СОДЕРЖАНИЕ

Том 62, номер 11, 2022 год

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

МАТЕМАТИЧЕСКАЯ ФИЗИКА

ИНФОРМАТИКА

EDN: NUXJIW
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1763–1803

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.633

Светлой памяти Бориса Вадимовича Рогова посвящается

КОМБИНИРОВАННЫЕ ЧИСЛЕННЫЕ СХЕМЫ[1)]

© 2022 г. М. Д. Брагин[1,2,*], О. А. Ковыркина[3,], М. Е. Ладонкина[1,***], В. В. Остапенко[3,****], В. Ф. Тишкин[1,*****], Н. А. Хандеева[3,******]**

[1] 125047 Москва, Миусская пл., 4, ИПМ РАН, Россия
[2] 141700 Долгопрудный, М.о., Институтский пер., 9, МФТИ, Россия
[3] 630090 Новосибирск, пр-т акад. Лаврентьева, 15, ИГиЛ СО РАН, Россия
**e-mail: michael@bragin.cc*
***e-mail: olyana@ngs.ru*
****e-mail: ladonkina@imamod.ru*
*****e-mail: ostapenko_vv@ngs.ru*
******e-mail: v.f.tishkin@mail.ru*
*******e-mail: nzyuzina1992@gmail.com*

Поступила в редакцию 05.02.2022 г.
Переработанный вариант 05.02.2022 г.
Принята к публикации 08.06.2022 г.

Представлен обзор работ по численным методам повышенной точности, предназначенным для сквозного расчета разрывных решений гиперболических систем законов сохранения. Сформулированы основные проблемы, возникающие в теории таких методов, и предложены подходы к их решению. Основное внимание уделяется принципиально новым численным методам сквозного счета (получившим название комбинированные схемы), которые монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния. Приведены тестовые расчеты, демонстрирующие существенные преимущества комбинированных схем по сравнению со стандартными NFC-схемами при расчете разрывных решений с ударными волнами. Библ. 99. Фиг. 18.

Ключевые слова: гиперболические системы законов сохранения, ударные волны, численные методы сквозного счета повышенной точности, комбинированные схемы.

DOI: 10.31857/S0044466922100027

1. ВВЕДЕНИЕ

В работе [1], широко известной в связи со схемой распада разрыва, было введено понятие монотонности численной схемы и показано, что среди линейных двухслойных по времени схем нет монотонных схем повышенного порядка аппроксимации. Дальнейшее развитие теории численных методов сквозного счета для гиперболических систем законов сохранения было направлено на преодоление этого запрета Годунова. В результате были разработаны различные классы конечно-разностных, конечно-объемных и проекционных схем, в которых повышенный порядок аппроксимации на гладких решениях и монотонность достигались за счет нелинейной коррекции потоков, приводящей к нелинейности этих схем при аппроксимации линейного уравнения переноса. Перечислим основные классы таких схем, которые будем сокращенно называть NFC (Nonlinear Flux Correction) схемами: MUSCL (см. [2]), TVD (см. [3]), ENO (см. [4]), CU (Central Upwind) (см. [5]), WENO (см. [6], [7]), DG (Discontinuous Galerkin) (см. [8]), CABARET (см. [9]), MBiC (Monotonized BiCompact) (см. [10], [11]). Основное достоинство этих схем заключается в том, что они с высокой точностью локализуют ударные волны при отсутствии существенных нефизических осцилляций на их фронтах.

[1)]Работа выполнена при частичной финансовой поддержке РФФИ и ГФЕН в рамках научного проекта № 21-51-53012 (разд. 4–7, 9), а также РНФ проект № 21-11-00198 (разд. 8, 10). The reported study was partially funded by RFBR and NSFC, project number 21-51-53012 (sect. 4–7, 9), and also RSF project number 21-11-00198 (sect. 4-7, 9).

При построении NFC-схем повышенный порядок аппроксимации понимается в смысле тейлоровского разложения на гладких решениях, что не гарантирует аналогичного повышения точности при расчете разрывных решений. Несмотря на это, в течение длительного времени преобладала ошибочная точка зрения, что эти схемы должны сохранять повышенную точность (соответствующую порядку их классической аппроксимации) во всех гладких частях рассчитываемых обобщенных решений. Способствовало распространению этого ошибочного мнения то, что в подавляющем числе работ тестирование разностных схем сквозного счета в значительной степени проводится на различных вариантах задачи Римана о распаде разрыва, точное решение которой представляет собой набор простых волн, соединенных областями постоянных течений (причем тестирование обычно ограничивается графическим сравнением численного решения с точным). Такое тестирование позволяет эффективно оценить разрешимость схемой сильных и слабых разрывов, а именно, ширину их размазывания и наличие или отсутствие осцилляций на фронтах ударных волн. Однако оно не может дать достаточной информации о реальной точности схемы в областях влияния ударных волн, поскольку точное решение за их фронтами является постоянным. Кроме того, эту точность нельзя оценить при расчете ударных волн, возникающих при решении скалярного закона сохранения, поскольку в этом случае область влияния устойчивой ударной волны совпадает с линией ее фронта.

Для корректного определения точности схемы в областях влияния ударных волн необходимо рассчитывать разрывные решения квазилинейных систем законов сохранения с ударными волнами, за фронтами которых формируется непостоянное решение. Такое решение для систем законов сохранения, как правило, не описывается точными формулами, и для определения скорости сходимости к нему разностного решения необходимо проведение серии расчетов на последовательности сжимающихся сеток. В [12]–[15] указанным способом было показано, что различные типы NFC-схем имеют не более чем первый порядок локальной сходимости в областях влияния ударных волн и, тем самым, по существу схемами повышенной точности не являются. Такое снижение порядков сходимости свидетельствует о том, что в этих схемах происходит потеря точности при передаче условий Гюгонио через размазанные фронты ударных волн. Однако свидетельствует опосредованно.

Для непосредственной оценки точности передачи схемой условий Гюгонио необходимо исследовать сходимость интегралов от разностного решения по областям, содержащим фронт ударной волны. Причем эти интегралы должны допускать потенциальную возможность получения повышенного (как минимум, второго) порядка сходимости для схем сквозного счета, в силу чего такая сходимость не может быть сильной, например, в нормах L_1 или L_2. Связано это с тем, что в схемах сквозного счета в нескольких узлах в окрестности фронта ударной волны отсутствует локальная сходимость разностного решения к точному, и поэтому порядок сходимости разностного решения в сильной норме, содержащей линию разрыва, в принципе не может быть выше первого.

В этой связи в [16] для TVD-схемы Хартена из [3], в [17] для различных вариантов WENO-схемы из [7], в [18] для DG-метода Кокбурна из [8], в [19] для монотонной модификации схемы CABARET из [9], в [20] для CU-схемы из [5] и в [21] для MBiC-схемы из [11] точность передачи схемой условий Гюгонио через фронт ударной волны оценивается путем определения порядка сходимости интеграла от разностного решения, что соответствует сходимости в соответствующей негативной норме. В [16]–[21] показано, что в рассмотренных NFC-схемах такой порядок интегральной сходимости снижается до первого на интервалах интегрирования, одна или обе границы которых находятся в области влияния ударной волны. Одна из причин такого снижения точности заключается в том, что коррекция потоков, характерная для этих схем, приводит к снижению гладкости разностных потоков, что в свою очередь приводит к снижению порядка аппроксимации ε-условий Гюгонио на фронтах ударных волн (см. [22]). В то же время, как показано в [15], [16], некоторые немонотонные схемы повышенной точности (в частности, схема Русанова из [23] и компактная схема из [15]), имеющие аналитические функции численных потоков и, как следствие, с повышенной точностью аппроксимирующие ε-условия Гюгонио, сохраняют повышенный порядок сходимости в негативной норме при интегрировании по областям, содержащим сильные разрывы. Эти немонотонные схемы, в отличие от NFC-схем, сохраняют повышенный порядок сходимости в областях влияния ударных волн, несмотря на заметные схемные осцилляции на их фронтах. Далее для таких немонотонных схем повышенной точности мы будем использовать аббревиатуру HASIA (High Accuracy Shock Influence Area).

В результате в теории численных схем сквозного счета сложилась следующая альтернатива: невозможно одновременно с высокой точностью локализовать сильные разрывы и сохранить

повышенный порядок сходимости в областях их влияния. При этом на практике NFC-схемы (особенно схемы WENO и DG) широко применяются при численном моделировании сложных газо- и гидродинамических течений с большим числом ударных волн различной амплитуды, в силу чего все такие расчеты имеют лишь первый порядок точности. Первая попытка решения этой проблемы была связана с применением методики построения гибридных схем (см. [24]–[28]), при которой на каждом временном слое численное решение сначала строится с помощью внешней немонотонной HASIA-схемы, имеющей заметные осцилляции на ударной волне. После этого в окрестности фронта ударной волны численное решение стандартным образом корректируется с помощью одной из NFC-схем, и на новом временном слое получается монотонизированное численное решение без заметных нефизических осцилляций. Однако тестовые расчеты показали, что в построенной таким образом гибридной схеме теряется основное преимущество внешней HASIA-схемы – ее повышенная точность в областях влияния ударных волн; эта точность снижается приблизительно так же, как и в стандартных NFC-схемах.

Данный недостаток как NFC-схем, так и гибридных схем, непосредственно связан с их главным преимуществом – монотонной локализацией фронтов ударных волн, поскольку любая конечная сумма ряда Фурье для разрывной функции не является монотонной. С учетом этого осцилляции, возникающие на фронтах ударных волн в немонотонных HASIA-схемах, несут информацию о волновой структуре фурье-разложения разрывной функции в окрестности сильного разрыва, что позволяет этим схемам с повышенной точностью передавать условия Гюгонио и, как следствие, сохранять повышенную точность в областях влияния ударных волн. NFC-схемы и гибридные схемы в результате искусственного сглаживания численных ударных волн эту информацию теряют, что приводит к снижению их точности при аппроксимации условий Гюгонио.

В [29] была предложена методика построения принципиально нового класса численных методов сквозного счета (получивших название комбинированные схемы), которые сочетают достоинства как NFC-схем, так и немонотонных HASIA-схем, а именно, монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния. В комбинированной схеме применяется базисная немонотонная HASIA-схема, по которой разностное решение строится во всей расчетной области. В окрестностях больших градиентов, где это решение имеет нефизические осцилляции, оно корректируется путем численного решения внутренних начально-краевых задач по одной из NFC-схем. Причем внутренняя NFC-схема (в отличие от случая гибридных схем) не влияет на решение, получаемое по базисной схеме, что позволяет комбинированной схеме сохранять повышенную точность в областях влияния ударных волн. В [29] была рассмотрена комбинированная схема, в которой в качестве базисной HASIA-схемы использовалась компактная схема третьего порядка слабой аппроксимации (см. [15]), а в качестве внутренней NFC-схемы – монотонная модификация схемы CABARET (см. [9]) второго порядка точности на гладких решениях. Далее для компактной схемы из [15] будем использовать аббревиатуру CWA (Compact Weak Approximation), а для схемы CABARET, предложенной в [30], будем применять аббревиатуру CABARETM.

Недостаток комбинированной схемы, построенной в [29], заключался в том, что соответствующие ей базисная и внутренняя схемы имели существенно различные типы: базисная CWA-схема являлась неявной и трехслойной по времени, в то время как внутренняя схема CABARETM – явной и двухслойной по времени, что приводило к определенным сложностям при численной реализации такого алгоритма. Поэтому в [31] был предложен новый вариант комбинированной разностной схемы, в которой как немонотонная базисная HASIA-схема, так и внутренняя NFC-схема являлись явными и двухслойными по времени. А именно, в качестве базисной была использована схема Русанова третьего порядка (см. [23]), а в качестве внутренней – схема CABARETM второго порядка (см. [30]). В [29] и [31] приведены тестовые расчеты разрывных решений с ударными волнами, демонстрирующие существенные преимущества построенных в них комбинированных разностных схем по сравнению со схемой WENO5 из [7] пятого порядка по пространству и третьего порядка по времени.

Следующий этап в развитии теории комбинированных схем был связан с разработкой согласованных численных алгоритмов, в которых базисная и внутренняя схемы являются схемами одного класса, а именно, внутренняя схема получается из базисной схемы в результате применения соответствующей NFC-процедуры. В [32] такая комбинированная схема была построена на основе DG-метода из [8], при этом в качестве базисной схемы использовался немонотонный вариант DG-метода третьего порядка без применения NFC-процедуры, а в качестве внутренней схемы использовался монотонный вариант этого метода, в котором коррекция потоков осуществ-

лялась с помощью ограничителя Кокбурна (см. [8]). Построенная комбинированная DG-схема показала существенные преимущества по сравнению со стандартной NFC-схемой DG-метода при расчете нестационарных ударных волн.

В [33] согласованная комбинированная схема была построена на основе бикомпактных разностных схем. В качестве внутренней схемы применялась бикомпактная схема четвертого порядка аппроксимации по пространству с интегрированием по времени неявным методом Эйлера. Эта схема устойчива при любых соотношениях шагов сетки и монотонна при числах Куранта, не меньших 1/4. В отличие от работ [29], [31], [32] вычисления по внутренней схеме велись во всей пространственно-временной расчетной области. Решение базисной схемы определялось посредством глобальной (пассивной в терминологии [34]) экстраполяции Ричардсона второго порядка по решениям внутренней схемы, рассчитанным на двух вложенных сетках. Локальные немонотонности у решения базисной схемы устранялись с помощью процедуры, близкой по своей сути к гибридной схеме из [10]. Однако эта процедура имела характер пост-обработки, т.е. применялась лишь после завершения вычислений по базисной и внутренней схемам, а не при каждом переходе со слоя на слой. Комбинированная бикомпактная схема из [33] продемонстрировала высокий порядок точности, который не достигается MBiC-схемой (см. [11]).

В настоящей работе более детально излагаются теория комбинированных схем, методы их построения и результаты тестовых расчетов, демонстрирующие преимущества этого нового класса схем по сравнению со стандартными NFC-схемами. В разд. 2 описываются методы оценки точности схем сквозного счета, аппроксимирующих гиперболическую систему законов сохранения. В разд. 3 приводится основная тестовая задача для уравнений теории мелкой воды, в результате численного решения которой проводится сравнительный анализ точности изучаемых схем. В разд. 4 описываются численные алгоритмы разностных схем Русанова, CWA, CABARETM и WENO5, а в разд. 5 приводятся результаты расчетов по этим схемам основной тестовой задачи, из которых следует, что HASIA-схемы Русанова и CWA имеют существенно более высокую точность в областях влияния ударных волн, чем NFC-схемы CABARETM и WENO5. В разд. 6 излагается общая методика построения гибридных численных схем и изучаются две гибридные схемы, в которых в качестве базисной используется схема Русанова или CWA-схема, а в качестве внутренней применяется схема CABARETM. Тестовые расчеты показали, что в этих гибридных схемах теряется основное преимущество исходных HASIA-схем – повышенная точность в областях влияния ударных волн, которая становится сравнимой с точностью NFC-схем.

В разд. 7 описывается методика построения комбинированных схем, которые монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния. Построены и протестированы две комбинированные схемы, в которых базисной является схема Русанова или CWA-схема, а внутренней – схема CABARETM. В разд. 8 изучаются согласованные комбинированные схемы, в которых базисная и внутренняя схемы являются схемами одного класса, а именно, внутренняя схема получается из базисной HASIA-схемы в результате применения соответствующей NFC-процедуры. В п. 8.1 такая комбинированная схема построена на основе DG-метода из [8], а в п. 8.2 – на основе бикомпактных разностных схем из [10]. В разд. 9 проведен сравнительный анализ точности комбинированных разностных схем, построенных в разд. 7, со схемой WENO5 при численном моделировании задачи о многократном взаимодействии ударных волн. В разд. 10 построена двумерная комбинированная бикомпактная схема, которая при расчете пространственно двумерных разрывных решений обеспечивает второй порядок локальной сходимости в областях влияния ударных волн. В разд. 11 дается общая характеристика полученных результатов, приводится их сравнение с современным мировым уровнем данного научного направления и формулируются основные перспективные направления дальнейшего развития теории комбинированных схем.

2. МЕТОДЫ ОЦЕНКИ ТОЧНОСТИ СХЕМ СКВОЗНОГО СЧЕТА

Рассмотрим квазилинейную строго гиперболическую систему законов сохранения (см. [35], [36])

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = \mathbf{0}, \tag{2.1}$$

где $\mathbf{u}(x,t)$ – искомая, а $\mathbf{f}(\mathbf{u})$ – заданная гладкие вектор-функции, содержащие m компонент. Строгая гиперболичность системы (2.1) означает, что все собственные значения $\lambda_i(\mathbf{u})$ матрицы Якоби $A(\mathbf{u}) = \mathbf{f}_\mathbf{u}(\mathbf{u})$ действительны и различны, в силу чего соответствующие им системы левых

$\mathbf{l}^i(\mathbf{u})$ и правых $\mathbf{r}^i(\mathbf{u})$ собственных векторов матрицы $A(\mathbf{u})$ формируют базисы в пространстве $\mathbb{R}^m$. Поставим для системы (2.1) задачу Коши с периодическими начальными данными

$$\mathbf{u}(x,0) = \mathbf{u}_0(x) = \mathbf{u}_0(x+X), \tag{2.2}$$

где $\mathbf{u}_0(x)$ – заданная гладкая вектор-функция, X – длина периода. Предположим, что задача (2.1), (2.2) имеет единственное слабое решение $\mathbf{u}(x,t)$, которое является ограниченным, и в котором в результате градиентных катастроф при $t > 0$ возникают ударные волны.

Явные численные схемы, аппроксимирующие задачу (2.1), (2.2), будем строить на равномерной прямоугольной сетке

$$S = \{(x_j,t_n) : x_j = jh,\ t_n = n\tau, \quad n \geq 0\}, \tag{2.3}$$

где $h = X/M$ – шаг сетки по пространству, M – заданное целое положительное число,

$$\tau = zh/\max_{k,j,n}\left|\lambda_k(\mathbf{v}_h(x_{j+\alpha},t_n))\right| \tag{2.4}$$

есть шаг сетки по времени, выбирамый из условия устойчивости Куранта, в котором $z \in (0,1)$ – коэффициент запаса, $\mathbf{v}_h(x_{j+\alpha},t_n)$ – численное решение в пространственном узле $x_{j+\alpha} = (j+\alpha)h$, где $\alpha = 0$ или $\alpha = 1/2$ в зависимости от вида конкретной разностной схемы. В случае проведения прикладных расчетов для экономии компьютерного времени более естественно использовать неравномерные по времени численные сетки

$$\bar{S} = \{(x_j,t_n)\colon x_j = jh,\ t_{n+1} = t_n + \tau_n,\ t_0 = 0\},$$

в которых шаг по времени τ_n определяется по формуле

$$\tau_n = zh/\max_{k,j}\left|\lambda_k(\mathbf{v}_h(x_{j+\alpha},t_n))\right|.$$

Однако в настоящей работе, целью которой является экспериментальное изучение точности численных схем на последовательности сжимающихся сеток, более удобно применение равномерной численной сетки (2.3) с постоянным шагом по времени (2.4).

Для приближенной оценки порядков локальной сходимости численного решения $\mathbf{v}_h$, построенного на равномерной сетке (2.3), зафиксируем на этой сетке некоторый узел $(x_{j+\alpha},t_n)$, где $n \geq 1$, и введем для него новое обозначение $(\tilde{x},\tilde{t})$, где $\tilde{x} = (j+\alpha)h$ и $\tilde{t} = n\tau > 0$. Рассмотрим последовательность сгущающихся сеток

$$S_i = \{(x_j^i,t_n^i) : x_j^i = jh_i,\ t_n^i = n\tau_i,\ n \geq 0\}, \quad i = 1,2,\ldots, \tag{2.5}$$

где $h_i = h/k^{i-1}$, $\tau_i = \tau/k^{i-1}$; k – целое положительное число, удовлетворяющее условию $k \geq 2$ при $\alpha = 0$ и условию $k = 3l \geq 3$, l – натуральное число, при $\alpha = 1/2$. Предположим, что на последовательности сеток (2.5), получаемых путем сжатия базисной сетки (2.3), численное решение $\mathbf{v}_{h_i}$ в точке $(\tilde{x},\tilde{t})$ сходится к точному решению $\mathbf{u}$ с порядком r. Это означает, что в точке $(\tilde{x},\tilde{t})$ с точностью $o(h_i^r)$ выполнено условие

$$\mathbf{v}_{h_i} - \mathbf{u} = \mathbf{B}h_i^r \ \Rightarrow\ \left|\mathbf{v}_{h_i} - \mathbf{u}\right| = \left|\mathbf{B}\right|h_i^r, \tag{2.6}$$

где $\mathbf{B}$ – векторная величина, не зависящая от h_i и такая, что $\left|\mathbf{B}\right| > 0$. Беря отношение вторых равенств (2.6) при $i = 1,2$, имеем

$$\frac{\left|\mathbf{v}_{h_1} - \mathbf{u}\right|}{\left|\mathbf{v}_{h_2} - \mathbf{u}\right|} = \left(\frac{h_1}{h_2}\right)^r = k^r.$$

Отсюда следует формула Рунге

$$r = r(\tilde{x},\tilde{t}) = \log_k\frac{|\,\mathbf{v}_{h_1} - \mathbf{u}\,|}{|\,\mathbf{v}_{h_2} - \mathbf{u}\,|} = \log_{1/k}\frac{|\,\mathbf{v}_{h_2} - \mathbf{u}\,|}{|\,\mathbf{v}_{h_1} - \mathbf{u}\,|}, \tag{2.7}$$

которую можно применять для приближенного определения порядков локальной сходимости численного решения к точному.

В случае дискретных конечно-разностных и конечно-объемных схем, для которых численное решение определено в целых или в полуцелых пространственных узлах численной сетки (2.3), для приближенного определения порядков интегральной сходимости численного решения к точному зададим целое число $N \geq 2$ и момент времени $T = N\tau = N\tau_i k^{i-1} > 0$, для которого путем линейной или квадратичной (параболической) интерполяции доопределим дискретные численные решения $\mathbf{v}_{h_i}(x^i_{j+\alpha}, t^i_n)$ до непрерывных по x функций $\mathbf{v}_{h_i}(x,T)$. В случае проекционной схемы DG-метода [8] численное решение на сетке (2.5) строится как функция $\mathbf{v}_{h_i}(x,t^i_n)$, зависящая от непрерывно изменяющегося аргумента x, которая в пространственных узлах x^i_j в общем случае имеет сильные разрывы. Поэтому для DG-метода соответствующая функция $\mathbf{v}_{h_i}(x,T)$ непосредственно получается из самого численного решения.

Зафиксируем отрезок $[a,b]$ на оси x и зададим интегралы

$$\mathbf{U}(a,b,T) = \int\limits_a^b \mathbf{u}(x,T)dx, \quad \mathbf{V}_{h_i}(a,b,T) = \int\limits_a^b \mathbf{v}_{h_i}(x,T)dx,$$

где в случае дискретных схем интеграл $\mathbf{V}_{h_i}$ вычисляется по формуле трапеций или формуле парабол в зависимости от способа интерполяции численного решения. Следуя [16], будем говорить, что последовательность численных решений $\mathbf{v}_{h_i}(x,T)$ сходится на отрезке $[a,b]$ с порядком ρ к точному решению $\mathbf{u}(x,T)$, если с точностью $o(h_i^\rho)$ выполнено условие

$$\mathbf{V}_{h_i}(a,b,T) - \mathbf{U}(a,b,T) = \mathbf{C}h_i^\rho, \tag{2.8}$$

где $\mathbf{C}$ – векторная величина, не зависящая от h_i и такая, что $|\mathbf{C}| > 0$. При выполнении условия (2.8), по аналогии с (2.7), мы получаем следующую формулу

$$\rho = \rho(a,b,T) = \log_k \frac{|\mathbf{V}_{h_1}(a,b,T) - \mathbf{U}(a,b,T)|}{|\mathbf{V}_{h_2}(a,b,T) - \mathbf{U}(a,b,T)|} = \log_{1/k} \frac{|\mathbf{V}_{h_2}(a,b,T) - \mathbf{U}(a,b,T)|}{|\mathbf{V}_{h_1}(a,b,T) - \mathbf{U}(a,b,T)|} \tag{2.9}$$

для приближенного определения порядков интегральной сходимости численного решения к точному. Обоснование метода интегральной сходимости для исследования точности конечно-разностных схем сквозного счета было дано в [37].

Если интегралы $\mathbf{V}_{h_i}(t,a,b)$ в случае дискретных схем вычисляются по формуле трапеций (которая имеет второй порядок точности на гладких функциях), то порядок интегральной сходимости в общем случае удовлетворяет условию $\rho \leq 2$; если эти интегралы вычисляются по формуле парабол (которая имеет четвертый порядок точности на гладких функциях), то в общем случае $\rho \leq 4$. Для DG-метода указанное ограничение на порядок интегральной сходимости отсутствует и при расчете гладких решений этот порядок будет совпадать с формальным порядком аппроксимации DG-метода. При расчете разрывных решений порядок интегральной сходимости на отрезках $[a,b]$, содержащих ударные волны, может снижаться за счет потери точности при передаче схемой условий Гюгонио через размазанные фронты ударных волн.

Поскольку в рассматриваемых далее тестовых задачах Коши (2.1), (2.2) точное решение $\mathbf{u}$ заранее неизвестно, то для приближенного вычисления порядков сходимости разностного решения мы не можем непосредственно воспользоваться формулами (2.7) и (2.9), которые зависят от этого точного решения. Это затруднение можно преодолеть двумя различными способами. При первом способе точное решение $\mathbf{u}$ в формулах (2.7) и (2.9) заменяется на некоторое квазиточное численное решение $\mathbf{v}_{h_*}$, получаемое по одной из схем повышенной точности на равномерной сетке с пространственным шагом $h_* \ll h$, где h – пространственный шаг базисной сетки (2.3). При втором способе необходимо выполнить три численных расчета на сетках (2.5) с пространственными шагами $h_1 = h$, $h_2 = h/k$ и $h_3 = h/k^2$. Вычитая из первой формулы (2.6) эту же формулу, в которой индекс i заменен на $i+1$, получаем

$$\mathbf{v}_{h_i} - \mathbf{v}_{h_{i+1}} = \mathbf{B}\left(h_i^r - h_{i+1}^r\right) \quad \Rightarrow \quad \left|\mathbf{v}_{h_i} - \mathbf{v}_{h_{i+1}}\right| = |\mathbf{B}|h_i^r\left(1 - k^{-r}\right). \tag{2.10}$$

Беря отношение вторых равенств (2.10) при $i = 1, 2$, имеем

$$\frac{|\mathbf{v}_{h_1} - \mathbf{v}_{h_2}|}{|\mathbf{v}_{h_2} - \mathbf{v}_{h_3}|} = \left(\frac{h_1}{h_2}\right)^r = k^r. \tag{2.11}$$

Отсюда следует формула Рунге

$$r = r(\tilde{x}, \tilde{t}) = \log_k \frac{|\mathbf{v}_{h_1} - \mathbf{v}_{h_2}|}{|\mathbf{v}_{h_2} - \mathbf{v}_{h_3}|} = \log_{1/k} \frac{|\mathbf{v}_{h_2} - \mathbf{v}_{h_3}|}{|\mathbf{v}_{h_1} - \mathbf{v}_{h_2}|}, \tag{2.12}$$

не зависящая от точного решения $\mathbf{u}$. Аналогичным образом для приближенного определения порядка интегральной сходимости получаем формулу

$$\rho = \rho(a, b, T) = \log_k \frac{|\mathbf{V}_{h_1}(a, b, T) - \mathbf{V}_{h_2}(a, b, T)|}{|\mathbf{V}_{h_2}(a, b, T) - \mathbf{V}_{h_3}(a, b, T)|} = \log_{1/k} \frac{|\mathbf{V}_{h_2}(a, b, T) - \mathbf{V}_{h_3}(a, b, T)|}{|\mathbf{V}_{h_1}(a, b, T) - \mathbf{V}_{h_2}(a, b, T)|}. \tag{2.13}$$

Следует отметить, что первый способ вычисления порядков сходимости, использующий квазиточное численное решение, формально является более точным, но и существенно более трудоемким по сравнению со вторым способом, при котором применяются формулы (2.12) и (2.13). Поэтому в приводимых далее расчетах порядки сходимости в основном определяются с помощью второго способа, корректность которого предварительно проверяется путем сравнения на контрольных тестах с результатами, получаемыми на основе первого способа. Численные расчеты показывают (см. [12], [15]), что порядки локальной сходимости, вычисляемые по формуле (2.12), могут сильно осциллировать в областях влияния ударных волн, особенно у NFC-схем, что не позволяет с необходимой точностью определить значения этих порядков; в [15], [18], [19] для подавления таких осцилляций перед применением формулы (2.12) использовались различные методы локального осреднения получаемого разностного решения. Поэтому в данной работе мы в основном будем приводить порядки интегральной сходимости (2.13), позволяющие оценить точность, с которой численное решение аппроксимирует условия Гюгонио на фронте ударной волны.

Пусть $w = w(\mathbf{u})$ – гладкая скалярная функция векторного решения $\mathbf{u}$, точность вычисления которой необходимо оценить. Для этого так же, как при определении порядков сходимости численного решения, можно применять два различных способа. При использовании квазиточного численного решения $\mathbf{v}_{h_*}$ сразу получается формула

$$\delta w_h = w(\mathbf{u}_h) - w(\mathbf{u}_{h_*}) \tag{2.14}$$

для дисбаланса (ошибки) вычисления функции w на базисной сетке (2.3). В случае применения второго способа, связанного с использованием формул (2.10)–(2.12), мы с учетом (2.6) предполагаем, что с точностью $o(h_i^r)$ выполнено условие

$$w_{h_i} - w = Dh_i^r \;\Rightarrow\; w_{h_i} - w_{h_{i+1}} = D\left(h_i^r - h_{i+1}^r\right) = Dh_i^r\left(1 - k^{-r}\right), \tag{2.15}$$

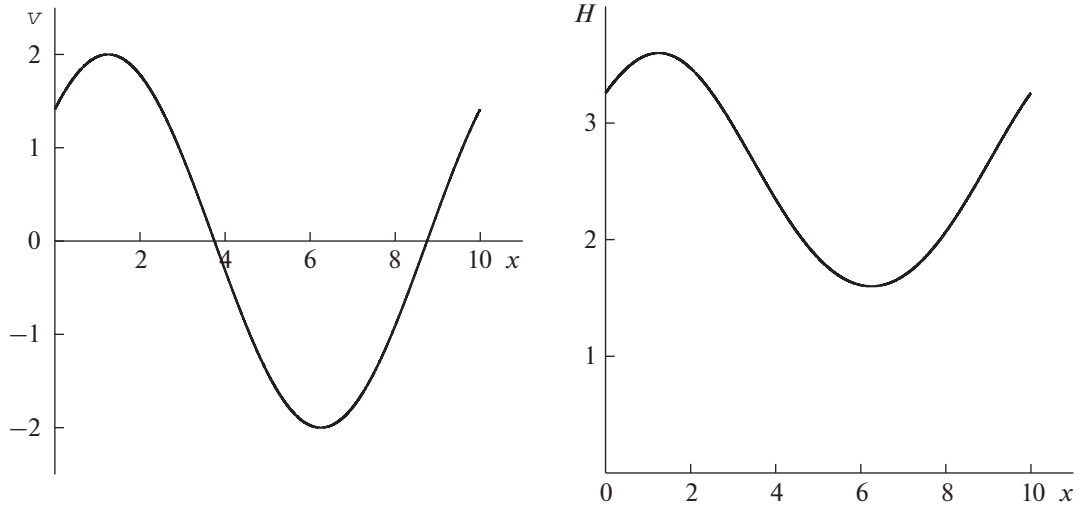
где $w_{h_i} = w(\mathbf{u}_{h_i})$ и D – скалярная величина, не зависящая от h_i. Из формул (2.15) при $i = 1$ имеем

$$D = \frac{w_{h_1} - w_{h_2}}{h_i^r\left(1 - k^{-r}\right)} \;\Rightarrow\; w_{h_1} - w = \frac{w_{h_1} - w_{h_2}}{1 - k^{-r}}. \tag{2.16}$$

Подставляя во вторую формулу (2.16) выражение для порядка сходимости r, задаваемое уравнением (2.12), и учитывая, что $h = h_1$, получаем приближенную формулу

$$\delta w_h = w_h - w = \left(w_{h_1} - w_{h_2}\right)\left(1 - \frac{|\mathbf{v}_{h_2} - \mathbf{v}_{h_3}|}{|\mathbf{v}_{h_1} - \mathbf{v}_{h_2}|}\right)^{-1} \tag{2.17}$$

для дисбаланса вычисления функции w.



Фиг. 1. Начальные значения скорости жидкости v и глубины жидкости H, задаваемые формулами (3.2).

В приводимых далее тестовых расчетах применяются обе формулы (2.14) и (2.17). При этом на графиках мы будем показывать относительные дисбалансы вычисления функции w, определяемые по формуле

$$\Delta w_h = \lg\frac{|\delta w_h|}{|w_h|} = \lg\frac{|w_h - w|}{|w_h|}. \tag{2.18}$$

3. ОСНОВНАЯ ТЕСТОВАЯ ЗАДАЧА

В качестве конкретной гиперболической системы законов сохранения выберем систему уравнений первого приближения теории мелкой воды (см. [38]), дивергентная форма записи которой в случае прямоугольного горизонтального русла без учета влияния донного трения имеет вид (2.1), где

$$\mathbf{u} = \begin{pmatrix} H \\ q \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} q \\ \dfrac{q^2}{H} + \dfrac{gH^2}{2} \end{pmatrix}. \tag{3.1}$$

Здесь $H(x,t)$ и $q(x,t)$ – глубина и расход жидкости, g – ускорение свободного падения (в расчетах $g = 10$). Рассмотрим для системы (2.1), (3.1), задачу Коши (2.2) с начальными данными (фиг. 1)

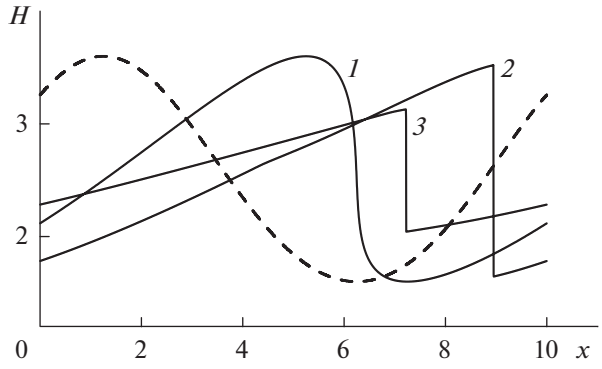
$$v(x,0) = a\sin\left(\frac{2\pi x}{X} + \frac{\pi}{4}\right), \quad H(x,0) = \frac{\bigl(v(x,0)+b\bigr)^2}{4g} = \frac{1}{4g}\left[a\sin\left(\frac{2\pi x}{X} + \frac{\pi}{4}\right) + b\right]^2, \tag{3.2}$$

где $v = q/H$ – горизонтальная скорость жидкости, $a = 2$, $X = b = 10$. Начальным данным (3.2) соответствуют следующие начальные значения инвариантов $w_1 = v - 2c$ и $w_2 = v + 2c$:

$$w_1(x,0) = -b, \quad w_2(x,0) = 2v(x,0) + b = 2a\sin\left(\frac{2\pi x}{X} + \frac{\pi}{4}\right) + b, \tag{3.3}$$

где $c = \sqrt{gH}$ – скорость распространения малых возмущений. Из формул (3.3) следует, что в начальный момент времени инвариант w_1, соответствующий характеристическому направлению $\lambda_1 = v - c$, является постоянным, а инвариант w_2, соответствующий характеристическому направлению $\lambda_1 = v + c$, является периодической функцией пространственной координаты.

В точном решении задачи (2.1), (3.1), (3.2) в момент времени $t \approx 0.54$ в результате градиентных катастроф формируется последовательность изолированных ударных волн, которые распространяются друг за другом с одинаковыми скоростями в положительном направлении оси x, в силу чего расстояние между соседними ударными волнами остается постоянным и равным длине пе-



Фиг. 2. Глубины жидкости H в квазиточном решении задачи Коши (2.1), (3.1), (3.2), получаемые в моменты времени $t = 0$ (штриховая линия), $t = 0.5$ (линия *1*), $t = 1$ (линия *2*) и $t = 2.5$ (линия *3*).

риода X. На фиг. 2 на отрезке $[0, X]$ длины периода штриховой линией показана начальная глубина жидкости, задаваемая второй формулой (3.2), а сплошными линиями изображены квазиточные профили глубины, получаемые в результате численного расчета по схеме CABARETM из [30] на достаточно мелкой сетке в моменты времени $t = 0.5$, $t = 1$ и $t = 2.5$. К моменту времени $t = 0.5$ в точном решении начинают формироваться области больших градиентов, но решение еще остается гладким (линия *1* на фиг. 2). Ударные волны, которые в момент времени $t \approx 0.54$ возникают как сильные разрывы первоначально бесконечно малой амплитуды, в момент времени $t = 1$ имеют конечную амплитуду (линия *2* на фиг. 2), но область их влияния, расположенная внутри интервала $(4,9) \subset [0, X]$, еще не заполняет всю расчетную область. К моменту времени $t = 2.5$ ударные волны проходят расстояние, большее длины периода X, и вся расчетная область становится их областью влияния. С учетом этого сильный разрыв, расположенный на линии *3* фиг. 2, соответствует ударной волне, которая сформировалась при $t \approx 0.54$ внутри интервала $[-X, 0]$.

4. ЧИСЛЕННЫЕ СХЕМЫ РУСАНОВА, CWA, CABARETM И WENO5

В данном разделе приводятся численные алгоритмы схем Русанова из [23], CWA из [15], CABARETM из [30] и WENO5 из [7], аппроксимирующих гиперболическую систему (2.1). Результаты тестовых расчетов по этим схемам задачи Коши (2.1), (3.1), (3.2), приведенные в следующем разд. 5, лежат в основе методики построения комбинированных схем. Отметим, что в схемах Русанова, CWA и WENO5 численные решения определяются в целых пространственных узлах $x_j = jh$ базисной сетки (2.3), в силу чего, наряду с $\mathbf{v}_h$, мы будем использовать для них сокращенные обозначения $\mathbf{v}_j^n = \mathbf{v}_h(x_j, t_n)$.

4.1. Схема Русанова

Исторически схема Русанова была первой явной разностной схемой третьего порядка, в которой для аппроксимации по времени использовался соответствующий метод Рунге–Кутты (краткая информация об этой схеме опубликована в [23], ее детальный анализ проведен в [39]). Если на n-м временном слое известно численное решение $\mathbf{v}_j^n$, то в схеме Русанова на $(n+1)$-м временном слое решение $\mathbf{v}_j^{n+1}$ определяется по формулам

$$\mathbf{v}_{j+1/2}^{(1)} = \frac{\mathbf{v}_j^n + \mathbf{v}_{j+1}^n}{2} - \frac{R\left(\mathbf{f}_{j+1}^n - \mathbf{f}_j^n\right)}{3}, \quad \mathbf{v}_j^{(2)} = \mathbf{v}_j^n - \frac{2R\left(\mathbf{f}_{j+1/2}^{(1)} - \mathbf{f}_{j-1/2}^{(1)}\right)}{3}, \tag{4.1}$$

$$\mathbf{v}_j^{n+1} = \mathbf{v}_j^n - \frac{R\left(7\left(\mathbf{f}_{j+1}^n - \mathbf{f}_{j-1}^n\right) - 2\left(\mathbf{f}_{j+2}^n - \mathbf{f}_{j-2}^n\right)\right)}{24} - \frac{3R\left(\mathbf{f}_{j+1}^{(2)} - \mathbf{f}_{j-1}^{(2)}\right)}{8} - \frac{C\mathbf{w}_j^n}{24}, \tag{4.2}$$

в которых $R = \tau/h$, $\mathbf{f}_j^n = \mathbf{f}(\mathbf{v}_j^n)$, $\mathbf{f}_{j\pm1/2}^{(1)} = \mathbf{f}(\mathbf{v}_{j\pm1/2}^{(1)})$, $\mathbf{f}_{j\pm1}^{(2)} = \mathbf{f}(\mathbf{v}_{j\pm1}^{(2)})$;

$$\mathbf{w}_j^n = \mathbf{v}_{j+2}^n - 4\mathbf{v}_{j+1}^n + 6\mathbf{v}_j^n - 4\mathbf{v}_{j-1}^n + \mathbf{v}_{j-2}^n \tag{4.3}$$

есть искусственная вязкость четвертого порядка дивергентности, аппроксимирующая четвертую пространственную производную $h^4\mathbf{u}_{xxxx}$, где $Ch^3/(24r)$ – коэффициент вязкости.

Поскольку схему Русанова (4.1)–(4.3) можно представить в виде явной двухслойной по времени разностной схемы с аналитической вектор-функцией численных потоков, то она удовлетворяет основной теореме работы [22], в силу которой ее порядок точности на гладких решениях совпадает с порядком аппроксимации этой схемой ε-условий Гюгонио, что обеспечивает повышенную точность при передаче этих условий через размазанные фронты ударных волн. При $C = 0$ схема (4.1)–(4.3) имеет порядок аппроксимации $O(h^4 + \tau^3)$. Однако в этом случае она является неустойчивой в линейном приближении (см. [23], [39]); для ее устойчивости необходимо, чтобы коэффициент C был положителен и удовлетворял неравенствам

$$3 \geq C \geq z^2(4 - z^2),$$

что приводит к третьему порядку схемы Русанова как по времени, так и по пространству. Тестовые расчеты разрывных решений уравнений газовой динамики проводились при близких значениях $C = 2.5$ в [23] и $C = 2.8$ в [39]. В настоящей работе мы будем использовать значение $C = 2.5$.

4.2. Компактная схема третьего порядка слабой аппроксимации (CWA-схема)

При построении практически всех реально используемых численных схем сквозного счета, как немонотонных (см. [23], [40], [41]), так и NFC-схем (см. [2]–[10]), повышенный порядок аппроксимации понимается в смысле тейлоровского разложения на гладких решениях, что не гарантирует аналогичного повышения точности при расчете обобщенных решений, поскольку классическое понятие аппроксимации становится неопределенным в окрестностях сильных разрывов. В то же время именно аппроксимацией в окрестности ударной волны определяется точность, с которой разностная схема передает условия Гюгонио через ее фронт. В связи с этим в [42] было введено понятие слабой численной аппроксимации гиперболической системы (2.1) на классе кусочно-непрерывных ограниченных функций и получены необходимые и достаточные условия такой аппроксимации (в том числе с повышенным порядком); причем в число необходимых условий входит консервативность разностной схемы, эквивалентность различных определений которой изучалась в [43]. Однако эти условия приводят к достаточно жестким ограничениям на вид численной схемы (для k-го порядка слабой аппроксимации необходимо, чтобы схема имела не менее k временных слоев). Причина этого заключается в том, что в [42] аппроксимация изучалась не на слабых решениях, а на классе кусочно-непрерывных ограниченных функций и, тем самым, в ней речь, по существу, шла о слабой аппроксимации разностным оператором дивергентного дифференциального оператора, из которой следует соответствующая слабая аппроксимация схемой системы (2.1).

В то же время, поскольку большинство реально используемых схем сквозного счета (см. [2]–[10], [23], [40], [41]) являются явными и двухслойными по времени, в них повышение порядка достигается за счет использования дифференциальных следствий аппроксимируемой системы на ее гладких решениях, в силу чего аппроксимационные свойства этих схем на разрывных решениях требуют специального, более детального изучения. С этой целью в [44] на примере явных двухслойных по времени консервативных схем была проанализирована возможность использования дифференциальных и интегральных следствий законов сохранения для повышения порядка слабой аппроксимации на разрывных решениях. Было показано, что (в отличие от линейного случая) дифференциальные следствия квазилинейного закона сохранения (за счет которых происходит повышение порядка на гладких решениях) в общем случае не имеют интегральных аналогов и поэтому для нелинейных схем повышение порядка аппроксимации на гладких решениях в общем случае не приводят к аналогичному повышению порядка слабой аппроксимации на разрывных решениях. В [22], [45] была изучена точность, с которой явные двухслойные по времени консервативные схемы аппроксимируют (ε, δ)-условия Гюгонио, представляющие собой соотношения, связывающие значения точного разрывного решения в точках $(x(t) + \varepsilon, t - \delta)$ и $(x(t) - \varepsilon, t + \delta)$ по обе стороны от линии фронта $x = x(t)$ нестационарной ударной волны, для которой $x'(t) > 0$. В [22] показано, что при $\delta = 0$ на ударных волнах, линии фронтов которых являются достаточно гладкими, разностные схемы с достаточно гладкими функциями численных потоков аппроксимируют $(\varepsilon, 0)$-условия (ε-условия) Гюгонио с тем же порядком, который они

имеют на гладких решениях. В [45] показано, что при $\delta \neq 0$ эти схемы аппроксимируют условия Гюгонио лишь с первым порядком, независимо от их точности на гладких решениях.

Одним из недостатков проведенного в [22], [44], [45] анализа является то, что в нем аппроксимация рассматривается на разрывных решениях аппроксимируемой системы (2.1), в силу чего никак не учитывается процесс численного размазывания фронтов ударных волн. В то же время характер такого размазывания может оказывать существенное влияние на точность передачи условий Гюгонио (см. [46]). Поэтому в [15] был применен другой подход, при котором слабая аппроксимация определяется не на разрывных решениях $\mathbf{u}$ системы (2.1), а на сходящихся к ним численных решениях $\mathbf{v}_h$ (этот подход естественным образом согласован с основной теоремой работы [40] о слабой сходимости численных решений консервативных разностных схем). Было показано, что среди явных двухслойных по времени численных схем отсутствуют схемы повышенного порядка слабой аппроксимации, а в симметричных по времени и пространству компактных разностных схемах (см. [47]–[50]) порядки классической и слабой аппроксимации совпадают. Однако в таких компактных схемах отсутствует внутренний диссипативный механизм, и в случае аппроксимации гиперболической системы (2.1) они становятся неустойчивыми при расчете ударных волн. Поэтому для расчета разрывных решений уравнений Эйлера или резко меняющихся решений уравнений Навье–Стокса симметричную компактную схему модифицируют, вводя несимметричные компактные аппроксимации третьего порядка (см. [50]) или добавляя специальную искусственную вязкость первого порядка дивергентности (см. [49]). К сожалению, оба эти способа стабилизации симметричной компактной схемы приводят к потере основного ее достоинства – повышенного порядка слабой аппроксимации.

С учетом этого в [15], [51] был применен другой подход, при котором для обеспечения устойчивости симметричной компактной схемы в нее добавляется соответствующая искусственная вязкость повышенного порядка дивергентности. Простейшая из таких компактных схем (CWA-схема) получается из трехточечной по пространству и трехслойной по времени симметричной схемы четвертого порядка в результате добавления в нее искусственной вязкости четвертого порядка дивергентности, аналогичной (4.3). Операторная форма записи CWA-схемы (третьего порядка как классической, так и слабой аппроксимации) имеет вид (см. [15])

$$A_h \circ \Delta^\tau \circ \mathbf{v}_h + RA^\tau \circ \Delta_h \circ \mathbf{f}(\mathbf{v}_h) = C\left(\Delta_{h/2}\right)^4 \circ T^{-\tau} \circ \mathbf{v}_h, \tag{4.4}$$

где

$$\Delta_h = T_h - T_{-h}, \quad \Delta^\tau = T^\tau - T^{-\tau}, \quad A_h = T_h + 4E + T_{-h}, \quad A^\tau = T^\tau + 4E + T^{-\tau}, \tag{4.5}$$

$$\left(\Delta_{h/2}\right)^4 = \left(T_{h/2} - T_{-h/2}\right)^4 = T_{2h} - 4T_h + 6E - 4T_{-h} + T_{-2h} \tag{4.6}$$

есть пространственный разностный оператор четвертого порядка дивергентности, $T_{jh}^{n\tau}$ – оператор сдвига, действие которого на каждую функцию $\mathbf{u}(x,t)$ определяется тождеством

$$T_{jh}^{n\tau} \circ \mathbf{u}(x,t) = \mathbf{u}(x + jh, t + n\tau), \tag{4.7}$$

с учетом которого $T_{jh} = T_{jh}^0$, $T^{n\tau} = T_0^{n\tau}$, $E = T_0 = T^0 = T_0^0$; $C < 0$ – коэффициент вязкости. В настоящей работе мы будем использовать значение $C = -1/8$, при котором достигается максимальное подавление осцилляций на фронтах ударных волн.

Поскольку CWA-схема (4.4)–(4.7) является трехслойной по времени, то для получения численного решения задачи Коши (2.1), (2.2) на первом шаге по времени применяется явная схема МакКормака (см. [41]) второго порядка, что сохраняет третий порядок аппроксимации компактной схемы внутри расчетной области. Поскольку CWA-схема является неявной, то для нахождения ее численных решений на верхнем временном слое необходимо использовать трехточечные прогонки с итерациями по нелинейности. В случае аппроксимации системы законов сохранения теории мелкой воды (2.1), (3.1) эта процедура сводится (см. [15]) к двум скалярным прогонкам относительно глубины H и расхода q жидкости.

4.3. Схема CABARETM

Для численного решения гиперболических уравнений была предложена (см. [52]) трехслойная по времени и двухточечная по пространству схема Upwind Leapfrog, которая имеет второй порядок аппроксимации на гладких решениях, является явной и условно устойчивой при числах

Куранта $z \in (0,1]$. Детальный анализ этой схемы при аппроксимации линейного уравнения переноса был проведен в [53], [54], где, с учетом кососимметричности своего пространственного шаблона, она была названа схемой Кабаре. Основные достоинства этой схемы связаны с тем, что она задана на компактном пространственном шаблоне, является обратимой по времени и точной при двух различных числах Куранта $z = 0.5, 1$, что наделяет ее уникальными диссипативными и дисперсионными свойствами (см. [54]). Для численного решения квазилинейных гиперболических систем законов сохранения (2.1) были разработаны NFC-варианты схемы КАБАРЕ (см. [55]), в которых нелинейная коррекция потоков проводится на основе принципа максимума (см. [36]); для данных схем в [9] была предложена аббревиатура CABARET. Различные варианты схемы CABARET эффективно применяются для численного моделирования различных прикладных задач математической физики, в частности, пространственно многомерных газодинамических течений (см. [56]), мезомасштабных течений в океане (см. [57]) и волновых течений мелкой воды над неровным дном (см. [58]).

Монотонность схемы CABARET при аппроксимации линейного уравнения переноса в одномерном случае изучалась в [59], [60], в двумерном случае – в [61]. Условия монотонности этой схемы при аппроксимации однородного квазилинейного скалярного закона сохранения с выпуклым потоком исследовались в [62], [63], с невыпуклым потоком – в [64]. Монотонность схемы CABARET, аппроксимирующей неоднородный скалярный закон сохранения, исследовалась в [65]. Монотонность модифицированной схемы CABARETM при аппроксимации гиперболической системы законов сохранения (2.1), допускающей вектор инвариантов $\mathbf{w} = W(\mathbf{u})$, изучалась в [30].

В схеме CABARETM наряду с потоковыми переменными $\mathbf{v}_j^n = \mathbf{v}_h(x_j, t_n)$, заданными в целых пространственных узлах базисной сетки (2.3), используются консервативные переменные $\mathbf{v}_{j+1/2}^n = \mathbf{v}_h(x_{j+1/2}, t_n)$, заданные в полуцелых пространственных узлах этой сетки. В схеме CABARETM по известным значениям $\mathbf{v}_j^n$ и $\mathbf{v}_{j+1/2}^n$ определяются консервативные переменные

$$\mathbf{v}_{j+1/2}^{n+1/2} = \mathbf{v}_{j+1/2}^n - \frac{R}{2}\left(\mathbf{f}_{j+1}^n - \mathbf{f}_j^n\right), \quad \mathbf{v}_{j+1/2}^{n+1} = \mathbf{v}_{j+1/2}^n - R\left(\mathbf{f}_{j+1}^{n+1/2} - \mathbf{f}_j^{n+1/2}\right), \tag{4.8}$$

где $\mathbf{f}_j^n = \mathbf{f}(\mathbf{v}_j^n)$, $\mathbf{f}_j^{n+1/2} = \mathbf{f}(\mathbf{v}_j^{n+1/2})$, $\mathbf{v}_j^{n+1/2} = W^{-1}(\mathbf{w}_j^{n+1/2})$; W^{-1} – оператор, обратный к оператору W; $\mathbf{w}_j^{n+1/2}$ – вектор численных инвариантов, каждая i-я компонента которого $(\mathbf{w}_i)_j^{n+1/2}$ вычисляется по i-м компонентам векторов инвариантов $\mathbf{w}_j^n = W(\mathbf{v}_j^n)$ и $\mathbf{w}_{j+1/2}^n = W(\mathbf{v}_{j+1/2}^n)$ с помощью формулы

$$(w_i)_j^{n+1} = \Phi\left((w_i)_{j-1}^n, (w_i)_{j-1/2}^n, (w_i)_j^n, (w_i)_{j+1/2}^n, (w_i)_{j+1}^n\right), \tag{4.9}$$

где Φ – функция, построение которой описано в [30].

После определения консервативных переменных $\mathbf{v}_{j+1/2}^{n+1}$ находятся потоковые переменные $\mathbf{w}_j^{n+1}$ и $\mathbf{v}_j^{n+1} = W^{-1}(\mathbf{w}_j^{n+1})$, где каждая i-я компонента $(w_i)_j^{n+1}$ вектора инвариантов $\mathbf{w}_j^{n+1}$ определяется по следующим формулам, при записи которых индекс i опущен:

$$w_j^{n+1} = F\left(\tilde{w}_j^{n+1}, m_j^{n+1}, M_j^{n+1}\right), \quad \tilde{w}_j^{n+1} = 2w_j^{n+1/2} - w_j^n, \tag{4.10}$$

$$m_j^{n+1} = \min\left(w_{j-1/2}^{n+1}, w_{j+1/2}^{n+1}\right), \qquad M_j^{n+1} = \max\left(w_{j-1/2}^{n+1}, w_{j+1/2}^{n+1}\right), \tag{4.11}$$

$$F(u,m,M) = \begin{cases} m, & u \le m, \\ u, & m \le u \le M, \\ M, & u \ge M, \end{cases} \tag{4.12}$$

где $w_{j\pm1/2}^{n+1}$ суть i-е компоненты векторов инвариантов $\mathbf{w}_{j\pm1/2}^{n+1} = W(\mathbf{v}_{j\pm1/2}^{n+1})$.

Схема CABARETM (4.8)–(4.12) имеет второй порядок сходимости на гладких решениях. С учетом коррекции потоков, основанной на принципе максимума, эта схема с повышенной точностью локализует сильные разрывы, а с учетом дополнительной коррекции потоков (4.10)–(4.12) она сохраняет монотонность разностного решения относительно инвариантов линейного приближения аппроксимируемой системы (2.1).

4.4. Схема WENO5 третьего порядка по времени

WENO-схемы (см. [6], [7]) были построены в результате модификации ENO-схем из [4], которые, в свою очередь, были получены путем модификации TVD-схем из [3] с целью сохранения повышенной точности при аппроксимации локальных экстремумов в гладких частях рассчитываемых точных решений; при этом в схемах ENO и WENO свойство TVD при аппроксимации квазилинейного скалярного закона сохранения выполняется приближенно. Ключевая идея ENO-схем, связанная с выбором наиболее “гладкого” шаблона из нескольких кандидатов при аппроксимации потоков, в WENO-схемах видоизменяется следующим образом: используется выпуклая комбинация всех подходящих шаблонов, каждому из которых присваивается весовой коэффициент, определяющий его вклад в окончательную аппроксимацию численного потока.

Рассмотрим схему WENO5 третьего порядка по времени (см. [7]), при реализации которой используется глобальное расщепление потоков Лакса–Фридрихса. В этой схеме для аппроксимации по времени применяется метод Рунге–Кутты третьего порядка, в силу чего разностное решение $\mathbf{v}_j^{n+1}$ определяется по известным значениям $\mathbf{v}_j^n$ с помощью формул

$$\mathbf{v}_j^{(1)} = \mathbf{v}_j^n - R\mathbf{L}[\mathbf{v}^n]_j, \quad \mathbf{v}_j^{(2)} = \frac{1}{4}\left(3\mathbf{v}_j^n + \mathbf{v}_j^{(1)} - R\mathbf{L}[\mathbf{v}^{(1)}]_j\right), \quad \mathbf{v}_j^{n+1} = \frac{1}{3}\left(\mathbf{v}_j^n + 2\mathbf{v}_j^{(2)} - 2R\mathbf{L}[\mathbf{v}^{(2)}]_j\right), \tag{4.13}$$

где оператор $\mathbf{L}[\mathbf{v}]$ в каждом узле j вычисляется следующим образом:

$$\mathbf{L}[\mathbf{v}]_j = \hat{\mathbf{f}}_{j+1/2} - \hat{\mathbf{f}}_{j-1/2}, \quad \hat{\mathbf{f}}_{j\pm1/2} = \hat{\mathbf{f}}^+_{j\pm1/2} + \hat{\mathbf{f}}^-_{j\pm1/2}. \tag{4.14}$$

Положительная часть $\hat{\mathbf{f}}^+_{j+1/2}$ численного потока в каждом узле j вычисляется по следующим формулам, при записи которых индекс “+” для краткости опускается:

$$\hat{\mathbf{f}}_{j+1/2} = \sum_{i=1}^{m} \tilde{f}^i_{j+1/2}\mathbf{r}^i\left(\mathbf{v}_{j+1/2}\right), \quad \tilde{f}^i_{j+1/2} = \sum_{k=0}^{2} w^i_{kj} q^i_{kj}, \quad \mathbf{v}_{j+1/2} = \frac{\mathbf{v}_j + \mathbf{v}_{j+1}}{2}, \tag{4.15}$$

$$w^i_{kj} = w_k\left(f^i_{j-2}, f^i_{j-1}, f^i_j, f^i_{j+1}, f^i_{j+2}\right), \quad q^i_{kj} = q_k\left(f^i_{k+j-2}, f^i_{k+j-1}, f^i_{k+j}\right), \tag{4.16}$$

$$f^i_j = \mathbf{l}^i\left(\mathbf{v}_j\right)\left(\mathbf{f}(\mathbf{v}_j) + \gamma_i\mathbf{v}_j\right), \quad \gamma_i = \max_j|\lambda_i(\mathbf{v}_j)|, \tag{4.17}$$

где $\mathbf{r}^i$ и $\mathbf{l}^i$ – правый и левый собственные векторы матрицы Якоби системы (2.1), отвечающие собственному значению λ_i,

$$w_k\left(f_{-2}, f_{-1}, f_0, f_1, f_2\right) = \frac{\alpha_k}{\alpha_0 + \alpha_1 + \alpha_2}, \quad \alpha_k = \frac{C_k}{\left(\varepsilon + \varphi_k\right)^2}, \tag{4.18}$$
$$C_0 = 1, \quad C_1 = 6, \quad C_2 = 3, \quad \varepsilon = 10^{-9},$$

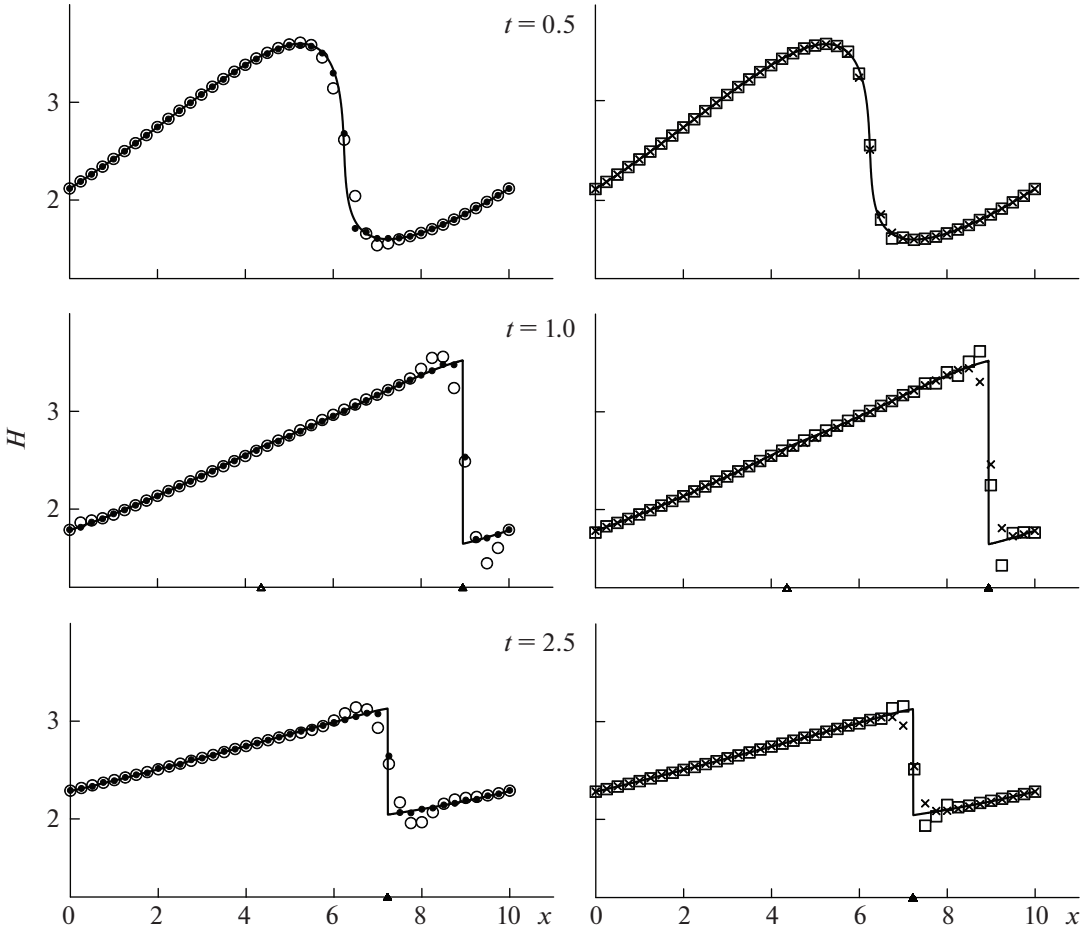
$$\varphi_0 = a\psi_{-1}^2 + b\left(f_{-2} - 4f_{-1} + 3f_0\right)^2, \quad \varphi_1 = a\psi_0^2 + b\left(f_{-1} - f_{+1}\right)^2, \quad \varphi_2 = a\psi_1^2 + b\left(3f_0 - 4f_1 + f_2\right)^2, \tag{4.19}$$

$$\psi_l = f_{l-1} - 2f_l + f_{l+1}, \quad q_k(f_0, f_1, f_2) = \frac{1}{6}(a_{k0}f_0 + a_{k1}f_1 + a_{k2}f_2), \tag{4.20}$$

$$a = 13/12, \quad b = 1/4, \quad a_{00} = a_{12} = a_{20} = 2, \quad a_{10} = a_{22} = -1,$$
$$a_{11} = a_{21} = 5, \quad a_{01} = -7, \quad a_{02} = 11. \tag{4.21}$$

Отрицательная часть $\hat{\mathbf{f}}^-_{j+1/2}$ численного потока, с учетом замены в первом равенстве (4.17) знака плюс на знак минус, вычисляется по формулам, которые симметричны формулам (4.15)–(4.21) относительно точки $x_{j+1/2} = (j + 1/2)h$.

В отличие от других классов NFC-схем, в которых для монотонизации разностного решения используются различные типы минимаксной коррекции потоков, в WENO-схемах такая коррекция достигается за счет введения весовых параметров (4.18), что позволяет сохранить повышенную гладкость функций численных потоков. В настоящее время WENO-схемы, особенно их современные модификации, широко применяются для численного моделирования различных прикладных задач газо- и гидродинамики.



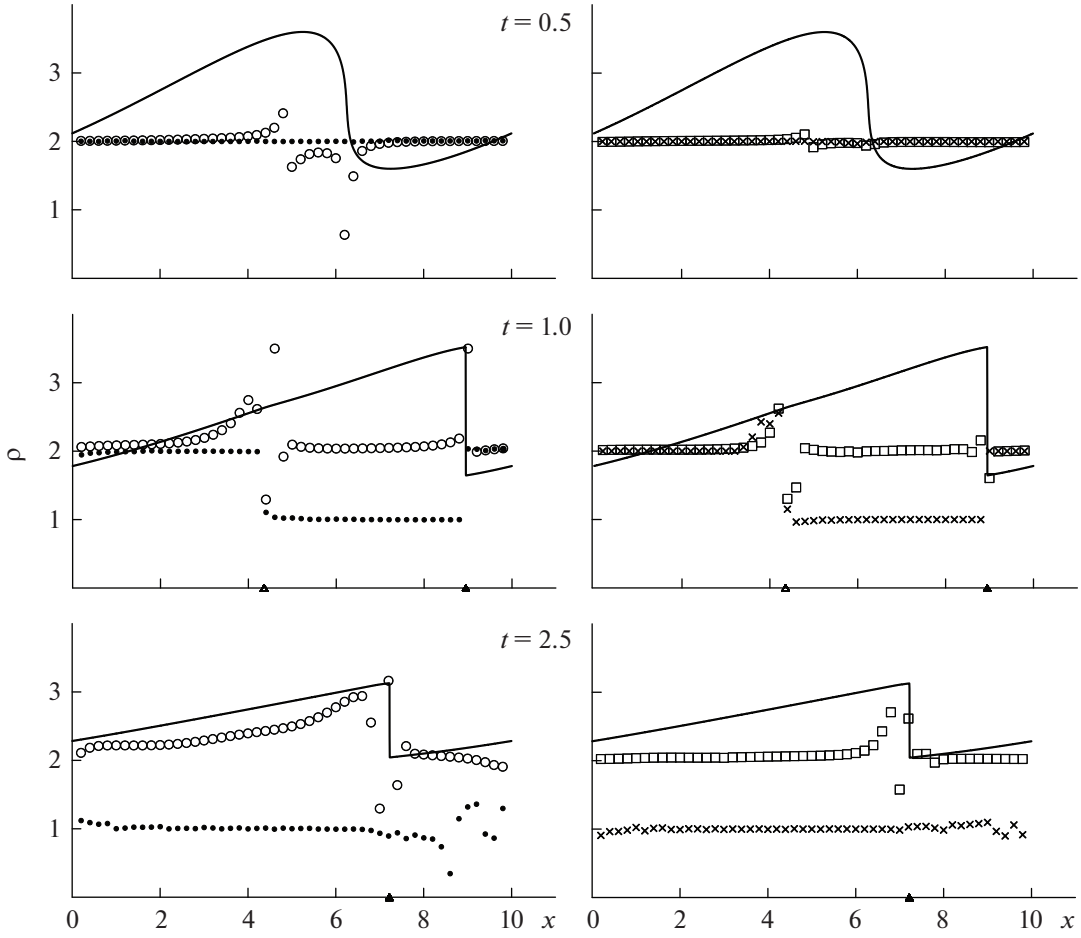
Фиг. 3. Глубины жидкости H, получаемые при численном решении задачи Коши (2.1), (3.1), (3.2) по схемам Русанова (кружки), CWA (квадратики), CABARETM (точки) и WENO5 (крестики). Сплошными линиями изображены квазиточные значения глубины.

5. РЕЗУЛЬТАТЫ РАСЧЕТА ОСНОВНОЙ ТЕСТОВОЙ ЗАДАЧИ ПО СХЕМАМ РУСАНОВА, CWA, CABARETM И WENO5

В этом разделе на три момента времени $t = 0.5, 1, 2.5$ приведены результаты численных расчетов основной тестовой задачи (2.1), (3.1), (3.2) по схемам Русанова (4.1)–(4.3), CWA (4.4)–(4.7), CABARETM (4.8)–(4.12) и WENO5 (4.13)–(4.21), проведенные на равномерной сетке (2.3) с коэффициентом запаса $z = 0.45$, что обеспечивает выполнение условия устойчивости (2.4). На фиг. 3–6 результаты расчетов по схеме Русанова показаны кружками, по CWA-схеме – квадратиками, по схеме CABARETM – точками и по схеме WENO5 – крестиками.

На фиг. 3 приведены значения глубины жидкости, получаемые при численном расчете на сетке (2.3) с пространственным шагом $h = 0.25$. Сплошными линиями на этой фигуре изображены квазиточные профили глубины, получаемые в результате численного расчета по схеме CABARETM на достаточно мелкой сетке. Из фиг. 3 следует, что, в отличие от NFC-схем CABARETM и WENO5, схемы Русанова и CWA имеют заметные осцилляции в окрестностях ударных волн. Следует также отметить, что схема CABARETM заметно меньше размазывает фронт ударной волны, чем схема WENO5 более высокого порядка аппроксимации.

На фиг. 4 показаны порядки интегральной сходимости $\rho(x_j, X, t)$, определяемые по формуле (2.13), в которой интегралы $\mathbf{V}_{h_i}$ вычисляются по формуле трапеций, а на фиг. 5 и 6 – относительные локальные дисбалансы $\Delta w_{ih}(x_j, t)$ вычисления инвариантов w_i, задаваемые формулами (2.17) и (2.18); причем в формуле (2.13) для схем Русанова, CWA и WENO5 параметр $k = 2$, а для схемы CABARETM параметр $k = 3$. Расчеты для фиг. 4–6 проводились на базисной сетке (2.3) с

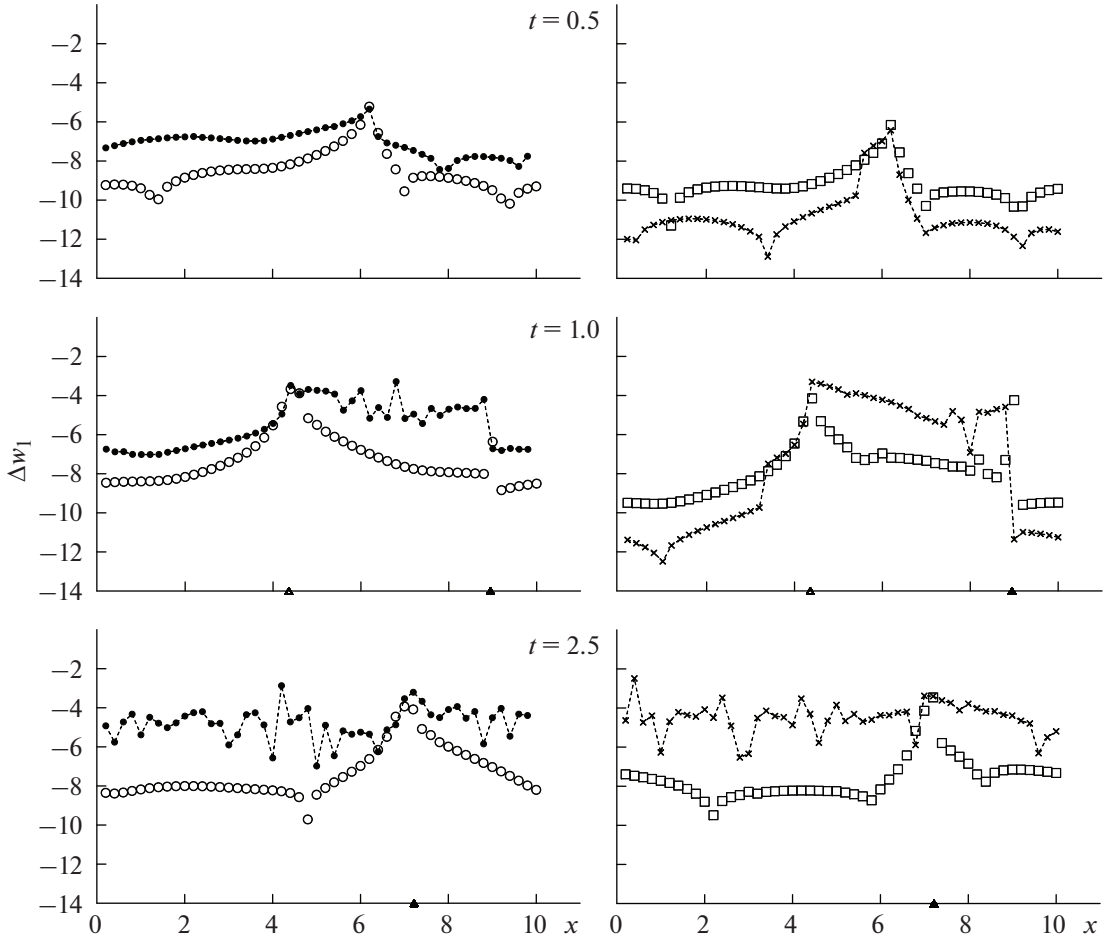


Фиг. 4. Порядки интегральной сходимости ρ, получаемые по схемам Русанова (кружки), CWA (квадратики), CABARETM (точки) и WENO5 (крестики). Сплошными линиями изображены квазиточные значения глубины.

пространственным шагом $h = 0.005$, что соответствует 2000 пространственным ячейкам сетки на отрезке $[0, X]$ длины периода; результаты этих расчетов показаны для каждого 40-го пространственного узла $j = 40i$ численной сетки. На фиг. 4–6 при $t = 1, 2.5$ темными треугольниками на оси x обозначены положения фронтов ударных волн, а при $t = 1$ светлым треугольником показана левая граница области влияния ударной волны.

Из фиг. 4 следует, что в момент времени $t = 0.5$, когда точное решение является гладким, и ударная волна еще не сформировалась, все схемы имеют приблизительно второй порядок интегральной сходимости ρ_j почти на всех отрезках $[x_j, X] \in [0, X]$. Заметные колебания порядков интегральной сходимости, получаемых по схеме Русанова в окрестности области больших градиентов точного решения, свидетельствуют о том, что эта схема более тонко реагирует на начальный этап формирования ударной волны. При $t = 1, 2.5$ из фиг. 4 следует, что в отличие от немонотонных HASIA-схем Русанова и CWA, обеспечивающих повышенный порядок интегральной сходимости, схемы CABARETM и WENO5 имеют приблизительно первый порядок интегральной сходимости на интервалах $[x_j, X]$, левая граница которых лежит в области влияния ударной волны. В момент времени $t = 2.5$, когда вся расчетная область становится областью влияния ударной волны, порядки интегральной сходимости NFC-схем CABARETM и WENO5 также снижаются приблизительно до первого порядка почти во всей расчетной области.

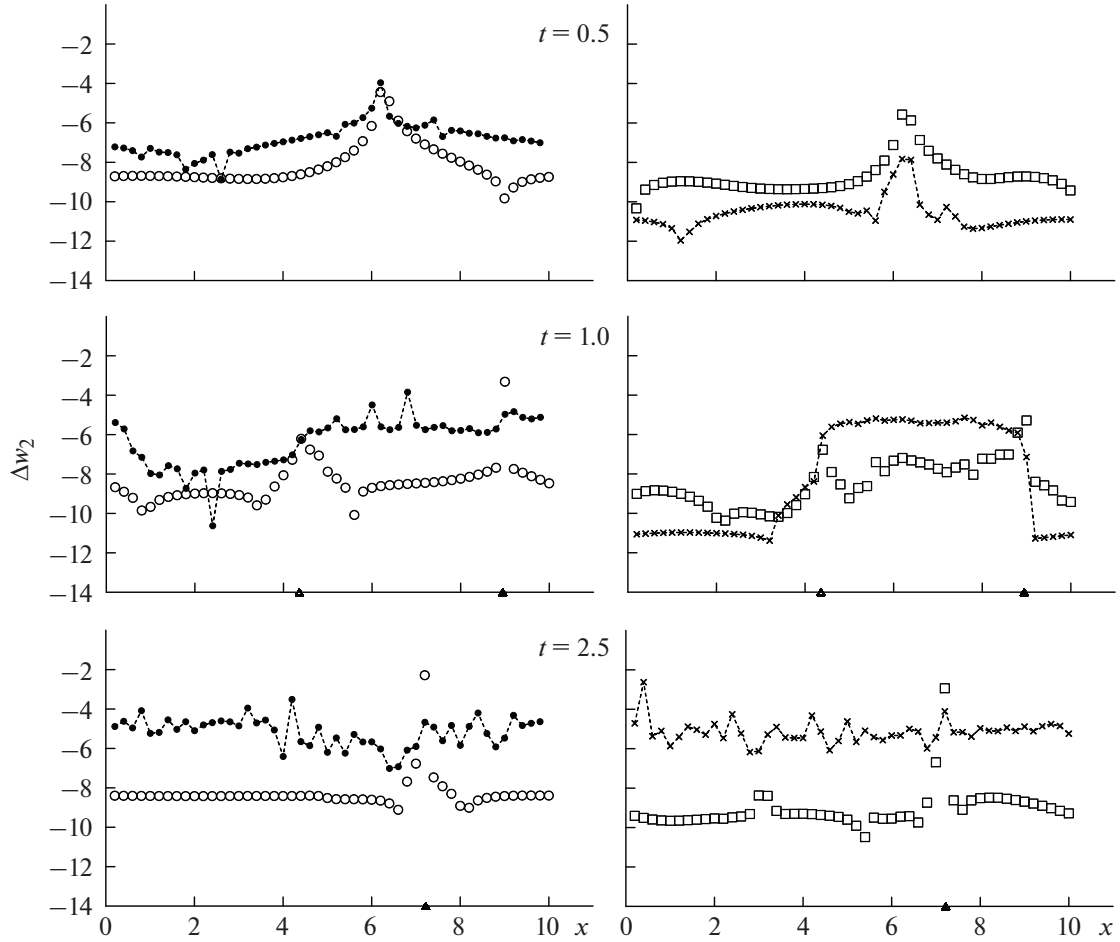
Из фиг. 5 и 6 следует, что вне области влияния ударной волны точность вычисления инвариантов во всех схемах является различной и согласуется с порядком их аппроксимации на гладких решениях. Однако в области влияния ударной волны, где порядок интегральной сходимости



Фиг. 5. Относительные локальные дисбалансы $\Delta w_{1h}(x_j,t)$ вычисления инварианта $w_1 = v - 2c$, получаемые по схемам Русанова (кружки), CWA (квадратики), CABARETM (точки) и WENO5 (крестики).

NFC-схем CABARETM и WENO5 снижается до первого, точность вычисления инвариантов по этим схемам резко падает, становится сравнимой и на несколько порядков меньшей, чем в HASIA-схемах Русанова и CWA, сохраняющих второй порядок интегральной сходимости. При этом значения функций $\Delta w_{ih}(x_j,t)$, получаемые по NFC-схемам, заметно осциллируют (у схемы CABARETM существенно больше, чем у схемы WENO5 формально более высокого порядка), что связано с мелкомасштабными колебаниями численного решения в областях влияния ударных волн, характерными для NFC-схем. Следует также отметить, что в момент времени $t = 1$, когда области влияния ударных волн еще не заполнили всю расчетную область, во всех схемах точность вычисления в этих областях инварианта $w_2 = v + 2c$, приходящего в них из гладких частей точного решения, существенно выше, чем инварианта $w_1 = v - 2c$, приходящего с фронта ударной волны и приносящего информацию о точности, с которой схема аппроксимирует условия Гюгонио. Причем максимальное снижение точности вычисления инварианта w_1 происходит не на правой границе области влияния, которая примыкает к фронту ударной волны, а в окрестности ее левой границы, где точное решение является достаточно гладким.

Из результатов численных расчетов, приведенных на фиг. 3–6, следует, что в теории численных схем сквозного счета сложилась следующая альтернатива: NFC-схемы, монотонно локализующие фронты ударных волн, теряют свою точность в областях их влияния, в то время как немонотонные HASIA-схемы, имеющие заметные нефизические осцилляции на фронтах ударных волн, в областях их влияния сохраняют повышенную точность.



Фиг. 6. Относительные локальные дисбалансы $\Delta w_{2h}(x_j,t)$ вычисления инварианта $w_2 = v + 2c$, получаемые по схемам Русанова (кружки), CWA (квадратики), CABARETM (точки) и WENO5 (крестики).

6. ГИБРИДНЫЕ СХЕМЫ

Первая попытка построения численных схем, которые монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния, была связана с применением методики построения гибридных схем (см. [24]–[28]), при которой на каждом временном слое численное решение сначала строится с помощью внешней немонотонной HASIA-схемы, имеющей заметные осцилляции на ударной волне. После этого в окрестности фронта ударной волны численное решение стандартным образом корректируется с помощью одной из NFC-схем, и на новом временном слое получается монотонизированное численное решение без заметных нефизических осцилляций. Опишем эту методику более детально при аппроксимации гибридной схемой задачи Коши (2.1), (2.2).

Предположим, что численное решение $\mathbf{v}_j^n$ гибридной схемы, определяемое в целых узлах базисной сетки (2.3), известно на первых n временных слоях этой сетки. Численное решение $\mathbf{v}_j^{n+1}$ на $(n+1)$-м временном слое находится следующим образом. По внешней немонотонной HASIA-схеме строится предварительное решение $\hat{\mathbf{v}}_j^{n+1}$ во всей однослойной расчетной области

$$S_{n+1} = \{(x_j, t_{n+1}): x_j = jh,\ t_{n+1} = t_n + \tau\}$$

на $(n+1)$-м временном слое. После этого выделяется подобласть $S_{n+1}^* \subset S_{n+1}$ больших пространственных градиентов численного решения $\hat{\mathbf{v}}_j^{n+1}$, где это решение может иметь нефизические осцилляции. В этой подобласти, которая в общем случае является многосвязной, решение $\hat{\mathbf{v}}_j^{n+1}$ за-

меняется на численное решение $\breve{\mathbf{v}}_j^{n+1}$ одношаговой по времени начально-краевой задачи, полученное по одной из NFC-схем на двухслойной сетке

$$S^*_{n,n+1} = \{(x_j,t_m)\colon m = n, n+1,\ (x_j,t_{n+1}) \in S^*_{n+1}\}$$

с начальными условиями $\mathbf{v}_j^n$, заданными на множестве

$$S^*_n = \{(x_j,t_n)\colon (x_j,t_{n+1}) \in S^*_{n+1}\}$$

и граничными условиями $\hat{\mathbf{v}}_j^{n+1}$ на границах области S^*_{n+1}. В результате решение $\mathbf{v}_j^{n+1}$ гибридной схемы определяется по формуле

$$\mathbf{v}_j^{n+1} = \begin{cases} \hat{\mathbf{v}}_j^{n+1}, & (x_j,t_{n+1}) \in S_{n+1} \backslash S^*_{n+1}, \\ \breve{\mathbf{v}}_j^{n+1}, & (x_j,t_{n+1}) \in S^*_{n+1}. \end{cases}$$

Из предыдущего раздела, в котором приведены результаты численного решения задачи Коши (2.1), (3.1), (3.2) для системы уравнений мелкой воды, следует, что схема CABARETM, имеющая компактный шаблон, с наибольшей точностью локализует фронты ударных волн (фиг. 3), в то время как HASIA-схемы Русанова и CWA, второго порядка интегральной сходимости (фиг. 4), имеют существенно более высокую точность в областях влияния ударных волн (фиг. 5, 6) по сравнению с NFC-схемами CABARETM и WENO5. С учетом этого возникает идея построения гибридных схем, в которых внешней схемой является одна из HASIA-схем Русанова или CWA, а в качестве внутренней NFC-схемы применяется схема CABARETM; далее для этих гибридных схем будем использовать аббревиатуры HR-схема (Hybrid Rusanov scheme) и HC-схема (Hybrid Compact scheme). В схемах HR и HC подобласть S^*_{n+1}, в которой применяется схема CABARETM, выделяется простейшим градиентным методом, предложенным в [29], [31]. В рамках этого метода

$$S^*_{n+1} = \{(x_j,t_{n+1})\colon j_n - m \le j \le j_n + m + 1\}, \tag{6.1}$$

где значения узлов j_n определяются из условия

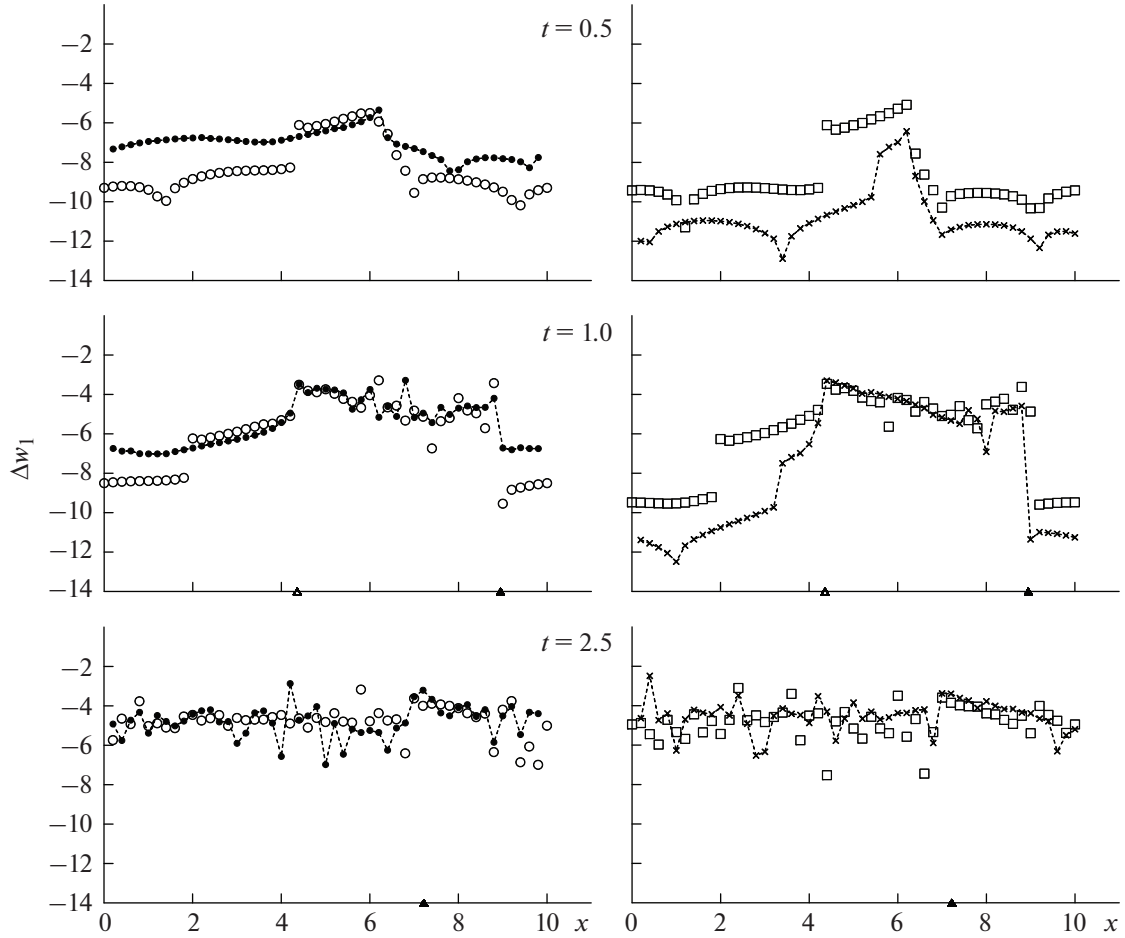
$$\left|\tilde{H}^{n+1}_{j_n+1/2}\right| = \max_j \left|\tilde{H}^{n+1}_{j+1/2}\right| \ge p, \quad \tilde{H}^{n+1}_{j+1/2} = \frac{H^{n+1}_{j+1} - H^{n+1}_j}{h}, \tag{6.2}$$

m и p – входные парамеры; в приводимых далее расчетах $m = 6$ и $p = 1.5$.

На фиг. 7, 8 показаны относительные локальные дисбалансы $\Delta w_{ih}(x_j,t)$ вычисления инвариантов w_i, полученные при расчете основной тестовой задачи (2.1), (3.1), (3.2) по схемам HR, HC, CABARETM и WENO5 на равномерной сетке (2.3) с коэффициентом запаса $z = 0.45$. Эти дисбалансы определялись по формулам (2.17) и (2.18), в которых $h_i = h/k^{i-1}$, где $k = 3$ для схем HR, HC, CABARETM и $k = 2$ для схемы WENO5. Расчеты для фиг. 7, 8 проводились на базисной сетке (2.3) с пространственным шагом $h = 0.005$, и их результаты показаны для каждого 40-го пространственного узла $j = 40i$ численной сетки.

Из результатов расчетов, приведенных на фиг. 7, 8, следует, что гибридные схемы HR и HC теряют основное преимущество исходных HASIA-схем Русанова и CWA – повышенный порядок интегральной сходимости на интервалах $[x_j, X]$, левая граница которых лежит в области влияния ударной волны. Это приводит к снижению точности схем HR и HC в областях влияния ударных волн, которая становится сравнимой с точностью NFC-схем CABARETM и WENO5 и существенно более низкой, чем в схемах Русанова и CWA. Поскольку алгоритм гибридизации (6.1), (6.2) в схемах HR и HC начинает работать в момент времени $t \approx 0.3$, т.е. раньше, чем возникают ударные волны в точном решении, то снижение точности этих схем происходит в областях влияния больших градиентов численного решения, которые содержат внутри себя и заметно превосходят области влияния ударных волн.

Далее будет изложена методика построения принципиально новых численных методов сквозного счета (получивших название комбинированные схемы), которые сочетают достоинства как NFC-схем, так и HASIA-схем, а именно, монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния.

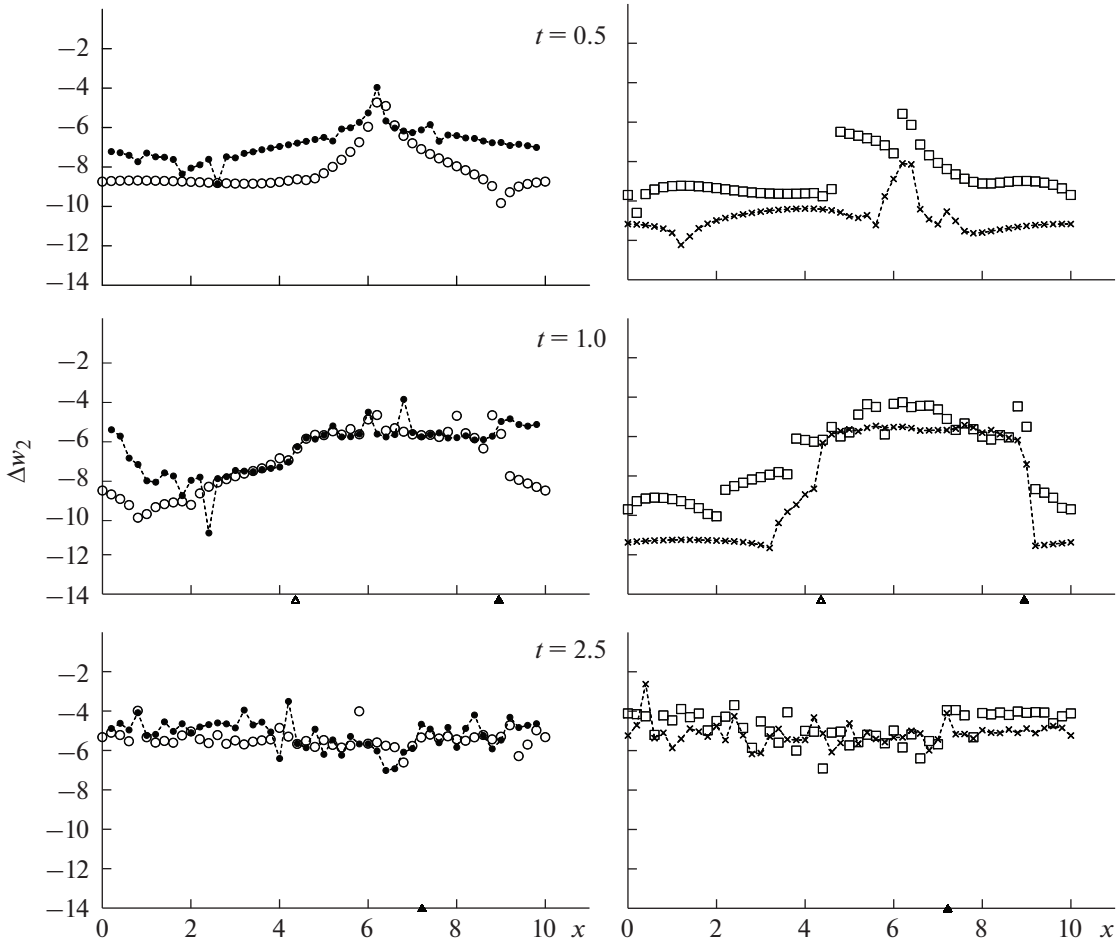


Фиг. 7. Относительные локальные дисбалансы $\Delta w_{1h}(x_j,t)$ вычисления инварианта $w_1 = v - 2c$, получаемые по схемам HR (кружки), HC (квадратики), CABARETM (точки) и WENO5 (крестики).

7. ПЕРВЫЕ КОМБИНИРОВАННЫЕ СХЕМЫ

Из результатов предыдущего раздела следует, что гибридные схемы, в которых в качестве внешней применяется одна из HASIA-схем, не сохраняют основного преимущества этих HASIA-схем – повышенную точность в областях влияния ударных волн. Данный недостаток как NFC-схем, так и гибридных схем, непосредственно связан с их главным преимуществом – монотонной локализацией фронтов ударных волн, поскольку любая конечная сумма ряда Фурье для разрывной функции не является монотонной. С учетом этого осцилляции, возникающие на фронтах ударных волн в немонотонных HASIA-схемах, несут информацию о волновой структуре фурье-разложения разрывной функции в окрестности сильного разрыва, что позволяет этим схемам с повышенной точностью передавать условия Гюгонио и, как следствие, сохранять повышенную точность в областях влияния ударных волн. NFC-схемы и гибридные схемы в результате искусственного сглаживания численных ударных волн эту информацию теряют, что приводит к снижению их точности при аппроксимации условий Гюгонио.

В [29], [31] была предложена методика построения комбинированных схем, которые сочетают достоинства как NFC-схем, так и немонотонных HASIA-схем, а именно, монотонно локализуют фронты ударных волн и одновременно сохраняют повышенную точность в областях их влияния. В комбинированной схеме применяется базисная немонотонная HASIA-схема, по которой численное решение $\hat{\mathbf{v}}_j^n$ задачи Коши (2.1), (2.2) строится во всей расчетной области S, задаваемой формулой (2.3). В подобласти $S^* \subset S$, где решение $\hat{\mathbf{v}}_j^n$ имеет большие градиенты, приводящие к нефизическим осцилляциям, оно корректируется путем построения численного решения $\breve{v}_j^n$



Фиг. 8. Относительные локальные дисбалансы $\Delta w_{2h}(x_j,t)$ вычисления инварианта $w_2 = v + 2c$, получаемые по схемам HR (кружки), HC (квадратики), CABARETM (точки) и WENO5 (крестики).

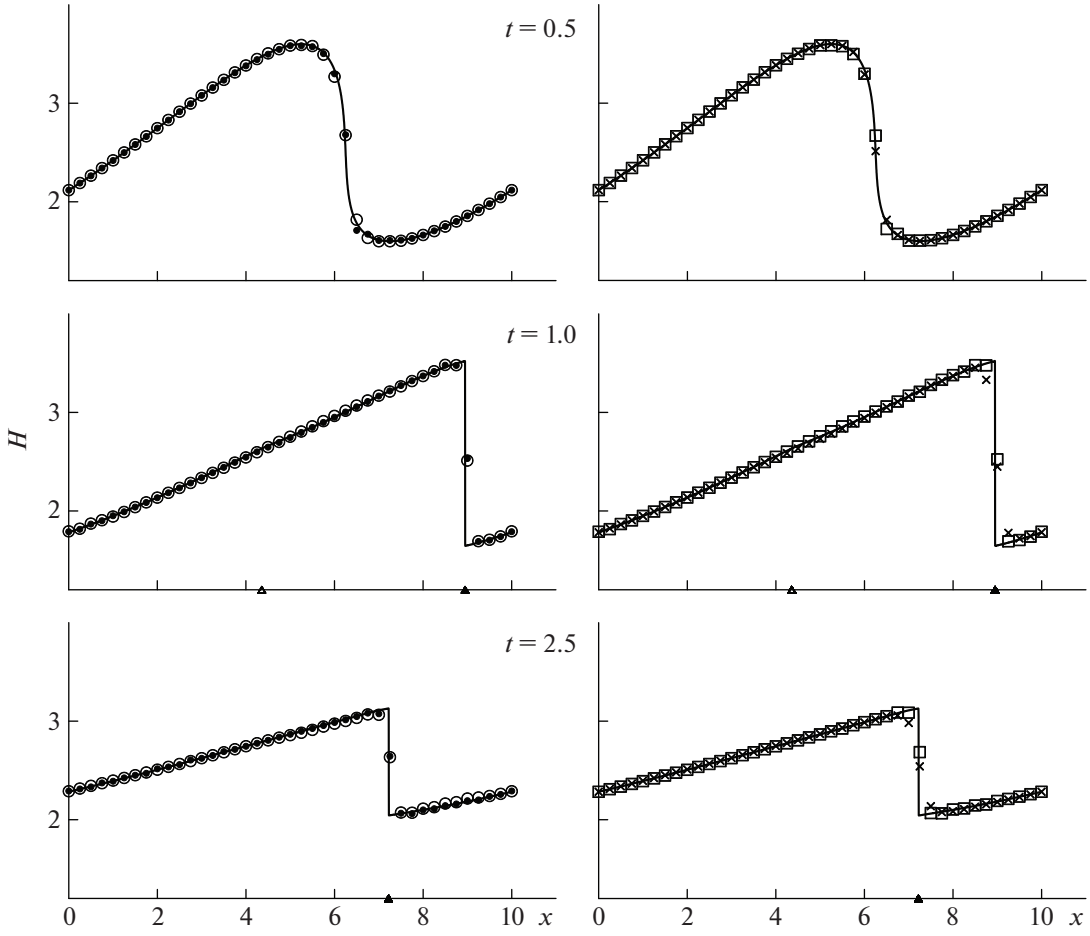
внутренней начально-краевой задачи по одной из NFC-схем с начальными и граничными условиями, получаемыми на границе области $S\backslash S^*$ из решения $\hat{\mathbf{v}}_j^n$. В результате решение комбинированной схемы в расчетной области (2.3) определяется по формуле

$$\mathbf{v}_j^n = \begin{cases} \hat{\mathbf{v}}_j^n, & (x_j,t_n) \in S\backslash S^*, \\ \check{\mathbf{v}}_j^n, & (x_j,t_{n+1}) \in S^*. \end{cases}$$

Принципиальное отличие комбинированной схемы от гибридной заключается в том, что внутренняя NFC-схема не влияет на решение, получаемое по базисной HASIA-схеме в области $S\backslash S^*$, что позволяет комбинированной схеме сохранять повышенную точность в областях влияния ударных волн. Далее для комбинированных схем, в которых базисной является схема Русанова или CWA-схема, а в качестве внутренней применяется схема CABARETM, будем использовать аббревиатуры CR-схема (Combined Rusanov scheme) и CCWA-схема (Combined Compact Weak Approximation scheme). При численном решении по комбинированным схемам CR и CCWA задачи Коши (2.1), (3.1), (3.2) для системы уравнений мелкой воды внутренняя область S^*, в которой применяется схема CABARETM, выделяется следующим образом:

$$S^* = \{(x_j,t_n) \in S\colon j_n - m \le j \le j_n + m + 1\}, \tag{7.1}$$

где значения узлов j_n определяются из условия (6.2); в приводимых далее расчетах $m = 6$ и $p = 1.5$.



Фиг. 9. Глубины жидкости H, получаемые при численном решении задачи Коши (2.1), (3.1), (3.2) по схемам CR (кружки), CCWA (квадратики), CABARETM (точки) и WENO5 (крестики). Сплошными линиями изображены квазиточные значения глубины.

На фиг. 9 приведены значения глубины жидкости, получаемые при численном расчете задачи Коши (2.1), (3.1), (3.2) по схемам CR (кружки), CCWA (квадратики), CABARETM (точки) и WENO5 (крестики) на сетке (2.3) с пространственным шагом $h = 0.25$. Из фиг. 9 следует, что в комбинированных схемах CR и CCWA эффективно подавляются нефизические осцилляции, присущие базисным HASIA-схемам Русанова и CWA на фронтах ударных волн (фиг. 3). Поскольку внутренней схемой для данных комбинированных схем является схема CABARETM, с высокой точностью локализующая ударные волны, то схемы CR и CCWA (подобно схеме CABARETM) заметно меньше размазывают фронт ударной волны по сравнению со схемой WENO5 более высокого порядка аппроксимации. Так как внутренняя схема CABARETM не влияет на решения базисных схем Русанова и CWA в области $S \backslash S^*$, то комбинированные схемы CR и CCWA (в отличие от гибридных схем HR и HCWA) сохраняют повышенную точность вычисления инвариантов в областях влияния ударных волн, которая вне области (7.1) совпадает с точностью вычисления инвариантов по схемам Русанова и CWA (фиг. 5, 6), имеющих повышенный порядок интегральной сходимости (фиг. 4).

8. СОГЛАСОВАННЫЕ КОМБИНИРОВАННЫЕ СХЕМЫ

Основной недостаток комбинированных схем CR и CCWA, построенных в предыдущем разделе, заключается в том, что соответствующие им базисная и внутренняя схемы имеют существенно различный тип, что приводит к определенным сложностям при реализации численного алгоритма на границе внутренней расчетной области (7.1). Поэтому следующий этап в развитии теории комбинированных схем был связан с разработкой согласованных численных алгоритмов,

в которых базисная и внутренняя схемы являются схемами одного класса, а именно, внутренняя схема получается из базисной схемы в результате применения соответствующей NFC-процедуры. В [32] такая комбинированная схема была построена на основе DG-метода из [8], а в [33] – на основе бикомпактных разностных схем (см. [10]).

8.1. Комбинированная схема DG-метода

В отличие от описанных в разд. 4 конечно-разностных схем, DG-схема представляет собой (см. [8]) проекционно-разностный метод (проекционный по пространственной переменной x и разностный по временной переменной t), в котором численное решение ищется в виде кусочно-полиномиальной разрывной функции относительно пространственной переменной x. Впервые предложенный в [66] и детально изученный в [8], DG-метод в настоящее время активно развивается (см. [67]–[70]) и применяется для решения сложных многомасштабных задач математической физики (см. [71]–[75]). Одним из основных достоинств данного метода является компактность пространственного шаблона, что позволяет обеспечить повышенный порядок аппроксимации на многомерных неструктурированных сетках с произвольной формой ячеек (см. [71]–[77]). Для монотонизации численного решения в окрестностях сильных разрывов в DG-схемах применяются различные методы коррекции потоков (см. [68], [73], [76], [78], [79]); наиболее распространенными из них являются NFC-лимитеры (см. [8], [70], [76], [77], [80], [81], которые, однако, могут приводить к снижению точности получаемого численного решения (см. [18]). При этом DG-метод показал (см. [82]) заметные преимущества по сравнению с MUSCL-схемами из [2] при численном расчете различных тестовых задач как на декартовых сетках, так и на неподвижных и движущихся сетках Вороного.

Поскольку применение DG-схемы для расчета многомерных прикладных задач требует выполнения большего числа достаточно громоздких вычислений, то эффективное использование этой схемы связано с применением всех возможностей современной вычислительной техники (см. [83], [84]), что диктует необходимость создания программных комплексов, достаточно легко адаптируемых для работы на различных (в том числе гибридных) параллельных компьютерных архитектурах. С учетом этого при решении DG-методом многомерных уравнений Навье–Стокса был разработан новый сеточно-операторный подход к программированию задач математической физики (см. [74]), позволяющий единообразно компактно записывать и эффективно вычислять сложные математические формулы на разных типах сеток и для различных вычислительных архитектур, в том числе для графических ускорителей CUDA (см. [84], [85]).

Далее в этом разделе приводятся результаты работы [32], в которой исследуется точность различных модификаций пространственно-одномерной DG-схемы, заданной на равномерной численной сетке (2.3). Рассмотрим сначала проекционно-дифференциальную форму записи DG-схемы (проекционную по x и дифференциальную по t), в рамках которой численное решение $\mathbf{v}_h(x,t)$ в каждой пространственной ячейке $[x_j, x_{j+1})$ сетки (2.3) представляет собой полином степени не выше p относительно переменной x:

$$\mathbf{v}_h(x,t) = \mathbf{v}_{hj}(x,t) = \sum_{i=0}^{p} \mathbf{v}_{ji}(t)\varphi_{hji}(x), \quad x \in [x_j, x_{j+1}), \tag{8.1}$$

где $\mathbf{v}_{ji}(t)$ – искомые вектор-функции,

$$\varphi_{hji}(x) = \left(\frac{x - x_{j+1/2}}{h}\right)^i \tag{8.2}$$

есть базисные многочлены. В приводимых далее формулах индекс h у функций $\mathbf{v}_{hj}$ и φ_{hji} для краткости опускается.

При аппроксимации DG-схемой гиперболической системы (2.1) каждая функция $\mathbf{v}_j(x,t)$ удовлетворяет векторному дифференциальному уравнению (см. [8])

$$\frac{d}{dt}\int_{x_j}^{x_{j+1}} \mathbf{v}_j \varphi_{jk}\, dx = \mathbf{\Psi}_{jk}, \tag{8.3}$$

в котором

$$\mathbf{\Psi}_{jk} = \int\limits_{x_j}^{x_{j+1}} \mathbf{f}(\mathbf{v}_j)\varphi'_{jk}dx - \mathbf{F}_{j+1}\varphi_{jk}(x_{j+1}) + \mathbf{F}_j\varphi_{jk}(x_j), \tag{8.4}$$

где

$$\mathbf{F}_j = \frac{\mathbf{f}(\mathbf{v}_{j-1}(x_j,t)) + \mathbf{f}(\mathbf{v}_j(x_j,t))}{2} - \frac{a_j\left(\mathbf{v}_j(x_j,t) - \mathbf{v}_{j-1}(x_j,t)\right)}{2} \tag{8.5}$$

есть численные потоки, определяемые по формуле Русанова–Лакса–Фридрихса,

$$a_j = \max_m \max\left(|\lambda_m(\mathbf{v}_{j-1}(x_j,t))|, |\lambda_m(\mathbf{v}_j(x_j,t))|\right),$$

λ_m – собственные значения матрицы Якоби $\mathbf{f}_\mathbf{u}$ системы (2.1). С учетом формулы (8.1)

$$\frac{d}{dt}\int\limits_{x_j}^{x_{j+1}} \mathbf{v}_j\varphi_{jk}dx = \sum_{i=0}^{p} a_{ki}^j \frac{d\mathbf{v}_{ji}}{dt}, \quad a_{ki}^j = \int\limits_{x_j}^{x_{j+1}} \varphi_{jk}\varphi_{ji}dx,$$

где матрицы $A_j = \left(a_{ki}^j\right)$ являются невырожденными, векторное уравнение (8.3) эквивалентно следующей системе векторных уравнений:

$$\frac{d\mathbf{v}_{ji}}{dt} = \sum_{k=0}^{p} b_{ik}^j \mathbf{\Psi}_{jk}, \quad i = \overline{0,p}, \tag{8.6}$$

в которой $B_j = \left(b_{ik}^j\right)$ – матрица, обратная к A_j.

Следуя [32], мы будем рассматривать DG-метод (8.1)–(8.6), для которого $p = 1$, в силу чего численное решение (8.1) ищется в виде кусочно-линейной по x вектор-функции

$$\mathbf{v}_h(x,t) = \mathbf{v}_{j0}(t) + \frac{x - x_{j+1/2}}{h}\,\mathbf{v}_{j1}(t)\,, \qquad x \in [x_j, x_{j+1}). \tag{8.7}$$

Для такого DG-метода будем использовать аббревиатуру DG1. Поскольку DG1-метод (8.1)–(8.7) имеет не менее, чем второй порядок точности на гладких решениях, и не принадлежит классу NFC-схем, то получаемое на его основе численное решение может иметь заметные нефизические осцилляции на фронтах ударных волн. Для подавления этих осцилляций применим ограничитель Кокбурна (см. [8]), который каждую компоненту v_{j1} векторной функции $\mathbf{v}_{j1}$, входящей в формулу (8.7), заменяет на величину

$$w_{j1} = M\left(v_{j1}, \alpha(v_{j+1,0} - v_{j0}), \alpha(v_{j0} - v_{j-1,0})\right), \tag{8.8}$$

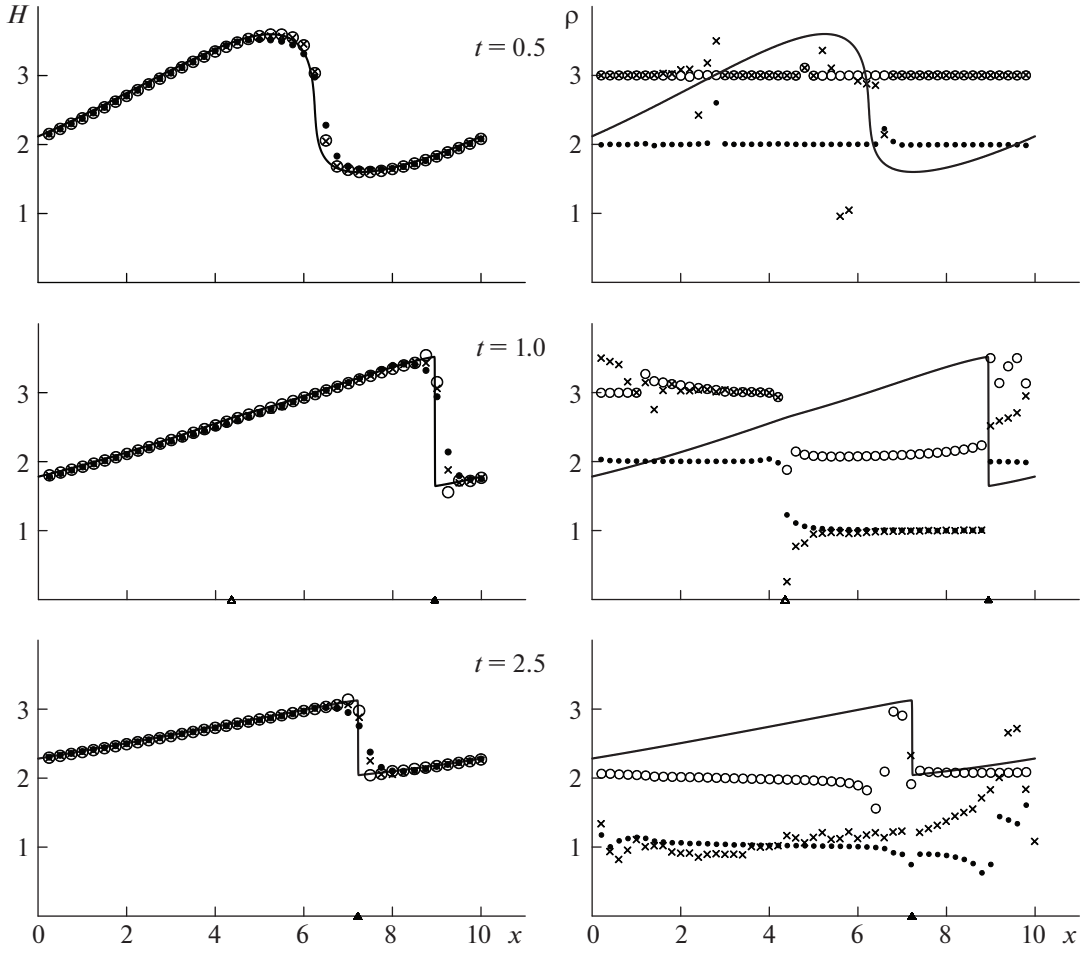
где $\alpha \in [1,2]$ – эвристический параметр, v_{l0} – соответствующие компоненты векторов $\mathbf{v}_{l0}$,

$$M(v_1, v_2, v_3) = s\min\left(|v_1|, |v_2|, |v_3|\right) \tag{8.9}$$

есть стандартный оператор minmod, в котором $s = \mathrm{sign}(v_i)$ при условии, что все числа v_i, входящие в соотношение (8.9), имеют одинаковый знак, и $s = 0$, если это условие не выполнено. Для DG1-методов, коррекция которых по формуле (8.8) происходит при параметрах $\alpha = 1$ и $\alpha = 2$, будем использовать обозначения DG1A1 и DG1A2 соответственно.

Для DG-схемы (8.1)–(8.9) полудискретное численное решение $\mathbf{v}_h(x, n\tau)$ получается путем решения системы обыкновенных дифференциальных уравнений (8.6) методом Рунге–Кутты третьего порядка по формулам, аналогичным (4.13). При этом тестовые расчеты показывают, что увеличение параметра α, входящего в оператор коррекции (8.8), приводит к уменьшению схемной вязкости, что может вызвать возникновение небольших осцилляций на фронтах ударных волн. В то же время уменьшение параметра α, связанное с увеличением схемной вязкости, будет приводить к более сильному размазыванию фронтов ударных волн. С учетом этого метод DG1A1 полностью подавляет численные осцилляции, возникающие на фронтах ударных волн в DG1-методе, в то время как метод DG1A2 подавляет эти осцилляции лишь частично.

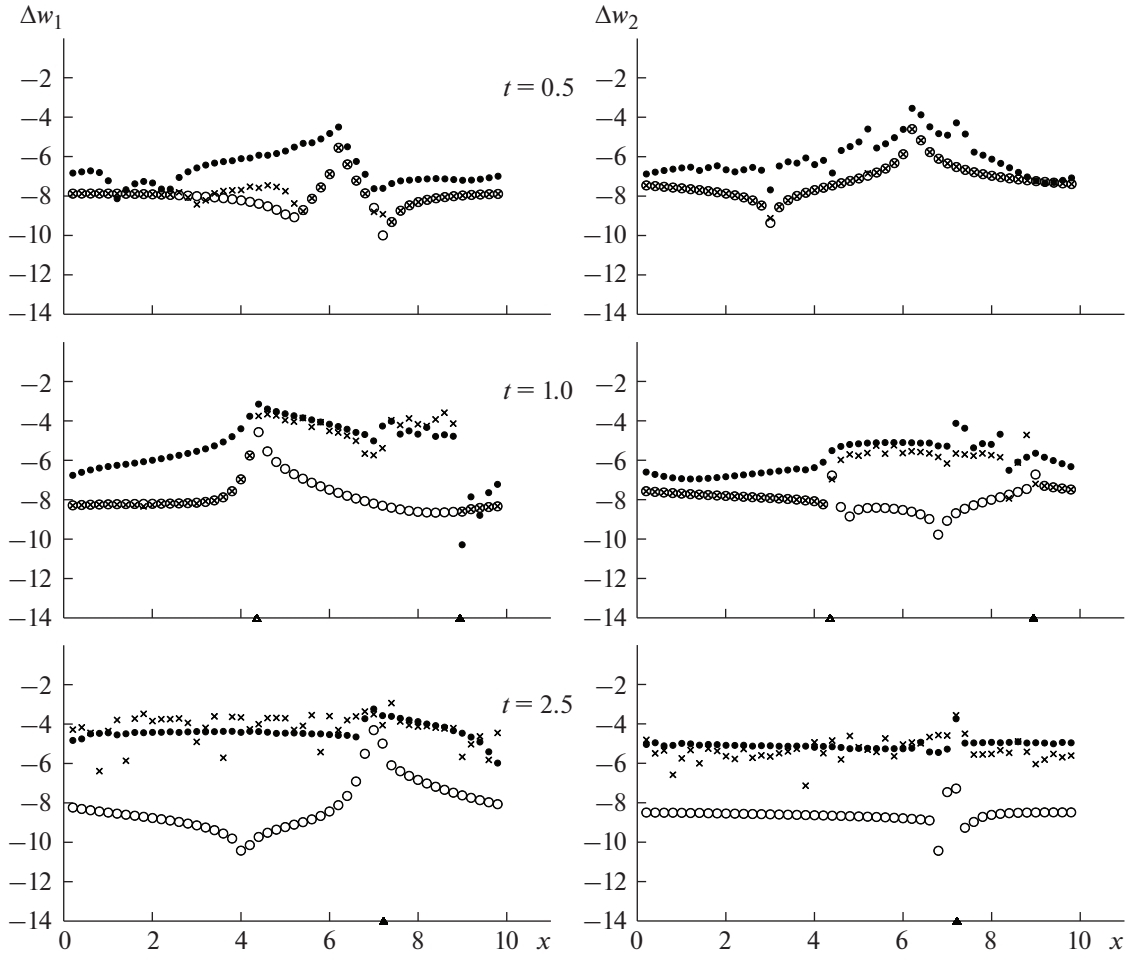
На фиг. 10–12 приведены результаты численных расчетов основной тестовой задачи (2,1), (3.1), (3.2) по трем DG-схемам: DG1, DG1A1 и DG1A2, проведенные на равномерной сетке (2.3)



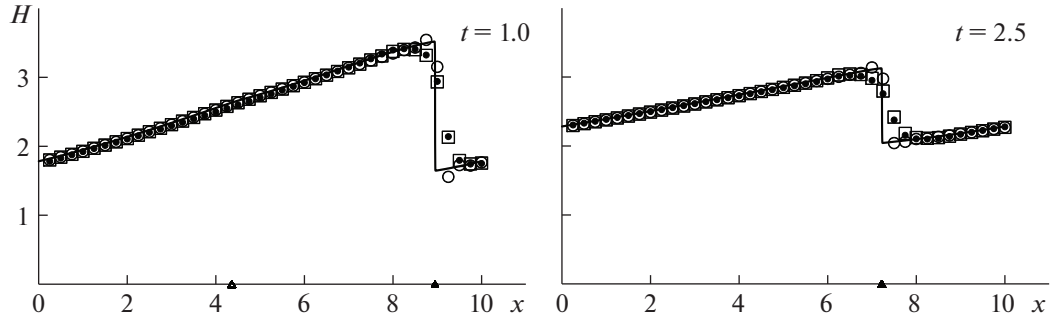
Фиг. 10. Глубины жидкости H и порядки интегральной сходимости ρ, получаемые при численном решении задачи Коши (2.1), (3.1), (3.2) по схемам DG1 (кружки), DG1A1 (точки) и DG1A2 (крестики). Сплошными линиями изображены квазиточные значения глубины.

с коэффициентом запаса $z = 0.45$, что обеспечивает выполнение условия устойчивости (2.4); результаты расчетов по схеме DG1 показаны кружками, по схеме DG1A1 – точками и по схеме DG1A2 – крестиками. При $t = 1{,}2.5$ темными треугольниками на оси x обозначены положения фронтов ударных волн, а при $t = 1$ светлым треугольником показана левая граница области влияния ударной волны. Сплошными линиями на фиг. 10 и 12 изображены квазиточные профили глубины, получаемые в результате численного расчета по схеме DG1A1 на достаточно мелкой сетке. На фиг. 10 показаны численные значения глубины жидкости $H(x_{j+1/2}, t)$ и порядки интегральной сходимости $\rho(x_j, X, t)$, определяемые по формуле (2.13), в которой $k = 3$, а на фиг. 11 – относительные локальные дисбалансы $\Delta w_{ih}(x_{j+1/2}, t)$, вычисления инвариантов w_i, задаваемые формулами (2.17) и (2.18). Глубины жидкости H вычислялись на сетке (2.3) с пространственным шагом $h = 0.25$, а величины ρ и Δw_{ih} были вычислены на базисной сетке (2.3) с шагом $h = 0.005$ и показаны для каждой 40-й ячейки этой сетки.

На графиках, приведенных слева на фиг. 10, видно, что в отличие от NFC-схем DG1A1 и DG1A2, схема DG1 имеет заметные осцилляции в окрестностях ударных волн; при этом схемы DG1 и DG1A2 заметно меньше размазывают фронт ударной волны, чем схема DG1A1. Из графиков, показанных справа на фиг. 10, следует, что немонотонная схема DG1, несмотря на нефизические осцилляции на фронте ударной волны, обеспечивает второй порядок интегральной сходимости на отрезках $[x_j, X]$, левая граница которых расположена внутри области влияния ударной волны. Схемы DG1A1 и DG1A2, полученные путем монотонизации схемы DG1, подоб-



Фиг. 11. Относительные локальные дисбалансы $\Delta w_{ih}(x_{j+1/2}, t)$ вычисления инвариантов w_i, получаемые по схемам DG1 (кружки), DG1A1 (точки) и DG1A2 (крестики).



Фиг. 12. Глубины жидкости H, получаемые при численном решении задачи Коши (2.1), (3.1), (3.2) по схемам DG1 (кружки), DG1A1 (точки) и CDG (квадратики). Сплошными линиями изображены квазиточные значения глубины.

но конечно-разностным NFC-схемам, снижают скорость этой сходимости приблизительно до первого порядка, что приводит к заметному снижению их точности по сравнению со схемой DG1 при вычислении инвариантов в области влияния ударной волны (фиг. 11).

На графиках, приведенных справа на фиг. 10, видно, что при $t = 0.5, 1$ схемы DG1 и DG1A2 имеют третий порядок интегральной сходимости на интервалах $[x_j, X]$, левая граница которых

лежит вне области влияния ударной волны, в то время как схема DG1A1 имеет на этих интервалах второй порядок интегральной сходимости. Объясняется это тем, что схема DG1 обеспечивает третий порядок локальной аппроксимации на гладких решениях и коррекция потоков в схеме DG1A2 этот порядок сохраняет, в то время как более сильная коррекция потоков в схеме DG1A1 снижает точность этой аппроксимации до второго порядка. В результате вне области влияния ударной волны инварианты вычисляются по схеме DG1A1 с существенно более низкой точностью, чем по схемамам DG1 и DG1A2 (фиг. 11).

Проведенный анализ точности схем DG1 и DG1A1 позволяет использовать их для построения методом, изложенным в разд. 7, комбинированной схемы разрывного метода Галеркина, в которой в качестве базисной HASIA-схемы применяется немонотонная схема DG1, а в качестве внутренней NFC-схемы – схема DG1A1. Для получаемой таким образом комбинированной DG-схемы будем использовать аббревиатуру CDG-схема (Combined Discontinuous Galerkin scheme). На фиг. 12 при $t = 1, 2.5$ приведены значения глубины жидкости $H(x_{j+1/2}, t)$, получаемые при численном расчете задачи Коши (2.1), (3.1), (3.2) по схемам CDG (квадратики), DG1 (кружки) и DG1A1 (точки) на сетке (2.3) с пространственным шагом $h = 0.25$. На фиг. 12 видно, что в комбинированной CDG-схеме эффективно подавляются нефизические осцилляции на фронтах ударных волн, присущие базисной DG1-схеме (фиг. 10). Поскольку внутренняя схема DG1A1 не влияет на решение базисной схемы DG1 вне многосвязной области (7.1), внутри которой расположены ударные волны, то комбинированная CDG-схема сохраняет повышенную точность вычисления инвариантов в областях влияния ударных волн, которая совпадает с точностью вычисления инвариантов по схеме DG1 (фиг. 11), имеющей повышенный порядок интегральной сходимости (фиг. 10).

8.2. Бикомпактные комбинированные схемы

Определяющей чертой бикомпактных схем является высокоточная компактная аппроксимация пространственных производных на шаблоне, который размещается в одной ячейке сетки. По каждому пространственному направлению этот шаблон включает в себя лишь два целых узла сетки, что и дало название этому классу схем. Высокий четный порядок аппроксимации по пространству в бикомпактных схемах достигается за счет введения дополнительных искомых сеточных функций, определенных на множестве либо уже имеющихся целых, либо вспомогательных дробных узлов. Для отыскания этих дополнительных функций привлекаются дифференциальные следствия аппроксимируемых уравнений. Преимущество бикомпактных схем заключается в сочетании нескольких положительных свойств: это слабые ограничения по устойчивости (благодаря неявности временной дискретизации), экономичная реализация, совпадение числа граничных условий в разностной и дифференциальной постановках задачи (благодаря минимальности шаблона), высокое спектральное разрешение (см. [86], [87]).

Применительно к нестационарным уравнениям в частных производных бикомпактные схемы были впервые построены: для одномерного линейного уравнения теплопроводности в [88], для одномерного линейного уравнения переноса в [10], для одномерных гиперболических систем законов сохранения в [89]. В дальнейшем бикомпактные схемы получили обобщение на различные многомерные уравнения и системы уравнений: гиперболические уравнения (см. [90]) и системы (см. [28], [91]), уравнения Эйлера для многокомпонентных химически реагирующих газов (см. [92]), линейное уравнение конвекции-диффузии (см. [93]), уравнения Навье–Стокса для несжимаемой жидкости (см. [94]). Для эффективной реализации многомерных бикомпактных схем было предложено применять локально-одномерное расщепление (см. [95]) и метод итерируемой приближенной факторизации (см. [96]). В работах [86], [87] были исследованы диссипативные и дисперсионные свойства бикомпактных схем. Было показано в [86], что бикомпактные схемы имеют лучшее спектральное разрешение по сравнению с классическими компактными схемами такого же порядка аппроксимации по пространству. В [11] был предложен метод консервативной монотонизации бикомпактных схем, в котором устранен главный недостаток метода гибридной схемы (см. [10], [28], [89]–[91]) – нарушение свойства консервативности. Вопросы точности бикомпактных схем в областях влияния ударных волн и построения комбинированных бикомпактных схем рассматривались в [21], [33], [97]. В настоящем разделе дается обзор результатов из [21], [33].

Простейшая бикомпактная схема, аппроксимирующая систему (2.1), имеет вид

$$\begin{aligned}&\frac{1}{\tau}A_0^x(\mathbf{v}^*_{j+1/2}-\mathbf{v}^n_{j+1/2})+\Lambda_1^x\mathbf{f}^+(\mathbf{v}^*_{j+1/2})=\mathbf{0},\quad \frac{1}{\tau}A_0^x(\mathbf{v}^{n+1}_{j+1/2}-\mathbf{v}^*_{j+1/2})+\Lambda_1^x\mathbf{f}^-(\mathbf{v}^{n+1}_{j+1/2})=\mathbf{0},\\ &\frac{1}{\tau}\Lambda_1^x(\mathbf{v}^{n+1}_{j+1/2}-\mathbf{v}^*_{j+1/2})+\Lambda_2^x\mathbf{f}^-(\mathbf{v}^{n+1}_{j+1/2})=\mathbf{0};\quad \frac{1}{\tau}\Lambda_1^x(\mathbf{v}^*_{j+1/2}-\mathbf{v}^n_{j+1/2})+\Lambda_2^x\mathbf{f}^+(\mathbf{v}^*_{j+1/2})=\mathbf{0};\end{aligned}\tag{8.10}$$

где искомые функции $\mathbf{v}^n_j$ и $\mathbf{v}^n_{j+1/2}$, так же как в схеме CABARETM, вычисляются в целых и полуцелых пространственных узлах однородной разностной сетки (2.3), действие сеточных операторов A_0^x, Λ_1^x, Λ_2^x определяется формулами

$$A_0^x\mathbf{v}^n_{j+1/2}=\frac{\mathbf{v}^n_j+4\mathbf{v}^n_{j+1/2}+\mathbf{v}^n_{j+1}}{6},\quad \Lambda_1^x\mathbf{v}^n_{j+1/2}=\frac{\mathbf{v}^n_{j+1}-\mathbf{v}^n_j}{h},\quad \Lambda_2^x\mathbf{v}^n_{j+1/2}=\frac{4(\mathbf{v}^n_j-2\mathbf{v}^n_{j+1/2}+\mathbf{v}^n_{j+1})}{h^2},$$

а функции $\mathbf{f}^{\pm}(\mathbf{u})$ порождаются глобальным расщеплением потоков Лакса–Фридрихса

$$\mathbf{f}^{\pm}(\mathbf{u})=\frac{1}{2}\mathbf{f}(\mathbf{u})\pm C_2\mathbf{u},\quad C_2=\frac{1+2\delta}{2}\max_{k,j,\alpha}\left|\lambda_k(\mathbf{v}^n_{j+\alpha})\right|,\tag{8.11}$$

в котором параметр $\delta>0$ отвечает за запас положительной либо отрицательной определенности матриц Якоби $A^{\pm}(\mathbf{u})=\mathbf{f}^{\pm}_{\mathbf{u}}(\mathbf{u})$. Бикомпактная схема (8.10) является консервативной и безусловно устойчивой, имеет четвертый порядок аппроксимации по пространству и первый порядок аппроксимации по времени, поскольку в этой схеме для дискретизации по времени применяется неявный метод Эйлера первого порядка. При аппроксимации линейного уравнения переноса схема (8.10) монотонна при числах Куранта, не меньших 1/4.

Далее мы рассматриваем три бикомпактные схемы на основе (8.10): MBiC, NFC-схему формально третьего порядка аппроксимации по времени (см. [11]); RBiC, HASIA-схему с пассивной (в терминологии [34]) или глобальной экстраполяцией Ричардсона (см. [21]) второго порядка по времени; CBiC, комбинированную схему второго порядка аппроксимации по времени (см. [33]).

Дадим краткое описание схемы MBiC. Ее построение начинается с того, что порядок аппроксимации схемы (8.10) по времени повышается до третьего. Для этого временные производные аппроксимируются не разностями назад, а диагонально-неявным методом Рунге–Кутты соответствующего порядка из работы [98]:

$$\begin{array}{c|ccc}a & a & & \\ a_{21}+a & a_{21} & a & \\ 1 & a_{31} & a_{32} & a\\ \hline & a_{31} & a_{32} & a\end{array},\quad a=1+\sqrt{2}\cos\left(\frac{4\pi}{3}+\frac{1}{3}\arccos\frac{2\sqrt{2}}{3}\right),$$

$$a_{21}=-\frac{a-1/3}{2(a^2-2a+1/2)},\quad a_{32}=-\frac{2(a^2-2a+1/2)^2}{a-1/3},\quad a_{31}=1-a-a_{32}.$$

Добавим, что данный метод является жестко-точным и L-устойчивым. Полученную таким образом бикомпактную схему третьего порядка аппроксимации по времени будем называть схемой B. По теореме Годунова из [1] схема B является немонотонной. Генерируемые ею нефизические осцилляции подавляются с помощью метода консервативной монотонизации, предложенного в [11]. Его конструкция предполагает использование второй схемы – монотонной схемы A. Последняя не обязана быть бикомпактной и может быть выбрана исходя из любых нужных критериев (экономичность, устойчивость, количество аппроксимационной вязкости). В качестве схемы A возьмем схему (8.10).

Рассмотрим переход со слоя t_n на слой t_{n+1} по схеме MBiC. Искомое решение $\mathbf{v}^{n+1}_h$ находится в три этапа. На первом этапе по схемам A и B вычисляются два решения на слое t_{n+1}: $\mathbf{v}^A_h$ и $\mathbf{v}^B_h$ соответственно. Общим промежуточным начальным условием для обеих схем полагается решение $\mathbf{v}^n_h$.

На втором этапе решение $\mathbf{v}_h^B$ ограничивается локально в каждой ячейке $K_{j+1/2} = [x_j, x_{j+1}]$ около своего интегрального среднего $\overline{\mathbf{v}}_{j+1/2}^B = A_0^x \mathbf{v}_{j+1/2}^B$:

$$\tilde{v}_h(x; K_{j+1/2}) = \overline{v}_{j+1/2}^B + \beta_{j+1/2}[v_h^B(x) - \overline{v}_{j+1/2}^B], \quad x = x_j, x_{j+1/2}, x_{j+1}, \tag{8.12}$$

где

$$\beta_{j+1/2} = \frac{1}{1 + d_{j+1/2}^2}, \quad d_{j+1/2} = \frac{C_1 \max\limits_{x=x_j, x_{j+1/2}, x_{j+1}} \left| v_h^A(x) - v_h^B(x) \right|}{\max\limits_{x=x_j, x_{j+1/2}, x_{j+1}} \left| v_h^A(x) \right|}. \tag{8.13}$$

Формулы (8.12), (8.13) применяются покомпонентно, т.е. $v \equiv v^i$, $\beta \equiv \beta^i$, $d = d^i$, $i = \overline{1, m}$. На третьем этапе устраняется многозначность решения $\tilde{\mathbf{v}}_h$ на границах между ячейками, после чего мы получаем результирующее решение на слое t_{n+1}:

$$\mathbf{v}_h(x_j, t_{n+1}) = \frac{\tilde{\mathbf{v}}_h(x_j; K_{j-1/2}) + \tilde{\mathbf{v}}_h(x_j; K_{j+1/2})}{2}, \quad \mathbf{v}_h(x_{j+1/2}, t_{n+1}) = \tilde{\mathbf{v}}_h(x_{j+1/2}; K_{j+1/2}). \tag{8.14}$$

Число $C_1 \geqslant 0$ является параметром метода консервативной монотонизации (см. [11]).

Суть формул (8.12)–(8.14) сводится к следующему. В областях гладкости точного решения

$$\left| v_h^A(x) - v_h^B(x) \right| = O(\tau), \quad d_{j+1/2} = O(\tau), \quad \left| \beta_{j+1/2} - 1 \right| = O(\tau^2)$$

и решение $\mathbf{v}_h^B$ корректируется лишь на величину $O(h^3)$. Вблизи сильных разрывов точного решения

$$\left| v_h^A(x) - v_h^B(x) \right| = O(1), \quad d_{j+1/2} = O(1), \quad \left| \beta_{j+1/2} - 1 \right| = O(1)$$

и решение $\mathbf{v}_h^B$ сглаживается на величину $O(1)$. Из формул (8.12) и (8.13) следует, что монотонизация тем интенсивнее, чем больше параметр C_1; в предельном случае при $C_1 = 0$ монотонизация отсутствует.

Перейдем к схеме RBiC и основанной на ней комбинированной схеме CBiC. Чтобы построить решение по схеме RBiC, необходимо сначала провести полный расчет задачи по бикомпактной схеме (8.10) на двух сетках: с шагами $h_1 = h$, $\tau_1 = \tau$ и шагами $h_2 = h/2$, $\tau_2 = \tau/2$. Затем на любом слое t_n^1 по двум полученным решениям с помощью пассивной экстраполяции Ричардсона второго порядка вычисляется искомое решение

$$\mathbf{v}_{h_1}(x_{j+\alpha}^1, t_n^1; B) = 2\mathbf{v}_{h_2}(x_{j+\alpha}^1, t_n^1; A) - \mathbf{v}_{h_1}(x_{j+\alpha}^1, t_n^1; A), \tag{8.15}$$

где ради краткости мы обозначили схему (8.10) буквой A (как в MBiC), а схему RBiC – буквой B. Подчеркнем, что пассивная (глобальная) экстраполяция Ричардсона, в отличие от активной (локальной), реализуется не при каждом переходе со слоя на слой (подобно стадиям методов Рунге–Кутты), а только после полного завершения счета.

Решение, задаваемое формулой (8.15), содержит осцилляции, однако их количество невелико, и они локализованы в окрестностях ударных волн. Поскольку схема RBiC реализуется пассивно, эти осцилляции разумно устранять некоторой простой пост-обработкой. Применяя для этого формулы, аналогичные формулам гибридной схемы из [10], мы приходим к комбинированной схеме CBiC из [33]:

$$\begin{gathered} v_h(x_{j+\alpha}^1, t_n^1; \mathrm{CBiC}) = v_{h_1}(x_{j+\alpha}^1, t_n^1; B) + \omega_{j+\alpha} D_{j+\alpha}, \\ D_{j+\alpha} = v_{h_2}(x_{j+\alpha}^1, t_n^1; A) - v_{h_1}(x_{j+\alpha}^1, t_n^1; B), \quad \omega_{j+\alpha} = \begin{cases} 0 & \text{при} \quad \left| D_{j+\alpha} \right| \leqslant d_* Q, \\ 1 - \dfrac{d_* Q}{\left| D_{j+\alpha} \right|} & \text{иначе}, \end{cases} \\ Q = \max_{j,\alpha} v_{h_2}(x_{j+\alpha}^1, t_n^1; A) - \min_{j,\alpha} v_{h_2}(x_{j+\alpha}^1, t_n^1; A), \end{gathered} \tag{8.16}$$

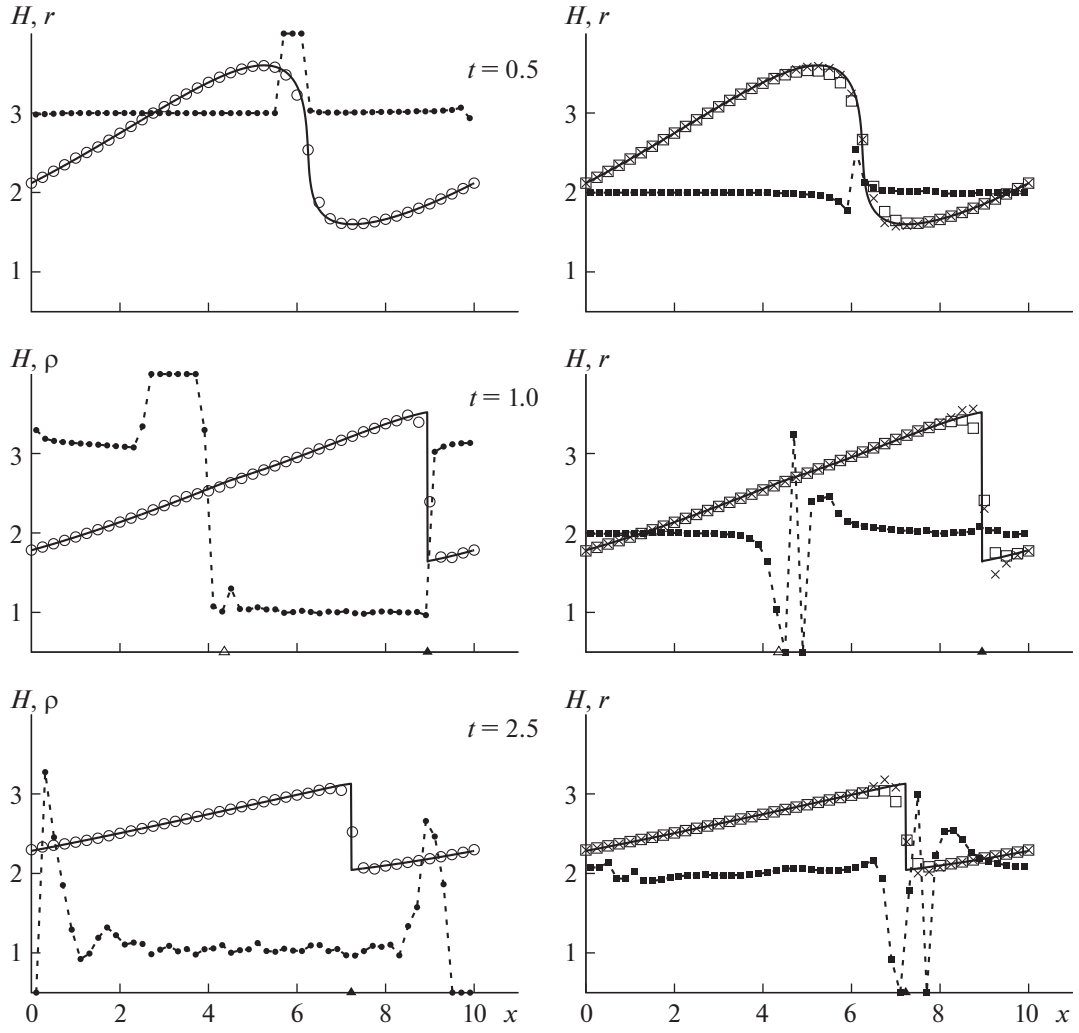
где $d_* \geqslant 0$ – входной параметр схемы CBiC. Как и формулы (8.12)–(8.14), формула (8.16) применяется покомпонентно. В комбинированной схеме CBiC базисной HASIA-схемой является схема RBiC (схема B), а внутренней монотонной схемой – схема (8.10) (схема A). Отметим, что для реализации комбинированной схемы CBiC не нужно явного выделения фронтов ударных волн вида (6.1), (6.2) и последующего решения внутренних начально-краевых задач. Обратим внимание на одно важное свойство базисной схемы RBiC. Предположим, что ударная волна большой амплитуды распространяется по фону, близкому к границам области применимости данной физической модели. Например, в случае уравнений мелкой воды такой границей является нулевая глубина жидкости ($H = 0$), в случае уравнений газовой динамики – нулевые плотность, температура и давление газа. При сквозном расчете такого процесса по какой-либо HASIA-схеме может возникнуть осцилляция, которая выведет решение в область физически недопустимых значений, что, в свою очередь, приведет к аварийному завершению счета. В схеме же RBiC такая ситуация исключена, так как решение этой базисной схемы строится пассивно из монотонных решений внутренней бикомпактной схемы (8.10).

Обсудим результаты расчетов основной тестовой задачи, описанной в разд. 3, по бикомпактным схемам MBiC, RBiC и CBiC. Укажем значения параметров схем: во всех схемах $\delta = 0.2$; в схеме MBiC $\tau/h = 0.09$ и $C_1 = 10$; в схемах RBiC и CBiC $\tau/h = 0.05$, в схеме CBiC $d_* = 0.01$.

На фиг. 13 показаны численные значения глубины жидкости H, а также порядки локальной сходимости r, определяемые по формуле (2.7) при $k = 2$, и порядки интегральной сходимости ρ, определяемые по формуле (2.9), в которой $k = 2$ и интегралы $\mathbf{V}_{h_i}$ вычисляются по формуле парабол. Порядки интегральной сходимости приводятся в том случае (графики слева на фиг. 13 при $t = 1{,}2.5$), когда порядки локальной сходимости разностных решений сильно осциллируют и не дают необходимой информации о точности этих решений. На фиг. 14 приведены относительные локальные дисбалансы Δw_{ih} вычисления инвариантов w_i, задаваемые формулами (2.14) и (2.18). Глубины жидкости H определялись на сетке (2.3) с пространственным шагом $h = 0.25$, а величины r, ρ и Δw_{ih} были вычислены на базисной сетке (2.3) с шагом $h = 0.005$ и показаны для каждой 40-й ячейки этой сетки. В качестве квазиточного решения в формулах (2.7) и (2.9) использовалось численное решение, полученное по схеме Русанова на достаточно мелкой сетке.

Из фиг. 13 следует, что профили глубины, рассчитанные по схемам MBiC и CBiC, практически не отличаются друг от друга. Разница между схемами проявляется только к моменту времени $t = 2.5$: NFC-схема MBiC размазывает фронт ударной волны на две ячейки, а комбинированная схема CBiC – на три. Немонотонность HASIA-схемы RBiC заметна при $t = 1.0$, 2.5. Как уже было отмечено выше, осцилляции численного решения локализованы непосредственно перед и за фронтом ударной волны. В комбинированной схеме CBiC эти осцилляции отсутствуют. Из фиг. 13 также следует, что бикомпактная схема MBiC, как и другие NFC-схемы, снижает свою точность в области влияния ударной волны: порядок сходимости снижается с третьего до первого. Комбинированная бикомпактная схема CBiC имеет второй порядок локальной сходимости во всей расчетной области, за исключением некоторых окрестностей ударных волн. Это означает, что бикомпактная схема RBiC с пассивной экстраполяцией Ричардсона по времени является HASIA-схемой.

Из фиг. 14 следует, что при $t = 0.5$, когда начинают формироваться области больших градиентов, но решение еще остается гладким, схемы RBiC и CBiC предсказуемо проигрывают NFC-схемам MBiC и WENO5 по реальной точности. Это объясняется более низким, вторым порядком аппроксимации по времени у схем RBiC и CBiC. Однако при $t = 1.0$ в области влияния ударной волны схемы RBiC и CBiC, наоборот, превосходят NFC-схемы MBiC и WENO5: погрешности инвариантов для первой пары схем в среднем на порядок меньше. Тем не менее схемы MBiC и WENO5 по-прежнему демонстрируют лучшую точность вне области влияния. По мере расширения этой области данное преимущество исчезает. К моменту времени $t = 2.5$, когда область влияния ударной волны занимает всю расчетную область, схемы RBiC и CBiC оказываются в среднем на порядок точнее всюду, кроме относительно небольших окрестностей ударных волн. При этом схемы MBiC и WENO5, будучи по существу разными схемами, имеют примерно одну и ту же точность в области влияния ударной волны.



Фиг. 13. Глубины жидкости H, порядки локальной сходимости r и интегральной сходимости ρ, получаемые по схемам MBiC (кружки полые и закрашенные), CBiC (квадратики полые и закрашенные) и RBiC (крестики). Сплошными линиями изображены квазиточные значения глубины.

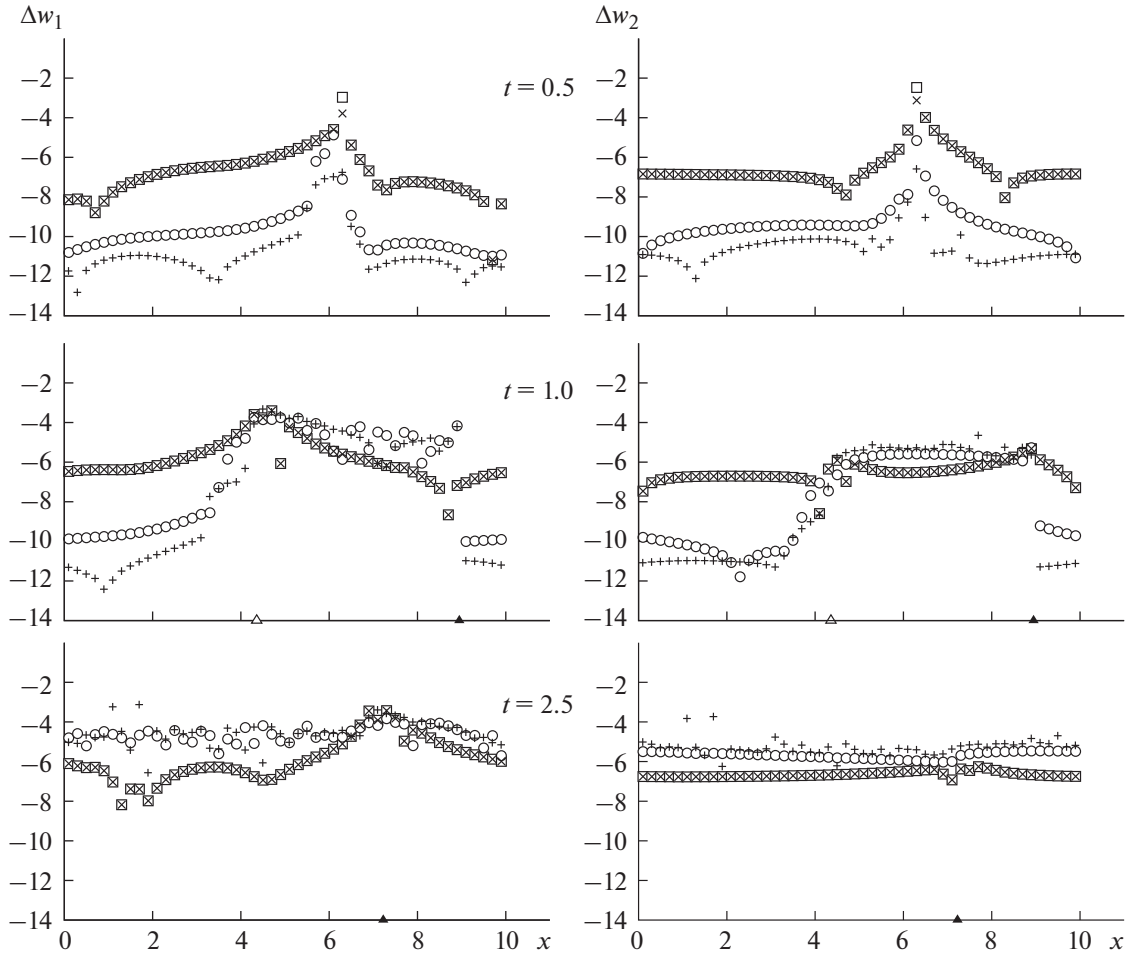
9. ЗАДАЧА О МНОГОКРАТНОМ ВЗАИМОДЕЙСТВИИ УДАРНЫХ ВОЛН

В этом разделе приводятся результаты работы [99], в которой проведен сравнительный анализ точности комбинированных разностных схем CR и CCWA, построенных в разд. 7, со схемой WENO5 при численном моделировании задачи о многократном взаимодействии ударных волн на равномерной сетке (2.3) с коэффициентом запаса $z = 0.45$ в условии устойчивости (2.4).

Рассмотрим для системы уравнений мелкой воды (2.1), (3.1) задачу Коши (2.2) с периодическими начальными данными

$$h(x,0) = 2\sin\frac{\pi(2x+5)}{X} + 3, \quad v(x,0) = 0, \tag{9.1}$$

где $X = 10$ – длина периода; на фиг. 15 начальное значение глубины жидкости $h(x,0)$ показано штриховой линией. Квазиточное решение задачи (2.1), (3.1), (9.1) моделируется численным расчетом по CR-схеме на мелкой сетке с пространственным шагом $h = 0.005$. Профили глубины, получаемые в этом расчете в моменты времени $t = 0.5, 1, 2, 3$, изображены на фиг. 15 сплошными линиями на отрезке $[0, X]$ длины периода. В результате решения данной задачи впадина, расположенная в начальный момент времени на интервале $[0, X]$ (штриховая линия), постепенно заполняется жидкостью, что приводит к подъему ее уровня в окрестности точки $x = X/2$ (линия *1*).

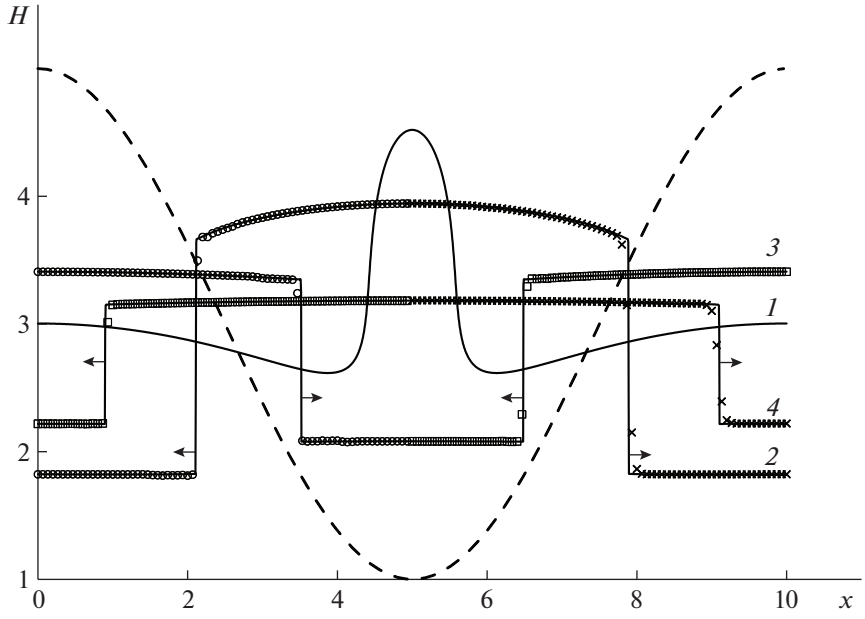


Фиг. 14. Относительные локальные дисбалансы $\Delta w_{ih}(x_j, t)$ вычисления инвариантов w_i, получаемые по схемам MBiC (кружки), CBiC (квадратики), RBiC (косые крестики) и WENO5 (прямые крестики).

Этот подъем уровня вызывает формирование двух расходящихся ударных волн (линия *2*), которые на границах интервала $[0, X]$ взаимодействуют с ударными волнами, двигающимися им навстречу. В результате этого взаимодействия на интервале $[0, X]$ образуются две сходящиеся ударные волны (линия *3*). Эти ударные волны, соударяясь в точке $x = X/2$, формируют две новые расходящиеся ударные волны (линия *4*). Далее этот процесс повторяется, приводя к возникновению новых пар сходящихся и расходящихся ударных волн, амплитуда которых постепенно уменьшается, что связано с потерей полной энергии потока на фронтах этих волн.

На фиг. 15 в моменты времени $t = 1, 2, 3$ приведены глубины жидкости, получаемые в результате численных расчетов задачи (2.1), (3.1), (9.1) по схемам CR, CCWA и WENO5 с пространственным шагом $h = 1/15$. Поскольку эти разностные решения (так же, как и точное решение) симметричны относительно точки $x = X/2$, то на фиг. 15 каждое из них в фиксированный момент времени приводится на одном из отрезков $[0, X/2]$ или $[X/2, X]$ длины полупериода. Из фиг. 15 следует, что комбинированные схемы CR и CCWA размазывают фронт ударной волны существенно меньше, чем схема WENO5. Это объясняется тем, что в этих комбинированных схемах в качестве внутренней используется схема CABARETM, которая с высокой точностью локализует фронты ударных волн.

На фиг. 16 в моменты времени $t = 0.5, 1, 2, 3$ показаны относительные локальные дисбалансы $\Delta w_{1h}(x_j, t)$, вычисления инварианта $w_1 = v - 2c$, задаваемые формулами (2.14) и (2.18) при $h = 0.02$, в которых квазиточное значение инварианта w_1 моделируется расчетом по CR-схеме при $h = 0.0002$. В силу симметрии начальных данных (9.1) графики относительных локальных



Фиг. 15. Глубины жидкости H, получаемые при численном решении задачи Коши (2.1), (3.1), (9.1) в моменты времени $t = 0$ (штриховая линия), $t = 0.5$ (линия *1*), $t = 1$ (линия *2*), $t = 2$ (линия *3*) и $t = 3$ (линия *4*). Кружками показаны результаты численного расчета по CR-схеме, квадратиками – по CCWA-схеме и крестиками – по схеме WENO5.

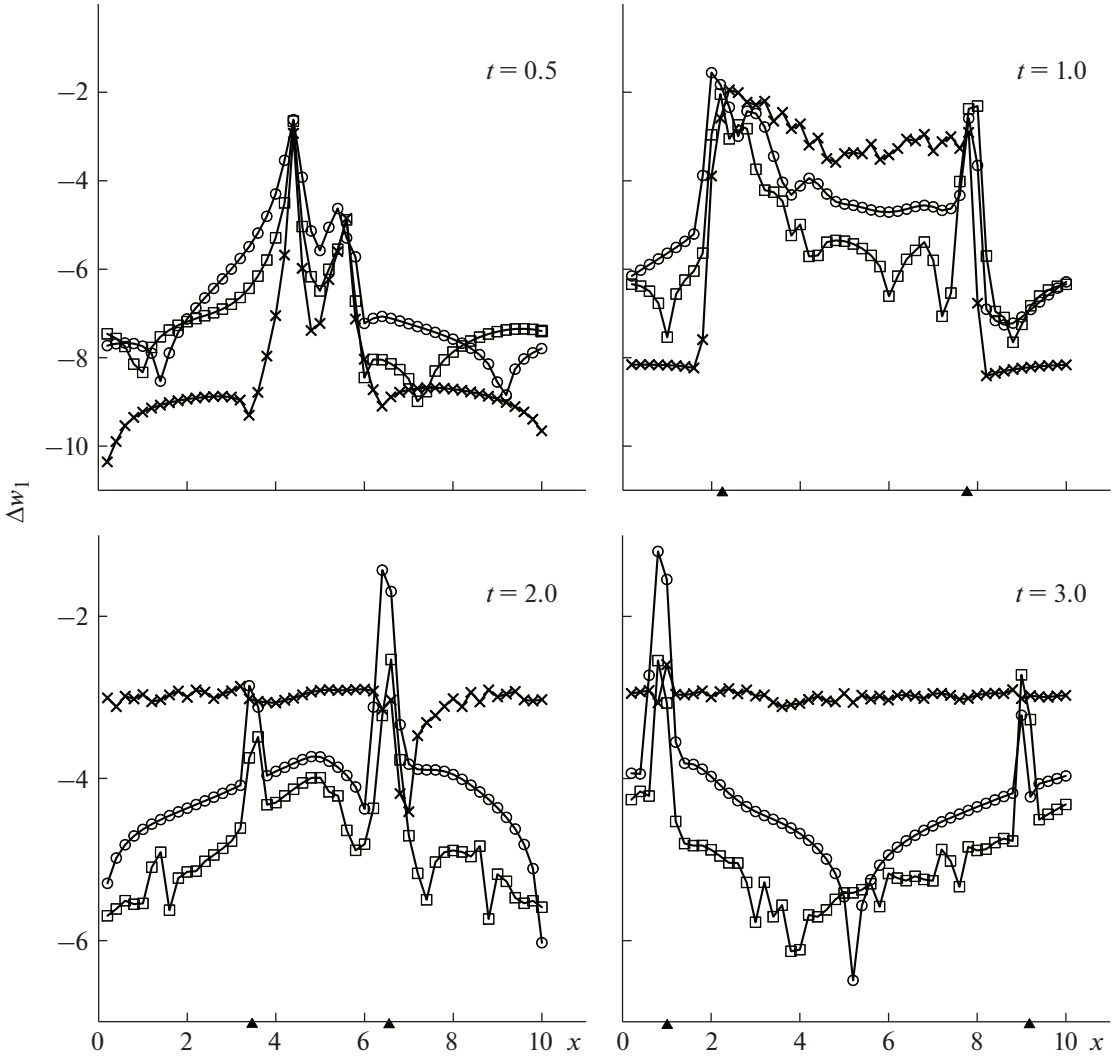
дисбалансов $\Delta w_{2h}(x_j, t)$ вычисления инварианта $w_2 = v + 2c$ получаются из графиков, приведенных на фиг. 16, путем их отражения относительно вертикальной прямой $x = X/2$. Из фиг. 16 при $t = 0.5, 1$ следует, что на интервалах, которые не входят в области влияния ударных волн, точность вычисления инвариантов в схеме WENO5 существенно выше, чем в комбинированных схемах; это объясняется более высокой точностью схемы WENO5 на гладких решениях. Однако на интервалах, расположенных между расходящимися ударными волнами, на фиг. 16 при $t = 1$ точность схемы WENO5 резко падает и становится ниже, чем в комбинированных схемах. С течением времени, когда вся расчетная область становится областью влияния взаимодействующих ударных волн (фиг. 16 при $t = 2, 3$), точность вычисления инвариантов в комбинированных схемах становится существенно выше, чем в схеме WENO5, везде за исключением малых окрестностей ударных волн, где сходимость разностного решения к точному отсутствует. Отметим также, что CCWA-схема, в которой базисной является компактная схема (4.4)–(4.7), демонстрирует более высокую точность, чем CR-схема, в которой базисной является схема Русанова (4.1)–(4.3). Это объясняется тем, что компактная схема (4.4)–(4.7) имеет третий порядок как классической аппроксимации на гладких решениях, так и слабой аппроксимации на разрывных решениях, в то время как схема Русанова третьего порядка классической аппроксимации имеет лишь первый порядок слабой аппроксимации.

10. ПРИМЕНЕНИЕ КОМБИНИРОВАННОЙ БИКОМПАКТНОЙ СХЕМЫ ДЛЯ РАСЧЕТА ДВУМЕРНЫХ ГИДРОДИНАМИЧЕСКИХ ТЕЧЕНИЙ

В этом разделе приводятся результаты работы [97], в которой комбинированная бикомпактная схема, предложенная в [33], обобщается на многомерный случай.

Двумерные уравнения первого приближения теории мелкой воды в случае плоского горизонтального дна без учета влияния донного трения имеют вид

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x + \mathbf{g}(\mathbf{u})_y = \mathbf{0}, \tag{10.1}$$



Фиг. 16. Относительные локальные дисбалансы $\Delta w_{1h}(x_j,t)$ вычисления инварианта $w_1 = v - 2c$, получаемые при численном решении задачи Коши (2.1), (3.1), (9.1) по схемам Русанова (кружки), CWA (квадратики) и WENO5 (крестики).

$$\mathbf{u} = \begin{pmatrix} H \\ q_x \\ q_y \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} q_x \\ q_x^2/H + gH^2/2 \\ q_x q_y/H \end{pmatrix}, \quad \mathbf{g}(\mathbf{u}) = \begin{pmatrix} q_y \\ q_x q_y/H \\ q_y^2/H + gH^2/2 \end{pmatrix}, \tag{10.2}$$

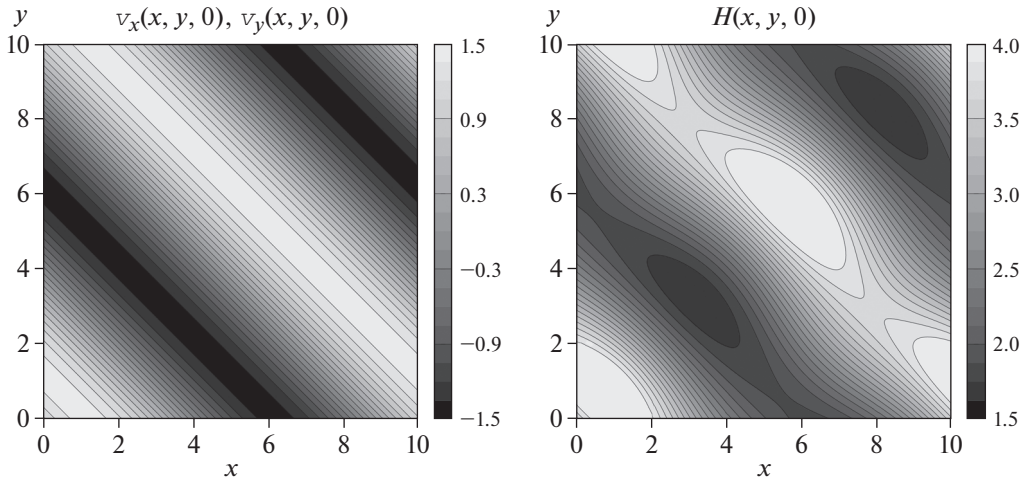
где $H(x,y,t)$ и $q_x(x,y,t)$, $q_y(x,y,t)$ – глубина и горизонтальные компоненты импульса жидкости соответственно, g – ускорение свободного падения (в расчетах $g = 10$). Рассмотрим для системы (10.1), (10.2) задачу Коши со следующими периодическими начальными данными (фиг. 17):

$$v_x(x,y,0) = v_y(x,y,0) = \frac{a}{\sqrt{2}}\sin\left(\frac{2\pi x}{X} + \frac{2\pi y}{Y} + \frac{\pi}{4}\right), \tag{10.3}$$

$$H(x,y,0) = H_1(x,y)H_2(x,y), \tag{10.4}$$

в которых

$$\begin{aligned} H_1(x,y) &= \frac{1}{4g}\left[a\sin\left(\frac{2\pi x}{X} + \frac{2\pi y}{Y} + \frac{\pi}{4}\right) + b\right]^2, \\ H_2(x,y) &= 1 + \frac{1}{10}\left[\sin^2\left(\frac{2\pi x}{X} + \frac{\pi}{4}\right) + \sin^2\left(\frac{2\pi y}{Y} + \frac{\pi}{4}\right)\right], \end{aligned} \tag{10.5}$$



Фиг. 17. Начальные значения скоростей жидкости v_x, v_y и глубины жидкости H, задаваемые формулами (10.3) и (10.4).

где $v_x = q_x/H$, $v_y = q_y/H$ – компоненты скорости жидкости, $a = 2$, $X = Y = b = 10$. Двумерные начальные условия (10.3) и (10.4) соответствующим образом согласованы с одномерными начальными условиями (3.2) основной тестовой задачи и с учетом множителя (10.5) в начальном условии (10.4) являются существенно двумерными. При решении двумерной задачи Коши (10.1)–(10.4), аналогично решению одномерной тестовой задачи (2.1), (3.1), (3.2), в момент времени $t \approx 0.5$ в результате градиентных катастроф начинает формироваться последовательность изолированных ударных волн, которые распространяются друг за другом с одинаковыми скоростями; при $t \geq 1$ области влияния этих ударных волн заполняют всю расчетную область.

Комбинированная бикомпактная схема CBiC, аппроксимирующая задачу (10.1)–(10.4) и представляющая собой двумерный вариант схемы A из п. 8.2, строится на равномерной прямоугольной разностной сетке

$$S_2 = \{(x_i, y_j, t_n)\colon x_i = ih_x,\ y_j = jh_y,\ t_n = n\tau,\ n \geq 0\}, \tag{10.6}$$

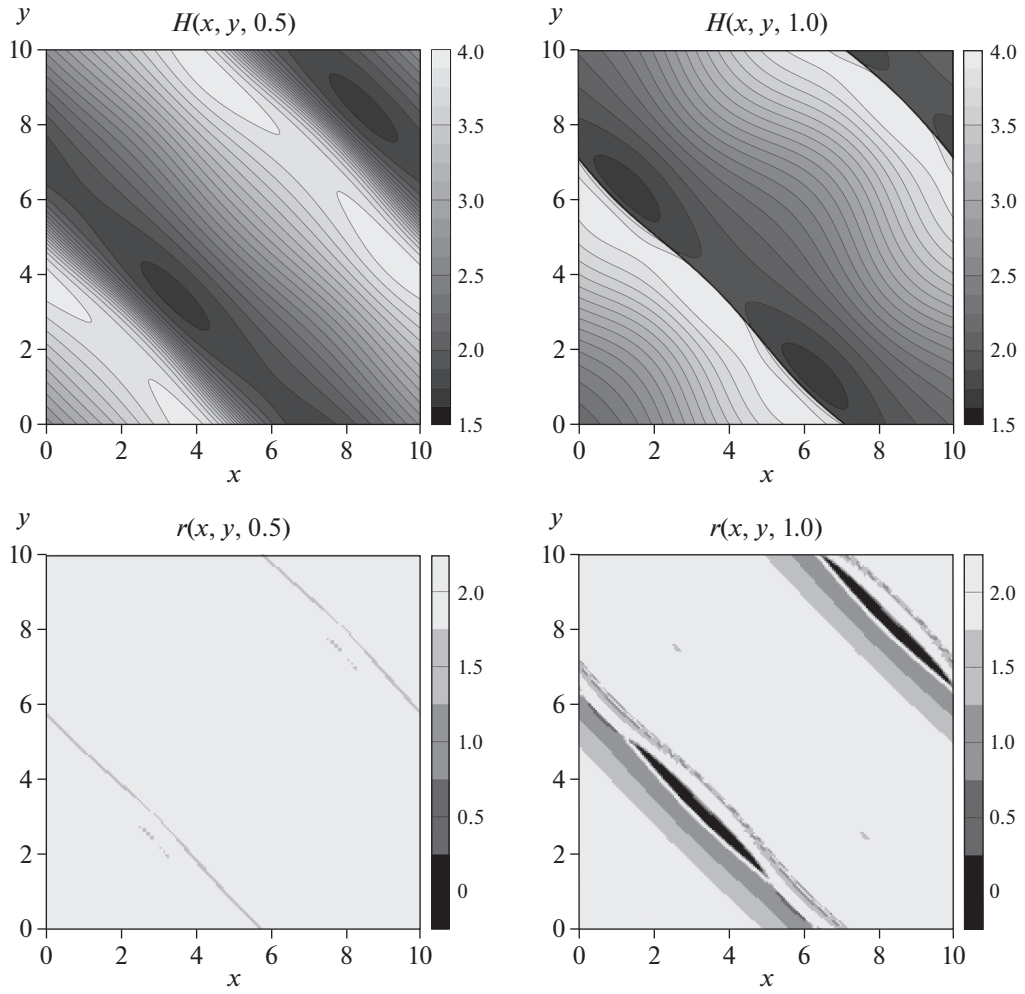
в которой $h_x = X/M_x$ и $h_y = Y/M_y$ – постоянные шаги сетки по пространственным переменным x и y, где M_x и M_y – заданные целые положительные числа; τ – постоянный шаг сетки по времени. Так же, как в пространственно одномерных бикомпактных схемах, в пространственно двумерной схеме CBiC численное решение определяется как в целых x_i, y_j, так и в полуцелых $x_{i+1/2} = (i + 1/2)h_x$, $y_{j+1/2} = (j + 1/2)h_y$ пространственных узлах разностной сетки (10.6). Для этого решения на n-м временном слое введем сокращенное обозначение $\mathbf{v}^n_{\mathbf{h}}$, где $\mathbf{h} = (h_x, h_y)$.

Применяя методы локально-одномерного расщепления из [95] и глобального расщепления потоков Лакса–Фридрихса, представим систему (10.1) в виде суммы четырех одномерных систем

$$\frac{1}{4}\mathbf{u}_t + \mathbf{f}^{\pm}(\mathbf{u})_x = \mathbf{0}, \quad \frac{1}{4}\mathbf{u}_t + \mathbf{g}^{\pm}(\mathbf{u})_y = \mathbf{0}, \tag{10.7}$$

в которых потоки $\mathbf{f}^{\pm}(\mathbf{u})$, $\mathbf{g}^{\pm}(\mathbf{u})$ определяются аналогично одномерному случаю (8.11). При переходе с временного слоя t_n на слой t_{n+1} системы (10.7) решаются последовательно с шагом $\tau/4$ с помощью одномерных схем A, построенных в п. 8.2. Например, одномерные разностные схемы, аппроксимирующие системы (10.7) с потоками $\mathbf{f}^{\pm}(\mathbf{u})$, имеют вид

$$\frac{1}{\tau}A_0^x\left(\mathbf{v}^{(2)}_{i+1/2,j+\alpha} - \mathbf{v}^{(1)}_{i+1/2,j+\alpha}\right) + \Lambda_1^x\mathbf{f}^{\pm}(\mathbf{v}^{(2)}_{i+1/2,j+\alpha}) = \mathbf{0},$$
$$\frac{1}{\tau}\Lambda_1^x\left(\mathbf{v}^{(2)}_{i+1/2,j+\alpha} - \mathbf{v}^{(1)}_{i+1/2,j+\alpha}\right) + \Lambda_2^x\mathbf{f}^{\pm}(\mathbf{v}^{(2)}_{i+1/2,j+\alpha}) = \mathbf{0},$$



Фиг. 18. Глубины жидкости H и порядки локальной сходимости r, получаемые при численном решении по комбинированной схеме CBiC двумерной задачи Коши (10.1)–(10.4).

где $\mathbf{v}^{(1)}$ и $\mathbf{v}^{(2)} = S_x^{\pm}(\tau) \circ \mathbf{v}^{(1)}$ – промежуточные численные решения, $S_x^{\pm}(\tau)$ – операторы перехода, $\alpha = 0, 1/2$. Аналогичным образом для одномерных бикомпактных схем A, аппроксимирующих системы (10.7) с потоками $\mathbf{g}^{\pm}(\mathbf{u})$, будем использовать операторную форму записи $\mathbf{v}^{(2)} = S_y^{\pm}(\tau) \circ \mathbf{v}^{(1)}$. С учетом этого двумерная бикомпактная схема A, аппроксимирующая систему (10.1), задается формулой

$$\mathbf{v}_{\mathbf{h}}^{n+1} = S_y^{-}(\tau) \circ S_y^{+}(\tau) \circ S_x^{-}(\tau) \circ S_x^{+}(\tau) \circ \mathbf{v}_{\mathbf{h}}^{n}.$$

Пассивная экстраполяции Ричардсона (8.15) и процедура пост-обработки комбинированной схемы (8.16) естественным образом обобщаются на двумерный случай, что позволяет построить двумерные схемы RBiC и CBiC. Отметим, что применение локально-одномерного расщепления в построенной двумерной бикомпактной схеме A является уместным, поскольку эта схема имеет первый порядок аппроксимации по времени, который повышается до второго порядка с помощью экстраполяции Ричардсона в схеме RBiC.

Рассмотрим результаты расчетов задачи Коши (10.1)–(10.4) по комбинированной бикомпактной схеме CBiC, которые выполнялись при параметрах $\delta = 0.2$, $\tau/h_x = \tau/h_y = 0.04$ и $d_* = 0.01$. На фиг. 18 в моменты времени $t = 0.5, 1.0$ приведены графики глубины жидкости, получаемые при расчете на сетке (10.6) с пространственными шагами $h_x = h_y = 0.025$, и графики порядков

локальной сходимости, определяемые по формуле, аналогичной (2.7), в которой $k = 2$ и в качестве базисной сетки используется разностная сетка (10.6) при $h_x = h_y = 0.05$. Из фиг. 18 следует, что в решении, получаемом по комбинированной схеме CBiC, отсутствуют нефизические осцилляции в окрестностях ударных волн, а порядки локальной сходимости $r \approx 2$ почти во всей расчетной области, как при $t = 0.5$, когда ударные волны только начинают формироваться, так и при $t = 1$, когда области влияния ударных волн занимают всю расчетную область. Отклонение от второго порядка сходимости происходит только в малых окрестностях ударных волн, где отсутствует локальная сходимость разностного решения к точному.

11. ЗАКЛЮЧЕНИЕ

Рассмотренные в данной работе комбинированные схемы сквозного счета: CR (Combined Rusanov), CCWA (Combined Compact Weak Approximation), CDG (Combined Discontinuous Galerkin) и CBiC (Combined BiCompact), при численном расчете основной тестовой задачи (2.1), (3.1), (3.2) обеспечивают существенно более высокую точность в областях влияния ударных волн по сравнению со стандартными NFC-схемами (MUSCL, TVD, CU, WENO5, DG1A2, CABARETM, MBiC), которые в настоящее время широко применяются при численном моделировании прикладных задач математической физики. Причем, как показали тестовые расчеты, точность комбинированных схем при расчете разрывных решений может быть на несколько порядков выше точности NFC-схем (независимо от формального порядка аппроксимации этих схем на гладких решениях), в силу чего изложенные в данной работе результаты заметно превосходят современный мировой уровень развития данного научного направления и, по нашему мнению, в значительной степени будут определять дальнейшее развитие этого научного направления. Поэтому безусловно актуальной представляется необходимость дальнейшего всестороннего развития теории комбинированных схем сквозного счета и изучение возможностей применения этих схем для численного расчета различных многомерных прикладных задач. Сформулируем основные перспективные направления развития теории комбинированных схем.

В комбинированных схемах CR, CCWA и CDG в качестве базисных применяются HASIA-схемы третьего порядка, в то время как в комбинированной бикомпактной схеме CBiC базисной является HASIA-схема второго порядка. В результате точность вычисления инвариантов в области влияния ударных волн по схеме CBiC приблизительно на порядок выше, чем в NFC-схеме WENO5 (фиг. 14), но одновременно приблизительно на порядок ниже, чем в схемах CR, CCWA и CDG (фиг. 5, 6, 11). Поэтому несомненный интерес представляет разработка новых комбинированных бикомпактных схем, в которых в качестве базисных будут использованы бикомпактные HASIA-схемы не ниже третьего порядка, что позволит существенно повысить точность этих комбинированных схем, по сравнению со схемой CBiC.

В настоящей работе тестирование разностных схем сквозного счета проводилось на задачах Коши с гладкими периодическими начальными условиями, в точных решениях которых ударные волны возникают в результате градиентных катастроф строго внутри расчетных областей. Сделано это было для того, чтобы на первом этапе построения теории комбинированных схем избежать проблемы, связанной с сохранением повышенной точности этих схем при аппроксимации разрывных начальных и граничных условий. Поэтому на следующих этапах развития этой теории необходимо будет разработать методику построения комбинированных схем, имеющих повышенную точность при численном расчете обобщенных решений задач Коши и начально-краевых задач с разрывными начальными и граничными условиями, в частности, при расчете различных задач Римана о распаде начального разрыва.

К настоящему времени преимущества комбинированных схем были продемонстрированы на тестах, связанных с расчетом разрывных решений системы уравнений мелкой воды, представляющей собой простейшую сильно нелинейную гиперболическую систему законов сохранения (см. [35]), не имеющую контактных разрывов и допускающую запись в форме инвариантов. Поэтому в дальнейшем предполагается провести изучение точности комбинированных схем при расчете разрывных решений гиперболических систем законов сохранения общего вида, в частности, законов сохранения неизоэнтропической газовой динамики (см. [36]), допускающих контактные разрывы.

В данной работе при построении комбинированных схем область расчета по внутренней NFC-схеме определяется простейшим градиентным методом (6.2), что затрудняет применение этих схем для моделирования более сложных (в том числе многомерных) задач, в которых происходит многократное взаимодействие сильных разрывов. В то же время в рамках теории по-

строения гибридных схем (см. [24]–[28]) разработаны более эффективные методы выделения областей, внутри которых сосредоточены основные особенности рассчитываемого точного решения. Наиболее распространенными являются методы, использующие коэффициенты фурье-разложения (см. [24]) или вейвлет-разложения (см. [25]) разностного решения, основанные на определении численного производства энтропии (см. [26]), а также использующие слабую локальную невязку разностного решения, получаемую путем подстановки непрерывного аналога этого решения в аппроксимируемую систему законов сохранения, записанную в слабо интегральной форме (см. [27]). С учетом этого в дальнейшем планируется разработать новые классы комбинированных схем сквозного счета, в которых для выделения областей больших градиентов, где расчет проводится по внутренней NFC-схеме, будут применены наиболее эффективные методы, развитые в теории гибридных схем.

СПИСОК ЛИТЕРАТУРЫ

1. *Годунов С.К.* Разностный метод численного расчета разрывных решений уравнений гидродинамики // Мат. сб. 1959. Т. 47. № 3. С. 271–306.
2. *Van Leer B.* Toward the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method // J. Comput. Phys. 1979. V. 32. № 1. P. 101–136. https://doi.org/10.1016/0021-9991(79)90145-1
3. *Harten A.* High resolution schemes for hyperbolic conservation laws // J. Comput. Phys. 1983. V. 49. P. 357–393. https://doi.org/10.1016/0021-9991(83)90136-5
4. *Harten A., Osher S.* Uniformly high-order accurate nonoscillatory schemes // SIAM J. Numer. Analys. 1987. V. 24. № 2. P. 279–309. https://doi.org/10.1007/978-3-642-60543-7_11
5. *Nessyahu H., Tadmor E.* Non-oscillatory central differencing for hyperbolic conservation laws // J. Comput. Phys. 1990. V. 87. № 2. P. 408–463. https://doi.org/10.1016/0021-9991(90)90260-8
6. *Liu X.D., Osher S., Chan T.* Weighted essentially non-oscillatory schemes // J. Comput. Phys. 1994. V. 115. № 1. P. 200–212. https://doi.org/10.1006/jcph.1994.1187
7. *Jiang G.S., Shu C.W.* Efficient implementation of weighted ENO schemes // J. Comput. Phys. 1996. V. 126. P. 202–228. https://doi.org/10.1006/jcph.1996.0130
8. *Cockburn B.* An introduction to the discontinuous Galerkin method for convection-dominated problems // Lect. Not. Math. 1998. V. 1697. P. 150–268. https://doi.org/10.1007/BFb0096353
9. *Karabasov S.A., Goloviznin V.M.* Compact accurately boundary-adjusting high-resolution technique for fluid dynamics // J. Comput. Phys. 2009. V. 228. P. 7426–7451. https://doi.org/10.1016/j.jcp.2009.06.037
10. *Рогов Б.В., Михайловская М.Н.* Монотонные бикомпактные схемы для линейного уравнения переноса // Матем. моделирование. 2011. Т. 23. № 6. С. 98–110. https://doi.org/10.1134/S2070048212010103
11. *Bragin M.D., Rogov B.V.* Conservative limiting method for high-order bicompact schemes as applied to systems of hyperbolic equations // Appl. Numer. Math. 2020. V. 151. P. 229–245. https://doi.org/10.1016/j.apnum.2020.01.005
12. *Остапенко В.В.* О сходимости разностных схем за фронтом нестационарной ударной волны // Ж. вычисл. матем. и матем. физ. 1997. Т. 37. № 10. С. 1201–1212.
13. *Casper J., Carpenter M.H.* Computational consideration for the simulation of shock-induced sound // SIAM J. Sci. Comput. 1998. V. 19. № 1. P. 813–828. https://www.jstor.org/stable/2587267
14. *Engquist B., Sjogreen B.* The convergence rate of finite difference schemes in the presence of shocks // SIAM J. Numer. Anal. 1998. V. 35. P. 2464–2485. https://www.jstor.org/stable/2587267
15. *Остапенко В.В.* О построении разностных схем повышенной точности для сквозного расчета нестационарных ударных волн // Ж. вычисл. матем. и матем. физ. 2000. Т. 40. № 12. С. 1857–1874.
16. *Ковыркина О.А., Остапенко В.В.* О сходимости разностных схем сквозного счета // Докл. АН. 2010. Т. 433. № 5. С. 599–603. https://doi.org/10.1134S1064562410040265
17. *Михайлов Н.А.* О порядке сходимости разностных схем WENO за фронтом ударной волны // Матем. моделирование. 2015. Т. 27. № 2. С. 129–138. https://doi.org/10.1134/S2070048215050075

18. *Ладонкина М.Е., Неклюдова О.А., Остапенко В.В., Тишкин В.Ф.* О точности разрывного метода Галеркина при расчете ударных волн // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 8. С. 148–156. https://doi.org/10.1134/S0965542518080122
19. *Ковыркина О.А., Остапенко В.В.* О монотонности и точности схемы КАБАРЕ при расчете обобщенных решений с ударными волнами // Вычисл. техн. 2018. Т. 23. № 2. С. 37–54.
20. *Ковыркина О.А., Остапенко В.В.* О точности схем типа MUSCL при расчете ударных волн // Докл. РАН. Матем., информ., процессы управл. 2020. Т. 492. С. 43–48. https://doi.org/10.1134/S1064562420030126
21. *Брагин М.Д., Рогов Б.В.* О точности бикомпактных схем при расчете нестационарных ударных волн // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 5. С. 884–899. https://doi.org/10.1134/S0965542520050061
22. *Остапенко В.В.* О конечно-разностной аппроксимации условий Гюгонио на фронте ударной волны, распространяющейся с переменной скоростью // Ж. вычисл. матем. и матем. физ. 1998. Т. 38. № 8. С. 1355–1367.
23. *Русанов В.В.* Разностные схемы третьего порядка точности для сквозного счета разрывных решений // Докл. АН СССР. 1968. Т. 180. № 6. С. 1303–1305.
24. *Gelb A., Tadmor E.* Adaptive edge detectors for piecewise smooth data based on the minmod limiter // J. Sci. Comput. 2006. V. 28. P. 279–306. https://doi.org/10.1007/s10915-006-9088-6
25. *Arandiga F., Baeza A., Donat R.* Vector cell-average multiresolution based on Hermite interpolation // Adv. Comput. Math. 2008. V. 28. P. 1–22. https://doi.org/10.1007/s10444-005-9007-7
26. *Guermond J.L., Pasquetti R., Popov B.* Entropy viscosity method for nonlinear conservation laws // J. Comput. Phys. 2011. V. 230. P. 4248–4267. https://doi.org/10.1016/j.jcp.2010.11.043
27. *Dewar J., Kurganov A., Leopold M.* Pressure-based adaption indicator for compressible Euler equations // Numer. Meth. Part. Diff. Eq. 2015. V. 31. P. 1844–1874. https://doi.org/10.1002/num.21970
28. *Брагин М.Д., Рогов Б.В.* Гибридные бикомпактные схемы с минимальной диссипацией для уравнений гиперболического типа // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 6. С. 958–972. https://doi.org/10.1134/S0965542516060099
29. *Ковыркина О.А., Остапенко В.В.* О построении комбинированных разностных схем повышенной точности // Докл. АН. 2018. Т. 478. № 5. С. 517–522. https://doi.org/10.1134/S1064562418010246
30. *Ковыркина О.А., Остапенко В.В.* О монотонности схемы КАБАРЕ, аппроксимирующей гиперболическую систему законов сохранения // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 9. С. 1488–1504. https://doi.org/10.1134/S0965542518090129
31. *Зюзина Н.А., Ковыркина О.А., Остапенко В.В.* Монотонная разностная схема, сохраняющая повышенную точность в областях влияния ударных волн // Докл. АН. 2018. Т. 482. № 6. С. 639–643. https://doi.org/10.1134/S2070048219010186
32. *Ладонкина М.Е., Неклюдова О.А., Остапенко В.В., Тишкин В.Ф.* Комбинированная схема разрывного метода Галеркина, сохраняющая повышенную точность в областях влияния ударных волн // Докл. АН. 2019. Т. 489. № 2. С. 119–124. https://doi.org/10.1134/S106456241906005X
33. *Брагин М.Д., Рогов Б.В.* Комбинированная монотонная бикомпактная схема, имеющая повышенную точность в областях влияния ударных волн // Докл. АН. 2020. Т. 492. С. 79–84. https://doi.org/10.1134/S1064562420020076
34. *Faragó I., Havasi A., Zlatev Z.* Efficient implementation of stable Richardson Extrapolation algorithms // Comput. Math. Appl. 2010. V. 60. № 8. P. 2309–2325. https://doi.org/10.1016/j.camwa.2010.08.025
35. *Lax P.D.* Hyperbolic systems of conservation laws and the mathematical theory of shock waves. Philadelphia: Soc. Industr. Appl. Math. 1972. 48 p.
36. *Рождественский Б.Л., Яненко Н.Н.* Системы квазилинейных уравнений. М.: Наука, 1978.
37. *Остапенко В.В., Хандеева Н.А.* К обоснованию метода интегральной сходимости исследования точности конечно-разностных схем сквозного счета // Матем. моделирование. 2021. Т. 33. № 6. С. 1028–1037. https://doi.org/10.1134/S207004822106017X
38. *Стокер Дж. Дж.* Волны на воде. М.: Изд-во иностр. лит., 1959.
39. *Burstein S.Z., Mirin A.A.* Third order difference methods for hyperbolic equations // J. Comput. Phys. 1970. V. 5. № 3. P. 547–571. https://doi.org/10.1016/0021-9991(70)90080-X

40. *Lax P., Wendroff B.* Systems of Conservation Laws // Commun. Pure Appl. Math. 1960. V. 13. P. 217–237. https://doi.org/10.1002/cpa.3160130205

41. *MacCormack R.W.* The effect of viscosity in hypervelocity impact cratering // AIAA. 1969. P. 69–354. https://doi.org/10.2514/6.1969-354

42. *Остапенко В.В.* Об аппроксимации законов сохранения разностными схемами сквозного счета // Ж. вычисл. матем. и матем. физ. 1990. Т. 30. № 9. С. 1405–1417. https://doi.org/10.1016/0041-5553(90)90165-O

43. *Остапенко В.В.* Об эквивалентных определениях понятия консервативности для конечно-разностных схем // Ж. вычисл. матем. и матем. физ. 1989. Т. 29. № 8. С. 1114–1128. https://doi.org/10.1016/0041-5553(89)90124-9

44. *Остапенко В.В.* О повышении порядка слабой аппроксимации законов сохранения на разрывных решениях // Ж. вычисл. матем. и матем. физ. 1996. Т. 36. № 10. С. 146–157.

45. *Остапенко В.В.* Аппроксимация условий Гюгонио явными консервативными разностными схемами на нестационарных ударных волнах // Сиб. журн. вычисл. матем. 1998. Т. 1. № 1. С. 77–88.

46. *Остапенко В.В.* О локальном выполнении законов сохранения на фронте размазанной ударной волны // Матем. моделирование 1990. Т. 2. № 7. С. 129–138.

47. *Hirsh R.* Higher order accurate difference solutions of a fluid mechanics problems by a compact differencing technique // J. Comput. Phys. 1975. V. 19. № 1. P. 90–109. https://doi.org/10.1016/0021-9991(75)90118-7

48. *Berger A.E., Solomon J.M., Ciment M., Leventhal S.H., Weinberg B.C.* Generalized OCI schemes for boundary layer problems // Math. Comput. 1980. V. 35. № 6. P. 695–731. https://doi.org/10.1090/S0025-5718-1980-0572850-8

49. *Белоцерковский О.М., Быркин А.П., Мазуров А.П., Толстых А.И.* Разностный метод повышенной точности для расчета течений вязкого газа // Ж. вычисл. матем. и матем. физ. 1982. Т. 22. № 6. С. 1480–1490. https://doi.org/10.1016/0041-5553(82)90110-0

50. *Толстых А.М.* Компактные разностные схемы и их применение в задачах аэрогидродинамики. М.: Наука, 1990.

51. *Остапенко В.В.* Симметричные компактные схемы с искусственными вязкостями повышенного порядка дивергентности // Ж. вычисл. матем. и матем. физ. 2002. Т. 42. № 7. С. 1019–1038.

52. *Iserles A.* Generalized leapfrog methods // IMA Journal of Numerical Analysis. 1986. V. 6. № 4. P. 381–392. https://doi.org/10.1093/imanum/6.4.381

53. *Головизнин В.М., Самарский А.А.* Разностная аппроксимация конвективного переноса с пространственным расщеплением временной производной // Матем. моделирование. 1998. Т. 10. № 1. С. 86–100.

54. *Головизнин В.М., Самарский А.А.* Некоторые свойства разностной схемы "Кабаре" //Матем. моделирование. 1998. Т. 10. № 1. С. 101–116.

55. *Головизнин В.М., Зайцев М.А., Карабасов С.А., Короткин И.А.* Новые алгоритмы вычислительной гидродинамики для многопроцессорных вычислительных комплексов. М.: Изд-во МГУ, 2013.

56. *Karabasov S.A., Goloviznin V.M.* New efficient high-resolution method for nonlinear problems in aeroacoustics // AIAA J. 2007. V. 45. № 12. P. 2861–2871. https://doi.org/10.2514/1.29796

57. *Karabasov S.A., Berloff P.S., Goloviznin V.M.* Cabaret in the ocean gyres // Ocean Modelling. 2009. V. 30. № 2. P. 155–168. https://doi.org/10.1016/j.ocemod.2009.06.009

58. *Головизнин В.М., Исаков В.А.* Применение балансно-характеристической схемы для решения уравнений мелкой воды над неровным дном // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 7. С. 1142–1160. https://doi.org/10.1134/S0965542517070089

59. *Ковыркина О.А., Остапенко В.В.* О монотонности двухслойной по времени схемы Кабаре // Матем. моделирование 2012. Т. 24. № 9. С. 97–112. https://doi.org/10.1134/S2070048213020051

60. *Ковыркина О.А., Остапенко В.В.* О монотонности схемы КАБАРЕ, аппроксимирующей гиперболическое уравнение со знакопеременным характеристическим полем // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 5. С. 796–815. https://doi.org/10.1134/S0965542516050122

61. *Ковыркина О.А., Остапенко В.В.* О монотонности схемы КАБАРЕ в многомерном случае // Докл. АН. 2015. Т. 462. № 4. С. 385–390. https://doi.org/10.1134/S1064562415030217

62. *Зюзина Н.А., Остапенко В.В.* О монотонности схемы КАБАРЕ, аппроксимирующей скалярный закон сохранения с выпуклым потоком // Докл. АН. 2016. Т. 466. № 5. С. 513–517. https://doi.org/10.1134/S1064562416010282

63. *Зюзина Н.А., Остапенко В.В.* Монотонная аппроксимация схемой КАБАРЕ скалярного закона сохранения в случае знакопеременного характеристического поля // Докл. АН. 2016. Т. 470. № 4. С. 375–379. https://doi.org/10.1134/S1064562416050185

64. *Остапенко В.В., Черевко А.А.* Применение схемы КАБАРЕ для расчета разрывных решений скалярного закона сохранения с невыпуклым потоком // Докл. АН. 2017. Т. 476. № 5. С. 518–522. https://doi.org/10.1134/S1028335817100056

65. *Зюзина Н.А., Остапенко В.В., Полунина Е.И.* Метод расщепления при аппроксимации схемой CABARET неоднородного скалярного закона сохранения // Сиб. журн. вычисл. матем. 2018. Т. 21. № 2. С. 185–200. https://doi.org/10.1134/S1995423918020052

66. *Reed W.H., Hill T.R.* Triangular mesh methods for the neutron transport equation // Los Alamos Scientific Laboratory Report LA-UR-73-79, 1973. USA. https://www. osti.gov/servlets/purl/4491151

67. *Arnold D.N., Brezzi F., Cockburn B., Marini L.D.* Unified analysis of discontinuous Galerkin methods for elliptic problems // SIAM J. Numer. Analys. 2002. V. 39. № 5. P. 1749–1779. https://doi.org/10.1137/S0036142901384162

68. *Peraire J., Persson P.O.* High-order discontinuous Galerkin methods for CFD // Adv. in CFD: Adaptive High-Order Meth. Comput. Fluid Dyn. 2011. V. 2. P. 119–152. https://doi.org/10.1142/9789814313193_0005

69. *Ладонкина М.Е., Неклюдова О.А., Тишкин В.Ф.* Использование разрывного метода Галеркина при решении задач газовой динамики // Матем. моделирование. 2014. Т. 26. № 1. С. 17–32. https://doi.org/10.1134/S207004821404005X

70. *Shu C.W.* High order WENO and DG methods for time-dependent convection-dominated PDEs: A brief survey of several recent developments // J. Comput. Phys. 2016. V. 316. P. 598–613. https://doi.org/10.1016/j.jcp.2016.04.030

71. *Волков А.В.* Особенности применения метода Галеркина к решению пространственных уравнений Навье–Стокса на неструктурированных гексаэдральных сетках // Уч. записки ЦАГИ. 2009. Т. 40. № 6. С. 41–59. https://www.elibrary.ru/item.asp?id_065602

72. *Yasue K., Furudate M., Ohnishi N., Sawada K.* Implicit discontinuous Galerkin method for RANS simulation utilizing pointwise relaxation algorithm // Commun. Comput. Phys. 2010. V. 7. № 3. P. 510–533. https://doi.org/10.4208/cicp.2009.09.055

73. *Dumbser M.* Arbitrary high order $P_N P_M$ schemes on unstructured meshes for the compressible Navier–Stokes equations // Comput. Fluid. 2010. V. 39. № 1. P. 60–76. https://doi.org/10.1016/j.compfluid.2009.07.003

74. *Краснов М.М., Кучугов П.А., Ладонкина М.Е., Тишкин В.Ф.* Разрывный метод Галеркина на трехмерных тетраэдральных сетках. Использование операторного метода программирования // Матем. моделир. 2017. Т. 29. № 2. С. 3–22. https://doi.org/10.1134/S2070048217050064

75. *Luo H., Baum J.D. and Lohner R.* Fast p-multigrid discontinuous Galerkin method for compressible flow at all speeds // AIAA J. 2008. V. 46. № 3. P. 635–652. https://doi.org/10.2514/1.28314

76. *Luo H., Baum J.D., Lohner R.A.* Hermite WENO-based limiter for discontinuous Galerkin method on unstructured grids // J. Comput. Phys. 2007. V. 225. P. 686–713. https://doi.org/10.1016/j.jcp.2006.12.017

77. *Zhu J., Zhong X., Shu C.W., Qiu J.* Runge-Kutta discontinuous Galerkin method using a new type of WENO limiters on unstructured meshes // J. Comput. Phys. 2013. V. 248. P. 200–220. https://doi.org/10.1016/j.jcp.2013.04.012

78. *Волков А.В., Ляпунов С.В.* Монотонизация метода конечного элемента в задачах газовой динамики // Уч. записки ЦАГИ. 2009. Т. 40. № 4. С. 15–28. https://www.elibrary.ru/item.asp?id_904664

79. *Krivodonova L., Xin J., Remacle J.F., Chevogeon N., Flaherty J.* Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws // Appl. Numer. Math. 2004. V. 48. № 3. P. 323–338. https://doi.org/10.1016/j.apnum.2003.11.002

80. *Krivodonova L.* Limiters for high-order discontinuous Galerkin methods // J. Comput. Phys. 2007. V. 226. № 1. P. 879–896. https://doi.org/10.1016/j.jcp.2007.05.011

81. *Ладонкина М.Е., Неклюдова О.А., Тишкин В.Ф.* Построение лимитера для разрывного метода Галеркина на основе усреднения решения // Матем. моделирование. 2018. Т. 30. № 5. С. 99–116. https://doi.org/10.1134/S2070048219010101

82. *Mocz P., Vogelsberger M., Sijacki D., Pakmor R., Hernquist L.* A discontinuous Galerkin method for solving the fluid and MHD equations in astrophysical simulations // Mon. Not. R. Astron. Soc. 2014. V. 437. № 1. P. 397–414. https://doi.org/10.1093/mnras/stt1890

83. *Klockner A., Warburton T., Hesthaven J.S.* Nodal discontinuous Galerkin methods on graphics processors // J. Comput. Phys. 2009. V. 228. № 21. P. 7863–7882. https://doi.org/10.1016/j.jcp.2009.06.041

84. *Chan J., Wang Z., Modave A., Remacle J., Warburton T.* GPU-accelerated discontinuous Galerkin methods on hybrid meshes // J. Comput. Phys. 2016. V. 318. P. 142–168. https://doi.org/10.1016/j.jcp.2016.04.003

85. *Краснов М.М., Ладонкина М.Е.* Разрывный метод Галеркина на трехмерных тетраэдральных сетках. Применение шаблонного метапрограммирования языка C++ // Программирование. 2017. № 3. С. 41–53.

86. *Rogov B.V.* Dispersive and dissipative properties of the fully discrete bicompact schemes of the fourth order of spatial approximation for hyperbolic equations // Appl. Numer. Math. 2019. V. 139. P. 136–155. https://doi.org/10.1016/j.apnum.2019.01.008

87. *Chikitkin A.V., Rogov B.V.* Family of central bicompact schemes with spectral resolution property for hyperbolic equations // Appl. Numer. Math. 2019. V. 142. P. 151–170. https://doi.org/10.1016/j.apnum.2019.03.007

88. *Рогов Б.В., Михайловская М.Н.* О сходимости компактных разностных схем // Матем. моделирование. 2008. Т. 20. № 1. С. 99–116. https://doi.org/10.1134/S2070048209010104

89. *Михайловская М.Н., Рогов Б.В.* Монотонные компактные схемы бегущего счета для систем уравнений гиперболического типа // Ж. вычисл. матем. и матем. физ. 2012. Т. 52. № 4. С. 672–695. https://doi.org/10.1134/S0965542512040124

90. *Рогов Б.В.* Высокоточная монотонная компактная схема бегущего счета для многомерных уравнений гиперболического типа // Ж. вычисл. матем. и матем. физ. 2013. Т. 53. № 2. С. 264–274. https://doi.org/10.1134/S0965542513020097

91. *Chikitkin A.V., Rogov B.V., Utyuzhnikov S.V.* High-order accurate monotone compact running scheme for multidimensional hyperbolic equations // Appl. Numer. Math. 2015. V. 93. P. 150–163. https://doi.org/10.1016/j.apnum.2014.02.008

92. *Брагин М.Д., Рогов Б.В.* Высокоточные бикомпактные схемы для численного моделирования течений многокомпонентных газов с несколькими химическими реакциями // Матем. моделирование. 2020. Т. 32. № 6. С. 21–36. https://doi.org/10.1134/S2070048221010063

93. *Брагин М.Д., Рогов Б.В.* Бикомпактные схемы для многомерного уравнения конвекции-диффузии // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 4. С. 625–643. https://doi.org/10.1134/S0965542521040023

94. *Брагин М.Д., Рогов Б.В.* О точности бикомпактных схем в задаче о распаде вихря Тейлора–Грина // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 11. С. 1759–1778. https://doi.org/10.1134/S0965542521110051

95. *Брагин М.Д., Рогов Б.В.* О точном пространственном расщеплении многомерного скалярного квазилинейного гиперболического закона сохранения // Докл. АН. 2016. Т. 469. № 2. С. 143–147. https://doi.org/10.1134/S1064562416040086

96. *Брагин М.Д., Рогов Б.В.* Метод итерируемой приближенной факторизации операторов высокоточной бикомпактной схемы для систем многомерных неоднородных уравнений гиперболического типа // Докл. АН. 2017. Т. 473. № 3. С. 263–267. https://doi.org/10.1134/S1064562417020107

97. *Брагин М.Д., Рогов Б.В.* Комбинированная многомерная бикомпактная схема, имеющая повышенную точность в областях влияния нестационарных ударных волн // Докл. АН. 2020. Т. 494. С. 9–13. https://doi.org/10.1134/S1064562420050282

98. *Alexander R.* Diagonally implicit Runge–Kutta methods for stiff O.D.E.'s // SIAM J. Numer. Anal. 1977. V. 14. № 6. P. 1006–1021. http://www.jstor.org/stable/2156678

99. *Остапенко В.В., Хандеева Н.А.* О точности разностных схем при расчете взаимодействия ударных волн // Докл. АН. 2019. Т. 485. № 6. С. 40–45. https://doi.org/10.1134/S1028335819040128

EDN: GANDKY
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1804–1821

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 512.643.8

ПОИСК РАЗРЕЖЕННЫХ РЕШЕНИЙ ДЛЯ СВЕРХБОЛЬШИХ СИСТЕМ, ОБЛАДАЮЩИХ ТЕНЗОРНОЙ СТРУКТУРОЙ[1)]

© 2022 г. Д. А. Желтков[1,*], Н. Л. Замарашкин[1,], С. В. Морозов[1,***]**

[1] *119333 Москва, ул. Губкина, 8, Институт вычислительной математики им. Г.И. Марчука РАН, Россия*

**e-mail: dmitry.zheltkov@gmail.com*

***e-mail: nikolai.zamarashkin@gmail.com*

****e-mail: stanis-morozov@yandex.ru*

Поступила в редакцию 30.12.2021 г.
Переработанный вариант 06.06.2022 г.
Принята к публикации 07.07.2022 г.

Задача поиска разреженного решения для больших систем линейных уравнений возникает во многих приложениях, связанных с обработкой сигналов. В некоторых случаях размеры возникающих систем оказываются столь велики, что известные методы становятся неэффективными. Решение таких систем возможно только при наличии в них дополнительной структуры. В настоящей работе предлагается эффективный метод поиска разреженных решений сверхбольших систем линейных уравнений, обладающих тензорной структурой определенного вида. Приведенный теоретический анализ и экспериментальные результаты позволяют судить об эффективности предложенного метода. Библ. 14. Фиг. 7.

Ключевые слова: метод наименьших квадратов, разреженное решение, тензорная структура оператора.

DOI: 10.31857/S0044466922110151

1. ВВЕДЕНИЕ

Восстановление разреженного сигнала по небольшому числу линейных измерений (возможно, искаженных аддитивным шумом) является фундаментальной задачей в обработке сигналов. Остановимся на следующей модели

$$y = X\beta + \varepsilon,$$

где $\beta \in \mathbb{R}^n$ – сигнал, $y \in \mathbb{R}^m$ – вектор наблюдений, $X \in \mathbb{R}^{m\times n}$ – матрица измерений, а $\varepsilon \in \mathbb{R}^m$ – вектор шума. В некоторых прикладных задачах интерес представляет ситуация, когда размерность сигнала n велика и существенно превосходит число измерений m (т.е. $n \gg m$), а вектор сигнала β *разреженный* (имеет малое число ненулевых компонент).

Обозначим множество ненулевых компонент вектора $\beta \in \mathbb{R}^n$ через

$$\operatorname{supp}\beta = \{i : \beta_i \neq 0\}$$

и будем искать k-разреженное решение β задачи наименьших квадратов

$$\|y - X\beta\|_2 \to \min, \quad \|\beta\|_0 = |\operatorname{supp}\beta| \leq k. \tag{1}$$

Среди методов, используемых для решения (1), наиболее популярным является алгоритм OMP (Orthogonal Matching Pursuit [1]). Сложность OMP оценивается как $\mathbb{O}(mnk)$. Это приемлемо при n, сравнимых с 10^6, но уже недостаточно эффективно при n порядка 10^9.

В данной работе основное внимание уделяется решению сверхбольших систем линейных уравнений, в которых n имеет порядок 10^{13} и более. Очевидно, что при таких размерах линейной

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

системы требуются дополнительные предположения о структуре задачи. В работе таких предположений три:

1) пространство параметров тензоризовано $\mathbb{R}^n = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \mathbb{R}^{n_d}$, $n = n_1 \times n_2 \times \cdots \times n_d$, а действие матрицы X на пространстве $\mathbb{R}^n$ соответствует действию линейного оператора $\mathcal{A}$ на тензоризованном пространстве. Относительно структуры оператора $\mathcal{A}$ будем предполагать, что он допускает возможность эффективного применения к тензорам ранга 1;

2) нормальное решение задачи наименьших квадратов

$$v = \arg\min_u \|\mathcal{A}u - y\|_2$$

может быть приближено в $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$ тензорами малого тензорного ранга;

3) строки матрицы измерений задают отображение ограниченной изометрии из пространства размерности m в пространство размерности n

$$(1-\delta)(h,l) < \left(\mathcal{A}^{\mathrm{T}}h, \mathcal{A}^{\mathrm{T}}l\right) < (1+\delta)(h,l), \quad \text{где} \quad h,l \in \mathbb{R}^m. \tag{2}$$

В работе изучается метод, позволяющий при изложенных выше предположениях находить разреженное решение для систем со сверхбольшой размерностью пространства параметров сигнала n. Новый метод является адаптацией метода OMP, в которой учитывается структура задачи. По сути, предлагается новый алгоритм выбора существенных (ненулевых) компонент разреженного решения. Реализованная идея выглядит несколько парадоксально. Поиск существенных компонент разреженного решения (1) осуществляется с помощью анализа компонент нормального решения. На первый взгляд может показаться, что в алгоритме происходит замена задачи меньшей вычислительной сложности на задачу большей вычислительной сложности. Это, как будет показано далее, не так, если допустить существование тензорной структуры в нормальном решении и свойства квазиизометрии оператора измерений $\mathcal{A}$.

Оставшаяся часть текста организована следующим образом. В разд. 2 рассмотрен OMP алгоритм поиска разреженных решений. В разд. 3 дается описание предлагаемого метода, который мы называем *тензоризованным OMP*. В разд. 4 приведен теоретический анализ различных этапов нового метода. Разд. 5 содержит результаты численных экспериментов, сравнение работы классического и тензоризованного алгоритмов, а также ряд дополнительных эвристик, позволяющих повысить качество работы предлагаемого метода. Разд. 6 содержит заключительные замечания.

2. АЛГОРИТМ OMP

Одним из наиболее эффективных методов поиска разреженных решений для задачи наименьших квадратов является Orthogonal Matching Pursuit (OMP) алгоритм [1]. Это жадная итерационная процедура [2], получившая значительное внимание благодаря простоте записи алгоритма и его эффективности в большом числе реальных приложений [3]–[6]. На каждом шаге OMP увеличивает список выбранных компонент на одну так, чтобы добиться наибольшего падения невязки. Стандарное описание OMP метода приводится ниже в алгоритме 1.

Алгоритм 1. Orthogonal Matching Pursuit

Вход: матрица $A \in \mathbb{R}^{m\times n}$, вектор правой части $b \in \mathbb{R}^m$, порог ошибки ε_0.

Инициализация:

- $k = 0$;
- $x^0 = 0$;
- $r^0 = b - Ax^0 = b$;
- множество выбранных столбцов $S^0 = \emptyset$;

while $\left\|r^k\right\|_2 \geq \varepsilon_0$ **do**

1. Вычислить ошибки для каждого из столбцов $\epsilon(j) = \min\limits_{z_j} \left\|a_j z_j - r^k\right\|_2^2$ для всех j. Решением задачи минимизации является $z_j^* = \dfrac{a_j^{\mathrm{T}} r^k}{\|a_j\|_2^2}$.
2. Найти столбец j_0, сильнее всего уменьшающий невязку, а именно $j_0 : \forall j \notin S^k$, $\epsilon(j_0) \le \epsilon(j)$. Затем обновляем множество выбранных элементов: $S^{k+1} = S^k \cup \{j_0\}$.
3. Найти оптимальные коэффициенты при столбцах из S^{k+1}, т.е. вычислить x^{k+1}, минимизирующий $\|Ax - b\|_2^2$ при условии $\operatorname{supp} x^{k+1} = S^{k+1}$.
4. Вычислить невязку $r^{k+1} = b - Ax^{k+1}$.
5. $k = k + 1$.

end while

return S^k

Сложность итерации OMP оценивается через $\mathbb{O}(mn)$ арифметических операций. При размерах n, по порядку больших 10^9, алгоритм становится неэффективным. Цель данной работы адаптировать OMP алгоритм на случай больших n, при условии, что пространство $\mathbb{R}^n$ тензоризовано, а в операторе $\mathcal{A}$ системы и нормальном решении задачи наименьших квадратов присутствует тензорная структура.

3. ОБЩАЯ СХЕМА ТЕНЗОРИЗОВАННОГО OMP

Рассмотрим линейный оператор

$$\mathcal{A} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_d} \to \mathbb{R}^m.$$

Будем считать, что столбцы матрицы оператора $\mathcal{A}$ имеют одинаковую длину. От оператора потребуем возможности быстрого умножения на векторы, представимые тензорами ранга 1, т.е. на векторы вида $v_1 \otimes v_2 \otimes \ldots \otimes v_d$, где $v_i \in \mathbb{R}^{n_i}$ (здесь $\otimes$ обозначает кронекерово произведение).

Кроме того, предположим, что нормальное решение задачи наименьших квадратов

$$v = \arg\min_u \|\mathcal{A}u - y\|_2$$

приближается (возможно довольно грубо) тензором малого ранга. Воспользуемся этим предположением, чтобы снизить сложность стандартного OMP.

Начнем со следующего простого наблюдения: наиболее трудоемкий этап OMP относится к вычислению невязок

$$\epsilon(j) = \min_{z_j} \|a_j z_j - r\|_2 \tag{3}$$

для всех столбцов $\mathcal{A}$. Действительно, для j-го столбца оптимальное значение z_j^* дается выражением $z_j^* = \dfrac{a_j^{\mathrm{T}} r}{\|a_j\|_2^2}$, где r обозначает текущий вектор невязки. Прямое вычисление z_j^* имеет сложность $\mathbb{O}(m)$. Всего таких вычислений n. Таким образом, только прямое вычисление всех z_j^* уже имеет сложность $\mathbb{O}(mn)$.

Поскольку $a_j z_j^*$ и $a_j z_j^* - r$ ортогональны, то решение задачи (3) эквивалентно поиску максимума среди величин

$$c_j = \frac{(a_j, y)}{\|a_j\|_2}$$

с последующим выбором столбца с максимальным c_j. Учитывая, что все длины столбцов $\|a_j\|_2$ совпадают, наилучшая компонента на текущей итерации OMP дается максимумом модуля скалярного произведения $|(a_j, y)|$. Таким образом, снизить алгоритмическую сложность OMP возможно, предъявив алгоритм малой сложности для вычисления величин $|(a_j, y)|$.

Сделаем элементарное, но важное наблюдение. Для оператора $\mathcal{A}^{\mathrm{T}}$, обладающего свойством ограниченной изометрии (см. предположение 3), оценки на величины $|(a_j, y)|$ можно получить, рассматривая нормальное решение задачи наименьших квадратов. Действительно, нормальное решение $\tilde{u}$ представляется в виде

$$\tilde{u} = \mathcal{A}^{\dagger} y = \mathcal{A}^{\mathrm{T}} \left(\mathcal{A}\mathcal{A}^{\mathrm{T}}\right)^{-1} y. \tag{4}$$

Тогда для компоненты $\tilde{u}_i$

$$|\tilde{u}_i| = \left|a_i^{\mathrm{T}} (\mathcal{A}\mathcal{A}^{\mathrm{T}})^{-1} y\right| = \left|\left(a_i, (\mathcal{A}\mathcal{A}^{\mathrm{T}})^{-1} y\right)\right| \le \frac{1}{1-\delta}\left|\left(\mathcal{A}^{\mathrm{T}} a_i, \mathcal{A}^{\mathrm{T}} (\mathcal{A}\mathcal{A}^{\mathrm{T}})^{-1} y\right)\right| = \frac{1}{1-\delta}|(a_i, y)|.$$

Аналогично получается оценка снизу

$$|\tilde{u}_i| \ge \frac{1}{1+\delta}|(a_i, y)|.$$

Имея в виду полученные оценки, будем выбирать существенные компоненты в OMP на основе модуля компонент нормального решения задачи наименьших квадратов.

На первый взгляд делается парадоксальный шаг. Действительно, поиск нормального решения задачи наименьших квадратов превосходит по сложности OMP алгоритм. Напомним, однако, что нами сделано предположение о тензорной структуре нормального решения. В этом случае, как будет показано далее, существует эффективный алгоритм, который является незначительной модификацией алгоритма ALS.

Остается еще вопрос о быстром поиске больших компонент в полученном приближении к нормальному решению. Для тензорных представлений ответ хорошо известен, а соответствующие алгоритмы описаны, например в [7], [8]. Однако в наших численных экспериментах тензорный ранг T всегда выбирался равным 1. В этом случае выбор C максимальных элементов тензора может быть еще упрощен. А именно, пусть после решения задачи наименьших квадратов для тензора v_T получается представление в виде

$$v_T = u_1 \otimes u_2 \otimes \ldots \otimes u_d.$$

Выберем из u_1 не более C максимальных по модулю элементов и обозначим получившееся множество через G_1. Затем перемножим попарно все элементы из G_1 с элементами из u_2 и выберем не более C максимальных по модулю элементов, а полученное множество обозначим через G_2. Продолжая процедуру, в конце получим множество G_d, содержащее C максимальных по модулю элементов тензора v_T. Сложность такого поиска при простейшей реализации составляет $\mathbb{O}(C(n_1 + \ldots + n_d + d\log C))$.

Основываясь на сказанном, предлагается следующий порядок вычислений в тензоризованном алгоритме 2.

Algorithm 2. Тензоризованный OMP

Вход: тензоризованный линейный оператор $\mathcal{A}$, вектор правой части b, количество компонент разреженного решения K, ранг тензора для малоранговой аппроксимации T, количество кандидатов C.

Инициализация:

- $x^0 = 0$;
- $r^0 = b - \mathcal{A}x^0 = b$;
- множество выбранных столбцов $S^0 = \emptyset$;
- множество столбцов для ортогонализации $Q^0 = \emptyset$;

for $k = 0,\ldots,K-1$ **do**

1. Вычислить решение ранга T для решения задачи наименьших квадратов

$$v_T = \underset{\substack{u\in\mathbb{R}^{n_1\times\ldots\times n_d}:\\ \operatorname{rank} u\le T}}{\arg\min} \left\|Q^k\mathcal{A}u - Q^k b\right\|_2.$$

Для простоты мы обозначили Q^k и множество столбцов, и матрицу ортогонализации к этим столбцам.

2. Найти C элементов с максимальными по модулю элементами в тензоре v_T. Обозначим полученное множество $\hat{S}$.

3. Построим матрицу $\hat{A}$, состоящую из столбцов с номерами из $\hat{S}$; i-й столбец может быть получен действием оператора $\mathcal{A}$ на единичный вектор.

4. Применим один шаг метода OMP к задаче с матрицей $Q^k\hat{A}$ и правой частью $Q^k b$. В результате получим некоторый номер столбца j_0 для матрицы исходного оператора $\mathcal{A}$.

5. Добавим столбец j_0 в множество выбранных столбцов: $S^{k+1} = S^k \cup \{j_0\}$, а также столбец $Q^k a_{j_0}$ в множество столбцов для ортогонализации: $Q^{k+1} = Q^k \cup \{Q^k a_{j_0}\}$.

end for

return S^K

Остается вопрос об эффективном поиске решения малого тензорного ранга для ортогонализованной задачи наименьших квадратов

$$v_T = \underset{\substack{u\in\mathbb{R}^{n_1\times\ldots\times n_d}:\\ \operatorname{rank} u\le T}}{\arg\min} \left\|Q^k\mathcal{A}u - Q^k b\right\|_2.$$

Для этого предлагается использовать простую модификацию алгоритма ALS. Действительно, для решаемой задачи

$$\left\|Q\mathcal{A}\left(\sum_{t=1}^{\mathrm{T}} u_{1t}\otimes u_{2t}\otimes\ldots\otimes u_{dt}\right) - Qb\right\|_2 \to \min. \tag{5}$$

выберем некоторое i от 1 до d и фиксируем все u_{jt} при $j \ne i$. Решим задачу относительно u_{it}. Пусть $U_i = \left[u_{i1}^{\mathrm{T}}\ u_{i2}^{\mathrm{T}}\ \ldots\ u_{iT}^{\mathrm{T}}\right]^{\mathrm{T}}$, перепишем (5) в виде

$$\left\|QM_iU_i - Qb\right\|_2 \to \min,$$

где $M_i \in \mathbb{R}^{m\times Tn_i}$ размерность новой задачи наименьших квадратов. Для получения решения алгоритм ALS многократно итерирует по всем i от 1 до d.

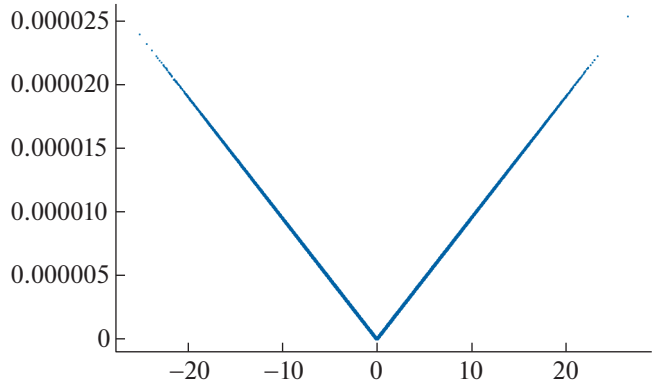
Заметим также, что наличие матрицы ортогонализации Q не существенно влияет на оценку сложности. Действительно, Q имеет вид

$$Q = (I - w_1w_1^{\mathrm{T}})(I - w_2w_2^{\mathrm{T}})\ldots(I - w_kw_k^{\mathrm{T}}),$$

где k соответствует количеству векторов, к которым происходит ортогонализация. Таким образом, умножение на Q соответствует последовательности из k одноранговых преобразований и имеет сложность $\mathbb{O}(kmTn_i)$, которая сопоставима со сложностью решения задачи наименьших квадратов $\mathbb{O}(mT^2n_i^2)$. Таким образом, решение ортогонализованной задачи оказывается ненамного сложнее классического алгоритма ALS для нахождения малорангового решения.

Пусть алгоритм совершает N_{ALS} полных итераций (т.е. при которых i пробегает от 1 до d). Тогда полная сложность поиска решения малого ранга равна

$$\mathbb{O}\left(mN_{ALS}\left(kT\sum_{i=1}^{d} n_i + T^2\sum_{i=1}^{d} n_i^2\right)\right).$$



Фиг. 1. График зависимости модуля компонент решения $|v_j|$ от скалярного произведения (a_j, y) для случайной матрицы из нормального распределения. Размер матрицы 20×2^{20}.

Из приведенных рассуждений легко видеть, что алгоритм имеет полиномиальную сложность относительно параметров $m, n_1, \ldots, n_d$. Более того, сложность линейна по m.

4. ТЕОРЕТИЧЕСКИЙ АНАЛИЗ

В этом разделе мы не ставим целью доказать строгие результаты о сходимости метода тензоризованного OMP. Мы лишь предполагаем дать более глубокое понимание различных шагов алгоритма.

Тензоризованное решение задачи наименьших квадратов малого тензорного ранга представляет собой довольно сложный для анализа объект. Кроме того, согласно одному из базовых предположений, нормальное решение задачи наименьших квадратов приближается тензором малого ранга. Поэтому вместо анализа решений малого ранга начнем с анализа нормального решения общей задачи наименьших квадратов для случайных матриц. Для достаточно богатого класса случайных матриц, в частности, матриц из нормального распределения, можно показать наличие свойства ограниченной изометрии [9]. Как было показано выше, для матриц, обладающих свойством ограниченной изометрии, нормальное решение задачи наименьших квадратов

$$v = \arg\min_u \|\mathcal{A}u - y\|_2$$

является хорошей оценкой на величину

$$c_j = \frac{(a_j, y)}{\|a_j\|_2}.$$

В качестве подтверждения этому был проведен численный эксперимент. Были сгенерированы случайная матрица A из нормального распределения размера $m \times 2^m$ при $m = 20$ и случайный вектор правой части y. Для каждого $j = 1, \ldots, 2^m$ были вычислены значение v_j компоненты решения задачи наименьших квадратов и величина (a_j, y). На фиг. 1 изображены пары точек $((a_j, y), |v_j|)$. Отсюда легко видеть, что компоненты $|v_j|$ являются хорошими оценками на величины $|c_j|$ при малых значениях δ.

4.1. Влияние ортогонализации

Опишем, что изменится, если вместо задачи наименьших квадратов для $Ax = y$ решается задача

$$\|QAx - Qy\|_2 \to \min.$$

Рассуждения этого пункта верны для любых матриц. Пусть имеются матрица $A \in \mathbb{R}^{m\times n}$ и вектор правой части y, и на первом шаге метода мы некоторым образом нашли вектор $a^{(1)}$ и ортогонализуем к нему матрицу. Матрица ортогонализации может быть представлена в виде

$$Q_1 = I - a^{(1)}(a^{(1)})^{\mathrm{T}} / \left\|a^{(1)}\right\|_2^2.$$

Таким образом, решается система $Q_1 A = Q_1 y$. Пусть далее мы нашли столбец $a^{(2)}$ и ортогонализуем систему к нему. Поскольку $a^{(2)}$ выбирался из столбцов матрицы $Q_1 A$, векторы $a^{(1)}$ и $a^{(2)}$ будут ортогональны. Матрица Q_2 строится аналогичным образом:

$$Q_2 = I - a^{(2)}(a^{(2)})^{\mathrm{T}} / \left\|a^{(2)}\right\|_2^2$$

и решается система $Q_2 Q_1 A = Q_2 Q_1 y$. Верна следующая

Теорема 1. *Пусть $A \in \mathbb{R}^{m\times n}$ и $y \in \mathbb{R}^m$. Пусть $Q_1, Q_2, \ldots, Q_k$ – набор матриц ортогонализации, порожденных попарно ортогональными векторами. В этом случае решения задач наименьших квадратов систем*

$$Q_n Q_{n-1} \ldots Q_1 A = Q_n Q_{n-1} \ldots Q_1 y \tag{6}$$

и

$$A = Q_n Q_{n-1} \ldots Q_1 y \tag{7}$$

совпадают.

Доказательство. Докажем, что псевдообратная матрица к матрице ортогонализации Q_i совпадает с ней самой, т.е. $Q_i^{\dagger} = Q_i$. Легко проверить, что

$$Q_i^2 = Q_i, \quad Q_i^{\mathrm{T}} = Q_i,$$

откуда следуют все аксиомы псевдообратной матрицы. Далее покажем, что матрицы Q_i и Q_j коммутируют. Действительно,

$$Q_i Q_j = \left(I - a^{(i)}\left(a^{(i)}\right)^{\mathrm{T}} / \left\|a^{(i)}\right\|_2^2\right)\left(I - a^{(j)}\left(a^{(j)}\right)^{\mathrm{T}} / \left\|a^{(j)}\right\|_2^2\right) =$$
$$= I - a^{(i)}\left(a^{(i)}\right)^{\mathrm{T}} / \left\|a^{(i)}\right\|_2^2 - a^{(j)}\left(a^{(j)}\right)^{\mathrm{T}} / \left\|a^{(j)}\right\|_2^2 + \frac{1}{\left\|a^{(i)}\right\|_2^2 \left\|a^{(j)}\right\|_2^2} a^{(i)}\left(a^{(i)}\right)^{\mathrm{T}} a^{(j)}\left(a^{(j)}\right)^{\mathrm{T}}.$$

Но последнее слагаемое равно 0 в силу того, что $(a^{(i)})^{\mathrm{T}} a^{(j)} = 0$, откуда видно, что матрицы Q_i и Q_j коммутируют.

Решение системы (6) представляется в виде

$$\begin{aligned}(Q_n Q_{n-1} \ldots Q_1 A)^{\dagger} Q_n Q_{n-1} \ldots Q_1 y = A^{\dagger} Q_1^{\dagger} Q_2^{\dagger} \ldots Q_n^{\dagger} Q_n Q_{n-1} \ldots Q_1 y = \\ = A^{\dagger} Q_1 Q_2 \ldots Q_n Q_n Q_{n-1} \ldots Q_1 y = A^{\dagger} Q_n^2 Q_{n-1}^2 \ldots Q_1^2 y = A^{\dagger} Q_n Q_{n-1} \ldots Q_1 y,\end{aligned} \tag{8}$$

что в точности совпадает с решением задачи наименьших квадратов (7). Теорема доказана.

Из этого видно, что решения задач

$$\left\|QAx - Qy\right\|_2 \to \min$$

и

$$\left\|Ax - Qy\right\|_2 \to \min$$

совпадают, т.е. достаточно ортогонализовывать только правую часть системы. Отсюда следует, что рассуждения о ранжировании по величине компонент нормального решения верны на любом шаге метода, поскольку опираются на свойство ограниченной изометрии матрицы.

Заметим, что это утверждение верно лишь для нормального решения задачи наименьших квадратов и может оказаться неверным для решения задачи наименьших квадратов малого ранга.

Проведенный анализ показывает, что выбор кандидатов на основе модулей компонент решения задачи наименьших квадратов не ограничивает классического алгоритма OMP в том смысле,

что для широкого класса случайных матриц алгоритм тензоризованного OMP строит множество столбцов, которое не может сильно отличаться от ранжирования по величинам

$$c_j = \frac{(a_j, r)}{\|a_j\|_2}.$$

Однако на практике мы не можем найти точное решение задачи наименьших квадратов для больших систем, поэтому алгоритм тензоризованного OMP выбирает C кандидатов на основе компонент малорангового решения. С учетом предположения, что точное решение задачи наименьших квадратов может быть хорошо приближено тензором малого ранга, это дает возможность надеяться, что в множество из C кандидатов попадет столбец, который бы выбрал алгоритм OMP.

4.2. О падении невязки на различных шагах метода

Изучим вопрос о скорости падения невязки для матрицы A, имеющей вид

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & 1 & 1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & -1 & -1 & \dots & -1 & -1 \\ 1 & -1 & 1 & -1 & \dots & 1 & -1 \end{bmatrix}. \tag{9}$$

Эта матрица содержит всевозможные столбцы из ± 1, взятые в лексикографическом порядке. Легко показать, что эта матрица обладает свойством ограниченной изометрии с $\delta = 0$. Пусть правая часть задана в виде суммы k столбцов матрицы A

$$b = \sum_{j=1}^{k} \alpha_j a_{i_j} + \varepsilon.$$

Пусть произошла ортогонализация к вектору p, где $\|p\|_2 = 1$. Разложим столбцы матрицы A, входящие в правую часть, на компоненты, коллинеарные p и ортогональные ему:

$$a_{i_j} = a_{i_j}^{\parallel} + a_{i_j}^{\perp},$$

откуда $a_{i_j}^{\parallel} = \beta_j p$. Тогда

$$b = \sum_{j=1}^{k} \alpha_j a_{i_j}^{\parallel} + \varepsilon^{\parallel} + \sum_{j=1}^{k} \alpha_j a_{i_j}^{\perp} + \varepsilon^{\perp}.$$

Тогда после ортогонализации

$$b_1 = \left(I - pp^{\mathrm{T}}\right) b = \sum_{j=1}^{k} \alpha_j a_{i_j}^{\parallel} + \varepsilon^{\parallel} + \sum_{j=1}^{k} \alpha_j a_{i_j}^{\perp} + \varepsilon^{\perp} - \left(\sum_{j=1}^{k} \alpha_j (p, a_{i_j}^{\parallel}) + (p, \varepsilon^{\parallel}) \right) p.$$

Откуда, используя $a_{i_j}^{\parallel} = \beta_j p$ и $\varepsilon^{\parallel} = \gamma p$, получаем

$$b_1 = \sum_{j=1}^{k} \alpha_j a_{i_j}^{\perp} + \varepsilon^{\perp}.$$

Добавим и вычтем коллинеарную составляющую

$$b_1 = \sum_{j=1}^{k} \alpha_j a_{i_j}^{\perp} + \varepsilon^{\perp} + \sum_{j=1}^{k} \alpha_j a_{i_j}^{\parallel} + \varepsilon^{\parallel} - \sum_{j=1}^{k} \alpha_j \beta_j p - \gamma p = \sum_{j=1}^{k} \alpha_j a_{i_j} + \tau p + \varepsilon.$$

Заметим, что p с точностью до нормировки является одним из столбцов матрицы A. Таким образом, после шага метода в правой части становится на 1 столбец больше.

Оценим норму вектора правой части b_1 после ортогонализации. По теореме Пифагора имеем

$$\|b\|_2^2 = \|b_1\|_2^2 + \frac{(p,b)^2}{\|p\|_2^2},$$

откуда

$$\frac{\|b_1\|_2^2}{\|b\|_2^2} = 1 - \frac{(p,b)^2}{\|p\|_2^2 \|b\|_2^2},$$

что равносильно

$$\frac{\|b_1\|_2}{\|b\|_2} = |\sin \angle(p,b)|.$$

Оценим для начала худший случай. Выберем нормированный вектор b так, что $\max\limits_p (p,b)^2$ минимально. Поскольку p состоит из ± 1, можно, не ограничивая общности, считать, что все компоненты вектора b неотрицательны и вектор p состоит из всех 1. Тогда нужно найти вектор b такой, что $\|b\|_2 = 1$, все элементы неотрицательны и их сумма минимальна. Очевидно, это вектор $b = e_1$, имеющий единицу в 1 позиции и нули в остальных.

Тогда в худшем случае имеем

$$\frac{\|b_1\|_2^2}{\|b\|_2^2} = 1 - \frac{1}{m}.$$

Заметим, что этот случай вовсе не является нереалистичным:

$$b = \frac{1}{2}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}-1\\1\\\vdots\\1\end{bmatrix}.$$

Однако в среднем все не так плохо. Были проведены следующие численные эксперименты. Для правой части выбиралось k случайных столбцов матрицы и k коэффициентов из нормального распределения. Находилось падение квадрата нормы правой части и результат усреднялся по 10^6 экспериментов. На фиг. 2 приведен график зависимости $1 - \|b_1\|_2^2 / \|b\|_2^2$ от количества столбцов k. Легко видеть, что среднее падение всегда больше 0.5 и очень слабо зависит от числа строк матрицы m.

Оценим величину

$$\max_p \frac{(p,b)^2}{\|p\|^2 \|b\|^2}. \tag{10}$$

Во-первых, ясно, что $\|p\|_2^2 = m$. Пусть

$$b = \sum_{j=1}^{k} \alpha_j a_{i_j}.$$

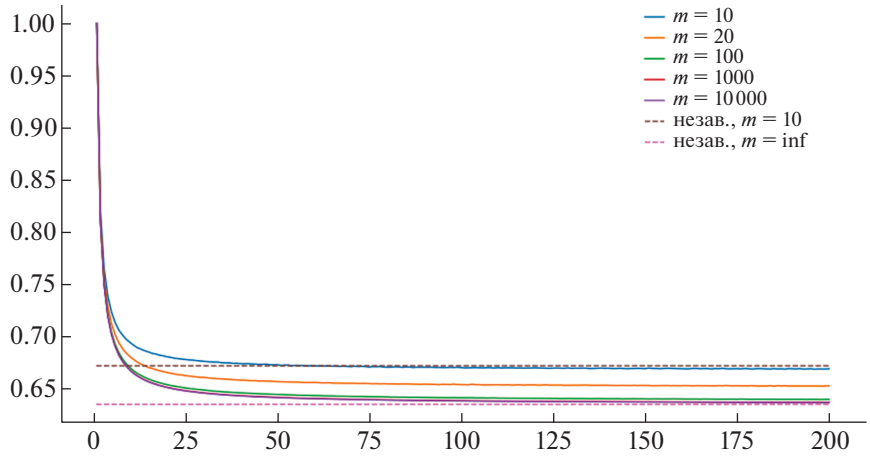
Тогда все компоненты имеют вид

$$b_i = \pm\alpha_1 \pm \alpha_2 \pm \ldots \pm \alpha_k,$$

где $\alpha_j \sim \mathcal{N}(0,1)$, и знаки + и – выбираются с равной вероятностью. Легко видеть, что $\pm\alpha_j \sim \mathcal{N}(0,1)$, следовательно,

$$b_i \sim \mathcal{N}(0,k).$$

Однако компоненты вектора b не являются независимыми.



Фиг. 2. График зависимости квадрата падения нормы невязки от количества компонент в правой части.

Далее, из (10) легко посчитать максимум

$$\max_p \frac{(p,b)^2}{\|p\|^2\|b\|^2} = \frac{\|b\|_1^2}{m\|b\|_2^2}.$$

Нетрудно понять, что, чем больше слагаемых в b (т.е. чем больше k), тем более независимы между собой будут компоненты b_i, и наоборот – чем меньше k, тем более зависимы компоненты. Так, если $k = 1$, то будет всего один вариант $|\alpha_1|$, если $k = 2$, то 2 варианта: $|\alpha_1 + \alpha_2|$ и $|\alpha_1 - \alpha_2|$. В общем случае 2^{k-1} вариантов. Величина $\dfrac{\|b\|_1^2}{\|b\|_2^2}$ будет максимальна, если все компоненты равны, т.е. наиболее зависимы. Из этого наблюдения и из фиг. 2 можно видеть, что с ростом k (по оси x) величина (10) убывает. Отсюда вытекает гипотеза, что оценкой снизу на (10) будет ситуация, когда все компоненты независимы. Доказательством этой гипотезы мы не обладаем, однако, численные эксперименты хорошо согласуются с ней. В случае если все компоненты b_i независимые и одинаково распределенные с $b_i \sim \mathcal{N}(0,1)$ (в случае $b_i \sim \mathcal{N}(0,k)$ можно отнормировать числитель и знаменатель), имеем

$$\mathbb{E}\left(\frac{\|b\|_1^2}{\|b\|_2^2}\right) = 1 + \frac{2}{\pi}(m-1).$$

Для сравнения эти оценки изображены на фиг. 2 штриховыми линиями. Эту оценку можно доказать аналитически.

Действительно,

$$\|b\|_1^2 = (|b_1| + |b_2| + \ldots + |b_m|)^2 = \sum_{j=1}^{m} b_j^2 + \sum_{i,j=1, i\neq j}^{k} |b_i b_j|.$$

Тогда

$$\frac{\|b\|_1^2}{\|b\|_2^2} = 1 + \sum_{i,j=1, i\neq j}^{k} \frac{|b_i b_j|}{b_1^2 + \ldots + b_m^2},$$

откуда имеем, что

$$\mathbb{E}\left(\frac{\|b\|_1^2}{\|b\|_2^2}\right) = 1 + m(m-1)\mathbb{E}\left(\frac{|b_1 b_2|}{b_1^2 + \ldots + b_m^2}\right).$$

Последнее матожидание равно

$$\mathbb{E}\left(\frac{|b_1b_2|}{b_1^2+\ldots+b_m^2}\right)=\frac{1}{(2\pi)^{m/2}}\int\limits_{-\infty}^{\infty}\ldots\int\limits_{-\infty}^{\infty}\frac{|x_1x_2|}{x_1^2+\ldots+x_m^2}\exp\left(-\frac{x_1^2+\ldots+x_m^2}{2}\right)dx_1\ldots dx_m=$$
$$=\frac{2^{m/2}}{\pi^{m/2}}\int\limits_{0}^{\infty}\ldots\int\limits_{0}^{\infty}\frac{x_1x_2}{x_1^2+\ldots+x_m^2}\exp\left(-\frac{x_1^2+\ldots+x_m^2}{2}\right)dx_1\ldots dx_m.$$

Вычислим при $\varepsilon>0$ интеграл

$$\int\limits_{0}^{\infty}\ldots\int\limits_{0}^{\infty}\int\limits_{\varepsilon}^{\infty}\int\limits_{\varepsilon}^{\infty}\frac{x_1x_2}{x_1^2+\ldots+x_m^2}\exp\left(-\frac{x_1^2+\ldots+x_m^2}{2}\right)dx_1\ldots dx_m.$$

Сделаем замену переменных

$$\begin{aligned}y_1&=x_1^2,\\ y_2&=x_2^2,\\ y_3&=x_3,\\ &\vdots\\ y_m&=x_m.\end{aligned}$$

Легко видеть, что якобиан такого преобразования равен $\dfrac{1}{4x_1x_2}$, откуда последний интеграл равен

$$\frac{1}{4}\int\limits_{0}^{\infty}\ldots\int\limits_{0}^{\infty}\int\limits_{\varepsilon^2}^{\infty}\int\limits_{\varepsilon^2}^{\infty}\frac{\exp\left(-\dfrac{y_1+y_2+y_3^2+\ldots+y_m^2}{2}\right)}{y_1+y_2+y_3^2+\ldots+y_m^2}dy_1\ldots dy_m.$$

Сделаем еще одну замену

$$\begin{aligned}t_1&=\frac{y_1+y_2+y_3^2+\ldots+y_m^2}{2},\\ t_2&=y_2,\\ t_3&=y_3,\\ &\vdots\\ t_m&=y_m.\end{aligned}$$

В этом случае якобиан равен 2 и интеграл принимает вид

$$\frac{1}{4}\int\limits_{0}^{\infty}\ldots\int\limits_{0}^{\infty}\int\limits_{\varepsilon^2}^{\infty}\int\limits_{\frac{\varepsilon^2+t_2+t_3^2+\ldots+t_m^2}{2}}^{\infty}\frac{e^{-t_1}}{t_1}dt_1\ldots dt_m. \tag{11}$$

Введем обозначение для экспоненциального интеграла

$$E_n(x)=\int\limits_{1}^{\infty}\frac{e^{-xt}}{t^n}dt.$$

Заметим, что

$$E_1(x)=\int\limits_{1}^{\infty}\frac{e^{-xt}}{t}dt$$

и сделаем замену $y=xt$ при $x>0$. Тогда получим

$$E_1(x)=\int\limits_{x}^{\infty}\frac{e^{-y}}{y}dy.$$

В этих обозначениях интеграл (11) равен

$$\frac{1}{4}\int_0^\infty \dots \int_0^\infty \int_{\varepsilon^2}^\infty E_1\left(\frac{\varepsilon^2 + t_2 + t_3^2 + \dots + t_m^2}{2}\right) dt_2 \dots dt_m.$$

Вычислим производную

$$\left(E_n\left(\frac{x}{2} + c\right)\right)'_x = \int_1^\infty \frac{e^{-\left(\frac{x}{2}+c\right)t}\left(-\frac{t}{2}\right)}{t^n} dt = -\frac{1}{2} E_{n-1}\left(\frac{x}{2} + c\right).$$

$$E_n\left(\frac{x}{2} + c\right) = -2\left(E_{n+1}\left(\frac{x}{2} + c\right)\right)'_x.$$

Посчитаем неопределенный интеграл

$$\begin{aligned}\int E_1\left(\frac{\varepsilon^2 + t_2 + t_3^2 + \dots + t_m^2}{2}\right) dt_2 &= -2\int\left(E_2\left(\frac{\varepsilon^2 + t_2 + t_3^2 + \dots + t_m^2}{2}\right)\right)'_{t_2} dt_2 = \\ &= -2\int_1^\infty \frac{e^{-\frac{\varepsilon^2 + t_2 + t_3^2 + \dots + t_m^2}{2} t_1}}{t_1^2} dt_1 + C.\end{aligned} \tag{12}$$

Откуда имеем, что

$$\int_{\varepsilon^2}^\infty E_1\left(\frac{\varepsilon^2 + t_2 + t_3^2 + \dots + t_m^2}{2}\right) dt_2 = 2\int_1^\infty \frac{e^{-\left(\varepsilon^2 + \frac{t_3^2 + \dots + t_m^2}{2}\right)t_1}}{t_1^2} dt_1.$$

Переходя к пределу при $\varepsilon \to 0$, получаем, что

$$\int_0^\infty \dots \int_0^\infty \frac{x_1 x_2}{x_1^2 + \dots + x_m^2} \exp\left(-\frac{x_1^2 + \dots + x_m^2}{2}\right) dx_1 \dots dx_m =$$

$$= \frac{1}{2}\int_0^\infty \dots \int_0^\infty \int_1^\infty \frac{e^{-\left(\frac{t_3^2 + \dots + t_m^2}{2}\right)t_1}}{t_1^2} dt_1 dt_3 \dots dt_m = \frac{1}{2}\int_1^\infty \frac{1}{t_1^2}\left(\int_0^\infty e^{-\frac{t_3^2 t_1}{2}} dt_3\right) \dots \left(\int_0^\infty e^{-\frac{t_m^2 t_1}{2}} dt_m\right) dt_1.$$

Простой заменой получаем, что

$$\int_0^\infty e^{-\frac{t_3^2 t_1}{2}} dt_3 = \sqrt{\frac{\pi}{2t_1}}.$$

Тогда имеем

$$\int_0^\infty \dots \int_0^\infty \frac{x_1 x_2}{x_1^2 + \dots + x_m^2} \exp\left(-\frac{x_1^2 + \dots + x_m^2}{2}\right) dx_1 \dots dx_m = \frac{\pi^{\frac{m-2}{2}}}{2^{\frac{m}{2}}} \int_1^\infty \frac{dt_1}{t^{\frac{m+2}{2}}} = \frac{\pi^{\frac{m-2}{2}}}{2^{\frac{m-2}{2}} m}.$$

И в итоге получаем

$$\mathbb{E}\frac{|b_1 b_2|}{b_1^2 + \dots + b_m^2} = \frac{2^{\frac{m}{2}}}{\pi^{\frac{m}{2}}} \frac{\pi^{\frac{m-2}{2}}}{2^{\frac{m-2}{2}} m} = \frac{2}{\pi m},$$

$$\mathbb{E}\left(\frac{\|b\|_1^2}{\|b\|_2^2}\right) = 1 + m(m-1)\frac{2}{\pi m} = 1 + \frac{2}{\pi}(m-1),$$

$$\mathbb{E}\left(\max_p \frac{(p,b)^2}{\|b\|^2\|b\|^2}\right) = \frac{2}{\pi} + \left(1-\frac{2}{\pi}\right)\frac{1}{m}.$$

Тогда матожидание падения нормы правой части оказывается равно

$$\mathbb{E}\frac{\|b_1\|_2^2}{\|b\|_2^2} = \left(1-\frac{2}{\pi}\right)\left(1-\frac{1}{m}\right),$$

где b_1 – правая часть после первого шага метода.

Таким образом, для матриц вида (9) на первом шаге невязка будет падать в среднем более чем в 2 раза. Это позволяет надеяться на высокую скорость сходимости метода. Однако такое быстрое падение невязки является особенностью того, что столбцы матрицы (9) очень плотно покрывают все пространство. Для случайных матриц, не обладающих такой особенностью, падение невязки ожидается более медленным.

5. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

В этом разделе приведены результаты численных экспериментов для предложенного метода.

5.1. Описание модели

Модель, рассматриваемая в этом разделе, носит название модель EMP-Вольтерра. Введем обозначения. Пусть $X, Y \in \mathbb{C}^N$ и обозначают входы и выходы модели соответственно. Здесь N обозначает длину сигнала, $P = \left\lceil \frac{\max X_i}{2^K} \right\rceil$, где K является нормировочным параметром модели. Матрица $F \in \mathbb{C}^{N\times(P+1)}$ является некоторой функцией от входа X и считается известной, e обозначает вектор из единиц. Параметрами модели, подлежащими определению, являются $u_{rd} \in \mathbb{C}^{S_X}$, $v_{rd} \in \mathbb{C}^{S_X}$, $w_{rd} \in \mathbb{C}^{S_L}$, $a_{rd} \in \mathbb{C}^{P+1}$, $\alpha_{rd}, \beta_{rd} \in \mathbb{C}$. Тогда модель EMP-Вольтерра может быть записана следующим образом:

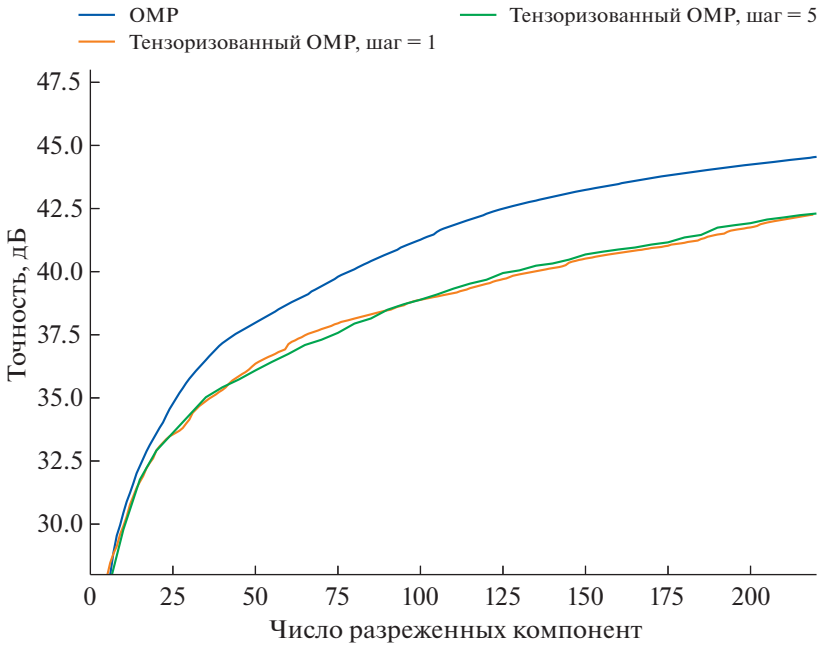
$$Y \approx \sum_{r=1}^{R}\left(\bigodot_{d=1}^{\left\lceil \frac{D_X}{2}\right\rceil}\left(X \circledast u_{rd} + \alpha_{rd}e\right)\right) \odot \left(\bigodot_{d=1}^{\left\lfloor \frac{D_X}{2}\right\rfloor}\left(\bar{X} \circledast v_{rd} + \beta_{rd}e\right)\right) \odot \left(\bigodot_{d=1}^{D_L}\left((F\alpha_{rd}) \circledast w_{rd}\right)\right).$$

Здесь символом $\odot$ обозначено адамарово произведение, а символом $\circledast$ свертка между двумя векторами.

Несложно видеть, что эту модель можно записать как тензоризованный линейный оператор относительно $[u_{rd}, \alpha_{rd}]$, $[v_{rd}, \beta_{rd}]$, w_{rd} и a_{rd}. Число столбцов матрицы такого оператора может быть вычислено по формуле

$$N_{COL} = (S_X+1)^{D_X} \times S_L^{D_L} \times (P+1)^{D_L}.$$

В качестве входного сигнала X и выходного сигнала Y были использованы реальные данные для решения задачи цифрового предыскажения (Digital Predistortion, DPD). Длина сигнала $N \sim 5\times10^4$. Наименьшие размеры модели, которая обеспечивает достаточное для практического применения качество цифрового предыскажения для данного сигнала, равны $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 1$. В этом случае число столбцов получающейся матрицы равно 39,424. Чуть более сложная модель получается выбором $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 3$, что соответствует 233,744,896 столбцам. Максимальный размер задачи, для которой проводились эксперименты, соответствует параметрам $S_X = 13$, $S_L = 13$, $D_X = 3$, $D_L = 3$, где матрица содержит приблизительно 8×10^9 столбцов и 5×10^4 строчек.



Фиг. 3. График сравнения методов OMP и тензоризованного OMP с различным числом шагов в пункте 4 Алгоритма 2. Модель $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 1$. Для метода тензоризованного OMP отбиралось $C = 400$ столбцов.

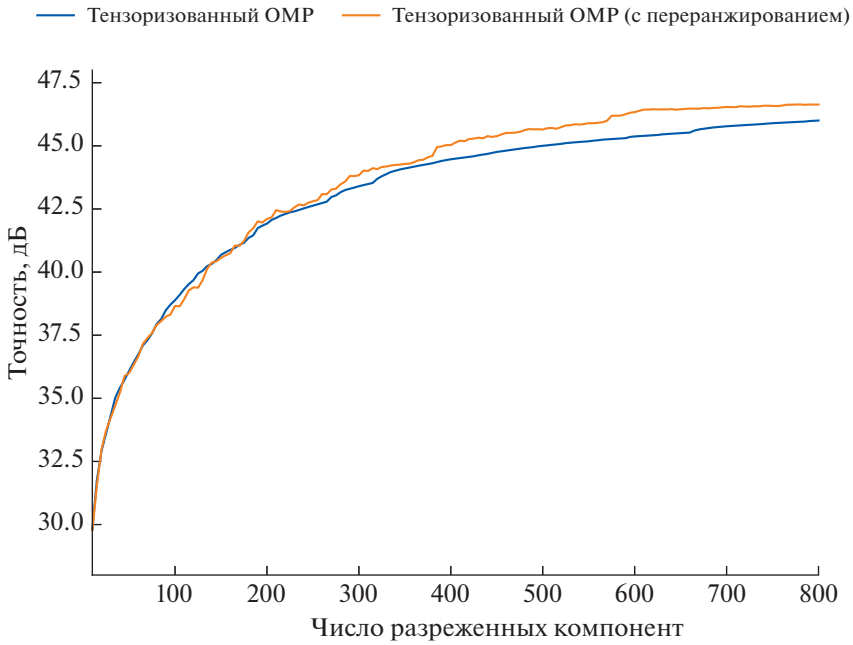
В качестве метрики качества работы была взята классическая для теории обработки сигналов единица измерения – децибелы. Таким образом, точность вычисляется по формуле

$$-20\lg_{10}\frac{\|\mathcal{A}u - Y\|_2}{\|Y\|_2},$$

где u – найденное разреженное решение.

5.2. Результаты

Ниже приведены результаты применения метода тензоризованного OMP к описанной выше модели. На фиг. 3 приведен график сравнения алгоритма OMP и метода тензоризованного OMP для модели с параметрами $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 1$. Здесь и далее алгоритм тензоризованного OMP отбирает $C = 400$ столбцов-кандидатов для множества $\hat{S}$. В качестве метода поиска малорангового решения вместо алгоритма ALS использовался метод Левенберга-Марквардта [10], [11]. Как отмечалось выше, на каждой итерации метода тензоризованного OMP в пункте 4 Алгоритма 2 можно выбирать не 1 столбец, а несколько. Поскольку нахождение малорангового решения занимает большую часть времени работы, такой прием позволяет существенно ускорить время вычислений. Для сравнения на фиг. 3 приведен также график для метода тензоризованного OMP с выбором 5 столбцов на каждой итерации. Из графиков, во-первых, можно видеть, что выбор 5 столбцов вместо одного практически не ухудшает точности построенного решения, но позволяет почти в 5 раз ускорить работу программы. Поэтому во всех дальнейших экспериментах в методе тензоризованного OMP выбирает 5 столбцов за итерацию. Во-вторых, из графиков видно, что метод тензоризованного OMP работает ненамного хуже классического OMP, хотя имеет полиномиальную сложность, в отличие от OMP, имеющего экспоненциальную сложность. На данном этапе качество и не могло получиться лучше, чем у OMP, поскольку метод тензоризованного OMP в некотором смысле является его аппроксимацией, предназначенной для работы со сверхбольшими тензоризованными матрицами.

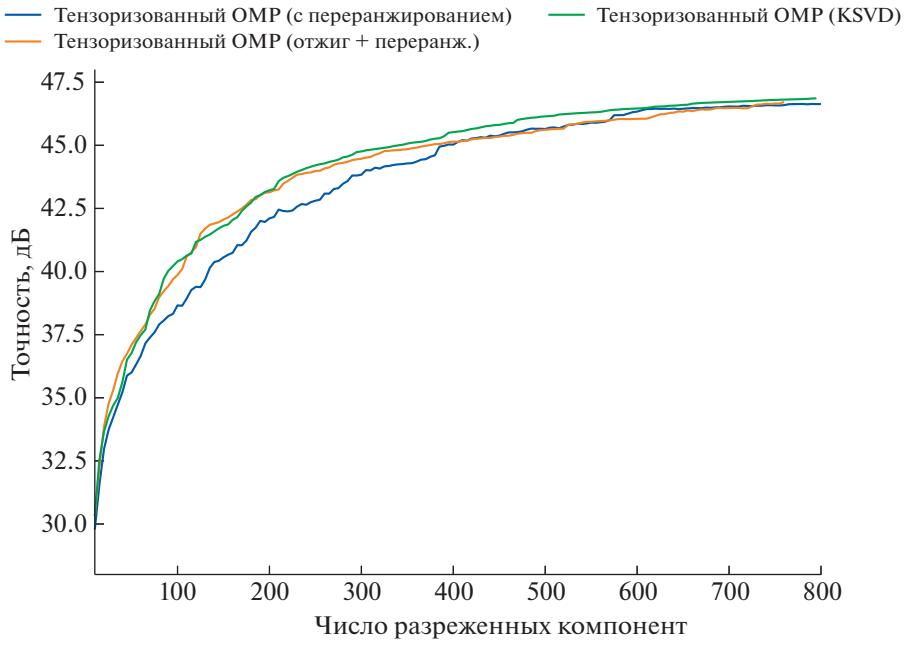


Фиг. 4. График сравнения методов простого тензоризованного OMP и тензоризованного OMP с переранжированием на каждом шаге. Модель $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 1$. Для метода тензоризованного OMP отбиралось $C = 400$ столбцов.

Метод тензоризованного OMP может быть дополнительно улучшен с помощью следующего подхода. На шагах 3–5 Алгоритма 2 будем строить матрицу $\hat{A}$ содержащей не только столбцы из $\hat{S}$, но также и все столбцы из множества уже отобранных ранее столбцов S. Пусть на некоторой итерации Алгоритма 2 уже выбрано множество S. Получим множество новых столбцов-кандидатов $\hat{S}$. Построим матрицу $\hat{A}$, содержащую столбцы и $S \cup \hat{S}$, и отберем из нее (после ортогонализации) $|S| + 5$ столбцов с помощью классического алгоритма OMP. Такой подход гарантированно не ухудшит качества работы, но может помочь найти хорошее решение с меньшим числом разреженных компонент, что можно видеть из фиг. 4.

Заметим, что матрица $Q^k \hat{A}$, для которой применяется OMP в пункте 4 Алгоритма 2 имеет всего $C = 400$ столбцов, что позволяет применять здесь не только жадный алгоритм OMP, но и более продвинутые стратегии оптимизации. В качестве примера были использованы метод имитации отжига и алгоритм K-SVD [12], [13]. Метод имитации отжига представляет собой метод глобальной оптимизации, который на каждом шаге переходит либо в состояние, уменьшающее значение функционала, либо, с некоторой вероятностью, в состояние, ухудшающее значение функционала, причем вероятность последнего зависит от температурного коэффициента, значение которого убывает в процессе работы алгоритма [14]. Алгоритм K-SVD отличается от OMP своей возможностью не только жадно добавлять столбцы, но и выбрасывать ненужные. На фиг. 5 приведены результаты сравнения методов. Алгоритм K-SVD по построению всегда пересматривает множество ранее выбранных столбцов, поэтому для OMP и метода имитации отжига было использовано переранжирование всех столбцов на каждой итерации. Из графиков можно видеть, что на первых итерациях метод имитации отжига выбирает столбцы лучше, поскольку является менее жадным методом, однако затем лидерство переходит к K-SVD, поскольку задача оптимизации становится слишком сложной для метода имитации отжига.

Одной из проблем метода тензоризованного OMP является борьба с влиянием уже выбранных столбцов. Если столбец j уже был выбран ранее, то после ортогонализации он будет нулевым, поэтому коэффициент на j-й позиции может быть выбран любым, и это не скажется на невязке. В случае тензоризованного OMP ситуация несколько менее критична, поскольку малоранговость решения представляет собой своего рода регуляризацию. Однако возможность ставить произвольные значения в некоторые позиции может позволить находить лучшее с точки



Фиг. 5. График сравнения различных методов отбора столбцов в пункте 4 Алгоритма 2. Использованные методы: OMP (с переранжированием после каждого шага), метод имитации отжига (с переранжированием после каждого шага) и K-SVD [12], [13]. Модель $S_X = 7$, $S_L = 7$, $D_X = 3$, $D_L = 1$. Для метода тензоризованного OMP отбиралось $C = 400$ столбцов.

зрения невязки решение, что в корне противоречит тому, как дальше будут использованы компоненты решения. Большая по модулю компонента должна соответствовать более полезному для решения столбцу. Для того чтобы добиться этого, можно добавить простейший l_2–регуляризатор с коэффициентом λ. Минимизируемая функция в этом случае записывается как

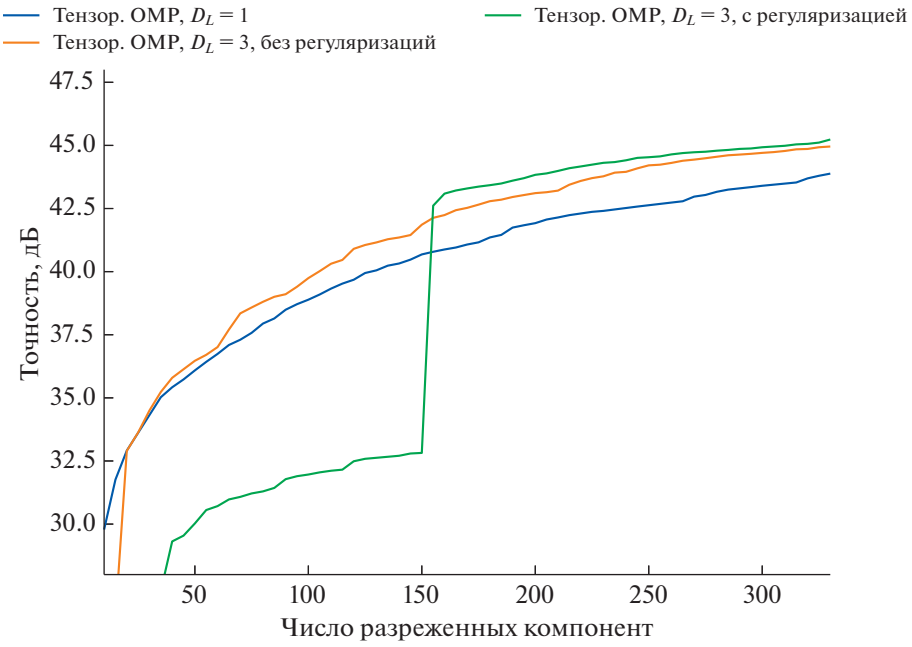
$$\left\| Q\mathcal{A}\left(\sum_{t=1}^{\tau} u_{1t} \otimes u_{2t} \otimes \ldots \otimes u_{dt} \right) - Qb \right\|_2^2 + \lambda \sum_{k,t} \|u_{kt}\|_2^2 \to \min. \tag{13}$$

На фиг. 6 приведен график сравнения метода с нерегуляризованной моделью при $D_L = 1$ и $D_L = 3$ и регуляризованной модели с $D_L = 3$. Из графика можно видеть, что увеличение размера модели закономерно приводит к улучшения качества. Из уравнения (13) несложно заметить, что коэффициент регуляризации λ необходимо выбирать адаптивно. Действительно, на первых итерациях метода невязка будет большой и необходима большая λ для получения эффекта от регуляризации. В то же время слишком большая λ будет приводить к нахождению нулевого решения в качестве оптимального. Поэтому λ выбирается адаптивно: изначально выбирается достаточно большое значение коэффициента регуляризации, но когда поиск малорангового решения начинает давать большую невязку, коэффициент λ уменьшается в несколько раз. Результат такого уменьшения можно видеть на графике в виде скачка.

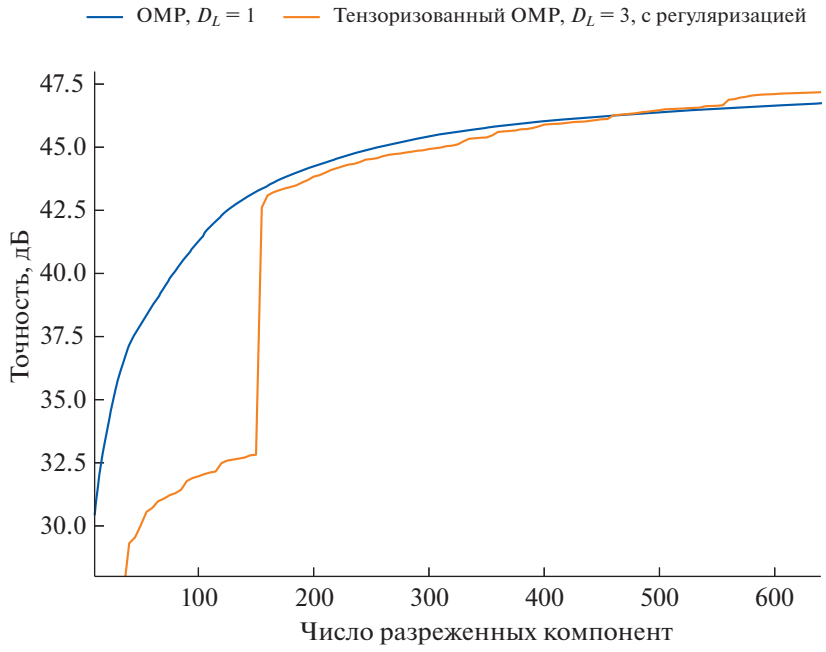
Классический метод OMP может быть применен к модели с $D_L = 1$, но его использование для модели с $D_L = 3$ уже затруднительно. На фиг. 7 приведен график сравнения OMP для модели с $D_L = 1$ и тензоризованного OMP с $D_L = 3$. Можно видеть, что при большом числе компонент тензоризованный OMP начинает обходить классический OMP, что доказывает его эффективность для поиска разреженных решений сверхбольших систем линейных уравнений.

6. ЗАКЛЮЧЕНИЕ

В работе предложен метод нахождения разреженного решения для сверхбольших систем линейных уравнений, обладающих тензорной структурой. Проведенный теоретический анализ позволяет понять, почему метод тензоризованного OMP способен успешно находить решения



Фиг. 6. График сравнения качества работы на различных моделях, а также эффект от применения регуляризации. Во всех моделях $S_X = 7$, $S_L = 7$, $D_X = 3$. Для метода тензоризованного OMP отбиралось $C = 400$ столбцов.



Фиг. 7. График сравнения методов OMP и тензоризованного OMP. Во всех моделях $S_X = 7$, $S_L = 7$, $D_X = 3$. Для метода тензоризованного OMP отбиралось $C = 400$ столбцов.

систем линейных уравнений, а также проследить отдельные этапы работы алгоритма. Представленные экспериментальные результаты демонстрируют хорошую масштабируемость метода тензоризованного OMP, и его способность эффективно находить разреженные решения сверхбольших систем линейных уравнений.

СПИСОК ЛИТЕРАТУРЫ

1. *Mallat S.G., Zhang Z.* Matching pursuits with time-frequency dictionaries // IEEE Transactions on Signal Proc. 1993. V. 41. № 12. P. 3397–3415.
2. *Elad M.* Sparse and redundant representations: from theory to applications in signal and image processing. Springer Science & Business Media. 2010.
3. *Cai T.T., Wang L.* Orthogonal matching pursuit for sparse signal recovery with noise // IEEE Transactions on Information Theory. 2011. V. 57. № 7. P. 4680–4688.
4. *Candes E.J., Tao T.* Decoding by linear programming // IEEE Transactions on Information Theory. 2005. V. 51. № 12. P. 4203–4215.
5. *Davenport M.A., Wakin M.B.* Analysis of orthogonal matching pursuit using the restricted isometry property // IEEE Transactions on Information Theory. 2010. V. 56. № 9. P. 4395–4401.
6. *Tropp J.A.* Greed is good: Algorithmic results for sparse approximation // IEEE Transactions on Information theory. 2004. V. 50. № 10. P. 2231–2242.
7. *Lebedeva O.* Tensor conjugate-gradient-type method for rayleigh quotient minimizationin block qtt-format // Russian Journal of Numerical Analysis and Math. Modelling. 2011. V. 26. № 5. P. 465–489.
8. *Zheltkov D.A., Osinsky A.* Global optimization algorithms using tensor trains. Internat. Conference on Large-Scale Scientific Computing. Springer. 2019. P. 197–202.
9. *Vershynin R.* Introduction to the non-asymptotic analysis of random matrices //arXiv preprint arXiv:1011.3027. 2010.
10. *Levenberg K.*A method for the solution of certain non-linear problems in least squares // Quarterly of Applied Math. 1944. V. 2. № 2. P. 164–168.
11. *Marquardt D.W.* An algorithm for least-squares estimation of nonlinear parameters // J. of the Society for Industrial and Applied Math. 1963. V. 11. № 2. P. 431–441.
12. *Aharon M., Elad M., Bruckstein A.* K-svd: An algorithm for designing overcomplete dictionaries for sparse representation // IEEE Transactions on Signal Proc. 2006. V. 54. № 11. P. 4311–4322.
13. *Rubinstein R., Bruckstein A.M., Elad M.* Dictionaries for sparse representation modeling // Proc. of the IEEE. 2010. V. 98. № 6. P. 1045–1057.
14. *Kirkpatrick S., Gelatt Jr, C., Vecchi M.* Optimization by Simulated Annealing // Science. 1983. V. 220. № 4598. P. 671–680.

EDN: IUGRWP
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1822–1839

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.63

КОНТРОЛЬ ТОЧНОСТИ ПРИБЛИЖЕННЫХ РЕШЕНИЙ ОДНОГО КЛАССА СИНГУЛЯРНО ВОЗМУЩЕННЫХ КРАЕВЫХ ЗАДАЧ

© 2022 г. С. И. Репин[1, 2, *]

[1] 191023 С.-Петербург, Фонтанка, 27, С.-Петербургское отделение Матем. ин-та им. В.А. Стеклова РАН, Россия

[2] 195251, С.-Петербург, ул. Политехническая, 29, С.-Петербургский политехн. ун-т Петра Великого, Россия

**e-mail: repin@pdmi.ras.ru*

Поступила в редакцию 22.06.2022 г.
Переработанный вариант 22.06.2022 г.
Принята к публикации 07.07.2022 г.

Рассматриваются уравнения реакции–конвекции–диффузии с малым параметром при старшей производной, и изучается вопрос о том, как эффективно контролировать точность приближенных решений таких задач с помощью апостериорных оценок. Полученные оценки не зависят от способа построения приближенного решения и работоспособны в широком диапазоне значений параметра. Основой для получения оценок являются специальные (апостериорные) тождества, левая часть которых представляет собой меру отклонения приближенного решения от точного, а правая содержит данные задачи и известное приближенное решение. В серии примеров показано, что тождества и вытекающие из них оценки позволяют эффективно вычислять погрешность как грубых, так и весьма точных аппроксимаций задач при различных значениях малого параметра. Библ. 38. Фиг. 6. Табл. 3.

Ключевые слова: сингулярно возмущенные уравнения, краевые задачи, тождества для мер отклонения от точного решения, апостериорные оценки функционального типа.

DOI: 10.31857/S0044466922110096

1. ВВЕДЕНИЕ

Класс краевых задач, которые принято называть сингулярно возмущенными, связан с физическими моделями, содержащими пограничные слои. Как правило, они связаны с наличием малого параметра при старшей производной и возникают в задачах диффузии (см. [1]), гидроаэродинамики (см. [2]) и электромагнетизма (см. [3]). Вероятно, первые исследования этих математических моделей были предприняты Л. Прандтлем (см. [4]), который ввел понятие пограничного слоя как их типичную особенность. Соответствующие решения имеют совершенно разный характер в области пограничного слоя и вне его. Например, в задачах гидродинамики с малой вязкостью пограничный слой может образоваться вблизи неподвижной границы, где скорости малы и решение зависит от величины вязкости. В остальной области решение практически не зависит от вязкости и несущественно отличается от течения идеальной жидкости. Аналогичную структуру имеют решения сингулярно возмущенных задач конвекции–реакции–диффузии (4.1), которое рассматривается в данной статье.

Попытки решения сингулярно возмущенных задач с помощью стандартных численных методов могут сталкиваться с серьезными трудностями (например, возникновением неустойчивости). Поэтому значительные усилия были направлены на разработку специальных методов, которые учитывают специфические особенности данного класса задач (см. [5]–[14] и цитированную литературу). Эти методы часто используют специальные (layer-resolving) неравномерные сетки (так называемые сетки Шишкина и Бахвалова). Обзоры результатов, относящихся к сингулярно возмущенным задачам, можно найти в [15]–[17].

В настоящей статье мы не обсуждаем методы построения численного решения, а изучаем вопрос о том, как эффективно контролировать приближенные решения сингулярно возмущенных задач. Ясно, что наличие малого параметра и специфика точного решения порождают серьезные трудности и в решении этой проблемы. Конечно, вопросы контроля точности приближенных

решений ранее изучались, и есть ряд работ, посвященных этой теме. Одними из первых изучались сеточные аппроксимации, и в [5] были получены оценки погрешности в сеточных нормах для системы обыкновенных дифференциальных уравнений и для уравнения эллиптического типа с малым параметром при одной из частных производных. Эти вопросы исследовались и в контексте других задач и различных методов аппроксимаций (см. [13], [18], [19] и многие другие работы). Например, в [12] получены асимптотические оценки вида $\|\!|u - u_h|\!\| \le Ch^q$ для энергетической нормы разности между точным решением (4.1) и его приближением на сетке размера h. Такие оценки гарантируют глобальную сходимость, равномерную относительно ε, если C и q не зависят от ε. Для уравнения $-\varepsilon\Delta u + a(x,\varepsilon)u = f(x,\varepsilon)$ оценки ошибок в норме $\|u - u_h\|_{\infty,\Omega}$ получены в статье [20]. Подобные (асимптотические) оценки погрешности аппроксимаций изучались разными авторами для многих сингулярно возмущенных задач. Они служат обоснованием численного метода и его устойчивости по отношению к малому параметру ε.

Другая группа исследований связана с получением апостериорных оценок и индикаторов погрешности для адаптивных вычислительных методов (см., например, [15], [21]). В [22], [23] были получены апостериорные оценки для обыкновенного дифференциального уравнения и эллиптической краевой задачи с малым параметром при старшей производной. В этих работах изучались конечно-разностные схемы на неравномерных сетках, а оценки построены в терминах максимальной разности между точным решением и линейным (или полилинейным) восполнением точного решения конечно-разностного уравнения. В [24] аналогичные оценки получены для аппроксимаций, построенных с помощью сплайнов и метода коллокаций.

Нас интересуют оценки другого рода, а именно те, что давали бы погрешность любого приближенного решения независимо от метода его получения. Такие оценки должны быть явно вычисляемыми и гарантированными мажорантами (минорантами) соответствующих мер погрешности и не должны содержать констант, зависящих от сетки и других специфических свойств аппроксимаций. Ранее подобные оценки были построены для широкого круга задач математической физики (см. [25]–[27]) путем анализа соответствующей краевой задачи чисто функциональными методами без использования специальных свойств аппроксимаций. Поэтому их часто называют апостериорными оценками функционального типа. Они отражают наиболее общие зависимости между мерой отклонения от точного решения и невязками функциональных соотношений, которые определяют конкретную задачу. В данной статье этот подход применяется к сингулярно возмущенной краевой задаче. Основой анализа являются тождества особого рода, в которых левая часть представляет собой некоторую меру отклонения приближенного решения от точного, а правая зависит от приближенного решения, данных задачи и некоторых других величин. Их можно назвать *апостериорными тождествами*. Может показаться, что такие тождества можно получать с помощью несложных формальных преобразований соответствующего уравнения. Например, для задачи $\Delta u = f$ в выпуклой области Ω, с $f \in L^2(\Omega)$ с условием $u = 0$ на границе Γ легко получить тождество

$$\|\Delta e\| = \|\Delta v - f\|, \tag{1.1}$$

где $\|\cdot\|$ обозначает норму $L^2(\Omega)$, а $e := v - u$ – функция, которая показывает отклонение приближенного решения v от точного u. Однако (1.1) мало пригодно для практических целей. Прежде всего оно выполняется только для аппроксимаций, обладающих повышенной регулярностью ($\Delta v \in L^2(\Omega)$). Кроме того, последовательности аппроксимаций, построенные с помощью стандартных численных методов (например, МКЭ), как правило, не обладают тем свойством, что невязка уравнения, записанного в классической форме, стремится к нулю при увеличении размерности конечномерного пространства. Обычно имеет место более слабая сходимость относительно нормы функционального пространства, содержащего это решение. При этом для последовательности сходящихся аппроксимаций правая часть (1.1) может не убывать, даже если сами аппроксимации (или их регуляризации) удовлетворяют требованиям повышенной регулярности. Естественно, что в этом случае тождество не имеет практического значения.

Тождества для отклонений от точного решения, которые соответствуют требованиям, естественным для большинства численных методов, были получены на основе теории двойственности вариационного исчисления в [26]. Подробное изложение соответствующей теории и приложения к широкому кругу задач для уравнений эллиптического типа содержится в [28, гл. 2] и в [29]. Недавно апостериорные тождества были получены для некоторых линейных и нелинейных параболических задач (см. [30]).

В настоящей статье апостериорные тождества отклонений от точного решения получены для стационарной задачи реакции–конвекции–диффузии

$$-\operatorname{div} p^* + a \cdot \nabla u + \rho^2 u = f \quad \text{в} \quad \Omega, \tag{1.2}$$

$$p^* = A\nabla u \quad \text{в} \quad \Omega, \tag{1.3}$$

$$u = u_0 \quad \text{на} \quad \Gamma \tag{1.4}$$

в ограниченной области $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ с липшицевой границей Γ. Здесь и далее $\cdot$ обозначает скалярное произведение векторов, а функция u_0 задает краевое условие (в смысле следов). Мы считаем, что $u_0 \in V := H^1(\Omega)$, $V_0 := \overset{\circ}{H}{}^1(\Omega)$ является подпространством V, содержащим функции, обращающиеся в нуль на границе. Предполагается, что A – это симметричная матрица с вещественными коэффициентами, и что выполнены следующие условия:

$$f \in L^2(\Omega), \quad a \in L^\infty(\Omega, \mathbb{R}^d), \quad \operatorname{div} a \in L^\infty(\Omega), \tag{1.5}$$

$$c_1^2|\xi|^2 \leq A\xi \cdot \xi \leq c_2^2|\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \tag{1.6}$$

$$\rho \in L^\infty(\Omega), \quad 0 \leq \rho \leq \rho_\oplus, \tag{1.7}$$

$$\rho^2 - \frac{1}{2}\operatorname{div} a := \sigma_a^2 \geq 0. \tag{1.8}$$

Обобщенное решение u определяется как функция из множества

$$V_0 + u_0 := \{w = w_0 + u_0 \,|\, w_0 \in V_0\},$$

удовлетворяющая интегральному тождеству

$$\int_\Omega (A\nabla u \cdot \nabla w + (a \cdot \nabla u)w + \rho^2 uw)dx = \int_\Omega fw dx \quad \forall w \in V_0. \tag{1.9}$$

Вопросы существования и единственности функции u, удовлетворяющей (1.9), а также регулярность и другие качественные свойства решения хорошо изучены (см., например, [31], [32]). Также имеется большое количество литературы, посвященной построению численных аппроксимаций этой задачи (см. обзор методов в [33]). Далее мы будем считать, что функция $v \in V_0 + u_0$ является такой аппроксимацией, полученной каким-либо способом. Также в рассмотрении участвует аппроксимация y^* точного потока p^*. Эти аппроксимации могут быть очень хорошими приближениями u и p^*, а могут быть и весьма грубым. Целью работы является получение универсальных оценок, которые контролируют отклонение таких приближенных решений от точных. При этом никаких специальных свойств типа галеркинской ортогональности для v или различных вариантов полного или частичного уравновешивания потока y^* не используется. Также не используются свойства сетки и сеточные константы.

Основой анализа являются два тождества для мер отклонения приближенных решений от точного решения задачи (1.2)–(1.4). Первое тождество (2.1) получено при минимальных предположениях относительно коэффициентов уравнения (1.2). Второе тождество (2.12) требует выполнения дополнительного условия (2.10). Это сужает область его применимости, однако важной особенностью (2.12) является то, что его правая часть зависит только от приближенных решений и может быть вычислена непосредственно. В разд. 3 тождества (2.1) и (2.12) используются для получения полностью вычисляемых и гарантированных апостериорных оценок. Задачам с малым параметром при старшей производной посвящен разд. 4. Здесь тождества (2.1) и (2.12) и соответствующие оценки преобразуются с учетом специфики задачи. В разд. 5 теоретические результаты проверяются на серии модельных задач для различных аппроксимаций и различных значениях малого параметра.

В статье используются следующие обозначения. Средние значения функций обозначаются символом $\{|\cdot|\}$, например, $\{|g|\}_\Omega := \dfrac{1}{|\Omega|}\displaystyle\int_\Omega g dx$. Нормы скалярных и векторных функций в $L^2(\Omega)$ обозначаются $\|\cdot\|$, а $\|\cdot\|_\rho$ соответствует норме с весом, т.е. $\|w\|_\rho^2 := \displaystyle\int_\Omega \rho^2 w^2 dx$. Векторнозначные функ-

ции, компоненты которых интегрируемы с квадратом, образуют пространство $Q^* := L^2(\Omega, \mathbb{R}^d)$, в котором можно задать две нормы:

$$\|q\|_A^2 := \int_\Omega Aq \cdot q dx \text{ и } \|q\|_{A^{-1}}^2 := \int_\Omega A^{-1} q \cdot q dx, \|q\|.$$

Пространство Q^*_{div} является подпространством Q^*. Оно содержит такие векторные функции, которые имеют интегрируемую с квадратом дивергенцию. Это пространство является гильбертовым относительно скалярного произведения

$$(p,q)_{\mathrm{div}} := \int_\Omega (p \cdot q + \operatorname{div} p \operatorname{div} q) dx.$$

2. ТОЖДЕСТВА ДЛЯ МЕР ОТКЛОНЕНИЙ ОТ ТОЧНОГО РЕШЕНИЯ ЗАДАЧИ (1.2)–(1.4)

Пусть $v \in V_0 + u_0$ и $y^* \in Q^*_{\mathrm{div}}$ являются аппроксимациями u и p^* соответственно. Функции $e := v - u$ и $e^* := y^* - p^*$ можно назвать *функциями отклонений* от точных решений (или функциями ошибок), а функция

$$\mathcal{R}_f(v, y^*) := \operatorname{div} y^* - a \cdot \nabla v + f - \rho^2 v$$

является невязкой уравнения (1.2). Она зависит только от известных коэффициентов, v и y^*, и поэтому может быть вычислена непосредственно.

Теорема. *Для $v \in V_0 + u_0$ и $y^* \in Q^*_{\mathrm{div}}$ выполняется тождество*

$$\mu_1^2(e, e^*) = \|A\nabla v - y^*\|_{A^{-1}}^2 - 2\int_\Omega \mathcal{R}_f(v, y^*) e\, dx, \tag{2.1}$$

где мера отклонения задается соотношением

$$\mu_1^2(e, e^*) := |||e|||^2 + \|e^*\|_{A^{-1}}^2,$$

а

$$|||e|||^2 := \|\nabla e\|_A^2 + 2\int_\Omega \sigma_a e^2 dx.$$

Доказательство. Вследствие (1.3) мы имеем тождество

$$\|A\nabla v - y^*\|_{A^{-1}}^2 = \|A\nabla e - e^*\|_{A^{-1}}^2.$$

Поэтому

$$\|A\nabla v - y^*\|_{A^{-1}}^2 + 2\int_\Omega e^* \cdot \nabla e dx = \|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2. \tag{2.2}$$

Преобразуем интеграл в (2.2) следующим образом:

$$\begin{aligned}\int_\Omega (y^* - p^*) \cdot \nabla e dx &= \int_\Omega (a \cdot \nabla u - f + \rho^2 u - \operatorname{div} y^*) e dx = \int_\Omega (a \cdot \nabla v - f + \rho^2 v - \operatorname{div} y^*) e dx + \\ &+ \int_\Omega (a \cdot \nabla(u - v) + \rho^2 (u - v)) e dx = \int_\Omega (a \cdot \nabla(u - v) + \rho^2 (u - v)) e dx - \int_\Omega \mathcal{R}_f(y^*, v) e dx.\end{aligned} \tag{2.3}$$

Так как

$$2\int_\Omega (a \cdot \nabla e) e dx = \int_\Omega a \cdot \nabla(e^2) dx = -\int_\Omega \operatorname{div} a e^2 dx, \tag{2.4}$$

мы имеем равенство

$$2\int_\Omega (a \cdot \nabla(u - v) + \rho^2 (u - v)) e dx = -2\int_\Omega (a \cdot \nabla e + \rho^2 e) e dx = \int_\Omega (\operatorname{div} a - 2\rho^2) e^2 dx. \tag{2.5}$$

Из (2.2), (2.3) и (2.5) следует тождество

$$\|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 + \int_\Omega (2\rho^2 - \operatorname{div} a)e^2\,dx = \|A\nabla v - y^*\|_{A^{-1}}^2 - 2\int_\Omega \mathcal{R}(v, y^*)e\,dx, \tag{2.6}$$

которое с учетом (1.8) совпадает с (2.1). Теорема доказана.

Тождество (2.1) содержит в левой части меру $\mu_1(e, e^*)$, которая является естественной характеристикой того, насколько хорошо v и y^* приближают точное решение u и точный поток p^* соответственно. Первый член в правой части вычисляется непосредственно, а второй содержит известную функцию невязки $\mathcal{R}_f(v, y^*)$ и неизвестную функцию e. Далее мы обсудим, как получить полностью вычисляемые оценки этого интеграла.

Замечание 1. Возьмем инфимум от обеих частей (2.1) по $y^* \in Q^*_{\mathrm{div}}$. Нетрудно видеть, что левая часть (2.1) достигает минимума, если $y^* = p^*$. В этом случае правая часть (2.1) также достигает минимума. Действительно,

$$\|A\nabla v - p^*\|_{A^{-1}}^2 = \|A\nabla(v - u)\|_{A^{-1}}^2 = \|\nabla e\|_A^2$$

и поскольку $\operatorname{div} p^* + f = a \cdot \nabla u + \rho^2 u$, то $\mathcal{R}_f(v, p^*) = -a \cdot \nabla e - \rho^2 e$. Поэтому

$$2\int_\Omega \mathcal{R}_f(v, p^*)e\,dx = -2\int_\Omega ((a \cdot \nabla e)e + \rho^2 e^2)dx = 2\int_\Omega (\operatorname{div} a - 2\rho^2)e^2\,dx = -2\int_\Omega \sigma_a^2 e^2\,dx.$$

Таким образом, мы получаем тождество для одной части $\mu_1(e, e^*)$:

$$|||e|||^2 = \inf_{y^* \in Q^*_{\mathrm{div}}} \left\{ \|A\nabla v - y^*\|_{A^{-1}}^2 - 2\int_\Omega \mathcal{R}_f(v, y^*)e\,dx \right\}. \tag{2.7}$$

Замечание 2. Априори известно, что обобщенное решение задачи содержится в $V_0 + u_0$, а соответствующий поток принадлежит пространству Q^*_{div}. Тождество (2.1) выполняется для любых функций $(v, y^*) \in \mathcal{H} := (V_0 + u_0) \times Q^*_{\mathrm{div}}$, т.е. позволяет оценивать отклонения от точных решений в пределах естественных энергетических множеств задачи. Множество $\mathcal{H}$ можно сузить таким образом, чтобы интеграл в правой части (2.1) исчез. Для этого надо потребовать, чтобы v и y^* принадлежали множеству

$$\mathcal{H}_0 := \left\{ (v, y^*) \in \mathcal{H} \,|\, \int_\Omega (y^* \cdot \nabla w + (a \cdot \nabla v)w + \rho^2 v w - fw)dx = 0 \quad \forall w \in V_0 \right\}.$$

В этом случае последний интеграл в (2.3) равен нулю, и мы получаем упрощенное тождество:

$$\|\nabla e\|_A^2 + 2\int_\Omega \sigma_a e^2\,dx + \|e^*\|_{A^{-1}}^2 = \|A\nabla v - y^*\|_{A^{-1}}^2, \tag{2.8}$$

которое, однако, выполняется только для $(v, y^*) \in \mathcal{H}_0$.

В случае $a = 0$ и $\rho = 0$ тождество (2.8) совпадает с хорошо известным “равенством гиперциклов” (см. [34])

$$\|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 = \|A\nabla v - y^*\|_{A^{-1}}^2, \tag{2.9}$$

где множество $\mathcal{H}_0$ задается уравнением $\operatorname{div} y^* + f = 0$ и накладывает ограничения только на переменную y^*. В литературе, посвященной методам апостериорного контроля точности, имеется целое направление, основанное на (2.9) и его аналогов в других задачах (например, для уравнений линейной упругости). В рамках этого подхода авторы стараются использовать так называемые уравновешенные (equilibrated) аппроксимации, которые содержатся в $\mathcal{H}_0$. Тождество (2.8) показывает, что в задачах более сложных, чем простое уравнение диффузии, множество $\mathcal{H}_0$ накладывает совместные ограничения на обе переменные. Их точное выполнение может приводить к серьезным техническим трудностям в практических вычислениях. Надлежащее использование полного тождества (2.1) позволяет избежать эти трудности и использовать простые аппроксимации для v и y^* (например, стандартные конечноэлементные аппроксимации Куранта и Равьяра–Тома соответственно).

Если

$$\rho(x) > 0 \quad \text{и} \quad \rho_a^2(x) := \rho^2(x) - \operatorname{div} a > 0 \quad \forall x \in \Omega, \tag{2.10}$$

то (2.1) можно преобразовать так, чтобы правая часть тождества оказалась полностью вычисляемой. Нетрудно видеть, что $\operatorname{div} e^* - a \cdot \nabla e = \mathcal{R}_f(v, y^*) + \rho^2 e$. Поэтому

$$(\operatorname{div} e^* - a \cdot \nabla e)^2 = \mathcal{R}_f^2(v, y^*) + \rho^4 e^2 + 2\rho^2 \mathcal{R}_f(v, y^*) e$$

и, следовательно,

$$\left\|(\operatorname{div} e^* - a \cdot \nabla e)\right\|_{\rho^{-1}}^2 - \left\|\mathcal{R}_f(v, y^*)\right\|_{\rho^{-1}}^2 - \int_\Omega \rho^2 e^2 dx = 2\int_\Omega \mathcal{R}_f(v, y^*) e dx. \tag{2.11}$$

Из (2.1) и (2.11) вытекает тождество

$$\mu_2^2(e, e^*) = \left\|A\nabla v - y^*\right\|_{A^{-1}}^2 + \left\|\mathcal{R}_f(v, y^*)\right\|_{\rho^{-1}}^2, \tag{2.12}$$

где мера отклонения μ_2 задается равенством

$$\mu_2^2(e, e^*) := \|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 + \int_\Omega \rho_a^2 e^2 dx + \left\|a \cdot \nabla e - \operatorname{div} e^*\right\|_{\rho^{-1}}^2.$$

Ясно, что мера $\mu_2(e, e^*)$ обращается в нуль, только если e и e^* тождественно равны нулю. Правая часть (2.12) легко вычисляется, так что тождество позволяет просто контролировать точность приближенных решений. Особенностью тождества (2.12) является наличие весовой функции $1/\rho^2$ в обеих частях. С вычислительной точки зрения это является недостатком, если ρ мало. При очень малых ρ нормы, содержащие эти весовые функции в левой и правой частях тождества, становятся доминирующими и почти равными друг другу, что делает тождество малоинтересным.

Замечание 3. Если $\operatorname{div} a = 0$, то $a \cdot \nabla u - \operatorname{div} p^* = \operatorname{div}(au - A\nabla u)$. Таким образом, векторнозначная функция $au - A\nabla u$ представляет собой полный поток p_{tot}^*, который состоит из диффузионного потока $-A\nabla u$ и потока, связанного с адвекцией $p_{\text{adv}}^* := au$. Соответственно $-y^*$ и av являются приближениями этих потоков. Поэтому

$$a \cdot \nabla e - \operatorname{div} e^* = \operatorname{div}(av - y^*) - \operatorname{div} p_{\text{tot}}^*,$$

и величина $\left\|a \cdot \nabla e - \operatorname{div} e^*\right\|_{\rho^{-1}}^2$ представляет собой взвешенную норму ошибки аппроксимации дивергенции полного потока. Если эту ошибку исключить, то из тождества (2.12) следует оценка

$$\|\nabla e\|_A^2 + \|e\|_{\rho_a}^2 + \|e^*\|_{A^{-1}}^2 \le \left\|A\nabla v - y^*\right\|_{A^{-1}}^2 + \left\|\mathcal{R}_f(v, y^*)\right\|_{\rho^{-1}}^2, \tag{2.13}$$

в которой мера отклонения почти такая же, как в (2.1).

Если $\operatorname{div} a = 0$, то тождество (2.12) приобретает вид

$$\|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 + \|e\|_\rho^2 + \left\|\operatorname{div} e^* - a \cdot \nabla e\right\|_{\rho^{-1}}^2 = \left\|A\nabla v - y^*\right\|_{A^{-1}}^2 + \left\|\mathcal{R}_f(v, y^*)\right\|_{\rho^{-1}}^2. \tag{2.14}$$

Если $a = 0$, то (2.14) еще более упрощается:

$$\|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 + \|e\|_\rho^2 + \left\|\operatorname{div} e^*\right\|_{\rho^{-1}}^2 = \left\|A\nabla v - y^*\right\|_{A^{-1}}^2 + \left\|\mathcal{R}_f(v, y^*)\right\|_{\rho^{-1}}^2.$$

Это апостериорное тождество было получено ранее для задачи реакции–диффузии в [25], [26] (см. также [28, гл. 2]).

3. ОЦЕНКА $\int_\Omega \mathcal{R}_f(v, y^*) e dx$

Для получения полностью вычисляемых оценок мы преобразуем интеграл $\int_\Omega \mathcal{R}_f(v, y^*) e dx$, входящий в (2.1), двумя разными способами. Первый основывается на простой оценке

$$\int_\Omega \mathcal{R}_f(v, y^*) e dx \le C_\Omega \left\|\mathcal{R}_f(v, y^*)\right\| \|\nabla e\| \le \frac{\alpha}{2} C_\Omega^2 \left\|\mathcal{R}_f(v, y^*)\right\|^2 + \frac{1}{2\alpha} \|\nabla e\|^2,$$

где $\alpha > 0$. Она приводит к следующему заключению: для любых $y^* \in Q^*_{\mathrm{div}}$, $v \in V_0 + u_0$ и $\alpha \geq 1/c_1^2$ выполняется неравенство

$$\mu_3^2(e, e^*, \alpha) \leq M_3^2(v, y^*, \alpha) := \|A\nabla v - y^*\|^2_{A^{-1}} + \alpha C_\Omega^2 \|\mathcal{R}_f(v, y^*)\|^2, \tag{3.1}$$

где

$$\mu_3^2(e, e^*, \alpha) := \left(1 - \frac{1}{c_1^2\alpha}\right)\|\nabla e\|^2_A + 2\int_\Omega \sigma_a e^2\, dx + \|e^*\|^2_{A^{-1}}.$$

Перейдя к инфимуму по y^* в обеих частях (3.1), мы получаем оценку для той части меры отклонения, которая связана с e:

$$\left(1 - \frac{1}{c_1^2\alpha}\right)\|\nabla e\|^2_A + 2\int_\Omega \sigma_a e^2\, dx \leq \inf_{y^* \in Q^*_{\mathrm{div}}} \left\{\|A\nabla v - y^*\|^2_{A^{-1}} + \alpha C_\Omega^2 \|\mathcal{R}_f(v, y^*)\|^2\right\}. \tag{3.2}$$

Как будет видно из примеров, простые оценки (3.1) и (3.2) иногда работают достаточно хорошо, но в ряде случаев могут сильно переоценивать меру отклонения. Поэтому далее мы рассмотрим другой метод.

Более точные оценки можно получить, если интеграл $\int_\Omega \mathcal{R}_f(v, y^*)e\,dx$ преобразуется с помощью специально построенной вспомогательной задачи. Пусть $u_{\mathcal{R}}$ и $p^*_{\mathcal{R}}$ являются решениями задачи

$$\operatorname{div} p^*_{\mathcal{R}} + \mathcal{R}_f(v, y^*) = 0,$$
$$p^*_{\mathcal{R}} = \nabla u_{\mathcal{R}},$$
$$u_{\mathcal{R}} = 0 \quad \text{на} \quad \Gamma.$$

Тогда

$$\int_\Omega \mathcal{R}_f(v, y^*)e\,dx = -\int_\Omega \nabla u_{\mathcal{R}} \cdot \nabla e\,dx.$$

Эта задача намного проще, чем (1.2)–(1.4), и не содержит малого параметра. Она обладает и другим важным свойством. Поскольку

$$\int_\Omega \nabla u_{\mathcal{R}} \cdot \nabla w\,dx = \int_\Omega \mathcal{R}_f(v, y^*)w\,dx \qquad \forall w \in V_0,$$

мы заключаем, что

$$\|p^*_{\mathcal{R}}\|^2 = \|\nabla u_{\mathcal{R}}\|^2 = \int_\Omega \mathcal{R}_f(v, y^*)u_{\mathcal{R}}\,dx = \int_\Omega \left((p^* - y^*) \cdot \nabla u_{\mathcal{R}} - a \cdot \nabla(v - u)u_{\mathcal{R}} - \rho^2(v - u)u_{\mathcal{R}}\right)dx \leq$$
$$\leq \left(\|p^* - y^*\| + C_\Omega(\|a\|_\infty \|\nabla(v - u)\| + \|v - u\|^2_\rho)\right)\|\nabla u_{\mathcal{R}}\|.$$

Отсюда следует, что $\|\nabla u_{\mathcal{R}}\|$ и $\|p^*_{\mathcal{R}}\|$ стремятся к нулю, если $y^* \to p^*$ в $L^2(\Omega)$ и $v \to u$ в V, что является совершенно естественным требованием относительно сходимости последовательности приближенных решений.

Конечно, точное решение $u_{\mathcal{R}}$ неизвестно. Однако можно использовать конечномерный аналог вспомогательной задачи и получить соответствующее приближение $u_{\mathcal{R},h}$. Эта идея была реализована в [29], где применялись классические конечноэлементные аппроксимации, а для оценки нормы разности между $u_{\mathcal{R}}$ и ее конечноэлементной аппроксимацией $u_{\mathcal{R},h}$ использовались стандартные интерполяционные оценки. Последние основаны на повышенной регулярности точного решения $u_{\mathcal{R}}$, что ограничивает область применимости данного метода. Здесь мы рассматриваем другую реализацию этой идеи, которая основана на аппроксимации $p^*_{\mathcal{R}}$ и не связана с требованиями повышенной регулярности.

Рассмотрим разбиение Ω на подобласти (элементы) T_i, так что $\overline{\Omega} = \bigcup_{i=1}^N \overline{T}_i$, $T_i \cap T_j = \emptyset$, если $i \neq j$. Элементы имеют характерный размер H. Это разбиение не зависит от способа дискретизации, использованного при построении приближенного решения v и соответствующего по-

тока y^*. В частности, v может быть конечноэлементной аппроксимацией u_h, построенной на сетке $\mathcal{T}_h$ с характерным размером элементов h, а сетка $\mathcal{T}_\text{H}$ во вспомогательной задаче может совпадать с $\mathcal{T}_h$, а может и быть другой (крупнее или мельче в зависимости от конкретных обстоятельств).

На элементе T определим простейший интерполяционный оператор $\pi_\text{H} : L^2(T) \to P^0(T)$ с помощью соотношения $\pi_\text{H} w\big|_T = \{|w|\}_T$. Тогда $\pi_\text{H} w$ – это кусочно-постоянная функция, принимающая на каждом T_i среднее значение функции w. Разобьем интеграл на два слагаемых:

$$\int\limits_\Omega \mathcal{R}_f(v, y^*) e\,dx = \int\limits_\Omega \mathcal{R}_f(v, y^*)\pi_\text{H} e\,dx + \int\limits_\Omega \mathcal{R}_f(v, y^*)(e - \pi_\text{H} e)\,dx. \tag{3.3}$$

Для преобразования первого интеграла в правой части (3.3) используем вспомогательную задачу: найти $p_\text{H}^* \in \hat{Q}_\text{H}^*$ и $u_\text{H} \in \hat{V}_\text{H}$ такие, что

$$\int\limits_\Omega (u_\text{H} \operatorname{div} y_\text{H}^* + y_\text{H}^* p_\text{H}^*)\,dx = 0 \quad \forall y_\text{H}^* \in \hat{Q}_\text{H}^*, \tag{3.4}$$

$$\int\limits_\Omega (\operatorname{div} p_\text{H}^* + \mathcal{R}_f(v, y^*) w_\text{H})\,dx = 0 \quad \forall w_\text{H} \in \hat{V}_\text{H}, \tag{3.5}$$

где $\hat{Q}_\text{H}^* \subset Q_\text{div}^*$, а $\hat{V}_\text{H}^*$ состоит из кусочно-постоянных функций, принимающих постоянные значения на каждом T_i. Система уравнений (3.4), (3.5) определяет двойственные смешанные аппроксимации краевой задачи $\Delta u_\mathcal{R} + \mathcal{R}_f(v, y^*) = 0$ с однородными краевыми условиями Дирихле. Эти аппроксимации хорошо изучены (см. монографии [35], [36]). Для симплициальных сеток пространство $\hat{Q}^*$ часто формируют с помощью элементов Равьяра–Тома (Raviart–Thomas elements RT_0). Для сеток, использующих полигональные ячейки, можно использовать соответствующие макроэлементы (например, такие как в [37]).

Из (3.4) и (3.5) следует, что

$$\|p_\text{H}^*\|_\Omega^2 = -\int\limits_\Omega u_\text{H} \operatorname{div} p_\text{H}^*\,dx = \int\limits_\Omega \mathcal{R}_f(v, y^*) u_\text{H}\,dx.$$

Отсюда видно, что $\|p_\text{H}^*\|_\Omega$ стремится к нулю, если $\mathcal{R}_f(v, y^*)$ стремится к нулю слабо в $L^2(\Omega)$.

Соотношение (3.5) позволяет переписать (3.3) в виде

$$\int\limits_\Omega \mathcal{R}_f(v, y^*) e\,dx = -\int\limits_\Omega \operatorname{div} p_\text{H}^* \pi_\text{H} e\,dx + \int\limits_\Omega \mathcal{R}_f(v, y^*)(e - \pi_\text{H} e)\,dx. \tag{3.6}$$

Оценим первое слагаемое в правой части (3.6). Поскольку $\int_{T_i} \pi_\text{H} w\,dx = \{|w|\}_{T_i} |T_i| = \int_{T_i} w\,dx$ и $\operatorname{div} p_\text{H}^* \in P^0(T_i)$ для любого T_i, мы получаем оценку

$$\begin{aligned}\int\limits_\Omega \operatorname{div} p_\text{H}^* \pi_\text{H} e\,dx &= \sum_i \int\limits_{T_i} \operatorname{div} p_\text{H}^* \pi_\text{H} e\,dx = \sum_i (\operatorname{div} p_\text{H}^*)_{T_i} \int\limits_{T_i} \pi_\text{H} e\,dx = \int\limits_\Omega \operatorname{div} p_\text{H}^* e\,dx \le \\ &\le \|p_\text{H}^*\|_\Omega \|\nabla e\|_\Omega \le \frac{1}{2\alpha c_1^2}\|\nabla e\|_A^2 + \frac{\alpha}{2}\|p_\text{H}^*\|_\Omega^2.\end{aligned} \tag{3.7}$$

Второе слагаемое оценивается с помощью неравенства Пуанкаре

$$\left|\int\limits_\Omega \mathcal{R}(v, y^*)(e - \pi_\text{H} e)\,dx\right| \le \sum_{i=1}^N C_\text{P}(T_i)\|\mathcal{R}(v, y^*) - \xi_i\|_{T_i} \|\nabla e\|_{T_i} \le S(v, y^*)\|\nabla e\| \le \frac{1}{2\beta c_1^2}\|\nabla e\|_A^2 + \frac{\beta}{2} S(v, y^*)^2, \tag{3.8}$$

где

$$S^2(v, y^*) = \sum_{i=1}^N C_\text{P}^2(T_i)\|\mathcal{R}_f(v, y^*) - \xi_i\|_{T_i}^2, \quad \xi_i = \{|\mathcal{R}(v, y^*)|\}_{T_i}.$$

Для выпуклого T_i мы имеем оценку $C_{\mathrm{P}}(T_i) \le \operatorname{diam} T_i/\pi$ (см. [38]). Таким образом,

$$S^2(v, y^*) \le \overline{S}^2(v, y^*) := \sum_{i=1}^{N} \frac{(\operatorname{diam} T_i)^2}{\pi^2} \left\| \mathcal{R}(v, y^*) - \xi_i \right\|_{T_i}^2. \tag{3.9}$$

Поскольку предполагается, что $\operatorname{diam} T_i \le d\mathrm{H}$, где d – некоторая положительная постоянная, то множители в сумме не превосходят $d^2\mathrm{H}^2/\pi^2$.

Равенство (3.3) с учетом (3.9) и (3.8) приводит к оценке

$$2\left| \int_\Omega \mathcal{R}(v, y^*) e dx \right| \le K_{\alpha\beta} \|\nabla e\|_A^2 + \beta \overline{S}^2(v, y^*) + \alpha \|p_{\mathrm{H}}^*\|_\Omega^2, \tag{3.10}$$

где $K_{\alpha\beta} = (\alpha + \beta)/(\beta\alpha c_1^2)$. Здесь α и β – это положительные постоянные такие, что $1/\alpha + 1/\beta \le c_1^2$. С учетом (3.10) тождество (2.1) дает оценку

$$\mu_4^2(e, e^*, \alpha, \beta) \le M_4^2(v, y^*, \alpha, \beta) := \|A\nabla v - y^*\|_{A^{-1}}^2 + \beta \overline{S}^2(v, y^*) + \alpha \|p_{\mathrm{H}}^*\|_\Omega^2, \tag{3.11}$$

где

$$\mu_4^2(e, e^*, \alpha, \beta) := (1 - K_{\alpha\beta}) \|\nabla e\|_A^2 + 2\int_\Omega \sigma_a e^2 dx + \|e^*\|_{A^{-1}}^2,$$

так что эта мера зависит от выбора параметров α и β. Выбирая параметры, можно увеличивать или уменьшать множитель при $\|\nabla e\|_A^2$. Если взять $\alpha = \beta = 2/c_1^2$, то $K_{\alpha\beta} = 1$, и первое слагаемое меры исчезает.

Замечание 4. Рассуждая аналогичным образом, можно получить оценку меры отклонения снизу

$$(1 + K_{\alpha\beta}) \|\nabla e\|_A^2 + 2\int_\Omega \sigma_a e^2 dx + \|e^*\|_{A^{-1}}^2 \ge \|A\nabla v - y^*\|_{A^{-1}}^2 - \beta \overline{S}^2(v, y^*) - \alpha \|p_{\mathrm{H}}^*\|_\Omega^2. \tag{3.12}$$

4. ОЦЕНКИ ДЛЯ СИНГУЛЯРНО ВОЗМУЩЕННЫХ ЗАДАЧ

Соотношения, полученные в предыдущем разделе, носят общий характер и применимы для любых задач, удовлетворяющих условиям (1.5), (1.6), (1.7) и (1.8). Применим их к задачам с $A = \varepsilon A_0$, где A_0 – это положительно-определенная симметричная матрица с наименьшим собственным значением порядка единицы, а $\varepsilon > 0$ – малый параметр. В этом случае (1.2) и (1.3) приобретают вид

$$-\varepsilon \operatorname{div} A_0 \nabla u + a \cdot \nabla u + \rho^2 u = f, \tag{4.1}$$

$$p^* = \varepsilon A_0 \nabla u. \tag{4.2}$$

Нетрудно видеть, что

$$\|e^*\|_{A^{-1}}^2 = \varepsilon^{-1} \|e^*\|_{A_0^{-1}}^2, \quad \|\nabla e\|_A^2 = \varepsilon \|\nabla e\|_{A_0}^2, \quad \|A\nabla v - y^*\|_{A^{-1}}^2 = \varepsilon^{-1} \|\varepsilon A_0 \nabla v - y^*\|_{A_0^{-1}}^2.$$

Поэтому (2.1) заменяется тождеством

$$\mu_1^2(e, e^*, \varepsilon) = \frac{1}{\varepsilon} \|\varepsilon A_0 \nabla v - y^*\|_{A_0^{-1}}^2 - 2\int_\Omega \mathcal{R}_f(v, y^*) e dx, \tag{4.3}$$

где

$$\mu_1^2(e, e^*, \varepsilon) := \varepsilon \|\nabla e\|_{A_0}^2 + \frac{1}{\varepsilon} \|e^*\|_{A_0^{-1}}^2 + \int_\Omega (2\rho^2 - \operatorname{div} a) e^2 dx.$$

Тождество (2.12) приобретает вид

$$\mu_2^2(e, e^*) = \frac{1}{\varepsilon} \|\varepsilon A_0 \nabla v - y^*\|_{A_0^{-1}}^2 + \left\| \rho^{-1} \mathcal{R}_f(v, y^*) \right\|^2, \tag{4.4}$$

где

$$\mu_2^2(e,e^*) := \varepsilon\|\nabla e\|^2_{A_0} + \frac{1}{\varepsilon}\|e^*\|^2_{A_0^{-1}} + \int\limits_\Omega (\rho_a^2 e^2 + \rho^{-2}(\operatorname{div} e^* - a\cdot\nabla e)^2)dx,$$

и предполагается, что условие (2.10) выполнено. Подчеркнем, что апостериорные тождества (4.3) и (4.4) *выполняются при любых* $\varepsilon > 0$. Этот факт создает основу для контроля точности приближенных решений сингулярно возмущенных уравнений. В первую очередь это относится к задачам, в которых условие (2.10) выполнено и ρ не принимает очень малых значений. В этом случае явно вычисляемая правая часть тождества (4.4) дает верхнюю границу отклонения для комбинированной меры ошибки $\mu_2(e,e^*,\varepsilon)$. Также следует отметить, что весовые множители правильно балансируют компоненты меры в соответствии с поведением точного решения, которое равномерно ограничено по ε в L^2 норме и не может расти быстрее $\varepsilon^{-1/2}$ по норме градиента.

Для случая $\operatorname{div} a = 0$ из (4.4) следует оценка

$$\varepsilon\|\nabla e\|^2_{A_0} + \frac{1}{\varepsilon}\|e^*\|^2_{A_0^{-1}} + \|e\|^2_\rho \le \frac{1}{\varepsilon}\|\varepsilon A_0\nabla v - y^*\|^2_{A_0^{-1}} + \|\mathcal{R}_f(v,y^*)\|^2_{\rho^{-1}}. \tag{4.5}$$

Чтобы получить полностью вычисляемые оценки с помощью (4.3), используем те же рассуждения, что и в разд. 3. Так как

$$2\int\limits_\Omega \mathcal{R}_f(v,y^*)e dx \le \frac{\alpha}{\varepsilon}C_\Omega^2\|\mathcal{R}_f(v,y^*)\|^2 + \frac{\varepsilon}{\alpha}\|\nabla e\|^2 \le \frac{\alpha}{\varepsilon}C_\Omega^2\|\mathcal{R}_f(v,y^*)\|^2 + \frac{\varepsilon}{c_1^2(A_0)\alpha}\|\nabla e\|^2_{A_0},$$

где $c_1(A_0)$ – это наименьшее собственное значение матрицы A_0, мы получаем соответствующий вариант оценки (3.1)

$$\mu_3^2(e,e^*,\alpha,\varepsilon) \le \frac{1}{\varepsilon}\|\varepsilon A_0\nabla v - y^*\|^2_{A_0^{-1}} + \frac{\alpha}{\varepsilon}C_\Omega^2\|\mathcal{R}_f(v,y^*)\|^2, \tag{4.6}$$

где

$$\mu_3^2(e,e^*,\alpha,\varepsilon) := \varepsilon\left(1 - \frac{1}{c_1^2(A_0)\alpha}\right)\|\nabla e\|^2_{A_0} + \frac{1}{\varepsilon}\|e^*\|^2_{A_0^{-1}} + 2\int\limits_\Omega \sigma_a^2 e^2 dx.$$

Из (4.6) следует аналог оценки (3.2)

$$\varepsilon\left(1 - \frac{1}{c_1^2(A_0)\alpha}\right)\|\nabla e\|^2_{A_0} + 2\int\limits_\Omega \sigma_a^2 e^2 dx \le \inf_{y^*\in Q^*_{\operatorname{div}}}\left\{\frac{1}{\varepsilon}\|\varepsilon A_0\nabla v - y^*\|^2_{A_0^{-1}} + \frac{\alpha}{\varepsilon}C_\Omega^2\|\mathcal{R}_f(v,y^*)\|^2\right\}. \tag{4.7}$$

Преобразуем оценку (3.11), где $K_{\alpha\beta} = (\alpha+\beta)/(\varepsilon\beta\alpha c_1^2(A_0))$. Путем замены $\alpha = \overline{\alpha}/\varepsilon$, $\beta = \overline{\beta}/\varepsilon$ определим новую константу $K_{\overline{\alpha}\overline{\beta}} = (\overline{\alpha}+\overline{\beta})/(\overline{\beta}\overline{\alpha}c_1^2(A_0))$. Тогда оценка трансформируется следующим образом:

$$\mu_4^2(e,e^*,\overline{\alpha},\overline{\beta},\varepsilon) \le \frac{1}{\varepsilon}\left\{\|\varepsilon A_0\nabla v - y^*\|^2_{A^{-1}} + \mathscr{E}(v,y^*,p_{\mathrm{H}}^*,\overline{\alpha},\overline{\beta})\right\}, \tag{4.8}$$

где

$$\mathscr{E}(v,y^*,p_{\mathrm{H}}^*,\overline{\alpha},\overline{\beta}) := \overline{\beta}\overline{S}^2(v,y^*) + \overline{\alpha}\|p_{\mathrm{H}}^*\|^2_\Omega,$$

$$\mu_4^2(e,e^*,\overline{\alpha},\overline{\beta},\varepsilon) := (1 - K_{\overline{\alpha}\overline{\beta}})\varepsilon\|\nabla e\|^2_{A_0} + 2\int\limits_\Omega \sigma_a e^2 dx + \frac{1}{\varepsilon}\|e^*\|^2_{A^{-1}}.$$

Далее мы покажем как апостериорные тождества (4.3), (4.4) и оценки (3.11), (3.12) позволяют контролировать приближенные решения сингулярно возмущенных задач.

5. ПРИМЕРЫ

В этом разделе апостериорные тождества и вытекающие из них оценки проверяются на примере сингулярно возмущенной двухточечной задачи

$$-\varepsilon u'' + au' + \rho^2 u = f, \quad u(0) = 0, \quad u(1) = b \tag{5.1}$$

с постоянными ρ и a. В этом случае Ω является интервалом $I := (0,1)$,

$$\|e\|_{A_0}^2 = \int_0^1 |e'|^2\,dx, \quad c_1(A_0) = 1, \quad C_\Omega = \frac{1}{\pi}, \quad \mathcal{R}_f(v, y^*) = (y^*)' + f - av' - \rho^2 v,$$

$$\mu_1^2(e, e^*, \varepsilon) = \varepsilon\|e'\|^2 + \frac{1}{\varepsilon}\|e^*\|^2 + 2\int_0^1 \rho^2 e^2 dx,$$

$$\mu_2^2(e, e^*, \varepsilon) = \varepsilon\|e'\|^2 + \frac{1}{\varepsilon}\|e^*\|^2 + \int_0^1 (\rho^2|e|^2 + \rho^{-2}[(e^*)' - ae']^2)dx.$$

Тождества (4.3) и (4.4) приобретают вид

$$\mu_1^2(e, e^*, \varepsilon) = \frac{1}{\varepsilon}\|\varepsilon v' - y^*\|^2 - 2\int_0^1 \mathcal{R}_f(v, y^*)e\,dx \tag{5.2}$$

и

$$\mu_2^2(e, e^*, \varepsilon) = \frac{1}{\varepsilon}\|\varepsilon v' - y^*\|^2 + \|\mathcal{R}_f(v, y^*)\|_{\rho^{-1}}^2 =: M^2(v, y^*, \varepsilon) \tag{5.3}$$

соответственно. Приведем также оценку (4.5):

$$\varepsilon\|e'\|^2 + \frac{1}{\varepsilon}\|e^*\|^2 + \|e\|_\rho^2 \le M^2(v, y^*, \varepsilon). \tag{5.4}$$

Тождество (5.2) проверялось и во всех примерах выполнялось с точностью до ошибок численного интегрирования (которые всегда можно уменьшить). Однако это тождество содержит неизвестную функцию e в правой части. Поэтому с практической точки зрения интересно не столько оно само, сколько оценки, которые из него следуют. Напротив, величина $M(v, y^*, \varepsilon)$ в тождестве (5.3) вычисляется явно. Для того чтобы проверить это тождество и оценить эффективность оценки (5.4), введем коэффициенты

$$I_1 := \frac{M(v, y^*, \varepsilon)}{\mu_2(e, e^*, \varepsilon)} \quad \text{и} \quad I_2 := \frac{M(v, y^*, \varepsilon)}{\left(\varepsilon\|e'\|^2 + \frac{1}{\varepsilon}\|e^*\|^2 + \|e\|_\rho^2\right)^{1/2}}.$$

Представленная в таблицах и графиках величина I_1 являет собой отношение правой части (5.3) к левой (см. табл. 1–3). Во всех случаях она была равна 1 с точностью до малых погрешностей, обу-

Таблица 1. Сравнение оценок (5.3), (5.4), (5.5) и (5.6) для интерполянтов точного решения в примере 1

ε	1.0000	0.5000	0.2500	0.1250	0.0625	0.0312	0.0156	0.0078
I_1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
I_2	1.0453	1.0860	1.1560	1.2600	1.3801	1.4763	1.5366	1.5874
I_3	1.4286	1.4286	1.4287	1.4291	1.4303	1.4337	1.4424	1.4648
I_4	1.7320	1.7320	1.7320	1.7321	1.7322	1.7326	1.7335	1.7361
I_5	1.4274	1.4647	1.5953	1.9762	2.8047	4.1593	6.1088	8.9940

Таблица 2. Сравнение оценок (5.3), (5.4), (5.5) и (5.6) в примере 2 на сетке с 500 интервалами

ε	1.0000	0.5000	0.2500	0.1250	0.0625	0.0312	0.0156	0.0078
I_1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
I_2	2.9227	5.0951	9.5666	17.9884	32.4251	52.7068	70.5289	74.3708
I_3	1.4284	1.4277	1.4251	1.4158	1.3889	1.3426	1.3444	1.4623
I_4	1.7318	1.7309	1.7274	1.7152	1.6821	1.6375	1.6791	1.8602
I_5	2.2485	4.7141	12.1877	32.2792	81.5810	181.9772	323.3232	453.9166

Таблица 3. Сравнение оценок (5.3), (5.4), (5.5) и (5.6) в примере 3 на сетке с 500 интервалами

ε	1.0000	0.5000	0.2500	0.1250	0.0625	0.0312	0.0156	0.0078
I_1	1.0001	1.0001	1.0002	1.0002	1.0002	1.0001	1.0000	1.0000
I_2	1.4737	1.4576	1.5201	1.7132	2.1667	3.1023	4.8093	7.4144
I_3	1.4286	1.4284	1.4280	1.4269	1.4240	1.4158	1.3941	1.3523
I_4	1.7322	1.7318	1.7312	1.7296	1.7256	1.7149	1.6879	1.6438
I_5	2.5047	3.1943	4.5965	7.6472	14.7498	31.7016	71.2819	153.8980

словленных использованием численных квадратур. Величина I_2 больше единицы и может оказаться достаточно большой, если вклад слагаемого $\rho^{-2}[(e^*)' - ae']^2$ значителен. Как следует из примеров, при малых ε это вполне возможно.

Обратимся к оценке (3.11), которая была построена с помощью вспомогательной конечномерной задачи (3.4), (3.5). Разобьем интервал I на подынтервалы $T_i := (x_i, x_{i+1})$ так, что $|x_{i+1} - x_i| = \mathrm{H}_i$, и обозначим

$$\zeta_i = \frac{1}{\mathrm{H}} \int_{x_i}^{x_{i+1}} \mathcal{R}_f(v, y^*)dx = \{|\mathcal{R}|\}(q_h^*)_{I_i}.$$

В (3.4), (3.5) функция p_{H}^* является непрерывной кусочно-аффинной, а функция u_{H} является разрывной функцией, принимающей постоянные значения на каждом T_i. Равенство (3.5) означает, что

$$\int_0^1 ((p_{\mathrm{H}}^*)' + \mathcal{R}_f(v, y^*))w_{\mathrm{H}}dx = 0$$

для кусочно-постоянной тест-функции w_{H}, значения которой на T_i равны некоторым константам. Поскольку эти константы могут быть любыми, мы заключаем, что на каждом подынтервале должно выполняться равенство

$$\int_{x_i}^{x_{i+1}} ((p_{\mathrm{H}}^*)' + H\zeta_i)dx = 0,$$

откуда следует, что $(p_{\mathrm{H}}^*)' = -\zeta_i$ на T_i. Определим функцию $æ(x) = -\zeta_i$ для $x \in I_i$. Таким образом, искомым решением является кусочно-аффинная непрерывная функция p_{H}^* такая, что $p_{\mathrm{H}}^* = \int_0^x æ(t)dt + c$. В силу свойств двойственной вариационной формулировки (которая порождает систему (3.4), (3.5)) интеграл $\int_0^1 |p_{\mathrm{H}}^*|^2 dx$ должен быть минимален. Это дает условие на постоянную c:

$$c = \hat{c} := -\int_0^1 \int_0^x æ(t)dtdx.$$

При этом

$$(v, y^*, p_{\mathrm{H}}^*, \overline{\alpha}, \overline{\beta}) := \overline{\beta} \sum_{i=1}^N \frac{\mathrm{H}_i^2}{\pi^2} \|\mathcal{R}_f(v, y^*) - \zeta_i\|_{T_i}^2 + \overline{\alpha}\|p_{\mathrm{H}}^*\|_\Omega^2,$$

и мы получаем оценку

$$\mu_4^2(e, e^*, \overline{\alpha}, \overline{\beta}, \varepsilon) \le \frac{1}{\varepsilon} \int_0^1 |\varepsilon v' - y^*|^2 dx + \frac{1}{\varepsilon}(v, y^*, p_{\mathrm{H}}^*, \overline{\alpha}, \overline{\beta}) =: M_{\overline{\alpha},\overline{\beta}}(v, y^*, \varepsilon), \tag{5.5}$$

где

$$\mu_4^2(e,e^*,\bar\alpha,\bar\beta,\varepsilon) := \int_0^1 ((1-K_{\bar\alpha\bar\beta})\varepsilon|e'|^2 + 2\sigma_a e^2 dx + \frac{1}{\varepsilon}|e^*|^2)dx.$$

В примерах качество этой оценки характеризуется величинами

$$I_3 := \frac{M_{\bar\alpha,\bar\beta}(v,y^*,\varepsilon)}{\mu_4(e,e^*,\bar\alpha,\bar\beta,\varepsilon)}, \quad \bar\alpha = \bar\beta = 3, \quad \text{и} \quad I_4 := \frac{M_{\bar\alpha,\bar\beta}(v,y^*,\varepsilon)}{\mu_4(e,e^*,\bar\alpha,\bar\beta,\varepsilon)}, \quad \bar\alpha = 2, \quad \bar\beta = 100.$$

Эти оценки оказались наиболее работоспособными при малых ε, что ясно видно из приведенных далее таблиц и графиков.

В заключение приведем оценку (4.6), которая для задачи (5.1) записывается в виде:

$$\mu_3^2(e,e^*,\alpha,\varepsilon) \le \frac{1}{\varepsilon}\|\varepsilon v' - y^*\|^2 + \frac{\alpha}{\varepsilon\pi^2}\|\mathcal{R}_f(v,y^*)\|^2 =: M_\alpha^2(v,y^*,\varepsilon), \tag{5.6}$$

где

$$\mu_3^2(e,e^*,\alpha,\varepsilon) := \varepsilon\left(1-\frac{1}{\alpha}\right)\|e'\|^2 + \frac{1}{\varepsilon}\|e^*\|^2 + 2\int_\Omega \rho^2 e^2 dx.$$

Эффективность этой оценки характеризует число

$$I_5 := \frac{M_\alpha(v,y^*,\varepsilon)}{\mu_3(e,e^*,\alpha,\varepsilon)}.$$

При $\varepsilon \sim 1$ эта оценка вполне применима, но при малых значениях ε она может приводить к большой переоценке.

Пример 1. Проверку эффективности оценок начнем с наиболее простой задачи, в которой конвективный член отсутствует. Рассмотрим двухточечную задачу

$$-\varepsilon u'' + \rho^2 u = f, \quad u(0) = 0, \quad u(1) = 0 \tag{5.7}$$

с $f = \text{const}$. Точное решение

$$u = -\frac{1}{\rho^2 K}\left(\mu_1 e^{-\delta x} + \mu_2 e^{\delta x}\right) + \frac{1}{\rho^2}, \quad u' = -\frac{\delta}{\rho^2 K}\left(-\mu_1 e^{-\delta x} + \mu_2 e^{\delta x}\right),$$

$$u'' = -\frac{\delta^2}{\rho^2 K}\left(\mu_1 e^{-\delta x} + \mu_2 e^{\delta x}\right),$$

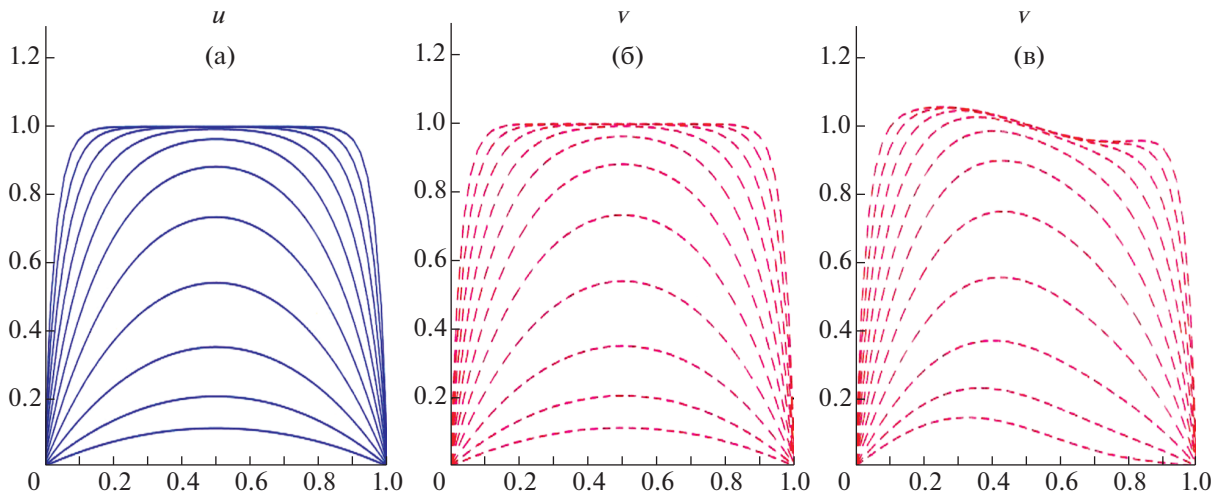
$$p^* = \varepsilon u', \quad (p^*)' = -\frac{1}{K}\left(\mu_1 e^{-\delta x} + \mu_2 e^{\delta x}\right).$$

Здесь

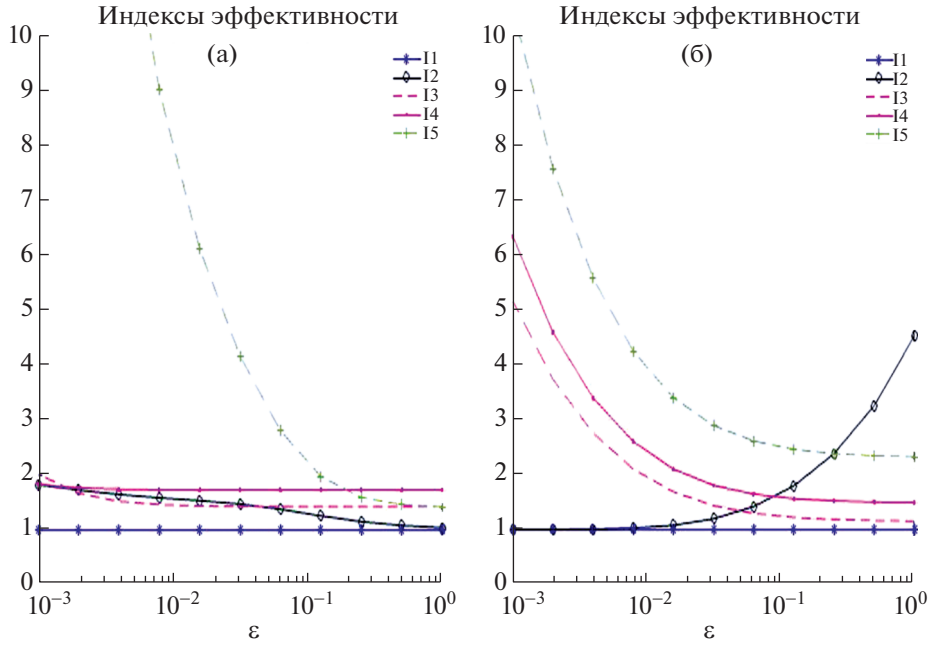
$$\delta = \frac{\rho}{\sqrt{\varepsilon}}, \quad \mu_1 = e^\delta - 1, \quad \mu_2 = 1 - e^{-\delta}, \quad K = \mu_1 + \mu_2 = e^\delta - e^{-\delta}.$$

В расчетах было взято $\rho = 1$, $f = 1$. Изменение точного решения при уменьшении ε показано на фиг. 1а. В качестве v был взят интерполянт точного решения на равномерной сетке с 500 интервалами. Соответствующие функции изображены на фиг. 1б. Визуально они практически не отличаются от точного решения и их можно считать очень хорошими аппроксимациями точного решения (фактически они совпадают с галеркинскими аппроксимациями u_h на данной сетке). В качестве y^* была взята простейшая регуляризация $\varepsilon u_h'$, которая строится путем усреднения значений в узлах сетки. В табл. 1 приведены результаты, которые показывают, как работают различные оценки. Соответствующие графики изображены на фиг. 2а.

Из табл. 1 и рисунков следует, что тождество (5.3) (которое является частным случаем (2.12)) выполняется точно при любом ε (первая строка табл. 1). Для хороших аппроксимаций оценки (5.4) и (5.5) также работают прекрасно, однако при уменьшении ε эффективность простой оценки (5.6) ухудшается. Для грубых аппроксимаций тождество (5.3) также выполняется для любых ε. При уменьшении ε хорошо работает оценка (5.4), но эффективность остальных оценок снижается. Тем не менее обе оценки (5.5) позволяют правильно оценить порядок ошибки и установить, что соответствующие аппроксимации являются весьма грубыми. Это обстоятельство мо-



Фиг. 1. Точные решения в задаче (5.2) при различных ε (а), их интерполянты (б) и грубые аппроксимации (в).

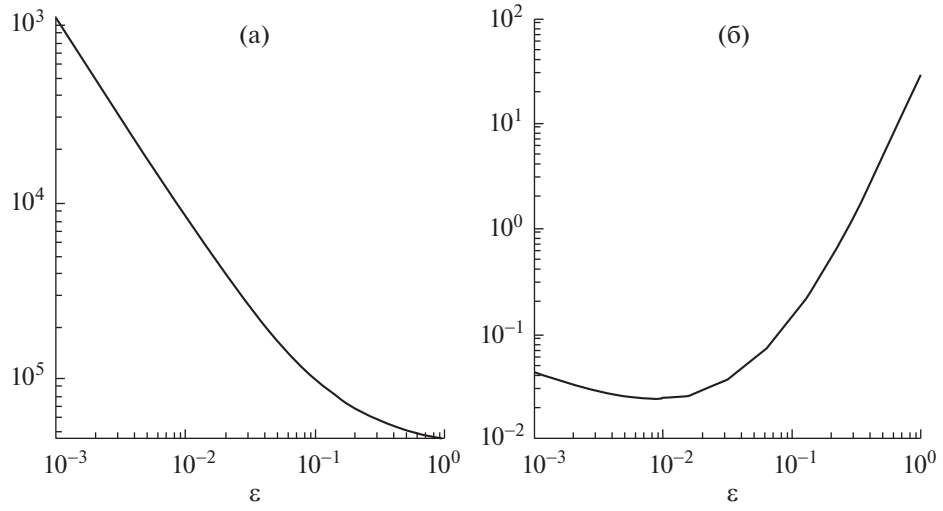


Фиг. 2. Индексы эффективности оценок при различных ε для интерполянта точного решения (а) и грубой аппроксимации (б).

жет оказаться полезным, если для решения используются численные схемы, в которых нельзя исключить возникновение неустойчивости.

Решения сингулярно возмущенных задач сильно зависят от ε. Меры μ_i, $i = 1,2,3,4$, также зависят от ε. Поэтому в примерах полезно рассматривать не только их абсолютные значения, но и значения, нормализованные относительно норм градиента точного решения ∇u и соответствующего потока p^*. В частности, определим нормализованный вариант меры μ_2 следующим образом:

$$\hat{\mu}_2^2(e,e^*,\varepsilon) = \frac{\mu_i^2(e,e^*,\varepsilon)}{\varepsilon\|\nabla u\|_{A_0}^2 + \frac{1}{\varepsilon}\|p^*\|_{A_0^{-1}}^2} = \frac{1}{2\varepsilon}\frac{\mu_1^2(e,e^*,\varepsilon)}{\|\nabla u\|_{A_0}^2}.$$



Фиг. 3. Относительная погрешность аппроксимаций при различных ε для интерполянта точного решения (а) и грубой аппроксимации (б).

На фиг. 3 приведены величины этой меры в зависимости от ε для точных (фиг. 3а) и грубых (фиг. 3б) аппроксимаций. Видно, что относительная погрешность аппроксимаций изменяется в широких пределах. Таким образом, работоспособность предлагаемых методов апостериорного контроля точности не зависит ни от абсолютной, ни от относительной погрешностей приближенных решений.

Пример 2. Если $f = at - \varepsilon s + (as + \rho^2 t)x + \frac{1}{2}\rho^2 s x^2$, то точное решение (5.1) задается соотношениями

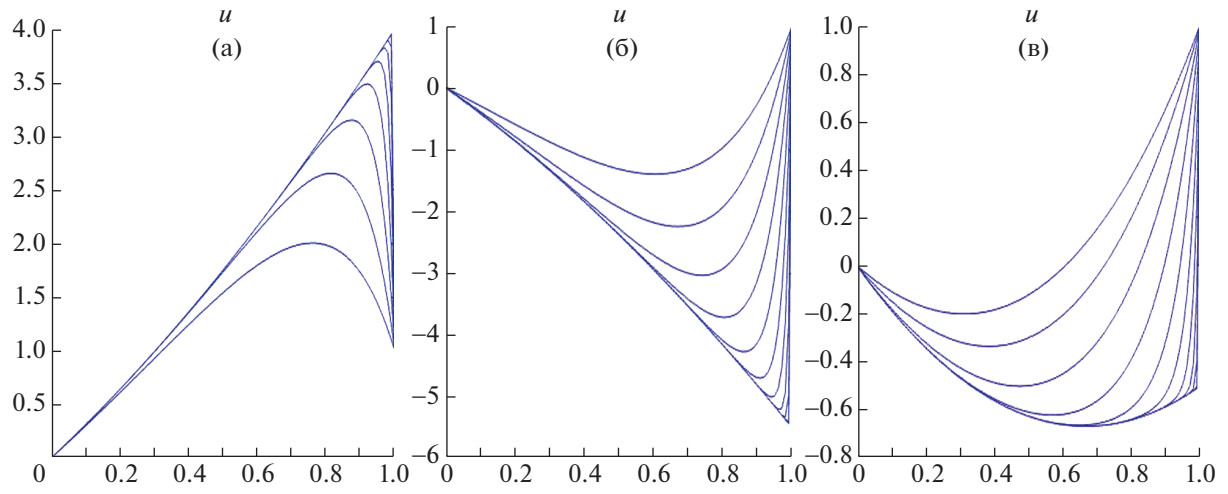
$$u = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x} + tx + \frac{1}{2}sx^2, \quad p^* = \varepsilon(\lambda_1 C_1 e^{\lambda_1 x} + \lambda_2 C_2 e^{\lambda_2 x} + t + sx),$$

где

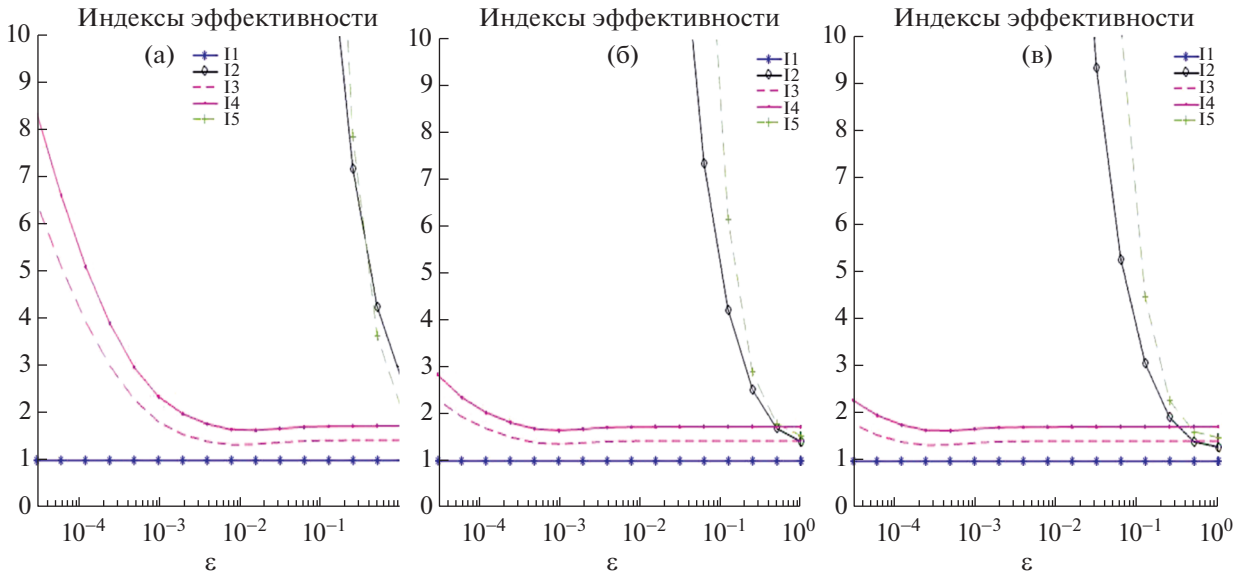
$$\lambda_{1,2} = \frac{a}{2\varepsilon} \mp \frac{1}{\sqrt{\varepsilon}}\sqrt{\rho^2 + \frac{a^2}{4\varepsilon}}, \quad C_1 = \left(1 - t - \frac{s}{2}\right)\frac{e^{-\lambda_1}}{(1 - e^{\lambda_2 - \lambda_1})}, \quad C_2 = -C_1.$$

В этом примере исследовалась задача с параметрами $a = 5$, $\rho = 1$, $t = 3$ и $s = 2$, которые моделируют ситуацию с доминирующей конвекцией. Точные решения при различных ε изображены на фиг. 4а. Хорошо видно формирование пограничного слоя при уменьшении ε. В качестве приближенного решения v рассматривался интерполянт точного решения на равномерной сетке с $n = 50$, 500, 5000 и 10000 интервалами. Функция y^* строится как кусочно-аффинная непрерывная функция, являющаяся усреднением $\varepsilon v'$ на той же самой сетке. Соответствующие результаты представлены в табл. 2 и на фиг. 5.

Результаты показывают, что апостериорное тождество (5.3) (которое является частным случаем (2.12)) точно выполняется при любых ε. Простые оценки (5.4) и (5.6) (с $\alpha = 2$) хорошо работают только при значениях ε порядка 1. При уменьшении ε качество этих оценок быстро ухудшается, и они приводят к существенной переоценке меры ошибки. Однако оценка (5.5) работает устойчиво вплоть до весьма малых ε. Так же, как и в примере 1, оценка (5.5) дает лучшие результаты для более точных аппроксимаций. Для точных аппроксимаций оценки дают лучшие результаты, чем для грубых, что вполне ожидаемо, поскольку более точные аппроксимации дают меньшее $\mathcal{R}_f(v, y^*)$ и уменьшают общую переоценку, которая возникает в слагаемом $\mathcal{E}(v, y^*, p_H^*, \overline{\alpha}, \overline{\beta})$. Для $n = 10000$ оценка не превышает истинное значение меры отклонения более чем в 2 раза вплоть до значений $\varepsilon = 10^{-5}$. Эти результаты получены на равномерных сетках, но позволяют предположить, что на сгущающихся сетках апостериорная оценка (5.5) будет весьма точной. Действительно, при увеличении n в медленно меняющейся части решения практически ничего



Фиг. 4. Точные решения в примерах 2 (а), 3 (б) и 4 (в) при различных ε.

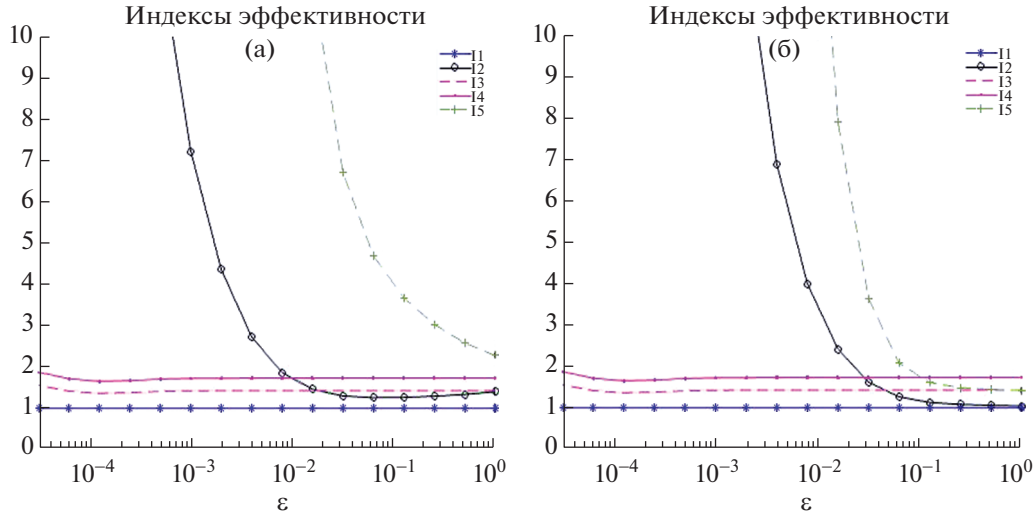


Фиг. 5. Эффективность оценок в примере 2 при различных ε для $n = 500$ (а) и $n = 5000$ (б) и $n = 10\,000$ (в).

не меняется, так что изменение точности оценок на фиг. 5 объясняется тем, что при более точной аппроксимации пограничного слоя вблизи $x = 1$ оценка также дает более точный результат.

Пример 3. В этом примере $a = 1$, $\rho = 3$, $t = -4$ и $s = -3$, что соответствует случаю, когда реакция превалирует над конвекцией. Точные решения для различных ε представлены на фиг. 4б в центре. Как и в примере 2, видно образование пограничного слоя с большими значениями производной вблизи точки $x = 1$. Как и в предыдущем примере, приближенное решение построено как кусочно-аффинный интерполянт точного решения на сетке из 5000 интервалов. В этом примере (5.4) работает лучше и обеспечивает хорошую оценку вплоть до $\varepsilon = 0.03$. Однако при меньших значениях ε возникает существенная переоценка меры отклонения. Как видно из табл. 3, оценка (5.5) дает очень хорошие результаты (строки I_3 и I_4) вплоть до значений $\varepsilon = 10^{-3}$.

Пример 4. В этом примере рассматривается случай, когда $a = \rho = 1$, $t = -2$ и $s = 3$. Точные решения, соответствующие различным значениям ε, изображены на фиг. 4а. В качестве v был взят интерполянт точного решения на сетке с 5000 интервалами. На фиг. 6 показано поведение ин-



Фиг. 6. Эффективность оценок при различных ε в примерах 3 (а) и 4 (б).

дексов эффективности I_1–I_5. Картина очень похожа на то, что мы наблюдали ранее. Индекс I_1 показывает точное выполнение апостериорного тождества (5.3) при всех значениях ε, простые оценки (5.4) и (5.6) можно использовать, но лишь для $\varepsilon > 10^{-1}$, а оценка (5.5) устойчива и отлично работает, по крайней мере, до значений $\varepsilon = 10^{-5}$. В заключение отметим, что приведенные выше примеры относятся к области обыкновенных дифференциальных уравнений, которые являются достаточно простыми, но типичными представителями сингулярно возмущенных краевых задач и адекватно отражают возникающие трудности (неслучайно исследование математических свойств и разностных аппроксимаций этого класса задач начиналось с подобных одномерных моделей). Поэтому можно предполагать, что при переходе к уравнениям в частных производных поведение полученных в статье апостериорных оценок принципиально не изменится, хотя несомненно этот вопрос требует дальнейшего изучения.

СПИСОК ЛИТЕРАТУРЫ

1. *Farrell P.A., Hegarty A.F., Miller J.J.H., Ó Riordan E., Shishkin G.I.* Robust computational techniques for boundary layers. CRC Press, Taylor&Frencis, Boca Raton, 2018.
2. *Oleinik O.A., Samokhin V.N.* Mathematical models in boundary layer theory. CRC Press, Taylor&Frencis, Boca Raton, 1999.
3. *Schilders W.H.A., Polak S.J., van Welij J.S.* Singular perturbation theory and its application to the computation of electromagnetic fields // IEEE Trans. on Magn. MAG–21. 1985. V. 6. P. 2211–2216.
4. *Prandtl L.* Uber Flussigkeits – bewegung bei kleiner Reibung, in: Verhandlungen, III, Int. Math. Kongresses, Tuebner, Leipzig, 1905. P. 484–491.
5. *Бахвалов Н.С.* К оптимизации методов решения краевых задач при наличии пограничного слоя // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. № 4. Р. 841–859.
6. *Дулан Э., Миллер Дж., Шилдерс У.* Равномерные численные методы решения задач с пограничным слоем. М.: Мир, 1983.
7. *Шишкин Г.И.* Первая краевая задача для уравнения второго порядка с малыми параметрами при производных // Дифференц. ур-ния. 1977. Т. 13. № 2. Р. 376–378.
8. *Шишкин Г.И.* Решение краевой задачи для эллиптического уравнения с малым параметром при старших производных // Ж. вычисл. матем. и матем. физ. 1986. Т. 26. № 7. Р. 1019–1031.
9. *Шишкин Г.И.* Аппроксимация решений и производных сингулярно возмущенного эллиптического уравнения конвекции–диффузии // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 5. Р. 672–689.
10. *Шишкин Г.И., Шишкина Л.П.* Улучшенные аппроксимации решения и производных сингулярно возмущенного уравнения реакции-диффузии на основе метода декомпозиции решения // Ж. вычисл. матем. и матем. физ. 2011. Т. 51. № 6. Р. 1091–1120.
11. *Ó Riordan E.* Singular perturbation finite element methods // Numer. Math. 1984. V. 44. P. 425–434.

12. *Ó Riordan E., Stynes M.* A uniformly accurate finite-element method for a singularly perturbed one-dimensional reaction-diffusion problem // Math. Comput. 1986. V. 47. № 176. P. 555–570.

13. *Kellogg R.B., Tsan A.* Analysis of Some Difference Approximations for a Singular Perturbation Problem Without Turning Points // Math. Comput. 1978. V. 32. № 144. P. 1025–1039.

14. *Miller J.J.H., O'Riordan E., Shishkin G.I.* Solution of Singularly Perturbed Problems with ε-uniform Numerical Methods. Introduction to the Theory of Linear Problems in One and Two Dimensions, World Scientific, Singapore, 1996.

15. *Roos H.-G., Stynes M., Tobiska L.* Robust numerical methods for singularly perturbed differential equations: convection-diffusion-reaction and flow problems. Springer Series in Computational Mathematics, 2008.

16. *Kadalbajoo M.K., Patidar K.C.* Singularly perturbed problems in partial differential equations: a survey // Appl. Math. Comput. 2003. V. 134. P. 371–429.

17. *Linβ T.* Layer-adapted meshes for reaction-convection-diffusion problems. In: Lecture Notes in Mathematics, vol. 1985. Berlin: Springer, 2010.

18. *Ильин А.М.* Разностная схема для дифференциального уравнения с малым параметром при старшей производной // Матем. заметки. 1969. Т. 6. № 2. Р. 237–248.

19. *Андреев В.Б.* Равномерная сеточная аппроксимация негладких решений смешанной краевой задачи для сингулярно возмущенного уравнения реакции–диффузии в прямоугольнике // Ж. вычисл. матем. и матем. физ. 2008. Т. 48. № 1. Р. 90–114.

20. *Schatz A.H., Wahlbin L.B.* On the finite element method for singularly perturbed reaction-diffusion problems in two and one dimensions // Math. Comp. 1983. V. 40. P. 47–89.

21. *Verfürth R.* A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques, Stuttgart: Wiley-Teubner, 1996.

22. *Kopteva N.* Maximum norm a posteriori error estimates for a one-dimensional convection- diffusion problem // SIAM J. Numer. Anal. 2002. V. 39. № 2. P. 423–441.

23. *Kopteva N.* Maximum norm a posteriori error estimate for a 2D singularly perturbed semilinear reaction-diffusion problem // SIAM J. Numer. Anal. 2008. V. 46. № 3. P. 1602–1618.

24. *Linβ T., Radojev G., Zarin H.* Approximation of singularly perturbed reaction-diffusion problems by quadratic C^1-splines // Numer Algor. 2012. V. 61. P. 35–55.

25. *Repin S.* A posteriori error estimation for variational problems with uniformly convex functionals // Math. Comp. 2000. V. 69. P. 481–500.

26. *Repin S.* Two-sided estimates of deviation from exact solutions of uniformly elliptic equations // Amer. Math. Soc. Transl. Ser. 2. 2003. V. 209. P. 143–171.

27. *Repin S.* A posteriori estimates for partial differential equations, Berlin: Walter de Gruyter GmbH & Co. KG, 2008.

28. *Repin S., Sauter S.* Accuracy of mathematical models. Dimension reduction, simplification, and homogenization // EMS Tracts in Math. 2020. V. 33.

29. *Репин С.И.* Тождество для отклонений от точного решения задачи $A^*\Lambda Au + \ell = 0$ и его следствия // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 12. С. 22–45.

30. *Repin S.* Error identities for parabolic initial boundary value problems // Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI). 2021. V. 508. P. 147–172.

31. *Gilbarg D., Trudinger N.S.* Elliptic partial differential equations of second order. Berlin: Springer-Verlag, 1977.

32. *Ladyzhenskaya O.A., Uraltseva N.N.* Linear and Quasilinear Elliptic equations. New York: Acad. Press, 1968.

33. *Morton K.W.* Numerical Solution of Convection-Diffusion Problems. New York: Taylor&Francis, 1996.

34. *Prager W., Synge J.L.* Approximation in elasticity based on the concept of function space // Quart. Appl. Math. 1947. V. 5. P. 241–269.

35. *Brezzi F., Fortin M.* Mixed and Hybrid Finite Element Methods. 15. New York: Springer Series in Computational Mathematics, 1991.

36. *Roberts J.E., Thomas J.-M.* Mixed and hybrid methods. In: Handbook of Numerical Analysis. Vol. II. Amsterdam: North-Holland, 1991. P. 523–639.

37. *Kuznetsov Yu., Repin S.* New mixed finite element method on polygonal and polyhedral meshes // Russian J. Numer. Anal. Math. Model. 2003. V. 18. № 3. P. 261–278.

38. *Payne L.E., Weinberger H.F.* An optimal Poincaré inequality for convex domains // Arch. Rational Mech. Anal. 1960. V. 5. P. 286–292.

EDN: IRAXMB
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1840–1850

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 517.977

К ЗАДАЧЕ РЕКОНСТРУКЦИИ ПРИ ДЕФИЦИТЕ ИНФОРМАЦИИ В КВАЗИЛИНЕЙНОМ СТОХАСТИЧЕСКОМ ДИФФЕРЕНЦИАЛЬНОМ УРАВНЕНИИ

© 2022 г. В. Л. Розенберг[1,*]

[1] 620990 Екатеринбург, ул. С. Ковалевской, 16, Ин-т матем. и механики им. Н.Н. Красовского УрО РАН, Россия

**e-mail: rozen@imm.uran.ru*

Поступила в редакцию 01.07.2021 г.
Переработанный вариант 09.06.2022 г.
Принята к публикации 07.07.2022 г.

Задача восстановления неизвестных внешних воздействий в квазилинейном стохастическом дифференциальном уравнении рассматривается в рамках подхода теории динамического обращения. Реконструкция возмущений в детерминированном и стохастическом членах уравнения базируется на дискретной информации о некотором количестве реализаций части координат случайного процесса. Задача сводится к обратной задаче для системы обыкновенных нелинейных дифференциальных уравнений, которой удовлетворяют математическое ожидание и ковариационная матрица исходного процесса. Для конечношагового программно реализуемого алгоритма решения, основанного на методе вспомогательных управляемых моделей, получена оценка точности относительно количества доступных измерению реализаций. Приведен модельный пример. Библ. 22. Фиг. 2.

Ключевые слова: квазилинейное стохастическое дифференциальное уравнение, динамическая реконструкция, неполная входная информация, управляемая модель.

DOI: 10.31857/S0044466922110114

1. ВВЕДЕНИЕ

Рассматриваемая задача реконструкции вкладывается в проблематику обратных задач динамики управляемых систем, которые, как правило, являются некорректными и требуют применения регуляризирующих процедур. Для ее решения используется ставший классическим подход, предложенный в работах Ю.С. Осипова и его коллег (см. [1]–[6] и библиографию в них) и получивший название метода динамического обращения. Он основан на сочетании принципов теории позиционного управления (см. [7]) и идей теории некорректных задач (см. [8]). Суть его в сведении задачи реконструкции к задаче управления по принципу обратной связи (часто на основе принципа экстремального прицеливания Н.Н. Красовского, см. [7]) вспомогательной динамической системой (моделью). Адаптация модельного управления к результатам текущих наблюдений обеспечивает аппроксимацию (в подходящем смысле) неизвестного входа. Метод динамического обращения был многократно реализован для различных детерминированных систем, описываемых обыкновенными дифференциальными уравнениями (ОДУ), дифференциально-функциональными уравнениями, уравнениями и вариационными неравенствами с распределенными параметрами и др. Были созданы устойчивые алгоритмы, работающие для некоторых классов частично наблюдаемых систем (в случае конечномерной системы роль входного сигнала могли играть измерения части координат фазового вектора, а в случае бесконечномерной – значения решения на некоторых подмножествах области определения). Такие задачи были сформулированы и решены, например, в [4], [5]. Касательно стохастических объектов отметим, что впервые задача позиционного моделирования неизвестного стохастического управляющего воздействия в системе, описываемой ОДУ, была рассмотрена в [9]. Впоследствии поле применения методов динамической регуляризации было расширено для систем с неопределенными входами в [10], где рассмотрена задача об устойчивой аппроксимации неизвестного входа управляемой системы по результатам неточных наблюдений ее фазовых состояний при условии, что помехи в канале наблюдения подчинены некоторому вероятностному распределению с ма-

лым математическим ожиданием. Впервые в рамках теории динамического обращения использован вероятностный критерий качества и построен динамический алгоритм реконструкции нормального входа, доставляющий сколь угодно высокую точность среднеквадратической аппроксимации с вероятностью, сколь угодно близкой к единице.

Настоящая работа фактически продолжает исследования задач реконструкции неизвестных параметров линейных и квазилинейных стохастических дифференциальных уравнений (СДУ), начатые в [11] с рассмотрения задачи динамического восстановления возмущения, входящего в интеграл Ито и характеризующего амплитуду случайной помехи, на основе измерения реализаций всего фазового вектора линейного СДУ. Идея решения состояла в переходе к задаче для линейной системы ОДУ, которой удовлетворяют математическое ожидание и ковариационная матрица исходного процесса. В [12] исследовалась достаточно общая постановка обратной задачи для линейного СДУ, предполагающая реконструкцию возмущений, входящих и в детерминированный, и в стохастический члены уравнения, на основе дискретной по времени информации о некотором количестве реализаций части координат случайного процесса. Обоснована применимость алгоритма восстановления неизвестных параметров, разработанного ранее для частично наблюдаемой системы ОДУ (см. [4], [5]), и предложена соответствующая модификация. Анализ задачи реконструкции неизвестных входов квазилинейного СДУ при измерении реализаций всего фазового вектора был проведен в [13]. Отличительной особенностью задачи для квазилинейного СДУ является нелинейность системы ОДУ, описывающей динамику математического ожидания и ковариационной матрицы процесса. Настоящая статья призвана дополнить исследование задач динамической реконструкции для квазилинейных СДУ случаем неполной информации, когда измеряется лишь часть координат фазового состояния. Отметим, что обратная задача для системы гибридного типа, состоящей из квазилинейного СДУ и ОДУ, при условии измерения реализаций случайного процесса рассмотрена в [14].

2. ПОСТАНОВКА ЗАДАЧИ

Рассматривается квазилинейное (в соответствии с терминологией [15] и других работ этих авторов) СДУ с диффузией, зависящей от фазового состояния:

$$\begin{gathered} dx(t,\omega) = (A(t)x(t,\omega) + B(t)u_1(t) + f(t))dt + U_2(t)x(t,\omega)d\xi(t,\omega), \\ x(0,\omega) = x_0, \quad t \in T = [0,\vartheta]. \end{gathered} \tag{2.1}$$

Здесь $x = (x_1,\dots,x_n) \in \mathbb{R}^n$, $\xi \in \mathbb{R}$; x_0 – известный детерминированный или случайный (нормально распределенный, с независимыми координатами) вектор начальных условий; $\omega \in \Omega$, (Ω, F, P) – вероятностное пространство (см., например, [16]); $\xi(t,\omega)$ – стандартный скалярный винеровский процесс (т.е. выходящий из нуля процесс с нулевым математическим ожиданием и дисперсией, равной t); $f(t) = \{f_i(t)\}$, $A(t) = \{a_{ij}(t)\}$ и $B(t) = \{b_{ij}(t)\}$ – непрерывные матричные функции размерности $n\times 1$, $n\times n$, и $n\times r$ соответственно. На уравнение действуют два внешних возмущения: вектор $u_1(t) = (u_{11}(t), u_{12}(t),\dots,u_{1r}(t)) \in \mathbb{R}^r$ и диагональная матрица $U_2(t) = \{u_{21}(t), u_{22}(t),\dots,u_{2n}(t)\} \in \mathbb{R}^{n\times n}$. Воздействие u_1 входит в детерминированную компоненту и влияет на математическое ожидание искомого процесса. Поскольку $U_2 x d\xi = (u_{21}x_1 d\xi, u_{22}x_2 d\xi,\dots,u_{2n}x_n d\xi)$, то можно считать, что вектор $u_2 = (u_{21}, u_{22},\dots,u_{2n})$ характеризует амплитуду случайных помех. Полагаем, что возмущения $u_1(\cdot) \in L_2(T;\mathbb{R}^r)$ и $u_2(\cdot) \in L_2(T;\mathbb{R}^n)$ принимают значения из заданных выпуклых компактов S_{u_1} и S_{u_2}.

Решение уравнения (2.1) определяется как случайный процесс, удовлетворяющий при любом t с вероятностью 1 соответствующему интегральному тождеству, содержащему в правой части стохастический интеграл Ито. Как известно, при сделанных предположениях существует единственное решение, являющееся нормальным марковским процессом с непрерывными реализациями (см. [17]). Отметим, что уравнения типа (2.1) описывают простейшие линеаризованные модели, например, изменения численности многовидовой биологической популяции в стохастической среде или динамики цен на товарных рынках при влиянии случайных факторов.

Изучаемая задача состоит в следующем. В дискретные, достаточно частые, моменты времени $\tau_i \in T$, $\tau_i = i\delta$, $\delta = \vartheta/l$, $i \in [0 : (l-1)]$, поступает информация о некотором количестве N реализаций случайного процесса $x(\tau_i)$, причем измерению доступны только q ($q \le n$) первых координат, т.е. вектор $(x_1,\dots,x_q)$. Полагаем, что $l = l(N)$ и существуют оценки m_{qi}^N q-подвектора

$m_q(t) = \{m_j(t)\}$, $j \in [1:q]$, вектора математического ожидания процесса $m(t) = Mx(t)$ и D_{qi}^N $(q \times q)$-подматрицы $D_q(t) = \{d_{jp}(t)\}$, $j, p \in [1:q]$, ковариационной матрицы $D(t) = M(x(t) - m(t))(x(t) - m(t))'$ (штрих означает транспонирование) такие, что выполняется соотношение

$$P\left(\max_{i\in[1:((l(N)-1)]}\left\{\left\|m_{qi}^N - m_q(\tau_i)\right\|_{\mathbb{R}^q}, \left\|D_{qi}^N - D_q(\tau_i)\right\|_{\mathbb{R}^{q\times q}}\right\} \le h(N)\right) = 1 - g(N), \tag{2.2}$$

причем $h(N), g(N) \to 0$ при $N \to \infty$. Стандартные статистические процедуры (см. [18]) построения оценок m_{qi}^N и D_{qi}^N допускают модификации, обеспечивающие выполнение (2.2).

Необходимо построить конструктивный алгоритм динамического восстановления неизвестных возмущений $u_1(t)$ и $u_2(t)$, фактически определяющих случайный процесс $x(t)$, по неполной дискретной информации о реализациях части его координат, причем вероятность сколь угодно малого отклонения приближений от искомых входов в метрике соответственно пространств $L_2(T;\mathbb{R}^r)$ и $L_2(T;\mathbb{R}^n)$ должна быть близка к 1 при достаточно большом N и специальным образом согласованном с N шаге временной дискретизации $\delta = \delta(N) = \vartheta/l(N)$. Такую обратную задачу можно трактовать как реконструкцию помех в условиях дефицита информации в случае, когда динамика уравнения допускает одновременные измерения достаточно большого количества траекторий.

3. ПЕРЕХОД К ЗАДАЧЕ ДЛЯ СИСТЕМЫ ОДУ

Применяя ставший классическим метод моментов из [19], по аналогии с [12], [13], сведем задачу восстановления для СДУ к задаче для системы ОДУ. Введем обозначения $m_0 = Mx_0$, $D_0 = M(x_0 - m_0)(x_0 - m_0)'$. Отметим, что матрица D_0 является диагональной (ввиду независимости координат вектора x_0). В силу квазилинейности исходного уравнения и равенства нулю математического ожидания интеграла Ито величина $m(t)$ зависит только от $u_1(t)$; ее динамика описывается уравнением

$$\dot m(t) = A(t)m(t) + B(t)u_1(t) + f(t), \quad m \in \mathbb{R}^n, \quad m(0) = m_0. \tag{3.1}$$

Напомним, что измерению доступны первые q координат исходного n-мерного вектора x, что обеспечивает получение оценки (2.2) для первых q координат вектора m; количество неизмеряемых координат равно $n - q$. Используя стандартную схему теории динамического обращения (см. [4], [5]), запишем уравнение (3.1) в виде системы с разделением измеряемой и неизмеряемой компонент. Введем следующие матрицы:

$$C_1^m = \{c_{1ps}^m\}, \quad p \in [1:q], \quad s \in [1:n], \quad c_{1ps}^m = 1, \quad \text{если} \quad s = p,$$

$$c_{1ps}^m = 0 \quad \text{в противном случае;}$$

$$D_1^m = \{d_{1ps}^m\}, \quad p \in [1:(n-q)], \quad s \in [1:n], \quad d_{1ps}^m = 1, \quad \text{если} \quad s = q + p,$$

$$d_{1ps}^m = 0 \quad \text{в противном случае;}$$

$$C_2^m = (C_1^m)'; \quad D_2^m = (D_1^m)'.$$

Тогда q-мерный вектор $C_1^m m$ – измеряемая часть m, а $(n-q)$-мерный вектор $D_1^m m$ – его неизмеряемая часть. Обозначим $y_m = C_1^m m$, $z_m = D_1^m m$. Умножая уравнение (3.1) поочередно на матрицы C_1^m и D_1^m, учитывая равенство $m = C_2^m y_m + D_2^m z_m$, приходим к системе с разделением измеряемой и неизмеряемой компонент:

$$\dot y_m(t) = A_{1y}^m(t)y_m(t) + A_{1z}^m(t)z_m(t) + B_1^m(t)u_1(t) + C_1^m f(t), \quad y_m(0) = y_{m0}, \tag{3.2}$$

$$\dot z_m(t) = A_{2y}^m(t)y_m(t) + A_{2z}^m(t)z_m(t) + B_2^m(t)u_1(t) + D_1^m f(t), \quad z_m(0) = z_{m0}, \tag{3.3}$$

где $A_{1y}^m = C_1^m A C_2^m$, $A_{1z}^m = C_1^m A D_2^m$, $B_1^m = C_1^m B$, $y_{m0} = C_1^m m_0$, $A_{2y}^m = D_1^m A C_2^m$, $A_{2z}^m = D_1^m A D_2^m$, $B_2^m = D_1^m B$, $z_{m0} = D_1^m m_0$.

Ковариационная матрица $D(t)$ не зависит от $u_1(t)$ явно. Для описания ее динамики используется схема вывода уравнения метода моментов, примененная в [12], [13]. В результате для $D \in \mathbb{R}^{n\times n}$ имеем следующее уравнение:

$$\dot{D}(t) = A(t)D(t) + D(t)A'(t) + U_2(t)(D(t) + m(t)m'(t))U_2'(t), \quad D(0) = D_0. \tag{3.4}$$

Матричное уравнение (3.4) переписываем в виде более традиционного для рассматриваемых задач векторного уравнения, размерность которого, с учетом симметричности матрицы $D(t)$, определяется как $n_{yz} = (n^2 + n)/2$. Вводится вектор $d(t) = \{d_s(t)\}$, $s \in [1 : n_{yz}]$, состоящий из последовательно записанных и пронумерованных элементов матрицы $D(t)$, взятых построчно, начиная с элементов, расположенных на главной диагонали; его координаты находятся по элементам матрицы $D(t) = \{d_{ij}(t)\}$, $i, j \in [1 : n]$:

$$d_s(t) = d_{ij}(t), \quad i \le j, \quad s = (n - i/2)(i-1) + j. \tag{3.5}$$

Отметим, что соотношения (3.5) между индексами s и i, j взаимно однозначны ввиду способа нумерации элементов части матрицы $D(t)$. Процедура, аналогичная подробно описанной в [20], позволяет преобразовать уравнение (3.4) к виду

$$\dot{d}(t) = \bar{A}(t)d(t) + \bar{B}(d(t), \bar{m}(t))\bar{u}(t), \quad d(t_0) = d_0, \tag{3.6}$$

где матрица $\bar{A}(t) : T \to \mathbb{R}^{n_{yz}\times n_{yz}}$ выписывается явно (см. [20]) $\bar{m}(t) = F(m(t))$, $\bar{u}(t) = F(u_2(t))$, $F : \mathbb{R}^n \to \mathbb{R}^{n_{yz}}$, $\forall a \in \mathbb{R}^n F(a) = \bar{a} = \{\bar{a}_s\}$, $s \in [1 : n_{yz}]$, $\bar{a}_s = a_i a_j$, $i \le j$, $s = (n - i/2)(i-1) + j$ (фактически отображение F удлиняет n-мерный вектор до размерности вектора $d(t) \in \mathbb{R}^{n_{yz}}$), диагональная матрица $\bar{B}(d(t), \bar{m}(t)) : T \to \mathbb{R}^{n_{yz}\times n_{yz}}$ находится из формул

$$\bar{b}_{ss} = d_s + \bar{m}_s, \quad \bar{b}_{sr} = 0, \quad s \ne r, \quad s, r \in [1 : n_{yz}]. \tag{3.7}$$

Из существования, единственности и вида решения уравнения (2.1) следуют существование и единственность решения системы (3.1), (3.6). Очевидно, что вектор $\bar{u}(\cdot) \in L_2(T; \mathbb{R}^{n_{yz}})$ принимает значения из некоторого выпуклого компакта $S_{\bar{u}} \in \mathbb{R}^{n_{yz}}$, который естественным образом строится по компакту $S_{u_2} \in \mathbb{R}^n$ с учетом того, что координаты вектора $\bar{u}(\cdot)$ являются попарными произведениями координат вектора $u_2(\cdot)$. Полагаем, что начальное состояние d_0 и измерения d_i^N получаются из D_0 и D_i^N в соответствии с формулой (3.5). Таким образом, формально по информации о векторе $d(t)$ будем восстанавливать вектор $\bar{u}(t)$. Затем, ограничиваясь рассмотрением тех его координат, которые равны u_{2i}^2, $i \in [1 : n]$, мы покажем, что реальная величина $u_2(t)$ может быть восстановлена при дополнительных, достаточно естественных, предположениях.

Измерения первых q координат вектора x обеспечивают оценку типа (2.2) для $(q^2 + q)/2$ координат вектора d, номера которых находятся из (3.5) и, в общем случае, не идут подряд. Обозначим $n_y = (q^2 + q)/2$, $n_z = n_{yz} - n_y$. Определим I_y как n_y-мерный упорядоченный по возрастанию массив индексов, соответствующих измеряемым координатам вектора d:

$$I_y[p] = s_p, \quad p \in [1 : n_y], \quad s_p \in [1 : n_{yz}],$$

$$p = (q - i/2)(i-1) + j, \quad i, j \in [1 : q], \quad i \le j, \quad s_p = (n - i/2)(i-1) + j,$$

а также, аналогичным образом, I_z как n_z-мерный упорядоченный по возрастанию массив индексов, соответствующих неизмеряемым координатам вектора d.

Введем следующие матрицы:

$$C_1^d = \{c_{1ps}^d\}, \quad p \in [1 : n_y], \quad s \in [1 : n_{yz}], \quad c_{1ps}^d = 1, \quad \text{если} \quad s = I_y[p],$$

$$c_{1ps}^d = 0 \quad \text{в противном случае;}$$

$$D_1^d = \{d_{1ps}^d\}, \quad p \in [1 : n_z], \quad s \in [1 : n_{yz}], \quad d_{1ps}^d = 1, \quad \text{если} \quad s = I_z[p],$$

$$d^d_{1ps} = 0 \quad \text{в противном случае;}$$

$$C^d_2 = (C^d_1)'; \quad D^d_2 = (D^d_1)'.$$

Тогда n_y-мерный вектор $C^d_1 d$ – измеряемая часть d, а n_z-мерный вектор $D^d_1 d$ – его неизмеряемая часть. Обозначим $y_d = C^d_1 d$, $z_d = D^d_1 d$. Умножая уравнение (3.6) поочередно на матрицы C^d_1 и D^d_1, учитывая равенство $d = C^d_2 y_d + D^d_2 z_d$, приходим к системе с разделением измеряемой и неизмеряемой компонент:

$$\dot{y}_d(t) = A^d_{1y}(t)y_d(t) + A^d_{1z}(t)z_d(t) + B^d_1(y_d(t), y_m(t))\overline{u}(t), \quad y_d(0) = y_{d0}, \tag{3.8}$$

$$\dot{z}_d(t) = A^d_{2y}(t)y_d(t) + A^d_{2z}(t)z_d(t) + B^d_2(z_d(t), z_m(t))\overline{u}(t), \quad z_d(0) = z_{d0}, \tag{3.9}$$

где $A^d_{1y} = C^d_1 \overline{A} C^d_2$, $A^d_{1z} = C^d_1 \overline{A} D^d_2$, $y_{d0} = C^d_1 d_0$, $A^d_{2y} = D^d_1 \overline{A} C^d_2$, $A^d_{2z} = D^d_1 \overline{A} D^d_2$, $z_{d0} = D^d_1 d_0$, $B^d_1 = C^d_1 \overline{B}(d, \overline{m})$, $B^d_2 = D^d_1 \overline{B}(d, \overline{m})$, $\overline{m} = F(m) = F(C^m_2 y_m + D^m_2 z_m)$.

Отметим, что ввиду специфики матрицы $\overline{B}$, B^d_1 зависит только от $y_d(t)$, $y_m(t)$, а B^d_2 – только от $z_d(t)$, $z_m(t)$.

Теперь для систем (3.2), (3.3) и (3.8), (3.9) можно переформулировать исходную задачу восстановления. По ходу развития процесса в дискретные моменты времени $\tau_i \in T$, $\tau_i = i\delta$, $\delta = \vartheta/l(N)$, $i \in [0 : (l(N) - 1)]$ поступает информация, позволяющая оценить части фазового состояния указанных систем, соответственно, векторы $y_m(\tau_i)$ и $y_d(\tau_i)$. Полагаем, что выполняется следующее соотношение, соответствующее (2.2):

$$P\left(\max_{i \in [1:(l(N)-1)]} \left\{ \left\| \xi^N_{mi} - y_m(\tau_i) \right\|_{\mathbb{R}^q}, \left\| \xi^N_{di} - y_d(\tau_i) \right\|_{\mathbb{R}^{n_y}} \right\} \le h(N) \right) = 1 - g(N), \tag{3.10}$$

где оценочные векторы $\xi^N_{mi} \in \mathbb{R}^q$ и $\xi^N_{di} \in \mathbb{R}^{n_y}$ естественным образом получаются из оценок m^N_{qi} и D^N_{qi}, а $h(N) \to 0$ и $g(N) \to 0$ при $N \to \infty$.

Требуется указать алгоритм динамического восстановления неизвестных возмущений $u_1(t)$ и $\overline{u}(t)$ по информации (3.10), причем вероятность сколь угодно малого отклонения приближений от искомых входов в метрике соответственно пространств $L_2(T;\mathbb{R}^r)$ и $L_2(T;\mathbb{R}^{n_{yz}})$ должна быть близка к 1 при достаточно большом N и специальным образом согласованном с N шаге временной дискретизации $\delta = \delta(N) = \vartheta/l(N)$.

В такой формулировке задача соответствует задаче, рассмотренной в [4] для случая измерения части координат ОДУ. В настоящей работе показано, что алгоритм, предложенный в [4], применим к решению задачи, полученной для систем (3.2), (3.3) и (3.8), (3.9), поскольку допускает конструктивное согласование своих параметров с количеством доступных измерению реализаций исходного случайного процесса, при этом легко проверяемые достаточные условия разрешимости задачи фактически формулируются в терминах исходного уравнения (2.1).

4. АЛГОРИТМ ВОССТАНОВЛЕНИЯ НЕИЗВЕСТНЫХ ВОЗМУЩЕНИЙ

Разрешающий алгоритм будем строить в случае выполнения следующих условий.

Условие 1. Размерность неизвестной вектор-функции $u_1(\cdot)$ не превосходит размерности компоненты $y_m(\cdot)$ $(r \le q)$, и при всех $t \in T$ матрица $B^m_1(t)$ размерности $q \times r$ имеет ранг, равный r, т.е. является матрицей полного ранга.

Условие 2. Неизвестная вектор-функции $u_2(\cdot)$ имеет следующую структуру: $\exists k \le q$ $\forall i \in [(k+1) : n]$ $\exists j \in [1 : k] : u_{2i} = u_{2j}$.

Таким образом, если количество неизвестных компонент функции $u_2(\cdot)$ не превосходит q, тогда количество неизвестных компонент функции $\overline{u}(\cdot)$, равное $n_k = (k^2 + k)/2$, не превосходит размерности компоненты $y_d(\cdot)$, т.е. $n_k \le n_y$.

Отметим, что, в отличие от условий, гарантирующих реконструкцию возмущения в частично наблюдаемой системе линейных СДУ (см. [12]), в случае квазилинейного СДУ можно потребовать наличие полного ранга матрицы B_1^m, но не матрицы B_1^d, которая зависит от решений систем (3.2), (3.3) и (3.8), (3.9).

Лемма 1. *Пусть все структурные элементы исходного уравнения* (2.1) *(именно, матрицы* A, B, f, вектор x_0, *множества* S_{u_1} и S_{u_2}*) таковы, что решение* $m(t)$ *уравнения* (3.1) *покоординатно строго больше нуля на всем промежутке времени* T, *а решение* $d(t)$ *уравнения* (3.6) *покоординатно неотрицательно на всем промежутке времени* T. *Тогда существует* $\bar{N}$ *такое, что* $\forall N \geq \bar{N}$ *матрица* $B_1^d(\xi_{di}^N, \xi_{mi}^N)$ *с вероятностью* $1 - g(N)$ *является матрицей полного ранга* $\forall i \in [1 : (l(N) - 1)]$.

Доказательство. Отметим, что условие леммы выполняется, например, в случае, когда все элементы матриц A, B, f неотрицательны, все координаты начального вектора m_0 строго больше нуля (тогда как элементы матрицы D_0 неотрицательны вследствие предположения о независимости координат начального вектора x_0), и возмущения u_1 и u_2 принимают неотрицательные значения. Полагаем, что существует величина $\beta > 0$ такая, что $\forall t \in T$, $\forall j \in [1 : n]$ $m_j(t) \geq \beta$.

Полнота ранга матрицы $B_1^d(\xi_{di}^N, \xi_{mi}^N)$ размерности $n_y \times n_{yz}$, ввиду ее структуры, описанной выше, определяется тем, что на местах, соответствующих n_y единицам матрицы C_1^d, расположены ненулевые элементы, являющиеся элементами диагональной матрицы $\bar{B}(d(\tau_i), \bar{m}(\tau_i))$ с индексами, пробегающими множество I_y (см. (3.7)), и при подстановке ξ_{di}^N и ξ_{mi}^N. Фактически, для доказательства леммы следует проверить выполнение соотношений

$$\bar{b}_{ss}(\xi_{di}^N, \xi_{mi}^N) = \xi_{dis}^N + \bar{\xi}_{mis}^N \neq 0, \quad i \in [1 : (l(N) - 1)], \quad s \in I_y, \quad \bar{\xi}_{mi}^N = F(\xi_{mi}^N). \tag{4.1}$$

Интересующие нас элементы матрицы $\bar{B}(d(\tau_i), \bar{m}(\tau_i))$ можно разделить на два типа, согласно (3.7): первый представляет собой сумму дисперсии измеряемой координаты и квадрата ее математического ожидания, т.е.

$$\bar{b}_1(\tau_i) = d_{jj}(\tau_i) + m_j^2(\tau_i) = M(x_j^2(\tau_i)), \quad j \in [1 : q],$$

второй – сумму ковариации двух измеряемых координат и произведения их математических ожиданий, т.е.

$$\bar{b}_2(\tau_i) = d_{jp}(\tau_i) + m_j(\tau_i)m_p(\tau_i) = M(x_j(\tau_i)x_p(\tau_i)), \quad j, p \in [1 : q], \quad j \neq p.$$

Очевидно, что все $\bar{b}_1(\tau_i) \geq \beta^2$ и $\bar{b}_2(\tau_i) \geq \beta^2$, $i \in [1 : (l(N) - 1)]$. Для доказательства утверждения леммы рассмотрим, к примеру, соотношение, получаемое для элементов второго типа (индекс s соответствует паре (j, p), см. (3.7), (4.1)):

$$\begin{aligned}\left|d_{jp}(\tau_i) + m_j(\tau_i)m_p(\tau_i) - \xi_{dis}^N - \xi_{mij}^N\xi_{mip}^N\right| &\leq \left|d_{jp}(\tau_i) - \xi_{dis}^N\right| + \left|m_j(\tau_i)m_p(\tau_i) - \xi_{mij}^N\xi_{mip}^N\right| \leq \\ &\leq \left|d_{jp}(\tau_i) - \xi_{dis}^N\right| + \left|m_j(\tau_i)(m_p(\tau_i) - \xi_{mip}^N)\right| + \left|(m_j(\tau_i) - \xi_{mij}^N)\xi_{mip}^N\right| \leq \bar{K}h(N).\end{aligned}$$

Данное неравенство получено из (3.10) и выполняется с вероятностью $1 - g(N)$. Здесь $\bar{K}$ – некоторая константа, которая может быть выписана явно. Отсюда выводим

$$\xi_{dis}^N + \xi_{mij}^N\xi_{mip}^N \geq \beta^2 - \bar{K}h(N) = \beta^2/2.$$

Полагаем, что последнее равенство выполняется $\forall N \geq \bar{N}$; $\bar{N}$ существует, поскольку $h(N) \to 0$ при $N \to \infty$. Соотношение (4.1) выполняется, следовательно, лемма доказана.

Выполнение условий 1, 2 и леммы 1 гарантирует единственность функций $u_1(\cdot)$ и $\bar{u}(\cdot)$, определяющих решения $(y_m(\cdot), z_m(\cdot))$ и $(y_d(\cdot), z_d(\cdot))$ систем (3.2), (3.3) и (3.8), (3.9) соответственно. Обоснование этого утверждения опирается на свойства псевдообратной матрицы для матрицы полного ранга (см. [4], [21]), т.е. в данной постановке на свойства псевдообратных матриц для B_1^m и B_1^d, и, за вычетом малозначительных деталей, следует аналогичным рассуждениям из [4], [5].

Алгоритм, приведенный ниже, является приложением вычислительной процедуры из [4] к системам (3.2), (3.3) и (3.8), (3.9). В начальный момент $\tau_0 = 0$ фиксируется значение N, опреде-

ляются величины $l^N = l(N)$, $h^N = h(N)$ и $g^N = g(N)$ (см. (3.10)) и строится равномерное разбиение промежутка T с шагом $\delta^N = \vartheta/l^N$: $\tau_i \in T$, $\tau_i = i\delta^N$, $i \in [0 : l^N]$. Вводится управляемая система-модель, фактически содержащая два блока, относящихся к системам (3.2), (3.3) и (3.8), (3.9). Фазовый вектор модели обозначим через $w(t)$; он состоит из двух троек: (i) q-мерного вектора $w_{my}(t)$, $(n-q)$-мерного вектора $w_{mz}(t)$, q-мерного вектора $w_{mu}(t)$ и (ii) n_y-мерного вектора $w_{dy}(t)$, n_z-мерного вектора $w_{dz}(t)$, n_y-мерного вектора $w_{dv}(t)$. Динамика модели и ее начальное состояние определяются соотношениями

$$\dot w_{my}(t) = \overline{u}_{1i}^N, \quad \dot w_{dy}(t) = \overline{v}_i^N,$$

$$\dot w_{mz}(t) = A_{2y}^m(\tau_i)\xi_{mi}^N + A_{2z}^m(\tau_i)w_{mz}(\tau_i) + B_2^m(\tau_i)B_1^{m+}(\tau_i)(\overline{u}_{1i}^N - A_{1y}^m(\tau_i)\xi_{mi}^N - A_{1z}^m(\tau_i)w_{mz}(\tau_i) - C_1^m f(\tau_i)) + D_1^m f(\tau_i),$$

$$\dot w_{dz}(t) = A_{2y}^d(\tau_i)\xi_{di}^N + A_{2z}^d(\tau_i)w_{dz}(\tau_i) + B_2^d(w_{dz}(\tau_i), w_{mz}(\tau_i))B_1^{d+}(\xi_{di}^N, \xi_{mi}^N)(\overline{v}_i^N - A_{1y}^d(\tau_i)\xi_{di}^N - A_{1z}^d(\tau_i)w_{dz}(\tau_i)),$$

$$\dot w_{mu}(t) = A_{1y}^m(\tau_i)\xi_{mi}^N + A_{1z}^m(\tau_i)w_{mz}(\tau_i) + B_1^m(\tau_i)\hat u_{1i}^N + C_1^m f(\tau_i), \tag{4.2}$$

$$\dot w_{dv}(t) = A_{1y}^d(\tau_i)\xi_{di}^N + A_{1z}^d(\tau_i)w_{dz}(\tau_i) + B_1^d(\xi_{di}^N, \xi_{mi}^N)\hat v_i^N,$$

$$t \in (\tau_i, \tau_{i+1}], \quad i \in [0 : (l^N - 1)],$$

$$w_{my}(\tau_0) = y_{m0}, \quad w_{mz}(\tau_0) = z_{m0}, \quad w_{mu}(\tau_0) = y_{m0}, \quad w_{dy}(\tau_0) = y_{d0}, \quad w_{dz}(\tau_0) = z_{d0}, \quad w_{dv}(\tau_0) = y_{d0}.$$

Здесь B_1^{m+}, B_1^{d+} – псевдообратные матрицы, $\overline{u}_{1i}^N$, $\overline{v}_i^N$, $\hat u_{1i}^N$, $\hat v_i^N$ – управляющие воздействия соответствующих размерностей, вычисляемые позиционно в момент τ_i по правилам, которые конкретизируются ниже.

Динамика (4.2) выбирается из следующих соображений. Движение вспомогательных компонент $w_{my}(t)$ и $w_{dy}(t)$ при подходящем выборе модельных управлений

$$\overline{u}_1^N(t) = \overline{u}_{1i}^N, \quad \overline{v}^N(t) = \overline{v}_i^N, \quad t \in (\tau_i, \tau_{i+1}], \quad i \in [0 : (l^N - 1)],$$

обеспечивает аппроксимацию координат $y_m(t)$ и $y_d(t)$. Тогда, пользуясь оценкой (3.10) и формальным выражением возмущений $u_1(t)$ и $\overline{u}(t)$ из уравнений (3.2) и (3.8) с заменой $\dot y_m(t)$ и $\dot y_d(t)$ на $\overline{u}_{1i}^N$ и $\overline{v}_i^N$ соответственно, ожидаем близость $w_{mz}(t)$ к $z_m(t)$ и $w_{dz}(t)$ к $z_d(t)$, что, в свою очередь, делает возможным отслеживание компонентами $w_{mu}(t)$ и $w_{dv}(t)$ координат $y_m(t)$ и $y_d(t)$ посредством выбора модельных управлений

$$\hat u_1^N(t) = \hat u_{1i}^N, \quad \hat v^N(t) = \hat v_i^N, \quad t \in (\tau_i, \tau_{i+1}], \quad i \in [0 : (l^N - 1)],$$

приближающих в нужном смысле функции $u_1(t)$ и $\overline{u}(t)$. Очевидно, нетрудно записать дискретный аналог модели (4.2).

Работа алгоритма разбивается на l^N однотипных шагов. На i-м шаге, который выполняется на интервале $(\tau_i, \tau_{i+1}]$, исходными данными для вычислений служат оценки ξ_{mi}^N, ξ_{di}^N и сформированное к этому моменту состояние модели $w(\tau_i)$. Предполагая покоординатную ограниченность правых частей уравнений (3.2) и (3.8) константой $\overline{K}$ (ее существование очевидно), находим s-ю координату $\overline{u}_{1is}^N$ вектора $\overline{u}_{1i}^N$ и s-ю координату $\overline{v}_{is}^N$ вектора $\overline{v}_i^N$ из соотношений

$$\begin{aligned} \overline{u}_{1is}^N &= -\overline{K}\,\mathrm{sign}(w_{mys}(\tau_i) - \xi_{mis}^N), \quad s \in [1 : q], \\ \overline{v}_{is}^N &= -\overline{K}\,\mathrm{sign}(w_{dys}(\tau_i) - \xi_{dis}^N), \quad s \in [1 : n_y]. \end{aligned} \tag{4.3}$$

Вторая пара модельных управлений определяется следующим образом: $\hat u_i^N$ и $\hat v_i^N$ суть единственные решения экстремальных задач

$$\begin{aligned} \hat u_{1i}^N &= \arg\min\left\{2\langle w_{mu}(\tau_i) - \xi_{mi}^N, B_1^m(\tau_i)u\rangle + \alpha^N \|u\|_{\mathbb{R}^r}^2 : u \in S_{u_1}\right\}, \\ \hat v_i^N &= \arg\min\left\{2\langle w_{dv}(\tau_i) - \xi_{di}^N, B_1^d(\xi_{di}^N, \xi_{mi}^N)v\rangle + \alpha^N \|v\|_{\mathbb{R}^{n_{yz}}}^2 : v \in S_{\overline{u}}\right\}, \end{aligned} \tag{4.4}$$

где $\langle\cdot,\cdot\rangle$ – скалярное произведение в соответствующем евклидовом пространстве, $\alpha^N = \alpha(h^N)$ – параметр регуляризации. После вычисления управлений по формулам (4.3) и (4.4) состояние модели $w(\tau_{i+1})$ пересчитывается согласно дискретному аналогу (4.2). Процесс заканчивается в конечный момент времени ϑ.

Теорема 1. *Пусть выполняются условия согласования параметров*

$$h^N \to 0, \quad g^N \to 0, \quad \delta^N \to 0, \quad \alpha^N \to 0, \quad \frac{\delta^N + h^N}{\alpha^N} \to 0 \quad \text{при} \quad N \to \infty. \tag{4.5}$$

Тогда для модельных управлений $\hat{u}_1^N(\cdot)$ и $\hat{v}^N(\cdot)$, формируемых согласно (4.4), *имеет место сходимость при $N \to \infty$*:

$$P\left(\max\left\{\left\|\hat{u}_1^N(\cdot) - u_1(\cdot)\right\|_{L_2(T;\mathbb{R}^r)}, \left\|\hat{v}^N(\cdot) - \overline{u}(\cdot)\right\|_{L_2(T;\mathbb{R}^{n_{yz}})}\right\} \to 0\right) \to 1. \tag{4.6}$$

При дополнительных предположениях об ограниченности вариации исходных возмущений $u_1(\cdot)$ и $u_2(\cdot)$ справедлива следующая оценка точности алгоритма относительно количества реализаций процесса, доступных измерению:

$$P\left(\max\left\{\left\|\hat{u}_1^N(\cdot) - u_1(\cdot)\right\|_{L_2(T;\mathbb{R}^r)}, \left\|\hat{v}^N(\cdot) - \overline{u}(\cdot)\right\|_{L_2(T;\mathbb{R}^{n_{yz}})}\right\} \le C_1\left(\frac{1}{N}\right)^{2/13}\right) = 1 - C_2\left(\frac{1}{N}\right)^{2/13}, \tag{4.7}$$

где C_1 и C_2 – некоторые константы, не зависящие от N, $u_1(\cdot)$ и $\overline{u}(\cdot)$.

Доказательство теоремы, за вычетом некоторых технических деталей, повторяет доказательство соответствующего утверждения для алгоритма реконструкции двух возмущений в случае измерения части координат линейного СДУ (см. [12]). В частности, там показано, что стандартные оценки m_i^N математического ожидания $m(\tau_i)$ и D_i^N ковариационной матрицы $D(\tau_i)$, построенные по N $(N > 1)$ реализациям $x^1(\tau_i), x^2(\tau_i), \ldots, x^N(\tau_i)$ случайных величин $x(\tau_i)$, $i \in [1 : l^N]$, по следующим правилам (см. [18]):

$$m_i^N = \frac{1}{N}\sum_{r=1}^{N} x^r(\tau_i), \qquad D_i^N = \frac{1}{N-1}\sum_{r=1}^{N}(x^r(\tau_i) - m_i^N)(x^r(\tau_i) - m_i^N)', \tag{4.8}$$

с учетом измерения q первых координат, обеспечивают выполнение свойства (2.2) (и, следовательно, (3.10)).

Теорема 2. *Если искомый вектор $u_2(\cdot)$ является единственным, и его координаты неотрицательны, т.е. S_{u_2} таково, что $u_{2i}(t) \ge 0$ $\forall t \in T$ $\forall i \in [1 : n]$, то алгоритм* (4.2)–(4.5) *восстанавливает $u_2(\cdot)$ в смысле $L_2(T;\mathbb{R}^n)$-метрики.*

Доказательство. Отметим, что ограничение $u_{2i}(t) \ge 0$ представляется естественным для шума с нулевым средним. Рассмотрим n-мерный вектор $\overline{\overline{u}}(\cdot)$, состоящий из координат вектора $\overline{u}(\cdot)$, которые равны u_{2i}^2, $i \in [1 : n]$, и вектор $\overline{\overline{v}}^N(\cdot)$, состоящий из соответствующих координат вектора $\hat{v}^N(\cdot)$. Из структуры множества $S_{\overline{u}}$ и формул (4.4) для нахождения $\hat{v}^N(\cdot)$ следует, что $\overline{\overline{v}}_i^N(t) \ge 0$ $\forall t \in T$ $\forall i \in [1 : n]$. Поэтому возможно положить $v_{2i}^N(t) = \sqrt{\overline{\overline{v}}_i^N(t)}$. Тогда из (4.6) получаем

$$P\left(\int_T \sum_{i=1}^{n}\left(\overline{\overline{v}}_i^N(s) - \overline{\overline{u}}_i(s)\right)^2 ds \to 0\right) \to 1,$$

$$P\left(\int_T \sum_{i=1}^{n}\left(v_{2i}^N(s) - u_{2i}(s)\right)^2\left(v_{2i}^N(s) + u_{2i}(s)\right)^2 ds \to 0\right) \to 1,$$

$$P\left(\int_T \sum_{i=1}^{n}\left(v_{2i}^N(s) - u_{2i}(s)\right)^4 ds \to 0\right) \to 1 \quad \text{при} \quad N \to \infty.$$

Используя неравенства

$$\left(\int_T \left(v_{2i}^N(s) - u_{2i}(s)\right)^2 ds\right)^2 \le \overline{C}_{15} \int_T \left(v_{2i}^N(s) - u_{2i}(s)\right)^4 ds,$$

выводим

$$P\left(\int_T \sum_{i=1}^n \left(v_{2i}^N(s) - u_{2i}(s)\right)^2 ds \to 0\right) \to 1.$$

Таким образом, имеем сходимость, соответствующую (4.6) и доказывающую аппроксимацию в $L_2(T;\mathbb{R}^n)$-метрике возмущения $u_2(\cdot)$ вектором $v_2^N(\cdot)$. Порядок полученной при дополнительных предположениях оценки точности алгоритма относительно величины $1/N$ (см. (4.7)) сохраняется.

5. ЧИСЛЕННЫЙ ПРИМЕР

Перейдем к описанию модельного примера. Рассмотрим квазилинейное уравнение, описывающее случайный процесс, который можно трактовать как "испорченный" средневозвратный процесс Орнштейна–Уленбека (см. [17]):

$$\begin{gathered} dx_1(t) = (-2x_1(t) + 0.1x_2(t) + u_1(t) + 1)dt + u_2(t)x_1(t)d\xi(t), \\ dx_2(t) = (0.1x_1(t) - x_2(t) + 2u_1(t) + 2)dt + u_2(t)x_2(t)d\xi(t), \\ t \in T = [0,1], \quad x_1(0) = 1, \quad x_2(0) = 1, \quad u_1 \in [0,1], \quad u_2 \in [1,2]. \end{gathered} \tag{5.1}$$

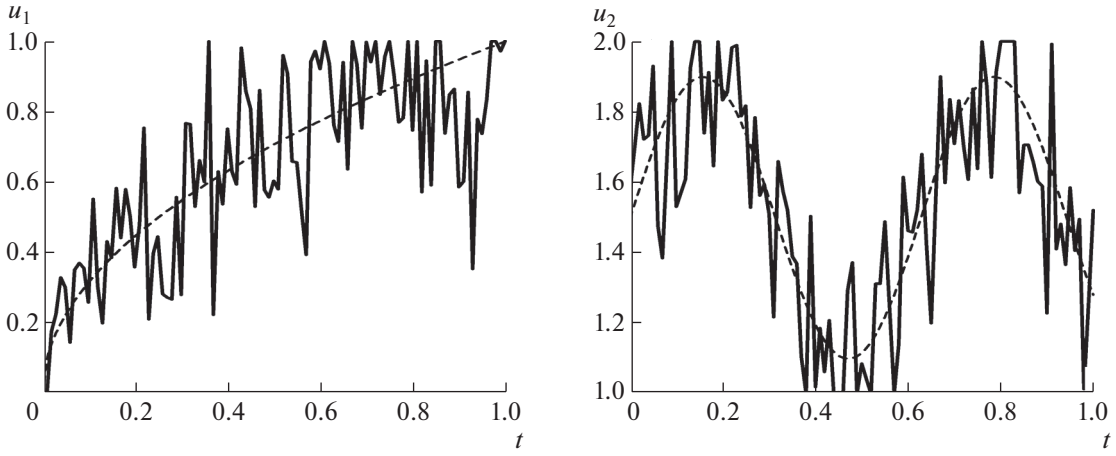
Уравнения типа (5.1) используются, в частности, в некоторых простейших моделях, описывающих динамику численности относительно устойчивых популяций живых организмов (см. [17]). В таком случае величины $x_1(t)$ и $x_2(t)$ представляют текущие численности (в условных единицах) двух взаимодействующих популяций; структура детерминированной части уравнения определяет численности (например, средние за некоторую предысторию), стремление вернуться к которым "подсознательно" присутствует у популяции. Такому возврату могут препятствовать как влияние соседей, так и воздействие внешних факторов через функцию $u_1(t)$ в детерминированной части и функцию $u_2(t)$, входящую в амплитуду случайных колебаний. Именно внешние возмущения $u_1(t)$ и $u_2(t)$ подлежат восстановлению. Измерению в дискретные моменты времени доступны траектории координаты $x_1(t)$, соответствующие различным колониям организмов первого типа.

Запишем "редуцированные" системы вида (3.2), (3.3) и (3.8), (3.9), описывающие, соответственно, динамику математического ожидания $m(t)$ и ковариационной матрицы $D(t) = \begin{pmatrix} d_1 & d_2 \\ d_2 & d_3 \end{pmatrix}$:

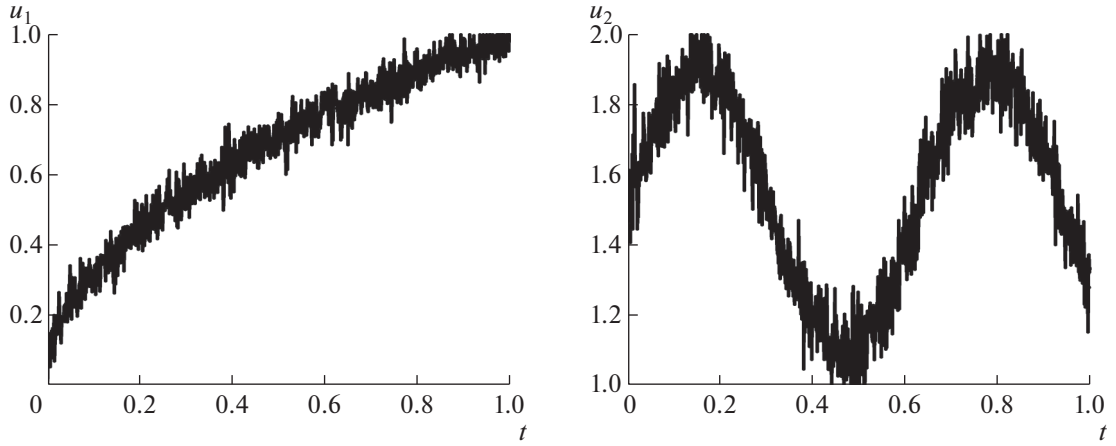
$$\begin{aligned} \dot{m}_1(t) &= -2m_1(t) + 0.1m_2(t) + u_1(t) + 1, \quad m_1(0) = 1, \\ \dot{m}_2(t) &= 0.1m_1(t) - m_2(t) + 2u_1(t) + 2, \quad m_2(0) = 1; \end{aligned} \tag{5.2}$$

$$\begin{gathered} \dot{d}_1(t) = -4d_1(t) + 0.2d_2(t) + (d_1(t) + m_1^2(t))u_2^2(t), \quad d_1(0) = 0, \\ \dot{d}_2(t) = 0.1d_1(t) - 3d_2(t) + 0.1d_3(t) + (d_2(t) + m_1(t)m_2(t))u_2^2(t), \quad d_2(0) = 0, \\ \dot{d}_3(t) = 0.2d_2(t) - 2d_3(t) + (d_3(t) + m_2^2(t))u_2^2(t), \quad d_3(0) = 0. \end{gathered} \tag{5.3}$$

Фактически в дискретные моменты времени в системе (5.2) измеряется координата $m_1(t)$, в системе (5.3) – координата $d_1(t)$ (в соответствии с оценкой (3.10)), восстанавливаются величины $u_1(t)$ и $u_2(t)$. В данном случае $\overline{u}(t) = u_2^2(t)$, $\overline{u} \in [1,4]$, и выполняются условия реконструкции самой функции $u_2(t)$. Аналог модели (4.2) для систем (5.2), (5.3) выписывается естественным образом. Кроме того, условие 1 выполняется, так как $B_1^m(t) \equiv 1$, а справедливость леммы 1 для $B_1^d(\xi_{di}^N, \xi_{mi}^N) = \xi_{d1i}^N + (\xi_{m1i}^N)^2$ обеспечивается положительностью решений (5.2), (5.3) и достаточной



Фиг. 1. Параметры: $N = 10^4$, $\delta^N = 0.01$, $\alpha^N = 0.04$, $\delta_s = \delta^N/10^4$; погрешности: $\left\|\hat{u}_1^N(\cdot) - u_1(\cdot)\right\|_{L_2} = 0.953$, $\left\|v_2^N(\cdot) - u_2(\cdot)\right\|_{L_2} = 0.912$.



Фиг. 2. Параметры: $N = 10^6$, $\delta^N = 0.001$, $\alpha^N = 0.009$, $\delta_s = \delta^N/10^4$; погрешности: $\left\|\hat{u}_1^N(\cdot) - u_1(\cdot)\right\|_{L_2} = 0.096$, $\left\|v_2^N(\cdot) - u_2(\cdot)\right\|_{L_2} = 0.104$.

точностью оценок (3.10). В вычислительном эксперименте были выбраны следующие реализации неизвестных функций:

$$u_1(t) = \sqrt{t}, \quad u_2(t) = 1.5 + 0.4\sin 10t.$$

Для моделирования динамики системы (5.1) использовался метод Эйлера с заменой винеровского процесса последовательностью случайных импульсов (подробное описание аппроксимационной схемы можно найти, например, в [22]); его среднеквадратический порядок точности для квазилинейного СДУ равен $O(\delta_s)$, где δ_s – шаг моделирования. Этот шаг, как правило, очень мал относительно шага δ^N, с которым поступает информация о траекториях системы, обрабатываемая согласно (4.8), и, соответственно, работает процедура восстановления неизвестных функций $u_1(t)$ и $u_2(t)$. Для получения оценок (3.10) необходимо отслеживание N независимых траекторий системы (5.1).

Результаты восстановления функций $u_1(t)$ и $u_2(t)$, полученные для различных наборов параметров алгоритма (4.2)–(4.5), приведены на фиг. 1 и фиг. 2, где реальные функции $u_1(t)$ и $u_2(t)$ изображены штриховой линией, а результаты их восстановления – сплошной. Они согласуются

с основным утверждением статьи, подтверждая сходимость типа (4.6)–(4.7): чем больше количество N измеряемых траекторий, тем меньше (с вероятностью, близкой к 1) погрешность восстановления.

6. ЗАКЛЮЧЕНИЕ

Таким образом, исследована постановка обратной задачи для квазилинейного СДУ, предполагающая динамическую реконструкцию двух неизвестных неслучайных возмущений, входящих в детерминированный и стохастический члены уравнения. В качестве входной информации используются точные измерения некоторого количества реализаций части координат случайного процесса в дискретные моменты времени. Решение задачи состоит в переходе к обратной задаче для двух систем ОДУ специального вида с последующим использованием модификации алгоритма восстановления, разработанного ранее для частично наблюдаемых систем ОДУ. Доказанная оценка точности алгоритма относительно количества доступных измерению реализаций проиллюстрирована на модельном примере. Ряд вопросов, связанных с рассматриваемой постановкой, остается открытым: например, об улучшении порядка точности оценки (4.7) при дополнительных предположениях и о возможности введения погрешности измерения траекторий исходного СДУ.

СПИСОК ЛИТЕРАТУРЫ

1. *Кряжимский А.В., Осипов Ю.С.* О моделировании управления в динамической системе // Изв. АН СССР. Техн. кибернетика. 1983. № 2. С. 51–60.
2. *Osipov Yu.S., Kryazhimskii A.V.* Inverse problems for ordinary differential equations: dynamical solutions. London: Gordon and Breach, 1995.
3. *Максимов В.И.* Задачи динамического восстановления входов бесконечномерных систем. Екатеринбург: Изд-во УрО РАН, 2000.
4. *Кряжимский А.В., Осипов Ю.С.* Об устойчивом позиционном восстановлении управления по измерениям части координат // Некоторые задачи управления и устойчивости. Свердловск: УрО АН СССР, 1989. С. 33–47.
5. *Осипов Ю.С., Кряжимский А.В., Максимов В.И.* Методы динамического восстановления входов управляемых систем. Екатеринбург: Изд-во УрО РАН, 2011.
6. *Сурков П.Г.* Задача динамического восстановления правой части системы дифференциальных уравнений нецелого порядка // Дифференц. ур-ния. 2019. Т. 55. № 6. С. 865–874.
7. *Красовский Н.Н., Субботин А.И.* Позиционные дифференциальные игры. М.: Наука, 1984.
8. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М.: Наука, 1978.
9. *Осипов Ю.С., Кряжимский А.В.* Позиционное моделирование стохастического управления в динамических системах // Докл. междунар. конф. по стохастической оптимизации. Киев, 1984. С. 43–45.
10. *Кряжимский А.В., Осипов Ю.С.* О динамической регуляризации при случайных помехах // Тр. Матем. ин-та им. В.А. Стеклова. 2010. Т. 271. С. 134–147.
11. *Розенберг В.Л.* Задача динамического восстановления неизвестной функции в линейном стохастическом дифференциальном уравнении // Автоматика и телемехан. 2007. № 11. С. 76–87.
12. *Розенберг В.Л.* Восстановление внешних воздействий при дефиците информации в линейном стохастическом уравнении // Тр. Ин-та матем. и механ. УрО РАН. 2016. Т. 22. № 2. С. 236–244.
13. *Розенберг В.Л.* Динамическая реконструкция возмущений в квазилинейном стохастическом дифференциальном уравнении // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 7. С. 1121–1131.
14. *Rozenberg V.L.* On a problem of dynamical input reconstruction for a system of special type under conditions of uncertainty // AIMS Math. 2020. V. 5. № 5. P. 4108–4120.
15. *Румянцев Д.С., Хрусталев М.М.* Оптимальное управление квазилинейными системами диффузионного типа при неполной информации о состоянии // Изв. РАН. Теория и системы управления. 2006. № 5. С. 43–51.
16. *Ширяев А.Н.* Вероятность, статистика, случайные процессы. М.: Изд-во МГУ, 1974.
17. *Оксендаль Б.* Стохастические дифференциальные уравнения. Введение в теорию и приложения. М.: Мир, 2003.
18. *Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф.* Справочник по теории вероятностей и математической статистике. М.: Наука, 1985.
19. *Черноусько Ф.Л., Колмановский В.Б.* Оптимальное управление при случайных возмущениях. М.: Наука, 1978.
20. *Rozenberg V.L.* A control problem under incomplete information for a linear stochastic differential equation // Ural Math. J. 2015. V. 1. № 1. P. 68–82.
21. *Гантмахер Ф.Р.* Теория матриц. М.: Наука, 1988.
22. *Мильштейн Г.Н.* Численное интегрирование стохастических дифференциальных уравнений. Свердловск: Изд-во УрГУ, 1988.

EDN: PLJCWH
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1851–1860

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.956.4

АСИМПТОТИЧЕСКОЕ РЕШЕНИЕ ЗАДАЧИ ГРАНИЧНОГО УПРАВЛЕНИЯ ДЛЯ УРАВНЕНИЯ ТИПА БЮРГЕРСА С МОДУЛЬНОЙ АДВЕКЦИЕЙ И ЛИНЕЙНЫМ УСИЛЕНИЕМ[1)]

© 2022 г. В. Т. Волков[1],*, Н. Н. Нефедов[1],**

[1] *119991 Москва, Ленинские горы, МГУ, физический факультет, Россия*

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

***e-mail: nefedov@phys.msu.ru*

Поступила в редакцию 15.10.2021 г.
Переработанный вариант 04.04.2022 г.
Принята к публикации 08.06.2022 г.

Рассмотрена сингулярно возмущенная периодическая задача для параболического уравнения реакция–диффузия–адвекция типа Бюргерса с модульной адвекцией и линейным усилением. Получены условия существования, единственности и асимптотической устойчивости по Ляпунову периодического решения с внутренним переходным слоем и построено его асимптотическое приближение. Асимптотический анализ применен при решении задачи граничного управления для достижения требуемого закона движения фронта. Сформулировано понятие асимптотического решения этой задачи, получены достаточные условия существования и единственности решения, построено асимптотическое приближение ее решения. Библ. 22.

Ключевые слова: сингулярно возмущенные параболические уравнения, периодические задачи, уравнения реакция–диффузия, контрастные структуры, внутренние слои, фронты, асимптотические методы, дифференциальные неравенства, асимптотическая устойчивость по Ляпунову, уравнения Бюргерса с модульной адвекцией, коэффициентная обратная задача, асимптотическое решение обратной задачи.

DOI: 10.31857/S0044466922110138

1. ВВЕДЕНИЕ

В настоящей работе асимптотический анализ применен при решении задачи граничного управления для сингулярно возмущенного уравнения типа Бюргерса (уравнения реакция–диффузия–адвекция) с модульной адвекцией и линейным источником (линейным усилением). Рассмотрен случай решения с внутренним переходным слоем и построено асимптотическое приближение такого решения. Доказано существование решения с построенной асимптотикой и его асимптотическая устойчивость по Ляпунову. Под задачей граничного управления понимается задача определения граничных условий, при которых достигается заданный режим движения фронта. Отметим, что сформулированная задача, несмотря на определенную аналогичность, отличается от обратной задачи определения параметров модели по наблюдению за движущимся слоем, так как в этом случае наблюдаемый режим входит в число реализуемых (достижимых) в рамках условий прямой задачи, определяющих модель. В рассматриваемой задаче это требует исследования, что определяет новизну развиваемого подхода.

Подобные задачи возникают в газовой динамике, в нелинейной теории волн, биофизике, химической кинетике и многих других практических приложениях и описываются нелинейными параболическими уравнениями с малыми параметрами при производных (см., например, [1] и ссылки в этой работе). К этому классу задач относятся уравнения Бюргерса (см. [2], [3]) и уравнения типа Бюргерса (см. [4]–[8]), которые интенсивно изучаются в связи с тем, что они выступают в качестве математических моделей, выявляющих основные механизмы, определяющие поведение и более сложных моделей нелинейной теории волн. Особенностью задач указанного типа является то, что их решения могут содержать узкие пограничные и/или внутренние слои –

[1)]Работа выполнена при финансовой поддержке РНФ (проект 18-11-00042).

стационарные и/или движущиеся фронты. Поэтому как прямые, так и соответствующие обратные задачи чрезвычайно сложны для численного решения, что требует развития специальных методов исследования (см., например, обзор [9], где изложены базовые идеи, развиваемые в этих классах задач, и более позднюю работу [10]). Асимптотический анализ сингулярно возмущенных периодических краевых задач типа реакция–диффузия–адвекция может быть найден, например, в [11]–[19]. Настоящая работа посвящена исследованию нового класса обратных коэффициентных задач для таких уравнений. Обратные коэффициентные задачи широко исследуются в связи со многими приложениями (см., например, [20]–[22] и ссылки в этих работах). Но заметим, что развиваемые подходы не являются эффективными для сингулярно возмущенных уравнений. Поэтому применение асимптотического анализа, помимо повышения устойчивости стандартных методов решения обратных задач, позволяет получить совершенно новый подход к решению таких задач. Важной особенностью асимптотического подхода к исследованию нелинейных дифференциальных уравнений с малыми параметрами является то, что асимптотический анализ позволяет свести исходную нелинейную сингулярно возмущенную задачу к набору более простых задач, дающих возможность установить более простые связи между входными данными и искомыми параметрами обратной задачи, некорректность которых существенно ниже исходной задачи или вовсе отсутствует. Эти идеи были применены в [13]–]15].

В настоящей работе получила дальнейшее развитие концепция асимптотического решения обратных задач, предложенная авторами в [16], [17], и состоящая в том, что асимптотический анализ позволяет свести исходную коэффициентную задачу к более простой некорректно поставленной, а в ряде случаев – к корректно поставленной задаче. При этом отметим, что в указанных работах решалась обратная задача, а уравнение, описывающее главный член в асимптотике движения слоя, было конечным (алгебраическим).

Структура работы такова. В разд. 2 и 3 приведена постановка прямой задачи, сформулирована теорема существования решения, в рамках условий которой получено асимптотическое приближение решения прямой задачи и решается обратная задача граничного управления. В разд. 4 приводится и обсуждается асимптотическое решение задачи граничного управления.

2. ПОСТАНОВКА ПРЯМОЙ И ОБРАТНОЙ ЗАДАЧИ

Рассмотрим задачу для нелинейного сингулярно возмущенного уравнения типа реакция–диффузия–адвекция, называемого в приложениях уравнением типа Бюргерса и применяемого, например, в нелинейной теории волн для описания нелинейных волн в среде без дисперсии с линейным усилением (см. [5]–[8]). А именно,

$$\begin{gathered}\varepsilon\frac{\partial^2 u}{\partial x^2}-\frac{\partial u}{\partial t}+\frac{\partial |u|}{\partial x}-K\cdot u=0,\quad (x,t)\in D:=\{x\in(-1,1);\ t\in\mathbb{R}\},\\ u(-1,t;\varepsilon)=u^{(-)}(t),\quad u(1,t;\varepsilon)=u^{(+)}(t),\quad t\in\mathbb{R},\\ u(x,0;\varepsilon)=u(x,t+T;\varepsilon),\quad x\in[-1,1],\quad t\in\mathbb{R},\end{gathered}\tag{1}$$

где ε – малый параметр ($0<\varepsilon\ll 1$), а функции $u^{(-)}(t)$ и $u^{(+)}(t)$ – достаточно гладкие и T-периодические по переменной t, $K>0$ – постоянная.

Для этой задачи (*прямая задача*) будет сформулирована теорема существования решения, в рамках которой будет поставлена обратная задача граничного управления и получено ее асимптотическое решение.

Рассматриваемая нами прямая задача (1) является частным случаем изученной в [10]:

$$\begin{gathered}\varepsilon\frac{\partial^2 u}{\partial x^2}-\frac{\partial u}{\partial t}+\frac{\partial |u|}{\partial x}-B(u,x,t)=0,\quad (x,t)\in D:=\{x\in(-1,1);\ t\in\mathbb{R}\},\\ u(-1,t;\varepsilon)=u^{(-)}(t),\quad u(1,t;\varepsilon)=u^{(+)}(t),\quad t\in\mathbb{R},\\ u(x,0;\varepsilon)=u(x,t+T;\varepsilon),\quad x\in[-1,1],\quad t\in\mathbb{R},\end{gathered}\tag{2}$$

где $0<\varepsilon\ll 1$ – малый параметр, а функции $B(u,x,t)$, $u^{(-)}(t)$ и $u^{(+)}(t)$ – достаточно гладкие и T-периодические по переменной t.

Введем определения: $D_T:=(t,x)\in(\mathbb{R}\times(-1;1))$, $D_T^{(-)}:=(t,x)\in(\mathbb{R}\times(-1;x_{\mathrm{tr}}))$, $D_T^{(+)}:=(t,x)\in(\mathbb{R}\times\times(x_{\mathrm{tr}};1))$, где $x_{\mathrm{tr}}=x_{\mathrm{tr}}(t;\varepsilon)$ – T-периодическая функция, причем $-1<x_{\mathrm{tr}}(t;\varepsilon)<1$ при всех $t\in\mathbb{R}$.

Решением задачи (2) назовем T-периодическую функцию $u(x,t;\varepsilon) \in C(\bar{D}_T) \cap C^1(D_T) \cap \cap C^{1,2}(D_T^{(-)} \cup D_T^{(+)})$, удовлетворяющую уравнению (2) в каждой из подобластей $D_T^{(-)}$ и $D_T^{(+)}$, а также граничным условиям.

В [10], [19] при определенных условиях на входящие в задачу функции было доказано существование периодического решения, его асимптотическая устойчивость по Ляпунову, а также получено асимптотическое приближение решения по параметру ε. В этих работах получены условия, при которых задача (2) имеет T-периодическое по переменной t решение вида движущегося фронта: на интервале $(-1,1)$ существует точка $x_{\mathrm{tr}}(t;\varepsilon)$, движущаяся по периодическому во времени закону, в окрестности которой наблюдается узкий внутренний переходный слой, т.е. слева от указанной точки (при $-1 < x < x_{\mathrm{tr}}(t;\varepsilon)$) решение близко к некоторому уровню $\varphi^{(-)}(x,t)$, а справа (при $x_{\mathrm{tr}}(t;\varepsilon) < x < 1$) – к $\varphi^{(+)}(x,t)$. Указанные функции $\varphi^{(-)}(x,t)$ и $\varphi^{(+)}(x,t)$ являются решениями вырожденного уравнения и определены ниже в условии 2.

Положение точки перехода $x_{\mathrm{tr}}(t;\varepsilon)$ заранее неизвестно и определяется (при фиксированном t) условием пересечения решения $u(x,t;\varepsilon)$ и некоторого уровня между $\varphi^{(-)}(x,t)$ и $\varphi^{(+)}(x,t)$ (в рассматриваемой задаче – нулевого уровня). Таким образом, положение точки перехода $x_{\mathrm{tr}}(t;\varepsilon)$ определим условием $u(x_{\mathrm{tr}}(t;\varepsilon),t;\varepsilon) = 0$.

Обратная задача для (1) (*задача граничного управления*), рассмотренная в данной работе, заключается в нахождении одного из граничных условий (для определенности – правого $u^{(+)}(t)$), при котором фронт будет двигаться по заданному временному закону. Второе граничное условие $u^{(-)}(t)$ при этом считается известным.

Асимптотическим решением задачи граничного управления мы называем такое решение задачи определения граничного условия, при котором скорость или положение фронта получается как асимптотическое приближение по малому параметру к заданному. Показано, что для рассматриваемого класса уравнений задача граничного управления сводится к решению алгебраических уравнений, связывающих наблюдаемое положение и скорость движущегося фронта с коэффициентами в уравнении и граничными условиями. Аналогично может быть рассмотрена задача определения граничного условия по наблюдению траектории движения фронта при условии, что погрешность в измерении как положения, так и скорости движения фронта мала.

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ АСИМПТОТИЧЕСКОГО ИССЛЕДОВАНИЯ ПРЯМОЙ ЗАДАЧИ

Приведем основной результат работ [10], [19] для задачи (2), который будет применен при формулировке условий теоремы существования решения прямой задачи (1), а также использован для решения задачи граничного управления для (1).

3.1. Условия и теорема существования решения задачи (2)

Если положить $\varepsilon = 0$ в уравнении (2), получим вырожденное уравнение

$$\frac{\partial |u|}{\partial x} - \frac{\partial u}{\partial t} = B(u,x,t). \tag{3}$$

Уравнение (3) – обыкновенное дифференциальное уравнение первого порядка. Оно рассматривается с одним из дополнительных условий из задачи (2):

$$u(-1,t) = u^{(-)}(t), \quad t \in \mathbb{R}, \tag{4}$$

$$u(1,t) = u^{(+)}(t), \quad t \in \mathbb{R}. \tag{5}$$

Относительно этих задач предполагаются следующие условия.

Условие 1. Функции $B(u,x,t)$, $u^{(-)}(t)$ и $u^{(+)}(t)$ – достаточно гладкие и T-периодические по переменной при t, причем

$$u^{(-)}(t) < 0, \quad u^{(+)}(t) > 0 \quad \text{при} \quad t \in \mathbb{R}. \tag{6}$$

Условие 2. Задачи (3), (4) и (3), (5) имеют решения $u = \varphi^{-}(x,t)$ и $u = \varphi^{(+)}(x,t)$, определенные при $x \in [-1,1]$, $t \in \mathbb{R}$, T-периодические по переменной t и удовлетворяющие неравенствам

$$\varphi^{(-)}(x,t) < 0 < \varphi^{(+)}(x,t) \quad \text{для всех} \quad x \in [-1,1], \quad t \in \mathbb{R}.$$

Отметим, что $\varphi^{(+)}(x,t) - \varphi^{(-)}(x,t) > 0$ при всех $x \in [-1,1]$, $t \in \mathbb{R}$, и введем функцию

$$G(x,t) := \frac{\varphi^{(+)}(x,t) + \varphi^{(-)}(x,t)}{\varphi^{(+)}(x,t) - \varphi^{(-)}(x,t)}. \tag{7}$$

Условие 3. Пусть задача

$$\frac{dx}{dt} = -G(x,t), \quad x(t) = x(t+T) \tag{8}$$

имеет T-периодическое решение $x = x_0(t)$, причем

$$-1 < x_0(t) < 1 \quad \text{при} \quad t \in \mathbb{R}.$$

Замечание 1. Из теоремы сравнения для периодической задачи (8) (см., например, [18]) следует, что простым достаточным условием выполнения условия 3 является следующее:

$$G(-1,t) < 0 < G(1,t).$$

Замечание 2. В силу условий 2 и 3 для всех $t \in \mathbb{R}$ имеет место важная для дальнейшего оценка

$$\left|\frac{dx_0}{dt}\right| < 1. \tag{9}$$

Условие 4. Функция $x_0(t)$ такова, что

$$\int_0^T G_x(x,t)dt\Big|_{x=x_0(t)} > 0 \quad \text{при} \quad t \in \mathbb{R}.$$

Условие 4 обеспечивает локальную единственность и устойчивость решения $x_0(t)$ задачи (8).

Основным результатом работ [10], [19] является следующая теорема.

Теорема 1. *Пусть выполнены условия* 1–4. *Тогда для достаточно малых* ε *существует асимптотически устойчивое по Ляпунову решение* $u(x,t;\varepsilon)$ *задачи* (2) *такое, что для любого сколь угодно малого, но фиксированного* ν *имеют место предельные соотношения*

$$\lim_{\varepsilon\to 0} u(x,t;\varepsilon) = \begin{cases} \varphi^{(-)}(x,t), & x \in [-1, x_0(t) - \nu], \quad t \in \mathbb{R}, \\ \varphi^{(+)}(x,t), & x \in [x_0(t) + \nu, 1], \quad t \in \mathbb{R}. \end{cases}$$

Более того, $x_{\mathrm{tr}}(t;\varepsilon) - x_0(t) = O(\varepsilon)$, $t \in \mathbb{R}$, $u(x,t,\varepsilon) - \varphi^{(-)}(x,t) = O(\varepsilon)$ *для* $x \in [-1, x_0(t) - \nu]$, $t \in \mathbb{R}$, *и* $u(x,t,\varepsilon) - \varphi^{(+)}(x,t) = O(\varepsilon)$ *для* $x \in [x_0(t) + \nu, 1]$, $t \in \mathbb{R}$.

Замечание 3. В [10], [19] получено более подробное описание структуры переходного слоя и более точное асимптотическое приближение решения.

3.2. Теорема существования решения задачи (1)

Вырожденное уравнение в задаче (1)

$$\frac{\partial |u|}{\partial x} - \frac{\partial u}{\partial t} = K \cdot u.$$

Функции $\varphi^{(-)}(x,t)$ и $\varphi^{(+)}(x,t)$ определяются как решения задач Коши

$$\begin{aligned} &\varphi^{(-)}(x,t): \quad \frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = -K \cdot u, \quad u(-1,t) = u^{(-)}(t), \\ &\varphi^{(+)}(x,t): \quad \frac{\partial u}{\partial x} - \frac{\partial u}{\partial t} = K \cdot u, \quad u(1,t) = u^{(+)}(t). \end{aligned} \tag{10}$$

Они находятся в явном виде и являются T-периодическими по t при каждом фиксированном $-1 < x < 1$:

$$\varphi^{(-)}(x,t) = u^{(-)}(t-(1+x))e^{-K(1+x)}, \quad \varphi^{(+)}(x,t) = u^{(+)}(t-(1-x))e^{-K(1-x)}. \tag{11}$$

Заметим, что так как $u^{(-)}(t) < 0$ и $u^{(+)}(t) > 0$ при $t \in \mathbb{R}$ (условие 1), то условие 2 также выполнено и

$$\varphi^{(-)}(x,t) < 0, \quad \varphi^{(+)}(x,t) > 0 \quad \text{при} \quad (x,t) \in \bar{D} := \{x \in [-1,1],\ t \in \mathbb{R}\}.$$

Для выполнения условий теоремы 1 в рассматриваемом случае функция

$$G(x,t) = \frac{u^{(+)}(t-1+x)e^{Kx} + u^{(-)}(t-1-x)e^{-Kx}}{u^{(+)}(t-1+x)e^{Kx} - u^{(-)}(t-1-x)e^{-Kx}},$$

введенная в уравнении (8), должна удовлетворять условиям 3 и 4, которые необходимо проверять в конкретных задачах. В частности, ниже в п. 3.3 показано, что при достаточно больших значениях коэффициента усиления $K > 0$ выполнение условий 3 и 4 гарантировано.

Имеет место теорема существования решения прямой задачи (1), являющаяся следствием теоремы 1 при перечисленных выше условиях.

Теорема 2. *Пусть выполнены условия* 3 *и* 4. *Тогда для достаточно малых* ε *существует асимптотически устойчивое по Ляпунову решение* $u(x,t;\varepsilon)$ *задачи* (1) *такое, что для любого сколь угодно малого, но фиксированного* ν *выполнены предельные соотношения*

$$\lim_{\varepsilon\to 0} u(x,t;\varepsilon) = \begin{cases} \varphi^{(-)}(x,t), & x \in [-1, x_0(t)-\nu], \quad t \in \mathbb{R}, \\ \varphi^{(+)}(x,t), & x \in [x_0(t)+\nu, 1], \quad t \in \mathbb{R}. \end{cases}$$

Более того, при всех $t \in \mathbb{R}$ *имеют место оценки* $x_{\mathrm{tr}}(t;\varepsilon) - x_0(t) = O(\varepsilon)$, *а также* $u(x,t;\varepsilon) - \varphi^{(-)}(x,t) = O(\varepsilon)$ *для* $x \in [-1, x_0(t)-\nu]$ *и* $u(x,t;\varepsilon) - \varphi^{(+)}(x,t) = O(\varepsilon)$ *для* $x \in [x_0(t)+\nu, 1]$.

Сделаем важное для формулировки основного результата обратной задачи замечание.

Замечание 4. Из доказательства теорем 1 и 2 также следует непрерывная зависимость решения задачи (1) от малых возмущений граничных условий, т.е. если в задаче (1) заменить, например, $u^{(+)}(t)$ на функцию $\tilde{u}^{(+)}(t;\varepsilon)$, зависящую гладким образом от малого параметра ε, причем $\tilde{u}^{(+)}(t;\varepsilon) - u^{(+)}(t) = O(\varepsilon)$, то имеет место следующий результат (аналог теоремы 2): $x_{\mathrm{tr}}(t;\varepsilon) - x_0(t) = O(\varepsilon)$, а также $u(x,t;\varepsilon) - \varphi^{(-)}(x,t) = O(\varepsilon)$ для $x \in [0, x_0(t)-\nu]$ и $u(x,t;\varepsilon) - \varphi^{(+)}(x,t) = O(\varepsilon)$ для $x \in [x_0(t)+\nu, 1]$ при всех $t \in \mathbb{R}$.

3.3. Существование решения задачи (1) при больших K

В случае больших коэффициентов усиления $K > 0$ условие 3 выполнено, так как

$$G(1,t) = \frac{u^{(+)}(t) + u^{(-)}(t-2)e^{-2K}}{u^{(+)}(t) - u^{(-)}(t-2)e^{-2K}}$$

и

$$G(-1,t) = \frac{u^{(+)}(t-2)e^{-2K} + u^{(-)}(t)}{u^{(+)}(t-2)e^{-2K} - u^{(-)}(t)}$$

и, следовательно, при достаточно больших $K > 0$

$$G(1,t) > 0, \quad G(-1,t) < 0.$$

Поэтому в случае достаточно больших значений коэффициента усиления $K > 0$ условия существования решения задачи (8) (см. замечание 1), удовлетворяющего неравенствам $-1 < x_0(t) < 1$, выполняются. В других случаях условие 3 требует проверки, и задачу (8) необходимо решать для конкретных входных данных.

Для проверки условия 4 найдем производную

$$G_x(x,t) = -2\left[\frac{2K\cdot u^{(+)}(t-1+x)u^{(-)}(t-1-x)}{(u^{(+)}(t-1+x)e^{Kx} - u^{(-)}(t-1-x)e^{-Kx})^2} + \right.$$
$$\left. + \frac{u'^{(+)}(t-1+x)u^{(-)}(t-1-x) + u^{(+)}(t-1+x)u'^{(-)}(t-1-x)}{(u^{(+)}(t-1+x)e^{Kx} - u^{(-)}(t-1-x)e^{-Kx})^2}\right].$$

Видим, что первое слагаемое в числителе дроби в квадратных скобках отрицательно, так как $K > 0$, а функции $u^{(+)}$ и $u^{(-)}$ имеют разные знаки; второе слагаемое ограничено ввиду гладкости функций $u^{(+)}$ и $u^{(-)}$. Следовательно, при достаточно больших $K > 0$ функция $G(x,t)$ монотонна по переменной x, причем $G_x(x,t) > 0$ при всех $(x,t) \in ((-1;1)\times\mathbb{R})$, и условие 4 выполняется. Таким образом, достаточно большое значение коэффициента K обеспечивает выполнение условий 3 и 4 и гарантирует однозначную разрешимость прямой задачи (1).

С учетом формул (11) и (7) задача (8)

$$-\frac{dx}{dt}(\varphi^{(+)}(x,t) - \varphi^{(-)}(x,t)) = \varphi^{(+)}(x,t) + \varphi^{(-)}(x,t), \quad x(t) = x(t+T),$$

определяющая главный член асимптотики положения фронта $x_0(t)$, преобразуется к виду

$$-\frac{dx}{dt}(u^{(+)}(t-1+x)e^{Kx} - u^{(-)}(t-1-x)e^{-Kx}) = \\ = u^{(+)}(t-1+x)e^{Kx} + u^{(-)}(t-1-x)e^{-Kx}, \quad x(t) = x(t+T). \tag{12}$$

Следовательно, функция $x_0(t)$ удовлетворяет соотношению

$$-\frac{dx_0}{dt}(u^{(+)}(t-1+x_0(t))e^{Kx_0(t)} - u^{(-)}(t-1-x_0(t))e^{-Kx_0(t)}) = \\ = u^{(+)}(t-1+x_0(t))e^{Kx_0(t)} + u^{(-)}(t-1-x_0(t))e^{-Kx_0(t)}. \tag{13}$$

4. АСИМПТОТИЧЕСКОЕ РЕШЕНИЕ ЗАДАЧИ ГРАНИЧНОГО УПРАВЛЕНИЯ

Пусть задан желаемый закон движения фронта, т.е. задана функция $f(t) = x_{\mathrm{tr}}(t;\varepsilon)$ на интервале времени, равном периоду T. Задача граничного управления заключается в нахождении одного из граничных условий (для определенности – правого $u^{(+)}(t)$), при котором фронт будет двигаться по заданному временному закону. Второе граничное условие $u^{(-)}(t)$ при этом считается известным, и мы предполагаем, что выполнены условия существования решения задачи (1).

Постановку обратной задачи можно записать в операторном виде

$$A(u^{(+)}) = f. \tag{14}$$

В данной работе точный оператор A задачи (14) мы заменяем на приближенный оператор A_0, определяемый соотношением (13). Из теоремы 2 следует, что $\|A - A_0\|_C = O(\varepsilon)$. В результате решается задача

$$A_0(u^{(+)}) = f, \tag{15}$$

решение которой $u^{(+)}(t) = h_0(t)$ и есть асимптотическое решение задачи граничного управления, так как подстановка его в (14) обеспечивает желаемый закон движения фронта $f(t) = x_{\mathrm{tr}}(t;\varepsilon)$ с заданной точностью $O(\varepsilon)$.

Из (13) получим

$$u^{(+)}(t-1+x_0(t)) = u^{(-)}(t-1-x_0(t))e^{-2Kx_0(t)}\frac{x_0'(t)-1}{x_0'(t)+1}.$$

Обозначим $t - 1 + x_0(t) = \tau$. Тогда из уравнения $\tau = t - 1 + x_0(t) \equiv g(t)$ найдем $t = g^{(-1)}(\tau)$. Разрешимость этого уравнения относительно t обеспечивается тем, что $|x_0'(t)| < 1$ (см. (9)), следовательно, $g'(t) = 1 + x_0'(t) \neq 0$ при всех $t \in \mathbb{R}$.

Тогда

$$u^{(+)}(\tau) = \left[u^{(-)}(\tau - 2x_0(t)) e^{-2Kx_0(t)} \frac{x_0'(t) - 1}{x_0'(t) + 1} \right]_{t = g^{(-1)}(\tau)}. \tag{16}$$

Видим, что $u^{(+)}(\tau) > 0$ при всех $\tau \in \mathbb{R}$, так как $u^{(-)}(t) < 0$ и $|x_0'(t)| < 1$ при всех $t \in \mathbb{R}$.

Покажем, что функция $u^{(+)}(\tau)$ является T-периодической по τ, если $x_0(t)$ и $u^{(-)}(t)$ – T-периодические по t. Действительно, в силу T-периодичности $x_0(t)$, определений τ и функции $g(t)$ справедливо равенство $\tau + T = g(t + T)$. Таким образом, $g^{-1}(\tau + T) = t + T = g^{-1}(\tau) + T$. Это с учетом T-периодичности функции $u^{(-)}(t)$ обеспечивает T-периодичность $u^{(+)}(t)$.

Заметим, что мы строим *асимптотическое решение* задачи граничного управления, т.е. искомое граничное условие (в нашем случае – на правой границе) должно быть определено так, чтобы требуемый режим движения фронта был бы реализован с заданной точностью.

Одним из результатов асимптотического анализа, проведенного в [10], являются равномерные по $t \in \mathbb{R}$ оценки

$$|x_{\mathrm{tr}}(t;\varepsilon) - x_0(t)| = O(\varepsilon), \quad \left| \frac{dx_{\mathrm{tr}}(t;\varepsilon)}{dt} - \frac{dx_0(t)}{dt} \right| = O(\varepsilon). \tag{17}$$

Если желаемый закон движения фронта задан, т.е. определена достаточно гладкая функция $f(t) = x_{\mathrm{tr}}(t;\varepsilon)$, удовлетворяющая условию $|f'(t)| < 1$, то заменив в (16) главный член асимптотики положения фронта $x_0(t)$ на заданную функцию $f(t)$ и учитывая, что $f(t) = x_{\mathrm{tr}}(t;\varepsilon) = x_0(t) + O(\varepsilon)$, получим

$$u^{(+)}(\tau) = \left[u^{(-)}(\tau - 2f(t)) e^{-2Kf(t)} \frac{f'(t) - 1}{f'(t) + 1} \right]_{t = g^{(-1)}(\tau)} = h_0(\tau) + O(\varepsilon). \tag{18}$$

Здесь введено обозначение

$$\left[u^{(-)}(\tau - 2x_0(t)) e^{-2Kx_0(t)} \frac{x_0'(t) - 1}{x_0'(t) + 1} \right]_{t = g^{(-1)}(\tau)} \equiv h_0(\tau).$$

Формула (18) дает асимптотическое решение задачи граничного управления с точностью $O(\varepsilon)$.

Таким образом, обратная задача граничного управления для уравнения Бюргерса с модульной адвекцией и линейным усилением сводится к линейному алгебраическому уравнению (13), связывающему наблюдаемые параметры движущегося фронта с коэффициентами в уравнении и граничными условиями. Решение (16) этого уравнения относительно функции $u^{(+)}(t)$ в силу замечания 4 является асимптотическим решением задачи граничного управления для (1) с точностью $O(\varepsilon)$, т.е. обеспечивает реализацию заданного режима движения фронта (скорость и положение) с точностью $O(\varepsilon)$. Отметим также, что заданный режим движения, для которого не выполняется условие $|f'(t)| < 1$, не является реализуемым, так как не выполнены условия прямой задачи. Это легко иллюстрируется на примере: если $f'(t) > 1$, то $u^{(+)}(\tau) < 0$, что противоречит условию постановки прямой задачи.

Проведенные выше исследования позволяют сформулировать следующий результат.

Теорема 3. *Пусть задан требуемый закон движения внутреннего слоя $x_{\mathrm{tr}}(t;\varepsilon) = f(t)$, где $f(t)$ – достаточно гладкая T-периодическая функция, удовлетворяющая условию $|f'(t)| < 1$. Тогда при достаточно большом коэффициенте усиления $K > 0$ и достаточно малых ε существует асимптотическое решение задачи граничного управления для уравнения* (1) *, задаваемое формулой* (18).

В данной задаче можно получить более точное асимптотическое приближение по параметру граничного управления. Проиллюстрируем это на примере построения граничной функции,

обеспечивающей заданный режим движения фронта (скорость и положение) с точностью $O(\varepsilon^2)$. Для этого будем строить граничное управление в виде

$$u^{(+)}(t) = h_0(t) + \varepsilon h_1(t),$$

и запишем исходную задачу (1)

$$\begin{gathered}\varepsilon\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} + \frac{\partial |u|}{\partial x} - K\cdot u = 0, \quad (x,t)\in D := \{x\in(-1,1);\, t\in\mathbb{R}\},\\ u(-1,t;\varepsilon) = u^{(-)}(t), \quad u(1,t;\varepsilon) = u^{(+)}(t) = h_0(t) + \varepsilon h_1(t), \quad t\in\mathbb{R},\\ u(x,0;\varepsilon) = u(x,t+T;\varepsilon), \quad x\in[-1,1], \quad t\in\mathbb{R}.\end{gathered} \tag{19}$$

Для построения асимптотического приближения граничного управления воспользуемся алгоритмом, предложенным в [10].

Главный член асимптотического приближения граничного управления – функция $h_0(t)$ – определен выше и обеспечивает реализацию требуемого режима движения фронта с точностью $O(\varepsilon)$. Функцию $h_1(t)$ будем подбирать так, чтобы граничное управление $u^{(+)}(t) = h_0(t) + \varepsilon h_1(t)$ обеспечивало реализацию требуемого режима движения фронта с точностью $O(\varepsilon^2)$.

Регулярные члены первого приближения по ε – функции $U_1^{(-)}(x,t)$ и $U_1^{(+)}(x,t)$ – определяются как T-периодические решения задач

$$\begin{aligned}&U_1^{(-)}(x,t)\colon \quad \frac{\partial U_1^{(-)}}{\partial x} + \frac{\partial U_1^{(-)}}{\partial t} = -K\cdot U_1^{(-)} + F_1^{(-)}(x,t), \quad U_1^{(-)}(-1,t) = 0,\\ &U_1^{(+)}(x,t)\colon \quad \frac{\partial U_1^{(+)}}{\partial x} - \frac{\partial U_1^{(+)}}{\partial t} = K\cdot U_1^{(+)} + F_1^{(+)}(x,t), \quad U_1^{(+)}(1,t) = h_1(t),\end{aligned} \tag{20}$$

где $F_1^{(\pm)}(x,t) = \partial^2\varphi^{(\pm)}(x,t)/\partial x^2$ – известные T-периодические функции.

Решения задач (20) находятся в явном виде и являются T-периодическими по t при каждом фиксированном $-1 < x < 1$, а именно,

$$U_1^{(-)}(x,t) = \int_{-1}^{x} e^{-K(x-\xi)} F_1^{(-)}(\xi, \xi - x + t)\,d\xi, \tag{21}$$

$$U_1^{(+)}(x,t) = h_1(t + x - 1)e^{K(x-1)} + \int_{1}^{x} e^{K(x-\xi)} F_1^{(+)}(\xi, x + t - \xi)\,d\xi. \tag{22}$$

Асимптотическое приближение положения фронта (см. [10]) имеет вид

$$x_{\mathrm{tr}}(t;\varepsilon) = x_0(t) + \varepsilon x_1(t) + O(\varepsilon^2),$$

где $x_0(t)$ определена в (13), а $x_1(t)$ является решением задачи

$$\frac{dx_1}{dt} + G_x(x_0(t),t)x_1 = \frac{H_1(t)}{\varphi^{(-)}(x_0(t),t) - \varphi^{(+)}(x_0(t),t)}, \quad x_1(t) = x_1(t+T), \tag{23}$$

$$\begin{aligned}H_1(t) = {}& \frac{\partial\varphi^{(+)}}{\partial x}(x_0(t),t) - \frac{\partial\varphi^{(-)}}{\partial x}(x_0(t),t) + (1 + x_0'(t))U_1^{(+)}(x_0(t),t) +\\ &+ (1 - x_0'(t))U_1^{(-)}(x_0(t),t) - \int_0^{+\infty} q_1^{(+)}(\xi,t)\,d\xi - \int_{-\infty}^{0} q_1^{(-)}(\xi,t)\,d\xi.\end{aligned} \tag{24}$$

Функции $q_1^{(\pm)}(\xi,t)$ известны и выражаются через найденные на предыдущем шаге асимптотической процедуры, а $U_1^{(+)}(x,t)$ определена в (22).

Тогда из (22) получим

$$h_1(t + x_0(t) - 1) = U_1^{(+)}(x_0(t), t)e^{K(1-x_0(t))} - \int_1^{x_0(t)} e^{K(1-\xi)} F_1^{(+)}(\xi, x_0(t) + t - \xi)d\xi, \tag{25}$$

где $U_1^{(+)}(x_0(t), t)$ находится в явном виде из (24) с учетом (23), причем в (23)

$$x_1(t) = \frac{x_{\text{tr}}(t;\varepsilon) - x_0(t)}{\varepsilon}, \quad \frac{dx_1}{dt} = \frac{x'_{\text{tr}}(t;\varepsilon) - x'_0(t)}{\varepsilon}.$$

Введя, как и выше, переменную $\tau = t + x_0(t) - 1 \equiv g(t)$, получим

$$h_1(\tau) = U_1^{(+)}(\tau + 1 - t)\, e^{K(t-\tau)}\Big|_{t=g^{(-1)}(\tau)} - \int_1^{\tau+1-t} e^{K(1-\xi)} F_1^{(+)}(\xi, \tau + 1 - \xi)d\xi\Big|_{t=g^{(-1)}(\tau)}. \tag{26}$$

Если требуемый закон движения фронта $x_{\text{tr}}(t;\varepsilon) = f(t)$ задан, то, заменяя в (25) главный член асимптотики положения фронта $x_0(t)$ на функцию $f(t)$ и учитывая, что $f(t) = x_{\text{tr}}(t;\varepsilon) = x_0(t) + O(\varepsilon)$, получим приближение граничного управления

$$u^{(+)}(\tau) = h_0(\tau) + \varepsilon h_1(\tau) + O(\varepsilon^2).$$

Аналогичным образом могут быть построены асимптотики граничного управления более высокого порядка.

Очевидно, что изложенные результаты могут быть применены к близкой обратной задаче определения граничного условия на основе наблюдения положения внутреннего переходного слоя в случае, если имеется априорная информация о приближенном значении положения $x_{\text{tr}}(t;\varepsilon)$ и скорости движения внутреннего слоя $x'_{\text{tr}}(t;\varepsilon)$.

5. ЗАКЛЮЧЕНИЕ

В работе получил дальнейшее развитие асимптотико-численный подход к решению прямых и обратных задач для уравнений, решения которых содержат пограничные и внутренние слои. Концепция асимптотического решения обратных коэффициентных задач, предложенная авторами, применена к новому классу периодических по времени задач типа реакция–диффузия–адвекция с внутренними переходными слоями. Развиваемый подход продемонстрирован на примере задачи граничного управления для уравнения Бюргерса с модульной адвекцией и линейным усилением. Показано, что для этого класса уравнений асимптотическое решение задачи граничного управления сводится к существенно более простой задаче, связывающей скорость движущегося фронта с коэффициентами в исходном уравнении и граничными условиями. Предлагаемый подход может быть применен к достаточно широкому классу задач с пограничными и внутренними слоями.

Авторы выражают благодарность рецензенту за внимательное прочтение статьи и ряд ценных замечаний.

СПИСОК ЛИТЕРАТУРЫ

1. *Nefedov N.* Comparison principle for reaction-diffusion-advection problems with boundary and internal layers // Lect. Not. Comput. Sci. 2013. 8236. P. 62–72.
2. *Burgers J.M.* A mathematical model illustrating the theory of turbulence // Adv. Appl. Mech. 1948. V. 1. P. 171–199.
3. *Cole J.D.* On a quasilinear parabolic equation occurring in aerodynamics // Quart. Appl. Math. 1951. V. 9. P. 225–236.
4. *Rudenko O.V., Gurbatov S.N., Hedberg C.M.* Nonlinear Acoustics Through Problems and Examples. Victoria, BC, Canada: Trafford, 2011.
5. *Руденко О.В.* Линеаризуемое уравнение для волн в диссипативных средах с модульной, квадратичной и квадратично-кубичной нелинейностями // Докл. АН. 2016. Т. 471. № 1. С. 23–27.
6. *Руденко О.В.* Модульные солитоны // Докл. АН. 2016. Т. 471. 6. С. 451–454.
7. *Нефедов Н.Н., Руденко О.В.* О движении фронта в уравнении типа Бюргерса с квадратичной и модульной нелинейностью при нелинейном усилении // Докл. АН. 2018. Т. 478. № 3. С. 274–279.

8. *Нефедов Н.Н., Руденко О.В.* О движении, усилении и разрушении фронтов в уравнениях типа Бюргерса с квадратичной модульной нелинейностью // Докл. АН. 2020. Т. 493. № 1. С. 26–31.
9. *Бутузов В.Ф., Васильева А.Б., Нефедов Н.Н.* Асимптотическая теория контрастных структур (обзор) // Автомат. и телемехан. 1997. № 7. С. 4–32; Autom. Remote Control, 58:7 (1997), P. 1068–1091.
10. *Нефедов Н.Н.* Развитие методов асимптотического анализа переходных слоев в уравнениях реакция–диффузия–адвекция: теория и применение // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 12. С. 2074–2094.
11. *Nefedov N., Recke L., Schneider K.* Existence and asymptotic stability of periodic solutions with an interior layer of reaction-advection-diffusion equations // J. Math. Analys. Appl. 2013. V. 405. № 1. P. 90–103.
12. *Nefedov N.* Existence and asymptotic stability of periodic solutions with an interior layer of Burgers type equation with modular advection // Math. Model. Nat. Phenom. 2019. V. 14. № 4. P. 401.
13. *Lukyanenko D.V., Grigorev V.B., Volkov V.T., Shishlenin M.A.* Solving of the coefficient inverse problem for a nonlinear singularly perturbed two-dimensional reaction-diffusion equation with the location of moving front data // Comput. Math. Appl. 2019. V. 77. № 5. P. 1245–1254.
14. *Лукьяненко Д.В., Волков В.Т., Нефедов Н.Н., Ягола А.Г.* Применение асимптотического анализа для решения обратной задачи определения коэффициента линейного усиления в уравнении типа Бюргерса // Вестник МГУ. Сер. 3: Физика. Астрономия. 2019. № 2. С. 38–43.
15. *Lukyanenko D.V., Shishlenin M.A., Volkov V.T.* Asymptotic analysis of solving an inverse boundary value problem for a nonlinear singularly perturbed time-periodic reaction-diffusion-advection equation // J. Inverse Ill-Posed Problem. 2019. V. 27. № 5. P. 745–758.
16. *Волков В.Т., Нефедов Н.Н.* Асимптотическое решение коэффициентных обратных задач для уравнений типа Бюргерса // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 6. С. 975–084.
17. *Nefedov N.N., Volkov V.T.* Asymptotic solution of the inverse problem for restoring the modular type source in Burgers' equation with modular advection // J. Inverse and Ill-Posed Problem. 2020. V. 28. № 5. P. 633–639.
18. *Hess P.* Periodic-Parabolic Boundary Value Problems and Positivity. New York: Pitman Res. Not. Math. Ser., 1991. 139 p.
19. *Nefedov N.* The periodic solutions with an interior layer of Burgers type equations with modular advection: Asymptotic approximation and asymptotic solutions of some inverse coefficient problems // Современные проблемы математики и механики. Материалы междунар. конф., посвященной 80-летию академика В.А. Садовничего. V. 2. М.: Макс Пресс, 2019. P. 427–429.
20. *Kabanikhin S.I.* Definitions and examples of inverse and ill-posed problems // J. Inverse and Ill-Posed Problem. 2008. V. 16. № 4. P. 317–357.
21. *Beilina L., Klibanov M.V.* A globally convergent numerical method for a coefficient inverse problem // SIAM J. Sci. Comput. 2008. V. 31. № 1. P. 478–509.
22. *Kabanikhin S.I., Sabelfeld K.K., Novikov N.S., Shishlenin M.A.* Numerical solution of an inverse problem of coefficient recovering for a wave equation by a stochastic projection methods // Monte Carlo Meth. Appl. 2015. V. 21. № 3. P. 189–203.

EDN: NGFLDR
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1861–1866

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 517.95

ЗАДАЧА ЛИНЕЙНОГО СОПРЯЖЕНИЯ ДЛЯ ОБОБЩЕННОГО УРАВНЕНИЯ КОШИ–РИМАНА С СВЕРХСИНГУЛЯРНЫМИ ТОЧКАМИ НА ПОЛУПЛОСКОСТИ

© 2022 г. И. Н. Дорофеева[1,*], А. Б. Расулов[1,]**

[1] *111250 Москва, ул. Красноказарменная, 14, ФГБУ ВО МЭИ, Россия*

**e-mail: idoro224@gmail.com*

***e-mail: rasulzoda55@gmail.com*

Поступила в редакцию 20.05.2021 г.
Переработанный вариант 20.05.2021 г.
Принята к публикации 07.07.2022 г.

Для уравнения с оператором Коши–Римана с сильными изолированными точечными особенностями в младшем коэффициенте найдено интегральное представление решения в классе ограниченных функций на бесконечности и исследована задача линейного сопряжения на полуплоскости. Библ. 15.

Ключевые слова: оператор Коши–Римана, сильные особенности в точках, оператор Помпейю–Векуа, полуплоскость, задача типа линейного сопряжения.

DOI: 10.31857/S0044466922110084

Рассмотрим на комплексной плоскости $\mathbb{C}$ вне множества заданных конечных точек $l = \{z_1,\dots,z_m,\ \mathrm{Re}\, z_j \neq 0,\ j = \overline{1,m}\}$ уравнение

$$\partial_{\bar z} u - Au = f, \quad A(z) = \sum_{j=1}^{m} \frac{(z - z_j)a_j(z)}{|z - z_j|^{n_j+1}}, \quad n_j > 1. \tag{1}$$

Предполагается, что функции $a_1,\dots,a_m$ и f непрерывны и удовлетворяют условиям

$$a_j(z) = a_j(z_j) + O(|z|^{n_j-2/p}) \quad \text{при} \quad z \to z_j, \quad j = \overline{1,m}; \quad f(z) = o(|z|^{-1-\alpha}) \quad \text{при} \quad z \to \infty, \tag{2}$$

где $a_j(z_j) \neq 0,\ n_j > 1,\ j = \overline{1,m},\ p > 2$ и $\alpha = \min(n_1,\dots,n_m)$.

Обобщенная система Коши–Римана

$$\frac{\partial u}{\partial \bar z} + au + b\bar u = f \tag{3}$$

ранее рассматривалась, в основном, в конечной области $D \subseteq \mathbb{C}$, с комплекснозначными функциями $a(z)$, $b(z)$, $f(z)$ – заданными в ограниченной области D, $u(z)$ – неизвестная функция.

Хорошо известно, сколь важную роль в приложениях играет теория обобщенных аналитических функций уравнения (3), созданная И.Н. Векуа [1], в случае, когда $a, b, f \in L_p(D)$, $p > 2$. Она имеет глубокие связи со многими разделами анализа, геометрии и механики, включая квазиконформные отображения, теорию поверхностей, теорию оболочек, газовую динамику. В частности, она широко используется при моделировании трансзвуковых течений газа, состояний безмоментного напряженного равновесия выпуклых оболочек и многих других процессов.

В этой теории ключевую роль играет интеграл Помпейю

$$(Tf)(z) = -\frac{1}{\pi}\int_G \frac{f(\zeta)d_2\zeta}{\zeta - z}, \tag{4}$$

с плотностью $f \in L^p(D)$, $p > 2$, где здесь и ниже $d_2\zeta$ означает элемент площади. Хорошо известно, что этот оператор ограничен из $L^p(D)$ в соболевское пространство $W^{1,p}(D)$ и имеет место вложение $W^{1,p}(D) \subseteq C^{\mu}(\overline{D})$ с показателем Гёльдера $\mu = (p-2)/p$ и $\partial_{\overline{z}} Tf = f$. Следовательно, если $A, f \in L^p(D)$, $p > 2$, и $\Omega = TA$, то общее решение уравнения (1) дается формулой [1]:

$$u = e^{\Omega}[\phi + T(e^{-\Omega} f)],$$

где ϕ – произвольная аналитическая функция.

Уравнение (3) с коэффициентами $a = O(\overline{z}^{-1})$, $b = O(\overline{z}^{-1})$, при $z \to 0 \in D$ в связи с его многочисленными приложениями рассматривалось многими авторами. В монографии Л.Г. Михайлова [2] разрешимость интегрального уравнения, к которому сводится уравнение (3), доказывается при определенных условиях малости этих коэффициентов и на основе полученного решения исследована граничная задача Гильберта.

З.Д. Усмановым [3] построена теория уравнения (3) при $a = 0$, $b(z) = \overline{z}^{-1}\beta e^{ik\varphi}$, $k \in Z$. Однако случай, когда $b(z) = \overline{z}^{-1}(\beta_1 e^{ik\varphi} + \beta_2 e^{im\varphi})$, где $\beta_1 \neq \beta_2$, приводит к бесконечным системам обыкновенных дифференциальных уравнений, которые не поддаются исследованию. Поэтому в дальнейшем основной целью исследований З.Д. Усманова [4] является нахождение связи решений уравнений (3) с коэффициентами $a(z) = 0$, $b(z) = \overline{z}^{-1}\beta$ и с коэффициентами $a(z) = 0$, $b(z) = O(\overline{z}^{-1})$ при $z \to 0 \in D$ посредством линейного интегрального уравнения с вполне непрерывным оператором, чтобы какой-либо вопрос для общего уравнения редуцировать к аналогичному вопросу для уравнения частного вида с постоянными коэффициентами. Доказано, что существуют решения уравнения, допускающие особенности порядка $\nu > 0$ в точке $z = 0$ в виде рядов Фурье, коэффициенты которого определяются через функции Бесселя (функции Макдональда).

H. Begehr и Dao-Qing Dai [5] изучили уравнение

$$\frac{\partial u}{\partial \overline{z}} = \frac{Q(z)}{P(z)} + au + b\overline{u}, \tag{5}$$

с коэффициентами $a = O(\overline{z}^{-1})$, при $z \to 0 \in D$, $b(z) \in L^p(D)$, $p > 2$, где $Q(z)$ и $P(z)$ – многочлены комплексной переменной z; и полином P имеет только простые корни в замкнутом единичном диске D. Было установлено, что число решений задачи Римана–Гильберта для уравнения (5) зависит не только от ее индекса, но и от местоположения и типа особенностей.

Класс непрерывных решений уравнения (3) при $a(z) = O(\overline{z}^{-1})$, $b = O(\overline{z}^{-1})$ изучен в работах А.Б. Тунгатарова [6]. Решение уравнения (3) с сингулярными коэффициентами в виде рядов также изучается в работе А. Мезиани [7].

По мнению многих исследователей класс уравнений (3) (при $a = O(\overline{z}^{-1})$, $b = O(\overline{z}^{-1})$ при $z \to 0 \in D$), исследованный Л.Г. Михайловым, является простейшим и, пожалуй, единственным представителем класса обобщенных систем Коши–Римана с квазисуммируемыми коэффициентами, относительно которого ряд результатов удается получить путем использования основного оператора (4) теории регулярных обобщенных систем Коши–Римана [8].

Поэтому пришли к другим методам исследования уравнения (3), минуя оператор Помпейю-Векуа (4). Например, А. Тунгатаровым и др. [6] уравнения (3), имеющие особенности первого порядка в точке или на линии, рассмотрены в бесконечной угловой области произвольного раствора. Кратко изложим схему построения решения. Используя сочетания метода Фукса и метода Фурье разделения переменных, уравнения (1) приводятся к сингулярным интегральным уравнениям. Далее с помощью модифицированного метода последовательных приближений из этих сингулярных интегральных уравнений получены представления решений, где в правой части имеется n кратный интеграл, содержащий неизвестную функцию. При $n \to \infty$ этот член стремится к нулю, и поэтому интегральные представления превращаются в многообразия непрерывных решений.

Нам удалось применить оператор Помпейю (4) к исследованию уравнения (3) в случае , когда коэффициенты a,b имеют сильные особенности в точках и линиях области D (см., например, [11], [12]).

В работах А.П. Солдатова [9], [10] даны оценки классического интеграла Помпейю (4), рассматриваемого на всей комплексной плоскости с особыми точками $z = 0$ и $z = \infty$ в семействах различных весовых пространств, некоторые из которых мы используем в данной работе.

Для более подробного ознакомления с обзором проделанных работ можно обратиться к работам [2], [3], [6] и другим источникам.

Следовательно, если в ранее изложенных работах исследование обобщенной системы уравнений Коши–Римана велось, в основном, в конечной области, в нашем случае рассматривается расширенная комплексная плоскость, дополненная по сравнению с обычной бесконечно удаленной точкой: $\mathbb{C} \cup \{\infty\} = \overline{\mathbb{C}}$. Геометрически точка $z = \infty$ изображается точкой сферы Римана (ее “северный полюс”). Следовательно, к m-конечным точкам $z_j, j = \overline{1,m}$, плоскости, в которых коэффициент A имеет сильные особенности, добавится еще одна особая точка $z = \infty$.

Пусть $\mathbb{C} = \mathbb{C}^+ \cup \mathbb{R} \cup \mathbb{C}^-$, где $\mathbb{C}^+$ и $\mathbb{C}^-$ – соответственно, верхная и нижная плоскость, $\mathbb{R}$ – вещественная ось.

В рассматриваемом случае интегральный оператор T понимается по отношению к неограниченной области, в том числе и по отношению к областям $\mathbb{C}^+$, $\mathbb{C}^-$ или $\mathbb{C}$. Хорошо известно [1], что если функция f непрерывно дифференцируема и $f(z) = O(|z|^{\delta})$ при $z \to \infty$ с некоторым $\delta < -1$, то функция

$$(Tf)(z) = -\frac{1}{\pi}\int\limits_{\mathbb{C}} \frac{f(\zeta)d_2\zeta}{\zeta - z}, \quad \zeta, z \in \mathbb{C}, \tag{6}$$

непрерывно дифференцируема и является решением уравнения (1) при $A = 0$. В монографии [1] И.Н. Векуа описал условие на функцию f, обеспечивающее принадлежность Tf классу $C^{\mu}(\mathbb{C})$ в терминах введенного им пространства $L^{p,\nu}(\mathbb{C})$, $p > 2$. Под $C^{\mu}(\mathbb{C})$ здесь понимается класс непрерывных функций $f(z)$, которые вместе с $f(1/z)$ принадлежат $C^{\mu}(D)$ в единичном круге D. По определению пространство $L^{p,\nu}(\mathbb{C})$ состоит из всех функций f, для которых $f(z)$ и $f_{\nu}(z) = |z|^{-\nu} f(1/z)$ принадлежат $L^p(D)$. В этих обозначениях если $f \in L^{p,2}(\mathbb{C})$, $p > 2$, то функция $Tf \in C^{\mu}(\mathbb{C})$, $\mu = 1 - 2/p$, и обращается в нуль на бесконечности (см. теоремы 1.24, 1.25 в монографии [1]). В частности, $(Tf)(z) = o(|z|^{\mu-1})$ при $z \to \infty$.

Обычно под обобщенным решением уравнения (1) понимается функция u, которая в области $\mathbb{C}^+\backslash\{l\}$ допускает обобщенную производную по $\overline{z}$, причем

$$u_{\overline{z}} \in L^{p,2}(\mathbb{C}^+_{\varepsilon}), \quad \mathbb{C}^+_{\varepsilon} = \{z, |z - z_j| > \varepsilon,\ j = 1,\ldots,m\},$$

для любого $\varepsilon > 0$, ограниченная на бесконечности и удовлетворяющая уравнению (1) почти всюду.

В дальнейшем для компактного изложения при $n_j > 1$, $j = \overline{1,m}$, введем следующие обозначения:

$$\omega_j = \frac{2}{(n_j - 1)|z - z_j|^{n_j - 1}}, \quad A_0(z) = \sum_{j=1}^{m} \frac{(z - z_j)[a_j(z) - a_j(z_j)]}{|z - z_j|^{n_j+1}},$$

где ω_j – решение уравнения $u_{\overline{z}}(z) = -(z - z_j)|z - z_j|^{-1-n}$, $a_j \in C(\mathbb{C}^+ \cup \mathbb{R})$, $j = \overline{1,m}$.

Введем сингулярный интеграл

$$\Omega(z) = \lim_{\varepsilon \to 0}(T_{\varepsilon}A)(z) \equiv -\lim_{\varepsilon \to 0}\frac{1}{\pi}\int\limits_{\mathbb{C}^+_{\varepsilon}} \frac{A(\zeta)d_2\zeta}{\zeta - z},$$

где интегральный оператор T_{ε} определяется аналогично (2) по отношению к области $\mathbb{C}^+_{\varepsilon}$.

Теорема 1. *Пусть* $n_j > 1$, $j = \overline{1,m}$, *и* $A_0 \in L^{p,2}(\mathbb{C}^+)$. *Тогда функция* $\Omega(z)$, $z \neq z_j$, $j = \overline{1,m}$; *существует и представима в виде*

$$\Omega(z) = -\sum_{j=1}^{m} a_j(z_j)\omega_j + h(z), \tag{7}$$

где $h(z) \in C^{\mu}$, *определяется равенством*

$$h(z) = (TA_0)(z) + \frac{1}{\pi i}\sum_{j=1}^{m}\frac{a_j(z_j)}{(n_j-1)}\int_{\mathbb{R}}\frac{1}{|z-z_j|^{n_j-1}}\frac{d\zeta}{\zeta - z}$$

и удовлетворяет уравнению $\Omega_{\bar{z}} = A$.

Соответственно в предположении $e^{-\Omega}f \in L^{p,2}(\mathbb{C}^+)$, *обобщенное решение уравнения* (1) *дается формулой*

$$u = e^{\Omega}[\phi + T(e^{-\Omega}f)], \tag{8}$$

где $\phi \in C^{\mu}(\overline{\mathbb{C}^+}\backslash\{l\})$ *— произвольная аналитическая в области* $\mathbb{C}^+\backslash\{l\}$ *функция и* $\phi(z) = o(|z|^{-2/p})$ *при* $|z| \to \infty$.

Доказательство теоремы проводится аналогично доказательству теоремы об интегральном представлении системы Коши–Римана с одной сверхсингулярной точкой, которое приведено в [15] и базируется на тождестве $\partial_{\bar{z}}\Omega = A$ и формуле Грина

$$\int_B \frac{\partial U}{\partial \bar{\zeta}} d_2\zeta = \frac{1}{2i}\int_{\partial B} U d\zeta,$$

в круге $B = \{|z| < R\}$ достаточно большого радиуса R. Тогда при $R \to \infty$ убеждаемся в справедливости теоремы.

Замечание 1. Заметим, что при $0 < \delta < 1$ условие

$$u(z) = O(|z-z_j|^{-\delta})\exp\left[-\frac{\operatorname{Re} 2a(z_j)}{(n-1)|z-z_j|^{n-1}}\right],$$

при $z \to z_j$, $j = \overline{1,m}$, равносильно тому, что в этом представлении функция ϕ имеет $z = z_j$, $j = \overline{1,m}$, устранимую особую точку и, следовательно, аналитична во всей области $\mathbb{C}^+$ и по условию $\phi(z) = o(|z|^{-2/p})$ при $|z| \to \infty$.

Поэтому фактически функция u относится к классу функций, для которых $e^{-\Omega}u \in H(\overline{\mathbb{C}^+})$. Этот класс функций, удовлетворяющий условию Гёльдера с некоторым показателем, удобно обозначить через $H(\overline{\mathbb{C}^+}, e^{\Omega})$, где $\overline{\mathbb{C}}^+ = \mathbb{C}^+ \cup \mathbb{R}$.

Замечание 2. Заметим, что $A_{0j}(z) \in L^p(G)$, $p > 2$, если

$$a_j(z) - a_j(z_j) = O(|z-z_j|^{\gamma_j}), \quad \gamma_j > n_j - 2p^{-1}, \quad n_j > 1, \quad j = \overline{1,m}.$$

Заметим, что теорема 1 сохраняет свою силу, если условие $A_0(z) \in L^{p,2}(\mathbb{C})$ заменено на (2), или, более точно, на условие $a_j(z) - a_j(z_j) = o(|z|^{-\alpha})$ при $z \to \infty$.

2. ЗАДАЧА ТИПА ЛИНЕЙНОГО СОПРЯЖЕНИЯ

Требуется найти решение уравнения (1) *в полуплоскостях* $\mathbb{C}^+$, $\mathbb{C}^-$, соответственно принадлежащее классам $L^{p,2}(\mathbb{C}^{\pm}) \cap H(\overline{\mathbb{C}^+}, e^{\Omega})$ и такое, что для функций $(e^{-\Omega}u)^{\pm}$, ограниченных в $\mathbb{C}^+$, и $\mathbb{C}^-$, предельные значения на контуре $\mathbb{R}$ удовлетворяют следующему граничному условию:

$$(e^{-\Omega}u)^+(t) = G(t)(e^{-\Omega}u)^-(t) + g(t), \quad t \in \mathbb{R}, \tag{9}$$

где $G(t)$, $g(t)$ – заданные на $\mathbb{R}$ функции, удовлетворяющие условию Гёльдера, как в конечных точках, так и в окрестности бесконечно удаленной точки контура, причем $G(t) \neq 0$, $t \in \mathbb{R}$ и $g(t) = o(|t|^{-\delta})$, $\delta > 0$, $t \to \infty$.

Используя интегральное представление (4) и условие задачи (9), мы приходим к следующей задаче линейного сопряжения теории аналитических функций для полуплоскости:

$$\phi^+(t) = G(t)\phi^-(t) + \tilde{g}(t), \quad t \in \mathbb{R}, \tag{10}$$

где

$$\tilde{g} = T^+(e^{-\Omega}f) + g - GT^-(e^{-\Omega}f), \quad T^{\pm}(e^{-\Omega}f) = \left\{\frac{1}{\pi}\int\limits_{\mathbb{C}^+}\frac{e^{-\Omega}(\zeta)f(\zeta)}{\zeta - z}d_2\zeta\right\}^{\pm}.$$

Из (9) и (10) следует, что индекс $æ = \operatorname{Ind} G(t)$ этих задач совпадает.

Таким образом, задача (9) приводится к задаче линейного сопряжения для полуплоскости теории аналитических функций, решение которой явным образом находится (см. [13, с. 120]). Следовательно, если $æ = \operatorname{Ind} G(t) \geq 0$, то задача (1), (9) разрешима, ее общее решение дается формулой (4), в которой функция $\phi(z)$ имеет вид

$$\phi(z) = \begin{cases} \dfrac{X(z)}{2\pi i}\displaystyle\int\limits_{\mathbb{R}}\frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{\tau - z} + X(z)\frac{P_æ(z)}{(z+i)^æ} & \text{при} \quad æ \geq 0; \\ X(z)\left[\dfrac{1}{2\pi i}\displaystyle\int\limits_{\mathbb{R}}\frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{\tau - z} + C\right] & \text{при} \quad æ < 0, \end{cases} \tag{11}$$

где

$$X^+(z) = e^{\Gamma^+(z)}, \quad X^-(z) = \left(\frac{z-i}{z+i}\right)^{-æ} e^{\Gamma^-(z)},$$

$$\Gamma(z) = \frac{1}{2\pi i}\int\limits_{\mathbb{R}}\frac{\ln\left[\left(\dfrac{z-i}{z+i}\right)^{-æ} G(\tau)\right]}{\tau - z}d\tau,$$

$P_æ(z)$ – полином степени не выше æ с произвольными комплексными коэффициентами. При $æ < 0$ функция $X(z)$ в точке $z = -i$ имеет полюс порядка $-æ$, поэтому для разрешимости задач нужно положить $C = -\psi(-i)$. При $æ < -1$, кроме того, должно выполняться еще следующее условие:

$$\int\limits_{\mathbb{R}}\frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{(\tau + i)^k} = 0, \quad k = 2, \ldots - æ. \tag{12}$$

Формула (11) выражает решение задачи Римана в полуплоскости для аналитических функций, которое является ограниченным на бесконечности [13], [14].

Как следует из интегрального представления (4), решение изчезает в бесконечно-удаленной точке. Подставляя в краевое условие (10) $\phi^{\pm}(\infty) = 0$, получим $g(\infty) = 0$. Следовательно, чтобы задача линейного сопряжения для полуплоскости имела решение, исчезающее на бесконечности, свободный член краевого условия должен на бесконечности обращаться в нуль. Поэтому предположим, что $g(t) = o(|t|^{-\delta})$, $\delta > 0$, $t \to \infty$. Для получения решения в данном случае нужно в (11) вместо $P_æ$ взять $P_{æ-1}$, а постоянную C приравнять к нулю. Таким образом,

$$\phi(z) = \begin{cases} \dfrac{X(z)}{2\pi i}\displaystyle\int\limits_{\mathbb{R}}\frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{\tau - z} + X(z)\frac{P_{æ-1}(z)}{(z+i)^æ} & \text{при} \quad æ > 0; \\ \dfrac{X(z)}{2\pi i}\displaystyle\int\limits_{\mathbb{R}}\frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{\tau - z} & \text{при} \quad æ \leq 0. \end{cases} \tag{13}$$

При æ ≤ 0 в этой формуле нужно положить $P_{æ} \equiv 0$. К условиям разрешимости нужно добавить еще условие

$$\Psi(i) = \frac{1}{2\pi i}\int\limits_{\mathbb{R}} \frac{\widetilde{g(\tau)}}{X^+(\tau)}\frac{d\tau}{\tau - i} = 0.$$

Таким образом, эти условия примут вид

$$\int\limits_{\mathbb{R}} \frac{\widetilde{g(\tau)}}{X^+(\tau)(\tau + i)^k}\frac{d\tau}{} = 0, \quad k = 1, 2, \ldots -æ. \tag{14}$$

Терема 2. *Пусть* $æ = \operatorname{Ind} G(t) > 0$. *Тогда задача* (1), (9) *для полуплоскости безусловно разрешима и ее общее решение дается формулой* (4), *в которой функция* $\phi(z)$ *определяется формулой* (13), *причем это решение зависит линейно от* æ *произвольных постоянных. Если* $æ \le 0$, *то задача имеет единственное решение. При* $æ < 0$ *разрешима однозначно лишь при выполнении* –æ *условий разрешимости* (14).

СПИСОК ЛИТЕРАТУРЫ

1. *Векуа И.Н.* Обобщенные аналитические функции. М.: Физматгиз, 1959. 628 с.
2. *Михайлов Л.Г.* Новые классы особых интегральных уравнений и их применение к дифференциальным уравнениям с сингулярными коэффициентами. Душанбе: Таджик НИИНТИ, 1963. 183 с.
3. *Усманов З.Д.* Обобщенные системы Коши–Римана с сингулярной точкой. Душанбе: Изд-во АН ТаджССР, 1993. 244 с.
4. *Усманов З.Д.* Связь многообразия решений общей и модельной обобщенных систем Коши - Римана с сингулярной точкой // Матем. заметки. 1999. Т. 66. № 2. С. 302–307.
5. *Begehr, Dai D.Q.* On continuous solutions of a generalized Cauchy–Riemann system with more than one singularity // J. Differen. Eq. 2004. V. 196. P. 67–90.
6. *Abdymanapov S.A., Tungatarov A.B.* Some classes of elliptic systems in the plane with singular coefficients. Almaty: Nauka, 2005. 169 p.
7. *Meziani A.* Representation of solutions of a singular CR equation in the plane // Complex Var. and Elliptic Eq. 2008. V. 53. P. 1111–1130.
8. *Abdymanapov S.A., Begehr H., Harutugian G., Tungatarov A.* Four boundary value problems for the Cauchy–Riemann equation in a quarter plane // More Progresses in analysis. Pro. of the 5th Interna. ISAAC Congress. Catania, Italy, 2005. P. 1137–1147.
9. *Солдатов А.П.* Сингулярные интегральные операторы и эллиптические краевые задачи, I //Современ. матем. Фундамент. напр. Функц. анализ. 2017. Т. 63. № 1. С. 1–189.
10. Об интеграле Помпею и некоторых его обобщениях // Вестник ЮУрГУ ММП. 2021. Т. 14. № 1. С. 53–67.
11. *Расулов А.Б., Солдатов А.П.* Краевая задача для обобщенного уравнения Коши–Римана с сингулярными коэффициентами// Дифференц. ур-ния 2016. Т. 52. № 5. С. 637–650.
12. *Раджабов Н.Р., Расулов А.Б.* О корректной постановке задач для системы Бицадзе со сверхсингулярной точкой и окружностью // Научные Ведомости БелГУ серия математика, физика. 2011. № 23(118). Вып. 25. С. 93–101.
13. *Гахов Ф.Д.* Краевые задачи. М: Наука, 1977. 640 с.
14. *Мусхелишвили Н.И.* Сингулярные интегральные уравнения. М.: Наука, 1968. 511 с.
15. *Расулов А.Б., Дорофеева И.Н.* Задача Дирихле для обобщенного уравнения Коши–Римана с сверхсингулярной точкой на полуплоскости // Ж. вычисл. матем и матем. физ. 2020. Т. 60. № 10. С. 82–88.

EDN: JGZDXJ
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1867

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.633

NUMERICAL SOLUTION OF TWO AND THREE-DIMENSIONAL FRACTIONAL HEAT CONDUCTION EQUATIONS VIA BERNSTEIN POLYNOMIALS[1)]

© 2022 г. M. Gholizadeh[1,*], M. Alipour[1,], M. Behroozifar[1,***]**

[1] *Department of Mathematics, Faculty of Basic Science, Babol Noshirvani University of Technology, Shariati Ave., Babol, 47148-71167, Iran*

**e-mail: gholizadeh.g.1363@gmail.com*

***e-mail: m.alipour2323@gmail.com*

****e-mail: m_behroozifar@nit.ac.ir*

Поступила в редакцию 11.10.2021 г.
Переработанный вариант 10.06.2022 г.
Принята к публикации 07.07.2022 г.

Численное решение двумерных и трехмерных уравнений теплопроводности с дробными производными с помощью полиномов Бернштейна. Разработана новая схема численного решения двумерных и трехмерных уравнений теплопроводности с дробными производными в прямоугольной области. Исследованы операционные матрицы Бернштейна для производной и первообразной в двумерном и трехмерном случаях, которые применяются для решения поставленных задач. В результате проблема сводится к решению системы алгебраических уравнений. Представленный метод применяется для решения ряда модельных задач. Сравнение построенного метода с некоторыми точными решениями показывает малую погрешность разработанного метода.

Ключевые слова: полиномы Бернштейна, уравнения теплопроводности с дробными производными, операционные матрицы.

DOI: 10.31857/S0044466922110035

[1)]Полный текст статьи печатается в английской версии журнала.

EDN: VZNQOV
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1868–1882

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.6

ИНТЕРПОЛЯЦИОННАЯ БАЛАНСНО-ХАРАКТЕРИСТИЧЕСКАЯ СХЕМА С УЛУЧШЕННЫМИ ДИСПЕРСИОННЫМИ СВОЙСТВАМИ ДЛЯ ЗАДАЧ ВЫЧИСЛИТЕЛЬНОЙ ГИДРОДИНАМИКИ[1)]

© 2022 г. Н. А. Афанасьев[1,*], Н. Э. Шагиров[1,], В. М. Головизнин[1,***]**

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

**e-mail: vmnaf@cs.msu.ru*

***e-mail: nikkey.shagirov@yandex.ru*

****e-mail: gol@ibrae.ac.ru*

Поступила в редакцию 09.04.2022 г.
Переработанный вариант 09.04.2022 г.
Принята к публикации 07.07.2022 г.

Балансно-характеристические схемы для численного решения систем гиперболических уравнений объединяют достоинства консервативных методов улавливания скачка и метода характеристик. Они оперируют двумя типами переменных – консервативными и потоковыми. Консервативные переменные имеют смысл средних величин, относятся к серединам ячеек и вычисляются на основе метода конечного объема. Потоковые переменные определяют потоки на гранях расчетных ячеек и рассчитываются с использованием характеристической формы уравнений и локальных инвариантов Римана. Эта часть алгоритма допускает различные реализации, от которых зависят диссипативные и дисперсионные свойства алгоритмов. Так, в схеме КАБАРЕ потоковые величины вычисляются линейной экстраполяцией локальных инвариантов, но существуют и схемы с интерполяцией инвариантов и последующим переносом их по характеристикам (схемы с активными потоками). В последнем случае также возможны различные варианты. Результатам исследования одного из возможных вариантов балансно-характеристических схем интерполяционного типа для систем уравнений гиперболического типа и посвящена эта статья. Библ. 15. Табл. 1. Фиг. 13.

Ключевые слова: вычислительная гидродинамика, балансно-характеристические методы, гиперболические уравнения, инварианты Римана.

DOI: 10.31857/S0044466922110023

1. ВВЕДЕНИЕ

Вычислительная гидродинамика, как правило, имеет дело с системами законов сохранения гиперболического типа с возмущенной правой частью [1], [2]. К таким системам, в частности, относятся уравнения динамики стратифицированной жидкости со свободной границей, описывающие циркуляцию морей и океанов [3]. Сложная природа таких задач требует разработки численных методов высокого порядка точности, обладающих минимальным вычислительным шаблоном, применимых на сетках с произвольной топологией ячеек и эффективных с точки зрения параллелизации.

Одним из активно развивающихся подходов к решению систем уравнений гиперболического типа являются балансно-характеристические методы [4]. Такие методы позволяют учитывать не только дивергентную, но и характеристическую природу уравнений, реконструируя потоки с использованием локальных инвариантов Римана. Для балансно-характеристической схемы КАБАРЕ [5] был проведен полный цикл исследований, начиная от решения одномерного линейного уравнения переноса [6] и заканчивая решением многомерных задач газовой динамики [5] и вычислительной океанологии [7], [8].

Недостатком схемы КАБАРЕ является ограниченность сферы ее применения расчетными сетками с четырехугольными и гексагональными ячейками. Известные обобщения схемы КАБАРЕ на треугольные двумерные сетки [9], [10] обладают рядом недостатков, затрудняющих

[1)]Работа выполнена при финансовой поддержке РНФ (грант 18-11-00163).

их применение для решения прикладных задач. Разработка балансно-характеристических схем с хорошими диссипативными и дисперсионными свойствами на расчетных сетках с более общей структурой ячеек является, таким образом, достаточно актуальной задачей.

В работе [2] был предложен балансно-характеристический метод решения систем гиперболического типа, основанный на параболической реконструкции инвариантов Римана внутри ячейки и последующим их переносом на новый временной слой. Метод был сформулирован для одномерных задач, а для многомерного случая были сформулированы основные идеи его обобщения. В ходе экспериментов выяснилось, что используемые в данном методе процедуры монотонизации на основе принципа максимума [11] неэффективны. Дальнейшее развитие предложенного метода оказалось нецелесообразным.

В настоящей статье предлагается балансно-характеристический метод второго порядка аппроксимации, также основанный на параболической реконструкции инвариантов Римана, но учитывающий интегральный смысл консервативных переменных метода. Для нового метода приводятся дисперсионные и диссипативные поверхности, анализируются его свойства в применении к линейному уравнению переноса. Метод тестируется на одномерных уравнениях переноса, Хопфа и мелкой воды.

Текст организован следующим образом. В разд. 2 приведены метод [2] и новый метод для линейного уравнения переноса, проанализированы их свойства. В разд. 3 описывается новый балансно-характеристический метод для одномерных систем законов сохранения гиперболического типа, обсуждаются его особенности. Тестовые расчеты для уравнений переноса, Хопфа и мелкой воды приводятся в разд. 4. Статья завершается разд. 5 с заключительными замечаниями.

2. ОПИСАНИЕ МЕТОДА ДЛЯ УРАВНЕНИЯ ПЕРЕНОСА

Рассмотрим простейшее одномерное линейное однородное уравнение переноса

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0, \quad c = \text{const} > 0. \tag{1}$$

Здесь x – пространственная координата, t – время. Для определенности будем считать, что $(x, t) \in \Omega = [a, b]\times[0, T]$ и дополним наше уравнение начальным условием

$$u(x,0) = f(x), \quad a \le x \le b, \tag{2}$$

и некоторым граничным условием (например, условием периодичности).

Введем пространственно-временную сетку $\omega = \omega_h \times \omega_\tau$, где $\omega_h = \{x_j | a = x_0 < x_1 < \ldots < x_N = b;$ $x_{j+1} - x_j = h,\ j = \overline{0, N-1}\}$ – сетка по пространству, $\omega_\tau = \{t_n | 0 = t_0 < t_1 < \ldots < t_K = T;\ t_{n+1} - t_n = \tau,$ $n = \overline{0, K-1}\}$ – сетка по времени. Определим в центрах пространственных ячеек так называемые консервативные переменные: $U^n_{j+1/2}$ – на целых слоях по времени, $U^{n+1/2}_{j+1/2}$ – на полуцелых слоях по времени. В узлах сетки ω определим так называемые потоковые переменные: u^n_j.

2.1. Метод ICCh-1 для линейного уравнения переноса

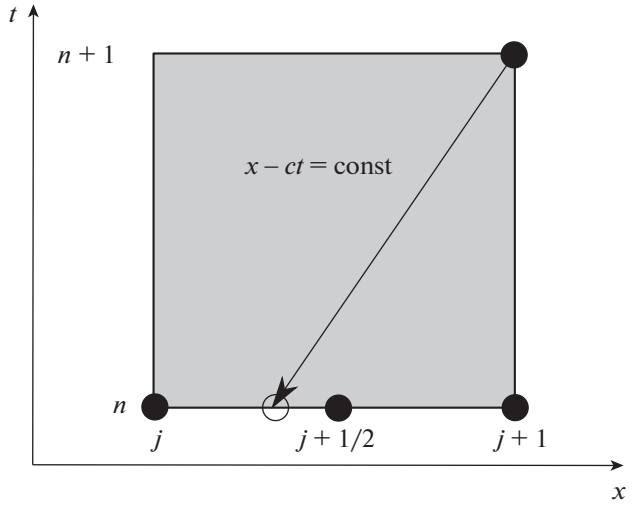
Ранее в работе [2] был предложен новый балансно-характеристический метод интерполяционного типа для решения уравнения (1):

$$\frac{U^{n+1/2}_{j+1/2} - U^n_{j+1/2}}{\tau/2} + c\frac{u^n_{j+1} - u^n_j}{h} = 0, \tag{3}$$

$$u^{n+1}_{j+1} = (1 - 3r + 2r^2)u^n_{j+1} + 4r(1-r)U^n_{j+1/2} + 2r(r-0.5)u^n_j, \tag{4}$$

$$\frac{U^{n+1}_{j+1/2} - U^{n+1/2}_{j+1/2}}{\tau/2} + c\frac{u^{n+1}_{j+1} - u^{n+1}_j}{h} = 0, \tag{5}$$

где $r = c\tau/h$ – число Куранта–Фридрихса–Леви. Уравнения (3) и (5) (консервативные фазы) есть консервативные аппроксимации (1) по методу конечного объема в ячейке $[x_j, x_{j+1}]$ на слоях n и $n+1$ соответственно. Уравнение (4) (характеристическая фаза) представляет собой перенос инварианта Римана u по характеристике $x - ct = \text{const}$, опущенной из точки (x_{j+1}, t_{n+1}) на слой по



Фиг. 1. Шаблон характеристической фазы алгоритма.

времени n (см. фиг. 1). При этом переносимое значение u восполняется по параболе $P_2(x)$, построенной по 3 значениям на нижнем слое: $u_{j+1}^{n+1} = P_2(x_{j+1} - c\tau)$,

$$P_2(x_{j+1}) = u_{j+1}^n, \tag{6}$$

$$\begin{aligned} P_2(x_j) &= u_j^n, \\ P_2(x_{j+1/2}) &= U_{j+1/2}^n. \end{aligned} \tag{7}$$

Метод *ICCh-1* (*interpolatory conservative-characteristic 1*) (3)–(5) является устойчивым при числах Куранта $r \in [0,1]$ и обладает вторым порядком аппроксимации, но, в отличие от балансно-характеристической схемы КАБАРЕ [5], не является обратимым по времени. Для получения монотонного решения метод также дополняется процедурой нелинейной коррекции потоков на основе принципа максимума [11] после характеристической фазы (4):

$$u_{j+1}^{n+1} = \begin{cases} \tilde{u}_{j+1}^{n+1}, & \text{если} \quad [\min u]_{j+1/2}^n \le \tilde{u}_{j+1}^{n+1} \le [\max u]_{j+1/2}^n, \\ [\min u]_{j+1/2}^n, & \text{если} \quad \tilde{u}_{j+1}^{n+1} < [\min u]_{j+1/2}^n, \\ [\max u]_{j+1/2}^n, & \text{если} \quad \tilde{u}_{j+1}^{n+1} > [\max u]_{j+1/2}^n, \end{cases} \tag{8}$$

где $\tilde{u}_{j+1}^{n+1}$ – потоковое значение, полученное в результате характеристической фазы (4), и

$$[\min u]_{j+1/2}^n = \min\{u_j^n, U_{j+1/2}^n, u_{j+1}^n\},$$

$$[\max u]_{j+1/2}^n = \max\{u_j^n, U_{j+1/2}^n, u_{j+1}^n\}.$$

Метод *ICCh-1* разрабатывался в первую очередь для того, чтобы распространить его на двумерные треугольные и трехмерные тетраэдральные сетки, на которых применение схемы КАБАРЕ достаточно проблематично [9], [10]. Но данный метод обладает нормальной дисперсией (т.е. бегущие волны всех гармоник запаздывают по отношению к аналитическому решению) при малых числах Куранта, что отражено на его дисперсионной поверхности (см. фиг. 3). Экспериментальные расчеты показывают, что нормальная дисперсия метода делает процедуру монотонизации (8) неэффективной, и затрудняет использование метода даже на простейших нелинейных уравнениях гиперболического типа. Актуальной становится разработка интерполяционного балансно-характеристического метода, обладающего схожим с методом *ICCh-1* вычислительным шаблоном и хотя бы частично аномальной дисперсией при малых числах Куранта.

2.2. Метод ICCh-2 для линейного уравнения переноса

Для построения модифицированного балансно-характеристического метода интерполяционного типа изменим характеристическую фазу метода *ICCh-1* (4), а консервативные фазы (3) и (5) оставим прежними. Как и ранее, потоковое значение u_{j+1}^{n+1} мы будем переносить по характеристике $x - ct = \text{const}$, опущенной из точки (x_{j+1}, t_{n+1}) на слой по времени n (см. фиг. 1), но функцию $u(x)$ на нижнем слое восполним с помощью другой параболы $\tilde{P}_2(x)$.

Так как консервативные фазы (3), (5) есть аппроксимации закона сохранения (1), то консервативные переменные $U_{j+1/2}^n$ приближают средние значения $u(x,t_n)$ по ячейкам $[x_j, x_{j+1}]$:

$$U_{j+1/2}^n \approx \frac{1}{h}\int\limits_{x_j}^{x_{j+1}} u(x,t_n)dx.$$

Воспользуемся этим свойством и построим на нижнем слое параболу $\tilde{P}_2(x)$, удовлетворяющую условиям (6), (7) и интегральному условию:

$$\frac{1}{h}\int\limits_{x_j}^{x_{j+1}} \tilde{P}_2(x)dx = U_{j+1/2}^n.$$

Тогда перенос потокового значения по характеристике $u_{j+1}^{n+1} = \tilde{P}_2(x_{j+1} - c\tau)$ приведет к отличной от (4) характеристической фазе:

$$u_{j+1}^{n+1} = (1 - 4r + 3r^2)u_{j+1}^n + 6r(1-r)U_{j+1/2}^n + r(3r-2)u_j^n. \tag{9}$$

Отметим, что способ параболической реконструкции инварианта Римана (9) также используется в *V*-схеме Ван Лира [12] и методе *Active Flux* третьего порядка [13]. Схему (3), (9), (5) назовем *ICCh-2*-методом (interpolatory conservative-characteristic 2).

2.3. Свойства метода ICCh-2

Для анализа диссипативных и дисперсионных свойств полученной разностной схемы подставим в уравнения (3), (9), (5) частное решение в виде бегущей волны:

$$u_j^n = A\exp\{i(\omega n\tau - kjh)\} = Aq^n\exp\{-i(kjh)\}, \quad q = e^{i\omega\tau},$$

$$U_{j+1/2}^n = B\exp\{i(\omega n\tau - k(j+0.5)h)\} = Bq^n\exp\{-i[k(j+0.5)h]\}.$$

Таким образом, получим следующее квадратное уравнение для определения q:

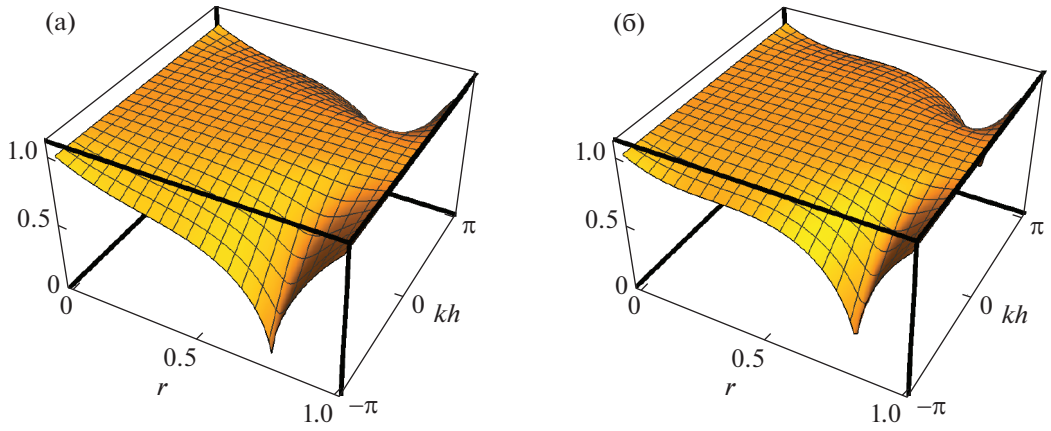
$$q^2 + q[4r - 3r^3 - 2 + e^{ikh}(3r^3 - 6r^2 + 2r)] + 1 - 4r + 6r^2 - 3r^3 + e^{ikh}(3r^3 - 2r) = 0.$$

Данное уравнение имеет два корня: $q_1 = q_1(r,kh)$, $q_2 = q_2(r,kh)$, которые зависят от приведенного волнового числа $kh \in [-\pi,\pi]$, а также от числа Куранта r. Модули этих корней определяют область устойчивости схемы: схема устойчива, если $|q_1| \le 1$ и $|q_2| \le 1$. Относительная дисперсия разностной схемы задается величиной

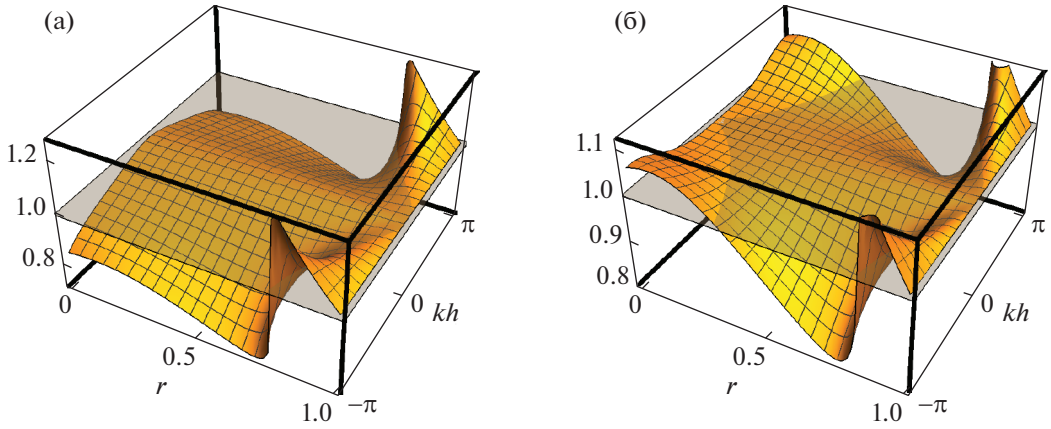
$$\gamma(r,kh) = \frac{\omega}{kc} = -\frac{i}{rkh}\ln[q(r,kh)]$$

и определяет степень отклонения фазовой скорости бегущей волны относительно аналитического решения уравнения (1). Дисперсия называется нормальной, если $|\gamma(r,kh)| < 1$ (т.е. бегущая волна отстает от аналитического решения), и аномальной, если $|\gamma(r,kh)| > 1$ (т.е. бегущая волна опережает аналитическое решение). Отметим, что один из двух корней (обозначим его как q_2) всегда является паразитным, и не несет в себе никакой полезной информации о свойствах схемы (кроме разве что определения области неустойчивости схемы при $|q_2| > 1$).

На фиг. 2 приведены диссипативные поверхности первых корней $z = |q_1(r,kh)|$ методов *ICCh-1* (а) и *ICCh-2* (б) для чисел Куранта $r \in [0,1]$, на фиг. 3 – дисперсионные поверхности первых корней $z = |\gamma_1(r,kh)|$ методов *ICCh-1* (а) и *ICCh-2* (б). На фиг. 4 также дополнительно приведены дис-



Фиг. 2. Диссипативные поверхности первых корней $z = |q_1(r, kh)|$ методов *ICCh-1* (а) и *ICCh-2* (б).



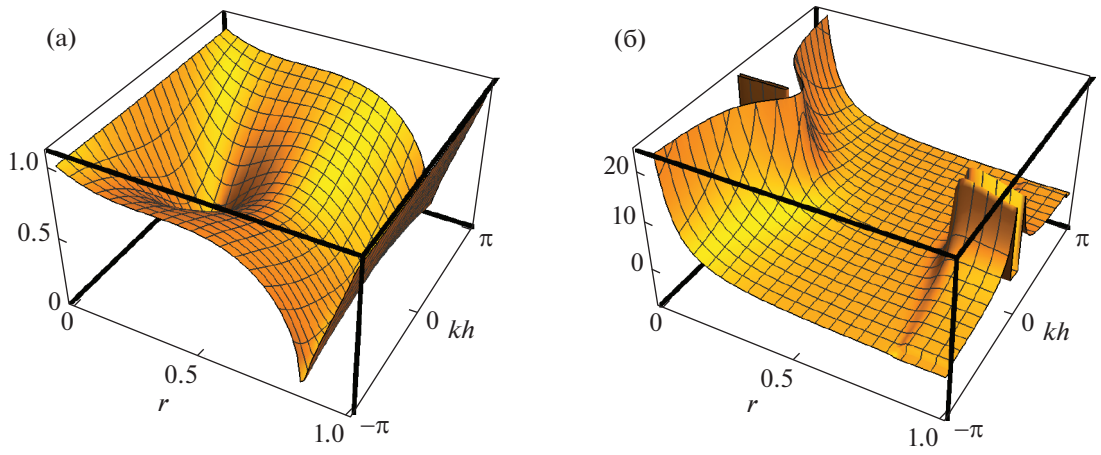
Фиг. 3. Дисперсионные поверхности первых корней $z = |\gamma_1(r, kh)|$ методов *ICCh-1* (а) и *ICCh-2* (б).

сипативная (а) и дисперсионная (б) поверхности второго корня метода *ICCh-2*. Поверхности для чисел Куранта $r \in (1, +\infty)$ не приводятся, так как в этой области оба метода являются неустойчивыми ($|q_1| > 1$).

Как видно из диссипативных поверхностей на фиг. 2, метод *ICCh-2*, как и метод *ICCh-1*, устойчив при $0 \le r \le 1$ и необратим по времени, но при этом обладает меньшей диссипацией. Дисперсионные поверхности на фиг. 3 показывают, что метод *ICCh-2* обладает аномальной дисперсией при малых числах Куранта, но только для высоких гармоник. Метод *ICCh-1* при этом обладает только нормальной дисперсией в области малых чисел Куранта. Именно наличие области аномальной дисперсии позволяет частично монотонизировать метод *ICCh-2*, что будет показано на тестовых расчетах.

В работе [2] было показано, что метод *ICCh-1* обладает вторым порядком аппроксимации как по времени, так и по пространству. Покажем, что и для метода *ICCh-2* это выполняется. Для этого просуммируем уравнения (3), (5) и подставим вместо потоковых значений u_j^{n+1} и u_{j+1}^{n+1} выражения, задаваемые (9):

$$\frac{U_{j+1/2}^{n+1} - U_{j+1/2}^{n}}{\tau} + \frac{c}{2}\frac{u_{j+1}^{n} - u_{j}^{n}}{h} + \frac{c}{2}\left[(1 - 4r + 3r^2)\frac{u_{j+1}^{n} - u_{j}^{n}}{h} + \right.$$
$$\left. + 6r(1-r)\frac{U_{j+1/2}^{n} - U_{j-1/2}^{n}}{h} + r(3r-2)\frac{u_{j}^{n} - u_{j-1}^{n}}{h}\right] = 0. \tag{10}$$



Фиг. 4. Диссипативная поверхность $z = |q_2(r, kh)|$ и дисперсионная поверхность $z = |\gamma_2(r, kh)|$ второго (паразитного) корня метода *ICCh-2*.

Заменяя в (10) значения сеточных функций на значения аналитического решения $u(x,t)$ в соответствующих точках и раскладывая $u(x,t)$ в ряд Тейлора в окрестности точки $(x_{j+1/2}, t_n)$, можно получить:

$$u_t(x_{j+1/2},t_n) + cu_x(x_{j+1/2},t_n) - \frac{c^2\tau}{2}u_{xx}(x_{j+1/2},t_n) + \frac{\tau}{2}u_{tt}(x_{j+1/2},t_n) + \\ + \left(\frac{ch^2}{24} - \frac{c^2\tau h}{8} + \frac{5c^3\tau^2}{24}\right)u_{xxx}(x_{j+1/2},t_n) + \ldots = 0. \tag{11}$$

Воспользовавшись в (11) тем, что $u_t + cu_x = 0$ и $u_{tt} = c^2u_{xx}$ (что получается дифференцированием (1) по времени), получим следующий вид для погрешности аппроксимации в точке $(x_{j+1/2}, t_n)$:

$$\psi^n_{j+1/2} = \left(\frac{ch^2}{24} - \frac{c^2\tau h}{8} + \frac{5c^3\tau^2}{24}\right)u_{xxx}(x_{j+1/2},t_n) + \ldots . \tag{12}$$

Учитывая, что при решении уравнений гиперболического типа явными методами τ и h есть величины одного порядка, (12) позволяет заключить, что метод *ICCh-2* обладает вторым порядком аппроксимации. Как и метод *ICCh-1*, для получения монотонного решения метод *ICCh-2* надо дополнять процедурой монотонизации (8).

3. ОПИСАНИЕ МЕТОДА ДЛЯ СИСТЕМ УРАВНЕНИЙ ГИПЕРБОЛИЧЕСКОГО ТИПА

Рассмотрим систему одномерных квазилинейных законов сохранения:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{u})}{\partial x} = \mathbf{Q}(x,t,\mathbf{u}), \tag{13}$$

где $\mathbf{u}$ – вектор неизвестных, $\mathbf{F}(\mathbf{u})$ – векторная функция потоков, $\mathbf{Q}(x,t,\mathbf{u})$ – векторная функция источников и стоков. Систему (13) будем называть *консервативной формой* уравнений. Раскрывая в (13) производную по x, получим так называемую *простую форму* уравнений:

$$\frac{\partial \mathbf{u}}{\partial t} + A(\mathbf{u})\frac{\partial \mathbf{u}}{\partial x} = \mathbf{Q}(x,t,\mathbf{u}), \tag{14}$$

где $A(\mathbf{u})$ – матрица Якоби $\mathbf{F}(\mathbf{u})$. Предполагая, что система (14) является гиперболической (т.е. все собственные значения $A(\mathbf{u})$ действительны и имеется полный базис из собственных векторов) и что все аналитические выражения для инвариантов Римана известны (что выполняется для уравнения Хопфа и уравнений мелкой воды), получаем *характеристическую форму* уравнений:

$$\frac{\partial \mathbf{I}}{\partial t} + \Lambda(\mathbf{u})\frac{\partial \mathbf{I}}{\partial x} = \widehat{\mathbf{Q}}(x,t,\mathbf{I}),$$

где $\mathbf{I} = \mathbf{I}(\mathbf{u}) = \{I_k\}$ – вектор инвариантов Римана, $\Lambda(\mathbf{u}) = \mathrm{diag}\{\lambda_k\}$ – диагональная матрица собственных значений матрицы $A(\mathbf{u})$, $\widehat{\mathbf{Q}}(x,t,\mathbf{I})$ – некоторая правая часть (точный вид которой нам неважен).

Метод *ICCh-2* для системы законов сохранения (13) на неравномерной пространственно-временной сетке с шагами $\tau_n = t_{n+1} - t_n$ и $h_{j+1/2} = x_{j+1} - x_j$ состоит из 3 фаз.

Фаза 1 *(консервативная)*. Явное вычисление консервативных переменных на полуцелом слое $\mathbf{U}_{j+1/2}^{n+1/2}$ с помощью аппроксимации консервативных уравнений (13):

$$\frac{\mathbf{U}_{j+1/2}^{n+1/2} - \mathbf{U}_{j+1/2}^{n}}{\tau_n/2} + \frac{\mathbf{F}(\mathbf{u}_{j+1}^{n}) - \mathbf{F}(\mathbf{u}_{j}^{n})}{h_{j+1/2}} = \mathbf{Q}_{j+1/2}^{n}.$$

После вычисления консервативных переменных на промежуточном слое находятся собственные значения $(\lambda_k)_{j+1/2}^{n+1/2}$, которые будут определять направление переноса инвариантов Римана в характеристической фазе алгоритма.

Фаза 2 *(характеристическая)*. Явное вычисление потоковых значений инвариантов Римана на новом временном слое $\mathbf{I}_j^{n+1}$ с помощью переноса по соответствующим характеристикам (см. обозначения на фиг. 5):

$$(\tilde{I}_k)_j^{n+1} = \begin{cases} \varphi[(I_k)_j^n, (I_k)_{cL}^n, (I_k)_L^n, r_L] + \tau_n(\hat{Q}_k)_{cL}^n, & \text{если} \quad (\lambda_k)_{cL}^{n+1/2} > 0 \quad \text{и} \quad (\lambda_k)_{cR}^{n+1/2} > 0, \\ \varphi[(I_k)_j^n, (I_k)_{cR}^n, (I_k)_R^n, r_R] + \tau_n(\hat{Q}_k)_{cR}^n, & \text{если} \quad (\lambda_k)_{cL}^{n+1/2} < 0 \quad \text{и} \quad (\lambda_k)_{cR}^{n+1/2} < 0, \\ 0.5[(I_k)_{cL}^{n+1/2} + (I_k)_{cR}^{n+1/2}] + 0.5\tau_n[(\hat{Q}_k)_{cL}^n + (\hat{Q}_k)_{cR}^n] & \text{иначе}, \end{cases}$$

$$r_L = (\lambda_k)_{cL}^{n+1/2}\frac{\tau_n}{h_{j-1/2}}, \quad r_R = \left|(\lambda_k)_{cR}^{n+1/2}\right|\frac{\tau_n}{h_{j+1/2}}, \tag{15}$$

$$\varphi[\alpha,\beta,\gamma,r] = (1 - 4r + 3r^2)\alpha + 6r(1-r)\beta + r(3r-2)\gamma,$$

$$(\hat{Q}_k)_{j+1/2}^n = \frac{(I_k)_{j+1/2}^{n+1/2} - (I_k)_{j+1/2}^n}{\tau_n / 2} + (\lambda_k)_{j+1/2}^{n+1/2}\frac{(I_k)_{j+1}^n - (I_k)_j^n}{h_{j+1/2}}.$$

Требуемые значения инвариантов Римана на слоях n и $n+1/2$ при этом вычисляются по известным аналитическим формулам $\mathbf{I} = \mathbf{I}(\mathbf{u})$. После переноса (15) инварианты Римана корректируются с помощью процедуры монотонизации на основе принципа максимума:

$$(I_k)_j^{n+1} = \begin{cases} \mathrm{Max}\,\mathrm{Min}[(\tilde{I}_k)_j^{n+1}]_{j-1/2}, & \text{если} \quad (\lambda_k)_{j-1/2}^{n+1/2} > 0 \quad \text{и} \quad (\lambda_k)_{j+1/2}^{n+1/2} > 0, \\ \mathrm{Max}\,\mathrm{Min}[(\tilde{I}_k)_j^{n+1}]_{j+1/2}, & \text{если} \quad (\lambda_k)_{j-1/2}^{n+1/2} < 0 \quad \text{и} \quad (\lambda_k)_{j+1/2}^{n+1/2} < 0, \\ (\tilde{I}_k)_j^{n+1} & \text{иначе}, \end{cases}$$

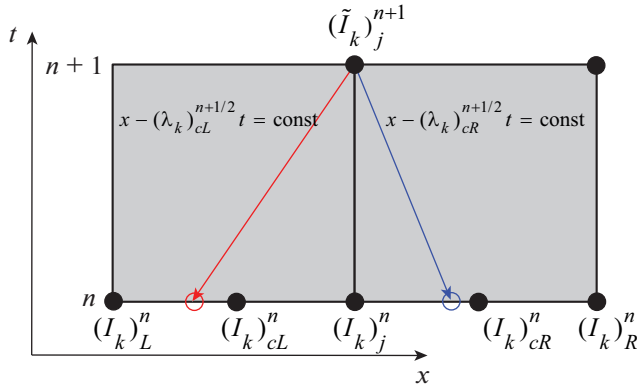
$$\mathrm{Max}\,\mathrm{Min}[(\tilde{I}_k)_j^{n+1}]_{j+1/2} = \begin{cases} (\tilde{I}_k)_j^{n+1}, & \text{если} \quad \min(I_k)_{j+1/2}^n \le (\tilde{I}_k)_j^{n+1} \le \max(I_k)_{j+1/2}^n, \\ \max(I_k)_{j+1/2}^n, & \text{если} \quad (\tilde{I}_k)_j^{n+1} > \max(I_k)_{j+1/2}^n, \\ \min(I_k)_{j+1/2}^n, & \text{если} \quad (\tilde{I}_k)_j^{n+1} < \min(I_k)_{j+1/2}^n, \end{cases} \tag{16}$$

$$\min(I_k)_{j+1/2}^n = \min\{(I_k)_j^n, (I_k)_{j+1/2}^n, (I_k)_{j+1}^n\} + \tau_n(\hat{Q}_k)_{j+1/2}^n,$$

$$\max(I_k)_{j+1/2}^n = \max\{(I_k)_j^n, (I_k)_{j+1/2}^n, (I_k)_{j+1}^n\} + \tau_n(\hat{Q}_k)_{j+1/2}^n.$$

После нахождения полного набора инвариантов Римана на следующем слое по времени $n+1$ по известным аналитическим формулам вычисляются потоковые значения исходных неизвестных $\mathbf{u}_j^{n+1} = \mathbf{u}(\mathbf{I}_j^{n+1})$.

Отметим, что в процессе проведения расчетов могут возникать так называемые звуковые точки, когда в один пространственно-временной узел приходят либо 0 ($(\lambda_k)_{j-1/2}^{n+1/2} < 0$ и $(\lambda_k)_{j+1/2}^{n+1/2} > 0$), либо 2 инварианта Римана ($(\lambda_k)_{j-1/2}^{n+1/2} > 0$ и $(\lambda_k)_{j+1/2}^{n+1/2} < 0$). В таких случаях требуется отдельный алгоритм обработки звуковых точек. В алгоритме (15) представлен самый простой способ такой об-



Фиг. 5. Шаблон переноса (15) инварианта Римана $(I_k)_j^{n+1}$ слева (красный) и справа (синий).

работки, когда на звуковых точках инвариант осредняется по значениям из центров левой и правой пространственно-временных ячеек. Более сложные алгоритмы обработки звуковых точек представлены в [14], [15].

Фаза 3 *(консервативная)*. Вычисление консервативных переменных на следующем слое $U_{j+1/2}^{n+1}$ с помощью аппроксимации консервативных уравнений (13):

$$\frac{\mathbf{U}_{j+1/2}^{n+1} - \mathbf{U}_{j+1/2}^{n+1/2}}{\tau_n/2} + \frac{\mathbf{F}(\mathbf{u}_{j+1}^{n+1}) - \mathbf{F}(\mathbf{u}_j^{n+1})}{h_{j+1/2}} = \mathbf{Q}_{j+1/2}^{n+1}. \tag{17}$$

Схема (17) является, вообще говоря, неявной в силу наличия в правой части члена $\mathbf{Q}_{j+1/2}^{n+1}$, зависящего от еще не известной консервативной переменной $\mathbf{U}_{j+1/2}^{n+1}$. Тем не менее данное уравнение удается разрешить явно при линейной зависимости $\mathbf{Q}(x,t,\mathbf{u})$ от $\mathbf{u}$.

В конце третьей фазы с помощью собственных значений на новом слое $(\lambda_k)_{j+1/2}^{n+1}$ по заданному числу Куранта $CFL \in (0,1]$ вычисляется величина следующего шага по времени:

$$\tau_{n+1} = CFL \cdot \min_k \min_j \left[\frac{h_{j+1/2}}{|(\lambda_k)_{j+1/2}^{n+1}|} \right].$$

4. ТЕСТОВЫЕ РАСЧЕТЫ

В данном разделе представлены результаты применения предложенного метода *ICCh-2* к решению некоторых уравнений и систем уравнений в частных производных гиперболического типа. Рассматривались начально-краевые задачи для линейного уравнения переноса, уравнения Хопфа, а также для системы уравнений мелкой воды над плоским дном. Для сравнения приводятся результаты расчетов по схеме КАБАРЕ и методу *ICCh-1*.

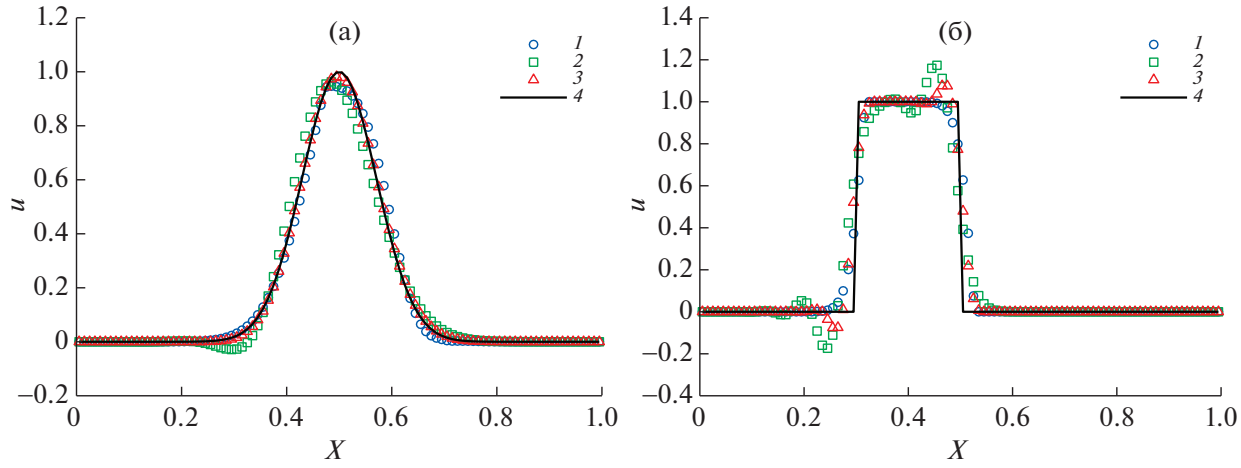
4.1. Линейное уравнение переноса

Предложенный метод *ICCh-2* был протестирован на задаче (1), (2) с периодическими граничными условиями на отрезке $x \in [0, 1]$ при двух видах начальных условий – достаточно гладком и разрывном. В качестве гладкого начального условия был взят “одиночный гауссиан”

$$f(x) = \exp[-(x - x_0)^2/\Delta^2]; \quad x_0 = 0.5, \quad \Delta = 0.1, \tag{18}$$

в качестве разрывного – прямоугольник:

$$f(x) = \begin{cases} 1, & 0.3 \le x \le 0.5, \\ 0 & \text{иначе}. \end{cases} \tag{19}$$



Фиг. 6. Перенос начальных профилей (18), (19) для уравнения переноса (1) на 100 ячейках, $CFL = 0.3$. (а) – результат переноса начального профиля (18) на момент времени $t = 7.0$; (б) – результат переноса начального профиля (19) на момент времени $t = 1.0$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-1*, *3* – решение, полученное методом *ICCh-2*, *4* – аналитическое решение.

Скорость переноса c принималась равной единице.

Начальная инициализация потоковых и консервативных переменных проводилась следующим образом. В качестве начальных значений потоковых переменных использовались значения начального профиля в соответствующих узлах сетки ω_h, значения консервативных переменных принимались равными полусуммам значений соседних потоковых переменных:

$$u_j^0 = f(x_j), \quad j = \overline{0, N}, \tag{20}$$

$$U_{j+1/2}^0 = 0.5(u_j^0 + u_{j+1}^0), \quad j = \overline{0, N-1}. \tag{21}$$

Сравнивались результаты, полученные по методам *ICCh-2* и *ICCh-1*, а также по схеме КАБАРЕ [4]. На фиг. 6 представлены результаты расчетов переноса начальных профилей (18), (19) на равномерной сетке из $N = 100$ ячеек при числе Куранта $CFL = 0.3$. Картинка (а) соответствует результату переноса начального профиля (18) и приводится на момент времени $t = 7.0$, что соответствует 7 проходам профиля по отрезку [0,1]. Картинка (б) соответствует результату переноса начального профиля (19) и приводится на момент времени $t = 1.0$, что соответствует 1 проходу профиля по отрезку [0,1]. Результаты на фиг. 6 показывают, что в случае переноса гладкого профиля при включенных процедурах монотонизации (8) метод *ICCh-2* обладает меньшей численной дисперсией и диссипацией, чем методы КАБАРЕ и *ICCh-1*. При переносе разрывного профиля как у метода *ICCh-1*, так и у метода *ICCh-2* возникают немонотонности, которые не удается сгладить процедурой монотоназиции. Тем не менее, так как метод *ICCh-2* при $CFL = 0.3$ обладает частичной аномальной дисперсией, возникающие при его использовании немонотонности минимальны.

В табл. 1 также приведены результаты численных исследований сходимости метода *ICCh-2* на переносе гладкого профиля (18) при отключенной процедуре монотонизации (8). В табл. 1 указаны ошибки по C-норме $\|u\|_c = \max\limits_j |u_{j+1/2}|$ и порядки сходимости (*OOC* – order of convergence) на момент времени $t = 1.0$ на разных сетках. Результаты в табл. 1 позволяют заключить, что метод *ICCh-2* обладает вторым порядком сходимости.

4.2. Уравнение Хопфа

Уравнение Хопфа имеет следующий вид:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = 0. \tag{22}$$

Таблица 1. Ошибки по C-норме и порядки сходимости (OOC) для переноса гладкого профиля (18) по методу *ICCh-2*

Ячейки	C-ошибка $CFL = 0.3$	OOC $CFL = 0.3$	C-ошибка $CFL = 0.6$	OOC $CFL = 0.6$	C-ошибка $CFL = 0.9$	OOC $CFL = 0.9$
100	4.57×10^{-3}	–	1.25×10^{-2}	–	2.5×10^{-2}	–
200	1.11×10^{-3}	2.04	3.17×10^{-3}	1.98	6.69×10^{-3}	1.90
400	2.71×10^{-4}	2.03	7.94×10^{-4}	2.00	1.7×10^{-3}	1.98
800	6.7×10^{-5}	2.02	1.99×10^{-4}	2.00	4.26×10^{-4}	2.00
1600	1.7×10^{-5}	1.98	5×10^{-5}	1.99	1.07×10^{-4}	1.99

Единственный инвариант Римана уравнения (22) и скорость его переноса совпадают с неизвестной функцией u.

Рассмотрим уравнение (22) в области $\Omega = [0, 1]\times[0, T]$ с периодическими граничными условиями и начальным условием в виде прямоугольника:

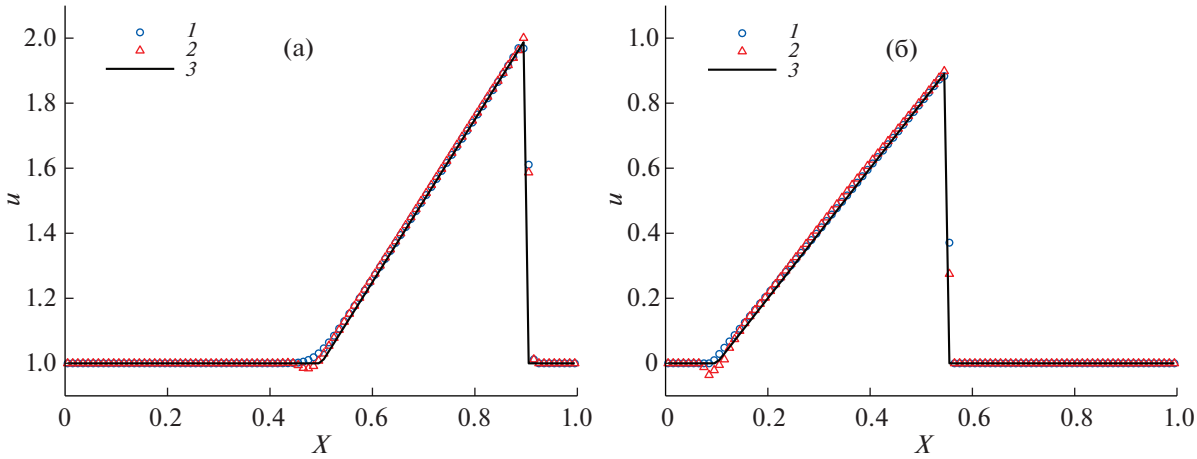
$$u_0(x) = \begin{cases} u_2, & 0.1 \le x \le 0.3, \\ u_1 & \text{иначе.} \end{cases} \tag{23}$$

На фиг. 7 представлены результаты расчетов переноса начального профиля (23) при $u_1 = 1$, $u_2 = 2$ на момент времени $t = 0.4$ (а) и при $u_1 = 0$, $u_2 = 1$ на момент времени $t = 0.5$ (б) на равномерной сетке из $N = 100$ ячеек при числе Куранта $CFL = 0.3$. Начальные значения потоковых и консервативных переменных задавались аналогично (20), (21). В обоих расчетах у метода *ICCh-2* возникают некоторые немонотонности слева от негладких участков решения. При этом для расчета задачи (б) это приводит к тому, что решение опускается ниже 0 и возникает звуковая точка, алгоритм обработки которой (15) не приводит к аварийному останову. Отметим, что схема КАБАРЕ не обладает данным недостатком, так как процедура монотонизации (16) не дает решению уйти ниже 0.

4.3. Уравнения мелкой воды

Рассмотрим систему одномерных уравнений мелкой воды над ровным дном:

$$\frac{\partial H}{\partial t} + \frac{\partial Hu}{\partial x} = 0, \quad \frac{\partial Hu}{\partial t} + \frac{\partial Hu^2}{\partial x} + \frac{g}{2}\frac{\partial H^2}{\partial x} = 0. \tag{24}$$



Фиг. 7. Перенос начального профиля (23) для уравнения Хопфа (22) на 100 ячейках при $CFL = 0.3$: (а) – $u_1 = 1$, $u_2 = 2$, профиль при $t = 0.4$; (б) – $u_1 = 0$, $u_2 = 1$, профиль при $t = 0.5$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.

Здесь $H(x,t)$ – глубина жидкости, $u(x,t)$ – горизонтальная скорость столба жидкости, g – ускорение свободного падения. Система (24) обладает следующими инвариантами Римана I_k и соответствующими им собственными значениями λ_k:

$$I_1 = u + 2\sqrt{gH}, \quad \lambda_1 = u + \sqrt{gH},$$
$$I_2 = u - 2\sqrt{gH}, \quad \lambda_2 = u - \sqrt{gH}.$$

Тестирование метода *ICCh-2* проводилось для различных задач Римана для дозвуковых, трансзвуковых и сверхзвуковых течений:

$$H_0(x) = \begin{cases} H_L, & a \le x \le x^*, \\ H_R, & x^* < x \le b, \end{cases} \quad u_0(x) = \begin{cases} u_L, & a \le x \le x^*, \\ u_R, & x^* < x \le b. \end{cases} \tag{25}$$

Граничные условия мы не приводим, так как все расчеты для следующих тестовых задач останавливаются до того, как какие-либо волны разрежения или ударные волны достигнут границ рассматриваемой области.

4.3.1. Дозвуковые течения. Ниже представлены результаты по методу *ICCh-2* и схеме КАБАРЕ для строго дозвуковых течений ($|u| < \sqrt{gH}$). Были рассмотрены следующие задачи Римана (25) на сегменте $x \in [0, 1]$:

1) волна разрежения слева, ударная волна справа

$$H_L = 2, \quad H_R = 1, \quad u_L = u_R = 0, \quad x^* = 0.5; \tag{26}$$

2) две волны разрежения

$$H_L = H_R = 1, \quad u_L = -1, \quad u_R = 1, \quad x^* = 0.5; \tag{27}$$

3) сталкивающиеся ударные волны

$$H_L = H_R = 1, \quad u_L = 1, \quad u_R = -1, \quad x^* = 0.5. \tag{28}$$

На фиг. 8–10 приведены результаты расчетов задач Римана (26), (27) и (28), соответственно, по методу *ICCh-2* и схеме КАБАРЕ для $CFL = 0.3$. Отметим основные свойства решений, полученных по методу *ICCh-2*: разрывы, как и в схеме КАБАРЕ, занимают 2–4 расчетные ячейки; в областях негладкости решений в силу частично аномальной дисперсии схемы возникают некритичные немонотонности; волны разрежения передаются достаточно точно и без артефактов. Отметим, что схема КАБАРЕ позволяет сохранить “острые” участки решения в верхних частях волн разрежения [15]. Метод *ICCh-2* таким свойством не обладает и сглаживает все области негладкости решения.

4.3.2. Сверхзвуковое течение. Здесь представлен пример расчета по методу *ICCh-2* и схеме КАБАРЕ для случая строго сверхзвукового течения ($|u| > \sqrt{gH}$). Рассматривалась следующая задача Римана (25) на сегменте $x \in [0, 50]$:

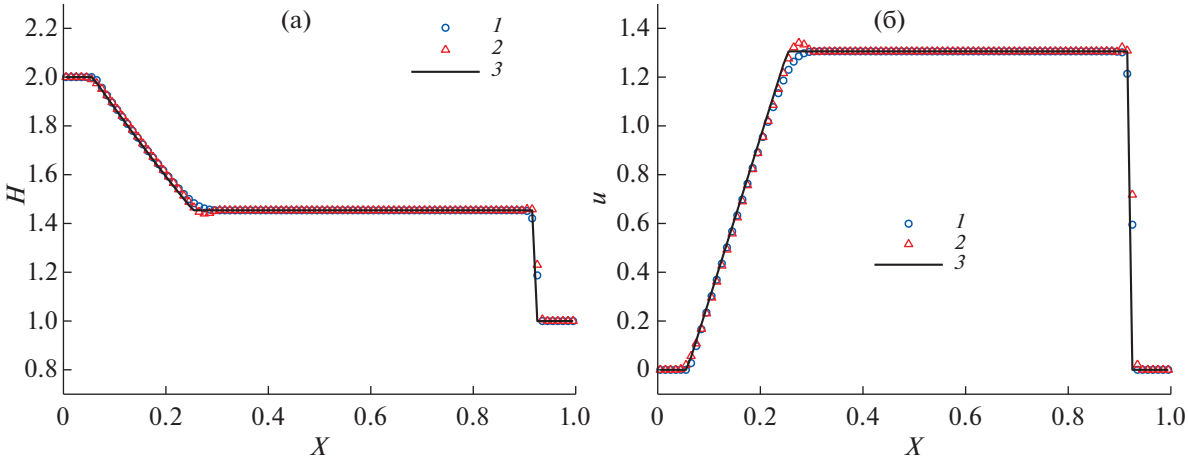
$$H_L = 1, \quad H_R = 0.1, \quad u_L = 5, \quad u_R = 2.5, \quad x^* = 10.0. \tag{29}$$

На фиг. 11 представлен результат расчета задачи (29) на равномерной сетке из $N = 101$ ячеек на момент времени $t = 5.0$. В случае строго сверхзвукового течения метод *ICCh-2* дает такой же удовлетворительный результат, как и для строго дозвуковых течений, и для него справедливы те же замечания.

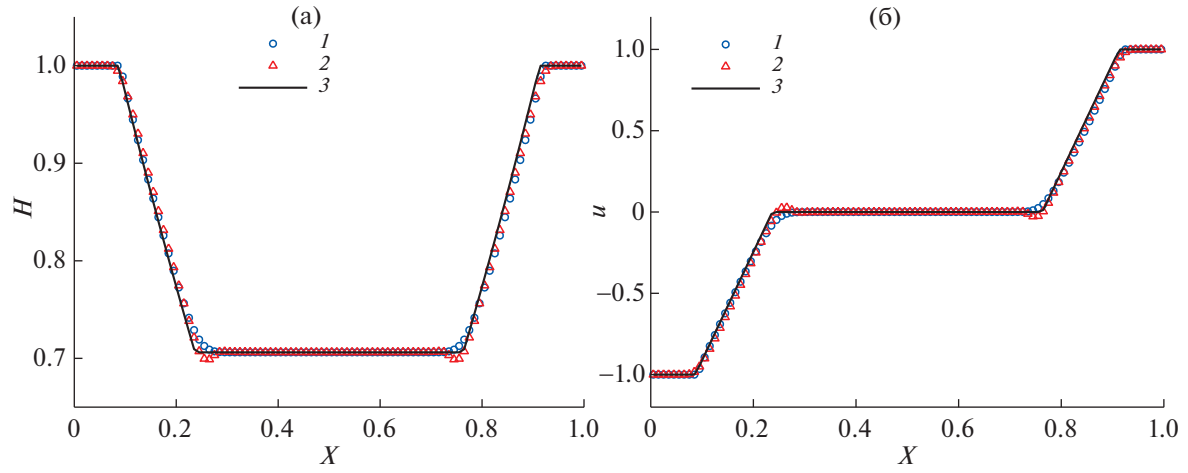
4.3.3 Трансзвуковые течения. Ниже представлены результаты расчетов нескольких задач Римана о трансзвуковом течении, для которых требуется применение алгоритма обработки звуковых точек. Все балансно-характеристические методы нуждаются в подобных алгоритмах при расчете трансзвуковых задач, и для каждой системы уравнений нужен, вообще говоря, отдельный алгоритм. В случае схемы КАБАРЕ для уравнений мелкой воды наилучший алгоритм обработки звуковых точек был представлен в [15]. В случае метода *ICCh-2* наиболее удовлетворительные результаты показал простейший алгоритм, представленный в (15), результаты расчетов для которого и приводятся в данном разделе.

Рассмотрим следующую трансзвуковую задачу Римана (25) на сегменте $x \in [0, 50]$ (так называемый первый тест Торо):

$$H_L = 1.0, \quad H_R = 0.1, \quad u_L = 2.5, \quad u_R = 0.0, \quad x^* = 10.0. \tag{30}$$



Фиг. 8. Профили глубины (а) и скорости (б) для задачи Римана (26) на сегменте [0, 1] на момент времени $t = 0.1$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.



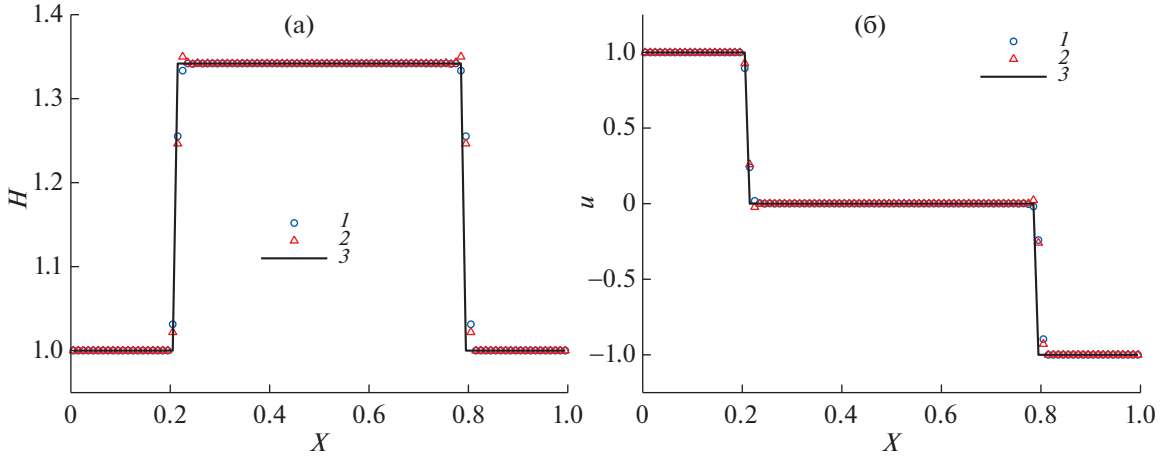
Фиг. 9. Профили глубины (а) и скорости (б) для задачи Римана (27) на сегменте [0, 1] на момент времени $t = 0.1$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.

На фиг. 12 представлен результат расчета задачи (30) на равномерной сетке из $N = 101$ ячеек на момент времени $t = 7.0$, число Куранта $CFL = 0.3$. В областях дозвукового и сверхзвукового течения метод *ICCh-2*, как и прежде, показывает хорошие результаты. Но в окрестности звуковой точки у решения возникает артефакт в виде немонотонностей. Ширина такого артефакта фиксирована (4–6 ячеек), а амплитуда зависит от величины разрывов в задаче Римана.

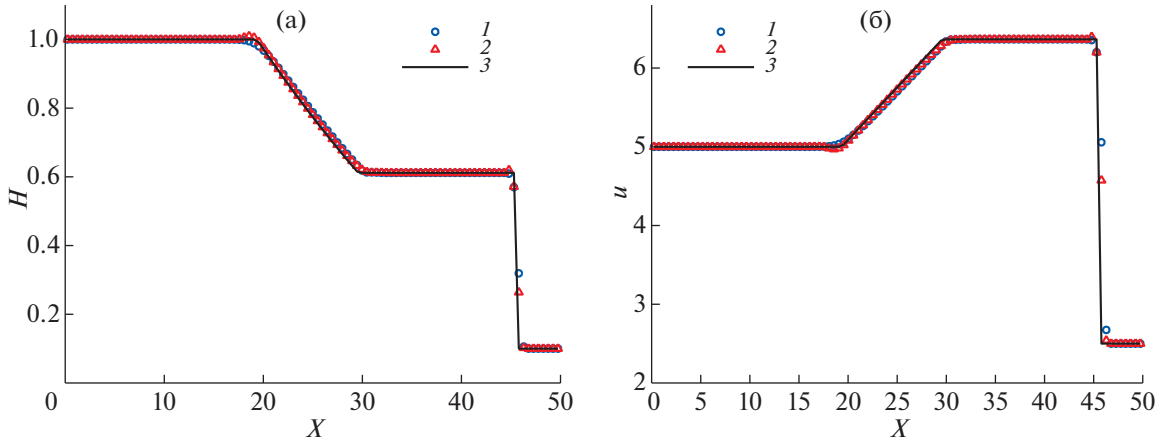
Для более подробной демонстрации вышеупомянутого артефакта рассмотрим другую трансзвуковую задачу Римана на сегменте $x \in [0, 1]$:

$$H_L = 100, \quad H_R = 1, \quad u_L = u_R = 0, \quad x^* = 0.5. \tag{31}$$

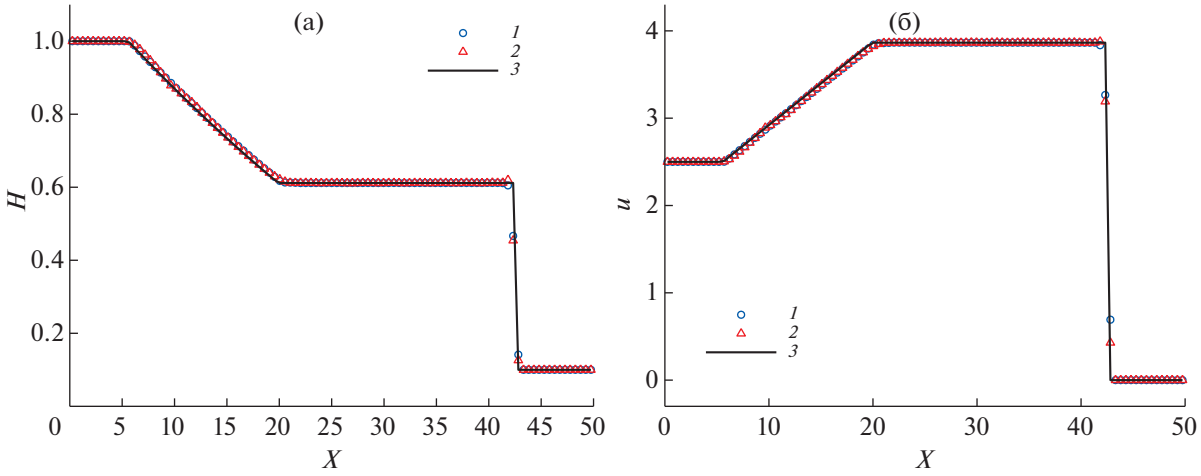
На фиг. 13 представлен результат расчета задачи (31) на равномерной сетке из $N = 101$ ячеек в момент времени $t = 0.012$, число Куранта $CFL = 0.3$. Артефакт метода *ICCh-2* находится в окрестности звуковой точки $x^* = 0.5$ и занимает около 6 расчетных ячеек. По-видимому, наличие немонотонностей вызвано лишь частично аномальной дисперсией метода и неполностью



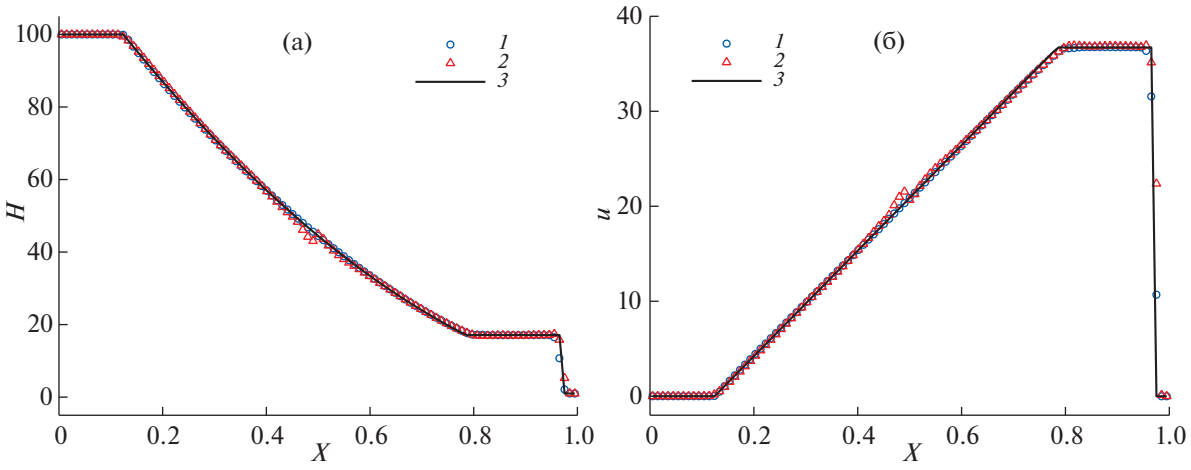
Фиг. 10. Профили глубины (а) и скорости (б) для задачи Римана (28) на сегменте [0, 1] на момент времени $t = 0.1$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.



Фиг. 11. Профили глубины (а) и скорости (б) для задачи Римана (29) на момент времени $t = 5.0$ на $N = 101$ ячейке. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.



Фиг. 12. Профили глубины (а) и скорости (б) для задачи Римана (30) на сегменте [0, 50] на $N = 101$ ячейках в момент времени $t = 7.0$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.



Фиг. 13. Профили глубины (а) и скорости (б) для задачи Римана (31) на сегменте [0, 1] на $N = 101$ ячейках на момент времени $t = 0.012$. *1* – решение, полученное по схеме КАБАРЕ, *2* – решение, полученное методом *ICCh-2*, *3* – аналитическое решение.

работающей процедурой монотонизации (16), что говорит о том, что способ монотонизации схемы требует доработки.

5. ЗАКЛЮЧЕНИЕ

В работе представлен новый явный балансно-характеристический метод решения систем нелинейных дифференциальных уравнений в частных производных гиперболического типа второго порядка аппроксимации. Характеристическая фаза метода основана на параболической реконструкции инвариантов Римана, учитывающей интегральную природу консервативных переменных, заданных в центрах ячеек. Построенная схема обладает частично аномальной дисперсией при малых числах Куранта, что позволяет избавиться от высокочастотных немонотонностей в решении. Метод был протестирован на некоторых задачах для уравнений переноса, Хопфа и мелкой воды. В случае гладких профилей решение, полученное по методу *ICCh-2*, обладает меньшей численной диссипацией и дисперсией по сравнению с методом *ICCh-1* и схемой КАБАРЕ при включенной процедуре монотонизации. В случае разрывных решений метод также позволяет получить качественное решение, обладающее лишь незначительными немонотонностями в областях негладкостей и звуковых точек.

В качестве дальнейшей работы предложенный метод следует обобщить на случай двумерных и трехмерных систем гиперболического типа на треугольных и тетраэдральных расчетных сетках. Основные идеи такого обобщения уже были заложены в работе [2]. Особый интерес может представлять сравнение метода с существующими обобщениями схемы КАБАРЕ на треугольные сетки [9], [10].

СПИСОК ЛИТЕРАТУРЫ

1. *Куликовский А.Г., Погорелов Н.В., Семенов А.Ю.* Математические вопросы численного решения гиперболических систем уравнений. М.: Физматлит, 2001. 607 с.
2. *Головизнин В.М., Четверушкин Б.Н.* Алгоритмы нового поколения в вычислительной гидродинамике // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 8. С. 20–29.
3. *Guerrero Fernandez E., Castro-Diaz M.J., Morales de Luna T.* A second-order well-balanced finite volume scheme for the multilayer shallow water model with variable density // Mathematics. 2020. V. 8. №. 5. P. 848.
4. *Головизнин В.М., Зайцев М.А., Карабасов С.А., Короткин И.А.* Новые алгоритмы вычислительной гидродинамики для многопроцессорных вычислительных комплексов. М.: Изд-во МГУ, 2013. 467 с.
5. *Karabasov S.A., Goloviznin V.M.* Compact accurately boundary-adjusting high-resolution technique for fluid dynamics // J. Comput. Phys. 2009. V. 228. № 19. P. 7426–7451.

6. *Головизнин В.М., Самарский А.А.* Разностная аппроксимация конвективного переноса с пространственным расщеплением временной производной // Матем. моделирование. 1998. Т. 10. № 1. С. 86–100.
7. *Goloviznin V.M., Mayorov P.A., Mayorov P.A.* Hyperbolic decomposition for hydrostatic approximation of free surface flow problems // J. of Physics: Conference Series. 2019. V. 1392. № 012035.
8. *Головизнин В.М., Майоров П.А., Соловьев А.В.* Новый численный алгоритм для уравнений многослойной мелкой воды на основе гиперболической декомпозиции и схемы КАБАРЕ // Морской гидрофизический ж. 2019. Т. 35. № 6. С. 600–620.
9. *Яковлев П.Г., Карабасов С.А., Головизнин В.М.* Прямое моделирование взаимодействия вихревых пар. // Матем. моделирование. 2011. Т. 23. № 11. С. 21–32.
10. *Gorbachev D.Y., Goloviznin V.M.* The Balance-Characteristic Numerical Method on Triangle Grids // J. of Physics: Conference Series. 2019. V. 1392. № 012036.
11. *Головизнин В.М., Карабасов С.А.* Нелинейная коррекция схемы Кабаре // Матем. моделирование. 1998. Т. 10. № 12. С. 107–123.
12. *Van Leer B.* Towards the Ultimate Conservative Difference Scheme. IV. A New Approach to Numerical Convection // J. Comput. Phys. 1977. V. 23. № 3. P. 276–299.
13. *Eymann T.A., Roe P.L.* Active Flux Schemes // 49th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. 2011.
https://doi.org/10.2514/6.2011-382
14. *Головизнин В.М., Исаков В.А.* Применение балансно-характеристической схемы для решения уравнений мелкой воды над неровным дном // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 7. С. 1142–1160.
15. *Afanasiev N.A., Goloviznin V.M.* A locally implicit time-reversible sonic point processing algorithm for one-dimensional shallow-water equations // J. Comput. Phys. 2021. V. 434. № 110220.

EDN: NDNHVT
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1883–1894

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.635

ЭФФЕКТИВНЫЙ МЕТОД РЕШЕНИЯ УРАВНЕНИЯ БОЛЬЦМАНА НА ОДНОРОДНОЙ СЕТКЕ

© 2022 г. А. Д. Беклемишев[1, 2, *], Э. А. Федоренков[1, 2, **]

[1] *630090 Новосибирск, пр-т Акад. Лаврентьева, 11, ИЯФ СО РАН, Россия*

[2] *630090, Новосибирск, ул. Пирогова, 1, НГУ, Россия*

**e-mail: bekl@bk.ru*

***e-mail: e.fedorenkov.nsu@yandex.ru*

Поступила в редакцию 05.02.2022 г.
Переработанный вартант 06.06.2022 г.
Принята к публикации 07.07.2022 г.

Предложен новый численный метод решения уравнения Больцмана на однородной сетке в пространстве скоростей. Асимптотическая сложность метода $O(N^3)$, где N – полное число узлов на трехмерной сетке. Алгоритм эффективен на небольших сетках за счет простоты операций и легкого распараллеливания. Метод сохраняет важнейшие свойства решения: неотрицательность, сохранение полной энергии, импульса и числа частиц. Библ. 30. Фиг. 7.

Ключевые слова: кинетическое уравнение, уравнение Больцмана, модели дискретных скоростей, 0D3V кинетический код.

DOI: 10.31857/S0044466922110059

1. ВВЕДЕНИЕ

Взаимодействие холодного газа с горячей плазмой играет важную роль в физике удержания высокотемпературной плазмы. Например, в открытых магнитных ловушках сложилось качественное представление о том, что нейтральный газ на торцах установки может приводить к ухудшению продольного удержания горячей плазмы. Грубые оценки показывают, что допустимая концентрация нейтрального газа для стационарного удержания не должна превышать 10^{12} частиц в кубическом сантиметре [1]. Это накладывает очень жесткие требования на вакуумную систему открытых ловушек. В экспериментах [2] влияние газа не столь сильно. Это может объясняться перераспределением остаточного газа в объеме установки при его взаимодействии с горячей плазмой и холодными стенками так, что его концентрация в плазме на порядки ниже, чем вблизи стенок. При этом длина свободного пробега в газе также меняется на порядки – от размера системы (в центре) до малых долей радиуса (у стенки). В таком переходном режиме можно ожидать сильного отклонения локальной функции распределения от максвелловской, что требует кинетических расчетов. Для более детального понимания роли нейтрального газа в физике удержания плазмы в открытых магнитных ловушках нами разрабатывается кинетический код. Он позволит учитывать различные элементарные процессы, происходящие при проникновении нейтрального газа в плазму. В статье мы продемонстрируем наш способ учета упругих столкновений.

Упругие столкновения газа описываются уравнением Больцмана. Его решение является серьезной численной задачей из-за нелинейности и многомерной структуры интеграла. Существуют два основных подхода к решению кинетических уравнений: стохастические численные методы, такие как метод прямого моделирования Монте-Карло (DSMC) [3]–[7], и детерминированные методы.

Прямой метод Монте-Карло заключается в моделировании движения пробных частиц, рассеивающихся на фоновом газе. Обычно детерминированные методы имеют более высокий порядок точности, однако не могут конкурировать со стохастическими методами в скорости вычислений. Для нашей задачи больший интерес представляют детерминированные методы, поскольку распределение фонового газа неизвестно. Это серьезно усложняет вычисления методом Монте-Карло из-за необходимости итераций.

Детерминированные методы, в свою очередь, делятся на два основных подхода: разложение функции распределения по гладким функциям и сеточные методы.

В первом подходе используют специальные наборы ортогональных полиномов или фурье-гармоники. В нелинейном случае этот метод получил развитие в работах Бернетта [8] и Града [9]. В качестве набора функций обычно используют полиномы Эрмита или Сонина. Такие методы хорошо подходят, когда распределения слабо отличаются от равновесного. Для более общих случаев используют спектральные методы (разложение по фурье-гармоникам). Быстрое преобразование Фурье позволяет эффективно использовать спектральные методы на практике. Первыми спектральный метод предложили независимо Л. Парески и Б. Пертхам [10] и А. Бобылев с С. Рясанов [11]. Метод получил развитие в большом числе других работ [12]–[15]. В том числе были разработаны быстрые спектральные алгоритмы [16]. Благодаря этому сложность решения уравнения Больцмана уменьшилась с $O(N^{2+\varepsilon})$ ($\varepsilon \sim 1/3$ для трехмерного пространства скоростей) до $O(N\log(N))$, где N – число узлов на сетке скоростей, что сделало спектральный метод сравнимым по сложности с линейными методами Монте-Карло.

Основной недостаток разложения по фурье-гармоникам или специальным полиномам в том, что не обеспечена неотрицательность функции распределения, возникающая из-за необходимости обрезать сумму по базисным функциям. В нашей задаче необходимо добавить различные элементарные процессы (ионизация электронным ударом, перезарядка, возбуждение и т.д.). Наличие отрицательных значений может привести к некорректной обработке при учете многих элементарных процессов вместе.

Во втором подходе предполагается, что пространство скоростей дискретно $\mathbf{v} = \Delta v\mathbf{n}$, где $\mathbf{n} \in \mathbb{Z}^3$. С таким допущением интеграл столкновений Больцмана представим в виде бесконечной суммы

$$St(f,f)(v_i) \simeq \sum_j \sum_{k,l} \Gamma_{ij}^{kl}(f_k f_l - f_i f_j), \tag{1}$$

где $f_i = f(v_i)$, индексы $i, j, k, l \in \mathbb{Z}^3$ – целочисленные векторы, нумерующие значения скоростей v_i, w_j, v_k', w_l' из уравнения (2) на дискретной сетке. Суммирование по j заменяет интегрирование по $\mathbf{w}$, а сумма по k, l заменяет интегрирование по углам рассеяния $\mathbf{n}$. Тензор Γ_{ij}^{kl} обладает симметриями, отражающими физические свойства интеграла столкновений Больцмана, такие как обратимость процесса и неизменность от перестановок частиц до и после столкновения

$$\Gamma_{ij}^{kl} = \Gamma_{kl}^{ij} = \Gamma_{ji}^{kl} = \Gamma_{ij}^{lk}.$$

Тензор Γ_{ij}^{kl} должен сохранять импульс и энергию в каждом элементарном акте столкновений, однако задача их сохранения на дискретной ограниченной сетке является очень нетривиальной.

Метод впервые предложил Аристов в 1985 г., и более детально его можно изучить в [17]. В дальнейшем теоретически были показаны сходимость (1) к интегралу Больцмана [18], и сходимость решения в модели дискретных скоростей к решению уравнения Больцмана [19]. Метод получил развитие в следующих работах [20], [21]. В том числе были разработаны различные способы ускорения вычислений [22], [23]. Следует отметить, что без использования различных ускорительных процедур сложность алгоритма ведет себя как $O(N^{2+\varepsilon})$, где N – полное число узлов сетки скоростей. Это сильно увеличивает время вычислений на трехмерной сетке. А также требуется большой объем памяти для хранения вычисленных заранее коэффициентов Γ_{ij}^{kl} даже с учетом их симметрии. Позднее был разработан новый консервативный метод дискретных скоростей, в котором количество коэффициентов, вычисленных заранее, удалось сократить до N^2 [24]. Из последних работ также следует отметить [25]. В работе предложен алгоритм со сложностью вычисления интеграла столкновений $O(N^{8/3})$.

В этой работе будет продемонстрирован новый сеточный метод аппроксимации интеграла столкновений. Метод быстрый и консервативный, а также не требующий большого количества памяти для хранения коэффициентов.

Статья организована следующим образом. В разд. 2 кратко представим уравнение Больцмана и опишем основные свойства этого уравнения. В разд. 3 и 4 мы расскажем наш новый метод ре-

шения уравнения Больцмана. В разд. 5 мы продемонстрируем работу нашего метода на численных тестах. В разд. 6 обсудим результаты и сделаем выводы.

2. УРАВНЕНИЕ БОЛЬЦМАНА

Уравнение Больцмана (2) описывает эволюцию функции распределения газа по скоростям, вызванную парными столкновениями:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = St(f, f), \tag{2}$$

где $f \equiv f(\mathbf{v}, \mathbf{r}, t)$ – функция распределения, зависящая от скорости частицы $\mathbf{v} \in \mathbb{R}^3$, ее положения в пространстве $\mathbf{r} \in \mathbb{R}^3$ и от времени $t \in \mathbb{R}$; $\mathbf{a}$ – ускорение частицы, возникающее при наличии внешних сил; $St(f, f)$ – интеграл столкновений:

$$St(f, f) = \int\limits_{\mathbb{R}^3} \int\limits_{\mathbb{S}(0,1)} B(u, \mathbf{n})(f(\mathbf{v}')f(\mathbf{w}') - f(\mathbf{v})f(\mathbf{w}))d^2n d^3w, \tag{3}$$

где $\mathbf{n}$ – направление рассеяния, $\mathbf{u}$ – относительная скорость сталкивающихся частиц, $\mathbf{v}$, $\mathbf{w}$ – скорости частиц до столкновения, $\mathbf{v}'$, $\mathbf{w}'$ – скорости частиц после столкновения. $B(u, \mathbf{n})$ – ядро интеграла:

$$B(u, \mathbf{n}) = u \frac{d\sigma}{d\Omega}(u, \mathbf{n}),$$

где $\frac{d\sigma}{d\Omega}(u, \mathbf{n})$ – дифференциальное сечение рассеяния, зависящее от относительной скорости и направления рассеяния. Интегрирование в (3) проводится по всем возможным направлениям рассеяния $\mathbf{n} \in \mathbb{S}(0,1)$ ($\mathbb{S}(0,1)$ – единичная сфера с центром в начале координат), и по всем возможным скоростям второй частицы до столкновения $\mathbf{w} \in \mathbb{R}^3$.

Скорости частиц до и после столкновения связаны законами сохранения импульса и энергии:

$$\mathbf{v} + \mathbf{w} = \mathbf{v}' + \mathbf{w}', \quad v^2 + w^2 = v'^2 + w'^2,$$

и могут быть выражены друг через друга следующим образом:

$$\mathbf{v}', \mathbf{w}' = \frac{1}{2}(\mathbf{v} + \mathbf{w}) \pm \frac{1}{2} u\mathbf{n}.$$

Хорошо известно, что интеграл столкновений (2) удовлетворяет законам сохранения числа частиц, импульса, энергии, а также приводит к не убыванию энтропии (H-теорема Больцмана):

$$\int\limits_{\mathbb{R}^3} St(f, f) \begin{pmatrix} 1 \\ \vec{v} \\ v^2 \end{pmatrix} d^3v = 0, \quad \int\limits_{\mathbb{R}^3} St(f, f) \ln(f) d^3v \leqslant 0. \tag{4}$$

Стационарное и не зависящее от пространственных координат решение уравнения (2) имеет вид максвелловской функции распределения:

$$f_M = \frac{n}{\pi^{3/2} v_T^3} \exp\left(-\frac{|\vec{v} - \vec{V}|^2}{v_T^2}\right), \quad v_T = \sqrt{\frac{2T}{m}}, \tag{5}$$

где n – плотность, T – температура, $\vec{V}$ – гидродинамическая скорость и m – масса молекулы газа.

Условия (4) и (5) являются важными физическими свойствами. Более детально познакомиться с другими физическими и математическими аспектами уравнения Больцмана можно в [26]. В этой статье мы будем рассматривать пространственно однородное уравнение Больцмана без внешних сил:

$$\frac{\partial f}{\partial t} = St(f, f). \tag{6}$$

Учесть пространственную неоднородность и внешние силы можно стандартным методом разделения бесстолкновительного транспорта и локальных столкновений на два последовательно чередующихся шага. Более подробно этот метод освещен в [27].

Уравнение (6) в общем случае не имеет известных аналитических решений. Далее мы представим наш численный метод решения этого уравнения.

3. ОБЩЕЕ ОПИСАНИЕ МЕТОДА

3.1. Локальные упругие столкновения

Пусть в области $\Omega \subset \mathbb{R}^3$ пространства скоростей задана однородная сетка $\mathbb{V}$ с шагом Δv.

$$\mathbb{V} = \left\{ v_{\mathbf{i}} \in \mathbb{R}^3 \,|\, v_{\mathbf{i}} = \Delta v \mathbf{i}, \text{ где } \mathbf{i} \in \mathbb{Z}^3 \right\}.$$

Обозначим через $f_{\mathbf{i}}$ значения функции распределения на сетке $\mathbb{V}$. Будем считать, что функция распределения между узлами сетки представима в виде линейной комбинации значений $f_{\mathbf{i}}$:

$$f(\mathbf{v}) = \sum_{\mathbf{k} \in \mathbb{Z}^3} \alpha_{\mathbf{k}}(\mathbf{v}) f_{\mathbf{k}}, \tag{7}$$

коэффициенты $\alpha_{\mathbf{k}}(\mathbf{v})$ зависят от скорости. Запись (7) охватывает случаи кусочно-постоянной функции распределения и трилинейной интерполяции набора $f_{\mathbf{i}}$. В первом случае реальное значение скорости $\mathbf{v}$ заменяется на ближайший узел сетки скоростей

$$\alpha_{\mathbf{k}}(\mathbf{v}) = \delta(\mathbf{v} - v_{\mathbf{k}}), \quad \mathbf{k} : |\mathbf{v} - v_{\mathbf{k}}| \to \min. \tag{8}$$

Во втором случае значение скорости $\mathbf{v}$ заменяется на восемь ближайших узлов кубической сетки с весами, определяемыми трилинейной интерполяцией набора $f_{\mathbf{i}}$:

$$\alpha_{\mathbf{k}}(\mathbf{v}) = \sum_{i=1}^{8} \frac{|v_x - v_{k_{x_i}}|\,|v_y - v_{k_{y_i}}|\,|v_z - v_{k_{z_i}}|}{\Delta v^3}, \quad k_{j_i} : |v_j - v_{k_{j_i}}| < \Delta v. \tag{9}$$

Подставим функцию распределения (7) в интеграл столкновений из уравнения (2)

$$St(f,f)(v_{\mathbf{i}}) \simeq \iint B(u,\mathbf{n}) \left(\alpha_{\mathbf{k}}(\mathbf{v}')\alpha_{\mathbf{l}}(\mathbf{w}') - \alpha_{\mathbf{k}}(\mathbf{v})\alpha_{\mathbf{l}}(\mathbf{w}) \right) d^2 n d^3 w f_{\mathbf{k}} f_{\mathbf{l}}. \tag{10}$$

Вычислив пятимерный интеграл (10), мы получим значение интеграла столкновений в i-м узле в виде свертки

$$St(f,f)(v_{\mathbf{i}}) \simeq G^{\mathbf{kl}}(v_{\mathbf{i}}) f_{\mathbf{k}} f_{\mathbf{l}} \equiv G_{\mathbf{i}}^{\mathbf{kl}} f_{\mathbf{k}} f_{\mathbf{l}}. \tag{11}$$

Тензор $G_{\mathbf{i}}^{\mathbf{kl}}$, очевидно, обладает симметрией по двум верхним индексам. Однако, даже с учетом этого, на грубой сетке размера $10 \times 10 \times 10$ нужно хранить порядка 5×10^8 элементов. Занимаемый объем памяти коэффициентами $G_{\mathbf{i}}^{\mathbf{kl}}$ меньше, чем коэффициентами $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$, однако, по-прежнему очень большой для хранения их в оперативной памяти. По сути, аппроксимация интеграла столкновений (11) пока отличается от (1) только избавлением от суммирования по индексу $\mathbf{j}$. Коэффициенты $G_{\mathbf{i}}^{\mathbf{kl}}$ можно получить из $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$ следующим образом:

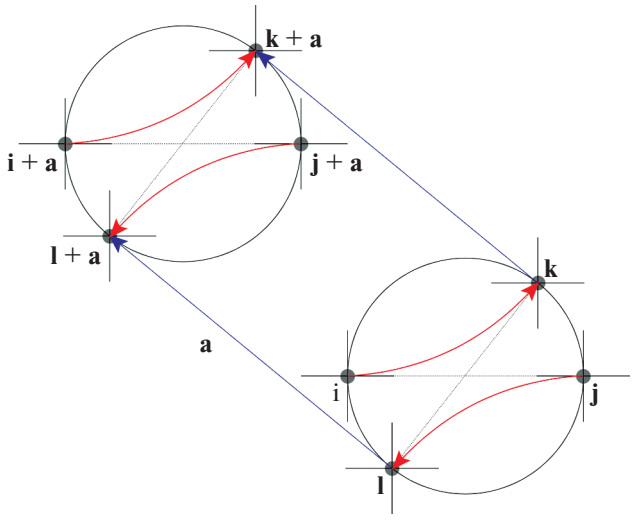
$$G_{\mathbf{i}}^{\mathbf{kl}} = \sum_{\mathbf{j} \in \mathbb{Z}^3} \Gamma_{\mathbf{ij}}^{\mathbf{kl}} - \sum_{\mathbf{n},\mathbf{m} \in \mathbb{Z}^3} \Gamma_{\mathbf{ik}}^{\mathbf{nm}} \delta_{\mathbf{il}}. \tag{12}$$

Для дальнейшего упрощения метода мы воспользуемся еще одной симметрией тензора $G_{\mathbf{i}}^{\mathbf{kl}}$.

3.2. Трансляционная симметрия тензоров $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$ и $G_{\mathbf{i}}^{\mathbf{kl}}$

Если область Ω, в которой задана однородная сетка скоростей, совпадает с $\mathbb{R}^3$, то тензор $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$ из (1) обладает трансляционной симметрией:

$$\Gamma_{\mathbf{i+a,j+a}}^{\mathbf{k+a,l+a}} = \Gamma_{\mathbf{ij}}^{\mathbf{kl}} \quad \forall \mathbf{a} \in \mathbb{Z}^3. \tag{13}$$



Фиг. 1. Иллюстрация трансляционной симметрии коэффициентов $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$ на бесконечной однородной сетке. $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, $\mathbf{l}$ – узлы сетки на сфере столкновений, $\mathbf{a}$ – вектор смещений по однородной сетке.

Это свойство можно понять, рассмотрев сферу рассеяния с диаметрально противоположными точками $\mathbf{i}$ и $\mathbf{j}$, которые в результате упругого столкновения перейдут в диаметрально противоположные точки $\mathbf{k}$ и $\mathbf{l}$ (фиг. 1). Законы сохранения энергии и импульса не позволяют находиться точкам $\mathbf{k}$ и $\mathbf{l}$ вне сферы рассеяния. Сферу рассеяния можно сдвинуть на любой вектор $\Delta v \cdot \mathbf{a}$, где $\mathbf{a} \in \mathbb{Z}^3$. Поскольку сетка однородна и бесконечна, то узлы с координатами $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, $\mathbf{l}$ перейдут в узлы с координатами $\mathbf{i}+\mathbf{a}$, $\mathbf{j}+\mathbf{a}$, $\mathbf{k}+\mathbf{a}$, $\mathbf{l}+\mathbf{a}$, лежащими на смещенной сфере рассеяния. Рассеяние $(\mathbf{i},\mathbf{j}) \to (\mathbf{k},\mathbf{l})$ происходит на тот же угол и с той же относительной скоростью, что и рассеяние $(\mathbf{i}+\mathbf{a},\mathbf{j}+\mathbf{a}) \to (\mathbf{k}+\mathbf{a},\mathbf{l}+\mathbf{a})$. Следовательно, имеет место равенство (13). Учитывая связи коэффициентов $G_{\mathbf{i}}^{\mathbf{kl}}$ и $\Gamma_{\mathbf{ij}}^{\mathbf{kl}}$ (12), тензор $G_{\mathbf{i}}^{\mathbf{kl}}$ тоже обладает трансляционной симметрией на бесконечной однородной сетке:

$$G_{\mathbf{i}+\mathbf{a}}^{\mathbf{k}+\mathbf{a},\mathbf{l}+\mathbf{a}} = G_{\mathbf{i}}^{\mathbf{kl}} \quad \forall \mathbf{a} \in \mathbb{Z}^3.$$

Это свойство позволяет вычислять интеграл столкновений с меньшим количеством коэффициентов. Достаточно вычислить только матрицу $G_0^{\mathbf{kl}}$:

$$St(f,f)(v_{\mathbf{i}}) \simeq \sum_{\mathbf{k},\mathbf{l}} G_0^{\mathbf{k}-\mathbf{i},\mathbf{l}-\mathbf{i}} f_{\mathbf{k}} f_{\mathbf{l}} \equiv \sum_{\mathbf{k},\mathbf{l}} G_0^{\mathbf{kl}} f_{\mathbf{k}+\mathbf{i}} f_{\mathbf{l}+\mathbf{i}}. \tag{14}$$

Трансляционная симметрия интеграла в пространстве скоростей прямо связана с (дискретизированным) законом сохранения импульса (теорема Нетер). В любой ограниченной области законы сохранения корректно выполняться не могут. Их выполнения можно добиться внесением специальных искажений интеграла столкновений, зависящих от реализации сетки.

3.3. Корректировка законов сохранения

Сконструируем добавку к интегралу столкновений, зависящую от пяти параметров, которые определяются из законов сохранения числа частиц, импульса и энергии. Помимо этого, добавка не должна нарушать неотрицательность функции распределения. Этого можно добиться, сделав добавку к интегралу столкновений пропорциональной функции распределения. Это гарантирует малое изменение функции распределения в той части фазового пространства, где она сама мала. Вышесказанное можно реализовать, в частности, при внесении в интеграл столкновений добавки вида

$$\hat{St}(\mathbf{v}) = St(\mathbf{v}) + (a + \mathbf{b}\cdot\mathbf{v} + cv^2) f(\mathbf{v}). \tag{15}$$

Коэффициенты a, $\mathbf{b}$, c определяются из следующих уравнений:

$$\int \hat{St}(\mathbf{v})d^3v = 0; \quad \int \mathbf{v}\hat{St}(\mathbf{v})d^3v = 0; \quad \int v^2\hat{St}(\mathbf{v})d^3v = 0.$$

На сетке при известном $St(\mathbf{v})$ эти уравнения являются линейной системой на искомые коэффициенты a, $\mathbf{b}$, c, что позволяет легко и быстро их находить.

На ограниченной сетке отличие суммы (15) от истинного значения состоит из двух частей: ошибка дискретизации, зависящая от размера ячейки сетки, и ошибка, связанная с положением границы счетной области, принимая во внимание, что значения функции распределения вне этой области не учитываются.

3.4. Алгоритм решения столкновительного шага

Учитывая все выше перечисленное, наш алгоритм решения уравнения (6) состоит из следующих шагов.

Шаг 1. Выделяем границы области счета и задаем в ней однородную сетку скоростей. Используя данные о сечении и заданную сетку, вычисляем матрицу $G_0^{\mathbf{kl}}$. Как именно это сделать, будет показано в разд. 4.

Шаг 2. В цикле по времени вычисляем по функции распределения $f(t_n) \equiv f^n$ в момент времени t_n интеграл столкновений $St(f^n, f^n)(v_i)$ согласно формуле (14).

Шаг 3. Корректируем интеграл столкновений $St(f^n, f^n)(v_i)$ согласно (15).

Шаг 4. Получаем функцию распределения $f(t_n + \Delta t)$ в момент времени t_{n+1}, как численное решение уравнения (6) по схеме Рунге–Кутты второго порядка точности.

4. АЛГОРИТМЫ ВЫЧИСЛЕНИЯ МАТРИЦЫ $G_0^{\mathbf{kl}}$

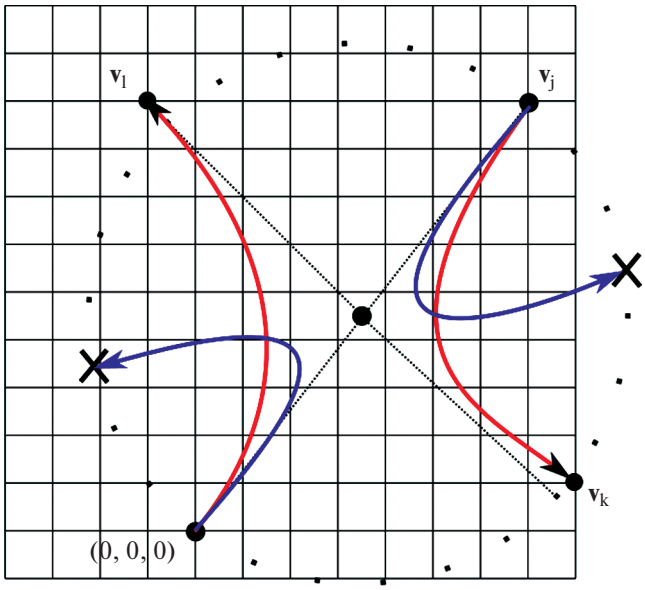
Пусть имеется трехмерная однородная сетка скоростей размера $N \times N \times N$ в кубе $[-V, V]^3$. Далее мы продемонстрируем наш алгоритм вычисления матрицы $G_0^{\mathbf{kl}}$ в такой области. От способа вычисления матрицы $G_0^{\mathbf{kl}}$ зависит порядок точности вычисления интеграла столкновений. Здесь мы приведем метод первого порядка точности.

4.1. Алгоритм первого порядка точности вычисления интеграла столкновений

Воспользуемся представлением функции распределения (7) с коэффициентами разложения (8). Одна из частиц до столкновения будет всегда находиться в центре сетки со скоростью $\mathbf{v}_i = (0,0,0)$. Вторая частица до столкновения пробегает все возможные N^3 значений скорости $\mathbf{v}_j$. Фиксированные значения $\mathbf{v}_i$ и $\mathbf{v}_j$ задают сферу столкновений. Согласно законам сохранения энергии и импульса, скорости частиц после столкновения, $\mathbf{v}_k$ и $\mathbf{v}_l$, должны лежать на этой сфере.

Для численного интегрирования по сфере рассеяния необходимо создать равномерный набор точек на сфере. Для этого мы использовали spiral-based sampling метод [28], который создает квазиравномерный набор точек на сфере. В этом алгоритме по заданному количеству точек M подбираются параметры спирали, которая лежит на сфере, соединяя два ее полюса. После чего набор из M точек укладывается на спираль. С ростом M заполнение сферы становится более равномерным. В качестве полюсов мы выбираем узлы $\mathbf{v}_i$ и $\mathbf{v}_j$. Набор точек на сфере задает набор направлений рассеяния $\mathbf{n}_{ij}$. Фиксируя направление, мы определяем пару скоростей частиц после столкновения $(\mathbf{v}_k, \mathbf{v}_l)$ на сфере рассеяния. Если ни одна из скоростей не вышла за пределы сетки, то мы, согласно модели (8), находим ближайшие к этим скоростям узлы сетки (фиг. 2). Обозначим индексы этих узлов через $\mathbf{k}$ и $\mathbf{l}$, и добавим в соответствующий элемент матрицы $G_0^{\mathbf{kl}}$ значение

$$G_0^{\mathbf{kl}} = \sum_j B(|\mathbf{v}_j|, \mathbf{n}_{ij})(\Delta v)^3 \frac{4\pi}{M},$$



Фиг. 2. Иллюстрация столкновений учитывающихся при вычислении матрицы $G_0^{\mathbf{kl}}$, и процессов, приводящих к вылету частиц за область вычислений, которые не учитываются.

где B – ядро интеграла Больцмана, Δv – шаг сетки, $(\Delta v)^3$ – элемент объема сетки скоростей, M – число точек на сфере рассеяния, $4\pi/M$ – элемент телесного угла. Если хоть одна из скоростей $\mathbf{v}_k$, $\mathbf{v}_l$ вышла за пределы сетки, то такой процесс учтен не будет.

В итоге алгоритм первого порядка точности состоит из следующих шагов.

Шаг 1. Для каждого узла сетки скоростей v_j создается набор возможных направлений рассеяния $\mathbf{n}_{ij}$.

Шаг 2. Для каждого направления из множества $\mathbf{n}_{ij}$ определяется пара скоростей после рассеяния $(\mathbf{v}_k, \mathbf{v}_l)$.

Шаг 3. Проверка попадания скоростей $(\mathbf{v}_k, \mathbf{v}_l)$ в область расчетов. Если хоть одна из скоростей вышла за пределы сетки, такой процесс не учитывается.

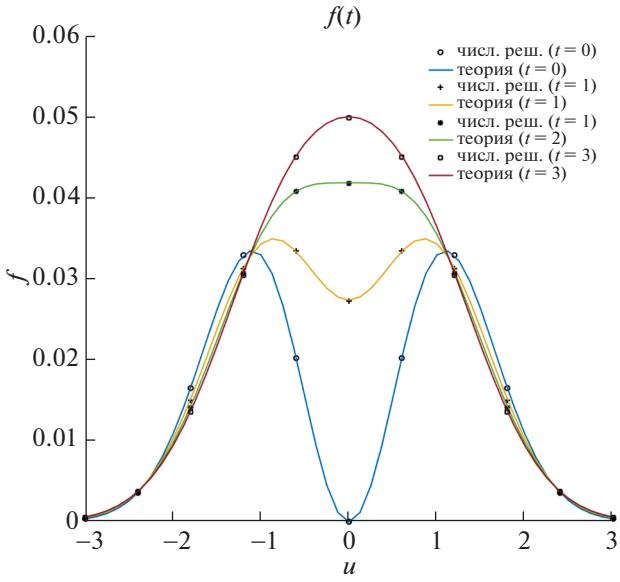
Шаг 4. Если скорости $(\mathbf{v}_k, \mathbf{v}_l)$ лежат в области счета, то находим ближайшие к ним узлы сетки. Обозначим их через $\mathbf{k}$ и $\mathbf{l}$.

Шаг 5. В элемент матрицы $G_0^{\mathbf{kl}}$ добавляем значение $B(|\mathbf{v}_j|, \mathbf{n}_{ij})(\Delta v)^3 4\pi/M$.

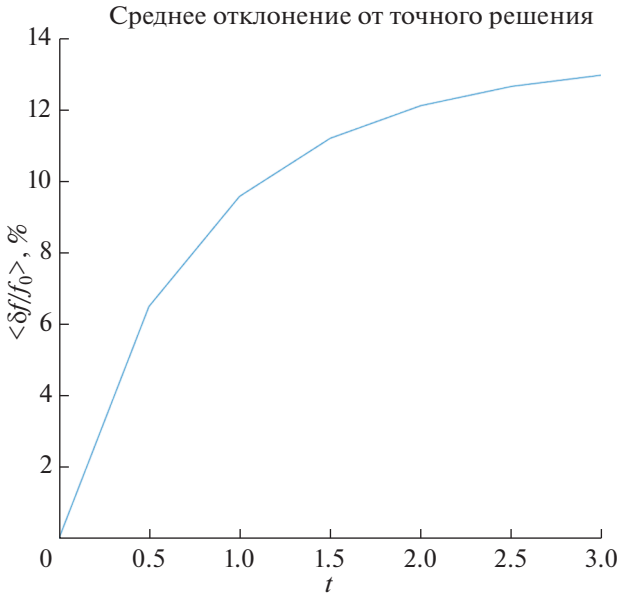
Таким образом, сложность алгоритма вычисления матрицы $G_0^{\mathbf{kl}}$ можно оценить как $O(MN^3)$. В вычислении интеграла столкновений есть два источника ошибок: ошибка, связанная с ограниченностью области счета, ошибка метода интегрирования в области счета. Используя представление функции распределения (7) с коэффициентами разложения (9), в принципе можно добиться повышения точности вычисления интеграла столкновений. При условии, что ошибки, связанные с границей счетной области, не являются доминирующими.

5. ЧИСЛЕННЫЕ ТЕСТЫ

Здесь мы приведем два теста кинетического кода, основанного на описанном выше алгоритме, и продемонстрируем результаты решения уравнения Больцмана. Вычисления выполнены на процессоре Intel Core i5-11600K (12 МБ кэш-память, до 4.9 ГГц). Вычисление матрицы G_0^{kl} выполняется примерно за 700 мс. Вычисление интеграла столкновений на одном шаге по времени занимает примерно 200 мс. Оба теста выполнены на сетке размера 11 × 11 × 11. Время выполнения составляет 63.5 и 99.8 с, соответственно.



Фиг. 3. Точное и численное решение уравнения Больцмана в зависимости от времени. Время отсчитывается от $t = 6\ln(2.5)$.

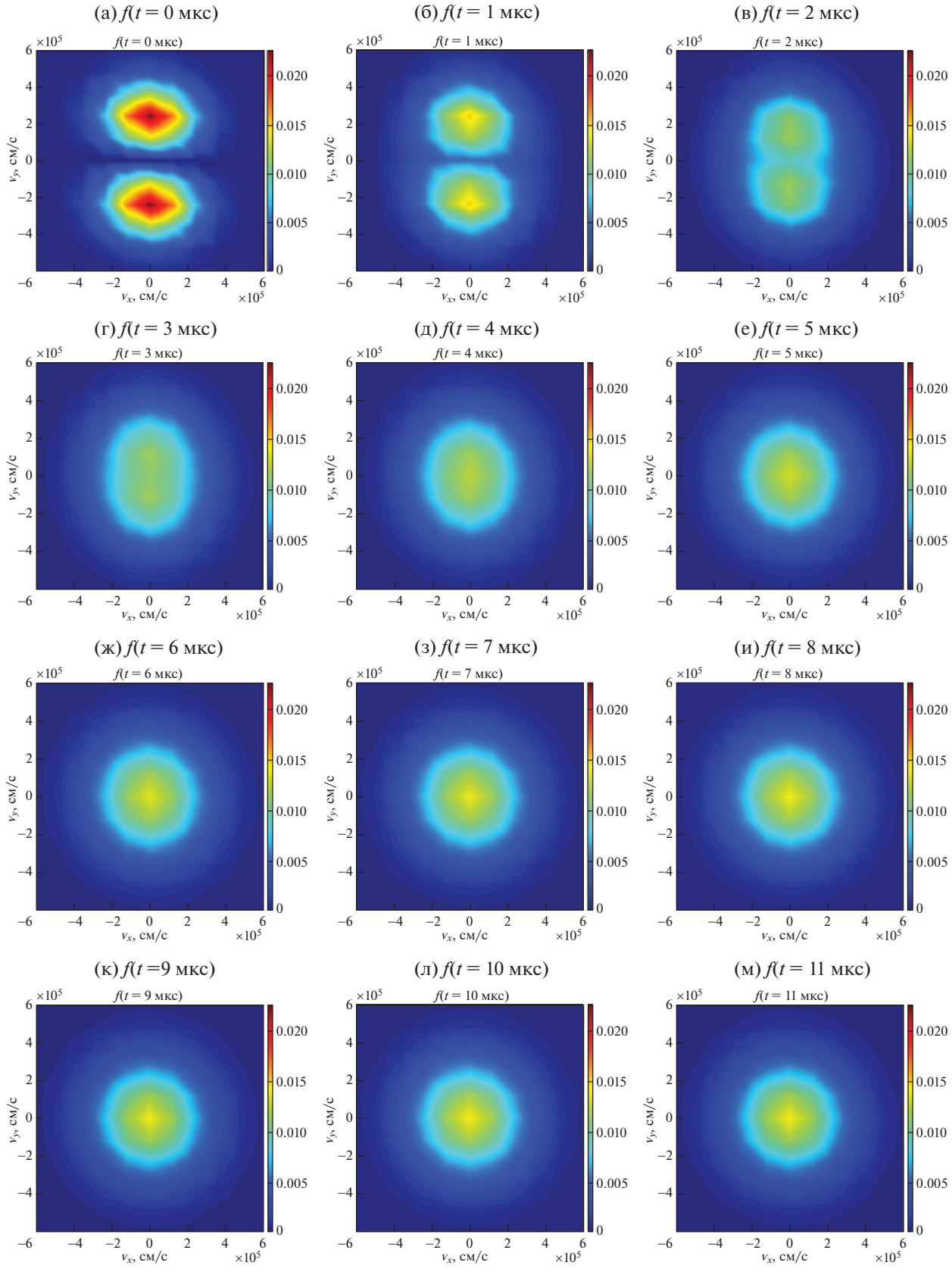


Фиг. 4. Среднее по сетке отклонение численного решения от точного в процентах. Время отсчитывается от $t = 6\ln(2.5)$.

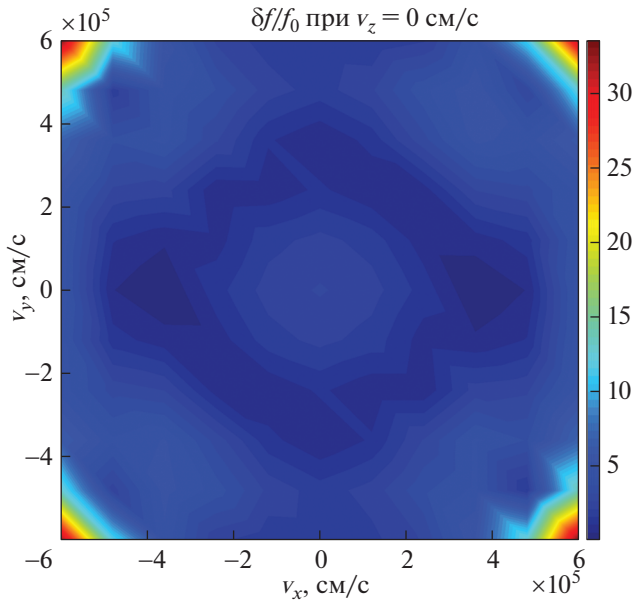
5.1. Тест 1: сравнение численного решения с точным решением уравнения Больцмана

Уравнение Больцмана имеет единственное известное аналитическое решение, найденное Бобылевым [29] и затем независимо Круком и Ву [30]. Решение получено для дифференциального сечения, имеющего специальный вид:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha}{u}, \tag{16}$$



Фиг. 5. Релаксация газа упругих шаров к равновесному распределению.



Фиг. 6. Относительное отклонение численного решения от равновесной функции распределения в конце расчетов.

где α – постоянная величина, u – относительная скорость сталкивающихся частиц. Решение уравнения Больцмана с дифференциальным сечением (16) в безразмерных переменных имеет следующий вид:

$$f_{BKW}(t,v) = \frac{1}{2K(2\pi K)^{3/2}}\left[5K - 3 + \frac{1-K}{K}v^2\right]\exp\left(-\frac{v^2}{2K}\right), \tag{17}$$

где

$$K = 1 - \exp\left(-\frac{t}{6}\right),$$

а время и скорость обезразмерены следующим образом:

$$t \to 4\pi\alpha n t, \quad v \to v\sqrt{\frac{m}{T}}.$$

В вычислениях мы приняли $n = 1$, $\sqrt{m/T} = 1$, $\alpha = 1/4\pi$, а в качестве начального условия:

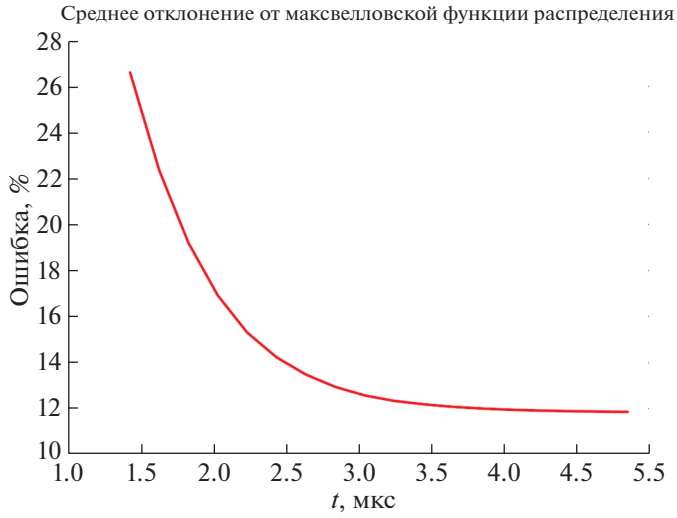
$$f(0,v) = f_{BKW}(t = 6\ln(2.5), v) = \frac{25\sqrt{5}}{9(6\pi)^{3/2}}v^2\exp\left(-v^2\frac{5}{6}\right).$$

Результаты теста в виде сравнения численного и точного решения (17) показаны на фиг. 3. При этом средняя относительная ошибка не превышает 14%, и локализована на краях сетки (фиг. 4).

5.2. Тест 2: релаксация к равновесию газа упругих шаров

В этом тесте пронаблюдаем релаксацию к максвелловскому распределению газа упругих шаров, $d\sigma/d\Omega = \text{const}$. В качестве начального условия было выбрано распределение

$$f_0 = \frac{n_0}{\pi^{3/2}v_{T_0}^3}\left(\frac{v_y}{v_{T_0}}\right)^2\exp\left(-\frac{v^2}{v_{T_0}^2}\right).$$



Фиг. 7. Эволюция среднего по сетке отклонения от максвелловского распределения.

С учетом законов сохранения это распределение релаксирует к максвелловскому с плотностью и температурой:

$$n = \frac{n_0}{2}, \quad T = \frac{5}{3}T_0$$

за характерное время $\tau \sim 1/(\sigma n v_T)$. Моделирование проведено со следующими параметрами:

$$n_0 = 4\times10^{15}\ \text{см}^{-3}, \quad T_0 = 0.026\ \text{эВ},$$
$$\sigma \simeq 1.3\times10^{-15}\ \text{см}^2, \quad \tau \sim 1.4\ \text{мкс}.$$

Результат численного решения представлен на фиг. 5. Из него видно, что для данного начального распределения равновесие устанавливается за время порядка 4τ. Распределение остается стационарным при дальнейшем вычислении. Относительное отклонение от максвелловской функции распределения представлено на фиг. 6. Из графика следует, что самая большая ошибка в углах области счета и она составляет порядка 30%. Далеко от границ области расчетов ошибка порядка 5–10%. Сходимость к равновесному распределению можно пронаблюдать на фиг. 7. Здесь показано среднее по всей сетке отклонение от максвелловского распределения в зависимости от времени.

6. ЗАКЛЮЧЕНИЕ

Разработан довольно быстрый и точный метод интегрирования уравнения Больцмана на грубой равномерной сетке, работающий даже на персональном компьютере. Эффективность метода достаточна для решения одномерной задачи для газа в расширителе. Алгоритм допускает обобщение на неупругие процессы с разными сортами частиц и легко распараллеливается, что позволяет надеяться на успешное совершенствование модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Soldatkina E.I. et al.* Axial plasma confinement in gas dynamic trap // Plasma and Fusion Research. 2019. Т. 14. С. 2402006–2402006.
2. *Soldatkina E.I. et al.* Measurements of axial energy loss from magnetic mirror trap // Nuclear Fusion. 2020. Т. 60. № 8. С. 086009.
3. *Ivanov M.S., Rogasinsky S.V.* Analysis of numerical techniques of the direct simulation Monte Carlo method in the rarefied gas dynamics. 1988.
4. *Babovsky H., Neunzert H.* On a simulation scheme for the Boltzmann equation // Math. Methods in the Applied Sciences. 1986. Т. 8. № 1. С. 223–233.

5. *Bird G.A.* Direct simulation and the Boltzmann equation // The Physics of Fluids. 1970. Т. 13. № 11. С. 2676–2681.

6. *Nanbu K.* Direct simulation scheme derived from the Boltzmann equation. I. Monocomponent gases // J. of the Physical Society of Japan. 1980. Т. 49. № 5. С. 2042–2049.

7. *Bird G.A.* Molecular gas dynamics and the direct simulation of gas flows. Oxford: Clarendon press, 1994. Т. 5.

8. *Burnett D.* The distribution of velocities in a slightly nonвЂЊuniform gas // Proc. of the London Math. Society. 1935. V. 2. № 1. С. 385–430.

9. *Grad H.* On the kinetic theory of rarefied gases // Commun. on Pure and Applied Math. 1949. Т. 2. № 4. С. 331–407.

10. *Pareschi L., Perthame B.* A Fourier spectral method for homogeneous Boltzmann equations // Transport Theory and Statistical Physics. 1996. Т. 25. № 3–5. С. 369–382.

11. *Bobylev A., Rjasanow S.* Difference scheme for the Boltzmann equation based on the Fast Fourier Transform // European journal of mechanics. B, Fluids. 1997. Т. 16. № 2. С. 293–306.

12. *Pareschi L., Russo G.* Numerical solution of the Boltzmann equation I: Spectrally accurate approximation of the collision operator // SIAM Journal on Numerical Analysis. 2000. Т. 37. № 4. С. 1217–1245.

13. *Gamba I.M., Haack J.R.* A conservative spectral method for the Boltzmann equation with anisotropic scattering and the grazing collisions limit // J. of Computational Physics. 2014. Т. 270. С. 40–57.

14. *Wu L. et al.* Deterministic numerical solutions of the Boltzmann equation using the fast spectral method // J. of Comput. Physics. 2013. Т. 250. С. 27–52.

15. *Wu L. et al.* A fast spectral method for the Boltzmann equation for monatomic gas mixtures // J. of Comput. Physics. 2015. Т. 298. С. 602–621.

16. *Filbet F., Mouhot C., Pareschi L.* Solving the Boltzmann equation in N log2 N // SIAM J. on Scientific Computing. 2006. Т. 28. № 3. С. 1029–1053.

17. *Aristov V.V.* Solving the Boltzmann equation for discrete velocities // Akademiia Nauk SSSR Doklady. 1985. Т. 283. С. 831–834.

18. *Schneider J.* Une méthode déterministe pour la résolution de l'équation de Boltzmann : Doctoral dissertation, Paris 6, 1993.

19. *Palczewski A., Schneider J.* Existence, stability, and convergence of solutions of discrete velocity models to the Boltzmann equation // J. of Statistical Physics. 1998. V. 91. № 1–2. С. 307–326.

20. *Vasiljevitch Bobylev A., Palczewski A., Schneider J.* On approximation of the Boltzmann equation by discrete velocity models // Comptes rendus de l'Académie des sciences. Série 1, Mathématique. 1995. V. 320. № 5. С. 639–644.

21. *Palczewski A., Schneider J., Bobylev A.V.* A consistency result for a discrete-velocity model of the Boltzmann equation // SIAM Journal on Numerical Analysis. 1997. Т. 34. № 5. С. 1865–1883.

22. *Buet C.* A discrete-velocity scheme for the Boltzmann operator of rarefied gas dynamics // Transport Theory and Statistical Physics. 1996. Т. 25. № 1. С. 33–60.

23. *Płatkowski T., Waluś W.* An acceleration procedure for discrete velocity approximation of the Boltzmann collision operator // Computers and Math. with Applicat. 2000. V. 39. № 5–6. С. 151–163.

24. *Panferov V.A., Heintz A.G.* A new consistent discrete-velocity model for the Boltzmann equation // Math. Methods in the Applied Sciences. 2002. Т. 25. № 7. С. 571–593.

25. *Alekseenko A., Josyula E.* Deterministic solution of the spatially homogeneous Boltzmann equation using discontinuous Galerkin discretizations in the velocity space // J. of Comput. Physics. 2014. Т. 272. С. 170–188.

26. *Cercignani C., Illner R., Pulvirenti M.* The mathematical theory of dilute gases. London: Springer Science and Business Media, 2013. Т. 106.

27. *Narayan A., Klöckner A.* Deterministic Numerical Schemes for the Boltzmann Equation // arXiv. 2009. С. arXiv: 0911.3589.

28. *Hardin D.P., Michaels T.J., Saff E.B.* A Comparison of Popular Point Configurations on $\mathbb{S}^2$ // arXiv preprint arXiv:1607.04590. 2016.

29. *Bobylev A.V.* Exact solutions of the Boltzmann equation // Akademiia Nauk SSSR Doklady. 1975. V. 225. С. 1296–1299.

30. *Krook M., Wu T.T.* Exact solutions of the Boltzmann equation // The Physics of Fluids. 1977. Т. 20. № 10. С. 1589–1595.

EDN: JMPEIF
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1895–1911

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 531.36

НЕСТАЦИОНАРНЫЙ ИЗГИБ ОРТОТРОПНОЙ КОНСОЛЬНО-ЗАКРЕПЛЕННОЙ БАЛКИ ТИМОШЕНКО С УЧЕТОМ РЕЛАКСАЦИИ ДИФФУЗИОННЫХ ПОТОКОВ

© 2022 г. А. В. Земсков[1, 2, *], Д. В. Тарлаковский[2, 1, **]

[1] *125993 Москва, Волоколамское ш. 4, МАИ, Россия*

[2] *119192 Москва, Мичуринский пр-т, 1, НИИ механ. МГУ, Россия*

**e-mail: azemskov1975@mail.ru*

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

Поступила в редакцию 17.03.2022 г.
Переработанный вариант 25.06.2022 г.
Принята к публикации 07.07.2022 г.

Рассматривается нестационарная задача об изгибе консольно-закрепленной упругодиффузионной ортотропной балки Тимошенко под действием нагрузки, приложенной к свободному концу балки. Модель учитывает конечную скорость распространения диффузионных возмущений вследствие релаксации диффузионных потоков. Физико-механические процессы описываются связанной системой уравнений изгиба балки Тимошенко с учетом диффузии. Решение задачи ищется с помощью метода эквивалентных граничных условий. Для этого рассматривается вспомогательная задача, решение которой получается с помощью интегрального преобразования Лапласа по времени и разложения в тригонометрические ряды Фурье. Далее строятся соотношения, связывающие правые части граничных условий исходной и вспомогательной задачи. Эти соотношения представляют собой систему интегральных уравнений Вольтерра I рода. Решение этой системы осуществляется численно с помощью квадратурных формул. На примере трехкомпонентного материала выполнено численное исследование взаимодействия нестационарных механического и диффузионного полей в ортотропной балке. В заключение приведены основные выводы о влиянии связанности полей на напряженно-деформированное состояние и массоперенос в стержне. Библ. 31. Фиг. 8.

Ключевые слова: нестационарная механодиффузия, балка Тимошенко, изгиб консоли, нестационарные задачи, преобразование Лапласа, метод эквивалентных граничных условий.

DOI: 10.31857/S004446692211014X

ВВЕДЕНИЕ

В работе исследуются нестационарные упругодиффузионные колебания балки Тимошенко. Эта модель является уточнением классической модели балки Бернулли–Эйлера, за счет учета деформаций сдвига и влияния инерционных сил при повороте нормали относительно срединной поверхности.

Классическая модель балки Бернулли–Эйлера является наиболее простой из всех балочных моделей и в ряде случаев обеспечивает удовлетворительную точность решения инженерных задач, связанных с расчетом конструкций на прочность. Однако учет деформаций сдвига, которых нет в модели Бернулли–Эйлера, может оказаться существенным, для расчетов стержней, изготовленных из анизотропного материала, у которых модуль сдвига много меньше модуля Юнга. Так же важное значение имеет учет деформаций сдвига в задачах устойчивости трехслойных стержней и пластин, где два несущих слоя выполнены из тонкого высокопрочного жесткого материала, между ними легкий и менее прочный заполнитель.

Балки, пластины и оболочки являются основой любой конструкции, поэтому их моделям посвящено очень большое количество научных работ. Среди наиболее современных работ можно отметить [1]–[6], где изложены общие принципы построения уравнений теории пластин и оболочек, основанные на вариационных принципах и асимптотических методах моделирования.

Здесь же изложены методы решения задач об изгибе пластин и оболочек при различных способах закрепления и различных видов механического нагружения.

Следует отметить, что расчет тонкостенных конструкций существенно усложняется в том случае, когда приходится учитывать взаимодействие полей различной физической природы: механических, диффузионных, температурных и пр. Поэтому чаще всего ограничиваются исследованием взаимодействия стационарных физических полей в сплошных средах и технических системах.

Анализу механодиффузионных процессов и оценке их влияния на несущую способность стержней, пластин и оболочек посвящены работы [7]–[14]. В работах [7], [8] исследуется влияние диффузионных процессов на несущую способность пологой трансверсально-изотропной оболочки. Контактное нестационарное взаимодействие стержня с упругим полупространством рассматривается в работах [9], [10]. Публикации [11]–[13] посвящены исследованию механодиффузионных процессов в пластинах. Расчет сферических оболочек с учетом диффузии рассмотрен в [14].

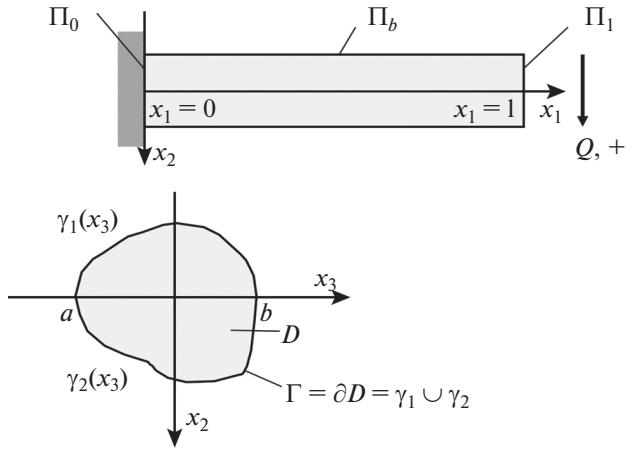
Моделирование нестационарных процессов как в тонкостенных элементах конструкций, так и в сплошных средах представляет собой достаточно сложную математическую задачу, связанную с обращением преобразования Лапласа, которое применяется при аналитических методах решения нестационарных начально-краевых задач. В большинстве случаев эта проблема решается с помощью специальных квадратурных формул, позволяющих приближенно вычислить интеграл Меллина. В качестве наиболее известных алгоритмов используются: метод Дурбина [15], [16], “Gaver-Stehfast algorithm” [17], “Zakian’s algorithm” [18], методы основанные на использовании ортогональных полиномов Лежандра [14] и пр. Подробное описание методов обращения преобразования Лапласа можно найти также в монографии [19]. Данные методы хорошо зарекомендовали себя при вычислении оригиналов в определенном классе функций. Однако для нахождения функций Грина эти алгоритмы не пригодны, так как функции Грина принадлежат к классу обобщенных функций, что затрудняет использование методов численного интегрирования.

В настоящей работе рассматривается задача о нестационарных колебаниях консольно-закрепленной балки с учетом диффузии. Уравнения упругодиффузионных колебаний балки получены из классической модели механодиффузии для сплошных сред с помощью вариационного принципа Даламбера [20], [21]. Основная сложность в решении данной задачи заключается в том, что краевые условия, соответствующие консольному закреплению, не позволяют воспользоваться методом Фурье разделения переменных. Это существенно осложняет обращение преобразования Лапласа, о котором говорилось ранее. Например, для свободно опертых балок [20], [21] применение метода Фурье и преобразования Лапласа позволяет свести проблему обращения преобразования Лапласа к проблеме нахождения оригиналов от рациональной функции. Такая задача решается аналитически с помощью вычетов и таблиц операционного исчисления.

Таким образом, задача об изгибе консольно-закрепленной балки с учетом диффузии намного сложнее аналогичной задачи для свободно опертой балки. Для ее решения предлагается использовать метод эквивалентных граничных условий, который заключается в том, что вместо исходной задачи рассматривается вспомогательная задача того же вида, но с граничными условиями, допускающими представление решений в виде рядов Фурье. Далее строятся соотношения, связывающие правые части граничных условий обеих задач. Эти соотношения записываются в виде системы интегральных уравнений Вольтерра I рода. Затем полученная система уравнений решается численно с помощью квадратурных формул. Предложенный метод был апробирован при решении аналогичной задачи для консольно-закрепленной балки Бернулли–Эйлера [22].

1. ПОСТАНОВКА ЗАДАЧИ

В работе рассматривается задача о нестационарном упругодиффузионном изгибе консольно-закрепленной ортотропной балки Тимошенко (фиг. 1).



Фиг. 1. Иллюстрация к постановке задачи.

Математическая постановка представляет собой замкнутую систему уравнений поперечных колебаний балки с учетом диффузии, которая получена из общей модели упругой диффузии для сплошных сред с помощью вариационного принципа Даламбера [20], [21]:

$$\ddot{v} - C_{66}k^2\left(v'' - \chi'\right) - \frac{Q}{F} = 0, \quad a = \frac{F}{J_3}, \quad q = \overline{1,N},$$
$$\ddot{\chi} - \chi'' - aC_{66}k^2\left(v' - \chi\right) - \sum_{j=1}^{N}\alpha_1^{(j)}H_j' - \frac{M}{J_3} = 0, \tag{1}$$
$$\sum_{k=0}^{K}\frac{\tau_q^k}{k!}\frac{\partial^k \dot{H}_q}{\partial\tau^k} - D_1^{(q)}H_q'' - \Lambda_{11}^{(q)}\chi''' - \frac{z^{(q)}}{J_3} = 0, \quad H_{N+1} = -\sum_{j=1}^{N}H_j.$$

Здесь точки обозначают производную по времени, штрихи – производную по координате x_1. Все величины в (1) являются безразмерными. Для них приняты следующие обозначения:

$$x_i = \frac{x_i^*}{l}, \quad v = \frac{v^*}{l}, \quad \tau = \frac{Ct}{l}, \quad C^2 = \frac{C_{11}^*}{\rho}, \quad C_{ij} = \frac{C_{ij}^*}{C_{11}^*}, \quad M = \frac{lM^*}{C_{11}^*}, \quad Q = \frac{lQ^*}{C_{11}^*},$$
$$D_\alpha^{(q)} = \frac{D_{\alpha\alpha}^{(q)}}{Cl}, \quad \alpha_\beta^{(q)} = \frac{\alpha_{\beta\beta}^{(q)}}{C_{11}^*}, \quad \Lambda_{\alpha\beta}^{(q)} = \frac{m^{(q)}D_{\alpha\alpha}^{(q)}\alpha_{\beta\beta}^{(q)}n_0^{(q)}}{\rho RT_0Cl}, \quad \tau_q = \frac{C\tau^{(q)}}{l},$$

где t – время; x_i^* – прямоугольные декартовы координаты; v^* – поперечный прогиб балки; χ – угол поворота сечения; l – длина балки; k – коэффициент сдвига Тимошенко, который зависит от формы балки (для прямоугольного сечения балки $k^2 = 5/6$); H_q – линейная плотность приращения концентрации q-й компоненты вещества в составе многокомпонентной среды; $n_0^{(q)}$ – начальная концентрация q-го вещества; C_{ij}^* – упругие постоянные; ρ – плотность; $\alpha_{ij}^{(q)}$ – коэффициенты, характеризующие объемное изменение среды за счет диффузии; $D_{ij}^{(q)}$ – коэффициенты диффузии; R – универсальная газовая постоянная; T_0 – температура среды; $m^{(q)}$ – молярная масса q-го вещества; F^* – площадь сечения; J_3^* – момент инерции сечения балки относительно оси Ox_3; $\tau^{(q)}$ – время релаксации диффузионных потоков; Q^* – распределенная поперечная нагрузка; M^* – распределенный изгибающий момент.

Компоненты тензора напряжений определяются с помощью равенств [20], [21]:

$$\sigma_{11} = -x_2\left(\chi' + \sum_{j=1}^{N}\alpha_1^{(j)}H_j\right), \quad \sigma_{22} = -x_2\left(C_{12}\chi' + \sum_{j=1}^{N}\alpha_2^{(j)}H_j\right), \quad \sigma_{12} = C_{66}\left(v' - \chi\right).$$

Описываемая здесь модель учитывает конечную скорость распространения диффузионных возмущений, что обусловлено релаксацией диффузионных потоков. Инерционность в уравнения теплопереноса первым ввел Максвелл, а в 1948 году Каттанео был предложен вариант закона Фурье с релаксационным членом. Предложенная идея была распространена на модели, описывающие диффузионные процессы. В настоящее время существуют различные обобщения законов Фурье и Фика, с которыми можно ознакомиться в работах [23]–[27]. Используемая в данной работе модель массопереноса основана на теории Грина–Нагди [23], [27], согласно которой компоненты вектора диффузионного потока должны удовлетворять соотношениям:

$$\sum_{k=0}^{K}\frac{\tau_q^k}{k!}\frac{\partial^k J_1^{(q)}}{\partial\tau^k} = -x_2\left(D_1^{(q)}H_q' + \Lambda_{11}^{(q)}\chi''\right), \quad \sum_{k=0}^{K}\frac{\tau_q^k}{k!}\frac{\partial^k J_2^{(q)}}{\partial\tau^k} = -D_2^{(q)}H_q - \Lambda_{21}^{(q)}\chi' \quad \left(q=\overline{1,N}\right). \tag{2}$$

Верхний предел суммирования K в уравнениях (1) и формулах (2) определяется на основе заданной точности вычислений. Однако, как показывают расчеты [27], практически всегда можно ограничиться значением $K = 2$, а в большинстве случаев приемлемая точность обеспечивается даже при $K = 1$ (модель Каттанео). Случай $K = 0$ соответствует классической модели массопереноса с бесконечной скоростью распространения диффузионных возмущений.

Выражения для изгибающих моментов M_0 и перерезывающих сил Q_0, заданных на свободном конце стержня, также получены в работах [20], [21]

$$M_0 = \left(\chi' + \sum_{q=1}^{N}\alpha_1^{(q)}H_q\right)J_3, \quad Q_0 = (v'-\chi)C_{66}F.$$

Начальные условия полагаем нулевыми. Граничные условия в соответствии с моделью изгиба консоли имеют вид ($x = x_1$):

$$\begin{gathered} v\big|_{x=0} = 0, \quad \chi\big|_{x=0} = 0, \quad H_q\big|_{x=0} = 0, \quad (v'-\chi)\big|_{x=1} = f_{12}, \\ \left(\chi' + \sum_{q=1}^{N}\alpha_1^{(q)}H_q\right)\Bigg|_{x=1} = 0, \quad \left(\Lambda_{11}^{(q)}\chi'' + D_1^{(q)}H_q'\right)\Big|_{x=1} = 0. \end{gathered} \tag{3}$$

2. МЕТОД РЕШЕНИЯ

Как было отмечено во введении, основная проблема в решении поставленной задачи заключается в невозможности использования метода разделения переменных. Это не так важно, когда речь идет о статическом или стационарном изгибе балки. В данном случае указанное обстоятельство существенно осложняет обращение преобразования Лапласа, которое используется при решении этой задачи. Для преодоления указанной проблемы предлагается использовать метод эквивалентных граничных условий, который был успешно применен для решения аналогичной задачи для балки Бернулли–Эйлера [22].

Суть алгоритма заключается в том, что на начальном этапе вместо исходной задачи рассматривается вспомогательная задача со следующими граничными условиями:

$$\begin{gathered} v\big|_{x=0} = 0, \quad \left(\chi' + \sum_{q=1}^{N}\alpha_1^{(q)}H_q\right)\Bigg|_{x=0} = f_{21}^*, \quad H_q\big|_{x=0} = 0, \\ (v'-\chi)\big|_{x=1} = f_{12}, \quad \chi\big|_{x=1} = f_{22}^*, \quad \left(\Lambda_{11}^{(q)}\chi'' + D_1^{(q)}H_q'\right)\Big|_{x=1} = 0, \end{gathered} \tag{4}$$

где функции $f_{21}^*(\tau)$, $f_{22}^*(\tau)$ подлежат определению.

Ее решение в интегральной форме записывается в виде

$$v(x,\tau)=\sum_{l=1}^{2}\int_0^\tau G_{12l}(x,\tau-t)\,f_{2l}^*(t)\,dt+\int_0^\tau G_{112}(x,\tau-t)\,f_{12}(t)\,dt,$$
$$\chi(x,\tau)=\sum_{l=1}^{2}\int_0^\tau G_{22l}(x,\tau-t)\,f_{2l}^*(t)\,dt+\int_0^\tau G_{212}(x,\tau-t)\,f_{12}(t)\,dt, \tag{5}$$
$$H_q(x,\tau)=\sum_{l=1}^{2}\int_0^\tau G_{q+2,2l}(x,\tau-t)\,f_{2l}^*(t)\,dt+\int_0^\tau G_{q+2,12}(x,\tau-t)\,f_{12}(t)\,dt.$$

Здесь G_{mkl} – функции Грина задачи (1), (4), которые являются решениями следующих задач:

$$\ddot G_{1kl}-C_{66}k^2(G''_{1kl}-G'_{2kl})=0,$$
$$\ddot G_{2kl}-G''_{2kl}-aC_{66}k^2(G'_{1kl}-G_{2kl})-\sum_{q=1}^{N}\alpha_1^{(q)}G'_{q+2,kl}=0, \tag{6}$$
$$\sum_{k=0}^{K}\frac{\tau_q^k}{k!}\frac{\partial^k\dot G_{q+2,kl}}{\partial\tau^k}-\sum_{r=1}^{N}D_1^{(qr)}G''_{r+2,kl}-\Lambda_{11}^{(q)}G'''_{2kl}=0;$$

$$G_{1kl}\big|_{x=0}=\delta_{1k}\delta_{1l}\delta(\tau),\quad \left(G'_{2kl}+\sum_{q=1}^{N}\alpha_1^{(q)}H_q\right)\Big|_{x=0}=\delta_{2k}\delta_{1l}\delta(\tau),$$
$$G_{q+2,kl}\big|_{x=0}=\delta_{q+2,k}\delta_{1l}\delta(\tau),\quad \left(G'_{1kl}-G_{2kl}\right)\Big|_{x=1}=\delta_{1k}\delta_{2l}\delta(\tau),\quad G_{2kl}\big|_{x=1}=\delta_{2k}\delta_{2l}\delta(\tau), \tag{7}$$
$$\left(\Lambda_{11}^{(q)}G''_{2kl}+D_qG'_{q+2,kl}\right)\Big|_{x=1}=\delta_{q+2,k}\delta_{2l}\delta(\tau).$$

Для нахождения функций G_{mkl} используются преобразование Лапласа по времени и разложение в ряды Фурье. Применяя указанные действия к вспомогательной задаче (6), (7), получаем следующую систему линейных алгебраических уравнений (s – параметр преобразования Лапласа, верхний индекс L обозначает трансформанту Лапласа):

$$k_{1n}(s)G_{1kln}^L(s)-C_{66}k^2\lambda_nG_{2kln}^L(s)=F_{1kln},$$
$$-a\mu k^2\lambda_nG_{1kln}^L(s)+k_{2n}(s)G_{2kln}^L(s)-\lambda_n\sum_{j=1}^{N}\alpha_1^{(j)}G_{j+2,kln}^L(s)=F_{2kln}, \tag{8}$$
$$-\Lambda_{11}^{(q)}\lambda_n^3G_{2kln}^L(s)+k_{q+2,n}(s)G_{q+2,kln}^L(s)=F_{q+2,kln};$$

$$\begin{Bmatrix}G_{mkl}^L(x,s)\\G_{q+2,kl}^L(x,s)\end{Bmatrix}=\sum_{n=1}^{\infty}\begin{Bmatrix}G_{mkln}^L(s)\\G_{q+2,kln}^L(s)\end{Bmatrix}\sin\lambda_nx,\quad G_{2kl}^L(x,s)=\sum_{n=1}^{\infty}G_{2kln}^L(s)\sin\lambda_nx,\quad \lambda_n=\pi\left(n-\frac12\right),$$
$$\begin{Bmatrix}G_{mkln}^L(s)\\G_{q+2,kln}^L(s)\end{Bmatrix}=2\int_0^1\begin{Bmatrix}G_{mkl}^L(x,s)\\G_{q+2,kl}^L(x,s)\end{Bmatrix}\sin\lambda_nx\,dx,\quad G_{2kln}^L(s)=2\int_0^1G_{2kl}^L(x,s)\cos\lambda_nx\,dx. \tag{9}$$

Коэффициенты $k_{1n}(s)$ и правые части F_{1kln} системы (8) приведены в Приложении (формулы (25)).

Решение системы (8) имеет вид $\left(q,p=\overline{1,N},\ k=\overline{1,N+1}\right)$

$$G_{1kln}^L(s)=\frac{P_{1kln}(s)}{P_n(s)},\quad G_{2kln}^L(s)=\frac{P_{2kln}(s)}{P_n(s)},\quad G_{q+2,1ln}^L(s)=\frac{P_{q+2,1ln}(s)}{P_n(s)},$$
$$G_{q+2,2ln}^L(s)=\frac{2\Lambda_{11}^{(q)}\lambda_n\left(\delta_{1l}-(-1)^{n+1}\lambda_n\delta_{2l}\right)}{k_{q+2,n}(s)}+\frac{P_{q+2,2ln}(s)}{Q_{qn}(s)}, \tag{10}$$
$$G_{q+2,p+2,ln}^L(s)=\frac{P_{q+2,p+2,ln}(s)}{Q_{qn}(s)}+\frac{2\lambda_n\left(D_1^{(q)}\delta_{pq}-\Lambda_{11}^{(q)}\alpha_1^{(p)}\right)\delta_{1l}+2(-1)^{n+1}\delta_{pq}\delta_{2l}}{k_{q+2,n}(s)},$$

где $P_n(s)$, $Q_{qn}(s)$ и $P_{jkln}(s)$ – многочлены от s, приведенные в Приложении (формулы (26) и (27)).

Оригиналы в (10) находятся с помощью вычетов и таблиц операционного исчисления [28]

$$
\begin{gathered}
G_{ikln}(\tau)=\sum_{j=1}^{\Sigma}A_{ikln}^{(j)}e^{s_{jn}\tau},\quad A_{ikln}^{(j)}=\frac{P_{ikln}(s_{jn})}{P_n'(s_{jn})}\quad (i=1,2,\quad \Sigma=(K+1)N+4),\\
G_{q+2,1ln}^{s}(\tau)=\sum_{j=1}^{\Sigma}A_{q+2,1ln}^{(j)}e^{s_{jn}\tau},\quad A_{q+2,1ln}^{(j)}=\frac{P_{q+2,1ln}(s_{jn})}{P_n'(s_{jn})},\\
G_{q+2,2ln}(\tau)=2\Lambda_{11}^{(q)}\lambda_n\left(\delta_{1l}-(-1)^{n+1}\lambda_n\delta_{2l}\right)\sum_{r=1}^{K+1}\frac{e^{\xi_{rqn}\tau}}{k'_{q+1,n}(\xi_{rqn})}+\sum_{j=1}^{\Sigma+K+1}A_{q+1,2ln}^{(j)}e^{s_{jn}\tau},\\
G_{q+2,p+2,ln}(\tau)=\sum_{j=1}^{\Sigma+K+1}A_{q+1,p+2,ln}^{(j)}e^{s_{jn}\tau}+\\
+2\left[\lambda_n\left(D_1^{(q)}\delta_{pq}-\Lambda_{11}^{(q)}\alpha_1^{(p)}\right)\delta_{1l}+(-1)^{n+1}\delta_{pq}\delta_{2l}\right]\sum_{r=1}^{K+1}\frac{e^{\xi_{rqn}\tau}}{k'_{q+1,n}(\xi_{rqn})},\\
A_{q+2,kln}^{(j)}=\frac{P_{q+2,kl}(s_{jn})}{Q'_{qn}(s_{jn})}\quad (k\geq 2).
\end{gathered}
\tag{11}
$$

Здесь $s_{jn}\ \left(j=\overline{1,\Sigma}\right)$ – нули многочлена $P_n(s)$, ξ_{rqn} – нули многочлена $k_{q+2,n}(s)$. При $K=1$ они имеют вид:

$$
\xi_{1qn}(\lambda_n)=\frac{-1-\sqrt{1-4\tau_q D_1^{(q)}\lambda_n^2}}{2\tau_q},\quad \xi_{2qn}(\lambda_n)=\frac{-1+\sqrt{1-4\tau_q D_1^{(q)}\lambda_n^2}}{2\tau_q}.
$$

На следующем этапе алгоритма строятся соотношения, связывающие правые части граничных условий обеих задач. Так как решение вспомогательной задачи должно удовлетворять граничным условиям (3), то с учетом представлений (5) указанные соотношения записываются в виде системы интегральных уравнений Вольтерра I рода, которая имеет вид [22]

$$
\sum_{j=1}^{2}\int_0^{\tau}a_{ij}(\tau-t)f_{2j}^{*}(t)\,dt=\varphi_i(\tau),
\tag{12}
$$

где

$$
\begin{gathered}
a_{11}(\tau)=G_{221}(0,\tau),\quad a_{12}(\tau)=G_{222}(0,\tau),\\
a_{21}(\tau)=G'_{221}(1,\tau)+\sum_{j=1}^{N}\alpha_1^{(j)}G_{j+2,21}(1,\tau),\quad a_{22}(\tau)=G'_{222}(1,\tau)+\sum_{j=1}^{N}\alpha_1^{(j)}G_{j+2,22}(1,\tau),\\
\varphi_1(\tau)=-\int_0^{\tau}G_{212}(0,\tau-t)f_{12}(t)\,dt,\\
\varphi_2(\tau)=-\int_0^{\tau}\left[G'_{212}(1,t-\tau)+\sum_{j=1}^{N}\alpha_1^{(j)}G_{j+2,12}(1,\tau-t)\right]f_{12}(t)\,dt.
\end{gathered}
$$

Как известно, функции Грина принадлежат к классу обобщенных функций. Поэтому ряды (9), коэффициенты которых определяются равенствами (10), (11), сходятся только в обобщенном смысле. Это затрудняет применение численных алгоритмов для решения системы (12). Для преодоления указанной трудности проинтегрируем по частям интегралы в (12). Получим систему уравнений относительно производных $\partial f_{2j}^{*}(\tau)/\partial\tau$

$$
\begin{gathered}
\sum_{j=1}^{2}\int_0^{\tau}A_{ij}(\tau-t)\frac{\partial f_{2j}^{*}(t)}{\partial t}dt=F_i(\tau),\quad F_i(\tau)=\varphi_i(\tau)-\sum_{j=1}^{2}A_{ij}(\tau)f_{2j}^{*}(0),\\
A_{ij}(\tau)=\int_0^{\tau}a_{ij}(t)\,dt,\quad A_{ij}(\tau-t)=\int_0^{\tau-t}a_{ij}(\varepsilon)\,d\varepsilon.
\end{gathered}
\tag{13}
$$

Решение системы (13) будет зависеть от значений $f_{2j}^*(0)\,(j=1,2)$, которые определим исходя из условия сопряжения начальных и граничных условий в угловых точках пространственно-временной области рассматриваемых задач. С учетом нулевых начальных условий будем далее полагать, что $f_{2j}^*(0)=0$. В случае часто встречающихся в вычислительной практике разрывных решений функции $f_{2j}^*(\tau)$ в точке $\tau=0$, вообще говоря, не определены и могут принимать любые значения. Поэтому здесь, исходя из вышеизложенного, также полагаем $f_{2j}^*(0)=0$.

Полученная система уравнений (13) решается численно с помощью квадратурных формул. Для этого разбиваем область $[0,T]$ изменения времени τ на N_τ отрезков точками $\tau_m = mh\left(m=\overline{0,N_\tau}\right)$ с равномерным шагом $h=T/N_\tau$ и вводим сеточные функции $y_m^j = \partial f_{2j}^*(\tau_m)/\partial\tau$, $A_m^{ij} = A_{ij}(\tau_m)$.

Каждый из интегралов в (13) при $\tau=\tau_m$ приближенно заменяем суммой, соответствующей формуле средних прямоугольников:

$$\int_0^\tau A_{ij}(\tau-t)\frac{\partial f_{2j}^*(t)}{\partial t}\,dt \approx hS_{m-1/2}^{ij} + hA_{1/2}^{ij}y_{m-1/2}^j, \quad S_{m-1/2}^{ij} = \sum_{l=1}^{m-1} A_{m-l+1/2}^{ij}y_{l-1/2}^j \quad \left(i,j=\overline{1,N+2}\right),$$

$$\tau_{m-1/2} = \frac{\tau_{m-1}+\tau_m}{2} = h\left(m-\frac{1}{2}\right), \quad \tau_{m-l+1/2} = \tau_m - \tau_{l-1/2} = h\left(m-l+\frac{1}{2}\right) \quad \left(m=\overline{1,N_\tau}\right).$$

В результате приходим к рекуррентной последовательности систем линейных алгебраических уравнений $(m\ge 1)$:

$$\mathbf{A}\mathbf{y}_{m-1/2} = \mathbf{b}_{m-1/2}, \tag{14}$$

где $\mathbf{y}_{m-1/2} = \left(y_{m-1/2}^i\right)_{2\times 1}$ – столбец неизвестных, а остальные величины определяются так

$$\mathbf{A} = \left(A_{1/2}^{ij}\right)_{2\times 2}, \quad \mathbf{b}_{m-1/2} = \left(b_{m-1/2}^i\right)_{2\times 1}, \quad b_{m-1/2}^i = \frac{F_i(\tau_m)}{h} - \sum_{j=1}^{2} S_{m-1/2}^{ij}.$$

Ее решение находится по формулам Крамера и имеет следующий вид:

$$y_{m-1/2}^1 = \frac{b_{m-1/2}^1 A_{1/2}^{22} - b_{m-1/2}^2 A_{1/2}^{12}}{A_{1/2}^{11}A_{1/2}^{22} - A_{1/2}^{12}A_{1/2}^{21}}, \quad y_{m-1/2}^2 = \frac{b_{m-1/2}^2 A_{1/2}^{11} - b_{m-1/2}^1 A_{1/2}^{21}}{A_{1/2}^{11}A_{1/2}^{22} - A_{1/2}^{12}A_{1/2}^{21}}. \tag{15}$$

Таким образом, решение исходной задачи (1), (3) получается путем численного вычисления сверток функций Грина вспомогательной задачи (1), (4) с сеточными функциями (15), полученными в результате численного решения системы уравнений (14). При этом с учетом преобразования (13) свертки (5) запишутся так:

$$v(x,\tau_i) = h\sum_{l=1}^{2}\sum_{j=1}^{i}\tilde G_{12l}(x,\tau_{i-j+1/2})\,y_{j-1/2}^l + \int_0^{\tau_i} G_{112}(x,\tau_i-t)\,f_{12}(t)\,dt,$$

$$\chi(x,\tau_i) = h\sum_{l=1}^{2}\sum_{j=1}^{i}\tilde G_{22l}(x,\tau_{i-j+1/2})\,y_{j-1/2}^l + \int_0^{\tau_i} G_{212}(x,\tau_i-t)\,f_{12}(t)\,dt, \tag{16}$$

$$H_q(x,\tau_i) = h\sum_{l=1}^{2}\sum_{j=1}^{i}\tilde G_{q+2,2l}(x,\tau_{i-j+1/2})\,y_{j-1/2}^l + \int_0^{\tau_i} G_{q+2,12}(x,\tau_i-t)\,f_{12}(t)\,dt, \quad \tilde G_{mkl}(x,\tau) = \int_0^{\tau} G_{mkl}(x,t)\,dt.$$

3. ПРЕДЕЛЬНЫЕ ПЕРЕХОДЫ

Полагая в (1) и (4) $\alpha_1^{(q)}=0$, $D_1^{(q)}=0$, из соотношений (11) получаем функции Грина $G_{ij}^v(x,\tau) = \lim_{\alpha_1^{(q)}\to 0} G_{1ij}(x,\tau)$ и $G_{ij}^\chi(x,\tau) = \lim_{\alpha_1^{(q)}\to 0} G_{2ij}(x,\tau)$ упругой задачи

$$G_{11}^{v}(x,\tau)=2C_{66}k^{2}\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{\lambda_{n}\left(\lambda_{n}^{2}-\gamma_{jn}^{2}\right)\sin\gamma_{jn}\tau\sin\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{21}^{v}(x,\tau)=-2C_{66}k^{2}\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{\lambda_{n}\sin\gamma_{jn}\tau\sin\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{12}^{v}(x,\tau)=2C_{66}k^{2}\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{(-1)^{n+1}\left(\lambda_{n}^{2}-\gamma_{jn}^{2}+aC_{66}k^{2}\right)\sin\gamma_{jn}\tau\sin\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{22}^{v}(x,\tau)=-2C_{66}k^{2}\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{(-1)^{n+1}\lambda_{n}^{2}\sin\gamma_{jn}\tau\sin\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{11}^{\chi}(x,\tau)=2C_{66}k^{2}a\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{\gamma_{jn}\sin\gamma_{jn}\tau\cos\lambda_{n}x}{\nu_{n}-2\gamma_{jn}^{2}},$$

$$G_{21}^{\chi}(x,\tau)=-2\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{\left(C_{66}k^{2}\lambda_{n}^{2}-\gamma_{jn}^{2}\right)\sin\gamma_{jn}\tau\cos\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{12}^{\chi}(x,\tau)=2a^{2}C_{66}^{2}k^{4}\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{(-1)^{n+1}\lambda_{n}\sin\gamma_{jn}\tau\cos\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)},$$

$$G_{22}^{\chi}(x,\tau)=2\sum_{n=1}^{\infty}\sum_{j=1}^{2}\frac{(-1)^{n+1}\lambda_{n}\left(C_{66}k^{2}\lambda_{n}^{2}-\gamma_{jn}^{2}\right)\sin\gamma_{jn}\tau\cos\lambda_{n}x}{\gamma_{jn}\left(\nu_{n}-2\gamma_{jn}^{2}\right)}.$$

Здесь нули многочлена $P_n(s)$ при $\alpha_1^{(q)}=0$ и $D_1^{(q)}=0$ представлены в виде

$$s_{1n}=i\gamma_{1n},\quad s_{2n}=i\gamma_{2n},\quad s_{3n}=-s_{1n},\quad s_{4n}=-s_{2n},$$

$$\gamma_{1n}=\sqrt{\frac{\nu_{n}+\sqrt{\nu_{n}^{2}-4C_{66}k^{2}\lambda_{n}^{4}}}{2}},\quad \gamma_{2n}=\sqrt{\frac{\nu_{n}-\sqrt{\nu_{n}^{2}-4C_{66}k^{2}\lambda_{n}^{4}}}{2}},$$

$$\nu_{n}=\lambda_{n}^{2}+C_{66}k^{2}\lambda_{n}^{2}+aC_{66}k^{2}.$$

Полагая в граничных условиях (3), (4)

$$f_{km}^{*}(\tau)=\tilde{f}_{km}^{*}H(\tau),\quad f_{km}(\tau)=\tilde{f}_{km}H(\tau),$$

и переходя к пределу при $\tau\to\infty$ можно получить решение задачи об изгибе консольно-закрепленной балки под действием статической нагрузки, приложенной к свободному концу. Здесь $H(\tau)$ – функция Хевисайда.

Функции Грина статической задачи $G_{mk}^{st}(x)$ выражаются через функции Грина $G_{mk}(x,\tau)$ динамической задачи с помощью соотношений [28]

$$G_{mk}^{st}(x)=\lim_{\tau\to\infty}\left[G_{mk}(x,\tau)*H(\tau)\right]=\lim_{s\to0}\left[sG_{mk}^{L}(x,s)\frac{1}{s}\right]=\lim_{s\to0}G_{mk}^{L}(x,s).\tag{17}$$

Преобразуя свертки (5) с помощью (17), получаем решение статической задачи

$$\begin{aligned}v^{(st)}(x)&=G_{121}^{(st)}(x)\tilde{f}_{21}^{*}+G_{122}^{(st)}(x)\tilde{f}_{22}^{*}+G_{112}^{(st)}(x)\tilde{f}_{12},\\ \chi^{(st)}(x)&=G_{221}^{(st)}(x)\tilde{f}_{21}^{*}+G_{222}^{(st)}(x)\tilde{f}_{22}^{*}+G_{212}^{(st)}(x)\tilde{f}_{12},\\ H_{q}^{(st)}(x)&=G_{q+2,21}^{(st)}(x)\tilde{f}_{21}^{*}+G_{q+2,22}^{(st)}(x)\tilde{f}_{22}^{*}+G_{q+2,12}^{(st)}(x)\tilde{f}_{12}.\end{aligned}\tag{18}$$

Вычисляя предел (17), получаем следующие выражения для функций $G_{r21}^{(st)}(x)$, $G_{r12}^{(st)}(x)$, $G_{r22}^{(st)}(x)$ ($r=\overline{1,N+2}$, суммы рядов найдены с помощью таблиц [29])

$$G_{121}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{121n}(0)}{P_n(0)}\sin\lambda_n x=-2\sum_{n=1}^{\infty}\frac{\sin\lambda_n x}{\lambda_n^3}=\frac{x^2}{2}-x,$$

$$G_{112}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{112n}(0)}{P_n(0)}\sin\lambda_n x=2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n^2}\sin\lambda_n x+$$

$$+2aC_{66}k^2\Phi\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n^4}\sin\lambda_n x=x-aC_{66}k^2\frac{\Phi}{2}\left(\frac{x^3}{3}-x\right),$$

$$G_{122}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{122n}(0)}{P_n(0)}\sin\lambda_n x=2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n^2}\sin\lambda_n x=x,$$

$$G_{221}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{221n}(0)}{P_n(0)}\cos\lambda_n x=-2\sum_{n=1}^{\infty}\frac{\cos\lambda_n x}{\lambda_n^2}=x-1,$$

$$G_{212}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{212n}(0)}{P_n(0)}\cos\lambda_n x=2a^2C_{66}k^2\Phi\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n^3}\cos\lambda_n x=-a^2C_{66}k^2\Phi\left(\frac{x^2}{2}-\frac{1}{2}\right),$$

$$G_{222}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{222n}(0)}{P_n(0)}\cos\lambda_n x=2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n}\cos\lambda_n x=1, \tag{19}$$

$$G_{q+2,21}^{(st)}(x)=\frac{2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}\frac{\sin\lambda_n x}{\lambda_n}+\sum_{n=1}^{\infty}\frac{P_{q+1,21n}(0)}{Q_{qn}(0)}\sin\lambda_n x=$$

$$=\frac{2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}\frac{\sin\lambda_n x}{\lambda_n}+\frac{-2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}\frac{\sin\lambda_n x}{\lambda_n}=0,$$

$$G_{q+2,12}^{(st)}(x)=\sum_{n=1}^{\infty}\frac{P_{q+1,12n}(0)}{Q_{qn}(0)}\sin\lambda_n x=$$

$$=2a^2C_{66}k^2\Lambda_{11}^{(q)}\Phi_q\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\lambda_n^2}\sin\lambda_n x=a^2C_{66}k^2\Lambda_{11}^{(q)}\Phi_q x,$$

$$G_{q+2,22}^{(st)}(x)=-\frac{2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}(-1)^{n+1}\sin\lambda_n x+\sum_{n=1}^{\infty}\frac{P_{q+1,22n}(0)}{Q_{qn}(0)}\sin\lambda_n x=$$

$$=-\frac{2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}(-1)^{n+1}\sin\lambda_n x+\frac{2\Lambda_{11}^{(q)}}{D_1^{(q)}}\sum_{n=1}^{\infty}(-1)^{n+1}\sin\lambda_n x=0.$$

Здесь введены следующие обозначения:

$$\Phi=\frac{\prod_{j=1}^{N}D_1^{(j)}}{\prod_{j=1}^{N}D_1^{(j)}-\sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\prod_{r=1,r\neq j}^{N}D_1^{(r)}},\quad \Phi_q=\frac{\prod_{r=1,r\neq q}^{N}D_1^{(r)}}{\prod_{j=1}^{N}D_1^{(j)}-\sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\prod_{r=1,r\neq j}^{N}D_1^{(r)}}. \tag{20}$$

С учетом предельного перехода (17) система уравнений (12) запишется в виде

$$\sum_{j=1}^{2}\tilde{a}_{ij}\tilde{f}_{1j}^{*}=\tilde{\varphi}_i, \tag{21}$$

где с учетом формул (19), (20)

$$\tilde{a}_{11} = G_{221}^{(st)}(0) = -1, \quad \tilde{a}_{12} = G_{222}^{(st)}(0) = 1,$$
$$\tilde{a}_{21} = G_{221}^{(st)\prime}(1) + \sum_{j=1}^{N} \alpha_1^{(j)} G_{j+2,21}^{(st)}(1) = 1, \quad \tilde{a}_{22} = G_{222}^{(st)\prime}(1) + \sum_{j=1}^{N} \alpha_1^{(j)} G_{j+2,22}^{(st)}(1) = 0, \tag{22}$$
$$\tilde{\varphi}_1 = -G_{212}^{(st)}(0)\tilde{f}_{12} = -\frac{a^2 C_{66} k^2 \Phi}{2}\tilde{f}_{12}, \quad \tilde{\varphi}_2 = -\left[G_{212}^{(st)\prime}(1) + \sum_{j=1}^{N} \alpha_1^{(j)} G_{j+1,12}^{(st)}(1)\right]\tilde{f}_{12} = a^2 C_{66} k^2 \tilde{f}_{12}.$$

Решение системы (21) находится по формулам (15). При этом используются следующие соответствия:

$$y_{m-1/2}^{i} \leftrightarrow \tilde{f}_{2j}^{*}, \quad A_{1/2}^{ij} \leftrightarrow \tilde{a}_{ij}, \quad b_{m-1/2}^{i} \leftrightarrow \tilde{\varphi}_i.$$

Используя равенства (22), получаем

$$\tilde{f}_{21}^{*} = a^2 C_{66} k^2 \tilde{f}_{12}, \quad \tilde{f}_{22}^{*} = a^2 C_{66} k^2 \left(1 - \frac{\Phi}{2}\right)\tilde{f}_{12}.$$

В результате решение статической задачи на основании равенств (18) записывается так

$$v^{(st)}(x) = \frac{a^2 C_{66} k^2}{2}\left(x^2 - \Phi\frac{x^3}{3}\right)\tilde{f}_{12} + x\tilde{f}_{12},$$
$$\chi^{(st)}(x) = a^2 C_{66} k^2 \tilde{f}_{12}\left(x - \Phi\frac{x^2}{2}\right), \quad H_q^{(st)}(x) = a^2 C_{66} k^2 \Lambda_{11}^{(q)} \Phi_q \tilde{f}_{12} x. \tag{23}$$

Для несвязанной задачи при $\alpha_1^{(q)} = 0$ (в этом случае $\Lambda_{11}^{(j)} = 0$) из соотношений (20) получаем, что $\Phi = 1$, $\Phi_q = 1/D_1^{(q)}$ и равенства (23) принимают вид

$$v^{(st)}(x) = \frac{a^2 C_{66} k^2}{2}\left(x^2 - \frac{x^3}{3}\right)\tilde{f}_{12} + x\tilde{f}_{12}, \quad \chi^{(st)}(x) = a^2 C_{66} k^2 \tilde{f}_{12}\left(x - \frac{x^2}{2}\right), \quad H_q^{(st)}(x) = 0. \tag{24}$$

Заметим, что функции прогиба $v^{(st)}(x)$ и угла поворота сечений $\chi^{(st)}(x)$, полученные с помощью предельного перехода (17), совпадают с классическим решением статической задачи об изгибе консольно-закрепленной балки Тимошенко [30].

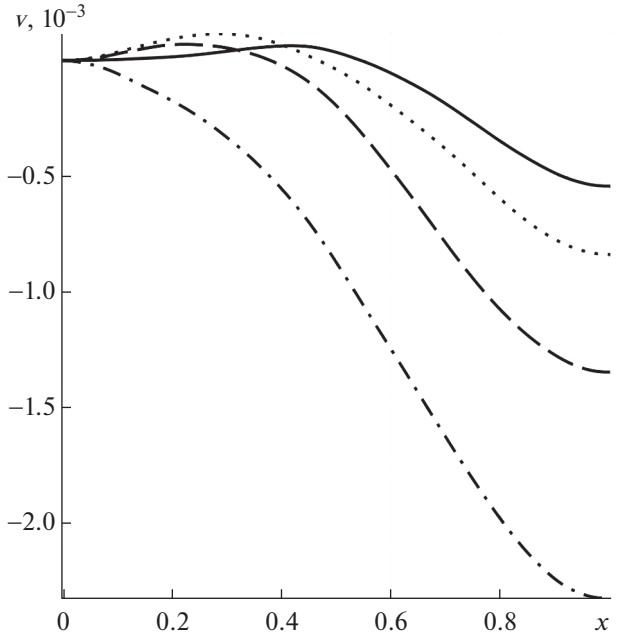
4. РАСЧЕТНЫЙ ПРИМЕР

Для расчета рассмотрим стержень длиной $l = 1$ см, прямоугольного сечения $h \times b = 0.05l \times 0.05l$ из трехкомпонентного материала (цинк, медь, алюминий), где в качестве независимых компонент выступают цинк (компонент 1) и медь (компонент 2) [31]:

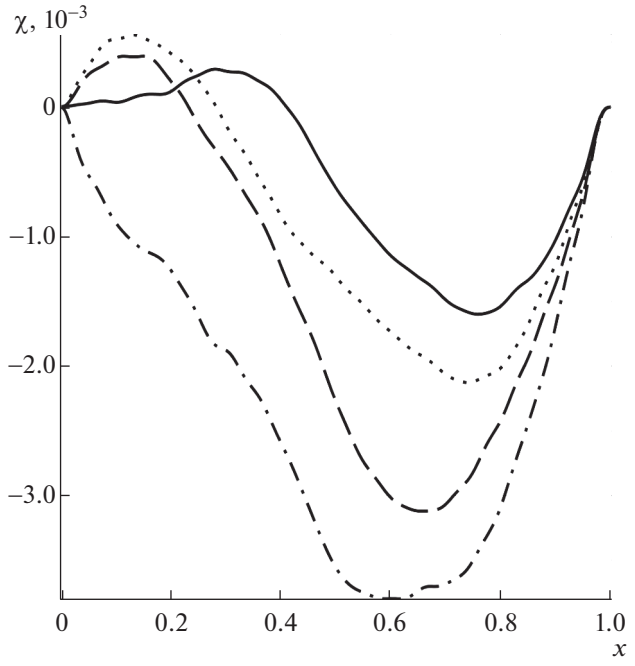
$$C_{12}^{*} = 6.93 \times 10^{10}\ \frac{\text{Н}}{\text{м}^2}, \quad C_{66}^{*} = 2.56 \times 10^{10}\ \frac{\text{Н}}{\text{м}^2}, \quad T_0 = 700\ \text{К}, \quad \rho = 2780\ \frac{\text{кг}}{\text{м}^3},$$
$$C_{11}^{*} = C_{12}^{*} + 2C_{66}^{*}, \quad l = 0.01\ \text{м}, \quad n_0^{(1)} = 0.01, \quad n_0^{(2)} = 0.045,$$
$$D_{11}^{*(1)} = 2.62 \times 0^{-12}\ \frac{\text{м}^2}{\text{с}}, \quad D_{11}^{*(2)} = 6.67 \times 10^{-14}\ \frac{\text{м}^2}{\text{с}}, \quad m^{(1)} = 0.065\ \frac{\text{кг}}{\text{моль}},$$
$$\alpha_{11}^{*(1)} = 6.55 \times 10^{7}\ \frac{\text{Дж}}{\text{кг}}, \quad \alpha_{11}^{*(2)} = 6.14 \times 10^{7}\ \frac{\text{Дж}}{\text{кг}}, \quad m^{(2)} = 0.064\ \frac{\text{кг}}{\text{моль}}.$$

Для того, чтобы сравнить полученные здесь результаты с результатами в работе [22], поперечную нагрузку на конце стержня $x = 1$ зададим в виде

$$f_{12}(\tau) = -\frac{J_3}{C_{66}F} H(\tau).$$

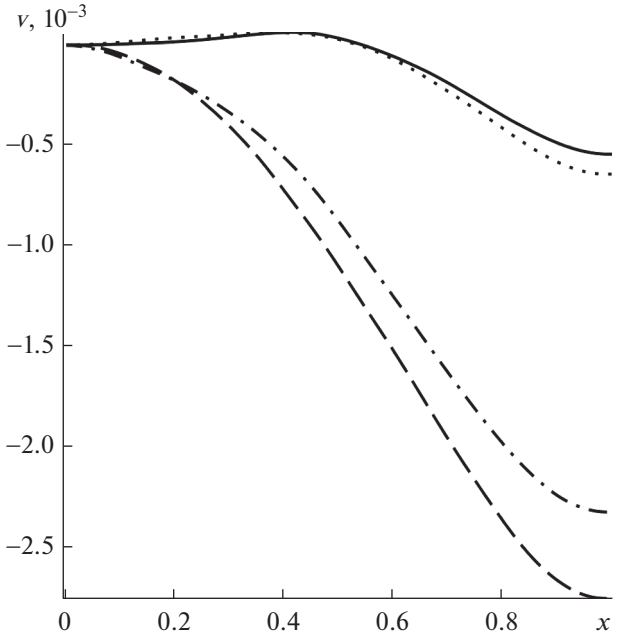


Фиг. 2. Прогибы балки $v(x,\tau)$. Сплошная линия соответствует $\tau = 3.3$, пунктирная линия – $\tau = 5.0$, штриховая линия – $\tau = 6.6$, штрихпунктирная линия – $\tau = 10.0$.

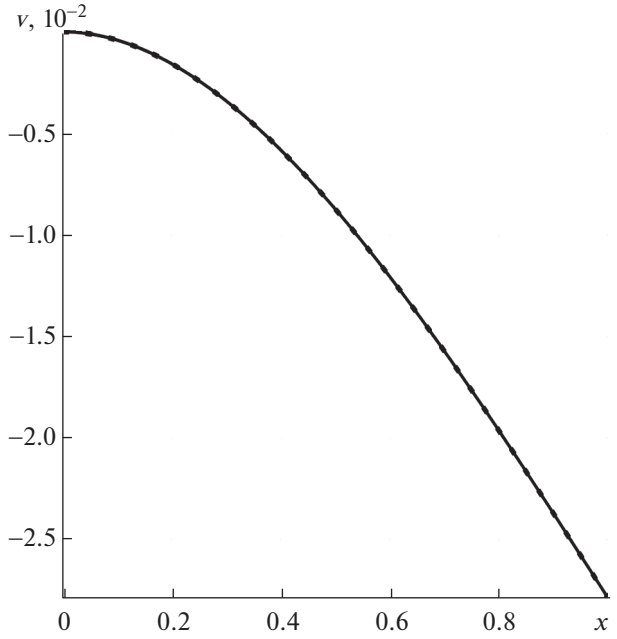


Фиг. 3. Повороты сечений $\chi(x,\tau)$. Сплошная линия соответствует $\tau = 3.3$, пунктирная линия – $\tau = 5.0$, штриховая линия – $\tau = 6.6$, штрихпунктирная линия – $\tau = 10.0$.

Решая численно систему (13) и подставляя найденные оттуда функции в свертки (16), получаем прогибы балки и повороты сечений, представленные на фиг. 2 и 3. Здесь для численного решения системы уравнений Вольтерра (13) использовалось $N_\tau = 40$ точек разбиения. Дальнейшее увеличение этого количества уже не приводит к какому-либо видимому изменению результатов.

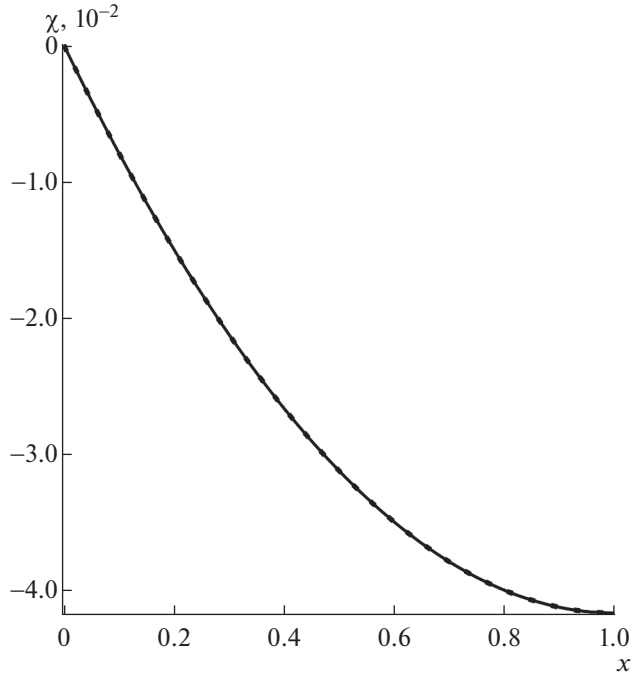


Фиг. 4. Прогибы балки $v(x,\tau)$. Решение для балки Тимошенко: сплошная линия соответствует $\tau = 3.3$, штрихпунктирная линия – $\tau = 10.0$. Решение для балки Бернулли–Эйлера: пунктирная линия – $\tau = 3.3$, штриховая линия – $\tau = 10.0$.

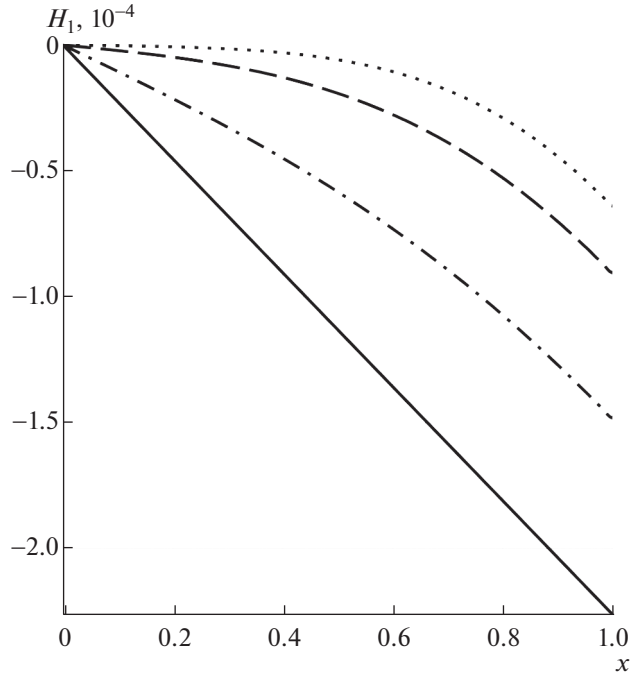


Фиг. 5. Прогибы балки $v^{(st)}(x)$ при статической нагрузке. Сплошная линия соответствует решению упругодиффузионной задачи; пунктирная соответствует решению упругой задачи при $\alpha_1^{(q)} = 0$.

Полученные результаты соответствуют классическим представлениям об изгибе консольно-закрепленных балок. На фиг. 4 можно видеть, как различаются прогибы балки, полученные с помощью модели Тимошенко и модели Бернулли–Эйлера. Решение аналогичной задачи для балки Бернулли–Эйлера было получено в работе [22].

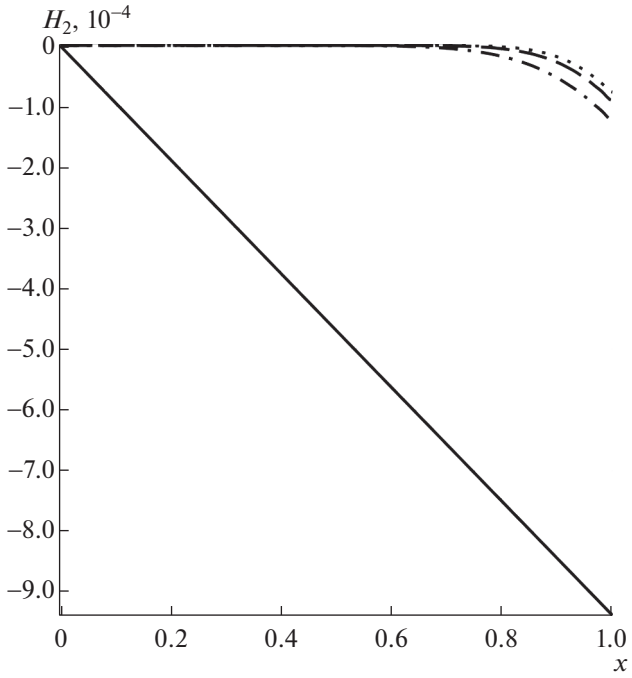


Фиг. 6. Повороты сечений $\chi(x)$ при статической нагрузке. Сплошная линия соответствует решению упруго-диффузионной задачи, пунктирная соответствует решению упругой задачи при $\alpha_1^{(q)} = 0$.



Фиг. 7. Линейная плотность приращения концентрации $H_1(x,\tau)$. Пунктирная линия соответствует $\tau = 1.5\times10^{12}$, штриховая линия – $\tau = 3.0\times10^{12}$, штрихпунктирная линия – $\tau = 8.3\times10^{12}$, сплошная линия соответствует решению статической задачи.

Прогибы и повороты сечений, соответствующие статическому режиму нагружения, получены по формулам (23) и представлены на фиг. 5 и 6. Видно, что при статических нагрузках массоперенос не влияет на поле перемещений.



Фиг. 8. Линейная плотность приращения концентрации $H_2(x,\tau)$. Пунктирная линия соответствует $\tau = 1.5\times10^{13}$, штриховая линия – $\tau = 3.0\times10^{13}$, штрихпунктиная линия – $\tau = 8.3\times10^{13}$, сплошная линия соответствует решению статической задачи.

Наконец, на фиг. 7 и 8 показано, как изменяется концентрация первого и второго компонентов в результате нестационарного изгиба консольно-закрепленного стержня. На фиг. 7 видно, что на рассматриваемом промежутке времени ($\tau = 8.3\times10^{12} \approx 5$ мес) приращение концентрации первого компонента (цинк) практически достигло предельного значения (сплошная линия), которое соответствует статическому режиму нагружения балки. Диффузия второго компонента (медь) происходит медленнее. На фиг. 8 видно, что приращение концентрации за время $\tau = 8.3\times10^{13} \approx 4.2$ года еще очень далеко от предельного значения (сплошная линия). На этих рисунках предельные значения приращений концентрации цинка и меди (сплошные линии) определяются по формуле (23).

ЗАКЛЮЧЕНИЕ

Предложен алгоритм, который позволяет найти решение нестационарной задачи механодиффузии для консольно-закрепленной балки Тимошенко.

На примере трехкомпонентной балки показано, что нестационарный изгиб консоли инициирует диффузионные потоки каждой из компонент. При этом массоперенос различных компонент осуществляется с различной интенсивностью. Также возникают вертикальные диффузионные потоки цинка и меди, которые компенсируются потоком третьей компоненты. Величина диффузионного потока увеличивается от закрепленного конца стержня к свободному концу.

Для верификации предложенного алгоритма проанализированы предельные переходы к классическим задачам теории упругости для консольно-закрепленных балок, а также сделано сравнение полученных здесь результатов с результатами аналогичной задачи для балки Бернулли–Эйлера, решение которой получено в работе [22]. Исследован предельный переход к статическим режимам. Выполнено сравнение полученного статического решения с известным классическим решением для консольно-закрепленной балки Тимошенко [30].

ПРИЛОЖЕНИЕ

1. Коэффициенты $k_{1n}(s)$ и правые части F_{1kln} системы (8)

$$k_{1n}(s) = C_{66}k^2\lambda_n^2 + s^2, \quad k_{2n}(s) = \lambda_n^2 + s^2 + aC_{66}k^2, \quad k_{q+2,n}(s) = \sum_{k=0}^{K}\frac{\tau_q^k}{k!}s^{k+1} + D_1^{(q)}\lambda_n^2,$$

$$\begin{gathered} F_{1kln} = 2C_{66}k^2\lambda_n\delta_{1k}\delta_{1l} + 2C_{66}k^2(-1)^{n+1}\delta_{1k}\delta_{2l}, \\ F_{2kln} = -2aC_{66}k^2\delta_{1k}\delta_{1l} - 2\delta_{2k}\delta_{1l} + 2(-1)^{n+1}\lambda_n\delta_{2k}\delta_{2l}, \\ F_{q+2,kln} = 2\Lambda_{11}^{(q)}\lambda_n\delta_{2k}\delta_{1l} - 2\Lambda_{11}^{(q)}(-1)^{n+1}\lambda_n^2\delta_{2k}\delta_{2l} + \\ + 2(-1)^{n+1}\delta_{q+2,k}\delta_{2l} + 2\lambda_n\left(D_1^{(q)}\delta_{q+2,k}\delta_{1l} - \Lambda_{11}^{(q)}\sum_{j=1}^{N}\alpha_1^{(j)}\delta_{j+2,k}\delta_{1l}\right). \end{gathered} \tag{25}$$

2. Многочлены $P_n(s)$, $Q_{qn}(s)$ и $P_{jkln}(s)$ к решениям (10)

$$\begin{gathered} P_n(s) = \left[k_{1n}(s)k_{2n}(s) - C_{66}^2k^4\lambda_n^2a\right]\Pi_n(s) - \lambda_n^4k_{1n}(s)\sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\Pi_{jn}(s), \\ Q_{qn}(s) = k_{q+2,n}(s)P_n(s), \quad \Pi_n(s) = \prod_{j=1}^{N}k_{j+2,n}(s), \quad \Pi_{jn}(s) = \prod_{k=1,k\neq j}^{N}k_{k+2,n}(s); \end{gathered} \tag{26}$$

$$\begin{gathered} P_{111n}(s) = 2C_{66}k^2\lambda_n\left[S_n^{(2)}(s) - aC_{66}k^2\Pi_n(s)\right], \\ P_{121n}(s) = -2C_{66}k^2\lambda_nS_n^{(1)}(s), \quad P_{1,q+2,1n}(s) = 2C_{66}k^2\alpha_1^{(q)}\lambda_n^3S_{qn}(s), \end{gathered} \tag{27}$$

$$\begin{gathered} P_{211n}(s) = 2C_{66}k^2a\Pi_n(s)\left[C_{66}k^2\lambda_n^2 - k_{1n}(s)\right], \\ P_{221n}(s) = -2k_{1n}(s)S_n^{(1)}(s), \quad P_{2,q+2,1n}(s) = 2\lambda_n^2\alpha_1^{(q)}k_{1n}(s)S_{qn}(s), \end{gathered}$$

$$\begin{gathered} P_{q+2,11n}(s) = 2C_{66}k^2\Lambda_{11}^{(q)}\lambda_n^3a\left[C_{66}k^2\lambda_n^2 - k_{1n}(s)\right]\Pi_{qn}(s), \\ P_{q+2,21n}(s) = -2\Lambda_{11}^{(q)}k_{1n}(s)\lambda_n^3S_n^{(1)}(s), \quad P_{q+2,p+2,1n}(s) = 2\lambda_n^4k_{1n}(s)\alpha_1^{(p)}\Lambda_{11}^{(q)}S_{qn}(s), \end{gathered}$$

$$\begin{gathered} P_{112n}(s) = 2C_{66}k^2(-1)^{n+1}S_n^{(2)}(s), \quad P_{122n}(s) = 2(-1)^{n+1}C_{66}k^2\lambda_n^2S_n^{(1)}(s), \\ P_{1,q+2,2n}(s) = 2(-1)^{n+1}C_{66}\alpha_1^{(q)}k^2\lambda_n^2\Pi_{qn}(s), \end{gathered}$$

$$\begin{gathered} P_{212n}(s) = 2(-1)^{n+1}a^2C_{66}^2k^4\lambda_n\Pi_n(s), \quad P_{222n}(s) = 2(-1)^{n+1}\lambda_nk_{1n}(s)S_n^{(1)}(s), \\ P_{2,q+2,2n}(s) = 2(-1)^{n+1}\alpha_1^{(q)}\lambda_nk_{1n}(s)\Pi_{qn}(s), \end{gathered}$$

$$\begin{gathered} P_{q+2,12n}(s) = 2(-1)^{n+1}a^2C_{66}^2k^4\Lambda_{11}^{(q)}\lambda_n^4\Pi_n(s), \quad P_{q+2,22n}(s) = 2(-1)^{n+1}\Lambda_{11}^{(q)}\lambda_n^4k_{1n}(s)S_n^{(1)}(s), \\ P_{q+2,p+2,2n}(s) = 2(-1)^{n+1}\alpha_1^{(p)}\Lambda_{11}^{(q)}\lambda_n^4k_{1n}(s)\Pi_{pn}(s), \end{gathered}$$

$$\begin{gathered} S_n^{(1)}(s) = \Pi_n(s) - \lambda_n^2\sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\Pi_{jn}(s), \quad S_{qn}(s) = \Pi_{qn}(s)D_1^{(q)} - \sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\Pi_{jn}(s), \\ S_n^{(2)}(s) = k_{2n}(s)\Pi_n(s) - \lambda_n^4\sum_{j=1}^{N}\alpha_1^{(j)}\Lambda_{11}^{(j)}\Pi_{jn}(s). \end{gathered}$$

СПИСОК ЛИТЕРАТУРЫ

1. *Le K.C.* Vibrations of shells and rods. Berlin: Springer Verlag, 1999. 419 p.
2. *Le K.C., Yi J.H.* An asymptotically exact theory of smart sandwich shells // Int. J. Engng. Sci. 2016. V. 106. P. 179–198.
3. *Михайлова Е.Ю., Тарлаковский Д.В., Федотенков Г.В.* Общая теория упругих оболочек. М.: МАИ, 2018. 112 с.
4. *Mindlin R.D., Yang J.* An Introduction to the Mathematical Theory of Vibrations of Elastic Plates. World Scientific Publishing, 2006. 212 p.
5. *Плескачевский Ю.М., Старовойтов Э.И., Леоненко Д.В.* Механика трехслойных стержней и пластин, связанных с упругим основанием. М.: Физматлит, 2011. 560 с.
6. *Mansfield E.H.* The Bending and Stretching of Plates. Cambridge University Press, 2005. 244 p.
7. *Швец Р.Н., Флячок В.М.* Уравнения механодиффузии анизотропных оболочек с учетом поперечных деформаций // Матем. методы и физико-механ. поля. 1984. № 20. С. 54–61.
8. *Швец Р.Н., Флячок В.М.* Вариационный подход к решению динамических задач механотермодиффузии анизотропных оболочек // Матем. физ. и нелинейн. механ. 1991. № 16. С. 39–43.
9. *Aouadi M., Copetti M.I.M.* Analytical and numerical results for a dynamic contact problem with two stops in thermoelastic diffusion theory // ZAMM Z. Angew. Math. Mech. 2015. V. 2015. https://doi.org/10.1002/zamm.201400285
10. *Copetti M., Aouadi M.* A quasi-static contact problem in thermoviscoelastic diffusion theory // Applied Numerical Mathematics. 2016. V. 109. P. 157–183. https://doi.org/10.1051/m2an/201603
11. *Aouadi M., Miranville A.* Smooth attractor for a nonlinear thermoelastic diffusion thin plate based on Gurtin-Pipkin's model // Asymptotic Analysis. 2015. V. 95. P. 129–160.
12. *Aouadi M.* On thermoelastic diffusion thin plate theory // Appl. Math. Mech. Engl. Ed. 2015. V. 36. № 5. P. 619–632.
13. *Aouadi M., Miranville A.* Quasi-stability and global attractor in nonlinear thermoelastic diffusion plate with memory // Evolution Equations and Control Theory. 2015. V. 4. № 3. P. 241–263.
14. *Bhattacharya D., Kanoria M.* The influence of two temperature generalized thermoelastic diffusion inside a spherical shell // Internat. Journal of Eng. and Technical Research (IJETR). 2014. V. 2. № 5. P. 151–159.
15. *Aouadi M.* A generalized thermoelastic diffusion problem for an infinitely long solid cylinder // Intern. J. Mathem. and Mathem. Sci. 2006. V. 6. P. 1–16. https://doi.org/10.1155/IJMMS/2006/25976
16. *Elhagary M.A.* Generalized thermoelastic diffusion problem for an infinitely long hollow cylinder for short times // Acta Mech. 2011. V. 218. P. 5–15.
17. *Tripathi J.J., Kedar G.D., Deshmukh K.C.* Generalized thermoelastic diffusion in a thick circular plate including heat source // Alexandria Eng. Journal. 2016. V. 55. № 3. P. 2241–2249.
18. *Zakian V.* Numerical inversions of Laplace transforms // Electron. Lett. 1969. V. 5. P. 120–121.
19. *Крылов В.И., Скобля Н.С.* Методы приближенного преобразования Фурье и обращения преобразования Лапласа. М.: ФИЗМАТЛИТ, 1974. 224 с.
20. *Zemskov A.V., Tarlakovskii D.V.* Modelling of unsteady elastic diffusion oscillations of a Timoshenko beam // Nonlinear Wave Dynamics of Materials and Structures. Advanced Structured Materials, V. 122. Springer Nature Switzerland AG 2020. P. 447–461.
21. *Вестяк А.В., Земсков А.В.* Модель нестационарных упругодиффузионных колебаний шарнирно закрепленной балки Тимошенко // Известия Российской академии наук. Механика твердого тела. 2020. № 5. С. 107–119. https://doi.org/10.31857/S0572329920030174
22. *Земсков А.В., Тарлаковский Д.В., Файкин Г.М.* Нестационарный изгиб консольно-закрепленной балки Бернулли–Эйлера с учетом диффузии // Вычисл. механ. сплошных сред. 2021. Т. 14. № 1. С. 40–50.
23. *Zenkour A.M.* Thermoelastic diffusion problem for a half-space due to a refined dual-phase-lag Green-Naghdi model // Journal of Ocean Engineering and Science. 2020. V. 5. № 3. P. 214–222. https://doi.org/10.1016/j.joes.2019.12.001
24. *Ailawaliar P., Budhiraja S.* Dynamic Problem in Thermoelastic Solid Using Dual-Phase-Lag Model with Internal Heat Source // J. of Math. Sci. and App. 2014. V. 2. № 1. P. 10–16.

25. *Формалев В.Ф.* Теплоперенос в анизотропных твердых телах. Численные методы, тепловые волны, обратные задачи. М.: ФИЗМАТЛИТ, 2015. 280 с.
26. *Abbas A.I.* The effect of thermal source with mass diffusion in a transversely isotropic thermoelastic infinite medium // J. of Measurements in Engng. 2014. V. 2. № 4. P. 175–184.
27. *Davydov S.A., Zemskov A.V.* Thermoelastic Diffusion Phase-Lag Model for a Layer with Internal Heat and Mass Sources // Internat. Journal of Heat and Mass Transfer. 2022. V. 183. Part C. 122213. https://doi.org/10.1016/j.ijheatmasstransfer.2021.122213
28. *Диткин В.А., Прудников А.П.* Справочник по операционному исчислению. М.: Высшая школа, 1965. 568 с.
29. *Прудников А.П., Брычков Ю.А., Маричев О.И.* Интегралы и ряды. Том 1. Элементарные функции. М.: Наука, 1981. 797 с.
30. *Тимошенко С.П.* Сопротивление материалов. Часть 1. Элементарная теория и задачи. М.: Наука. Главная редакция физико-математической литературы. 1965. 364 с.
31. Физические величины: Справочник *Бабичев А.П., Бабушкина Н.А., Братковский А.М., и др.* Под общей редакцией Григорьева И.С., Мейлихова И.З. М.: Энергоатомиздат, 1991. 1232 с.

EDN: DYYLSB
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1912–1926

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 533.6.011

ОБТЕКАНИЕ ПРЯМОУГОЛЬНОГО ЦИЛИНДРА ДОЗВУКОВЫМ ПОТОКОМ РАЗРЕЖЕННОГО ГАЗА[1)]

© 2022 г. О. И. Ровенская[1,*]

[1] 119333 Москва, ул. Вавилова, 40, ВЦ ФИЦ ИУ РАН, Россия

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

Поступила в редакцию 23.12.2020 г.
Переработанный вариант 12.12.2021 г.
Принята к публикации 07.06.2022 г.

Исследуется обтекание дозвуковым потоком разреженного газа прямоугольного цилиндра бесконечного размаха, используя численный метод, основанный на решении S-модельного кинетического уравнения. Изучается влияние числа Рейнольдса Re_∞ в диапазоне от 10 до 200 на возникающее вокруг прямоугольного цилиндра поле течения. При $\mathrm{Re}_\infty = 200$ исследуется влияние геометрии цилиндра на поле течения, изменяя соотношение между высотой цилиндра и его длиной AR от 1 до 8. Полученные результаты представлены в виде коэффициентов сопротивления, подъемной силы, давления и числа Струхаля. Картины течения в области за цилиндром демонстрируют возникновение рециркуляционной зоны за цилиндром, размер и форма которой зависят как от числа Рейнольдса Re_∞, так и от соотношения AR. Обнаружено, что при стационарном течении коэффициенты, характеризующие течение, сильно зависят от числа Рейнольдса. В то же время данная зависимость становится слабее, когда течение становится нестационарным. С увеличением соотношения AR зона рециркуляции за цилиндром сужается, что приводит к уменьшению коэффициента сопротивления. Кроме того, оценивается надежность используемого подхода для решения подобного класса задач путем сравнения полученных результатов с данными, приведенными в литературе. Библ. 23. Фиг. 15. Табл. 3.

Ключевые слова: прямоугольный цилиндр, внешнее течение, кинетические уравнения.

DOI: 10.31857/S0044466922110102

1. ВВЕДЕНИЕ

Течение газа вокруг различных элементов конструкций является неотъемлемой частью приложений в авиастроении, ветроэнергетике, электротехнике (охлаждение). Картина течения вокруг таких тел зависит от числа Рейнольдса, соотношения между его сторонами и его ориентации по отношению к набегающему потоку (см. [1]–[15]).

Картина течения около прямоугольного цилиндра близка к картине, формирующейся при обтекании кругового цилиндра (см. [16]). Однако механизм отрыва потока и возникающая зависимость подъемной силы, сопротивления и числа Струхаля от числа Рейнольдса существенно отличаются для потока, обтекающего прямоугольный цилиндр. Для прямоугольного цилиндра точки отрыва потока возникают либо на его передней кромке, либо на задней кромке, в зависимости от значения числа Рейнольдса. Кроме того, область рециркуляции за прямоугольным цилиндром значительно шире и длиннее по сравнению с круговым (см. [9], [16], [17]).

В зависимости от значения числа Рейнольдса течение около прямоугольного цилиндра демонстрирует несколько основных картин: (1) – ползущее ламинарное устойчивое течение без отрыва; (2) – устойчивое течение с отрывом, возникающим на задней кромке цилиндра (при этом за цилиндром образуются присоединенные вихри); (3) – нестационарный срыв потока с задней кромки цилиндра с вихреобразованием; (4) – отрыв потока на передней кромке цилиндра и его присоединение к боковым граням цилиндра; (5) – отрыв потока на передней кромке без присоединения к граням цилиндра. В данной работе рассматриваются (1)–(3) режимы течения.

[1)]Работа выполнена при финансовой поддержке РФФИ (проект № 18-01-00899).

В настоящее время работы, посвященные исследованию обтекания тел плохо обтекаемой формы, особенно течений, около круговых цилиндров широко представлены в литературе (см., например, [16]). Меньшее внимание уделяется обтеканию прямоугольного цилиндра, хотя такие течения представляют большой интерес, как с практической, так и с академической точек зрения. Поэтому изучение поведения потока около прямоугольного цилиндра и аккуратное предсказание его свойств являются актуальной проблемой.

Экспериментальные работы для определения аэродинамических нагрузок на цилиндры различного поперечного сечения и частоты формирования вихрей в следе за цилиндром проводились в аэродинамических трубах или водных каналах. В некоторых работах измерения сочетаются с визуализацией потока, обеспечивающей феноменологическое понимание движения потока в окрестности тела (см. [1]–[3]).

Несмотря на геометрическую простоту задачи, численное моделирование такого течения не является тривиальной задачей, что обусловлено рядом факторов: наличием неблагоприятных градиентов давления, зон отрыва и присоединения потока, рециркуляционных областей, сильно изогнутых линий течения и вихревых взаимодействий. Следует отметить, что в большей части работ, посвященных моделированию обтекания прямоугольного цилиндра, применяются сжимаемые и несжимаемые уравнения Навье–Стокса или метод моделирования больших вихрей (LES) с использованием классической подсеточной модели Смагоринского (см., например, [4]–[15]).

В [4] для чисел Рейнольдса $\mathrm{Re} \geq 300$ и соотношении сторон цилиндра от 1 до 10 с помощью метода моделирования больших вихрей (LES) были исследованы картины течения и периодические особенности следа, возникающие за прямоугольным цилиндром. Изучены важные физические механизмы, определяющие частоту возникновения вихрей. Несжимаемые уравнения Навье–Стокса применялись для численного исследования двумерного течения около квадратного цилиндра для стационарного и нестационарного режимов течения в [5]. Анализ влияния числа Рейнольдса и угла атаки на силы, воздействующие на квадратный цилиндр, выполнен в [6] также с использованием несжимаемых уравнений Навье–Стокса.

В настоящей работе представлены результаты численного исследования нестационарного течения вокруг прямоугольного цилиндра с соотношением сторон от 1 до 8 и числе Рейнольдса в диапазоне от 10 до 200, используя S-модельное кинетическое уравнение (см. [18]). Прямой численный метод решения S-модельного уравнения был оптимизирован с помощью MPI (Message Passing Interface) для расчетов на кластере МВС-100К (см. [19]). Изучены закономерности течения и важные физические механизмы, определяющие частоту вихреобразования. При проектировании конструкций в виде прямоугольного цилиндра необходимо учитывать силы, вызванные возникающими вихревыми структурами, поэтому в работе рассчитаны и проанализированы коэффициенты сопротивления, подъема, давления и число Струхаля.

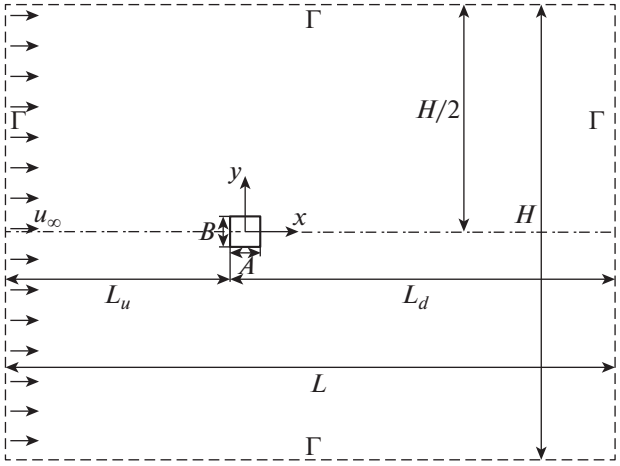
Следует отметить, что используемый численный метод уже эффективно применялся для моделирования как дозвуковых, так и сверхзвуковых течений разреженного газа в [20] и [21]. Однако для оценки надежности, используемого подхода для решения рассматриваемой задачи, выполнено также сравнение полученных результатов с данными, приведенными в литературе.

2. ПОСТАНОВКА ЗАДАЧИ

Рассматривается обтекание прямоугольного цилиндра длины A и высоты B с соотношением сторон $AR = A/B$ равномерным дозвуковым потоком моноатомного газа со скоростью u_∞ для значений числа Рейнольдса Re_∞ в диапазоне от 10 до 200. Расчетная область вокруг цилиндра, схематично показанная на фиг. 1, представляет собой прямоугольник размера $L \times H$. Размер расчетной области вверх по потоку до цилиндра L_u задается равной $30B$, за цилиндром вниз по потоку $L_d = 50B$. Полный продольный размер области равен $L = L_u + L_d = 80B$, а поперечный размер – $H = 60B$. Размер расчетной области был выбран таким образом, чтобы минимизировать влияние граничных условий на течение вокруг цилиндра (см. п. 3.1).

В качестве основных количественных характеристик, позволяющих оценить степень воздействия потока на прямоугольный цилиндр, были выбраны коэффициенты сопротивления, подъемной силы и давления:

$$C_D = \frac{F_D}{0.5\rho u_\infty^2 B}, \quad C_L = \frac{F_L}{0.5\rho u_\infty^2 B}, \quad C_P = \frac{p - p_\infty}{0.5\rho u_\infty^2}, \tag{1}$$



Фиг. 1. Геометрия расчетной области.

где F_D и F_l – сила сопротивления и подъемная сила, действующие на цилиндр, p – давление на поверхности цилиндра, p_∞ и u_∞ – давление и скорость набегающего потока.

Одной из важных характеристик нестационарного течения является частота срыва вихрей, в общем случае характеризующаяся безразмерным параметром – числом Струхаля ($\mathrm{St} = fB/u_\infty$). Для вычисления числа Струхаля использовалось быстрое преобразование Фурье временного ряда для коэффициента подъемной силы C_L.

Численный метод основан на решении S-модельного кинетического уравнения (см. [18]), которое можно записать как

$$\frac{\partial f}{\partial t} + \boldsymbol{\xi}\frac{\partial f}{\partial \mathbf{x}} = J_S(f,f), \tag{2}$$

$$J_S(f,f) = \frac{p}{\mu}\left(M\left(1 + \frac{2m^2\mathbf{qc}}{15\rho(kT)^2}\left(\frac{m\mathbf{c}^2}{2kT} - \frac{5}{2}\right)\right) - f\right) = \frac{p}{\mu}(S - f), \tag{3}$$

$$M(\rho, \mathbf{V}, T) = \frac{\rho}{(2\pi RT)^{3/2}}\exp\left(-\frac{\mathbf{c}^2}{2RT}\right). \tag{4}$$

Здесь $f = f(t, \mathbf{x}, \boldsymbol{\xi})$ – функция распределения молекул по скоростям, т.е. вероятность обнаружить частицу со скоростью $\boldsymbol{\xi} = (\xi_x, \xi_y, \xi_z) \in \mathrm{R}^3$ в точке двумерного пространства $\mathbf{x} = (x, y)$ в момент времени t. $S(\rho, \mathbf{c}, T)$ и $M(\rho, \mathbf{c}, T)$ – стандартные локальные функции распределения Шахова и Максвелла, $\mathbf{c} = \boldsymbol{\xi} - \mathbf{V}$ – собственная скорость молекул газа массы m, $\mathbf{V} = (u, v)$ – массовая скорость газа, $\mathbf{q}$ – вектор теплового потока, ρ, p и μ – плотность, давление и вязкость газа при температуре T, $\mathrm{R} = k/m$ – универсальная газовая постоянная, k – постоянная Больцмана.

В кинетическом подходе макровеличины, такие как плотность ρ, импульс $\rho\mathbf{V}$, температура T и вектор теплового потока $\mathbf{q}$ вычисляются интегрированием по всему скоростному пространству $R^3 \in [-\infty; \infty]$ в момент времени t:

$$\begin{gathered}\rho(\mathbf{x}) = \int f(\mathbf{x}, \boldsymbol{\xi})d\boldsymbol{\xi},\\ \rho(\mathbf{x})(u(\mathbf{x}), v(\mathbf{x}))^T = \int (\xi_x, \xi_y)^T f(\mathbf{x}, \boldsymbol{\xi})d\boldsymbol{\xi},\\ T(\mathbf{x}) = \frac{m}{3\rho(\mathbf{x})k}\int \mathbf{c}^2 f(\mathbf{x}, \boldsymbol{\xi})d\boldsymbol{\xi},\\ (q_x(\mathbf{x}), q_y(\mathbf{x}))^T = \frac{m}{2}\int (c_x, c_y)^T \mathbf{c}^2 f(\mathbf{x}, \boldsymbol{\xi})d\boldsymbol{\xi}.\end{gathered} \tag{5}$$

Начальное состояние потока одноатомного газа описывается функцией распределения Максвелла $M(\rho, \mathbf{c}, T)$ с плотностью $\rho = \rho_\infty$, температурой $T = T_\infty$, компонентами скорости $u = u_\infty$ и $v = 0$, где u_∞ – скорость набегающего потока.

На достаточно удаленном от цилиндра контуре Г (фиг. 1) для молекул, скорости которых направлены внутрь области (к цилиндру), задается функция распределения Максвелла $M(\rho_\infty, u_\infty, T_\infty)$, соответствующая однородному набегающему потоку со скоростью u_∞. На поверхности цилиндра задается диффузное отражение молекул с полной тепловой и импульсной аккомодацией:

$$f_w(\mathbf{x}, \boldsymbol{\xi}) = \kappa(\mathbf{x}) M(1, u_w, T_w), \tag{6}$$

где u_w и T_w – скорость и температура цилиндра, здесь $u_w = 0$, $T_w = T_\infty$. Плотность частиц, отраженных от поверхности цилиндра $\kappa(\mathbf{x})$, вычисляется из условия равенства падающего и отраженного от поверхности потоков (см. [22], [23]):

$$\kappa(\mathbf{x}) = -\frac{\int\limits_{\xi\cdot\vartheta(\mathbf{x})<0} f(t, \mathbf{x}, \xi)\xi \cdot \vartheta(\mathbf{x}) d\xi}{\int\limits_{\xi\cdot\vartheta(\mathbf{x})>0} M(1, u_w, T_w)\xi \cdot \vartheta(\mathbf{x}) d\xi}. \tag{7}$$

Уравнение (2) обезразмеривается с помощью параметров однородного набегающего потока газа: ρ_∞ и T_∞ и высоты цилиндра B. В работе используется модель твердых шаров, в этом случае безразмерная вязкость вычисляется как $\mu = T^{0.5}$.

Так как рассматриваемое течение является двумерным, можно аналитически исключить z-составляющую молекулярной скорости ξ_z. Для получения численного решения сначала решается уравнение переноса (левая часть уравнения (2)), а затем уравнение релаксации (столкновения) (правая часть уравнения (2)). Уравнение переноса аппроксимируется схемой второго порядка точности с использованием стандартного метода TVD с функцией ограничения потока minmod (см. [20]). Однородный этап столкновений между частицами аппроксимируется явно-неявным подходом (см. [20]).

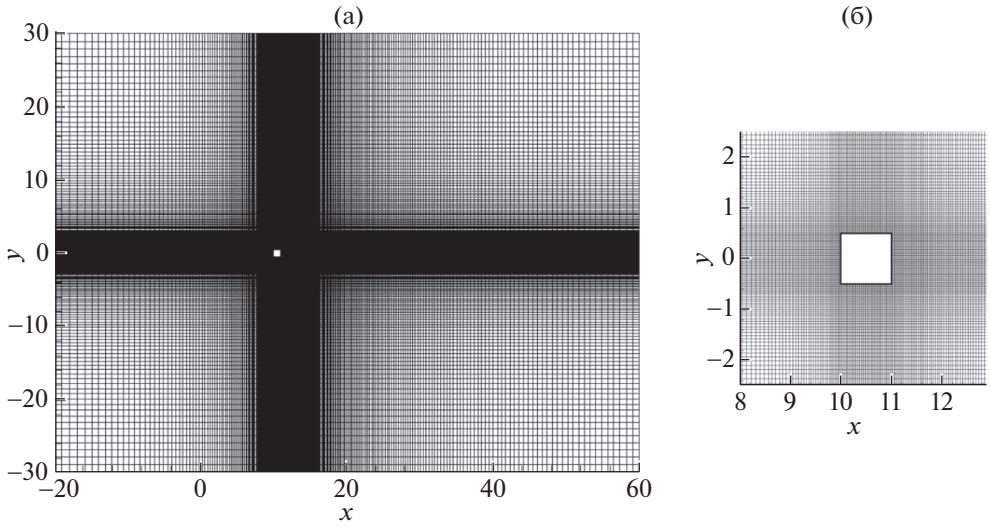
Для повышения эффективности расчетов программный код был распараллелен с помощью MPI (message passing protocol) (см. [19]). Каждому процессору присваивался свой собственный набор точек сетки в физическом пространстве. Этап релаксации (столкновения между молекулами) выполнялся независимо на различных процессорах. Перед этапом переноса процессоры обменивались данными между соседними сеточными узлами. Расчеты были выполнены на кластере MVS-100K.

3. ПАРАМЕТРЫ МОДЕЛИРОВАНИЯ

Численное моделирование обтекания прямоугольного цилиндра под нулевым углом атаки дозвуковым потоком моноатомного газа проводилось для числа Маха $\mathrm{Ma}_\infty = 0.2$, что соответствует скорости набегающего потока $u_\infty = 0.18257$, и изменении числа Рейнольдса Re_∞ от 10 до 200, т.е. изменении числа Кнудсена Kn_∞ в диапазоне от 0.033 до 0.00165. Также изучалось влияние геометрии прямоугольного цилиндра на поле течения, увеличивая соотношение его сторон AR от 1 до 8 при фиксированном значении числа Рейнольдса $\mathrm{Re}_\infty = 200$.

Для расчетов в физическом пространстве используется неоднородная, структурированная, криволинейная сетка с локальным измельчением в окрестности цилиндра. Типичная сетка, используемая в расчетах, а также ее увеличенное изображение в окрестности цилиндра показаны на фиг. 2. В табл. 1 приведены соотношение сторон цилиндра AR, соответствующий размер расчетной сетки (число узлов сетки, заданные в направлении по потоку и в поперечном направлении), число узлов сетки N_x вдоль длины цилиндра A и число узлов N_y вдоль его высоты B. Минимальный шаг сетки вдоль поверхности цилиндра составляет порядка 0.025 B и является компромиссом между требуемым разрешением течения и желаемой скоростью вычисления.

В пространстве скоростей задается однородная двумерная сетка. Размер скоростной сетки удовлетворяет условию $v_{\max} \geq \max(|u|,|v|) + 3.5T_{\max}^{0.5}$. Для большинства расчетов число узлов по каждому направлению скорости выбирается равным 24, а размер скоростного пространства ограничивается максимальным значением скорости $v_{\max} = 5$. Оптимальное число узлов в ско-



Фиг. 2. Сетка размера 344 × 248, используемая в расчетах: (а) – полный вид, (б) – ее увеличенное изображение в окрестности цилиндра.

ростном пространстве выбиралось таким образом, что удваивание их числа не изменяет значения основных характеристик течения (коэффициентов C_D, C_L и St) более чем на 1%.

Шаг по времени Δt удовлетворяет условию устойчивости Куранта–Фридрихса–Леви (CFL), где $CFL = 0.4$:

$$\Delta t = CFL/\max_{ij}\left(v_{\max}/\Delta x + v_{\max}/\Delta y\right). \tag{8}$$

3.1. Влияние размера расчетной области

В задачах с нестационарным следом и конвективными вихрями существенное влияние на точность расчетов оказывает размер расчетной области, в частности длина расчетной области за цилиндром L_d. Недостаточный размер области вниз по потоку может привести к искажению крупномасштабных структур из-за наличия конвективных вихрей.

Влияние размера расчетной области на поле течения оценивалось с помощью сравнения усредненного по времени коэффициента сопротивления $C_{D\,av}$, среднеквадратичного коэффициента подъемной силы $C_{L\mathrm{rms}}$ и значения числа Струхаля St, полученных для четырех последовательно увеличивающихся размеров расчетной области $L \times H$: 40 × 30, 60 × 45, 80 × 60 и 107 × 80, в которой расположен квадратный цилиндр ($AR = 1$). Число Рейнольдса Re_∞ фиксировано и равно 100. Полученные результаты приведены в табл. 2. В скобках указано изменение значений коэффициентов $C_{D\,av}$, $C_{L\mathrm{rms}}$ и числа Струхаля St в процентах.

Как следует из табл. 2, вариация значений коэффициентов $C_{D\,av}$, $C_{L\mathrm{rms}}$ и числа Струхаля St уменьшается с увеличением размера расчетной области. При переходе от размера 80 × 60 к размеру 107 × 80, расхождение между значениями коэффициентов становится меньше 1%.

Таблица 1. Геометрия цилиндра и размерность сетки

AR	Сетка	N_x	N_y
1	360 × 192	40	40
2	360 × 192	48	40
4	392 × 192	80	40
6	448 × 192	128	40
8	512 × 192	192	40

Таблица 2. Влияние размера расчетной области для $\mathrm{Re}_\infty = 100$

$L \times H$	$C_{D\,av}$	$C_{L\mathrm{rms}}$	St
40×30 ($L_u = 15$, $L_d = 25$)	1.425	0.123	0.142
60×45 ($L_u = 22.5$, $L_d = 37.5$)	1.439 (0.94)	0.149 (21)	0.149 (5)
80×60 ($L_u = 30$, $L_d = 50$)	1.45 (0.76)	0.154 (3.3)	0.1515 (1.68)
107×80 ($L_u = 40$, $L_d = 67$)	1.455 (0.34)	0.155 (0.65)	0.1527 (0.8)

Таблица 3. Влияние параметров сетки для расчетной области $L \times H = 80 \times 60$ ($L_u = 30$, $L_d = 50$) и $\mathrm{Re}_\infty = 100$

Сетка	$C_{D\,av}$	$C_{L\mathrm{rms}}$	St
172×124	1.41	0.15	0.1480
344×248	1.45 (2.8)	0.154 (2.6)	0.1515 (2.4)
516×372	1.46 (0.69)	0.155 (0.65)	0.1522 (0.46)

Таким образом, для последующих расчетов использовалась область размера $L \times H = 80 \times 60$, в этом случае размер области до цилиндра $L_u = 30$ и за цилиндром – $L_d = 50$. Заметим, что вариация значений коэффициента $C_{D\,av}$ оставалась незначительной при изменении размера расчетной области.

3.2. Влияние измельчения сетки на поле течения

Степень измельчения сетки также может оказывать влияние на результаты расчетов. Оценка влияния измельчения сетки проводилась для случая квадратного цилиндра ($AR = 1$) и фиксированного числа Рейнольдса $\mathrm{Re}_\infty = 100$. Для этого сравнивались значения усредненного по времени коэффициента сопротивления $C_{D\,av}$, среднеквадратичного коэффициента подъемной силы $C_{L\mathrm{rms}}$ и значения числа Струхаля St, полученные на трех последовательно измельченных сетках следующего размера: 172×124, 344×248 и 516×372. При этом число узлов сетки, равномерно распределенных вдоль стороны цилиндра, было равно 20, 40 и 70 соответственно. Полученные на разных сетках значения коэффициентов $C_{D\,av}$, $C_{L\mathrm{rms}}$ и числа Струхаля St суммируются в табл. 3. Изменение значений коэффициентов при изменении степени измельчения сетки указано в процентах внутри скобок.

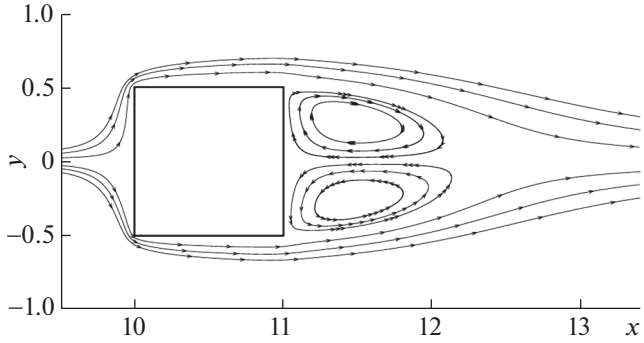
Как видно из табл. 3, вариация значений коэффициентов $C_{D\,av}$, $C_{L\mathrm{rms}}$ и числа Струхаля St не превышает 3% при переходе от самой грубой сетки к промежуточному уровню измельчения. При дальнейшем уточнении сетки расхождение между значениями коэффициентов уменьшается и становится меньше 1%. Таким образом, для последующих расчетов использовалась сетка размера 344×248.

4. АНАЛИЗ ЧИСЛЕННЫХ РЕЗУЛЬТАТОВ

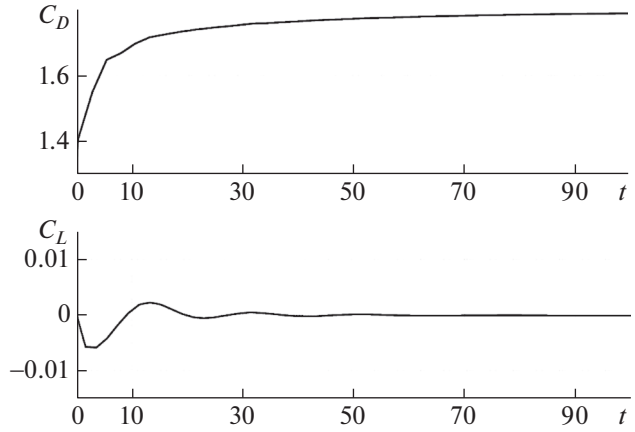
4.1. Квадратный цилиндр, $AR = 1$: $\mathrm{Re}_\infty \le 40$

Картина течения около квадратного цилиндра сильно зависит от значения числа Рейнольдса Re_∞. При малых Re_∞ течение является стационарным и в следе за цилиндром образуется устойчивая картина течения, состоящая из двух симметричных вихрей. На фиг. 3 при $\mathrm{Re}_\infty = 20$ отчетливо видна пара вращающихся против часовой стрелки вихрей, присоединенных к задней грани цилиндра и мешающих потоку газа за цилиндром. Эволюция соответствующих коэффициентов сопротивления C_D и подъемной силы C_L показана на фиг. 4.

На фиг. 5 показана полученная зависимость размера зоны рециркуляции за цилиндром L_r от значения числа Рейнольдса Re_∞ в сравнении с результатами [5] и [6]. Размер зоны рециркуляции за цилиндром L_r определялся как продольное расстояние от задней грани цилиндра до точки присоединения потока вдоль осевой линии следа. В [3] для стационарного течения, $\mathrm{Re}_\infty \le 40$,



Фиг. 3. Мгновенные линии тока для $\mathrm{Re}_\infty = 20$.



Фиг. 4. Изменение коэффициентов сопротивления и подъемной силы C_D и C_L во времени.

около квадратного цилиндра предложена эмпирическая зависимость размера зоны рециркуляции L_r от числа Рейнольдса Re:

$$L_r = 0.0672\,\mathrm{Re}. \tag{9}$$

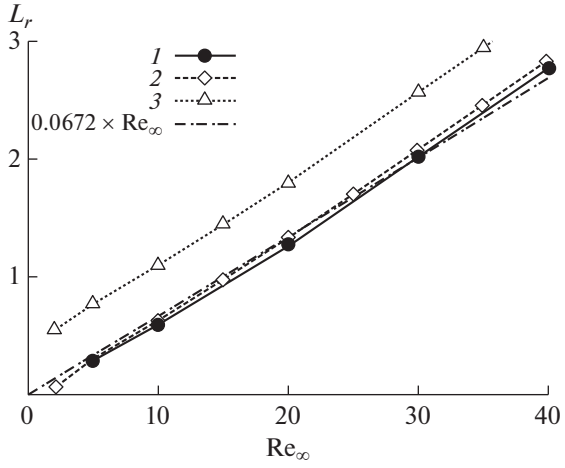
Как видно на фиг. 5, размер вихрей, образующихся за цилиндром, растет вместе с увеличением числа Рейнольдса Re_∞, хотя течение в следе остается устойчивым. В то же время (фиг. 4) коэффициенты сопротивления C_D и подъемной силы C_L стремятся к постоянным значениям. Кроме того, полученная зависимость размера зоны рециркуляции L_r от Re хорошо согласуется с результатами [5], [6] и эмпирической зависимостью (9).

Несмотря на то что в [5] и [6] для исследования течения около квадратного цилиндра применялись несжимаемые уравнения Навье–Стокса, в [5] и [6] использовались разные численные методы для их решения. Поэтому наблюдаемое отклонение полученных значений размера зоны L_r от данных [6] может быть связано с выбранным в [6] численным методом решения.

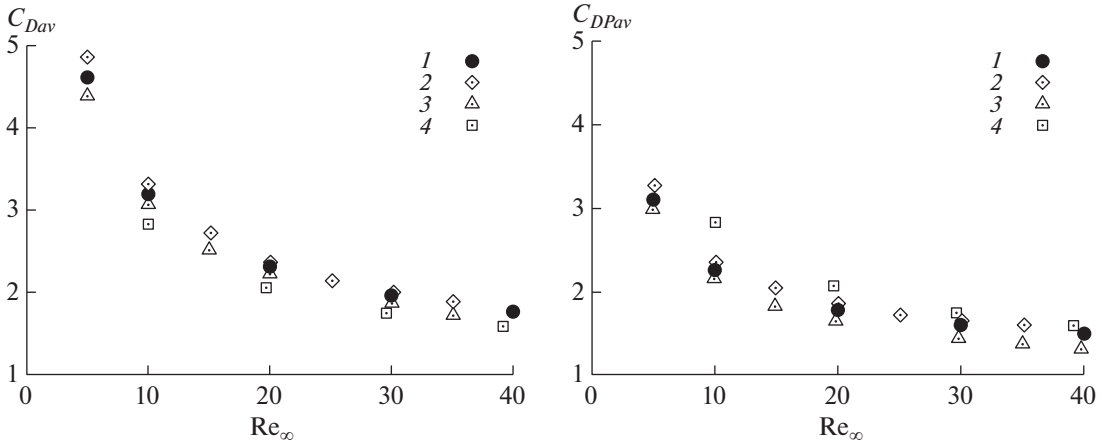
Полное сопротивление, которое испытывает цилиндр, включает в себя сопротивления, вызванные давлением и вязкими эффектами. Вязкое сопротивление создается трением между жидкостью и поверхностью, вдоль которой она течет. В то время как сопротивление давления возникает вследствие несимметричного распределения давления на верхней и нижней гранях цилиндра. Таким образом, коэффициент полного сопротивления может быть записан как

$$C_D = C_{Dv} + C_{DP}, \tag{10}$$

где C_{Dv} – коэффициент вязкостного сопротивления $C_{Dv} = F_{Dv}/0.5\rho u_\infty^2 B$, C_{DP} – коэффициент сопротивления давления $C_{DP} = F_{DP}/0.5\rho u_\infty^2 B$, F_{Dv} и F_{DP} – сила вязкого сопротивления и сила сопро-



Фиг. 5. Зависимость размера зоны рециркуляции L_r от числа Рейнольдса Re_∞ для устойчивого течения: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [6].

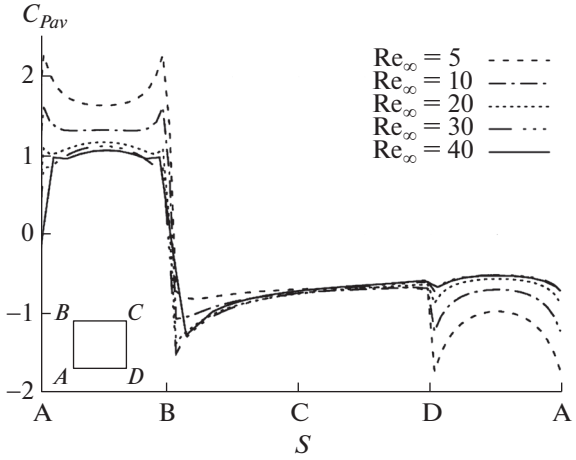


Фиг. 6. Зависимость усредненного коэффициента полного сопротивления $C_{D\,av}$ (справа) и усредненного коэффициента сопротивления давления $C_{DP\,av}$ от числа Рейнольдса Re_∞: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [6], *4* – из [7].

тивления давления, полученные интегрированием касательных и нормальных напряжений вдоль поверхности цилиндра соответственно.

На фиг. 6 видно, что усредненные по времени коэффициенты полного сопротивления $C_{D\,av}$ и сопротивления давления $C_{DP\,av}$ сильно зависят от числа Рейнольдса Re_∞ и уменьшаются с ростом Re_∞. При этом влияние на цилиндр вязкого сопротивления (коэффициент сопротивления C_{Dv}) значительно меньше, чем сопротивление давления (коэффициент C_{DP}) и уменьшается с увеличением Re_∞, однако, остается существенным в рассматриваемом диапазоне чисел Рейнольдса. Кроме того, полученные зависимости коэффициентов полного сопротивления $C_{D\,av}$ и сопротивления давления $C_{DP\,av}$ от числа Рейнольдса хорошо согласуются с результатами из [5]–[7].

Рассмотрим более детально распределение усредненного по времени коэффициента давления $C_{P\,av}$ вдоль поверхности цилиндра в случае стационарного течения (фиг. 7). Вблизи точки торможения потока, расположенной на передней кромке цилиндра A–B, коэффициент давления $C_{P\,av}$ падает с увеличением Re_∞, в то время как на задней кромке C–D (вблизи критической точки) коэффициент $C_{P\,av}$ растет с увеличением Re_∞.



Фиг. 7. Распределение усредненного по времени коэффициента давления вдоль поверхности цилиндра $C_{P\,av}$ при $\mathrm{Re}_\infty = 5$, 10, 20, 30 и 40.

В окрестности точки торможения поток разделяется вокруг прямоугольного цилиндра, что приводит к возникновению области высокого поверхностного давления на передней кромке цилиндра и низкого поверхностного давления на его задней кромке и, таким образом, к появлению значительного сопротивления давления. Поэтому коэффициент сопротивления $C_{D\,av}$, показанный на фиг. 6, уменьшается с увеличением Re_∞ в стационарном режиме.

4.2. Квадратный цилиндр, AR = 1: 50 ≤ Re_∞ < 200

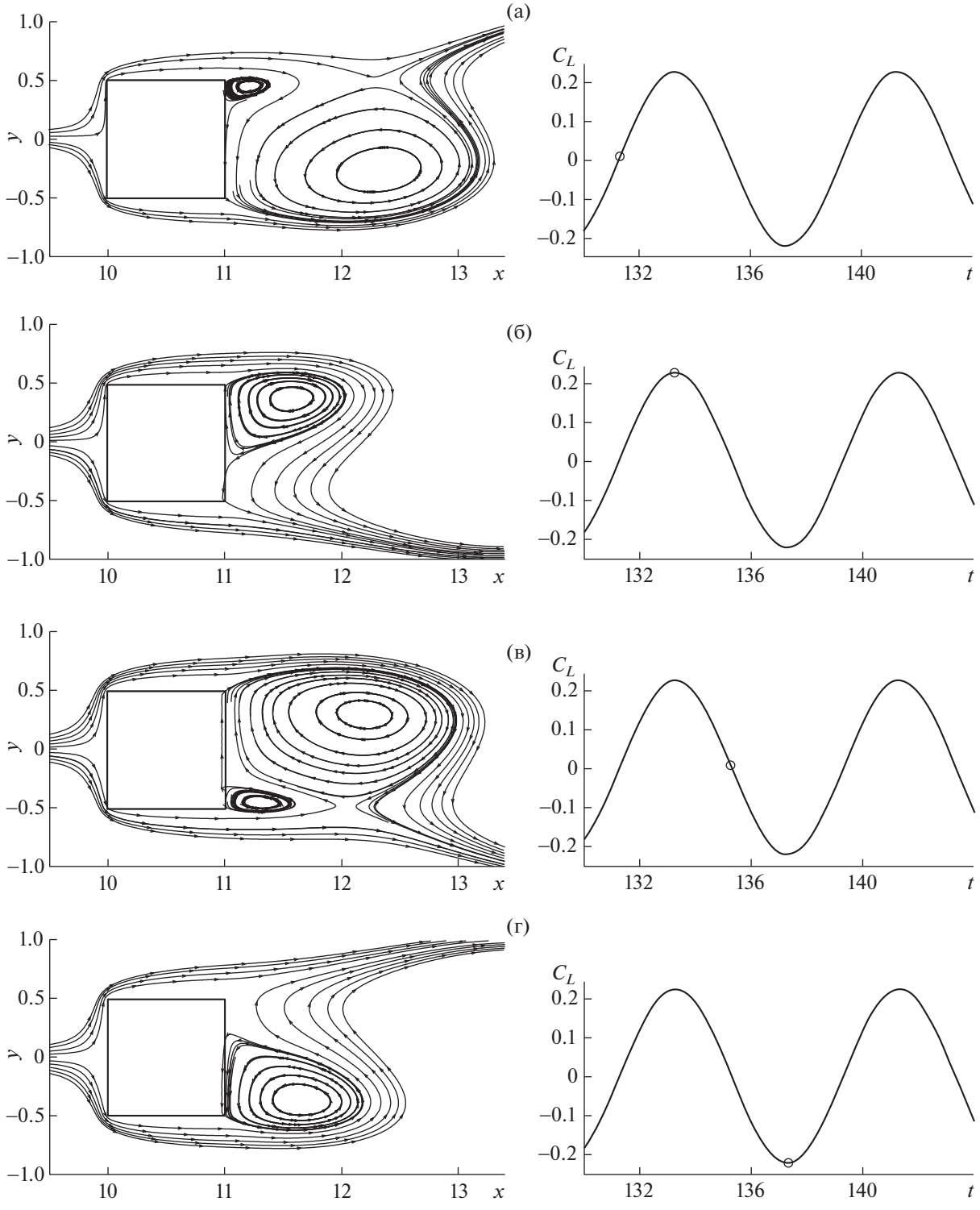
При дальнейшем увеличении числа Рейнольдса, $\mathrm{Re}_\infty \geq 50$, течение в следе за цилиндром становится неустойчивым и возникает процесс вихреобразования, приводящий к формированию вихревой дорожки, состоящей из последовательности вихрей, уносимых потоком от тела (дорожка Кармана). Мгновенные линии тока, демонстрирующие типичный процесс срыва вихрей для одного периода вихреобразования T, показаны на фиг. 8 для $\mathrm{Re}_\infty = 100$. На фиг. 8 можно наблюдать чередование основных вихрей, вращающихся в противоположном направлении. Два основных вращающихся в противоположных направлениях вихря, соответствующих 1/4 и 3/4 периодам вихреобразования T, показаны на фиг. 8а и 8в.

Коэффициент подъемной силы C_L, показанный на фиг. 8 справа, изменяется по синусоидальному закону. В нестационарном режиме течения гармоническое колебание подъемной силы является результатом срыва вихря из-за движения вихря от нижней грани цилиндра к его верхней грани и обратно. Периодический срыв вихрей с поверхности цилиндра вызывает периодическое изменение давления на цилиндре и в следе за цилиндром образуется дорожка вихрей (дорожка Кармана).

Распределение усредненного по времени коэффициента давления $C_{P\,av}$ на поверхности цилиндра для нестационарного режима течения показано на фиг. 9. В случае нестационарного режима течения коэффициент давления $C_{P\,av}$, усредненный за период вихреобразования T, монотонно уменьшается с увеличением Re_∞ вдоль всей поверхности цилиндра. Падение коэффициента давления $C_{P\,av}$ на B–C–D или на C–D менее глубокое, чем в стационарном случае, что приводит к более слабому изменению значений коэффициента сопротивления $C_{D\,av}$ с увеличением Re_∞ (фиг. 10).

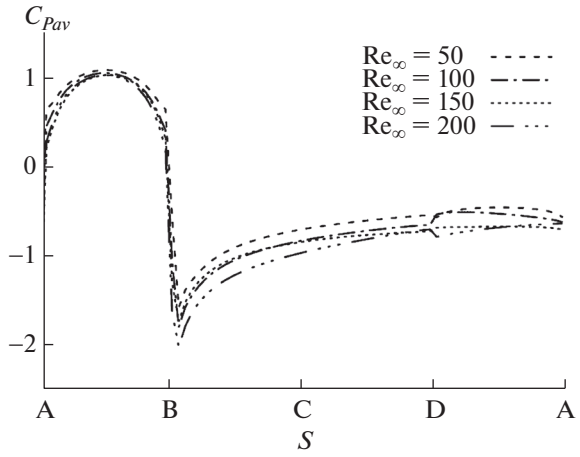
Среднеквадратичный коэффициент подъемной силы $C_{L\mathrm{rms}}$ позволяет оценить амплитуду колебаний в следе за цилиндром. Фигура 11 демонстрирует монотонный рост коэффициента $C_{L\mathrm{rms}}$ с увеличением числа Рейнольдса Re_∞.

Кроме того, как видно из фиг. 10 и 11, полученные значения коэффициентов полного сопротивления $C_{D\,av}$ и подъемной силы $C_{L\mathrm{rms}}$ достаточно хорошо воспроизводят поведение этих коэффициентов, представленное в литературе из [5]–[8].

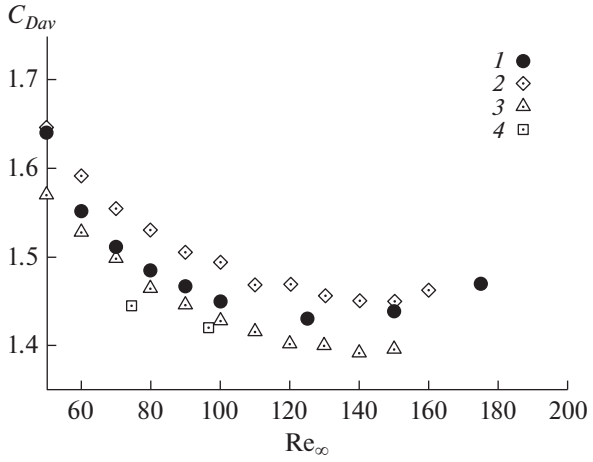


Фиг. 8. Мгновенные линии тока и изменение коэффициента подъемной силы для $\mathrm{Re}_\infty = 100$: (а) – $1/4T$, (б) – $2/4T$, (в) – $3/4T$, (г) – T, где T – период вихреобразования.

Зависимость размера зоны рециркуляции за цилиндром, полученной по усредненному по времени полю течения, от числа Рейнольдса Re_∞ показана на фиг. 12. Видно, что размер зоны рециркуляции монотонно убывает с ростом Re_∞ что согласуется с результатами, полученными в [5] и [9].



Фиг. 9. Распределение усредненного по времени коэффициента давления $C_{P\,av}$ вдоль поверхности цилиндра при $\mathrm{Re}_\infty = 50$, 100, 150 и 200.

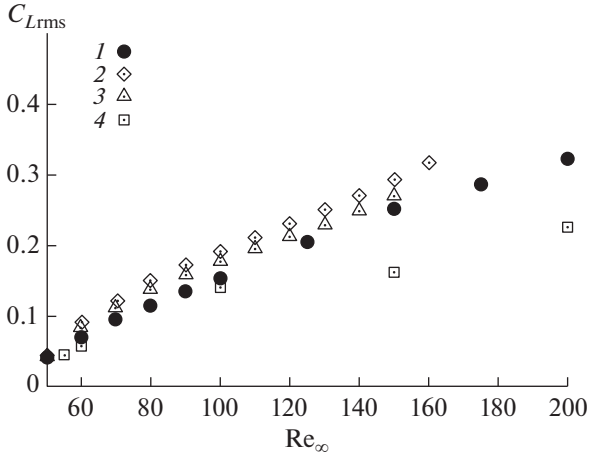


Фиг. 10. Зависимость усредненного по времени коэффициента полного сопротивления $C_{D\,av}$ от числа Рейнольдса Re_∞: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [6], *4* – из [7].

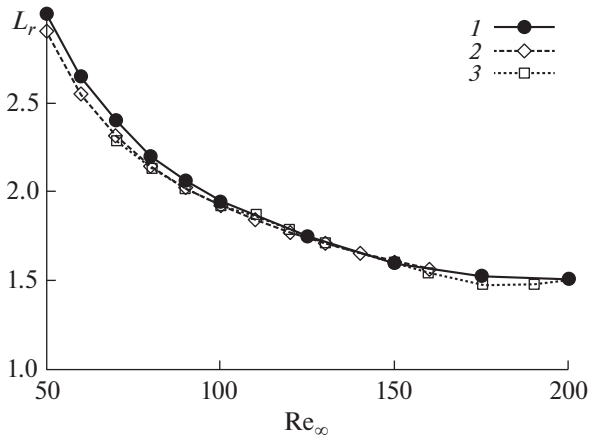
Чтобы оценить частоту вихреобразования, вычисляется число Струхаля St с помощью быстрого преобразования Фурье временного ряда для коэффициента подъемной силы C_L.

На фиг. 13 показано полученное число Струхаля St, а также соответствующие численные результаты из [5], [9] и экспериментальные данные из [8]. Можно видеть, что зависимость числа Струхаля St от числа Рейнольдса Re_∞, полученная в настоящей работе, хорошо воспроизводит тренд, показанный в [5], [9] и [8].

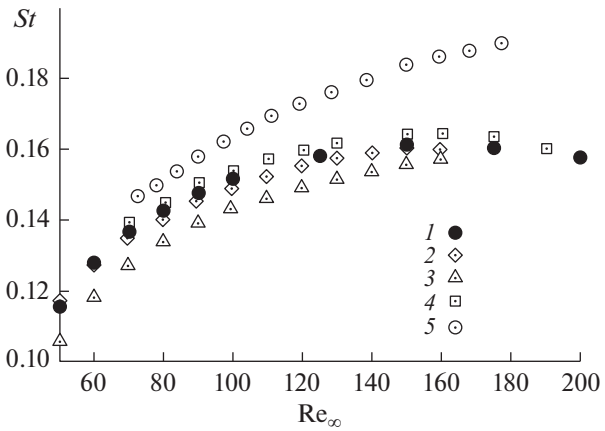
Кроме того, на фиг. 13 показано изменение числа Струхаля для кругового цилиндра (см. [17]). В отличие от поведения числа Струхаля для квадратного цилиндра, для кругового цилиндра число Струхаля монотонно возрастает. Различие между поведением числа Струхаля для кругового цилиндра и квадратного цилиндра может быть объяснено различной природой потоков, возникающих вокруг них. Квадратный цилиндр – менее обтекаемое тело, чем круговой цилиндр вследствие того, что передняя грань квадратного цилиндра более тупая, а также наличия острых передних кромок. Отрыв потока и появление областей рециркуляции сверху/снизу цилиндра приводят к увеличению эффективной ширины квадратного цилиндра. Следовательно, средний поток вверх по течению фактически обтекает тело с эффективным диаметром, большим, чем физический диаметр квадратного цилиндра. Поскольку число Струхаля прямо пропорциональ-



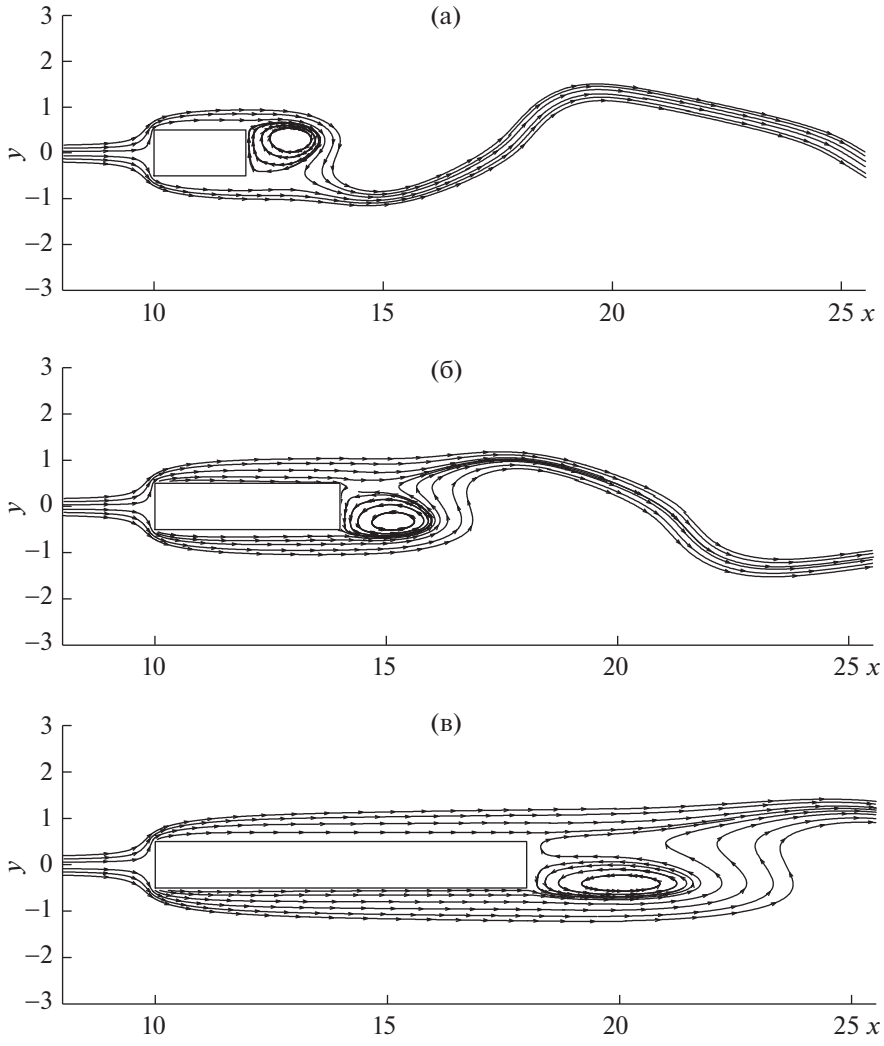
Фиг. 11. Зависимость коэффициента подъемной силы $C_{L\mathrm{rms}}$ от числа Рейнольдса Re_∞: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [6], *4* – из [8].



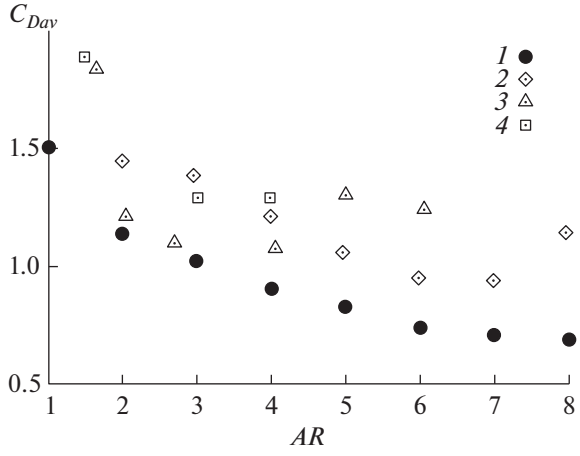
Фиг. 12. Зависимость размера зоны рециркуляции L_r от числа Рейнольдса Re_∞ для устойчивого течения: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [9].



Фиг. 13. Зависимость числа Струхаля от числа Рейнольдса Re_∞: *1* – результаты настоящей работы, *2* – результаты из [5], *3* – из [8], *4* – из [9], *5* – из [17].



Фиг. 14. Мгновенные линии тока при $\mathrm{Re}_\infty = 200$ для соотношений сторон цилиндра: (а) – $AR = 2$, (б) – $AR = 4$, (в) – $AR = 8$.



Фиг. 15. Зависимость усредненного по времени коэффициента полного сопротивления $C_{D\,av}$ от соотношения сторон цилиндра AR: *1* – результаты настоящей работы, *2* – результаты из [12], *3* – из [13], *4* – из [14].

но характерному размеру тела (высоте квадратного цилиндра B), использование эффективной высоты тела вместо высоты B приведет к увеличению числа Струхаля.

4.3. Влияние геометрии цилиндра

Геометрия цилиндра может оказывать существенное влияние на поле течения вокруг цилиндра. Для оценки данного эффекта в работе проведено исследование течения около прямоугольных цилиндров с соотношением сторон AR от 2 до 8 и фиксированном числе Рейнольдса $\mathrm{Re}_\infty = 200$.

На фиг. 14 показаны мгновенные линии тока около прямоугольного цилиндра с соотношениями сторон $AR = 2$, 4 и 8. Как и ожидалось, с задней кромки прямоугольного цилиндра возникает срыв потока. При этом увеличение AR приводит к сужению области вихревого следа и, следовательно, к уменьшению усредненного по времени коэффициента полного сопротивления $C_{D\,av}$ (фиг. 15), что согласуется с результатами из [12]–[14]. Некоторое отклонение значений полученного коэффициента $C_{D\,av}$ от данных [12]–[14] связано с тем, что в этих работах течение моделируется при существенно больших значениях числа Рейнольдса – порядка 10^3–10^4.

ВЫВОДЫ

В данной работе численно исследовалось обтекание дозвуковым потоком разреженного газа прямоугольного цилиндра бесконечного размаха, используя S-модельное кинетическое уравнение. Для обеспечения достоверности численных результатов проведен анализ влияния размера расчетной области и параметров сетки на получаемое поле течения.

В стационарном режиме течения $\mathrm{Re}_\infty \le 40$ наблюдалась сильная зависимость коэффициентов сопротивления и давления от значения числа Рейнольдса Re_∞, что согласуется с данными, представленными в литературе. Для нестационарного режима, $\mathrm{Re}_\infty \ge 50$, эта зависимость становилась слабее. С увеличением Re_∞ размер зоны рециркуляции увеличивался для стационарного режима, тогда как для нестационарного периодического режима размер зоны монотонно уменьшался. Геометрия цилиндра также влияла на картину течения: увеличение соотношения сторон цилиндра AR приводило к сужению зоны рециркуляции и к уменьшению коэффициента полного сопротивления.

Кроме того, сравнение с данными, приведенными в литературе, показало надежность используемого подхода, основанного на численном решении S-модельного уравнения, для задачи обтекания прямоугольного цилиндра дозвуковым потоком разреженного газа.

СПИСОК ЛИТЕРАТУРЫ

1. *Okajima A.* Strouhal numbers of rectangular cylinders // J. Fluid Mech. 1982. № 123. P. 379–398.
2. *Lindquist C.M.Sc.* Thesis, Universida de Estadual Paulista – UNESP, IlhaSolteira, Brazil, 2000.
3. *Zdravkovich M.M.* Smoke Observation of the Formation of a Karman Vortex Street // J. Fluid Mech. 1969. V. 37. P. 491–496.
4. *Almeida O., Mansur S.S., Silveira-Neto A.* On the flow past rectangular cylinders: physical aspects and numerical simulation // Therm. Eng. 2008. Iss. 7. P. 55–64.
5. *Sharma A., Eswaran V.* Heat and fluid flow across a square cylinder in the two-dimensional laminar flow regime // Numer. Heat Transfer. Part A. 2004. Iss. 45. P. 247.
6. *Yoon D.H., Yang K.-S., Choi C.-B.* Flow past a square cylinder with an angle of incidence// Phys. Fluids. 2010. Iss. 22. P. 043603.
7. *Okajima A., Yi D., Sakuda A., Nakano T.* Numerical study of blockage effects on aerodynamic characteristics of an oscillating rectangular cylinder // J. Wind Eng. Ind. Aerodyn. 1997. V. 67–68. P. 91–102.
8. *Sohankar A., Norberg C., Davidson L.* Low-Reynolds number flow around a square cylinder at incidence: study of blockage, onset of vortex shedding and outlet boundary condition // Inter. J. Numer. Methods Fluids. 1998. Iss. 26. P. 39.
9. *Robichaux J., Balachandar S., Vanka S.P.* Two-dimensional floquet instability of the wake of square cylinder// Phys. Fluids. 1999. V. 11. P. 560–578.
10. *Olawore A.S. 2D* Flow around a rectangular cylinder: A computational study // Inter. J. Sci. Technol. 2013. V. 2. Iss. 1. P. 1–26.
11. *Okajima A., Ueno H., Sakay H.* Numerical simulation of laminar and turbulent flows around rectangular cylinders// Inter. J. Num. Meth. Fluids. 1992. Iss. 15. P. 999–1012.

12. *Ying X., Xu F. Zhang Z.* Numerical simulation and visualization of flow around rectangular bluff bodies // Proc. of the 7th Inter. Colloquium on Bluff Body Aerodynamics and Applications (BBAA7) Shanghai, China, September 2–6, 2012, P. 272–281.
13. *Okajima A.* Numerical simulation of flow around rectangular cylinders // J. Wind Eng. Ind. Aerodyn. 1990. V. 33. P. 171–180.
14. *Yu D., Kareen A.* Parametric study of flow around rectangular prisms using LES // J. Ind. Aerodyn. 1998. V. 77–78. P. 653–662.
15. *Nakamura Y., Ohya Y., Ozono S., Nakayama R.* Reproducibility of flow past two-dimensional rectangular cylinders in a homogeneus turbulent flow by LES // J. Wind Eng. and Ind. Aerodyn. 1996. Iss. 65. P. 301–308.
16. *Zdravkovich M.M.* Flow Around circular cylinders. UK: Oxford Univ. Press Inc., 1997.
17. *Williamson C.H.K.* The existence of two stages in the transition to three-dimensionality of a cylinder wake // Phys. Fluids A. 1988. Iss. 31. P. 3165.
18. *Шахов Е.М.* Метод исследования движения разреженного газа. М.: Наука, 1974.
19. *Snir M., Dongarra J., Kowalik J.S., Huss-Lederman S., Otto S.W., Walker D.W.* MPI – The complete references. UK: Cambridge MIT Press, 2000.
20. *Aristov V.V., Rovenskaya O.I.* Kinetic description of the turbulence in the supersonic compressible flow over a backward/forward-facing step // Comput. Fluids. 2015. V. 111. P. 150–158.
21. *Rovenskaya O.I., Aristov V.V.* Kinetic simulation of a supersonic compressible flow over different geometry bodies // Eur. J. Mech. B Fluids. 2017. V. 64. P. 2017.
22. *Cercignani C.* Rarefied gas dynamics from basic concepts to actual calculations. UK: Cambridge Univ. Press; 1 ed., 2000.
23. *Rovenskaya O.I.* Numerical investigation of gas-surface scattering dynamics on the rarefied gas flow through a planar channel caused by a tangential temperature gradient // Inter. J. Heat and Mass Trans. 2015. V. 89. P. 1024–1033.

EDN: PSNFKI
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1927–1939

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.635

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ФИЛЬТРАЦИОННЫХ КОНЦЕНТРАЦИОННО-КОНВЕКТИВНЫХ ТЕЧЕНИЙ С КОНТРАСТОМ ВЯЗКОСТИ[1)]

© 2022 г. Е. Б. Соболева[1,*]

[1] 119526 Москва, пр-т Вернадского, 101, корп. 1, Институт проблем механики им. А.Ю. Ишлинского РАН, Россия

**e-mail: soboleva@ipmnet.ru*

Поступила в редакцию 22.11.2021 г.
Переработанный вариант 17.05.2022 г.
Принята к публикации 07.07.2022 г.

Выполняется численное моделирование концентрационно-конвективных течений в пористой среде с помощью двумерного кода, основанного на конечно-разностном методе. Гидродинамическая модель включает уравнения неразрывности, Дарси и переноса примеси. Модель описывает фильтрационные течения двухкомпонентной жидкой системы, состоящей из несжимаемой жидкости и растворенной примеси, которая распространяется за счет конвекции и диффузии. Получено численное решение задачи о развитии неустойчивости Рэлея–Тейлора в системе смешивающихся жидкостей разной вязкости. Рассмотрены системы с отношением коэффициентов вязкости слоев от 1 до 30. Исследовано влияние вязкости на структуру и интенсивность течения, на характеристики перемешивания. Библ. 36. Фиг. 6.

Ключевые слова: пористая среда, неустойчивость Рэлея–Тейлора, уравнение Дарси, уравнение конвекции-диффузии, контраст вязкости, конечно-разностный метод, схема Рунге–Кутты.

DOI: 10.31857/S0044466922110126

ВВЕДЕНИЕ

Изучение движения жидких субстанций в пористых средах актуально для решения технических, технологических и экологических задач. Знания о течениях нефти, воды, газа и их смесей через горные породы востребованы с точки зрения эффективного и рационального природопользования. Построение математических моделей для описания фильтрационных течений и развитие методов решения базовых уравнений составляют приоритетное направление исследований.

Наиболее часто используется представление гетерогенной многофазной среды как системы взаимодействующих взаимопроникающих континуумов [1], на базе которого развиваются гидродинамические модели разной степени сложности [2]–[4]. Описываются течения многокомпонентных реагирующих жидкостей и газов с фазовыми переходами в изотропных и анизотропных пористых средах с учетом инерционных эффектов, гидродинамической дисперсии и т.д. Решение сложных задач осуществляется численными методами с использованием коммерческих программных комплексов. Например, комплекс ANSYS FLUENT применялся для моделирования инфильтрации рассола из хранилища в окружающую пористую среду [5]. Имеются также отечественные некоммерческие комплексы программ, в частности MUFITS, с помощью которого изучалось вытеснение нефти водой и углеродным газом в подземных пористых пластах [6].

Настоящая работа посвящена исследованию фильтрационных течений подземных жидкостей с неоднородностями плотности под действием силы тяжести, которые с хорошей точностью описываются уравнением Дарси [2]–[4]. Благодаря тому что скорость таких течений невысока и число Рейнольдса, построенное по характерному размеру элемента твердой матрицы, много меньше единицы, инерционные члены отбрасываются; производной по времени также можно пренебречь в силу довольно больших временных масштабов происходящих процессов. Численное решение задач о конвективных фильтрационных течениях, особенно в классических поста-

[1)]Работа выполнена при финансовой поддержке РНФ (код проекта 21-11-00126).

новках, часто осуществляется с помощью оригинальных авторских кодов, отличающихся простотой, экономичностью и возможностью получать на выходе разнообразные характеристики процесса [7]–[11].

Рассматривается жидкость с растворенной примесью (например, вода с солями). Распространение примеси описывается нестационарным уравнением конвекции-диффузии (уравнение второго порядка, параболическое) и совместно с уравнениями неразрывности и Дарси (уравнения первого порядка) образует базовую систему уравнений. Для получения численных решений используется авторский двумерный вычислительный код, основанный на конечно-разностном методе. Код, предыдущие версии которого описаны в [12], [13] применялся для решения различных задач о концентрационной конвекции [14]–[17].

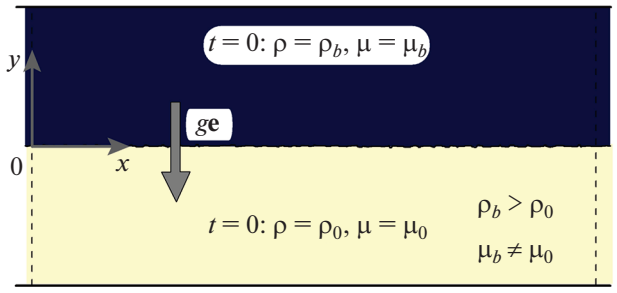
Принципы построения конечно-разностных аналогов уравнения конвекции-диффузии и их решения широко обсуждаются в литературе [18]–[20]. Примеры конкретных разностных схем, включая анализ их устойчивости и сходимости, можно найти в [8], [21]–[24]. В данной работе применяется аппроксимация конвективного члена по схеме Quadratic Upstream Interpolation for Convective Kinematics (QUICK) [25], которая является условно-устойчивой и имеет третий порядок точности на равномерной сетке. Схема сохраняет монотонность при значениях сеточного числа Пекле вплоть до нескольких десятков, что позволяет увеличить шаг пространственной сетки по сравнению с центрально-разностной схемой. Диффузионный член аппроксимируется центральными разностями. В результате дискретизации получается пятидиагональная линейная система уравнений. В предыдущей версии вычислительного кода [13] интегрирование по времени осуществлялось по двухслойной явной схеме, которая имеет первый порядок точности. На современном этапе код усовершенствован – интегрирование по времени проводится по двухшаговой схеме Рунге–Кутты, которая имеет второй порядок точности.

Решается задача о концентрационно-конвективном движении, развивающемся вследствие неустойчивости Рэлея–Тейлора. Данный вид неустойчивости возникает между контактирующими жидкостями разной плотности с границей раздела, перпендикулярной силе тяжести; более тяжелая жидкость находится сверху. Следует отметить, что задача о неустойчивости Рэлея–Тейлора является одной из классических в гидродинамике [26], поэтому используется для тестирования гидродинамических моделей и численных кодов [27]. Численное моделирование конвекции Рэлея–Тейлора в пористой среде проводилось в [28]–[30]. В настоящей работе, продолжая исследование [17], учитывается изменение вязкости раствора с количеством растворенной примеси. Анализируется влияние контраста вязкости (начального отношения коэффициентов вязкости слоев) на конвективное движение и перемешивание, которые оцениваются по различным характеристикам. Рассматриваются системы с контрастом вязкости в несколько единиц, что соответствует воде и водным растворам солей. Рассматривается также система с контрастом вязкости, равным 30, что ассоциируется с парой “нефть–вода”.

При моделировании конвекции Рэлея–Тейлора в пористой среде обнаружен новый эффект [30]: если отношение коэффициентов вязкости слоев больше 20, то конвективное течение теряет симметрию движения в направлении “верх–низ”. Считалось, что вязкость раствора возрастает с добавлением примеси, поэтому верхний слой был более тяжелым и вязким. Получено, что конвективные “пальцы” быстрее движутся вниз, чем вверх. Новизна настоящего исследования заключается в том, что рассматривается другая зависимость вязкости от количества примеси: коэффициент вязкости убывает с добавлением примеси, – поэтому верхний слой оказывается более тяжелым, но менее вязким. В статье будет показано, что в такой постановке при контрасте вязкости, равном 30, конвективные “пальцы” быстрее движутся вверх. Сравнение найденной в статье и описанной в [30] закономерностей позволяет сделать однозначный вывод о направлении асимметрии движения. Стоит заметить, что численное решение в [30] получено с помощью сложного гибридного метода, включающего псевдоспектральный и конечно-разностный методы. Применялись, соответственно, преобразование Хартли и конечно-разностные аппроксимации 4-го и 6-го порядков точности; интегрирование по времени производилось по одной из разновидностей полунеявного метода Адамса 3-го порядка. Новизной настоящей работы является также и то, что в ней показано: асимметричную конвекцию Рэлея–Тейлора можно моделировать с помощью относительно простого и экономичного конечно-разностного метода.

1. ПОСТАНОВКА ЗАДАЧИ И МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Имеется прямоугольная двумерная пористая область, заполненная двухслойной жидкостью: сверху однокомпонентной жидкости, имеющей плотность ρ_0 и вязкость μ_0, помещена двухком-



Фиг. 1. Постановка задачи.

понентная жидкость с примесью, имеющая плотность ρ_b и вязкость μ_b; $\rho_b > \rho_0$. В начальный момент система неподвижна и находится в состоянии гидростатического равновесия; граница между слоями горизонтальна (см. фиг. 1). На границах области для скорости задается условие проскальзывания, для давления сохраняются начальные значения, жидкость через границы не проникает. Примесь диффундирует из верхней части области в нижнюю, переходная зона между слоями увеличивается. На вертикальных границах задается диффузионное распределение примеси. В силу неустойчивого состояния равновесия такой системы в поле силы тяжести (неустойчивости Рэлея–Тейлора) со временем развивается естественное концентрационно-конвективное движение. Чтобы инициировать развитие конвекции, в начальный момент задаются флуктуации плотности на границе между слоями.

Описание гидродинамических процессов производится с помощью уравнений неразрывности несжимаемой жидкости, Дарси и переноса примеси [4]. Пористость ϕ и проницаемость твердого скелета k, а также коэффициент диффузии примеси D считаются постоянными; вязкость раствора μ – переменная. Исходную систему уравнений можно привести к следующему безразмерному виду [13]:

$$\nabla \cdot \mathbf{u} = 0, \tag{1.1}$$

$$\mathbf{u} = -\mathrm{Ra}\frac{\phi}{f_\mu}(\nabla\Pi - S\mathbf{e}), \tag{1.2}$$

$$\phi\frac{\partial S}{\partial t} + \mathbf{u}\cdot\nabla S = \nabla\cdot(\phi\nabla S). \tag{1.3}$$

Безразмерные переменные – это скорость фильтрации $\mathbf{u} = (u_x, u_y)$, давление $\Pi = (P - P_0)/((\rho_b - \rho_0)gH)$, плотность $S = (\rho - \rho_0)/(\rho_b - \rho_0)$. В качестве давления и плотности рассматриваются отклонения текущих значений P и ρ от значений P_0 и ρ_0 в жидкости, образующей нижний слой. Величина P_0 – гидростатическое распределение давления – линейно уменьшается с высотой. Характерными масштабами являются геометрический размер H, скорость D/H, время H^2/D, плотность $(\rho_b - \rho_0)$, давление $(\rho_b - \rho_0)gH$. Здесь g – ускорение свободного падения. В уравнении (1.2) содержится число Рэлея–Дарси

$$\mathrm{Ra} = \frac{(\rho_b - \rho_0)gHk}{\phi\mu_0 D}$$

и безразмерная вязкость $f_\mu = \mu/\mu_0$.

Плотность раствора ρ линейно растет с плотностью растворенной примеси ρ_c, что описывается уравнением состояния $\rho = \rho_0 + \alpha\rho_c$, используя которое можно получить другое выражение безразмерной переменной S, а именно: $S = \alpha\rho_c/(\rho_b - \rho_0)$. Видно, что S – это нормированная плотность примеси. Считается, что вязкость раствора увеличивается в зависимости от S экспоненциально:

$$f_\mu = \exp(\Gamma S). \tag{1.4}$$

Константа Γ определяет диапазон изменения вязкости.

Расчетная область Ω в безразмерном виде задается следующим образом:

$$\Omega = \{\mathbf{r}|\mathbf{r} = (x, y),\ 0 < x < h_x,\ -h_y < y < h_y\}.$$

Начальные условия имеют следующий вид:

$$\mathbf{u} = 0, \quad S = 0, \quad \Pi = \Pi^{\text{in}}, \quad 0 < x < h_x, \quad -h_y < y < 0,$$
$$\mathbf{u} = 0, \quad S = 1, \quad \Pi = \Pi^{\text{in}}, \quad 0 < x < h_x, \quad 0 < y < h_y, \quad t = 0.$$

На границе между слоями выполняется:

$$\mathbf{u} = 0, \quad S = 0.5 + \Delta s, \quad \Pi = \Pi^{\text{in}}, \quad 0 < x < h_x, \quad y = 0, \quad t = 0.$$

Здесь Δs – флуктуации плотности (малая величина); их определение в численном решении дано в (2.4). Распределение начального давления Π^{in} следует из условия гидростатического равновесия:

$$\Pi^{\text{in}} = 0, \quad 0 < x < h_x, \quad -h_y < y \le 0,$$
$$\Pi^{\text{in}} = -y, \quad 0 < x < h_x, \quad 0 < y < h_y, \quad t = 0.$$

На вертикальных границах изменение плотности примеси S соответствует аналитическому решению уравнения диффузии [31]. Граничные условия задаются следующими выражениями:

$$u_x = 0, \quad \frac{\partial u_y}{\partial x} = 0, \quad S = 0.5\left(1 - \operatorname{erf}\left(\frac{-y}{2t^{1/2}}\right)\right), \quad \Pi = \Pi^{\text{in}}, \quad x = 0, h_x, \quad -h_y < y < h_y,$$
$$\frac{\partial u_x}{\partial y} = 0, \quad u_y = 0, \quad \frac{\partial S}{\partial y} = 0, \quad \Pi = \Pi^{\text{in}}, \quad 0 < x < h_x, \quad y = -h_y, h_y, \quad t > 0. \tag{1.5}$$

Вязкость f_μ рассчитывается по полю плотности S в соответствии с (1.4). Исходно выполняются следующие условия:

$$f_\mu = 1, \quad 0 < x < h_x, \quad -h_y < y < 0,$$
$$f_\mu = F_\mu, \quad 0 < x < h_x, \quad 0 < y < h_y, \quad t = 0.$$

Здесь $F_\mu = \mu_b/\mu_0$ – начальное отношение коэффициентов вязкости верхнего и нижнего слоев, которое следует из (1.4) при $S = 1$. Контрастом вязкости обычно называют отношение большего коэффициента вязкости к меньшему. Контраст вязкости совпадает с F_μ, если $\mu_b > \mu_0$ и равен $1/F_\mu$, если $\mu_b < \mu_0$.

2. МЕТОД ЧИСЛЕННОГО РЕШЕНИЯ

Уравнения (1.1)–(1.3), записанные в плоской декартовой системе координат, заменяются конечно-разностными аналогами на разнесенной неравномерной пространственной сетке. Сначала определяются значения скорости и давления по дискретным уравнениям Дарси и неразрывности с использованием алгоритма типа SIMPLE [32]; полученные трехдиагональные системы уравнений для сеточных значений приращения давления решаются методом прогонки. Затем по дискретному уравнению конвекции-диффузии находятся значения плотности. После этого производится совместная корректировка скорости, давления и плотности итерационным методом Якоби. Когда плотность определена, по ней рассчитываются сеточные значения вязкости. Более подробно численный метод описан в [12], [13].

Формируется основная пространственная сетка

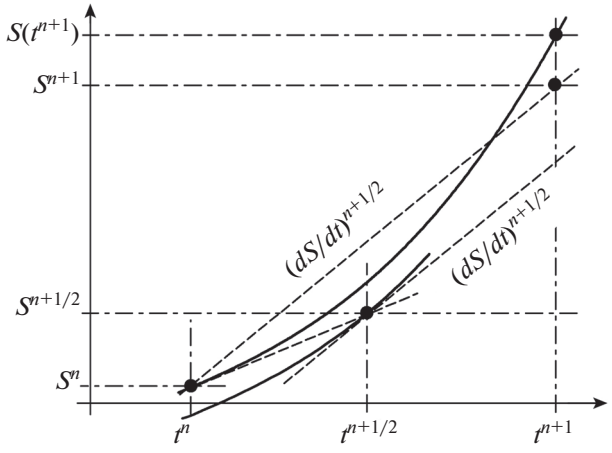
$$\Omega_h = \{(x_i, y_j),\ 0 = x_1 < x_2 \ldots < x_l = h_x,\ -h_y = y_1 < y_2 \ldots < y_m = h_y\}.$$

Вводятся дополнительные узлы

$$x_{i+1/2} = 0.5(x_{i+1} + x_i), \quad y_{j+1/2} = 0.5(y_{j+1} + y_j)$$

и формируются смещенные сетки. В узлах смещенной по x сетки

$$\Omega_{hx} = \{(x_{i+1/2}, y_j),\ -0.5x_2 = x_{1/2} < x_{3/2} \ldots < x_{l+1/2} = h_x + 0.5(x_l - x_{l-1}),\ -h_y = y_1 < y_2 \ldots < y_m = h_y\}$$



Фиг. 2. Пояснения к схеме Рунге–Кутты.

определяется вертикальная компонента скорости u_y. В узлах смещенной по y сетки

$$\Omega_{hy} = \{(x_i, y_{j+1/2}),\ 0 = x_1 < x_2 ... < x_l = h_x,\ -h_y - 0.5y_2 = y_{1/2} < y_{3/2} ... < y_{m+1/2} = h_y + 0.5(y_m - y_{m-1})\}$$

определяется горизонтальная компонента скорости u_x. В узлах смещенной по x и y сетки

$$\Omega_{hxy} = \{(x_{i+1/2}, y_{+1/2}),\ 1 - 0.5x_2 = x_{1/2} < x_{3/2} ... < x_{l+1/2} = h_x + 0.5(x_l - x_{l-1}),$$
$$-h_y - 0.5y_2 = y_{1/2} < y_{3/2} ... < y_{m+1/2} = h_y + 0.5(y_m - y_{m-1})\}$$

определяются все скалярные переменные. Узлы сетки Ω_{hxy} являются центрами ячеек сетки Ω_h.

Остановимся на методе интегрирования уравнения конвекции-диффузии, который на данном этапе изменен. Запишем пространственные производные уравнения (1.3) в конечно-разностном виде, используя схему QUICK [13], [25]. Полученное уравнение в каком-либо $(i + 1/2, j + 1/2)$-м узле можно представить следующим образом:

$$\frac{dS_{i+1/2,j+1/2}}{dt} = LS_{i+1/2,j+1/2}. \tag{2.1}$$

Здесь L – линейный сеточный оператор, шаблон этого оператора включает пять узлов сетки по каждому направлению. При попытке неявно проинтегрировать (2.1) по времени приходим к пятидиагональной системе уравнений, в которых в пяти узлах сеточные значения $S_{i+1/2,j+1/2}$ взяты с верхнего временного слоя. При решении такой системы, например, прогонками возникают определенные трудности при разрешении граничных условий. Поэтому рациональным выбором оказываются явные схемы. Использование простейшей двухслойной схемы (как это сделано в [13]) дает лишь первый порядок точности по времени. Чтобы повысить точность численного интегрирования, в текущей версии вычислительного кода применяется один из методов Рунге–Кутты [19], часто используемых для решения начально-краевых задач. Суть схемы поясняется на фиг. 2. Для простоты нижние индексы у $S_{i+1/2,j+1/2}$ опускаются. Обозначим шаг интегрирования по времени через τ. Сначала выполняется интегрирование на полушаге – определяется предварительное значение переменной $S^{n+1/2}$. Затем вычисляется окончательное значение S^{n+1} с использованием найденных промежуточных значений $S^{n+1/2}$:

$$S^{n+1/2} = S^n + \frac{\tau}{2}LS^n, \quad S^{n+1} = S^n + \tau LS^{n+1/2}. \tag{2.2}$$

Таким образом, по $S^{n+1/2}$ определяется производная $(dS/dt)^{n+1/2}$ в точке $t^{n+1/2}$, а по этой производной находится S^{n+1}.

Можно оценить точность данной схемы. Считается, что на нижнем временном слое сеточное значение S^n совпадает с точным значением $S(t^n)$. Вычислим погрешность ψ на шаге τ как раз-

ность между точным $S(t^{n+1})$ и сеточным S^{n+1} значениями на верхнем временном слое. Будем использовать схему (2.2) и разложения $S(t^{n+1})$ и $S(t^n)$ в ряд Тейлора около точки $t^{n+1/2}$:

$$S(t^{n+1}) = S(t^{n+1/2}) + \frac{\tau}{2}\frac{dS(t^{n+1/2})}{dt} + \frac{\tau^2}{8}\frac{d^2S(t^{n+1/2})}{dt^2} + O(\tau^3),$$

$$S(t^n) = S(t^{n+1/2}) - \frac{\tau}{2}\frac{dS(t^{n+1/2})}{dt} + \frac{\tau^2}{8}\frac{d^2S(t^{n+1/2})}{dt^2} + O(\tau^3).$$

Для ψ можно записать следующее:

$$\psi = S(t^{n+1}) - S^{n+1} = S(t^{n+1}) - S^n - \tau LS^{n+1/2} = S(t^{n+1/2}) + \frac{\tau}{2}\frac{dS(t^{n+1/2})}{dt} + \frac{\tau^2}{8}\frac{d^2S(t^{n+1/2})}{dt^2} -$$

$$- S(t^{n+1/2}) + \frac{\tau}{2}\frac{dS(t^{n+1/2})}{dt} - \frac{\tau^2}{8}\frac{d^2S(t^{n+1/2})}{dt^2} - \tau LS^{n+1/2} + O(\tau^3),$$

$$\psi = \tau\frac{dS(t^{n+1/2})}{dt} - \tau LS^{n+1/2} + O(\tau^3). \tag{2.3}$$

В силу равенства (2.1) первое и второе слагаемые справа в (2.3) равны друг другу, поэтому окончательно получаем $\psi = O(\tau^3)$, что свидетельствует о втором порядке точности.

Чтобы инициировать развитие конвекции, в начальный момент задаются флуктуации плотности на границе раздела жидкостей, которая проходит через узлы основной сетки при $j = (m+1)/2$ (m – нечетное число). Плотность определяется в узлах смещенной сетки, около границы раздела задаются условия:

$$S_{i+1/2,m/2} = \sigma R_{i+1/2}, \quad S_{i+1/2,m/2+1} = 1 - \sigma(1 - R_{i+1/2}).$$

Здесь σ – амплитуда флуктуаций ($\sigma \ll 1$), $R_{i+1/2}$ – ряд случайных чисел: $R_{i+1/2} \in [0,1]$. Распределение плотности в узлах основной сетки получается интерполяцией: $S_{i+1/2,(m+1)/2} = 0.5(S_{i+1/2,m/2} + S_{i+1/2,m/2+1})$. Легко найти, что

$$S_{i+1/2,(m+1)/2} = 0.5 + \Delta s_{i+1/2},\ \Delta s_{i+1/2} = \sigma(R_{i+1/2} - 0.5). \tag{2.4}$$

Здесь $\Delta s_{i+1/2}$ – сеточная функция флуктуаций плотности. Известно, что время начала конвекции зависит от амплитуды флуктуаций [17], [33], поэтому во всех вариантах берется одинаковое значение $\sigma = 0.01$. Если флуктуации физических переменных не заданы, то источником малых возмущений являются погрешности численного метода и округления при вычислениях.

При проведении счета на каждом временном слое по найденным значениям переменных определяются различные величины, характеризующие движение и массоперенос. Вычисляется высота зоны перемешивания слоев h_c. Для этого сначала рассчитывается средняя масса примеси Q на высоте $y_{j+1/2}$, $j = 1,2,\ldots m-1$:

$$Q(y_{j+1/2}) = \frac{1}{h_x}\sum_{i=1}^{l-1} S_{i+1/2,j+1/2}\Delta x_{i+1/2}. \tag{2.5}$$

Если перемешивания нет, то в нижней половине области должно быть $Q = 0$, а в верхней – $Q = 1$. Определяем нижнюю координату зоны перемешивания y_* как уровень, на котором $Q = 0.01$, т.е. $y_* = y_{j+1/2}$, если $Q(y_{j+1/2}) = 0.01$. Верхняя координата y^* находится из условия: $y^* = y_{j+1/2}$, если $Q(y_{j+1/2}) = 0.99$. Окончательно вычисляем $h_c = y^* - y_*$. Определяется также высота зоны диффузионного перемешивания h_d по аналитическому решению уравнения диффузии, записанному в (1.5); h_d характеризует чисто диффузионный перенос примеси в отсутствие конвекции.

Рассчитывается кинетическая энергия движения примеси, приходящаяся на единицу длины области,

$$K = \frac{1}{2h_x}\sum_{i=1}^{l-1}\sum_{j=1}^{m-1} S_{i+1/2,j+1/2}(u^2_{xi+1/2,j+1/2} + u^2_{yi+1/2,j+1/2})\Delta x_{i+1/2}\Delta y_{j+1/2}. \tag{2.6}$$

Компоненты скорости $u_{xi+1/2,j+1/2}$ и $u_{yi+1/2,j+1/2}$ находятся линейной интерполяцией соответствующих значений из соседних узлов сеток Ω_{hy} и Ω_{hx}. В (2.5), (2.6) входят $\Delta x_{i+1/2} = x_{i+1} - x_i$, $\Delta y_{j+1/2} = y_{j+1} - y_j$ – шаги сетки Ω_h по x и y.

Рассматривается завихренность скорости, которая по определению представляет собой вектор $\boldsymbol{\omega} = \nabla \times \mathbf{v}$, где $\mathbf{v} = \mathbf{u}/\phi$ – скорость движения жидкости в порах; $\mathbf{v} = (v_x, v_y)$. В двумерной системе координат вектор $\boldsymbol{\omega}$ имеет одну компоненту ω вдоль оси, перпендикулярной плоскости течения:

$$\omega = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}.$$

Сеточное значение $\omega_{i,j}$ находится в узле основной сетки Ω_h:

$$\omega_{i,j} = \frac{(v_{yi+1/2,j} - v_{yi-1/2,j})}{\Delta x_i} - \frac{(v_{xi,j+1/2} - v_{xi,j-1/2})}{\Delta y_j}.$$

Здесь $\Delta x_i = x_{i+1/2} - x_{i-1/2}$, $\Delta y_j = y_{j+1/2} - y_{j-1/2}$, – шаги сетки Ω_{hxy} по x и y; около горизонтальных границ задаются полушаги $\Delta y_1 = y_{3/2} - y_1$, $\Delta y_m = y_m - y_{m-1/2}$. Затем определяется суммарный модуль завихренности Σ_ω, приходящийся на единицу длины области:

$$\Sigma_\omega = \frac{1}{h_x} \sum_{i=1}^{l} \sum_{j=1}^{m} |\omega_{i,j}| \Delta x_i \Delta y_j.$$

Вычисляется также параметр неоднородности распределения примеси G вдоль линии начального положения границы между слоями жидкости, т.е. при $y_{(m+1)/2}$. Полагается, что G – это разность между максимальным $S_{\max}$ и минимальным $S_{\min}$ значениями плотности на этом уровне:

$$G = S_{\max} - S_{\min}, \quad S_{\max} = \max_i S_{i+1/2,(m+1)/2}, \quad S_{\min} = \min_i S_{i+1/2,(m+1)/2}.$$

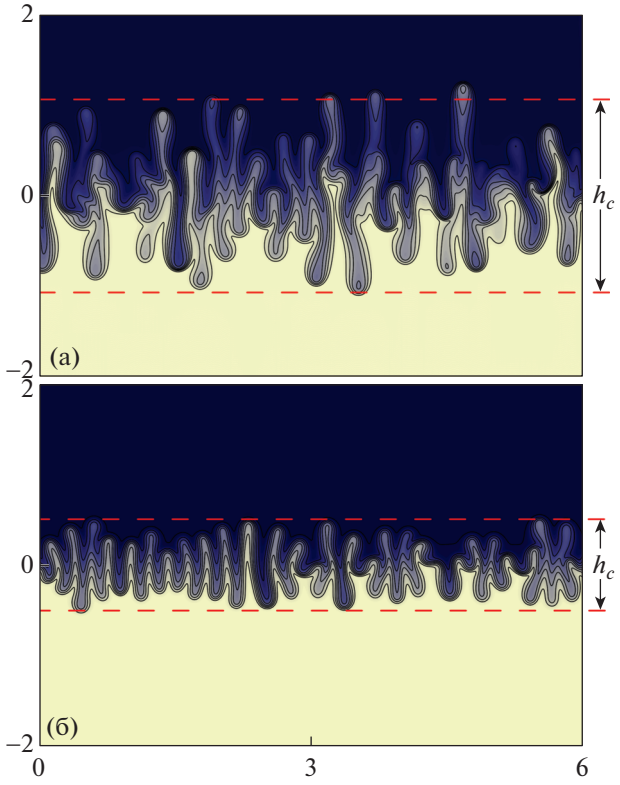
Чтобы обеспечить устойчивость и сходимость численного метода, численные параметры должны удовлетворять ряду критериев. Анализ устойчивости приводит к ограничению диффузионного числа Nd [34] и числа Куранта Co [19], а требование монотонности – к ограничению сеточного числа Пекле Pe [35]. В случае равномерной сетки критерии имеют вид

$$\mathrm{Nd} = \frac{\tau}{\Delta x^2} + \frac{\tau}{\Delta y^2} < 0.5, \quad \mathrm{Co} = \frac{\tau v_{x\max}}{\Delta x} + \frac{\tau v_{y\max}}{\Delta y} < 1, \qquad \mathrm{Pe} = v_{x\max}\Delta x + v_{y\max}\Delta y < 10. \tag{2.7}$$

Здесь $v_{x\max}$ и $v_{y\max}$ – максимальные значения компонент скорости движения жидкости v_x и v_y. При аппроксимации конвективного члена по центрально-разностной схеме следует использовать условие $\mathrm{Pe} < 2$; применение схемы QUICK позволило увеличить значение Pe.

3. МОДЕЛИРОВАНИЕ КОНВЕКЦИИ ПРИ НЕБОЛЬШОМ КОНТРАСТЕ ВЯЗКОСТИ

Проведено моделирование конвекции Рэлея–Тейлора при значении числа Рэлея–Дарси $\mathrm{Ra} = 10^3$, которое соответствует, например, водному раствору солей в пористой среде с физическими параметрами: $\rho_0 = 10^3$ кг/м^3, $\rho_b = 1.2\times10^3$ кг/м^3, $\mu_0 = 10^{-3}$ Па с, $D = 1.57\times10^{-9}$ м^2/с, $\alpha = 0.815$, $\phi = 0.2$, $k = 1.6\times10^{-14}$ м^2. Масштаб длины $H = 10$ м. Вязкость верхнего слоя больше, чем нижнего. Контраст вязкости, равный F_μ, варьируется. Рассмотрены четыре варианта, в которых взята константа $\Gamma = 0$; 0.405; 0.810; 1.216 и получены, соответственно, величины $F_\mu = 1.00$; 1.50; 2.25; 3.38. В [17] показано, что экспериментальные данные для водных растворов хлорида натрия с хорошей точностью описываются зависимостью (1.4) с константой $\Gamma = 1.0$; в настоящем исследовании взят диапазон Γ около этого значения.



Фиг. 3. Поле плотности примеси в момент времени $t = 6\times10^{-3}$ в среде с контрастом вязкости $F_\mu = 1.0$ (а); 3.38 (б). Здесь h_c – высота зоны перемешивания слоев.

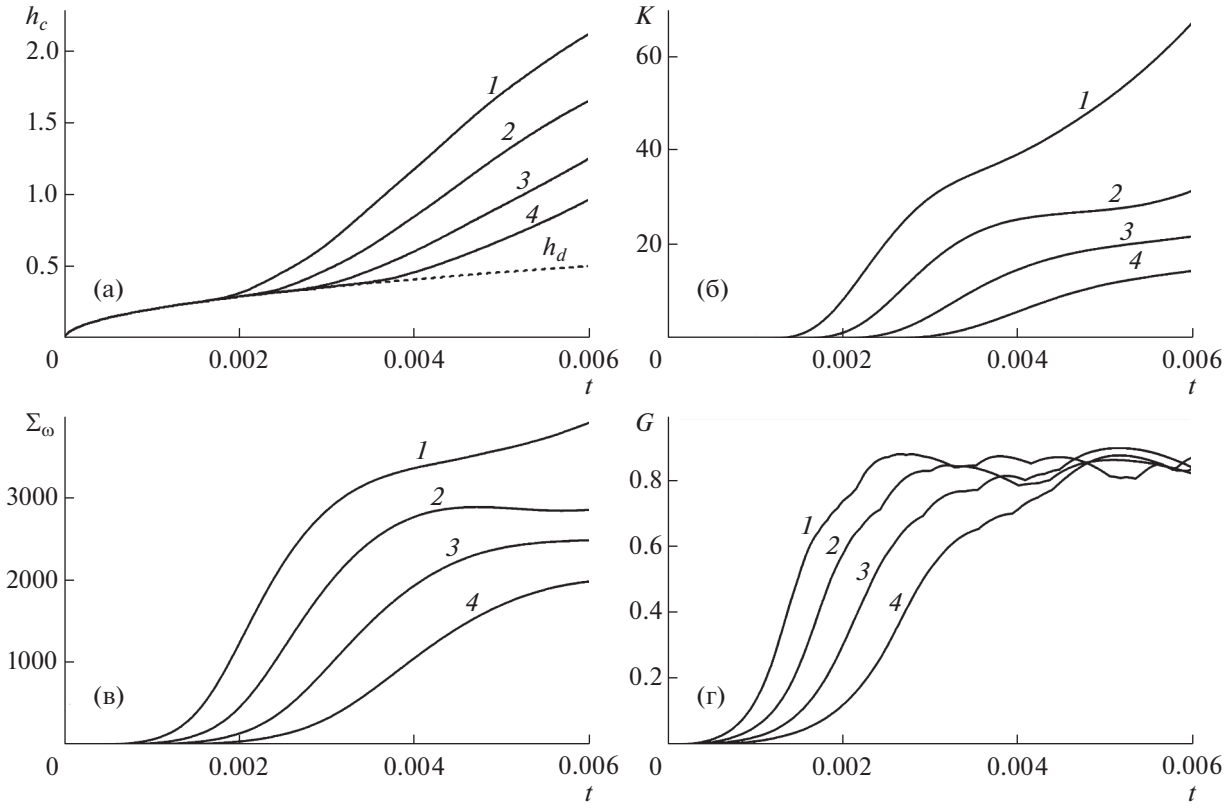
Рассматривается область 6×4, покрытая равномерной сеткой 2401×1601. Шаг интегрирования по времени τ менялся, однако, во всех вариантах $\tau \le 10^{-7}$. В расчетах ниже найдено для скорости движения жидкости: $v_{x\max}, v_{y\max} < 10^3$. Оценки показывают, что условия (2.7) соблюдаются.

Точность численного решения контролировалась по выполнению баланса массы примеси в расчетной области. На каждом временном слое вычислялась разность $\Delta M = M^n - M^0$, где M^n и M^0 – суммарная масса примеси в текущий t^n и начальный t^0 моменты времени. В силу того что на вертикальных границах $x = 0, h_x$ поддерживается определенное распределение плотности S (1.5), здесь могут возникать незначительные диффузионные потоки. Рассчитывается суммарный диффузионный поток массы M_J с начала процесса до текущего момента t^n. В точном решении выполняется $\Delta M = M_J$. В численном решении появляется невязка массы δM, которая определяется по выражению:

$$\delta M = \left|\frac{\Delta M - M_J}{\Delta M + M_J}\right|.$$

Получено, что во всех расчетах невязка δM не превышает нескольких долей процента, что свидетельствует о высокой точности численного решения.

Можно выделить три стадии процесса, начиная с нулевого момента времени [17], [28]. На первой стадии происходит диффузионный перенос примеси из верхней части области в нижнюю. При этом переходная зона между слоями расширяется, но остается горизонтальной; конвективное движение жидкости еще не появилось. Слабые неупорядоченные перемещения частей жидкости в переходной зоне, которые порождаются флуктуациями плотности, со временем перестраиваются, усиливаются и формируется согласованное движение. На второй стадии наблюдается периодическое конвективное движение, становится заметным волнообразное искривление

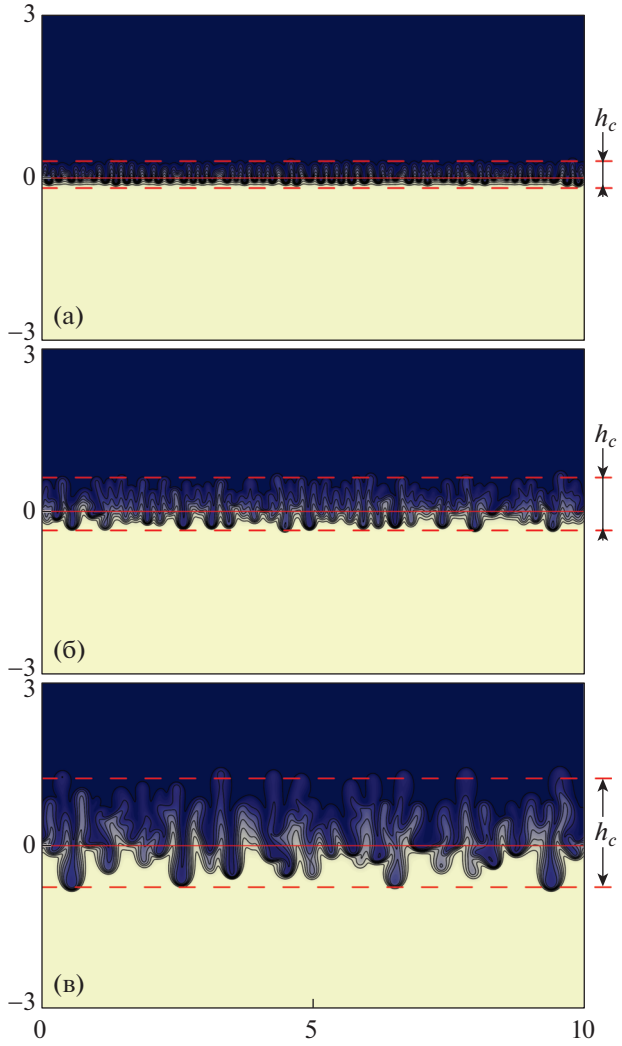


Фиг. 4. Высота зоны перемешивания слоев h_c (а), кинетическая энергия K (б), модуль завихренности Σ_ω (в) и параметр неоднородности распределения примеси G (г) в зависимости от времени в среде с контрастом вязкости $F_\mu = 1.00$ (*1*); 1.50 (*2*); 2.25 (*3*); 3.38 (*4*). Штриховая линия на фрагменте (а) – высота зоны диффузионного перемешивания h_d.

переходной зоны. Со временем конвективные “пальцы” удлиняются, искривляются, начинают сливаться друг с другом и процесс переходит в третью стадию, которая отличается наличием интенсивной стохастической конвекции.

Все стадии успешно воспроизводятся в численном моделировании. Учет увеличения вязкости раствора, как показано в [17], приводит к более позднему и медленному развитию конвективного течения. На фиг. 3 дано поле плотности примеси в момент времени, который соответствует третьей стадии процесса. Переход от светлого тона к темному означает увеличение плотности. Сравниваются варианты с постоянной ($F_\mu = 1.0$) и переменной ($F_\mu = 3.38$) вязкостью. Видно, что во втором случае зона конвективного перемешивания оказывается меньше, что объясняется замедленным развитием конвекции с увеличением вязкости и соответствует предыдущим результатам. На фиг. 4а показано, как растет высота h_c при разных F_μ. На первой стадии, когда конвекции нет, кривые h_c совпадают с h_d. По отрыву h_c от h_d можно судить о начале конвекции, которое при увеличении F_μ происходит позже.

В настоящей работе анализируются характеристики, которые ранее не рассматривались. На фиг. 4б, в представлены временные зависимости кинетической энергии K и модуля завихренности Σ_ω. Величина K ассоциируется, главным образом, с поступательным подъемно-опускным движением примеси, в то время как Σ_ω свидетельствует о перемешивании, поскольку локальная завихренность скорости – это удвоенная угловая скорость вращения [36]. По рисункам видно, что с началом конвекции K и Σ_ω растут почти линейно, затем кривые претерпевают изгиб, что связано с перестройкой периодического движения в стохастическое. Возрастание F_μ ведет к уменьшению K и Σ_ω. Полученные данные о параметре неоднородности распределения примеси G на фиг. 4г показывают, что неоднородность плотности при увеличении F_μ нарастает медлен-

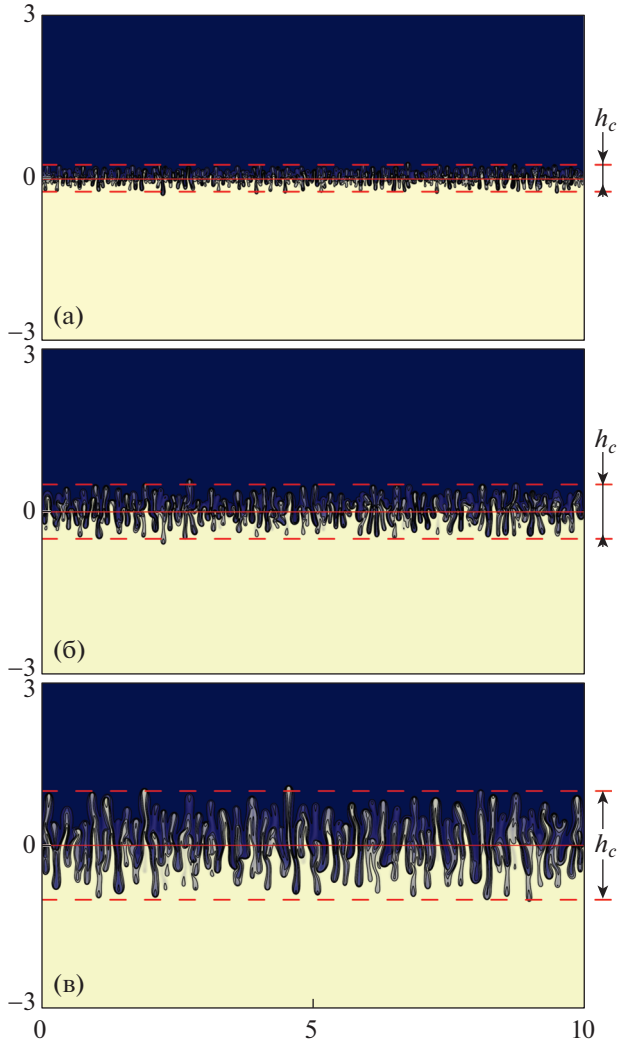


Фиг. 5. Поле плотности примеси в моменты времени $t = 4.10\times10^{-3}$ (а); 6.95×10^{-3} (б); 1.15×10^{-2} (в) в среде с контрастом вязкости $1/F_\mu = 30$.

нее. Однако все кривые G со временем выходят примерно на один уровень и в стадии развитой конвекции колеблются около значения $G \approx 0.85$. Предельная величина $G = 1$ не достигается из-за диффузионного рассеяния.

4. МОДЕЛИРОВАНИЕ КОНВЕКЦИИ ПРИ БОЛЬШОМ КОНТРАСТЕ ВЯЗКОСТИ

Исследуется конвекция Рэлея–Тейлора с контрастом вязкости слоев жидкости, равным 30, что ассоциируется с парой “нефть–вода”. В этом случае верхний слой (вода), более плотный, но менее вязкий, чем нижний слой (нефть), что задается с помощью отрицательного значения Γ в (1.4). Берется $\Gamma = -3.40$, получается $F_\mu = 0.0334$ и, следовательно, контраст вязкости $1/F_\mu = 30$. Для сравнения моделируется конвекция в среде постоянной вязкости, равной вязкости воды, т.е. при $F_\mu = 1$. В обоих вариантах число Рэлея–Дарси, построенное по коэффициенту вязкости воды, имеет значение $\mathrm{Ra} = 3\times10^3$. Нефть и вода представляют собой несмешивающиеся жидкости, разделенные резкой границей. Тем не менее модель с диффузией массы между слоями, принятая в настоящем исследовании, позволяет получить качественные результаты для несмешивающихся жидкостей на стадии развитой конвекции, когда конвективный перенос оказывается существенно больше, чем диффузионный перенос.



Фиг. 6. Поле плотности примеси в моменты времени $t = 4.73\times10^{-4}$ (а); 8.30×10^{-4} (б); 1.42×10^{-3} (в) в среде с контрастом вязкости $1/F_\mu = 1.0$.

Численное решение получено в области 10×6 на равномерной сетке 4001×2001. Шаг интегрирования по времени составляет $\tau = 0.5\times10^{-7}$; 0.125×10^{-7} при $F_\mu = 0.0334$; 1 соответственно. Компоненты максимальной скорости движения $v_{x\max}$, $v_{y\max}$ не превосходят значений 6×10^2 в первом и 2.2×10^3 во втором случаях, так что критерии (2.7) удовлетворяются.

На фиг. 5, 6 приведены поля плотности примеси в различные моменты времени в среде с переменной и постоянной вязкостью. Поскольку значения коэффициентов вязкости отличаются очень сильно, то процессы развиваются в существенно разных временных масштабах. Чтобы провести сравнение, были выбраны такие моменты, когда высота зоны перемешивания h_c на фрагментах (а) или (б), или (в) имеет одинаковое значение: $h_c = 0.5$ (а); 1.0 (б); 2.0 (в). Видны отличия: толщина конвективных “пальцев” на фиг. 5 заметно больше, чем на фиг. 6 в силу высокой вязкости нижнего слоя. На фиг. 5а структура течения еще близка к регулярной, в то время как на фиг. 6а уже преобразовалась в стохастическую.

На фиг. 5 можно заметить, что движение конвективных “пальцев” происходит несимметрично относительно начального положения границы раздела между слоями, т.е. уровня $y = 0$, выделенного сплошной красной линией. “Пальцы” более вязкой жидкости поднимаются вверх с большей скоростью, чем опускаются вниз “пальцы” менее вязкой жидкости. Как следствие, зо-

на конвективного перемешивания смещена вверх. Отношение верхней координаты зоны перемешивания y^* к нижней координате y_* (взятое по модулю, так как $y_* < 0$) имеет значение $|y^*/y_*| = 1.35$ (а); 1.74 (б); 1.63 (в). Однако, как показано на фиг. 6, в жидкой системе постоянной вязкости конвекция распространяется вверх и вниз одинаково; здесь $|y^*/y_*| = 1.0 \pm 0.03$ для всех фрагментов (а)–(в). Из сравнения фиг. 5 и 6 можно заключить, что несимметричное расширение зоны конвекции обуславливается разницей в значениях коэффициента вязкости слоев жидкости. Похожий эффект недавно описан в [30], где также проводилось численное моделирование конвекции Рэлея–Тейлора в пористой среде с экспоненциальным изменением вязкости раствора в зависимости от плотности примеси. В отличие от настоящего исследования, в [30] повышенную вязкость имеет верхний слой и преобладающее продвижение конвективных "пальцев" происходит вниз; эффект становится заметным при контрасте вязкости больше 20. В настоящей работе при тестировании вычислительного кода эффект, обнаруженный в [30], при аналогичной постановке задачи воспроизведен. Сравнивая новый результат о преобладающем распространении конвекции вверх, если вязкость раствора уменьшается при добавлении примеси, с результатом [30] можно установить следующую общую закономерность: быстрее движутся "пальцы" высоковязкой жидкости в зону жидкости низкой вязкости.

ЗАКЛЮЧЕНИЕ

Продемонстрирована эффективность авторского вычислительного кода при моделировании развития неустойчивости Рэлея–Тейлора в пористой среде на стадиях диффузионного переноса, периодической и стохастической конвекции. Легкий слой образован однокомпонентной жидкостью, тяжелый слой состоит из жидкости и растворенной примеси; вязкость раствора считается переменной. В текущей версии кода интегрирование по времени уравнения конвекции-диффузии производится по схеме Рунге–Кутты со вторым порядком точности.

Рассмотрен случай, когда верхний слой более вязкий, а контраст вязкости лежит в диапазоне от 1 до 3.38. Показано, что с увеличением контраста вязкости начало конвекции наступает позже, течение развивается медленнее, суммарная кинетическая энергия движения примеси и суммарный модуль завихренности скорости оказываются меньше. Однако параметр неоднородности распределения примеси, представляющий собой разность между максимальным и минимальным значениями плотности на линии начального положения границы между слоями, в стадии развитой конвекции выходит на один и тот же уровень независимо от контраста вязкости.

Рассмотрен случай, когда более вязким оказывается нижний слой, а контраст вязкости равен 30. Обнаружено, что распространение конвекции вверх-вниз происходит несимметрично, "пальцы" высоковязкой жидкости движутся вверх быстрее, чем опускаются "пальцы" менее вязкой жидкости. Полученный результат согласуется с описанным в литературе новым эффектом несимметричного развития конвекции при отношении коэффициентов вязкости слоев больше 20. Сравнивая полученную закономерность с приведенной в литературе, можно однозначно заключить, что асимметрия движения определяется только направлением изменения коэффициента вязкости – преимущественное распространение происходит в сторону его уменьшения. Результат работы имеет практическое применение для анализа системы, в которой слой воды располагается над слоем нефти. Хотя вода и нефть не смешиваются, но в стадии развитой конвекции поведение смешивающихся и несмешивающихся жидкостей качественно подобно. Контраст вязкости колеблется от нескольких единиц до примерно 200, поэтому можно предвидеть, что "пальцы" высоковязкой нефти будут подниматься в воде существенно быстрее, чем "пальцы" воды опускаться в нефти.

Автор благодарит Г.Г. Цыпкина за полезные обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. *Нигматулин Р.И.* Динамика многофазных сред. Ч. I, II. М.: Наука, 1987.
2. *Полубаринова-Кочина П.Я.* Теория движения грунтовых вод. М.: Наука, 1977.
3. *Nield D.A., Bejan A.* Convection in Porous Media. New York: Springer, 2006.
4. *Bear J., Cheng A.* Modeling Groundwater Flow and Contaminant Transport. New York: Springer, 2010.
5. *Любимова Т.П., Лепихин А.П., Паршакова Я.Н., Циберкин К.Б.* Численное моделирование инфильтрации жидких отходов из хранилища в прилегающие грунтовые воды и поверхностные водоемы // Вычисл. механ. сплошных сред. 2015. Т. 8. № 3. С. 310–318.

6. *Afanasyev A.A., Vedeneeva E.A.* Investigation of the efficiency of gas and water injection into an oil reservoir // Fluid Dynamics. 2020. V. 55. № 5. P. 621–630.
7. *Lyubimova T., Zubova N.* Nonlinear regimes of the Soret-induced convection of ternary fluid in a square porous cavity // Trans. in Porous Media. 2019. V. 127. P. 559–572.
8. *Абделхафиз М.А., Цибулин В.Г.* Численное моделирование конвективных движений в анизотропной среде и сохранение косимметрии // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 10. С. 1734–1747.
9. *Paoli M., Zonta F., Soldati A.* Dissolution in anisotropic porous media: Modeling convection regimes from onset to shutdown // Phys. of Fluids. 2017. V. 29. P. 026601.
10. *Hewitt D.R., Neufeld J.A., Lister J.R.* Ultimate regime of high Rayleigh number convection in a porous medium // Phys. Rev. Letters. 2012. V. 108. P. 224503.
11. *Pirozzoli S., Paoli M.De, Zonta F., Soldati A.* Towards the ultimate regime in Rayleigh–Darcy convection // J. Fluid Mech. 2021. V. 911. R4.
12. *Соболева Е.Б.* Метод численного исследования динамики соленой воды в почве // Матем. моделирование. 2014. Т. 26. № 2. С. 50–64.
13. *Соболева Е.Б.* Метод численного моделирования концентрационно-конвективных течений в пористых средах в приложении к задачам геологии // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 11. С. 162–173.
14. *Soboleva E.B., Tsypkin G.G.* Numerical simulation of convective flows in a soil during evaporation of water containing a dissolved admixture // Fluid Dynamics. 2014. V. 49. № 5. P. 634–644.
15. *Soboleva E.B., Tsypkin G.G.* Regimes of haline convection during the evaporation of groundwater containing a dissolved admixture // Fluid Dynamics. 2016. V. 51. № 3. P. 364–371.
16. *Soboleva E.B.* Density-driven convection in an inhomogeneous geothermal reservoir // Internat. Journal of Heat and Mass Transfer. 2018. V. 127 (part C). P. 784–798.
17. *Soboleva E.B.* Onset of Rayleigh–Taylor convection in a porous medium // Fluid Dynamics. 2021. V. 56. № 2. P. 200–210.
18. *Самарский А.А.* Теория разностных схем. М.: Наука, 1989.
19. *Калиткин Н.Н.* Численные методы. С.-Пб.: БХВ-Петербург, 2011.
20. *Самарский А.А., Вабищевич П.Н.* Численные методы решения задач конвекции-диффузии. М.: Либроком, 2015.
21. *Вабищевич П.Н., Захаров П.Е.* Схемы попеременно-треугольного метода для задач конвекции-диффузии // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 4. С. 587–604.
22. *Матус П.П., Хиеу Ле Минь.* Разностные схемы на неравномерных сетках для двумерного уравнения конвекции-диффузии // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 12. С. 2042–2052.
23. *Вабищевич П.Н.* Монотонные схемы для задач конвекции-диффузии с конвективным переносом в различной форме // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 1. С. 95–107.
24. *Брагин М.Д., Рогов Б.В.* Бикомпактные схемы для многомерного уравнения конвекции-диффузии // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 4. С. 625–643.
25. *Leonard B.P.* A stable and accurate convective modeling procedure based on quadratic upstream interpolation // Comp. Meth. in Applied Mech. and Engng. 1979. V. 19. № 1. P. 59–98.
26. *Drazin P.G., Reid W.H.* Hydrodynamic stability. Cambridge: Cambridge University Press, 1981.
27. *Елизарова Т.Г., Злотник А.А., Шильников Е.В.* Регуляризованные уравнения для численного моделирования течений гомогенных бинарных смесей вязких сжимаемых газов // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 11. С. 1899–1914.
28. *Paoli M. De, Giurgiu V., Zonta F., Soldati A.* Universal behavior of scalar dissipation rate in confined porous media // Phys. Rev. Fluids. 2019. V. 4. № 10. P. 101501.
29. *Elgahawy Y., Azaiez J.* Rayleigh–Taylor instability in porous media under sinusoidal time-dependent flow displacements // AIP Advances. 2020. V. 10. P. 075308.
30. *Sabet N., Hassanzadeh H., Wit A. De, Abedi J.* Scalings of Rayleigh–Taylor Instability at Large Viscosity Contrasts in Porous Media // Phys. Rev. Letters. 2021. V. 126. P. 094501.
31. *Ландау Л.Д., Лифшиц Е.М.* Гидродинамика. М.: Наука, 1986.
32. *Патанкар С.* Численные методы решения задач теплообмена и динамики жидкости. М.: Энергоиздат, 1984.
33. *Bestehorn M., Firoozabadi A.* Effect of fluctuations on the onset of density-driven convection in porous media // Phys. of Fluids. 2012. V. 24. P. 114102.
34. *Флетчер К.* Вычислительные методы в динамике жидкостей. В двух томах. Том 1. М.: Мир, 1991.
35. *Versteeg H.K., Malalasekera W.* An Introduction to Computational Fluid Dynamics: The Finite Volume Method. 2dn Ed. Glasgow: Bell & Bain Limited, 2007.
36. *Лойцянский Л.Г.* Механика жидкости и газа. М.: Наука, 1987.

EDN: CAKLTN
ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2022, том 62, № 11, с. 1940

ИНФОРМАТИКА

УДК 519.833

MULTI-CLUSTER COORDINATED MOVEMENT AND DYNAMIC REORGANIZATION[1)]

© 2022 г. Zhiqing Dang[1], Yang Yu[1], Zhaopeng Dai[1], Long Zhang[1], Ang Su[1], Zhihang You[1], Hongwei Gao[1,*]

[1] *School of Mathematics and statistics, Qingdao University, Qingdao 266071, Shandong, China*

**e-mail: gaohongwei@qdu.edu.cn*

Поступила в редакцию 07.01.2022 г.
Переработанный вариант 07.01.2022 г.
Принята к публикации 07.07.2022 г.

Многокластерное скоординированное движение и динамическая реорганизация. Рассматривается движение системы, состоящей из кластеров. Используя уравнения Гамильтона-Якоби-Беллмана, разрабатываются алгоритмы решения такой задачи, удовлетворяющие практическим требованиям быстродействия и точности. Основной проблемой исследуемой системы является изменяющийся состав членов кластера, которые могут добавляться или исчезать из него. Приводятся результаты численного моделирования движения системы, полученные на основе построенного алгоритма.

Ключевые слова: оптимальное управление, виртуальная траектория, предотвращение столкновений.

DOI: 10.31857/S0044466922110060

[1)]Полный текст статьи печатается в английской версии журнала.