = 1/2$, если реагенты — атомы одного сорта, и единице — если разного.

Реакция ассоциативной ионизации для *сталкивающихся* атомов реализуется с вероятностью $P_r(E_c)$, а с дополнительной вероятностью происходит столкновение без химической реакции. Реализация состоит в следующем. Создается ион AB^+ , имеющий скорость, равную скорости центра масс атомов-реагентов А и В. Импульсом свободного электрона мы пренебрегаем. Энергия атомов E_c уменьшается на величину теплового эффекта реакции, равного энергии активации E_a . Оставшаяся энергия распределяется между колебательными, вращательными модами образованного иона и свободным электроном пропорционально их степеням свободы: $E_c - E_a = E_i + E_e$. Полученное значение энергии электрона E_e накапливается в ячейке для подсчета величины источника энергии свободных электронов, который может быть использован при решении уравнения на энергию электронов (см. [14]).

1.2. Модель диссоциативной рекомбинации

В отличие от реализации реакций ассоциативной ионизации в методе ПСМ, рекомбинация ионов $AB^+ + e \rightarrow A + B$ требует учета взаимодействия с электронами, являющимися реагентом в реакции. В настоящей реализации столкновения с электронами не моделируются. Рекомбинация ионов и электронов выполнена в виде отдельной процедуры, следующей за столкновительной процедурой в ячейке (подробное обоснование модели приводится в работе [14]).

Число реакций $\nu = -\left(\frac{\partial n_i}{\partial t}\right)_r$, приходящихся на единицу объема и единицу времени, или частота реакций, полагается равным

$$\nu = n_i n_e \bar{\nu}_1 = n_i n_e \int_0^\infty \nu_1(E) f(E) dE, \quad (5)$$

где n_i и n_e — плотности ионов сорта i и электронов соответственно, а $\bar{\nu}_1$ — среднее значение частоты реакции, приходящееся на единичные концентрации реагентов. Таким образом, частота рекомбинации электрона с энергией E_e и иона с энергией возбуждения внутренних мод иона E_i предполагается зависящей от суммы $E = E_e + E_i$. Локальное распределение E описывается плотностью распределения $f(E)$, нормированной на единицу.

В случае локального равновесия $\bar{\nu}_1 = k(T)$, где $k(T) = AT^B \exp\left(\frac{-E_a}{k_B T}\right)$ — выражение Аррениуса. При температуре T в равновесном случае распределение $f_{E_e} \sim \sqrt{E_e} \exp\left(\frac{-E_e}{k_B T}\right)$, а распределение $f_{E_i} \sim E_i^{\zeta/2-1} \exp\left(\frac{-E_i}{k_B T}\right)$, где ζ — число внутренних степеней свободы иона. Тогда плотность функ-

ции распределения величины $E = E_e + E_i$

$$f_E^{\text{eq}} \sim \int_0^E (E - E_i)^{1/2} E_i^{\zeta/2-1} \exp\left(-\frac{E}{k_B T}\right) dE_i \sim E^{\zeta/2+1/2} \exp\left(-\frac{E}{k_B T}\right). \quad (6)$$

С учетом нормировки

$$f_E^{\text{eq}} = \frac{E^{\zeta/2+1/2}}{(k_B T)^{\zeta/2+3/2} \Gamma(\zeta/2 + 3/2)} \exp\left(-\frac{E}{k_B T}\right).$$

В настоящей работе используется модельная зависимость $v_1(E) = C_1 E^{C_2}$. Константы C_1 и C_2 определяются из условия равенства скорости реакции в равновесной плазме заданной температурной зависимости с параметрами (см. [19]), приведенными в табл. 1. Для этого зависимость Аррениуса $k(T)$ приравнивается интегралу из (5) с функцией $f(E) = f_E^{\text{eq}}$ (параметр $E_a = 0$ для всех трех реакций):

$$k(T) = AT^B = \int_0^\infty C_1 E^{C_2} f_E^{\text{eq}} dE = C_1 (k_B T)^{C_2} \frac{\Gamma\left(\frac{\zeta}{2} + C_2 + \frac{3}{2}\right)}{\Gamma\left(\frac{\zeta}{2} + \frac{3}{2}\right)},$$

$$C_1 = A \frac{\Gamma\left(\frac{\zeta}{2} + \frac{3}{2}\right)}{\Gamma\left(\frac{\zeta}{2} + B + \frac{3}{2}\right) k_B^B}, \quad C_2 = B.$$

Заметим, что аргумент гамма-функции должен быть положительным, и это является одним из ограничений модели ТСЕ на параметры скоростей реакций (см. [16]). Данные работы [19], используемые в настоящей работе, показывают, что такая ситуация может возникнуть при рекомбинации ионов кислорода $\text{O}_2^+ + e \rightarrow \text{O} + \text{O}$, поскольку значение параметра $B = -2.412$ для этой реакции достаточно малое. Для условий, рассматриваемых в настоящей работе, скорость образования O_2^+ на порядок меньше, чем N_2^+ и NO_2^+ (см. [20]). Поэтому реакции $\text{O} + \text{O} \leftrightarrow \text{O}_2^+ + e$ исключены из рассмотрения.

Приведем приближенную схему мажорантной частоты (см. [21]) для рекомбинации. Учитывая, что для всех рассматриваемых реакций $B < 0$ и функция $v_1(E) = C_1 E^B$ является убывающей, выберем $E_{\min}$ такое, что количеством реакций с $E < E_{\min}$ можно пренебречь. Для этого может быть использована приблизительная оценка, основанная на равновесной функции распределения из оценки интеграла (5):

$$\frac{\int_0^{E_{\min}} v_1(E) f_E^{\text{eq}}(E) dE}{\int_0^\infty v_1(E) f_E^{\text{eq}}(E) dE} < \varepsilon, \quad (7)$$

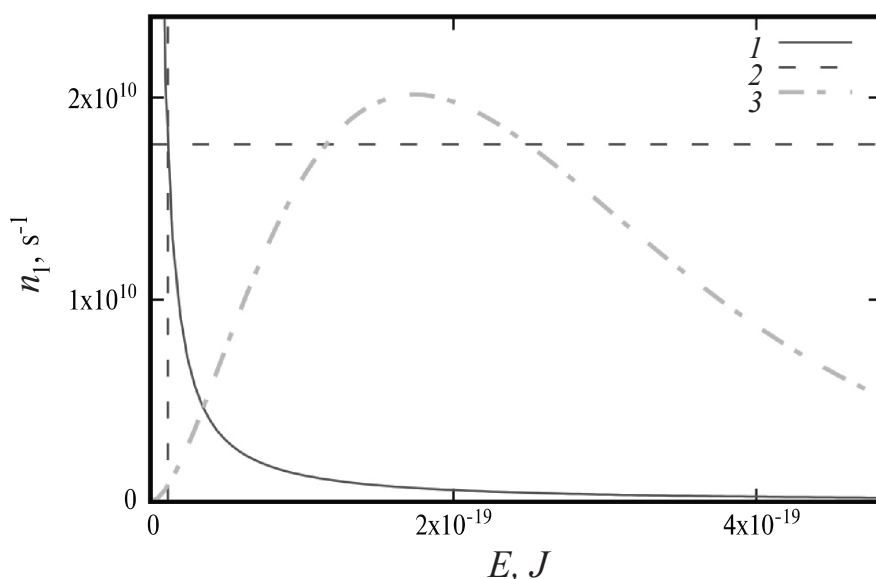
где ε — малое число. Для каждой реакции рекомбинации i -го иона из условия (7) определяется $E_{\min} = E_{\min,i}$ и рассчитывается значение функции $v_1(E = E_{\min,i}) = v_{1\max,i}$ для мажорирования подынтегрального выражения (5). На фиг. 1 на примере реакции $\text{NO}^+ + e \rightarrow \text{N} + \text{O}$ приведена зависимость $v_1(E) = C_1 E^B$ (линия 1). Величины $E = E_{\min,i}$ и $v_1 = v_{1\max,i}$ для значения $\varepsilon = 0.02$ и $T = 7000$ К показаны вертикальной и горизонтальной линиями 2. График функции $f^{\text{eq}}(E)$ (линия 3, выражение (6)) в относительных единицах приводится с иллюстративной целью.

Рассмотрим ячейку объемом V_c . Пусть ячейка содержит N_i молекулярных ионов сорта i . При этом число электронов определяется из условия локальной нейтральности плазмы, т.е. по числу расположенных в ячейке ионов. Проведение реакций в каждой ячейке на временном шаге Δt выполняется циклическим повторением следующих пунктов процедуры. Перед выполнением счетчик реакций $l = 0$.

1. Определяется число электронов в ячейке как сумма числа ионов $N_e = \sum N_i$.
2. Определяется частота рекомбинации для каждой реакции в ячейке

$$v_{m,i} = F_n N_i (N_e - 1) \frac{v_{1\max,i}}{V_c}, \quad (8)$$

а также суммарная частота $v_m = \sum v_{m,i}$.



Фиг. 1. 1 — Функция $v_1(E)$ для реакции $\text{NO}^+ + e \rightarrow \text{N} + \text{O}$; 2 — значение $E = E_{\min}$ (вертикальная линия) и $v_1 = v_{1\max,i}$ (горизонтальная линия); 3 — равновесное распределение $f^{\text{eq}}(E)$ в произвольных единицах. Пример для $T = 7000$ К.

3. $l = l + 1$. Время до следующей реакции τ_l разыгрывается случайным образом по плотности вероятности $v_m \exp(-v_m \tau_l)$.

4. Если $\sum \tau_l > \Delta t$, то процедура в данной ячейке прекращается. Иначе проводится переход по времени на τ_l :

- а) с вероятностью пропорциональной $v_{m,i}$ определяется одна из реакций i , выбирается один ион сорта i из N_i ;
- б) по значению температуры T_e в ячейке и распределению f_{E_e} выбирается энергия электрона E_e ;
- в) вычисляется $E = E_e + E_i$ — сумма энергии электрона и энергии внутренних степеней свободы иона, определяется $v_1 = CE^B$;
- г) реакция выполняется с вероятностью $\frac{v_1}{v_{1\max,i}}$, а с дополнительной вероятностью реагенты остаются без изменения.

2. ВЕСОВАЯ СХЕМА

Число моделирующих частиц в расчете методом ПСМ определяется параметром F_n , выражающим отношение числа моделирующих частиц к числу молекул газа (см. [21]). Необходимое число моделирующих молекул определяется исследованием сходимости по числу частиц (см. [22]). Значение параметра F_n при этом удовлетворяет различным условиям (см., например, [23], [24]), которые однако дают более-менее близкие друг к другу значения.

Введение различных значений F_n для различных сортов (т.н. весовые схемы метода ПСМ) обычно требуется для повышения объема статистики компонент газовой смеси, которые присутствуют в следовых количествах, без существенного увеличения временных затрат на проведение расчета. При использовании вышеописанной численной модели, основанной на отделенной от столкновений реализации реакций для слабоионизованных течений (степень ионизации — сотые и тысячные проценты), требования на F_n становятся очень строгими, поскольку ионы и электроны исчезают в процессе реакции. Если даже выбор параметра F_n обеспечивает корректное моделирование нейтральных компонентов, то для следовых заряженных частиц выбор возможных реагентов в ячейке будет существенно ограничен. Это приведет к занижению скорости реакции в сравнении с требуемыми значениями, а следовательно, к превышению концентрации плазмы в ударном слое.

При реализации весовой схемы предполагалось, что каждая компонента имеет значение F_{ni} , которое может отличаться от других. Во входных данных указывается во сколько раз F_{ni} меньше “глобального”, значения F_n , а программная реализация предполагает, что их отношение является целым числом — степенью двойки $\frac{F_n}{F_{ni}} = 2^{w_i}$, w_i — целое.

Рассмотрим сначала реализацию весовой схемы для столкновений. Затем опишем ее реализацию для бинарных химических реакций.

2.1. Весовая схема для упругих столкновений

Частота столкновений частицы сорта i в газе частиц сорта $j \neq i$ с плотностью n_j равна $v_{1ij} = n_j \overline{\sigma(g)} g$, где g — относительная скорость частиц, а черта сверху обозначает осреднение по скоростям частиц сорта j (см. [25]). Пусть в столкновительной ячейке объема V_c число частиц N_i и N_j для этих сортов. При одинаковых значениях F_n для этих сортов ожидаемое число столкновений N_{ij} моделирующих частиц в ячейке за время Δt будет соответствовать частоте столкновений молекул в газе как

$$\frac{F_n N_{ij}}{V_{cell}} = n_i v_{1ij} \Delta t = \frac{N_i F_n}{V_{cell}} \cdot \frac{N_j F_n}{V_{cell}} \overline{\sigma(g)} g \Delta t. \quad (9)$$

Здесь используется оценка концентрации частиц через их число в ячейке $n_i = N_i \frac{F_n}{V_{cell}}$, а $\overline{\sigma(g)} g$ теперь среднее по частицам сорта i и j .

Выражение (9) может быть обобщено для различных F_n как

$$F_{n,coll} \frac{N_{ij}}{V_{cell}} = n_i n_j \overline{\sigma(g)} g V_{cell} \Delta t = \frac{N_i F_{ni}}{V_{cell}} \cdot \frac{N_j F_{nj}}{V_{cell}} \overline{\sigma(g)} g \Delta t. \quad (10)$$

Из (10) можно получить выражение для числа столкновений моделирующих частиц

$$N_{ij} = N_i N_j \frac{F_{ni} F_{nj}}{F_{n,coll}} \overline{\sigma(g)} g \frac{\Delta t}{V_{cell}}. \quad (11)$$

Если положить $F_{n,coll} = \min(F_{ni}, F_{nj})$, то $\frac{F_{ni} F_{nj}}{F_{n,coll}} = \max(F_{ni}, F_{nj})$. Как обычно для схем мажорантной частоты, введем число потенциальных столкновений через оценку сверху

$$\{N_{ij}\}_{\max} = N_i N_j \max(F_{ni}, F_{nj}) \{\overline{\sigma(g)} g\}_{\max} \frac{\Delta t}{V_{cell}}. \quad (12)$$

Тогда, на основе (12), в ячейке для выполнения столкновений производится расчет мажорантой частоты столкновений сортов i и j , при этом вместо параметра F_n используется $\max(F_{ni}, F_{nj})$:

$$v_{m,ij} = \max(F_{ni}, F_{nj}) N_i N_j \frac{\{\overline{\sigma(g)} g\}_{\max}}{V_c},$$

а также суммарной частоты $v_m = \sum_{i,j \leq i} v_{m,ij}$. При $i = j$ число возможных пар $N_i N_j$ заменяется на $N_i \frac{N_i - 1}{2}$ (см. [21]):

$$v_{m,ii} = F_{ni} \left[N_i \frac{N_i - 1}{2} \right] \frac{\{\overline{\sigma(g)} g\}_{\max}}{V_c}.$$

Реализация столкновений в ячейке на временном шаге Δt состоит в повторении следующей процедуры.

1. Время до следующего столкновения τ разыгрывается случайным образом по плотности вероятности $v_m \exp(-v_m \tau)$.
2. Если время до конца шага больше τ , с вероятностью пропорциональной $v_{m,ij}$ определяются сорта i и j . Иначе — конец процедуры в ячейке.
3. Для случайных молекул сортов i и j определяется относительная скорость $g = |\mathbf{v}_i - \mathbf{v}_j|$.

4. С вероятностью $\sigma(g)g/\{\sigma(g)g\}_{\max}$ вычисляются параметры частиц после столкновения, но изменения параметров частиц не производятся. Иначе выполняется переход на пункт 1 процедуры.

5. Параметры частицы сорта i изменяются на параметры после столкновения с вероятностью $\frac{F_{n,coll}}{F_{ni}}$, с дополнительной вероятностью изменение параметров не проводится. Аналогично для частицы сорта j — с вероятностью $\frac{F_{n,coll}}{F_{nj}}$.

С точки зрения программной реализации эта процедура отличается от стандартной реализации столкновений в методе ПСМ изменением $v_{m,ij}$ и возможностью сохранения параметров частицы до столкновения, обеспечивая среднее число реальных (11) и потенциальных (12) столкновений, каждое из которых соответствует $F_{n,coll}$ реальных столкновений.

2.2. Химические реакции

В модели химических реакций ТСЕ после шага 4 вышеприведенной процедуры вычисляется вероятность химической реакции (3). Как одному столкновению моделирующих частиц соответствует $F_{n,coll}$ реальных столкновений, так и изменение числа моделирующих частиц в химической реакции соответствует изменению реальных частиц соответствующих сортов в $F_{n,coll}$ химических реакциях. Это достигается следующим образом.

- Если частица сорта i является реагентом в химической реакции, то она исчезает или изменяет свои параметры с вероятностью $\frac{F_{n,coll}}{F_{ni}}$, а с дополнительной вероятностью остается в расчете с параметрами, которые были у нее до столкновения.

- Если частица сорта i является продуктом, и $F_{ni} > F_{n,coll}$, то она появляется с вероятностью $\frac{F_{n,coll}}{F_{ni}}$.

- Если для продукта $F_{ni} < F_{n,coll}$ (производство следовой компоненты), то необходимо создание $\frac{F_{n,coll}}{F_{ni}}$ копий этой частицы.

Дробные части числа копий в нашей реализации не возникают, так как любые F_{ni} соотносятся между собой как целые числа.

Для диссоциативной рекомбинации иона сорта i число реакций соответствует рекомбинации F_{ni} реальных ионов. Поэтому в выражении (8) используется параметр F_{ni} , а $N_e = \sum_{j \in \text{ионы}} F_{nj} \frac{N_j}{F_{ni}}$.

В расчетах настоящей работы для ионов $F_{ni} < F_n$. Поэтому при ионизации возникают $\frac{F_n}{F_{ni}}$ копий ионов. Моделирование с идентичными ионами нарушает условие независимости частиц между собой (см. [24]) и приводит к большому отклонению результатов численного расчета от решаемых уравнений. В данном случае отклонение выражается в сильно заниженной плотности плазмы в газе, так как копии одного и того же иона, одновременно двигающиеся в расчете, с большой вероятностью рекомбинируют друг с другом. Для уменьшения влияния копий частиц друг на друга стандартно используется задержка частиц (molecule duplication buffer [1]). Аналогичные принципы используются и при реализации радиальных весов в методе ПСМ (см. [26], [27]).

3. РЕЗУЛЬТАТЫ РАСЧЕТОВ

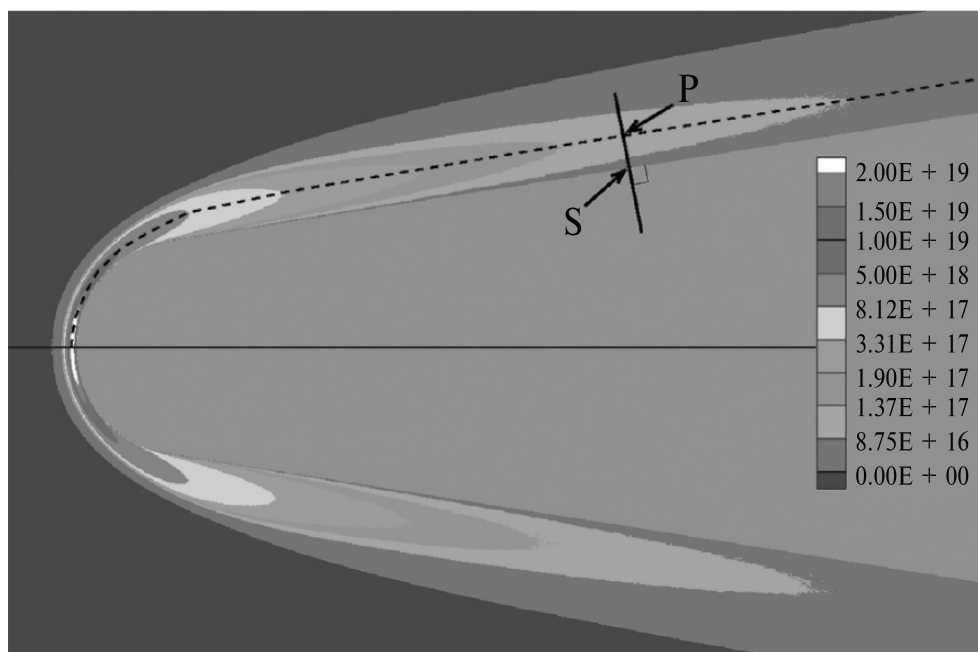
Рассматривалось хорошо исследованное слабоионизованное течение (см., например, [28], [29]), возникающее при обтекании капсулы RAM C-II на высотах 73 км и 81 км. Температура набегающего потока высоте 81 км составляла 194 К, давление 0.726 Па, а на высоте 73 км — 210 К и 2.81 Па. Молярные доли O_2 и N_2 предполагались равными 21.7% и 78.3% соответственно. Капсула представляет собой конус длиной 1.3 м с углом раствора 18° и сферическим затуплением радиусом 0.152 см. На поверхности капсулы реализовывались условия диффузного отражения молекул с полной аккомодацией импульса и энергии, температура поверхности капсулы была равна 1000 К. При отражении ионов от поверхности предполагалась их полная рекомбинация. Молекулярный ион (N_2^+ или NO^+), взаимодействуя со стенкой, становился нейтральной молекулой (N_2 или NO соответственно).

Течение рассмотрено в осесимметричной постановке, ось x направлена вдоль оси симметрии, а ось y — в радиальном направлении. Начало координат — в точке торможения потока. Для случая 81 км расчетная область имела размеры $-0.2 < x < 1.3$ и $y < 0.6$ и была разделена на 1200×480 ячеек. Для случая 73 км параметры были $-0.1 < x < 1.3$, $y < 0.6$ и 2240×960 ячеек. Эти ячейки использовались для накопления макропараметров, в то время как столкновительные ячейки получались в процессе адаптации — деления ячеек в каждом из двух направлений таким образом, чтобы количество частиц в ячейке было меньше восьми. Временной шаг был равен 5×10^{-8} с.

Рассматривается модель воздушной смеси, содержащая восемь компонент воздушной смеси (N, O, N₂, O₂, NO, N₂⁺, NO⁺, e) с 15 реакциями диссоциации, 4 реакциями обмена и 4 реакциями ассоциативной ионизации/диссоциативной рекомбинации. Описание используемых физико-химических моделей и их параметров приведено в работе [30].

Для исследования сходимости по весам заряженных компонент применялось приближение $T_e = T_v(N_2)$, которое использовалось для определения энергии электрона, когда проводились реакции диссоциативной рекомбинации ионов и электронов. В работе [14] ранее проводился анализ этого приближения и было показано, что реалистичное распределение T_e в целом в течении плохо удовлетворяется этим приближением.

В работе [31] рассчитано корректное распределение T_e , в частности, учитывающее теплопроводность свободных электронов. Такие расчеты требуют больших вычислительных ресурсов. Однако расчеты [31] показали, что для рассматриваемого течения результаты с $T_e = T_v(N_2)$ и с реалистичным распределением T_e дают близкие распределения концентрации электронов n_e (также показано на фиг. 2), что объясняется двумя факторами. Во-первых, в областях, где скорость рекомбинации значительна, эти две модели дают близкие значения $T_e \approx T_v(N_2)$. Во-вторых, рекомбинация слабо влияет на максимальные показатели концентрации плазмы в ударном слое, которые измерялись в летном эксперименте.



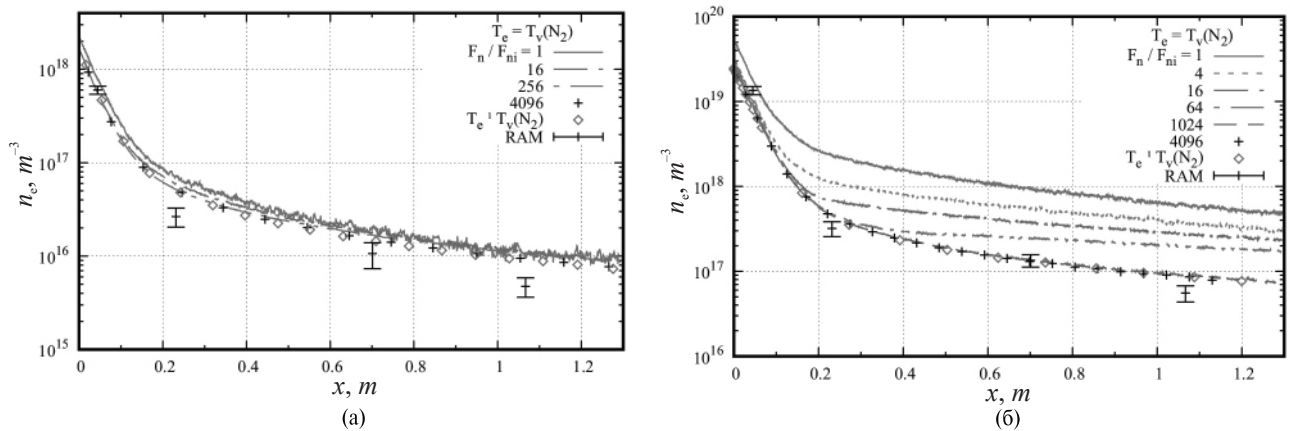
Фиг. 2. Поле концентрации электронов n_e (м⁻³). Вверху — приближение $T_e = T_v(N_2)$, внизу — расчет по процедуре [14] ($T_e \neq T_v(N_2)$).

Таким образом, использование сложной процедуры нахождения корректного распределения T_e не принципиально для исследования сходимости по весовым параметрам схемы, и такое исследование в настоящей работе проведено в приближении $T_e = T_v(N_2)$. Полученные значения весовых параметров использовались для сопряженного решения уравнения сохранения на энергию свободных электронов, когда $T_e \neq T_v(N_2)$, и расчета методом ПСМ.

Максимальная степень ионизации на высоте 81 км приблизительно равна 5×10^{-4} , следовательно, весовой множитель F_n/F_{ni} следует ожидать на уровне 2000. Проведены расчеты для значений F_n/F_{ni} от 1 (без весов) до 4096.

Схема построения графиков распределения концентрации электронов вдоль поверхности капсулы показана на фиг. 2. На схеме пунктиром показана кривая вдоль поверхности, на которой достигаются максимальные значения n_e . Для каждой точки Р этой кривой строится нормаль к поверхности капсулы и определяется соответствующая точка S. На графиках по оси абсцисс отложена координата x точки S. Максимальное вдоль нормали значение n_e (в точке Р) отложено по оси ординат.

Результаты для 81 км показаны на фиг. 3а. Как следует из графиков, использование весов для ионов практически не влияет на плотность электронов. Это объясняется тем, что при обтекании капсулы RAM С-II на данной высоте из-за малой степени ионизации скорость реакций рекомбинации чрезвычайно низкая. Таким образом, хотя использование весовой схемы исправляет скорость реакции рекомбинации, это не отражается на плотности электронов.



Фиг. 3. Концентрация электронов вдоль поверхности капсулы. Расчеты с различными весами ионов $\frac{F_n}{F_{ni}}$; RAM — эксперимент [32] (Radio Attenuation Measurement); (а) — высота 81 км, (б) — высота 73 км.

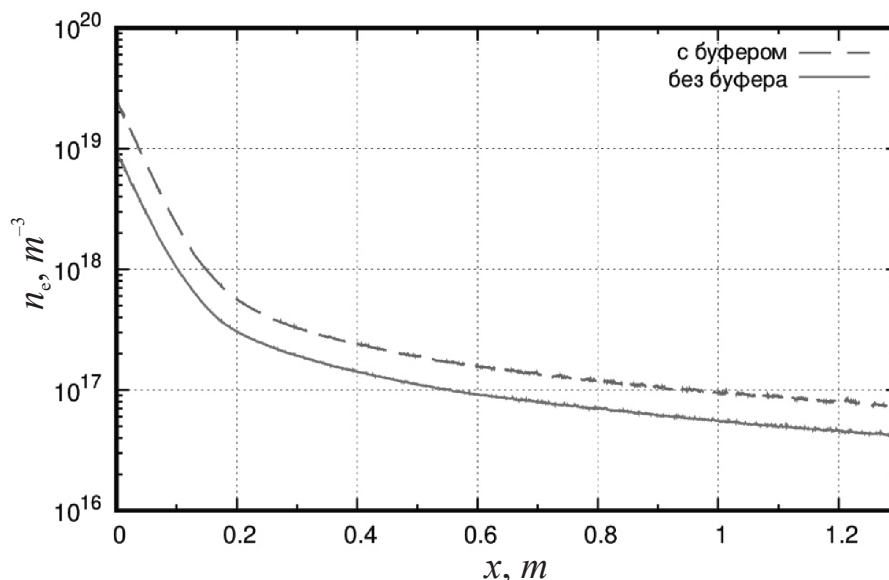
Также на фиг. 3а показан результат расчета, в котором распределение температуры электронов T_e определяется из решения континуального уравнения на энергию свободных электронов, а расчет плотности электронов методом ПСМ выполняется по процедуре, предложенной в работе [14]. Этот расчет обозначен на графике $T_e \neq T_v(N_2)$.

Результаты расчетов сравниваются с летным экспериментом (см. [32]), в котором проводились измерения максимальной плотности электронов с использованием микроволновых рефлектометров (обозначено на графике RAM — Radio Attenuation Measurement). Сравнение показывает хорошее совпадение рассчитанной плотности электронов с измерениями.

На высоте 73 км плотность набегающего потока в 3.6 раз больше, чем в случае 81 км. При этом максимальная концентрация электронов больше в 13.6 раз. Как следует из результатов расчета (фиг. 3б), концентрация электронов n_e заметно различается при разных значениях $\frac{F_n}{F_{ni}}$. Следовательно, влияние рекомбинации на концентрацию электронов для 73 км уже значительно. Из графиков следует, что при значении $\frac{F_n}{F_{ni}} = 1024$ достигается сходимость распределения n_e : график совпадает с $\frac{F_n}{F_{ni}} = 4096$. В расчете с распределением $T_e \neq T_v(N_2)$ по процедуре [14] получено распределение n_e , совпадающее с приближением $T_e = T_v(N_2)$ (ромбы на фиг. 3б), а сравнение с экспериментом вновь дает хорошее согласие.

На фиг. 4 сравниваются распределения n_e , полученные в расчетах с буфером задержки частиц и без буфера. Оба расчета выполнены при значении $\frac{F_n}{F_{ni}} = 1024$. В расчете с буфером, результаты которого показаны также на фиг. 3б, глубина буфера составляла 3 000 000 частиц. При этом на каждом из 63 про-

цессоров, на которых запускался расчет, создавался отдельный буфер. В среднем, для стационарного течения, каждый шаг выполнялось 38 112 замен во всех буферах. Сравнение графиков показывает, что без буфера плотность электронов получается приблизительно в 2 раза меньше, чем с буфером. Отличие объясняется повышенной скоростью рекомбинации, вызванной наличием в расчете копий ионов и свидетельствует о необходимости использования приемов по уменьшению влияния корреляций между копиями на моделируемый процесс рекомбинации.



Фиг. 4. Влияние буферизации дубликатов ионов на концентрацию электронов вдоль поверхности капсулы. Высота 73 км, $F_n/F_{ni} = 1024$.

ЗАКЛЮЧЕНИЕ

Представлено описание реализации весовой схемы метода ПСМ, позволяющей проводить расчеты течений, в которых определенные компоненты газовой смеси имеют низкую относительную концентрацию. Рассмотрено применение реализованной схемы для течения слабоионизованного воздуха, возникающего при обтекании спускаемых космических аппаратов, входящих в атмосферу Земли со скоростями порядка орбитальных (около 8000 м/с). Приведено подробное описание численных моделей релевантных плазмохимических процессов для метода ПСМ и показано, что особенности реализации этих моделей способствуют возникновению большой вычислительной ошибки при определении плотности плазмы, если не использовать весовые схемы.

Применение весовой схемы продемонстрировано на примере условий эксперимента RAM C-II (см. [32]). Рассмотрены случаи полета в условиях различного уровня разреженности течения, соответствующих высотам 81 и 73 км. Численные результаты показывают, что даже для близких по параметрам условий обтекания ошибка в определении плотности плазмы может либо не проявляться (случай 81 км), либо приводить к отклонению концентрации электронов на порядок величины (случай 73 км). Ошибка устраняется с использованием весовой схемы.

Рассматриваемое течение ярко демонстрирует необходимость применения весовых алгоритмов в методе ПСМ для исследования плазмы около космических аппаратов. Метод ПСМ интенсивно развивается, а спектр решаемых этим методом задач постоянно расширяется. Для задач аэротермодинамики спускаемых аппаратов метод ПСМ стал де-факто стандартом определения аэротермодинамических характеристик на верхних участках траектории спуска космических аппаратов. Особенную значимость результаты настоящей работы могут иметь для разработки возвращаемых космических аппаратов при проведении пилотируемых миссий, где важно предсказание участков траектории с экранированием радиосвязи. Одним из направлений дальнейшего развития модели может быть

учет возбуждения электронных уровней атомов и молекул, что, в частности, важно для расчета переноса теплового излучения в разреженном ударном слое.

СПИСОК ЛИТЕРАТУРЫ

1. *Bird G. A.* Molecular Gas Dynamics and the Direct Simulation of Gas Flows. Oxford: Clarendon Press, 1994.
2. *Kashkovsky A. V., Bondar Ye. A., Zhukova G. A., Ivanov M. S., Gimelshein S. F.* Object-oriented software design of real gas effects for the DSMC method // AIP Conf. Proc. 2005. V. 762. N 1. P. 583–588.
3. *Ivanov M. S., Kashkovsky A. V., Vashchenkov P. V., Bondar Ye. A.* Parallel object-oriented software system for DSMC modeling of high-altitude aerothermodynamic problems // AIP Conf. Proc. 2011. V. 1333. N 1. P. 211–218.
4. *Shoev G. V., Vashchenkov P. V., Bondar Ye. A.* Calculation of the heat flux and pressure on the cone surface in a high-enthalpy non-equilibrium flow of a binary nitrogen mixture (N_2/N) // AIP Conf. Proc. 2018. V. 2027. N 1. P. 040010.
5. *Wysong I., Gimelshein S., Bondar Ye., Ivanov M.* Comparison of direct simulation Monte Carlo chemistry and vibrational models applied to oxygen shock measurements // Phys. Fluid. 2014. V. 26. N 4. P. 043101.
6. *Galeyev A. Ye., Kashkovsky A. V., Shevyrin A. A., Bondar Ye. A.* Comparison of nonequilibrium dissociation models in the direct simulation Monte Carlo method // J. Phys.: Conf. Ser. 2019. V. 1404. P. 012107.
7. *Molchanova A. N., Kashkovsky A. V., Bondar Ye. A.* Surface recombination in the direct simulation Monte Carlo method // Phys. Fluid. 2018. V. 30. N 10. P. 107105.
8. *Litvintsev A. S., Molchanova A. N., Bondar Ye. A.* Effects of heterogeneous NO production on the aerothermodynamics of high-altitude re-entry // AIP Conf. Proc. 2020. V. 2288. N 1. P. 030090.
9. *Shevyrin A. A., Shkredov T. Yu., Shoev G. V., Bondar Ye. A.* Modeling of the plasma environment of re-entry space vehicles // AIP Conf. Proc. 2018. V. 2027. N 1. P. 030031.
10. *Shkredov T., Shoev G., Shevyrin A.* Implementation of the tangent-slab approximation for radiation modeling coupled with flowfield computation // AIP Conf. Proc. 2023. V. 2504. N 1. P. 030053.
11. *Lee J.-H.* Basic governing equation for flight regimes of Aeroassisted orbital transfer vehicles // AIAA 1984–1729.
12. *Gnoffo P. A., Gupta R. N., Shinn J. L.* Conservation equations and physical models for hypersonic air flows in thermal and chemical nonequilibrium // NASA TP V. 2867. 1989.
13. *Kim M. K., Gulhan A., Boyd I. D.* Modeling of electron temperature in hypersonic flows // AIAA 2011–1028.
14. *Shevyrin A., Bondar Ye.* On the calculation of the electron temperature flowfield in the DSMC studies of ionized re-entry flows // Adv. Aerodyn. 2020. V. 2. N 6.
15. *Bird G. A.* Nonequilibrium radiation during re-entry at 10 km/s // AIAA Paper 87–1543. 1987.
16. *Кусов А. Л.* Расчет ионизации методом прямого статистического моделирования Монте–Карло // Физ.-хим. кин. в газ. дин. 2016. Т. 17. Вып. 2.
17. *Parent B., Shneider M. N., Macheret S. O.* Detailed modeling of plasmas for computational aerodynamics // AIAA J. 2016. V. 54. N 3. P. 898–911.
18. *Kusov A. L., Levashov V. Y., Gerasimov G. Y., Kozlov P. V., Bykova N. G., Zabelinsky I. E.* Direct statistical Monte Carlo simulation of argon radiation behind the front of a strong shock wave // Fluid. Dyn. 2023. V. 58. P. 759–772.
19. *Boyd I. D.* Modeling backward chemical rate processes in the direct simulation Monte Carlo method // Phys. Fluid. 2007. V. 19. N 12. P. 126103.
20. *Gimelshein S. F., Wysong I. J.* Impact of the ionization reaction set in nonequilibrium hypersonic air flows // AIAA J. 2020. V. 58. N 3. P. 1255–1265.
21. *Иванов М. С., Рогазинский С. В.* Экономичные схемы прямого статистического моделирования течений разреженного газа // Матем. моделирование. 1989. Т. 1. № 7. С. 130–145.

22. *Shugalevskaia N. V., Shevyrin A. A., Bondar Ye. A.* Comparison of modern implementations of the direct simulation Monte Carlo method // *J. Phys.: Conf. Ser.* 2019. V. 1404. N 012123.
23. *Maltsev R. V.* On the Selection of the Number of Model Particles in DSMC Computations // *AIP Conf. Proc.* 2011. V. 1333. N 1. P. 289–294.
24. *Shevyrin A. A., Bondar Ye. A., Ivanov M. S.* Analysis of repeated collisions in the DSMC method // *AIP Conf. Proc.* 2005. V. 762. N 1. P. 565–570.
25. *Коган М. Н.* Динамика разреженного газа (кинетическая теория). М.: Наука, 1967. 440 с.
26. *Ivanov M. S., Markelov G. N., Kashkovsky A. V., Gimelshein S. F.* Statistical simulation of high-altitude aerodynamic problems // *Proc. 3rd Europ. Symp. on Aerothermodynamics for Space Vehicles, Noordwijk.* 1998. ESA SP-426. P. 245–252.
27. *Мальцев Р. В.* Численное исследование сверхзвуковых течений разреженных газовых смесей с сильно отличающимися массами компонент. Дис. ... канд. физ.-мат. наук. Новосибирск: ИТ СО РАН, 2014.
28. *Candler G. V., MacCormack R. W.* Computation of weakly ionized hypersonic flows in thermochemical nonequilibrium // *J. Thermophys. Heat Trans.* 1991. V. 5. N 3. P. 266–273.
29. *Суржиков С. Т.* Двумерный численный анализ ионизации потока в летном эксперименте RAM-C-II // *Хим. физика.* 2015. Т. 34. № 2. С. 24–42.
30. *Шевырин А. А., Бондарь Е. А., Калашников С. Т., Хлыбов В. И., Дегтярь В. Г.* Прямое статистическое моделирование разреженного высокоэнтальпийного течения около капсулы RAM C-II // *TBT.* 2016. Т. 54. № 3. С. 408–414.
31. *Shevyrin A. A., Bondar Ye. A.* Coupled kinetic-continuum modeling of reentry vehicle plasma environment // *AIP Conf. Proc.* 2024. V. 2996. N 1. P. 080015.
32. *Jones W. L., Cross A. E.* Electrostatic-probe measurements of plasma parameters for two reentry flight experiments at 25000 feet per second // *Tech. Rep. NASA, Washington, D.C., 1972. TN D-6617.*

USING MONTE CARLO WEIGHT SCHEMES FOR WEAKLY IONIZED RAREFIED GAS FLOWS

A. A. Shevyryan^{a,*}, E. A. Bondar^a

^a 630090 Novosibirsk, Institutskaya str., 4/1, S. A. Khristianovich Institute of Theoretical and Applied Mechanics, SB RAS, Russia

*e-mail: shevr@itam.nsc.ru

Received 23 November, 2023

Revised 18 March, 2024

Accepted 3 April, 2024

Abstract. The description of the procedures of the direct statistical Monte Carlo simulation of weakly ionized flows arising during the flow of returning spacecraft is presented. For ionization and recombination reactions, expressions are given for the model dependence of the probability of reactions on the velocities and energies of the reagents. The dissociative recombination algorithm is presented, the computational efficiency of which is achieved by bypassing the simulation of the interaction of electrons and heavy particles. An approach to the construction of a weight scheme for elastic collisions and chemical reactions is described, which significantly increases the computational efficiency of calculations. An example of using the described numerical models and procedures to study the weakly ionized flow near the return capsule in typical entry conditions of orbiting spacecraft is presented. The calculation results are compared with the data of measurements of plasma parameters in the flight experiment.

Keywords: weight schemes, high-altitude aerothermodynamics, ionization, direct statistical Monte Carlo simulation method, nonequilibrium gas flow, low-temperature plasma, charge recombination.