

СОДЕРЖАНИЕ

Том 62, номер 2, 2022 год

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

Convergence Analysis of the ADI Scheme for Parabolic Problems
using Discrete Harmonic Functions

B. Bialecki, M. Dryja, R. Fernandes

195

О конгруэнтных централизаторах блочно-диагональной матрицы
и матрицы Сергейчука–Хорна

Х. Д. Икрамов

196

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Инвариантные кривые некоторых дискретных динамических систем

В. П. Варин

199

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

Асимптотика решения задачи оптимального граничного управления
в двухсвязной области с различной интенсивностью на участках границы

А. Р. Данилин

217

Прямо-двойственный метод Ньютона с наискорейшим спуском
для линейной задачи полуопределенного программирования.
Ньютоновская система уравнений

В. Г. Жадан

232

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Стохастические алгоритмы решения первой краевой задачи
для некоторых эллиптических уравнений второго порядка
с разрывными коэффициентами

А. Н. Кузнецов, А. С. Сипин

249

МАТЕМАТИЧЕСКАЯ ФИЗИКА

Кинетические модели агрегации, приводящей к морфологической памяти
образовавшихся структур

С. З. Аджиев, В. В. Веденяпин, И. В. Мелихов

255

Математическое моделирование опасных явлений
природного характера в мелководном водоеме

А. М. Атаян, А. В. Никитина, А. И. Сухинов, А. Е. Чистяков

270

Быстрый алгоритм решения трехмерной обратной многочастотной задачи
скалярной акустики с данными в цилиндрической области

А. Б. Бакушинский, А. С. Леонов

289

Нахождение стационарных течений пуазейлевского типа
для несжимаемой полимерной жидкости методом установления

А. М. Блохин, Б. В. Семисалов

305

К численному решению одного класса интегральных уравнений III рода <i>Н. С. Габбасов, Э. Х. Галимова</i>	320
Сходимость аттракторов аппроксимации к аттракторам модифицированной модели Кельвина–Фойгта <i>М. В. Турбин, А. С. Устюжанинова</i>	330

ИНФОРМАТИКА

Ускоренные проксимальные оболочки: применение к покомпонентному методу <i>А. С. Аникин, В. В. Матюхин, Д. А. Пасечнюк</i>	342
--	-----

**ОБЩИЕ
ЧИСЛЕННЫЕ МЕТОДЫ**

УДК 519.633

**CONVERGENCE ANALYSIS OF THE ADI SCHEME FOR PARABOLIC
PROBLEMS USING DISCRETE HARMONIC FUNCTIONS¹⁾**

© 2022 г. **В. Bialecki^{1,*}, М. Dryja^{2,**}, R. Fernandes^{3,***}**

¹ *Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA*

² *Institute of Applied Mathematics and Mechanics, Warsaw University, Warsaw, Poland*

³ *Department of Mathematics, Khalifa University of Science and Technology, P.O. Box 2533,
Abu Dhabi, United Arab Emirates*

**e-mail: bbialeck@mines.edu*

***e-mail: dryja@mimuw.edu.pl*

****e-mail: ryan.fernandes@ku.ac.ae*

Поступила в редакцию 25.03.2021 г.

Переработанный вариант 25.03.2021 г.

Принята к публикации 12.10.2021 г.

Анализ сходимости схемы АДИ для параболических задач с использованием дискретных гармонических функций. Рассмотрена разностная неявная схема переменных направлений в прямоугольнике для уравнения теплопроводности. Используя результаты В.Б. Андреева по численной устойчивости схемы, доказан второй порядок сходимости этой схемы.

Ключевые слова: уравнение теплопроводности, конечные разности, неявная схема переменных направлений, анализ сходимости схемы.

DOI: 10.31857/S0044466922020041

¹⁾ Полный текст статьи печатается в английской версии журнала.

**ОБЩИЕ
ЧИСЛЕННЫЕ МЕТОДЫ**

УДК 512.643

**О КОНГРУЭНТНЫХ ЦЕНТРАЛИЗАТОРАХ БЛОЧНО-
ДИАГОНАЛЬНОЙ МАТРИЦЫ И МАТРИЦЫ СЕРГЕЙЧУКА–ХОРНА**

© 2022 г. **Х. Д. Икрамов**

119992 Москва, Ленинские горы, МГУ, ВМК, Россия

e-mail: ikratov@cs.msu.su

Поступила в редакцию 12.02.2021 г.
Переработанный вариант 12.02.2021 г.
Принята к публикации 12.10.2021 г.

Конгруэнтный централизатор матрицы A – это множество матриц X , удовлетворяющих соотношению $X^*AX = A$. Конструкция прямой суммы и матрица Сергейчука–Хорна суть два элемента в описании канонической формы произвольной комплексной матрицы относительно эрмитовых конгруэнций. В данной статье устраняются излишние предположения и неточности имеющихся описаний конгруэнтных централизаторов блочно-диагональных матриц и матрицы Сергейчука–Хорна. Библ. 4.

Ключевые слова: конгруэнция, конгруэнтный централизатор, коквадрат.

DOI: 10.31857/S0044466922020090

1. Пусть A – невырожденная комплексная $n \times n$ -матрица. Матрица

$$K_A = A^{-*}A$$

называется коквадратом матрицы A , а множество

$$\mathcal{C}_A^* = \{X \mid X^*AX = A\}$$

называется конгруэнтным централизатором для A . Этот термин объясняется тем, что множество \mathcal{C}_A^* есть аналог обычного централизатора матрицы A для случая, когда вместо подобий группа $GL_n(\mathbb{C})$ действует на матричном пространстве $M_n(\mathbb{C})$ конгруэнциями, т.е. преобразованиями вида $A \rightarrow P^*AP$ с произвольными невырожденными матрицами P .

Пусть A – матрица блочно-диагонального вида:

$$A = \begin{pmatrix} B & O \\ O & C \end{pmatrix}. \tag{1}$$

Блоки B и C суть квадратные матрицы порядков соответственно k и l ($k + l = n$). Для таких A в [1] было доказано следующее утверждение:

Теорема 1. Пусть A – невырожденная матрица вида (1), а X – произвольная матрица из централизатора \mathcal{C}_A^* . Представим X в блочной форме, согласованной с прямой суммой (1):

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}. \tag{2}$$

Предположим, что коквадраты матриц B и C не имеют общих собственных значений. Тогда, если в (2) подматрицы X_{11} и X_{22} невырождены, то и матрица X является прямой суммой:

$$X = X_{11} \oplus X_{22}. \tag{3}$$

Как следствие, X_{11} и X_{22} принадлежат конгруэнтным централизаторам соответственно матриц B и C .

В п. 2 настоящей статьи мы показываем, что условие невырожденности блоков X_{11} и X_{22} в этой теореме является лишним, и в действительности *все* матрицы из \mathcal{C}_A^* имеют блочно-диагональную форму.

Матрицей Сергейчука–Хорна мы называем матрицу

$$\Delta_n = \begin{pmatrix} & & & 1 \\ & & \cdots & i \\ & 1 & \cdots & \\ 1 & i & & \end{pmatrix}.$$

Матрица Δ_n представляет один из трех типов блоков, прямой суммой которых является каноническая форма произвольной квадратной матрицы относительно конгруэнций.

В [2] было доказано, что, с точностью до диагональных сомножителей специального вида, все матрицы из конгруэнтного централизатора \mathcal{C}_Δ^* матрицы Δ_n являются верхнетреугольными и тёплицевыми. В п. 3 мы показываем, что в этом утверждении оговорка относительно диагональных сомножителей не нужна, и верна следующая

Теорема 2. *Всякая матрица $X \in \mathcal{C}_\Delta^*$ тёплицева и имеет верхнетреугольную форму.*

2. Для произвольных квадратных матриц A и X одинакового порядка справедливо равенство

$$(X^* - I)A(X + I) + (X^* + I)A(X - I) = 2(X^*AX - A).$$

Если $X \in \mathcal{C}_A^*$, то правая часть обращается в нуль:

$$(X^* - I)A(X + I) + (X^* + I)A(X - I) = 0. \quad (4)$$

Предположим, что X не имеет собственного значения 1, и умножим (4) слева на матрицу $(X^* - I)^{-1}$, а справа – на $(X - I)^{-1}$. Полагая

$$Y = (X + I)(X - I)^{-1}, \quad (5)$$

получаем уравнение

$$AY + Y^*A = 0 \quad (6)$$

для матрицы Y . Пользуясь невырожденностью матрицы A , выводим отсюда

$$Y = -A^{-1}Y^*A$$

и

$$Y^* = -A^*YA^{-*}.$$

Подставляя это выражение для Y^* в (6), имеем

$$AY = A^*YA^{-*}A$$

или

$$(A^{-*}A)Y = Y(A^{-*}A). \quad (7)$$

В проведенных до сих пор выкладках A могла быть произвольной невырожденной матрицей. Пусть теперь A – блочно-диагональная матрица (1). Ее коквадрат $K_A = A^{-*}A$ также является прямой суммой:

$$A^{-*}A = B^{-*}B \oplus C^{-*}C.$$

По предположению, матрицы $B^{-*}B$ и $C^{-*}C$ не имеют общих собственных значений. Поэтому матрица Y , коммутирующая с $A^{-*}A$, должна быть блочно-диагональной с диагональными блоками порядков k и l .

Нетрудно показать, что, как и X , матрица Y не имеет собственного значения 1. Разрешая относительно X соотношение (5), имеем

$$X = (Y + I)(Y - I)^{-1}.$$

Матрица X сохраняет блочно-диагональную форму матрицы Y . Записывая X в виде (3), из условия $X^*AX = A$ выводим

$$X_{11}^*BX_{11} = B, \quad X_{22}^*CX_{22} = C.$$

Предположение, что X не имеет собственного значения 1, не является ограничительным. Очевидно, что вместе с X централизатору \mathcal{C}_A^* принадлежат и все матрицы вида αX , где $|\alpha| = 1$. Поэтому, если в спектре матрицы X есть 1, то X нужно заменить подходящим кратным $\alpha_0 X$, уже не имеющим этого собственного значения.

3. Пусть теперь $A = \Delta_n$, а X — матрица из централизатора \mathcal{C}_{Δ}^* , не имеющая собственного значения 1. Определяя матрицу Y формулой (5), приходим, как и в п. 2, к уравнению

$$(\Delta_n^{-*}\Delta_n)Y = Y(\Delta_n^{-*}\Delta_n). \quad (8)$$

Матрица $\Pi = \Delta_n^{-*}\Delta_n$ есть коквадрат матрицы Δ_n . Она описана в [3, задача 4.5.P15] и представляет собой верхнетреугольную тёплицеву матрицу с единичной главной диагональю и числом $2i$ на первой наддиагонали. Прочие детали устройства этой матрицы для нас не важны.

Обозначим через \mathcal{T}_n алгебру верхнетреугольных тёплицевых матриц порядка n . Этой алгебре, очевидно, принадлежат матрицы

$$I_n, \Pi - I_n, (\Pi - I_n)^2, \dots, (\Pi - I_n)^{n-1}. \quad (9)$$

Матрица $(\Pi - I_n)^k$ имеет нулевую главную диагональ и $k - 1$ первых нулевых наддиагоналей; в то же время k -я наддиагональ ненулевая. Отсюда следует, что матрицы (9) образуют (линейный) базис алгебры \mathcal{T}_n .

Соотношение (8) означает, что матрица Y коммутирует с Π , а потому Y коммутирует со всеми матрицами (9). Жорданова клетка $J_n(0)$ с нулем на главной диагонали принадлежит алгебре \mathcal{T}_n и, значит,

$$J_n(0)Y = YJ_n(0).$$

Это равенство означает (см. [4, гл. VIII, § 2]), что и матрица Y принадлежит \mathcal{T}_n , т.е. является верхнетреугольной тёплицевой матрицей.

Как и в п. 2, матрица Y не имеет собственного значения 1. Разрешая относительно X соотношение (5), имеем

$$X = (Y + I)(Y - I)^{-1}.$$

Отсюда заключаем, что матрица X имеет ту же верхнетреугольную тёплицеву форму, что и Y .

Предположение, что X не имеет собственного значения 1, снова не является ограничительным: если в спектре матрицы X есть 1, то X нужно заменить подходящим кратным $\alpha_0 X$, не имеющим этого собственного значения.

СПИСОК ЛИТЕРАТУРЫ

1. Икрамов Х.Д. Конгруэнтный централизатор блочно-диагональной матрицы // Зап. научн. сем. ПОМИ. 2016. Т. 453. С. 96–103.
2. Икрамов Х.Д. Конгруэнтный централизатор матрицы Сергейчука–Хорна // Зап. научн. сем. ПОМИ. 2016. Т. 453. С. 104–113.
3. Horn R.A., Johnson C.R. Matrix Analysis. Second Edition. — Cambridge: Cambridge University Press, 2013.
4. Гантмахер Ф.Р. Теория матриц. М.: Наука, 1966.

**ОБЫКНОВЕННЫЕ
ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ**

УДК 519.624

**ИНВАРИАНТНЫЕ КРИВЫЕ НЕКОТОРЫХ ДИСКРЕТНЫХ
ДИНАМИЧЕСКИХ СИСТЕМ**

© 2022 г. В. П. Варин

*125047 Москва, Миусская пл., 4, Институт прикладной математики
им. М.В. Келдыша РАН, Россия
e-mail: varin@keldysh.ru*

Поступила в редакцию 12.01.2021 г.
Переработанный вариант 23.03.2021 г.
Принята к публикации 04.08.2021 г.

Классическая задача о построении непрерывных итераций аналитического отображения рассматривается как задача о построении инвариантных кривых дискретной динамической системы. Такие системы часто изучаются как редукции непрерывных динамических систем (отображение Пуанкаре). Существование аналитических инвариантных кривых в дискретной динамической системе влечет (локально) существование дополнительного аналитического первого интеграла в непрерывной динамической системе. Однако доказать его существование удается крайне редко, так как доказательство обычно выводят из сходимости формальных разложений инвариантных кривых. Строятся примеры дискретных динамических систем, инвариантные кривые которых даются заведомо расходящимися рядами, но в то же время являются аналитическими. В частности, дается пример интегрируемой дискретной динамической системы, имеющей хаотические траектории. Библ. 15. Фиг. 4.

Ключевые слова: инвариантные кривые, непрерывные итерации, расходящиеся ряды, динамические системы, интегрируемость.

DOI: 10.31857/S0044466921120164

1. ВВЕДЕНИЕ

Данная статья продолжает работу [1], где некоторые важные вопросы были оставлены для дальнейшего исследования.

В статье [1] рассматривалась задача интерполяции итерационных последовательностей, что можно интерпретировать как восстановление непрерывной динамической системы (НДС) по ее дискретной редукции. В самом деле, любую итерационную последовательность можно рассматривать как дискретную динамическую систему (ДДС), т.е. дискретное отображение (возможно, с запаздыванием). Таким образом, в [1] рассматривалась задача, в некотором смысле обратная той, которая возникает при редукции НДС к дискретным системам.

За последние примерно полвека интерес к ДДС значительно возрос во многом благодаря знаменитой статье Хенона и Хейлеса (см. [2]). Видимо, начиная с этой публикации ДДС стали рассматриваться как весьма содержательные модельные задачи, которые выражают самое существенное в динамике НДС.

В самом деле, отображение Пуанкаре (отображение последования) осуществляет редукцию НДС к дискретной. Поэтому такие явления в НДС, как периодические или квазипериодические решения, бифуркации решений, а также динамический хаос, могут изучаться в дискретной модели. Это же относится и к весьма важному вопросу об интегрируемости (локальной интегрируемости) динамической системы или ее отсутствию.

Вопрос о (локальной) интегрируемости динамической системы (точнее, о существовании дополнительного интеграла) связан с существованием аналитических инвариантных кривых дискретного отображения. Так, в [2] было отмечено, что динамическая система, описывающая звездную динамику в галактике, является с большой вероятностью неинтегрируемой, так как обнаружить дополнительный аналитический интеграл не удастся. В то же время астрономические наблюдения показывают, что динамика в системе такова, как будто система интегрируема. Ста-

тья [2] посвящена изучению этого парадокса на примере упрощенной модели звездной динамики, т.е. системы Хенона–Хейлеса (СХХ).

Хенону и Хейлесу удалось обнаружить численно с большой точностью, что дискретные итерации отображения Пуанкаре в СХХ при небольших значениях энергии образуют инвариантные кривые в сечении Пуанкаре. Если бы удалось доказать аналитичность этих кривых, то СХХ была бы (локально) интегрируемой, так как существовал бы второй аналитический первый интеграл в гамильтоновой системе с двумя степенями свободы.

Доказать существование аналитических инвариантных кривых в СХХ не удастся. Это связано с тем, что инвариантные кривые обычно вычисляются с помощью приведения СХХ к нормальной форме. При этом нормализующее преобразование либо заведомо расходится, либо доказать его сходимости не представляется возможным.

Инвариантные кривые можно вычислить и непосредственно из СХХ с помощью систем компьютерной алгебры (CAS), т.е. получить (локально) второй первый интеграл в виде формального ряда. Однако и в этом случае полученный ряд скорее всего расходится.

В настоящее время СХХ считается неинтегрируемой, а значит, как полагают, инвариантные кривые в сечении Пуанкаре не могут быть аналитическими. Это же относится к многочисленным дискретным моделям, изучаемым в связи с этой тематикой (см. обзор [3]).

Наряду с этим существует ряд работ (см. [4] и ссылки там), посвященных доказательству существования аналитических инвариантных кривых в некоторых модельных ДДС, как консервативных, так и диссипативных.

Во всех без исключения упомянутых работах аналитичность инвариантных кривых выводится из доказательства сходимости формальных рядов, представляющих указанные кривые. При этом получение формальных рядов, приближающих инвариантные кривые, является обычно чисто технической задачей, в то время как доказательство их сходимости является совершенно нетривиальным и занимает основную часть работы.

Доказательства сходимости формальных рядов (или нормализующих преобразований) часто опираются на некоторые весьма тонкие диофантовы свойства рациональных коэффициентов формальных рядов. Поэтому такие доказательства довольно редки. Видимо, поэтому “интегрируемых” дискретных отображений сравнительно немного (см. обзор [3]).

В то же время можно с уверенностью утверждать, что в случае общего положения так называемые *стандартные отображения* неинтегрируемы в том смысле, что удастся доказать заведомую расходимость формальных рядов, приближающих инвариантные кривые вблизи неподвижной точки. Хотя доказательства расходимости этих рядов также весьма нетривиальны (см. обзор [3]).

Тот факт, что расходимость указанных формальных рядов является типичной, хорошо виден на примере итераций (композиций) аналитического отображения

$$y \rightarrow F(y) = y + \sum_{n=1}^{\infty} c_n y^{n+1}, \quad (1)$$

где $F(y)$ является голоморфной функцией в некоторой окрестности нуля.

Отображение (1) дает пример простейшей ДДС $y_{n+1} = F(y_n) = F^{(n+1)}(y_0)$, $n \in \mathbb{Z}$, $y \in \mathbb{R}$ или $y \in \mathbb{C}$. Как известно, формальные ряды *непрерывной итерации* (НИ) $F^{(x)}(y)$ (см. разд. 2) почти всегда расходятся, если $x \notin \mathbb{Z}$ (см. [1] и ссылки там).

Однако численные эксперименты в СХХ, а также со стандартным отображением Чирикова и в других задачах (см. [5]) показывают, что приближенные инвариантные кривые “идентичны натуральным”, т.е. ведут себя так же, как будто они являются аналитическими. Этот факт неоднократно упоминается в литературе в том плане, что численно установить отличие инвариантных кривых от аналитических не представляется возможным (см. [2]).

Это же наблюдение можно сделать при изучении НИ $F^{(x)}(y)$. Как отмечалось в [1], НИ $F^{(x)}(y)$, $x \in \mathbb{R}$, хорошо интерполируют дискретную последовательность $y_{n+1} = F(y_n)$, $n \in \mathbb{Z}$, несмотря на заведомую расходимость рядов $F^{(x)}(y)$, $x \notin \mathbb{Z}$.

Таким образом, мы видим, что аналитичность инвариантных кривых напрямую связывается со сходимостью их формальных разложений, каковая случается очень редко. Некоторые авторы идут еще дальше и связывают расходимость этих рядов с несуществованием инвариантных кривых (или отсутствием их аналитичности), хотя такие утверждения редки.

Насколько нам известно, этот “когнитивный диссонанс”, когда инвариантные кривые вроде бы не существуют, хотя, с другой стороны, они как бы есть, — нигде в литературе не обсуждается, во всяком случае в плане его разрешения.

В данной работе мы попытаемся частично прояснить этот парадокс, построив пример ДДС, инвариантные кривые которой даются заведомо расходящимися рядами, с одной стороны, но в то же время являются аналитическими. Таким образом, реализуется некоторый аналог теоремы Бореля–Ритта, утверждающей, что любой формальный степенной ряд (одной переменной) является асимптотическим разложением некоторой аналитической функции в некотором секторе комплексной плоскости.

2. ИНВАРИАНТНЫЕ КРИВЫЕ НЕПРЕРЫВНОЙ ИТЕРАЦИИ

Напомним некоторые факты о НИ отображения (1), которые нам понадобятся (см. [1] и ссылки там).

Рассмотрим обычную композицию (итерацию) отображения (1)

$$F^{(n)}(y) = F \circ F \circ \dots \circ F(y), \quad n \text{ раз.}$$

Очевидно, что

$$F^{(1)}(y) = F(y), \quad F^{(0)}(y) = y; \quad F^{(n+m)}(y) = F^{(n)}(F^{(m)}(y)), \quad n, m \in \mathbb{N}.$$

НИ $F^{(x)}(y)$ обобщает это полугрупповое свойство на произвольные комплексные числа $x \in \mathbb{C}$, т.е. порождает непрерывную группу отображений.

НИ отображения (1) можно записать в виде

$$F^{(x)}(y) = y + \sum_{n=1}^{\infty} a_n(x)y^{n+1}, \quad (2)$$

где $a_n(x)$ — однозначно определяемые полиномы степени не выше n .

Другая форма записи НИ использует *генератор* $g_1(y) = \omega(y)$ указанной группы отображений (см. [6]), т.е.

$$F^{(x)}(y) = \sum_{n=0}^{\infty} g_n(y) \frac{x^n}{n!}, \quad g_0(y) = y, \quad g_n(y) = g_1(y)g_{n-1}(y), \quad n \in \mathbb{N}. \quad (3)$$

Вычисление НИ является классической задачей. Систематическое изложение методов их вычисления, существовавших ранее, можно найти в [7, с. 314].

В [1] предложен весьма простой метод вычисления полиномов $a_n(x)$. Пусть

$$b_n = \left. \frac{d}{dx} a_n(x) \right|_{x=0},$$

тогда

$$b_n = c_n - d_n(1), \quad a_n(x) = b_n x + d_n(x), \quad (4)$$

где

$$d_n(x) = \int_0^x \left(\sum_{m=1}^{n-1} (m+1)a_m(t)b_{n-m} \right) dt, \quad n \in \mathbb{N}.$$

Формулу (4) можно также использовать для вычисления НИ непосредственно по ее генератору $\omega(y)$ без громоздкого рекурсивного дифференцирования; или для быстрого вычисления генератора $\omega(y)$. Поскольку, согласно (3), имеем

$$\frac{d}{dx} F^{(x)}(y) = \omega(y) \frac{d}{dy} F^{(x)}(y), \quad (5)$$

то, подставив $x = 0$ в эту формулу, получим

$$\omega(y) = \sum_{n=1}^{\infty} b_n y^{n+1},$$

где величины b_n вычисляются в (4); либо b_n считаются известными, и тогда вычисляются величины c_n одновременно с вычислением полиномов $a_n(x)$.

Приведем еще одну полезную формулу (см. [6, (84)]), связывающую НИ с ее генератором:

$$\frac{d}{dx} F^{(x)}(y) = \omega(F^{(x)}(y)). \quad (6)$$

Как отмечалось выше, отображение (1) порождает ДДС $y_{n+1} = F(y_n)$, $n \in \mathbb{Z}$. При этом НИ $F^{(x)}(y)$, $x \in \mathbb{C}$, порождает НДС при условии, что формальные ряды $F^{(x)}(y)$ можно интерпретировать как некоторые функции. Если подразумевать под этим сходимость формальных рядов НИ, то это удается сделать крайне редко.

Насколько нам известно, все существующие примеры аналитических НИ даются формулой (см. [8])

$$F^{(x)}(y) = y(1 - kxy^k)^{-1/k}, \quad k \in \mathbb{N}, \quad (7)$$

соответствующей генератору $\omega(y) = y^{k+1}$ (см. также [6, (72)]).

НИ дает пример динамической системы, инвариантные кривые которой можно сразу выписать в виде формальных рядов:

$$y = F^{(x)}(C), \quad \text{или} \quad C = F^{(-x)}(y), \quad (8)$$

где $C = y_0$ – начальное значение (и первый интеграл), и $y \in \mathbb{R}$ или $y \in \mathbb{C}$. В последнем случае непрерывное “время” $x \in \mathbb{R}$ можно исключить из уравнений и рассматривать инвариантные кривые как функции $u = u(v)$ или $v = v(u)$, где $u = \operatorname{Re}(y)$, $v = \operatorname{Im}(y)$.

Например, рассмотрим простейшее из всех возможных нетривиальных отображений (1), $F(y) = y - y^2$, т.е. частный случай логистического отображения ($F(y) = y + Ay^2$ приводится к этому виду перенормировкой):

$$y_{n+1} = y_n - y_n^2, \quad n \in \mathbb{N}_0, \quad y_0 \in \mathbb{C}. \quad (9)$$

Вычисления по формуле (4) дают генератор:

$$\omega(y) = -y^2 - y^3 - \frac{3}{2}y^4 - \frac{8}{3}y^5 - \frac{31}{6}y^6 - \frac{157}{15}y^7 - \frac{649}{30}y^8 + \dots, \quad (10)$$

а также формальный ряд, представляющий инвариантные кривые НИ:

$$C = (y - y^2)^{(-x)} = y + xy^2 + xy^3 + x^2y^3 + \frac{3}{2}xy^4 + \frac{5}{2}x^2y^4 + \frac{8}{3}xy^5 + x^3y^4 + \\ + 6x^2y^5 + \frac{31}{6}xy^6 + \frac{13}{3}x^3y^5 + \frac{157}{15}xy^7 + \frac{175}{12}x^2y^6 + \dots$$

Напомним, что оба этих ряда расходятся (см. [1] и ссылки там).

Тем не менее можно легко проверить численно, что даже этот скромный отрезок формального ряда $(y - y^2)^{(-x)}$ хорошо приближает инвариантную кривую $y = y(x)$ для $y(0) = C = 0.1$ по крайней мере в интервале $x \in [-2, 1]$.

Для меньших значений $y(0) = C$, либо для большего отрезка ряда $(y - y^2)^{(-x)}$ этот интервал можно значительно расширить в обе стороны (см. [1]).

С другой стороны, нет смысла просто увеличивать количество членов формальных разложений, так как ряды расходятся.

Данная ситуация напоминает известный парадокс: приближенные решения, полученные с помощью нормальных форм, хорошо приближают точные решения, несмотря на заведомую расходимость нормализующих преобразований.

Таким образом, нужно либо каким-то способом просуммировать эти формальные ряды, либо доказать существование аналитической функции, имеющей данные асимптотические разложения.

В этой работе мы намереваемся сделать и то и другое.

3. ПЕРВЫЙ ИНТЕГРАЛ НЕПРЕРЫВНОЙ ИТЕРАЦИИ

К сожалению, применить теорему Бореля–Ритга напрямую к формальным рядам, соответствующим инвариантным кривым НИ, не представляется возможным, так как эти ряды зависят от двух переменных. Такие ряды ведут себя значительно сложнее обычных уже в случае их схожести, а расходящиеся ряды от двух и более переменных вообще мало изучены.

Найдем функцию $H(x, y)$ (формальный ряд) такую, что ее линии уровня $H(x, y) = C$ будут соответствовать инвариантным кривым НИ. Положим формально

$$0 = H(x + t, F^{(t)}(y)) - H(x, y) = A(t), \quad (11)$$

где “время” x соответствует начальному значению $y = y(x)$. Сдвиг по времени t будем считать малым.

Теорема 1. Для любого формального ряда (1) существует аналитический первый интеграл $H(x, y)$, линии уровня которого соответствуют инвариантным кривым непрерывной итерации при достаточно малых $y \in \mathbb{C}$.

Доказательство. Нужно доказать, что функция $A(t)$ в (11) является плоской, т.е. имеет тождественный ноль в качестве степенного асимптотического разложения (в некотором секторе \mathbb{C}) для подходящей функции $H(x, y)$. Таким образом, первый интеграл (если существует) не является единственным. Однако он не может быть единственным даже в случае, если все ряды заведомо сходятся, поэтому неединственность $H(x, y)$ нас не должна беспокоить.

Запишем функцию $A(t)$ в (11) в виде

$$A(t) = H\left(x + t, y + g_1(y)t + g_2(y)\frac{t^2}{2!} + \dots\right) - H(x, y),$$

используя формулу (3), выражающую НИ с помощью ее генератора. Тогда

$$A(t) = \left(\frac{\partial H}{\partial x} + \frac{\partial H}{\partial y} \omega(y)\right)t + O(t^2).$$

Поэтому в первом приближении необходимо

$$H(x, y) = G\left(x - \int \frac{dy}{\omega(y)}\right), \quad (12)$$

где G — это произвольная регулярная функция. Имея в виду неединственность первого интеграла, положим $G() = 1$.

Далее необходимо найти поправки к $A(t)$ более высокого порядка по t . Однако в этом нет необходимости, так как, правильно выбрав пределы интегрирования в (12), мы получаем точное решение:

$$H(x, y) = x - \int_y^{F^{(x)}(y)} \frac{dt}{\omega(t)} + \text{const.} \quad (13)$$

Доказательство этой формулы можно вывести из представления НИ (3) и формулы (6), однако это уже сделано ранее (см. [6, (88)]).

Интеграл (13) является пока формальным. Запишем его в форме, соответствующей обозначениям в (8), т.е. положим $y_0 = C$ при времени $x = 0$ и $y = y(x)$ при времени x . Тогда получим другое выражение для инвариантной кривой (8), где переменные разделены:

$$x = \int_C^y \frac{dt}{\omega(t)}. \quad (14)$$

Поскольку подынтегральное выражение в (14) является формальным рядом одной переменной (независимо от сходимости или расходимости ряда (1)), ему соответствует некоторая голоморфная функция в некотором секторе комплексной плоскости (теорема Бореля–Ритта, см. [9, с. 43]). Почленно интегрируя асимптотическое разложение $1/\omega(t)$ в (14) по пути $t \in [C, y]$, который следует выбрать в этом секторе, получим аналитическую функцию $x = x(y)$, а также утверждение теоремы. Теорема доказана.

Например, взяв генератор $\omega(y)$ (10) для логистического отображения (9), получим численно

$$1 \approx \int_y^{y-y^2} \frac{dt}{\omega(t)},$$

причем тем точнее, чем меньше $|y|$, $y \in \mathbb{C}$. Дифференцируя эту формулу, получаем функциональное уравнение для генератора (10):

$$\omega(y - y^2) = \omega(y)(1 - 2y), \quad (15)$$

откуда можно получить рекуррентную формулу для его коэффициентов.

Как мы отмечали, все существующие примеры аналитических НИ являются алгебраическими и даются формулой (7). Дадим пример трансцендентной НИ, соответствующей генератору $\omega(y) = y^2 + y^3$. Формула (14) дает:

$$x = \log \left(\frac{(1+y)C}{(1+C)y} \right) + \frac{1}{C} - \frac{1}{y}.$$

Вернемся к обозначениям в (13) и обратим эту формулу. Тогда получим

$$F^{(x)}(y) = - \left(1 + W_s \left(- \frac{1+y}{y} \exp \left(x - \frac{1+y}{y} \right) \right) \right)^{-1}, \quad (16)$$

где $W_s(\cdot)$ – это функция Ламберта, а s – это ее ветвь: $s = 0$ при $y < 0$ и $s = -1$ при $y > 0$.

Разложить формулу (16) в ряд по y непосредственно не удастся. Следует сперва разложить ее в ряд по x , а затем воспользоваться тождеством $y = W(y \exp(y))$. Тогда получим ряд, соответствующий генератору по формуле (4).

Существует другой подход к вычислению первого интеграла $H(x, y)$, соответствующего инвариантным кривым отображения (1) (см. [1]). При этом мы не используем концепцию НИ, а рассматриваем отображение (1) как ДДС:

$$\left\{ x_{n+1} = x_n + 1, \quad y_{n+1} = y_n + \sum_{k=1}^{\infty} c_k y_n^{k+1} \right\}, \quad (17)$$

где переменная x_n играет роль дискретного времени.

Очевидно, задача вычисления первого интеграла $H(x, y)$ такого, что $H(x_{n+1}, y_{n+1}) = H(x_n, y_n)$, $n \in \mathbb{Z}$, эквивалентна вычислению НИ отображения (1), т.е. восстановлению НДС по ее дискретной редукции.

Сделаем преобразование масштабирования в системе (17):

$$x \rightarrow \frac{x}{h}, \quad y \rightarrow hy,$$

где h – это малый параметр. Затем умножим первое уравнение (17) на h , а второе уравнение (17) разделим на h . Тогда получим ДДС:

$$\left\{ x_{n+1} = x_n + h, \quad y_{n+1} = y_n + \sum_{k=1}^{\infty} c_k h^k y_n^{k+1} \right\}. \quad (18)$$

Вычислим первый интеграл $H(x, y) = \text{const}$ для системы (18), а затем проделаем обратное масштабирование, что эквивалентно подстановке $h = 1$ в полученные ниже формулы.

Положим формально $x_n = x$, $y_n = y$ и рассмотрим тождество

$$0 = H\left(x + h, y + \sum_{k=1}^{\infty} c_k h^k y^{k+1}\right) - H(x, y) = \left(\frac{\partial H(x, y)}{\partial x} + c_1 y^2 \frac{\partial H(x, y)}{\partial y}\right) h + O(h^2). \quad (19)$$

Выбирая простейшее частное решение, получаем

$$H(x, y) = c_1 x + \frac{1}{y} + h H_1(x, y) + O(h^2).$$

Используем эту формулу в (19). Тогда, опуская довольно громоздкие, но очевидные выкладки, получаем

$$0 = \left((c_1^2 - c_2)y + \frac{\partial H_1(x, y)}{\partial x} + c_1 y^2 \frac{\partial H_1(x, y)}{\partial y}\right) h^2 + O(h^3).$$

Далее для простоты будем считать $c_1 \neq 0$. В случае $c_1 = 0$ масштабирование следовало делать по-другому. Тогда получим

$$H(x, y) = c_1 x + \frac{1}{y} + \frac{h}{2} \left(\frac{c_2}{c_1} - c_1\right) \log(y^2) + h^2 H_2(x, y) + O(h^3).$$

Используем эту формулу в (19). Получим

$$H_2(x, y) = \frac{y}{2c_1^2} (2c_1 c_3 - c_1^2 c_2 + c_1^4 - 2c_2^2).$$

Далее, как можно убедиться с помощью математической индукции, в разложении $H(x, y)$ по параметру h получим формальный ряд по y . Поэтому далее положим $h = 1$ и будем искать функцию $H(x, y)$ в виде

$$H(x, y) = c_1 x + \frac{1}{y} + \frac{1}{2} \left(\frac{c_2}{c_1} - c_1\right) \log(y^2) + U(y), \quad (20)$$

где

$$U(y) = H_2(x, y) + O(y^2).$$

Тождество $H(x + 1, F(y)) = H(x, y)$, где $F(y)$ – это формальный ряд (1), дает функциональное уравнение для функции $U(y)$:

$$U(y) - U(F(y)) = c_1 - \frac{1}{y} + \frac{1}{F(y)} + \left(\frac{c_2}{c_1} - c_1\right) \log\left(\frac{F(y)}{y}\right). \quad (21)$$

Предполагая, что функцию $U(y)$ можно каким-либо образом вычислить по уравнению (21), мы получили первый интеграл (20), т.е. восстановили НДС по ее дискретной редукции.

Однако на самом деле мы не получили ничего нового, кроме уравнения (21).

Предложение 1. *Первый интеграл (20) совпадает с интегралом (14).*

Доказательство. Положим $x = x(y)$ в формулах (14) и (20), затем продифференцируем их y и приравняем результаты. Получим

$$x'(y) = \frac{1}{\omega(y)} = \frac{1}{c_1 y^2} + \frac{c_1^2 - c_2}{c_1^2 y} - \frac{1}{c_1} U'(y). \quad (22)$$

Но это не что иное, как асимптотическое разложение подынтегрального выражения в (14), так как по формуле (4) получим

$$\omega(y) = c_1 y^2 + (c_2 - c_1^2) y^3 + \left(c_3 - c_1^3 - \frac{5}{2}(c_2 - c_1^2) c_1\right) y^4 + \dots$$

Поэтому если определить функцию $U(y)$ как формальный ряд, полученный из ОДУ (22), т.е.

$$U(y) = -\frac{1}{y} + \left(c_1 - \frac{c_2}{c_1}\right) \log y - c_1 \int \frac{dy}{\omega(y)}, \quad (23)$$

то получим требуемое утверждение. Предложение доказано.

Для общих отображений (1) масштабирование (т.е. искусственное введение малого параметра) следует подбирать индивидуально. Например, для итераций синуса $y_{n+1} = \sin y_n$, $n \in \mathbb{Z}$, следует взять

$$x \rightarrow \frac{x}{h}, \quad y \rightarrow \sqrt{hy},$$

а затем умножить первое уравнение (17) на h , а второе уравнение (17) разделить на \sqrt{h} .

Однако проще сразу использовать асимптотическое разложение функции $1/\alpha(y) = \dots + U'(y)$ в формуле (14), где $U(y)$ – это формальный ряд по натуральным степеням y .

Например, для итераций синуса оба подхода дадут результат:

$$H(x, y) = -\frac{1}{3}x + \frac{1}{y^2} + \frac{1}{5}\log(y^2) + U(y),$$

где $U(y) = 79/3150y^2 + 29/7875y^4 + \dots$ удовлетворяет функциональному уравнению

$$U(y) - U(\sin(y)) = -\frac{1}{3} + \frac{y^2 - \sin^2(y)}{y^2 \sin^2(y)} - \frac{2}{5} \log\left(\frac{y}{\sin(y)}\right).$$

Эти формулы могут быть использованы для вычисления асимптотики итераций синуса. Можно проверить, что получится в точности то же, что мы получили эмпирически в [1].

Функциональное уравнение (21) для функции $U(y)$, введенное таким образом, дает, как мы увидим, эффективный способ ее вычисления (см. разд. 5).

4. АНАЛИТИЧЕСКИЕ СВОЙСТВА ПЕРВОГО ИНТЕГРАЛА НЕПРЕРЫВНОЙ ИТЕРАЦИИ

Как отмечалось в [10], существование аналитической функции, соответствующей асимптотическому разложению, представленному формальным степенным рядом, и вычисление этой функции – это две совершенно разные задачи.

Конструктивное вычисление аналитической функции $H(x, y)$, соответствующей НИ отображения (1), подразумевает также изучение аналитических свойств этой функции. В частности, изучение ее особенностей.

Ясно, что в общем случае это сделать невозможно, поэтому далее мы сосредоточимся на изучении аналитического интеграла $H(x, y)$ для функции $F(y) = y - y^2$, т.е. для логистического отображения (9).

Для этого отображения $c_1 = -1$, $c_n = 0$, $n > 1$. Поэтому первый интеграл (20) принимает вид

$$H(x, y) = -x + \frac{1}{y} + \frac{1}{2}\log(y^2) + U(y), \quad (24)$$

а функция $U(y)$ удовлетворяет функциональному уравнению

$$U(y) - U(y - y^2) = \frac{y}{1 - y} + \frac{1}{2}\log(1 - y)^2. \quad (25)$$

Это те же уравнения, которые были получены в [1].

Напомним, что для коэффициентов формального ряда $U(y)$ существует рекуррентное соотношение:

$$U(y) = \sum_{n=1}^{\infty} u_n (-y)^n = \frac{1}{2}y + \frac{1}{3}y^2 + \frac{13}{36}y^3 + \frac{113}{240}y^4 + \frac{1187}{1800}y^5 + \frac{877}{945}y^6 + \dots, \quad (26)$$

где

$$u_n = \frac{(-1)^n}{n+1} - \frac{1}{n} \sum_{k=\lfloor (n+1)/2 \rfloor}^{n-1} C(k, n+1-k) u_k, \quad n \in \mathbb{N},$$

где $[x]$ – это целая часть x , и $C(n, m)$ – это биномиальный коэффициент.

Из результатов разд. 3 следует, что расходимость формального разложения (26) функции $U(y)$ равносильна расходимости ряда (10) для генератора $\omega(y)$ НИ отображения (9). А расходимость ряда (10) влечет, в свою очередь, расходимость рядов, представляющих инвариантные кривые НИ этого отображения.

Доказательство расходимости рядов НИ для логистического отображения (9) являлось предметом диссертации (см. [11]). Однако мы не будем опираться на эти результаты, а докажем расходимость ряда для функции $U(y)$ прямо из уравнения (25).

Теорема 2. *Формальный ряд (26) всюду расходится, так как точка $y = 0$ является предельной точкой особенностей функции $U(y)$.*

Доказательство. Напомним, что функция $U(y)$ является аналитической в некотором секторе, примыкающем к точке $y = 0$ (см. теорему 1).

Точка $y_0 = 1$, очевидно, является особой для функции $U(y)$, что следует из уравнения (25). По той же причине точки

$$y_1^\pm = \frac{1}{2} \pm \frac{i}{2}\sqrt{3}, \quad i^2 = -1,$$

являются особыми как прообразы точки y_0 , т.е. $y - y^2 = y_0$, $y = y_1^\pm$.

Далее будем учитывать только точки в верхней полуплоскости \mathbb{C} . Поэтому получим, что все точки

$$y_{n+1} = F^{(-1)}(y_n) = \frac{1}{2} - \sqrt{\frac{1}{4} - y_n}, \quad n \in \mathbb{N}, \quad (27)$$

являются особыми.

Например, $y_{10} \approx -0.12356942653420 + 0.037052550159283i$.

Но последовательность точек (27) стремится к нулю, так как имеет асимптотику

$$y_n = -\frac{1}{n} + \frac{C - \log n}{n^2} - \frac{\frac{1}{2} + C + C^2 - (1 + 2C)\log n + \log^2 n}{n^3} + O\left(\frac{1}{n^4}\right), \quad (28)$$

которая вычисляется методом неопределенных коэффициентов.

На самом деле, это в точности та же асимптотика, которую мы нашли в [1, (21)] с помощью первого интеграла (24). Нужно лишь заменить там t на $-1/n$ и правильно подобрать константу C . Для последовательности (27), $C \approx -0.46947082 + 2.50176142i$. Терма доказана.

Продолжим изучение особенностей функции $U(y)$.

Рассуждая так же, как в доказательстве теоремы 2, получаем, что все точки в комплексной плоскости, которые за *конечное* число итераций $y \rightarrow y - y^2$ попадают в ноль, являются особыми для функции $U(y)$.

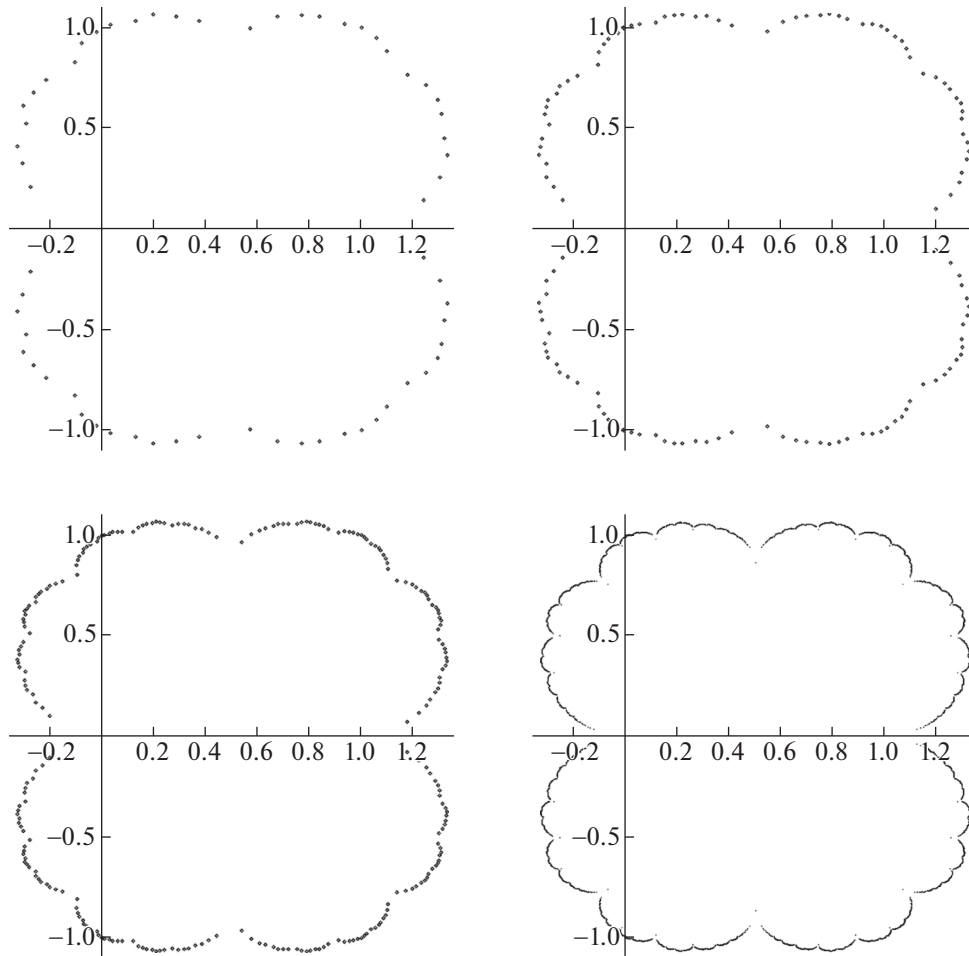
Но это еще далеко не все ее особенности. Например, подставим в уравнение (25) $y = 1 + i$, затем $y = 1 - i$ и сложим эти уравнения. Тогда получим $0 = -2 + i\pi$. Поэтому точки $y = 1 \pm i$ также являются особыми.

Рассуждая аналогично, получим, что все периодические орбиты отображения (9), $y = F^{(n)}(y)$, $n \in \mathbb{N}$, состоят из особых точек функции $U(y)$. А следовательно, и все точки, которые за конечное число итераций $y \rightarrow y - y^2$ попадают в одну из точек орбиты, также являются особыми.

На самом деле множество всех особых точек функции $U(y)$ — это не что иное, как *множество Жюлиа* \mathcal{J} отображения (9) (см. [12, с. 53]), так как замыкание объединения указанных выше особых множеств также состоит из особенностей функции $U(y)$.

На фиг. 1 показаны корни полиномов $y = (y - y^2)^{(n)}$, для $n = 6, 7, 8$, и множество Жюлиа \mathcal{J} для примерно 45000 особых точек.

Если расположить корни y_j полинома $y = (y - y^2)^{(n)}$ в порядке возрастания аргумента $\arg(y_j - 1/2)$, то можно посчитать периметр P_n многоугольника с данными вершинами. Оказалось, что $P_{n+1}/P_n \approx 1.022$, $n > 6$. Поэтому периметр фигур на фиг. 1 возрастает неограниченно.



Фиг. 1. Корни полиномов $y = (y - y^2)^n$, $n = 6, 7, 8$, и множество Жюлиа \mathcal{J} .

А это означает, что множество Жюлиа \mathcal{J} устроено фрактально, подобно звезде фон Коха. В частности, оно не является спрямляемой жордановой кривой.

Аналогичное множество Жюлиа для отображения

$$z \rightarrow z^2 + (0.99 + 0.14i)z,$$

показано более подробно в [13, р. 42, фиг. 5а]. Там оно было названо “простой замкнутой кривой” и “довольно дикой жордановой кривой”. На наш взгляд, это множество также является фракталом.

Разумеется, точное вычисление фрактальной размерности множества \mathcal{J} является отдельной и весьма сложной задачей.

Множество \mathcal{J} является по определению замкнутым. Открытое множество $\mathbb{C} \setminus \mathcal{J}$ распадается на две отдельные компоненты: \mathcal{F}_0 и \mathcal{F}_∞ . Это множества Фату (см. [13, р. 39]), в которых любая точка y попадает, соответственно, в ноль или бесконечность за бесконечное число итераций $y \rightarrow y - y^2$.

Множества Фату \mathcal{F}_0 и \mathcal{F}_∞ — это, соответственно, внутренняя и внешняя части множества Жюлиа (см. фиг. 1).

По теореме Бореля–Ритта функция $U(y)$ является голоморфной в некотором секторе, примыкающем к началу координат. Однако из уравнения (25) можно извлечь гораздо больше.

Теорема 3. Функция $U(y)$ является голоморфной в области Фату \mathcal{F}_0 .

Доказательство. Рассмотрим любую точку $y_0 \in \mathcal{F}_0$. По определению,

$$\lim_{n \rightarrow \infty} y_n = 0,$$

где $y_n = y_{n-1} - y_{n-1}^2$. Перепишем уравнение (25) в виде

$$U(y_n) = U(y_{n+1}) + \frac{y_n}{1 - y_n} + \log(1 - y_n), \quad n \in \mathbb{N}_0,$$

где логарифм мы упростили, так как он здесь однолистен. Затем просуммируем эти уравнения по $n \in \mathbb{N}_0$. Таким образом, имеем (пока формально)

$$U(y_0) = \sum_{n=0}^{\infty} \left(\frac{y_n}{1 - y_n} + \log(1 - y_n) \right) = \sum_{n=0}^{\infty} \left(\frac{1}{2} y_n^2 + \frac{2}{3} y_n^3 + O(y_n^4) \right). \quad (29)$$

В последней сумме существенно лишь то, что общий член имеет порядок $O(y_n^2)$. Согласно [1, (21)], асимптотика последовательности $\{y_n\}$, $n \in \mathbb{N}$, дается формулой

$$y_n = \frac{1}{n} + \frac{C - \log n}{n^2} + \frac{\frac{1}{2} + C + C^2 - (1 + 2C) \log n + \log^2 n}{n^3} + O\left(\frac{1}{n^4}\right), \quad (30)$$

где константа $C = C(y_0)$. Таким образом, общий член в сумме (29) имеет порядок $O(1/n^2)$, поэтому этот ряд сходится для любой точки $y_0 \in \mathcal{F}_0$.

Обозначим $y = y_0$ и перепишем формулу (29) в виде

$$U(y) = \sum_{n=0}^{\infty} G_n(y), \quad (31)$$

где $G_n(y)$ — это голоморфная функция в некоторой окрестности точки y . Согласно данной выше оценке, этот ряд сходится равномерно. Согласно теореме Вейерштрасса (см. [14, р. 249]), функция $U(y)$ также является голоморфной. Теорема доказана.

Формула (22) для отображения (9) имеет вид

$$U'(y) = \frac{1}{\omega(y)} + \frac{1 - y}{y^2}, \quad (32)$$

где $\omega(y)$ дана в (10). Поэтому вычислим также функцию $U'(y)$.

Ряд (31) (он же ряд (29)) можно почленно дифференцировать. Имеем

$$U'(y_0) = \sum_{n=0}^{\infty} \frac{d}{dy_n} \left(\frac{y_n}{1 - y_n} + \log(1 - y_n) \right) \frac{dy_n}{dy_0} = \sum_{n=0}^{\infty} \frac{y_n}{(1 - y_n)^2} \frac{dy_n}{dy_0}.$$

По правилу дифференцирования сложной функции получаем

$$U'(y_0) = \sum_{n=0}^{\infty} \frac{y_n}{(1 - y_n)^2} \prod_{k=0}^{n-1} (1 - 2y_k). \quad (33)$$

Оценим асимптотику произведения в формуле (33). Согласно асимптотике (30), это произведение расходится к нулю. Точнее,

$$\prod_{k=m}^n (1 - 2y_k) \asymp \exp(-2\Psi(n) + 2\Psi(m)) = O\left(\frac{1}{n^2}\right),$$

где мы оставили только главный член асимптотики, $y_k \asymp 1/k$. Поэтому ряд (33) также сходится (хотя мы знали это заранее по теореме Вейерштрасса, (см. [14, р. 249])).

Заметим, что мы только что еще раз доказали теорему 3, так как показали существование комплексной производной функции $U(y)$ в области Фату \mathcal{F}_0 .

Формула (33) позволяет вычислить функцию $U'(y)$ в некоторых точках в явном виде, так как ряд (33) обрывается при $y = 1/2$ и во всех ее прообразах. Например, $U'(1/2) = 2$. Поскольку уравнение $y - y^2 = 1/2$ имеет корни $y_{\pm} = 1/2 \pm i/2$, то

$$U'\left(\frac{1}{2} \pm \frac{i}{2}\right) = -1 \mp i, \quad \text{и т.д.}$$

Теорема 4. Генератор (10) $\omega(y)$ имеет счетное множество простых полюсов в области Фату \mathcal{F}_0 и не имеет там других особенностей. Эти полюса накапливаются вблизи множества Жюлиа \mathcal{J} . Кроме того, $\omega(y_*) = 0$ на счетном всюду плотном множестве точек $y_* \in \mathcal{J}$.

Доказательство. По формуле (32) получаем, что $\omega(1/2) = \infty$, т.е. точка $y = 1/2$ является особой для генератора (10). Поэтому все прообразы точки $y = 1/2$ также являются особыми для $\omega(y)$, что видно из функционального уравнения (15), которое можно записать в виде

$$\omega(y_0) = \omega(y_n) \left(\prod_{k=0}^{n-1} (1 - 2y_k) \right)^{-1}. \quad (34)$$

Все прообразы точки $y = 1/2$ являются простыми полюсами по этой формуле. Кроме того, счетное множество этих полюсов функции $\omega(y)$ в области Фату \mathcal{F}_0 не может иметь точку сгущения внутри этой области. Иначе это была бы существенно особая точка для $\omega(y)$, а стало быть и для функции $U'(y)$, которая, согласно теореме 3, является голоморфной в \mathcal{F}_0 .

По той же причине $\omega(y)$ не может иметь других существенных особенностей либо точек ветвления внутри \mathcal{F}_0 .

Других полюсов, кроме указанных, т.е. прообразов точки $y = 1/2$, у функции $\omega(y)$ внутри \mathcal{F}_0 быть не может. Иначе под действием итераций $y \rightarrow y - y^2$ эти полюса неизбежно попадут в любой сектор комплексной плоскости (с любым раствором), примыкающий к точке $y = 0$, что противоречит теореме Бореля–Ритта.

Это последнее утверждение следует из предложения 2, доказанного ниже.

Осталось найти нули функции $\omega(y)$.

Так как $\omega(0) = 0$ по определению, то из формулы (34) получаем, что все точки $y_* \in \mathcal{J}$, которые за конечное число итераций попадают в ноль, также являются нулями $\omega(y)$.

Для периодических орбит в \mathcal{J} система уравнений (34) имеет единственное решение $\omega(y) = 0$. Это же справедливо для прообразов этих точек.

Но эти множества точек всюду плотны в множестве Жюлиа \mathcal{J} . Терма доказана.

Разумеется, множество \mathcal{J} также является особым для функции $\omega(y)$. Причем, по-видимому, функция $\omega(y)$ может принимать на нем любое комплексное значение.

Это видно из формулы (34): для каждой точки $y_* \in \mathcal{J}$ выберем близкую к ней точку $y_0 \in \mathcal{F}_0$. Имеем $y_n \rightarrow 0$ внутри \mathcal{F}_0 , причем $y_n = O(1/n)$. Поэтому $\omega(y_n) = O(1/n^2)$. Но произведение в (34) расходится к нулю с той же асимптотикой.

Таким образом, генератор НИ (10) простейшего из всех возможных нетривиальных отображений (1) устроен весьма диким образом.

Тем не менее в следующем разделе мы покажем, что генератор $\omega(y)$, а также функции $U(y)$ и $U'(y)$ эффективно вычисляются в своих регулярных точках внутри \mathcal{F}_0 .

В завершение этого раздела приведем асимптотическую формулу для итераций отображения $y \rightarrow y - y^2$ в координатах $u = \operatorname{Re}(y)$, $v = \operatorname{Im}(y)$.

Для этого в формулу [1, (21)], т.е. в (30), нужно подставить $n = 1/t$, $C = A + iB$. Затем отделить вещественную и мнимую части разложения, что даст $u = t + O(t^2)$, $v = Bt^2 + O(t^3)$. Затем нужно обратить ряд для $u = u(t)$, что можно сделать методом неопределенных коэффициентов, либо применить алгоритм обращения степенно-логарифмических рядов, основанный на НИ (см. [1]).

Получим степенно-логарифмический ряд $t = u + O(u^2)$. Затем этот последний нужно подставить в ряд для $v = v(t)$ и переразложить его по u . В результате получим

$$v = Bu^2 + Bu^3 + \frac{1}{2}B(3 + 2B^2)u^4 + \frac{1}{3}B(8 + 11B^2)u^5 + O(u^6). \quad (35)$$

Непосредственно из интеграла (24) извлечь эту асимптотику не представляется возможным.

Другой способ вычислить эту асимптотику – это использовать первый интеграл (14). Запишем его в виде

$$\frac{dy(x)}{dx} = \omega(y(x)), \quad (36)$$

и используем разложение (10). Затем подставим в него $y(x) = u(x) + iv(x)$ и отделим вещественную и мнимую части. Затем запишем (весьма громоздкое) уравнение $dv/du = \dots$. Далее убеждаемся с помощью CAS, что это ОДУ имеет решение (35).

Любопытно, что логарифм и вещественная часть константы C не входят в разложение (35), что можно доказать следующим образом.

Предложение 2. *Разложение (35) продолжается неограниченно и является степенным с одним вещественным параметром.*

Доказательство. Мы дадим то, что называют “computer assisted proof”, так как без CAS воспроизвести это доказательство затруднительно.

Имеем, $y_n = u_n + iv_n$, $y_{n+1} = u_{n+1} + iv_{n+1} = y_n - y_n^2$. Поэтому

$$(a): u_{n+1} = u_n - u_n^2 + v_n^2,$$

$$(б): v_{n+1} = v_n - 2u_nv_n.$$

Мы уже имеем $v = Bu^2 + O(u^3)$, где B – это произвольная вещественная константа. Положим $v = Bu^2 + Xu^3 + O(u^4)$, где X – это неизвестный коэффициент. Тогда имеем

$$(в): v_n = Bu_n^2 + Xu_n^3 + \dots,$$

$$(г): v_{n+1} - (Bu_{n+1}^2 + Xu_{n+1}^3 + \dots) = 0.$$

Подставим (а) и (б) в уравнение (г), затем подставим туда (в) и $u_n = u$. Получим выражение, которое должно обращаться в ноль до любого порядка по u . Поэтому разложим его в ряд и найдем X из линейного уравнения при первой ненулевой степени u (т.е. при u^4). Получим $X = B$.

Далее положим $v = Bu^2 + Bu^3 + Xu^4 + O(u^5)$, и т.д.

Поскольку мы имеем дело только с умножением полиномов, то все полученные линейные уравнения для X однозначно разрешимы. Предложение доказано.

Формула (35) описывает как стремление $y \rightarrow 0$ внутри области Фату \mathcal{F}_0 под действием итераций $y \rightarrow y - y^2$, так и асимптотику множества Жюлиа \mathcal{J} вблизи нуля (для, возможно, канторова множества констант B).

5. КОНСТРУКТИВНОЕ ВЫЧИСЛЕНИЕ ПЕРВОГО ИНТЕГРАЛА НЕПРЕРЫВНОЙ ИТЕРАЦИИ

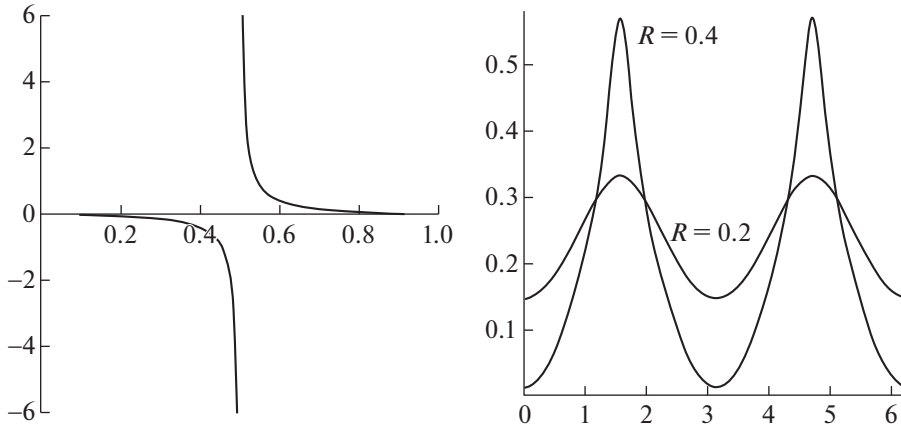
В предыдущем разделе мы показали, что функция $U(y)$, а также и $U'(y)$, и $\omega(y)$, однозначно вычисляются в своих регулярных точках в области Фату \mathcal{F}_0 с помощью сходящихся рядов. Область Фату \mathcal{F}_∞ мы далее рассматривать не будем, так как там указанные функции будут заведомо многолиственны.

Функция $\omega(y)$, очевидно, не может быть выражена в виде конечных комбинаций из известных специальных и элементарных функций ввиду особого характера ее множества особенностей. Тем не менее можно считать эту функцию, в некотором смысле, известной, так как мы знаем ее особенности и умеем ее вычислять в регулярных точках.

Практическое вычисление этих функций, разумеется, не предполагает фактического суммирования рядов. Вместо этого можно воспользоваться подходящим алгоритмом ускорения сходимости рядов (см., например, [15]). Однако есть гораздо более эффективный способ суммирования.

Для этого вычисляется лишь конечное и, как оказалось, небольшое количество членов ряда до члена, где величина $|y_n|$ достаточно мала. Затем используются соответствующие асимптотические разложения функций для вычисления остаточного члена.

Эта последняя операция также может быть улучшена с помощью Паде–аппроксимаций или другого подходящего преобразования (см. [10]).



Фиг. 2. Графики функций $\omega(y)$, $y \in [0, 1]$, и $|\omega(1/2 + R\exp(i\phi))|$, $R = 0.2$ и $R = 0.4$.

Заметим, что большинство из известных специальных (и даже элементарных) функций практически (т.е. в компьютерной процедуре) вычисляются именно этим способом (пример: эллиптические функции и преобразование Ландена).

Проще всего указанный алгоритм реализуется для функции $\omega(y)$ с помощью формулы (34). Для любой неособой точки $y_0 \in \mathcal{F}_0$ вычисляются итерации (9), т.е. последовательность точек y_n до номера n , где величина $|y_n|$ достаточно мала. Для точек, не слишком близких к границе области, количество итераций $y \rightarrow y - y^2$ невелико. Разумеется, следует предусмотреть остановку программы, если количество итераций велико (скажем, больше 1000). По мере вычисления точек y_n накапливается произведение в (34). Затем вычисляется величина $\omega(y_n)$ по разложению (10) и применяется (34).

Этим способом функция $\omega(y)$ (а значит, и $U'(y)$, см. (32)) может быть вычислена (в принципе) с произвольной и контролируемой точностью.

На фиг. 2 показаны график функции $\omega(y)$ на вещественном интервале $y \in [0, 1]$, и график функции $|\omega(1/2 + R\exp(i\phi))|$ на окружности вокруг особой точки $y = 1/2$ в комплексной плоскости при $R = 0.2$ и $R = 0.4$.

Тот факт, что функция $\omega(y)$ оказалась нечетной относительно своей особенности $y = 1/2$, виден из функционального уравнения (15), которое можно преобразовать к виду

$$\lambda\left(-\frac{1}{4} - z^2\right) = -2z\lambda(z), \quad \omega(z) = \lambda\left(z - \frac{1}{2}\right), \quad z = y - \frac{1}{2}. \tag{37}$$

Эта замена ($y_n = z_n + 1/2$) приводит отображение (9) к виду $z_{n+1} = -1/4 - z_n^2$, что объясняет симметрию множества Жюлиа на фиг. 1.

Вычисления дают приближение функции $\omega(y)$ вблизи ее полюса $y = 1/2$:

$$\omega(y) \approx 0.044583633/(y - 0.5) - 0.453007107(y - 0.5) + O((y - 0.5)^3).$$

Предложение 3. Функция $y(x)$ непрерывной итерации отображения (9) имеет в \mathbb{C} счетное множество точек ветвления второго порядка.

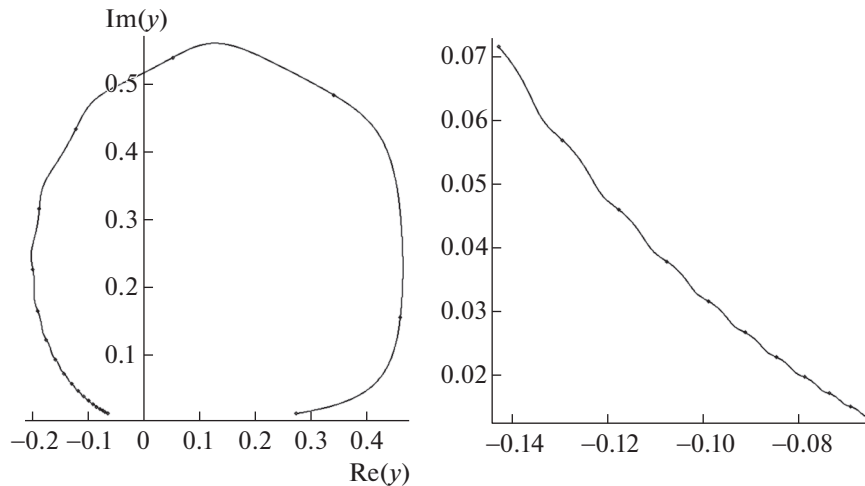
Доказательство. Функция $\omega(y)$ имеет счетное множество простых полюсов (см. теорему 4). Запишем ОДУ (36) в локальных координатах в окрестности одного из ее полюсов

$$y'(x) = \frac{a^2}{2y(x)} + b + cy(x) + \dots,$$

где a, b, c – это некоторые комплексные константы. Тогда локально

$$y(x) = ax^{1/2} + \frac{2}{3}bx + \frac{9a^2c + 2b^2}{18a}x^{3/2} + O(x^2).$$

Предложение доказано.



Фиг. 3. Непрерывные и дискретные итерации функции $y(x) = (y_0 - y_0^2)(x)$.

Уравнение (36) может быть использовано для эффективного вычисления НИ логистического отображения для комплексных $y \in \mathcal{F}_0$ при некоторых ограничениях, связанных с предложением 3. Дело в том, что нас интересует главным образом вещественное “время” x . Поэтому кривая НИ должна избегать точек ветвления $y(x)$. Это возможно, например, при $\text{Re}(y_0) < 1/2$, но невозможно при $\text{Re}(y_0) > 1/2$ в связи с наличием полюса $y = 1/2$ у функции $\omega(y)$.

На фиг. 3 показан график НИ отображения (9) для комплексного начального значения $y_0 = -0.065 + i0.0135$, которое расположено вблизи границы области Фату \mathcal{F}_0 , для вещественного времени $x \in [0, 20]$. Показаны также дискретные итерации $y \rightarrow y - y^2$. Правый фиг. 3 – это фрагмент левого фиг. 3, на котором видно, что кривая НИ ведет себя контринтуитивно, демонстрируя осцилляции. Это, разумеется, вызвано близостью множества Жюлиа, где накапливаются особенности функции $\omega(y)$.

Если продолжить кривую НИ на фиг. 3 для бóльших значений x так, что $|y(x)|$ станет достаточно мало, затем ввести новое локальное время x , т.е. начать с $x = 0$, то кривые НИ снова можно будет вычислять по их разложению, как мы это делали в [1] (см. также разд. 2).

Разумеется, аналитическая кривая на фиг. 3 вычисляется также из первого интеграла (24), но для этого нужно сначала вычислить функцию $U(y)$.

Следуя [1], введем новую функцию $V(y)$, которая не имеет особенностей на вещественной оси. Сделаем замену

$$U(y) = \log(1 - y) + \frac{y}{1 - y} - V(y)$$

в уравнении (25). Получим уравнение для функции $V(y)$:

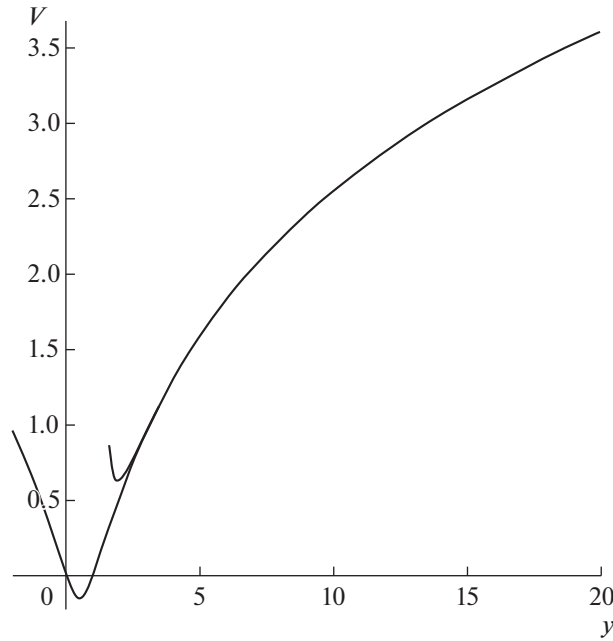
$$V(y - y^2) - V(y) = \log(y^2 - y + 1) + \frac{y(1 - y)}{y^2 - y + 1}. \quad (38)$$

Тогда интеграл (24) принимает вид

$$H(x, y) = -x + \log(y - y^2) + \frac{1}{y - y^2} - V(y). \quad (39)$$

Следуя доказательству теоремы 3, запишем уравнение (38) в виде

$$V(y_n) - V(y_{n-1}) = \log(1 - y_n) + \frac{y_n}{1 - y_n},$$



Фиг. 4. График функции $V(y)$ и ее асимптотики при $y \rightarrow +\infty$.

и просуммируем его по n . Тогда получим

$$V(y_0) = V(y_n) - \sum_{k=1}^n \left(\log(1 - y_k) + \frac{y_k}{1 - y_k} \right). \tag{40}$$

Функция $V(y)$ вычисляется по этой формуле внутри области Фату \mathcal{F}_0 точно так же, как описано в этом разделе для функции $\omega(y)$. Отличие состоит в том, что накапливается сумма, а не произведение, и $V(y)$ голоморфна внутри \mathcal{F}_0 .

Функцию $V(y)$ можно также вычислить при $y < 0$. Запишем уравнение (38) в виде

$$V(z_n) = V(z_{n+1}) + \log(1 - z_n) + \frac{z_n}{1 - z_n}, \quad z_{n+1} = \frac{1}{2}(1 - \sqrt{1 - 4z_n}),$$

и просуммируем его по n . Получим уравнение, аналогичное (40), по которому $V(y)$ вычисляется при $y < 0$ (см. (28) и (30)).

Функция $V(y)$ удобна тем, что она ограничена на интервале $y \in [0, 1]$ и является четной по отношению к точке $y = 1/2$. Это видно после замены $y \rightarrow 1 - y$ в уравнении (38) (см. также (37)). Поэтому (38) можно записать в виде

$$V(y^2 - y + 1) - V(y) = \log(y^2 - y + 1) + \frac{y(1 - y)}{y^2 - y + 1}.$$

Это уравнение позволяет вычислить асимптотику функции $V(y)$ при $y \rightarrow +\infty$ (см. [1]):

$$V(y) = A + 2 \log y - \frac{\log \log y}{\log 2} - \frac{1}{y} \left(1 - \frac{1}{2 \log 2 \log y} \right) - \frac{1}{y^2} \left(\frac{3}{2} - \frac{1}{8 \log 2 \log^2 y} \right) + O\left(\frac{1}{y^3}\right), \tag{41}$$

где $A \approx -0.7641492$ – это однозначно определяемая константа.

Таким образом, функция $V(y)$ однозначно определена на всей вещественной оси (хотя в области Фату \mathcal{F}_∞ эта функция будет многолистной).

На фиг. 4 показаны график функции $V(y)$, вычисленной по описанному алгоритму, а также график ее асимптотики (41), которую мы посчитали до члена $O(1/y^6)$. Видно, что график асимптотики $V(y)$, $y \rightarrow +\infty$, практически совпадает с графиком $V(y)$, но не продолжается до малых y .

Асимптотика (41) была использована в [1] для вычисления асимптотики отображения (9) при $|y| \rightarrow \infty$.

В заключение этого раздела приведем значение функции $V(y)$ в точке ее глобального минимума при $y = 1/2$:

$$V(0.5) = -0.1542881472560446692780986496974,$$

где, как мы полагаем, все знаки верны. Эти цифры совпадают для различных настроек алгоритма суммирования, а порядок погрешности ($\approx 10^{-32}$) соответствует учитываемой погрешности в асимптотическом разложении $V(y)$ в нуле.

6. ЗАКЛЮЧЕНИЕ

Итак, мы показали, что для любого формального ряда (1) существует аналитический первый интеграл для НИ (2). В частности, это означает, что всегда существуют аналитические инвариантные кривые НИ вне зависимости от сходимости или расходимости их формальных разложений.

Даже для простейшего из всех нетривиальных случаев отображения (1), т.е. для логистического отображения (9), формальный ряд генератора НИ, т.е. $\omega(y)$ (10) расходится. Мы это показали независимо от известных результатов (см. [11]) и весьма простым способом (см. теорему 2).

Тем не менее первый интеграл (14) определяет голоморфную функцию $x = x(y)$ во всей области Фату \mathcal{F}_0 , при условии, что $C \in \mathcal{F}_0$. В частности, тождество

$$\int_{y_0}^{y_n} \frac{dt}{\omega(t)} = n, \quad n \in \mathbb{Z},$$

справедливо при любом $y_0 \in \mathcal{F}_0$ и может быть проверено численно с любой (в принципе) точностью.

Таким образом, ДДС (9) является интегрируемой, хотя аналитический интеграл (39) выражается с помощью неизвестной ранее функции.

Кроме того, ДДС (9) дает пример интегрируемой системы, в которой реализуется динамический хаос. Это следует из хорошо известного хаотического поведения итераций дискретных отображений на их множествах Жюлиа (см. [13, с. 41]). В частности, любая точка $y_* \in \mathcal{F}$ с вероятностью единица близка к некоторой точке периодической орбиты на \mathcal{F} с неограниченным периодом.

Таким образом, интегрируемость в ДДС (а значит, и в НДС) не противоречит существованию хаоса в системе.

Система (9) также является в некотором роде контрпримером к широко распространенному тезису о том, что изучение нормальной формы системы заведомо проще, чем изучение системы в исходных переменных.

Дело в том, что нормализация системы всегда начинается с ее комплексификации. Но в данном случае “нормальная форма” ДДС (9) – это та же система, но в комплексных координатах.

Однако вещественная динамика в (9) совершенно тривиальна, в то время как комплексная динамика, как мы видели, совершенно нетривиальна.

И, наконец, система (9) дает еще один пример того, что так называемые “локальные методы” исследования особенностей применимы лишь к сравнительно простым случаям. В то время как исследование достаточно нетривиальной локальной задачи неизбежно приводит к необходимости изучать задачу глобально. См. также [10], где даны несколько аналогичных примеров.

СПИСОК ЛИТЕРАТУРЫ

1. *Варин В.П.* Об интерполяции некоторых рекуррентных последовательностей // Ж. вычисл. матем. и матем. физ. (в наборе). (2021).
2. *Henon M., Heiles C.* The applicability of the third integral of motion: some numerical experiments // *Astronomical J.*, 1964. V. 69. P. 73–79.
3. *Гельфрейх В.Г., Лазуткин В.Ф.* Расщепление сепаратрис: теория возмущений, экспоненциальная малость // *Успехи матем. наук.* 2001. Т. 56. Вып. 3(339). С. 79–142. (2001).

4. *Zhao H., Tian L.* Analytic invariant curves for an iterative equation related to dissipative standard map // *Math. Meth. Appl. Sci.* (2016). (wileyonlinelibrary.com) <https://doi.org/10.1002/mma.4260>
5. *Петров А.Г.* Асимптотические методы решения уравнений Гамильтона с помощью параметризации канонических преобразований // *Дифференц. ур-ния.* 2004. Т. 40. № 5. С. 626–638.
6. *Labelle G.* Sur l’Inversion et l’Iteration Continue des Series Formelles // *Europ. J. Combinatorics.* V. 1. P. 113–138 (1980).
7. *Comtet L.* *Advanced Combinatorics. The Art of Finite and Infinite Expansions.* (3rd ed.) Dordrecht-Holland: D. Reidel publ. company. (1974).
8. *Knuth D. E.* Convolution Polynomials // *arXiv:math/9207221v1.* (1992). (<http://arXiv.org/abs/math/9207221v1>).
9. *Wasow W.* *Asymptotic expansions for ordinary differential equations.* New York: Dover Publications, 1987.
10. *Варин В.П.* Факториальное преобразование некоторых классических комбинаторных последовательностей // *Ж. вычисл. матем. и матем. физ.* Т. 59. № 6. С. 1747–1770. (2018).
11. *Levin M.* MSc. Thesis. Israel Institute of Technology. (1960).
12. *Carleson L., Gamelin T.W.* *Complex Dynamics.* New York: Springer Verlag, 1992.
13. *Milnor J.* *Dynamics in One Complex Variable.* 3rd ed. Oxford: Princeton university press, 2006.
14. *Remmert R.* *Theory of Complex Functions.* New York: Springer Verlag, 1991.
15. *Weniger E. J.* Interpolation between sequence transformations // *Numerical Algorithms.* 1992. V. 3. P. 477–486.

**ОПТИМАЛЬНОЕ
УПРАВЛЕНИЕ**

УДК 517.977

**АСИМПТОТИКА РЕШЕНИЯ ЗАДАЧИ ОПТИМАЛЬНОГО
ГРАНИЧНОГО УПРАВЛЕНИЯ В ДВУХСВЯЗНОЙ ОБЛАСТИ
С РАЗЛИЧНОЙ ИНТЕНСИВНОСТЬЮ НА УЧАСТКАХ ГРАНИЦЫ**

© 2022 г. А. Р. Данилин

*620990 Екатеринбург, ул. С. Ковалевской, 16, Институт математики и механики
им. Н.Н. Красовского УрО РАН, Россия*

e-mail: dar@imm.uran.ru

Поступила в редакцию 24.03.2021 г.
Переработанный вариант 24.03.2021 г.
Принята к публикации 12.10.2021 г.

Рассматривается задача оптимального граничного управления решениями уравнения эллиптического типа в ограниченной области с гладкой границей с малым коэффициентом при операторе Лапласа и интегральными ограничениями на управление. На каждой из компонент границы интенсивность управления своя. Получено полное асимптотическое разложение по степеням малого параметра решения рассматриваемой задачи. Библ. 15.

Ключевые слова: сингулярные задачи, оптимальное управление, краевые задачи для систем уравнений в частных производных, асимптотические разложения.

DOI: 10.31857/S0044466922020077

ВВЕДЕНИЕ

Задачи оптимального управления, описываемые уравнениями в частных производных, особенно в постановках [1], давно привлекают внимание исследователей. Исследование этих задач не теряет своей актуальности и в настоящее время (см., например, [2]–[4] и библиографию в них). Однако исследований таких задач, содержащих малый параметр, очень мало, особенно, когда ищется полное асимптотическое разложение по малому параметру их решений. Одними из первых работ, где строилась такая асимптотика, были работы [5], [6], в которых на управление накладывались геометрические ограничения.

В научной школе А.М. Ильина по асимптотическому анализу исследованы некоторые задачи оптимального управления, описываемые краевыми задачами для линейных уравнений эллиптического типа с интегрально-квадратичным критерием качества (как с распределенным управлением (см. [7], [8]), так и с граничным (см. [9], [10])) и с различного рода сингулярностями (малый параметр при старших производных, малые полости в области определения уравнения, наличие угловых точек на границе). Условия оптимальности в таких задачах описываются краевыми задачами для систем двух уравнений эллиптического типа с дополнительным параметром (когда ограничения на управление по существу) и дополнительным соотношением на этот параметр. В ряде случаев такие задачи являются бисингулярными и для построения полного асимптотического разложения применяется метод согласования (см. [11]).

Данная работа является обобщением работы [12] на случай управления потоком через границу двухсвязной области с различной интенсивностью управления на каждой из частей границы.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть $\Omega := \Omega_1 \setminus \bar{\Omega}_2 \subset \mathbb{R}^n$ ($n = 2, 3$) – ограниченная двухсвязная область ($\bar{\Omega}_2 \subset \Omega_1$) с границей $\Gamma := \partial\Omega = \partial\Omega_1 \cup \partial\Omega_2 =: \Gamma_1 \cup \Gamma_2$, удовлетворяющей условию: граница Γ области Ω есть бесконечно дифференцируемое многообразие размерности $n - 1$, расположенное локально по одну сторону от Γ (иными словами, мы рассматриваем $\bar{\Omega}$ как многообразие с краем Γ класса C^∞).

Рассматривается следующая задача граничного оптимального управления (см. [1, гл. 2, соотношения (2.41), (2.9)]):

$$\begin{aligned} \mathcal{L}_\varepsilon z_\varepsilon &:= -\varepsilon^2 \Delta z_\varepsilon + a(x)z_\varepsilon = f(x), \quad x \in \Omega, \quad z_\varepsilon \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= g_1(x) + \mu_1 u_\varepsilon(x), \quad x \in \Gamma_1, \quad u_\varepsilon \in \mathcal{U}, \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= g_2(x) + \mu_2 u_\varepsilon(x), \quad x \in \Gamma_2, \end{aligned} \tag{1.1}$$

$$\mathcal{U} := \mathcal{U}_1, \quad \text{где} \quad \mathcal{U}_r := \{u(\cdot) \in L_2(\Gamma) : \|u\| \leq r\}, \tag{1.2}$$

$$J(u_\varepsilon) := \|z_\varepsilon - z_d\|^2 + v^{-1} \|u_\varepsilon(\cdot)\|^2 \rightarrow \inf, \quad u \in \mathcal{U}. \tag{1.3}$$

Здесь $v > 0$, $H^1(\Omega)$ – соболевское пространство функций, $\partial/\partial n$ – производная по внешней нормали к Γ , а через $\|\cdot\|$, $\|\cdot\|$, $\|\cdot\|_i$, $i = 1, 2$, обозначены нормы в пространствах $L_2(\Omega)$, $L_2(\Gamma)$ и $L_2(\Gamma_i)$ соответственно. Скалярные произведения в этих пространствах будем обозначать через (\cdot, \cdot) , $\langle \cdot, \cdot \rangle$ и $\langle \cdot, \cdot \rangle_i$ соответственно.

Для функций, определенных на границе области, будем использовать следующие обозначения: если $g \in L_2(\Gamma)$, то $g_i := g|_{\Gamma_i} \in L_2(\Gamma_i)$, $i = 1, 2$. И наоборот, если $g_i \in L_2(\Gamma_i)$, $i = 1, 2$, то через $g \in L_2(\Gamma)$ будем обозначать функцию такую, что $g_i := g|_{\Gamma_i} \in L_2(\Gamma_i)$, $i = 1, 2$.

Предполагается, что выполнены следующие условия:

$$\begin{aligned} a(\cdot), f(\cdot), z_d(\cdot) &\in C^\infty(\bar{\Omega}), \quad g_1(\cdot) \in C^\infty(\Gamma_1), \quad g_2(\cdot) \in C^\infty(\Gamma_2), \\ \forall x \in \bar{\Omega} \quad a(x) &\geq \alpha^2, \quad \alpha > 0 \quad \mu_1 > 0, \quad \mu_2 > 0, \quad 0 < \varepsilon \ll 1, \end{aligned} \tag{1.4}$$

а решение краевой задачи (1.1) понимается в обобщенном смысле (см., например, [1, гл. 1, § 3, п. 3.4]): для любого $\varphi \in H^1(\Omega)$ (с учетом (1.2) и (1.4)) справедливо равенство

$$\varepsilon^2 (\nabla z_\varepsilon, \nabla \varphi) + (a(\cdot)z_\varepsilon, \varphi) - \langle g_1 + \mu_1 u_\varepsilon, \varphi \rangle_1 - \langle g_2 + \mu_2 u_\varepsilon, \varphi \rangle_2 = (f, \varphi). \tag{1.5}$$

2. ОПРЕДЕЛЯЮЩИЕ СООТНОШЕНИЯ И АПРИОРНЫЕ ОЦЕНКИ

В этом случае единственное оптимальное управление $u_\varepsilon(\cdot)$ и соответствующее ему $z_\varepsilon(\cdot)$ находятся как решение следующей задачи (см. [1, гл. 2, соотношения (2.41), (2.36), (2.49)]):

$$\begin{aligned} \mathcal{L}_\varepsilon z_\varepsilon &= f(x), \quad \mathcal{L}_\varepsilon p_\varepsilon - z_\varepsilon = -z_d, \quad x \in \Omega, \quad z_\varepsilon, p_\varepsilon \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= g_1(x) + \mu_1 u_\varepsilon(x), \quad \varepsilon^2 \frac{\partial p_\varepsilon}{\partial n} = 0, \quad x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= g_2(x) + \mu_2 u_\varepsilon(x), \quad \varepsilon^2 \frac{\partial p_\varepsilon}{\partial n} = 0, \quad x \in \Gamma_2, \end{aligned} \tag{2.1}$$

$$\forall v(\cdot) \in \mathcal{U} \quad \langle p_\varepsilon + v^{-1} \mu u_\varepsilon, \mu v - \mu u_\varepsilon \rangle \geq 0. \tag{2.2}$$

Здесь

$$\mu(x) := \begin{cases} \mu_1, & x \in \Gamma_1, \\ \mu_2, & x \in \Gamma_2. \end{cases} \tag{2.3}$$

Лемма 1. Пусть $p \in L_2(\Gamma)$, $u \in \mathcal{U}_r$, $\mu \in L_2(\Gamma)$ и $\mu(x) \neq 0$ почти всюду на Γ . Тогда условие

$$\forall v(\cdot) \in \mathcal{U} \quad \langle p + v^{-1} \mu u, \mu v - \mu u \rangle \geq 0 \tag{2.4}$$

эквивалентно следующему:

$$\exists \lambda \geq 0 \quad (\|\hat{\lambda}(\cdot; \lambda)p\| \leq r) \wedge (\lambda(r - \|\hat{\lambda}(\cdot; \lambda)p\|) = 0), \quad \hat{\lambda}(x; \lambda) := \frac{v\mu(x)}{v\lambda + \mu^2(x)}. \tag{2.5}$$

При этом

$$u = -\hat{\lambda}(\cdot; \lambda) p|_{\Gamma} \quad u \quad \|u\| < r \Leftrightarrow \lambda = 0. \tag{2.6}$$

Доказательство. Поскольку

$$0 \leq \langle p + v^{-1}\mu u, \mu v - \mu u \rangle = \langle \mu p + v^{-1}\mu^2 u, v - u \rangle = \langle \mu p + v^{-1}\mu^2 u, v \rangle - \langle \mu p + v^{-1}\mu^2 u, u \rangle,$$

то, взяв минимум по $v \in \mathcal{O}u_r$ – шару радиуса r с центром в нуле, получим

$$-r \|\mu p + v^{-1}\mu^2 u\| \geq \langle \mu p + v^{-1}\mu^2 u, u \rangle,$$

что в силу неравенства Коши–Буняковского дает

$$-r \|\mu p + v^{-1}\mu^2 u\| = \langle \mu p + v^{-1}\mu^2 u, u \rangle. \tag{2.7}$$

Тем самым (2.4) эквивалентно равенству (2.7).

Если $\mu p + v^{-1}\mu^2 u = 0$, то (2.7) выполняется и $u = -vp/\mu$. В противном случае

$$u = -\frac{(\mu p + v^{-1}\mu^2 u)r}{\|\mu p + v^{-1}\mu^2 u\|}.$$

Пусть $\lambda := \|\mu p + v^{-1}\mu^2 u\|/r$. Тогда $u = -\lambda p$, откуда находим

$$u = -\frac{v\mu(x)}{v\lambda + \mu^2(x)} p,$$

т.е. (2.5) выполняется и, если $\|u\| < r$, то $\lambda = 0$.

Пусть теперь $\lambda \geq 0$ и удовлетворяет (2.5). Положим

$$u_\lambda := -\hat{\lambda}(x; \lambda) p = -\frac{v\mu(x)}{v\lambda + \mu^2(x)} p.$$

В этом случае

$$p = -\frac{v\lambda + \mu^2(x)}{v\mu(x)} p, \quad \text{а} \quad \mu p + v^{-1}\mu^2 u = -\lambda u_\lambda.$$

Таким образом, вектора $\mu p + v^{-1}\mu^2 u$ и u_λ противоположно направлены и поэтому

$$\langle \mu p + v^{-1}\mu^2 u, u_\lambda \rangle = -\lambda \|u_\lambda\|^2 - \|u_\lambda\| \cdot \|\mu p + v^{-1}\mu^2 u\|.$$

Отметим, что если $\mu p + v^{-1}\mu^2 u \neq 0$, то $\lambda \neq 0$ и в силу (2.5) $\|u_\lambda\| = r$, т.е. (2.7) выполнено.

В силу леммы 1 задача (2.1)–(2.3) (а значит, и задача (1.1)–(1.3)) эквивалентна краевой задаче

$$\begin{aligned} \mathcal{L}_\varepsilon z_\varepsilon &= f(x), \quad \mathcal{L}_\varepsilon p_\varepsilon - z_\varepsilon = -z_d, \quad x \in \Omega, \quad z_\varepsilon, p_\varepsilon \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} + \hat{\lambda}_{1,\varepsilon} p_\varepsilon(x) &= g_1(x), \quad \varepsilon^2 \frac{\partial p_\varepsilon}{\partial n} = 0, \quad x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} + \hat{\lambda}_{2,\varepsilon} p_\varepsilon(x) &= g_2(x), \quad \varepsilon^2 \frac{\partial p_\varepsilon}{\partial n} = 0, \quad x \in \Gamma_2, \\ \hat{\lambda}_{1,\varepsilon} &= \frac{v\mu_1}{v\lambda_\varepsilon + \mu_1^2}, \quad \hat{\lambda}_{2,\varepsilon} = \frac{v\mu_2}{v\lambda_\varepsilon + \mu_2^2}, \quad \lambda_\varepsilon \geq 0, \end{aligned} \tag{2.8}$$

зависящей от скалярного параметра $\lambda_\varepsilon \geq 0$ с дополнительным соотношением (2.5) при $r = 1$. Оптимальное управление u_ε определяется функцией p_ε по формуле (2.6), принимающей вид $u_\varepsilon = -\hat{\lambda}(\cdot; \lambda) p|_{\Gamma}$.

При этом, если в задаче (1.1)–(1.3) ограничение на управление не по существу, то

$$\lambda_\varepsilon = 0, \quad \hat{\lambda}_{1,\varepsilon} = \frac{\nu}{\mu_1} =: \hat{\lambda}_{1,\nu}, \quad \hat{\lambda}_{2,\varepsilon} = \frac{\nu}{\mu_2} =: \hat{\lambda}_{2,\nu}. \quad (2.9)$$

Отметим, что в любом случае

$$0 < \hat{\lambda}_{1,\varepsilon} \leq \frac{\nu}{\mu_1}, \quad 0 < \hat{\lambda}_{2,\varepsilon} \leq \frac{\nu}{\mu_2}. \quad (2.10)$$

Цель работы – изучить поведение z_ε , p_ε и λ_ε при $\varepsilon \rightarrow 0$.

В дальнейшем различные положительные константы, зависящие только от области Ω и коэффициента $a(\cdot)$, часто будем обозначать одной и той же буквой K (возможно с индексами).

В [9, лемма 2] доказано, что если выполнено условие (1.4), $f \in L_2(\Omega)$, $q \in L_2(\Gamma)$ и y_ε есть решение задачи

$$\mathcal{L}_\varepsilon y_\varepsilon = f(x), \quad x \in \Omega, \quad y_\varepsilon \in H^1(\Omega), \quad \varepsilon^2 \frac{\partial y_\varepsilon}{\partial n} = q(x), \quad x \in \Gamma, \quad (2.11)$$

то существует $K > 0$ такое, что

$$\max\{\varepsilon^{1/2} \|y\|, \varepsilon \|y\|, \varepsilon^{3/2} \|\nabla y\|\} \leq K(\|q\| + \varepsilon^{1/2} \|f\|) =: KD(f, q). \quad (2.12)$$

Поскольку z_ε – решение задачи (1.1), (1.2) есть решение задачи (2.11) с функцией q такой, что $q|_{\Gamma_i} = g_i + \mu_i u_\varepsilon$, $i = 1, 2$, то

$$\|g + u\| \leq \mu + \|g\|, \quad \mu := \mu_1 + \mu_2, \quad \|g\| := \|g_1\|_1^2 + \|g_2\|_2^2.$$

Применяя (2.12), получаем, что

$$\varepsilon^{1/2} \|z_\varepsilon\| \leq K(\|g\| + \mu + \varepsilon^{1/2} \|f\|). \quad (2.13)$$

Но p_ε удовлетворяет (2.11) с $f = -z_d$ и $q = 0$, поэтому в силу (2.12) получим, что

$$D(-z_d, 0) = \varepsilon^{1/2} \|z_\varepsilon + z_d\| \leq \varepsilon^{1/2} \|z_\varepsilon\| + \varepsilon^{1/2} \|z_d\| \stackrel{(2.13)}{\leq} K(\|g\| + \mu + \varepsilon^{1/2} \|f\|) + \varepsilon^{1/2} \|z_d\|.$$

Таким образом, если $\{z_\varepsilon, p_\varepsilon\}$ – решение задачи (2.8), (1.1)–(1.3), то существует $K > 0$ такое, что

$$\begin{aligned} \max\{\varepsilon^{1/2} \|z_\varepsilon\|, \varepsilon \|z_\varepsilon\|, \varepsilon^{3/2} \|\nabla z_\varepsilon\|\} &\leq K(\|g\| + \mu + \varepsilon^{1/2} \|f\| + \varepsilon^{1/2} \|z_d\|), \\ \max\{\varepsilon^{1/2} \|p_\varepsilon\|, \varepsilon \|p_\varepsilon\|, \varepsilon^{3/2} \|\nabla p_\varepsilon\|\} &\leq K(\|g\| + \mu + \varepsilon^{1/2} \|f\| + \varepsilon^{1/2} \|z_d\|). \end{aligned}$$

Тем самым, для z_ε – решения задачи (1.1)–(1.3), получим следующие асимптотические оценки:

$$\|z_\varepsilon\| = O(\varepsilon^{-1/2}), \quad \|p_\varepsilon\| = O(\varepsilon^{-1/2}), \quad \|z_\varepsilon\| = O(\varepsilon^{-1}), \quad \|p_\varepsilon\| = O(\varepsilon^{-1}), \quad \varepsilon \rightarrow 0.$$

В [10, теорема 1] показано, что если

$$f_j \in L_2(\Omega), \quad g_{j,i} \in H^{1/2}(\Gamma_i), \quad j, i = 1, 2, \quad (2.14)$$

то при любых $\hat{\lambda}_1 > 0$ и $\hat{\lambda}_2 > 0$ задача

$$\begin{aligned} \mathcal{L}_\varepsilon z &= f_1(x), \quad \mathcal{L}_\varepsilon p - z = f_2(x), \quad x \in \Omega, \quad z, p \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z}{\partial n} + \hat{\lambda}_1 p &= g_{1,1}(x), \quad \varepsilon^2 \frac{\partial p}{\partial n} = g_{2,1}(x), \quad x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z}{\partial n} + \hat{\lambda}_2 p &= g_{1,2}(x), \quad \varepsilon^2 \frac{\partial p}{\partial n} = g_{2,2}(x), \quad x \in \Gamma_2, \end{aligned} \quad (2.15)$$

разрешима единственным образом и функции $\{z, p\}$ – ее решение, удовлетворяют соотношению $z, p \in H^2(\Omega)$.

При этом, если $f_j \in C^\infty(\bar{\Omega})$, а $g_{j,i} \in C^\infty(\Gamma_i)$, $j, i = 1, 2$, то $z, p \in C^\infty(\bar{\Omega})$.

Лемма 2. Пусть $\{z, p\}$ — решение задачи (2.15). Тогда

$$\|z\|^2 + \hat{\lambda}_1 \|p\|_1^2 + \hat{\lambda}_2 \|p\|_2^2 = (f_1, p) - (f_2, z) + \langle g_{1,1}, p \rangle_1 - \langle g_{2,1}, z \rangle_1 + \langle g_{1,2}, p \rangle_2 - \langle g_{2,2}, z \rangle_2. \quad (2.16)$$

Доказательство. В силу определения обобщенного решения задачи (2.15) (см. (1.5)) для любых $\varphi, \psi \in H^1(\Omega)$ справедливы равенства

$$\begin{aligned} \varepsilon^2(\nabla z, \nabla \varphi) + (a(\cdot)z, \varphi) + \langle \hat{\lambda}_1 p - g_{1,1}, \varphi \rangle_1 + \langle \hat{\lambda}_2 p - g_{1,2}, \varphi \rangle_2 &= (f, \varphi), \\ \varepsilon^2(\nabla p, \nabla \psi) + (a(\cdot)p, \psi) - (z, \psi) - \langle g_{2,1}, \psi \rangle_1 - \langle g_{2,2}, \psi \rangle_2 &= (f_2, \psi). \end{aligned} \quad (2.17)$$

Положив $\varphi = p$ в (2.17), а $\psi = z$, и вычитая из первого получившегося равенства второе, получим (2.16).

Теорема 1. Пусть выполнены условия (1.4) и (2.14). Если $\{z, p\}$ — решение задачи (2.15), то существует $K > 0$ такое, что справедливы оценки

$$\begin{aligned} \max\{\varepsilon^{3/2} \|z\|, \varepsilon^2 \|z\|, \varepsilon^{5/2} \|\nabla z\|\} &\leq K(\varepsilon + \hat{\lambda}_1 + \hat{\lambda}_2)D_1(f, g), \\ \max\{\varepsilon \|p\|, \varepsilon^{3/2} \|p\|, \varepsilon^2 \|\nabla p\|\} &\leq K(\sqrt{\varepsilon} + \sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2})D_1(f, g), \end{aligned}$$

где $D_1(f, g) := \varepsilon^{1/2}(\|f\|_1 + \|f_2\|) + \|g_1\| + \|g_2\|$.

Доказательство. Сначала рассмотрим z_1 и p_1 — решение задачи

$$\begin{aligned} \mathcal{L}_\varepsilon z_1 &= f_1(x), & \mathcal{L}_\varepsilon p_1 - z_1 &= f_2(x), & x \in \Omega, & z, p \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z}{\partial n} &= g_{1,1}(x), & \varepsilon^2 \frac{\partial p}{\partial n} &= g_{2,1}(x), & x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z}{\partial n} &= g_{1,2}(x), & \varepsilon^2 \frac{\partial p}{\partial n} &= g_{2,2}(x), & x \in \Gamma_2. \end{aligned}$$

Тогда в силу (2.12)

$$\max\{\varepsilon^{1/2} \|z_1\|, \varepsilon \|z_1\|, \varepsilon^{3/2} \|\nabla z_1\|\} \leq K(\|g_1\| + \varepsilon^{1/2} \|f_1\|) \leq KD_1(f, g). \quad (2.18)$$

Так как p_1 удовлетворяет (2.11) с $f = f_2 + z_1$ и $q = g_2$, то

$$D(f_2 + z_1, g_2) \leq \|g_2\| + \varepsilon^{1/2} \|f_2\| + \varepsilon^{1/2} \|z_1\| \stackrel{(2.18)}{\leq} \|g_2\| + \varepsilon^{1/2} \|f_1\| + KD(f_1, g_1).$$

Таким образом, в силу (2.12) для функции p_1 тоже справедливы оценки вида (2.18):

$$\max\{\varepsilon^{1/2} \|p_1\|, \varepsilon \|p_1\|, \varepsilon^{3/2} \|\nabla p_1\|\} \leq K_1 D_1(f, g). \quad (2.19)$$

Теперь функции $z_2 := z - z_1$ и $p_2 := p - p_1$ удовлетворяют системе

$$\begin{aligned} \mathcal{L}_\varepsilon z_2 &= 0, & \mathcal{L}_\varepsilon p_2 - z_2 &= 0, & x \in \Omega, & z, p \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z}{\partial n} + \hat{\lambda}_1 p_2 &= -\hat{\lambda}_1 p_1, & \varepsilon^2 \frac{\partial p}{\partial n} &= 0, & x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z}{\partial n} + \hat{\lambda}_2 p_2 &= -\hat{\lambda}_2 p_1, & \varepsilon^2 \frac{\partial p}{\partial n} &= 0, & x \in \Gamma_2. \end{aligned} \quad (2.20)$$

Соотношение (2.16), примененное к z_2 и p_2 , с учетом вида системы (2.20) дает неравенство

$$\|z_2\|^2 + \hat{\lambda}_1 \|p_2\|_1^2 + \hat{\lambda}_2 \|p_2\|_2^2 \leq \hat{\lambda}_1 \|p_2\|_1 \cdot \|p_1\|_1 + \hat{\lambda}_2 \|p_2\|_2 \cdot \|p_1\|_2. \quad (2.21)$$

Отсюда следует, что

$$\hat{\lambda}_1 \|p_2\|_1^2 + \hat{\lambda}_2 \|p_2\|_2^2 \leq \hat{\lambda}_1 \|p_2\|_1 \cdot \|p_1\|_1 + \hat{\lambda}_2 \|p_2\|_2 \cdot \|p_1\|_2.$$

Последнее неравенство есть квадратичное неравенство относительно $\|p_2\|_1$ и $\|p_2\|_2$. Применяя элементарную оценку решения таких неравенств (получающуюся методом выделения полных квадратов), получаем

$$\|p_2\|_1 \leq \left(1 + \frac{\sqrt{\hat{\lambda}_2}}{2\sqrt{\hat{\lambda}_1}}\right) \|p_1\|, \quad \|p_2\|_2 \leq \left(1 + \frac{\sqrt{\hat{\lambda}_1}}{2\sqrt{\hat{\lambda}_2}}\right) \|p_1\|.$$

Отсюда получим

$$\hat{\lambda}_1 \|p_2\|_1 + \hat{\lambda}_2 \|p_2\|_2 \leq (\hat{\lambda}_1 + \sqrt{\hat{\lambda}_1 \hat{\lambda}_2} + \hat{\lambda}_2) \|p_1\| \leq (\sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2})^2 \|p_1\|. \tag{2.22}$$

Из (2.21) и (2.22) следует, что

$$\|z_2\| \leq (\sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2}) \|p_1\|. \tag{2.23}$$

Задача для p_2 есть задача (2.11) с $f = z_2$ и $q = 0$. Применив оценки (2.12) (с учетом (2.23) и (2.18)) для p_2 , получим

$$D(z_2, 0) = \varepsilon^{1/2} \|z_2\| \leq (\sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2}) \varepsilon^{-1/2} K D_1(f, g).$$

Тем самым,

$$\max\{\varepsilon \|p_2\|, \varepsilon^{3/2} \|p_2\|, \varepsilon^2 \|\nabla p_2\|\} \leq K (\sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2}) D_1(f, g).$$

Аналогично, задача для z_2 есть задача (2.11) с $f = 0$ и $q: q|_{\Gamma_i} = -\hat{\lambda}_i(p_1 + p_2)$, $i = 1, 2$. При этом в силу (2.22) и (2.19)

$$\begin{aligned} \|q\| &\leq \hat{\lambda}_1 \|p_1\|_1 + \hat{\lambda}_1 \|p_2\|_1 + \hat{\lambda}_2 \|p_1\|_1 + \hat{\lambda}_2 \|p_2\|_2 \leq 2(\hat{\lambda}_1 + \hat{\lambda}_2) + (\sqrt{\hat{\lambda}_1} + \sqrt{\hat{\lambda}_2})^2 \|p_1\| \leq \\ &\leq 4(\hat{\lambda}_1 + \hat{\lambda}_1) \|p_1\| \leq 4(\hat{\lambda}_1 + \hat{\lambda}_1) \|p_1\| \varepsilon^{-1} K_1 D_1. \end{aligned}$$

Применив оценки (2.12) для z_2 , получим

$$\max\{\varepsilon^{3/2} \|z_2\|, \varepsilon^2 \|z_2\|, \varepsilon^{5/2} \|\nabla z_2\|\} \leq K_2 (\hat{\lambda}_1 + \hat{\lambda}_1) D_1(f, g).$$

Теперь для получения итоговых оценок осталось применить неравенство треугольника для норм функций $z = z_1 + z_2$ и $p = p_1 + p_2$.

3. АППРОКСИМАЦИОННЫЕ ТЕОРЕМЫ

Для обоснования асимптотических разложений решений задачи (2.8), (2.5) при $r = 1$ нужны теоремы об оценке уклонения точного решения $\{z_\varepsilon, p_\varepsilon, \lambda_\varepsilon\}$ этой задачи от решений аппроксимационной задачи

$$\begin{aligned} \mathcal{L}_\varepsilon z_{\varepsilon,\gamma} &= f(x) + f_{1,\varepsilon,\gamma}(x), & \mathcal{L}_\varepsilon p_{\varepsilon,\gamma} - z_{\varepsilon,\gamma} &= -z_d + f_{2,\varepsilon,\gamma}(x), & x \in \Omega, \\ \varepsilon^2 \frac{\partial z_{\varepsilon,\gamma}}{\partial n} + \hat{\lambda}_{1,\varepsilon,\gamma} p_{\varepsilon,\gamma} &= g_{1,1,\varepsilon,\gamma}(x), & \varepsilon^2 \frac{\partial p_{\varepsilon,\gamma}}{\partial n} &= g_{2,1,\varepsilon,\gamma}(x), & x \in \Gamma_1, \\ \varepsilon^2 \frac{\partial z_{\varepsilon,\gamma}}{\partial n} + \hat{\lambda}_{2,\varepsilon,\gamma} p_{\varepsilon,\gamma} &= g_{1,2,\varepsilon,\gamma}(x), & \varepsilon^2 \frac{\partial p_{\varepsilon,\gamma}}{\partial n} &= g_{2,2,\varepsilon,\gamma}(x), & x \in \Gamma_2, \end{aligned} \tag{3.1}$$

в случае, когда при $\varepsilon \rightarrow 0$

$$\begin{aligned} f_{j,\varepsilon,\gamma} &\in C^\infty(\bar{\Omega}), & g_{j,i,\varepsilon,\gamma} &\in C^\infty(\Gamma_i), & \|f_{j,\varepsilon,\gamma}\| &= O(\varepsilon^\gamma), \\ \|g_{j,i,\varepsilon,\gamma}\|_j &= O(\varepsilon^\gamma), & j, i &= 1, 2, \end{aligned} \tag{3.2}$$

и аппроксимации условия (2.5) при $r = 1$.

Отметим, что если при всех достаточно малых $\varepsilon > 0$ выполнено неравенство

$$\|\hat{\lambda}(\cdot; \lambda_\varepsilon) p_\varepsilon\| < 1, \tag{3.3}$$

то в этом случае условие (2.5) при $r = 1$ переходит в условие (2.9).

При выполнении (3.3) теорема 1 дает необходимые оценки погрешности аппроксимаций.

Теорема 2. Пусть выполнены условия (1.4), (3.2) и (3.3). Если $\{z_\varepsilon, p_\varepsilon, \lambda_\varepsilon\}$ – решение задачи (2.8), (2.5) при $r = 1$, а $\{z_{\varepsilon,\gamma}, p_{\varepsilon,\gamma}\}$ – решение задачи (3.1) с $\lambda_{\varepsilon,\gamma} = 0$, т.е. $\hat{\lambda}_{i,\varepsilon,\gamma} = \hat{\lambda}_{i,\gamma}$, $i = 1, 2$, то

$$\begin{aligned} \max\{\varepsilon^{3/2} \|z_\varepsilon - z_{\varepsilon,\gamma}\|, \varepsilon^2 \|z_\varepsilon - z_{\varepsilon,\gamma,v}\|, \varepsilon^{5/2} \|\nabla(z_\varepsilon - z_{\varepsilon,\gamma,v})\|\} &= O(\varepsilon^\gamma), \\ \max\{\varepsilon \|p_\varepsilon - p_{\varepsilon,\gamma}\|, \varepsilon^{3/2} \|p_\varepsilon - p_{\varepsilon,\gamma,v}\|, \varepsilon^2 \|\nabla(p_\varepsilon - p_{\varepsilon,\gamma,v})\|\} &= O(\varepsilon^\gamma) \end{aligned}$$

при $\varepsilon \rightarrow 0$.

В случае, когда при всех достаточно малых $\varepsilon > 0$ ограничения на управление по существу, т.е.

$$\|\hat{\lambda}(\cdot; \lambda_\varepsilon) p_\varepsilon\| = 1, \tag{3.4}$$

аппроксимация условия (2.5) при $r = 1$ имеет вид

$$\|\hat{\lambda}(\cdot; \lambda_{\varepsilon,\gamma}) p_{\varepsilon,\gamma}\| = 1 + O(\varepsilon^\gamma), \tag{3.5}$$

и для получения доказательства аппроксимационной теоремы требуется вспомогательное утверждение о зависимости от r оптимального $u_{\varepsilon,r}$ в задаче (1.1)–(1.3) при условии $\|u_{\varepsilon,r}\| = r$.

Лемма 3. Пусть выполнены условия (1.4), а $u_{\varepsilon,r}$ – решение задачи (1.1)–(1.3) с $\mathcal{U} = \mathcal{U}_r$ и $\|u_{\varepsilon,r}\| = r$ при всех $r \in [r_*; r^*]$. Тогда при некоторых $K > 0$ и $\varepsilon_0 > 0$

$$\forall r, r' \in [r_*; r^*], \quad \forall \varepsilon \in (0; \varepsilon_0] \quad \|u_r - u_{r'}\| \leq K\varepsilon^{-3} |r - r'|. \tag{3.6}$$

Доказательство. Пусть $z_{\varepsilon,0}$ – решение задачи (1.1) с $u = 0$, а оператор $A : L_2(\Gamma_1) \rightarrow L_2(\Omega)$ ставит в соответствие функции u_ε решение задачи (1.1) с $f = 0$ и $g = 0$. Тогда $z_\varepsilon = z_{\varepsilon,0} + Au_\varepsilon$ и функционал качества примет вид

$$J(u_\varepsilon) = \|Au_\varepsilon + v_0\|^2 + v^{-1} \|u_\varepsilon\|,$$

где $v_0 := z_{\varepsilon,0} - z_d$. По теореме 3 из [12]

$$\|u_r - u_{r'}\| \leq K \cdot |r - r'| \cdot \|A\|^2 \cdot (\|A\| + \|v_0\|)^4.$$

По определению $\|A\|$ в силу оценок (2.13) получим $\|A\| \leq K(\varepsilon^{-1/2}(1 + \mu) + 0) \leq K_1\varepsilon^{-1/2}$. При этом $\|v_0\| \leq \|z_{\varepsilon,0}\| + \|z_d\| \stackrel{(2.12)}{\leq} K(\varepsilon^{-1/2}(\|g\| + \mu) + \|f\|) + \|z_d\|$. Тем самым

$$\|u_r - u_{r'}\| \leq K_2\varepsilon^{-3} |r - r'|$$

при всех достаточно малых $\varepsilon > 0$.

Теорема 3. Пусть выполнены условия (1.4), (3.2) и (3.4). Если $\{z_\varepsilon, p_\varepsilon, \lambda_\varepsilon\}$ – решение задачи (2.8), (2.5) при $r = 1$, а $\{z_{\varepsilon,\gamma}, p_{\varepsilon,\gamma}\}$ – решение задачи (3.1) с (3.2), (3.5), то

$$\begin{aligned} \max\{\varepsilon^{3/2} \|z_\varepsilon - z_{\varepsilon,\gamma}\|, \varepsilon^2 \|z_\varepsilon - z_{\varepsilon,\gamma}\|, \varepsilon^{5/2} \|\nabla z_\varepsilon - z_{\varepsilon,\gamma}\|\} &= O(\varepsilon^{\gamma-3}), \\ \max\{\varepsilon \|p_\varepsilon - p_{\varepsilon,\gamma}\|, \varepsilon^{3/2} \|p_\varepsilon - p_{\varepsilon,\gamma}\|, \varepsilon^2 \|\nabla p_\varepsilon - p_{\varepsilon,\gamma}\|, \|\hat{\lambda}(\cdot; \lambda_\varepsilon) - \hat{\lambda}(\cdot; \lambda_{\varepsilon,\gamma})\|_{C(\Gamma)}\} &= O(\varepsilon^{\gamma-9/2}) \end{aligned}$$

при $\varepsilon \rightarrow 0$, и $\gamma > 4$.

Доказательство. Функции $\hat{z}_{\varepsilon,\gamma} := z_{\varepsilon,\gamma} - z_\varepsilon$ и $\hat{p}_{\varepsilon,\gamma} := p_{\varepsilon,\gamma} - p_\varepsilon$ являются решением системы

$$\begin{aligned} \mathcal{L}_\varepsilon \hat{z}_{\varepsilon,\gamma} &= f_{1,\varepsilon,\gamma}(x), \quad \mathcal{L}_\varepsilon \hat{p}_{\varepsilon,\gamma} - \hat{z}_{\varepsilon,\gamma} = f_{2,\varepsilon,\gamma}(x), \quad x \in \Omega, \\ \varepsilon^2 \frac{\partial \hat{z}_{\varepsilon,\gamma}}{\partial n} &= g_{1,\varepsilon,\gamma}(x) + \hat{\lambda}(x; \lambda_\varepsilon) p_\varepsilon - \hat{\lambda}(x; \lambda_{\varepsilon,\gamma}) p_{\varepsilon,\gamma}, \quad \varepsilon^2 \frac{\partial p}{\partial n} = g_{2,\varepsilon,\gamma}(x), \quad x \in \Gamma. \end{aligned} \tag{3.7}$$

Поскольку

$$\|\hat{\lambda}(x; \lambda_\varepsilon) p_\varepsilon - \hat{\lambda}(x; \lambda_{\varepsilon,\gamma}) p_{\varepsilon,\gamma}\| \stackrel{(3.6)}{\leq} K\varepsilon^{-3} \|\|\hat{\lambda}(\cdot; \lambda_\varepsilon) p_\varepsilon\| - \|\hat{\lambda}(\cdot; \lambda_{\varepsilon,\gamma}) p_{\varepsilon,\gamma}\|\| \stackrel{(3.4),(3.5)}{=} O(\varepsilon^{\gamma-3})$$

при $\varepsilon \rightarrow 0$, то, применяя к решению системы (3.7) оценки (2.18) и (2.19), получаем все оценки доказываемой теоремы, кроме оценки величины $\|\hat{\lambda}(\cdot; \lambda_\varepsilon) - \hat{\lambda}(\cdot; \lambda_{\varepsilon, \gamma})\|_{C(\Gamma)}$.

Так как

$$1 = \overset{(3.4)}{\|\hat{\lambda}(\cdot; \lambda_\varepsilon) p_\varepsilon\|} \overset{(2.10)}{\leq} \frac{v}{\tilde{\mu}} \|p_\varepsilon\|, \quad \tilde{\mu} := \min\{\mu_1, \mu_2\} > 0,$$

то

$$\|p_\varepsilon\| \geq \tilde{\mu}/v. \tag{3.8}$$

Из (2.10), (3.8) и уже полученных оценок следует, что

$$\frac{\tilde{\mu}}{v} \|p_\varepsilon\| \leq \|\hat{\lambda}(\cdot; \lambda_\varepsilon) - \hat{\lambda}(\cdot; \lambda_{\varepsilon, \gamma})\|_{C(\Gamma)} \cdot \|p_\varepsilon\| \leq \|\hat{\lambda}(\cdot; \lambda_\varepsilon) p_\varepsilon - \hat{\lambda}(\cdot; \lambda_{\varepsilon, \gamma}) p_{\varepsilon, \gamma}\| + \|\hat{\lambda}(\cdot; \lambda_{\varepsilon, \gamma})(p_\varepsilon - p_{\varepsilon, \gamma})\| = O(\varepsilon^{\gamma-9/2})$$

при $\varepsilon \rightarrow 0$.

4. ПОСТРОЕНИЕ АСИМПТОТИКИ

В силу теорем 2 и 3 для построения асимптотического разложения рассматриваемой задачи нужно построить ее *формальное асимптотическое решение* (см., например, [11]). Его построение осуществляется аналогично тому, как это делается в случае одного уравнения (см., например, [14], [15]).

Внешнее разложение ищем в виде рядов

$$z_{\text{out}}(x, \varepsilon) := \sum_{k=0}^{\infty} \varepsilon^{2k} z_{2k}(x), \quad p_{\text{out}}(x, \varepsilon) := \sum_{k=0}^{\infty} \varepsilon^{2k} p_{2k}(x), \quad \varepsilon \rightarrow 0, \tag{4.1}$$

коэффициенты которых находятся из соответствующей рекуррентной системы

$$z_0(x) = \frac{f(x)}{a(x)}, \quad p_0(x) = \frac{z_0 - z_d}{a(x)}, \quad z_{2k}(x) = \frac{\Delta z_{2k-2}}{a(x)}, \quad p_{2k} = \frac{\Delta p_{2k-2}}{a(x)}, \quad k \geq 1. \tag{4.2}$$

В силу (1.4) все $z_{2k}, p_{2k} \in C^\infty(\bar{\Omega})$ и ряды (4.1) хорошо аппроксимируют уравнения из (2.8), но, вообще говоря, не аппроксимируют граничных условий (и дополнительного условия (2.5) в случае (3.4)).

Для аппроксимации граничных условий (и дополнительного условия (2.5)) в малых окрестностях границ Γ_i (пограничные слои) вводятся новые переменные (это можно сделать в силу гладкости границ) (s_i, τ_i) , где s_i – координата на многообразии Γ_i , а τ_i – расстояние по нормали к Γ_i , исходящей из точки на Γ_i с координатой s_i .

В пограничных слоях стандартно (см., например, [14], [11, с. 31–34]) перейдем к растянутым переменным $\xi_i := \tau_i \varepsilon^{-1}$ и к следующему виду внутреннего разложения:

$$Z_{i,\text{in}}(s_i, \tau_i \varepsilon) := \varepsilon^{m_0} \sum_{m=0}^{\infty} \varepsilon^m Z_{i,m}(s_i), \quad P_{i,\text{in}}(s_i, \tau_i \varepsilon) := \varepsilon^{m_0} \sum_{m=0}^{\infty} \varepsilon^m P_{i,m}(s_i), \quad i = 1, 2, \tag{4.3}$$

аппроксимирующему однородную систему из (2.8) и подправляющему граничные условия.

При переходе к новым координатам (s_i, ξ_i) оператор \mathcal{L}_ε перейдет в оператор

$$\tilde{\mathcal{L}}_{\varepsilon,i} Y_i = -\frac{\partial^2 Y_i}{\partial \xi_i^2} + \varepsilon L_{i,1} \frac{\partial Y_i}{\partial \xi_i} + \varepsilon^2 L_{i,2} Y_i + \tilde{a}_i(s_i, \varepsilon \xi_i) Y_i, \quad i = 1, 2.$$

Здесь $L_{i,1}$ и $L_{i,2}$ – дифференциальные операторы 1-го и 2-го порядков, содержащие лишь дифференцирование по переменной s_i , с гладкими коэффициентами от s_i и $\tau_i = \varepsilon \xi_i$, а волна над функцией, определенной в переменных x , означает выражение этой функции в переменных s_i и τ_i .

Подставляя в однородную систему, соответствующую системе из (2.8), ряды (4.3) и разлагая коэффициенты в уравнениях системы и операторов $\tilde{\mathcal{L}}_{\varepsilon,i}$ в ряды Тейлора по переменной $\tau_i = \varepsilon \xi_i$, получим следующую систему:

$$\begin{aligned} \tilde{\mathcal{L}}_{0,i} Z_{i,0} &:= -\frac{\partial^2}{\partial \xi_i^2} Z_{i,0} + \tilde{a}_{i,0}(s_i) Z_{i,0} = 0, & \tilde{\mathcal{L}}_{0,i} P_{i,0} - Z_{i,0} &= 0, \\ \tilde{\mathcal{L}}_{0,i} Z_{i,m} &= F_{i,m}(s_i, \xi_i), & \tilde{\mathcal{L}}_{0,i} P_{i,m} - Z_{i,m} &= G_{i,m}(s_i, \xi_i), \quad m \geq 1, \quad i = 1, 2, \end{aligned} \tag{4.4}$$

где $F_{i,m}(s_i, \xi_i)$ и $G_{i,m}(s_i, \xi_i)$ линейно выражаются через предыдущие $Z_{i,m}$, $P_{i,m}$ и их производные и полиномиально зависят от ξ_i и гладко от s_i , а функция

$$a(x) = \tilde{a}_i(s_i, \varepsilon \xi_i) = \sum_{m=0}^{\infty} \varepsilon^m \xi_i^m \tilde{a}_{i,m}(s_i)$$

разложена в ряд по степеням малого параметра ε в окрестности границы Γ_i , $i = 1, 2$.

Сначала рассмотрим построение асимптотики решения задачи (2.8), (2.9).

Отметим, что в данном случае задача для главных членов внутреннего разложения имеет вид

$$\begin{aligned} \tilde{\mathcal{L}}_{0,i} Z_{i,0} &= 0, & \tilde{\mathcal{L}}_{0,i} P_{i,0} - Z_{i,0} &= 0, \\ -\varepsilon^{1+m_0} \frac{\partial}{\partial \xi_i} Z_{i,0}(s_i, 0) + \varepsilon^2 \left(\frac{\partial z_0}{\partial n} \right)_i(s_i, 0) + \hat{\lambda}_{i,v}(\varepsilon^{m_0} P_{i,0}(s_i, 0) + \tilde{p}_{0i}(s_i, 0)) &= \tilde{g}_i(s_i), \\ -\varepsilon^{1+m_0} \frac{\partial}{\partial \xi_i} P_{i,0}(s, 0) + \varepsilon^2 \left(\frac{\partial p_0}{\partial n} \right)_i(s_i, 0) &= 0, \quad i = 1, 2. \end{aligned} \tag{4.5}$$

Отсюда следует, что $m_0 = 0$ и, тем самым, граничные условия в (4.5) имеют вид

$$\hat{\lambda}_{i,v} P_{i,0}(s_i, 0) = \tilde{g}_i(s_i) - \hat{\lambda}_{i,v} \tilde{p}_{0i}(s_i, 0), \quad \frac{\partial}{\partial \xi_i} P_{i,0}(s, 0) = 0. \tag{4.6}$$

В классе экспоненциально убывающих при $\xi_i \rightarrow +\infty$ функций система (4.5) имеет единственное решение

$$\begin{aligned} Z_{i,0}(s_i, \xi_i) &= \frac{2\tilde{a}_{i,0}(s_i)(\tilde{g}_i(s_i) - \hat{\lambda}_{i,v} \tilde{p}_{0i}(s_i, 0))}{\hat{\lambda}_{i,v}} e^{-\sqrt{\tilde{a}_{i,0}(s_i)} \xi_i}, \\ P_{i,0}(s, \xi) &= \frac{\tilde{g}_i(s_i) - \hat{\lambda}_{i,v} \tilde{p}_{0i}(s_i, 0)}{\hat{\lambda}_{i,v}} e^{-\sqrt{\tilde{a}_{i,0}(s)} \xi_i} + \frac{\sqrt{\tilde{a}_{i,0}(s)}(\tilde{g}_i(s_i) - \hat{\lambda}_{i,v} \tilde{p}_{0i}(s_i, 0))}{\hat{\lambda}_{i,v}} \xi_i e^{-\sqrt{\tilde{a}_{i,0}(s)} \xi_i}, \end{aligned} \tag{4.7}$$

и, тем самым,

$$\hat{\lambda}(x; 0)(P_{i,0}(s, \xi) + \tilde{p}_{0i}(s_i, 0)) = \tilde{g}_i(s_i), \quad i = 1, 2. \tag{4.8}$$

Граничные условия в рассматриваемом случае таким образом имеют вид

$$\hat{\lambda}_{i,v} P_{i,m}(s_i, 0) = \hat{F}_{i,m}(s_i), \quad \frac{\partial}{\partial \xi} P_{i,m}(s, 0) = \hat{G}_{i,m}(s_i), \tag{4.9}$$

где $\hat{F}_{i,m}(s_i)$ и $\hat{G}_{i,m}(s_i)$ однозначно определяются предыдущими коэффициентами внешнего и соответствующего внутреннего разложений.

Задачи (4.4), (4.9) в классе экспоненциально убывающих при $\xi_i \rightarrow +\infty$ функций имеют единственное решение. Каждая из функций $Z_{i,m}$ и $P_{i,m}$ с учетом (4.7) имеет вид $Q(\xi_i : s_i) e^{-\sqrt{\tilde{a}_{i,0}(s_i)} \xi_i}$, где $Q(\xi_i : s_i)$ – полином по ξ_i с коэффициентами, гладко зависящими от s_i .

Таким образом, в рассматриваемом случае все коэффициенты рядов (4.3) однозначно находятся и в силу теоремы 2 ряды

$$\begin{aligned} & \sum_{k=0}^{\infty} \varepsilon^{2k} z_{2k}(x) + \sum_{i=1}^2 \eta_i(s_i, \tau_i) \varepsilon^{m_0} \sum_{m=0}^{\infty} \varepsilon^m Z_{i,m}(s_i, \tau_i/\varepsilon), \\ & \sum_{k=0}^{\infty} \varepsilon^{2k} p_{2k}(x) + \sum_{i=1}^2 \eta_i(s_i, \tau_i) \varepsilon^{m_0} \sum_{m=0}^{\infty} \varepsilon^m P_{i,m}(s_i, \tau_i/\varepsilon) \end{aligned} \tag{4.10}$$

при $\varepsilon \rightarrow 0$ суть асимптотические разложения компонент $\{z_{\varepsilon,v}, p_{\varepsilon,v}\}$ решения задачи (2.15) с $\lambda = 0$, $f_1 = f$, $f_2 = -z_d$, $g_{1,1} = g$ и $g_{j,i} = 0$ для остальных пар j, i .

Здесь $\eta_i(s_i, \tau_i)$ – срезающие функции пограничных слоев, т.е. бесконечно дифференцируемые функции, равные 1 в некоторой окрестности границ Γ_i и равные 0 вне чуть больших окрестностей границ Γ_i . Тем самым доказана теорема.

Теорема 4. Пусть выполнены условия (1.4). Тогда ряды (4.10) с $m_0 = 0$ и коэффициентами, однозначно определенными из (4.2) для внешнего разложения, и как решения задач (4.4), (4.6) и (4.9) для внутреннего разложения, при $\varepsilon \rightarrow 0$ суть асимптотические разложения компонент $\{z_{\varepsilon,v}, p_{\varepsilon,v}\}$ – решения задачи (2.8), (2.9) при $r = 1$.

В частности, в силу (4.8)

$$\|\hat{\lambda}(\cdot, 0) p_{\varepsilon,v} 1\| \rightarrow \|g\|, \quad \varepsilon \rightarrow 0. \tag{4.11}$$

Таким образом, при выполнении неравенства

$$\|g\| < 1 \tag{4.12}$$

при всех достаточно малых $\varepsilon > 0$ выполняется неравенство $\|\hat{\lambda}(\cdot, 0) p_{\varepsilon,v}\| < 1$ и, тем самым, $z_{\varepsilon,v} = z_\varepsilon$, а $p_{\varepsilon,v} = p_\varepsilon$. Поэтому при выполнении (4.12) теорема 4 есть теорема об асимптотическом разложении решений задачи (2.8), (2.5) при $r = 1$.

Пусть теперь выполнено неравенство

$$\|g\| > 1. \tag{4.13}$$

Тогда в силу (4.11) при всех достаточно малых $\varepsilon > 0$ справедливо следующее неравенство: $\|\hat{\lambda}(\cdot, 0) p_{\varepsilon,v}\| > 1$, и поэтому реализуется случай (3.4).

Пусть $\hat{\lambda}_{i,\varepsilon} = \varepsilon^{m_\lambda} \hat{\lambda}_{i,0}$, $i = 1, 2$. Тогда задача для главных членов внутреннего разложения имеет вид

$$\begin{aligned} & \tilde{\mathcal{L}}_{0,i} Z_{i,0} = 0, \quad \tilde{\mathcal{L}}_{0,i} P_{i,0} - Z_{i,0} = 0, \\ & -\varepsilon^{1+m_0} \frac{\partial}{\partial \xi_i} Z_{i,0}(s_i, 0) + \varepsilon^2 \left(\frac{\partial z_0}{\partial n} \right)_i (s_i, 0) + \varepsilon^{m_\lambda} \hat{\lambda}_{i,0} (\varepsilon^{m_0} P_{i,0}(s_i, 0) + \tilde{p}_{0i}(s_i, 0)) = \tilde{g}_i(s_i), \\ & -\varepsilon^{1+m_0} \frac{\partial}{\partial \xi} P_{i,0}(s, 0) + \varepsilon^2 \left(\frac{\partial p_0}{\partial n} \right)_i (s_i, 0) = 0, \quad i = 1, 2. \end{aligned}$$

В предыдущем случае слагаемое $\frac{\partial}{\partial \xi_i} Z_{i,0}(s_i, 0)$ не использовалось, поэтому теперь естественно взять $m_0 = -1$, что влечет $m_\lambda = 1$, т.е. $\hat{\lambda}_{i,\varepsilon} = O(\varepsilon)$ при $\varepsilon \rightarrow +0$. Поэтому в рассматриваемом случае удобно в определении $\hat{\lambda}(x; \lambda_\varepsilon)$ взять вместо параметра λ_ε параметр $\Lambda_\varepsilon := \lambda_\varepsilon^{-1}$ и, тем самым,

$$\hat{\lambda}(x; \Lambda_\varepsilon^{-1}) = \frac{v\mu(x)\Lambda_\varepsilon}{v + \mu^2(x)\Lambda_\varepsilon}, \quad x \in \Gamma. \tag{4.14}$$

При этом асимптотическое разложение $\hat{\lambda}_{i,\varepsilon}$ будем искать в виде

$$\hat{\lambda}(i, \varepsilon) \sim \varepsilon \sum_{m=0}^{\infty} \varepsilon^m \hat{\lambda}_{i,m}, \quad \varepsilon \rightarrow 0, \quad i = 1, 2. \tag{4.15}$$

В силу (4.14) величины $\hat{\lambda}(i, \varepsilon)$ и коэффициенты $\hat{\lambda}_{i,m}$ их разложений связаны соотношениями

$$\begin{aligned} \mu_2 \hat{\lambda}(1, \varepsilon) &= \mu_1 \hat{\lambda}(2, \varepsilon) + \frac{\mu_2^2 - \mu_1^2}{\nu} \hat{\lambda}(1, \varepsilon) \hat{\lambda}(2, \varepsilon), \\ \mu_2 \hat{\lambda}_{1,0} &= \mu_1 \hat{\lambda}_{2,0}, \\ \mu_2 \hat{\lambda}_{1,m} &= \mu_1 \hat{\lambda}_{2,m} + q_m, \quad m > 0, \end{aligned} \tag{4.16}$$

где q_m однозначно определяются предыдущими $\hat{\lambda}_{i,\tilde{m}}$ ($\tilde{m} < m$).

Исходя из (3.4), естественно взять в качестве $\hat{\lambda}_{i,0}$ решение уравнения

$$1 = \|\hat{\lambda}_0 P_0(\cdot, 0)\|^2 = \|\hat{\lambda}_{1,0} P_{1,0}(\cdot, 0)\|^2 + \|\hat{\lambda}_{2,0} P_{2,0}(\cdot, 0)\|^2. \tag{4.17}$$

Решая систему

$$\begin{aligned} \tilde{\mathcal{L}}_{0,i} Z_{i,0} &= 0, \quad \tilde{\mathcal{L}}_{0,i} P_{i,0} - Z_{i,0} = 0, \\ -\frac{\partial}{\partial \xi_i} Z_{i,0}(s_i, 0) + \hat{\lambda}_{i,0} P_{i,0}(s_i, 0) &= \tilde{g}_i(s_i), \\ -\frac{\partial}{\partial \xi} P_{i,0}(s_i, 0) &= 0, \quad i = 1, 2, \end{aligned}$$

получаем

$$\begin{aligned} Z_{i,0}(s_i, \xi_i) &= \frac{2\tilde{a}_{i,0}(s_i)\tilde{g}_i(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}} e^{-\sqrt{\tilde{a}_{i,0}(s_i)}\xi_i}, \\ P_{i,0}(s_i, \xi_i) &= \frac{\tilde{g}_i(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}} e^{-\sqrt{\tilde{a}_{i,0}(s_i)}\xi_i} + \frac{\sqrt{\tilde{a}_{i,0}(s_i)}\tilde{g}_i(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}} \xi_i e^{-\sqrt{\tilde{a}_{i,0}(s_i)}\xi_i}, \end{aligned} \tag{4.18}$$

и, тем самым, уравнение (4.17) принимает вид

$$1 = \left\| \frac{\hat{\lambda}_{1,0}\tilde{g}_1(s_1)}{2\tilde{a}_{1,0}^{2/3}(s_1) + \hat{\lambda}_{1,0}} \right\|^2 + \left\| \frac{\hat{\lambda}_{2,0}\tilde{g}_2(s_2)}{2\tilde{a}_{2,0}^{2/3}(s_2) + \hat{\lambda}_{2,0}} \right\|^2 =: \mathcal{F}(\hat{\lambda}_{1,0}, \hat{\lambda}_{2,0}). \tag{4.19}$$

С учетом (4.16) рассмотрим функцию $\mathcal{F}_1(\hat{\lambda}) := \mathcal{F}(\hat{\lambda}, \mu_2\mu_1^{-1}\hat{\lambda})$. Она непрерывна, строго возрастает, $\mathcal{F}_1(0) = 0$ и $\mathcal{F}_1(+\infty) = \|\tilde{g}\| > 1$. Поэтому существует единственное $\hat{\lambda}_0 > 0$ такое, что $\mathcal{F}_1(\hat{\lambda}_0) = 1$. Положим

$$\hat{\lambda}_{1,0} = \hat{\lambda}_0, \quad \hat{\lambda}_{2,0} = \mu_2\mu_1^{-1}\hat{\lambda}_0. \tag{4.20}$$

Для таких $\hat{\lambda}_{1,0}, \hat{\lambda}_{2,0}$ равенство (4.19) выполнено.

При $m > 0$ граничные условия в этом случае имеют вид

$$\begin{aligned} -\frac{\partial}{\partial \xi_i} Z_{i,m}(s_i, 0) + \hat{\lambda}_{i,0} P_{i,m}(s_i, 0) + \hat{\lambda}_{i,m} P_{i,0}(s_i, 0) &= \hat{F}_{i,m}(s_i), \\ -\frac{\partial}{\partial \xi} P_{i,m}(s_i, 0) &= \hat{G}_{i,m}(s_i), \quad i = 1, 2. \end{aligned} \tag{4.21}$$

Здесь $\hat{F}_{i,m}$ и $\hat{G}_{i,m}$ однозначно определяются коэффициентами разложений (4.3), (4.15) с меньшими индексами и являюся гладкими функциями на Γ_i .

При известных $\hat{\lambda}_{i,m}$ задачи (4.4), (4.21) в классе экспоненциально убывающих при $\xi_i \rightarrow +\infty$ функций имеют единственное решение. Каждая из функций $Z_{i,m}$ и $P_{i,m}$ в силу (4.18) имеет вид $Q(\xi_i : s_i)e^{-\sqrt{\hat{\alpha}_{i,0}(s_i)}\xi_i}$, где $Q(\xi_i : s_i)$ — полином по ξ_i с коэффициентами, гладко зависящими от s_i .

Для нахождения $\hat{\lambda}_{i,m}$ при $m > 0$ используется асимптотическое равенство, соответствующее (3.4):

$$1 = \sum_{i=1}^2 \lambda_{i,\varepsilon}^2 \|p_\varepsilon\|_i^2 \sim \sum_{i=1}^2 \left\| \left(\sum_{m=0}^{\infty} \varepsilon^m \lambda_{i,m} \right) \cdot \left(\sum_{m=0}^{\infty} \varepsilon^m \hat{P}_{i,m} \right) \right\|_i^2, \tag{4.22}$$

где $\hat{P}_{i,m} = P_{i,m} + \overline{P_{i,m-1}}$.

Уравнение для $\lambda_{i,m}$ при $m > 0$ из (4.22) с учетом (4.20) — выбора $\lambda_{i,0}$, имеет вид

$$\sum_{i=1}^2 (\lambda_{i,0}^2 \langle P_{i,0}, P_{i,m} \rangle_i + \lambda_{i,0} \lambda_{i,m} \|P_{i,0}\|_i^2) = h_m, \tag{4.23}$$

где константы h_m однозначно определяются предыдущими членами разложений.

Функции $Z_{i,m}$ и $P_{i,m}$ удобно представить в виде

$$Z_{i,m} = Z_{i,m,1} + \lambda_m \tilde{Z}_i, \quad P_{i,m} = P_{i,m,1} + \lambda_m \tilde{P}_i, \tag{4.24}$$

где $\{Z_{i,m,1}, P_{i,m,1}\}$ — решение задачи

$$\begin{aligned} \tilde{\mathcal{L}}_{0,i} Z_{i,m,1} &= F_{i,m}(s_i, \xi_i), & \tilde{\mathcal{L}}_{0,i} P_{i,m,1} - Z_{i,m,1} &= G_m(s_i, \xi_i), \\ -\frac{\partial}{\partial \xi_i} Z_{i,m,1}(s_i, 0) + \lambda_{i,0} P_{i,m,1}(s_i, 0) &= \hat{F}_{i,m}(s_i), & \frac{\partial}{\partial \xi_i} P_{i,m,1}(s_i, 0) &= \hat{G}_{i,m}(s_i), \end{aligned} \tag{4.25}$$

а $\{\tilde{Z}_i, \tilde{P}_i\}$ — решение задачи

$$\begin{aligned} \tilde{\mathcal{L}}_0 \tilde{Z}_i &= 0, & \tilde{\mathcal{L}}_0 \tilde{P}_i - \tilde{Z}_i &= 0, \\ -\frac{\partial}{\partial \xi_i} \tilde{Z}_i(s_i, 0) + \lambda_{i,0} \tilde{P}_i(s_i, 0) + P_{i,0}(s_i, 0) &= 0, & \frac{\partial}{\partial \xi_i} \tilde{P}_i(s_i, 0) &= 0. \end{aligned} \tag{4.26}$$

Отметим, что $Z_{i,m,1}$ и $P_{i,m,1}$ однозначно определяются предыдущими членами разложений, а $\tilde{P}_i(s_i, 0)$ находится по формуле (4.18):

$$\tilde{P}_i(s_i, 0) = -\frac{P_{i,0}(s_i, 0)}{2\hat{\alpha}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}}. \tag{4.27}$$

С учетом (4.24) уравнение (4.23) принимает вид

$$\sum_{i=1}^2 (\lambda_{i,0}^2 \lambda_{i,m} \langle P_{i,0}, P_{i,m} \rangle_i + \lambda_{i,0} \lambda_{i,m} \|P_{i,0}\|_i^2) = h_{m,1}, \tag{4.28}$$

где константы $h_{m,1}$ однозначно определяются предыдущими членами разложений.

Выражая $\lambda_{2,m}$ в силу (4.16) через $\lambda_{1,m}$ и подставляя это выражение в (4.28), приходим к уравнению для нахождения $\lambda_{1,m}$, а следовательно, и $\lambda_{2,m}$:

$$\lambda_{1,0}^2 \lambda_{1,m} \langle P_{1,0}, \tilde{P}_1 \rangle_1 + \lambda_{1,0} \lambda_{1,m} \|P_{1,0}\|_1^2 + \mu_2^3 \mu_1^{-3} \lambda_{1,0}^2 \lambda_{1,m} \langle P_{2,0}, \tilde{P}_2 \rangle_2 + \mu_2^2 \mu_1^{-2} \lambda_{1,0} \lambda_{1,m} \|P_{2,0}\|_2^2 = h_{m,2}, \tag{4.29}$$

где константы $h_{m,2}$ однозначно определяются предыдущими членами разложений.

Лемма 4. Пусть выполнены условия (1.4). Тогда

$$C := \lambda_{1,0}^2 \langle P_{1,0}, \tilde{P}_1 \rangle_1 + \lambda_{1,0} \|P_{1,0}\|_1^2 + \mu_2^3 \mu_1^{-3} \lambda_{1,0}^2 \langle P_{2,0}, \tilde{P}_2 \rangle_2 + \mu_2^2 \mu_1^{-2} \lambda_{1,0} \|P_{2,0}\|_2^2 \neq 0.$$

Доказательство. В силу (4.16) и (4.27) имеем

$$C = \sum_{i=1}^2 \left(- \left\langle \frac{\hat{\lambda}_{i,0}}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}}, \hat{\lambda}_{i,0} P_{i,0}^2 \right\rangle_i + \left\langle 1, \hat{\lambda}_{i,0} P_{i,0}^2 \right\rangle_i \right) = \sum_{i=1}^2 \left\langle \frac{2\tilde{a}_{i,0}^{2/3}(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}}, \hat{\lambda}_{i,0} P_{i,0}^2 \right\rangle_i.$$

Поскольку $\frac{2\tilde{a}_{i,0}^{2/3}(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}}$ и $P_{i,0}^2$ непрерывны, $\frac{2\tilde{a}_{i,0}^{2/3}(s_i)}{2\tilde{a}_{i,0}^{2/3}(s_i) + \hat{\lambda}_{i,0}}$ положительна, а $\hat{\lambda}_{i,0} P_{i,0}^2$ неотрицательна на Γ_i , то равенство $C = 0$ равносильно соотношению $P_{i,0} \equiv 0$, что в силу (4.17) невозможно.

В силу леммы 4 из (4.29) находятся однозначно $\hat{\lambda}_{i,m}$ и все коэффициенты рядов (4.3) и (4.15). Таким образом, справедлива

Теорема 5. Пусть выполнены условия (1.4) и (4.13). Тогда ряды (4.10) с $m_0 = -1$ и коэффициентами, однозначно определенными из (4.2) для внешнего разложения, и как решения задач (4.4), (4.21), (4.23)–(4.26) и (4.29) для внутреннего разложения, при $\varepsilon \rightarrow 0$ суть асимптотические разложения компонент $\{z_\varepsilon, p_\varepsilon\}$ – решения задачи (2.8) при $r = 1$, а ряды из (4.15) есть асимптотические разложения величин $\hat{\lambda}_{i,\varepsilon}$, $i = 1, 2$.

5. ПРИМЕР

Проиллюстрируем описанные конструкции на следующем примере:

$$\begin{aligned} \mathcal{L}_\varepsilon z_\varepsilon &:= -\varepsilon^2 \Delta z_\varepsilon + z_\varepsilon = x_1, & x \in \Omega, & z_\varepsilon \in H^1(\Omega), \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= q + 2u_\varepsilon(x), & |x| = 1, & u_\varepsilon \in \mathcal{U}, \\ \varepsilon^2 \frac{\partial z_\varepsilon}{\partial n} &= u_\varepsilon(x), & |x| = 2, & \end{aligned}$$

где

$$\begin{aligned} \Omega &= \{x \in \mathbb{R}^2 : 1 < |x| < 2\}, & |x| &:= \sqrt{x_1^2 + x_2^2}, \\ J(u_\varepsilon) &:= \|z_\varepsilon\|^2 + \|u_\varepsilon(\cdot)\|^2 \rightarrow \inf, & u &\in \mathcal{U}, \end{aligned}$$

т.е. $a = 1$, $f(x_1, x_2) = x_1$, $\mu_1 = 2$, $\mu_2 = 1$, $g_1 = q - \text{const}$, $g_2 = 0$ и $v = 1$.

В этом случае внешнее разложение имеет только одно слагаемое:

$$z_0(x_1, x_2) = x_1, \quad p_0(x_1, x_2) = x_1, \quad z_{2k} = 0, \quad p_{2k} = 0, \quad k > 0,$$

и, тем самым, вне малой окрестности границы $z_\varepsilon = x_1 + O(\varepsilon^{+\infty})$ при $\varepsilon \rightarrow +0$.

Координаты s_i — это полярный угол по модулю 2π , $\tau_1 = \rho - 1$, $\tau_2 = 2 - \rho$, где ρ — полярный радиус. Поэтому в силу вида оператора Лапласа в полярных координатах получим

$$\begin{aligned} \tilde{\mathcal{L}}_{\varepsilon,1} Y_1 &= -\frac{\partial^2 Y_1}{\partial \xi_1^2} + \frac{\varepsilon}{1 + \varepsilon \xi_1} \frac{\partial Y_1}{\partial \xi_1} + \frac{\varepsilon^2}{(1 + \varepsilon \xi_1)^2} \frac{\partial^2 Y_1}{\partial \varphi_1^2} + Y_1, \\ \tilde{\mathcal{L}}_{\varepsilon,1} Y_2 &= -\frac{\partial^2 Y_2}{\partial \xi_2^2} - \frac{\varepsilon}{2 - \varepsilon \xi_2} \frac{\partial Y_2}{\partial \xi_2} + \frac{\varepsilon^2}{(2 - \varepsilon \xi_2)^2} \frac{\partial^2 Y_2}{\partial \varphi_1^2} + Y_2. \end{aligned}$$

Раскладывая $(1 + \varepsilon \xi_1)^{-1}$, $(1 + \varepsilon \xi_1)^{-2}$, $(2 - \varepsilon \xi_1)^{-1}$ и $(2 - \varepsilon \xi_1)^{-2}$ в степенные ряды по ε , получаем степенные разложения операторов $\tilde{\mathcal{L}}_{\varepsilon,i}$, в частности,

$$\begin{aligned} \tilde{\mathcal{L}}_{\varepsilon,1} Y_1 &= -\frac{\partial^2 Y_1}{\partial \xi_1^2} + Y_1 + \varepsilon \frac{\partial Y_1}{\partial \xi_1} - \varepsilon^2 \xi_1 \frac{\partial Y_1}{\partial \xi_1} + \varepsilon^2 \frac{\partial^2 Y_1}{\partial \varphi_1^2} + \dots, \\ \tilde{\mathcal{L}}_{\varepsilon,1} Y_2 &= -\frac{\partial^2 Y_2}{\partial \xi_2^2} + Y_2 - \frac{\varepsilon}{2} \frac{\partial Y_2}{\partial \xi_2} - \frac{\varepsilon^2 \xi_2}{4} \frac{\partial Y_2}{\partial \xi_2} + \frac{\varepsilon^2}{4} \frac{\partial^2 Y_2}{\partial \varphi_1^2} + \dots \end{aligned}$$

Поскольку z_0 и p_0 в полярных координатах имеют вид $\rho \cos \varphi$, то

$$\left(\frac{\partial \widetilde{z_0}}{\partial n}\right)_i(\varphi, 0) = (-1)^i \cos \varphi, \quad \left(\frac{\partial \widetilde{p_0}}{\partial n}\right)_i(\varphi, 0) = (-1)^i \cos \varphi,$$

$$\widetilde{p}_{0_1}(\varphi, 0) = \cos \varphi, \quad \widetilde{p}_{0_2}(\varphi, 0) = 2 \cos \varphi.$$

Отметим, что $\|g\| = \sqrt{2\pi q}$ и, если $\sqrt{2\pi q} < 1$, то ограничения на управления не по существу, $\hat{\lambda}_{1,v} = 1/2$, $\hat{\lambda}_{2,v} = 1$ и граничные условия (4.6) имеют вид

$$P_{1,0}(\varphi, 0) = 2q - \cos \varphi, \quad P_{2,0}(\varphi, 0) = -\cos \varphi, \quad \frac{\partial}{\partial \xi_i} P_{i,0}(\varphi, 0) = 0$$

и, тем самым,

$$Z_{1,0}(\varphi, \xi_1) = 2(2q - \cos \varphi)e^{-\xi_1}, \quad P_{1,0}(\varphi, \xi_1) = (2q - \cos \varphi)e^{-\xi_1} + (2q - \cos \varphi)\xi_1 e^{-\xi_1},$$

$$Z_{2,0}(\varphi, \xi_2) = -2 \cos \varphi e^{-\xi_2}, \quad P_{2,0}(\varphi, \xi_2) = -2 \cos \varphi e^{-\xi_2} - 2 \cos \varphi \xi_2 e^{-\xi_2}.$$

Таким образом, в полярных координатах z_ε (в малой окрестности границы Γ) и u_ε имеют при $\varepsilon \rightarrow +0$ следующее асимптотическое представление:

$$z_\varepsilon = \rho \cos \varphi + 2(2q - \cos \varphi)e^{-(\rho-1)/\varepsilon} - 2 \cos \varphi e^{-(2-\rho)/\varepsilon} + O(\varepsilon),$$

$$u_{1,\varepsilon} = -q + \frac{1}{2} \cos \varphi + O(\varepsilon), \quad u_{2,\varepsilon} = \cos \varphi + O(\varepsilon).$$

Заметим, что в рассматриваемой области срезающие функции при внутреннем разложении можно опустить.

Если $\sqrt{2\pi q} > 1$, то ограничения на управление по существу и уравнение (4.19) примут вид

$$1 = \left\| \frac{\hat{\lambda}_{1,0} \cdot q}{2 + \hat{\lambda}_{1,0}} \right\|^2 = 2\pi \left(\frac{\hat{\lambda}_{1,0} \cdot q}{2 + \hat{\lambda}_{1,0}} \right)^2.$$

Поэтому $\hat{\lambda}_{1,0} = 2/(\sqrt{2\pi q} - 1)$, а в силу (4.16) имеем $\hat{\lambda}_{2,0} = 1/(\sqrt{2\pi q} - 1)$. Таким образом, в силу (4.18)

$$Z_{1,0}(\varphi, \xi_1) = \frac{2}{\sqrt{2\pi}} e^{-\xi_1}, \quad P_{1,0}(\varphi, \xi_1) = \frac{1}{\sqrt{2\pi}} e^{-\xi_1} + \frac{2}{\sqrt{2\pi}} \xi_1 e^{-\xi_1}, \quad Z_{2,0}(\varphi, \xi_2) = 0, \quad P_{2,0}(\varphi, \xi_2) = 0$$

и в полярных координатах z_ε (в малой окрестности границы Γ) и u_ε имеют при $\varepsilon \rightarrow +0$ следующее асимптотическое представление:

$$z_\varepsilon = \varepsilon^{-1} \frac{2}{\sqrt{2\pi}} e^{-(\rho-1)/\varepsilon} + O(1), \quad u_{1,\varepsilon} = -\frac{1}{\sqrt{2\pi}} + O(\varepsilon), \quad u_{2,\varepsilon} = O(\varepsilon).$$

СПИСОК ЛИТЕРАТУРЫ

1. Лионс Ж.-Л. Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972. 414 с.
2. Casas E. A review on sparse solutions in optimal control of partial differential equations // SeMA J. 2017. V. 74. P. 319–344.
3. Lou H., Yong J. Second-order necessary conditions for optimal control of semilinear elliptic equations with leading term containing controls // Math. Control Relat. Fields. 2018. V. 8. № 1. P. 57–88. <https://doi.org/10.3934/mcrf.2018003>
4. Betz L.M. Second-order sufficient optimality conditions for optimal control of nonsmooth, semilinear parabolic equations // SIAM J. Control Optim. 2019. V. 57. № 6. P. 4033–4062.
5. Капустян В.Е. Асимптотика ограниченных управлений в оптимальных эллиптических задачах // Докл. АН Украины. Сер. Матем., естествозн., техн. науки. 1992. № 2. С. 70–74.
6. Капустян В.Е. Оптимальные бисингулярные эллиптические задачи с ограниченным управлением // ДАН Украины. 1993. № 6. С. 81–85.
7. Данилин А.Р. Асимптотика ограниченных управлений для сингулярной эллиптической задачи в области с малой полостью // Матем. сб. 1998. Т. 189. № 11. С. 27–60.

8. Данилин А.Р. Аппроксимация сингулярно возмущенной эллиптической задачи оптимального управления // Матем. сб. 2000. Т. 191. № 10. С. 3–12.
9. Данилин А.Р., Зорин А.П. Асимптотика решения задачи оптимального граничного управления // Тр. ИММ УрО РАН. 2009. Т. 15. № 4. С. 95–107.
10. Данилин А.Р. Оптимальное граничное управление в области с малой полостью // Уфимский матем. ж. 2012. Т. 4. № 2. С. 87–100.
11. Ильин А.М. Согласование асимптотических разложений решений краевых задач. М.: Наука, 1989. 336 с.
12. Данилин А.Р. Асимптотика решения задачи оптимального граничного управления потоком через часть границы // Тр. ИММ УрО РАН. 2014. Т. 20. № 4. С. 116–127.
13. Люстерник Л.А., Соболев В.И. Элементы функционального анализа М.: Наука, 1965. 520 с.
14. Вишик М.И., Люстерник Л.А. Регулярное вырождение и пограничный слой для линейных дифференциальных уравнений с малым параметром // Успехи матем. наук. 1957. Т. 12. Вып. 5. С. 3–122.
15. Ильин А.М. Пограничный слой // Современ. проблемы матем. Фундамент. направления. Т. 34. М.: ВНИТИ, 1988. С. 175–214. (Итоги науки и техники ВНИТИ АН СССР.)

**ОПТИМАЛЬНОЕ
УПРАВЛЕНИЕ**

УДК 519.658

ПРЯМО-ДВОЙСТВЕННЫЙ МЕТОД НЬЮТОНА С НАИСКОРЕЙШИМ СПУСКОМ ДЛЯ ЛИНЕЙНОЙ ЗАДАЧИ ПОЛУОПРЕДЕЛЕННОГО ПРОГРАММИРОВАНИЯ. НЬЮТОНОВСКАЯ СИСТЕМА УРАВНЕНИЙ¹⁾

© 2022 г. В. Г. Жадан^{1,2}

¹ 117333 Москва, ул. Вавилова, 40, ВЦ ФИЦ ИУ РАН, Россия

² 141700 Долгопрудный, М. о., Институтский пер., 9, МФТИ, Россия

e-mail: zhadan@ccas.ru

Поступила в редакцию 02.04.2021 г.

Переработанный вариант 02.04.2021 г.

Принята к публикации 12.10.2021 г.

Рассматривается линейная задача полуопределенного программирования. Для ее решения предлагается допустимый прямо-двойственный метод, основанный на решении методом Ньютона системы уравнений, описывающих условия оптимальности в задаче. Обсуждается вопрос, как строить ньютоновские направления перемещения в случае принадлежности текущих точек итерационного процесса границам допустимых множеств. При этом существенным образом используется разбиение пространства симметричных матриц на подпространства. Библ. 14. Фиг. 1.

Ключевые слова: линейная задача полуопределенного программирования, условия оптимальности, прямо-двойственный метод Ньютона, наискорейший спуск.

DOI: 10.31857/S0044466922020132

1. ВВЕДЕНИЕ

Линейная задача полуопределенного программирования в стандартной постановке заключается в минимизации на конусе симметричных положительно полуопределенных матриц линейной целевой функции при линейных ограничениях типа равенства. Численным методам решения таких задач уделяется большое внимание [1]–[7], причем наиболее популярными из них являются методы внутренней точки, главным образом, прямо-двойственные методы центрального пути [8]. При построении таких методов существенным образом используется метод Ньютона, который применяется для решения системы уравнений, описывающих условия оптимальности в задаче.

В настоящей работе также строится прямо-двойственный метод, однако в отличие от упомянутых методов в нем используется наискорейший спуск при выборе шага перемещения и допускается выход на границы допустимых множеств. Это приводит к определенной специфике в вычислении ньютоновских направлений. Чтобы найти эти направления, предлагается в пространстве симметричных матриц выделить подпространство симметричных блочно-диагональных матриц, состоящее из четырех блоков, и определять направления перемещений отдельно в каждом из этих блоков. Размеры блоков зависят от рангов прямой переменной и слабой двойственной переменной, а также от симметризованного произведения этих матриц. Данный прием применялся ранее для решения задач линейного программирования [9], [10].

При разработке предложенного метода используются условия оптимальности, причем условие комплементарности берется в виде равенства нулевой матрице симметризованного произведения прямой переменной со слабой двойственной переменной. Главная идея метода заключалась в том, чтобы постепенно от итерации к итерации добиваться выполнения условия комплементарности. Делается это за счет уменьшения или по крайней мере не увеличения ранга симметризованного произведения. Это приводит к важному отличию предложенного метода от допустимого прямо-двойственного метода внутренней точки, например, [5]. Данное отличие состоит в том, что в ходе вычислений при приближении к решению задач полуопределенного программирования уменьшаются размерности вспомогательных задач, которые требуется решать для определения ньютоновских направлений.

¹⁾ Работа выполнена при финансовой поддержке РФФ (проект 21-71-30005).

Работа состоит из четырех разделов. В разд. 1 дается постановка задачи и приводятся вспомогательные сведения из теории полуопределенного программирования. В разд. 2 рассматривается симметризованное произведение прямой и слабой двойственной переменных и исследуется вопрос об образе этого произведения и его нуль-пространстве. Выясняется, каким образом образ и нуль-пространство симметризованного произведения связаны с образами и нуль-пространствами прямой и слабой двойственной переменных. Здесь даются также некоторые дополнительные определения, в частности, вводится понятие регулярной согласованной пары прямой и слабой двойственной переменных. В таких парах возможно разбиение пространства симметричных матриц на четыре ортогональных между собой подпространства. В разд. 3 показывается, каким образом при ряде дополнительных предположений о прямых и слабых двойственных переменных можно представить эти переменные в виде симметричных блочно-диагональных матриц. Четвертый раздел посвящен построению ньютоновской системы для определения ньютоновских направлений. Подход к решению этой системы дается в отдельной статье, где приводятся точные формулы, позволяющие вычислить данные направления.

Ниже через 0_k обозначается нулевой k -мерный вектор, через 0_{kl} – нулевая матрица размером $k \times l$, через I_k – единичная матрица порядка k . Знак точки с запятой, разделяющий векторы или компоненты векторов, показывает, что эти векторы или компоненты помещаются одни под другими. Остальные обозначения вводятся по ходу изложения.

1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Пусть \mathbb{S}^n обозначает пространство симметричных матриц порядка n . Пусть, кроме того, \mathbb{S}_+^n и \mathbb{S}_{++}^n – подмножества из \mathbb{S}^n , состоящие соответственно из положительно полуопределенных и положительно определенных матриц. Множество \mathbb{S}_+^n является конусом в \mathbb{S}^n , множество \mathbb{S}_{++}^n является его внутренностью. Для указания на то, что матрица $M \in \mathbb{S}^n$ положительно полуопределена (положительно определена), будем пользоваться также неравенством $M \succeq 0$ ($M \succ 0$). Конус \mathbb{S}_+^n не является полиэдральным, его размерность совпадает с размерностью пространства \mathbb{S}^n и равняется так называемому “треугольному” числу $n^\Delta = \frac{1}{2}n(n+1)$.

Скалярное (внутреннее) произведение между двумя квадратными матрицами L и M порядка n определяется как след матрицы $L^T M$ и обозначается в виде

$$L \bullet M = \text{tr}(L^T M) = \sum_{i,j=1}^n l_{ij} m_{ij},$$

где l_{ij} и m_{ij} суть (ij) -е элементы соответственно матриц L и M . Если L и M – две положительно полуопределенные матрицы из \mathbb{S}^n , то $L \bullet M \geq 0$ и $L \bullet M = 0$ в том и только в том случае, когда $LM = ML = 0_{nn}$. Более того, согласно теореме Фейера (о следе), матрица $L \in \mathbb{S}^n$ положительно полуопределена тогда и только тогда, когда $L \bullet M \geq 0$ для всех $M \succeq 0$, т.е. конус \mathbb{S}_+^n является самосопряженным.

Рассмотрим задачу полуопределенного программирования:

$$\begin{aligned} \min C \bullet X, \\ A_i \bullet X = b^i, \quad i \in J^m, \\ X \succeq 0. \end{aligned} \tag{1}$$

Здесь и ниже J^k – множество индексов от 1 до k , матрицы C , X и A_i , $i \in J^m$, принадлежат пространству \mathbb{S}^n .

Двойственной к (1) является задача

$$\begin{aligned} & \max \langle b, u \rangle, \\ & \sum_{i=1}^m u^i A_i + Y = C, \\ & Y \succeq 0, \end{aligned} \quad (2)$$

в которой $b = (b^1; \dots; b^m) \in \mathbb{R}^m$, $Y \in \mathbb{S}^n$. Для слабой двойственной переменной Y в дальнейшем используется также обозначение $Y(u) = C - \sum_{i=1}^m u^i A_i$.

Предполагается, что задачи (1) и (2) имеют решения и что матрицы A_i , $i \in J^m$, линейно независимы. Считаем также, что C – ненулевая матрица, и b – ненулевой вектор.

Обозначим допустимые множества в прямой и двойственной задачах соответственно через \mathcal{F}_P и \mathcal{F}_D , т.е.

$$\begin{aligned} \mathcal{F}_P &= \{X \in \mathbb{S}_+^n : A_i \bullet X = b^i, i \in J^m\}, \\ \mathcal{F}_D &= \{[u, Y] \in \mathbb{R}^m \times \mathbb{S}_+^n : Y = Y(u)\}. \end{aligned}$$

Через $\mathcal{F}_{D,u}$ и $\mathcal{F}_{D,Y}$ обозначим также проекции множества \mathcal{F}_D на пространство \mathbb{R}^m и конус \mathbb{S}_+^n соответственно:

$$\begin{aligned} \mathcal{F}_{D,u} &= \{u \in \mathbb{R}^m : [u, Y] \in \mathcal{F}_D \text{ для некоторого } Y \in \mathbb{S}_+^n\}, \\ \mathcal{F}_{D,Y} &= \{Y \in \mathbb{S}_+^n : [u, Y] \in \mathcal{F}_D \text{ для некоторого } u \in \mathbb{R}^m\}. \end{aligned}$$

Для любых $X \in \mathcal{F}_P$ и $u \in \mathcal{F}_{D,u}$ выполняется соотношение *слабой двойственности*: $C \bullet X \geq \langle b, u \rangle$.

Введем ряд определений, которые потребуются в дальнейшем.

Определение 1. Пара $[X, Y] \in \mathbb{S}^n \times \mathbb{S}^n$ называется *внутренней*, если $X \succeq 0$, $Y \succeq 0$, и *строго внутренней*, если $X \succ 0$, $Y \succ 0$.

Определение 2. Пара $[X, Y] \in \mathbb{S}^n \times \mathbb{S}^n$ называется *допустимой*, если $X \in \mathcal{F}_P$ и $Y \in \mathcal{F}_{D,Y}$.

Определение 3. Пара $[X, Y] \in \mathbb{S}^n \times \mathbb{S}^n$ называется *комплементарной*, если $X \bullet Y = 0$.

Всякая допустимая пара $[X, Y]$ есть внутренняя пара. Допустимую пару $[X, Y]$ назовем *оптимальной*, если X и Y (вместе с некоторым $u \in \mathbb{R}^m$) являются решениями соответственно задач (1) и (2). Пара $[X, Y]$ оптимальная, если она допустимая и комплементарная.

Возьмем произвольную внутреннюю пару $[X, Y]$. Так как матрицы X и Y симметричные, то для них имеют место разложения:

$$X = QD(\eta)Q^T, \quad Y = HD(\theta)H^T, \quad (3)$$

где Q и H – ортогональные матрицы, $D(\eta)$ и $D(\theta)$ – диагональные матрицы с векторами собственных значений $\eta = [\eta^1; \dots; \eta^n]$ и $\theta = [\theta^1; \dots; \theta^n]$ на диагоналях.

Пусть ранг матрицы X равен $0 < r_X < n$. Тогда матрицу Q и вектор η можно разбить на две части. Не умаляя общности, считаем, что

$$Q = [Q_B, Q_N], \quad \eta = [\eta_B; \eta_N], \quad \eta_B > 0_{r_X}, \quad \eta_N = 0_{d_X}, \quad d_X = n - r_X, \quad (4)$$

причем матрица Q_B имеет размер $n \times r_X$, матрица Q_N – размер $n \times d_X$. Аналогично, если матрица Y имеет ранг $0 < r_Y < n$, то H и θ можно представить соответственно в виде:

$$H = [H_B, H_N], \quad \theta = [\theta_B; \theta_N], \quad \theta_B = 0_{d_Y}, \quad \theta_N > 0_{r_Y}, \quad d_Y = n - r_Y, \quad (5)$$

где H_B есть $(n \times d_Y)$ -матрица, H_N есть $(n \times r_Y)$ -матрица.

Ниже для произвольной матрицы M через $\mathcal{R}(M)$ обозначается ее образ, через $\mathcal{N}(M)$ – ее ядро (нуль-пространство). Тогда $\mathcal{R}(X) = \mathcal{R}(Q_B)$, $\mathcal{N}(X) = \mathcal{R}(Q_N)$. Для образа и ядра Y имеем соответственно: $\mathcal{R}(Y) = \mathcal{R}(H_N)$, $\mathcal{N}(Y) = \mathcal{R}(H_B)$.

Проведем дальнейшее разбиение матриц Q и H на подматрицы. Данное разбиение зависит от пересечений подпространств $\mathcal{R}(Q_B)$, $\mathcal{R}(Q_N)$, $\mathcal{R}(H_B)$ и $\mathcal{R}(H_N)$ друг с другом, при этом некоторые из них могут оказаться пустыми.

В матрице Q_B выделяем две подматрицы. Подматрица Q_{BB} состоит из тех столбцов q_i из Q_B , которые одновременно принадлежат $\mathcal{R}(H_B)$. Оставшиеся столбцы матрицы Q_B помещаем в матрицу Q_{BN} . Точно так же поступаем с матрицей H_N . В ней выделяем подматрицу H_{NN} , содержащую те и только те столбцы h_j из H_N , которые одновременно принадлежат подпространству $\mathcal{R}(Q_N)$. Столбцы h_j , не обладающие этим свойством, составляют подматрицу H_{NB} матрицы H_N .

Обратимся теперь к подматрицам Q_N и H_B и также каждую из них разобьем на две подматрицы (опять отметим, что некоторые из подматриц могут быть пустыми). К подматрице Q_{NN} отнесем те столбцы q_i матрицы Q_N , которые принадлежат подпространству $\mathcal{R}(H_N)$, а остальные столбцы поместим в матрицу Q_{NB} . Аналогично, если столбец h_j матрицы H_B принадлежит подпространству $\mathcal{R}(Q_B)$, то его помещаем в подматрицу H_{BB} , иначе – в подматрицу H_{BN} . Заметим, что при таком разбиении матриц Q_N и H_B не оказывается столбцов матриц H_N и Q_B , которые принадлежат соответственно подпространствам $\mathcal{R}(Q_{NB})$ и $\mathcal{R}(H_{BN})$. В самом деле, если, например, вектор h_j , являющийся столбцом матрицы H_N , оказался в $\mathcal{R}(Q_{NB})$, то он принадлежал бы некоторому подпространству пространства $\mathcal{R}(Q_{NB})$. В этом случае можно так повернуть векторы из Q_{NB} , образующие базис в указанном подпространстве, чтобы они, оставаясь в нем, были ортогональными друг к другу, и чтобы один из векторов совпадал с h_j . По построению, мы должны такой вектор h_j поместить в качестве столбца в матрицу Q_{NN} .

Отметим, что столбцы матриц Q_N и H_B можно выбирать достаточно произвольно, важно только, чтобы они образовывали базис в соответственно подпространствах $\mathcal{N}(X)$ и $\mathcal{N}(Y)$. Если какие-то столбцы q_i матрицы Q_B принадлежат подпространству $\mathcal{R}(H_B)$ (столбцы матрицы Q_{BB}), то, поскольку все они ортогональны между собой, их можно включить в матрицу H_B , доопределив остальные столбцы H_B так, чтобы в совокупности все они образовывали ортогональный базис в $\mathcal{N}(Y)$. Соответствующим образом поступим и с матрицей Q_N , т.е. считаем, что столбцы матрицы H_{NN} являются одновременно и столбцами матрицы Q_{NN} . Таким образом, получаем фактически: $\mathcal{R}(Q_{BB}) = \mathcal{R}(H_{BB})$ и $\mathcal{R}(Q_{NN}) = \mathcal{R}(H_{NN})$, что с учетом вышесказанного означает:

$$Q_{BB} = H_{BB}, \quad Q_{NN} = H_{NN}. \tag{6}$$

В дальнейшем, не ограничивая общности, считаем для определенности, что

$$Q_B = [Q_{BB} Q_{BN}], \quad Q_N = [Q_{NB} Q_{NN}]. \tag{7}$$

Точно так же поступим и с матрицами H_B и H_N , разбив их на подматрицы:

$$H_B = [H_{BB} H_{BN}], \quad H_N = [H_{NB} H_{NN}]. \tag{8}$$

Разбиениям матриц Q_B и H_N соответствуют разбиения векторов собственных значений η_B и θ_N , а именно: $\eta_B = [\eta_{BB}; \eta_{BN}]$, $\theta_N = [\theta_{NB}; \theta_{NN}]$.

По своему построению все четыре подпространства $\mathcal{R}(Q_{BB})$, $\mathcal{R}(Q_{BN})$, $\mathcal{R}(Q_{NB})$ и $\mathcal{R}(Q_{NN})$ ортогональны друг другу. Аналогично, подпространства $\mathcal{R}(H_{BB})$, $\mathcal{R}(H_{BN})$, $\mathcal{R}(H_{NB})$ и $\mathcal{R}(H_{NN})$ также ортогональны друг другу.

Утверждение 1. Пусть допустимая пара $[X, Y]$ такова, что

$$\dim \mathcal{R}(Q_{BB}) + \dim \mathcal{R}(H_{NN}) = n. \tag{9}$$

Тогда пара $[X, Y]$ оптимальная.

Доказательство. В силу (9) ортогональные друг другу подпространства $\mathcal{R}(Q_{BB})$ и $\mathcal{R}(H_{NN})$ в совокупности составляют все пространство \mathbb{R}^n . Пара $[X, Y]$ является допустимой и комплементарной. Такая пара, как уже отмечалось, является оптимальной.

При выполнении равенства (9) подпространства $\mathcal{R}(Q_{BN})$, $\mathcal{R}(Q_{NB})$ и $\mathcal{R}(H_{BN})$, $\mathcal{R}(H_{NB})$ оказываются пустыми. Таким образом, если будет найдена допустимая пара $[X, Y]$, в которой ранг матрицы $[Q_{BB}, Q_{NN}]$ или, что одно и то же, ранг матрицы $[H_{BB}, H_{NN}]$ равняется n , то данная пара будет оптимальной.

2. СИММЕТРИЗОВАННОЕ ПРОИЗВЕДЕНИЕ МАТРИЦ X И Y

Ниже помимо матриц $X \in \mathbb{S}^n$ и $Y \in \mathbb{S}^n$ нам потребуется их симметризованное произведение $X \circ Y$, определяемое как

$$X \circ Y = (XY + YX)/2. \quad (10)$$

Данное произведение обладает следующим свойством.

Утверждение 2. *Внутренняя пара матриц $[X, Y]$ удовлетворяет равенству*

$$X \circ Y = 0_{nn} \quad (11)$$

в том и только том случае, когда $XY = YX = 0_{nn}$.

Таким образом, если выполнено равенство (11), то матрицы X и Y коммутируют между собой. Однако в общем случае, когда для внутренней пары $[X, Y]$ равенство (11) нарушено, и матрицы X и Y не коммутируют между собой, матрица $X \circ Y$ не обязательно является знакоопределенной.

Определение 4. Внутренняя пара $[X, Y]$ называется *согласованной*, если

$$X \circ Y \succeq 0.$$

Строго внутренняя неоптимальная пара $[X, Y]$ такая, что $XY = \mu I_n$, где $\mu > 0$, будет согласованной, причем $X \circ Y \succ 0$. Оптимальная пара $[X, Y]$ также является согласованной.

Так как $X \circ Y$ – симметричная матрица, то для нее справедливо разложение

$$X \circ Y = UD(\lambda)U^T, \quad (12)$$

где U – ортогональная матрица, λ – вектор собственных значений $X \circ Y$. Пусть ранг матрицы $X \circ Y$ равняется $r_{X \circ Y}$ и пусть $U = [U_P, U_O]$, где U_P – подматрица матрицы U , состоящая из ее первых $r_{X \circ Y}$ столбцов. Считаем, что столбцы U_P являются собственными векторами матрицы $X \circ Y$, соответствующие ее ненулевым собственным значениям. Считаем также, что эти ненулевые собственные значения собраны в вектор λ_P размерности $r_{X \circ Y}$, который является первым подвектором вектора $\lambda = [\lambda_P; \lambda_O]$. Второй подвектор λ_O , имеющий размерность $d_{X \circ Y} = n - r_{X \circ Y}$, нулевой. Столбцы матрицы U_O принадлежат нуль-пространству матрицы $X \circ Y$, размерность которого равна $d_{X \circ Y}$.

Рассмотрим подпространства $\mathcal{R}(X \circ Y)$ и $\mathcal{N}(X \circ Y)$. Матрица $X \circ Y$ симметричная, поэтому подпространство $\mathcal{N}(X \circ Y)$ ортогонально $\mathcal{R}(X \circ Y)$. Выясним, каким образом данные подпространства $\mathcal{R}(X \circ Y)$ и $\mathcal{N}(X \circ Y)$ связаны с подпространствами, определяемыми подматрицами матриц Q и H из (6)–(8). Имеет место следующее утверждение.

Лемма 1. *Матрицы XY и YX могут быть представлены соответственно в виде*

$$XY = Q_{BN} D(\eta_{BN}) Q_{BN}^T H_{NB} D(\theta_{NB}) H_{NB}^T, \quad (13)$$

$$YX = H_{NB} D(\theta_{NB}) H_{NB}^T Q_{BN} D(\eta_{BN}) Q_{BN}^T, \quad (14)$$

если, разумеется, Q_{BN} и H_{NB} – непустые матрицы.

Доказательство. Рассмотрим сначала матрицу XY , имеющую согласно разложениям (3) и разбиениям (4), (5) представление

$$XY = Q_B D(\eta_B) Q_B^T H_N D(\theta_N) H_N^T. \quad (15)$$

Отсюда следует, что образ матрицы XY принадлежит образу матрицы Q_B . Кроме того, если найдется столбец q_i матрицы Q_B такой, что $H_N^T q_i = 0$, то появляется нулевая строка в матрице $D(\eta_B)Q_B^T H_N$ и, значит, столбец q_i не может принадлежать образу матрицы XY , его можно удалить. Но равенство $H_N^T q_i = 0$ указывает на то, что столбец q_i принадлежит $\mathcal{N}(Y) = \mathcal{R}(H_B)$. Такой столбец, как уже говорилось, мы помещаем в матрицу Q_{BB} . Таким образом, вместо (15) можем записать следующее:

$$XY = Q_{BN} D(\eta_{BN}) Q_{BN}^T H_N D(\theta_N) H_N^T.$$

Далее, если обратиться к матрице H_N , то для ее столбцов h_j , входящих в подматрицу H_{NN} , выполняется $Q_B^T h_j = 0$. Тем более выполняется равенство $Q_{BN}^T h_j = 0$. Поэтому в матрице H_N можно оставить только столбцы, входящие в подматрицу H_{NB} , а остальные столбцы (входящие в подматрицу H_{NN}) могут быть удалены. В результате приходим к представлению (13) матрицы XY .

Справедливость представления (14) матрицы YX обосновывается аналогичным образом. Лемма доказана.

Ниже для сокращения записи через $\mathcal{R}^I(\cdot)$ обозначается пересечение образов нескольких матриц, которые перечислены в скобках. Через $\mathcal{R}^S(\cdot)$ обозначается их сумма (по Минковскому), например,

$$\mathcal{R}^I(X, Y, Z) = \mathcal{R}(X) \cap \mathcal{R}(Y) \cap \mathcal{R}(Z), \quad \mathcal{R}^S(X, Y, Z) = \mathcal{R}(X) + \mathcal{R}(Y) + \mathcal{R}(Z).$$

Оценим ранг матрицы $X \circ Y$. Имеем $\mathcal{R}(X \circ Y) \subseteq \mathcal{R}^S(XY, YX)$. Но $YX = (XY)^T$, поэтому

$$\mathcal{R}(X \circ Y) \subseteq \mathcal{R}(XY) + \mathcal{N}^\perp(XY),$$

где $\mathcal{N}^\perp(XY)$ – ортогональное дополнение к $\mathcal{N}(XY)$, или, что то же самое,

$$\mathcal{R}(X \circ Y) \subseteq \mathcal{R}(YX) + \mathcal{N}^\perp(YX).$$

Более точное включение для $\mathcal{R}(X \circ Y)$ можно получить с помощью леммы 1.

Утверждение 3. Для образа $\mathcal{R}(X \circ Y)$ матрицы $X \circ Y$ имеет место включение

$$\mathcal{R}(X \circ Y) \subseteq \mathcal{R}^S(Q_{BN}, H_{NB}), \tag{16}$$

а для его размерности, равной рангу $r_{X \circ Y}$ матрицы $X \circ Y$, – неравенство

$$r_{X \circ Y} \leq 2 \min\{r_{Q_{BN}}, r_{H_{NB}}\}, \tag{17}$$

где $r_{Q_{BN}}$ и $r_{H_{NB}}$ – ранги матриц Q_{BN} и H_{NB} соответственно.

Доказательство. Как следует из определения (10) симметризованного произведения, образ матрицы $X \circ Y$ принадлежит сумме образов матриц XY и YX соответственно. Но согласно утверждению леммы 1 имеем $\mathcal{R}(XY) \subseteq \mathcal{R}(Q_{BN})$, $\mathcal{R}(YX) \subseteq \mathcal{R}(H_{NB})$. Отсюда приходим к включению (16).

Кроме того, для ранга матрицы XY получаем оценку сверху

$$r_{XY} \leq \min\{r_{Q_{BN}}, r_{H_{NB}}\}. \tag{18}$$

Матрица YX является транспонированной к матрице XY . Поэтому ранг YX совпадает с рангом матрицы XY и для него также справедлива оценка (18). Неравенство (17) следует из (18) и неравенства для ранга суммы двух матриц. Утверждение доказано.

Следствие 1. Для матрицы U_p имеет место разложение

$$U_p = [Q_{BN}, H_{NB}] W_p,$$

где W_p – матрица полного ранга размером $(r_{Q_{BN}} + r_{H_{NB}}) \times r_{X \circ Y}$.

В случае, когда подпространства $\mathcal{R}(Q_{NB})$ или $\mathcal{R}(H_{BN})$ непустые, можно уточнить оценку сверху (17) на ранг матрицы $X \circ Y$.

Утверждение 4. Пусть

$$\dim \mathcal{R}^I(Q_N, H_B) > 0. \quad (19)$$

Тогда обязательно $\dim \mathcal{R}(Q_{NB}) > 0$, $\dim \mathcal{R}(H_{BN}) > 0$ и

$$\mathcal{R}^I(Q_N, H_B) = \mathcal{R}^I(Q_{NB}, H_{BN}). \quad (20)$$

Доказательство. Согласно неравенству (19) оба подпространства $\mathcal{R}(Q_N)$ и $\mathcal{R}(H_B)$ непустые и отличны от нулевых. Возьмем произвольный ненулевой вектор $z \in \mathbb{R}^n$ такой, что $z \in \mathcal{R}^I(Q_N, H_B)$. Покажем, что в этом случае обязательно $z \in \mathcal{R}(Q_{NB})$. Отсюда, в частности, следует ненулевая размерность пространства $\mathcal{R}(Q_{NB})$.

Так как $z \in \mathcal{R}(Q_N)$ и $\mathcal{R}(Q_N) = \mathcal{R}^S(Q_{NB}, Q_{NN})$, то для z справедливо разложение $z = z_1 + z_2$, где $z_1 \in \mathcal{R}(Q_{NB})$, $z_2 \in \mathcal{R}(Q_{NN})$. Оба подпространства $\mathcal{R}(Q_{NB})$ и $\mathcal{R}(Q_{NN})$ ортогональны друг другу. Считаем для общности, что второе подпространство $\mathcal{R}(Q_{NN})$ непустое и $z_2 \neq 0_n$.

Но одновременно $z \in \mathcal{R}(H_B) = \mathcal{R}^S(H_{BB}, H_{BN})$, причем оба подпространства $\mathcal{R}(H_{BB})$ и $\mathcal{R}(H_{BN})$ также ортогональны друг другу. Поэтому для z справедливо может быть разложение только в виде суммы двух компонент, одна из которых принадлежит подпространству $\mathcal{R}(H_{BB})$, а другая – подпространству $\mathcal{R}(H_{BN})$. Согласно (6) имеем $Q_{NN} = H_{NN}$, что влечет $\mathcal{R}(Q_{NN}) = \mathcal{R}(H_{NN})$. Таким образом, предположение, что $z_2 \in \mathcal{R}(Q_{NN})$, приводит к тому, что $z_2 \in \mathcal{R}(H_{NN})$, т.е. z_2 принадлежит ортогональному дополнению к подпространству $\mathcal{R}(H_B)$, что невозможно. Поэтому обязательно $z \in \mathcal{R}(Q_{NB})$.

Схожим образом убеждаемся, что $\dim \mathcal{R}(H_{BN}) > 0$ и $z \in \mathcal{R}(H_{BN})$. Следовательно, справедливо равенство (20) Утверждение доказано.

Пусть непустое подпространство $\mathcal{R}^I(Q_{NB}, H_{BN})$ ненулевое. Обозначим через Q_{NB}^Z матрицу, для которой подпространство $\mathcal{R}(Q_{NB}^Z)$ совпадает с $\mathcal{R}^I(Q_{NB}, H_{BN})$. Аналогично, пусть H_{BN}^Z – матрица, для которой подпространство $\mathcal{R}(H_{BN}^Z)$ совпадает с $\mathcal{R}^I(Q_{NB}, H_{BN})$. Так как обе матрицы Q_{NB}^Z , H_{BN}^Z состоят из собственных векторов матриц X и Y , которые одновременно принадлежат их нуль-пространствам, то в качестве Q_{NB}^Z и H_{BN}^Z можно взять одинаковые матрицы, определив нужным образом соответствующие собственные векторы из нуль-пространств. Другими словами, можно считать, что $Q_{NB}^Z H_{BN}^Z$. Оставшиеся подматрицы матриц Q_{NB} и H_{BN} обозначим соответственно через Q_{NB}^P и H_{BN}^P . Опять отмечаем, что некоторые из них могут оказаться пустыми. Если подпространство $\mathcal{R}^I(Q_{NB}, H_{BN})$ нулевое, то в качестве Q_{NB}^P и H_{BN}^P берутся сами матрицы Q_{NB} и H_{BN} . Таким образом, выполняются следующие разложения:

$$\mathcal{R}(Q_{NB}) = \mathcal{R}^S(Q_{NB}^P, Q_{NB}^Z), \quad \mathcal{R}(H_{BN}) = \mathcal{R}^S(H_{BN}^P, H_{BN}^Z). \quad (21)$$

Поскольку $\mathcal{R}(Q_N) = \mathcal{R}^S(Q_{NN}, Q_{NB})$, то из (21) следует, что если подпространство $\mathcal{N}(Q)$ ненулевое, то в общем случае его можно представить как сумму трех подпространств, а именно,

$$\mathcal{N}(Q) = \mathcal{R}(Q_N) = \mathcal{R}^S(Q_{NN}, Q_{NB}^P, Q_{NB}^Z),$$

причем все три составляющие подпространства $\mathcal{R}(Q_{NN})$, $\mathcal{R}(Q_{NB}^P)$ и $\mathcal{R}(Q_{NB}^Z)$ ортогональны между собой. То же самое можно сказать и про подпространство $\mathcal{N}(H)$, для которого справедлива формула

$$\mathcal{N}(H) = \mathcal{R}(H_B) = \mathcal{R}^S(H_{BB}, H_{BN}^P, H_{BN}^Z).$$

Имеем в силу вышесказанного $\mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z)$. Кроме того, согласно своим определениям $\mathcal{R}(Q_{BB}) = \mathcal{R}(H_{BB})$ и $\mathcal{R}(Q_{NN}) = \mathcal{R}(H_{NN})$. Поэтому

$$\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z) = \mathcal{R}^S(H_{NN}, H_{BB}, H_{BN}^Z). \quad (22)$$

Q_B		Q_N		
Q_{BB}	Q_{BN}	Q_{NB}		Q_{NN}
Q_{BB}	Q_{BN}	Q_{NB}^Z	Q_{NB}^P	Q_{NN}
H_{BB}	H_{BN}^P	H_{BN}^Z	H_{NB}	H_{NN}
H_{BB}	H_{BN}		H_{NB}	H_{NN}
H_B			H_N	

Фиг. 1. Разбиение матриц Q и H на подматрицы.

Отсюда, в частности, следует, что

$$\mathcal{R}^S(Q_{BN}, Q_{NB}^P) = \mathcal{R}^S(H_{NB}, H_{BN}^P). \tag{23}$$

Это равенство вытекает из того, что подпространства, стоящие в левой и правой частях (23), являются дополнениями подпространств (22) до всего пространства \mathbb{R}^n .

Схематично разбиение матриц Q и H на подматрицы (6)–(8), а также разбиение матриц Q_{NB} и H_{BN} на соответственно подматрицы Q_{NB}^P, Q_{NB}^Z и H_{BN}^P, H_{BN}^Z представлено на фиг. 1. Там же опять схематично указано на равенство между образами матриц Q_{BB} и Q_{NN} с образами соответственно матриц H_{BB} и H_{NN} . Обращено также внимание на совпадение образов матриц Q_{NB}^Z и H_{BN}^Z .

Утверждение 5. Пусть $\dim \mathcal{R}^I(H_{NB}, Q_{NB}) > 0$. Тогда

$$\mathcal{R}^I(H_{NB}, Q_{NB}) = \mathcal{R}^I(H_{NB}, Q_{NB}^P), \tag{24}$$

Аналогично,

$$\mathcal{R}^I(Q_{BN}, H_{BN}) = \mathcal{R}^I(Q_{BN}, H_{BN}^P), \tag{25}$$

если $\dim \mathcal{R}^I(Q_{BN}, H_{BN}) > 0$.

Доказательство. Убедимся только в равенстве (24). При этом считаем, не умаляя общности, что подпространство $\mathcal{R}(Q_{NB}^Z)$ непустое, иначе, так как $Q_{NB} = Q_{NB}^P + Q_{NB}^Z$, равенство (24) очевидно.

Возьмем произвольный ненулевой вектор $y \in \mathcal{R}^I(H_{NB}, Q_{NB})$ и покажем, что обязательно $y \in \mathcal{R}(Q_{NB}^P)$. Допустим, что $y = y_1 + y_2$, где $y_1 \in \mathcal{R}(Q_{NB}^P)$, $y_2 \in \mathcal{R}(Q_{NB}^Z)$ и $y_2 \neq 0$. Имеем $\mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z)$. Отсюда с учетом того, что $\mathcal{R}(H_{BN}) \perp \mathcal{R}(H_{NB})$ получаем $y_2 \notin \mathcal{R}(H_{NB})$. Так как $y \in \mathcal{R}(H_{NB})$, то $y_2 = 0$. Мы пришли к противоречию. Следовательно, $y = y_1 \in \mathcal{R}(Q_{NB}^P)$. Утверждение доказано.

Лемма 2. Пусть матрицы X и Y таковы, что

$$k_{XY} = \dim \mathcal{R}^I(H_{NB}, Q_{NB}^P) > 0. \tag{26}$$

Тогда $r_{XY} = r_{H_{NB}} - k_{XY}$.

Доказательство. Воспользуемся формулой для ранга произведения XY :

$$r_{XY} = r_Y - \dim(\mathcal{R}(Y) \cap \mathcal{N}(X)),$$

справедливой, вообще говоря, для произведения произвольных матриц X и Y . Если применить ее к матрице XY вида (13), то приходим к следующему значению для ее ранга:

$$r_{XY} = r_{H_{NB}} - \dim(\mathcal{R}(H_{NB}) \cap \mathcal{N}(Q_{BN}^T)).$$

Учтем далее, что ядро $\mathcal{N}(Q_{BN}^T)$ матрицы Q_{BN}^T есть ортогональное дополнение к подпространству $\mathcal{R}(Q_{BN})$. Тогда $\mathcal{N}(Q_{BN}^T) = \mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB})$ или, если воспользоваться (21),

$$\mathcal{N}(Q_{BN}^T) = \mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^P, Q_{NB}^Z).$$

Но, как уже отмечалось, выполняются следующие равенства для подпространств:

$$\mathcal{R}(Q_{BB}) = \mathcal{R}(H_{BB}), \quad \mathcal{R}(Q_{NN}) = \mathcal{R}(H_{NN}), \quad \mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z).$$

Более того, все эти подпространства $\mathcal{R}(H_{BB})$, $\mathcal{R}(H_{NN})$ и $\mathcal{R}(H_{BN}^Z)$ ортогональны $\mathcal{R}(H_{NB})$. Поэтому остается только пересечение $\mathcal{R}(H_{NB}) \cap \mathcal{N}(Q_{BN}^T) = \mathcal{R}^I(H_{NB}, Q_{NB}^P)$. Отсюда приходим к требуемому утверждению.

Схожим образом доказывается

Лемма 3. Пусть матрицы X и Y таковы, что

$$k_{YX} = \dim \mathcal{R}^I(Q_{BN}, H_{BN}^P) > 0. \quad (27)$$

Тогда $r_{YX} = r_{Q_{BN}} - k_{YX}$.

Замечание 1. Любое подпространство $\mathcal{R}^S(h_1, \dots, h_k)$, где h_i , $1 \leq j \leq k$, – столбцы матрицы H_{NB} , не может принадлежать $\mathcal{R}(Q_{NB})$. Действительно, иначе эти столбцы были бы включены в матрицу H_{NN} . Аналогично, любое подпространство $\mathcal{R}^S(q_1, \dots, q_l)$, где q_i , $1 \leq j \leq l$, – столбцы матрицы Q_{BN} , не может принадлежать $\mathcal{R}(H_{NB})$.

Обратимся теперь к нуль-пространству матрицы $X \circ Y$. Оно совпадает с нуль-пространством матрицы U_P , а его размерность $d_{X \circ Y}$ равняется рангу матрицы U_O . Подпространство $\mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z)$, если непустое, в силу своего определения обязательно принадлежит $\mathcal{R}(U_O)$.

Утверждение 6. Пусть подпространства $\mathcal{R}(Q_{BB})$, $\mathcal{R}(Q_{NN})$ и $\mathcal{R}(Q_{NB}^Z)$ непустые. Тогда

$$\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z) \subseteq \mathcal{R}(U_O). \quad (28)$$

Доказательство. Прежде всего отметим, что подпространства $\mathcal{R}(Q_{BB})$, $\mathcal{R}(Q_{NN})$ и $\mathcal{R}(Q_{NB}^Z)$ согласно своим определениям ортогональны подпространству $\mathcal{R}(Q_{BN})$ и, следовательно,

$$\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z) \subseteq \mathcal{N}(Q_{BN}^T). \quad (29)$$

Кроме того, имеет место совпадение подпространств

$$\mathcal{R}(Q_{BB}) = \mathcal{R}(H_{BB}), \quad \mathcal{R}(Q_{NN}) = \mathcal{R}(H_{NN}), \quad \mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z).$$

Поскольку опять, согласно своим определениям, все три подпространства $\mathcal{R}(H_{BB})$, $\mathcal{R}(H_{NN})$ и $\mathcal{R}(H_{BN}^Z)$ ортогональны подпространству $\mathcal{R}(H_{NB})$, то верно следующее:

$$\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z) \subseteq \mathcal{N}(H_{NB}^T). \quad (30)$$

Из (29) и (30) с учетом представлений (13), (14) приходим к выводу, что

$$\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z) \subseteq \mathcal{N}(X \circ Y),$$

что приводит к включению (28). Утверждение доказано.

Следствие 2. Пусть выполнены предположения утверждения 6. Тогда подпространства $\mathcal{R}(H_{BB})$, $\mathcal{R}(H_{NN})$, $\mathcal{R}(H_{BN}^Z)$ непустые и имеет место включение

$$\mathcal{R}^S(H_{BB}, H_{NN}, H_{BN}^Z) \subseteq \mathcal{R}(U_O). \quad (31)$$

Доказательство. Включение (31) равносильно включению (28) из-за совпадения подпространства $\mathcal{R}^S(H_{BB}, H_{NN}, H_{BN}^Z)$ с подпространством $\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z)$.

Из утверждения 6 и следствия к нему вытекает, что

$$\mathcal{R}(U_P) \subseteq \mathcal{R}^S(Q_{BN}, Q_{NB}^P), \quad \mathcal{R}(U_P) \subseteq \mathcal{R}^S(H_{NB}, H_{BN}^P),$$

причем согласно (23) правые части в обоих включениях совпадают между собой.

Составим матрицы

$$Q_P = [Q_{BN}, Q_{NB}^P], \quad H_P = [H_{NB}, H_{BN}^P],$$

в которых все столбцы нормированы и ортогональны друг к другу. Подпространство $\mathcal{R}(Q_P)$ является ортогональным дополнением к подпространству $\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z)$, которое согласно утверждению 6 принадлежит подпространству $\mathcal{R}(U_O)$. В свою очередь, подпространство $\mathcal{R}(H_P)$ является ортогональным дополнением к подпространству $\mathcal{R}^S(H_{NN}, H_{BB}, H_{BN}^Z)$. Последнее подпространство, как было показано, совпадает с подпространством $\mathcal{R}^S(Q_{BB}, Q_{NN}, Q_{NB}^Z)$, что позволяет говорить о равенстве $\mathcal{R}(Q_P) = \mathcal{R}(H_P)$ и включениях $\mathcal{R}(U_P) \subseteq \mathcal{R}(Q_P)$, $\mathcal{R}(U_P) \subseteq \mathcal{R}(H_P)$.

Определение 5. Внутренняя пара $[X, Y]$ называется *регулярной*, если имеет место равенство $\mathcal{R}(Q_P) = \mathcal{R}(H_P) = \mathcal{R}(U_P)$.

В регулярной паре $[X, Y]$ подпространство $\mathcal{N}(X \circ Y)$ есть пространство столбцов матрицы U_O . Учтем также, что столбцы матриц Q_{BB} и H_{NN} образуют часть базиса пространства $\mathcal{R}(U_O)$. Тогда для матрицы U_O без ограничения общности справедливо разложение

$$U_O = [U_B U_N U_Z], \tag{32}$$

где

$$U_B = Q_{BB} = H_{BB}, \quad U_N = Q_{NN} = H_{NN}, \tag{33}$$

и U_Z — произвольная матрица, задающая ортонормированный базис в подпространстве $\mathcal{R}(Q_{NB}^Z) = \mathcal{R}(H_{BN}^Z)$, т.е. совпадающая с $Q_{NB}^Z = H_{BN}^Z$, если, разумеется, эта матрица непустая. Кроме того, в случае, когда матрица U_P непустая, всю матрицу U опять, не ограничивая общность, можно записать в виде

$$U = [U_P, U_B, U_N, U_Z]. \tag{34}$$

Ранги матриц U_P , U_B , U_N и U_Z в дальнейшем будем обозначать соответственно через n_P , n_B , n_N и n_Z . Если какая-то из матриц оказывается пустой, то считаем, что она нулевого ранга.

Утверждение 7. Все непустые подпространства $\mathcal{R}(U_B)$, $\mathcal{R}(U_N)$ и $\mathcal{R}(U_Z)$ ортогональны друг другу, а также ортогональны подпространству $\mathcal{R}(U_P)$, если оно непустое.

Доказательство. Ортогональность данных подпространств следует из их определения.

Утверждение 8. Пусть пара $[X, Y]$ регулярная. Тогда для матрицы U_P справедливы следующие разложения:

$$U_P = Q_P F_{X_P}^T, \quad U_P = H_P F_{Y_P}^T, \tag{35}$$

где F_{X_P} и F_{Y_P} — ортогональные матрицы порядка $n_P \times n_{X \circ Y}$.

Доказательство. Справедливость разложений (35) следует из определения регулярной пары. Кроме того, поскольку $U_P^T U_P I_{n_P}$ и

$$(Q_P)^T Q_P = I_{n_P}, \quad (H_P)^T H_P = I_{n_P},$$

то для матриц F_{X_P} и F_{Y_P} получаем $F_{X_P} F_{X_P}^T = I_{n_P}$, $F_{Y_P} F_{Y_P}^T = I_{n_P}$, т.е. обе эти матрицы ортогональные. Утверждение доказано.

Замечание 2. Требование внутренности пары $[X, Y]$ по существу в приведенных доказательствах нигде не использовалось.

Введем еще одно определение.

Определение 6. Внутренняя пара $[X, Y]$ называется *неособой*, если $n_Z = 0$. В противном случае, когда $n_Z > 0$, она называется *особой*.

В неособой паре $[X, Y]$ матрица U_Z в представлении (34) матрицы U пропадает, матрица Q_P переходит в $Q_P = [Q_{BN}, Q_{NB}]$, а H_P – в матрицу $H_P = [H_{NB}, H_{BN}]$. В дальнейшем, пока не оговорено противное, считаем, что рассматриваемые пары неособые.

3. ДОПУСТИМЫЕ РЕГУЛЯРНЫЕ СОГЛАСОВАННЫЕ ПАРЫ

Если внутренняя пара помимо того, что она регулярная, является еще согласованной, то можно уточнить вид пространства столбцов матрицы U_P , а именно, $\mathcal{R}(U_P)$ определяется только пространствами столбцов матрицы Q_{BN} или пространством столбцов матрицы H_{NB} .

Теорема 1. Пусть допустимая пара $[X, Y]$ регулярная и согласованная. Пусть, кроме того, $X \bullet Y > 0$. Тогда матрицы Q_{NB} и H_{BN} пустые и верно

$$\mathcal{R}(U_P) = \mathcal{R}(Q_{BN}) = \mathcal{R}(H_{NB}). \quad (36)$$

Доказательство. Прежде всего отметим, что поскольку пара $[X, Y]$ не комлементарная, то матрица U_P непустая, и, следовательно, имеет положительный ранг $n_P > 0$. Согласно определению регулярной неособой пары $\mathcal{R}(U_P) = \mathcal{R}(Q_P)$, где $Q_P = [Q_{BN}, Q_{NB}]$. Покажем, что на самом деле $\mathcal{R}(U_P) = \mathcal{R}(Q_{BN})$.

Допустим противное, что матрица Q_{NB} непустая. Положим $\eta_P = [\eta_{BN}; \eta_{NB}]$, где η_{NB} – нулевой вектор, размерность которого совпадает с числом столбцов в матрице Q_{NB} . Так как в этом случае $Q_{BN}D(\eta_{BN})Q_{BN}^T = Q_P D(\eta_P)Q_P^T$, то в силу леммы 1 и разложения (12) имеем

$$Q_P D(\eta_P)Q_P^T H_{NB} D(\theta_{NB})H_{NB}^T + H_{NB} D(\theta_{NB})H_{NB}^T Q_P D(\eta_P)Q_P^T = 2U_P D(\lambda_P)U_P^T,$$

где $\lambda_P > 0_{n_P}$. Воспользуемся утверждением 8 и умножим данное равенство слева и справа на матрицы $F_{X_P}^T U_P^T$ и $U_P F_{X_P}$ соответственно. С учетом равенств (35) получаем

$$D(\eta_P)Q_P^T H_{NB} D(\theta_{NB})H_{NB}^T U_P F_{X_P} + F_{X_P}^T U_P^T H_{NB} D(\theta_{NB})H_{NB}^T Q_P D(\eta_P) = 2F_{X_P}^T D(\lambda_P)F_{X_P}. \quad (37)$$

В диагональной матрице $D(\eta_P)$ имеется по крайней мере один нулевой диагональный элемент, который входит в вектор η_{NB} . Пусть, для определенности, это будет элемент с номером k , т.е. $\eta_P^k = 0$. Тогда k -я строка и k -й столбец матрицы $D(\eta_P)$ нулевые. Отсюда следует, что k -й диагональный элемент матрицы, стоящей в левой части равенства (37), нулевой.

С другой стороны, если обратиться к матрице $F_{X_P}^T D(\lambda_P)F_{X_P}$, то она представима в виде

$$F = F_{X_P}^T D(\lambda_P)F_{X_P} = \sum_{i=1}^{n_P} \lambda_P^i f_i f_i^T,$$

где f_i есть i -я строка ортогональной матрицы F_{X_P} , $1 \leq i \leq n_P$. Поэтому k -й диагональный элемент матрицы F равен

$$F^{kk} = \sum_{i=1}^{n_P} \lambda_P^i (f_i^k)^2.$$

Так как $\lambda_P > 0_{n_P}$, то для того, чтобы равенство $F^{kk} = 0$ имело место, должно выполняться следующее: $f_i^k = 0$, $1 \leq i \leq n_P$, т.е. k -й столбец матрицы F_{X_P} нулевой. Это противоречит ортогональности матрицы F_{X_P} .

Схожим образом убеждаемся, что в такой паре $[X, Y]$ матрица H_{BN} отсутствует и имеет место равенство $\mathcal{R}(U_P) = \mathcal{R}(H_{NB})$. Утверждение доказано.

Следствие 3. В регулярной согласованной паре $[X, Y]$ матрица U_P представима в виде $U_P = Q_{BN} F_{X_P}^T$ или в виде $U_P = H_{NB} F_{Y_P}^T$.

Утверждение 9. Пусть $[X, Y]$ – допустимая регулярная согласованная пара. Для того, чтобы она была оптимальной, необходимо и достаточно, чтобы $n_p = 0$.

Доказательство. Достаточность. При $n_p = 0$ получаем, что матрица $X \circ Y$ нулевая. Так как $X \succeq 0, Y \succeq 0$, то отсюда следует выполнение условия комплементарности $X \bullet Y = 0$. Выполнение данного условия вместе с допустимостью точек X и Y гарантирует оптимальность пары $[X, Y]$.

Необходимость. Оптимальная пара $[X, Y]$ комплементарная. Но, поскольку она внутренняя, то имеет место равенство $XY = YX = 0_{nn}$, откуда $X \circ Y = 0_{nn}$. Поэтому подматрица U_p обязательно должна быть пустой, т.е. $n_p = 0$. Утверждение доказано.

Из утверждения 9 следует, что любая допустимая пара $[X, Y]$, в которой $n_p > 0$, не может быть оптимальной.

Возьмем произвольную допустимую регулярную согласованную пару $[X, Y]$, для которой справедливо разложение (12) с ортогональной матрицей U . Если пара $[X, Y]$ не является оптимальной, то в ортонормированном базисе, задаваемым столбцами матрицы U , матрицы X и Y оказываются блочно-диагональными с блоками соответственно порядков n_p, n_B и n_N , причем некоторые из них могут оказаться пустыми.

Имеем в общем случае следующее:

$$X_U = \begin{bmatrix} X_{U_p} & 0 & 0 \\ 0 & X_{U_B} & 0 \\ 0 & 0 & X_{U_N} \end{bmatrix}, \quad Y_U = \begin{bmatrix} Y_{U_p} & 0 & 0 \\ 0 & Y_{U_B} & 0 \\ 0 & 0 & Y_{U_N} \end{bmatrix}. \quad (38)$$

Здесь в первой матрице диагональные блоки согласно (33) и следствию к теореме 1 имеют вид

$$\begin{aligned} X_{U_p} &= U_p^T X U_p = F_{X_p} Q_{BN}^T Q_{BN} D(\eta_{BN}) Q_{BN} Q_{BN} F_{X_p}^T = F_{X_p} D(\eta_{BN}) F_{X_p}^T, \\ X_{U_B} &= U_B^T X U_B = Q_{BB}^T Q_{BB} D(\eta_{BB}) Q_{BB} Q_{BB} = D(\eta_{BB}), \\ X_{U_N} &= U_N^T X U_N = 0. \end{aligned} \quad (39)$$

Для второй матрицы Y_U выполняется соответственно следующее:

$$\begin{aligned} Y_{U_p} &= U_p^T Y U_p = F_{Y_p} H_{NB}^T H_{NB} D(\theta_{NB}) H_{NB} H_{NB} F_{Y_p}^T = F_{Y_p} D(\theta_{NB}) F_{Y_p}^T, \\ Y_{U_N} &= U_N^T Y U_N = H_{NN}^T H_{NN} D(\theta_{NN}) H_{NN} H_{NN} = D(\theta_{NN}), \\ Y_{U_B} &= U_B^T Y U_B = 0. \end{aligned} \quad (40)$$

Обозначим через Пусть \mathbb{S}_U^n подпространство пространства \mathbb{S}^n , образованное блочно-диагональными матрицами с блоками соответственно размеров n_p, n_B и n_N . Размерность подпространства \mathbb{S}_U^n равна

$$n_U = \dim \mathbb{S}_U^n = n_p^\Delta + n_B^\Delta + n_N^\Delta. \quad (41)$$

Из вышесказанного следует, что X_U и Y_U принадлежат \mathbb{S}_U^n . Пусть

$$\eta_U = [\eta_{BN}; \eta_{BB}; 0_{n_N}], \quad \theta_U = [\theta_{NB}; 0_{n_B}; \theta_{NN}].$$

Согласно (38)–(40) имеем $X_U = F_{X_U} D(\eta_U) F_{X_U}^T, Y_U = F_{Y_U} D(\theta_U) F_{Y_U}^T$, где

$$F_{X_U} = \begin{bmatrix} F_{X_p} & 0 & 0 \\ 0 & I_{n_B} & 0 \\ 0 & 0 & I_{n_N} \end{bmatrix}, \quad F_{Y_U} = \begin{bmatrix} F_{Y_p} & 0 & 0 \\ 0 & I_{n_B} & 0 \\ 0 & 0 & I_{n_N} \end{bmatrix}.$$

Обе матрицы F_{X_U} и F_{Y_U} ортогональные.

Рассмотрим дополнительно матрицы $C_U = U^T C U \in \mathbb{S}^n$ и $A_{i,U} = U^T A_i U \in \mathbb{S}^n$, являющиеся представлением матриц C и $A_i, i \in J^m$, в базисе, задаваемым столбцами ортогональной матрицы U .

Кроме того, произведем “сужение” этих матриц до блочно-диагональных матриц \tilde{C}_U и $\tilde{A}_{i,U}$ из \mathbb{S}_U^n , оставив только в C_U и $A_{i,U}$ их диагональные блоки, а остальные блоки взяв нулевыми, например,

$$\tilde{A}_{i,U} = \begin{bmatrix} A_{i,U_P} & 0 & 0 \\ 0 & A_{i,U_B} & 0 \\ 0 & 0 & A_{i,U_N} \end{bmatrix}, \quad i \in J^m, \tag{42}$$

где $A_{i,U_P} = U_P^T A_i U_P$, $A_{i,U_B} = U_B^T A_i U_B$, $A_{i,U_N} = U_N^T A_i U_N$.

Матрицы $\tilde{A}_{1,U}, \dots, \tilde{A}_{m,U}$ порождают в \mathbb{S}_U^n подпространство, которое обозначим через $\mathcal{E}_{\tilde{A}}$. Через $\mathcal{E}_{\tilde{A}}^\perp$ обозначим его ортогональное дополнение в \mathbb{S}_U^n . Размерность подпространства $\mathcal{E}_{\tilde{A}}^\perp$ равна $l_U = n_U - m$. Пусть $\tilde{K}_{j,U}$, $j \in J^{l_U}$, – произвольные матрицы из \mathbb{S}_U^n , задающие базис в подпространстве $\mathcal{E}_{\tilde{A}}^\perp$. Все матрицы $\tilde{K}_{j,U}$ имеют вид, аналогичный (42), а именно:

$$\tilde{K}_{j,U} = \begin{bmatrix} K_{j,U_P} & 0 & 0 \\ 0 & K_{j,U_B} & 0 \\ 0 & 0 & K_{j,U_N} \end{bmatrix}, \quad j \in J^{l_U},$$

где $K_{j,U_P} = U_P^T K_j U_P$, $K_{j,U_B} = U_B^T K_j U_B$, $K_{j,U_N} = U_N^T K_j U_N$.

Введем определение невырожденности для допустимой пары $[X_U, Y_U] \in \mathbb{S}_U^n \times Y_U^n$, считая, по-прежнему, что она согласованная и регулярная.

Определение 7. Допустимая пара $[X_U, Y_U] \in \mathbb{S}_U^n \times S_U^n$ называется *X-невырожденной*, если матрицы

$$B_{i,U} = \begin{bmatrix} A_{i,U_P} & 0 \\ 0 & A_{i,U_B} \end{bmatrix}, \quad i \in J^m, \tag{43}$$

линейно независимы. Соответственно, пара $[X_U, Y_U] \in \mathbb{S}_U^n \times S_U^n$ называется *Y-невырожденной*, если матрицы

$$B_{j,U} = \begin{bmatrix} K_{j,U_P} & 0 \\ 0 & K_{j,U_N} \end{bmatrix}, \quad j \in J^{l_U}, \tag{44}$$

линейно независимы. В случае, когда пара $[X_U, Y_U] \in \mathbb{S}_U^n \times Y_U^n$ является одновременно X-невырожденной и Y-невырожденной, о ней будем говорить как о *невырожденной* паре.

Принимая во внимание размерности диагональных блоков в матрицах (43) и (44), получаем, что пара $[X, Y]$ может быть невырожденной в том и только том случае, когда выполняются неравенства:

$$n_P^\Delta + n_B^\Delta \geq m, \quad n_P^\Delta + n_N^\Delta \geq l_U. \tag{45}$$

Если учесть, что $l_U = n_U - m$, то второе неравенство принимает вид $n_B^\Delta \leq m$.

Утверждение 10. Пусть допустимая регулярная согласованная пара $[X_U, Y_U] \in \mathbb{S}_U^n \times S_U^n$ является невырожденной. Тогда, если $n_N > 0$, то матрицы K_{j,U_N} , $j \in J^{l_U}$, порождают пространство \mathbb{S}^{n_N} . С другой стороны, если $n_B > 0$, то матрицы A_{i,U_B} , $i \in J^m$, порождают пространство \mathbb{S}^{n_B} .

Доказательство. Предположим, что $n_N > 0$ и докажем первое утверждение. Введем в рассмотрение подпространство $\mathcal{T}_{\tan}(X_U | \mathbb{S}_{U,+}^n)$ пространства \mathbb{S}_U^n , определив его следующим образом:

$$\mathcal{T}_{\tan}(X_U | \mathbb{S}_{U,+}^n) = F_{X_U} \{Z_U \in \mathbb{S}_U^n : Z_{U_N} = 0_{n_N n_N}\} F_{X_U}^T,$$

где $\mathbb{S}_{U,+}^n = \mathbb{S}_U^n \cap \mathbb{S}_+^n$. Множество $\mathcal{T}_{\tan}(X_U | \mathbb{S}_{U,+}^n)$ является касательным подпространством к конусу $\mathbb{S}_{U,+}^n$ в X_U . Ортогональным дополнением $\mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n)$ к $\mathcal{T}_{\tan}(X_U | \mathbb{S}_{U,+}^n)$ является множество

$$\mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n) = F_{X_U} \{Z_U \in \mathbb{S}_U^n : Z_{U_P} = 0_{n_P n_P}, Z_{U_B} = 0_{n_B n_B}\} F_{X_U}^\top.$$

Кроме того, через $\mathcal{R}_{\tilde{A},U}$ обозначим подпространство пространства \mathbb{S}_U^n , порожденное матрицами $\tilde{A}_{i,U}$, $i \in J^m$. Его ортогональным дополнением в \mathbb{S}_U^n будет множество

$$\mathcal{R}_{\tilde{A},U}^\perp = \{Z_U \in \mathbb{S}_U^n : \tilde{A}_{i,U} \bullet Z_U = 0, i \in J^m\}.$$

Положим

$$B_{i,U} = F_{X_U}^\top \tilde{A}_{i,U} F_{X_U} - M_{i,U}, \quad M_{i,U} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A_{i,U_N} \end{bmatrix}.$$

Отсюда получаем, что для матриц $\tilde{A}_{i,U}$, $i \in J^m$, справедливо разложение

$$\tilde{A}_{i,U} = \hat{B}_{i,U} + \hat{M}_{i,U}, \tag{46}$$

где

$$\hat{B}_{i,U} = F_{X_U} B_{i,U} F_{X_U}^\top \in \mathcal{T}_{\tan}(X_U | \mathbb{S}_{U,+}^n), \quad \hat{M}_{i,U} = F_{X_U} M_{i,U} F_{X_U}^\top \in \mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n). \tag{47}$$

Так как матрицы (43) линейно независимы, то матрицы $\hat{B}_{i,U}$, $i \in J^m$, также линейно независимы.

Покажем, что выполняется равенство

$$\mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n) \cap \mathcal{R}_{\tilde{A},U} = 0_{nn}. \tag{48}$$

От противного, пусть данное пересечение содержит ненулевую матрицу \hat{Z} . Тогда найдутся коэффициенты α_i , $i \in J^m$, в совокупности не равные нулю, такие, что

$$\hat{Z} = \sum_{i=1}^m \alpha_i \tilde{A}_{i,U} \in \mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n).$$

Поэтому на основании (46) и (47) имеем $\sum_{i=1}^m \alpha_i \hat{B}_{i,U} = 0_{nn}$, что противоречит линейной независимости матриц $\hat{B}_{i,U}$, $i \in J^m$.

Пусть \mathcal{L}_{K,U_N} – подпространство в \mathbb{S}^{n_N} , порождаемое матрицами K_{j,U_N} , $j \in J^{l_U}$. Если оно не совпадает с \mathbb{S}^{n_N} , то ортогональное в \mathbb{S}^{n_N} к \mathcal{L}_{K,U_N} подпространство содержит ненулевую матрицу Z_{U_N} порядка n_N . Имеем $Z_{U_N} \bullet K_{j,U_N} = 0$, $j \in J^{l_U}$. Данные равенства можно переписать в виде

$$\tilde{Z}_U \bullet \tilde{K}_{j,U} = 0, \quad j \in J^{l_U}, \tag{49}$$

где

$$\tilde{Z}_U = F_X \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Z_{U_N} \end{bmatrix} F_X^\top \in \mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n),$$

причем ранг матрицы \tilde{Z}_U совпадает с рангом матрицы Z_{U_N} , т.е. \tilde{Z}_U – ненулевая матрица. Кроме того, из (49) следует, что $\tilde{Z}_U \in \mathcal{R}_{\tilde{A},U}$. Таким образом,

$$\tilde{Z}_U \in \mathcal{T}_{\tan}^\perp(X_U | \mathbb{S}_{U,+}^n) \cap \mathcal{R}_{\tilde{A},U},$$

что противоречит (48). Поэтому на самом деле матрицы K_{j,U_N} , $j \in J^{l_U}$, порождают все пространство \mathbb{S}^{n_N} .

Второе утверждение доказывается аналогично.

4. НЬЮТОНОВСКАЯ СИСТЕМА

Из предположения о существовании решений у пары задач (1) и (2) следует, что эти решения должны удовлетворять условиям оптимальности:

$$\begin{aligned} X \bullet Y &= 0, \\ A_i \bullet X &= b^i, \quad i \in J^m, \\ Y &= C - \sum_{i=1}^m u^i A_i, \\ X &\succeq 0, \quad Y \succeq 0. \end{aligned} \quad (50)$$

Но равенство $X \bullet Y = 0$ при $X \succeq 0, Y \succeq 0$ выполняется тогда и только тогда, когда $XY = YX = 0_{nm}$. Поэтому первое равенство из (50) может быть заменено на $X \circ Y = 0_{nm}$ и решение задач (1) и (2), таким образом, сводится к отысканию решений системы

$$\begin{aligned} X \circ Y &= 0_{nm}, \\ A_i \bullet X &= b^i, \quad i \in J^m, \\ Y &= C - \sum_{i=1}^m u^i A_i, \end{aligned} \quad (51)$$

причем X и Y должны принадлежать конусу \mathbb{S}_+^n .

Наша цель заключается в решении системы (51) методом Ньютона при условии, что стартовые точки X и Y берутся допустимыми и согласованными. Более того, все последующие точки в ходе итерационного процесса остаются такими же.

Далее для простоты считаем, что $[X, Y]$ – допустимая регулярная согласованная пара. Кроме того, считаем, что она неоптимальная и неособая. Построим ньютоновскую систему уравнений для определения направлений перемещений ΔX и ΔY (ньютоновских направлений) в этой паре. Воспользуемся базисом, определяемым столбцами ортогональной матрицы U из разложения (12). Тогда условия оптимальности (51) в этом базисе запишутся в виде

$$\begin{aligned} X_U \circ Y_U &= 0_{nm}, \\ A_{i,U} \bullet X_U &= b^i, \quad i \in J^m, \\ Y_U &= C_U - \sum_{i=1}^m u^i A_{i,U}, \end{aligned} \quad (52)$$

где $C_U = U^T C U$, $A_{i,U} = U^T A_i U$, $i \in J^m$, и $X_U = U^T X U \succeq 0$, $Y_U = U^T Y U \succeq 0$.

Согласно (38) текущие матрицы X_U и Y_U в базисе, задаваемом матрицей U , являются блочно-диагональными. Кроме того, они допустимые, следовательно, удовлетворяют второму и третьему равенствам из (51). Эти равенства с учетом блочной диагональности X_U и Y_U можно переписать в виде

$$\tilde{A}_{i,U} \bullet X_U = b^i, \quad i \in J^m; \quad Y_U = \tilde{C}_U - \sum_{i=1}^m u^i \tilde{A}_{i,U}. \quad (53)$$

Так как пара $[X, Y]$ неоптимальная, то первое равенство из (52), разумеется, не выполняется.

С целью упрощения дальнейшего изложения, хотя это и не принципиально, освободимся от двойственной переменной u во втором равенстве (53). Для этого воспользуемся введенными ранее матрицами $\tilde{K}_{j,U}$, $j \in J^{l_U}$, задающими базис в подпространстве $\mathcal{E}_{\tilde{A}}^\perp$. Тогда, умножая скалярным образом равенство $Y_U = \tilde{C}_U - \sum_{i=1}^m u^i \tilde{A}_{i,U}$ последовательно на матрицы $\tilde{K}_{j,U}$, получаем

$$\tilde{K}_{j,U} \bullet Y_U = c^j, \quad j \in J^{l_U}, \quad (54)$$

где $c^j = \tilde{K}_{j,U} \bullet \tilde{C}_U$. Заметим также, что матрицы $\tilde{K}_{j,U}$, $j \in J^{l_U}$, порождают в \mathbb{S}_U^n подпространство, которое обозначим через $\mathcal{E}_{\tilde{K}}$. По своему определению $\mathcal{E}_{\tilde{K}}$ совпадает с $\mathcal{E}_{\tilde{A}}^\perp$. Подставляя (54) вместо

третьего равенства в условиях (51), приходим к следующей системе уравнений относительно X_U и Y_U :

$$\begin{aligned} X_U \circ Y_U &= 0_{nm}, \\ \tilde{A}_{i,U} \bullet X_U &= b^i, \quad i \in J^m, \\ \tilde{K}_{j,U} \bullet Y_U &= c^j, \quad j \in J^l. \end{aligned} \tag{55}$$

Составим теперь систему для нахождения ньютоновских направлений ΔX_U и ΔY_U , решающих систему (55). Так как матрицы X_U и Y_U симметричные, то чтобы сохранить их симметричность при перемещении вдоль направлений ΔX_U и ΔY_U будем также в качестве последних брать симметричные матрицы. Кроме того, поскольку матрицы X_U , Y_U и все матрицы \tilde{C}_U , $\tilde{A}_{i,U}$, $\tilde{K}_{j,U}$, где $i \in J^m$ и $j \in J^l$, принадлежат подпространству \mathbb{S}_U^n , то целесообразно и ньютоновские направления ΔX_U и ΔY_U брать из этого подпространства, т.е. полагать

$$\Delta X_U = \begin{bmatrix} \Delta X_{U_P} & 0 & 0 \\ 0 & \Delta X_{U_B} & 0 \\ 0 & 0 & \Delta X_{U_N} \end{bmatrix}, \quad \Delta Y_U = \begin{bmatrix} \Delta Y_{U_P} & 0 & 0 \\ 0 & \Delta Y_{U_B} & 0 \\ 0 & 0 & \Delta Y_{U_N} \end{bmatrix}. \tag{56}$$

Тогда линейная система алгебраических уравнений для нахождения ньютоновских направлений ΔX_U и ΔY_U в паре $[X, Y]$ имеет вид

$$\begin{aligned} Y_U \circ \Delta X_U + X_U \circ \Delta Y_U &= -X_U \circ Y_U, \\ \tilde{A}_{i,U} \bullet \Delta X_U &= 0, \quad i \in J^m, \\ \tilde{K}_{j,U} \bullet \Delta Y_U &= 0, \quad j \in J^l. \end{aligned} \tag{57}$$

Более того, если принять во внимание, что $X_U \circ Y_U D(\lambda)$, то первое уравнение (57) можно также переписать следующим образом:

$$Y_U \circ \Delta X_U + X_U \circ \Delta Y_U = -D(\lambda). \tag{58}$$

Из (57) следует, что матрицы ΔX_U и ΔY_U должны принадлежать подпространствам, ортогональным соответственно к подпространствам $\mathcal{E}_{\tilde{A}}$ и $\mathcal{E}_{\tilde{K}}$. Другими словами, $\Delta X_U \in \mathcal{E}_{\tilde{K}}$, $\Delta Y_U \in \mathcal{E}_{\tilde{A}}$.

Так как матрицы ΔX_U и ΔY_U имеют вид (56), то матричное уравнение (58) распадается на три уравнения, а именно:

$$\begin{aligned} Y_{U_P} \circ \Delta X_{U_P} + X_{U_P} \circ \Delta Y_{U_P} &= -D(\lambda_P), \\ D(\eta_{BB}) \circ \Delta Y_{U_B} &= 0_{n_B n_B}, \\ D(\theta_{NN}) \circ \Delta X_{U_N} &= 0_{n_N n_N}, \end{aligned} \tag{59}$$

причем из двух последних сразу получаем

$$\Delta X_{U_N} = 0_{n_N n_N}, \quad \Delta Y_{U_B} = 0_{n_B n_B}. \tag{60}$$

Второе и третье уравнения из (55) в развернутом виде переписываются следующим образом:

$$\begin{aligned} A_{i,U_P} \bullet \Delta X_{U_P} + A_{i,U_B} \bullet \Delta X_{U_B} + A_{i,U_N} \bullet \Delta X_{U_N} &= 0, \\ K_{j,U_P} \bullet \Delta Y_{U_P} + K_{j,U_B} \bullet \Delta Y_{U_B} + K_{j,U_N} \bullet \Delta Y_{U_N} &= 0, \end{aligned}$$

где $i \in J^m$, $j \in J^l$. Кроме того, если учесть (60), то они принимают следующий более простой вид:

$$A_{i,U_P} \bullet \Delta X_{U_P} + A_{i,U_B} \bullet \Delta X_{U_B} = 0, \quad i \in J^m, \tag{61}$$

$$K_{j,U_P} \bullet \Delta Y_{U_P} + K_{j,U_N} \bullet \Delta Y_{U_N} = 0, \quad j \in J^l. \tag{62}$$

Объединяя первое уравнение из (59) с (61) и (62), окончательно приходим к системе

$$\begin{aligned} Y_{U_p} \circ \Delta X_{U_p} + X_{U_p} \circ \Delta Y_{U_p} &= -D(\lambda_p), \\ A_{i,U_p} \bullet \Delta X_{U_p} + A_{i,U_B} \bullet \Delta X_{U_B} &= 0, \quad i \in J^m, \\ K_{j,U_p} \bullet \Delta Y_{U_p} + K_{j,U_N} \bullet \Delta Y_{U_N} &= 0, \quad j \in J^l. \end{aligned} \quad (63)$$

Так как $l_U = n_U - m$ и число n_U определяется (41), то общее число уравнений в системе (63) равно $2n_p^\Delta + n_B^\Delta + n_N^\Delta$. Оно совпадает с общим числом неизвестных.

Разрешая систему (63), можно найти приращения ΔX_{U_p} , ΔY_{U_p} , ΔX_{U_B} и ΔY_{U_N} и перейти к новым точкам. Конкретные выражения для этих приращений будут получены в отдельной статье. Там же будет дано описание итерационного процесса и предложен способ выбора шага перемещения.

СПИСОК ЛИТЕРАТУРЫ

1. *Nesterov Yu.E., Nemirovski A.S.* Interior Point Polynomial Algorithms in Convex Programming. SIAM Publications, SIAM, Philadelphia. 1994.
2. *Alizadeh F.* Interior point methods in semidefinite programming with applications to combinatorial optimization // SIAM J. Optim. 1995. V. 5. № 1. P. 13–51.
3. *Vandenberghe L., Boyd S.* Semidefinite programming // SIAM Rev. 1996. V. 38. № 1. P. 49–95.
4. *Todd M.J.* Semidefinite optimization // Acta Numer. 2001. V. 10. P. 515–560.
5. *Wolkowicz H., Saigal R., Vandenberghe L.* (eds). Handbook of Semidefinite Programming Kluwer Academic Publishers. Boston, Dordrecht, London. 2000.
6. *Laurent M., Rendl F.* Semidefinite Programming and Integer Programming. Handbooks on Operations Research and Management Science. 2005. V. 12. P. 393–514.
7. *De Klerk E.* Aspects of Semidefinite Programming. Interior Point Algorithms and Selected Applications. Kluwer Academic Publishers, 2004.
8. *Monteiro R.D.C.* First- and second-order methods for semidefinite programming // Mathematical Programming, Series B. 2003. V. 97. № 1–2. P. 209–244.
9. *Жадан В.Г.* Прямо-двойственный метод Ньютона для задач линейного программирования // Ж. вычисл. матем. и матем. физ. 1999. Т. 39. № 1. С. 17–32.
10. *Жадан В.Г.* О сходимости прямо-двойственного метода Ньютона для задачи линейного программирования // Ж. вычисл. матем. и матем. физ. 1999. Т. 39. № 5. С. 431–445.
11. *Арнольд В.И.* О матрицах, зависящих от параметров // Успехи матем. наук. 1971. Т. 26. В. 2 (158). С. 101–114.
12. *Alizadeh F., Haeblerly J.-P.F., Overton M.L.* Complementarity and nondegeneracy in semidefinite programming // Mathematical Programming, Series B. 1997. V. 77. № 2. P. 129–162.
13. *Magnus J.R., Neudecker H.* The elimination matrix: some lemmas and applications // SIAM J. Alg. Disc. Meth. 1980. V. 1. № 4. P. 422–449.
14. *Магнус Я.Р., Нейдеккер Ч.* Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М.: Физматлит, 2002.

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.245, 519.632.4

СТОХАСТИЧЕСКИЕ АЛГОРИТМЫ РЕШЕНИЯ ПЕРВОЙ КРАЕВОЙ ЗАДАЧИ ДЛЯ НЕКОТОРЫХ ЭЛЛИПТИЧЕСКИХ УРАВНЕНИЙ ВТОРОГО ПОРЯДКА С РАЗРЫВНЫМИ КОЭФФИЦИЕНТАМИ¹⁾

© 2022 г. А. Н. Кузнецов^{1,*}, А. С. Сипин^{1,**}¹ 160000 Вологда, ул. Ленина, 15, Вологодский государственный университет, Россия*e-mail: pt_kan@mail.ru**e-mail: cac1909@mail.ruПоступила в редакцию 11.02.2021 г.
Переработанный вариант 23.06.2021 г.
Принята к публикации 12.10.2021 г.

Рассматриваются стохастические алгоритмы решения первой краевой задачи для эллиптического уравнения второго порядка, коэффициенты которого имеют разрыв на гладкой поверхности. Предполагаются непрерывность решения и согласованность его нормальных производных с разных сторон поверхности разрыва коэффициентов. Предложена и доказана формула среднего значения в шаре (или эллипсоиде), которая позволяет определить случайное блуждание внутри области и статистические оценки (на его траекториях) для определения решения краевой задачи в начальной точке блуждания методом Монте-Карло. Библиография: 7.

Ключевые слова: эллиптический оператор, краевая задача, формула среднего значения, случайное блуждание, стохастический алгоритм, несмещенная оценка.

DOI: 10.31857/S0044466922020107

ВВЕДЕНИЕ

Для решения краевых задач используются различные численные методы, в том числе методы статистического моделирования, т.е. методы Монте-Карло (см., например, [1], [2]). Разработаны эффективные процедуры статистического моделирования для решения уравнений переноса излучения, уравнений газовой динамики, ряда задач в области электростатики, теории упругости и др. Статистические алгоритмы позволяют решать краевые задачи как внутри, так и вне ограниченной области, граница которой может иметь сложную структуру. Для широкого класса задач вычислительная работа в таких алгоритмах линейно зависит от размерности области.

При построении статистического алгоритма решение краевой задачи записывается в виде математического ожидания некоторой случайной величины ξ , т.е. случайная величина ξ является несмещенной оценкой решения задачи. Обычно несмещенные оценки для решения краевых задач строятся на траекториях марковского случайного процесса с дискретным временем (случайного блуждания). Само случайное блуждание происходит в замкнутой ограниченной области $D \subset \mathbb{R}^m$, в которой решается краевая задача. Вероятность перехода $P(x, A)$ из точки $x \in D$ в множество $A \subset D$ является субстохастическим ядром. Исходная краевая задача при этом заменяется эквивалентным ей интегральным уравнением в пространстве непрерывных функций $C(D)$:

$$u(x) = \int_D P(x, dy)u(y) + F(x), \quad x \in D, \quad (0.1)$$

где правая часть $F(x)$ уравнения (0.1) определяется через правую часть дифференциального уравнения и граничные условия. Пусть $x(0) = x, x(1), x(2), \dots$ — цепь Маркова с переходными вероятностями $P(x, B)$. Вообще говоря, это обрывающаяся цепь. На траекториях цепи стандартным образом строится последовательность $\xi(n), n = 0, 1, \dots$, несмещенных оценок для $u(x)$. Пусть N — мо-

¹⁾ Работа выполнена при финансовой поддержке РФФИ (грант 19-11-00020).

мент обрыва цепи ($N = \infty$, если цепь не обрывается), $N \wedge n$ – минимум из N и n , $\eta(i)$ – несмещенная оценка для $F(x(i))$, тогда

$$\xi(n) = u(x(N \wedge n)) + \sum_{i=0}^{N \wedge n - 1} \eta(i), \quad n = 1, 2, \dots \quad (0.2)$$

Из марковского свойства цепи и уравнения (0.1) следует

Лемма 1. Пусть $F(x) \geq 0$, $u(x)$ – ограниченное решение уравнения (0.1) и оценки $\eta(i)$ неотрицательны. Тогда последовательность несмещенных оценок (0.2) образует мартингал. Если он равномерно интегрируем, то он сходится с вероятностью 1 к случайной величине $\xi(\infty)$, имеющей конечное математическое ожидание. Для любого марковского момента τ случайная величина $\xi(\tau)$ является несмещенной оценкой $u(x)$.

Поведение траекторий процесса характеризует

Лемма 2. Пусть координатные функции x_i , $i = 1, 2, \dots, m$, являются супергармоническими для ядра $P(x, A)$, тогда цепь $x(n)$ сходится с вероятностью 1 к некоторому случайному вектору $x(\infty)$.

Доказательство лемм можно найти в [3]. При этом мы пользуемся терминологией и результатами из книги П. Мейера [4]. Вектор $x(\infty)$, как правило, распределен на границе области D . В качестве марковского момента τ обычно выбирается момент первого попадания траектории блуждания в ε -окрестность границы области D .

Интегральное уравнение (0.1) обычно получают по формуле Грина, используя функцию Грина или функцию Леви для семейства областей $T(x) \subset D$, $x \in D$. При этом носителем меры $P(x, A)$ будет либо граница области, либо сама область $T(x)$. Вид носителя меры $P(x, A)$ определяет название блуждания и название статистического метода решения задачи: *блуждание по сферам*, *блуждание по шарам*, *блуждание по эллипсоидам*, *блуждание по полусферам*, *блуждание по границе*.

Таким образом, в статье рассматриваются стохастические алгоритмы вычисления значения решения краевой задачи $u(x)$ в точке x ограниченной области по известным значениям этой функции в граничных точках.

1. ПОСТАНОВКА ЗАДАЧИ И ТЕОРЕМЫ О СРЕДНЕМ ЗНАЧЕНИИ

Пусть D – ограниченная область в \mathbb{R}^m , $m \geq 3$, с границей Γ , которая состоит из двух подобластей D_1 , D_2 и их общей кусочно-гладкой границы γ . Для функций $u \in C^2(D_1 \cup D_2) \cap C(\bar{D})$ определим оператор M равенством

$$Mu = \sum_{i,j=1}^m \frac{\partial}{\partial x_j} \left(a_{ij}(x) \frac{\partial}{\partial x_i} u \right), \quad (1.1)$$

где матрица $\|a_{i,k}(x)\|_{i,k=1}^m$ – симметричная, а ее элементы $a_{i,k} \in C^1(D_1 \cup D_2)$, причем их производные первого порядка непрерывны вплоть до границы в каждой подобласти. Предположим также, что матрица коэффициентов – положительно-определенная, т.е. для некоторой постоянной $\nu > 0$ при всех $x \in D$ и всех $z \in \mathbb{R}^m$ справедливо неравенство

$$\sum_{i,j=1}^m a_{ij} z_i z_j \geq \nu \|z\|^2.$$

Таким образом, M – эллиптический оператор.

Рассмотрим краевую задачу

$$Mu(x) = 0, \quad x \in D_1 \cup D_2, \quad (1.2)$$

$$\lim_{y \in D_1, y \rightarrow x} \sum_{i,j=1}^m n_i(x) a_{ij}(y) \frac{\partial}{\partial y_j} u(y) = \lim_{y \in D_2, y \rightarrow x} \sum_{i,j=1}^m n_i(x) a_{ij}(y) \frac{\partial}{\partial y_j} u(y), \quad x \in \gamma, \quad (1.3)$$

где $n(x) = (n_1(x), \dots, n_m(x))$ – внешняя нормаль к границе области D_1 . На внешней границе области D функция $u(x)$ известна и равна $\varphi(x)$.

Практически важным примером поставленной задачи является задача для уравнения

$$\operatorname{div}(\varepsilon(x) \operatorname{grad} u(x)) = 0, \quad x \in D_1 \cup D_2, \tag{1.4}$$

с условиями согласования нормальных производных

$$\lim_{y \in D_1, y \rightarrow x} \varepsilon_1(y) \frac{\partial}{\partial n} u(y) = \lim_{y \in D_2, y \rightarrow x} \varepsilon_2(y) \frac{\partial}{\partial n} u(y), \quad x \in \gamma. \tag{1.5}$$

Функции $\varepsilon_1(y)$ и $\varepsilon_2(y)$ предполагаются ограниченными снизу постоянной $\nu > 0$ и непрерывно дифференцируемыми в \bar{D}_1 и \bar{D}_2 соответственно.

В случае, когда функции $\varepsilon_1(y)$ и $\varepsilon_2(y)$ являются различными положительными постоянными ε_1 и ε_2 и граница раздела сред плоская, несмещенные оценки решения задачи (1.4), (1.5) построены в [3]. Аналогичные, но смещенные оценки, в областях с произвольной границей построены в работах Н.А. Симонова (см. [5], [6]).

Для построения несмещенных и малосмещенных оценок поставленной краевой задачи (1.4), (1.5) с переменными коэффициентами нам потребуется представление решения $u(x)$ для точек $x \in \gamma$ в интегральной форме. Для этого мы воспользуемся формулами Грина для подобласти T с кусочно-гладкой границей, расположенной в \bar{D}_1 или \bar{D}_2 :

$$\int_T (v(y)Mu(y) - u(y)Mv(y)) dy = \int_{\partial T} \varepsilon(y) \left(v(y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial v(y)}{\partial n} \right) d_y S,$$

справедливой для $u, v \in C^2(T) \cap C(\bar{T})$. Нормаль в этой формуле внешняя по отношению к области.

Пусть граница γ гладкая в окрестности $x \in \gamma$. Рассмотрим шар $T(x) \subset D$ с центром в точке x и радиуса $R = R(x)$. Рассмотрим два его полушара $T_1(x) = T(x) \cap D_1$ и $T_2(x) = T(x) \cap D_2$ и шар T_δ с центром в точке x и радиуса δ . В качестве функции $v(y)$ рассмотрим функцию Леви

$$\mathcal{L}(y, x) = \int_r^R \left(\frac{1}{r^{m-2}} - \frac{1}{\rho^{m-2}} \right) d\rho,$$

где $r = |y - x|$ – расстояние между точками x и y . Применяя формулу Грина к области $T = T_1 \setminus T_\delta$, при $\delta \rightarrow 0$ получаем вторую формулу Грина

$$-\int_{T_1} u(y)Mv(y)dy = \int_{\partial T_1} \varepsilon_1(y) \left(v(y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial v(y)}{\partial n} \right) d_y S - \sigma_m R(m-2)\varepsilon_1(x)u(x), \tag{1.6}$$

в которой σ_m обозначает “площадь поверхности” единичной сферы в \mathbb{R}^m . Аналогично получаем вторую формулу Грина для области T_2

$$-\int_{T_2} u(y)Mv(y)dy = \int_{\partial T_2} \varepsilon_2(y) \left(v(y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial v(y)}{\partial n} \right) d_y S - \sigma_m R(m-2)\varepsilon_2(x)u(x).$$

Заметим, что

$$\operatorname{grad}(v(y)) = (R - r) \operatorname{grad}(r^{2-m}) = (R - r)(2 - m)r^{-m}(y - x)$$

равен нулю на сферической части границы, сама функция $v(y)$ – тоже. Пусть $\gamma(x) = T(x) \cap \gamma$ – часть границы γ , лежащая в шаре. Упрощая формулы, получаем равенства

$$-\int_{T_1} u(y)Mv(y)dy = \int_{\gamma(x)} \varepsilon_1(y) \left(v(y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial v(y)}{\partial n} \right) d_y S - \frac{\sigma_m}{2} R(m-2)\varepsilon_1(x)u(x),$$

$$-\int_{T_2} u(y)Mv(y)dy = \int_{\gamma(x)} \varepsilon_2(y) \left(v(y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial v(y)}{\partial n} \right) d_y S - \frac{\sigma_m}{2} R(m-2)\varepsilon_2(x)u(x).$$

Нормали в этих формулах являются противоположными векторами. При сложении этих формул, используя условие на нормальные производные решения, получаем

$$\begin{aligned} & -\int_{\Gamma_1} u(y)Mv(y)dy - \int_{\Gamma_2} u(y)Mv(y)dy = \\ & = \int_{\gamma(x)} (\varepsilon_2(y) - \varepsilon_1(y))u(y) \frac{\partial v(y)}{\partial n} d_y S - \frac{\sigma_m}{2} R(m-2)(\varepsilon_1(x) + \varepsilon_2(x))u(x). \end{aligned} \quad (1.7)$$

Для плоской границы γ нормальная производная в этих формулах равна нулю. После несложных вычислений находим значение оператора

$$Mv(y) = \varepsilon(y)(m-2)r^{1-m}[1 - (R-r)(\text{grad}(\ln \varepsilon(y)), \text{grad}(r))]$$

и получаем формулу среднего значения

$$\begin{aligned} u(x) = \frac{2}{R\sigma_m(\varepsilon_1(x) + \varepsilon_2(x))} & \left(\int_{\Gamma_1} \varepsilon_1(y)r^{1-m}[1 - (R-r)(\text{grad} \ln \varepsilon_1(y), \text{grad} r)]u(y)dy + \right. \\ & \left. + \int_{\Gamma_2} \varepsilon_2(y)r^{1-m}[1 - (R-r)(\text{grad}(\ln(\varepsilon_2(y)), \text{grad}(r))]u(y)dy \right). \end{aligned} \quad (1.8)$$

Отметим, что осредняющее ядро в формуле (1.8) неотрицательно при выполнении условия

$$R \sup_{y \in D} |\text{grad}(\ln(\varepsilon(y)))| \leq 1,$$

которое справедливо при достаточно малых R в силу сделанных предположений о параметрах краевой задачи. Кроме того, оно стохастическое, так как после подстановки в формулу (1.8) решения краевой задачи $u(x) \equiv 1$ получим, что интеграл от ядра равен единице.

Для точек, не лежащих на границе $\Gamma \cup \gamma$, верны формулы среднего значения в шаре, а именно, справедлива следующая лемма.

Лемма 3. Если точка $x \in D_1$ и $T(x) \subset D_1$, то

$$u(x) = \frac{1}{\sigma_m R(m-2)\varepsilon_1(x)} \int_T u(y)Mv(y)dy, \quad x \in D_1. \quad (1.9)$$

Аналогично, если $x \in D_2$ и $T(x) \subset D_2$, то

$$u(x) = \frac{1}{\sigma_m R(m-2)\varepsilon_2(x)} \int_T u(y)Mv(y)dy, \quad x \in D_2. \quad (1.10)$$

Доказательство леммы непосредственно следует из формулы (1.6) и равенства нулю на границе шара как самой функции $v(y)$, так и ее градиента.

2. СТОХАСТИЧЕСКИЙ АЛГОРИТМ В СЛУЧАЕ ПЛОСКОЙ ГРАНИЦЫ РАЗДЕЛА ОБЛАСТИ

В случае переменных ε_1 и ε_2 для построения малосмещенной оценки решения краевой задачи предлагается следующий алгоритм.

1. Задаем произвольное $\delta > 0$, которое будет определять порядок погрешности приближенного решения задачи.

2. Определяем две δ -окрестности границы $\gamma_{1,\delta}$ и $\gamma_{2,\delta}$ как множества точек из D_1 и D_2 соответственно, отстоящих от границы γ не более, чем на δ . Также определяем Γ_δ – окрестность внешней границы.

3. Для каждой точки x из произвольной δ -окрестности границы определяем точку x^* как ближайшую к x точку границы.

4. Фиксируем начальную точку x , для которой будет строиться статистическая оценка $\xi(x)$ для значения $u(x)$ и полагаем $x_0 := x$.

5. Если $x_0 \in \Gamma_\delta$, то полагаем $\xi(x) := \varphi(x^*)$ и останавливаемся.

6. Если $x_0 \in \gamma_{1,\delta}$ или $x_0 \in \gamma_{2,\delta}$, то $x_0 := x_0^*$ и переходим из точки x_0 в точку x_1 , распределенную внутри области $T(x_0)$ с плотностью, которая является ядром в формуле (1.8). Полагаем $x_0 := x_1$, переходим к пункту 5.

7. Переходим из точки x_0 в точку x_1 , распределенную внутри области $T(x_0)$ с плотностью, которая является ядром в формуле (1.9) или формуле (1.10). Полагаем $x_0 := x_1$, переходим к пункту 5.

Назовем построенное случайное блуждание *блужданием по шарам*. Все переходы в этом блуждании легко моделируются методом отбора. Процедура моделирования подробно описана в [2].

Работоспособность алгоритма обеспечивается следующими его свойствами.

– Блуждание по шарам в области D_1 (в области D_2) достигает ее δ границы за конечное число шагов. При этом среднее число шагов конечно. (Эти свойства доказаны для любого блуждания по эллипсоидам, в том числе и для блуждания по шарам в [2].)

– Среднее число посещений границы блужданием по шарам конечно. Этот факт является очевидным следствием строго марковского свойства блуждания и ограниченности снизу вероятности события, состоящего в том, что процесс попадет впервые в Γ_δ раньше, чем в γ_δ . (Асимптотика этих средних при $\delta \rightarrow 0$ конечно важна, но ее удастся определить крайне редко.)

Замечание 1. Как известно (см. [3]), в случае постоянных коэффициентов ϵ_1 и ϵ_2 для обеспечения выхода блуждания на плоскую границу раздела γ можно использовать обычную функцию Грина для оператора Лапласа в полусфере. В алгоритме блуждания по полусферам существенно меньше смещение статистической оценки, поскольку оно определяется лишь заменой точного значения функции в последней точке блуждания на значение в ближайшей к ней точке границы Γ .

3. СТОХАСТИЧЕСКИЙ АЛГОРИТМ В СЛУЧАЕ ГЛАДКОЙ ГРАНИЦЫ РАЗДЕЛА ОБЛАСТИ

В случае гладкой границы γ вместо формулы (1.8) приходится использовать формулу (1.7), которая не имеет вероятностного смысла в связи с присутствием интеграла

$$\int_{\gamma(x)} (\epsilon_2(y) - \epsilon_1(y))u(y) \frac{\partial v(y)}{\partial n} d_y S.$$

Чтобы его убрать, нужно его оценить сверху величиной

$$\sup_{y \in \gamma} |(\epsilon_2(y) - \epsilon_1(y))u(y)| \int_{\gamma(x)} \left| \frac{\partial v(y)}{\partial n} \right| d_y S,$$

которая должна быть меньше $\delta \frac{\sigma_m}{2} R(m-2)(\epsilon_1(x) + \epsilon_2(x))$. Фактически получается неравенство на величину телесного угла, под которым граница $\gamma(x)$ видна из точки x . Выбор радиуса шара теперь зависит от кривизны поверхности в окрестности точки x . После выбрасывания интеграла получаем формулу

$$u(x) \approx \frac{2}{R\sigma_m(\epsilon_1(x) + \epsilon_2(x))} \left(\int_{T_1} \epsilon_1(y)r^{1-m} [1 - (R-r)(\text{grad} \ln \epsilon_1(y), \text{grad} r)] u(y) dy + \int_{T_2} \epsilon_2(y)r^{1-m} [1 - (R-r)(\text{grad}(\ln(\epsilon_2(y))), \text{grad}(r))] u(y) dy \right),$$

применение которой позволяет построить статистическую оценку решения $u(x)$.

В случае постоянных ϵ_1 и ϵ_2 можно использовать функцию Леви

$$\mathcal{L}(y, x) = \frac{1}{r^{m-2}} - \frac{1}{R^{m-2}}.$$

Тогда получится процесс *блуждания по сферам*.

Замечание 2. Используя формулу (1.6), можно совсем легко построить и обосновать для уравнения (1.4) алгоритм решения задачи со смешанными граничными условиями в областях с кусочно-гладкой границей, на одной части которой задано условие Неймана, а на другой – условие Дирихле. Такие алгоритмы рассматривались для уравнений Пуассона и Лапласа в [3], [5]–[7].

4. ЗАДАЧА ДЛЯ ПРОИЗВОЛЬНОГО ОПЕРАТОРА

Вернемся теперь к задаче для произвольного самосопряженного эллиптического оператора (1.1). Введем необходимые обозначения. Матрицу коэффициентов оператора будем обозначать через $A(x)$, а ее определитель — $|A(x)|$. Рассмотрим квадратичную форму $\sigma^2(y, x) = (y - x)'A^{-1}(x)(y - x)$, где $(y - x)'$ обозначает строку, транспонированную к столбцу $(y - x)$. Для точек $x \in D_1 \cup D_2$ определим функцию $\mathcal{L}(y, x)$ равенством

$$\mathcal{L}(y, x) = \int_{\sigma}^R (\sigma^{2-m} - \rho^{2-m}) d\rho.$$

Рассматривая в качестве множеств $T(x)$ области $T(x) = \{y \in D : \sigma^2(y, x) < R^2\}$, можно определить процесс блуждания по эллипсоидам и на его траекториях уже описанным методом построить статистические оценки для решения задачи (1.2), (1.3).

Авторы выражают благодарность рецензенту за замечания, способствовавшие улучшению работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Sabelfeld K.K.* Monte Carlo methods in boundary value problems. Berlin Heidelberg: Springer-Verlag, 1991.
2. *Ермаков С.М., Некруткин В.В., Сипин А.С.* Случайные процессы для решения классических уравнений математической физики. М.: Наука, 1984. *Ermafov S.M., Nekrutkin V.V., Sipin A.S.* Random processes for classical equations of mathematical physics. Dordrecht/Boston/London: Kluwer Acad. Publ., 1989.
3. *Ермаков С.М., Сипин А.С.* Процесс “блуждания по полусферам” и его применение к решению краевых задач // Вестн. СПбГУ. Сер. 1. 2009. Вып. 3. С. 9–18.
4. *Мейер П.А.* Вероятность и потенциалы. М.: Мир, 1973.
5. *Симонов Н.А.* Методы Монте-Карло для решения эллиптических уравнений с граничными условиями, включающими в себя нормальную производную // Докл. АН. 2006. Т. 410. № 2. С. 1–4.
6. *Симонов Н.А.* Алгоритмы случайного блуждания по сферам для решения смешанной краевой задачи и задачи Неймана // Сиб. журн. вычисл. матем. 2007. Т. 10. № 2. С. 209–220.
7. *Сипин А.С.* О стохастических алгоритмах решения краевых задач для оператора Лапласа // Вероятность и статистика. 23. Зап. научн. сем. ПОМИ. СПб.: ПОМИ, 2015. Т. 442. С. 133–142; *Sipin A.S.* J. Math. Sci. (N.Y.). 2017. V. 225:5. P. 812–817.

МАТЕМАТИЧЕСКАЯ
ФИЗИКА

УДК 517.958

КИНЕТИЧЕСКИЕ МОДЕЛИ АГРЕГАЦИИ, ПРИВОДЯЩЕЙ
К МОРФОЛОГИЧЕСКОЙ ПАМЯТИ ОБРАЗОВАВШИХСЯ СТРУКТУР

© 2022 г. С. З. Аджиев^{1,*}, В. В. Веденяпин^{2,3,**}, И. В. Мелихов^{1,***}

¹ 119991 Москва, Ленинские Горы, МГУ им. М.В. Ломоносова, хим. факультет, Россия

² 125047 Москва, Миусская пл., 4, ИПМ РАН, Россия

³ 117198 Москва, ул. Миклухо-Маклая, 6, РУДН, Россия

*e-mail: sergeyadzhiev@yandex.ru

**e-mail: vicveden@yahoo.com

***e-mail: melikhov@radio.chem.msu.ru

Поступила в редакцию 20.06.2021 г.

Переработанный вариант 27.07.2021 г.

Принята к публикации 12.10.2021 г.

Обсуждаются кинетические уравнения эволюции частиц дисперсного вещества, различающиеся по свойствам (размерам, скоростям, координатам центра масс и т.д.). Цель настоящей работы – это создание априорной математической модели и определение коэффициентов полученных уравнений по функциям распределения, получаемым в экспериментах. Задача, которая возникает, – получение правильных (физико-химически обоснованных) уравнений агрегации. Из системы уравнений эволюции дискретной функции распределения частиц дисперсного вещества мы получаем континуальные уравнения типа Фоккера–Планка, или типа Эйнштейна–Колмогорова, или диффузное приближение на функцию распределения агрегирующих частиц, различающихся по уровню агрегации и числу составляющих их молекул. Мы рассматриваем функции распределения, аппроксимирующие экспериментальные данные, и определяем по ним коэффициенты уравнения типа Фоккера–Планка. Библ. 38. Фиг. 2. Табл. 1.

Ключевые слова: дисперсное вещество, функция распределения частиц по свойствам, зарождение и рост частиц, агрегация и агломерация, система уравнений Смолуховского, уравнения Беккера–Дёринга, уравнение Фоккера–Планка, диффузное приближение, морфологическая память.

DOI: 10.31857/S0044466922020028

ВВЕДЕНИЕ

Задача настоящей работы, которая решалась и решается, – это создание оптимизированного математического описания кинетики зарождения, роста и агрегирования частиц по их экспериментальным функциям распределения по свойствам.

Структуру статьи определяет следующая основная идея математического описания результатов экспериментов. Сначала возникает постановка задачи: определяются параметры функции распределения. На этом этапе экспериментаторы предлагают параметры, которые они могут измерить, т.е. какие параметры выбирать для модели. Затем первый шаг математического моделирования – это создание априорной модели, содержащей неизвестные функции как параметры, подлежащие в дальнейшем определению. Априорная модель – это написание правильных (физико-химически обоснованных) уравнений, содержащих неизвестные параметры. Модель должна отвечать на вопросы: каковы условия эксперимента, как и в каком диапазоне их следует менять, и что следует определять в эксперименте, чтобы получить эти неизвестные функции априорной модели. Мы отвечаем на эти вопросы для нашей задачи. На основе экспериментов эти неизвестные функции определяются: возникает апостериорная модель. Так получается математическое описание экспериментов. После того исследование предполагается повторить, расширив перечень и интервалы варьирования учитываемых условий. Цель – предсказывать, как синтезировать новые материалы с наибольшей экологической безопасностью и с наименьшими энергетическими затратами.

Поэтому возникает вопрос: какую информацию о кинетике элементарных процессов сохраняет тело к моменту завершения процесса? Частичный ответ на этот вопрос дают данные о зависимости формы тел и некоторых других их свойств от условий синтеза, которую стали рассматривать как память тел (морфологическая память) о том, как они образовались. Это свойство тел, которое было названо морфологической памятью, можно результативно использовать при разработке технологий.

1. ПОСТАНОВКА ЗАДАЧИ

В работах Дж.К. Максвелла (см. [1]), Л. Больцмана (см. [2]) и др. исследователей кинетики различных процессов исследуется функция распределения по скоростям и координатам, зависящая от времени. В [2] Больцманом впервые была рассмотрена теорема, о которой мы будем неоднократно говорить в настоящей работе. Теперь же возникает многопараметрическое описание, которое связано с тем, что оно стало доступно для экспериментального наблюдения. Задача состоит в том, чтобы написать уравнения на функцию распределения частиц по этим многим параметрам.

В настоящей статье мы продолжаем линию работ [3], [4] и др., посвященных математическому моделированию зарождения и роста иерархических структур с помощью кинетического подхода.

Типичная структура представляет собой иерархическую структуру. Виды тел, отличающиеся уровнем иерархической структуры (номером поколения), будем обозначать параметром $k \in N$: частицы каждого k -го поколения укрупняются путем присоединения частиц всех $j < k$ видов, причем, как показывает эксперимент, происходит парное взаимодействие частиц.

При описании образования тел целесообразно различать несколько стадий процесса. Сначала зарождаются и растут первичные частицы, далее наступает стадия агрегации — образование агломератов из первичных частиц в результате их столкновений друг с другом, а затем происходит и образование агрегатов из них. Несмотря на то что все стадии идут одновременно, сначала преобладает процесс образования первичных частиц, потом — их агрегатов, т.е. вторичных агломератов, затем — агрегатов из вторичных агломератов и т.д. Итак, образование дисперсного вещества происходит стадийно, и каждая стадия вносит вклад в структуру образующегося вещества. Этот факт используется при моделировании.

Чтобы создать наноструктуру (и вообще, какую-то структуру) с нужными свойствами, используют явления образования и распада зародышей первичных наночастиц, их роста и растворения, появления и ликвидации ростовых дефектов, захвата и выброса захваченных примесей, агрегации и разрушения агрегатов, перекристаллизации и топохимических реакций спекания и керамизации (см. [3]). При описании данных явлений учитываются свойства \mathbf{X} каждой частицы: внутренние и внешние. Внешние параметры состояния частицы характеризуют ее вещественно-энергетический обмен со средой, а внутренние — перераспределение вещества и энергии внутри частицы.

На всех стадиях процесса рассматриваемая система неоднородна, причем на каждом ее участке присутствуют молекулы, агрегаты молекул, укрупняющиеся за счет присоединения молекул, агрегаты второго поколения и т.д. В первом приближении можно считать, что стенки и мешалка обеспечивают однородность среды и фазообразующего вещества. В континуальном приближении каждый участок характеризуется собственной функцией распределения частиц по свойствам:

$$\Psi(\mathbf{X}, t) = \sum_k \Psi_k(\mathbf{X}, t), \quad N_{k, \mathbf{X}} = \int_{\mathbf{X}_0}^{\mathbf{X}} \Psi_k(\mathbf{X}', t) d\mathbf{X}'.$$

Здесь $\Psi_k(\mathbf{X}, t)$ — функция распределения частиц поколения k ; $N_{k, \mathbf{X}}$ — локальная концентрация частиц k -вида, у которых параметры состояния меньше соответствующих компонент вектора \mathbf{X} ; \mathbf{X}_0 — минимально возможное значение свойства \mathbf{X} . Вектор \mathbf{X} не включает пространственные координаты частиц, которые отождествляются с координатами участка.

Для решения нашей задачи (выписывания корректных уравнений с коэффициентами, подлежащих в дальнейшем определению на основе эксперимента) нам понадобится выразить функ-

цию от многих параметров $\Psi_k(\mathbf{X}, t)$ через однопараметричные функции распределения. Нам понадобится следующая

Лемма. Пусть функция $f(x_1, x_2) > 0$ такая, что $\int f(x_1, x_2) dx_2$ конечен. Тогда функция f может быть представлена единственным способом в следующем виде:

$$f(x_1, x_2) = g_1(x_1)g_2(x_1, x_2),$$

где $\int g_2(x_1, x_2) dx_2 = 1$.

Доказательство. Пусть $g_1(x_1) \equiv \int f(x_1, x_2) dx_2$. Тогда $g_2(x_1, x_2) = \frac{1}{g_1(x_1)} f(x_1, x_2)$.

Такое представление приводит к теоретико-вероятностному описанию второй переменной x_2 как случайной величины: $g_2(x_1, x_2)$ как функции распределения величины x_2 при условии x_1 , и записывается в теории вероятностей как $g_2(x_2|x_1)$.

Это представление допускает обобщение на произвольное число переменных. Пусть $f(x_1, \dots, x_r) > 0$ и все интегралы $\int f(x_1, \dots, x_r) dx_{r-k} \dots dx_r$ ($k = 0, \dots, s - 2$) существуют. Тогда разложение $f = g_1(x_1) \dots g_k(x_1, \dots, x_k) \dots g_r(x_1, \dots, x_r)$, где $\int g_k(x_1, \dots, x_k) dx_k = 1$, определяется единственным способом. Такие формулы в теории вероятностей называются формулами Байеса (см. [5]), а кинетическая теория придает смысл сокращенного описания, как отмечалось в [3].

Пример 1 (см. [1], [2]). В кинетической теории вводят функцию $f(\mathbf{v}, \mathbf{x}, t)$ распределения молекул по скоростям $\mathbf{v} \in R^3$ и пространству $\mathbf{x} \in R^3$ в момент времени t . Обычно считают величину \mathbf{v} вероятностной переменной, и происходит сокращение описания в результате столкновений до локально максвелловского распределения частиц массы M :

$$f(\mathbf{x}, \mathbf{v}, t) = n(\mathbf{x}, t) \left(\frac{M}{2\pi k_B T(\mathbf{x}, t)} \right)^{3/2} \exp \left(- \frac{M(\mathbf{v} - \mathbf{V}(\mathbf{x}, t))^2}{2k_B T(\mathbf{x}, t)} \right),$$

где $n(\mathbf{x}, t) \equiv \psi_1(\mathbf{x}, t)$ – концентрация частиц массы M , $T(\mathbf{x}, t)$ – температура, k_B – постоянная Больцмана, $\mathbf{V}(\mathbf{x}, t)$ – значение средней скорости частиц с массой M .

Основной параметр функции распределения частиц каждого из k видов определяется как первый параметр в приведенном ниже разложении. Целесообразно брать число молекул n , составляющих частицу, в качестве основного параметра: $x_1 = n$, и тогда

$$\Psi_k(\mathbf{X}, t) = \psi_k(n, t) f_{k,2}(x_2|n; t) \dots f_{k,s}(x_s|n, x_2, \dots, x_{s-1}; t), \tag{1.1}$$

где $\psi_k(n, t)$ – функция распределения частиц по числу n , составляющих их молекул, $f_{k,i}(x_i|n, x_2, \dots, x_{i-1}; t)$, где $i = 2, \dots, s$, – это плотности распределений частиц по свойству x_i при условии, что параметры, выписанные после |, фиксированы: $\int f_{k,i}(x_i|n, x_2, \dots, x_{i-1}; t) dx_i = 1$.

В силу того, что число молекул n , составляющих частицу, – основной параметр, в первую очередь, представляет интерес рассмотрение функций распределения по этому параметру.

Пример 1 мы привели в качестве классического примера. Рассмотрим еще пример распределения (1.1) из [6].

Пример 2 (см. [1], [2], [6]). Пример распределения (1.1) дает максвелловская функция распределения частиц с массой $M_n = nM$, где M – масса одной молекулы фазообразующего вещества (см. [6]):

$$\Psi_1(n, \mathbf{x}, \mathbf{p}, t) = \psi_1(n, \mathbf{x}, t) \frac{1}{(2\pi M_n k_B T(\mathbf{x}, t))^{3/2}} \exp \left(- \frac{(\mathbf{p} - M_n \mathbf{V}(\mathbf{x}, t))^2}{2M_n k_B T(\mathbf{x}, t)} \right), \tag{1.2}$$

где $\mathbf{p} \in R^3$ – значение импульса частицы с массой M_n .

Более точно, масса в (1.1) должна быть представлена как

$$M_n = nM_0 + n_L M_L, \tag{1.3}$$

где M_0 и M_L – массы отдельных молекул вещества, образующего твердую фазу, и среды системы, n_L – количество молекул среды, захваченных в объеме и на поверхности частицы.

Тогда

$$n = (v - n_L v_{0L}) / v_0, \quad (1.4)$$

где v – объем частицы, v_0 и v_{0L} – объемы, занимаемые отдельными молекулами вещества, образующего твердую фазу, и среды системы внутри частицы. В рамках оценочной модели целесообразно предположить, что объем и площадь поверхности каждой частицы могут быть представлены как функции эффективного размера:

$$l_M = \gamma_1 l, \quad s = \gamma_2 l^2, \quad v = \gamma_3 l^3, \quad (1.5)$$

где γ_1 , γ_2 и γ_3 – параметры формы, l_M – максимальное расстояние между точками на поверхности частицы. Ряд закономерностей в поведении частицы можно выявить, если предположить, что ее форма близка к шару ($\gamma_1 = 1$, $\gamma_2 = \pi$, $\gamma_3 = \pi/6$) или кубу ($\gamma_1 = \sqrt{3}$, $\gamma_2 = 6$, $\gamma_3 = 1$). Для более точного моделирования необходимо использовать параметры формы, определяемые экспериментально.

Если частицы имеют форму пластины, то целесообразно полагать, что $v = hs$, где h – эффективная толщина частицы. Если частицы представляют собой цепочки, тогда $v = h^2 l_M$. Поэтому в общем случае, если $v \gg n_L v_{0L}$, то из (1.4) и (1.5) получаем

$$n = \gamma_\delta h^{3-\delta} l^\delta / v_0, \quad (1.6)$$

где v_0 – объем одной молекулы фазообразующего вещества, $\delta = 1$ для цепочек из молекул, $\delta = 2$ для пластин, $\delta = 3$ для объемов. Мы также предполагаем, что параметры формы γ_δ не зависят от l .

Число молекул n , составляющих частицу, является основным параметром, и поэтому в первую очередь представляет интерес для функций распределения дисперсного вещества по нему. В эксперименте мы получаем интегральную функцию распределения в зависимости от размера частиц l , у которой измерен дискретный набор значений. Интегральные функции распределения агрегатов поколения k , моделирующие (аппроксимирующие) экспериментальное дискретное распределение, будем обозначать через $\theta_k(l)$.

Мы рассматриваем систему, которая является двухкомпонентной и состоит из растворителя или газа-носителя и кристаллообразующего вещества: пересыщенного раствора или пара.

Частотные функции для частиц, которые являются объемами из молекул, могут быть пропорциональны $n^{1/3}$, $n^{2/3}$ или не зависеть от n . Первый случай означает, что активные центры представляют собой одномерную структуру, второй – то, что число активных центров пропорционально площади поверхности, а последний, что число активных центров – фиксированное число. Например, один или два активных центра возникают в случае цепочек из полимерных молекул. Энергия активации ϵ рассчитывается по уравнению Аррениуса (см. [6]–[8]): частота присоединения или отрыва молекулы (или более крупной частицы) от другой частицы пропорциональна $\exp(-\epsilon/kT)$, где ϵ – энергия активации объединения или распада. Ее определяют экспериментально, согласно (1.3), по частотным функциям объединения и распада.

Мы будем рассматривать только два параметра, характеризующие частицу: ее массу и потенциальную энергию, которую мы учтем в коэффициентах уравнений.

Цель нашего исследования – определить коэффициенты уравнений на функцию распределения по тем, которые получаются в экспериментах. Поэтому первая задача – это получение этих уравнений, вторая – рассмотрение классов функций распределения частиц, хорошо аппроксимирующих данные, получаемые в эксперименте, а третья – получение соотношений между параметрами таких функций и коэффициентами уравнений.

Первая задача, которая возникает, – получение правильных уравнений, описывающих кинетику дисперсного вещества и, следовательно, вывод одних уравнений из других: рассмотрение связи между дискретными и континуальными уравнениями, переход от многопараметрического описания к сокращенному (например, вывод уравнений на функцию распределения частиц по размерам из обобщенных уравнений типа Больцмана на функцию распределения тел по размерам и скоростям их центра масс). Вторая задача начала рассматриваться в [6], а третья – в [9], и мы также будем ее обсуждать в следующем разделе. Рассматривается дискретная функция распределения, и получаются уравнения на нее.

В разд. 2 формулируются дискретные уравнения баланса первичных частиц, учитывающие возможность их укрупнения и распада, и показывается, как из дискретного описания получают-ся континуальные уравнения типа Фоккера–Планка. В теории вероятностей такое уравнение называют дифференциальным уравнением Эйнштейна–Колмогорова (см. [10]). В силу параболического вида полученного уравнения оно также называется диффузным приближением.

В разд. 3 мы обсудим параметры и классы функций распределения, хорошо аппроксимирующие данные, получаемые в эксперименте.

В разд. 4 рассмотрено точное решение диффузного приближения, которое в каждый момент времени совпадает с предложенными новыми аппроксимирующими функции распределения, и получены другие классы моделирующих функций на основе анализа этого уравнения и соотношения между параметрами аппроксимирующих функций и коэффициентами уравнений, описывающих кинетику дисперсного вещества.

2. ПЕРЕХОД К КОНТИНУАЛЬНОМУ ОПИСАНИЮ

Уравнения коагуляции частиц были впервые выведены и изучены для случая растворов Смолуховским (см. [11]) (без учета фрагментации – случай Смолуховского). Эта система уравнения, конечно, не учитывала разделение частиц на агрегаты различных поколений.

Первая стадия агрегирования описывается так называемыми уравнениями Беккера–Дёринга (см. [12]) – это случай, когда только одна молекула может присоединяться или отделяться от другой частицы.

Выпишем систему уравнений объединения–фрагментации с учетом различия поколений агрегатов. Она получается похожей на систему уравнений Смолуховского (см. [11]), которая в случае учета фрагментации называется системой коагуляции–дробления (см. [13, с. 56, и ссылки в ней]). В правой части такой системы возникает суммирование по числу молекул, составляющих частицу. А у нас (см. [3], [4]) возникает двойное суммирование, так как мы учитываем, что частицы представляют собой иерархические структуры и различаются по поколениям k (соответствующий индекс суммирования l):

$$\begin{aligned} \frac{\partial N_k(n,t)}{\partial t} = & \sum_{p=1}^{k-1} \left(\sum_{c=n_{p,0}}^{m_p} \alpha_{k,p}(n-c,c) N_p(c,t) N_k(n-c,t) - \sum_{d=n_{p,0}}^{m_p} \beta_{k,p}(n,d) N_k(n,t) \right) + \\ & + \sum_{p=1}^{k-1} \left(\sum_{d=n_{p,0}}^{m_p} \beta_{k,p}(n+d,d) N_k(n+d,t) - \sum_{c=n_{p,0}}^{m_p} \alpha_{k,p}(n,c) N_p(c,t) N_k(n,t) \right) + \Omega_k(n,t) \end{aligned} \quad (2.1)$$

для $n = n_{p,0}, n_{p,0} + 1, \dots, m_p$, где $n_{p,0}$ и m_p – соответственно минимально и максимально возможные числа молекул в частице поколения p . В (2.1) $N_k(n,t)$ – число частиц поколения k , состоящих из n молекул; $\alpha_{k,p}(n,c)$ – частотная функция (сечение) объединения частицы поколения k , состоящих из n молекул, с частицей вида p , состоящей из c молекул; $\beta_{k,p}(n,d)$ – частота распада частицы вида k , состоящей из n молекул, на частицу того же поколения, состоящую из $(n-d)$ молекул, и на частицу вида p , состоящую из d молекул; $\Omega_k(n,t)$ означает, что число частиц вида k может также меняться в результате объединения двух частиц поколения $k-1$ или процесса, обратного данному, вследствие объединения частицы вида k с частицами поколения, большего либо равного k , и вследствие отпада частицы вида k от частиц поколения, большего либо равного k .

Однако эксперименты и вычисления на компьютере требуют системы дискретных уравнений с шагом по времени Δt :

$$\begin{aligned} \Delta N_k(n,t,\Delta t) = & \Delta t \left[\sum_{l=1}^{k-1} \left(\sum_{c=n_{l,0}}^{m_l} \alpha_{k,l}(n-c,c) N_l(c,t) N_k(n-c,t) - \sum_{d=n_{l,0}}^{m_l} \beta_{k,l}(n,d) N_k(n,t) \right) + \right. \\ & \left. + \sum_{l=1}^{k-1} \left(\sum_{d=n_{l,0}}^{m_l} \beta_{k,l}(n+d,d) N_k(n+d,t) - \sum_{c=n_{l,0}}^{m_l} \alpha_{k,l}(n,c) N_l(c,t) N_k(n,t) \right) + \Omega_k(n,t) \right]. \end{aligned} \quad (2.2)$$

Здесь $\Delta N_k(n,t,\Delta t)$ – изменение числа частиц $N_k(n,t)$ поколения k , состоящих из n молекул, за интервал времени Δt . Мы должны определить $\Delta N_k(n,t,\Delta t)$ через $N(n,t)$. Если

$\Delta N(n, t, \Delta t) \equiv N(n, t + \Delta t) - N(n, t)$, то мы имеем явную дискретизацию системы обыкновенных дифференциальных уравнений (2.1). Система уравнений (2.1) принадлежит классу систем уравнений физико-химической кинетики (см. [14]), потому для нее H -теорема выполняется при выполнении условия динамического равновесия (the complex balance condition) (см. [15]–[17]), а для (2.1) с явной дискретизацией по времени она не верна в этом случае: в [17], [18] доказано, что H -теорема не выполняется в случае, когда рассматриваются только отдельные молекулы и димеры. Но она справедлива для неявной дискретизации по времени: когда $\Delta N(n, t, \Delta t) \equiv N(n, t) - N(n, t - \Delta t)$ (см. [19], [21]). И поэтому не следует использовать явную дискретизацию по времени для компьютерного моделирования.

Простая континуальная аналогия системы уравнений (2.1) имеет вид

$$\begin{aligned} \frac{\partial \psi_k(n, t)}{\partial t} = & \sum_{l=1}^{k-1} \left(\sum_{c=n_{l,0}}^{m_l} \alpha_{k,l}(n-c, c) N_l(c, t) \psi_k(n-c, t) - \sum_{d=n_{l,0}}^{m_l} \beta_{k,l}(n, d) \psi_k(n, t) \right) + \\ & + \sum_{l=1}^{k-1} \left(\sum_{d=n_{l,0}}^{m_l} \beta_{k,l}(n+d, d) \psi_k(n+d, t) - \sum_{c=n_{l,0}}^{m_l} \alpha_{k,l}(n, c) N_l(c, t) \psi_k(n, t) \right) + \Omega_k(n, t) \end{aligned} \quad (2.3)$$

для $n \in [n_{k,0}, m_k]$. Здесь вместо суммирования по c и d мы можем написать интегралы.

Уравнения (2.3) обосновываются тем, что подстановка суммы δ -функций

$$\psi_k(n, t) = \sum_{i=n_{k,0}}^{m_k} N_k(i, t) \delta(n-i)$$

для каждого номера поколения k дает (2.1). Действительно, подставляя в (2.3) сумму δ -функций и интегрируя для каждого натурального $n \in [n_{k,0}, m_k]$ обе части полученного уравнения на интервале, содержащем только одно целое значение n , мы получаем (2.1) для этого n .

Перепишем (2.3) в виде

$$\begin{aligned} \frac{\partial \psi_k(n, t)}{\partial t} = & \sum_{l=1}^{k-1} \left(\sum_{c=n_{l,0}}^{m_l} (\alpha_{k,l}(n-c, c) N_l(c, t) \psi_k(n-c, t) - \alpha_{k,l}(n, c) N_l(c, t) \psi_k(n, t)) \right) + \\ & + \sum_{l=1}^{k-1} \left(\sum_{d=n_{l,0}}^{m_l} (\beta_{k,l}(n+d, d) \psi_k(n+d, t) - \beta_{k,l}(n, d) \psi_k(n, t)) \right) + \Omega_k(n, t), \end{aligned}$$

разложим приращения функций в этом уравнении в ряды Тейлора и рассмотрим только два первых члена в разложениях в ряды Тейлора. Мы можем пренебречь остальными членами, если размер частиц поколений меньших k достаточно мал по сравнению с размерами агрегатов вида k . Тогда мы получаем уравнения типа Фоккера–Планка на функции $\psi_k = \psi_k(n, t)$, которые имеют вид

$$\frac{\partial \psi_k}{\partial t} = -\frac{\partial}{\partial n} J(n, t) + \Omega_k(n, t) = -\frac{\partial}{\partial n} (G(n, t) \psi_k) + \frac{\partial^2}{\partial n^2} (D(n, t) \psi_k) + \Omega_k(n, t), \quad (2.4)$$

где $J(n, t) = G(n, t) \psi_k - \frac{\partial}{\partial n} (D(n, t) \psi_k)$,

$$G(n, t) = \sum_{l=1}^{k-1} \left(\sum_{d=n_{l,0}}^{m_l} d \beta_{k,l}(n, d) - \sum_{c=n_{l,0}}^{m_l} N_l(c, t) c \alpha_{k,l}(n, c) \right), \quad (2.5)$$

$$D(n, t) = \frac{1}{2} \sum_{l=1}^{k-1} \left(\sum_{d=n_{l,0}}^{m_l} d^2 \beta_{k,l}(n, d) + \sum_{c=n_{l,0}}^{m_l} N_l(c, t) c^2 \alpha_{k,l}(n, c) \right). \quad (2.6)$$

В (2.5), (2.6) получилась линейная зависимость коэффициентов уравнения типа Фоккера–Планка (2.4) от $N_l(c, t)$, $l < k$. Так получилось из-за того, что мы предполагали в (2.1) парность взаимодействия частиц, что, вообще говоря, верно только для разреженных газов.

Итак, мы получили уравнения типа Фоккера–Планка (2.4) с источником или стоком $\Omega_k(n, t)$.

Наша дальнейшая цель – определить коэффициенты уравнения типа Фоккера–Планка (2.4) по функциям распределения.

3. МОДЕЛИРУЮЩИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

В настоящем и следующих разделах мы будем опускать в обозначениях функций распределения нижний индекс k , определяющий номер поколения частицы.

Дисперсный состав порошков описывают с помощью дифференциальных функций распределения $\varphi(l)$ или $\psi(n)$, связь между которыми проявляется в интегральной функции распределения

$$\theta(l, t) \equiv \frac{1}{N(t)} \int_{l_0}^l \varphi(l', t) dl' = \frac{1}{N(t)} \int_{n_0}^{n(l)} \psi(n', t) dn',$$

где $l_0 = l_0(t)$ – минимальный размер ($\varphi = 0$ при $l < l_0$),

$$N(t) \equiv \int_{l_0}^{+\infty} \varphi(l', t) dl' = \int_{n_0}^{+\infty} \psi(n, t) dn$$

есть полное число частиц, $n(l)$ – число молекул фазообразующего вещества в частице размера l (см. (1.6)), $n_0 \equiv n(l_0)$ – минимальное значение числа таких молекул. На практике в основном применяются четыре формулы:

- Гордена–Андреева $\theta(l) = Al^q$ (см. [21], [22]),
- Розина–Раммлера–Беннета $\theta(l) = 1 - \exp(-Al^q)$ (см. [21], [23]),
- нормальное (см. [23], [25], [29]) или
- логарифмически нормальное (см. [24]–[29]) распределения.

В [24], [28], [29] рассматривался вопрос о недостаточности двухпараметричных распределений. В [27], [30] возникают распределения с несколькими модами: разные моды соответствуют разным поколениям агрегатов.

Оказалось, что существует наименьший размер частиц l_0 , меньше которого агрегатов не наблюдается. Поэтому использование указанных распределений с минимальным размером $l_0 = 0$ является нецелесообразным. Кроме того, формулы Розана–Раммлера–Беннета (1934 г.) и Гордена–Андреева (1951 г.) (которая является первыми двумя членами разложения в ряд Маклорена первой из них) содержат, вообще говоря, дробный показатель степени q . Обобщение этих формул на случай $l_0 > 0$ было рассмотрено в [4]. Но чтобы избежать дробности, можно рассматривать линейные комбинации с целыми показателями степени. Целые значения q соответствовали бы размерности пространства или процесса, согласно (1.2).

Для описания полимерных частиц, представляющих собой поверхности с фиксированным числом активных центров присоединения и распада, оказалось целесообразным использовать следующий класс функций:

$$\theta = 1 - \exp(-a_1 Z(s)(1 + a_2 Z(s))), \quad \text{где} \quad Z(s) \equiv s/s_0 - 1. \quad (3.1)$$

Здесь a_1, a_2, s_0 – параметры функции распределения, $s = l^2$ – площадь изображения частицы нанодисперсного вещества, $s_0 = l_0^2$. В одномерном случае (для цепочек из молекул) площадь изображения пропорциональна l . Уравнение (3.1) хорошо аппроксимирует функции распределения частиц, являющихся поверхностями из исходных частиц (с фиксированным числом активных центров). В этом случае $\delta = 2$ в (1.6), и можно переписать (3.1) в следующем виде:

$$\theta(l(n)) = 1 - \exp(-a_1 (n/n_0 - 1)(1 + a_2 (n/n_0 - 1))). \quad (3.2)$$

Поэтому, согласно (1.6), мы имеем классы функций

$$\theta(l) = 1 - \exp(-a_1 (l^\delta/l_0^\delta - 1)(1 + a_2 (l^\delta/l_0^\delta - 1))), \quad (3.3)$$

где $\delta = 1, 2, 3$ соответственно для цепочек, поверхностей и объемов из исходных частиц.

4. СООТНОШЕНИЯ МЕЖДУ ПАРАМЕТРАМИ АППРОКСИМИРУЮЩИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ И КОЭФФИЦИЕНТАМИ УРАВНЕНИЯ ТИПА ФОККЕРА–ПЛАНКА

Теперь мы попытаемся, по возможности, связать простые феноменологические формулы из разд. 3 с уравнениями типа Фоккера–Планка из разд. 2 настоящей работы.

Уравнение типа Фоккера–Планка имеет вид

$$\frac{\partial \Psi}{\partial t} = -\frac{\partial}{\partial n}(G(n,t)\Psi) + \frac{\partial^2}{\partial n^2}(D(n,t)\Psi). \quad (4.1)$$

Уравнение (4.1) рассматривается в замкнутой области $n \in [n_0(t), +\infty)$, $t \in [t_0, +\infty)$. Минимальное число составляющих частицу молекул n_0 может зависеть от времени, например, через зависимость n_0 от температуры: при различных температурах значения n_0 могут различаться. При $n < n_0(t)$ функция распределения обращается в нуль.

Пусть

$$G(n,t) = h(n)g(t), \quad D(n,t) = h(n)d(t). \quad (4.2)$$

Для $h(n)$ мы возьмем выражение

$$h(n) = n^\sigma, \quad \text{где } 0 \leq \sigma < 1. \quad (4.3)$$

Уточним возможные значения показателя σ в (4.3).

Активные центры (места на поверхности частицы, где присоединяются молекулы) представляют собой либо одномерную структуру, либо число активных центров пропорционально площади поверхности, либо число активных молекул на поверхности тела фиксировано. В последнем случае $\sigma = 0$. Для цепочек из молекул, для поверхностей из молекул и для объемных тел может реализовываться этот случай. Кроме того, в случае поверхностей из молекул активные центры могут представлять собой одномерную структуру, тогда $\sigma = 1/2$. А для объемных тел, кроме случая $\sigma = 0$, возможно, что $\sigma = 1/3$, когда активные центры представляют собой одномерную структуру, или что $\sigma = 2/3$, когда число активных центров пропорционально площади поверхности.

Во всех случаях $h(n)$ таково, что для достаточно больших n

$$h' \equiv dh(n)/dn \ll 1, \quad h'' \equiv d^2h(n)/dn^2 \ll 1. \quad (4.4)$$

Уравнение (4.1) с учетом (4.2) перепишем в виде

$$\frac{\partial \Psi}{\partial t} = -h'g\Psi - G(n,t)\frac{\partial \Psi}{\partial n} + h''d\Psi + 2h'd\frac{\partial \Psi}{\partial n} + D(n,t)\frac{\partial^2 \Psi}{\partial n^2}. \quad (4.5)$$

Сделаем замену переменных

$$\tau = h(n)\left(\int_{t_0}^t d(x)dx + a\right) + A, \quad z = n + h(n)\left(-\int_{t_0}^t g(x)dx + b\right) + B. \quad (4.6)$$

Тогда для функции $\tilde{\Psi}(z, \tau) = \Psi(n, t)$ получим, что

$$\frac{\partial \Psi}{\partial t} = \frac{\partial \tilde{\Psi}}{\partial \tau}hd + \frac{\partial \tilde{\Psi}}{\partial z}(-hg), \quad (4.7)$$

$$\frac{\partial \Psi}{\partial n} = \frac{\partial \tilde{\Psi}}{\partial \tau}h'\left(\int_{t_0}^t d(x)dx + a\right) + \frac{\partial \tilde{\Psi}}{\partial z}\left(1 + h'\left(-\int_{t_0}^t g(x)dx + b\right)\right), \quad (4.8)$$

$$\begin{aligned} \frac{\partial^2 \Psi}{\partial n^2} = & \frac{\partial^2 \tilde{\Psi}}{\partial \tau^2}\left(h'\left(\int_{t_0}^t d(x)dx + a\right)\right)^2 + 2\frac{\partial^2 \tilde{\Psi}}{\partial z \partial \tau^2}h'\left(\int_{t_0}^t d(x)dx + a\right)\left(1 + h'\left(-\int_{t_0}^t g(x)dx + b\right)\right) + \\ & + \frac{\partial \tilde{\Psi}}{\partial z^2}\left(1 + h'\left(-\int_{t_0}^t g(x)dx + b\right)\right)^2 + \frac{\partial \tilde{\Psi}}{\partial \tau}\left(h''\left(\int_{t_0}^t d(x)dx + a\right)\right) + \frac{\partial \tilde{\Psi}}{\partial z}h''\left(-\int_{t_0}^t g(x)dx + b\right). \end{aligned} \quad (4.9)$$

Подставляя (4.7)–(4.9) в (4.5), имеем уравнение на функцию $\tilde{\Psi}(z, \tau) = \Psi(n, t)$. Воспользовавшись условиями (4.4) и пренебрегая производными от $h(n)$, получим существенно более простое уравнение

$$\frac{\partial \tilde{\Psi}}{\partial \tau} = \frac{\partial^2 \tilde{\Psi}}{\partial z^2}. \quad (4.10)$$

В случае, когда $\sigma = 0$, замена (4.6) имеет вид

$$\tau = \int_{t_0}^t D(x) dx + a + A, \quad z = n - \int_{t_0}^t G(x) dx + b + B,$$

и в точности приводит исходное уравнение (4.1) к уравнению (4.10) (см. [31, с. 309]).

Возьмем следующее решение уравнения (4.10) (см. [31, с. 233]):

$$\tilde{\Psi}(z, \tau) = C \frac{z}{\tau^{3/2}} \exp\left(-\frac{z^2}{4\tau}\right), \quad C > 0,$$

и получим выражение для интегральной функции распределения

$$\theta = \frac{1}{N(t)} \int_{n_0}^n \Psi(n, t) dn = \frac{1}{N(t)} \int_{z(n_0, t)}^{z(n, t)} \tilde{\Psi}(z', \tau) dz' = 1 - \exp\left(-\frac{(z(n_0, t))^2}{4\tau(n_0, t)}\right) \exp\left(-\frac{(z(n, t))^2}{4\tau(n, t)}\right). \quad (4.11)$$

Итак, мы рассматриваем решение не исходного уравнения (4.1), (4.2), а уравнения, которое имеет простой вид (4.10) в переменных τ, z . В исходных переменных (4.10) дает громоздкое (в случае $\sigma \neq 0$) уравнение вида

$$\frac{\partial \Psi}{\partial t} \frac{\partial t}{\partial \tau} + \frac{\partial \Psi}{\partial n} \frac{\partial n}{\partial \tau} = \frac{\partial^2 \Psi}{\partial t^2} \left(\frac{\partial t}{\partial z}\right)^2 + \frac{\partial \Psi}{\partial t} \frac{\partial^2 t}{\partial z^2} + \frac{\partial^2 \Psi}{\partial n^2} \left(\frac{\partial n}{\partial z}\right)^2 + \frac{\partial \Psi}{\partial n} \frac{\partial^2 n}{\partial z^2},$$

коэффициенты которого определяются, согласно замене (4.6). Оно дает уравнение (4.1), (4.2) существенно более простого вида, если пренебречь производными от $h(n)$. Важно отметить, что в случае, когда $\sigma = 0$, мы в точности решаем уравнение (4.1).

Пусть

$$b_1(t) = \int_{t_0}^t d(x) dx + a, \quad b_2(t) = \int_{t_0}^t g(x) dx - b. \quad (4.12)$$

Тогда

$$b_1(t_0) = a, \quad b_2(t_0) = -b. \quad (4.13)$$

В силу (4.11), (4.6) и (4.12) имеем, что

$$\theta = \frac{1}{N(t)} \int_{n_0}^n \Psi(n, t) dn = 1 - \exp\left(-\frac{(n_0 - h(n_0) b_2(t) + B)^2}{4(h(n_0) b_1(t) + A)}\right) \exp\left(-\frac{(n - h(n) b_2(t) + B)^2}{4(h(n) b_1(t) + A)}\right). \quad (4.14)$$

Учитывая (4.3), перейдем в (4.14) к следующему виду:

$$\theta = 1 - \exp\left(\frac{(1 - c_2(t) + C_2(t))^2}{4(c_1(t) + C_1(t))}\right) \exp\left(-\frac{\left(\frac{n}{n_0} - c_2(t) \left(\frac{n}{n_0(t)}\right)^\sigma + C_2(t)\right)^2}{4\left(c_1(t) \left(\frac{n}{n_0(t)}\right)^\sigma + C_1(t)\right)}\right), \quad (4.15)$$

где

$$c_1(t) \equiv b_1(t)/n_0^{2-\sigma}(t), \quad C_1(t) \equiv A/n_0^2(t), \quad c_2(t) \equiv b_2(t)/n_0^{1-\sigma}(t), \quad C_2(t) \equiv B/n_0(t). \quad (4.16)$$

Поскольку $\tau > 0$, то из (4.6) получаем $h(n)a + A > 0$. Следовательно, $a > 0$ в силу (4.3) и поэтому $A > -h(n)a$. Откуда имеем, что

$$c_1(t) \geq c_1(t_0) = a/n_0^{2-\sigma} > 0, \quad C_1(t) > -c_1(t_0). \quad (4.17)$$

Так как $z \geq 0$, то, согласно (4.6) и (4.3), получаем, что

$$h(n) \left(n_0/h(n_0) - \int_{t_0}^t g(x) dx + b \right) + B \geq 0.$$

Следовательно, в силу (4.3)

$$b \geq \int_{t_0}^t g(x) dx - n_0/h(n_0), \quad B \geq h(n_0) \int_{t_0}^t g(x) dx - h(n_0)b - n_0.$$

Откуда имеем, что

$$c_2(t) \leq 1, \quad C_2(t) \geq c_2(t) - 1. \quad (4.18)$$

Если $\sigma = 0$, то, обозначая

$$d_1(t) \equiv c_1(t) + C_1(t), \quad d_2(t) \equiv c_2(t) - C_2(t) - 1, \quad (4.19)$$

из (4.15) получаем

$$\theta = 1 - \exp\left(\frac{d_2^2(t)}{4d_1(t)}\right) \exp\left(-\frac{(Z(n) - d_2(t))^2}{4d_1(t)}\right), \quad Z(n) = n/n_0. \quad (4.20)$$

Если возьмем $a_1 \equiv -d_2/(2d_1)$, $a_1 a_2 \equiv 1/(4d_1)$, то получим функцию распределения (3.2) и

$$d_1 \equiv 1/(4a_1 a_2), \quad d_2 \equiv -1/(2a_2). \quad (4.21)$$

Причем, согласно (4.17)–(4.19), (4.21), имеем $a_1(t) > 0$ и $a_2(t) > 0$.

В силу (4.12), (4.13), (4.16), (4.19), (4.21) мы получаем связь между частотными функциями уравнения типа Фоккера–Планка и параметрами моделирующей функции распределения (3.2) в случае $\sigma = 0$:

$$\begin{aligned} \int_{t_0}^t d(x) dx &= n_0^2(t)/(4a_1(t)a_2(t)) - n_0^2(t_0)/(4a_1(t_0)a_2(t_0)), \\ \int_{t_0}^t g(x) dx &= n_0(t)(1 - 1/(2a_2(t))) - n_0(t_0)(1 - 1/(2a_2(t_0))). \end{aligned} \quad (4.22)$$

В общем случае, когда σ необязательно равно нулю, получается целесообразным использовать неупрощенную аппроксимирующую функцию распределения (4.15), для которой параметры связаны с частотными функциями уравнения типа Фоккера–Планка (4.1), (4.2) соотношениями

$$\begin{aligned} \int_{t_0}^t d(x) dx &= n_0^{2-\sigma}(t)c_1(t) - n_0^{2-\sigma}(t_0)c_1(t_0), \\ \int_{t_0}^t g(x) dx &= n_0^{1-\sigma}(t)c_2(t) - n_0^{1-\sigma}(t_0)c_2(t_0). \end{aligned} \quad (4.23)$$

Полученные формулы (4.22), (4.23) связывают феноменологию разд. 5 с уравнением типа Фоккера–Планка разд. 2 настоящей работы.

Согласно определениям (2.5) и (2.6) функций G и D , они зависят от времени t только через температуру $T(t)$ и числа частиц поколений l , меньших k : $N_l(c, t)$, $l < k$. Если $T(t)$ и $N_l(c, t)$ постоянны или их изменениями в течение интервала времени $[t, t + \Delta t]$ можно пренебречь, то, согласно (4.23) (или (4.22) в случае $\sigma = 0$), по двум измерениям в моменты времени t и $t + \Delta t$ мы можем найти значения функций g и d для заданной температуры T и для чисел частиц $N_l(c, t)$: $g = n_0(c_2(t + \Delta t) - c_2(t))/\Delta t$, $d = n_0^2(c_1(t + \Delta t) - c_1(t))/\Delta t$.

Меняя условия системы: рассматривая различные значения T и $N_l(c, t)$, и вычисляя в каждом случае значения функций g и d , мы можем получить сами эти функции как зависимости от T и проверить линейность зависимости G и D от $N_l(c, t)$ в (2.5) и (2.6).

5. ОЦЕНКИ ДЛЯ ПАРАМЕТРОВ АППРОКСИМИРУЮЩИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

Для вычисления оптимальных значений параметров на компьютере нужно задать априорный диапазон изменения параметров.

В разд. 4 мы получили, что для функции (3.2) (или (3.1))

$$a_1(t) > 0, \quad a_2(t) > 0, \quad (5.1)$$

а для параметров функции (4.15) имеют место оценки (4.17), (4.18).

В настоящем разделе мы уточним (улучшим) оценки (5.1).

Итак, рассмотрим функцию (3.2) (или (3.1)):

$$\theta = 1 - \exp(-a_1 Z (1 + a_2 Z)), \quad Z \equiv s/s_0 - 1 = n/n_0 - 1, \quad (5.2)$$

и обозначим через $Z_{1/2}$ медиану, т.е. точку, где она принимает значение $1/2$, т.е. половина всех частиц дает значение Z меньшее, чем $Z_{1/2}$, а половина – большее:

$$1/2 = 1 - \exp(-a_1 Z_{1/2} (1 + a_2 Z_{1/2})). \quad (5.3)$$

Из (5.3) находим, что

$$a_1 a_2 Z_{1/2}^2 + a_1 Z_{1/2} - \ln 2 = 0. \quad (5.4)$$

Пусть мы имеем оценку для $Z_{1/2}$:

$$Z_{1/2}^{\min} \leq Z_{1/2} \leq Z_{1/2}^{\max}, \quad (5.5)$$

$Z_{1/2}^{\min}$ можно взять равным, например, $s_m/(2s_0)$, где s_m – площадь изображения одной молекулы, а $Z_{1/2}^{\max}$ оценить, например, значением 10^2 : медианное значение площади частицы не превосходит минимальное значение больше, чем в сто раз. Более точные значения $Z_{1/2}^{\min}$ и $Z_{1/2}^{\max}$ можно получить из рассматриваемых экспериментальных данных.

В силу (5.4) и (5.5) имеем

$$a_1 a_2 (Z_{1/2}^{\min})^2 + a_1 Z_{1/2}^{\min} - \ln 2 \leq 0, \quad (5.6)$$

$$a_1 a_2 (Z_{1/2}^{\max})^2 + a_1 Z_{1/2}^{\max} - \ln 2 \geq 0. \quad (5.7)$$

Из формул (5.6), (5.7) видно, что проще и лучше рассматривать получение оценок в переменных $a_1, a_1 a_2$: в этих переменных все неравенства задают полуплоскости.

Из неравенств (5.1), (5.6) получаются следующие оценки параметров:

$$0 \leq a_1 \leq \ln 2 / Z_{1/2}^{\min}, \quad 0 \leq a_1 a_2 \leq \ln 2 / (Z_{1/2}^{\min})^2.$$

Неравенство (5.7) не улучшает полученных оценок.

Обозначим через \tilde{Z} точку перегиба функции (5.2), ее значение определяется равенством $\partial^2 \theta / \partial Z^2 = 0$:

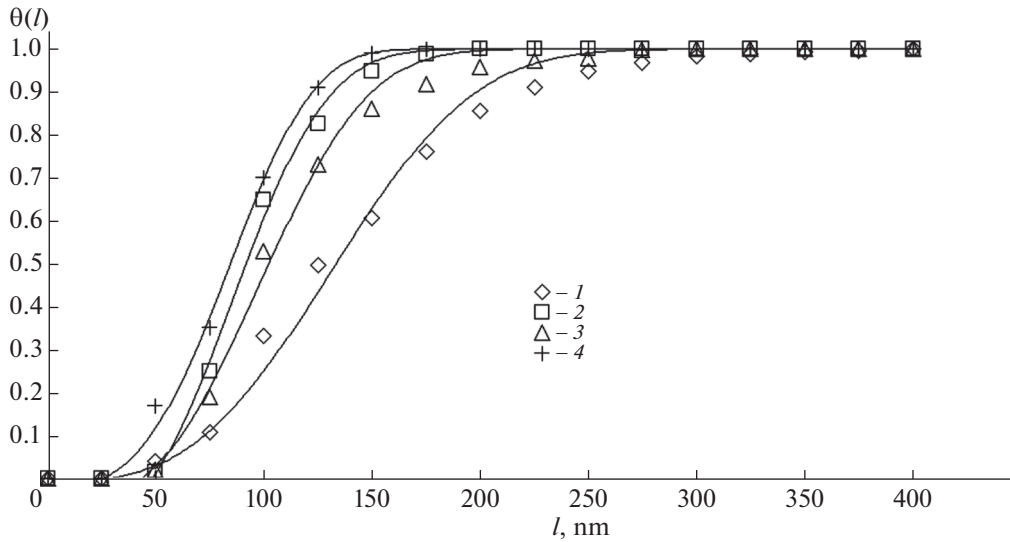
$$\tilde{Z} = \frac{(\sqrt{2a_1 a_2} - a_1)}{(2a_1 a_2)}. \quad (5.8)$$

Точка перегиба может быть у функции (5.2), когда $\tilde{Z} > 0$ или

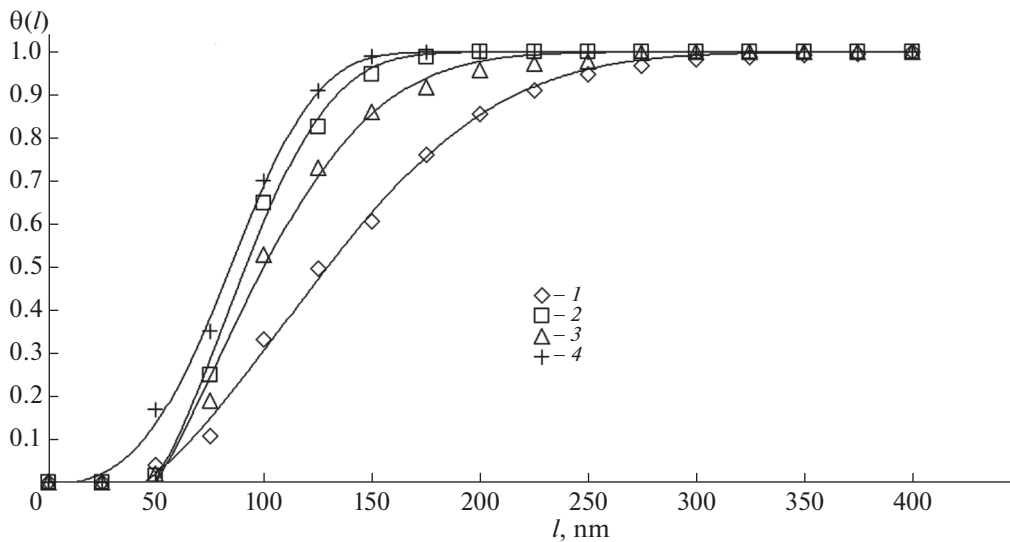
$$a_1 a_2 > a_1^2 / 2, \quad (5.9)$$

или отсутствовать при $\tilde{Z} \leq 0$ или

$$a_1 a_2 \leq a_1^2 / 2. \quad (5.10)$$



Фиг. 1. Аппроксимация с помощью (3.3) (или (4.15) с $\sigma = 0$) при $\delta = 3$ интегральной функции распределения носом по размерам $\theta(l)$ (l измеряется в нанометрах) при нагревании до температуры 303 (1), 313 (2), 323 (3) и 333 (4) К.



Фиг. 2. Аппроксимация с помощью (4.15) при $\delta = 3$ и $\sigma = 2/3$ интегральной функции распределения носом по размерам $\theta(l)$ (l измеряется в нанометрах) при нагревании до температуры 303 (1), 313 (2), 323 (3) и 333 (4) К.

В случае (5.9) имеем следующие улучшенные оценки:

$$0 \leq a_1 < (-1 + \sqrt{1 + 2 \ln 2}) / Z_{1/2}^{\min}, \quad (1 + \ln 2 - \sqrt{1 + 2 \ln 2}) / (Z_{1/2}^{\max})^2 < a_1 a_2 \leq \ln 2 / (Z_{1/2}^{\min})^2.$$

А в случае (5.10) имеем

$$(\sqrt{1 + 2 \ln 2} - 1) / Z_{1/2}^{\max} \leq a_1 \leq \ln 2 / Z_{1/2}^{\min}, \quad 0 \leq a_1 a_2 \leq (1 + \ln 2 - \sqrt{1 + 2 \ln 2}) / (Z_{1/2}^{\min})^2.$$

Аналогично, можно получить оценки параметров и для функции (4.15).

6. СРАВНЕНИЕ ТЕОРИИ С ЭКСПЕРИМЕНТОМ

В [32] анализ эффекта теплового разброса средних размеров наночастиц носом (представляющих собой поверхностно-активные вещества) обнаружил тенденцию к уменьшению их диа-

Таблица 1. Параметры функций (4.15) при $\delta = 3$ и $\sigma = 2/3$, аппроксимирующих интегральные функции распределения ниосом по размерам

Температура, К	303 (1)	313 (2)	323 (3)	333 (4)
$l_0, \pm 0.1$ нм	46.2	49.5	50.3	7.5
$1/(4c_1)$	$5 \times 10^{-6} \pm 10^{-6}$	$8.1 \times 10^{-3} \pm 5 \times 10^{-5}$	$1.5 \times 10^{-4} \pm 5 \times 10^{-5}$	$1.2 \times 10^{-5} \pm 10^{-6}$
$c_2, \pm 0.1$	-140.6	-3.8	-37.2	-10.1

метра при увеличении температуры. Ниосомы (поверхностно активные частицы, нерастворимые в воде), которые подвергли тепловой обработке, становились более устойчивыми и сохраняли средний размер частиц около 100 нм (фиг. 1 и 2). Этот подход может быть рекомендован для дальнейшего улучшения методов приготовления и хранения лекарственных форм.

Мы аппроксимировали экспериментальные данные из [32] (точки на фиг. 1 и 2) интегральными функциями распределения (3.3) или (4.15) с $\sigma = 0$ при $\delta = 3$ (фиг. 1) и (4.15) с $\sigma = 2/3$ при $\delta = 3$ (фиг. 2). Во втором случае мы пренебрегли отличием $C_1(t)$ и $C_2(t)$ от нуля, чтобы не увеличивать число параметров в (4.15) по сравнению с (3.3). Значения остальных параметров представлены в табл. 1. В первом случае мы получили плохую аппроксимацию, а во втором — хорошую. Это соответствует тому, что ниосомы представляют собой поверхностно-активные вещества. Итак, мы получили для этих экспериментальных данных, что коэффициенты уравнения типа Фоккера—Планка зависят не только от времени, но и от числа молекул n , составляющих частицы.

Оказалось, что доступными для определения стали трехпараметрические функции типа (4.15), (4.20), которые более информативны, чем обычно используемые нормальное и логнормальное распределения, а также распределения Розана—Рамллера—Беннета и Гордена—Андрева. На доступность трехпараметрических функций распределения указывают, в частности, опубликованные в [32] данные об изменении свойств нановезикул—ниосом при их нагревании до разных температур. Обработка этих данных показала, что изученные ниосомы были распределены по размеру (фиг. 2) в соответствии с функцией распределения (4.15) с $\delta = 3$ и $\sigma = 2/3$.

ЗАКЛЮЧЕНИЕ И ОБСУЖДЕНИЕ

В разд. 2 мы осуществили переход от дискретных уравнений к континуальному уравнению баланса числа частиц, учитывающему возможность их укрупнения и распада.

В разд. 3 мы рассмотрели основные известные и новые моделирующие (аппроксимирующие) функции распределения и их параметры. Такие функции, получаемые в экспериментах, не являются стационарными решениями рассматриваемой системы уравнений. Поэтому в разд. 4 мы нашли точные решения уравнения типа Фоккера—Планка, которые в каждый момент времени дают функции распределения частиц дисперсного вещества из разд. 3, и научились находить частотные функции по параметрам моделирующих функций распределения.

В разд. 5 получены оценки для параметров аппроксимирующих функций распределения. Они нужны для того, чтобы вычислить оптимальные значения параметров на компьютере, поскольку для этого требуется задать априорный диапазон изменения параметров.

Отметим, что рассмотрение H -теоремы для нелинейных систем с дискретным временем, в частности, даже для дискретных моделей уравнения Больцмана (см. [33], [34]) и для системы уравнений Беккера—Дёринга, является крайне важной задачей, поскольку компьютерное моделирование играет принципиальную роль при решении проблемы создания новых материалов. В линейном случае переход от непрерывного времени к дискретному дает переход от марковского процесса к марковской цепи, и H -теорема выполняется и изучена в этом случае (см. [35] и ссылки в ней, [37]). В нелинейном случае для явной дискретизации по времени она выполняется в редких случаях (см. [17], [18]), а для неявной она исследована в [17], [19], [37]. Диффузное приближение широко используется для моделирования процессов кристаллизации (см. [3], [4]). Однако мы доказали, что для рассматриваемого нелинейного уравнения типа Фоккера—Планка (2.4)—(2.6) H -теорема, вообще говоря, неверна (см. [37]). Но уравнения типа Фоккера—Планка (2.4)—(2.6), тем не менее, хорошо описывают эксперименты, поскольку в них функция распределения сдвигается и расплывается в пространстве свойств. Для расчетов на компьютере лучше использовать их дискретные аналоги, о которых мы говорили и для которых H -теорема исследована, как обсуждалось в настоящей работе (см. [18]—[20], [37]). Однако в эксперименте не удастся

измерить дискретную функцию распределения. Именно поэтому уравнения типа Фоккера–Планка так любят использовать химики. А надо использовать уравнения на гистограммные дискретные функции распределения, аналогичные с точностью до коэффициентов, рассмотренным дискретным аналогам, но это, возможно, тема для отдельной небольшой статьи.

Характерные особенности, которые возникли при решении задачи, обсуждаемой в настоящей работе, – это (1) существование минимального размера и (2) самоподобие функций распределения дисперсного вещества независимо от того, эволюция какого типа агрегатов рассматривается. Первый пункт мы учли, когда предлагали новые моделирующие функции распределения. Оба этих пункта нуждаются в обосновании.

В настоящей работе во всех разделах мы обобщаем результаты обзорной статьи [38].

СПИСОК ЛИТЕРАТУРЫ

1. *Максвелл Дж.К.* О динамической теории газов // Труды по кинетической теории. Пер. с англ. М.: Бинوم. Лаборатория знаний, 2011. С. 173–230; пер. с англ.: *Maxwell J.C.* On the Dynamical Theory of Gases // *Philosophical Magazine*. 1868. V. 35. № 235. P. 141–144.
2. *Больцман Л.* Дальнейшие исследования теплового равновесия между молекулами газа // Избранные труды. М.: Наука, 1984. С. 125–189, пер. с нем.: *Boltzmann L.* Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen // *Wien. Ber.* 1872. 66. P. 275–370; *Wissenschaftliche Abhandlungen*, v. 1, Barth, Leipzig. 1909. P. 316–402.
3. *Мелихов И.В.* Эволюционный подход к созданию наноструктур // *Наносистемы: физика, химия, математика*. 2010. Т. 1. № 1. С. 148–155.
4. *Мелихов И.В., Рудин В.Н., Козловская Э.Д., Аджиев С.З., Алексеева О.В.* Морфологическая память полимеров и возможности ее использования при разработке технологий новых материалов // *Теор. основы хим. технологии*. 2016. Т. 50. № 3. С. 268–268.
5. *Bayes Th., Price Th.* An Essay towards solving a Problem in the Doctrine of Chance // By the late Rev. Mr. Bayes communicated by Mr. Price, in a letter to John Canton, M.A. and F.R.S. *Philosophical Transactions of the Royal Society of London*. 1963. V. 53. P. 370–418.
6. *Аджиев С.З., Веденяпин В.В., Волков Ю.А., Мелихов И.В.* Обобщенные уравнения типа Больцмана для агрегации в газе // *Ж. вычисл. матем. и матем. физ.* 2017. Т. 57. № 12. С. 2065–2078.
7. *Arrhenius S.* Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren // *Z. Physik. Chemie (Leipzig)*. 1889. 4. P. 226–248.
8. *Штиллер В.* Уравнение Аррениуса и неравновесная кинетика. М.: Мир, 2000.
9. *Adzhiev S.Z., Melikhov I.V., Vedenyapin V.V.* The derivation of integration–fragmentation equations in the Becker–Döring case // *J. of Phys.: Conf. Ser.* 2019. V. 1205. № 012001. P. 1–6.
10. *Колмогоров А.Н.* Об аналитических методах в теории вероятностей // *Успехи матем. наук*. 1938. Вып. 5. С. 5–41.
11. *Smoluchowski M.* Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen // *Z. Phys. Chem.* 1917. 92. P. 129–168; пер. с англ.: *Смолуховский М.* Опыт математической теории кинетики коагуляции коллоидных растворов // *Коагуляция коллоидов*. Ред. А.И. Рабинович, П.С. Васильев. ОНТИ, 1936. С. 7–39.
12. *Becker R., Döring W.* Kinetische Behandlung der Keimbildung in übersättigten Dampfer // *Ann. Phys.* 1935. V. 24. P. 719–752.
13. *Веденяпин В.В.* Кинетические уравнения Больцмана и Власова. М.: Физматлит, 2001.
14. *Вольперт А.И., Худяев С.И.* Анализ в классах разрывных функций и уравнения математической физики. М.: Наука, 1975, 394 с.; англ. пер.: *Vol’pert A.I., Hudjaev S.I.* Analysis in classes of discontinuous functions and equations of mathematical physics, *Mech. Anal.*, 8. Dordrecht: Martinus Nijhoff Publ., 1985, xviii+678 pp.
15. *Horn F., Jackson R.* General mass action kinetics // *Arch. Ration. Mech. Anal.* 1972. V. 47. P. 87–116.
16. *Батищева Я.Г., Веденяпин В.В.* II-й закон термодинамики для химической кинетики // *Матем. моделирование*. 2005. Т. 17. № 8. С. 106–110.
17. *Мальшиев В.А., Пирогов С.А.* Обратимость и необратимость в стохастической химической кинетике // *Успехи матем. наук*. 2008. Т. 63. № 1. С. 3–36; англ. пер.: *Malyshev V.A., Pirogov S.A.* Reversibility and irreversibility in stochastic chemical kinetics // *Russ. Math. Surveys*. 2008. V. 63. № 1. P. 1–34.
18. *Adzhiev S., Melikhov I., Vedenyapin V.* The H-theorem for the chemical kinetic equations with discrete time and for their generalizations // *J. of Phys.: Conf. Ser.* 2017. V. 788. № 012001.
19. *Adzhiev S.Z., Melikhov I.V., Vedenyapin V.V.* The H-theorem for the physico-chemical kinetic equations with explicit time discretization // *Physica A: Statistic. Mech. and Appl.* 2017. V. 481. P. 60–69.
20. *Adzhiev S.Z., Melikhov I.V., Vedenyapin V.V.* The H-theorem for the physico-chemical kinetic equations with discrete time and for their generalizations // *Physica A: Statistic. Mech. and Appl.* 2017. V. 480. P. 39–50.

21. Коузов П.А. Основы анализа дисперсного состава промышленных пылей и измельченных материалов. Л.: Химия, 1987.
22. Андреев С.Е. О формулах среднего диаметра // Горн. журн. 1951. № 11. С. 32–36.
23. Rozin P., Rammlar E. Die Kornzusammensetzung des Mahlgutes im Lichte der Wahrscheinlichkeitslehre // Kolloid Zeitschrift. 1934. Bd. 67. H. I. S. 16–26.
24. Buchan G.D., Grewal K.S., Robson A.B. Improved models of particle-size distribution: an illustration of model comparison techniques // Soil Sci. Soc. Am. J. 1993. V. 57. P. 901–908.
25. Vargas-Ubera J., Aguilar J.F., Gale D.M. Reconstruction of particle-size distributions from light-scattering patterns using three inversion methods // Appl. Optics. 2007. V. 46. № 1. P. 124–132.
26. Dahi S.R., Clelland R., Hrenya C.M. The effects of continuous size distributions on the rapid flow of inelastic particles // Phys. of Fluids. 2002. V. 14. № 6. P. 1972.
27. Бакина О.В., Глазкова Е.А., Сваровская Н.В., Коровин М.С., Фоменко А.Н., Лернер М.И., Августинович А.В. Магнитные наночастицы Cu/Fe с противоопухолевой активностью // Сиб. онкологич. ж. 2018. Т. 17. № 1. С. 19–25.
<https://doi.org/10.21294/1814-4861-2018-17-1-19-25>
28. Чен Б.Б., Попель С.И., Филиппе Г., Адушкин В.В., Когай Г.А., Вайдлер П.Г., Фридрих Ф. Основные оптические и микрофизические свойства нано- и микромасштабного аэрозоля в атмосфере центрально-азиатского региона // Вестн. КРСУ. 2017. Т. 17. № 8. С. 28–35.
29. Wagner L.E., Ding D. Representing aggregate size distributions as modified lognormal distributions // Transact. of the ASAE. 1994. V. 37. № 3. P. 815–821.
30. Buchan G.D., Grewal K.S., Robson A.B. Improved models of particle-size distribution: an illustration of model comparison techniques // Soil Sci. Soc. Am. J. 1993. V. 57. P. 901–908.
31. Полянин А.Д., Зайцев В.Ф. Справочник по дифференциальным уравнениям с частными производными. Точные решения. Изд-во: Междунар. программа образования, 1996.
32. Diskaeva E.I., Vecher O.V., Bazikov I.A., Vakalov D.S. Particle size analysis of niosomes as a function of temperature // Nanosystems: Physics, Chemistry, Mathematics. 2018. V. 9. № 2. P. 290–294.
33. Аджиев С.З., Амосов С.А., Веденяпин В.В. Одномерные дискретные модели кинетических уравнений для смесей // Ж. вычисл. матем. и матем. физ. 2004. Т. 44. № 3. С. 553–558.
34. Аджиев С.З., Веденяпин В.В. О размерах дискретных моделей уравнения Больцмана для смесей // Ж. вычисл. матем. и матем. физ. 2007. Т. 47. № 6. С. 1045–1054.
35. Аджиев С.З., Веденяпин В.В. Временные средние и экстремали Больцмана для марковских цепей, дискретного уравнения Лиувилля и круговой модели Марка Каца // Ж. вычисл. матем. и матем. физ. 2011. Т. 51. № 11. С. 2063–2074.
36. Веденяпин В.В., Аджиев С.З. Энтропия по Больцману и Пуанкаре // Успехи матем. наук. 2014. Т. 69. № 6. С. 45–80; англ. пер.: Vedenyapin V.V., Adzhiev S.Z. Entropy in the sense of Boltzmann and Poincaré // Russ. Math. Surveys. 2014. V. 69. № 6. P. 995–1029.
37. Adzhiev S.Z., Melikhov I.V., Vedenyapin V.V. On the H-theorem for the Becker–Döring system of equations for the cases of continuum approximation and discrete time // Physica A: Statistic. Mech. and Appl. 2020. V. 553. № 124608. P. 1–13.
38. Adzhiev S.Z., Melikhov I.V., Vedenyapin V.V. Approaches to determining the kinetics for the formation of a nano-dispersed substance from the experimental distribution functions of its nanoparticle properties // Nanosystems: Physics, Chemistry, Mathematics. 2019. V. 10. № 5. P. 549–563.

МАТЕМАТИЧЕСКАЯ
ФИЗИКА

УДК 519.34

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОПАСНЫХ ЯВЛЕНИЙ
ПРИРОДНОГО ХАРАКТЕРА В МЕЛКОВОДНОМ ВОДОЕМЕ¹⁾

© 2022 г. А. М. Агаян^{1,2}, А. В. Никитина^{1,2,3,4,*}, А. И. Сухинов¹, А. Е. Чистяков^{1,2}

¹ 344000 Ростов-на-Дону, пл. Гагарина, 1, Донской государственный технический университет, Россия

² 347900 Сочи, Олимпийский пр-т, 1, Научно-технологический университет “Сириус”, Россия

³ 344006 Ростов-на-Дону, Большая Садовая ул., 105/42, Южный федеральный университет, Россия

⁴ 347900 Таганрог, пер. Итальянский, 106, Научно-исследовательский центр супер-ЭВМ
и нейрокомпьютеров, Россия

*e-mail: nikitina.vm@gmail.com

Поступила в редакцию 15.02.2021 г.

Переработанный вариант 17.07.2021 г.

Принята к публикации 04.08.2021 г.

Работа посвящена построению и исследованию взаимосвязанных математических моделей гидрофизики и биологической кинетики, используемых для прогнозирования опасных явлений природного характера, возникающих в мелководных водоемах. На распространение и трансформацию гидробионтов влияют такие физические факторы, как пространственно-трехмерное движение водной среды с учетом адвективного переноса и микротурбулентной диффузии, пространственно-неоднородное распределение температуры, солености и кислорода. Биогенные загрязняющие вещества вызывают рост водорослей, в том числе, токсичных и вредоносных, их массовое развитие может приводить к возникновению опасных явлений в водоеме, включая эвтрофикацию и заморные явления. Построена и исследована трехмерная математическая модель гидродинамики, используемая в работе для расчета поля скоростей водного потока. Для исследования опасных явлений мелководного водоема, связанных с заморными явлениями в нем, разработана пространственно-неоднородная трехмерная ихтиологическая модель динамики промысловой рыбы. Рассмотрены модели наблюдений, параметризованные на основе стехиометрических соотношений, законов Моно, Михаэлиса–Ментен и Митчерлиха–Бауле, описывающие потребление, накопление планктоном и промысловыми рыбами-детритофагами питательных веществ, а также рост гидробионтов в зависимости от пространственного распределения солености и температуры, кислородного режима. Для калибровки и верификации разработанных моделей использовались постоянно пополняемые базы экологических данных, полученные, в том числе, и с помощью экспедиционных исследований Азовского моря и Таганрогского залива. Для повышения точности прогнозного моделирования натурные данные были отфильтрованы на основе алгоритма Калмана. При решении задачи обработки гидрологической информации получены изолинии солености и температуры в поверхностном слое, для чего применен алгоритм распознавания. С помощью алгоритма интерполяции и путем наложения границ области получены более подробные карты глубин, солености и температуры Азовского моря. Разработаны численные методы решения поставленных задач, использующие конечно-разностные схемы, учитывающие степень заполненности контрольных ячеек расчетной области, реализованные на высокопроизводительных вычислительных системах, позволяющие уменьшить погрешность численного решения задачи и сократить время расчетов в несколько раз. На основе численной реализации разработанных моделей проведена реконструкция опасных явлений природного характера, возникающих в мелководном водоеме (связанных с распространением вредных загрязняющих веществ), эвтрофикацию, “цветение водорослей”, вызывающее заморные явления в водоеме. Библ. 21. Фиг. 12. Табл. 1.

Ключевые слова: математическая модель, гидродинамика, биологическая кинетика, планктон, промысловая рыба, программный комплекс, высокопроизводительная вычислительная система, опасные явления, мелководный водоем.

DOI: 10.31857/S0044466921120048

ВВЕДЕНИЕ

Решение многих задач водной экологии начинается с расчета гидрофизических параметров. Гидродинамические характеристики водоемов (поле скоростей водного потока, функция возвы-

¹⁾ Работа выполнена при финансовой поддержке РФФИ (проект № 19-31-51017).

шения уровня) используются в качестве входных данных для расчета концентраций загрязняющих веществ и изучения развития всего живого в водной среде, начиная от микроорганизмов, фито-, зоопланктона и заканчивая млекопитающими.

Для изучения процессов гидродинамики и биологической кинетики широко используются методы и подходы математического моделирования, позволяющие значительно расширить возможности исследователя, разработать предупредительные меры по локализации, устранению и уменьшению вреда опасных явлений, периодически возникающих в водных экосистемах в результате воздействия природных и техногенных факторов.

Изучению динамических процессов, ответственных за формирование циркуляции атмосферы и океана, посвящены работы Г.И. Марчука, В.П. Дымникова. В этих работах на основе последовательной разработки и анализа математических моделей исследуются процессы переноса примесей в атмосфере и океане, нелинейная передача энергии по спектру в квазидвумерном и квазигеострофическом приближениях, динамика крупномасштабных волн и их устойчивость, формирование пограничных струйных течений. Даны оценки точности аппроксимаций задач на базе использования неявных разностных аппроксимаций метода расщепления. Доказана абсолютная устойчивость разностных схем и установлены балансные соотношения в форме квадратичных функционалов. Введены в рассмотрение сопряженные уравнения гидродинамики и сформулирована теория возмущений, на основе которой делаются выводы о предсказуемости метеорологических элементов (см. [1]).

Для выявления общих свойств сложных пространственных течений широко используются численные методы (см. [2]). Известно, что сложно получить достаточно точное решение физической задачи, связанное с нахождением гидродинамических характеристик, если граничные условия не заданы адекватным образом. Уравнения Навье—Стокса и неразрывности представляют собой дифференциальные уравнения, описывающие удивительно широкий класс разнообразных течений, которые при изменении условий на границе могут отличаться друг от друга не только количественно, но и качественно.

В публикациях авторов R. Anson, T. Sonar представлены математические модели, включающие уравнения Навье—Стокса, Эйлера, уравнения пограничного слоя, модели турбулентности с целью получения качественного, а также количественного представления о процессах потоковых явлений (см. [3]).

Работа В. Roux и др. [4] посвящена изучению нелинейных гидродинамических процессов на основе 3D-модели RANS в широкой полузаткнутой лагунной экосистеме, а также условий восстановления погруженной водной растительности.

Изучению процессов биологической кинетики посвящен ряд работ Е.В. Якушева (см. [5]), Д.О. Логофета (см. [6]), Ю.В. Тютюнова (см. [7]), В.Г. Ильичева, занимающихся изучением влияния различных факторов как внешних природных, так и биотических, включая таксис, межвидовую конкуренцию, механизм эктокринной регуляции, на динамику изучаемых продукционно-деструкционных процессов.

Интерес к численному решению класса таких задач не ослабевает — в последние десятилетия появилось множество программных комплексов и отдельных модулей, которые широко используются для изучения гидрофизических процессов. В институте ИПМ им. М.В. Келдыша РАН Б.Н. Четверушкиным, В.А. Гасиловым, С.В. Поляковым, М.В. Якововским и др. разработан пакет прикладных программ GIMM для решения задач гидродинамики на многопроцессорных вычислительных системах (см. [8]).

Модуль Вычислительная гидродинамика, расширяющий возможности среды численного моделирования COMSOL Multiphysics, разработанный сотрудниками компании COMSOL, основанной еще в 1986 г. в Стокгольме (Швеция), объединяет несколько офисов и дистрибьюторскую сеть, раскинувшуюся по всему миру. Разработанное программное обеспечение может быть применимо для численного анализа систем (в которых гидродинамические процессы сопровождаются другими физическими явлениями), содержит инструменты для создания ключевых моделей течений, описывающих несжимаемые и сжимаемые среды, ламинарные и турбулентные течения, однофазные и многофазные потоки, течения в свободной и пористой среде, а также в открытых областях, течения в тонких пленках. Эти возможности реализованы в структурированных гидродинамических интерфейсах, предназначенных для постановки, решения и анализа стационарных и нестационарных задач в двумерных, двумерных осесимметричных и трехмерных областях. В модуле реализованы алгоритмы решения задач о течении неньютоновских жидкостей, течении во вращающемся оборудовании и течении при высоких числах Маха. Возможность описать в модели сразу несколько физических явлений является ценной при анализе гид-

родинамических процессов. С помощью модуля можно строить модели сопряженной теплопередачи и химически-реагирующих потоков. Имеются дополнительные мультифизические возможности, включая расчет взаимодействия потоков с механическими конструкциями. Несмотря на большое количество публикаций по данной проблематике часть вопросов, связанных с изучением механизмов возникновения опасных явлений в водных экосистемах, до сих пор остается открытой, прикладные задачи гидрофизики и биологической кинетики зачастую не рассматриваются в комплексе.

Существующие пакеты и программные комплексы нередко сложно настроить на конкретную водную экосистему ввиду существенного различия в гидрофизических параметрах и видовом составе гидробионтов. Требуемая высокая точность расчетов накладывает дополнительные условия на метод дискретизации исходных нелинейных непрерывных задач. Кроме того, возникает необходимость в обработке и фильтрации входных данных. Большой объем вычислительной работы по реализации рассматриваемых задач в реальных областях сложной формы при ограниченном времени построения прогнозов изменения экологической обстановки в водоеме, например, при возникновении в нем опасного явления, требует разработки эффективных алгоритмов, ориентированных на супервычислительные системы с распределенной памятью.

2. ОСНОВНЫЕ УРАВНЕНИЯ МОДЕЛЕЙ ГИДРОДИНАМИКИ И БИОЛОГИЧЕСКОЙ КИНЕТИКИ

Для описания процессов гидродинамики и биологической кинетики в мелководном водоеме, таком как Азовское море, характеризующимся значительным перепадом глубин, а также построения прогнозов изменения состояния водных экосистем при возникновении в них опасных явлений природного характера, к которым можно отнести штормовые нагоны, явления аноксии и гипоксии, в данной работе предложено использовать математические модели гидродинамики и биологической кинетики.

Гидродинамическая модель движения водной среды предназначена для расчета поля вектора скорости водного потока водоема со сложной батиметрией, к такому типу можно отнести мелководный водоем – Азовское море. Модель учитывает пространственно-трехмерную структуру движения водной среды, стоки рек, сгонно-нагонные явления, сложную геометрию рельефа дна и береговой линии.

Модель биологической кинетики представляет собой систему “фитопланктон–зоопланктон–биогенное вещество–детрит–рыба”, учитывает следующие физические процессы: движение водного потока, микротурбулентную диффузию, температуру, соленость, пространственное распределение кислорода и загрязняющих биогенных веществ, включая соединения азота, фосфора и кремния, их взаимодействие и осаждение, влияние на продукционно-деструкционные процессы водоема; нелинейное взаимодействие гидробионтов, включая фито-, зоопланктон и промысловую рыбу, межвидовую конкуренцию, вылов, таксис, механизм наружно-гормонального регулирования и аллелопатическое воздействие.

2.1. Модель гидродинамики

В основу разрабатываемой модели расчета трехмерных полей вектора скорости движения водной среды, температуры и солености положена математическая модель гидродинамики мелководных водоемов, учитывающая транспорт тепла и солей (см. [9]):

– уравнение движения (Навье–Стокса)

$$u'_t + uu'_x + vu'_y + wu'_z = -\frac{1}{\rho} P'_x + (\mu u'_x)'_x + (\mu u'_y)'_y + (v u'_z)'_z + 2\Omega(v \sin \vartheta - w \cos \vartheta), \quad (1)$$

$$v'_t + uv'_x + vv'_y + wv'_z = -\frac{1}{\rho} P'_y + (\mu v'_x)'_x + (\mu v'_y)'_y + (v v'_z)'_z + 2\Omega u \sin \vartheta, \quad (2)$$

$$w'_t + uw'_x + vw'_y + ww'_z = -\frac{1}{\rho} P'_z + (\mu w'_x)'_x + (\mu w'_y)'_y + (v w'_z)'_z + 2\Omega u \cos \vartheta + g, \quad (3)$$

– уравнение неразрывности в случае переменной плотности

$$\rho'_t + (\rho u)'_x + (\rho v)'_y + (\rho w)'_z = 0, \quad (4)$$

– уравнение транспорта тепла

$$T'_t + uT'_x + vT'_y + wT'_z = (\mu T'_x)'_x + (\mu T'_y)'_y + (v T'_z)'_z + f_T, \quad (5)$$

– уравнение транспорта солей

$$S'_t + uS'_x + vS'_y + wS'_z = (\mu S'_x)'_x + (\mu S'_y)'_y + (v S'_z)'_z + f_S, \quad (6)$$

где $V = \{u, v, w\}$ – компоненты вектора скорости, P – полное гидродинамическое давление, S и T – соленость и температура водной среды, ρ – плотность водной среды, μ, ν – горизонтальная и вертикальная составляющие коэффициента турбулентного обмена, $\Omega = \Omega \times (\cos \vartheta \times \mathbf{j} + \sin \vartheta \times \mathbf{k})$ – угловая скорость вращения Земли, ϑ – широта места, g – ускорение свободного падения, f_T, f_S – источники тепла и соли (находятся на границе области).

Для мелководных водоемов из полного гидродинамического давления условно можно выделить две составляющие: функцию давления столба жидкости и гидродинамическую часть:

$$P(x, y, z, t) = p(x, y, z, t) + \rho_0 g z,$$

где p – гидростатическое давление невозмущенной жидкости, ρ_0 – плотность пресной воды при нормальных условиях.

Уравнение состояния для плотности

$$\rho = \tilde{\rho} + \rho_0,$$

где ρ_0 – плотность пресной воды при нормальных условиях, $\tilde{\rho}$ определяется уравнением, рекомендованным UNESCO:

$$\tilde{\rho} = \tilde{\rho}_w + (8.24493 \times 10^{-1} - 4.0899 \times 10^{-3} T + 7.6438 \times 10^{-5} T^2 - 8.2467 \times 10^{-7} T^3 + 5.3875 \times 10^{-9} T^4) S + (-5.72466 \times 10^{-3} + 1.0227 \times 10^{-4} T - 1.6546 \times 10^{-6} T^2) S^{3/2} + 4.8314 \times 10^{-4} S^2, \quad (7)$$

где $\tilde{\rho}_w$ – плотность пресной воды, задаваемая полиномом

$$\tilde{\rho}_w = 999.842594 + 6.793952 \times 10^{-2} T - 9.095290 \times 10^{-3} T^2 + 1.001685 \times 10^{-4} T^3 - 1.120083 \times 10^{-6} T^4 + 6.536332 \times 10^{-9} T^5.$$

Уравнение (7) применимо для солености в диапазоне 0–42‰ и температуры от –2 до 40°C.

Система уравнений (1)–(7) рассматривается при следующих краевых условиях:

– на входе (устья рек Дон и Кубань)

$$\mathbf{V} = \mathbf{V}_1, \quad P'_n = 0, \quad T = T_1, \quad S = S_1,$$

– донная граница

$$\rho_v \mu (\mathbf{V}_\tau)'_n = -\tau, \quad \mathbf{V}_n = 0, \quad P'_n = 0, \quad T'_n = 0, \quad S'_n = 0, \quad f_T = 0, \quad f_S = 0,$$

– боковая граница

$$(\mathbf{V}_\tau)'_n = 0, \quad \mathbf{V}'_n = 0, \quad P'_n = 0, \quad T'_n = 0, \quad S'_n = 0, \quad f_T = 0, \quad f_S = 0,$$

– верхняя граница

$$\rho_v \mu (\mathbf{V}_\tau)'_n = -\tau, \quad w = -\omega - P'_t / \rho g, \quad P'_n = 0, \quad T'_n = 0, \quad S'_n = 0, \\ f_T = k(T_a - T), \quad f_S = \frac{\omega}{h_z - h_\omega} S,$$

– на выходе (Керченский пролив)

$$P'_n = 0, \quad \mathbf{V}'_n = 0, \quad T'_n = 0, \quad S'_n = 0, \quad f_T = 0,$$

где ω – интенсивность испарения жидкости, $\mathbf{V}_n, \mathbf{V}_\tau$ – нормальная и тангенциальная составляющие вектора скорости, $\tau = \{\tau_x, \tau_y, \tau_z\}$ – вектор тангенциального напряжения, ρ – плотность водной среды, ρ_v – плотность взвеси, T_a – температура атмосферы, k – коэффициент передачи тепла между атмосферой и водной средой, h_z – шаг расчетной сетки по глубине, $h_\omega = \omega \tau$ – слой жидкости, который испаряется за время τ .

Вектор тангенциального напряжения для свободной поверхности рассчитывается по формуле

$$\tau = \rho_a C d_s |W| W,$$

где W – вектор скорости ветра относительно воды, ρ_a – плотность атмосферы, $C d_s$ – безразмерный коэффициент поверхностного сопротивления, который зависит от скорости ветра и находится в диапазоне 0.0016–0.0032 (см. [5]).

Вектор тангенциального напряжения для дна имеет вид

$$\tau = \rho C d_b |V| V,$$

где $C d_b = g k^2 / h^{1/3}$, k – групповой коэффициент шероховатости в формуле Мэннинга, рассматривается в диапазоне 0.025–0.2, в процессе моделирования использовалось значение 0.025, обусловленное преимущественным покрытием дна Азовского моря илистыми отложениями, $h = H + \eta$ – глубина акватории, H – глубина до невозмущенной поверхности, η – высота свободной поверхности относительно геоида (уровень моря).

Система уравнений (1)–(7) рассматривается при начальных условиях

$$\mathbf{V} = \mathbf{V}_0, \quad T = T_0, \quad S = S_0.$$

Турбулентная вязкость ν определяется средним значением скорости диссипации энергии турбулентности ε , приходящейся на единицу объема: $\nu \sim \varepsilon^{1/3} \Delta^{4/3}$, где Δ – характерный масштаб сетки (для расчета коэффициента турбулентного обмена по вертикали $\Delta \equiv h_z$). Скорость диссипации $\bar{\varepsilon}$ может быть выражена через среднюю скорость деформации ячеечного масштаба: $\bar{\varepsilon} = 2\bar{s}_{ij}\bar{s}_{ij}$, где \bar{s}_{ij} – осредненный тензор скоростей деформации, \bar{s} : $\varepsilon^{2/3} \Delta^{-4/3}$. Выражение для турбулентной вязкости имеет вид

$$\nu = C_s^2 \Delta^2 \bar{s}^{-1/2}, \quad (8)$$

где C_s – безразмерная эмпирическая константа, значение которой определяется на основе расчета процесса затухания однородной изотропной турбулентности. Выбранное значение постоянной C_s должно обеспечивать соответствие с экспериментальными измерениями.

2.2. Модель биологической кинетики

Многовидовая математическая модель взаимодействия промысловой рыбы-детритофага, фито- и зоопланктона основана на системе нестационарных уравнений конвекции–диффузии–реакции параболического типа с нелинейными функциями источников и младшими производными, вид которых для каждого модельного блока S_i определяется следующим образом:

$$\begin{aligned} \frac{\partial S_i}{\partial t} + \operatorname{div}(\mathbf{v} S_i) &= \mu_i \Delta S_i + \frac{\partial}{\partial z} \left(\nu_i \frac{\partial S_i}{\partial z} \right) + \psi_i, \quad \psi_1 = g_1(S_1, S_3) - \delta_1 S_1 S_2 - \lambda_1 S_1 - \sigma_1 S_1 S_5, \\ \psi_2 &= g_2(S_1, S_2) - \lambda_2 S_2 - \delta_2 S_2, \quad \psi_3 = \gamma_3 \lambda_4 S_4 - g_3(S_1, S_3) + B(\tilde{S}_3 - S_3) + f, \\ \psi_4 &= \lambda_1 S_1 + \lambda_2 S_2 - \lambda_4 S_4 - g_4(S_4, S_5), \quad \psi_5 = g_5(S_1, S_4, S_5) - \lambda_5 S_5 - \delta_5 S_5, \end{aligned} \quad (9)$$

где S_i – концентрация i -й компоненты, $i = \overline{1, 5}$; ψ_i – химико-биологический источник (сток) или член, описывающий агрегирование (слипание-разлипание), если соответствующая компонента является взвесью, i – вид субстанции, $i = \overline{1, 5}$: 1 – концентрации фитопланктона (*Coscinodiscus*) (P), 2 – зоопланктона (*Copepoda*) (Z), 3 – биогенного вещества (M), 4 – детрита (D), 5 – промысловой рыбы-детритофага (F); $\mathbf{v} = \mathbf{V} + \mathbf{V}_{0i}$ – скорость конвективного переноса вещества; \mathbf{V}_{0i} – скорость осаждения i -й субстанции под действием силы тяжести, $i \in \overline{1, 4}$, g_i – трофические функции для субстанций $i \in \overline{1, 5}$, \tilde{S}_3 – предельно возможная концентрация биогенного вещества; $f = f(t, x, y, z)$ – функция источника загрязнения; B – удельная скорость поступления загрязняющего вещества; λ_2, λ_5 – коэффициенты элиминации (смертности) Z, F соответственно. Положим $g_1(S_1, S_3) = \gamma_1 \alpha_3 \phi_1 S_1 S_3$, $g_2(S_1, S_2) = \gamma_2 \phi_2 \delta_1 S_1 S_2$, $g_3(S_1, S_3) = \alpha_3 S_1 S_3$, $g_4(S_4, S_5) = \beta_4 S_4 S_5$, $g_5(S_1, S_4, S_5) = (\gamma_5 \beta_4 S_4 + \xi_5 \sigma_1 S_1) \phi_5 S_5$; α_3 – коэффициент потребления биогенного вещества фитопланктоном; $\gamma_1, \gamma_2, \gamma_5$ – передаточные коэффициенты трофических функций; γ_3 – доля питатель-

ного вещества, находящегося в биомассе фитопланктона; λ_1 – коэффициент, учитывающий смертность и метаболизм P ; δ_1 – убыль фитопланктона за счет выедания зоопланктоном; δ_2 – убыль зоопланктона за счет выедания рыбами; δ_5 – убыль рыбы за счет вылова; λ_4 – коэффициент разложения детрита; β_4 – скорость потребления органических остатков рыбой; σ_1 – коэффициент убыли фитопланктона в результате потребления его рыбой; ξ_5 – передаточный коэффициент роста концентрации рыбы за счет фитопланктона; μ_i, ν_i – диффузионные коэффициенты в горизонтальном и вертикальном направлениях i -й субстанции, $i \in \overline{1,5}$; Δ – двумерный оператор Лапласа.

При параметризации системы (9) использовались модели наблюдений, представляющие собой функциональные зависимости, параметризованные на основе стехиометрических соотношений, законов Моно, Михаэлиса–Ментен и Митчерлиха–Бауле, описывающие потребление, накопление планктоном и промысловыми рыбами-детритофагами питательных веществ, а также рост гидробионтов в зависимости от пространственного распределения солености и температуры, кислородного режима (см. [5]–[7]).

Функции зависимости скорости роста гидробионтов от температуры и солености задавались следующим образом:

$$\phi_m = \exp \left[-\alpha_m \left(\frac{T - T_{opt}}{T_{opt}} \right)^2 \right] \exp \left[-\beta_m \left(\frac{S - S_{opt}}{S_{opt}} \right)^2 \right], \quad m \in \{1, 2, 5\},$$

где T_{opt}, S_{opt} – температура и соленость, оптимальные для данного вида планктона и рыб; $\alpha_m > 0, \beta_m > 0$ – коэффициенты ширины интервала толерантности планктона к температуре и солености соответственно.

Расчетная область \bar{G} (Азовское море) представляет собой замкнутую область, ограниченную невозмущенной поверхностью водоема Σ_0 , дном $\Sigma_H = \Sigma_H(x, y)$ и боковой цилиндрической поверхностью σ для временного промежутка $0 < t \leq T_0$; Σ – кусочно-гладкая граница области G , $\Sigma = \Sigma_0 \cup \Sigma_H \cup \sigma$; \mathbf{n} – вектор внешней нормали к поверхности; \mathbf{v}_n – нормальная по отношению к Σ составляющая вектора скорости водного потока.

Начальные условия

$$S_i|_{t=0} = S_{i0}(x, y, z), \quad i = \overline{1,5}. \tag{10}$$

Граничные условия

$$S_i = 0 \text{ на } \sigma, \text{ если } \mathbf{v}_n < 0; \quad \frac{\partial S_i}{\partial \mathbf{n}} = 0 \text{ на } \sigma, \text{ если } \mathbf{v}_n \geq 0, \quad i = \overline{1,5};$$

$$\frac{\partial S_i}{\partial z} = -\varepsilon_i S_i \text{ на } \Sigma_H; \quad \frac{\partial S_i}{\partial z} = \varphi(S_i) \text{ на } \Sigma_0; \quad \varphi(S_i) = \begin{cases} k_i S_i, & i = 3; \\ 0, & i \neq 3, \end{cases} \tag{11}$$

где $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ – неотрицательные постоянные; $\varepsilon_1, \varepsilon_3, \varepsilon_5$ – учитывают опускание планктона и рыб на дно и их затопление; $\varepsilon_2, \varepsilon_4$ – учитывают поглощение биогенного вещества и детрита донными отложениями; k_3 – удельная скорость поступления загрязняющего биогенного вещества из воздушной среды.

Будем учитывать влияние кормового таксиса на функционирование системы “фитопланктон–зоопланктон–биогенное вещество–детрит–рыба”, для этого к системе (9)–(11) добавим следующие уравнения:

$$\frac{\partial \mathbf{v}_5}{\partial t} + \text{div}(\mathbf{V}_5 \otimes \mathbf{v}_5) = \mu_u \Delta \mathbf{v}_5 + \frac{\partial}{\partial z} \left(\nu_u \frac{\partial \mathbf{v}_5}{\partial z} \right) - \alpha_u \mathbf{v}_5 + k_1 \text{grad } S_1 + k_4 \text{grad } S_4, \tag{12}$$

где $\mathbf{v}_5 = \mathbf{V} + \mathbf{V}_5$ – скорость конвективного переноса рыбы; \mathbf{V}_5 – скорость движения рыбы относительно воды; k_1, k_4 – коэффициенты таксиса; μ_u, ν_u – коэффициенты горизонтальной и вертикальной составляющей диффузии скорости таксиса; α_u – коэффициент инерционного движения рыбы; \otimes – тензорное произведение векторов.

Поскольку рассматриваемый вид рыбы является детритофагом, то им перерабатывается значительное количество отложений органических илов – отмерших остатков планктона, рыб, бентосных животных, разложение которых в летний период при достаточном прогреве водной толщи вызывает нехватку кислорода в воде и приводит к заморным явлениям в экосистеме. Для учета распределения концентрации кислорода в водоеме к системе (9)–(12) добавим следующее уравнение:

$$C = C_0 - mS_4, \quad C|_{t=0} = \theta(x, y, z) \quad \forall (x, y, z) \in \bar{G}, \quad (13)$$

где C_0 – концентрация кислорода при отсутствии органических примесей; m – количество кислорода, необходимое для окисления детрита в толще водоема; θ – заданная функция.

3. МЕТОД РЕШЕНИЯ ЗАДАЧИ ГИДРОДИНАМИКИ

Согласно методу поправки к давлению, исходная модель гидродинамики разбивается на три подзадачи (см. [9]). Первая подзадача представлена уравнением диффузии–конвекции–реакции, с помощью которого вычисляются компоненты поля вектора скорости водного потока на промежуточном временном слое (см. [10]):

$$\frac{\tilde{u} - u}{\tau} + u\bar{u}'_x + v\bar{u}'_y + w\bar{u}'_z = (\mu\bar{u}'_x)'_x + (\mu\bar{u}'_y)'_y + (v\bar{u}'_z)'_z + 2\Omega(v \sin \theta - w \cos \theta), \quad (14)$$

$$\frac{\tilde{v} - v}{\tau} + u\bar{v}'_x + v\bar{v}'_y + w\bar{v}'_z = (\mu\bar{v}'_x)'_x + (\mu\bar{v}'_y)'_y + (v\bar{v}'_z)'_z - 2\Omega u \sin \theta, \quad (15)$$

$$\frac{\tilde{w} - w}{\tau} + u\bar{w}'_x + v\bar{w}'_y + w\bar{w}'_z = (\mu\bar{w}'_x)'_x + (\mu\bar{w}'_y)'_y + (v\bar{w}'_z)'_z + 2\Omega u \cos \theta + g \left(\frac{\rho_0}{\rho} - 1 \right), \quad (16)$$

где u, v, w – значения компонент вектора скорости на предыдущем слое по времени, $\tilde{u}, \tilde{v}, \tilde{w}$ – значения компонент вектора скорости на промежуточном слое по времени, $\bar{u} = \sigma\tilde{u} + (1 - \sigma)u$, $\bar{v} = \sigma\tilde{v} + (1 - \sigma)v$, $\bar{w} = \sigma\tilde{w} + (1 - \sigma)w$, $\sigma \in [0, 1]$ – вес схемы.

Следует отметить, что слагаемое $g(\rho_0/\rho - 1)$ описывает плавучесть (силу Архимеда). Многочисленными экспериментами по моделированию движения среды в мелководных водоемах, подобных Азовскому морю установлено, что данное слагаемое вносит незначительный вклад в решение задачи и им можно пренебречь. Для аппроксимации уравнения диффузии–конвекции–реакции по времени использованы схемы с весами. Здесь $\bar{u} = \sigma\tilde{u} + (1 - \sigma)u$, $\sigma \in [0, 1]$ – вес схемы.

Расчет распределения давлений (вторая подзадача) базируется на уравнении Пуассона

$$p''_{xx} + p''_{yy} + p''_{zz} = \frac{\bar{\rho} - \rho}{\tau^2} + \frac{(\bar{\rho}\tilde{u})'_x}{\tau} + \frac{(\bar{\rho}\tilde{v})'_y}{\tau} + \frac{(\bar{\rho}\tilde{w})'_z}{\tau}. \quad (17)$$

Значение поля скорости водного потока на верхней границе (поверхности водоема) задается следующим образом: $w = -\omega - p'_i/\rho g$. При данном условии уравнение (17) запишется в виде

$$\left(\frac{p'_i}{\tau g h_z} \right)_k = \frac{(p''_{xx} + p''_{yy})_k}{2} + \frac{p_{k+1} - p_k}{h_z^2} - \frac{1}{2} \left(\frac{\bar{\rho} - \rho}{\tau^2} + \frac{(\bar{\rho}\tilde{u}z)'_x}{\tau} + \frac{(\bar{\rho}\tilde{v})'_y}{\tau} \right)_k - \frac{(\bar{\rho}\tilde{w})_{k+1/2} + \bar{\rho}_k \omega}{\tau h_z}, \quad (18)$$

где k – индекс расчетной сетки по вертикальному координатному направлению.

В качестве начального приближения для данной задачи используется упрощенная гидростатическая модель движения водной среды, что значительно уменьшает время расчета.

Третья подзадача позволяет по явным формулам определить распределение скоростей водного потока на следующем временном слое:

$$\frac{\hat{u} - \tilde{u}}{\tau} = -\frac{1}{\rho} p'_x, \quad \frac{\hat{v} - \tilde{v}}{\tau} = -\frac{1}{\rho} p'_y, \quad \frac{\hat{w} - \tilde{w}}{\tau} = -\frac{1}{\rho} p'_z, \quad (19)$$

где $\hat{u}, \hat{v}, \hat{w}$ – значения компонент вектора скорости водного потока на текущем слое по времени.

Для программной реализации трехмерной математической модели гидродинамики вводится равномерная сетка. Обозначим через $\sigma_{i,j,k}$ степень “заполненности” ячейки (i, j, k) . Степень за-

Таблица 1. Погрешность решения задачи течения жидкости между двумя цилиндрами

Размеры сетки	11 × 21	21 × 41	41 × 81	81 × 161
Максимальное значение погрешности в случае гладкой границы, м/с	0.053	0.052	0.058	0.056
Среднее значение погрешности в случае гладкой границы, м/с	0.023	0.012	0.006	0.003
Максимальное значение погрешности в случае ступенчатой границы, м/с	0.272	0.731	0.717	0.75
Среднее значение погрешности в случае ступенчатой границы, м/с	0.165	0.132	0.069	0.056

полненности ячейки задается давлением водного столба на дно данной ячейки. В общем случае степень заполненности ячеек рассчитывается, исходя из выражения (см. [11])

$$o_{i,j,k} = \frac{P_{i,j,k} + P_{i-1,j,k} + P_{i,j-1,k} + P_{i-1,j-1,k}}{4\rho gh_z}. \tag{20}$$

Аппроксимация задачи расчета поля скорости движения водной среды по пространственным переменным выполнена на основе метода баланса с учетом коэффициентов заполненности контрольных областей.

3.1. Разностные схемы для операторов диффузии и конвекции, учитывающие функцию заполненности ячеек

Для построения и тестирования разностных схем операторов диффузии и конвекции, учитывающих функцию заполненности ячеек, вводится равномерная сетка:

$$w_h = \{t^n = n\tau, x_i = ih_x, y_j = jh_y; n = \overline{0, N_t}, i = \overline{0, N_x}, j = \overline{0, N_y}; N_t\tau = T, N_x h_x = l_x, N_y h_y = l_y\},$$

где τ – шаг по времени, h_x, h_y – шаги по пространству, N_t – наибольшее число шагов по времени, N_x, N_y – количество шагов пространственной сетки по направлению Ox и Oy соответственно.

Дискретные аналоги операторов конвективного uc'_x и диффузионного $(\mu c'_x)'_x$ переноса в случае частичной заполненности ячеек с граничными условиями III рода $c'_n(x, z, t) = \alpha_n c + \beta_n$ могут быть записаны в следующем виде (см. [11]):

$$(q_0)_{i,j}uc'_x \approx (q_1)_{i,j}\mu_{i+1/2,j} \frac{c_{i+1,j} - c_{i,j}}{2h_x} + (q_2)_{i,j}\mu_{i-1/2,j} \frac{c_{i,j} - c_{i-1,j}}{2h_x},$$

$$(q_0)_{i,j}(\mu c'_x)'_x \approx (q_1)_{i,j}\mu_{i+1/2,j} \frac{c_{i+1,j} - c_{i,j}}{h_x^2} - (q_2)_{i,j}\mu_{i-1/2,j} \frac{c_{i,j} - c_{i-1,j}}{h_x^2} - |(q_1)_{i,j} - (q_2)_{i,j}|\mu_{i,j} \frac{\alpha_x c_{i,j} + \beta_x}{h_x},$$

где q_0, q_1, q_2 – коэффициенты заполненности контрольных областей.

Тестирование разностных схем, учитывающих функцию заполненности ячеек, выполнено на основе численного решения задачи расчета течения вязкой жидкости между двумя соосными полуцилиндрами (течения Куэтга–Тейлора). Для данной задачи исследованы погрешности численного решения на сетках, учитывающих “заполненность” ячеек (случай гладкой границы), и на сетках со ступенчатой аппроксимацией границы. При использовании сеток, учитывающих “заполненность” ячеек, погрешность численного решения модельной задачи гидродинамики не превосходят 6%, а в случае ступенчатой аппроксимацией границы может достигать 70% от решения задачи. В табл. 1 представлены значения погрешности численного решения задачи течения жидкости между двумя соосными полуцилиндрами, полученные на последовательности сгущающихся расчетных сеток размерами 11 × 21, 21 × 41, 41 × 81 и 81 × 161 узлов в случаях гладкой и ступенчатой аппроксимации границы.

Анализ результатов расчетов погрешности численного решения задачи течения жидкости между двумя цилиндрами на последовательности сгущающихся сеток, представленных в табл. 1, позволяет сделать вывод об эффективности использования разностных схем, учитывающих заполненность ячеек. При решении модельной задачи дробление сетки в 8 раз по каждому из пространственных направлений не приводит к повышению точности, которой обладают решения, полученные на сетках, учитывающих заполненность ячеек.

3.2. Метод решения задачи восстановления донной поверхности

Для получения функции рельефа дна используем решение уравнения диффузии, к которому сводится уравнение Сен-Венана, описывающее транспорт донных материалов (см. [12]). Решение задачи диффузии на длительные интервалы времени сводится к решению уравнения Лапласа

$$\Delta H = 0, \quad (21)$$

где H – глубина водоема.

Данный подход обладает существенным недостатком, связанным с отсутствием гладкости в точках, где задаются значения поля глубин. Для устранения данной проблемы можно использовать решения следующего уравнения:

$$\Delta^2 H = 0. \quad (22)$$

К недостаткам этого подхода относят большие выбросы (отклонение от линейной функции). С помощью первых двух подходов можно получать функции, которые не обладают выделенностью направлений, но каждый из подходов обладает недостатками.

Для получения гладкой функции рельефа дна также можно использовать решение уравнения, применяемого для получения схем повышенного порядка точности для уравнения Лапласа:

$$\Delta H - \frac{h^2}{12} \Delta^2 H = 0. \quad (23)$$

Заметим, что оператор для третьей задачи можно записать как линейную комбинацию операторов для первой и второй задачи.

Фундаментальной системой решений уравнения (21) являются функции

$$H_1(x) = 1, \quad H_2(x) = x; \quad (24)$$

уравнения (22) –

$$H_1(x) = 1, \quad H_2(x) = x, \quad H_3(x) = x^2, \quad H_4(x) = x^3; \quad (25)$$

уравнения (23) –

$$H_1(x) = 1, \quad H_2(x) = x, \quad H_3(x) = ch(kx), \quad H_4(x) = sh(kx), \quad k = \sqrt{12}/h. \quad (26)$$

В первом случае интерполяция осуществляется отрезками прямых, проходящих через соседние точки. Во втором случае интерполяция выполняется на основе кубических сплайнов. В третьем случае – на основе сплайнов функции (26). Ниже описан алгоритм, предназначенный для одномерной интерполяции на основе функции (26), и проведено сравнение предложенных подходов. Результат работы предложенного алгоритма приведен на фиг. 1.

Из фиг. 1 видно, что интерполяция, полученная на основе предложенного алгоритма, имеет меньшее отклонение от линейной интерполяции по сравнению с кубическим сплайном и обладает достаточной степенью гладкости в точках склейки функций.

3.3. Модифицированный попеременно-треугольный итерационный метод

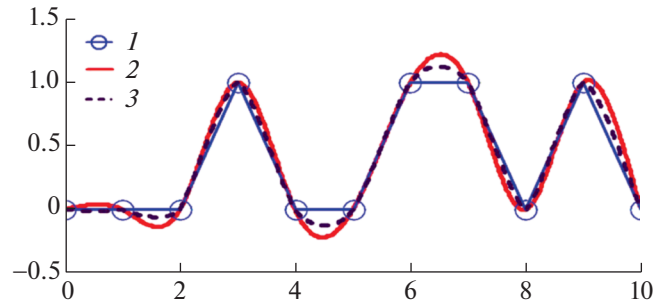
После получения разностных схем исходные задачи сводятся к решению систем сеточных уравнений, которые могут быть записаны в матричном виде. В конечномерном гильбертовом пространстве H рассматривается задача отыскания решения операторного уравнения (см. [13]–[15])

$$Ax = f, \quad A : H \rightarrow H, \quad (27)$$

где A – линейный, положительно-определенный оператор ($A > 0$). Для нахождения решения задачи (27) будем использовать неявный итерационный процесс

$$B \frac{x^{m+1} - x^m}{\tau} + Ax^m = f, \quad B : H \rightarrow H, \quad (28)$$

где m – номер итерации, τ – итерационный параметр ($\tau > 0$), а B – некоторый обратимый оператор. Обращение оператора B в (28) должно быть существенно проще, чем непосредственное об-



Фиг. 1. Результат интерполяции: 1 – линейная интерполяция, 2 – кубический сплайн, 3 – на основе предложенного алгоритма.

ращение исходного оператора A в (27). При построении B будем исходить из аддитивного представления оператора A_0 – симметричной части оператора A :

$$A_0 = R_1 + R_2, \quad R_1 = R_2^*, \tag{29}$$

где R_1, R_2 – ниже- и верхнетреугольные операторы.

Также здесь и далее будем использовать кососимметричную часть оператора A :

$$A_1 = \frac{A - A^*}{2}.$$

В силу (29) $(Ay, y) = (A_0y, y) = 2(R_1y, y) = 2(R_2y, y)$. Поэтому в (29) $R_1 > 0, R_2 > 0$. Пусть в (28)

$$B = (D + \omega R_1)D^{-1}(D + \omega R_2), \quad D = D^* > 0, \quad \omega > 0, \quad y \in H, \tag{30}$$

где D – некоторый оператор (например, диагональная часть оператора A).

Поскольку $A_0 = A_0^* > 0$, то вместе с (29) это дает $B = B^* > 0$. Соотношения (28)–(30) задают модифицированный попеременно-треугольный метод (МПТМ) решения задачи, если определены операторы R_1, R_2 и указаны способы определения параметров τ, ω и оператора D .

Алгоритм нахождения решений системы сеточных уравнений МПТМ вариационного типа запишется в виде

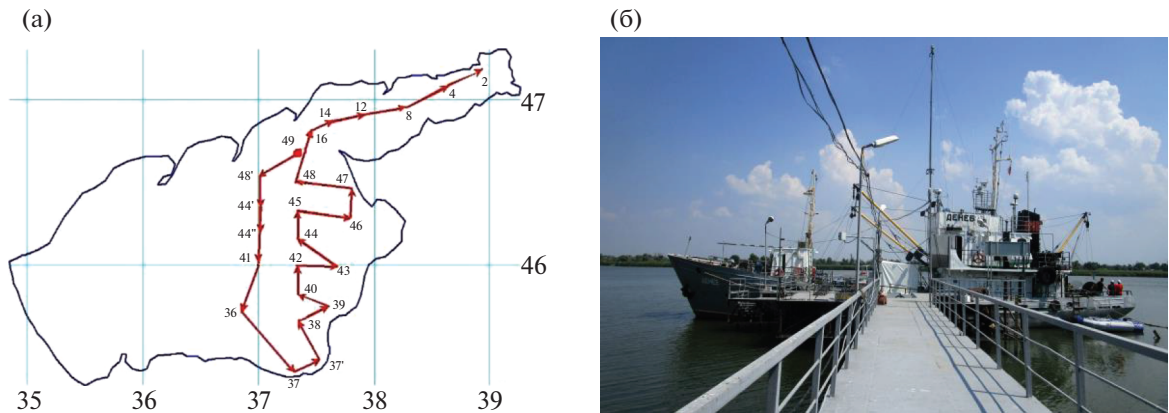
$$\begin{aligned} r^m &= Ax^m - f, \quad B(\omega_m)w^m = r^m, \quad \tilde{\omega}_m = \sqrt{\frac{(Dw^m, w^m)}{(D^{-1}R_2w^m, R_2w^m)}}, \\ s_m^2 &= 1 - \frac{(A_0w^m, w^m)^2}{(B^{-1}A_0w^m, A_0w^m)(Bw^m, w^m)}, \quad k_m^2 = \frac{(B^{-1}A_1w^m, A_1w^m)}{(B^{-1}A_0w^m, A_0w^m)}, \\ \theta_m &= \frac{1 - \sqrt{\frac{s_m^2 k_m^2}{(1 + k_m^2)}}}{1 + k_m^2(1 - s_m^2)}, \quad \tau_{m+1} = \theta_m \frac{(A_0w^m, w^m)}{(B^{-1}A_0w^m, A_0w^m)}, \quad x^{m+1} = x^m - \tau_{m+1}w^m, \quad \omega_{m+1} = \tilde{\omega}_m, \end{aligned} \tag{31}$$

где r^m – вектор невязки, w^m – вектор поправки, s_m – скорость сходимости метода, k_m – отношение нормы кососимметричной части оператора к норме симметричной части.

Скорость сходимости метода равна

$$\rho \leq \frac{v^* - 1}{v^* + 1}, \tag{32}$$

где $v^* = v(\sqrt{1 + k^2} + k)^2$, где v – число обусловленности матрицы $C_0, C_0 = B^{-1/2}A_0B^{-1/2}$.



Фиг. 2. Экспедиция в Азовском море: (а) – маршрут экспедиции, (б) – научно-исследовательское судно “Денеб”.

Значение ω оптимально, если

$$\omega = \sqrt{\frac{(Dw^m, w^m)}{(D^{-1}R_2w^m, R_2w^m)}}$$

и имеет место оценка числа обусловленности матрицы C_0

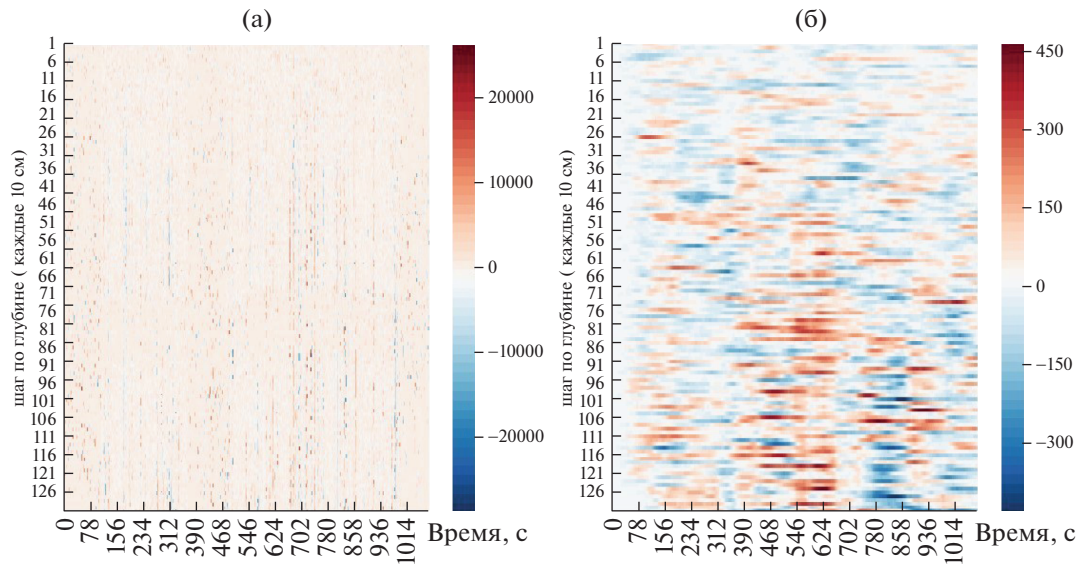
$$\nu = \max_{y \neq 0} \left(\frac{1}{2} + \frac{\sqrt{(Dy, y)(D^{-1}R_2y, R_2y)}}{(A_0y, y)} \right) \leq \frac{1}{2} \left(1 + \sqrt{\frac{\Delta}{\delta}} \right) = \frac{1 + \sqrt{\xi}}{2\sqrt{\xi}},$$

где $\xi = \frac{\delta}{\Delta}$, $D \leq \frac{1}{\delta} A_0$, $R_1 D^{-1} R_2 \leq \frac{\Delta}{4} A_0$.

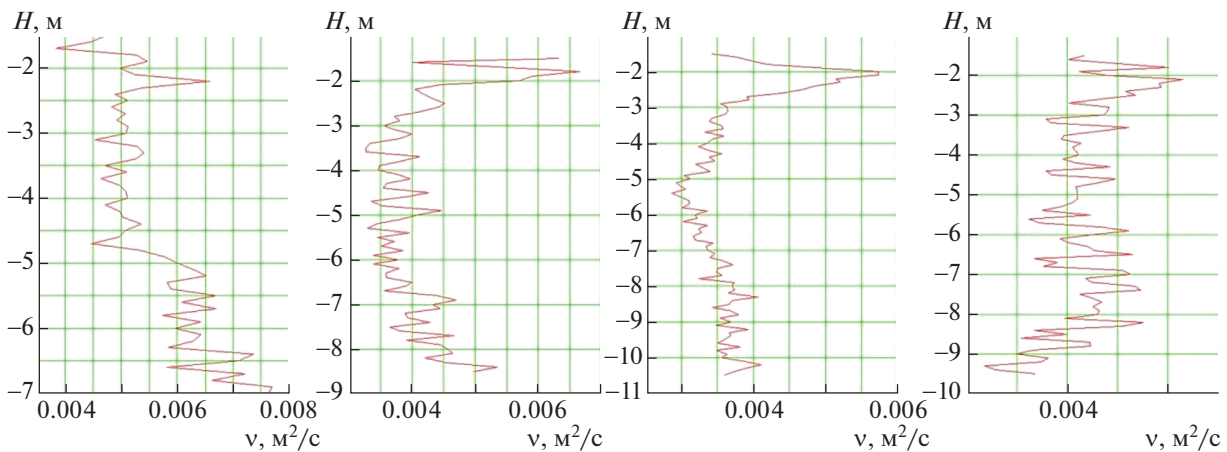
4. ЭКСПЕДИЦИОННЫЕ ИССЛЕДОВАНИЯ АЗОВСКОГО МОРЯ И ОБРАБОТКА НАТУРНЫХ ДАННЫХ

Исследования проводились на 17 станциях в центрально-восточной части Азовского моря на научно-исследовательском судне “Денеб” Южного научного центра РАН (фиг. 2). В процессе реализации расчета структуры течений в водоемах с различной батиметрией самая часто возникающая проблема – это параметризация процесса турбулентного обмена. Поскольку именно турбулентная структура течения водного потока определяет интенсивность и характер протекания таких процессов, как транспорт наносов и перенос различных загрязняющих примесей, размыв дна.

В ходе экспедиции были получены данные о пульсациях скоростей водного потока в некоторых точках мелководных систем на основе зонда ADCP (Acoustic Doppler Current Profiler) WNS600 Sentinel. Профилограф работает на доплеровском эффекте, передавая акустический сигнал на фиксированной частоте и принимая отраженный на неоднородностях водной среды сигнал в расположенной под излучателем (начиная с зоны чувствительности) толще водного столба. Изучено влияние изменения значений коэффициента вертикального турбулентного обмена на содержание растворенного кислорода в придонном слое мелководного водоема. Примерно на глубине 3 м и ниже значения коэффициента вертикального турбулентного обмена близки к нулю, что означает пониженный турбулентный обмен по вертикали в данной области и объясняет явление гипоксии в придонном слое центрально-восточной части Азовского моря в летний период. В качестве входных данных при решении задачи фильтрации поступают исходные данные о скоростях потоков течения водной среды. Исходные данные для решения поставленной задачи следующие: количество элементов разрешения по глубине составляло 128, шаг измерения по глубине был взят 10 см, период измерений 1 с, количество измерений на каждой станции менялось в пределах от 200 до 1000. Для решения задачи построения модели фильтрации за основу был взят двухэтапный алгоритм Калмана. На фиг. 3 представлен тестовый пример работы программного обеспечения, предназначенного для устранения зашумленности экспедиционных измерений одной из составляющих вектора скорости водного потока.



Фиг. 3. Применение фильтра Калмана для устранения зашумленности данных в двумерном случае: (а) — исходные данные с погрешностью измерительного прибора, (б) — данные, полученные после применения алгоритма фильтрации.



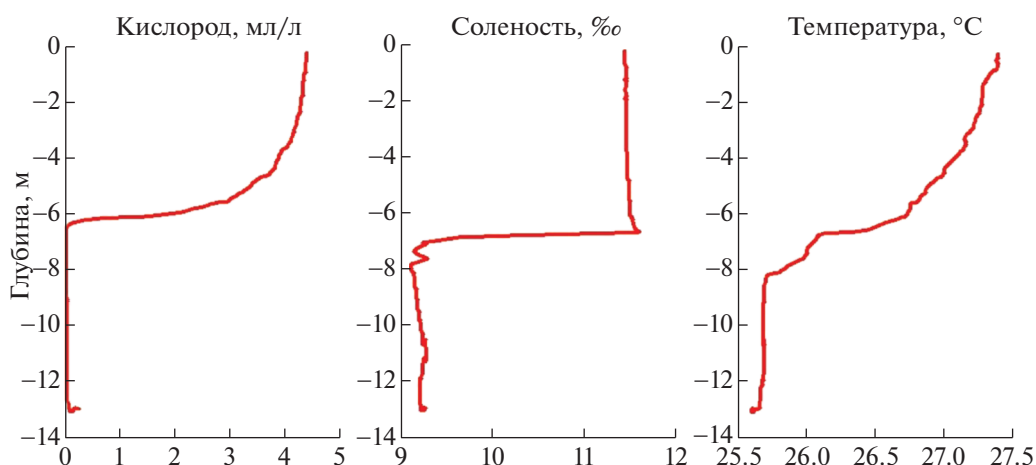
Фиг. 4. Коэффициент вертикального турбулентного обмена в точках 1, 2, 3, 12.

По вертикальной оси отмечена глубина водоема, по горизонтали — время. Цветом выделена скорость водного потока в мм/с в соответствии с приведенной на фиг. 3 шкалой.

Турбулентность напрямую рассчитывается непосредственно из уравнений Навье–Стокса и находится естественным путем при численном моделировании, если вертикальное разрешение сетки позволяет воспроизвести все механизмы до масштабов вязкой диссипации очень мелких вихрей.

Выполненные численные эксперименты на основе нескольких подходов для всех точек позволили получить распределение коэффициента вертикального турбулентного обмена в Азовском море (фиг. 4).

Основываясь на результатах экспедиционных измерений параметров водной среды, проведенных авторами в акватории Азовского моря, можно утверждать, что в его центрально-восточной части существует область, в которой происходят заморные явления биоты. Причем появление этой области носит сезонный характер.



Фиг. 5. Профили растворенного кислорода, солёности и температуры.



Фиг. 6. Восстановленное поле солёности Азовского моря.

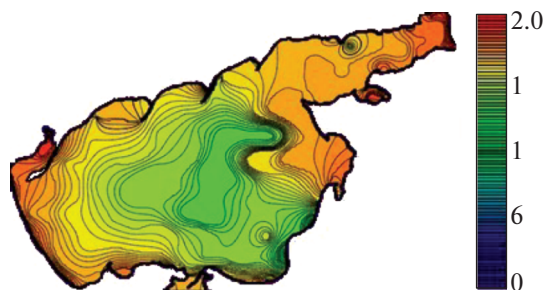
На фиг. 5 представлены результаты натуральных измерений профиля температуры, солёности и растворенного кислорода, полученные с помощью гидрофизического зонда “Sea Bird Electronics 19 Plus” на одной из экспедиционных станций.

На фиг. 5 продемонстрирован тот факт, что содержание кислорода в контрольной точке Азовского моря на глубине более 6.5 м близко к нулевой отметке. На данной станции можно было наблюдать явление аноксии. Явление аноксии периодически, преимущественно в летний период, (не каждый год) возникает в Азовском море, но сопровождается катастрофическими последствиями, связанными с массовой гибелью рыбы на обширной территории. Известно, что механизмы возникновения аноксии для мелководных и глубоководных морей различны.

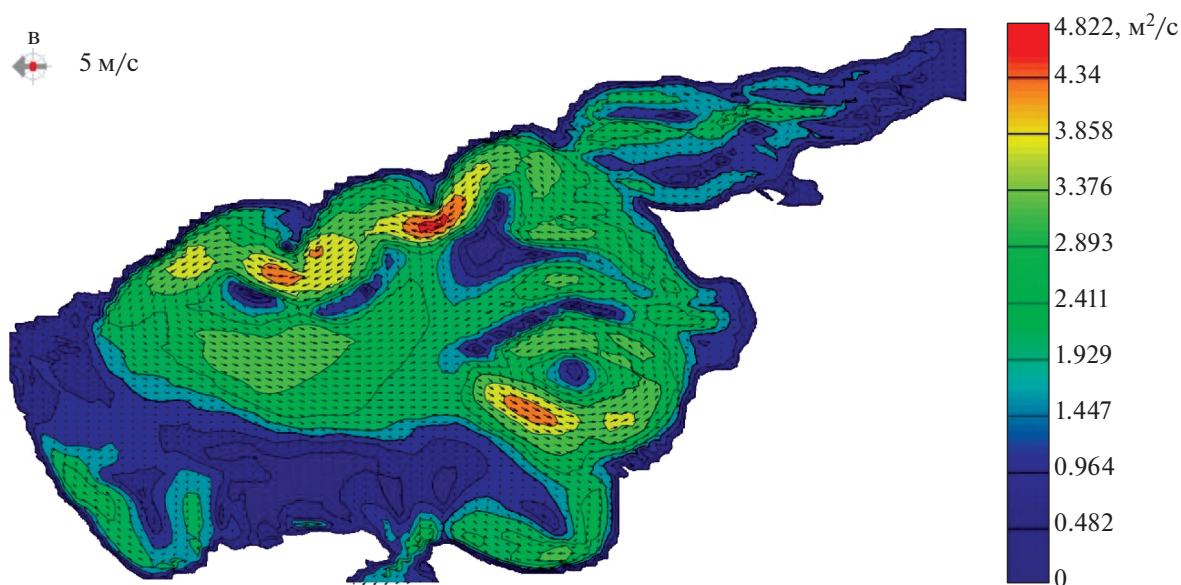
5. ОПИСАНИЕ ПРОГРАММНОГО КОМПЛЕКСА И РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ ПО РЕКОНСТРУКЦИИ ЭКОЛОГИЧЕСКИХ КАТАСТРОФ

При решении задачи обработки гидрологической информации получены изолинии солёности и температуры в поверхностном слое, для чего применен алгоритм распознавания. С помощью алгоритма интерполяции получены карты солёности и температуры Азовского моря в летний период (фиг. 6, 7) на основе имеющихся гидрологических данных (изолинии уровня, значения полей в отдельных точках). Данная информация является входной для моделей гидродинамики.

Программная реализация математических моделей учитывает силу Кориолиса, ветровые течения и трение о дно, турбулентный обмен, испарение, стоки рек, а также сложную геометрию дна и береговой линии. Расчетная область соответствует физическим размерам Азовского моря: длина 355 км, ширина 233 км, максимальная глубина 13.2 м, шаг по пространству в горизонтальном направлении 1000 и 0.5 м по вертикали, шаг по времени 6 мин.



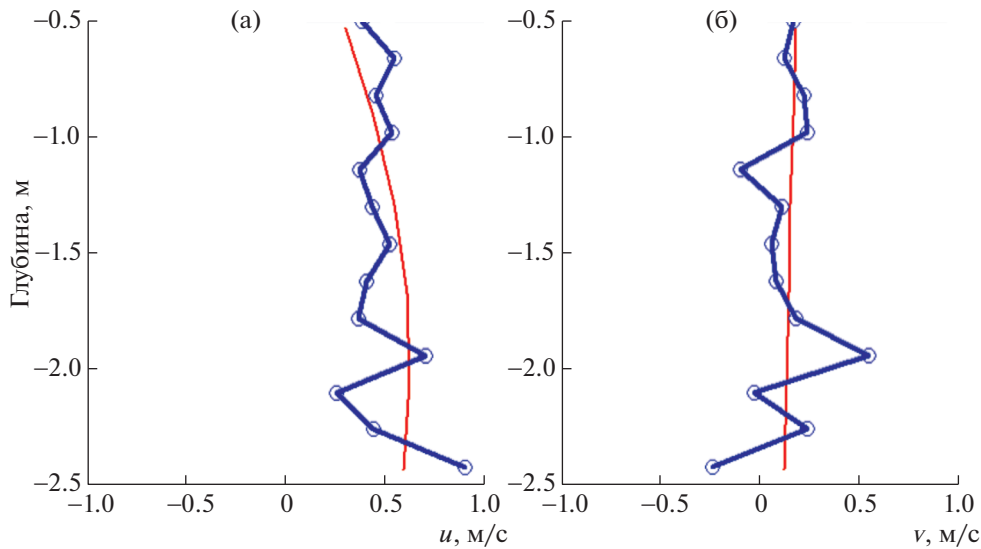
Фиг. 7. Восстановленное поле температуры Азовского моря.



Фиг. 8. Результаты математического моделирования движения водной среды.

Для реконструкции экологической катастрофы была численно реализована трехмерная математическая модель вида (1)–(8), описывающая гидрофизические процессы, происходящие в мелководных водоемах на примере Азовского моря. На фиг. 8 приведены результаты численного моделирования движения водной среды (водные потоки) в акватории Азовского моря на основе программного комплекса “Azov3d”. Рассмотрена задача на установление течений в Азовском море при восточном ветре интенсивностью 5 м/с. Изначально жидкость находилась в состоянии покоя. Расчетный интервал составлял 24 ч. На мелководье водные потоки преимущественно направлены вдоль ветра. В глубоководных участках Азовского моря, вследствие разности уровней водной среды, в толще водоема поток направлен против ветра. Для изучения полей течений в Азовском море удобно использовать интегральные характеристики: $U = \int_{\eta}^H u dz$, $V = \int_{\eta}^H v dz$. На фиг. 8 приведен пример расчета полей течений водного потока Азовского моря. Шкала справа отражает интегральную характеристику водного потока (U, V).

В восточной части Азовского моря можно наблюдать область с вихревой структурой течений (природной ловушкой), в которой может находиться достаточно большое количество отмершей органики, ранее вовлеченной в движение водотока. Предположительно, данные явления привели к экологической катастрофе в июле 2001 г. в Азовском море, когда на площади более 1000 кв. км возникла зона анаэробного (сероводородного) заражения и наблюдалась массовая гибель ихтиофауны.



Фиг. 9. Профили горизонтальной компоненты вектора скорости: (а) – составляющая вектора скорости, направленная с запада на восток (u), (б) – с севера на юг (v).

Проведенные комплексные экспедиционные измерения параметров водной среды в акватории Азовского моря и Таганрогского залива позволили обновить базы данных многолетних наблюдений за состоянием их водной среды. На фиг. 9 приведены профили горизонтальных составляющих вектора скорости водного потока (красная тонкая линия – результаты численных экспериментов, толстая синяя линия – натуральных). Относительное отклонение расчетных гидрофизических параметров, полученных в результате моделирования и натуральных опытов, находится в пределах от 15 до 20%.

Дальнейшее повышение точности модели осложнено тем, что результаты натуральных экспериментов при одних и тех же условиях по некоторым параметрам могут значительным образом отличаться от среднего значения.

Натурные исследования позволили провести калибровку и верификацию разработанной математической модели гидродинамики.

6. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ ПО МОДЕЛИРОВАНИЮ ПРОЦЕССОВ БИОЛОГИЧЕСКОЙ КИНЕТИКИ

Для решения поставленной задач гидродинамики (1)–(8) и биологической кинетики вида (9)–(11) были разработаны модули, входящие в программный комплекс (ПК) “Azov3D”, реализованные на высокопроизводительных вычислительных системах: МВС и графическом ускорителе (см. [16]–[18]). ПК включает следующие блоки: блок управления, океанологические и метеорологические базы данных, библиотека прикладных программ для решения задач гидробиологии; осуществляется интеграция с различными геоинформационными системами (ГИС), глобальной базой данных с ресурсами для геотаггинга и доступа к системам сбора спутниковых данных, базой данных реанализа NCEP/NCAR (фиг. 10).

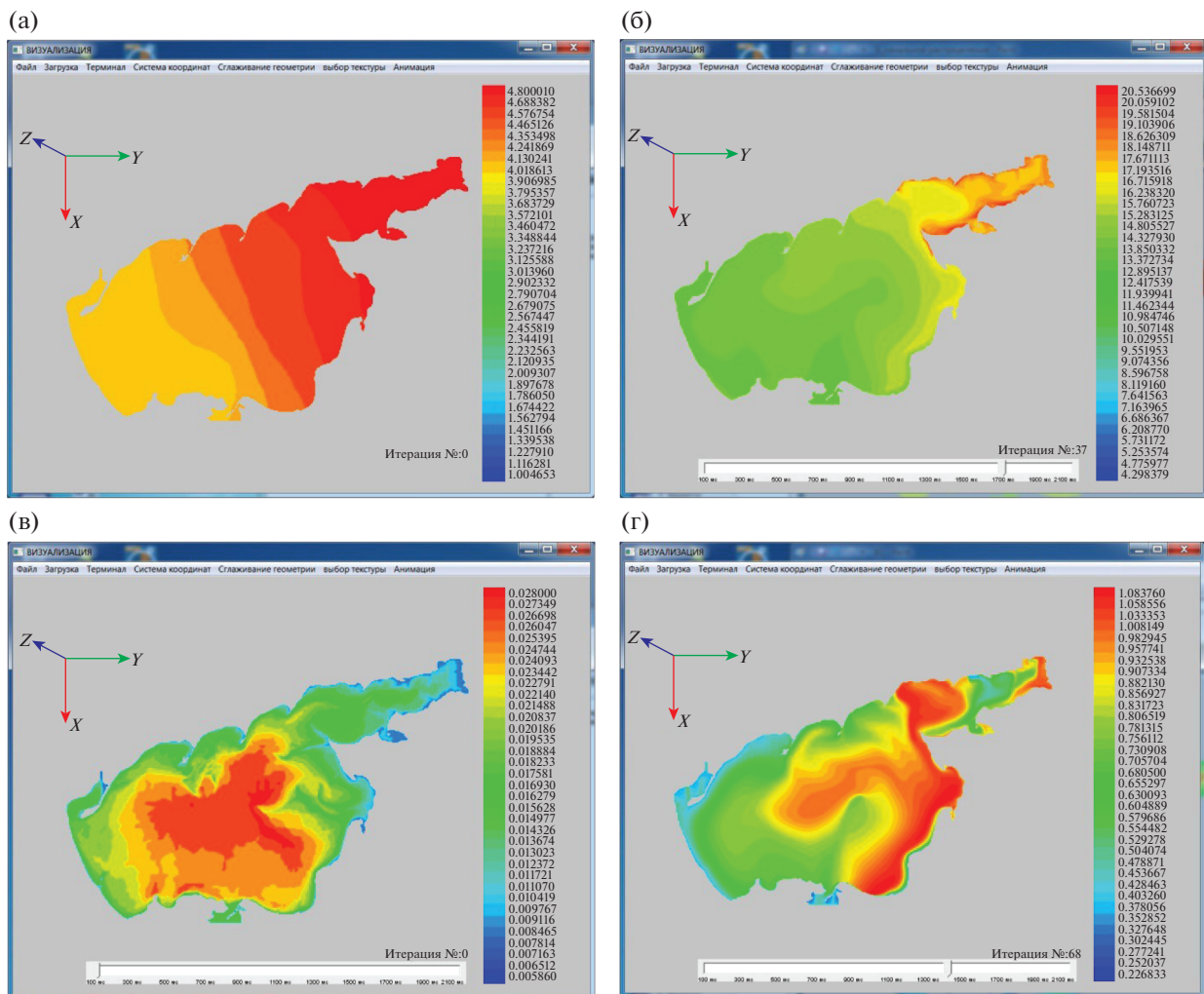
Использование ГИС предоставляет дополнительные возможности для более качественного и сложного пространственного анализа, а решения на его основе являются более точными.

На фиг. 11 представлены результаты расчета концентрации загрязняющего биогенного вещества и фитопланктона, усредненных по глубине, для модели биологической кинетики.

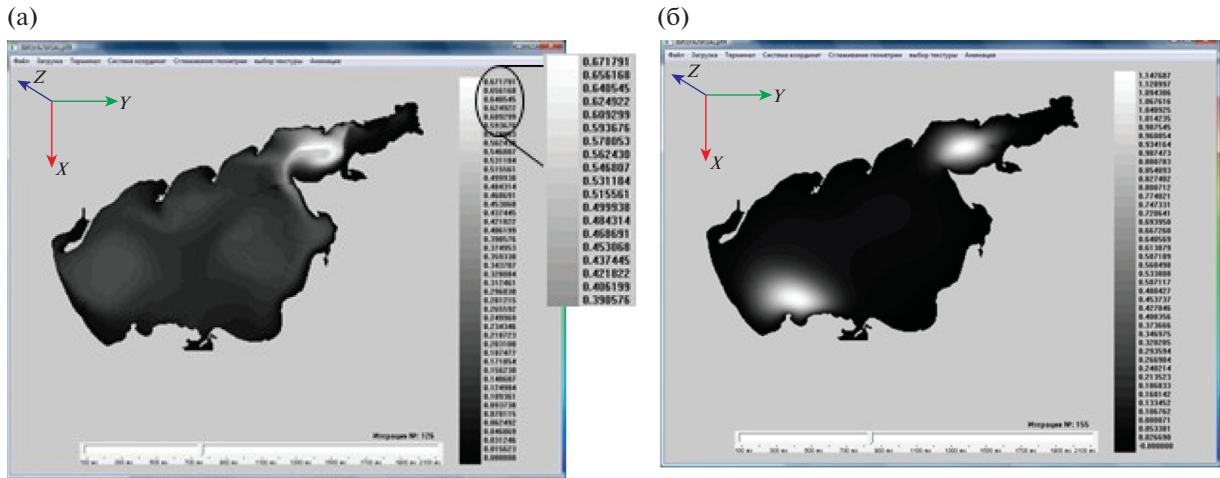
Значения коэффициентов: $\mu_3 = 5 \times 10^{-10}$; $\nu_3 = 10^{-10}$; $B = 0.001$; $\tilde{S}_3 = 1$; $f = 3$; $\tau_i = 0.1$; $i \in \{1, 5\}$; $\lambda_2 = 0.8$; $\mu_1 = 5 \times 10^{-11}$; $\nu_1 = 10^{-11}$; N – номер итерации (начальное распределение полей течений водного потока для северного ветра). Начальные данные для модели биологической кинетики были получены на основе экспедиционных данных, литературных источников, электронного атласа, подготовленного ЮНЦ РАН 2018 г., экологического атласа Черного и Азовского морей, подготовленного в 2019 г. Для калибровки и верификации разработанных моделей гидрофизики,



Фиг. 10. Схема программного комплекса.



Фиг. 11. Распределение концентраций: (а) – загрязняющих биогенных веществ в начальный момент времени (S_3); (б) – с течением времени, $T = 28$ дней; (в) – распределение концентрации фитопланктона (S_1) в начальный момент времени; (г) – с течением времени, $T = 64$ дня.



Фиг. 12. Распределение концентраций: (а) – детрита, временной интервал $T = 62$ дня; (б) – промысловой рыбы, временной интервал $T = 76$ дней.

входящих в ПК, использовались данные портала Единой государственной системы информации об обстановке в Мировом океане “ЕСИМО”, портала “Аналитические ГИС”, разработанного Институтом проблем передачи информации РАН (Москва). В качестве входных данных для моделирования гидрофизических процессов, помимо экспедиционных данных и литературных источников, использовались данные НИЦ “Планета”, данные Азовского научно-исследовательского института рыбного хозяйства (“АзНИИРХ”), ФГУ “Азовморинформцентра”, экспедиционные исследования.

С использованием разработанного комплекса программ был изучен механизм опасного явления – образования заморных зон в мелководном водоеме. На фиг. 12 представлены результаты моделирования возможных сценариев развития экосистемы Азовского моря (изменения концентрации популяции промысловой рыбы-детритофага пиленгас) (начальное распределение полей течений водного потока, полученное в результате натурального эксперимента, задано для северного направления ветра). Белым цветом выделены максимальные значения концентраций рыбы и детрита, начальное распределение полей течений в Азовском море при северном ветре. Значения коэффициентов: $\mu_4 = 3 \times 10^{-11}$; $\nu_4 = 10^{-11}$; $\varepsilon_4 = 1.9 \times 10^{-5}$; $\beta_4 = 0.1$; $\lambda_4 = 0.4$; $\mu_5 = 1.5 \times 10^{-3}$; $\nu_5 = 1.6 \times 10^{-3}$; $\gamma_5 = 0.125$; $\lambda_5 = 1.16 \times 10^{-3}$; $\xi_5 = 0.8$; $\varepsilon_5 = 0.47$; $\delta_5 = 0.05$.

Представленные результаты численного эксперимента, проведенного на основе сценарного подхода, позволяют исследовать возможный сценарий зарыбления акватории Азовского моря пеленгасом, участвующим в донной мелиорации мест скопления детрита. При значении временного интервала, начиная с 61-го дня, согласно представленным результатам, наблюдается снижение концентрации детрита, а значит, и концентрации донных отложений в центрально-восточной части Азовского моря, что в конечном счете приведет к уменьшению площадей заморных зон и улучшению качества вод в изучаемом водоеме.

С помощью разработанного программного комплекса можно исследовать пластичность пищевого режима и вопросы акклиматизации рыб-детритофагов в новых по экологическому режиму условиях, оценить специфику условий водоема, используемого для выпуска молоди, что позволит избежать непредвиденных отрицательных технологических влияний и планировать улучшение процессов воспроизводства промысловых рыб, активно участвующих в процессах самоочищения и донной мелиорации экосистемы. Анализ полученных результатов позволил сделать выводы о свойствах математической модели детрит–промысловая рыба и возможностях управления качеством вод мелководных водоемов, таких как Азовское море, с помощью методов и средств математического моделирования.

ЗАКЛЮЧЕНИЕ

Разработаны взаимосвязанные математические пространственно-трехмерные модели гидродинамики и биологической кинетики мелководного водоема, характеризующегося значитель-

ными перепадами глубин и солености, сложным распределением температуры водной среды и кислородным режимом. Результаты расчетов с помощью пространственно-трехмерной модели гидродинамики, учитывающей физические свойства водной среды, используются в качестве входных данных для разработки сценариев динамики процессов транспорта и трансформации загрязняющих биогенных элементов в водоеме, влияющих на развитие биоты, включая фито-, зоопланктон и рыбные популяции (см. [19]–[21]).

Построена и исследована трехмерная математическая модель динамики планктона и промысловых рыб для мелководного водоема, учитывающая конвективный и диффузионный переносы, поглощение и выделение питательных веществ фито- и зоопланктоном, соленость, температуру, кислородный режим, пространственное распределение детрита, нелинейные взаимодействия гидробионтов, таксис и др. Разработка численных методов решения поставленных модельных задач, использующих конечно-разностные схемы с учетом степени заполненности контрольных ячеек расчетной области, позволила уменьшить погрешность и сократить время расчетов в десятки раз. Задачи гидродинамики и биологической кинетики были решены с использованием высокопроизводительных вычислительных систем, способных выполнять большое количество сложных расчетов и обрабатывать огромные объемы данных за ограниченное время, что является актуальным при построении оперативных прогнозов изменения состояния водной экосистемы при возникновении в ней опасных явлений природного характера.

На основе разработанного программного комплекса проведена реконструкция опасных явлений мелководного водоема (природного и техногенного характеров), связанных с распространением вредных загрязняющих веществ, эвтрофикацией, “цветением водорослей”, вызывающим возникновение заморных зон, характеризующихся пониженной концентрацией кислорода. Изучение подобных ПК водоемов показало, что в результате его разработки удалось повысить точность прогнозов изменения концентраций планктона и промысловых рыб на 10–15% в зависимости от решаемой модельной задачи биологической кинетики. Мониторинг водной акватории, проводимый на регулярной основе, позволил разработать математические модели процессов гидрофизики и биологической кинетики, предназначенные для прогнозирования возможных сценариев развития экосистемы Азовского моря, а также изучения механизмов возникновения областей анаэробного заражения и принятия своевременных мер, направленных на их локализацию и уменьшение возможных негативных последствий.

СПИСОК ЛИТЕРАТУРЫ

1. Марчук Г.И., Дымников В.П., Залесный В.Б., Лыков В.Н., Бобылева И.М., Галин В.Я., Перов В.Л. Математическая модель общей циркуляции атмосферы и океана // Докл. АН СССР. 1980. Т. 253. № 3. С. 577–581.
2. Самарский А.А., Вабичевич П.Н. Численные методы решения задач конвекции–диффузии. М.: URSS, 2009. 248 с.
3. Ansgore R., Sonar Th. Mathematical models of fluid dynamics: modelling, theory, basic numerical facts –an introduction, 2nd, Updated Edition. ISBN: 978-3-527-62797-4. 2009. 242 p.
4. Alekseenko E., Roux B., Sukhinov A.I., Kotarba R., Fougere D. Coastal hydrodynamics in a windy lagoon // Computers and Fluids. 2013. V. 77. P. 24–35.
5. Yakushev E.V., Lukashev Yu.F., Skirta A.Yu., Sorokin P.Yu., Soldatova E.V., Yakubenko V.G., Sukhinov A.I., Sergeev N.E., Fomin S.Yu., Sapozhnikov F.V. Comprehensive oceanological studies of the sea of Azov during cruise 28 of r/v Akvanavt (july–august 2001) // Oceanology. 2003. V. 43. № 1. P. 39–47.
6. Logofet D.O. Stronger-than-Lyapunov notions of matrix stability, or how “flowers” help solve problems in mathematical ecology // Linear Algebra and its Appl. 2005. V. 398. P. 75–100.
7. Tyutyunov Yu.V., Titova L.I., Senina I.N. Prey-taxis destabilizes homogeneous stationary state in spatial Gause–Kolmogorov-type model for predator-prey system // Ecological Complexity. 2017. V. 31. P. 170–180.
8. Четверушкин Б.Н., Гасилов В.А., Поляков С.В., Карташева Е.Л., Яковлевский М.В., Абалакин И.В., Бобков В.Г., Болдарев А.С., Болдырев С.Н., Дьяченко С.В., Кринов П.С., Минкин А.С., Нестеров И.А., Ольховская О.Г., Попов И.В., Суков С.А. Пакет прикладных программ GIMM для решения задач гидродинамики на многопроцессорных вычислительных системах // Матем. моделирование. 2005. Т. 17. № 6. С. 58–74.
9. Давыдов А.А., Четверушкин Б.Н., Шильников Е.В. Моделирование течений несжимаемой жидкости и слабосжимаемого газа на многоядерных гибридных вычислительных системах // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 12. С. 2275–2284.
10. Четверушкин Б.Н. Пределы детализации и формулировка моделей уравнений сплошных сред // Матем. моделирование. 2012. Т. 24. № 11. С. 33–52.

11. *Сухинов А.И., Чистяков А.Е., Шишенин А.В., Тимофеева Е.Ф.* Предсказательное моделирование прибрежных гидрофизических процессов на многопроцессорной системе с использованием явных схем // Матем. моделирование. 2018. Т. 30. № 3. С. 83–100.
12. *Петров И.Б., Фаворская А.В., Санников А.В., Квасов И.Е.* Сеточно-характеристический метод с использованием интерполяции высоких порядков на тетраэдральных иерархических сетках с кратным шагом по времени // Матем. моделирование. 2013. Т. 25. № 2. С. 42–52.
13. *Белоцерковский О.М., Опарин А.М., Четкин В.М.* Турбулентность. Новые подходы. М.: Наука, 2003. 286 с.
14. *Коновалов А.Н.* Метод скорейшего спуска с адаптивным попеременно-треугольным переобуславливателем // Дифференц. уравнения. 2004. Т. 40. № 7. С. 953–963.
15. *Sukhinov A.I., Nikitina A.V., Chistyakov A.E., Semenov I.S.* Mathematical modeling of conditions for the formation of locks in shallow water reservoirs on a multiprocessor computer system // Comput. Meth. and Programming: New Comput. Technolog. 2013. V. 14. № 1. P. 103–112.
16. *Никитина А.В.* Модели биологической кинетики, стабилизирующие экологическую систему Таганрогского залива // Изв. ЮФУ. 2009. № 8 (97). С. 130–134.
17. *Никитина А.В., Семенов И.С.* Параллельная реализация модели динамики токсичной водоросли в Азовском море с применением многопоточности в операционной системе Windows // Изв. ЮФУ. Технические науки. 2013. № 1 (138). С. 130–135.
18. *Sukhinov A.I., Chistyakov A.E., Ugol'nitskii G.A., Usov A.B., Nikitina A.V., Puchkin M.V., Semenov I.S.* Game-theoretic regulations for control mechanisms of sustainable development for shallow water ecosystems // Automat. and Remote Control. 2017. V. 78 (6). P. 1059–1071.
19. *Sukhinov A.I., Chistyakov A.E., Nikitina A.V., Belova Yu.V., Sumbaev V.V., Semenyakina A.A.* Supercomputer modeling of hydrochemical condition of shallow waters in summer taking into account the influence of the environment // Comm. in Computer and Informat. Sci. 2018. V. 910. P. 336–351.
20. *Nikitina A.V., Kravchenko L., Semenov I., Belova Y., Semenyakina A.* Modeling of production and destruction processes in coastal systems on a supercomputer // MATEC Web of Conferences, 2018, 226, 04025. <https://doi.org/10.1051/mateconf/201822604025>
21. *Гущин В.А., Никитина А.В., Семенякина А.А., Сухинов А.И., Чистяков А.Е.* Модель транспорта и трансформации биогенных элементов в прибрежной системе и ее численная реализация // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 8. С. 120–137.

БЫСТРЫЙ АЛГОРИТМ РЕШЕНИЯ ТРЕХМЕРНОЙ ОБРАТНОЙ МНОГОЧАСТОТНОЙ ЗАДАЧИ СКАЛЯРНОЙ АКУСТИКИ С ДАННЫМИ В ЦИЛИНДРИЧЕСКОЙ ОБЛАСТИ¹⁾

© 2022 г. А. Б. Бакушинский^{1,2,*}, А. С. Леонов^{3,**}

¹ 117312 Москва, пр-т 60-летия Октября, 9, ИСА ФИЦ ИУ РАН, Россия

² 424000 Йошкар-Ола, пл. Ленина, 1, МарГУ, Республика Марий Эл, Россия

³ 115409 Москва, Каширское ш., 31, НИЯУ «МИФИ», Россия

*e-mail: bakush@isa.ru

**e-mail: asleonov@mephi.ru

Поступила в редакцию 10.03.2021 г.
Переработанный вариант 10.03.2021 г.
Принята к публикации 04.08.2021 г.

Предлагается новый алгоритм устойчивого решения трехмерной скалярной обратной задачи акустического зондирования неоднородной среды в цилиндрической области. Данными для ее решения является комплексная амплитуда волнового поля, измеряемая вне области акустических неоднородностей в цилиндрическом слое. Обратная задача сводится с помощью преобразования Фурье и рядов Фурье к решению совокупности одномерных интегральных уравнений Фредгольма I рода, к последующему вычислению комплексной амплитуды волнового поля в области неоднородности и далее к нахождению искомого поля скоростей звука в этой области. Алгоритм позволяет решать обратную задачу на персональном компьютере средней производительности для достаточно мелких трехмерных сеток за десятки секунд. Проведено численное исследование точности предлагаемого алгоритма для решения модельных обратных задач на различных частотах и исследованы вопросы устойчивости алгоритма по отношению к возмущениям данных. Библ. 28. Фиг. 8.

Ключевые слова: трехмерное волновое уравнение, обратная коэффициентная задача, регуляризирующий алгоритм, быстрое преобразование Фурье.

DOI: 10.31857/S004446692112005X

1. ВВЕДЕНИЕ

Пусть скалярная функция $p(\mathbf{x}, t)$ определяет акустическое волновое поле, зависящее от координат $\mathbf{x} = (x, y, z)$ и времени $t \geq 0$ в области $Q \subset \mathbb{R}^3$. Область Q есть бесконечный цилиндр вида $Q = \{(x, y, z): x^2 + y^2 \leq b^2, z \in \mathbb{R}\}$. Поле создается источниками, локализованными в известной области S . Среда характеризуется локальной фазовой скоростью звука $c(\mathbf{x})$ и имеет постоянную плотность. При этом известно, что $c(\mathbf{x}) = c_0 = \text{const}$ вне заданной области $X \subset Q$, удовлетворяющей условию $X \cap S = \emptyset$. В самой области X функция $c(\mathbf{x})$ может быть переменной, и это интерпретируется как расположение там акустических неоднородностей. В этом случае для гармонического источника вида $f(\mathbf{x}, \omega)e^{i\omega t}$ с известной частотой ω поле $p(\mathbf{x}, t)$ в рамках приближения линейной акустики может быть найдено в форме $p(\mathbf{x}, t) = u(\mathbf{x}, \omega)e^{i\omega t}$, где комплексная амплитуда $u(\mathbf{x}, \omega)$ удовлетворяет уравнению

$$\Delta u(\mathbf{x}, \omega) + k_0^2 u(\mathbf{x}, \omega) = f(\mathbf{x}, \omega) + \omega^2 \xi(\mathbf{x})u(\mathbf{x}, \omega), \quad \mathbf{x} \in Q. \quad (1.1)$$

¹⁾ Работа выполнена при финансовой поддержке РФФИ (проект 20-11-20085) для первого автора в части обоснования численных алгоритмов и Программы повышения конкурентоспособности Национального исследовательского ядерного университета МИФИ (проект 02.а03.21.0005 от 27.08.2013) для второго автора.

Здесь $k_0 = \frac{\omega}{c_0}$, а $\xi(\mathbf{x}) = c_0^{-2} - c^{-2}(\mathbf{x})$. Будем считать, что выполнено также граничное условие

$$\mathbb{N}[u]_{\partial Q} = \left(\frac{\partial u}{\partial n} + \sigma(\mathbf{x})u \right)_{\partial Q} = 0$$

с известной функцией $\sigma(\mathbf{x})$ и условие излучения по координате z . Не вдаваясь в детализацию условий на коэффициенты $\sigma(\mathbf{x})$ и $f(\mathbf{x}, \omega)$, мы сделаем

Предположение 1. Функция $\xi(\mathbf{x})$ непрерывна с компактным носителем в X и соответствующая задача (1.1) с указанными дополнительными условиями имеет при каждом рассматриваемом ω единственное решение $u(\mathbf{x}, \omega) \in H_1(Q)$.

Нахождение такой функции $u(\mathbf{x}, \omega)$ составляет *прямую задачу*. Нас интересует следующая *обратная задача* для уравнения (1.1) с указанными дополнительными условиями: зная для некоторого набора частот ω комплексную амплитуду поля $u(\mathbf{x}, \omega)$ в области Y ($Y \subset Q$, $Y \cap X = \emptyset$, $Y \cap S = \emptyset$) найти коэффициент $\xi(\mathbf{x})$, т.е. функцию $c(\mathbf{x})$, определяющую акустические неоднородности в области X .

Вводя функцию Грина $G(\mathbf{x}, \mathbf{x}', \omega)$ для уравнения Гельмгольца в области Q для нерезонансных величин ω , можно при определенных предположениях о гладкости функций $u(\mathbf{x}, \omega)$, $f(\mathbf{x}, \omega)$, $c(\mathbf{x})$ (см., например, [1]–[4] и др.) свести обратную задачу к нелинейной относительно неизвестных $u(\mathbf{x}', \omega)$, $\xi(\mathbf{x}')$, $\mathbf{x}' \in X$, системе интегральных уравнений:

$$\begin{aligned} u(\mathbf{x}, \omega) &= u_0(\mathbf{x}, \omega) + \omega^2 \int_X G(\mathbf{x}, \mathbf{x}', \omega) \xi(\mathbf{x}') u(\mathbf{x}', \omega) d\mathbf{x}', \quad \mathbf{x} \in X, \\ \omega^2 \int_X G(\mathbf{x}, \mathbf{x}', \omega) \xi(\mathbf{x}') u(\mathbf{x}', \omega) d\mathbf{x}' &= w(\mathbf{x}, \omega), \quad \mathbf{x} \in Y. \end{aligned} \quad (1.2)$$

Входящие в (1.2) функции

$$u_0(\mathbf{x}, \omega) = \int_X G(\mathbf{x}, \mathbf{x}', \omega) f(\mathbf{x}', \omega) d\mathbf{x}', \quad w(\mathbf{x}, \omega) = u(\mathbf{x}, \omega) - u_0(\mathbf{x}, \omega), \quad \mathbf{x} \in Y,$$

суть известные (вычислимые) функции, а величины $u(\mathbf{x}', \omega)$, $\xi(\mathbf{x}')$, $\mathbf{x}' \in X$, подлежат определению.

Задача (1.2) достаточно хорошо исследована теоретически: изучены вопросы существования и единственности ее решения в различных областях Q , X , Y (см., например, [2]–[5] и др.), вопросы ее устойчивости по отношению к возмущениям данных и т.д. Например, известно, что для единственности решения такой обратной задачи в общем случае необходимо использовать по меньшей мере счетное число частот. Однако на практике реализовать это невозможно, и приходится ограничиваться сравнительно небольшим их числом, теоретически теряя единственность решения. Анализ этой проблемы дан, например, в [4]. Тем не менее такое изучение акустических рассеивателей на сравнительно небольшом числе частот часто встречается на практике. Для получения в этом случае определенного решения обратной задачи необходимо использовать дополнительные априорные ограничения для него, сужающие множество возможных решений.

Отметим, что цель этой статьи – не теоретическое исследование свойств обратной задачи (1.2). Мы лишь предлагаем в рамках рассматриваемой модели (1.2) эффективный с точки зрения быстродействия на достаточно “мелких” сетках численный метод ее решения в случае цилиндрических областей Q , X , Y . Предположение цилиндричности является принципиальным: именно оно определяет структуру метода и в конечном итоге его быстродействие. Имея в виду такие акценты, мы не проводим детальную редукцию обратной задачи к системе (1.2) с указанием всех требований к коэффициентам, а сделаем лишь

Предположение 2. Редукция обратной задачи к системе (1.2) возможна для данной функции $u_0(\mathbf{x}, \omega) \in L_2(Q)$. Система (1.2) разрешима и для всех рассматриваемых ω определяет некоторое решение обратной задачи: непрерывную функцию $\xi(\mathbf{x})$ с компактным носителем в X и функцию $u(\mathbf{x}, \omega) \in L_2(X)$.

Подчеркнем, что при таких предположениях система (1.2) может иметь неединственное решение. Выделение единственного решения требует дополнительных предположений о его характере, и это будет сделано ниже.

Во многих работах рассмотрены разнообразные численные методы решения задач типа (1.2) в двумерной и трехмерной постановках. Так, в [3] система (1.2) сводится к нелинейному операторному уравнению, которое затем решается специальным итерационным методом. При этом явно не учитывается, что это уравнение является некорректно поставленной задачей и итерационный метод не регуляризуется. Тем не менее метод работает для модельных “рассеивателей средней силы” [3, с. 101–105]. В работе [5] для решения системы уравнений (1.2) в трехмерном аксиально-симметричном случае использован регуляризованный метод Гаусса–Ньютона. В работе [6] для аналогичной задачи были применены специальный градиентный метод и метод Флетчера–Ривса. Другие градиентные методы были использованы в [7], [8]. Обзор этих и подобных подходов можно найти, например, в [4].

Отметим, что имеются и альтернативные подходы в решении обратной задачи акустического зондирования, не связанные с системой типа (1.2). В частности, оригинальный *метод граничного управления*, позволяющий решать трехмерные задачи типа (2), был предложен и развит в работах [9], [10]. В [11] для решения двумерной обратной задачи рассеяния был применен известный метод Р.Г. Новикова [12], а в последующей работе [13] проведен сравнительный анализ разновидности этого метода и некоторых других функционально-аналитических методов решения двумерных обратных задач акустического рассеяния (см. также [4]). Весьма многообещающими при обработке реальных экспериментальных данных оказались методы М.В. Клибанова, суммированные в монографии [14], а также методы из монографии [15]. Большой интерес представляют также недавние работы [16], [17].

Все упомянутые методы решения обратной задачи акустического зондирования требуют в трехмерной постановке значительных вычислительных ресурсов и, прежде всего, – большого времени решения. Поэтому здесь актуально создание быстродействующих алгоритмов. В связи с этим отметим статьи [18]–[20]. В них рассматривались не задачи (1.1) и (1.2), а трехмерное интегральное уравнение Фредгольма I рода

$$\int_X \frac{\zeta(x') dx'}{|x - x'|} = U^{(0)}(x) - U(x), \quad x \in Y,$$

к которому в различных областях X сводится обратная коэффициентная задача для волнового уравнения. Правая часть этого уравнения содержит специальные интегралы $U^{(0)}(x)$, $U(x)$ от регистрируемого в Y волнового поля, а искомая функция $\zeta(x)$ связана с $\xi(x)$. Принципиальной в этих статьях является область регистрации этого поля – плоский или цилиндрический тонкий слой. В результате предложенные в этих статьях методы решения интегрального уравнения оказались очень эффективным численно, и позволили решать трехмерные обратные задачи для достаточно мелких сеток на персональном компьютере (ПК) даже без распараллеливания за несколько минут. В дальнейшем мы применили в работе [21] аналогичный, но не тождественный, метод для решения уже другой задачи – задачи (1.2) в бесконечном пространстве с использованием плоских слоев X , Y . В предлагаемой теперь работе мы модифицируем метод решения указанного интегрального уравнения из [20] для его приложения к обратной задаче (2) в цилиндрической области, и в итоге получаем весьма быстродействующий алгоритм ее решения.

В статье мы придерживаемся следующей схемы решения нелинейной системы (1.2) как обратной задачи для каждой используемой частоты ω :

1) решаем второе уравнение (линейное интегральное уравнение Фредгольма I рода), записанное в виде

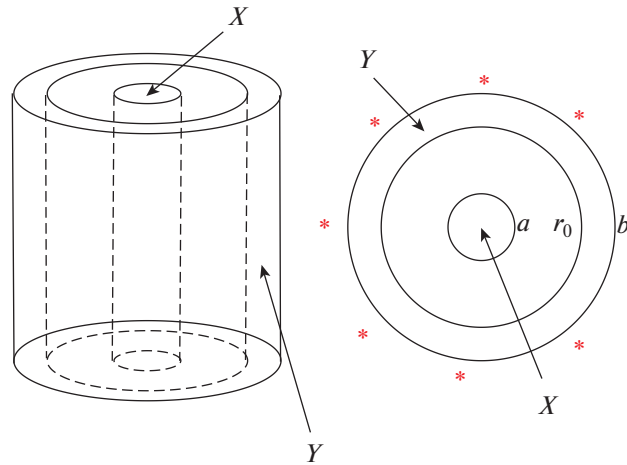
$$\omega^2 \int_X G(\mathbf{x}, \mathbf{x}', \omega) v(\mathbf{x}', \omega) d\mathbf{x}' = w(\mathbf{x}, \omega), \quad \mathbf{x} \in Y, \tag{1.3}$$

относительно функции $v(\mathbf{x}', \omega) = \xi(\mathbf{x}') u(\mathbf{x}', \omega)$, $\mathbf{x}' \in X$;

2) вычисляем функцию $u(\mathbf{x}, \omega)$, $\mathbf{x} \in X$, из первого равенства системы (1.2), записанного в форме

$$u(\mathbf{x}, \omega) = u_0(\mathbf{x}, \omega) + \omega^2 \int_X G(\mathbf{x}, \mathbf{x}', \omega) v(\mathbf{x}', \omega) d\mathbf{x}', \quad \mathbf{x} \in X; \tag{1.4}$$

3) находим функцию $\xi(\mathbf{x})$ из уравнения $v(\mathbf{x}, \omega) = \xi(\mathbf{x}) u(\mathbf{x}, \omega)$, $\mathbf{x} \in X$, используя найденные функции $v(\mathbf{x}, \omega)$, $u(\mathbf{x}, \omega)$.



Фиг. 1. Геометрическая схема регистрации данных обратной задачи: X — область рассеивателей волнового поля, Y — область регистрации данных, звездочки — условные положения источников.

Эта схема будет конкретизирована ниже. Близкий подход использовался и ранее. Например, он реализован в работе [22]. Однако там для решения некорректной задачи — трехмерного уравнения, аналогичного (1.3), не применялись методы регуляризации, а решение обратной задачи искалось на достаточно узком классе функций с “кусочно-постоянным током”. В итоге, для решения такой трехмерной обратной задачи на сравнительно мелких сетках также требуется значительное время.

Мы предлагаем численный алгоритм решения обратной задачи (1.2), основанный на схеме 1)–3) и специализированный для цилиндрических областей X , Y , Q . В этом смысле алгоритм не является универсальным. Однако используемая в нем геометрия области решений X и области регистрации данных Y приемлема в определенной мере для ряда технических систем. Ниже мы продемонстрируем быстрдействие полученного алгоритма. Этот новый алгоритм решения обратной задачи (1.2) в цилиндрической области, основанный на схеме 1)–3), и его численное исследование являются основными результатами данной работы.

2. ИСПОЛЬЗУЕМАЯ ГЕОМЕТРИЧЕСКАЯ СХЕМА И СВЕДЕНИЕ ОБРАТНОЙ ЗАДАЧИ К ОДНОМЕРНЫМ ИНТЕГРАЛЬНЫМ УРАВНЕНИЯМ

Везде далее области X , Y имеют вид цилиндров: область решения $X = \{0 \leq r \leq a\} \times \mathbb{R}_z$, $r = \sqrt{x^2 + y^2}$, и область наблюдения $Y = \{r_0 \leq r \leq b\} \times \mathbb{R}_z$. На фиг. 1 представлена схема расположения этих областей. Там же условно показано и возможное положение источников. Мы считаем цилиндры бесконечными по переменной z : $z \in (-\infty, +\infty)$, имея в виду финитность искомой функции ξ и по этой переменной в частности. В дальнейшем будем обозначать через X_{xy} , Y_{xy} сечения цилиндрических областей X , Y плоскостью, перпендикулярной оси Oz .

Будем далее считать, что $\sigma(\mathbf{x}) = \sigma = \text{const} \geq 0$. Функция Грина для задачи (1.1) с указанными дополнительными условиями в цилиндрической области Q хорошо известна (см., например, [23, с. 615]), и здесь мы не будем выписывать ее явно. Отметим важную особенность этой функции: в цилиндрических координатах $\mathbf{x} = (r, \varphi, z)$, $\mathbf{x}' = (r', \varphi', z')$ она имеет форму:

$$\begin{aligned} G(\mathbf{x}, \mathbf{x}', \omega) &= G\left(\sqrt{r^2 + r'^2 - 2rr' \cos(\varphi - \varphi') + (z - z')^2}; \omega\right) = \\ &= G_0(r, r', \varphi - \varphi', z - z'; \omega), \end{aligned}$$

а вид функции G_0 можно найти, например, в [23]. В этих же координатах представим функции $u(\mathbf{x}, \omega) = u(r, \varphi, z, \omega)$, $v(\mathbf{x}, \omega) = v(r, \varphi, z, \omega)$. Тогда уравнения (1.3), (1.4) можно записать в виде

$$\omega^2 \int_0^a \int_0^{2\pi} \int_{-\infty}^{+\infty} G_0(r, r', \varphi - \varphi', z - z'; \omega) v(r', \varphi', z'; \omega) r' dr' d\varphi' dz' = w(r, \varphi, z; \omega), \quad r \in [r_0, b], \quad (2.1)$$

$$\begin{aligned} & u(r, \varphi, z; \omega) - u_0(r, \varphi, z; \omega) = \\ & = \omega^2 \int_0^a \int_0^{2\pi} \int_{-\infty}^{+\infty} G_0(r, r', \varphi - \varphi', z - z'; \omega) v(r', \varphi', z'; \omega) r' dr' d\varphi' dz', \quad r \in [0, a], \end{aligned} \quad (2.2)$$

при $\varphi \in [0, 2\pi]$, $z \in \mathbb{R}$. Здесь $v(r', \varphi', z'; \omega) = \xi(r', \varphi', z') u(r', \varphi', z'; \omega)$.

Введем преобразования Фурье

$$F_z[\cdot] = \int_{-\infty}^{+\infty} [\cdot] e^{i\Omega z} dz$$

по переменной z (или z') для функций G_0, v, u, u_0, w как элементов пространства L_2 (см. предположение 2):

$$\begin{aligned} \tilde{G}(r, r', \varphi, \Omega; \omega) &= F_z[G_0(r, r', \varphi, z; \omega)](\Omega), \quad \tilde{w}(r, \varphi, \Omega; \omega) = F_z[w(r, \varphi, z; \omega)](\Omega), \\ \tilde{u}(r, \varphi, \Omega; \omega) &= F_z[u(r, \varphi, z; \omega)](\Omega), \quad \tilde{u}_0(r, \varphi, \Omega; \omega) = F_z[u_0(r, \varphi, z; \omega)](\Omega). \end{aligned}$$

Тогда по теореме о свертке равенства (2.1), (2.2) можно записать в виде

$$\omega^2 \int_0^a \int_0^{2\pi} \tilde{G}(r, r', \varphi - \varphi', \Omega; \omega) \tilde{v}(r', \varphi', \Omega; \omega) r' dr' d\varphi' = \tilde{w}(r, \varphi, \Omega; \omega), \quad r \in [r_0, b], \quad (2.3)$$

$$\begin{aligned} & \tilde{u}(r, \varphi, \Omega; \omega) - \tilde{u}_0(r, \varphi, \Omega; \omega) = \\ & = \omega^2 \int_0^a \int_0^{2\pi} \tilde{G}(r, r', \varphi - \varphi', \Omega; \omega) \tilde{v}(r', \varphi', \Omega; \omega) r' dr' d\varphi', \quad r \in [0, a], \end{aligned} \quad (2.4)$$

где $\tilde{v}(r', \varphi', \Omega; \omega) = F_z[v(r', \varphi', z'; \omega)](\Omega) = F_z[\xi(r', \varphi', z') u(r', \varphi', z'; \omega)]$. Введем также разложения функций $\tilde{G}, \tilde{v}, \tilde{u}, \tilde{u}_0, \tilde{w}$ по системе $\{e^{in\varphi}\}$, $n \in \mathbb{Z}$, в пространстве $L_2(0, 2\pi)$:

$$\begin{aligned} \tilde{G}(r, r', \varphi, \Omega; \omega) &= \sum_n G_n(r, r', \Omega; \omega) e^{in\varphi}, \quad \tilde{v}(r', \varphi', \Omega; \omega) = \sum_n v_n(r', \Omega; \omega) e^{in\varphi'}, \\ \tilde{u}(r, \varphi, \Omega; \omega) &= \sum_n u_n(r, \Omega; \omega) e^{in\varphi}, \quad \tilde{u}_0(r, \varphi, \Omega; \omega) = \sum_n u_{0n}(r, \Omega; \omega) e^{in\varphi}; \\ \tilde{w}(r, \varphi, \Omega; \omega) &= \sum_n w_n(r, \Omega; \omega) e^{in\varphi}; \quad \varphi, \varphi' \in [0, 2\pi] \end{aligned}$$

с коэффициентами

$$\begin{aligned} G_n(r, r', \Omega; \omega) &= \frac{1}{2\pi} \int_0^{2\pi} \tilde{G}(r, r', \varphi, \Omega; \omega) e^{-in\varphi} d\varphi, \\ v_n(r', \Omega; \omega) &= \frac{1}{2\pi} \int_0^{2\pi} \tilde{v}(r', \varphi', \Omega; \omega) e^{-in\varphi'} d\varphi', \quad w_n(r, \Omega; \omega) = \frac{1}{2\pi} \int_0^{2\pi} \tilde{w}(r, \varphi, \Omega; \omega) e^{-in\varphi} d\varphi, \\ u_n(r, \Omega; \omega) &= \frac{1}{2\pi} \int_0^{2\pi} \tilde{u}(r, \varphi, \Omega; \omega) e^{-in\varphi} d\varphi, \quad u_{0n}(r, \Omega; \omega) = \frac{1}{2\pi} \int_0^{2\pi} \tilde{u}_0(r, \varphi, \Omega; \omega) e^{-in\varphi} d\varphi. \end{aligned}$$

Тогда соотношения (2.3), (2.4) можно свести к следующей системе равенств, справедливой для любых $n \in \mathbb{Z}$, $\Omega \in \mathbb{R}$ и всех рассматриваемых ω :

$$\omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n(r', \Omega; \omega) r' dr' = \frac{1}{2\pi} w_n(r, \Omega; \omega), \quad r \in [r_0, b]; \quad (2.5)$$

$$\frac{1}{2\pi}(u_n(r, \Omega; \omega) - u_{0n}(r, \Omega; \omega)) = \omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n(r', \Omega; \omega) r' dr', \quad r \in [0, a]. \quad (2.6)$$

При этом

$$\begin{aligned} v_n(r', \Omega; \omega) &= \frac{1}{2\pi} \int_0^{2\pi} F_z [v(r', \varphi, z; \omega)] e^{-in\varphi} d\varphi = \frac{1}{2\pi} \int_0^{2\pi} F_z [\xi(r', \varphi, z) u(r', \varphi, z; \omega)] e^{-in\varphi} d\varphi = \\ &= \frac{1}{2\pi} \int_0^{2\pi} F_z \left[\xi(r', \varphi, z) \sum_m u_m(r', \Omega; \omega) e^{im\varphi} \right] e^{-in\varphi} d\varphi. \end{aligned} \quad (2.7)$$

Соотношения (2.5), (2.6) являются уравнениями для неизвестных функций $v_n(r', \Omega; \omega)$ и $u_n(r, \Omega; \omega)$ одной переменной r' или r . Другие аргументы этих функций, т.е. Ω , ω , играют роль параметров.

3. АЛГОРИТМЫ РЕШЕНИЯ ПРЯМОЙ И ОБРАТНОЙ ЗАДАЧИ ДЛЯ ЦИЛИНДРИЧЕСКОЙ ОБЛАСТИ

3.1. Прямая задача

Рассматриваемая ниже прямая задача заключается в нахождении из равенств (2.1), (2.2) функции $w(r, \varphi, z, \omega)$, $r \in [r_0, b]$, $\varphi \in [0, 2\pi]$, $z \in \mathbb{R}$, по известной финитной функции $\xi(r', \varphi', z')$ и заданной функции источников $u_0(r, \varphi, z; \omega)$ для рассматриваемого конечного набора частот ω . Для этого равенства (2.1), (2.2) сводятся к системе (2.3), (2.4) и далее к системе соотношений (2.5)–(2.7). Тогда вычисление функции w можно представить в виде следующего алгоритма.

Алгоритм 1

Шаг 1. Для набора рассматриваемых частот ω вычисляем преобразования Фурье по z известных функций G_0, u_0 :

$$\tilde{G}(r, r', \varphi, \Omega; \omega) = F_z [G_0(r, r', \varphi, z; \omega)](\Omega), \quad \tilde{u}_0(r, \varphi, \Omega; \omega) = F_z [u_0(r, \varphi, z; \omega)](\Omega)$$

и раскладываем полученные функции в ряды Фурье по φ :

$$\tilde{G}(r, r', \varphi, \Omega; \omega) = \sum_n G_n(r, r', \Omega; \omega) e^{in\varphi}, \quad \tilde{u}_0(r, \varphi, \Omega; \omega) = \sum_n u_{0n}(r, r', \Omega; \omega) e^{in\varphi}.$$

Обе эти процедуры реализуются с применением быстрого дискретного преобразования Фурье (БПФ).

Шаг 2. Для каждого из параметров ω, Ω реализуем следующий итерационный процесс решения уравнений (2.6), (2.7) относительно набора функции $\{u_n(r, \Omega; \omega)\}$:

$$v_n^{(k)}(r', \Omega; \omega) = \frac{1}{2\pi} \int_0^{2\pi} F_z \left[\xi(r', \varphi, z) \sum_m u_m^{(k)}(r', \Omega; \omega) e^{im\varphi} \right] e^{-in\varphi} d\varphi, \quad r' \in [0, a], \quad (3.1)$$

$$\begin{aligned} u_n^{(k+1)}(r, \Omega; \omega) &= u_{0n}(r, \Omega; \omega) + 2\pi\omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n^{(k)}(r', \Omega; \omega) r' dr', \quad r \in [0, a], \\ k &= 0, 1, 2, \dots, \end{aligned} \quad (3.2)$$

с начальным приближением в виде набора функций $\{u_n^{(0)}(r, \Omega; \omega)\} = \{u_{0n}(r, \Omega; \omega)\}$.

Шаг 3. Останавливаем процесс по некоторому правилу на итерации номер ν и получаем приближенное решение $\{u_n^{(\nu)}(r, \Omega; \omega)\}$ системы (2.6), (2.7) и соответствующие функции $\{v_n^{(\nu)}(r, \Omega; \omega)\}$.

Шаг 4. Согласно (2.5) вычисляем приближенный набор величин $\{w_n(r, \Omega; \omega)\}$:

$$w_n^{(\nu)}(r, \Omega; \omega) = 2\pi\omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n^{(\nu)}(r', \Omega; \omega) r' dr', \quad r \in [r_0, b], \quad (3.3)$$

и принимаем функцию

$$\tilde{w}^{(v)}(r, \varphi, \Omega; \omega) = \sum_n w_n^{(v)}(r', \Omega; \omega) e^{in\varphi} \quad (3.4)$$

(или ее обратное преобразование Фурье) как приближенное решение прямой задачи.

Мы не будем здесь проводить теоретический анализ сходимости алгоритма 1, так как формально он не используется нами при решении обратной задачи. Он нужен нам лишь для генерации ее модельных данных, и это лишь один из способов получения таких данных. Отметим лишь, что, как следует из общей теории решения интегральных уравнений II рода (см., например, [25]), алгоритм будет быстро сходиться по крайней мере для малых ω . Ниже будут продемонстрированы численные примеры, подтверждающие это положение.

3.2. Обратная задача

Она состоит в нахождении решения $\xi(r', \varphi', z')$ из системы (2.1), (2.2) по заданной функции $w(r, \varphi, z, \omega)$, $r \in [r_0, b]$, $\varphi \in [0, 2\pi]$, $z \in \mathbb{R}$, и заданной функции источников $u_0(r, \varphi, z; \omega)$ для каждой рассматриваемой частоты ω . Для этого мы сводим задачу (2.1), (2.2) к системе (2.5)–(2.7), считая заданными (вычисленными) наборы функций $\{w_n(r, \Omega; \omega)\}$, $\{u_{0n}(r, \Omega; \omega)\}$, и по этим наборам находим функцию ξ . Процедура решения представлена в виде следующего алгоритма.

Алгоритм 2

Шаг 1. Для всех рассматриваемых параметров ω , Ω и всех рассматриваемых n решаем одномерные интегральные уравнения I рода, соответствующие равенствам (2.5):

$$\omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n(r', \Omega; \omega) r' dr' = \frac{1}{2\pi} w_n(r, \Omega; \omega), \quad r \in [r_0, b]. \quad (3.5)$$

При этом используется подходящий метод регуляризации (регуляризирующий алгоритм, РА) этих некорректно поставленных задач. В итоге получается набор приближенных решений $\{v_n(r', \Omega; \omega)\}$.

Шаг 2. По найденному набору $\{v_n(r', \Omega; \omega)\}$ вычисляем, используя равенства (2.6), набор функций $\{u_n(r, \Omega; \omega)\}$:

$$u_n(r, \Omega; \omega) = u_{0n}(r, \Omega; \omega) + 2\pi\omega^2 \int_0^a G_n(r, r', \Omega; \omega) v_n(r', \Omega; \omega) r' dr', \quad r \in [0, a]. \quad (3.6)$$

Шаг 3. Восстанавливаем по наборам $\{v_n(r', \Omega; \omega)\}$ и $\{u_n(r, \Omega; \omega)\}$ функции $v(r, \varphi, z)$, $u(r, \varphi, z)$, $r', r \in [0, a]$, $\varphi \in [0, 2\pi]$, $z \in \mathbb{R}$, суммируя соответствующие ряды Фурье.

Шаг 4. Находим решение $\xi(r, \varphi, z)$ из уравнения $u(r, \varphi, z, \omega)\xi(r, \varphi, z) = v(r, z, \omega)$ для каждой точки $(r, \varphi, z) \in X$. Это можно сделать для каждого рассматриваемого ω , и результат будет, вообще говоря, зависеть от ω . Такая процедура соответствует исследованию акустических неоднородностей раздельно на каждой частоте. Далее можно, например, усреднить результаты тем или иным методом по величине ω . Альтернативный вариант – решить для каждой точки $(r, \varphi, z) \in X$ систему уравнений для всех рассмотренных ω : $u(r, \varphi, z, \omega)\xi(r, \varphi, z) = v(r, z, \omega)$, используя метод наименьших квадратов. Это также будет определенным усреднением по ω получаемых решений.

Шаг 5. Вычисляем функцию $c(r, \varphi, z)$ из равенства $\xi(r, \varphi, z) = c_0^{-2} - c^{-2}(r, \varphi, z)$.

Сделаем некоторые замечания об алгоритме 2.

Замечание 1. Для реализации ш. 1 необходимо детализировать предположения о функциях v , w . Используя предположение 2, мы считаем, что справедливы включения $v(\mathbf{x}, \omega) \in L_2(X)$, $w(\mathbf{x}, \omega) \in L_2(Y)$ для каждой рассматриваемой частоты ω . Первое включение вытекает из финитности функции ξ , а второе постулируется. В этом случае для уравнений (3.5) применимы известные методы решения линейных некорректно поставленных задач в гильбертовых пространствах (см., например, [5], [6], [26]–[28] и др.).

Замечание 2. В используемой схеме решения обратной задачи уравнение (1.3) (т.е. (2.1)) и порождаемые им уравнения (3.5) могут иметь неединственное решение для используемого конечного набора частот ω . Поэтому важно установить связь решений этих уравнений. Эту связь обосновывают следующие утверждения.

Теорема 1. 1. Пусть $w(r, \varphi, z; \omega) \in L_2(Y)$ при каждом ω . Тогда всякое решение уравнения (2.1) $v(r, \varphi, z; \omega) \in L_2(X)$ представимо в виде

$$v(r, \varphi, z; \omega) = F_{\Omega}^{-1} \left[\sum_n v_n(r, \Omega; \omega) e^{in\varphi} \right] (z), \quad (r, \varphi, z) \in X, \quad (3.7)$$

где функции $v_n(r, \Omega; \omega) \in L_2\{[0, a] \times \mathbb{R}_{\Omega}\}$ удовлетворяют интегральным уравнениям (2.5) при каждом ω . Обратное, если $v_n(r, \Omega; \omega) \in L_2\{[0, a] \times \mathbb{R}_{\Omega}\}$ – решения уравнений (2.5) такие, что $\sum_n \|v_n(r, \Omega; \omega)\|_{L_2\{[0, a] \times \mathbb{R}_{\Omega}\}}^2 < \infty$ при каждом ω , то функция вида (3.7) есть решение уравнения (2.1).

2. Для каждого ω справедливо равенство

$$\|v(r, \varphi, z; \omega)\|_{L_2(X)}^2 = \sum_n \|v_n(r, \Omega; \omega)\|_{L_2\{[0, a] \times \mathbb{R}_{\Omega}\}}^2 \quad (3.8)$$

Здесь $F_{\Omega}^{-1}[\cdot]$ – обратное преобразование Фурье по переменной z .

Следствие 1. Пусть функции $\bar{v}_n(r, \Omega; \omega) \in L_2\{[0, a] \times \mathbb{R}_{\Omega}\}$ – нормальные решения уравнений (2.5) (решения с минимальной нормой). Тогда функция

$$\bar{v}(r, \varphi, z; \omega) = F_{\Omega}^{-1} \left[\sum_n \bar{v}_n(r, \Omega; \omega) e^{in\varphi} \right] (z), \quad (r, \varphi, z) \in X,$$

есть единственное нормальное решение уравнения (2.1).

Доказательства этих утверждений проводятся так же, как в работе [18]. Там аналогичные утверждения доказывались для решений другого интегрального уравнения I рода, схожего с (2.1) по виду и свойствам и отличающегося ядром и правой частью. Для краткости мы опускаем повторение этих доказательств, заметив, что интегральное уравнение из [18] не имеет прямого отношения к обратной задаче, рассматриваемой в данной статье.

Замечание 3. В используемой схеме решения обратной задачи и в алгоритме 2 процедура обращения фактически используется только при решении уравнений I рода (2.5) (или (3.5)), которые представляют собой некорректно поставленные задачи. Для уравнения вида (2.5) применимы многие устойчивые методы решения линейных некорректно поставленных задач (например, [5], [6], [26]–[28] и др.). В силу возможной неединственности решений этих уравнения мы применяем для реализации ш. 1 алгоритма методы, ориентированные на поиск нормальных решений (тихоновская регуляризация, метод TSVD). В случае неединственного решения уравнения (2.1), т.е. неединственных решений уравнений (3.5), такой подход выделяет из всех возможных решений то, которое имеет минимальную норму “вторичных источников” $v(r, z, \omega) = u(r, \varphi, z, \omega) \xi(r, \varphi, z)$ (см. [4]), и это теоретически позволяет исключить или ослабить появления ложных акустических неоднородностей.

В случае единственности решения оно будет совпадать с вычисленным нормальным решением. Указанные методы регуляризации обоснованы и апробированы в ряде работ (например, в [18], [19] при решении другой обратной задачи). Отметим, что ш. 1 алгоритма является наиболее трудоемким пунктом при использовании алгоритма 2.

4. КОНЕЧНОМЕРНАЯ АППРОКСИМАЦИЯ И РЕШЕНИЕ МОДЕЛЬНЫХ ЗАДАЧ

Везде далее полагается, что уравнения (1.3), (1.4) и их следствия записаны в безразмерном виде с $c_0 = 1$, так что $k_0 = \omega$. Модельные области имеют вид

$$Q = \{(x, y, z): x^2 + y^2 \leq 4^2, |z| \leq 2\}, \quad X = \{(x, y, z): x^2 + y^2 \leq 1, |z| \leq 2\}, \\ Y = \{(x, y, z): 3^2 \leq x^2 + y^2 \leq 4^2, |z| \leq 2\}.$$

Считается также, что $\sigma(\mathbf{x}) = 0$ и модельные источники задаются в форме

$$f(\mathbf{x}) = \sum_{m=1}^{M_s} A_m \delta(\mathbf{x} - \mathbf{x}_m),$$

где \mathbf{x}_m – координаты точечных δ -образных источников. Тогда

$$u_0(\mathbf{x}, \omega) = \sum_{m=1}^{M_s} A_m G_0(\mathbf{x}, \mathbf{x}_m, \omega),$$

и преобразование Фурье этой функции, $\tilde{u}_0(r, \varphi, \Omega; \omega)$, можно заранее вычислить. Мы не ставили себе целью оптимизировать число, положения и амплитуды источников. Во всех расчетах считалось, что $M_s = 8$, $A_m = 1$, $\mathbf{x}_m = (r_m, \varphi_m, z_m)$, где $r_m = 4.01$ и

$$\varphi_m = \left[0, \frac{\pi}{2}, -\frac{\pi}{2}, \pi, 0, \frac{\pi}{2}, -\frac{\pi}{2}, \pi \right], \quad z_m = [-1, -1, -1, -1, 1, 1, 1, 1].$$

Для численного исследования предлагаемых алгоритмов рассмотрены две прямые и обратные модельные задачи.

Задача 1. Задача имеет решение $\xi(x, y, z)$ вида

$$\begin{aligned} \xi(x, y, z) &= \xi_1(x, y, z) + \xi_2(x, y, z), \quad (x, y, z) \in Q; \\ \xi_1(x, y, z) &= \{A_0 \exp\{-30R_1(x, y, z)\}, (x, y, z) \in Q_1; 0, (x, y, z) \in Q \setminus Q_1\}, \\ \xi_2(x, y, z) &= \{2A_0 \exp\{-30R_2(x, y, z)\}, (x, y, z) \in Q_2; 0, (x, y, z) \in Q \setminus Q_2\}, \end{aligned}$$

где

$$\begin{aligned} R_1(x, y, z) &= 5(x - 0.4)^2 + 5y^2 + 0.125(z + 0.1)^2, \\ R_2(x, y, z) &= 5(x + 0.4)^2 + 5(y - 0.4)^2 + 0.125(z - 0.2)^2 \end{aligned}$$

и

$$\begin{aligned} Q_1 &= \{(x, y, z): (x - 0.4)^2 + y^2 + 0.125(z + 0.1)^2 \leq 1.3^2\}, \\ Q_2 &= \{(x, y, z): (x + 0.4)^2 + (y - 0.4)^2 + 0.125(z - 0.2)^2 \leq 0.5^2\}. \end{aligned}$$

Эта функция моделирует небольшие локальные неоднородности среды, положение которых и соответствующие им распределения скоростей нужно найти. Именно на поиск таких неоднородностей “настраивается” алгоритм 2.

Величина A_0 определяет контраст

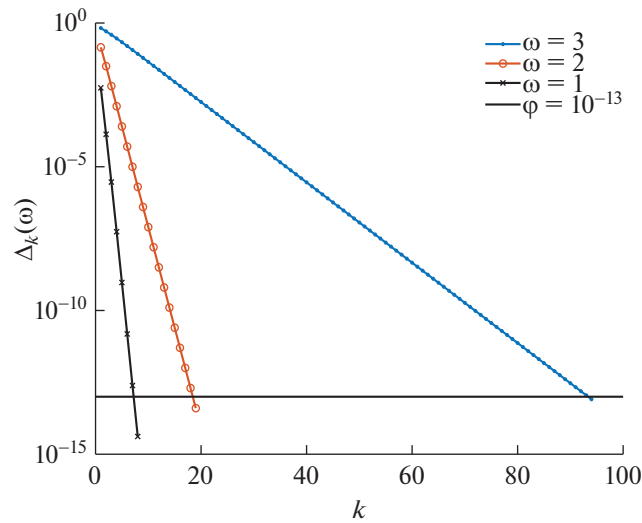
$$\frac{\Delta c}{c_0} = \mathbf{x} \max \left\{ \frac{1}{\sqrt{1 - c_0^2 \xi(\mathbf{x})}} \right\} - 1$$

искомого решения. В расчетах использовалось значение $A_0 = 0.545$, которое соответствует контрасту 15.95. По классификации из [3, с. 33] такой рассеиватель можно считать “сильным”, учитывая его характерные размеры $l \sim 0.2$ (см. фиг. 4а) и значения $c_0 = 1$, $\omega = 3: \frac{\Delta c}{c_0} \gg \frac{c_0}{l\omega}$. Подобные рассеиватели достаточно часто встречаются на практике. Вторая модельная задача будет описана ниже.

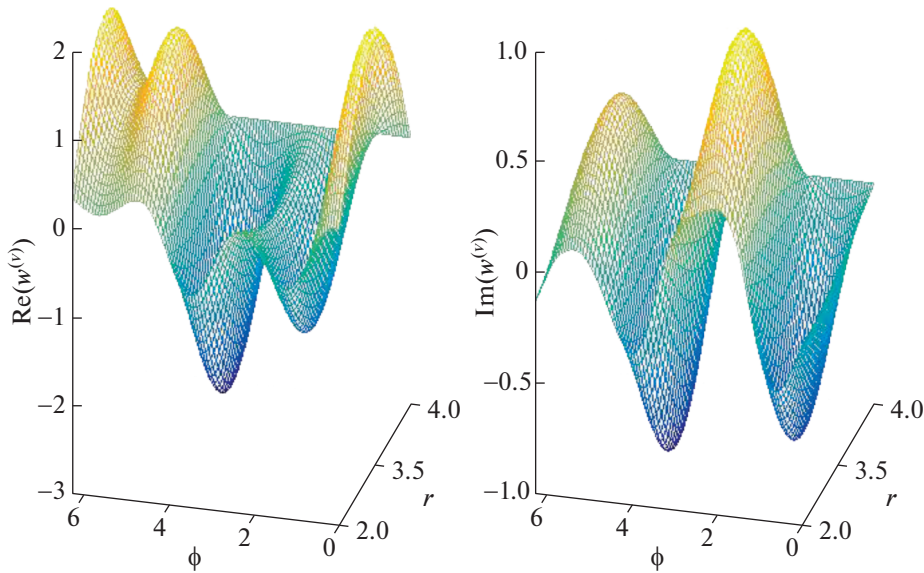
Исходные уравнения (1.3), (1.4) аппроксимировались в областях X, Y конечно-разностным методом на равномерных сетках по переменным r, r', φ, z . Размеры сеток определяются числами $N_r, N_{r'}, N_\varphi, N_z$. В области X сетка имеет размер $N_{r'} \times N_\varphi \times N_z$, а в области Y размер $N_r \times N_\varphi \times N_z$. Конкретные размеры будут указаны ниже для каждого примера. Дискретные аналоги функций $\tilde{G}(r, r', \varphi, \Omega; \omega)$, $\tilde{u}_0(r, \varphi, \Omega; \omega)$, использующиеся в алгоритмах 1 и 2, вычислялись для рассматриваемых частот ω по известным величинам $G_0(r, r', \varphi, z; \omega)$, $u_0(r, \varphi, z; \omega)$ с помощью быстрого преобразования Фурье с сеткой $\{\Omega^{(m)}\}_{m=1}^{N_z}$ по переменной Ω . Разложения в ряды Фурье реализовывались также с помощью БПФ с $n \in [0, N_\varphi - 1]$. Детали этих хорошо известных стандартных вычислений даны, например, в [26].

4.1. Получение модельных данных для решения обратной задачи с помощью алгоритма 1

Приведем типичные результаты численного исследования итерационного процесса (3.1), (3.2) для получения данных первой обратной задачи на сетках размера $N_r = 32$, $N_{r'} = 33$, $N_\varphi = 90$, $N_z = 64$. На фиг. 2 показаны в сравнении скорости сходимости процесса для различных величин $\omega = k_0$.



Фиг. 2. Скорость сходимости итераций (3.1), (3.2) для разных $\omega = k_0$.



Фиг. 3. Типичные данные $w^{(v)}(r, \varphi, 0; \omega = 3)$ для решения обратной задачи.

Останов итераций производился по величине

$$\Delta_k(\omega) = \left\{ \sum_n \|u_n^{(k)}(r, \Omega; \omega) - u_n^{(k-1)}(r, \Omega; \omega)\|_{L_2(\Pi)}^2 \right\}^{1/2} \left\{ \sum_n \|u_n^{(0)}(r, \Omega; \omega)\|_{L_2(\Pi)}^2 \right\}^{-1/2},$$

когда для номера итерации v выполнялось условие $\Delta_v(\omega) \leq 10^{-13}$. Здесь $\Pi = \{(r, \Omega) \in [0, a] \times \mathbb{R}_\Omega\}$. После этого по найденному набору функций $\{u_n^{(v)}(r, \Omega; \omega)\}$ по формуле (3.1) с $k = v$ вычислялись функции $\{v_n^{(v)}(r', \Omega; \omega)\}$, затем по формуле (3.3) находились функции $\{w_n^{(v)}(r, \Omega; \omega)\}$, которые далее преобразовывались, согласно (3.4), в функции $\tilde{w}^{(v)}(r, \varphi, \Omega; \omega)$. Обратное преобразование Фурье последней функции по переменной z , $w^{(v)}(r', \varphi, z; \omega)$, представляет данные для решения обратной задачи. Вид этой функции, найденной для $\omega = k_0 = 3$, показан на фиг. 3 для сечения $z = 0$.

Данные для решения обратной задачи задавались с некоторой ошибкой, которая интерпретируется как ошибка измерения. В наших расчетах это моделировалось наложением на функцию $w^{(v)}(r, \varphi, z; \omega)$ аддитивной нормально распределенной псевдослучайной помехи с нулевым средним так, что получаемая в итоге приближенная функция $w_{\delta}^{(v)}(r, \varphi, z; \omega)$ удовлетворяла бы условию

$$\|w_{\delta}^{(v)}(r, \varphi, z; \omega) - w^{(v)}(r, \varphi, z; \omega)\|_{L_2(Y)} \leq \delta \|w^{(v)}(r, \varphi, z; \omega)\|_{L_2(Y)} \stackrel{\text{def}}{=} \delta_0.$$

Это соответствует приближенному заданию данных с относительной точностью δ .

4.2. Реализация алгоритма 2

Первый шаг алгоритма 2 – решение уравнений I рода (3.5) методами регуляризации – обсуждался в работах [18], [19] в связи с решением другого интегрального уравнения, отличающегося видом ядра и правой части. В этих работах отмечено, что для каждой рассматриваемой частоты ω используемая дискретизация приводит уравнения (3.5) к системе линейных алгебраических уравнений (СЛАУ) вида $A_n^{(m)} V_n^{(m)} = W_n^{(m)}$ для каждого $n \in [0, N_{\varphi} - 1]$, $m \in [1, N_z]$. Здесь $A_n^{(m)} = [\mu_{ij} G_n(r_i, r'_j, \Omega^{(m)}; \omega)]_{i=1, j=1}^{N_r, N_r}$ – матрица системы, получаемая дискретизацией ядра уравнения (3.5) на рассматриваемой сетке размера $N_r \times N_r$, $\Omega^{(m)}$ – точки сетки величин Ω , а μ_{ij} – квадратурные коэффициенты для вычисления интегралов в (3.5). Правые части

$$W_n^{(m)} = \frac{1}{2\pi\omega^2} [w_n(r_i, \Omega^{(m)}; \omega)]_{i=1}^{N_r}$$

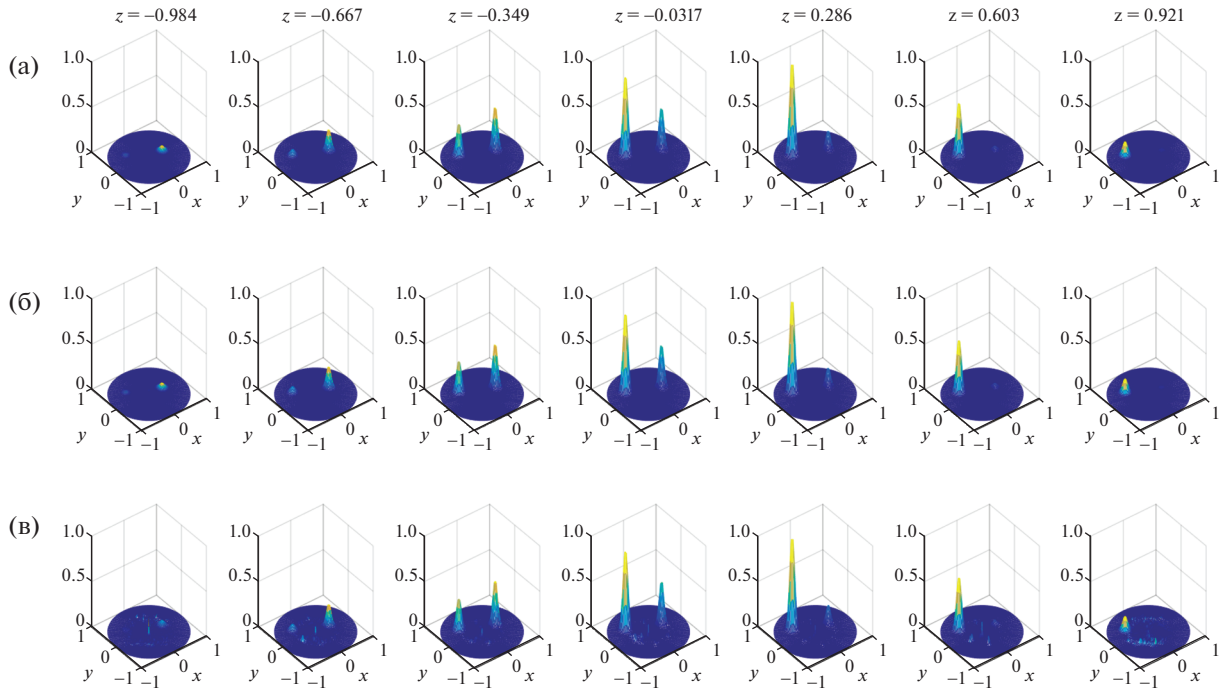
системы суть вектор-столбцы высоты N_r , а искомый вектор-столбец $V_n^{(m)}$ содержит неизвестные $v_n(r'_j, \Omega^{(m)}; \omega)$. Таким образом, при выполнении ш. 1 алгоритма 2 для каждого ω необходимо решить $N_{\varphi} \times N_z$ систем линейных уравнений с матрицами размера $N_r \times N_r$.

Мы решали указанные СЛАУ с помощью различных вариантов метода регуляризации А.Н. Тихонова [26], [27] и с помощью метода TSVD [28]. Обоснование применения этих методов дано с общих позиций теории регуляризации некорректных задач в [18], [19] для аналогичного интегрального уравнения, и здесь мы не повторяем это обоснование. Наилучшие результаты в расчетах получились для метода TSVD с “обрезанием” (см. [28]) сингулярных чисел матриц $A_n^{(m)}$ по уровню $\delta_0^{2/3}$. Эти результаты мы и будем приводить в дальнейшем.

Шаг 2 алгоритма 2 не вызывает сложностей для дискретизированной задачи, так как сводится к матричному умножению дискретных величин G_n , v_n и сложению результата с дискретным аналогом функции u_{0n} . Шаг 3 выполнялся с помощью обратного БПФ. Наконец, шаг 4 при решении обратной задачи для каждой из рассматриваемых частот ω реализовывался с помощью следующей процедуры нахождения нормального псевдорешения уравнения $u\xi = v$ в каждой точке $(r', \varphi', z) \in X$ по методу TSVD: $\xi = \left\{ \frac{v}{u}, |u| > \text{tol}; 0, |u| \leq \text{tol} \right\}$ с $\text{tol} = 10^{-12}$. Далее, при необходимости нетрудно пересчитать функцию $\xi(\mathbf{x})$ в $c(\mathbf{x})$. В приводимых ниже примерах мы для краткости этого не делаем, представляя на рисунках непосредственно величину $\xi(\mathbf{x})$.

4.3. Результаты решения обратной задачи

Первая серия экспериментов связана с решением с помощью предлагаемого алгоритма 2 первой модельной обратной задачи на сетках размера $N_r = 32$, $N_r = 33$, $N_{\varphi} = 90$, $N_z = 64$ при $\omega = k_0 = 3$. Задачи решались с точными (вычисленными по алгоритму 1) данными и приближенными данными с различным уровнем возмущения δ . На фиг. 4 представлены для качественного сравнения точное $\xi_{\text{exact}}(x, y, z)$ и приближенные решения $\xi_{\text{appr}}(x, y, z)$ обратной задачи в различных сечениях по z . Точное решение показано в первой строке рисунка. Во второй строке приведено приближенное решение обратной задачи для точных данных. В третьей строке изображено решение для возмущенных данных с $\delta = 10^{-8}$. Рисунок демонстрирует достаточно высокую чув-



Фиг. 4. Качественное сравнение точного решения $\xi_{\text{exact}}(x, y, z)$ и вычисленных приближенных решений $\xi_{\text{appr}}(x, y, z)$ обратной задачи в различных сечениях $z = \text{const}$; (а) – точное решение; (б) – приближенное решение, полученное для точных данных по алгоритму 2; (в) – приближенное решение для возмущенных данных с $\delta = 10^{-8}$.

ствительность приближенных решений к возмущениям данных. Более детальная информация о точности решения обратной задачи, т.е. о величине относительной ошибки

$$\Delta_{L_2}(z) = \frac{\|\xi_{\text{appr}}(x, y, z) - \xi_{\text{exact}}(x, y, z)\|_{L_2(X_{xy})}}{z \max \|\xi_{\text{exact}}(x, y, z)\|_{L_2(X_{xy})}}$$

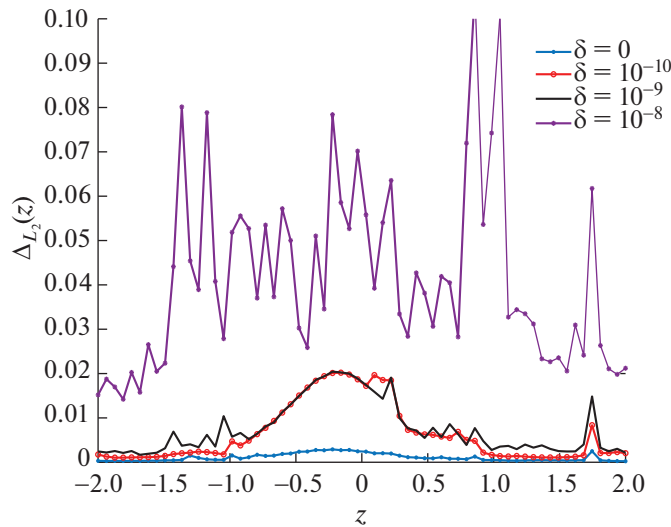
приближенного решения $\xi_{\text{appr}}(x, y, z)$ в слое $z = \text{const}$ от величины z при разных уровнях погрешности данных δ представлены на фиг. 5. Для сравнения на фиг. 6 приведены точности приближенных решений, полученных при разных δ для модельных задач с $\omega = 1$ и $\omega = 2$. Очевидно улучшение точности с возрастанием ω .

При решении рассматриваемой обратной задачи весьма важно, насколько точно алгоритм позволяет определять положения изучаемых локальных рассеивателей. Для иллюстрации была решена

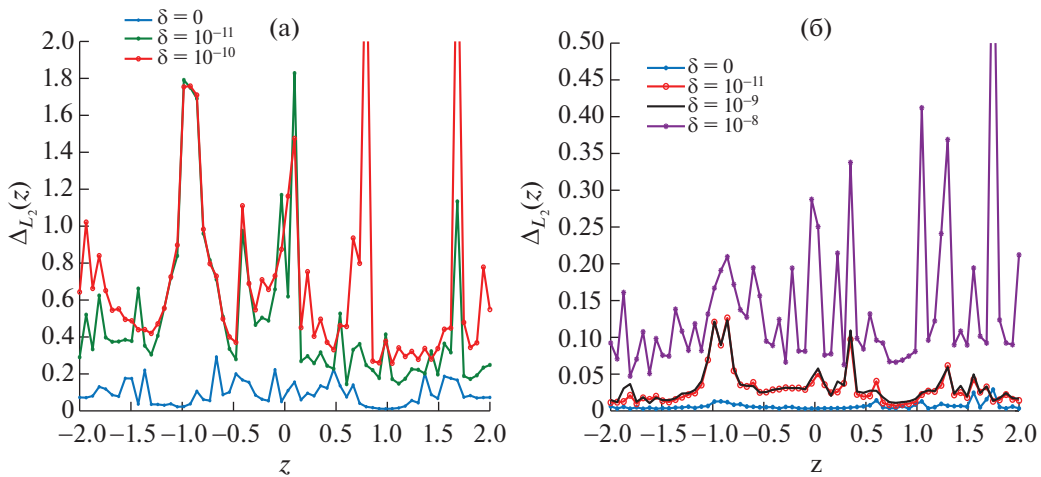
Задача 2. Задача отличается от первой лишь выражениями для $\xi_1(x, y, z)$ и $\xi_2(x, y, z)$:

$$\begin{aligned} \xi_1(x, y, z) &= \{A_0; (x, y, z) \in Q_1; 0, (x, y, z) \in Q \setminus Q_1\}, \\ \xi_2(x, y, z) &= \{2A_0; (x, y, z) \in Q_2; 0, (x, y, z) \in Q \setminus Q_2\}, \quad A_0 = 0.2. \end{aligned}$$

Это соответствует двум рассеивателям эллипсоидальной формы, лежащим в области X и заполненным веществом с различными показателями преломления. Фигура 7 показывает качественно влияние возмущения данных задачи на определение положения и геометрии неоднородностей. Контраст точного решения здесь равен 0.291. Видно, что положения могут быть достаточно уверенно определены в случае $\delta = 10^{-9}$, а при использовании подходящей фильтрации шумов в найденном решении и для $\delta = 10^{-8}$.



Фиг. 5. Первая модельная задача. Относительная точность $\Delta_{L_2}(z)$ приближенных решений для различных z при разных уровнях возмущения данных δ для $\omega = 3$.

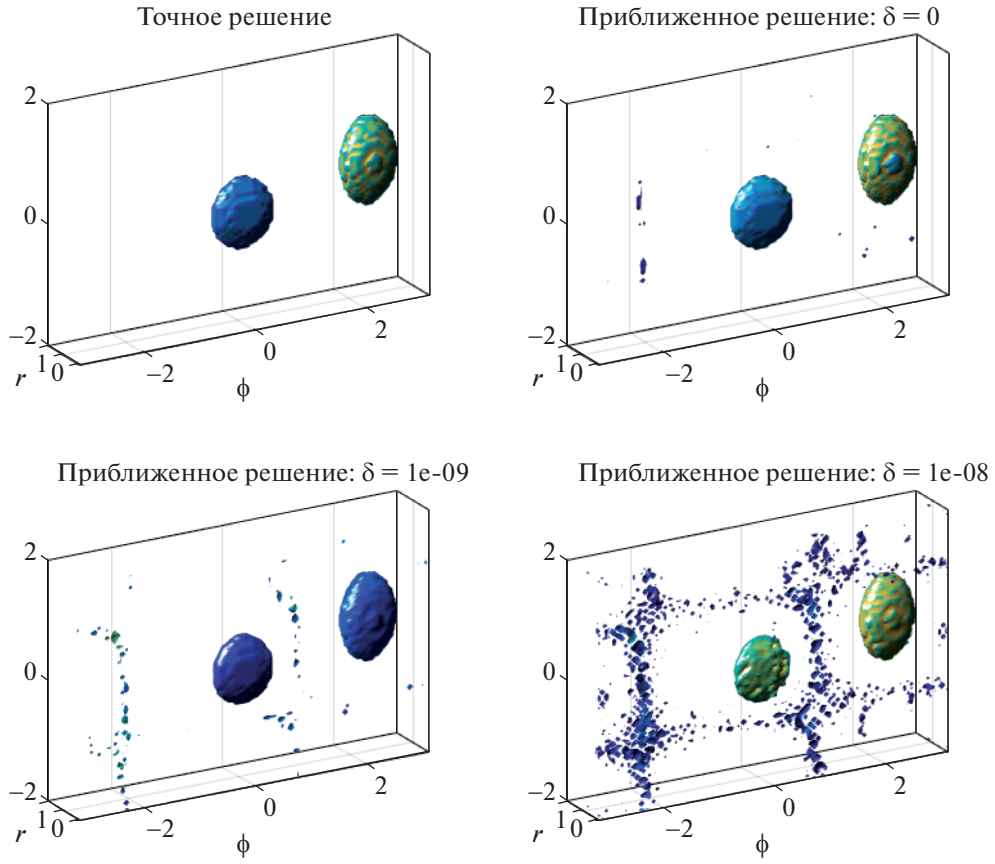


Фиг. 6. Первая модельная задача. Относительная точность $\Delta_{L_2}(z)$ приближенных решений для различных z при разных уровнях возмущения данных δ ; (а) – для $\omega = 1$; (б) – для $\omega = 2$.

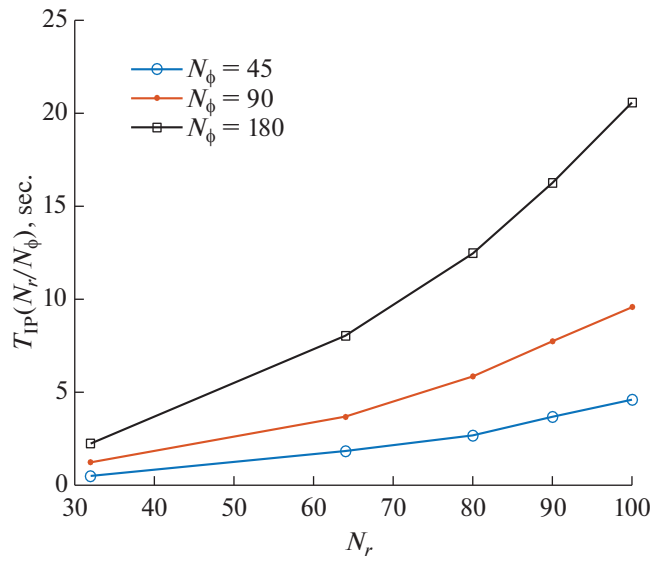
5. НЕКОТОРЫЕ СВОЙСТВА АЛГОРИТМА 2

Все вычисления проводились в системе МАТЛАБ на ПК с процессором Intel (R) Core (TM) i7-7700 CPU 3.60 GHz, ОЗУ 16 Гб (без распараллеливания). Алгоритм 2 решения обратной задачи оказался достаточно быстрым. Приведем данные соответствующих численных экспериментов для решения первой обратной задачи на частоте $\omega = 2$. В экспериментах считалось, что сетка по z фиксирована ($N_z = 64$), а меняются только размеры N_r, N_φ сеток по r и φ . Кроме того, полагалось, что $N_r = N_r + 1$. Тогда время решения обратной задачи есть функция вида $T_{IP}(N_r, N_\varphi)$. Эта зависимость приведена на фиг. 8. При изменении размера сетки по z время $T_{IP}(N_r, N_\varphi)$ меняется пропорционально числу N_z , так как оно определяется решением $N_z \times N_\varphi$ уравнений вида (3.5).

Еще раз отметим, что решаемая обратная задача весьма чувствительна к ошибкам входных данных. При решении интегрального уравнения (1.3) с двойной точностью внесение в правую часть эквивалентной системы уравнения (3.5) случайных ошибок с амплитудой порядка 10^{-8}



Фиг. 7. Вторая модельная задача с $\omega = 3$. Качественное сравнение положений и геометрии восстанавливаемой неоднородности $\xi(x)$ для различных δ .



Фиг. 8. Время решения обратной задачи $T_{IP}(N_r, N_\phi)$ для различных N_r, N_ϕ .

приводит к серьезным искажениям решения при использовании и метода TSVD, и метода регуляризации. Это связано с весьма быстрым убыванием сингулярных чисел матриц $A_n^{(m)}$ для СЛАУ, решаемых на ш. 1 алгоритма 2, и такое убывание является специфической особенностью решаемой обратной задачи, а точнее, интегрального уравнения (1.3). Аналогичное свойство обратной коэффициентной задачи для волнового уравнения отмечалось ранее в работах [18], [19]. Подтверждающие это соответствующие теоретические оценки устойчивости при различных априорных предположениях на точное решение можно найти в [5], [6].

6. ВЫВОДЫ

В этой работе мы не стремились построить наилучший по всем характеристикам численный алгоритм решения обратной задачи (1.2). В рамках именно этой модели, с учетом цилиндрической геометрии областей решений и наблюдений, мы предложили некоторый алгоритм, который позволяет быстро решать обратную задачу на достаточно мелких сетках. Исследуя свойства алгоритма в численных экспериментах, представленных в работе, мы можем сделать следующие выводы.

1. Рассматриваемая трехмерная обратная задача скалярной акустики в цилиндрической области может быть решена численно с помощью предлагаемого алгоритма для достаточно мелких сеток за время порядка десятков секунд на ПК средней производительности даже без распараллеливания. Для этого следует использовать указываемую в статье схему регистрации данных обратной задачи в цилиндрическом слое. Предлагаемый алгоритм легко распараллеливается.

2. Исходная обратная задача, рассматриваемая для конечного набора частот, имеет, вообще говоря, неединственное решение. Для выделения определенного решения обратной задачи в нашем алгоритме используется поиск нормального решения интегрального уравнения (1.3). В случае, когда уравнение (1.3) имеет единственное решение при рассматриваемом наборе частот, это решение совпадает с найденным нормальным.

3. Рассматриваемая обратная задача, основанная на уравнениях (1.2) и решаемая отдельно на каждой частоте, сама по себе весьма чувствительна к возмущениям данных: для получения детального приближенного решения требуются данные, измеренные с большой точностью. Эта особенность задачи связана с видом ядра интегрального уравнения (1.3), его сингулярными числами и не зависит от используемого алгоритма решения этого уравнения. Устойчивость исходной обратной задачи можно повышать, вводя дополнительные априорные ограничения на решения.

4. Алгоритм 2 позволяет достаточно надежно определять положения и форму небольших локальных неоднородностей акустической среды при данных с малыми ошибками.

Авторы благодарны А.Г. Яголе за полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Ramm A.G.* Multidimensional Inverse Scattering Problems. Pitman Monogr. Surv. Pure Appl. Math. 51. Harlow: Longman Scientific & Technical, 1992.
2. *Colton D., Kress R.* Inverse Acoustic and Electromagnetic Scattering Theory. 2nd ed. Appl. Math. Sci. 93. Berlin: Springer, 1998.
3. *Горюнов А.А., Сасковец А.В.* Обратные задачи рассеяния в акустике. М.: Изд-во МГУ, 1989.
4. *Буров В.А., Румянцева О.Д.* Обратные волновые задачи акустической томографии. Ч. 2. Обратные задачи акустического рассеяния. М.: ЛЕНАНД, 2020.
5. *Bakushinsky A., Goncharsky A.* Ill-Posed Problems: Theory and Applications. Dordrecht: Kluwer Academic Publishers, 1994.
6. *Bakushinsky A.B., Kokurin M.Yu.* Iterative methods for approximate solution of inverse problems. Mathematics and Its Applications. Dordrecht: Kluwer Academic Publishers, 2004.
7. *Гончарский А.В., Романов С.Ю.* О двух подходах к решению коэффициентных обратных задач для волновых уравнений // Ж. вычисл. матем. и матем. физ. 2012. Т. 52. № 2. С. 263–269.
8. *Гончарский А.В., Романов С.Ю.* Суперкомпьютерные технологии в разработке методов решения обратных задач в УЗИ-томографии // Вычисл. методы и программирование: новые вычисл. технологии. 2012. Т. 13. № 1. С. 235–238.
9. *Belishev M.I.* Recent progress in the boundary control method // Inverse Problems. 2007. V. 23. № 5. P. 1–67.

10. *Пестов Л.Н., Болгова В.М., Данилин А.Н.* Численная реконструкция трехмерной скорости звука методом граничного управления // Вестн. Югорского государственного университета. 2011. Вып. 3. С. 92–98.
11. *Буров В.А., Алексеев Н.В., Румянцев О.Д.* Многочастотное обобщение алгоритма Новикова для решения обратной двумерной задачи рассеяния // Акустический ж. 2009. Т. 55. № 6. С. 784–798.
12. *Новиков Р.Г.* Восстановление двумерного оператора Шредингера по амплитуде рассеяния при фиксированной энергии // Функцион. анализ и его прил. 1986. Т. 20. № 3. С. 90–91.
13. *Буров В.А., Вечерин С.Н., Морозов С.А., Румянцев О.Д.* Моделирование точного решения обратной задачи акустического рассеяния функциональными методами // Акустический ж. 2010. Т. 56. № 4. С. 516–536.
14. *Beilina L., Klibanov M.V.* Approximate Global Convergence and Adaptivity for Coefficient Inverse Problems. New York: Springer, 2012.
15. *Kabanikhin S.I., Satybaev A.D., Shishlenin M.A.* Direct Methods of Solving Multidimensional Inverse Hyperbolic Problems. Utrecht: VSP, 2004.
16. *Klibanov M.V., Kolesov A.E.* Convexification of a 3D coefficient inverse scattering problem // Computers and Mathematics with Applications. 2019. V. 77. P. 1681–1702.
17. *Klibanov M.V., Kolesov A.E., Nguyen Dinh-Liem.* Convexification method for an inverse scattering problem and its performance for experimental backscatter data for buried targets // SIAM J. Imaging Sciences. 2019. V. 12. № 1. P. 576–603.
18. *Vakushinsky A.V., Leonov A.S.* Fast numerical method of solving 3D coefficient inverse problem for wave equation with integral data // J. of Inverse and Ill-Posed Problems. 2018. V. 26. Issue 4. P. 477–492.
19. *Бакушинский А.Б., Леонов А.С.* Экономичный численный метод решения коэффициентной обратной задачи для волнового уравнения в трехмерном пространстве // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 4. С. 561–574.
20. *Бакушинский А.Б., Леонов А.С.* Численное решение трехмерной коэффициентной обратной задачи для волнового уравнения с интегральными данными в цилиндрической области // Сиб. ж. вычисл. матем. 2019. Т. 22. № 4. С. 381–396.
21. *Бакушинский А.Б., Леонов А.С.* К численному решению обратной многочастотной задачи скалярной акустики // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 6. С. 1013–1026.
22. *Евстигнеев Р.О., Медведик М.Ю., Смирнов Ю.Г., Цупак А.А.* Обратная задача восстановления неоднородностей тела для ранней диагностики заболеваний с помощью микроволновой томографии // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. 2017. Т. 44. № 4. С. 3–17.
23. *Будак Б.М., Самарский А.А., Тихонов А.Н.* Сборник задач по математической физике. М.: Наука, 1972.
24. *Владимиров В.С.* Обобщенные функции в математической физике. М.: Наука, 1976.
25. *Рисс Ф., Сёкефальви-Надь Б.* Лекции по функциональному анализу. М.: Мир, 1979.
26. *Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г.* Численные методы решения некорректных задач. М.: Наука, 1990.
27. *Леонов А.С.* Решение некорректно поставленных обратных задач: Очерк теории, практические алгоритмы и демонстрации в МАТЛАБ. Изд. 2. М.: Либроком, 2013.
28. *Engl H.W., Hanke M., Neubauer A.* Regularization of Inverse Problems. Dordrecht: Kluwer Academic Publishers, 1996.

НАХОЖДЕНИЕ СТАЦИОНАРНЫХ ТЕЧЕНИЙ ПУАЗЕЙЛЕВСКОГО ТИПА ДЛЯ НЕСЖИМАЕМОЙ ПОЛИМЕРНОЙ ЖИДКОСТИ МЕТОДОМ УСТАНОВЛЕНИЯ¹⁾

© 2022 г. **А. М. Блохин**^{1,2,*}, **Б. В. Семисалов**^{1,2,**}

¹ 630090 Новосибирск, ул. Пирогова, 1, НГУ, Россия

² 630090 Новосибирск, пр-т Коптюга, 4, ИМ СО РАН, Россия

*e-mail: blokhin@math.nsc.ru

**e-mail: vibis@ngs.ru

Поступила в редакцию 20.08.2020 г.
Переработанный вариант 20.08.2020 г.
Принята к публикации 17.09.2021 г.

Проведен численный анализ процесса установления стационарных течений несжимаемой вязкоупругой полимерной жидкости в канале с прямоугольным сечением под действием постоянного перепада давления. Для описания течений применяется реологическая мезоскопическая модель Покровского–Виноградова. При использовании интерполяции с узлами Чебышёва по пространственным переменным и неявной схемы по времени разработан алгоритм решения начально-краевых задач для нестационарных уравнений модели. Аналитически показано, что в стационарном случае модель допускает три решения высокой гладкости. Вопрос о том, какое из этих решений реализуется на практике, исследован с помощью расчетов предельного решения нестационарных уравнений. Установлено, что предельное решение с высокой точностью совпадает с одним из трех решений стационарной задачи, и рассчитаны значения параметров, при которых происходит переключение с одного решения на другое. Библ. 11. Фиг. 6. Табл. 3.

Ключевые слова: полимерная жидкость, мезоскопическая реологическая модель, устойчивое течение Пуазейля, метод без насыщения, переключение установившегося решения.

DOI: 10.31857/S0044466922020065

1. ВВЕДЕНИЕ

Работа посвящена численному исследованию устойчивости стационарных режимов течения несжимаемой вязкоупругой полимерной жидкости в канале с прямоугольным поперечным сечением. Такие течения по своим качественным особенностям схожи с решениями Пуазейля уравнений Навье–Стокса, для которых в некоторых простых случаях (например, течение в цилиндрической трубе) известны строгие результаты об экспоненциальной сходимости нестационарных решений к стационарным (см. [1]), что гарантирует устойчивость последних.

В отличие от ньютоновской жидкости течения растворов и расплавов полимерных материалов демонстрируют сложную градиентную зависимость элонгационной и сдвиговой вязкостей от скоростей деформации, эффекты запаздывания и зависимость потока от ориентации и размеров макромолекул (см. [2]–[4]). Вследствие этого получить в аналитическом виде решения уравнений, описывающих течения полимерной жидкости типа Пуазейля, и исследовать их устойчивость не представляется возможным. Далее для описания таких течений использована реологическая мезоскопическая модель Покровского–Виноградова (см. [2], [5]). Анализ квазилинейного уравнения, полученного в [6] для расчета стационарных течений в каналах под действием постоянного перепада давления, показал, что его коэффициенты и правая часть определяются неоднозначно. При определенных условиях это приводит к возникновению трех качественно различных решений высокой гладкости. В данной работе численно исследуется вопрос о том, какое из этих решений является предельным для исходной нестационарной постановки, т.е. какое из них может быть реализовано на практике.

¹⁾ Работа выполнена при финансовой поддержке РФФ (соглашение № 20-11-20036).

Для этого в статье разработан алгоритм, основанный на приближениях без насыщения и подходе к аппроксимации двумерных краевых задач для уравнений эллиптического типа, предложенном в [7]. При использовании этого алгоритма в совокупности с неявной схемой по временной переменной и задании начальных данных определенного вида получено предельное решение нестационарного уравнения, которое с высокой точностью совпадает с одним из трех решений стационарной задачи. Причем при определенных значениях параметров реализуется “переключение” предельного течения с одного решения стационарной задачи на другое. Таким образом, удалось прояснить структуру коэффициентов и правой части уравнения, описывающего устойчивые стационарные течения.

Отметим, что выход на полученный стационарный режим существенно зависит от начальных данных — распределения скорости течения в канале и его согласованности с граничными условиями. Если начальные значения скорости слишком велики или не согласованы с граничными условиями, стационарное течение не реализуется.

2. ПОСТАНОВКА ЗАДАЧИ

Следуя монографиям [1]–[3], [8], [9] и статье [5], запишем уравнения *реологической мезоскопической модели Покровского–Виноградова*, которая описывает течения несжимаемой вязкоупругой полимерной жидкости. Эти уравнения в безразмерной форме в прямоугольной декартовой системе координат (x_1, x_2, x_3) выглядят следующим образом:

$$\operatorname{div} \mathbf{u} = 0, \quad (1)$$

$$\frac{d\mathbf{u}}{dt} + \nabla P = \frac{1}{\operatorname{Re}} \operatorname{div}(\Pi), \quad (2)$$

$$\frac{da_{ij}}{dt} - \sum_{l=1}^3 \frac{\partial u_l}{\partial x_l} a_{ij} - \sum_{l=1}^3 \frac{\partial u_j}{\partial x_l} a_{li} - \frac{1}{W} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \mathcal{L}_{ij} = 0, \quad i = 1, 2, 3, \quad j = i, \dots, 3. \quad (3)$$

Здесь t — время; u_1, u_2, u_3 — компоненты вектора скорости \mathbf{u} ;

— P — давление;

— $a_{ij}, i, j = 1, 2, 3$, — компоненты симметрического тензора анизотропии Π второго ранга;

— $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ — столбцы симметрической матрицы $\Pi = (a_{ij}) = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$;

— $\|\mathbf{a}_i\|^2 = (\mathbf{a}_i, \mathbf{a}_i), i = 1, 2, 3$;

— $\operatorname{div}(\Pi) = (\operatorname{div}(\mathbf{a}_1), \operatorname{div}(\mathbf{a}_2), \operatorname{div}(\mathbf{a}_3))^T$;

— $\mathcal{L}_{ij} = \left(\frac{a_{ij}}{W} + \frac{\bar{k}(a_{11} + a_{22} + a_{33})}{3} a_{ij} + \beta(\mathbf{a}_i, \mathbf{a}_j) \right), i, j = 1, 2, 3$;

— $\bar{k} = k - \beta, k, \beta (0 < \beta < 1)$ — феноменологические параметры модели, характеризующие вклады, связанные с анизотропией (величина β учитывает ориентацию макромолекулярного клубка, число k — его размеры, см. [5], [10]);

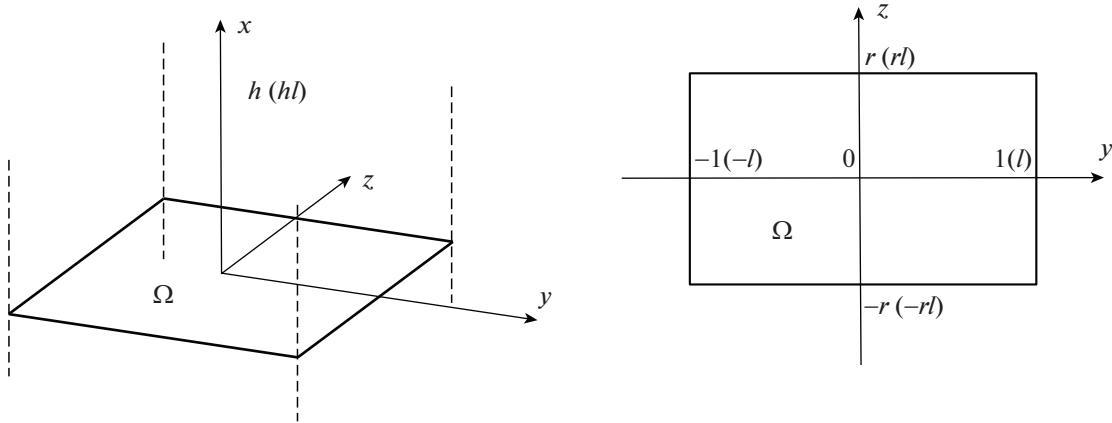
— $\operatorname{Re} = \frac{\rho u_H l}{\eta_0^*}$ — число Рейнольдса; $\rho (= \text{const})$ — плотность среды;

— $W = \frac{\tau_0^* u_H}{l}$ — число Вейсенберга;

— η_0^*, τ_0^* — начальные значения сдвиговой вязкости и времени релаксации при $T = T_0$, где T_0 — температура окружающей среды.

Далее для координат будем использовать обозначения x, y, z ($x = x_1, y = x_2, z = x_3$).

Замечание 1. В [5], [10] рассмотрены случаи $k = \beta, k = 1.2\beta$. В настоящей работе исследуется более общий случай: $k = c_k \beta$, где $c_k > 0$ — новый феноменологический параметр.



Фиг. 1. Канал с прямоугольным сечением Ω .

Система (1)–(3) записана в безразмерном виде: время t ; координаты x, y, z ; компоненты вектора скорости u_1, u_2, u_3 ; давление P ; компоненты тензора анизотропии a_{ij} отнесены к $\frac{l}{u_H}$; l ; u_H ; ρu_H^2 ; $W/3$ соответственно, где l – характерная длина, u_H – характерная скорость.

Пусть полимерная жидкость течет в канале Ω с прямоугольным поперечным сечением, лежащим в плоскости (y, z) (фиг. 1):

$$\Omega = \{(y, z) : -1 \leq y \leq 1, -r \leq z \leq r\}.$$

В канале вдоль оси x действует перепад давления $\Delta P(t)$. Будем искать *частное решение* исходной системы (1)–(3) следующего вида:

$$\begin{aligned} u_2 = u_3 &\equiv 0, & u_1 &= u(t, y, z), \\ P &= \mathcal{P}(t, y, z) - A(t)x, \\ a_{ij} &= a_{ij}(t, y, z), & i, j &= 1, 2, 3. \end{aligned} \tag{4}$$

Величина $A(t) = \frac{\Delta P(t)}{\rho u_H^2 h}$ есть безразмерный перепад давления на отрезке по x длины h , $\mathcal{P}(t, y, z)$ – компонента давления, зависящая только от времени и от координаты точки сечения.

Пусть

$$\begin{aligned} \alpha_{ij} &= \frac{a_{ij}}{\text{Re}}, & i, j &= 1, 2, 3, & \alpha_i &= \alpha_{ii} + \alpha^2, & \alpha^2 &= \frac{1}{W \text{Re}}; \\ K_I &= \text{Re} \left(\alpha^2 + \frac{\bar{k}}{3} I \right), & I &= \alpha_{11} + \alpha_{22} + \alpha_{33}, & \tilde{K}_I &= K_I + \beta \text{Re } I. \end{aligned}$$

Тогда с учетом (4) запишем систему (1)–(3) в виде двух систем:

$$\begin{aligned} u_t - (\alpha_{12})_y - (\alpha_{13})_z &= A(t), \\ (\alpha_{12})_t - \alpha_2 u_y - \alpha_{23} u_z + \tilde{K}_I \alpha_{12} + \beta \text{Re}(\alpha_{13} \alpha_{23} - \alpha_{12} \alpha_{33}) &= 0, \\ (\alpha_{13})_t - \alpha_3 u_y - \alpha_3 u_z + \tilde{K}_I \alpha_{13} + \beta \text{Re}(\alpha_{12} \alpha_{23} - \alpha_{13} \alpha_{22}) &= 0, \end{aligned} \tag{5}$$

$$\begin{aligned} (\alpha_{11})_t - 2(\alpha_{12} u_y + \alpha_{13} u_z) + K_I \alpha_{11} + \beta \text{Re}(\alpha_{11}^2 + \alpha_{12}^2 + \alpha_{13}^2) &= 0, \\ (\alpha_{22})_t + \tilde{K}_I \alpha_{22} + \beta \text{Re}(\alpha_{12}^2 - \alpha_{11} \alpha_{22} + \alpha_{23}^2 - \alpha_{22} \alpha_{33}) &= 0, \\ (\alpha_{33})_t + \tilde{K}_I \alpha_{33} + \beta \text{Re}(\alpha_{13}^2 - \alpha_{11} \alpha_{33} + \alpha_{23}^2 - \alpha_{22} \alpha_{33}) &= 0, \\ (\alpha_{23})_t + \tilde{K}_I \alpha_{23} + \beta \text{Re}(\alpha_{12} \alpha_{13} - \alpha_{11} \alpha_{23}) &= 0. \end{aligned} \tag{6}$$

Для определения давления получаем следующие выражения:

$$\mathcal{P}_y = (\alpha_{22})_y + (\alpha_{23})_z, \quad \mathcal{P}_z = (\alpha_{23})_y + (\alpha_{33})_z.$$

Дифференцируя первое уравнение (5) по t , второе уравнение по y и третье уравнение по z , а затем складывая их, приходим к уравнению второго порядка для функции $u(t, y, z)$:

$$u_{tt} - \tilde{\Delta}u - \tilde{A}(t, y, z)u_y - \tilde{B}(t, y, z)u_z + \tilde{K}_I[u_t - A(t)] + F(t, y, z) = A'(t), \quad (7)$$

где

$$\begin{aligned} \tilde{\Delta}u &= \alpha_2 u_{yy} + 2\alpha_{23} u_{yz} + \alpha_3 u_{zz}, & \tilde{A}(t, y, z) &= (\alpha_2)_y + (\alpha_{23})_z, & \tilde{B}(t, y, z) &= (\alpha_{23})_y + (\alpha_3)_z, \\ F(t, y, z) &= (\tilde{K}_I)_y \alpha_{12} + (\tilde{K}_I)_z \alpha_{13} + \beta \operatorname{Re}[(l_{22})_z + (l_{33})_y], \\ l_{22} &= \alpha_{12} \alpha_{23} - \alpha_{13} \alpha_{22}, & l_{33} &= \alpha_{13} \alpha_{23} - \alpha_{12} \alpha_{33}. \end{aligned}$$

Дополним уравнения (5)–(7) граничными условиями (условиями прилипания жидкости для скорости)

$$u(t, y, z) = 0 \quad \text{при} \quad t \geq 0, \quad (y, z) \in \partial\Omega, \quad (8)$$

и начальными данными в предположении, что в начальный момент времени $t = 0$ жидкость в канале покоится:

$$\alpha_{ij}(0, y, z) = 0, \quad u(0, y, z) = 0, \quad u_t(0, y, z) = 0 \quad \forall (y, z) \in \Omega. \quad (9)$$

3. СТАЦИОНАРНЫЙ СЛУЧАЙ

Следуя [5] (см. также [6]), заметим, что при реализации условий (4) в стационарном случае, когда переменные задачи не зависят от времени: $u = u(y, z)$, $a_{ij} = a_{ij}(y, z)$, $i, j = \overline{1, 3}$, $\mathcal{P} = \mathcal{P}(y, z)$, $A = \text{const}$, система уравнений (1)–(3) в канале с прямоугольным сечением сводится к одному разрешающему уравнению для определения продольной скорости течения:

$$\hat{a}u_{yy} - 2\hat{b}u_{yz} + \hat{c}u_{zz} = -\operatorname{Re} A \tilde{\mathcal{K}}, \quad (10)$$

где

$$\begin{aligned} \hat{a} &= 1 - u_y^2 \mathcal{L}, & \hat{c} &= 1 - u_z^2 \mathcal{L}, & \hat{b} &= u_y u_z \mathcal{L}, \\ \mathcal{L} &= \frac{1 - \Delta}{\lambda^2}, & \Delta &= \frac{\lambda}{\mu(\tilde{\mathcal{K}}_\mu \mu + \tilde{\mathcal{K}})}, & \lambda &= \sqrt{u_y^2 + u_z^2}, \\ \tilde{\mathcal{K}} &= \frac{1 + W\sigma + 2W(\bar{k}/3 + \beta)(\sigma(1 + W\sigma) + W\mu^2)}{(1 + W\sigma)^2}, & \sigma &= a_{22} + a_{33}, & \mu &= \sqrt{a_{12}^2 + a_{13}^2}, \\ \tilde{\mathcal{K}}_\mu &= \frac{W\sigma_\mu}{(1 + W\sigma)^2} [2(\bar{k}/3 + \beta) - 1] + \frac{4W^2\mu}{(1 + W\sigma)^3} (\bar{k}/3 + \beta) [W\sigma - W\mu\sigma_\mu + 1]. \end{aligned}$$

Используя формулы для $\tilde{\mathcal{K}}$ и выражение $\sigma(\mu)$, величину μ можно определить как решение нелинейного уравнения

$$\tilde{\mathcal{K}}(\mu)\mu = \lambda.$$

Функция $\sigma = \sigma(\mu)$ является решением кубического уравнения

$$\begin{aligned} a\sigma^3 + b\sigma^2 + c\sigma + d &= 0, & \sigma &\neq -1/W, \\ a &= \tilde{\beta}W, & b &= \tilde{\beta} + 1, & c &= W^{-1} + W\mu^2\tilde{\beta}, & d &= \beta\mu^2, & \tilde{\beta} &= \frac{2\bar{k}}{3} + \beta. \end{aligned} \quad (11)$$

Анализ этого уравнения, проведенный в [6], привел к заключению, что в прямоугольной области при достаточно малых градиентах давления A это уравнение имеет три действительных ре-

шения – функции $\sigma_1(\mu)$, $\sigma_2(\mu)$, $\sigma_3(\mu)$. Приведем далее выражения для этих решений. Следуя формулам Кардано для решения кубических уравнений, введем обозначения

$$p = \frac{3ac - b^2}{3a^2}, \quad q = \frac{2b^3 - 9abc + 27a^2d}{27a^3}, \quad Q = (p/3)^3 + (q/2)^2.$$

После преобразований получаем $p = \mu^2 - A$, $q = B\mu^2 + C$, где

$$A = \frac{1 - \tilde{\beta} + \tilde{\beta}^2}{3\tilde{\beta}^2 W^2}, \quad B = \frac{3\beta - \tilde{\beta} - 1}{3W\tilde{\beta}}, \quad C = \frac{(1 + \tilde{\beta})(\tilde{\beta} - 2)(2\tilde{\beta} - 1)}{27W^3\tilde{\beta}^3}.$$

Указанные три решения имеют вид

$$\sigma_1 = 2 \cos \alpha \left| \sqrt[6]{\frac{q^2}{4} - Q} \right| - \frac{b}{3a}, \quad \sigma_{2,3} = - \left| \sqrt[6]{\frac{q^2}{4} - Q} \right| (-\cos \alpha \pm \sqrt{3} \sin \alpha) - \frac{b}{3a}, \quad \alpha = \frac{\arctan(2\sqrt{-Q}/q)}{3}.$$

При этом выражения для производных $\frac{d\sigma_{1,2,3}}{d\mu}$, фигурирующих в формуле для $\tilde{\mathcal{K}}_\mu$, имеют вид

$$\begin{aligned} \frac{\partial \sigma_1}{\partial \mu} = \sigma'_1 &= \frac{2^{2/3}}{3\sqrt{-Q} |(q^2 - 4Q)^{5/6}} ((q \sin \alpha - 2\sqrt{-Q} \cos \alpha)Q' + (q\sqrt{-Q} \cos \alpha - 2Q \sin \alpha)q'), \\ \frac{\partial \sigma_{2,3}}{\partial \mu} = \sigma'_{2,3} &= \frac{-1}{3^3 \sqrt{2} \sqrt{-Q} |(q^2 - 4Q)^{5/6}} (\pm(\sqrt{3} \cos \alpha \pm \sin \alpha)(2Qq' - qQ') + \\ &+ (\cos \alpha \mp \sqrt{3} \sin \alpha)(2Q' - qq')\sqrt{-Q}), \end{aligned}$$

где для σ_2 выбираем верхний из знаков “±”, “∓”, для σ_3 – нижний;

$$q' = \frac{\partial q}{\partial \mu} = -2\mu \frac{1 - 3\beta + \tilde{\beta}}{3\tilde{\beta}W},$$

$$\begin{aligned} Q' = \frac{\partial Q}{\partial \mu} &= \frac{\mu}{27\tilde{\beta}^4 W^4} \{ 27\beta^2 \tilde{\beta}^2 \mu^2 W^2 + \tilde{\beta}(1 + \tilde{\beta} \mu^2 W^2) [\tilde{\beta}(\tilde{\beta}(6\mu^2 W^2 - 1) + 4) - 1] - \\ &- \beta(1 + \tilde{\beta}) [\tilde{\beta}(2\tilde{\beta}(9\mu^2 W^2 - 1) + 5) - 2] \}. \end{aligned}$$

Стационарное уравнение (10) дополняется граничными условиями

$$u(y, z) = 0 \quad \text{при} \quad (y, z) \in \partial\Omega. \tag{12}$$

В [5] описан алгоритм решения краевой задачи (10), (12), основанный на методе установления, приближениях без насыщения и методе коллокаций (см. также [7]). Этот алгоритм использован в настоящей работе для поиска решений стационарной задачи.

Возникает естественный вопрос: какое из трех возможных решений задачи (10), (12) реализуется на практике? Нестационарная формулировка модели, полученная в разд. 2, имеет целью исследование этого вопроса. Для этого в модели будет задан постоянный градиент давления и реализованы вычисления вплоть до выхода на стационарный режим течения при некотором $t = t_s$. Сопоставление скорости $u(t_s, y, z)$ нестационарного течения в момент времени $t = t_s$ и решения задачи (10), (12), полученного при использовании разных зависимостей $\sigma(\mu)$, покажет, какой режим реализуется на практике.

4. ОПИСАНИЕ ВЫЧИСЛИТЕЛЬНОГО АЛГОРИТМА

Для реализации расчетов введем сетку по времени с шагом τ и узлами $t_n = n\tau$, $n = 0, 1, \dots$. Обозначим $u^n = u^n(y, z) = u(t_n, y, z)$, $\alpha^n_{ij} = \alpha^n_{ij}(y, z) = \alpha_{ij}(t_n, y, z)$ и аппроксимируем в уравнениях (5)–(7) производные по времени разностными отношениями вида

$$u_t \approx \frac{u^{n+1} - u^n}{\tau}, \quad u_{tt} \approx \frac{u^{n+1} - 2u^n + u^{n-1}}{\tau^2}, \quad n = 1, 2, \dots$$

4.1. Линеаризация уравнений модели

Для реализации вычислений необходимо линеаризовать по Ньютону уравнения (5)–(7) относительно неизвестных функций α_{ij} и u и на каждом шаге по времени осуществлять итерации по нелинейности. В расчетах невязка итераций по нелинейности была установлена равной 10^{-10} . Запишем линеаризованные уравнения для функций α_{ij} в виде

$$\begin{aligned} \frac{\alpha_{11}^{n+1} - \alpha_{11}^n}{\tau} &= -\operatorname{Re} \left[\left(\frac{\bar{k}}{3} + 2\beta \right) \alpha_{11}^n + \frac{\bar{k}}{3} I^n + \alpha^2 \right] \alpha_{11}^{n+1} + \\ &+ 2(\alpha_{12}^n u_y^n + \alpha_{13}^n u_z^n) + \frac{\operatorname{Re} \bar{k}}{3} (\alpha_{11}^n)^2 - \beta \operatorname{Re} [(\alpha_{12}^n)^2 + (\alpha_{13}^n)^2 - (\alpha_{11}^n)^2], \\ \frac{\alpha_{22}^{n+1} - \alpha_{22}^n}{\tau} &= - \left[K_I^n + \operatorname{Re} \left(\frac{\bar{k}}{3} + 2\beta \right) \alpha_{22}^n \right] \alpha_{22}^{n+1} + \operatorname{Re} \left(\frac{\bar{k}}{3} (\alpha_{22}^n)^2 + \beta [(\alpha_{22}^n)^2 - (\alpha_{12}^n)^2 - (\alpha_{23}^n)^2] \right), \\ \frac{\alpha_{33}^{n+1} - \alpha_{33}^n}{\tau} &= - \left[K_I^n + \operatorname{Re} \left(\frac{\bar{k}}{3} + 2\beta \right) \alpha_{33}^n \right] \alpha_{33}^{n+1} + \operatorname{Re} \left(\frac{\bar{k}}{3} (\alpha_{33}^n)^2 + \beta [(\alpha_{33}^n)^2 - (\alpha_{13}^n)^2 - (\alpha_{23}^n)^2] \right), \\ \frac{\alpha_{12}^{n+1} - \alpha_{12}^n}{\tau} &= [\beta \operatorname{Re} \alpha_{33}^n - \tilde{K}_I^n] \alpha_{12}^{n+1} + \alpha_{2y}^n u_y^n + \alpha_{2z}^n u_z^n - \beta \operatorname{Re} \alpha_{13}^n \alpha_{23}^n, \\ \frac{\alpha_{13}^{n+1} - \alpha_{13}^n}{\tau} &= [\beta \operatorname{Re} \alpha_{22}^n - \tilde{K}_I^n] \alpha_{13}^{n+1} + \alpha_{2y}^n u_y^n + \alpha_{3z}^n u_z^n - \beta \operatorname{Re} \alpha_{12}^n \alpha_{23}^n, \\ \frac{\alpha_{23}^{n+1} - \alpha_{23}^n}{\tau} &= [\beta \operatorname{Re} \alpha_{11}^n - \tilde{K}_I^n] \alpha_{23}^{n+1} - \beta \operatorname{Re} \alpha_{12}^n \alpha_{13}^n, \end{aligned} \quad (13)$$

где $I^n = a_{11}^n + a_{22}^n + a_{33}^n$, $K_I^n = \operatorname{Re} \left(\alpha^2 + \frac{\bar{k}}{3} I^n \right)$, $\tilde{K}_I^n = K_I^n + \beta \operatorname{Re} I^n$, $\alpha_i^n = \alpha_{ii}^n + \alpha^2$, $i = 2, 3$, u_y^n, u_z^n – производные функции $u(t, y, z)$ при $t = t_n$.

Для поиска функции $u^{n+1}(y, z)$ из (7) несложно вывести следующее линеаризованное уравнение:

$$\begin{aligned} \frac{u^{n+1} - 2u^n + u^{n-1}}{\tau^2} + \operatorname{Re} \alpha^2 \left(\frac{u^{n+1} - u^n}{\tau} \right) - \alpha^2 (u_{yy}^{n+1} + u_{zz}^{n+1}) &= \alpha_{22}^n u_{yy}^n + \alpha_{33}^n u_{zz}^n + \\ &+ 2\alpha_{23}^n u_{yz}^n - \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) I^n \left(\frac{u^n - u^{n-1}}{\tau} \right) + \tilde{K}_I^n A(t_n) + F^n, \end{aligned} \quad (14)$$

где

$$F^n = [(\alpha_{12}^n)_y + (\alpha_{23}^n)_z] u_y^n + [(\alpha_{23}^n)_y + (\alpha_{12}^n)_z] u_z^n - \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) [I_y^n \alpha_{12}^n + I_z^n \alpha_{13}^n] - \beta \operatorname{Re} [(I_{22}^n)_z + (I_{33}^n)_y] + A'(t_n),$$

$I_{y,z}^n = (\alpha_{11}^n)_{y,z} + (\alpha_{22}^n)_{y,z} + (\alpha_{33}^n)_{y,z}$ – производные от функции I^n , $(I_{22}^n)_z = (\alpha_{12}^n \alpha_{23}^n - \alpha_{13}^n \alpha_{22}^n)_z$, $(I_{33}^n)_y = (\alpha_{13}^n \alpha_{23}^n - \alpha_{12}^n \alpha_{33}^n)_y$, $A'(t_n) = \left. \frac{dA(t)}{dt} \right|_{t=t_n}$. Отметим также, что при записи (14) использована аппроксимация

$$\tilde{K}_I u_t = \operatorname{Re} \alpha^2 u_t + \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) I^n u_t \approx \operatorname{Re} \alpha^2 \left(\frac{u^{n+1} - u^n}{\tau} \right) + \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) I^n \left(\frac{u^n - u^{n-1}}{\tau} \right).$$

4.2. Аппроксимация неизвестных функций и их производных

На каждом шаге итерационной схемы по времени будем искать решения уравнений (13), (14) с граничными условиями (8) в классе достаточно гладких в области Ω функций. Далее для записи вычислительной схемы необходимо аппроксимировать в уравнениях (13), (14) функции $u^n(y, z)$ и $\alpha_{ij}^n(y, z)$ и их производные по переменным y и z элементами конечномерных функциональных

пространств. В качестве таких пространств используем пространства алгебраических полиномов, а в качестве способа аппроксимации – интерполяционные полиномы в форме Лагранжа с узлами в нулях многочленов Чебышёва. Конкретно, для приближения $u^n(y, z)$ и $\alpha_{ij}^n(y, z)$ будем использовать прямые (тензорные) произведения таких полиномов в области Ω . Свойства таких приближений, а именно отсутствие их насыщения, логарифмический рост констант Лебега, а также оценки погрешности подробно обсуждаются в [11, гл. 3].

Введем в области Ω сетку с узлами (y_k, z_m) , где

$$y_k = \cos \frac{(2k-1)\pi}{2K}, \quad z_m = r \cos \frac{(2m-1)\pi}{2M}, \quad k = \overline{1, K}, \quad m = \overline{1, M}.$$

Сделаем для удобства преобразование координаты z : $\tilde{z} = z/r$, $\tilde{z}_m = z_m/r$, где $\tilde{z}, \tilde{z}_m \in [-1, 1]$.

Пусть $(\alpha_{ij}^n)_{km} = \alpha_{ij}^n(y_k, z_m)$, $u_{km}^n = u^n(y_k, z_m)$, тогда

$$\alpha_{ij}^n(y, z) \approx P(\alpha_{ij}^n, y, z) = \sum_{k=1}^K \sum_{m=1}^M l_{kK}(y) l_{mM}(\tilde{z}) (\alpha_{ij}^n)_{km}, \tag{15}$$

$$u^n(y, z) \approx P_\omega(u^n, y, z) = \sum_{k=1}^K \sum_{m=1}^M \frac{\omega(y, \tilde{z})}{\omega(y_k, \tilde{z}_m)} l_{kK}(y) l_{mM}(\tilde{z}) u_{km}^n, \tag{16}$$

где

$$l_{kK}(y) = \frac{T_K(y)}{(y - y_k) T_K'(y_k)}, \quad l_{mM}(\tilde{z}) = \frac{T_M(\tilde{z})}{(\tilde{z} - \tilde{z}_m) T_M'(\tilde{z}_m)}$$

суть фундаментальные многочлены интерполяции, $T_K(y)$, $T_M(\tilde{z})$ – многочлены Чебышёва степени K и M соответственно, множитель $\omega(y, \tilde{z}) = (1 - y^2)(1 - \tilde{z}^2)$ в формуле (16) обеспечивает автоматическое выполнение граничных условий (8) (см. также [11, гл. 9, § 5]).

Для аппроксимации уравнений (13), (14) далее будет использован метод коллокаций с узлами (y_k, z_m) , поэтому нам потребуются значения производных полиномов (15), (16) в этих узлах. Приведем формулы дифференцирования по переменной y (по переменной z формулы строятся аналогично):

$$\frac{\partial P}{\partial y} = \sum_{k=1}^K \sum_{m=1}^M S_{kK}(y) s_k^2 \left\{ \frac{T_K(y)}{y - y_k} - T_K'(y) \right\} l_{mM}(\tilde{z}) (\alpha_{ij}^n)_{km}, \tag{17}$$

$$\frac{\partial P_\omega}{\partial y} = \sum_{k=1}^K \sum_{m=1}^M (-1)^{K-1} S_{kK}(y) \lambda_{mM}(\tilde{z}) \left\{ \frac{2yy_k - y^2 - 1}{y - y_k} T_K(y) + (1 - y^2) T_K'(y) \right\} u_{km}^n. \tag{18}$$

Поскольку уравнение (14) имеет второй порядок, потребуются также аппроксимации второй производной от функции $u(y, z)$:

$$\frac{\partial^2 P_\omega}{\partial y^2} = \sum_{k=1}^K \sum_{m=1}^M (-1)^{K-1} S_{kK}(y) \lambda_{mM}(\tilde{z}) \left\{ \frac{2s_k^2 - K^2(y - y_k)^2}{(y - y_k)^2} T_K(y) - \frac{y^2 - 3yy_k + 2}{y - y_k} T_K'(y) \right\} u_{km}^n, \tag{19}$$

где

$$s_k = \sqrt{1 - y_k^2}, \quad S_{kK}(y) = \frac{(-1)^k}{K(y - y_k) s_k}, \quad \lambda_{mM}(\tilde{z}) = \frac{(1 - \tilde{z}^2)}{1 - \tilde{z}_m^2} l_{mM}(\tilde{z}).$$

Переходя в (17)–(19) к пределу при $y \rightarrow y_l$, $z \rightarrow z_q$, $l = \overline{1, K}$, $q = \overline{1, M}$, используя свойство фундаментальных многочленов $l_{mM}(\tilde{z}_q) = \delta_{mq}$ – символ Кронекера и правило Лопиталья, находим

$$\frac{\partial \alpha_{ij}^n}{\partial y}(y_l, z_q) \approx \lim_{(y,z) \rightarrow (y_l, z_q)} \frac{\partial P(y, z)}{\partial y} = \sum_{k=1, k \neq l}^K \frac{(-1)^{l+k} s_k}{s_l(y_l - y_k)} (\alpha_{ij}^n)_{kq} - \frac{y_l}{2s_l^2} (\alpha_{ij}^n)_{lq}, \tag{20}$$

$$\frac{\partial u^n}{\partial y}(y_l, z_q) \approx \lim_{(y,z) \rightarrow (y_l, z_q)} \frac{\partial P_\omega(y, z)}{\partial y} = \sum_{k=1, k \neq l}^K (-1)^{k+l} \frac{s_l}{s_k(y_l - y_k)} u_{kq}^n - \frac{3y_l}{2s_l^2} u_{lq}^n, \tag{21}$$

$$\frac{\partial^2 u^n}{\partial y^2}(y_l, z_q) \approx \lim_{(y,z) \rightarrow (y_l, z_q)} \frac{\partial^2 P_\omega(y, z)}{\partial y^2} = \sum_{k=1, k \neq l}^K (-1)^{k+l-1} \frac{2s_l^2 + 3y_l(y_l - y_k)}{s_k s_l (y_l - y_k)^2} u_{kq}^n - \frac{(K^2 + 5)s_l^2 + 3y_l^2}{3s_l^4} u_{lq}^n. \quad (22)$$

Пусть

$$\xi_{lk} = \frac{(-1)^{l+k} s_k}{s_l (y_l - y_k)}, \quad \eta_{lk} = \frac{(-1)^{l+k} s_l}{s_k (y_l - y_k)}, \quad \zeta_{lk} = (-1)^{l+k-1} \frac{2s_l^2 + 3y_l(y_l - y_k)}{s_k s_l (y_l - y_k)^2},$$

где $l, k = \overline{1, K}, l \neq k$,

$$v_l = -\frac{y_l}{2s_l^2}, \quad \hat{v}_l = -\frac{3y_l}{2s_l^2}, \quad \tilde{v}_l = -\frac{(K^2 + 5)s_l^2 + 3y_l^2}{3s_l^4}, \quad l = \overline{1, K},$$

$\Upsilon_{ij}^n, (\Gamma_{ij}^n)_y$ суть $K \times M$ -матрицы с элементами $\alpha_{ij}^n(y_k, z_m)$ и $\frac{\partial \alpha_{ij}^n}{\partial y}(y_k, z_m)$,

U^n, U_y^n, U_{yy}^n суть $K \times M$ -матрицы с элементами $u^n(y_k, z_m)$, $\frac{\partial u^n}{\partial y}(y_k, z_m)$, $\frac{\partial^2 u^n}{\partial y^2}(y_k, z_m)$.

Сформируем $K \times K$ -матрицы:

$$A_1 = \begin{pmatrix} v_1 & \xi_{12} & \dots & \xi_{1K} \\ \xi_{21} & v_2 & \dots & \xi_{2K} \\ \vdots & \vdots & \dots & \vdots \\ \xi_{K1} & \xi_{K2} & \dots & v_K \end{pmatrix}, \quad \mathcal{A}_1 = \begin{pmatrix} \hat{v}_1 & \eta_{12} & \dots & \eta_{1K} \\ \eta_{21} & \hat{v}_2 & \dots & \eta_{2K} \\ \vdots & \vdots & \dots & \vdots \\ \eta_{K1} & \eta_{K2} & \dots & \hat{v}_K \end{pmatrix}, \quad \mathcal{A}_2 = \begin{pmatrix} \tilde{v}_1 & \zeta_{12} & \dots & \zeta_{1K} \\ \zeta_{21} & \tilde{v}_2 & \dots & \zeta_{2K} \\ \vdots & \vdots & \dots & \vdots \\ \zeta_{K1} & \zeta_{K2} & \dots & \tilde{v}_K \end{pmatrix}.$$

Переписывая линейные комбинации в правых частях (20)–(22) в виде произведений матриц, для аппроксимации производных по y в уравнениях (13), (14) получим формулы

$$(\Upsilon_{ij}^n)_y \approx A_1 \Upsilon_{ij}^n, \quad U_y^n \approx \mathcal{A}_1 U^n, \quad U_{yy}^n \approx \mathcal{A}_2 U^n.$$

Дифференцируя приближения (15), (16) по переменной z и проводя выкладки, аналогичные описанным выше, сформируем $M \times M$ -матрицы $B_1, \mathcal{B}_1, \mathcal{B}_2$ для аппроксимации производных по переменной z :

$$(\Upsilon_{ij}^n)_z \approx \Upsilon_{ij}^n B_1^T, \quad U_z^n \approx U^n \mathcal{B}_1^T, \quad U_{zz}^n \approx U^n \mathcal{B}_2^T, \quad (\Upsilon_{ij}^n)_{yz} \approx A_1 \Upsilon_{ij}^n B_1^T, \quad U_{yz}^n \approx \mathcal{A}_1 U^n \mathcal{B}_1^T.$$

Здесь $K \times M$ -матрицы $(\Upsilon_{ij}^n)_z, U_z^n, U_{zz}^n, (\Upsilon_{ij}^n)_{yz}, U_{yz}^n$ определяются аналогично $(\Upsilon_{ij}^n)_y, U_y^n$.

Далее для построения алгоритмов воспользуемся спектральным разложением матриц $\mathcal{A}_2, \mathcal{B}_2$, аппроксимирующих вторые производные:

$$\mathcal{A}_2 = R_{\mathcal{A}} D_{\mathcal{A}} R_{\mathcal{A}}^{-1}, \quad \mathcal{B}_2 = R_{\mathcal{B}} D_{\mathcal{B}} R_{\mathcal{B}}^{-1}, \quad (23)$$

где $R_{\mathcal{A}}, R_{\mathcal{B}}$ – матрицы собственных векторов \mathcal{A}_2 и \mathcal{B}_2 ; $D_{\mathcal{A}}, D_{\mathcal{B}}$ – диагональные матрицы собственных значений \mathcal{A}_2 и \mathcal{B}_2 – $d_{\mathcal{A}}^k, d_{\mathcal{B}}^m$.

Замечание 2. Строго говоря, при наличии у матриц \mathcal{A}_2 и \mathcal{B}_2 пар комплексно-сопряженных собственных чисел, матрицы $D_{\mathcal{A}}, D_{\mathcal{B}}$ являются блочно-диагональными, содержащими на диагонали блоки размера 2×2 . Однако, как показали численные эксперименты, выбранный базис исключает такую возможность. Важным обстоятельством является также медленный рост обусловленностей матриц $R_{\mathcal{A}}, R_{\mathcal{B}}$ с ростом K, M , обеспечивающий устойчивость алгоритма к погрешностям округления (см. [7]).

4.3. Решение задачи линейной алгебры на каждом временном шаге

После подстановки описанных аппроксимаций в уравнения (13), (14) и отбрасывания погрешности приходим к линейным матричным уравнениям для выражения матриц $\Upsilon_{ij}^{n+1}, U^{n+1}$ че-

рез матрицы Υ_{ij}^n, U^n . Запишем для примера одно из уравнений для выражения Υ_{11}^n (см. первое уравнение в системе (13)):

$$\begin{aligned} \mathfrak{M}_{11} \cdot \Upsilon_{11}^{n+1} &= \left\{ \frac{1}{\tau} + \operatorname{Re} \left[\left(\frac{\bar{k}}{3} + 2\beta \right) \Upsilon_{11}^n + \frac{\bar{k}}{3} (\Upsilon_{11}^n + \Upsilon_{22}^n + \Upsilon_{33}^n) + \alpha^2 E_I \right] \right\} \cdot \Upsilon_{11}^{n+1} = \\ &= \frac{1}{\tau} \Upsilon_{11}^n + 2(\Upsilon_{12}^n \cdot (\mathcal{A}_1 U^n) + \Upsilon_{13}^n \cdot (U^n \mathcal{B}_1^T)) + \frac{\operatorname{Re} \bar{k}}{3} (\Upsilon_{11}^n \cdot \Upsilon_{11}^n) - \beta \operatorname{Re} (\Upsilon_{12}^n \cdot \Upsilon_{12}^n + \Upsilon_{13}^n \cdot \Upsilon_{13}^n - \Upsilon_{11}^n \cdot \Upsilon_{11}^n). \end{aligned} \quad (24)$$

Здесь точка означает поэлементное произведение матриц, E_I – матрица размера $K \times M$, все элементы которой равны единице, начальные данные задаются в соответствии с (9): $\Upsilon_{11}^0 = (0)_{K \times M}$ – нулевая матрица. Аппроксимации остальных пяти уравнений из системы (13) записываются аналогично, при этом в левой части составляются матрицы \mathfrak{M}_{ij} .

Аппроксимация уравнения (14) выглядит следующим образом:

$$\begin{aligned} \left(\frac{1}{\tau^2} + \frac{\operatorname{Re} \alpha^2}{\tau} \right) U^{n+1} - \alpha^2 \mathcal{A}_2 U^{n+1} - \alpha^2 U^{n+1} \mathcal{B}_2^T = \mathfrak{F}_U^n := \frac{1}{\tau^2} (2U^n - U^{n-1}) + \frac{\operatorname{Re} \alpha^2}{\tau} U^n + \Upsilon_{22}^n \cdot (\mathcal{A}_2 U^n) + \\ + \Upsilon_{33}^n \cdot (U^n \mathcal{B}_2^T) + 2\Upsilon_{23}^n \cdot (\mathcal{A}_1 U^n \mathcal{B}_1^T) - \frac{\operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right)}{\tau} (\Upsilon_{11}^n + \Upsilon_{22}^n + \Upsilon_{33}^n) \cdot (U^n - U^{n-1}) + \tilde{\mathcal{H}}_I^n(t_n) + \mathfrak{F}^n, \end{aligned} \quad (25)$$

где

$$\begin{aligned} \tilde{\mathcal{H}}_I^n &= \operatorname{Re} \alpha^2 E_I + \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) (\Upsilon_{11}^n + \Upsilon_{22}^n + \Upsilon_{33}^n), \\ \mathfrak{F}^n &= [A_1 \Upsilon_{22}^n + \Upsilon_{23}^n \mathcal{B}_1^T] \cdot (\mathcal{A}_1 U^n) + [A_1 \Upsilon_{23}^n + \Upsilon_{33}^n \mathcal{B}_1^T] \cdot (U^n \mathcal{B}_1^T) - \\ &- \operatorname{Re} \left(\frac{\bar{k}}{3} + \beta \right) [(A_1 (\Upsilon_{11}^n + \Upsilon_{22}^n + \Upsilon_{33}^n)) \cdot \Upsilon_{12}^n + ((\Upsilon_{11}^n + \Upsilon_{22}^n + \Upsilon_{33}^n) \mathcal{B}_1^T) \cdot \Upsilon_{13}^n] - \\ &- \beta \operatorname{Re} [(\Upsilon_{12}^n \cdot \Upsilon_{23}^n - \Upsilon_{13}^n \cdot \Upsilon_{22}^n) \mathcal{B}_1^T + A_1 (\Upsilon_{13}^n \cdot \Upsilon_{23}^n - \Upsilon_{12}^n \cdot \Upsilon_{33}^n)] + A'(t_n) \end{aligned}$$

суть $K \times N$ -матрицы.

Для решения матричного уравнения (24) на каждом шаге по времени необходимо выполнять деления элементов матрицы в правой части на элементы матрицы \mathfrak{M}_{11} . Составляя по аналогии с (24) матричные уравнения для остальных пяти уравнений системы (13) и осуществляя поэлементные деления их правых и левых частей, получаем приближенные значения решений α_{ij}^{n+1} в узлах коллокации.

Для решения (25) используем спектральные разложения (23) матриц $\mathcal{A}_2, \mathcal{B}_2^T$, стоящих в левых частях. Умножим (25) на матрицу $R_{\mathcal{A}}^{-1}$ слева и на матрицу $R_{\mathcal{B}}^{-1}$ справа, обозначим $V^{n+1} = R_{\mathcal{A}}^{-1} U^{n+1} R_{\mathcal{B}}^{-1}$, $\mathcal{G}^n = R_{\mathcal{A}}^{-1} \mathfrak{F}_U^n R_{\mathcal{B}}^{-1}$, $p_\tau = \frac{1}{\tau^2} + \frac{\operatorname{Re} \alpha^2}{\tau}$ и получим матричное уравнение:

$$p_\tau V^{n+1} - \alpha^2 D_{\mathcal{A}} V^{n+1} - \alpha^2 V^{n+1} D_{\mathcal{B}} = \mathcal{G}^n,$$

решение которого относительно элементов v_{km}^{n+1} матрицы V^{n+1} дается элементарными формулами

$$v_{km}^{n+1} = \frac{g_{km}^n}{p_\tau - \alpha^2 (d_{\mathcal{A}}^k + d_{\mathcal{B}}^m)}, \quad k = \overline{1, K}, \quad m = \overline{1, M},$$

где g_{km}^n – элементы матрицы \mathcal{G}^n . В данном случае шаг сетки по времени τ нужно выбирать, исходя из условия $p_\tau \neq \alpha^2 (d_{\mathcal{A}}^k + d_{\mathcal{B}}^m) \forall k, m$ или учитывая, что $\alpha^2 = (W \operatorname{Re})^{-1}$,

$$\tau \neq \frac{\operatorname{Re} \pm \sqrt{\operatorname{Re}^2 + 4W \operatorname{Re} (d_{\mathcal{A}}^k + d_{\mathcal{B}}^m)}}{2(d_{\mathcal{A}}^k + d_{\mathcal{B}}^m)} \quad \forall k, m.$$

Зная элементы матрицы V^{n+1} , несложно восстановить значения решения в узлах коллокации: $U^{n+1} = R_{\mathcal{A}} V^{n+1} R_{\mathcal{B}}$.

Отметим, что в силу нелинейности исходных уравнений для перехода от предыдущего шага по времени к следующему решению линейризованных матричных уравнений (24), (25) ищется с помощью итераций по нелинейности. Иначе говоря, при фиксированном $t = t_n$ реализуется цикл, на каждом шаге s которого решения $\Upsilon_{ij}^{n,s}, U^{n,s}$ выражаются через решения $\Upsilon_{ij}^{n,s-1}, U^{n,s-1}$ в соответствии с формулами, описанными выше, и начальными условиями $\Upsilon_{ij}^{n,0} = \Upsilon_{ij}^n, U^{n,0} = U^n$. При этом итерации останавливаются при реализации условия

$$\frac{\|U^{n,s} - U^{n,s-1}\|}{\|U^{n,s}\|} \leq \varepsilon_{NI}, \quad \frac{\|\Upsilon_{ij}^{n,s} - \Upsilon_{ij}^{n,s-1}\|}{\|\Upsilon_{ij}^{n,s}\|} \leq \varepsilon_{NI}, \quad i = 1, 2, 3, \quad j = i, \dots, 3.$$

Номер s в таком случае обозначим σ и положим $\Upsilon_{ij}^{n+1} = \Upsilon_{ij}^{n,\sigma}, U^{n+1} = U^{n,\sigma}$. Здесь и далее норма обозначает максимальный элемент матрицы. В наших расчетах задано $\varepsilon_{NI} = 10^{-10}$. При переходе на следующий шаг по времени меняются значения функций $A(t_n), A'(t_n)$, стоящих в правой части (25) и описывающих зависимость градиента давления в канале от времени.

Отметим, что предложенная аппроксимация уравнений модели в сравнении с классической схемой метода коллокаций приводит к существенному сокращению вычислительных затрат.

Действительно, для вычисления Υ_{ij}^{n+1} в соответствии с (24) необходимо совершить порядка KM операций (деления элементов матрицы в правой части на элементы матрицы в левой части).

Для расчета U^{n+1} по (25) наиболее затратными являются операции произведения матриц $\mathcal{G}^n = R_{sd}^{-1} \mathcal{F}_U^n R_{sb}^{-1}, U^{n+1} = R_{sd} V^{n+1} R_{sb}$. Порядок числа операций, необходимых для этого, есть $O(MK(M+K))$. Здесь предполагается, что спектральные разложения (23) сделаны заранее, и матрицы этих разложений для различных K, M занесены в базу данных. Классическая же схема метода коллокаций в данной задаче ведет к системам с большими разреженными матрицами размера $KM \times KM$, решение которых требует порядка $O((KM)^3)$ операций.

Отметим также, что указанные операции поэлементного деления матриц и произведения матриц относительно небольшого размера во многих современных вычислительных средах выполняются с повышенной скоростью за счет векторизации, что обеспечивает высокую вычислительную эффективность предложенного алгоритма.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

Разработанный метод использован для решения нестационарной задачи (5)–(9). При этом градиент давления был задан постоянным $A(t) \equiv A_0$. Наряду с нестационарной методом установления решалась также стационарная задача (10), (12) при $A = A_0$. При этом использован алгоритм, предложенный в [7]. Отметим, что сходимость метода установления достигается только при использовании определенных ветвей решения $\sigma(\mu)$ уравнения (11). Какие именно ветви обеспечивают сходимость, и какая ветвь соответствует предельному решению нестационарной задачи (5)–(9), зависит от значений параметров модели.

В табл. 1 приведены описание и диапазоны значений основных параметров модели и алгоритма, использованные в расчетах. Обозначим время выхода течения на стационарный режим величиной t^* , т.е. $t^* = n^* \tau$, где n^* – номер итерации, на которой впервые реализуются условия стабилизации течения:

$$\frac{\|U^{n^*} - U^{n^*-1}\|}{\tau} \leq \varepsilon_S, \quad \frac{\|\Upsilon_{ij}^{n^*} - \Upsilon_{ij}^{n^*-1}\|}{\tau} \leq \varepsilon_S, \quad i = 1, 2, 3, \quad j = 1, \dots, i,$$

ε_S – малое число. В проведенных экспериментах установлено, что t^* практически не зависит от параметров K, M и τ . Относительное отклонение нестационарной скорости течения в момент времени t^* от решения стационарной задачи (10), (12) обозначим через $D_u(y, z)$. В экспериментах эта величина рассчитывалась как относительная разность двух массивов, содержащих значения соответствующих скоростей в узлах (y_k, z_m) (см. п. 4.2).

В численных экспериментах установлено, что в случае $c_k = 1$ и $\beta < 0.5$ (именно такие значения использованы в [2], [10]) сходимость метода установления в стационарной задаче имеет место

Таблица 1. Параметры модели и численного метода

Параметр	Описание	Значение
β	Феноменологический параметр	0.01–0.9
c_k	Феноменологический параметр, $k = c_k \beta$	0.5–20
A_0	Безразмерный градиент давления	1–330
W	Число Вайсенберга	0.01–3
$2r$	Безразмерная ширина канала	0.02–2
K, M	Число узлов сетки вдоль осей y и z	21–31
τ	Шаг сетки по времени	0.0015–0.1
ε_S	Погрешность стабилизации течения	10^{-8}

Таблица 2. Значения t^* и $\|D_u(y, z)\|$ при $\beta = 0.1, c_k = 1$ и вариации параметров W, A_0, r

Вариация W при $A_0 = 1, r = 1$			Вариация градиента A_0 при $W = 0.01, r = 1$			Вариация r при $A_0 = 200, W = 0.05$		
W	t^*	$\ D_u(y, z)\ $	A_0	t^*	$\ D_u(y, z)\ $	r	t^*	$\ D_u(y, z)\ $
1.5	41.75	6.23×10^{-9}	200	9.1	1.32×10^{-8}	0.025	1.446	4.49×10^{-6}
2.5	61.95	2.62×10^{-8}	300	14.88	4.62×10^{-8}	0.05	1.434	1.1×10^{-5}
3	108.65	4.64×10^{-8}	330	19.09	5.31×10^{-8}	0.1	1.584	1.32×10^{-6}
$W^{\text{lim}} \approx 3.4$			$A_0^{\text{lim}} \approx 335$			$r^{\text{lim}} \approx 0.115$		

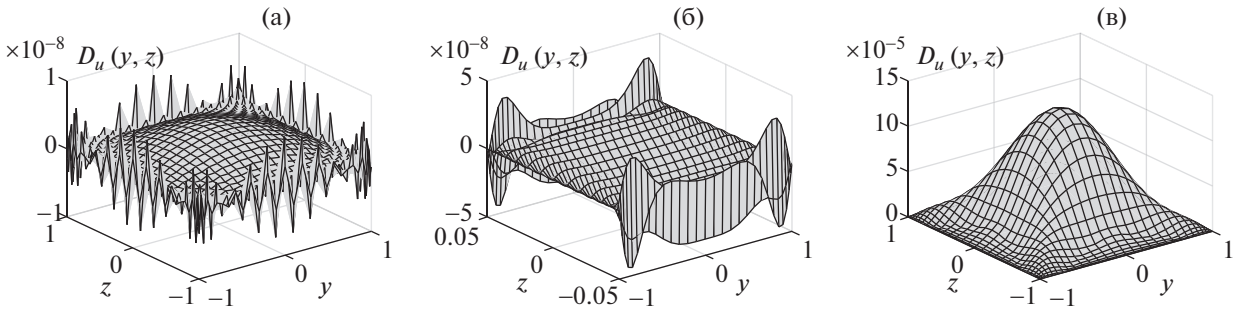
Таблица 3. Значения t^* и $\|D_u(y, z)\|$ при $W = 0.01, r = 1, A_0 = 20, 200$ и вариации параметров β, c_k

Вариация β при $A_0 = 200, c_k = 1$			Вариация c_k при $A_0 = 20, \beta = 0.37$		
β	t^*	$\ D_u(y, z)\ $	c_k	t^*	$\ D_u(y, z)\ $
0.15	10.4	2.58×10^{-8}	1	5.1	3.32×10^{-10}
0.2	13.32	4.39×10^{-8}	1.52	5.1	9.91×10^{-5}
0.26	18.12	6.69×10^{-8}	2	5.15	3.35×10^{-10}
$\beta^{\text{lim}} \approx 0.267$			неограничено		

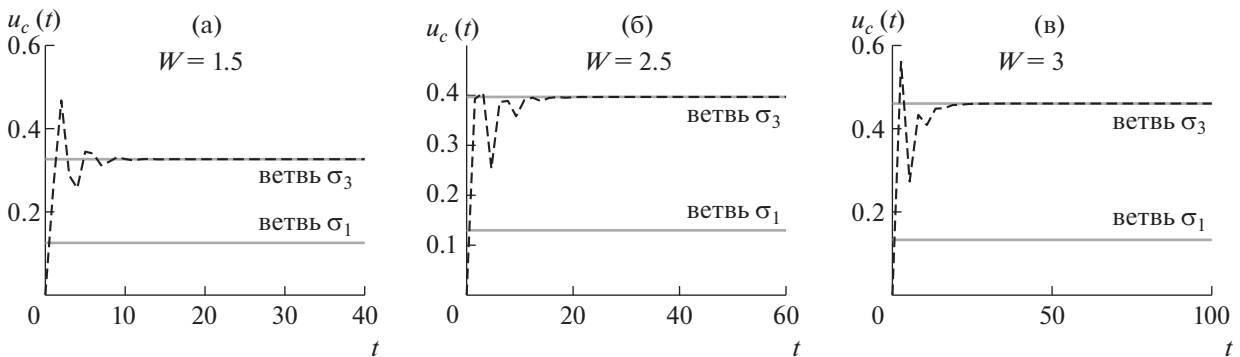
только при $\sigma = \sigma_{1,3}$, причем предельное решение нестационарной задачи соответствует ветви $\sigma = \sigma_3$. В табл. 2 для этого случая приведены значения t^* и $\|D_u(y, z)\|$ при вариации параметров W, A_0, r в некоторой окрестности предельных значений этих параметров. Под предельными здесь понимаются значения, при переходе через которые стационарный режим течения перестает существовать. Эти значения найдены приближенно в расчетах и приведены в последних строках табл. 2, 3 с индексом lim . В табл. 3 указаны значения t^* и $\|D_u(y, z)\|$ при вариации c_k и β .

Видно, что при значениях параметров W и A_0 , близких к предельным, зависимость t^* от этих значений носит нелинейный характер. Время выхода на стационарный режим наиболее чувствительно к значению W . При W , близких к W^{lim} , и малых t имеют место колебания значений скорости, и для установления течения требуется достаточно много времени (см. также фиг. 5). При уменьшении ширины канала r течение устанавливается намного быстрее и для больших значений W и A_0 , однако относительное отклонение $\|D_u(y, z)\|$ при этом возрастает (см. также фиг. 2б). При вариации r в окрестности малых значений ($r < 0.05$) время t^* практически не зависит от r .

Отметим, что зависимость t^* от β близка к линейной, и отклонение $\|D_u(y, z)\|$ при приближении β к β^{lim} мало меняется. Интересная картина наблюдается при вариации c_k (табл. 3): отклоне-



Фиг. 2. Относительные отклонения $D_u(y, z)$ при (а) $A_0 = 1$, $\beta = 0.1$, $c_k = 1$, $r = 1$, $W = 1.5$ (стандартная ситуация); (б) $A_0 = 200$, $\beta = 0.1$, $c_k = 1$, $W = 0.05$, $r = 0.05$ (узкий канал); (в) $A_0 = 20$, $W = 0.01$, $r = 1$, $\beta = 0.37$, $c_k = 1.52$ (переключение $3 \rightarrow 1$ установившегося решения нестационарной задачи).

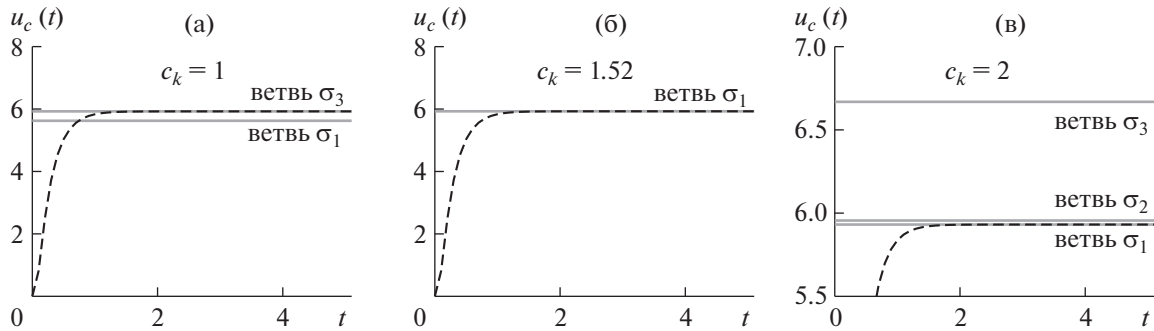


Фиг. 3. Значения скорости $u_c(t)$, полученные при решении нестационарной задачи (штрихи), и соответствующие значения решений стационарной задачи в точке $y = 0$, $z = 0$ (серые линии) при $A_0 = 1$, $\beta = 0.1$, $c_k = 1$, $r = 1$ и различных W : (а) $W = 1.5$, (б) $W = 2.5$, (в) $W = 3$.

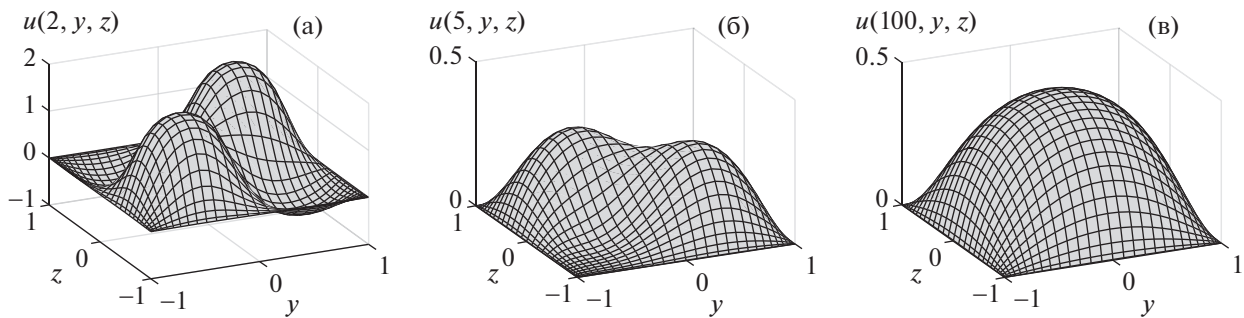
ния $\|D_u(y, z)\|$ при изменении c_k в широких диапазонах остаются малыми за исключением окрестности некоторой точки (в данном случае точки $c_k \approx 1.52$), где они резко возрастают (фиг. 2в). Такая особенность связана с *переключением установившегося решения нестационарной задачи с решения, соответствующего ветви σ_3 , на решение, соответствующее ветви σ_1 , стационарной задачи*. Этот важный эффект, который далее будем именовать “переключение $3 \rightarrow 1$ ”, обсудим чуть ниже.

На фиг. 2 для некоторых режимов течения показаны значения относительных отклонений $D_u(y, z)$ предельных решений нестационарной задачи от решений стационарной задачи. В большинстве случаев отклонения имеют высокочастотные колебания, амплитуда которых возрастает в окрестности границ, достигая значений порядка 10^{-10} – 10^{-8} (фиг. 2а). Проведенные эксперименты позволяют заключить, что такой характер поведения $D_u(y, z)$ связан с вычислительной погрешностью и значениями K , M и ε_s . Однако в некоторых случаях наблюдаются конкретные тенденции в поведении $D_u(y, z)$: например, рост этих значений при уменьшении ширины канала в окрестности стенок, имеющих большую длину (фиг. 2б); или рост отклонений в центре канала при переключении $3 \rightarrow 1$ (фиг. 2в).

На фиг. 3, 4 показаны зависимости скорости течения в центре канала от времени $u_c(t) = u(t, 0, 0)$ (штрихи), полученные при решении нестационарной задачи (5)–(9) с вариацией значений W и c_k . На этих же графиках серыми линиями отмечены значения решений стационарных уравнений (10), (12) в точке $y = 0$, $z = 0$, полученные при использовании тех ветвей функции $\sigma(\mu)$, которые обеспечивают сходимость метода установления. Значения параметров рассмотренных режимов соответствуют указанным в табл. 2, 3.



Фиг. 4. Значения скорости $u_c(t)$, полученные при решении нестационарной задачи (штрихи), и соответствующие значения решений стационарной задачи в точке $y = 0, z = 0$ (серые линии) при $W = 0.01, \beta = 0.37, A_0 = 20, r = 1$ и различных c_k : (а) $c_k = 1$, (б) $c_k = 1.52$, (в) $c_k = 2$.

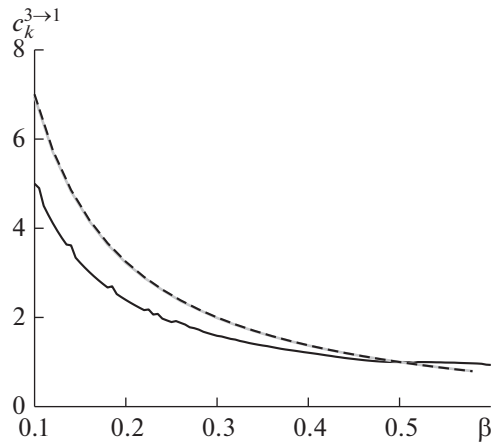


Фиг. 5. Значения скорости $u(t, y, z)$, рассчитанные при $A_0 = 1, \beta = 0.1, c_k = 1, r = 1$ и $W = 3$ в различные моменты времени: (а) $t = 2$, (б) $t = 5$, (в) $t = 100$.

Отметим, что колебания значений на фиг. 3 характеризуют сложный процесс установления течения при достаточно больших W . Более подробно этот процесс прослеживается на фиг. 5, где приведены распределения скорости $u(t, y, z)$ в различные моменты времени t . Видно, что на начальных этапах (при малых t) имеют место сильные колебания решения, а также течения в направлении, противоположном градиенту давления (области отрицательных значений скорости видны на фиг. 5а). При $t \approx 8$ сильные колебания затухают и течение плавно выходит на стационарный режим.

На фиг. 4 показано, что количество решений стационарной задачи (10), (12), которые удается получить с помощью метода установления, меняется в зависимости от параметров задачи. Существенное влияние при этом оказывают параметры β и c_k . Причем при определенном соотношении этих параметров происходит переключение $3 \rightarrow 1$.

Более конкретно, при значениях параметров $A_0 = 20, W = 0.01, r = 1, \beta = 0.37$ и $c_k < 1.515$ методом установления удастся получить два стационарных решения, соответствующих ветвям σ_1, σ_3 , при этом устойчивое решение соответствует ветви σ_3 (фиг. 4а). При $c_k \approx 1.515$ метод установления перестает сходиться для $\sigma = \sigma_3$, и происходит переключение $3 \rightarrow 1$. В диапазоне $1.515 < c_k < 1.528$ удастся получить единственное решение стационарной задачи для $\sigma = \sigma_1$ (фиг. 4б). В диапазоне $1.528 < c_k < 1.532$ – два решения для $\sigma = \sigma_{1,2}$. При $c_k > 1.532$ – три решения для всех $\sigma = \sigma_{1,2,3}$ (фиг. 4в). При этом для $c_k > 1.515$ устойчивым остается решение, соответствующее ветви σ_1 . Подчеркнем, что неравенства, указанные в этом примере, были получены в расчетах и являются приближенными. Заметим, что описанный характер исчезновения и возникновения численных решений стационарной задачи при увеличении c_k , а также переключения $3 \rightarrow 1$ сохраняется при вариации других параметров задачи.



Фиг. 6. Кривые $c_k^{3 \rightarrow 1}(\beta)$ при $A_0 = 1$, $r = 1$ и различных W : $W = 0.01$ (штрихи), $W = 0.1$ (серая линия), $W = 1$ (черная линия).

Обозначим $c_k^{3 \rightarrow 1}$ значение c_k , при котором происходит переключение $3 \rightarrow 1$. На фиг. 6 показан график зависимости $c_k^{3 \rightarrow 1}$ от β при различных значениях параметра W . Из фиг. 6 следует, что при малых значениях W кривые $c_k^{3 \rightarrow 1}(\beta)$ практически совпадают (серая и штриховая линии), однако при увеличении W кривая меняется. Аналогичная картина наблюдается при увеличении A_0 .

Обозначим символом $c_k^{1,2,3}$ минимальное значение параметра c_k , при котором стационарная задача имеет три решения, соответствующие ветвям $\sigma_{1,2,3}$, и рассмотрим зависимость $c_k^{1,2,3}(\beta)$. Из примера, приведенного выше, можно заключить, что $c_k^{1,2,3}(0.37) \approx 1.532$. При проведении вычислений установлено, что линия $c_k^{1,2,3}(\beta)$ практически не меняется при вариации других параметров задачи и с высокой точностью аппроксимируется гиперболой $c_k^{1,2,3}(\beta) \approx 3/(4\beta) - 0.5$. График этой гиперболы визуально неотличим от штриховой линии на фиг. 6. При уменьшении параметров W и A_0 кривые перехода $c_k^{3 \rightarrow 1}(\beta)$ стремятся к этой гиперболе снизу. При приближении значений W и A_0 к предельным, кривая $c_k^{3 \rightarrow 1}(\beta)$ существенно отдалится от $c_k^{1,2,3}(\beta)$ и меняет форму. При малых значениях W и A_0 и вариации r в широких диапазонах кривые $c_k^{3 \rightarrow 1}(\beta)$ остаются в окрестности $c_k^{1,2,3}(\beta)$.

В заключение отметим, что реализуемость стационарного течения существенно зависит от начальных данных. При проведении экспериментов здесь использованы нулевые начальные данные (9). Однако, если задать начальную функцию для скорости с достаточно большими значениями в центре канала, либо функцию, несогласованную с граничными условиями, стационарное течение типа Пуазейля не реализуется.

СПИСОК ЛИТЕРАТУРЫ

1. Лойцянский Л.Г. Механика жидкости и газа. М.: Наука, 1978 [Loitsyanskii L.G. Mechanics of liquids and gases. Oxford: Stewartson Pergamon Press; Moscow: Nauka, 1978].
2. Алтухов Ю.А., Гусев А.С., Пышнограй Г.В. Введение в мезоскопическую теорию текучести полимерных систем. Барнаул: АлтГПА, 2012 [Altukhov Yu.A., Gusev A.S., Pyshnograi G.V. Introduction to the mesoscopic theory of flowing polymer systems. Barnaul: AltGPA, 2012].
3. Pokrovskii V.N. The mesoscopic theory of polymer dynamics. 2nd ed. Berlin: Springer, 2010.
4. Vinogradov G.V., Pokrovskii V.N., Yanovsky Yu.G. Theory of viscoelastic behavior of concentrated polymer solutions and melts in one-molecular approximation and its experimental verification // Rheol. Acta. 1972. V. 7. P. 258–274.
5. Блохин А.М., Семисалов Б.В. Стационарное течение несжимаемой вязкоупругой полимерной жидкости в канале с эллиптическим сечением // Сиб. журн. индустриал. матем. 2014. Т. XVII. № 4 (60). С. 38–47

- [*Blokhin A.M., Semisalov B.V.* A stationary flow of an incompressible viscoelastic fluid in a channel with elliptic cross section // *J. Appl. Industr. Math.* 2015. V. 9. 1. P. 18–26].
6. *Блохин А.М., Семисалов Б.В.* Расчет стационарных неізотермических МГД течений полимерной жидкости в каналах с внутренними нагревательными элементами // *Сиб. журн. индустриал. матем.* 2020. Т. 23. № 2. С. 17–40 [*Blokhin A.M., Semisalov B.V.* Simulation of the stationary nonisothermal MHD flows of polymeric fluids in channels with interior heating elements // *J. Appl. Industr. Math.* 2020. V. 14. № 2. P. 222–241].
 7. *Семисалов Б.В.* Быстрый нелокальный алгоритм решения краевых задач Неймана–Дирихле с контролем погрешности // *Вычисл. методы и программирование.* 2016. Т. 17. № 4. С. 500–522 [*Semisalov B.V.* A fast nonlocal algorithm for solving Neumann-Dirichlet boundary value problems with error control // *Vych. Metody Programm.* 2016. V. 17. Is. 4. P. 500–522].
 8. *Седов Л.И.* Механика сплошной среды. Т. 1. М.: Наука, 1970 [*Sedov L.I.* Mechanics of continuous media. V. 1. Singapore: World Sci., 1997].
 9. *Pai Shih-I.* Introduction to the theory of compressible flow. Literary Licensing, LLC. 2013.
 10. *Зинович С.А., Головичёва И.Э., Пышнограй Г.В.* Влияние молекулярной массы на сдвиговую и продольную вязкость линейных полимеров // *Прикладн. механ. и техн. физ.* 2000. Т. 41. № 2. С. 154–160 [*Golovicheva S.A., Zinovich G.V., Pyshnograï G.V.* Effect of the molecular mass on the shear and longitudinal viscosity of linear polymers // *J. Appl. Mech. Tech. Phys.* 2000. V. 41. P. 347–352].
 11. *Бабенко К.И.* Основы численного анализа. М.–Ижевск: НИЦ “Регулярная и хаотическая динамика”, 2002 [*Babenko K.I.* Fundamentals of numerical analysis. Moscow–Izhevsk: Sci. Publ. Center “Regular and chaotic dynamics”, 2002].

**МАТЕМАТИЧЕСКАЯ
ФИЗИКА**

УДК 517.958

К ЧИСЛЕННОМУ РЕШЕНИЮ ОДНОГО КЛАССА ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ III РОДА

© 2022 г. Н. С. Габбасов^{1,*}, З. Х. Галимова^{2,**}

¹ 423810 Набережные Челны, пр-т Мира, 68/19, Набережночелнинский ин-т Казанского ун-та, Россия

² 423815 Набережные Челны, пр-т Вахитова, 53/02, Набережночелнинский филиал

Казанского инновационного университета им. В.Г. Тимирязова, Россия

*e-mail: gabbasovnazim@rambler.ru

**e-mail: galimovazh2020@mail.ru

Поступила в редакцию 09.06.2021 г.

Переработанный вариант 09.06.2021 г.

Принята к публикации 12.10.2021 г.

Исследовано линейное интегральное уравнение III рода с фиксированными особенностями в ядре. Для его приближенного решения в пространстве обобщенных функций предложен и обоснован специальный обобщенный вариант сплайн-метода. Установлена оптимальность по порядку точности построенного метода. Библ. 19.

Ключевые слова: интегральное уравнение III рода, пространство обобщенных функций, приближенное решение, теоретическое обоснование.

DOI: 10.31857/S0044466922020089

1. ВВЕДЕНИЕ

Объектом исследования является линейное интегральное уравнение III рода с фиксированными особенностями в ядре (УТРФО):

$$Ax \equiv x(t) \prod_{j=1}^l (t - t_j)^{m_j} + \int_{-1}^1 K(t, s) \left[(s+1)^{p_1} (1-s)^{p_2} \right]^{-1} x(s) ds = y(t), \quad (1.1)$$

где

$$t \in I \equiv [-1, 1], \quad t_j \in (-1, 1), \quad m_j \in N \quad (j = \overline{1, l}); \quad p_1, p_2 \in R^+, K$$

и y – известные непрерывные функции, обладающие определенными свойствами “гладкости” точечного характера, $x(t)$ – искомая функция, а интеграл понимается в смысле конечной части по Адамару [1, с. 144–150]. Уравнения вида (1.1) находят все более широкие применения как в теории, так и в приложениях. К такого рода уравнениям приводит ряд важных задач теорий упругости, переноса нейтронов, рассеяния частиц (см., например, [2], [3] и библиографию к [2] и [4]), а также теорий сингулярных интегральных уравнений с вырождающимся символом [5] и уравнений с частными производными смешанного типа [6]. При этом естественными классами решений УТР, как правило, являются специальные пространства обобщенных функций (ПОФ) типа D или V . Под D (соответственно, V) понимается ПОФ, построенное на основе функционала “дельта-функция Дирака” (соответственно “конечная часть интеграла по Адамару”). Исследуемые УТРФО точно решаются лишь в очень редких частных случаях, поэтому разработка теоретически обоснованных эффективных методов их приближенного решения в ПОФ является актуальным и активно развивающимся направлением математического анализа и вычислительной математики. Ряд результатов в этой области получен в работах [7]–[12], в которых предложены и обоснованы специальные прямые полиномиальные методы решения УТРФО вида (1.1) в ПОФ типа D и V .

В настоящей работе для приближенного решения УТРФО (1.1) в ПОФ типа D предложен новый вариант обобщенного метода коллокации, основанный на применении кубических сплайнов минимального дефекта. Проведено его теоретическое обоснование в смысле [13, гл. 1] и

установлено, что разработанный метод оптимален по порядку точности на некотором классе F гладких функций среди всех прямых проекционных методов решения исследуемых уравнений в ПОФ.

2. О ПРОСТРАНСТВАХ ОСНОВНЫХ И ОБОБЩЕННЫХ ФУНКЦИЙ

Пусть $C \equiv C(I)$ – пространство непрерывных на I функций с обычной \max -нормой и $m \in N$. Следуя [14], скажем, что функция $f \in C$ принадлежит классу $C\{m; 0\} \equiv C_0^{\{m\}}(I)$, если в точке $t = 0$ существует тейлоровская производная $f^{\{i\}}(0)$ порядка m (естественно считаем, что $C\{0; 0\} \equiv C$). Множество $C\{m; 0\}$ назовем классом *точечно-гладких* функций с *характеристическим* оператором $T : C\{m; 0\} \rightarrow C$, определенным по правилу

$$Tf \equiv \left[f(t) - \sum_{i=0}^{m-1} \frac{f^{\{i\}}(0)t^i}{i!} \right] t^{-m} \equiv F(t) \in C \quad (F(0) \equiv \lim_{t \rightarrow 0} F(t)).$$

По норме

$$\|f\|_{C\{m; 0\}} \equiv \|Tf\|_C + \sum_{i=0}^{m-1} |f^{\{i\}}(0)|$$

пространство $C\{m; 0\}$ полно и нормально вложено в C (см., например, [15, гл. 1, § 2]).

Далее, пусть $p \in R^+$ и $g \in C$. Также следуя [14], будем обозначать $g \in C\{p; 1\} \equiv C_1^{\{p\}}(I)$, если существуют левые тейлоровские производные $g^{\{j\}}(1)$ ($j = \overline{1, [p]}$) в точке $t = 1$, причем при $p \neq [p]$ ($[]$ – целая часть числа) конечен предел

$$\lim_{t \rightarrow 1^-} \left\{ \left[g(t) - \sum_{j=0}^{[p]} \frac{g^{\{j\}}(1)(t-1)^j}{j!} \right] (1-t)^{-p} \right\}.$$

Векторное пространство $C\{p; 1\}$ снабдим нормой

$$\|g\|_{\{p\}} \equiv \|g\|_{C\{p; 1\}} \equiv \|Sg\|_C + \sum_{i=0}^{\lambda} |g^{\{i\}}(1)|, \tag{2.1}$$

где

$$Sg \equiv \left[g(t) - \sum_{i=0}^{\lambda} \frac{g^{\{i\}}(1)(t-1)^i}{i!} \right] (1-t)^{-p} \equiv G(t) \in C, \tag{2.2}$$

$$\lambda = \lambda(p) \equiv [p] - (1 + \text{sign}([p] - p)), \quad G(1) \equiv \lim_{t \rightarrow 1^-} G(t).$$

Заметим, что элементы пространства $C\{p; 1\}$ – функции вида

$$g(t) = (1-t)^p G(t) + \sum_{i=0}^{\lambda} b_i (t-1)^i, \tag{2.3}$$

причем $Sg = G(t) \in C$, $g^{\{i\}}(1) = b_i i!$ ($i = \overline{0, \lambda}$). Отсюда ясно, что пространство $C\{p; 1\}$ с нормой (2.1) полно и вложено в C .

Теперь образуем основное для наших исследований пространство

$$Y \equiv C_{0;1}^{\{m\};\{p\}}(I) \equiv \{y \in C\{m; 0\} | Ty \in C\{p; 1\}\}.$$

В качестве нормы в нем выберем величину

$$\|y\|_Y \equiv \|Ty\|_{\{p\}} + \sum_{i=0}^{m-1} |y^{\{i\}}(0)| \quad (y \in Y). \tag{2.4}$$

Лемма 2.1 (см. [7]). 1). *Относительно структуры основных функций справедливо соотношение*

$$\varphi \in Y \Leftrightarrow \varphi(t) = (UV\Phi)(t) + t^m \sum_{j=0}^{\lambda} \alpha_j (t-1)^j + \sum_{i=0}^{m-1} e_i t^i, \quad (2.5)$$

где $\Phi \in C$, $\alpha_j \in R$, $j = \overline{0, \lambda}$, $e_i \in R$, $i = \overline{0, m-1}$, причем $ST\varphi = \Phi$, $(T\varphi)^{[j]}(1) = \alpha_j j!$ ($\forall j$), $\varphi^{[i]}(0) = e_i i!$ ($\forall i$); $Uf \equiv t^m f(t)$, $Vf \equiv (1-t)^p f(t)$;

2) по норме (2.4) пространство Y полно и вложено в $C\{m; 0\}$.

Пусть $h \in C(I^2)$ и при каждом фиксированном $s \in I$ функция $h(t, s)$ принадлежит пространству $C\{p; 1\}$. Будем говорить, что $h \in C_t^{\{p\}}(I^2)$, если $S_t h \in C$, где S_t обозначает оператор (2.2), примененный по переменной t . Аналогично определяем класс $C_s^{\{p\}}(I^2)$. Тогда

$$C_1^{\{p\}}(I^2) \equiv C_t^{\{p\}}(I^2) \cap C_s^{\{p\}}(I^2).$$

Теперь над пространством Y основных функций построим семейство $X \equiv D^{\{p\}}\{m; 0\}$ обобщенных функций $x(t)$ вида

$$x(t) \equiv z(t) + \sum_{i=0}^{m-1} \gamma_i \delta^{[i]}(t), \quad (2.6)$$

где $t \in I$, $z \in C\{p; 1\}$, $\gamma_i \in R$ – произвольные постоянные, а δ и $\delta^{[i]}$ – соответственно дельта-функция Дирака и ее тейлоровские производные, действующие на пространстве Y основных функций по следующему правилу:

$$(\delta^{[i]}, y) \equiv \int_{-1}^1 \delta^{[i]}(t) y(t) dt \equiv (-1)^i y^{[i]}(0), \quad y \in Y, \quad i = \overline{0, m-1}.$$

Ясно, что векторное пространство X по норме

$$\|x\|_X \equiv \|z\|_{\{p\}} + \sum_{i=0}^{m-1} |\gamma_i| \quad (2.7)$$

является банаховым.

3. О СПЛАЙНОВОМ ПРИБЛИЖЕНИИ ТОЧЕЧНО-ГЛАДКИХ ФУНКЦИЙ

Рассмотрим вопрос о приближении элементов основного пространства $Y \equiv C_{0;1}^{\{m\};\{p\}}(I)$ с использованием кубических сплайнов.

Зададим на I равномерную сетку

$$\Delta_n : -1 \equiv s_0 < s_1 < \dots < s_n \equiv 1, \quad n = 2, 3, \dots, \quad (3.1)$$

где $s_k \equiv -1 + 2k/n$, $k = \overline{0, n}$, и на ней рассмотрим кубический сплайн

$$z_n(t) \equiv \sum_{i=-1}^{n+1} c_i B_i(t), \quad c_i \in R,$$

удовлетворяющий краевым условиям

$$z_n^{(3)}(s_j - 0) = z_n^{(3)}(s_j + 0), \quad j = 1, n-1. \quad (3.2)$$

Здесь базисные функции $B_i(t)$ суть B -сплайны с носителем (s_{i-2}, s_{i+2}) (см., например, [16, гл. 3, § 8]). Для определения всех функций $B_i(t)$ сетку (3.1) дополним равномерно расположенными узлами:

$s_{-3} < s_{-2} < s_{-1} < s_0 \equiv -1, 1 \equiv s_n < s_{n+1} < s_{n+2} < s_{n+3}$. Обозначим через S_n^3 пространство всех кубических сплайнов $z_n(t)$ на сетке Δ_n , обладающих свойством (3.2), с нормой $\|z_n\|_C$. Далее, пусть

$P_n : C \rightarrow S_n^3$ означает оператор, который всякой функции $f \in C$ ставит в соответствие ее интер-

полюсионный кубический сплайн $P_n f \in S_n^3$ с условием (3.2) такой, что $(P_n f)(s_i) = f(s_i)$, $i = \overline{0, n}$. В книге [16, гл. 3, § 1, теорема 3.1] доказаны существование и единственность интерполяционного кубического сплайна при различных краевых условиях и указан алгоритм построения таких сплайнов. Там же (см. гл. 3, § 5) особо отмечается, что при приближении кубическими сплайнами выбор краевых условий (3.2) является наиболее удачным.

Из теорем 9, 10, 13 в [17, гл. 2, § 4] и соответствующего результата работы [18] (см. там лемму 2) как следствие вытекает

Лемма 3.1. Пусть $r = \overline{1, 4}$ и $f \in C^{(r)} \equiv C^{(r)}(J)$. Тогда

$$\|P_n f - f\|_C = O(n^{-r}), \quad n \rightarrow \infty. \tag{3.3}$$

Пусть $\Pi_q \equiv \text{span}\{t^i\}_0^q$ – класс всех алгебраических полиномов степени, не выше q . Обозначим через $Y_n \equiv \text{span}\{UVB_i\}_{-1}^{n+1} \oplus \Pi_{m+\lambda}$ $(n + m + \lambda + 4)$ -мерное подпространство пространства Y и введем в рассмотрение оператор $\Gamma_n \equiv \Gamma_{n+m+\lambda+4} : Y \rightarrow Y_n$, относящий к любой функции $y \in Y$ обобщенный сплайн $\Gamma_n y \in Y_n$, определяемый условиями

$$(ST\Gamma_n y)(s_i) = (STy)(s_i), \quad i = \overline{0, n},$$

$$(T\Gamma_n y)^{\{j\}}(1) = (Ty)^{\{j\}}(1), \quad j = \overline{0, \lambda},$$

$$(\Gamma_n y)^{\{i\}}(0) = y^{\{i\}}(0), \quad i = \overline{0, m-1},$$

$$(ST\Gamma_n y)^{\{3\}}(s_j - 0) = (ST\Gamma_n y)^{\{3\}}(s_j + 0), \quad j = 1, n-1.$$

Рассуждая так же, как и в [15, гл. 1, § 5, п. 5.3], несложно получить представление

$$\Gamma_n y \equiv \Gamma_{n+m+\lambda+4}(y; t) = (UVP_n STy)(t) + t^m \sum_{j=0}^{\lambda} (Ty)^{\{j\}}(1)(t-1)^j / j! + \sum_{i=0}^{m-1} y^{\{i\}}(0)t^i / i!. \tag{3.4}$$

Лемма 3.2. Γ_n – проектор в пространстве Y .

В силу (3.4) и $P_n^2 = P_n$ данная лемма доказывается аналогично лемме 1.5.1 в [15, гл. 1, § 5]. При этом роль операторов U и T в лемме 1.5.1 играют соответственно UV и ST .

Далее будем использовать следующее обозначение:

$$YC^{(r)} \equiv \{y \in Y \mid STy \in C^{(r)}\},$$

где $r = 0, 1, 2, \dots$; причем $YC^{(0)} \equiv Y$.

Следующее утверждение характеризует скорость сходимости обобщенных интерполяционных сплайнов к интерполируемой функции.

Теорема 1. Если $y \in YC^{(r)}$, $r = \overline{1, 4}$, то

$$\|\Gamma_n y - y\|_Y = O(n^{-r}), \quad n \rightarrow \infty. \tag{3.5}$$

Доказательство. В силу (2.5), (3.4), (2.4), (2.1) и леммы 3.1 последовательно находим

$$\|\Gamma_n y - y\|_Y = \|UV(P_n STy - STy)\|_Y \equiv \|V(P_n STy - STy)\|_{\{t, p\}} \equiv \|P_n STy - STy\|_C \equiv O(n^{-r}).$$

Замечание 1. Очевидно, что из оценки (3.5) и хорошо известной теоремы Банаха–Штейнгауза следует равномерная ограниченность норм операторов $\Gamma_n : \|\Gamma_n\| = O(1)$, $n \rightarrow \infty$.

4. ОБОБЩЕННЫЙ МЕТОД КОЛЛОКАЦИИ С КУБИЧЕСКИМИ СПЛАЙНАМИ (ОМККС)

Пусть задано УТРФО (1.1). Для сокращения громоздких выкладок и упрощения формулировок, не ограничивая общности методов и результатов, всюду в дальнейшем будем считать $l = 1$, $t_1 = 0$, $p_1 = 0$, т.е. рассмотрим уравнение вида

$$Ax \equiv t^m x(t) + \int_{-1}^1 K(t, s)(1 - s)^{-p} x(s) ds = y(t) \quad (t \in I), \tag{4.1}$$

в котором $m \in N$, $p \in R^+$; $y \in Y$, K – известная непрерывная функция, удовлетворяющая следующим условиям:

$$\begin{aligned} K &\in C_s^{\{p\}}(I^2), \quad K(t, \cdot) \in Y, \quad \psi_j(t) \equiv K_s^{\{j\}}(t, 1) \in Y, \\ \tau_i(t) &\equiv K_s^{\{i\}}(t, 0) \in Y, \quad u \equiv S_s K \in C_t^{\{m\}}(I^2), \quad \theta_i(s) \equiv u_t^{\{i\}}(0, s) \in C, \\ v &\equiv T_t u \in C_t^{\{p\}}(I^2), \quad \varphi_j(s) \equiv v_t^{\{j\}}(1, s) \in C, \quad j = \overline{0, \lambda}, \quad i = \overline{0, m-1}; \end{aligned} \tag{4.2}$$

а $x \in X$ – искомая обобщенная функция. Фредгольмовость и достаточные условия непрерывной обратимости оператора $A : X \rightarrow Y$ установлены в работе [7], там же указан метод отыскания точного решения УТРФО (4.1) в классе X .

Приближенное решение уравнения (4.1) построим в виде

$$\begin{aligned} x_n &\equiv x_n(t; \{c_k\}) \equiv f_n(t) + \sum_{i=0}^{m-1} c_{i+\lambda+n+3} \delta^{\{i\}}(t), \\ f_n(t) &\equiv (1-t)^p z_n(t) + \sum_{i=0}^{\lambda} c_{i+n+2} (t-1)^i, \end{aligned} \tag{4.3}$$

где $z_n(t) \equiv \sum_{i=-1}^{n+1} c_i B_i(t)$ – кубический сплайн, рассмотренный выше в разд. 3. Набор $\{c_k\}_{-1}^{n+m+\lambda+2}$ неизвестных параметров найдем, согласно нашему ОМККС, из квадратной системы линейных алгебраических уравнений (СЛАУ) $(n + m + \lambda + 4)$ -го порядка:

$$\begin{aligned} (ST\rho_n)(s_i) &= 0, \quad i = \overline{0, n}, \quad (T\rho_n)^{\{j\}}(1) = 0, \quad j = \overline{0, \lambda}, \quad \rho_n^{\{i\}}(0) = 0, \quad i = \overline{0, m-1}, \\ (STUx_n)^{(3)}(s_k - 0) &= (STUx_n)^{(3)}(s_k + 0), \quad k = 1, n-1, \end{aligned} \tag{4.4}$$

где $\rho_n(t) \equiv \rho_n^A(t) \equiv (Ax_n - y)(t)$ – невязка приближенного решения, а $\{s_i\}_0^n$ – использованная ранее система узлов коллокации, порождающая сетку (3.1).

При обосновании предложенного метода полезную роль играют функции

$$\Psi_i(t) \equiv \sum_{l=0}^i \binom{i}{l} \tau_l(t) \prod_{k=0}^{i-l-1} (p+k), \quad i = \overline{0, m-1}.$$

Для вычислительного алгоритма (4.1)–(4.4) справедлива

Теорема 2. Пусть однородное УТРФО $Ax = 0$ имеет в X лишь нулевое решение (например, в условиях теоремы 3 в [7]), а исходные данные таковы, что $u \equiv S_s K$ (по t), $\psi_j, \Psi_i, y \in YC^{(r)}$, $r = \overline{1, 4}$, $j = \overline{0, \lambda}$, $i = \overline{0, m-1}$. Тогда при всех $n \in N$, $n \geq n_0$ СЛАУ (4.4) обладает единственным решением $\{c_k^*\}$ и последовательность приближенных решений $x_n^* \equiv x_n(t; \{c_k^*\})$ сходится к точному решению $x^* = A^{-1}y$ по норме пространства X со скоростью

$$\Delta x_n^* \equiv \|x_n^* - x^*\| = O(n^{-r}), \quad r = \overline{1, 4}. \tag{4.5}$$

Доказательство. УТРФО (4.1) представляется в виде линейного операторного уравнения

$$Ax \equiv Ux + Kx = y \quad (x \in X \equiv D^{\{p\}}\{m; 0\}, \quad y \in Y \equiv C_{0;1}^{\{m\};\{p\}}), \tag{4.6}$$

в котором оператор $A : X \rightarrow Y$ непрерывно обратим.

Систему (4.3), (4.4) требуется записать также в операторной форме. С этой целью построим соответствующие конечномерные подпространства в виде

$$X \supset X_n \equiv V(S_n^3) \oplus \text{span}\{(t-1)^i\}_0^\lambda \oplus \text{span}\{\delta^{[i]}(t)\}_0^{m-1},$$

$$Y \supset Y_n \equiv UV(S_n^3) \oplus \Pi_{m+\lambda}.$$

Тогда, следуя рассуждениям при доказательстве теоремы 4.3.1 (см. [15, гл. 4, § 3]), несложно показать, что вычислительная схема (4.3), (4.4) ОМККС равносильна линейному операторному уравнению

$$A_n x_n \equiv \Gamma_n A x_n = \Gamma_n y \quad (x_n \in X_n, \Gamma_n y \in Y_n), \tag{4.7}$$

где $\Gamma_n : Y \rightarrow Y_n$ – “сплайновый” оператор, введенный и подробно изученный в разделе 3. Следовательно, для доказательства утверждений теоремы 2 достаточно установить существование, единственность и сходимости решений уравнений (4.7).

Уточним структуру аппроксимирующего уравнения (4.7). Поскольку в силу леммы 3.2 $\Gamma_n^2 = \Gamma_n$, имеем $\Gamma_n U x_n = U x_n \in Y_n$ при любом элементе $x_n \in X_n$. Таким образом, система (4.3), (4.4) эквивалентна линейному уравнению вида

$$A_n x_n \equiv U x_n + \Gamma_n K x_n = \Gamma_n y \quad (x_n \in X_n, \Gamma_n y \in Y_n). \tag{4.8}$$

Обсудим теперь вопрос близости операторов A и A_n на подпространстве X_n . Используя уравнения (4.6) и (4.8), представления (2.5) и (3.4), а также нормы (2.4) и (2.1), для произвольного элемента $x_n \in X_n$ находим, что

$$\|A x_n - A_n x_n\|_Y = \|K x_n - \Gamma_n K x_n\|_Y = \|STK x_n - P_n STK x_n\|_C. \tag{4.9}$$

На основании (4.1), (4.2) и (2.6) имеем

$$(Kx)(t) = \int_{-1}^1 u(t,s)z(s)ds + \sum_{j=0}^{\lambda} \mu_j(z)\psi_j(t) + \sum_{i=0}^{m-1} (-1)^i \gamma_i \Psi_i(t),$$

где

$$\mu_j(z) \equiv \int_{-1}^1 (Sz)(s)(s-1)^j \frac{1}{j!} ds + \sum_{k=0}^{\lambda} z^{\{k\}}(1)\beta_{jk},$$

$$\beta_{jk} \equiv \int_{-1}^1 (s-1)^{j+k} \frac{1}{j!k!} (1-s)^{-p} ds, \quad j, k = \overline{0, \lambda}.$$

Тогда, с учетом (4.3) и (4.2), получаем, что

$$STK x_n = \int_{-1}^1 h(t,s) f_n(s) ds + \sum_{j=0}^{\lambda} \mu_j(f_n) \alpha_j(t) + \sum_{i=0}^{m-1} (-1)^i c_{i+\lambda+n+3} \beta_i(t), \tag{4.10}$$

где $h \equiv S_V$, $\alpha_j \equiv ST\psi_j$, $\beta_i \equiv ST\Psi_i$, $j = \overline{0, \lambda}$, $i = \overline{0, m-1}$.

В силу (4.10), (3.3) и определения (2.7) последовательно выводим следующую аппроксимативную оценку (здесь и далее d_i , $i = \overline{1, 4}$ – определенные константы, не зависящие от натурального n):

$$\begin{aligned} \|STK x_n - P_n STK x_n\|_C &= \max_{t \in I} \left| \int_{-1}^1 (h - P_n^t h)(t,s) f_n(s) ds + \sum_j \mu_j(f_n)(\alpha_j - P_n \alpha_j)(t) + \right. \\ &+ \left. \sum_i (-1)^i c_{i+\lambda+n+3} (\beta_i - P_n \beta_i)(t) \right| \leq 2 \|f_n\|_C d_1 n^{-r} + \sum_j |\mu_j(f_n)| d_1 n^{-r} + \sum_i |c_{i+\lambda+n+3}| d_1 n^{-r} \leq \\ &\leq 2^{p+1} \|f_n\|_{\{p\}} d_1 n^{-r} + (2^{\lambda+1} + \beta)(\lambda + 1) \|f_n\|_{\{p\}} d_1 n^{-r} + m \|x_n\|_X d_1 n^{-r} \leq d_2 n^{-r} \|x_n\|_X, \end{aligned} \tag{4.11}$$

где $d_2 \equiv [2^{p+1} + (2^{\lambda+1} + \beta)(\lambda + 1) + m]d_1$, $\beta \equiv \max_{0 \leq j, k \leq \lambda} |\beta_{jk}|$.

Из равенства (4.9) и оценки (4.11) следует, что

$$\varepsilon_n \equiv \|A - A_n\|_{X_n \rightarrow Y} \leq d_2 n^{-r}, \quad r = \overline{1,4}. \quad (4.12)$$

Тогда, благодаря неравенствам (4.12) и (3.5), из теоремы 7 (см. [13, гл. 1, § 4]) следует утверждение теоремы 2 с оценкой (4.5). Требуемое доказано.

В дальнейшем при оптимизации прямых проекционных методов решения УТРФО (4.1) существенную роль будет играть

Теорема 3. Пусть УТРФО (4.1) имеет решение вида

$$x^*(t) \equiv z^*(t) + \sum_{i=0}^{m-1} \gamma_i^* \delta^{[i]}(t), \quad Sz^* = STUx^* \in C^{(r)}, \quad r = \overline{1,4}, \quad (4.13)$$

при данном $y \in Y$ и соответствующий аппроксимирующий оператор A_n в ОМККС непрерывно обратим. Тогда погрешность приближенного решения $x_n^* \in X_n$ для правой части $\Gamma_n y \in Y_n$ представима в виде

$$\Delta x_n^* = O\{\|Sz^* - P_n Sz^*\|_C\} = O(n^{-r}), \quad r = \overline{1,4}. \quad (4.14)$$

Доказательство. В силу теоремы 6 (см. [13, гл. 1, § 3]) и структуры приближенного уравнения (4.8) имеем

$$\Delta x_n^* = O\{\|\Gamma_n\| \|x^* - x_n\|\}, \quad (4.15)$$

где $x_n \in X_n$ – пока произвольный элемент. Выберем его следующим образом:

$$x_n(t) \equiv (VP_n STUx^*)(t) + \sum_{j=0}^{\lambda} (TUx^*)^{[j]}(1)(t-1)^j / j! + \sum_{i=0}^{m-1} \gamma_i^* \delta^{[i]}(t). \quad (4.16)$$

Тогда требуемая оценка (4.14) следует из (4.15), (4.13), (2.3), (4.16), (2.7), (2.1), леммы 3.1 с учетом замечания 1:

$$\Delta x_n^* \leq d_3 \|VSz^* - VP_n Sz^*\|_X \equiv d_3 \|VSz^* - VP_n Sz^*\|_{\{p\}} \equiv d_3 \|Sz^* - P_n Sz^*\|_C \leq d_4 n^{-r}, \quad r = \overline{1,4}.$$

5. К ОПТИМИЗАЦИИ ПРЯМЫХ ПРОЕКЦИОННЫХ МЕТОДОВ РЕШЕНИЯ УТРФО

Предварительно приведем необходимые определения и постановку задачи. Пусть X и Y – банаховы пространства, а X_n и Y_n – их соответствующие произвольные подпространства одинаковой размерности $N = N(n) < +\infty$, $n \in N$, причем $N \rightarrow \infty$ при $n \rightarrow \infty$. Обозначим через $\Lambda_n \equiv \{\lambda_n\}$ некоторое множество линейных операторов λ_n , отображающих Y на Y_n . Далее рассмотрим два класса однозначно разрешимых линейных операторных уравнений

$$Ax = y, \quad x \in X, \quad y \in Y, \quad (5.1)$$

и

$$\lambda_n Ax_n = \lambda_n y, \quad x_n \in X_n, \quad \lambda_n \in \Lambda_n, \quad n \in N, \quad (5.2)$$

соответственно. Пусть $x^* \in X$ и $x_n^* \in X_n$ – решения уравнений (5.1) и (5.2) соответственно, а $F \equiv \{f\}$ – класс коэффициентов (т.е. исходных данных) уравнения (5.1), порождающий класс $X^* \equiv \{x^*\}$ искомых элементов.

Следуя работе [13, гл. 2, § 1], величину

$$V_N(F) \equiv \inf_{X_n, Y_n} \inf_{\lambda_n \in \Lambda_n} V(F; \lambda_n; X_n, Y_n), \quad (5.3)$$

где

$$V(F; \lambda_n; X_n, Y_n) \equiv \sup_{f \in F} (f; \lambda_n; X_n, Y_n) = \sup_{x^* \in X^*} \|x^* - x_n^*\|_X,$$

назовем *оптимальной оценкой погрешности* всевозможных прямых проекционных методов ($\lambda_n \in \Lambda_n$) решения уравнения (5.1) на классе F .

Определение 1 (см. [13, гл. 2, § 1]). Пусть существуют подпространства $X_n^0 \subset X, Y_n^0 \subset Y$ размерности $N = N(n) < +\infty$ и операторы $\lambda_n^0 : Y \rightarrow Y_n^0, \lambda_n^0 \in \Lambda_n$, при которых выполняется условие

$$V_N(F) \succ\prec V(F; \lambda_n^0; X_n^0, Y_n^0) \quad (N \rightarrow \infty), \tag{5.4}$$

где символ $\succ\prec$ означает, как обычно, слабую эквивалентность. Тогда метод (5.1), (5.2) при $X_n = X_n^0, Y_n = Y_n^0$ и $\lambda_n = \lambda_n^0$ называется *оптимальным по порядку точности* на классе F среди всех прямых проекционных методов λ_n ($\lambda_n \in \Lambda_n$) решения уравнений (5.1).

Рассмотрим теперь оптимизацию по порядку точности на классе однозначно разрешимых (равномерно относительно $K \in F$) УТРО (4.1) в случае, когда исходные данные принадлежат семейству $YC^{(r)}$, т.е. при $u \equiv S_3 K$ (по t), $\psi_j, j = \overline{0, \lambda}, \Psi_i, i = \overline{0, m-1}, y \in YC^{(r)}, r = \overline{1, 4}$. Тогда на основании теоремы 3 из [7] имеем

$$X^* \equiv \{x^* \in X \mid Ax^* = y; u, \psi_j, \Psi_i, y \in YC^{(r)}\} = XC^{(r)},$$

где $XC^{(r)} \equiv \{x \in X \mid STUx \in C^{(r)}\}$.

Далее пусть

$$X_n^0 \equiv V(S_n^3) \oplus \text{span}\{(t-1)^i\}_0^\lambda \oplus \text{span}\{\delta^{[i]}(t)\}_0^{m-1},$$

$$Y_n^0 \equiv UV(S_n^3) \oplus \Pi_{m+\lambda},$$

а $\Lambda_n^0 \equiv \{\lambda_n\}$ – семейство всех линейных операторов $\lambda_n : Y \rightarrow Y_n^0$.

Теорема 4. Пусть $F = YC^{(r)}, \Lambda_n = \Lambda_n^0$. Тогда

$$V_N(F) \succ\prec N^{-r}, \quad N = n + m + \lambda + 4, \quad r = \overline{1, 4}, \tag{5.5}$$

и этот оптимальный по точности порядок реализует ОМККС.

Доказательство. Заметим, что из определения N -го колмогоровского поперечника $d_N(L, X)$ множества L в нормированном пространстве $X \equiv D^{[p]} \{m; 0\}$ (см., например, [19, гл. 1, § 1]) и теоремы 1.3.6 (см. [4, гл. 1, § 1.3]) следует равенство

$$d_N(L, X) = d_{N-m-\lambda}(STU(L), C), \quad N > m + \lambda,$$

откуда, с учетом $d_l(C^{(r)}, C) \succ\prec l^{-r}$ ($l \in N$) (см., например, [19, гл. 3, § 3]), вытекает слабая эквивалентность

$$d_N(XC^{(r)}, X) \succ\prec N^{-r}. \tag{5.6}$$

Далее, известно (см. [13, гл. 4, § 2]), что $V_N(F) \geq d_N(X^*, X)$. Следовательно, из (5.6) следует, что

$$V_N(F) \geq d_N(XC^{(r)}, X) \succ\prec N^{-r}. \tag{5.7}$$

С другой стороны, согласно (5.3) и теоремам 2 и 3 находим оценку

$$V_N(F) \leq \sup_{x^* \in XC^{(r)}} \|x^* - x_n^*\|_X = O(N^{-r}), \quad x_n^* = A_n^{-1} \Gamma_n y.$$

Отсюда и из соотношений (5.7), (5.4) получаем утверждение теоремы 4 с оценкой (5.5). Требуемое доказано.

6. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Замечание 2. В силу определения нормы в $X \equiv D^{[p]} \{m; 0\}$ нетрудно заметить, что из сходимости последовательности (x_n^*) приближенных решений к точному решению $x^* = A^{-1}y$ в метрике X следует обычная сходимость в пространстве обобщенных функций, т.е. слабая сходимость.

Замечание 3. При приближении решений операторных уравнений $Ax = y$ возникает естественный вопрос о скорости сходимости невязки $\rho_n^*(t) \equiv (Ax_n^* - y)(t)$ исследуемого метода. Один из результатов в этом направлении легко получить из теоремы 2, а именно, из нее вытекает простое следствие: если исходные данные уравнения (4.1) принадлежат классу $YC^{(r)}$, $r = \overline{1, 4}$, то в условиях теоремы 2 справедлива оценка

$$\|\rho_n^*\|_Y = O(n^{-r}), \quad r = \overline{1, 4}.$$

Замечание 4. Поскольку $C_{0;1}^{\{0\};\{p\}} \equiv C\{p;1\} \equiv D^{\{p\}}\{0;0\}$, то при $m = 0$ УТРФО (4.1) превращается в интегральное уравнение II рода в $C\{p;1\}$ с фиксированной особенностью в ядре, а предложенный метод (4.3), (4.4) – в соответствующий вариант метода коллокации с кубическими сплайнами, причем $h \equiv S_j S_s K$, $\alpha_j \equiv S\psi_j$, $j = \overline{0, \lambda}$, $STy \equiv Sy$. Следовательно, теорема 2 содержит в себе соответствующие результаты по обоснованию данного варианта метода коллокации для приближенного решения уравнений II рода с особенностью в ядре; при этом погрешность характеризуется неравенством $\|x_n^* - x^*\|_{\{p\}} = O(n^{-r})$, $r = \overline{1, 4}$.

Замечание 5. Если $m = p = 0$, то $C_{0;1}^{\{0\};\{0\}} \equiv C \equiv D^{\{0\}}\{0;0\}$ и из УТРФО (4.1) получается интегральное уравнение II рода в пространстве C . При этом метод (4.3), (4.4) становится соответствующим методом кубической сплайн-коллокации для уравнения II рода, причем $h \equiv K$, $STy \equiv y$. Поэтому теорема 2 охватывает обоснование указанного метода при приближенном решении уравнений II рода в C . Соответствующая оценка погрешности имеет вид $\|x_n^* - x^*\|_C = O(n^{-r})$, $r = \overline{1, 4}$.

Замечание 6. Так как в условиях теоремы 2 соответствующие аппроксимирующие операторы A_n обладают свойством вида

$$\|A_n^{-1}\| = O(1), \quad A_n^{-1} : Y_n \rightarrow X_n, \quad n \geq n_1,$$

то (см. [13, гл. 1, § 5]) очевидно, что предложенный в настоящей работе прямой метод для УТРФО (4.1) устойчив относительно малых возмущений исходных данных. Это позволяет найти численное решение исследуемых уравнений на ЭВМ с любой наперед заданной степенью точности. Более того, если УТРФО (4.1) хорошо обусловлено, то хорошо обусловленной является также СЛАУ (4.4).

СПИСОК ЛИТЕРАТУРЫ

1. Адамар Ж. Задача Коши для линейных уравнений с частными производными гиперболического типа. М.: Наука, 1978. 352 с.
2. Bart G.R., Warnock R.L. Linear integral equations of the third-kind // SIAM J. Math. Anal. 1973. V. 4. № 4. P. 609–622.
3. Кейз К.М., Цвайфель П.Ф. Линейная теория переноса. М.: Мир, 1972. 384 с.
4. Замалиев Р.Р. О прямых методах решения интегральных уравнений третьего рода с особенностями в ядре: Дисс. ... канд. физ.-матем. наук. Казань: КФУ, 2012. 114 с.
5. Расламбеков С.Н. Сингулярное интегральное уравнение первого рода в исключительном случае в классах обобщенных функций // Изв. вузов. Математика. 1983. № 10. С. 51–56.
6. Бжухатлов Х.Г. Об одной краевой задаче со смещением // Дифференц. ур-ния. 1973. Т. 9. № 1. С. 162–165.
7. Габбасов Н.С. Методы решения интегрального уравнения третьего рода с фиксированными особенностями в ядре // Дифференц. ур-ния. 2009. Т. 45. № 9. С. 1341–1348.
8. Габбасов Н.С., Замалиев Р.Р. Новый вариант метода подобластей для интегральных уравнений третьего рода с особенностями в ядре // Изв. вузов. Математика. 2011. № 5. С. 12–18.
9. Габбасов Н.С. Новый вариант метода коллокации для одного класса интегральных уравнений в особом случае // Дифференц. ур-ния. 2013. Т. 49. № 9. С. 1178–1185.
10. Габбасов Н.С. Специальный прямой метод решения интегральных уравнений в особом случае // Дифференц. ур-ния. 2014. Т. 50. № 9. С. 1245–1252.
11. Габбасов Н.С., Галимова З.Х. К численному решению интегральных уравнений третьего рода с особенностями в ядре // Изв. вузов. Математика. 2016. № 12. С. 36–45.

12. *Габбасов Н.С., Галимова З.Х.* Специальный вариант метода коллокации для интегральных уравнений третьего рода с неподвижными особенностями в ядре // Изв. вузов. Математика. 2018. № 5. С. 20–27.
13. *Габдулхаев Б.Г.* Оптимальные аппроксимации решений линейных задач. Казань: Изд-во Казанск. ун-та, 1980. 232 с.
14. *Пресдорф З.* Сингулярное интегральное уравнение с символом, обращающимся в нуль в конечном числе точек // Матем. исследования. 1972. Т. 7. № 1. С. 116–132.
15. *Габбасов Н.С.* Методы решения интегральных уравнений Фредгольма в пространствах обобщенных функций. Казань.: Изд-во Казанск. ун-та, 2006. 176 с.
16. *Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.* Методы сплайн-функций. М.: Наука, 1980. 352 с.
17. *Стечкин С.Б., Субботин Ю.Н.* Сплаины в вычислительной математике. М.: Наука, 1976. 248 с.
18. *Педас А., Тимак Э.* Метод кубической сплайн-коллокации для решения слабо сингулярных интегральных уравнений // Дифференц. ур-ния. 2001. Т. 37. № 10. С. 1415–1424.
19. *Даугавет И.К.* Введение в теорию приближения функций. Л.: Изд-во ЛГУ, 1977. 184 с.

СХОДИМОСТЬ АТТРАКТОРОВ АППРОКСИМАЦИИ К АТТРАКТОРАМ МОДИФИЦИРОВАННОЙ МОДЕЛИ КЕЛЬВИНА–ФОЙГТА¹⁾

© 2022 г. М. В. Турбин^{1,*}, А. С. Устюжанинова^{1,**}

¹ 394018 Воронеж, Университетская пл., 1, Воронежский государственный ун-т, Россия

*e-mail: mrmike@mail.ru

**e-mail: nastyzhka@gmail.com

Поступила в редакцию 30.04.2021 г.
Переработанный вариант 10.07.2021 г.
Принята к публикации 12.10.2021 г.

Работа посвящена исследованию качественного поведения решений модифицированной модели Кельвина–Фойгта. Для этой модели рассматривается аппроксимация и доказываются существование минимального траекторного и глобального аттракторов как для самой модели, так и для ее аппроксимации. Затем показывается, что траекторные и глобальные аттракторы аппроксимации сходятся к траекторным и глобальным аттракторам исходной модели в смысле полуклонения в соответствующих пространствах при стремлении параметра аппроксимации к нулю. Библ. 24.

Ключевые слова: аттракторы, сходимость аттракторов, пространство траекторий, модифицированная модель Кельвина–Фойгта, слабое решение.

DOI: 10.31857/S0044466922020120

1. ВВЕДЕНИЕ

Теория аттракторов динамических систем создана уже давно и довольно хорошо развита. Она получила применение в ряде задач гидродинамики, для которых имеет место теорема единственности решения. В качестве наиболее известных примеров можно привести результаты О.А. Ладыженской для системы Навье–Стокса в двумерном случае [1], [2] и Г.А. Серегина для модели Бингама в двумерном случае [3]. При этом требование единственности решения является очень ограничительным и не выполняется для подавляющего большинства моделей гидродинамики. В связи с этим М.И. Вишиком и В.В. Чепыжовым [4], а также независимо от них Дж. Селлом [5], была создана теория инвариантных траекторных аттракторов и с ее помощью удалось исследовать аттракторы системы Навье–Стокса в трехмерном случае и ряда других систем. В дальнейшем эта теория была усовершенствована В.Г. Звягиным и Д.А. Воротниковым в работе [6] (подробное изложение см. в монографии [7]). А именно, на основе аппроксимационно-топологического метода исследования задач гидродинамики удалось отказаться от требования трансляционной инвариантности пространства траекторий. На основе этой теории аттракторов был исследован целый ряд моделей неньютоновской гидродинамики (см., например, обзорную статью [8] и имеющуюся там библиографию).

Отметим, что с прикладной точки зрения также важно уметь приближенно находить аттракторы изучаемых систем. Однако это вызывает сложности в связи с “плохими” математическими свойствами рассматриваемых моделей. Поэтому представляет интерес вопрос об аппроксимации аттракторов рассматриваемых моделей аттракторами моделей с “хорошими” свойствами. А именно, рассматривается некоторая аппроксимационная задача, для которой имеют место теорема единственности решений и свойство непрерывной зависимости решений от данных задачи. Следовательно, для этой аппроксимационной задачи можно будет воспользоваться различными численными методами. После чего показывается, что аттракторы этой аппроксимационной задачи сходятся к аттракторам исходной модели в смысле полуклонения. Это должно позволить получить численно представление об аттракторах изучаемой модели.

¹⁾Работа М.В. Турбина выполнена при финансовой поддержке РФФИ (проект 20-01-00051), работа А.С. Устюжаниновой выполнена при финансовой поддержке Минобрнауки РФ (проект FZGU-2020-0035).

В работе рассматриваются вопросы сходимости к аттракторам модифицированной модели Кельвина–Фойгта аттракторов ее аппроксимационной задачи.

2. ОПИСАНИЕ РАССМАТРИВАЕМОЙ МОДЕЛИ

В ограниченной области $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) с границей $\partial\Omega$ класса C^3 рассматривается следующая система уравнений:

$$\frac{\partial v}{\partial t} - \nu \Delta v + \sum_{i=1}^n v_i \frac{\partial v}{\partial x_i} - \kappa \frac{\partial \Delta v}{\partial t} - \kappa \sum_{k=1}^n v_k \frac{\partial \Delta v}{\partial x_k} + \nabla p = f; \quad (2.1)$$

$$\operatorname{div} v = 0. \quad (2.2)$$

Здесь $v(x, t)$ – вектор скорости частицы жидкости, находящейся в точке x в момент времени t ; $p(x, t)$ – давление жидкости в точке x в момент времени t ; $f(x, t)$ – вектор плотности внешних сил; $\nu > 0$, $\kappa > 0$ – вязкость жидкости и время релаксации соответственно. Неизвестными функциями являются v и p .

Рассматриваемая модель впервые была введена В.А. Павловским [9]. Она была подтверждена рядом экспериментальных исследований растворов полиэтиленоксида, полиакриламида и гуаровой смолы [10], [11].

Для системы (2.1), (2.2) рассмотрим начально-краевую задачу с начальным и граничным условиями

$$v|_{t=0} = a, \quad (2.3)$$

$$v|_{\partial\Omega} = 0. \quad (2.4)$$

В задаче (2.1)–(2.4) известные параметры ν , κ , а также плотность внешних сил f мы считаем раз и навсегда зафиксированными.

Для рассматриваемой модифицированной модели Кельвина–Фойгта в работе [12] было доказано существование слабого решения на произвольном конечном промежутке времени $[0, T]$. Существование траекторного и глобального аттракторов доказано в [13]. Также в работах [14], [15] было доказано существование решения задачи оптимального управления с обратной связью для этой модели.

Важно отметить, что для слабых решений начально-краевой задачи (2.1)–(2.4) не доказана теорема единственности. При этом для сильных решений рассматриваемой задачи теорема единственности имеет место [11], но существование сильных решений начально-краевой задачи (2.1)–(2.4) до сих пор не установлено.

3. НЕОБХОДИМЫЕ ОПРЕДЕЛЕНИЯ И ВСПОМОГАТЕЛЬНЫЕ УТВЕРЖДЕНИЯ

Приведем необходимые нам определения.

Определение 1. Пусть на линейном пространстве X определена не более чем счетная система полунорм \mathcal{P}_X такая, что из условия $p(x) = 0$ для всех $p \in \mathcal{P}_X$ следует, что $x = \theta$. Тогда пара (X, \mathcal{P}_X) называется *счетно-нормированным* пространством.

Нам потребуется ряд фактов из теории аттракторов (подробнее см. [7], [8]).

Через E , E_0 будем обозначать два банаховых пространства. Будем предполагать, что пространство E рефлексивно и вложение $E \subset E_0$ – непрерывно. Через \mathbb{R}_+ будем обозначать неотрицательную полуось числовой прямой \mathbb{R} .

Пространство $C(\mathbb{R}_+; E_0)$ состоит из непрерывных функций, определенных на \mathbb{R}_+ и принимающих значения в E_0 . Так как полуось \mathbb{R}_+ некомпактна, то в линейном пространстве $C(\mathbb{R}_+; E_0)$ нельзя задать обычную норму пространства непрерывных функций. Рассмотрим в пространстве $C(\mathbb{R}_+; E_0)$ семейство полунорм:

$$\|u\|_n = \|u\|_{C([0, n], E_0)}, \quad n = 1, 2, \dots, \quad (3.1)$$

и зададим топологию в $C(\mathbb{R}_+; E_0)$ при помощи определения сходимости последовательностей относительно введенных полунорм. А именно, последовательность $\{u_m\}$ из $C(\mathbb{R}_+; E_0)$ сходится к

функции u при $m \rightarrow \infty$, если $\|u_m - u\|_n \rightarrow 0$ при любом $n = 1, 2, \dots$. Пространство $C(\mathbb{R}_+; E_0)$ с семейством полунорм (3.1) является счетно-нормированным пространством. Топология локальной равномерной сходимости в пространстве $C(\mathbb{R}_+; E_0)$ является метризуемой относительно метрики

$$\rho(u, v) = \|u - v\|_{C(\mathbb{R}_+; E_0)} = \sum_{n=1}^{\infty} 2^{-n} \frac{\|v\|_{C([0, n], E_0)}}{1 + \|v\|_{C([0, n], E_0)}}.$$

Полученное метрическое пространство полно, т.е. является пространством Фреше.

В работе мы используем уже устоявшееся в работах по аттракторам неинвариантных пространств траекторий обозначение $\|u - v\|_{C(\mathbb{R}_+; E_0)}$ для метрики в $C(\mathbb{R}_+; E_0)$. Это связано с использованием абстрактных понятий и утверждений из работ [7], [8], [19], в которых используется данное обозначение (см. также замечания 1 и 2). Отметим, что функционал $\|\cdot\|_{C(\mathbb{R}_+; E_0)}$ не является нормой, так как $\|\lambda v\|_{C(\mathbb{R}_+; E_0)} \neq |\lambda| \|v\|_{C(\mathbb{R}_+; E_0)}$ при $\lambda \neq \pm 1$.

Обозначим через Π_M ($M \geq 0$) оператор сужения функций, заданных на \mathbb{R}_+ , на отрезок $[0, M]$. Имеет место следующий критерий относительной компактности множеств из $C(\mathbb{R}_+; E_0)$ (напомним, что множество P называется относительно компактным, если его замыкание компактно).

Лемма 1. Для того чтобы множество $P \subset C(\mathbb{R}_+; E_0)$ было относительно компактно в $C(\mathbb{R}_+; E_0)$ необходимо и достаточно, чтобы при любом $M > 0$ множество $\Pi_M P$ было относительно компактно в $C([0, M], E_0)$.

Обозначим через $L_\infty(\mathbb{R}_+; E)$ пространство всех существенно ограниченных функций, определенных на \mathbb{R}_+ и принимающих значение в E (т.е. для любой $u \in L_\infty(\mathbb{R}_+; E)$ найдется число $M_u < \infty$ такое, что $\|u(t)\|_E \leq M_u$ при почти всех $t \in \mathbb{R}_+$), с нормой $\|u\|_{L_\infty(\mathbb{R}_+; E)} = \operatorname{vrai} \max_{t \in \mathbb{R}_+} \|u(t)\|_E$ (здесь $\operatorname{vrai} \max_{t \in \mathbb{R}_+} \|u(t)\|_E$ — это нижняя грань всех M_u). Пространство $L_\infty(\mathbb{R}_+; E)$ является банаховым относительно указанной нормы (см. [17]).

Определение 2. Пусть J — конечный или бесконечный интервал вещественной оси и \bar{J} — его замыкание. Далее, пусть Y — банахово пространство. Функция $u : \bar{J} \rightarrow Y$ называется слабонепрерывной, если из $t_n \rightarrow t$, $t_n \in \bar{J}$ следует, что $u(t_n) \rightarrow u(t)$ слабо в Y . Множество слабонепрерывных функций $u : \bar{J} \rightarrow Y$ мы будем обозначать через $C_w(\bar{J}, Y)$.

Также нам потребуется одна известная (см., например, [18])

Теорема 1. Пусть E и E_0 — два банаховых пространства таких, что $E \subset E_0$, причем вложение непрерывно. Если функция v принадлежит $L_\infty(0, T; E)$ и непрерывна как функция со значениями в E_0 , то w слабонепрерывна как функция со значениями в E , т.е. $v \in C_w([0, T], E)$.

Следовательно, функции, принадлежащие классу $C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$, слабонепрерывны со значениями в E (и потому их значения принадлежат пространству E при всех $t \in \mathbb{R}_+$); они ограничены со значениями в E , и для $u \in C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$ верно равенство

$$\|u\|_{C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)} = \sup_{t \in \mathbb{R}_+} \|u(t)\|_E.$$

Рассмотрим операторы сдвигов $T(h)$ ($h \geq 0$), каждый из которых функции f ставит в соответствие функцию $T(h)f$ такую, что $T(h)f(t) = f(t + h)$. Отметим, что имеет место тождество $T(h_1)T(h_2) = T(h_1 + h_2)$, а также что $T(0)$ — тождественный оператор.

Рассмотрим непустое семейство функций

$$\mathcal{H}^+ \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E).$$

Множество \mathcal{H}^+ будем называть *пространством траекторий*, а его элементы — *траекториями*. Будем предполагать, что \mathcal{H}^+ непусто.

Приведем основные определения.

Определение 3. Полуотклонением множества C от множества D в метрическом пространстве (X, ρ) называется число

$$h(C, D) = \sup_{c \in C} \text{dist}(c, D) = \sup_{c \in C} \inf_{d \in D} \rho(c, d).$$

Определение 4. Множество $P \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$ называется притягивающим (для пространства траекторий \mathcal{H}^+), если для всякого множества $B \subset \mathcal{H}^+$, ограниченного в $L_\infty(\mathbb{R}_+; E)$, выполняется условие

$$\sup_{u \in B} \inf_{v \in P} \|T(h)u - v\|_{C(\mathbb{R}_+; E_0)} \rightarrow 0 \quad (h \rightarrow \infty). \quad (3.2)$$

Замечание 1. Отметим, что в силу определения сходимости в $C(\mathbb{R}_+; E_0)$ условие (3.2) эквивалентно условию

$$\sup_{u \in B} \inf_{v \in P} \|T(h)u - v\|_{C([0, n], E_0)} \rightarrow 0 \quad (h \rightarrow \infty)$$

для любого $n = 1, 2, \dots$

Замечание 2. Отметим, что в (3.2) величина $\inf_{v \in P} \|T(h)u - v\|_{C(\mathbb{R}_+; E_0)}$ представляет собой расстояние от точки $T(h)u$ до множества P в пространстве $C(\mathbb{R}_+; E_0)$. В свою очередь величина $\sup_{u \in B} \inf_{v \in P} \|T(h)u - v\|_{C(\mathbb{R}_+; E_0)}$ является полуотклонением множества $T(h)u$ от множества P в $C(\mathbb{R}_+; E_0)$.

Определение 5. Множество $P \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$ называется поглощающим (для пространства траекторий \mathcal{H}^+), если для всякого множества $B \subset \mathcal{H}^+$, ограниченного в $L_\infty(\mathbb{R}_+; E)$, существует $h \geq 0$ такое, что при всех $t \geq h$ имеет место включение $T(t)B \subset P$.

Отметим, что любое поглощающее множество является притягивающим.

Определение 6. Множество $P \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$ называется траекторным аттрактором (пространства траекторий \mathcal{H}^+), если оно удовлетворяет следующим условиям:

- (i) множество P компактно в $C(\mathbb{R}_+; E_0)$ и ограничено в $L_\infty(\mathbb{R}_+; E)$;
- (ii) имеет место равенство $T(t)P = P$ для всех $t \geq 0$;
- (iii) множество P является притягивающим в смысле определения 4.

Определение 7. Минимальным траекторным аттрактором пространства траекторий \mathcal{H}^+ называется наименьший по включению траекторный аттрактор.

Определение 8. Множество $\mathcal{A} \subset E$ называется глобальным аттрактором (в E_0) пространства траекторий \mathcal{H}^+ , если оно удовлетворяет следующим условиям:

- (i) множество \mathcal{A} компактно в E_0 и ограничено в E ;
- (ii) для всякого ограниченного в $L_\infty(\mathbb{R}_+; E)$ множества $B \subset \mathcal{H}^+$ выполняется условие притягивания

$$\sup_{u \in B} \inf_{y \in \mathcal{A}} \|u(t) - y\|_{E_0} \rightarrow 0 \quad (t \rightarrow \infty);$$

- (iii) множество \mathcal{A} является наименьшим по включению множеством, удовлетворяющим условиям (i) и (ii).

Замечание 3. Если существует минимальный траекторный аттрактор или глобальный аттрактор, то он единственный.

Теорема 2. Пусть существует минимальный траекторный аттрактор \mathcal{U} пространства траекторий \mathcal{H}^+ . Тогда существует глобальный аттрактор \mathcal{A} пространства \mathcal{H}^+ , и справедливо соотношение $\mathcal{A} = \mathcal{U}(t)$, $t \geq 0$.

Сходимость аттракторов для конкретных пространств траекторий мы будем доказывать на основе следующих абстрактных утверждений (см. [19]).

Квадратными скобками в дальнейшем будем обозначать замыкание в топологии локальной сходимости пространства $C(\mathbb{R}_+; E_0)$.

Определение 9. Для множества $P \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E)$, ограниченного в $L_\infty(\mathbb{R}_+; E)$, назовем ω -предельным множеством следующее множество $\omega(P) = \bigcap_{t \geq 0} \left[\bigcup_{s \geq t} T(s)P \right]$.

Лемма 2. Пусть существует поглощающее множество $P \subset \mathcal{H}^+$, относительно компактное в $C(\mathbb{R}_+; E_0)$, ограниченное в $L_\infty(\mathbb{R}_+; E)$ и трансляционно инвариантное ($T(h)P \subset P(h \geq 0)$). Тогда $\omega(P)$ – минимальный траекторный аттрактор пространства траекторий \mathcal{H}^+ .

Лемма 3. Пусть каждому λ из некоторого метрического пространства Λ поставлено в соответствие пространство траекторий

$$\mathcal{H}_\lambda^+ \subset C(\mathbb{R}_+; E_0) \cap L_\infty(\mathbb{R}_+; E),$$

пусть \mathcal{H}_λ^+ имеет минимальный траекторный аттрактор вида $\mathcal{U}_\lambda = \omega(P_\lambda)$, где P_λ – трансляционно-инвариантное множество и $P_\lambda \subset \mathcal{H}_\lambda^+ \cap P$, где P – некоторое множество, относительно компактное в $C(\mathbb{R}_+; E_0)$ и ограниченное в $L_\infty(\mathbb{R}_+; E)$; также пусть выполнено условие

$$\text{если } \lambda_m \rightarrow \lambda_0, \quad v_m \in P_{\lambda_m}, \quad v_m \rightarrow v_0 \text{ в } C(\mathbb{R}_+; E_0), \text{ то } v_0 \in [P_{\lambda_0}]. \tag{3.3}$$

Тогда имеют место следующие предельные соотношения:

$$h_{C(\mathbb{R}_+; E_0)}(\mathcal{U}_{\lambda_m}, \mathcal{U}_{\lambda_0}) = \sup_{u \in \mathcal{U}_{\lambda_m}} \inf_{v \in \mathcal{U}_{\lambda_0}} \|u - v\|_{C(\mathbb{R}_+; E_0)} \rightarrow 0 \quad (\lambda_m \rightarrow \lambda_0), \tag{3.4}$$

$$h_{E_0}(\mathcal{A}_{\lambda_m}, \mathcal{A}_{\lambda_0}) = \sup_{u \in \mathcal{A}_{\lambda_m}} \inf_{v \in \mathcal{A}_{\lambda_0}} \|u - v\|_{E_0} \rightarrow 0 \quad (\lambda_m \rightarrow \lambda_0). \tag{3.5}$$

Доказательство приведенных выше утверждений проводится в несколько этапов в работе [19]. Для ясности отметим, что существование глобального аттрактора при доказательстве последней леммы следует из теоремы 2.

4. СЛАБАЯ ПОСТАНОВКА ЗАДАЧИ И АППРОКСИМАЦИЯ

Обозначим через $C_0^\infty(\Omega)^n$ пространство функций на Ω со значениями в \mathbb{R}^n класса C^∞ с компактным носителем, содержащимся в Ω . Пусть $\mathcal{V} = \{v(x) = (v_1, \dots, v_n) \in C_0^\infty(\Omega)^n : \operatorname{div} v = 0\}$. Определим V^0 и V^1 как пополнение \mathcal{V} по нормам $L_2(\Omega)^n$ и $H^1(\Omega)^n$ соответственно и положим $V^2 = H^2(\Omega)^n \cap V^1$.

Пусть $\pi : L_2(\Omega)^n \rightarrow V^0$ – проектор Лере. Напомним, что в силу разложения Вейля $L_2(\Omega)^n = V^0 \oplus \nabla H^1(\Omega)$. Рассмотрим в пространстве \mathcal{V} оператор $A = -\Delta$, который продолжается в пространстве V^0 до замкнутого оператора и является самосопряженным положительным оператором с вполне непрерывным обратным (см., например, в [20], [21]). Область определения A совпадает с V^2 . В силу теоремы Гильберта о спектральном разложении вполне непрерывных операторов, собственные функции $\{e_j\}$ оператора A образуют ортонормированный базис в V^0 .

Пусть $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_k \leq \dots$ – собственные значения оператора A . Обозначим через E_∞ множество конечных линейных комбинаций, составленных из e_j , и определим пространство V^α , $\alpha \in \mathbb{R}$, как пополнение E_∞ по норме

$$\|v\|_{V^\alpha} = \left(\sum_{k=1}^\infty \lambda_k^\alpha |v_k|^2 \right)^{1/2}.$$

В [22], [23] показано, что указанные нормы в пространствах V^1, V^2, V^3 эквивалентны следующим нормам: $\|v\|_{V^1} = \|A^{1/2}v\|_{V^0}$, $\|v\|_{V^2} = \|Av\|_{V^0}$, $\|v\|_{V^3} = \|A^{3/2}v\|_{V^0}$.

Для определения слабого решения на отрезке введем следующие пространства:

$$W_1[0, T] = \{v : v \in L_\infty(0, T; V^2), v' \in L_\infty(0, T; V^1)\}, \quad v' = \frac{\partial v}{\partial t},$$

$$W_2[0, T] = \{v : v \in C([0, T], V^3), v' \in L_\infty(0, T; V^3)\}, \quad v' = \frac{\partial v}{\partial t},$$

с соответствующими нормами

$$\|v\|_{W_1[0, T]} = \|v\|_{L_\infty(0, T; V^2)} + \|v'\|_{L_\infty(0, T; V^1)}; \quad \|v\|_{W_2[0, T]} = \|v\|_{C([0, T], V^3)} + \|v'\|_{L_\infty(0, T; V^3)}.$$

Для определения слабого решения на полуоси \mathbb{R}_+ введем пространство $W_1^{\text{loc}}(\mathbb{R}_+)$, состоящее из функций v , определенных почти всюду на \mathbb{R}_+ и принимающих значения в V^2 таких, что ограничение v на любой отрезок $[0, T]$ принадлежит $W_1[0, T]$. Также введем пространство $W_2^{\text{loc}}(\mathbb{R}_+)$, состоящее из функций v класса $C(\mathbb{R}_+; V^3)$ таких, что ограничение v на любой отрезок $[0, T]$ принадлежит $W_2[0, T]$.

Будем предполагать, что $a \in V^2, f \in V^0$.

Определение 10. Слабым решением задачи (2.1)–(2.4) на отрезке $[0, T]$ будем называть функцию $v \in W_1[0, T]$ такую, что тождество

$$\int_{\Omega} \frac{\partial v}{\partial t} \varphi dx + v \int_{\Omega} \nabla v : \nabla \varphi dx - \sum_{i,j=1}^n \int_{\Omega} v_i v_j \frac{\partial \varphi_j}{\partial x_i} dx + \kappa \int_{\Omega} \nabla \left(\frac{\partial v}{\partial t} \right) : \nabla \varphi dx + \kappa \sum_{i,j=1}^n \int_{\Omega} v_i \Delta v_j \frac{\partial \varphi_j}{\partial x_i} dx = \int_{\Omega} f \varphi dx \tag{4.1}$$

выполнено почти всюду на $(0, T)$ для любой функции $\varphi \in V^1$ и функция v удовлетворяет начальному условию:

$$v(0) = a. \tag{4.2}$$

Здесь символ $:$ обозначает покомпонентное произведение матриц.

Определение 11. Слабым решением задачи (2.1)–(2.4) на полуоси \mathbb{R}_+ будем называть функцию $v \in W_1^{\text{loc}}(\mathbb{R}_+)$ такую, что при каждом $T > 0$ ограничение v на отрезок $[0, T]$ является слабым решением задачи (2.1)–(2.4) на отрезке $[0, T]$.

Введем следующие операторы:

$$A : V^1 \rightarrow V^{-1}, \quad \langle Au, \varphi \rangle = \int_{\Omega} \nabla u : \nabla \varphi dx \quad \forall u, \varphi \in V^1,$$

$$B_1 : L_4(\Omega)^n \rightarrow V^{-1}, \quad \langle B_1(u), \varphi \rangle = \sum_{i,j=1}^n \int_{\Omega} u_i u_j \frac{\partial \varphi_j}{\partial x_i} dx \quad \forall u \in L_4(\Omega)^n, \quad \varphi \in V^1,$$

$$B_2 : V^2 \rightarrow V^{-1}, \quad \langle B_2(u), \varphi \rangle = \sum_{i,j=1}^n \int_{\Omega} u_i \Delta u_j \frac{\partial \varphi_j}{\partial x_i} dx \quad \forall u \in V^2, \quad \varphi \in V^1,$$

$$J : V^1 \rightarrow V^{-1}, \quad \langle Ju, \varphi \rangle = \int_{\Omega} u \varphi dx \quad \forall u, \varphi \in V^1.$$

Тогда задача о поиске слабых решений начально-краевой задачи (2.1)–(2.4) эквивалентна задаче о поиске решения $v \in W_1[0, T]$ операторного уравнения

$$(J + \kappa A)v' + vAv - B_1(v) + \kappa B_2(v) = f, \tag{4.3}$$

удовлетворяющего начальному условию (4.2).

Рассмотрим следующее аппроксимационное операторное уравнение:

$$(J + \kappa A + \varepsilon e^{-\alpha t} A^2)v' + vAv - B_1(v) + \kappa B_2(v) = f, \tag{4.4}$$

где $\varepsilon \in (0, 1]$, а оператор A^2 определяется следующим образом:

$$A^2 : V^3 \rightarrow V^{-1}, \quad \langle A^2 u, \varphi \rangle = - \int_{\Omega} \nabla(\Delta u) : \nabla \varphi dx \quad \forall u \in V^3, \quad \varphi \in V^1.$$

Операторное уравнение (4.4) будем рассматривать с начальным условием

$$v(0) = b, \quad b \in V^3. \quad (4.5)$$

Определение 12. Решением уравнения (4.4) на отрезке $[0, T]$ будем называть функцию $v \in W_2[0, T]$ такую, что уравнение (4.4) выполнено в $L_{\infty}(0, T; V^{-1})$. Решением (4.4) на полуоси \mathbb{R}_+ будем называть функцию $v \in W_2^{\text{loc}}(\mathbb{R}_+)$ такую, что при каждом $T > 0$ ограничение v на отрезок $[0, T]$ является решением уравнения (4.4) на этом отрезке.

5. СХОДИМОСТЬ АТТРАКТОРОВ АППРОКСИМАЦИЙ

В настоящем разделе вводятся пространства траекторий для исходной и аппроксимационной задач и показывается, что их траекторные и глобальные аттракторы сходятся к соответствующим аттракторам исходной задачи. При определении пространств траекторий используются пространства $E = V^2$ и $E_0 = V^1$.

Введем постоянную

$$\alpha = \frac{\nu \kappa}{K_0^2 K_1^2 + 2\kappa K_1^2 + \kappa^2} = \frac{\nu \kappa}{K_2}, \quad K_2 = K_0^2 K_1^2 + 2\kappa K_1^2 + \kappa^2.$$

Здесь ν и κ — параметры задачи (2.1)–(2.4), а K_0 и K_1 — постоянные из соответствующих непрерывным вложениям $V^1 \subset V^0$ и $V^2 \subset V^1$ неравенств:

$$\|u\|_{V^0} \leq K_0 \|u\|_{V^1}, \quad u \in V^1; \quad \|u\|_{V^1} \leq K_1 \|u\|_{V^2}, \quad u \in V^2.$$

Ниже приведем теоремы, необходимые нам в дальнейшем.

Теорема 3. При любом $b \in V^3$ задача (4.4), (4.5) имеет единственное решение на полуоси \mathbb{R}_+ , причем при почти всех $t \in [0, T]$ справедливо неравенство

$$\begin{aligned} \kappa \|v(t)\|_{V^2} + e^{-\alpha t/2} \sqrt{\varepsilon} \kappa \|v(t)\|_{V^3} + \varepsilon e^{-\alpha t} \|v'(t)\|_{V^3} + \kappa \|v'(t)\|_{V^1} \leq \\ \leq C_1 \left(1 + e^{-\alpha t} \left(K_2 \|v(0)\|_{V^2}^2 + \varepsilon \|v(0)\|_{V^2}^2 + \varepsilon \kappa \|v(0)\|_{V^3}^2 \right) \right), \end{aligned} \quad (5.1)$$

где константа

$$C_1 = 2 \frac{K_1^2 K_2^2 + \kappa^2}{\alpha \nu \kappa} \|f\|_{V^0}^2$$

не зависит от ε , t и v .

Доказательство существования решения задачи (4.4), (4.5) и оценки (5.1) может быть найдено в [13]. Доказательство единственности решения проводится стандартным образом. Предполагается наличие двух решений $u, v, u \neq v$ задачи (4.4), (4.5). Затем из равенства (4.4) для u вычитается равенство (4.4) для v и полученное равенство применяется к функции $(J + \kappa A)w$, где $w = (u - v)$. Далее оцениваются сверху слагаемые с операторами B_1 и B_2 (здесь существенно используется принадлежность функций u и v пространству V^3), а остальные слагаемые преобразуются аналогично доказательству неравенства (5.1). К полученному в итоге неравенству применяется лемма Гронуолла–Беллмана, из которой следует, что $w \equiv 0$ и, следовательно, получаем $u = v$.

Доказательство следующих двух утверждений может быть найдено в [13].

Теорема 4. При любом $a \in V^2$ задача (4.3), (4.2) имеет решение на полуоси \mathbb{R}_+ , удовлетворяющее при почти всех $t > 0$ неравенству

$$\|v(t)\|_{V^2} + \|v'(t)\|_{V^1} \leq C_2 \left(1 + e^{-\alpha t} K_2 \|v(0)\|_{V^2}^2 \right).$$

Здесь $C_2 = \left(\frac{C_1}{\kappa} + \frac{C_1 C_4}{\kappa} + C_1 \right)$ — постоянная, не зависящая от v , t , ε .

Лемма 4. Пусть $\{v_m\}$ – ограниченная последовательность в пространстве $L_\infty(0, T; V^2)$, а последовательность производных $\{v'_m\}$ ограничена в $L_\infty(0, T; V^1)$. Тогда имеют место следующие утверждения.

1. Существует подпоследовательность $\{v_{m_k}\}$, сходящаяся к предельной функции v_* в пространстве $C([0, T], V^1)$, причем имеют место сходимости

$$\begin{aligned} Jv'_{m_k} &\rightarrow Jv'_* \quad \text{слабо в } L_2(0, T; V^{-1}), \\ Av'_{m_k} &\rightarrow Av'_* \quad \text{слабо в } L_2(0, T; V^{-1}), \\ \nu Av_{m_k} &\rightarrow \nu Av_* \quad \text{слабо в } L_2(0, T; V^{-1}), \\ B_1(v_{m_k}) &\rightarrow B_1(v_*) \quad \text{сильно в } L_\infty(0, T; V^{-1}), \\ B_2(v_{m_k}) &\rightarrow B_2(v_*) \quad \text{слабо в } L_2(0, T; V^{-1}). \end{aligned}$$

2. Пусть $\varepsilon_m \rightarrow 0$ – числовая последовательность, и последовательность $\{\varepsilon_m v'_m\}$ ограничена в норме пространства $L_\infty(0, T, V^3)$, то найдется подпоследовательность $\{\varepsilon_{m_k} v'_{m_k}\}$ такая, что $\varepsilon_{m_k} e^{-\alpha t} A^2 v'_{m_k} \rightarrow 0$ слабо в $L_2(0, T; V^{-1})$.

Определение 13. В качестве пространства траекторий \mathcal{H}_0^+ уравнения (4.3) будем рассматривать множество решений этого уравнения, определенных на \mathbb{R}_+ , существенно ограниченных со значениями в V^2 и удовлетворяющих оценке

$$\|v(t)\|_{V^2} + \|v'(t)\|_{V^1} \leq C_2 \left(1 + e^{-\alpha t} K_2 \|v\|_{L_\infty(\mathbb{R}_+, V^2)}^2\right) \tag{5.2}$$

при почти всех $t > 0$ или оценке

$$\|v\|_{L_\infty(\mathbb{R}_+, V^2)} + \|v'\|_{L_\infty(\mathbb{R}_+, V^1)} \leq 2C_2(1 + K_2) = 2C_3. \tag{5.3}$$

Рассмотрим множество

$$P = \left\{v \in C(\mathbb{R}_+, V^1) \cap L_\infty(\mathbb{R}_+, V^2) : v' \in L_\infty(\mathbb{R}_+, V^1), \|v\|_{L_\infty(\mathbb{R}_+, V^2)} + \|v'\|_{L_\infty(\mathbb{R}_+, V^1)} \leq 2C_2(1 + K_2) = 2C_3\right\}.$$

Исходя из определения множества P имеем, что оно трансляционно-инвариантно, ограничено в $L_\infty(\mathbb{R}_+, V^2)$ и относительно компактно в $C(\mathbb{R}_+, V^1)$ (относительная компактность получается, например, по теореме Обена–Дубинского–Симона [24]). Покажем, что P на самом деле компактно в $C(\mathbb{R}_+, V^1)$. Если последовательность $\{v_m\} \subset P$ сходится к v_0 в $C(\mathbb{R}_+, V^1)$, то в силу ограниченности этой последовательности в $L_\infty(\mathbb{R}_+, V^2)$ она сходится к своей предельной функции $*$ -слабо в $L_\infty(\mathbb{R}_+, V^2)$. Аналогичным образом, последовательность производных $\{v'_m\}$ сходится к v'_0 в смысле распределений, а будучи ограниченной в $L_\infty(\mathbb{R}_+, V^1)$, она сходится к v'_0 также $*$ -слабо в $L_\infty(\mathbb{R}_+, V^1)$, поэтому

$$\|v_0\|_{L_\infty(\mathbb{R}_+, V^2)} + \|v'_0\|_{L_\infty(\mathbb{R}_+, V^1)} \leq \liminf_{m \rightarrow \infty} \|v_m\|_{L_\infty(\mathbb{R}_+, V^2)} + \liminf_{m \rightarrow \infty} \|v'_m\|_{L_\infty(\mathbb{R}_+, V^1)} \leq \liminf_{m \rightarrow \infty} \left(\|v_m\|_{L_\infty(\mathbb{R}_+, V^2)} + \|v'_m\|_{L_\infty(\mathbb{R}_+, V^1)}\right) \leq 2C_3.$$

Таким образом, предельная функция v_0 принадлежит множеству P , следовательно, P компактно.

Пусть $P_0 = \mathcal{H}_0^+ \cap P$.

Лемма 5. Пространство траекторий \mathcal{H}_0^+ имеет минимальный траекторный аттрактор $\mathcal{U}_0 = \omega(P_0)$.

Доказательство. Покажем, что множество P_0 удовлетворяет условиям леммы 2. Поскольку множество P компактно в $C(\mathbb{R}_+, V^1)$ и ограничено в $L_\infty(\mathbb{R}_+, V^2)$, то получаем, что P_0 относительно компактно в $C(\mathbb{R}_+, V^1)$ и ограничено в $L_\infty(\mathbb{R}_+, V^2)$. Теперь покажем, что множество P_0 трансляционно-инвариантно. Пусть $v \in P_0$ и $h \geq 0$. Тогда функция v удовлетворяет уравнению (4.3). В силу

автономности уравнения (4.3) функция $T(h)v$ является его решением и при этом удовлетворяет неравенству:

$$\|T(h)v\|_{L_\infty(\mathbb{R}_+;V^2)} + \|T(h)v'\|_{L_\infty(\mathbb{R}_+;V^1)} \leq \|v\|_{L_\infty(\mathbb{R}_+;V^2)} + \|v'\|_{L_\infty(\mathbb{R}_+;V^1)} \leq 2C_3.$$

Откуда получаем, что $T(h)v \in P_0$.

Осталось показать, что P_0 – поглощающее множество. Пусть $B \subset \mathcal{H}_0^+$ – некоторое множество, ограниченное в $L_\infty(\mathbb{R}_+;V^2)$, и пусть $\|v\|_{L_\infty(\mathbb{R}_+;V^2)} \leq R$ для любого $v \in B$. Выберем такое t_0 , что $R^2 e^{-\alpha t_0} \leq 1$. Возможны два случая.

1. Если $v \in B$ и $v \notin P_0$, то в силу определения множества P_0 получаем, что $v \notin P$, и из определения пространства траекторий \mathcal{H}_0^+ следует, что v удовлетворяет оценке (5.2). Отметим, что оценка (5.3) не может быть выполнена, так как она влечет принадлежность функции v множеству P . При $t \geq t_0$ имеем

$$\|v(t)\|_{V^2} + \|v'(t)\|_{V^1} \leq C_2 \left(1 + e^{-\alpha t} K_2 \|v\|_{L_\infty(\mathbb{R}_+;V^2)}^2\right) \leq C_2(1 + K_2) = C_3. \quad (5.4)$$

Так как

$$\operatorname{vrai\,max}_{t \in \mathbb{R}_+} a(t) + \operatorname{vrai\,max}_{t \in \mathbb{R}_+} b(t) \leq 2 \operatorname{vrai\,max}_{t \in \mathbb{R}_+} (a(t) + b(t)),$$

для любых $a(t), b(t) \geq 0$, то, переходя к $\operatorname{vrai\,max}$ по $t \in \mathbb{R}_+$ в неравенстве (5.4), получаем, что

$$\begin{aligned} & \|T(t)v\|_{L_\infty(\mathbb{R}_+;V^2)} + \|T(t)v'\|_{L_\infty(\mathbb{R}_+;V^1)} \leq \|v\|_{L_\infty(\mathbb{R}_+;V^2)} + \|v'\|_{L_\infty(\mathbb{R}_+;V^1)} = \\ & = \operatorname{vrai\,max}_{t \in \mathbb{R}_+} \|v(t)\|_{V^2} + \operatorname{vrai\,max}_{t \in \mathbb{R}_+} \|v'(t)\|_{V^1} \leq 2 \operatorname{vrai\,max}_{t \in \mathbb{R}_+} (\|v(t)\|_{V^2} + \|v'(t)\|_{V^1}) \leq 2C_2(1 + K_2) = 2C_3. \end{aligned}$$

Это означает, что $T(t)v \in P$, а поскольку функция $T(t)v$ является решением уравнения (4.3), то $T(t)v \in \mathcal{H}_0^+$ и, таким образом, $T(t)v \in P_0$ при $t \geq 0$.

2. Если $v \in B$ и $v \in P_0$, то в силу трансляционной инвариантности множества P_0 имеем $T(t)v \in P_0$ при любом $t \geq 0$.

Таким образом, множество P_0 является поглощающим.

В итоге получаем, что множество P_0 удовлетворяет предположениям леммы 2. И, следовательно, \mathcal{H}_0^+ имеет минимальный траекторный аттрактор $\mathcal{O}_0 = \omega(P_0)$.

Перейдем к аттракторам аппроксимационного уравнения. Сначала введем пространство траекторий для аппроксимационного уравнения (4.4).

Определение 14. В качестве пространства траекторий $\mathcal{H}_\varepsilon^+$ уравнения (4.4) будем рассматривать множество, состоящее из решений этого уравнения, определенных на \mathbb{R}_+ , существенно ограниченных со значениями в V^2 и удовлетворяющих оценке

$$\varepsilon \|v(0)\|_{V^3}^2 \leq 1, \quad (5.5)$$

а также из функций вида $T(h)v$, где v – решение уравнения (4.4) на \mathbb{R}_+ , $h \geq 0$, и $T(h)v$ удовлетворяет оценке

$$\|T(h)v\|_{L_\infty(\mathbb{R}_+;V^2)} + \|T(h)v'\|_{L_\infty(\mathbb{R}_+;V^1)} \leq 2C_3.$$

Неравенство (5.5) оправдывается тем, что именно такие решения аппроксимационного уравнения были использованы при предельном переходе к решению исходного уравнения при доказательстве теоремы 4.

Отметим, что пространство траекторий определено корректно. В самом деле, поскольку решения аппроксимационного уравнения (4.4) принадлежат пространству $C(\mathbb{R}_+;V^3)$ и мы требуем принадлежность траекторий пространству $L_\infty(\mathbb{R}_+;V^2)$, то включение

$$\mathcal{H}_\varepsilon^+ \subset C(\mathbb{R}_+;V^3) \cap L_\infty(\mathbb{R}_+;V^2)$$

имеет место. Далее, пространство $\mathcal{H}_\varepsilon^+$ непусто, так как любая функция $b \in V^3$ такая, что $\varepsilon \|b\|_{V^3}^2 \leq 1$, служит началом новой траектории (в силу теоремы 3). Таким образом, пространство $\mathcal{H}_\varepsilon^+$ не только определенно корректно, но и достаточно “богато” на траектории (содержит большое количество траекторий).

Рассмотрим множество $P_\varepsilon = \mathcal{H}_\varepsilon^+ \cap P$.

Лемма 6. *Пространство траекторий $\mathcal{H}_\varepsilon^+$ ($\varepsilon > 0$) имеет минимальный траекторный аттрактор $\mathcal{U}_\varepsilon = \omega(P_\varepsilon)$.*

Доказательство. Покажем, что множество P_ε удовлетворяет условиям леммы 2. Так как P_ε содержится в P , то P_ε относительно компактно в $C(\mathbb{R}_+; V^1)$ и ограничено в $L_\infty(\mathbb{R}_+; V^2)$. Любая траектория, принадлежащая P_ε , имеет вид $T(h)v$, где v – решение уравнения (4.4). Функция $T(s)T(h)v$ принадлежит P в силу трансляционной инвариантности этого множества. Следовательно, эта функция принадлежит $P_\varepsilon = \mathcal{H}_\varepsilon^+ \cap P$. Таким образом, множество P_ε трансляционно-инвариантно.

Теперь докажем, что множество P_ε является поглощающим. Возьмем множество $B \subset \mathcal{H}_\varepsilon^+$, ограниченное в $L_\infty(\mathbb{R}_+; V^2)$. Для определенности предположим, что $\|v\|_{L_\infty(\mathbb{R}_+; V^2)} \leq R$ для каждого $v \in B$. Выберем t_0 такое, что $R^2 e^{-\alpha t_0} \leq 1$. Рассмотрим произвольную траекторию $v \in B$. Исходя из определения $\mathcal{H}_\varepsilon^+$, существует два возможных варианта.

1. Функция v является решением уравнения (4.4) и удовлетворяет условию (5.5). По теореме 3 функция v удовлетворяет при почти всех $t > 0$ неравенству

$$\kappa \|v(t)\|_{V^2} + \kappa \|v'(t)\|_{V^1} \leq C_1 \left(1 + e^{-\alpha t} \left(K_2 \|v(0)\|_{V^2}^2 + \varepsilon \|v(0)\|_{V^2}^2 + \varepsilon \kappa \|v(0)\|_{V^3}^2 \right) \right).$$

Используя непрерывность вложения $V^3 \subset V^2$, преобразуем это неравенство следующим образом:

$$\kappa \|v(t)\|_{V^2} + \kappa \|v'(t)\|_{V^1} \leq C_1 \left(1 + e^{-\alpha t} \left(K_2 \|v(0)\|_{V^2}^2 + C_4 \varepsilon \|v(0)\|_{V^3}^2 + \varepsilon \kappa \|v(0)\|_{V^3}^2 \right) \right).$$

Используя (5.5), получаем

$$\begin{aligned} \|v(t)\|_{V^2} + \|v'(t)\|_{V^1} &\leq \frac{C_1}{\kappa} \left(1 + e^{-\alpha t} \left(C_4 + \kappa + K_2 \|v(0)\|_{V^2}^2 \right) \right) \leq \\ &\leq \frac{C_1}{\kappa} \left(1 + C_4 + \kappa + e^{-\alpha t} K_2 \|v(0)\|_{V^2}^2 \right) \leq \left(\frac{C_1}{\kappa} + \frac{C_1 C_4}{\kappa} + C_1 \right) \left(1 + e^{-\alpha t} K_2 \|v(0)\|_{V^2}^2 \right) \leq C_2 \left(1 + e^{-\alpha t} K_2 \|v(0)\|_{V^2}^2 \right). \end{aligned}$$

Таким образом,

$$\|v(t)\|_{V^2} + \|v'(t)\|_{V^1} \leq C_2 \left(1 + e^{-\alpha t} K_2 \|v(0)\|_{V^2}^2 \right).$$

Поэтому, для любого $t \geq t_0$ имеем

$$\|v(t)\|_{V^2} + \|v'(t)\|_{V^1} \leq C_2(1 + K_2) = C_3.$$

Откуда следует, что для любого $s \geq t_0$ справедливо

$$\|T(s)v\|_{L_\infty(\mathbb{R}_+; V^2)} + \|T(s)v'\|_{L_\infty(\mathbb{R}_+; V^1)} \leq 2C_3.$$

Это означает, что $T(s)v \in P$. Отсюда следует, что $T(s)v \in P_\varepsilon$ при $s \geq t_0$.

2. Если v – функция вида $T(s)w$, удовлетворяющая неравенству

$$\|v\|_{L_\infty(\mathbb{R}_+; V^2)} + \|v'\|_{L_\infty(\mathbb{R}_+; V^1)} \leq 2C_3,$$

тогда $v \in P_\varepsilon$ и $T(s)v \in P_\varepsilon$ для любого $s \geq 0$, поскольку P_ε – трансляционно-инвариантное множество.

Из вышесказанного следует, что в любом из этих случаев $T(s)v \in P_\varepsilon$ при $s \geq t_0$, т.е. P_ε – поглощающее множество.

Таким образом получили, что все условия леммы 2 выполнены. Следовательно, траекторное пространство $\mathcal{H}_\varepsilon^+$ имеет минимальный траекторный аттрактор $\mathcal{U}_\varepsilon = \omega(P_\varepsilon)$.

6. ОСНОВНОЙ РЕЗУЛЬТАТ О СХОДИМОСТИ АТТРАКТОРОВ

Имеет место следующая теорема о сходимости аттракторов.

Теорема 5. *Минимальные траекторные аттракторы \mathcal{U}_ε аппроксимационного уравнения (4.4) сходятся к минимальному траекторному аттрактору \mathcal{U}_0 пространства траекторий \mathcal{H}_0^+ уравнения (4.3) в смысле полуотклонения в пространстве $C(\mathbb{R}_+; V^1)$, т.е.*

$$\sup_{u \in \mathcal{U}_\varepsilon} \inf_{v \in \mathcal{U}_0} \|u - v\|_{C(\mathbb{R}_+; V^1)} \rightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (6.1)$$

Глобальные аттракторы \mathcal{A}_ε аппроксимационного уравнения (4.4) сходятся к глобальному аттрактору \mathcal{A}_0 пространства траекторий \mathcal{H}_0^+ уравнения (4.3) в смысле полуотклонения в пространстве V^1 , т.е.

$$\sup_{y \in \mathcal{A}_\varepsilon} \inf_{z \in \mathcal{A}_0} \|y - z\|_{V^1} \rightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (6.2)$$

Доказательство. Проверим выполнение условий леммы 3. Необходимо доказать, что если $\varepsilon_m \rightarrow 0$, $v_m \in P_{\varepsilon_m}$, $v_m \rightarrow v_0$ в $C(\mathbb{R}_+; V^1)$, то v_0 принадлежит замыканию множества P_0 в топологии пространства $C(\mathbb{R}_+; V^1)$.

В силу включений $P_{\varepsilon_m} \subset P$ последовательность $\{v_m\}$ содержится в множестве P , компактном в $C(\mathbb{R}_+; V^1)$. Следовательно, $v_0 \in P$.

Покажем, что v_0 – решение уравнения (4.3) на \mathbb{R}_+ . В общем случае функции v_m имеют вид $T(h_m)w_m$, где w_m – решение уравнения (4.4) на полуоси \mathbb{R}_+ (т.е. при почти всех $t \geq 0$ функция w_m удовлетворяет уравнению (4.4)). Для любого $t \geq 0$ имеем $t + h_m \geq 0$, поэтому при почти всех $t \geq 0$ имеет место равенство

$$(J + \kappa A + \varepsilon_m e^{-\alpha(t+h_m)} A^2)w_m'(t+h_m) + \nu A w_m(t+h_m) - B_1(w_m)(t+h_m) + \kappa B_2(w_m)(t+h_m) = f.$$

По определению $w_m(t+h_m) = T(h_m)w_m(t) = v_m(t)$. Таким образом, получили, что

$$(J + \kappa A + \varepsilon_m e^{-\alpha h_m} e^{-\alpha t} A^2)v_m'(t) + \nu A v_m(t) - B_1(v_m)(t) + \kappa B_2(v_m)(t) = f, \quad (6.3)$$

при почти всех $t \in \mathbb{R}_+$.

Для того чтобы доказать, что v_0 – решение уравнения (4.3) на \mathbb{R}_+ , достаточно установить, что $\Pi_T v_0$ является решением уравнения (4.3) на отрезке $[0, T]$ при любом $T > 0$.

Из сходимости последовательности $\{v_m\}$ к v_0 в $C(\mathbb{R}_+; V^1)$ следует сходимость $\{\Pi_T v_m\}$ к $\Pi_T v_0$ в $C([0, T]; V^1)$. Так как v_m – решение уравнения (6.3) на \mathbb{R}_+ , то $\Pi_T v_m$ удовлетворяет (6.3) на отрезке $[0, T]$.

Так как функция w_m – решение уравнения (4.4) на полуоси, то оно удовлетворяет оценке (5.1). Следовательно, последовательность $\varepsilon_m v_m'(t) = \varepsilon_m w_m'(t+h_m)$ ограничена в $L_\infty(0, T; V^3)$. Откуда получаем, что функции $\varepsilon_m \Pi_T v_m'(t)$ ограничены в $L_\infty(0, T; V^3)$.

Далее, последовательность $\{\Pi_T v_m\}$ ограничена в $L_\infty(0, T; V^2)$, а $\{\Pi_T v_m'\}$ ограничена в $L_\infty(0, T; V^1)$. Поэтому по лемме 4, переходя к пределу в равенстве (6.3) при $\varepsilon \rightarrow 0$, получаем, что $\Pi_T v_0$ удовлетворяет при почти всех $t \in [0, T]$ следующему равенству:

$$(J + \kappa A)\Pi_T v_0' + \nu A \Pi_T v_0 - B_1(\Pi_T v_0) + \kappa B_2(\Pi_T v_0) = f.$$

В силу произвольности выбора отрезка $[0, T]$ функция v_0 является решением задачи (4.3), (4.2) на \mathbb{R}_+ .

Поскольку $v_0 \in P$, то v_0 удовлетворяет неравенству (5.3), и поэтому $v_0 \in \mathcal{H}_0^+$. Получаем, что $v_0 \in \mathcal{H}_0^+ \cap P = P_0$. Поэтому условия леммы 3 выполнены, и имеют место требуемые сходимости (6.1) и (6.2). Это завершает доказательство.

СПИСОК ЛИТЕРАТУРЫ

1. *Ладыженская О.А.* О динамической системе, порожаемой уравнениями Навье–Стокса // Зап. научн. сем. ЛОМИ. 1972. Т. 27. С. 91–115.
2. *Ладыженская О.А.* О нахождении минимальных глобальных аттракторов для уравнений Навье–Стокса и других уравнений с частными производными // Успехи матем. наук. 1987. Т. 42. № 6(258). С. 25–60.
3. *Серегин Г.А.* О динамической системе, порожденной двумерными уравнениями движения среды Бингама // Зап. научн. сем. ЛОМИ. 1991. Т. 188. С. 128–142.
4. *Chepyzhov V.V., Vishik M.I.* Trajectory attractors for evolution equations // C. R. Acad. Sci. Paris. Serie I. 1995. V. 321. P. 1309–1314.
5. *Sell G.* Global attractors for the three-dimensional Navier–Stokes equations // J. Dyn. Diff. Eq. 1996. V. 8. № 1. P. 1–33.
6. *Vorotnikov D.A., Zvyagin V.G.* Trajectory and global attractors of the boundary value problem for autonomous motion equations of viscoelastic medium // J. Math. Fluid Mech. 2008. V. 10. P. 19–44.
7. *Zvyagin V., Vorotnikov D.* Topological approximation methods for evolutionary problems of nonlinear hydrodynamics. Berlin: Walter de Gruyter, 2008. 248 p.
8. *Звягин В.Г., Кондратьев С.К.* Аттракторы уравнений неньютоновской гидродинамики // Успехи матем. наук. 2014. Т. 69. № 5(419). С. 81–156.
9. *Павловский В.А.* К вопросу о теоретическом описании слабых водных растворов полимеров // Докл. АН СССР. 1971. Т. 200. № 4. С. 809–812.
10. *Амфилохиев В.Б., Войткунский Я.И., Мазаева Н.П., Ходорковский Я.С.* Течения полимерных растворов при наличии конвективных ускорений // Тр. Ленинградского ордена Ленина кораблестроительного института. 1975. Т. 96. С. 3–9.
11. *Амфилохиев В.Б., Павловский В.А.* Экспериментальные данные о ламинарно-турбулентном переходе при течении полимерных растворов в трубах // Тр. Ленинградского ордена Ленина кораблестроительного института. 1976. Т. 104. С. 3–5.
12. *Турбин М.В., Устюжанинова А.С.* Теорема существования слабого решения начально-краевой задачи для системы уравнений, описывающей движение слабых водных растворов полимеров // Известия вузов. Математика. 2019. № 8. С. 62–78.
13. *Устюжанинова А.С., Турбин М.В.* Траекторные и глобальные аттракторы для модифицированной модели Кельвина–Фойгта // Сиб. ж. индустриальной матем. 2021. Т. 24. № 1. С. 126–138.
14. *Плотников П.И., Турбин М.В., Устюжанинова А.С.* Теорема существования слабого решения задачи оптимального управления с обратной связью для модифицированной модели Кельвина–Фойгта слабо концентрированных водных растворов полимеров // Докл. АН. 2019. Т. 488. № 2. С. 133–136.
15. *Ustuzhaninova A., Turbin M.* Feedback Control Problem for Modified Kelvin-Voigt Model // J. of Dynamical and Control Systems. 2021. <https://doi.org/10.1007/s10883-021-09539-0>
16. *Осколков А.П.* О единственности и разрешимости в целом краевых задач для уравнений движения водных растворов полимеров // Зап. науч. сем. ЛОМИ. 1973. Т. 38. С. 98–136.
17. *Гаевский Х., Грегер К., Захариас К.* Нелинейные операторные уравнения и операторные дифференциальные уравнения. М.: Мир, 1978. 336 с.
18. *Темам Р.* Уравнения Навье–Стокса. Теория и численный анализ. М.: Мир, 1981. 408 с.
19. *Zvyagin V.G., Kondratyev S.K.* Approximating topological approach to the existence of attractors in fluid mechanics // J. Fixed Point Theory Appl. 2013. V. 13. P. 359–395.
20. *Солонников В.А.* Оценки тензоров Грина для некоторых граничных задач // Докл. АН СССР. 1960. Т. 130. № 5. С. 988–991.
21. *Ворович И.И., Юдович В.И.* Стационарные течения вязкой несжимаемой жидкости // Матем. сборник. 1961. Т. 53. № 4. С. 393–428.
22. *Звягин В.Г., Турбин М.В.* Математические вопросы гидродинамики вязкоупругих сред. М.: КРАСАНД, 2012. 416 с.
23. *Фурсиков А.В.* Оптимальное управление распределенными системами. Теория и приложения. Новосибирск: Науч. книга, 1999. 352 с.
24. *Simon J.* Compact sets in the space $L^p(0, T; B)$ // Ann. Mat. Pura Appl. 1987. № 146. P. 65–96.

УДК 519.85

УСКОРЕННЫЕ ПРОКСИМАЛЬНЫЕ ОБОЛОЧКИ: ПРИМЕНЕНИЕ К ПОКОМПОНЕНТНОМУ МЕТОДУ¹⁾

© 2022 г. А. С. Аникин¹, В. В. Матюхин^{2,*}, Д. А. Пасечнюк²

¹ 664033 Иркутск, ул. Лермонтова, 134, а/я 292, Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения РАН, Россия

² 141701 Долгопрудный, М.о., Институтский пер., 9, Московский физико-технический институт (национальный исследовательский университет), Россия

*e-mail: vladmatyukh@gmail.com

Поступила в редакцию 16.02.2021 г.
Переработанный вариант 16.02.2021 г.
Принята к публикации 04.08.2021 г.

Статья посвящена одному частному случаю применения универсальных ускоренных проксимальных оболочек для получения вычислительно эффективных ускоренных вариантов методов, использующихся для решения различных частных постановок оптимизационных задач. В данной работе предлагается проксимально ускоренный покомпонентный градиентный метод с эффективной алгоритмической сложностью итерации, позволяющий существенно учитывать разреженность решаемой задачи, и рассматривается пример применения предлагаемого подхода для решения задачи оптимизации функции вида SoftMax, для которой описываемый метод позволяет ослабить зависимость вычислительной сложности решения от размерности n задачи в $\mathcal{O}(\sqrt{n})$ раз, и демонстрирует на практике более быструю по сравнению со стандартными методами сходимость. Библи. 22. Фиг. 2.

Ключевые слова: проксимальный ускоренный метод, каталист, ускоренный покомпонентный метод, SoftMax, LogSumExp.

DOI: 10.31857/S004446692202003X

1. ВВЕДЕНИЕ

Одним из важнейших теоретических результатов в выпуклой оптимизации является разработка ускоренных методов оптимизации [1]. На начальном этапе развития этой концепции было предложено множество ускоренных вариантов различных методов, применяющихся к решению многих задач выпуклой оптимизации, однако каждый такой случай требовал отдельного, частного рассмотрения возможности ускорения, ввиду чего предлагаемые конструкции были существенно различны и не позволяли предполагать способ их обобщения. Важным шагом к разработке универсальной схемы ускорения методов оптимизации стали работы, в которых предлагался и исследовался метод, названный каталист, основанный на идее ускоренного проксимального градиентного метода [2], [3] и позволяющий ускорять другие методы оптимизации, используя их для последовательного решения ряда регуляризованных по Моро-Иосиде вспомогательных задач [4]–[6]. В продолжение этих идей в дальнейшем было предложено множество вариантов применения данного метода и его модификаций [7]–[9]. Среди наиболее свежих, на момент написания данной статьи, результатов были также описаны обобщения обсуждаемого подхода на тензорные методы [10]–[13]. Соответствующее представление ускоренной проксимальной оболочки, если опираться на известные авторам сведения, является наиболее общим из описанных в литературе, и потому в данной работе внимание будет обращено прежде всего именно на методы, предложенные в последней из цитируемых выше работ.

¹⁾ Исследования Д.А. Пасечнюка были поддержаны стипендией А.М. Райгородского в области оптимизации и при финансовой поддержке РФФИ (код проекта 19-31-51001) Научное наставничество. Исследования А.С. Аникина выполнены при финансовой поддержке РФФИ (код проекта 18-29-03071 мк). Работа В.В. Матюхина выполнена при поддержке Минобрнауки РФ (госзадание) № 075-00337-20-03, номер проекта 0714-2020-0005.

Основной мотив данной работы состоит в том, чтобы описать возможности практического применения универсальных ускоренных проксимальных оболочек для конструирования вычислительно и оракульно-эффективных методов оптимизации. Рассмотрим классический покомпонентный метод [14], итерация которого для выпуклой функции $f : \mathbb{R}^n \rightarrow \mathbb{R}$ имеет вид:

$$x_{k+1}^i = x_k^i - \eta \nabla_i f(x_k), \quad i \sim \mathcal{U}\{1, \dots, n\}, \quad \eta > 0.$$

Одним из многих приложений данного метода является оптимизация функционалов, вычисление одной компоненты градиента которых значимо эффективнее, чем вычисление полного вектора градиента; в частности, многие задачи в случае разреженных постановок удовлетворяют данному условию. Однако оракульная сложность данного метода при условии остановки метода при достижении ε -малости невязки по значению функции составляет $\mathcal{O}\left(n \frac{\bar{L} R^2}{\varepsilon}\right)$, где $R^2 = \|x_0 - x_*\|_2^2$, $\bar{L} = \frac{1}{n} \sum_{i=1}^n L_i$ есть среднее констант Липшица компонент градиента, притом эта оценка не является оптимальной для класса выпуклых задач. Рассмотрим теперь ускоренный покомпонентный метод, предложенный Ю.Е. Нестеровым [15]. То есть оракульная сложность данного метода соответствует оптимальной оценке: $\mathcal{O}\left(n \sqrt{\frac{\bar{L} R^2}{\varepsilon}}\right)$, где $\sqrt{\bar{L}} = \frac{1}{n} \sum_{i=1}^n \sqrt{L_i}$ – среднее квадратных корней из констант Липшица компонент градиента. Вместе с тем ситуация кардинально меняется в случае рассмотрения алгоритмической сложности метода: пусть даже вычисление одной компоненты градиента имеет сложность $\mathcal{O}(s)$, $s \ll n$, сложность итерации ускоренного покомпонентного метода будет составлять $\mathcal{O}(n)$, в отличие от стандартного метода, сложность итерации которого есть $\mathcal{O}(s)$, – содержательно это означает, что степень разреженности задачи при применении ускоренного покомпонентного метода не влияет существенно на сложность алгоритма, и, кроме того, сложность в таком случае квадратично зависит от размерности задачи: вместе это в некоторой степени обесценивает применение покомпонентного метода в данном случае. Таким образом, интересной задачей является построение ускоренного покомпонентного метода, сложность итерации которого, как и в стандартном варианте метода, составляет $\mathcal{O}(s)$, при сохранении оптимальной оракульной сложности – в данной работе это удается осуществить благодаря применению универсальной ускоренной проксимальной оболочки “ускоренный метаалгоритм” [13].

Данная статья состоит из двух основных разделов. В разд. 2 описываются теоретические результаты о сходимости и алгоритмической сложности покомпонентного метода, ускоренного путем применения оболочки “ускоренный метаалгоритм”. В разд. 3 на примере задачи оптимизации функционала вида SoftMax экспериментально проверяется эффективность метода в отношении времени его работы, описываются возможности его вычислительно-эффективной имплементации и осуществляется сравнение со стандартными методами.

2. ТЕОРЕТИЧЕСКИЕ ГАРАНТИИ

Рассмотрим задачу оптимизации вида $\min_{x \in \mathbb{R}^n} f(x)$ функции $f : \mathbb{R}^n \rightarrow \mathbb{R}$, удовлетворяющей следующим условиям:

- 1) f дифференцируема на \mathbb{R}^n ;
- 2) f выпукла на \mathbb{R}^n ;
- 3) $\nabla_i f$ удовлетворяет условию Липшица:

$$\exists L_i \in \mathbb{R} : |\nabla_i f(x + ue_i) - \nabla_i f(x)| \leq L_i |u|, \quad \forall x \in \mathbb{R}^n, \quad u \in \mathbb{R},$$

где e_i есть i -й орт базиса, $i \in \{1, \dots, n\}$;

4) ∇f удовлетворяет условию Липшица с константой L .

Листинг 1. Ускоренный метаалгоритм для метода \mathcal{M}

Вход: $H > 0, x_0 \in \mathbb{R}^n$;
 $\lambda \leftarrow 1/2H$;
 $A_0 \leftarrow 0; v_0 \leftarrow x_0$;
for $k = 0, \dots, \tilde{N} - 1$ **do**
 $a_{k+1} \leftarrow \frac{\lambda + \sqrt{\lambda^2 + 4\lambda A_k}}{2}$;
 $A_{k+1} \leftarrow A_k + a_{k+1}$;
 $\tilde{x}_k \leftarrow \frac{A_k v_k + a_{k+1} x_k}{A_{k+1}}$;
 Посредством запуска метода \mathcal{M}
 найти с точностью ε по аргументу
 решение вспомогательной задачи:
 $v_{k+1} \in \text{Arg} \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{H}{2} \|y - \tilde{x}_k\|_2^2 \right\}$;
 $x_{k+1} \leftarrow x_k - a_{k+1} \nabla f(v_{k+1})$;
end
return $v_{\tilde{N}}$.

Обратимся к содержанию работы [13], где предложен общий вариант “ускоренного метаалгоритма” решения задач выпуклой оптимизации для композитных функционалов вида $F(x) = f(x) + g(x)$. Для рассматриваемой постановки задачи такая общность не требуется, достаточно применить частный случай описанной схемы при $p = 1, f \equiv 0$ (используются обозначения соответствующей работы), в котором описанная оболочка принимает вид ускоренного проксимального метода. Псевдокод используемого метода представлен в листинге 1.

Листинг 2. Покомпонентный метод

Вход: $y_0 \in \mathbb{R}^n$;
 $Z \leftarrow \sum_{i=1}^n (H + L_i)$;
 $p_i \leftarrow (H + L_i)/Z, i \in \{1, \dots, n\}$;
 Дискретное вероятностное
 распределение π с вероятностями p_i ;
for $k = 0, \dots, N - 1$ **do**
 $i \sim \pi\{1, \dots, n\}$;
 $y_{k+1} \leftarrow y_k$;
 $y_{k+1}^i = y_k^i - \frac{1}{H + L_i} \nabla_i F(y_k)$;
end
return y_N .

Прежде чем сформулировать какие-либо результаты о сходимости, следует подробнее рассмотреть вопрос о решении вспомогательной задачи — ее аналитическое решение доступно лишь в редких случаях, и потому необходимо решать ее численными методами, а значит, неточно.

Вспомогательную задачу допустимо решать до выполнения следующего условия останова ([16], Appendix B):

$$\left\| \nabla \left\{ F(y_*) := f(y_*) + \frac{H}{2} \|y_* - \tilde{x}_k\|_2^2 \right\} \right\|_2 \leq \frac{H}{2} \|y_* - \tilde{x}_k\|_2. \quad (1)$$

Ввиду того, что $\|\nabla F(y_*)\|_2 = 0$, а также ввиду $(L + H)$ -липшицевости ∇F , имеем

$$\|\nabla F(y_*)\|_2 \leq (L + H) \|y_* - y_k\|_2. \quad (2)$$

Выписав неравенство треугольника: $\|\tilde{x}_k - y_*\|_2 - \|y_* - y_k\|_2 \leq \|y_* - \tilde{x}_k\|_2$, и воспользовавшись неравенствами (1), (2), получаем окончательное условие останова:

$$\|y_* - y_k\|_2 \leq \frac{H}{3H + 2L} \|\tilde{x}_k - y_*\|_2. \quad (3)$$

Содержательно отсюда следует, что необходимая точность решения вспомогательной задачи по аргументу не зависит от требуемой точности решения общей задачи, что позволяет значимо упростить получение дальнейших результатов.

Рассмотрим теперь основной применяемый для решения вспомогательных задач метод: содержание покомпонентного метода [17] (в частном случае $\gamma = 1$) представлено в листинге 2. Для данного метода в случае рассматриваемых вспомогательных задач справедлива

Теорема 2.1 (см. [14, теорема 6.8]). Пусть F является H -сильно выпуклой относительно $\|\cdot\|_2$. Тогда для последовательности $\{y_k\}_{k=1}^N$, генерируемой покомпонентным методом, выполняется

$$\begin{aligned} \mathbb{E}[F(y_N)] - F(y_*) &\leq \left(1 - \frac{1}{\kappa}\right)^N (F(y_0) - F(y_*)), \\ \kappa &= \frac{H}{Z}, \quad Z = \sum_{i=1}^n (H + L_i). \end{aligned} \quad (4)$$

Используя данный результат, сформулируем утверждение о числе итераций покомпонентного метода, достаточном для выполнения полученного выше условия останова.

Следствие 2.1. Математическое ожидание $\mathbb{E}[y_N]$ точки, являющейся результатом работы покомпонентного метода, удовлетворяет условию (3) достижения достаточной точности решения вспомогательной задачи ускоренного метаалгоритма в том случае, если для числа итераций метода выполнено

$$\begin{aligned} N \geq N(\tilde{\epsilon}) &= \left\lceil \frac{Z}{H} \ln \left\{ \left(1 + \frac{L}{H}\right) \left(3 + \frac{2L}{H}\right)^2 \right\} \right\rceil, \\ \tilde{\epsilon} &= \frac{H}{2} \left(\frac{H}{3H + 2L} \right)^2 \|y_0 - y_*\|_2^2. \end{aligned}$$

Доказательство. Ввиду $(H + L)$ -липшицевой гладкости функции F верно:

$$F(y_0) - F(y_*) \leq \frac{H + L}{2} \|y_0 - y_*\|_2^2.$$

Используя это неравенство вместе с оценкой (4), можем выписать условие достижения заданной точности $\tilde{\epsilon}$ по функции:

$$\frac{H + L}{2} \left(1 - \frac{1}{\kappa}\right)^N \|y_0 - y_*\|_2^2 \leq \tilde{\epsilon}.$$

Также верно $1 - 1/\kappa \leq \exp\{-1/\kappa\}$, и, значит:

$$\frac{H + L}{2} \exp\{\kappa/N\} \|y_0 - y_*\|_2^2 \leq \tilde{\epsilon}.$$

Логарифмируя и подставляя выражение для κ , получаем выражение для числа итераций от $\tilde{\varepsilon}$:

$$N(\tilde{\varepsilon}) = \left\lceil \frac{Z}{H} \ln \left\{ \frac{(H+L)\|y_0 - y_*\|_2^2}{2\tilde{\varepsilon}} \right\} \right\rceil. \quad (5)$$

Ввиду H -сильной выпуклости функции F верно $\|\bar{y}_N - y_*\|_2^2 \leq \frac{2}{H}(F(\bar{y}_N) - F(y_*))$, где $\bar{y}_N = \mathbb{E}[y_N]$. Функция F выпукла, следовательно, по неравенству Йенсена, $F(\bar{y}_N) \leq \mathbb{E}[F(y_N)]$, и отсюда, вместе с (3) получаем достаточное условие достижения решения вспомогательной задачи:

$$\mathbb{E}[F(y_N)] - F(y_*) \leq \frac{H}{2} \left(\frac{H}{3H+2L} \right)^2 \|y_0 - y_*\|_2^2.$$

Подставляя в формулу (5) вместо $\tilde{\varepsilon}$ выражение из правой части данного неравенства, непосредственно приходим к выражению из утверждения.

Теперь, когда полностью прояснен вопрос о требуемой точности и оракульной сложности решения вспомогательной задачи с помощью покомпонентного метода, можно перейти к результатам о сходимости ускоренного метаалгоритма. Для используемого условия останова (3) метода, решающего вспомогательную задачу, справедлив следующий результат о сходимости ускоренного метаалгоритма:

Теорема 2.2 (см. [13, теорема 1]). *При $H > 0$ для последовательности $\{v_k\}_{k=1}^{\tilde{N}}$, генерируемой ускоренным метаалгоритмом, использующим для решения вспомогательной задачи некоторый не стохастический метод, выполняется*

$$f(v_{\tilde{N}}) - f(x_*) \leq \frac{48}{5} \frac{H \|x_0 - x_*\|_2^2}{\tilde{N}^2}. \quad (6)$$

На основании последнего утверждения можно сформулировать теорему о сходимости ускоренного метаалгоритма в случае применения стохастического метода и, в частности, покомпонентного градиентного спуска.

Теорема 2.3. *При $H > 0$ для некоторого $0 < \delta < 1$ точка $v_{\tilde{N}}$, являющаяся результатом работы ускоренного метаалгоритма, использующего для решения вспомогательной задачи покомпонентный метод, решающий вспомогательную задачу N_δ итераций, удовлетворяет условию*

$$\Pr(f(v_{\tilde{N}}) - f(x_*) < \varepsilon) \geq 1 - \delta$$

в случае, если

$$N_\delta \geq N \left(\frac{\tilde{\varepsilon}\delta}{\tilde{N}} \right) = \left\lceil \frac{Z}{H} \ln \left\{ \frac{\tilde{N}}{\delta} \left(1 + \frac{L}{H} \right) \left(3 + \frac{2L}{H} \right)^2 \right\} \right\rceil, \quad (7)$$

$$\tilde{N} \geq \left\lceil \frac{4\sqrt{15}}{5} \sqrt{\frac{H \|x_0 - x_*\|_2^2}{\varepsilon}} \right\rceil.$$

Доказательство. В следствии 2.1 представлена оценка числа итераций, достаточного для выполнения следующего условия на математическое ожидание значения функционала в результирующей точке метода:

$$\mathbb{E}[F(y_{N(\tilde{\varepsilon})})] - F(y_*) \leq \tilde{\varepsilon}.$$

Воспользуемся неравенством Маркова и получим формулировку данного условия в терминах оценки вероятности больших отклонений [18]: заведомо выберем допустимое значение вероятности невыполнения поставленного условия, так чтобы $0 < \delta/\tilde{N} < 1$, где \tilde{N} выражается из (6); тогда

$$\Pr(F(y_{N(\tilde{\varepsilon}/\tilde{N})}) - F(y_*) \geq \tilde{\varepsilon}) \leq \frac{\delta}{\tilde{N}} \frac{\mathbb{E}[F(y_{N(\tilde{\varepsilon}/\tilde{N})})] - F(y_*)}{\tilde{\varepsilon} \cdot \delta/\tilde{N}} = \frac{\delta}{\tilde{N}}.$$

Поскольку вероятность того, что полученное решение некоторой отдельно взятой вспомогательной задачи не будет удовлетворять поставленному условию, равна δ/\tilde{N} , значит, вероятность то-

го, что за \tilde{N} итераций ускоренного метаалгоритма условие будет не выполнено хотя бы для одной из задач, есть $\tilde{N} \cdot \delta / \tilde{N} = \delta$, откуда и следует доказываемое утверждение.

Далее, объединяя оценки, приводимые в теореме 9, можем получить асимптотическую оценку на общее число итераций покомпонентного метода, достаточное для решения рассматриваемой оптимизационной задачи с некоторой заданной точностью, а также оценку оптимального значения параметра H :

Следствие 2.2. Для того чтобы точка $v_{\tilde{N}}$, являющаяся результатом работы ускоренного метаалгоритма, удовлетворяла условию

$$Pr(f(v_{\tilde{N}}) - f(x_*) < \epsilon) \geq 1 - \delta,$$

достаточно выполнить в сумме

$$\hat{N} \geq \tilde{N} \cdot N_\delta = \mathcal{O}\left(\frac{Z \|x_0 - x_*\|_2}{\sqrt{H}} \frac{1}{\epsilon^{1/2}} \log\left\{\frac{1}{\epsilon^{1/2}\delta}\right\}\right) \tag{8}$$

итераций покомпонентного метода для решения вспомогательной задачи. При этом оптимально значение параметра H регуляризации вспомогательной задачи следует выбирать как $H \simeq \frac{1}{n} \sum_{i=1}^n L_i$ (\simeq обозначает равенство с точностью до малого множителя порядка \log).

Доказательство. Выражение для \hat{N} можно получить путем прямой подстановки одной из оценок, приводимых в (7), в другую, и их последующего умножения. Если исключить из рассмотрения малый множитель порядка $\log(L/H)$, константа в оценке будет зависеть от H как:

$$\sqrt{H} \cdot \frac{Z/n}{H} = \sqrt{H} \left(1 + \frac{\frac{1}{n} \sum_{i=1}^n L_i}{H}\right).$$

Минимизируя представленное выражение по H , получаем указанный результат.

Рассмотрим теперь подробнее вопрос об алгоритмической сложности предложенного ускоренного метода покомпонентного градиентного спуска. Очевидным является

Утверждение 2.1. Алгоритмическая сложность рассматриваемого метода составляет

$$T = \mathcal{O}(\tilde{N}(T_{\text{out}} + N_\delta T_{\text{inn}})),$$

где T_{out} – амортизированная оценка сложности вычислений, производимых на итерации ускоренного метаалгоритма, T_{inn} – амортизированная оценка сложности итерации покомпонентного метода.

Оттолкнувшись от него, сформулируем результат о вычислительной сложности метода.

Теорема 2.4. Пусть сложность вычисления одной компоненты градиента f составляет $\mathcal{O}(s)$. Тогда алгоритмическая сложность рассматриваемого метода есть

$$T = \mathcal{O}\left(sn \cdot \sqrt{\frac{\bar{L} \|x_0 - x_*\|_2^2}{\epsilon}} \log\left\{\frac{1}{\epsilon^{1/2}\delta}\right\}\right), \quad \bar{L} = Z/n = \frac{1}{n} \sum_{i=1}^n L_i.$$

Доказательство. Переформулируем оценку из утверждения 2.1 следующим образом:

$$T = \mathcal{O}(\hat{N} \cdot T_{\text{iter}}),$$

где T_{iter} – амортизированная оценка сложности элементарной итерации ускоренного метаалгоритма, т.е. итерации, которая может быть и внутренней итерацией покомпонентного метода, и основной итерацией метаалгоритма. Сложность основной итерации метаалгоритма определяется в первую очередь вычислением полного градиента f , а сложность этой процедуры (из условия теоремы) есть $\mathcal{O}(sn)$. В то же время, ввиду $Z = n\bar{L}$, верно также $N_\delta = \tilde{\mathcal{O}}(n)$, где символ $\tilde{\mathcal{O}}(\cdot)$ означает то же, что и $\mathcal{O}(\cdot)$, но с возможным присутствием множителей порядка $\log(\cdot)$. Поскольку основная итерация метаалгоритма выполняется через каждые N_δ элементарных итераций, где N_δ постоянно, из этого с помощью любого из методов амортизационного анализа тривиально получается амортизированная оценка сложности основной итерации метаалгоритма, составляющая $\mathcal{O}(s)$. Сложность итерации покомпонентного метода (если вместо копирования значений точки вы-

Таблица 1. Сравнение эффективности методов

Метод	Итеративная сложность	Вычислительная сложность	Источник
FGM	$\mathcal{O}(sn)$	$\mathcal{O}\left(sn \cdot \frac{1}{\varepsilon^{1/2}} \cdot \sqrt{L}\right)$	[1]
CDM	$\mathcal{O}(s)$	$\mathcal{O}\left(sn \cdot \frac{1}{\varepsilon} \cdot \bar{L}\right)$	[14]
ACDM	$\mathcal{O}(n)$	$\mathcal{O}\left(n^2 \cdot \frac{1}{\varepsilon^{1/2}} \cdot \sqrt{\bar{L}}\right)$	[15]
Catalyst CDM	$\mathcal{O}(s)$	$\tilde{\mathcal{O}}\left(sn \cdot \frac{1}{\varepsilon^{1/2}} \cdot \sqrt{\bar{L}}\right)$	данная работа

полнять операции “на месте”, что для данной конструкции вполне допустимо) определяется вычислением одной компоненты градиента, и составляет также $\mathcal{O}(s)$. Отсюда получаем $T_{\text{iter}} = \mathcal{O}(s)$. Используя оценку (8) и подставляя оптимальное значение H , получаем приведенную сложность метода.

Заметим также, что сложность метода по памяти при этом составляет $\mathcal{O}(n)$, также как и сложность предварительных вычислений (для покомпонентного метода нет необходимости выполнять их каждый раз заново).

Сравним полученные для предложенного метода оценки с оценками других методов, которые могут быть использованы для решения задач в описываемой постановке: быстрого градиентного метода (FGM), классического покомпонентного спуска (CDM), ускоренного покомпонентного спуска в варианте Ю.Е. Нестерова (ACDM) и предложенного в данной работе подхода (Catalyst CDM). Оценки приведены в табл. 1. Как можно видеть из приведенных асимптотических оценок вычислительной сложности, предложенный метод позволяет достигать эффективной в отношении характера зависимости от размерности задачи n и требуемой точности ε скорости сходимости, не уступающей другим методам, при некоторой плате за это в виде логарифмического множителя в оценке.

Заметим, кроме того, что несмотря на существенное сходство оценок, между двумя наиболее эффективными методами в табл. 1 (FGM и Catalyst CDM) существует также различие в константах, характеризующих гладкость функции: L – в FGM и \bar{L} – в Catalyst CDM, тем самым поведение рассматриваемого метода для различных задач напрямую зависит от характера их покомпонентной гладкости. В общем случае нельзя утверждать, что одна из констант асимптотически существенно выгоднее другой, однако в ряде частных случаев возможно выписать оценки значений констант явно, и часто оказывается, что \bar{L} “меньше” L . Наиболее существенен выигрыш в том случае, если справедливы соотношения $L = \mathcal{O}(n)$, $\bar{L} = \mathcal{O}(1)$ – тогда в оценке вычислительной сложности предлагаемого метода удастся редуцировать фактор порядка $\mathcal{O}(\sqrt{n})$ по сравнению с быстрым градиентным методом. В следующем разделе будет рассмотрен пример постановки задачи, в которой данный случай имеет место.

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

В данном разделе описывается характер практического поведения метода, на примере следующей оптимизационной задачи для функционала, имеющего вид SoftMax-функции:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \gamma \ln \left(\sum_{j=1}^m \exp \left(\frac{[Ax]_j}{\gamma} \right) \right) - \langle b, x \rangle \right\}, \quad (9)$$

где $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$. Задачи такого вида существенно важны для многих приложений, в частности, они возникают в задачах энтропийно-линейного программирования в качестве двойственной задачи [4], [22], в том числе в задаче оптимального транспорта, а также исполняют роль сглаженной аппроксимации функции \max (что и дало функционалу название SoftMax) и, соответ-

ственно, нормы $\|\cdot\|_\infty$, что может быть востребовано в некоторых постановках задачи PageRank или при решении системы линейных уравнений. Притом во всех описанных задачах важным частным случаем являются разреженные постановки, в случае которых матрица A разрежена, т.е. такова, что среднее число ненулевых элементов в строке A_j не превосходит некоторого $s \ll n$ (будет удобно также предполагать возможность для одной из строк A_j являться полностью неразрезанной).

Сформулируем свойства, которыми обладает функция f [19]:

- 1) f дифференцируема;
- 2) ∇f удовлетворяет условию Липшица с константой $L = \max_{j=1,\dots,m} \|A_j\|_2^2$;
- 3) $\nabla_i f$ удовлетворяют покомпонентному условию Липшица с константами $L_i = \max_{j=1,\dots,m} |A_{ji}|$.

Начнем уточнение свойств с первого пункта. Выпишем выражение для i -й компоненты градиента данного функционала:

$$\nabla_i f(x) = \frac{\sum_{j=1}^m A_{ji} \exp([Ax]_j)}{\sum_{j=1}^m \exp([Ax]_j)}.$$

Как можно видеть, наивное вычисление этого выражения может занимать время, сравнимое с вычислением градиента в целом, что будет значительно влиять на вычислительную сложность, а значит, и на время работы метода. Однако в то же время многие члены в этом выражении могут перевычисляться либо редко, либо покомпонентно, и использоваться при совершении шага метода как члены его дополнительной последовательности, так что сложность итерации будет оставаться эффективной, и применение покомпонентного метода будет оправдано. Для удобства описания используемых вычислительных приемов запишем шаг покомпонентного метода в виде

$$y_{k+1} = y_k + \delta e_i,$$

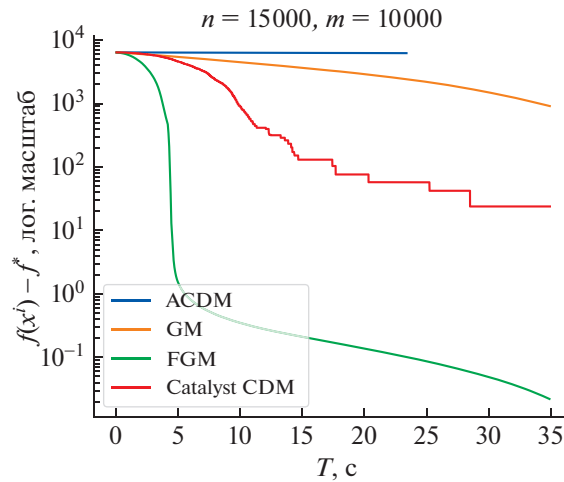
где δ – размер шага, умноженный на соответствующую компоненту градиента, e_i есть i -й орт базиса.

1. Будем хранить набор значений $\{\exp([Ay_k]_j)\}_{j=1}^m$, используемых для вычисления суммы в числителе. Обновление этих значений после осуществления шага метода имеет сложность $\mathcal{O}(s)$, ввиду того что $Ay_{k+1} = Ay_k + \delta A_i$, и вместе с тем A_i имеет не более s ненулевых компонент, а значит, потребуется вычисление s корректирующих множителей и умножение на них соответствующих значений из набора.

2. Как можно видеть из п. 1, производить умножение разреженных векторов следует за $\mathcal{O}(s)$, учитывая лишь ненулевые компоненты. В смысле программной реализации это означает необходимость использования для кэшируемых значений и для строк матрицы A разреженного представления, т.е. хранение лишь пар индекс-значение ненулевых элементов. Тогда, очевидно, сложность арифметических операций для таких векторов будет пропорциональна сложности цикла с элементарными арифметическими операциями, число итераций которого равно числу ненулевых элементов (в языке программирования python, например, такой формат хранения реализуется в методе `scipy.sparse.csr_matrix` [20]).

3. Аналогично, будем хранить значение $\sum_{j=1}^m \exp([Ay_k]_j)$, являющееся знаменателем. Его обновление осуществляется с той же сложностью, что и обновление последовательности из п. 1 (путем вычисления суммы ненулевых слагаемых, прибавляемых к каждому значению из набора).

4. Поскольку вычисление указанного выражения требует вычисления значений экспонент, может происходить переполнение типов. Для решения этой проблемы стандартно применяется `exp-normalize trick` [21]. Однако для его применения следует также хранить значение $\max_{j=1,\dots,m} [Ay_k]_j$. Вместе с тем нет необходимости знать именно это значение, или, иначе, знать его точно – достаточно лишь его приближения, чтобы значения в показателях экспонент были малы, так что перевычислять эту величину можно гораздо реже: например, раз в m итераций метода, в результате чего амортизированная сложность будет равна $\mathcal{O}(s)$.



Фиг. 1. Сходимость методов для задачи SoftMax (9) с равномерно разреженной случайной матрицей.

Итак, в дальнейших рассуждениях можно полагать, что итерация покомпонентного метода при решении соответствующей вспомогательной задачи имеет амортизированную сложность $\mathcal{O}(s)$.

Далее, рассмотрим подробнее вопрос о величине констант гладкости данного функционала.

Можно выписать асимптотические формулы для L и $\bar{L} = \frac{1}{n} \sum_{i=1}^n L_i$:

$$L = \max_{j=1, \dots, m} \|A_j\|_2^2 = \mathcal{O}(n), \quad \bar{L} = \frac{1}{n} \sum_{i=1}^n \max_{j=1, \dots, m} |A_{ji}| = \mathcal{O}(1).$$

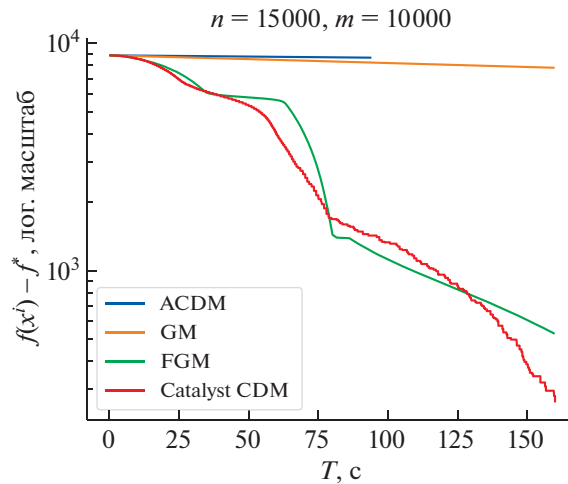
Используя эти оценки, уточним вычислительную сложность методов FGM и Catalyst CDM в применении к данной задаче:

$$T_{\text{FGM}} = \mathcal{O}\left(sn^{3/2} \cdot \frac{1}{\epsilon^{1/2}}\right), \quad T_{\text{CCDM}} = \tilde{\mathcal{O}}\left(sn \cdot \frac{1}{\epsilon^{1/2}}\right).$$

Таким образом, в теории, применение метода Catalyst CDM для решения данной задачи позволяет, по сравнению с FGM, редуцировать в асимптотической оценке вычислительной сложности множитель порядка $\mathcal{O}(\sqrt{n})$. Практически, это означает, что предложенный метод разумно применять для задач большой размерности.

Сравним теперь работу предложенного в статье метода (Catalyst CDM) с рядом альтернативных подходов: градиентным спуском (GM), быстрым градиентным методом (FGM), покомпонентным спуском (CDM) и ускоренным покомпонентным спуском (ACDM), на примере задачи (9) с искусственно сгенерированной двумя различными способами матрицей A . На фиг. 1 и 2 представлены графики сходимости рассматриваемых методов: по оси абсцисс отложено время работы методов в секундах, по оси ординат — невязка по функции в логарифмическом масштабе (f_* находится путем поиска соответствующей точки x_* с помощью метода FGM, настроенного на точность, заведомо значительно превосходящую возможную для достижения на выбранном временном промежутке).

На фиг. 1 представлен случай, для которого все элементы матрицы A являются нормально распределенными случайными величинами из дискретного равномерного распределения $A_{ji} \in \mathcal{U}\{0, 1\}$, при этом число ненулевых элементов составляет $s \approx 0.2m$, и параметр $\gamma = 0.6$ (так же как и во втором случае). В такой постановке предложенный метод демонстрирует более быструю сходимость по сравнению со всеми сравниваемыми методами, за исключением FGM. В то же время, в постановке, отраженной на фиг. 2, при которой число ненулевых элементов, по сравнению с первым случаем, увеличено до $s \approx 0.75m$, а матрица генерируется неравномерно в соответствии с правилом: $0.9m$ строк с $0.1n$ ненулевых элементов и $0.1m$ строк с $0.9n$ ненулевых элементов, а также одна и строк матрицы является полностью неразрезанной, предложенный метод



Фиг. 2. Сходимость методов для задачи SoftMax (9) с неоднородно разреженной случайной матрицей.

сходится быстрее FGM. Это объясняется тем, что в этом случае $L = n$, тогда как \bar{L} по-прежнему остается достаточно мало, в результате чего константа в предложенном методе оказывает заметно меньшее влияние на вычислительную сложность, чем в случае FGM. Из результатов эксперимента также можно отметить, что гораздо существеннее степени разреженности задачи на эффективность предложенного метода влияет характер ее покомпонентной гладкости.

4. ЗАКЛЮЧЕНИЕ

В данной работе предлагается вариант покомпонентного метода, ускоренного с помощью универсальной проксимальной оболочки “ускоренный метаалгоритм”. Проведенный теоретический анализ предложенного метода позволяет утверждать, что зависимость его вычислительной сложности от размерности задачи и требуемой точности решения не уступает прочим методам, используемым для оптимизации выпуклых липшицево-гладких функций, а оценка вычислительной сложности сравнима с оценкой быстрого градиентного метода. При этом в предложенной схеме сохраняются свойства, характерные для классического покомпонентного метода, в том числе возможность использования свойств покомпонентной гладкости функции. Приведенные численные эксперименты подтверждают практическую эффективность метода, и также подчеркивают особенную релевантность предложенного подхода для часто возникающей в различных приложениях задачи оптимизации функции вида SoftMax.

Данная статья представляет результат работы над проектом, предложенным А.В. Гасниковым в рамках проектной смены (Ссылка на сайт проектной смены: <https://sochisirius.ru/obucheniye/graduates/smena673/3258>) “Современные методы теории информации, оптимизации и управления” Сириус 2–23 августа 2020 г. Авторы выражают благодарность организатору проектной смены А.С. Ненашеву за создание комфортных условий для работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Nesterov Y.E.* Lectures on convex optimization // Springer. 2018. Volume 137.
2. *Parikh N., Boyd S.* Proximal algorithms // Foundations and Trends in Optimization. 2014. V. 1. № 3. P. 127–239.
3. *Rockafellar R.T.* Monotone operators and the proximal point algorithm // SIAM Journal on Control and Optimization. 1976. V. 14. № 5. P. 877–898.
4. *Чернов А.В.* Прямо-двойственный метод решения задачи энтропийно-линейного программирования // Интеллектуальные системы. Теория и приложения. 2016. Т. 20. № 1. С. 39–59.
5. *Lin H., Mairal J., Harchaoui Z.* A universal catalyst for first-order optimization // Advances in Neural Information Processing Systems. 2015. V. 28. P. 3384–3392.
6. *Lin H., Mairal J., Harchaoui Z.* Catalyst acceleration for first-order convex optimization: from theory to practice // The Journal of Machine Learning Research. 2017. V. 18. № 1. P. 7854–7907.

7. *Ivanova A., Pasechnyuk D., Grishchenko D., Shulgin E., Gasnikov A., Matyukhin V.* Adaptive catalyst for smooth convex optimization // arXiv preprint arXiv:1911.11271, 2019.
8. *Kulunchakov A., Mairal J.* A generic acceleration framework for stochastic composite optimization // *Advances in Neural Information Processing Systems*. 2019. P. 12556–12567.
9. *Paquette C., Lin H., Drusvyatskiy D., Mairal J., Harchaoui Z.* Catalyst acceleration for gradient-based non-convex optimization // arXiv preprint arXiv:1703.10993, 2017
10. *Bubeck S., Jiang Q., Lee Y.T., Li Y., Sidford A.* Near-optimal method for highly smooth convex optimization // *Conference on Learning Theory*. 2019. P. 492–507.
11. *Doikov N., Nesterov Y.* Contracting proximal methods for smooth convex optimization // *SIAM Journal on Optimization*. 2020. V. 30. № 4. P. 3146–3169.
12. *Monteiro R., Svaiter B.F.* An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods // *SIAM Journal on Optimization*. 2013. V. 23. № 2. P. 1092–1125.
13. *Гасников А.В., Двинских Д.М., Двуреченский П.Е., Камзолов Д.И., Матюхин В.В., Пасечнюк Д.А., Тупица Н.К., Чернов А.В.* Ускоренный метаалгоритм для задач выпуклой оптимизации // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61 № 1. С. 20–31.
14. *Bubeck S.* Convex optimization: Algorithms and complexity // arXiv preprint arXiv:1405.4980, 2014.
15. *Nesterov Y., Stich S.U.* Efficiency of the accelerated coordinate descent method on structured optimization problems // *SIAM Journal on Optimization*. 2017. V. 27. № 1. P. 110–123.
16. *Kamzolov D., Gasnikov A., Dvurechensky P.* On the optimal combination of tensor optimization methods // arXiv preprint arXiv:2002.01004, 2020.
17. *Nesterov Yu.* Efficiency of coordinate descent methods on huge-scale optimization problems // *SIAM Journal on Optimization*. 2012. V. 22. № 2. P. 341–362.
18. *Anikin A., Dvurechensky P., Gasnikov A., Golov A., Gornov A., Maximov Yu., Mendel M., Spokoiny V.* Modern efficient numerical approaches to regularized regression problems in application to traffic demands matrix calculation from link loads // *Proc. of International Conference ITAS-2015*. 2015.
19. *Гасников А.В.* Современные численные методы оптимизации. Метод универсального градиентного спуска. МФТИ, 2018. С. 181.
20. Python Scipy documentation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.csr_matrix.html
21. *Blanchard Pierre, Higham Desmond J., Higham Nicholas J.* Accurately Computing the Log-Sum-Exp and Softmax Functions // *MIMS Preprint*. 2019.
22. *Гасников А.В., Гасникова Е.В., Нестеров Ю.Е., Чернов А.В.* Об эффективных численных методах решения задач энтропийно-линейного программирования // *Ж. вычисл. матем. и матем. физ.* 2016. Т. 56. № 4. С. 523–534.