ISSN 0044-4669

Том 64, Номер 5

Май 2024

ФИЦ ИУ РАН

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ

НАУКА

1727

СОДЕРЖАНИЕ

Том 64, номер 5, 2024

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

МАТЕМАТИЧЕСКАЯ ФИЗИКА

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 699–712

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.72

ЧИСЛЕННО-АНАЛИТИЧЕСКИЙ МЕТОД ДЕКОМПОЗИЦИОННО-АВТОКОМПЕНСАЦИОННОГО РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ СИГНАЛОВ ПО РЕЗУЛЬТАТАМ НЕКОРРЕКТНЫХ НАБЛЮДЕНИЙ

© 2024 г. Ю. Г. Булычев[1,*]

[1]344000 Ростов-на-Дону, пр-т Соколова, 96, АО «Всероссийский НИИ "Градиент"», Россия
**e-mail: ProfBulychev@yandex.ru*

Поступила в редакцию 10.11.2023 г.
Переработанный вариант 20.12.2023 г.
Принята к публикации 06.02.2024 г.

Развивается численно-аналитический метод решения задачи оптимального распознавания совокупности возможных сигналов, наблюдаемых в виде аддитивной смеси, содержащей не только флуктуационную погрешность наблюдений (с неизвестным статистическим законом распределения), но и сингулярную помеху (с параметрической неопределенностью). Он позволяет не только обнаруживать сигналы, присутствующие в смеси, но и оценивать их параметры, в рамках заданного критерия качества и сопутствующих ограничений. Предлагаемый метод, реализованный на идее обобщенного инвариантно-несмещенного оценивания значений линейных функционалов, обеспечивает декомпозицию вычислительной процедуры и автокомпенсацию сингулярной помехи, не прибегая к традиционному расширению пространства состояний. Для параметрического конечномерного представления сигналов и помехи используются линейные спектральные разложения в заданных функциональных базисах, для описания погрешности наблюдений достаточно знания лишь ее корреляционной матрицы. Анализируются случайные и методические погрешности, приводится иллюстративный пример. Библ. 35.

Ключевые слова: уравнение наблюдения, флуктуационная погрешность, сингулярная помеха, корреляционная матрица ошибок измерений, метод множителей Лагранжа, некорректное наблюдение, сингулярная помеха, оптимальное оценивание, условия несмещенности и инвариантности, декомпозиция, автокомпенсация, вычислительные алгоритмы распознавания.

DOI: 10.31857/S0044466924050011, **EDN**: YDNZJY

ВВЕДЕНИЕ

В различных областях человеческой деятельности (передача сообщений, локация, навигация, связь, радиотехническая разведка, радиоастрономия, техническая и медицинская диагностика, информационная безопасность и многих др.) возникает необходимость распознавания совокупности сигналов из заданного ансамбля с помощью различных информационно-измерительных систем. Под распознаванием понимается решение задач, связанных с оцениванием, обнаружением, различением и разрешением сигналов при различных уровнях априорной неопределенности (см., например, [1]–[23]).

Традиционно любая из задач распознавания решается в рамках статистического подхода (см. [1]–[12]), предполагающего знание соответствующих плотностей вероятности, функций распределения и отношений правдоподобия с учетом существенных и несущественных параметров и, как правило, вычисления порогов сравнения (для выбора оптимальных решений применительно к множеству возможных гипотез). При наличии априорной статистической информации осуществляется усреднение данных плотностей, функций и отношений по указанным параметрам, а при ее отсутствии используется процедура расширения пространства состояний, сопровождающаяся предварительной оценкой этих параметров по результатам наблюдений. В условиях значительной статистической неопределенности вводятся в рассмотрение классы возможных распределений, а при наличии априорной информации о вероятностях возможных альтернатив и возможных рисках (от применения тех или иных решений) вводятся байесовские решающие правила (см., например, [10], [13]). При наличии сведений о весах измерений, которые зависят от характеристик сенсоров, обработка данных может

осуществляться в рамках оптимизационного подхода (см. [14]). Очевидно, что в условиях указанной неопределенности возможности получения таких сведений зачастую весьма ограничены.

В настоящее время весьма плодотворно развиваются адаптивные методы распознавания сигналов применительно к интеллектуальным системам обработки и анализа многомерной информации (см., например, [11], [12], [15]–[21]), которые используют процедуры фильтрации, кластеризации, нечеткого представления данных, аппроксимации на основе нейронных сетей и др. Однако известные проблемы, связанные со сходимостью и ее скоростью, сложностью разбиения на классы и выбором оптимальных параметров кластеризации, обучаемостью, наличием достаточной базы знаний и др., не позволяют применять такие методы для условий существенной неопределенности и жестких требований к оперативности некоторых классов информационно-измерительных систем. Данные проблемы еще более усугубляются, если в наблюдениях присутствует сингулярная помеха с большим числом степеней свободы (речь идет о помехе, которая имеет конечномерное параметрическое представление в заданном функциональном пространстве и может иметь на интервале наблюдения точки разрыва первого рода). В этом случае требуется предварительная оценка всех параметров такой помехи. В ряде работ такую помеху еще называют динамической или сигналоподобной, а сами наблюдения некорректными.

Существует широкий круг задач распознавания сигналов, для которых рассмотренные выше подходы к распознаванию сигналов трудно реализуемы, особенно применительно к классу информационно-измерительных систем, которые должны функционировать в реальном времени и условиях существенной неопределенности (в том числе, и с некорректными наблюдениями при минимуме заданных статистических данных). Для задач оценивания с указанными ограничениями широко применяется метод наименьших квадратов (МНК), оперирующий лишь с известной корреляционной матрицей ошибок наблюдений и обеспечивающий, в соответствии с известной теоремой Гаусса-Маркова, построение наилучшей линейной оценки (см. [23]). Одним из классических подходов к решению задачи распознавания для указанных систем может служить расширенный МНК (РМНК), который основан на расширении пространства состояний и предполагает включение в общий вектор оцениваемых параметров не только искомых спектральных коэффициентов линейных разложений сигналов, но и аналогичных коэффициентов сингулярной помехи. Известно (см. [22], [23]), что применение РМНК на практике зачастую приводит к эффекту “размазывания точности”, который наиболее выражен в многомерных задачах оценивания, оперирующих с плохо обусловленными матрицами. Кроме того, такое расширение приводит к существенному росту вычислительных затрат и, как следствие, к снижению оперативности вычислений.

В [24], [25] развита автокомпенсационная параллельная процедура обобщенного инвариантно-несмещенного оценивания (ОИНО) значений линейных функционалов, которая является альтернативой РМНК и позволяет строить алгоритмы оптимального оценивания параметров одного сигнала в условиях некорректных наблюдений. Данная процедура, основанная на декомпозиции, не требует расширения и обеспечивает автокомпенсацию сингулярной помехи, не прибегая к оцениванию ее спектральных коэффициентов, а также позволяет организовать параллельные вычисления. Показан существенный вычислительный эффект.

В настоящей статье идея ОИНО получила дальнейшее развитие, направленное на решение гораздо более сложных задач, связанных с распознаванием совокупности сигналов в некорректных условиях наблюдения, при минимуме статистических данных о погрешности наблюдений. При этом термин “автокомпенсация” рассматривается в более широком смысле, поскольку каждый сигнал в уравнении наблюдения по отношению к другому сигналу этого же уравнения рассматривается как помеховый. Метод ориентирован на параллельные вычисления с учетом достигаемой декомпозиции разрабатываемых алгоритмов распознавания сигналов. Достижения в области таких вычислений позволяют добиться обработки наблюдений в реальном времени (см. [26]–[29]).

В статье не используются стохастические сигналы, например, марковские, которые характерны для теории линейной и нелинейной фильтраций, оперирующей с большим объемом достоверной статистической информации и приводящей во многих случаях к построению алгоритмов текущего оценивания с плохой сходимостью и(или) длительными переходными процессами, что не соответствует классу рассматриваемых ниже информационно-измерительных систем.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим задачу распознавания сигналов из заданного ансамбля $\{s_i(t)\}_{i=1}^{K}$, наблюдаемых в виде аддитивной смеси

$$h(t) = \sum_{i=1}^{K} q_i s_i(t) + \theta(t) + \xi(t), \quad q_i \in \{0,1\}, \quad t \in [0,T], \tag{1.1}$$

где t — непрерывное время, K — число возможных сигналов ансамбля, $\theta(t)$ — сингулярная помеха, $\xi(t)$ — флуктуационная погрешность, q_i — параметр, характеризующий отсутствие ($q_i = 0$) или присутствие ($q_i = 1$) сигнала $s_i(t)$ в смеси.

При $K = 1$ имеем задачу обнаружения одного сигнала, если в (1.1) при $K > 1$ присутствует только один сигнал, то речь идет о задаче различения, а при наличии в (1.1) любого числа сигналов — о задаче разрешения. Сюда входят частные случаи: $q_i = 0\ \forall i = \overline{1,K}$ (когда все сигналы ансамбля отсутствуют) и $q_i = 1\ \forall i = \overline{1,K}$ (когда все сигналы ансамбля присутствуют).

Для элементов смеси (1.1) используем следующие линейные конечномерные модели:

$$s_i(t) = A_i^{\mathrm{T}}\Psi_i(t), \tag{1.2}$$

$$\theta(t) = B^{\mathrm{T}}\Omega(t), \tag{1.3}$$

где $A_i = \bigl[a_{im}, m = \overline{1,M_i}\bigr]^{\mathrm{T}}$ и $B = \bigl[b_j, j = \overline{1,J}\bigr]^{\mathrm{T}}$ — неизвестные коэффициенты, $\Psi_i(t) = \bigl[\psi_{im}(t), m = \overline{1,M_i}\bigr]^{\mathrm{T}}$ и $\Omega(t) = \bigl[\omega_j(t), j = \overline{1,J}\bigr]^{\mathrm{T}}$ — заданные базисные функции.

При необходимости упомянутые задачи распознавания сигналов требуют еще оценки векторных параметров A_i и B. В нашем случае относительно статистических характеристик этих параметров никаких предположений не делается.

В дальнейшем будем использовать наиболее распространенное на практике векторное уравнение наблюдения для дискретного времени

$$H = \sum_{i=1}^{K} q_i S_i + \Theta + \Xi, \tag{1.4}$$

где $H = \bigl[h_n, n = \overline{1,N}\bigr]^{\mathrm{T}}$, $S_i = \bigl[s_{in}, n = \overline{1,N}\bigr]^{\mathrm{T}}$, $\Theta = \bigl[\theta_n, n = \overline{1,N}\bigr]^{\mathrm{T}}$, $\Xi = \bigl[\xi_n, n = \overline{1,N}\bigr]^{\mathrm{T}}$, $h_n = h(t_n)$, $s_{in} = s_i(t_n)$, $\theta_n = \theta(t_n)$, $\xi_n = \xi(t_n)$.

Полагаем, что погрешность Ξ характеризуется нулевым математическим ожиданием и соответствующей корреляционной матрицей K^{Ξ}. Закон распределения для погрешности Ξ далее не используется (по аналогии с МНК из [23]).

Для дискретного времени имеем

$$S_i = \Psi_i A_i, \tag{1.5}$$

$$\Theta = \Omega B, \tag{1.6}$$

где $\Psi_i = \bigl[\psi_{imn}, n = \overline{1,N}, m = \overline{1,M_i}\bigr]$ — базисная матрица сигнала S_i, $\psi_{imn} = \psi_{im}(t_n)$, $\Omega = \bigl[\omega_{jn}, n = \overline{1,N}$, $j = \overline{1,J}\bigr]$ — базисная матрица помехи Θ, $\omega_{jn} = \omega_j(t_n)$.

Также полагаем, что расширенный функциональный базис $\{\Psi_1(t), \ldots, \Psi_K(t), \Omega(t)\}$ линейно независим на сетке узлов $\{t_n\}_{n=1}^{N}$ (по аналогии с [24], [25]).

В самом общем случае задача распознавания сигналов предполагает оптимальное обнаружение каждого сигнала $s_i(t)$ из заданного ансамбля (т.е. вынесение оценки q_i^* для параметра q_i) и, в случае его обнаружения, нахождение оценки A_i^* для вектора A_i и оценки B^* для вектора B (если задача оценивания должна решаться). В таких условиях (многоальтернативных решений) возможно семейство гипотез Γ_l, $l \in \{1, \ldots, L\}$ (где $L = 2^K$), характеризующих все возможные варианты присутствия и отсутствия сигналов ансамбля в смеси (1.1). Под $\Gamma^0 \in \{\Gamma_1, \ldots, \Gamma_L\}$ далее понимается истинная гипотеза.

Каждой гипотезе Γ_l можно поставить в соответствие модельное наблюдение

$$H_l = \sum_{i=1}^{K_L} S_{il} + \Theta + \Xi, \quad l \in \{1, \ldots, L\}, \quad S_{il} \in \{S_1, \ldots, S_K\}, \tag{1.7}$$

где K_l — количество сигналов ансамбля, присутствующих в смеси согласно Γ_l.

С учетом (1.7) задача распознавания в рамках РМНК решается с использованием расширенного вектора спектральных коэффициентов $W_l = \bigl[A_{1l}^{\mathrm{T}}, \ldots, A_{K_l l}^{\mathrm{T}}, B^{\mathrm{T}}\bigr]^{\mathrm{T}}$ и критерия минимума квадратичной формы $\chi^{\text{рмнк}}(W_l)$:

$$\bigl(l^*, W_{l^*}\bigr) = \arg\min_{l, W_l} \chi^{\text{рмнк}}(W_l) = \arg\min_{l, W_l} \bigl[\Delta^{\text{рмнк}}(W_l)\bigr]^{\mathrm{T}} (K^{\Xi})^{-1} \Delta^{\text{рмнк}}(W_l), \tag{1.8}$$

где $\Delta^{\text{рмнк}}(W_l) = H - H^{\text{рмнк}}(W_l)$ — невязка.

Критерий (1.8) позволяет обеспечить минимизацию с учетом невязок $\Delta^{\text{рмнк}}(W_l)$ и весовой матрицы $(K^{\Xi})^{-1}$, при этом под W_{l^*} понимается оценка для W_l применительно к оптимальной гипотезе Γ_{l^*}, $l^* \in \{1, \ldots, L\}$.

Очевидно, что размерность такой задачи достаточно высока, и при работе с плохо обусловленными матрицами погрешности оценивания могут существенно превосходить методическую погрешность и обесценивать результаты оптимальной обработки наблюдений (это наглядно продемонстрировано в иллюстративном примере). Кроме того, критерий (1.8) не предусматривает возможности распараллеливания вычислительной процедуры.

Преодолеть недостатки РМНК во многом удается, если использовать модифицированную процедуру ОИНО, ориентированную на задачу распознавания сигналов. Для этого, применительно к фиксированному значению k, запишем наблюдение (1.7) в двух формах:

$$H_l = \begin{cases} S_{kl} + X_{kl} + \Xi, & k \in \{1, \ldots, K_l\}, \\ \Theta + X_l + \Xi, \end{cases} \tag{1.9}$$

где $S_{kl} = S_{kl}(A_{kl})$, $X_l = X_l(A_l)$, $A_L = \left[A_{1l}^{\mathrm{T}}, \ldots, A_{1K_l}^{\mathrm{T}}\right]^{\mathrm{T}}$,

$$X_{kl} = \sum_{\substack{i=1 \\ i \neq k}}^{K_l} S_{il} + \Theta, \quad X_l = \sum_{i=1}^{K_L} S_{il}.$$

Первая форма позволяет рассматривать X_{kl} как составляющую, мешающую оцениванию полезного сигнала S_{kl}, а вторая форма позволяет рассматривать X_l как составляющую, мешающую оцениванию помехи Θ.

Далее нам потребуются матрицы оптимального линейного оценивания $P_{kl}^S = \left[p_{krnl}^S, r = \overline{1,N}, n = \overline{1,N}\right]$, $P_{kl}^A = \left[p_{kmnl}^A, m = \overline{1,M_{kl}}, n = \overline{1,N}\right]$ и $P_l^\Theta = \left[p_{rnl}^\Theta, r = \overline{1,N}, n = \overline{1,N}\right]$, $P_l^B = \left[p_{jnl}^B, j = \overline{1,J}, n = \overline{1,N}\right]$ для формирования оптимальных оценок (на основе ОИНО для фиксированного значения l) применительно к сигналу S_{kl} и вектору A_{kl} его спектральных коэффициентов, а также к помехе Θ и вектору B ее спектральных коэффициентов

$$S_{kl}^* = P_{kl}^S H_l, \quad A_{kl}^* = P_{kl}^A H_l, \quad k = \overline{1,K_l}, \quad \Theta_l^* = P_l^\Theta H_l, \quad B_l^* = P_l^B H_l.$$

Матрицы P_{kl}^S, P_{kl}^A и P_l^Θ, P_l^B для гипотезы Γ_l должны обеспечивать выполнение следующих равенств:

$$\begin{aligned} H_{kl}^S &= P_{kl}^S H_l = P_{kl}^S S_{kl} + P_{kl}^S X_{kl} + P_{kl}^S \Xi = S_{kl} + \Xi_{kl}^S, \quad k = \overline{1,K_l}, \\ H_{kl}^A &= P_{kl}^A H_l = P_{kl}^A S_{kl} + P_{kl}^A X_{kl} + P_{kl}^A \Xi = A_{kl} + \Xi_{kl}^A, \quad k = \overline{1,K_l}, \\ H_l^\Theta &= P_l^\Theta H_l = P_l^\Theta \Theta + P_l^\Theta X_l + P_l^\Theta \Xi = \Theta + \Xi_l^\Theta, \\ H_l^B &= P_l^B H_l = P_l^B \Theta + P_l^B X_l + P_l^B \Xi = B + \Xi_l^B, \end{aligned} \tag{1.10}$$

где $\Xi_{kl}^S = P_{kl}^S \Xi$ и $\Xi_l^\Theta = P_l^\Theta \Xi$ — шумы с нулевыми математическими ожиданиями

$$\begin{aligned} M\{\Xi_{kl}^S\} = M\{P_{kl}^S \Xi\} = P_{kl}^S M\{\Xi\} = [0]_{N\times 1}, \quad M\{\Xi_{kl}^A\} = M\{P_{kl}^A \Xi\} = P_{kl}^A M\{\Xi\} = [0]_{M_{kl}\times 1}, \\ M\{\Xi_l^\Theta\} = M\{P_l^\Theta \Xi\} = P_l^\Theta M\{\Xi\} = [0]_{N\times 1}, \quad M\{\Xi_l^B\} = M\{P_l^B \Xi\} = P_l^B M\{\Xi\} = [0]_{J\times 1} \end{aligned}$$

(здесь $M\{\cdot\}$ — символ математического ожидания, $[0]_{N\times 1}$ и $[0]_{M_{kl}\times 1}$ — нулевые вектор-столбцы соответствующей размерности, которая указывается квадратными скобками с нижними индексами).

Кроме того, для корреляционных матриц $K_{kl}^{\Xi S}$, $K_{kl}^{\Xi A}$ и $K_l^{\Xi\Theta}$, $K_l^{\Xi B}$ случайных векторов Ξ_{kl}^S, Ξ_{kl}^A и Ξ_l^Θ, Ξ_l^B должны обеспечиваться соответствующие условия минимума

$$\begin{aligned} SpK_{kl}^{\Xi S} &\to \min_{P_{kl}^S}, \quad k = \overline{1,K_l}, \\ SpK_{kl}^{\Xi A} &\to \min_{P_{kl}^A}, \quad k = \overline{1,K_l}, \\ SpK_l^{\Xi\Theta} &\to \min_{P_l^\Theta}, \\ SpK_l^{\Xi B} &\to \min_{P_l^B}, \end{aligned} \tag{1.11}$$

где под Sp понимается оператор нахождения следа матрицы.

Формула (1.10) отражает свойства несмещенности (по отношению к параметрам полезных сигналов и сингулярной помехи)

$$\begin{aligned} &P_{kl}^S S_{kl} = S_{kl}, \quad k = \overline{1, K_l}, \\ &P_{kl}^A S_{kl} = A_{kl}, \quad k = \overline{1, K_l}, \\ &P_l^\Theta \Theta = \Theta, \\ &P_l^B \Theta = B, \end{aligned} \tag{1.12}$$

а также инвариантности (по отношению к мешающим составляющим X_{kl} и X_l)

$$\begin{aligned} &P_{kl}^S X_{kl} = [0]_{N\times 1}, \quad k = \overline{1, K_l}, \\ &P_{kl}^A X_{kl} = [0]_{M_{kl}\times 1}, \quad k = \overline{1, K_l}, \\ &P_l^\Theta X_l = [0]_{N\times 1}, \\ &P_l^B X_l = [0]_{J\times 1}. \end{aligned} \tag{1.13}$$

Если матрицы P_{kl}^S, P_{kl}^A и P_l^Θ, P_l^B применять непосредственно к наблюдению (1.4), то получаем набор оптимальных оценок всех параметров задачи распознавания сигналов для фиксированного значения l:

$$\begin{aligned} S_{kl}^* &= P_{kl}^S H = \sum_{i=1}^K q_i P_{kl}^S S_i + P_{kl}^S \Theta + \Xi_{kl}^S, \quad k = \overline{1, K_l}, \\ A_{kl}^* &= P_{kl}^A H = \sum_{i=1}^K q_i P_{kl}^A S_i + P_{kl}^A \Theta + \Xi_{kl}^A, \quad k = \overline{1, K_l}, \\ \Theta_l^* &= P_l^\Theta H = \sum_{i=1}^K q_i P_l^\Theta S_i + P_l^\Theta \Theta + \Xi_l^\Theta, \\ B_l^* &= P_l^B H = \sum_{i=1}^K q_i P_l^B S_i + P_l^B \Theta + \Xi_l^B. \end{aligned} \tag{1.14}$$

В отличие от РМНК оценки параметров сигналов и помехи (1.14), формируемые в рамках ОИНО, осуществляются раздельно, что позволяет организовать параллельные вычисления, кроме того, выполнение условий инвариантности позволяет существенно снизить размерность обращаемых матриц (по аналогии с [24], [25]). Это хорошо продемонстрировано далее в иллюстративном примере.

Для истинной гипотезы Γ^0 с номером $l^0 \in \{1, 2, \ldots, L\}$, учитывая (1.12)–(1.14), получим

$$\begin{aligned} &S_{kl^0}^* = P_{kl^0}^S H = q_k S_k + \Xi_{kl^0}^S, \quad k = \overline{1, K_{l^0}}, \\ &A_{kl^0}^* = P_{kl^0}^A H = q_k A_k + \Xi_{kl^0}^A, \quad k = \overline{1, K_{l^0}}, \\ &\Theta_{l^0}^* = P_{l^0}^\Theta H = \Theta + \Xi_{l^0}^\Theta, \\ &B_{l^0}^* = P_{l^0}^B H = B + \Xi_{l^0}^B. \end{aligned} \tag{1.15}$$

Если Γ_l не соответствует истинной гипотезе Γ^0, то условия (1.12) и (1.13) нарушаются, что приводит к невязке

$$\Delta^{\text{оино}}(l) = H - H^{\text{оино}}\big(S_l^*, \Theta_l^*\big) = H - \sum_{i=1}^{K_L} S_{il}^* - \Theta_l^*,$$

где $S_l^* = \big[(S_{il}^*)^{\mathrm{T}}, i = \overline{1, K_l}\big]^{\mathrm{T}}$.

В этом случае задача распознавания в рамках ОИНО решается на основе критерия минимума квадратичной формы:

$$(l^*) = \arg\min_l \chi^{\text{оино}}(l) = \arg\min_l \big[\Delta^{\text{оино}}(l)\big]^{\mathrm{T}} \Big(K^\Xi\Big)^{-1} \Delta^{\text{оино}}(l), \tag{1.16}$$

при этом результирующие оценки параметров сигналов и сингулярной помехи находятся как $S_{l^*}^* = S_{l=l^*}^*$, $A_{l^*}^* = A_{l=l^*}^*$ и $\Theta_{l^*}^* = \Theta_{l=l^*}^*$, $B_{l^*}^* = B_{l=l^*}^*$, где $A_l^* = \big[(A_{kl}^*)^{\mathrm{T}}, k = \overline{1, K_l}\big]^{\mathrm{T}}$.

Формулы (1.1)–(1.16) полностью задают все модели, ограничения и критерии, необходимые для разработки нового метода разрешения сигналов в условиях существенной априорной неопределенности, а также его сравнения с РМНК. Требуется: построить матрицы P^S_{kl}, P^A_{kl} и P^Θ_l, P^B_l для фиксированного l; с учетом построенных матриц и принятого критерия оптимальности в декомпозированном виде решить задачу оптимального распознавания сигналов без традиционного расширения пространства состояний в условиях минимума априорной статистической информации (используя лишь матрицу K^Ξ), обеспечив автокомпенсацию сингулярной помехи и параллельную обработку наблюдений; привести формулы для случайных и методических ошибок результирующего оценивания; сравнить разработанный метод с РМНК в плане вычислительной эффективности; продемонстрировать возможность его сравнения с известными статистическими методами распознавания сигналов; привести иллюстративный пример, подтверждающий преимущества разработанного метода по сравнению с РМНК.

2. ПОСТРОЕНИЕ МАТРИЦ ОПТИМАЛЬНОГО ЛИНЕЙНОГО АВТОКОМПЕНСАЦИОННО-ДЕКОМПОЗИЦИОННОГО ОЦЕНИВАНИЯ

Предположим сначала, что в смеси (1.1) присутствуют все K сигналов ансамбля, т.е. $q_1 = \cdots = q_K = 1$ (в этом случае индекс l можно опустить). Тогда решение задачи распознавания можно искать в классе линейных оценок в виде K параллельных алгоритмов вычислений:

$$A^*_k = P^A_k H, \quad k = \overline{1,K}, \tag{2.1}$$

где A^*_k — оценка вектора A_k, $P^A_k = \left[p^A_{kmn}, m = \overline{1,M}, n = \overline{1,N}\right]$ — матрица неизвестных весовых коэффициентов оптимального оценивания.

Корреляционная матрица ошибок оценивания на основе (2.1) находится по правилу

$$K^A_k = P^A_k K^\Xi \left(P^A_k\right)^{\mathrm{T}}, \quad k = \overline{1,K}. \tag{2.2}$$

Критерий качества, необходимый для нахождения P^A_k, соответствует минимизации следа матрицы SpK^A_k (см. (1.11)). Кроме того, должны выполняться сопутствующие условия несмещенности (1.12) и инвариантности (1.13).

Для нахождения матрицы P^A_k преобразуем (1.9) к следующему виду:

$$H = S_k + X_k + \Xi = S_k + Y_k C_k + \Xi, \tag{2.3}$$

где $X_k = \left[x_{kn}, n = \overline{1,N}\right]^{\mathrm{T}}$,

$Y_k = \left[\Psi_1 \vdots \ldots \vdots \Psi_{k-1} \vdots \ldots \vdots \Psi_{k+1} \vdots \ldots \vdots \Psi_K \vdots \Omega\right]$ — матрица размером $N \times \left(\overline{M}_k + J\right)$,

$C_k = \left[A^{\mathrm{T}}_1 \vdots \ldots \vdots A^{\mathrm{T}}_{k-1} \vdots \ldots \vdots A^{\mathrm{T}}_{k+1} \vdots \ldots \vdots A^{\mathrm{T}}_K \vdots B^{\mathrm{T}}\right]^{\mathrm{T}}$ — вектор размером $\left(\overline{M}_k + J\right) \times 1$,

$\overline{M}_k = M_1 + \cdots + M_{k-1} + M_{k+1} + \cdots + M_K$.

Применительно к рассматриваемому случаю условия несмещенности и инвариантности (с учетом (1.12), (1.13) и (2.3)) можно записать так:

$$P^A_k \Psi_k - [E]_{M_k \times M_k} = [0]_{M_k \times M_k}, \tag{2.4}$$

$$P^A_k Y_k = [0]_{M_k \times \left(\overline{M}_k + J\right)}, \tag{2.5}$$

где $[0]_{M_k \times M_k}$ и $[E]_{M_k \times M_k}$ — нулевая и единичная матрицы соответственно, $[\cdot]_{M_k \times M_k}$ — обозначение размерности матрицы, стоящей в квадратных скобках (такое обозначение используется далее по всей статье).

Для дальнейшего изложения потребуется вектор $P^A_{km} = \left[P^A_{kmn}, n = \overline{1,N}\right]^{\mathrm{T}}$ который состоит из элементов m-й строки матрицы P^A_k. Он позволяет найти скалярную оценку a^*_{km} коэффициента a_{km} для фиксированных значений k и m. Очевидно, что выполняются условия несмещенности $\left(P^A_{km}\right)^{\mathrm{T}} S_k = a_{km}$ и инвариантности $\left(P^A_{km}\right)^{\mathrm{T}} X_k = 0$. Или, по аналогии с (2.4) и (2.5), имеем

$$\left(P^A_{km}\right)^{\mathrm{T}} \Psi_k - E^{\mathrm{T}}_{km} = [0]_{1 \times M_k}, \tag{2.6}$$

$$Y^{\mathrm{T}}_k P^A_{km} = [0]_{\left(\overline{M}_{kl} + J\right) \times 1}, \tag{2.7}$$

где E_{km} — вектор-столбец, состоящий из нулей, но на m-й позиции стоит единица.

Задачу оптимального оценивания коэффициента a_{km} будем решать методом множителей Лагранжа путем нахождения экстремума следующей функции:

$$F\big(P^A_{km}, \gamma_{km}, \eta_{km}\big) = \big(P^A_{km}\big)^{\mathrm{T}} K^{\Xi} P^A_{km} + \gamma^{\mathrm{T}}_{km} Y^{\mathrm{T}}_k P^A_{km} + \Big[\big(P^A_{km}\big)^{\mathrm{T}} \Psi_k - E^{\mathrm{T}}_{km}\Big]\eta_{km}, \tag{2.8}$$

$\gamma_{km} = \big[\gamma_{kmn}, n = \overline{1, \overline{M}_k + J}\big]^{\mathrm{T}}$ и $\eta_{km} = \big[\eta_{kmn}, n = \overline{1, M_k}\big]^{\mathrm{T}}$ — векторные множители Лагранжа, соответствующие условиям несмещенности (2.6) и инвариантности (2.7).

Дифференцируя функцию (2.8) по всем аргументам, получаем систему линейных алгебраических уравнений

$$\begin{aligned}
&\partial F/\partial P^A_{km} = 2K^{\Xi} P^A_{km} + Y_k \gamma_{km} + \Psi_k \eta_{km} = [0]_{N\times 1},\\
&\partial F/\partial \gamma_{km} = Y^{\mathrm{T}}_k P^A_{km} = [0]_{(\overline{M}_k + J)\times 1},\\
&\partial F/\partial \eta_{km} = \Psi^{\mathrm{T}}_k P^A_{km} - E_{km} = [0]_{M_k\times 1},
\end{aligned}$$

Введем следующие матрицы:

$$V_k = \big(K^{\Xi}\big)^{-1}\Psi_k,\ \overline{V}_k = \big(K^{\Xi}\big)^{-1} Y_k,\ \boldsymbol{\Phi}_k = \Psi^{\mathrm{T}}_k V_k,\ \bar{\boldsymbol{\Phi}}_k = Y^{\mathrm{T}}_k \overline{V}_k,\ Z_k = Y^{\mathrm{T}}_k V_k,\ \bar{Z}_k = \Psi^{\mathrm{T}}_k \overline{V}_k.$$

С учетом полученной системы уравнений и принятых обозначений находим строку весовых коэффициентов

$$P^A_{km} = 2^{-1}\big(V_k \eta_{km} - \overline{V}_k \gamma_{km}\big). \tag{2.9}$$

Умножая левую и правую части (2.9) слева на матрицу Y^{T}_k и учитывая условие инвариантности (2.7), после несложных преобразований находим

$$\gamma_{km} = \bar{\boldsymbol{\Phi}}^{-1}_k Z_k \eta_{km}. \tag{2.10}$$

Аналогичным умножением (2.9) на матрицу Ψ^{T}_k и учитывая условие несмещенности (2.6) получаем

$$\eta_{km} = 2\boldsymbol{\Phi}_k\big(E_{km} + 2^{-1}\bar{Z}_k \gamma_{km}\big). \tag{2.11}$$

Разрешая (2.10) и (2.11) относительно γ_{km} и η_{km}, имеем

$$\gamma_{km} = 2\bar{\boldsymbol{\Phi}}^{-1}_k Z_k \big([E]_{M_k\times M_k} - \boldsymbol{\Phi}^{-1}_k \bar{Z}_k \bar{\boldsymbol{\Phi}}^{-1}_k Z_k\big)^{-1} \boldsymbol{\Phi}^{-1}_k E_{km}, \tag{2.12}$$

$$\eta_{km} = 2\big([E]_{M_k\times M_k} - \boldsymbol{\Phi}^{-1}_k \bar{Z}_k \bar{\boldsymbol{\Phi}}^{-1}_k Z_k\big)^{-1} \boldsymbol{\Phi}^{-1}_k E_{km}. \tag{2.13}$$

Подставляя (2.12) и (2.13) в (2.9), получаем

$$P^A_{km} = \big(V_k - \overline{V}_k \bar{\boldsymbol{\Phi}}^{-1}_k Z_k\big)\big([E]_{M_k\times M_k} - \boldsymbol{\Phi}^{-1}_k \bar{Z}_k \bar{\boldsymbol{\Phi}}^{-1}_k Z_k\big)^{-1} \boldsymbol{\Phi}^{-1}_k E_{km}. \tag{2.14}$$

Вводя обозначение $\Lambda_k = [E]_{N\times N} - \overline{V}_k \bar{\boldsymbol{\Phi}}^{-1}_k Y^T_k$, формулу (2.14) запишем в следующем компактном виде:

$$P^A_{km} = \Lambda_k V_k \big(\Psi^{\mathrm{T}}_k \Lambda_k V_k\big)^{-1} E_{km}. \tag{2.15}$$

С учетом (2.15) скалярная оценка a^*_{km} коэффициента a_{km} находится так:

$$a^*_{km} = H^{\mathrm{T}} P^A_{km} = H^{\mathrm{T}} \Lambda_k V_k \big(\Psi^{\mathrm{T}}_k \Lambda_k V_k\big)^{-1} E_{km}. \tag{2.16}$$

Переходя от скалярного коэффициента a_{km} к вектору A_k с учетом (2.15) получаем матрицу оптимального оценивания

$$P^A_k = \Big[\Lambda_k V_k \big(\Psi^{\mathrm{T}}_k \Lambda_k V_k\big)^{-1}\Big]^{\mathrm{T}}, \quad k = \overline{1, K}. \tag{2.17}$$

Подставляя (2.17) в (2.1), находим искомые оценки

$$A^*_k = P^A_k H = \Big[\Lambda_k V_k \big(\Psi^{\mathrm{T}}_k \Lambda_k V_k\big)^{-1}\Big]^{\mathrm{T}} H, \quad k = \overline{1, K}. \tag{2.18}$$

В свою очередь, матрицу оптимального оценивания сигналов S_{kl} находим как

$$P^S_k = \Psi_k P^A_k, \tag{2.19}$$

а саму оценку в виде

$$S_k^* = \Psi_k A_k^* = \Psi_k \left[\Lambda_k V_k \left(\Psi_k^{\mathrm{T}} \Lambda_k V_k\right)^{-1}\right]^{\mathrm{T}} H, \quad k = \overline{1,K}. \tag{2.20}$$

Оценки (2.18) и (2.20) являются оптимальными в смысле несмещенности, эффективности (обеспечивают минимальную дисперсию) и инвариантности (по отношению к результирующей сингулярной помехе).

Для решения ряда задач распознавания сигналов, помимо матриц P_k^A и P_k^S, возникает необходимость построения матриц P^B и P^Θ оптимального оценивания параметров помехи с соблюдением условий несмещенности и инвариантности (по аналогии с (2.6) и (2.7)). Теперь вместо Y_{kl} и C_{kl} строятся матрицы Y_k^Θ и C_k^Θ с учетом того, что при нахождении оценок B^* и Θ^* все полезные сигналы рассматриваются как мешающие составляющие. Формулы для матриц P^B, P^Θ и оценок B^*, Θ^* записываются по аналогии с (2.17), (2.18) и (2.19), (2.20). Для задач распознавания с учетом возможных гипотез Γ_l строится семейство матриц P_k^A, P_k^S и P_k^B, P_k^Θ оптимального оценивания для всех значений l.

Рассмотренных в данном разделе матриц оценивания необходимо и достаточно для решения широкого круга задач, связанных с распознаванием сигналов, в условиях существенной априорной неопределенности. В следующем разделе рассматриваются наиболее распространенные задачи распознавания: оценивание (известно какие сигналы в наблюдении присутствуют и требуется только оценить их параметры); обнаружение (надо установить присутствует ли в наблюдении полезный сигнал); различение (в наблюдении присутствует только один сигнал из заданного ансамбля, и надо установить какой именно); разрешение (в наблюдении могут присутствовать те или иные сигналы ансамбля, и надо установить какие). Кроме того, задачи обнаружения, различения и разрешения могут сопровождаться оцениванием параметров сигналов и помехи.

3. ВЫЧИСЛИТЕЛЬНЫЕ АЛГОРИТМЫ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ СИГНАЛОВ

3.1. Алгоритм оценивания сигналов 1

Пусть в смеси (1.1) присутствуют все сигналы ансамбля, т.е. $q_1 = \cdots = q_K = 1$. Требуется дать оценку всех коэффициентов A_k и самих сигналов S_k, не прибегая к оценке коэффициента B помехи Θ, т.е. без расширения пространства состояний. В данной задаче гипотезы не используются, поэтому индекс l опускаем.

Шаг 1.1. Строим матрицы P_k^A весовых коэффициентов.

Шаг 1.2. Находим оценки $A_k^* = P_k^A H$ векторных коэффициентов A_k, $k = \overline{1,K}$.

Шаг 1.3. Строим матрицы $P_k^S = \Psi_k P_k^A$, $k = \overline{1,K}$.

Шаг 1.4. Находим оценки $S_k^* = P_k^S H$ самих сигналов S_k, $k = \overline{1,K}$.

3.2. Алгоритм оценивания сигналов и помехи 2

Пусть в смеси (1.1) присутствуют все сигналы ансамбля, т.е. $q_1 = \cdots = q_K = 1$. Требуется дать оценки всех коэффициентов A_k и самих сигналов S_k, а также оценки коэффициента B и самой помехи Θ. В данной задаче гипотезы не используются, поэтому индекс опускаем.

Шаг 2.1. Строим матрицы P_k^A, P_k^S и P^B, P^Θ.

Шаг 2.2. Находим результирующие оценки A_k^*, S_k^* и B^*, Θ^*.

3.3. Алгоритм совместного обнаружения-оценивания 3

В этом случае $K = 1$, $L = 2$, и, следовательно, можно записать $H = qS + \Theta + \Xi$, т.е. в зависимости от значения коэффициента q возможны две гипотезы (Γ_1, если $q = 0$ и Γ_2, если $q = 1$). Надо дать оценку $l^* \in \{1,2\}$ для параметра $l \in \{1,2\}$, а также построить оценки A_l^*, S_l^* и B_l^*, Θ_l^*. В этом случае индекс k можно опустить.

Шаг 3.1. Строим матрицы $P_{l=1}^A$, $P_{l=2}^A$ и $P_{l=1}^S$, $P_{l=2}^S$ для гипотез Γ_1 и Γ_2 соответственно.

Шаг 3.2. Находим частные оценки $A_{l=1}^*$, $S_{l=1}^*$ (для гипотезы Γ_1) и $A_{l=2}^*$, $S_{l=2}^*$ (для гипотезы Γ_2).

Шаг 3.3. Строим матрицы $P_{l=1}^B$, $P_{l=1}^\Theta$ и $P_{l=2}^B$, $P_{l=2}^\Theta$ для гипотез Γ_1 и Γ_2 соответственно.

Шаг 3.4. Находим частные оценки $B_{l=1}^*$, $\Theta_{l=1}^*$ и $B_{l=2}^*$, $\Theta_{l=2}^*$.

Шаг 3.5. Находим невязки $\Delta^{\text{оино}}(l) = H - S_l^* - \Theta_l^*$ и номер наилучшей гипотезы с использованием критерия

$$l^* = \arg\min_l \chi^{\text{оино}}(l) = \arg\min_l \left[\Delta^{\text{оино}}(l)\right]^{\mathrm{T}} \left(K^\Xi\right)^{-1} \Delta^{\text{оино}}(l), \quad l^* \in \{1,2\}.$$

Шаг 3.6. Находим результирующие оценки $A_{l^*}^*$, $S_{l^*}^*$ и $B_{l^*}^*$, $\Theta_{l^*}^*$.

3.4. Алгоритм совместного различения и оценивания параметров сигналов и помехи 4

В этом случае K произвольное, $L = K$, $H = S_l + \Theta + \Xi$, $l \in \{1, 2, \ldots, L\}$. Надо установить, какой сигнал из заданного ансамбля $\{S_1, S_2, \ldots, S_L\}$ присутствует в наблюдении, т.е. вынести оценку $l^* \in \{1, 2, \ldots, L\}$ для параметра l а также построить оценки $A^*_{l^*}$, $S^*_{l^*}$ и $B^*_{l^*}$, $\Theta^*_{l^*}$. В этом случае индекс k опускаем.

Шаг 4.1. Строим матрицы P^A_l, P^S_l и P^B_l, P^Θ_l для всех гипотез $\mathit{\Gamma}_l$.

Шаг 4.2. Находим частные оценки A^*_l, S^*_l и B^*_l, Θ^*_l для всех гипотез $\mathit{\Gamma}_l$.

Шаг 4.3. Находим невязки $\Delta^{\text{оино}}(l) = H - S^*_l - \Theta^*_l$ для всех гипотез $\mathit{\Gamma}_l$.

Шаг 4.4. Находим номер наилучшей гипотезы с использованием следующего критерия:

$$l^* = \arg\min_l \chi^{\text{оино}}(l) = \arg\min_l \left[\Delta^{\text{оино}}(l)\right]^{\mathrm{T}} \left(K^{\Xi}\right)^{-1} \Delta^{\text{оино}}(l).$$

Шаг 4.5. Находим результирующие оценки $A^*_{l^*}$, $S^*_{l^*}$ и $B^*_{l^*}$, $\Theta^*_{l^*}$ для гипотезы $\mathit{\Gamma}_{l^*}$.

3.5. Алгоритм совместного разрешения и оценивания параметров сигналов и помехи 5

Рассматривается общий случай (1.1).

Шаг 5.1. Строим матрицы P^A_{kl}, P^S_{kl} и P^B_l, P^Θ_l для всех гипотез $\mathit{\Gamma}_l$ и k.

Шаг 5.2. Находим частные оценки A^*_{kl}, S^*_{kl} и B^*_l, Θ^*_l для всех гипотез $\mathit{\Gamma}_l$ и k.

Шаг 5.3. Находим невязки $\Delta^{\text{оино}}(l) = H - \sum_{k=1}^{K_l} S^*_{kl} - \Theta^*_l$ для всех гипотез $\mathit{\Gamma}_l$.

Шаг 5.4. Находим номер наилучшей гипотезы с использованием следующего критерия:

$$l^* = \arg\min_l \chi^{\text{оино}}(l) = \arg\min_l \left[\Delta^{\text{оино}}(l)\right]^{\mathrm{T}} \left(K^{\Xi}\right)^{-1} \Delta^{\text{оино}}(l).$$

Шаг 5.5. Находим результирующие оценки $A^*_{kl^*}$, $S^*_{kl^*}$ (где $k = \overline{1, K_{l^*}}$) и $B^*_{l^*}$, $\Theta^*_{l^*}$ для гипотезы $\mathit{\Gamma}_{l^*}$.

Замечание. Использованные в алгоритмах критерии оптимизации обеспечивают минимизацию влияния друг на друга соседних сигналов ансамбля, автокомпенсацию сингулярной помехи и сглаживание шума (потенциально сравнимое с возможностями РМНК).

Рассмотренные алгоритмы лишь иллюстрируют некоторые возможности развиваемого метода. Возможны и другие постановки задачи распознавания сигналов в условиях неопределенности с учетом особенностей назначения и применения рассматриваемой информационно-измерительной системы. Очевидно, что предложенный метод можно комплексировать и с традиционными вероятностными подходами в зависимости от условий функционирования системы.

Необходимость обработки наблюдений для множества гипотез приводит к необходимости организации 2^L каналов параллельных вычислений. Для современных информационно-измерительных систем, ориентированных на режим функционирования в реальном времени, предлагаемый метод может оказаться весьма перспективным.

4. К АНАЛИЗУ ХАРАКТЕРИСТИК РАСПОЗНАВАНИЯ

В силу линейности предлагаемого метода существенно упрощается нахождение корреляционных матриц ошибок оценивания. Так, с учетом (2.17) и (2.18) находим выражение для корреляционной матрицы оценки

$$K^A_{kl^*} = \left[\Lambda_{kl^*} V_{kl^*} \left(\Psi^{\mathrm{T}}_{kl^*} \Lambda_{kl^*} V_{kl^*}\right)^{-1}\right]^{\mathrm{T}} K^{\Xi} \left[\Lambda_{kl^*} V_{kl^*} \left(\Psi^{\mathrm{T}}_{kl^*} \Lambda_{kl^*} V_{kl^*}\right)^{-1}\right], \quad k = \overline{1, K_{l^*}}. \tag{4.1}$$

Выражение (4.1) позволяет в каждом конкретном случае оценить потенциальные возможности метода с учетом требований, задаваемых к системе. По аналогии с (4.1) можно записать математические формулы для корреляционных матриц $K^S_{kl^*}$, $K^B_{l^*}$ и $K^\Theta_{l^*}$, характеризующих точности оценивания не только отсчетов сигналов и их спектральных коэффициентов, но и сингулярной помехи. Дисперсии ошибок оценивания скалярных координат a_{km} вектора A_k и отсчетов s_{kn} сигнала S_k находятся так:

$$\left(\sigma^A_{kml^*}\right)^2 = \left(P^A_{kml^*}\right)^{\mathrm{T}} K^{\Xi} P^A_{kml^*}, \quad m = \overline{1, M_{kl^*}},$$
$$\left(\sigma^S_{knl^*}\right)^2 = \left(P^S_{knl^*}\right)^{\mathrm{T}} K^{\Xi} P^S_{knl^*}, \quad n = \overline{1, N}.$$

Для оценки характеристик обнаружения, различения и разрешения сигналов, а также возможности сравнения развитого и известных методов, достаточно задаться наиболее подходящим (для рассматриваемых условий наблюдения) законом распределения флуктуационных погрешностей (как правило, гауссовским), а также при

необходимости некоторыми априорными вероятностями (например, появления полезных сигналов в наблюдении) и значениями рисков от неправильных решений. Это позволяет исследовать характеристики метода в рамках известных статистических подходов (см. [1], [2], [4], [5], [10]), для чего строятся соответствующие функции правдоподобия и на их основе находятся значения наиболее важных (для конкретной задачи) линейных функционалов (например, вероятностей ложной тревоги, правильного обнаружения и т.д.).

Анализ выражений (2.14) и (2.17) показывает, что в развиваемом методе требуется обращение матрицы $\bar{\Phi}_{kl}$ размером $\left(\overline{M}_{kl}+J\right)\times\left(\overline{M}_{kl}+J\right)$, а также матрицы $\Psi_{kl}^{\mathrm{T}}\Lambda_{kl}V_{kl}$ размером $M_{kl}\times M_{kl}$. Применительно к РМНК для фиксированного k предполагается обращение матрицы большего размера $\left(M_{1l}+\cdots+M_{K_l l}+J\right)\times\left(M_{1l}+\cdots\right.$ $\left.\cdots+M_{K_l l}+J\right)$. Очевидно, что при плохой обусловленности обращаемых матриц предлагаемый метод может оказаться весьма эффективным в вычислительном плане.

Пусть модельное уравнение наблюдения имеет вид

$$H_l=(S_{kl}+\Delta S_{kl})+(X_{kl}+\Delta X_{kl})+\Xi,$$

где ΔS_{kl} и ΔX_{kl} — добавки к сигналу и помехе, обусловленные учетом “хвостов” используемых функциональных рядов.

В этом случае для оценки a_{kml}^{*} (вычисленной без учета этих добавок) имеем

$$a_{kml}^{*}=P_{kml}^{\mathrm{T}}H_l=P_{kml}^{\mathrm{T}}(S_{kl}+\Delta S_{kl})+P_{kml}^{\mathrm{T}}(X_{kl}+\Delta X_{kl})+P_{kml}^{\mathrm{T}}\Xi,$$

Соответственно для истинного значения a_{kml} справедливо представление

$$a_{kml}=(P_{kml}+\Delta P_{kml})^{\mathrm{T}}(S_{kl}+\Delta S_{kl})+(P_{kml}+\Delta P_{kml})^{\mathrm{T}}(X_{kl}+\Delta X_{kl}),$$

где ΔP_{kml} — добавка к столбцу весовых коэффициентов, необходимая для учета указанных “хвостов”.

Если использовать символ математического ожидания $M\{\cdot\}$, то в качестве среднего значения методической ошибки можно принять величину

$$\Delta_{kml}=M\{a_{kml}-a_{kml}^{*}\}=\Delta P_{kml}^{\mathrm{T}}(S_{kl}+\Delta S_{kl})+\Delta P_{kml}^{\mathrm{T}}(X_{kl}+\Delta X_{kl}),\tag{4.2}$$

где учтено, что $M\{\Xi\}=0$.

Используя (4.2) совместно с (4.1), можно подобрать необходимые параметры развиваемого метода, обеспечивающие минимизацию результирующей ошибки оценивания в каждом конкретном случае.

Необходимые и достаточные условия существования и единственности решения задачи оценивания по аналогии с [24], [25] требуют невырожденности и соблюдения некоторых ограничений на ранги ряда матриц. Выполнение заданных условий на практике обеспечивается рациональным выбором используемых функциональных базисов, числа степеней свободы в моделях сигналов и сингулярной помехи, а также заданием соответствующих условий наблюдения. Все эти вопросы относятся к планированию вычислительного эксперимента и далее не рассматриваются, поскольку требуют отдельных исследований в каждом конкретном случае.

Для оценки вычислительной эффективности развитого метода достаточно воспользоваться результатами работы [24], в которой демонстрируется возможность реализации процедуры ОИНО на основе распределенной обработки данных. В качестве показателя вычислительной эффективности метода можно принять время, затрачиваемое на получение искомых оценок. Данное время определяется быстродействием распределенной среды, общим числом операций, необходимых при реализации метода, и способом программирования. В [24] показано, что поскольку процедура ОИНО не требует расширения пространства состояний, то реализуемые на ее основе методы оценивания позволяют достичь значительного выигрыша в вычислительной эффективности. В [24] также дана количественная оценка достигаемого выигрыша на конкретном примере.

5. ИЛЛЮСТРАТИВНЫЕ ПРИМЕРЫ

Пример 1 (для алгоритма 1). Пусть в смеси (1.1) присутствуют все сигналы ансамбля, т.е. $q_1=\cdots=q_k=1$. Требуется дать оценку всех коэффициентов A_k и самих сигналов S_k, не прибегая к оценке коэффициента B помехи Θ, т.е. без расширения пространства состояний. В данной задаче гипотезы не используются, поэтому индекс l опускаем.

Пример искусственно выбран таким, чтобы он был достаточно простым и, в то же время, наглядно демонстрировал достигаемый вычислительный эффект развиваемого метода по сравнению с РМНК. Для этой цели использованы исходные данные, приводящие к задаче оценивания параметров сигналов с плохо обусловленными матрицами. Пусть $K=2$, $H=S_1+S_2+\Theta+\Xi$, где $A_1=[a_{11},a_{12},a_{13},a_{14}]^{\mathrm{T}}$, $\Psi_1(t)=[1,t^2,t^3,t^5]^{\mathrm{T}}$, $M_1=4$,

$S_1 = A_1^{\mathrm{T}}\Psi_1$, $A_2 = [a_{21}, a_{22}, a_{23}]^{\mathrm{T}}$, $\Psi_2(t) = [t, t^4, t^6]^{\mathrm{T}}$, $M_2 = 3$, $S_2 = A_2^{\mathrm{T}}\Psi_2$, $B = [b_1, b_2]^{\mathrm{T}}$, $\Omega(t) = \big[\omega_1(t), \omega_2(t)\big]^{\mathrm{T}}$, $J = 2$,

$$\Psi_1 = \begin{bmatrix} 1 & t_1^2 & t_1^3 & t_1^5 \\ 1 & t_2^2 & t_2^3 & t_2^5 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_N^2 & t_N^3 & t_N^5 \end{bmatrix}, \quad \Psi_2 = \begin{bmatrix} t_1 & t_1^4 & t_1^6 \\ t_2 & t_2^4 & t_2^6 \\ \vdots & \vdots & \vdots \\ t_N & t_N^4 & t_N^6 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \omega_1(t_1) & \omega_2(t_1) \\ \omega_1(t_2) & \omega_2(t_2) \\ \vdots & \vdots \\ \omega_1(t_N) & \omega_2(t_N) \end{bmatrix}, \quad X_1 = Y_1C_1, \quad X_2 = Y_2C_2,$$

$Y_1 = [\Psi_2 \vdots \Omega]$ — матрица размером $N \times 5$, $Y_2 = [\Psi_1 \vdots \Omega]$ — матрица размером $N \times 6$,
$C_1 = [a_{21}, a_{22}, a_{23}, b_1, b_2]^{\mathrm{T}}$, $C_2 = [a_{11}, a_{12}, a_{13}, a_{14}, b_1, b_2]^{\mathrm{T}}$.

При формировании уравнения наблюдения полагалось $a_{11} = 10^3$, $a_{12} = 50$, $a_{13} = 10$, $a_{14} = 3$. $a_{21} = -2\times10^3$, $a_{22} = -2$, $a_{23} = 1$, $K^{\Xi} = \mathrm{diag}\big[\sigma_n^2, n = \overline{1,N}\big]$, $\sigma_n^2 = \sigma^2 = 4 \times 10^{-2}$, $t_{n+1} - t_n = 0.5$, $N = 300$. Рассматривается пример с большой по объему выборкой, обеспечивающей хорошее сглаживание шума наблюдения.

Мерой качества оценивания (в процентах) служит величина $\delta a_{km} = 10^2|a_{km}^* - a_{km}|/|\bar{a}_{km}^*|$, где $\bar{a}_{km}^* = \max\{a_{km}^*, |a_{km}|\}$. Сингулярная помеха задавалась в виде линейной комбинации из двух базисных функций

$$\theta(t, b_1, b_2) = b_1\omega_1(t) + b_2\omega_2(t) = b_1\sin(\alpha_1 t) + b_2\sin(\alpha_2 t),$$

где α_1 и α_2 — произвольные числа.

Параметры помехи выбирались случайным образом (их конкретные значения не принципиальны, поскольку в данном методе достигается полная автокомпенсация помехи при любых значениях параметров).

Все расчеты проводились с точностью до 15-ти разрядов путем усреднения результатов, полученных в ходе тысячи экспериментов. Применительно к развиваемому методу получены следующие оценки для координат векторов $\delta A_1 = \big[\delta a_{1m}, m = \overline{1,4}\big]^{\mathrm{T}}$ и $\delta A_2 = \big[\delta a_{2m}, m = \overline{1,3}\big]^{\mathrm{T}}$:

$$\delta a_{11} = 1.585, \quad \delta a_{12} = 0.621, \quad \delta a_{13} = 0.082, \quad \delta a_{14} = 1.967 \times 10^{-5},$$
$$\delta a_{21} = 1.643 \times 10^{-4}, \quad \delta a_{22} = 1.409 \times 10^{-7}, \quad \delta a_{23} = 2.756 \times 10^{-11}.$$

В свою очередь, для РМНК

$$\delta a_{11} = 43.053, \quad \delta a_{12} = 24.902, \quad \delta a_{13} = 2.686, \quad \delta a_{14} = 0.138,$$
$$\delta a_{21} = 2.968 \times 10^{-3}, \quad \delta a_{22} = 2.906 \times 10^{-6}, \quad \delta a_{23} = 3.846 \times 10^{-10}.$$

Видим, что достигается существенный выигрыш по точности, а оценки для РМНК (в рассматриваемых условиях) оказываются малопригодными.

Сравнительный анализ показывает, что разработанный метод позволяет также существенно снизить вычислительную погрешность оценивания, что обусловлено декомпозицией и снижением размерности вычислительной процедуры. В ходе расчетов с использованием евклидовой нормы определялись числа обусловленности (ν_1 и ν_2) обращаемых матриц $\Psi_1^{\mathrm{T}}\Lambda_1 V_1$ и $\Psi_2^{\mathrm{T}}\Lambda_2 V_2$, а также число ν обусловленности соответствующей объединенной матрицы РМНК. В итоге $\nu_1 = 5.283 \times 10^{15}$, $\nu_2 = 1.276 \times 10^{14}$ и $\nu = 4.537 \times 10^{26}$.

Кроме того, подсчитывалось количество сложений и умножений, необходимых для реализации РМНК и разработанного метода. Относительный вычислительный выигрыш составил 1.35 раза, что также подтверждает эффективность нового метода.

В ходе численного эксперимента рассчитывались корреляционные матрицы K_1^A и K_2^A (по аналогии с (5.1)). В целях сокращения записей приводятся лишь диагональные элементы этих матриц:

$$\big(\sigma_{11}^A\big)^2 = 0.031, \quad \big(\sigma_{12}^A\big)^2 = 3.631 \times 10^{-6}, \quad \big(\sigma_{13}^A\big)^2 = 2.306 \times 10^{-9}, \quad \big(\sigma_{14}^A\big)^2 = 0,$$
$$\big(\sigma_{21}^A\big)^2 = 2.287 \times 10^{-5}, \quad \big(\sigma_{22}^A\big)^2 = 0, \quad \big(\sigma_{23}^A\big)^2 = 0.$$

В тоже время для РМНК

$$\big(\sigma_{11}^A\big)^2 = 0.755, \quad \big(\sigma_{12}^A\big)^2 = 7.002 \times 10^{-4}, \quad \big(\sigma_{13}^A\big)^2 = 1.236 \times 10^{-6}, \quad \big(\sigma_{14}^A\big)^2 = 4.781 \times 10^{-14},$$
$$\big(\sigma_{21}^A\big)^2 = 0.073 \times 10^{-6}, \quad \big(\sigma_{22}^A\big)^2 = 5.096 \times 10^{-10}, \quad \big(\sigma_{23}^A\big)^2 = 0.$$

Поскольку оба сравниваемых метода являются линейными и оптимальными, то дисперсии ошибок оценивания не должны отличаться (если не учитывать погрешностей вычислений), т.е. методы наследуют одну и ту

же потенциальную точность оценивания. Однако, как показывают расчеты, в силу указанных выше причин оценки дисперсий для разработанного метода существенно меньше аналогичных оценок для РМНК. Это также связано с тем, что погрешности вычислений по заданным формулам существенно зависят от размерности обращаемых плохо обусловленных матриц.

Пример 2 (для алгоритма совместного обнаружения и оценивания 3, когда $K = 1$ и $L = 2$). В этом примере с целью большей наглядности рассмотрим малую по объему выборку, но с малым уровнем шума наблюдения. Принимаем $T = 4$, $M_1 = 1$, $\psi_1(t) \equiv 1\ \forall t \in [0,4]$ т.е. сигнал $s_1(t) = s(t) = a_{11} = a$ рассматривается как константа. Сингулярная помеха, для случая $J = 2$, представляется как кусочно-линейная функция $\Theta(t) = b_1\omega_1(t) + b_2\omega_2(t)$ с разрывом первого рода в точке $t = 2$, где $\omega_1(t) \equiv t\ \forall t \in [0,2]$ и $\omega_1(t) \equiv 0\ \forall t \in (2,4]$, $\omega_2(t) \equiv 0\ \forall t \in [0,2)$ и $\omega_2(t) \equiv t\ \forall t \in [2,4]$. Надо дать оценку $l^* \in \{1,2\}$ для параметра $l \in \{1,2\}$, а также построить соответствующие оценки параметров сигнала и помехи.

Поскольку $K = 1$, $M_1 = M = 1$ и $\overline{M}_{1l} = \overline{M}_l = 0$, то матрицы в (2.3) выглядят так:

$$\Psi_1 = \Psi = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},\ \Omega = \begin{bmatrix} t_1 & 0 \\ t_2 & 0 \\ 0 & t_3 \\ 0 & t_4 \end{bmatrix},\ Y_1 = Y = \Omega = \begin{bmatrix} t_1 & 0 \\ t_2 & 0 \\ 0 & t_3 \\ 0 & t_4 \end{bmatrix},\ C_1 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},\ X_1 = X = \begin{bmatrix} b_1 t_1 \\ b_1 t_2 \\ b_2 t_3 \\ b_2 t_4 \end{bmatrix}.$$

Формируя уравнение наблюдения $H = q_1 S_1 + \Theta + \Xi = qS + \Theta + \Xi$, примем, что $q = 1$, $a = 10$, $b_1 = -2$ и $b_2 = 4$, шум Ξ считаем нормальным процессом с нулевым математическим ожиданием, некоррелированными отсчетами и дисперсией $\sigma^2 = 10^{-2}$. Для простоты все величины полагаем безразмерными. Усреднение результатов осуществлялось по 100 экспериментам.

После выполнения всех шагов алгоритма 3 получены следующие результаты: $l^* = 1$ (для всех экспериментов) — оценка для номера наилучшей гипотезы (соответствует $q = 1$, т.е. сигнал обнаружен во всех случаях), $a^* = 10.039$ — результирующая оценка сигнала (относительная погрешность $\delta a = 0.4\%$), $a^* = 15.586$ — аналогичная результирующая оценка сигнала для РМНК (относительная погрешность $\delta a = 56\%$), $\Theta^*_{l^*} = [-2.004, -4.007, 12.085, 16.133]^{\mathrm{T}}$ — результирующая оценка сингулярной помехи (относительная погрешность $\delta\Theta = [0.199\%, 0.174\%, 0.703\%, 0.824\%]^{\mathrm{T}}$), $\Upsilon^* = 1.354$ — относительный показатель вычислительной эффективности в сравнении с РМНК (в разах, с учетом операций сложения и умножения).

Рассмотренные примеры подтверждает эффективность разработанного метода распознавания сигналов в условиях сингулярных помех наблюдения.

ЗАКЛЮЧЕНИЕ

Развитый метод может эффективно сочетаться с алгоритмами ортогональных разложений (см. [30]) и решениями некорректных задач (см. [31], [32]). Возможность декомпозиции и распараллеливания вычислительных процедур позволяет более эффективно решать целый круг прикладных задач, связанных с параллельной обработкой измерений в различных областях. Полученные алгоритмы распознавания сигналов в условиях сингулярных помех несложно реализовать на специализированных ЭВМ, ориентированных на информационно-измерительные комплексы, предназначенные для функционирования в реальном времени.

Получены компактные аналитические выражения, позволяющие заранее, под конкретную прикладную задачу, подобрать необходимые модели сигналов и помех, а также количественные значения их параметров, при которых развитый метод обеспечит достижение своих потенциальных возможностей. Метод относится к классу линейных, поэтому все вычислительные процедуры сводятся к простейшим математическим операциям над векторами и матрицами, а также допускает возможность комбинирования с традиционными статистическими подходами к решению прикладных задач, связанных с оптимальной и квазиоптимальной обработкой наблюдений.

Полученные результаты можно применять и к классу динамических систем с измеряемым выходом, если воспользоваться известным комбинированным методом опорных интегральных кривых и обобщенного инвариантно-несмещенного оценивания (см. [33]–[35]). В этом случае состояние и выход таких систем можно также представлять в виде конечной линейной оболочки заданного функционального базиса, если использовать заранее построенное семейство опорных кривых или поверхностей необходимого объема.

Кроме того, можно обобщить развитый метод на тот случай, когда используемый базис зависит от некоторых неизвестных параметров (например, временных задержек). Для решения задачи разрешения сигналов в этих условиях необходимо задавать сетку узлов в области изменения значений этих параметров и реализовывать рассмотренные выше вычислительные процедуры для каждого многомерного узла в отдельности с последующим выбором оптимального узла.

СПИСОК ЛИТЕРАТУРЫ

1. *Френкс Л.* Теория сигналов. М.: Сов. радио, 1969.
2. *Ширман Я.Д.* Разрешение и сжатие сигналов. М.: Сов. радио, 1974.
3. *Богданович В.А., Вострецов А.Г.* Теория устойчивого обнаружения и оценивания сигналов. М.: Физматлит, 2004.
4. *Сосулин Ю.Г., Костров В.В., Паршин Ю.Н.* Оценочно-корреляционная обработка сигналов и компенсация помех. М.: Радиотехника, 2014.
5. *Сергиенко А.Б.* Цифровая обработка сигналов. СПб.: БХВ-Петербург, 2011.
6. *Мельников В.В.* Безопасность информации в автоматизированных системах. Альтернативный подход // Защита информации. 2005. № 6. С. 40–45.
7. *Шувалов А.В.* Синтез и анализ компенсационного алгоритма подавления структурно детерминированных помех // Радиотехника. 2005. № 7. С. 43–49.
8. *Булычев Ю.Г., Манин А.П.* Математические аспекты определения движения летательных аппаратов. М.: Машиностроение, 2000.
9. *БулычевЮ.Г., Васильев В.В., Джуган Р.В.* и др. Информационно-измерительное обеспечение натурных испытаний сложных технических комплексов. М.: Машиностроение – Полет, 2016.
10. *Репин В.Г., Тартаковский Г.П.* Статистический синтез при априорной неопределенности и адаптация информационных систем. М.: Сов. радио, 1977.
11. *Татузов А.Л.* Нейронные сети в задачах радиолокации. М.: Радиотехника, 2009.
12. *Иванов Н.М.* Адаптивные методы обнаружения и пеленгования сигналов // Радиотехника и электроника. 2016. Т. 61. № 10. С. 979–983.
13. *Бакулин М.Г., Крейнделин В.Б., Григорьев В.А.* и др. Байесовское оценивание с последовательным отказом и учетом априорных знаний // Радиотехника и электроника. 2020. Т. 65. № 3. С. 257–266.
14. *Парфенов В.И., Калининский А.А.* Совместное обнаружение и классификация объектов при комплексировании решений, выносимых сенсорами в беспроводных сенсорных сетях // Радиотехника и электроника. 2020. Т. 65. № 3. С. 257–266.
15. *Абрамова Т.В., Ваганова Е.В., Горбачев С.В.* и др. Нейро-нечеткие методы в интеллектуальных системах обработки и анализа многомерной информации. Томск: Изд-во Томского ун-та, 2014.
16. *Бобырь М.В.* Проектирование нейронных и нечетких моделей в области вычислительной техники и систем управления. М.: Аграмак Медиа, 2018.
17. *Булычев Ю.Г.* Оптимизация кластерно-вариационного метода построения многопозиционной пеленгационной системы для условий априорной неопределенности // Автоматика и телемеханика. 2023. № 4. С. 96–114.
18. *Zekavat S., Buehrer R.* Handbook of Position Location: Theory Practice and Advances. Second ed. Hoboken. New Jersey: Wiley-IEEE Press 2019. https://doi.org/10.1002/9781119434610.
19. *Zhao J., Renzhou G., Xudong D.* A new measurement association mapping strategy for DOA tracking // Digital Signal Processing. 2021. V. 118. P. 103–228. ISSN 1051-2004. https://doi.org/10.1016/j.dsp.2021.103228. (https://www.sciencedirect.com/science/article/pii/S1051200421002670).
20. *L. Peng, W. Wenhui, Q. Junda, Y. Congzhe, S. Zhenqiu.* Robust Generalized Labeled Multi-Bernoulli Filter and Smoother for Multiple Target Tracking using Variational Bayesian // KSII Transactions on Internet and Information Systems. 2022. V. 16. № 3. P. 908–928. https://doi.org/10.3837/tiis.2022.03.009.
21. *Wang X., Wang A., Wang D., Xiong Y., Liang B., Qi Y.* A modified Sage-Husa adaptive Kalman filter for state estimation of electric vehicle servo control system // Energy Rep. 2022. V. 8. № 5. P. 20–27. ISSN 2352-4847. https://doi.org/10.1016/j.egyr.2022.02.105 (https://www.sciencedirect.com/science/article/pii/S2352484722003523).
22. *Леонов В.А., Поплавский Б.К.* Фильтрация ошибок измерений при оценивании линейного преобразования полезного сигнала // Техн. кибернетика. 1992. № 1. С. 163–170.
23. *Жданюк Б.Ф.* Основы статистической обработки траекторных измерений. М.: Сов. радио, 1978.

24. *Булычев Ю.Г., Елисеев А.В.* Вычислительная схема инвариантно-несмещенного оценивания значений линейных операторов заданного класса // Ж. вычисл. матем. и матем. физ. 2008. Т. 48. № 4. С. 580–592.

25. *Булычев Ю.Г.* Применение методов опорных интегральных кривых и обобщенного инвариантно-несмещенного оценивания для исследования многомерной динамической системы // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 7. С. 1151–1169.

26. *Ежова Н.А., Соколинский Л.Б.* Обзор моделей параллельных вычислений // Вестник ЮУрГУ. Серия "Вычислительная математика и информатика". 2019. Т. 8. № 3. С. 58–91.

27. *Иванов А.И., Шпилевая С.Г.* О квантовых параллельных вычислениях // Вестник Балтийского федер-го унив-та им. И. Канта. Серия «Физико-математические и технические науки». 2021. № 2. С. 95–99.

28. *Sutti C.* Lokal and global optimization by parallel algorithms for MIMD systems // Ann. of Operat. Res. 1984. V. 1.

29. *Price W.L.* Global optimization algorithms for a CAD workstation // J. Optimiz. Theory and Applic. 1987. V. 55. № 1.

30. *Лоусон Ч., Хенсон Р.* Численное решение задач метода наименьших квадратов. М.: Наука, 1986.

31. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М.: Наука, 1986.

32. *Бакушинский А.Б., Гончарский А.В.* Некорректные задачи. Численные методы и приложения. М.: Изд-во московского ун-та. 1989.

33. *Булычев Ю.Г.* Метод опорных интегральных кривых решения задачи Коши для обыкновенных дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 1988. Т. 28. № 10. С. 1482–1490.

34. *Булычев Ю.Г.* Методы численно-аналитического интегрирования дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 1991. Т. 31. № 9. С. 1305–1319.

35. *Булычев Ю.Г.* Численно-аналитического интегрирование дифференциальных уравнений с использованием обобщенной интерполяции // Ж. вычисл. матем. и матем. физ. 1994. Т. 34. № 4. С. 520–532.

NUMERICAL-ANALYTICAL METHOD OF DECOMPOSITION-AUTOCOMPENSATION FOR SOLVING THE SIGNAL RECOGNITION PROBLEM BASED ON INACCURATE OBSERVATIONS

Y. G. Bulychev*

JSC All-Russian Research Institute «Gradient», Sokolova Ave., 96, Rostov-on-Don, 344000, Russia
**e-mail: ProfBulychev@yandex.ru*
Received 10 November, 2023
Revised 20 December, 2023
Accepted 06 February, 2024

Abstract. A numerical-analytical method is developed to solve the problem of optimal recognition of a set of possible signals observed as an additive mixture containing not only a fluctuation observation error (with an unknown statistical distribution) but also a singular interference (with parametric uncertainty). The method allows for both the detection of signals present in the mixture and the estimation of their parameters within a specified quality criterion and associated constraints. The proposed method, based on the idea of generalized invariant-unbiased estimation of linear functional values, enables decomposition of the computational procedure and autocompensation of singular interference without resorting to the traditional expansion of the state space. Linear spectral decompositions in specified functional bases are used for the parametric finite-dimensional representation of signals and interference, while knowledge of the correlation matrix of the observation error is sufficient for error description. Random and systematic errors are analyzed, and an illustrative example is provided.

Keywords: observation equation, fluctuation error, singular interference, correlation matrix of measurement errors, Lagrange multiplier method, inaccurate observation, singular interference, optimal estimation, conditions of unbiasedness and invariance, decomposition, autocompensation, computational recognition algorithms.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 713–728

УДК 519.62

СПЕКТРАЛЬНЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ И ФУНКЦИОНАЛЬНЫХ УРАВНЕНИЙ

© 2024 г. В.П. Варин[1,*]

[1] 125047 Москва, Миусская пл., 4, Институт прикладной математики им. М.В. Келдыша РАН, Россия
**e-mail: varin@keldysh.ru*

Поступила в редакцию 16.10.2023 г.
Переработанный вариант 16.10.2023 г.
Принята к публикации 14.01.2024 г.

Операторный подход, развитый ранее для спектрального метода, использующего полиномы Лежандра, здесь обобщается на любые системы базисных функций (необязательно ортогональных), удовлетворяющих всего двум условиям: результат операции умножения на x либо дифференцирования по x выражается в тех же функциях. Все системы классических ортогональных полиномов удовлетворяют этим условиям. В частности, построен спектральный метод, использующий полиномы Чебышёва, который наиболее эффективен для численных расчетов. Этот метод применяется для численного решения линейных функциональных уравнений, которые возникают в задачах обобщенного суммирования рядов, а также в задачах аналитического продолжения дискретных отображений. Показано также, как этими методами решаются нестандартные и нелинейные краевые задачи, для которых обычные алгоритмы не применимы. Библ. 9.

Ключевые слова: спектральные методы, полиномы Чебышёва, краевые задачи, функциональные уравнения, высокоточные вычисления.

DOI: 10.31857/S0044466924050022, **EDN**: YDNGIE

1. ВВЕДЕНИЕ

Спектральные методы решения краевых задач подразумевают разложение решений в ряды по некоторым наборам базисных функций (или пробных, в контексте методов Галеркина), в качестве которых часто используются полиномы, ортогональные на данном интервале с некоторым весом.

В [1] был предложен спектральный метод решения краевых задач для голономных ОДУ на интервале $[0, 1]$, в котором неизвестные функции раскладываются в ряды по смещенным полиномам Лежандра.

Было показано, что любая линейная краевая задача для голономного ОДУ аппроксимируется с помощью всего двух операторов, действующих в конечномерном линейном пространстве коэффициентов Фурье–Лежандра решений этих ОДУ: оператора X умножения на независимую переменную x и оператора D дифференцирования по x. Некоторые дополнительные операторы, введенные в [1], играли лишь вспомогательную роль.

Представление функций в виде разложений по (практически) произвольным наборам базисных функций оказывается вполне аналогично такому представлению в виде разложений по полиномам Лежандра или Чебышёва, если для этих наборов функций существуют аналогичные операторы X и D, действующие в линейном пространстве коэффициентов разложений функций.

Для периодических функций, которые аппроксимируются с помощью тригонометрических полиномов, роль оператора X играет оператор умножения на основную гармонику.

В этой статье рассматриваются аналитические функции на отрезке $[0, 1]$, которые могут иметь особенности на концах интервала. Назовем это множество функций $\mathcal{H}$.

Предположим, что имеется алгоритм, который каждой функции $y(x) \in \mathcal{H}$ ставит в соответствие ее формальное разложение по некоторому набору базисных функций $\{p_n(x) \in \mathcal{H},\ n = 0, 1, \dots\}$. При этом всегда $p_0(x) = 1$ (или $p_0(x) = \mathrm{const}$), так как единица принадлежит $\mathcal{H}$.

Существование такого алгоритма означает, что определено линейное отображение множества $\mathcal{H}$ в пространство коэффициентов разложений этих функций, которое мы назовем $\mathcal{A}$. Это отображение аналогично дискретному преобразованию Фурье периодических функций, а коэффициенты разложения $\{a_n,\ n = 0, 1, \dots\}$ – это аналог обычных коэффициентов Фурье. Например, в [1] рассматривалось преобразование Фурье–Лежандра.

Отличие от обычного преобразования Фурье состоит в самом алгоритме, который теперь не обязан опираться на ортогональность функций в каком-либо пространстве, а также в том, что никаких условий сходимости полученных разложений (пока) не требуется.

Поскольку реально вычисления всегда проводятся с конечными отрезками разложений, то наряду с отображением $\mathcal{H} \to \mathcal{A}$ рассматриваются проекции $\mathcal{H} \to \mathcal{A}_N$, где N – это размерность аппроксимации. То есть функция $y(x)$ аппроксимируется (в каком-то пока не определенном смысле) своим разложением

$$y(x) = \sum_{n=0}^{N-1} a_n\, p_n(x), \tag{1}$$

где равенство понимается в проективном смысле и чисто формально, т.е. никакой близости функции и ее разложения (в среднем, поточечно, равномерно, и т.п.) априори не требуется.

Иными словами, мы разделяем задачи вычисления разложения функции и интерпретации полученного разложения как аппроксимации этой функции. Например (см. разд. 2), ряд может быть асимптотическим и расходящимся, но породившая его аналитическая функция (решение голономного ОДУ) при этом вполне определена.

Множество функций $\mathcal{H}$ замкнуто относительно операций дифференцирования и умножения на независимую переменную x, поэтому необходимо определить, какие линейные отображения эти операции индуцируют в пространствах коэффициентов конечномерных аппроксимаций $\mathcal{A}_N$. Иными словами, для аппроксимации (1) необходимо определить конечномерные отображения

$$X\colon \{a_n\} \to \{b_n\} \quad \text{и} \quad D\colon \{a_n\} \to \{c_n\},$$

где

$$x\,y(x) = \sum_{n=0}^{N-1} b_n\, p_n(x) \quad \text{и} \quad y'(x) = \sum_{n=0}^{N-1} c_n\, p_n(x).$$

Существование таких операторов – это как раз те два требования к набору базисных функций $\{p_n(x)\}$, о которых говорилось в аннотации.

В случае, когда наборы базисных функций – это (классические) ортогональные полиномы, матрица X – это всегда транспонированная матрица Якоби, ассоциированная с данной системой полиномов (см. [1, 3]).

В разд. 2 мы покажем, как эта общая конструкция работает при решении голономных ОДУ для некоторых конкретных наборов базисных функций. Это делается так же, как и в [1], с использованием полиномов Лежандра.

В случае если задача не сводится к решению голономного ОДУ, предложенного формализма недостаточно для решения задачи. Это видно уже для линейного неголономного ОДУ, так как операция умножения на известную функцию $v(x)$ соответствует оператору $V = v(X)$, где $v(X)$ – это функция от матрицы (предполагая, что она существует).

Для рациональных функций $r(x)$ функция от матрицы $r(X)$ получается формальной подстановкой матрицы X вместо переменной x (см. [4]). Именно поэтому голономные ОДУ решаются относительно просто. Но для линейных неголономных ОДУ, для нелинейных ОДУ и для функциональных уравнений функции от матриц придется вычислять другим способом.

В случае если набор базисных функций – это ортогональные полиномы, функция от матрицы $v(X)$ однозначно определяется значениями функции $v(x)$ на спектре матрицы X (т.е. на спектре матрицы Якоби) с помощью интерполяционной формулы Лагранжа–Сильвестра (см. [4]), так как спектр матрицы X размерности N – это корни полинома $p_N(x)$.

Поэтому корни ортогональных полиномов – это, как правило, самые удобные узлы коллокации. Однако в общем случае это не так. Поэтому мы накладываем единственное очевидное ограничение на узлы коллокации $x_n \in [0, 1]$, $n = 1, 2, \ldots, N$ – их несовпадение, т.е. $x_n \neq x_m$ при $n \neq m$.

Таким образом, имеем еще одно конечномерное представление функций $y(x) \in \mathcal{H}$,

$$y(x) = \{y_1, y_2, \ldots, y_N\}, \quad y_n = y(x_n) = \sum_{k=0}^{N-1} a_k\, p_k(x_n). \tag{2}$$

Здесь, как и ранее в (1), мы используем способ обозначения, который в программировании называется «overloading», т.е. когда смысл символа (в данном случае $y(x)$) определяется в зависимости от контекста. Так же как оператор и его матрица обычно обозначаются одним символом.

Интерполяционная формула (1), примененная в (2), – это не что иное, как аналог обратного (дискретного) преобразования Фурье (которое мы обозначим через F^{-1}), так как это линейный оператор, действующий в

пространстве «коэффициентов Фурье» $\mathcal{A}_N$ и восстанавливающий функцию $y(x)$, как таблицу ее (приближенных) значений в выбранных узлах. Таким образом, матрица этого преобразования всегда известна в явном виде для любого набора узлов x_n,

$$F^{-1} = [p_{k-1}(x_j)]_{1\leqslant j,k\leqslant N}, \quad F^{-1}\colon \mathcal{A}_N \to \mathcal{H}_N, \tag{3}$$

где $\mathcal{H}_N$ обозначает пространство конечномерных аппроксимаций функций $y(x) \in \mathcal{H}$ в виде таблиц их (приближенных) значений в узлах x_n.

Набор базисных функций $\{p_n(x) \in \mathcal{H},\, n \in \mathbb{N}_0\}$ теперь должен удовлетворять еще одному очевидному требованию: матрица F^{-1} должна быть обратимой для любой размерности аппроксимации N. Тогда мы всегда имеем аналог обычного (дискретного) преобразования Фурье, $F = (F^{-1})^{-1}$, которое преобразует функцию $y(x)$, представленную таблицей ее значений в узлах x_n, в таблицу ее «коэффициентов Фурье» $\{a_0, \ldots, a_{N-1}\}$.

Представление функций $y(x) \in \mathcal{H}$ двумя способами, т.е. в виде таблиц их значений в узлах коллокации и в виде таблиц их *коэффициентов Фурье* (кавычки далее опускаем), с возможностью менять эти представления по мере надобности, позволяет использовать преимущества обоих представлений.

Голономные дифференциальные операторы, действующие в пространстве коэффициентов Фурье, допускают весьма простой учет произвольных краевых условий, осуществляемый точно так же, как это делалось в [1] для аппроксимаций полиномами Лежандра. В то же время коллокационный подход позволяет обобщить эти результаты на произвольные линейные дифференциальные операторы, а также на линейные функциональные уравнения.

Например, функция от матрицы $v(X)$, нужная для оператора умножения на функцию $v(x)$, вычисляется (аппроксимируется) как

$$v(X) = F\,.\,\mathrm{Diag}\,[v(x_1), \ldots, v(x_N)]\,.\,F^{-1}, \tag{4}$$

где точка означает умножение матриц, а Diag[] обозначает диагональную матрицу.

Преимущество ортогональных полиномов над любыми другими наборами базисных функций состоит в том, что представление функции от матрицы (4) является точным, если в качестве узлов используется спектр матрицы X, т.е. корни N-го ортогонального полинома. А преимущество полиномов Чебышёва над любыми другими системами ортогональных полиномов состоит в том, что спектр матрицы X известен в явном виде, а также в том, что обе матрицы, F и F^{-1}, даются явными формулами (см. разд. 3).

Это, как мы полагаем, закрывает дискуссию о том, какие полиномы, Чебышёва или Лежандра, лучше в численных расчетах (см. [5]).

В разд. 2 мы покажем, как предложенный формализм работает при вычислении формальных степенных разложений решений голономных ОДУ. Эти разложения всегда асимптотические в силу самого способа их вычисления (так как метод неопределенных коэффициентов по сути асимптотический).

В то же время функция (решение ОДУ), породившая степенное разложение, является аналитической и может быть восстановлена, например, с помощью преобразования степенного ряда в факториальный (см. [6]). Оказывается (см. разд. 2), факториальное разложение решения можно вычислять непосредственно из ОДУ при подходящем наборе базисных функций.

В разд. 3 даются явные формулы для операторов X, D, F и F^{-1}, а также для некоторых других для аппроксимаций полиномами Чебышёва. Эти результаты применяются для численного решения некоторых линейных функциональных уравнений в разд. 4,5.

В разд. 6 показано, как результаты разд. 3 применяются при решении нестандартных и/или нелинейных краевых задач, для которых обычные численные методы неприменимы.

2. СТЕПЕННЫЕ И ФАКТОРИАЛЬНЫЕ РАЗЛОЖЕНИЯ

Рассмотрим простейший набор базисных функций

$$p_n(x) = x^n, \quad n \in \mathbb{N}_0, \tag{5}$$

и покажем, что наш минималистский подход к вычислению разложений является содержательным.

В данном случае пространство коэффициентов Фурье $\mathcal{A}$ – это просто коэффициенты формальных степенных разложений в нуле тех функций из $\mathcal{H}$, которые имеют эти разложения.

Операторы X и D здесь очевидно существуют, а их матрицы устроены особенно просто. Приведем эти матрицы для размерности $N = 5$:

$$X = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Структура этих матриц для любой размерности N очевидна.

Фиксируем размерность аппроксимации N и введем вектор $e \in \mathcal{A}_N$, соответствующий функции $e(x) = 1$, т.е. $e = \langle 1, 0, \ldots, 0 \rangle^t$.

Предложение 1. *Для любой регулярной в нуле функции $f(x)$ коэффициенты ее степенного разложения в нуле до номера $N-1$ включительно даются вектором $f(X).e$.*

Доказательство, на самом деле, очевидно. Вектор $X.e$ соответствует функции x, т.е. это коэффициенты Фурье функции x. Это же справедливо для любой степени X^n, т.е. вектор $X^n.e$ соответствует функции x^n. Но регулярная функция $f()$ от матрицы X дается отрезком ее тейлоровского разложения в нуле до номера $N-1$ включительно, так как матрица X нильпотентна. Поэтому $f(X).e$ – это коэффициенты Фурье функции $f(x)$, т.е. коэффициенты ее степенного разложения. Что требовалось доказать.

Аналогичное утверждение справедливо для решения любого ОДУ, которое сводится к голономному ОДУ, регулярному в нуле. Например, рассмотрим линейное ОДУ:

$$(1+x)\,y''(x) + x^2\,y'(x) - y(x) = 1 + x, \tag{6}$$

которое не интегрируемо (по крайней мере в CAS Maple), и найдем степенное разложение в нуле его решения $y(x)$, такого, что $y(0) = a$ и $y'(0) = b$.

Для этого составим матрицу $A = (E + X).D^2 + X^2.D - E$, где E – это единичная матрица. Затем вычислим вектор правой части, $r = (E + X).e$. Затем заменим две последние строки матрицы A на строки

$$A[N, n] = p(n-1, 0), \quad A[N-1, n] = p'(n-1, 0), \quad n = 1, 2, \ldots, N,$$

для учета начальных значений функции $y(x)$. Затем положим $r[N] = a$ и $r[N-1] = b$ (в том же порядке, как это делалось для матрицы A). Тогда вектор $y = A^{-1}.r$ дает коэффициенты степенного разложения (до $(N-1)$-го включительно) решения $y(x)$ ОДУ (6) с этими начальными данными, т.е.

$$y(x) = a + b\,x + \frac{1}{2}\,(1+a)\,x^2 - \frac{1}{6}\,(a-b)\,x^3 + \frac{1}{24}\,(1 + 3\,a - 4\,b)\,x^4 + \ldots.$$

Гибкость этого подхода к вычислению (регулярного) степенного разложения решения ОДУ состоит в том, что мы можем наложить на решение произвольные (линейные) начальные условия. Например, потребовать, чтобы третья производная функции в нуле была равна нулю. Тогда окажется, что необходимо $a = b$.

Если же потребовать выполнения несовместных начальных условий, то матрица A просто окажется вырожденной.

Наложение краевых условий в точке $x = 1$, как мы это делали для полиномов Лежандра в [1], возможно, но довольно бессмысленно, так как аппроксимация решения краевой задачи отрезками степенных разложений крайне неэффективна даже в случае их сходимости. В разд. 3 мы используем для этого полиномы Чебышёва.

По той же причине коллокационный подход здесь не имеет смысла, хотя матрица (3) здесь – это матрица Вандермонда и всегда обратима. То есть базисные функции (5) нужны только для вычисления формальных степенных разложений в нуле решений голономных ОДУ.

Также можно вычислять степенные разложения решений ОДУ, сингулярных в нуле. Такие разложения могут как сходиться, так и расходиться. Общим для них является то, что наложение n начальных условий на решение ОДУ n-го порядка (вообще говоря) невозможно, так как не все решения в нуле раскладываются в степенные ряды. Приведем два таких примера.

Рассмотрим сингулярную задачу Коши $y(0) = 0$ для уравнения

$$x^2\,y'(x) + y(x) = x. \tag{7}$$

Его общим решением является сингулярная в нуле функция, имеющая при $x > 0$ всюду расходящееся степенное разложение,

$$y(x) = \exp\left(\frac{1}{x}\right)\left(C + \mathrm{Ei}\left(1, \frac{1}{x}\right)\right), \quad y(x) = \sum_{n=1}^{\infty} (-1)^{n-1}\,(n-1)!\,x^n, \tag{8}$$

где $\mathrm{Ei}()$ – это интегральная экспонента.

Единственное ограниченное в нуле ($x > 0$) решение ОДУ (7) имеет константу интегрирования $C = 0$ и степенное разложение (8), поэтому конечномерная аппроксимация строится просто: берутся матрица $A = X^2.D + E$ и правая часть $r = X.e$. Никакой корректировки матрицы A и вектора r делать не нужно, так как это лишь приведет к тому, что полученная матрица будет сингулярна.

В результате для размерности аппроксимации N получим разложение (8) до степени $N-1$ включительно.

Рассмотрим уравнение

$$x^2\,y''(x) + (1 + 2\,x^2)\,y'(x) + 2\,y(x) = x, \tag{9}$$

которое интегрируемо в Maple. Однако общая формула первого интеграла весьма громоздка и не позволяет определить (встроенными средствами Maple) асимптотики решений в нуле. Также имеющиеся в Maple процедуры не позволяют найти степенное разложение решения, которое легко находится нашим способом.

Для этого составим матрицу $A = X^2.D^2 + (E + 2\,X^2).D + 2\,E$ и вектор правой части, $r = X.e$. Затем заменим последюю строку матрицы A на строку $\langle 1, 0, \ldots, 0\rangle$ и положим $r[N] = a$ для учета начального значения $y(0) = a$. Тогда вектор $y = A^{-1}.r$ дает коэффициенты степенного разложения (до $(N-1)$-го включительно) решения ОДУ (9), т.е.

$$y(x) = a - 2\,a\,x + \frac{1}{2}\,(1 + 4\,a)\,x^2 - \frac{2}{3}\,(1 + 2\,a)\,x^3 + \frac{1}{6}\,(5 + 4\,a)\,x^4 - \frac{1}{15}\,(23 + 4\,a)\,x^5 + \ldots.$$

Сравнение этого решения с полученной общей квадратурой (которую мы опускаем), а также с решением однородного уравнения (9) дает асимптотику в нуле той части квадратуры, для которой имеющиеся средства Maple не применимы. Нужно просто подставить $a = 0$ в это решение.

Рассмотрим теперь набор базисных функций,

$$\left\{q_n(x) = \frac{1}{(1 + 1/x)_n}\right\} = \left\{1, \frac{x}{x+1}, \frac{x^2}{(x+1)\,(2\,x+1)}, \ldots\right\}, \quad n \in \mathbb{N}_0, \tag{10}$$

где $()_n$ – это символ Почхаммера. Ряд Фурье по этим функциям – это классическое (но малоизвестное) факториальное разложение.

Эти ряды асимптотически эквивалентны степенным в нуле, но, как правило, сходятся на всем интервале $x \in (0, 1]$ (см. [6] и ссылки там).

Существует простое преобразование ряда Фурье по функциям (5) в ряд Фурье по функциям (10) и наоборот. Это нижнетреугольные матрицы

$$P = \left[(-1)^{n+m}\,S^{(1)}_{n-1,m-1}\right]_{1 \leqslant m \leqslant n \leqslant N}, \quad Q = \left[(-1)^{n+m}\,S^{(2)}_{n-1,m-1}\right]_{1 \leqslant m \leqslant n \leqslant N},$$

где P преобразует степенной ряд в факториальный, а $Q = P^{-1}$ – наоборот, и $S^{(1)}$ и $S^{(2)}$ – это числа Стирлинга первого и второго рода.

Можно легко проверить, что строить график функции (8) по ее степенному разложению совершенно бесполезно, в то время как применение оператора P к вектору коэффициентов этого разложения дает вектор коэффициентов разложения по функциям (10). Это разложение приближает функцию (8) на всем интервале $[0, 1]$ достаточно хорошо уже при небольших N.

Правда, диагональные Паде–аппроксимации в данном случае лучше. Однако разложения по функциям (10) дают рациональные приближения с заранее известными полюсами.

Обозначим через $\tilde{X}$ и $\tilde{D}$ матрицы соответствующих преобразований для функций (10), в то время как X и D – это матрицы для функций (5), как и ранее в этом разделе. Тогда можно проверить, что

$$\tilde{X} = P.X.Q, \quad \tilde{D} \approx P.D.Q,$$

т.е. матрица $\tilde{X}$ вычисляется точно, а матрица $\tilde{D}$ – с точностью до последней строки. Существуют и явные формулы для этих матриц, которые мы опускаем.

Таким образом, факториальные разложения решений голономных ОДУ можно получать точно так же, как мы получали степенные в этом разделе.

В заключение этого раздела отметим, что наш подход работает и для других типов разложений, например, по полиномам Бернулли. Хотя здесь пока даже неясно, в каком смысле понимать близость функции и ее разложения.

3. АППРОКСИМАЦИЯ ПОЛИНОМАМИ ЧЕБЫШЁВА

Здесь мы даем сводку формул, необходимых (и достаточных) для численного решения задач из обозначенных классов. Часть этих формул в той или иной форме уже встречалась в литературе, однако в различных других контекстах.

Под численным решением задачи мы понимаем предъявление алгоритма, позволяющего вычислить решение, в принципе, с произвольной заданной точностью. А точность полученного решения оценивается с помощью анализа коэффициентов Фурье решения.

Само же решение задачи понимается как таблица значений функции $y(x)$ в выбранных узлах либо таблица коэффициентов Фурье функции $y(x)$, по которым сама функция может быть восстановлена (интерполирована) с заданной точностью в любой точке интервала $[0,1]$ (или $(0,1)$).

Классические полиномы Чебышёва, обозначаемые $T(n,x)$, относятся к классу ортогональных полиномов Якоби и имеют весовую функцию $w(x)=(1-x)^{\alpha}(1+x)^{\beta}$, $\alpha=\beta=-1/2$. Эти полиномы наиболее популярны, и мы будем использовать именно их. Полиномы Якоби с показателями $\alpha=\beta=1/2$ обозначаются — $U(n,x)$ и также используются в численном анализе. Значительно реже, но также используются полиномы Якоби с показателями $\alpha=1/2$, $\beta=-1/2$ и $\alpha=-1/2$, $\beta=1/2$.

Все эти четыре типа полиномов Якоби называют полиномами Чебышёва, и все они имеют тригонометрическое представление и все преимущества, перечисленные в разд. 1. Так что явные формулы, приведенные в этом разделе для полиномов $T(n,x)$, существуют и для всех остальных типов полиномов Чебышёва.

Однако численные эксперименты не выявили существенных преимуществ полиномов Чебышёва над полиномами Лежандра (т.е. $\alpha=\beta=0$) при решении ряда модельных задач. Поэтому мы ограничимся здесь одним набором формул для полиномов $T(n,x)$. Аналогичные формулы имеются для полиномов Лежандра (см. [1]), но получаются со значительно большими затратами.

Далее в этом разделе (и статье) мы используем систему базисных функций, состоящую из смещенных полиномов Чебышёва, т.е.

$$p_0(x)=\frac{1}{2},\quad p_n(x)=T(n,2\,x-1),\quad n\in\mathbb{N}. \tag{11}$$

Коэффициент $1/2$ здесь наследуется от обычных рядов Фурье и их тесной связи с рядами Чебышёва. Избавиться от него никак невозможно, так как если взять $p_0(x)=1$, то коэффициент $1/2$ появится далее в других местах.

Это свойство базисных функций (11) приводит к тому, что единичный вектор в пространстве коэффициентов Фурье $\mathcal{A}$, соответствующий функции $e(x)=1$, имеет вид

$$e=\langle 2,0,\ldots,0\rangle^{t}.$$

Матрицы операторов X, D, а также $X2\approx X.X$ для системы базисных функций (11) устроены очень просто. Матрица $X2$ – это несколько улучшенный вариант матрицы X^2. Эти матрицы отличаются только в одном элементе, $16\,(X2-X.X)[N,N]=1$. Однако, как и для полиномов Лежандра в [1], оператор $X2$ все же лучше в численных расчетах, чем $X.X$, но не настолько, чтобы вводить отдельные операторы $X3, X4$ и т.д.

Структура этих матриц совершенно очевидна для любых N и не требует формализации. Для $N=6$ они имеют вид

$$X=\frac{1}{4}\begin{bmatrix}2&2&0&0&0&0\\1&2&1&0&0&0\\0&1&2&1&0&0\\0&0&1&2&1&0\\0&0&0&1&2&1\\0&0&0&0&1&2\end{bmatrix},\ D=\begin{bmatrix}0&4&0&12&0&20\\0&0&8&0&16&0\\0&0&0&12&0&20\\0&0&0&0&16&0\\0&0&0&0&0&20\\0&0&0&0&0&0\end{bmatrix},\ X2=\frac{1}{16}\begin{bmatrix}6&8&2&0&0&0\\4&7&4&1&0&0\\1&4&6&4&1&0\\0&1&4&6&4&1\\0&0&1&4&6&4\\0&0&0&1&4&6\end{bmatrix}.$$

Матрица X – это транспонированная матрица Якоби, ассоциированная с системой полиномов (11), а структура матрицы D следует из свойств полиномов Чебышёва (см. [7]).

Осталось привести вектор-строки для учета краевых условий, как мы это делали в [1]. Эти векторы размерности N имеют вид

$$i_m=\left\langle p_{n-1}^{(m)}(0)\right\rangle,\quad b_m=\left\langle p_{n-1}^{(m)}(1)\right\rangle,\quad n=1,\ldots,N,\ m\in\mathbb{N}_0,$$

где m обозначает номер производной полинома $p_n(x)$ по x. Векторы i_m отвечают за граничные значения функции при $x = 0$, а b_m – за граничные значения при $x = 1$. Эти векторы устроены весьма просто:

$$\begin{aligned} i_0 &= \left\langle \tfrac{1}{2}, -1, 1, -1, \dots \right\rangle, & b_0 &= \left\langle \tfrac{1}{2}, 1, 1, 1, \dots \right\rangle, \\ i_1 &= \left\langle 2\,(-1)^n\,(n-1)^2 \right\rangle, & b_1 &= \left\langle 2\,(n-1)^2 \right\rangle, \\ i_2 &= \left\langle \tfrac{4}{3}\,(-1)^{n-1}\,n\,(n-1)^2\,(n-2) \right\rangle, & b_2 &= \left\langle \tfrac{4}{3}\,n\,(n-1)^2\,(n-2) \right\rangle. \\ &\dots \end{aligned}$$

Как и в [1], последние строки матриц формальных дифференциальных или разностных операторов заполняются нужными строками i_m или b_m, а также их линейными комбинациями для учета краевых условий. В том же порядке последние элементы вектора правой части краевой задачи заменяются на нужные краевые условия.

Гибкость этого подхода к учету краевых условий задачи состоит в том, что он полностью формализован и универсален, т.е. любые линейные краевые условия учитываются одинаково. А также можно учитывать произвольные линейные условия, наложенные на решение внутри интервала (этого мы не нашли в литературе). Например, для того чтобы решение принимало заданное значение $y(x) = y_0$ в какой-либо точке $x_0 \in [0, 1]$, нужно заменить одну из последних строк матрицы формального дифференциального оператора, составленного из матриц X, D и единичной матрицы E, на строку

$$\langle p_{n-1}(x_0) \rangle, \quad n = 1, \dots, N,$$

и заменить соответствующий элемент вектора правой части задачи на y_0.

Решение голономных ОДУ этим методом полностью идентично тому, что мы делали в [1] с использованием полиномов Лежандра. Результаты по точности аппроксимации также вполне аналогичны. В частности, в рациональной арифметике метод дает точные коэффициенты разложения, где это возможно. Рассмотрим краевую задачу

$$x\,y'(x) - y(x) = x, \quad y(1) = 0,$$

решением которой является функция $y(x) = x \ln x$.

Действуя описанным способом, для каждого N получим все коэффициенты разложения решения по смещенным полиномам Чебышёва, кроме последнего, $y[N]$, что связано с методом вычисления, и кроме двух первых, поскольку они иррациональны. Таким образом, получаем

$$x\,\ln x = -\frac{1}{4} + \left(\frac{3}{2} - 2\ln 2\right) x + \sum_{n=2}^{\infty} \frac{(-1)^n\,p_n(x)}{n\,(n^2-1)}, \quad x \in [+0, 1].$$

Перейдем к изложению коллокационного подхода, который необходим для решения более общих задач, чем решение голономных ОДУ.

В качестве узлов коллокации, очевидно, следует использовать корни смещенных полиномов Чебышёва. То есть для размерности аппроксимации N,

$$x_n = \frac{1}{2} + \frac{1}{2}\,\cos\left(\frac{(2\,n-1)\pi}{2\,N}\right), \quad n = 1, 2, \dots, N. \tag{12}$$

Тот факт, что ноль и единица не входят в число узлов, является, на самом деле, преимуществом, так как это позволяет решать задачи с особенностями по краям интервала.

Матрицы преобразований Фурье, т.е. оператора (3) и его обращения, здесь известны в явном виде:

$$F^{-1} = \left[\varkappa(n)\cos\left(\frac{(n-1)\,(2\,m-1)\pi}{2\,N}\right)\right],\ F = \frac{2}{N}\,\left[\cos\left(\frac{(m-1)\,(2\,n-1)\pi}{2\,N}\right)\right],$$

где m и n нумеруют строки и столбцы соответственно, и

$$\varkappa(n) = \begin{cases} \frac{1}{2}, & n = 1, \\ 1, & n > 1. \end{cases}$$

Заметим, что F и F^{-1} – это практически одна и та же матрица с точки зрения вычислительных затрат. И они даны явными формулами. В то время как для полиномов Лежандра необходимо сперва вычислить корни N-го полинома численно, затем вычислить матрицу (3), а затем ее обратить.

Уже приведенных формул вполне достаточно для того, чтобы решать (теперь только в плавающей арифметике) произвольные линейные ОДУ (имеющие решения в $\mathcal{H}$) точно так же, как мы решали голономные, так как функции от матрицы X теперь вычисляются точно по формуле (4).

Альтернативный способ вычисления функций от матрицы X опирается на общую формулу Лагранжа–Сильвестра (см. [4]). Для функции $v(x)$, определенной в узлах (12), функция от матрицы X дается формулой

$$v(X) = \sum_{n=1}^{N} v(x_n)\, l_{N,n}(X), \quad l_{N,n}(x) = \prod_{k\neq n}^{N} \frac{x - x_k}{x_n - x_k}, \tag{13}$$

где $\{x_n,\ n = 1, \ldots, N\}$ даны в (12), а $l_{N,n}(x)$ – это фундаментальные полиномы лагранжевой интерполяции.

Предложение 2. *Для любой функции $v(x) \in \mathcal{H}$ матрицы $v(X)$, посчитанные по формулам* (4) *и* (13)*, совпадают.*

Доказательство. Любая функция $v(x) \in \mathcal{H}$ определена на спектре матрицы X (12). Матрица $v(X)\colon \mathcal{A}_N \to \mathcal{A}_N$ в (4) преобразуется в матрицу $F^{-1}.v(X).F = \mathrm{Diag}\,[v(x_1), \ldots, v(x_N)] : \mathcal{H}_N \to \mathcal{H}_N$, т.е. матрица $v(X)$ всегда диагонализируема. Оператор умножения на функцию $v(x)$ в пространстве $\mathcal{H}_N$ – это поточечное умножение функций. Но в пространстве $\mathcal{A}_N$ ему соответствует оператор $v(X)$, т.е. каноническая функция от матрицы (13), полученная по формуле Лагранжа–Сильвестра (см. [4]). Что требовалось доказать.

Ну и, разумеется, еще один способ вычислить функцию от матрицы – это просуммировать матричный ряд Тейлора этой функции.

Перейдем к изложению метода решения линейных функциональных уравнений, которые возникают в задачах обобщенного суммирования рядов и в задачах аналитического продолжения дискретных отображений. Эти уравнения имеют вид

$$y(f(x)) \pm y(x) = g(x), \tag{14}$$

где $x \in [0, 1]$, $f\colon [0, 1] \to [0, 1]$, и $g(x) \in \mathcal{H}$. Возможны также некоторые модификации уравнения (14), которые существенно не влияют на способ его решения.

Здесь необходимо применять как спектральное, т.е. $y(x) \in \mathcal{A}_N$, так и коллокационное, т.е. $y(x) \in \mathcal{H}_N$, представления функций.

Дело в том, что при знаке минус в (14) функция $y(x)$ определена лишь с точностью до константы. Поэтому конечномерная аппроксимация уравнения (14) даст вырожденную матрицу, так же как и при решении задачи Коши для линейного ОДУ. Поэтому необходимо учитывать краевые условия решения, а это возможно сделать эффективно только при спектральном представлении функций.

С другой стороны, необходимо вычислить значения функции $f(x)$ в узлах (12) для определения линейного оператора $L\colon \mathcal{A}_N \to \mathcal{A}_N$, соответствующего функции $y(f(x))$. Этот оператор имеет вид

$$L = F\,.\,[p_{k-1}(f(x_j))]_{1\leqslant j,k\leqslant N}\,, \tag{15}$$

т.е. представляется в виде суперпозиции оператора, аналогичного (3), и преобразования Фурье.

Далее очевидным образом составляется матрица конечномерной аппроксимации левой части (14) и вектор правой части

$$r = g(X).e = F.\,\langle g(x_1), \ldots, g(x_n)\rangle^t\,.$$

Затем учитываются краевые условия (или условия внутри интервала), как это делалось ранее.

Заметим, что решения $y(x)$ уравнений (14) обычно принадлежат $\mathcal{H}$ и даже могут иметь степенные разложения в нуле. Однако эти разложения, как правило, всюду расходятся и непригодны для аппроксимации функций. Так что наш метод (или его аналоги) численного решения уравнения (14), по-видимому, является безальтернативным. В следующих разделах мы приводим ряд примеров решения этих уравнений.

В заключение этого раздела заметим, что приведенный формализм имеет более широкие приложения, чем те, которые мы обозначили. В частности, мы не обсуждали функции от матрицы D. Например, экспонента от этой матрицы, которая равна

$$\exp(D) = \sum_{n=0}^{N-1} \frac{D^n}{n!}$$

в силу нильпотентности оператора D, дает оператор сдвига аргумента на единицу, т.е. дает коэффициенты Фурье функции $y(x + 1)$. Таким образом, уравнение

$$(\exp(D) - E).y = n\, X^{n-1}.e$$

имеет в качестве решения полином Бернулли $B(n, x)$ при $n < N$, при условии, что учтено начальное значение $y(0) = B_n$, как это делалось при решении голономных ОДУ.

4. АНАЛИТИЧЕСКОЕ ПРОДОЛЖЕНИЕ ДИСКРЕТНЫХ ОТОБРАЖЕНИЙ

Здесь мы рассмотрим один пример численного решения функциональных уравнений, возникающих в задачах аналитического продолжения дискретных отображений (см. [8]). Как мы полагаем, этот пример вполне типичен.

В [8] было показано, что существует аналитический первый интеграл

$$H(x,y) = -x + \log(y - y^2) + \frac{1}{y - y^2} - V(y), \tag{16}$$

дающий решение задачи об аналитическом продолжении логистического отображения

$$y_{n+1} = y_n - y_n^2, \quad y_0 \in \mathbb{C}, \quad n \in \mathbb{N}_0, \tag{17}$$

где непрерывная переменная x соответствует дискретному «времени» n, а функция $V(y)$ является голоморфной в области Фату $\mathcal{F}_0$, лежащей внутри множества Жюлиа отображения (17), и удовлетворяет там уравнению

$$V(y - y^2) - V(y) = \log(y^2 - y + 1) + \frac{y\,(1-y)}{y^2 - y + 1}, \quad V(0) = 0. \tag{18}$$

Уравнение (17) приводилось в [8] в качестве примера интегрируемой динамической системы, в которой реализуется динамический хаос.

Функция $V(y)$ является четной относительно точки $y = 1/2$, что видно после замены $y \to 1 - y$ в уравнении (18), но мы не будем этим пользоваться.

В [8] был предложен алгоритм, позволяющий вычислять функцию $V(y)$ в любой точке $y \in \mathcal{F}_0$ с заданной точностью. Однако этот алгоритм опирался на вычисления с использованием всюду расходящегося ряда функции $V(y)$ в нуле, что вполне аналогично применению формулы Эйлера–Маклорена. Так что альтернативный метод решения задачи здесь служит двоякой цели: демонстрирует эффективность предложенных методов и взаимно их проверяет.

Решение функционального уравнения (18) на интервале $y \in [0,1]$ получается описанным в предыдущем разделе методом. Мы выбрали для этого QD-арифметику (т.е. примерно 64 десятичных разряда), чтобы погрешности вычислений не влияли на оценку точности аппроксимации.

В [8] мы привели значение функции $V(y)$ в точке ее глобального минимума при $y = 1/2$,

$$V(0.5) = -0.154288147256044669278098649697 4,$$

с большой долей уверенности, что все десятичные знаки здесь верны. Теперь мы можем подтвердить это предположение.

На фиг. 1а показан график функции $-V(y)$, а на фиг. 1б приведены абсолютные величины четных коэффициентов Фурье полученного разложения

$$V(y) = \sum_{n=0}^{N-1} s_n\, p_n(y)$$

в логарифмической шкале для $N = 80$. Все нечетные коэффициенты $s_{2\,n-1}$ близки к машинному нулю в силу четности функции $V(y)$ относительно точки $y = 1/2$.

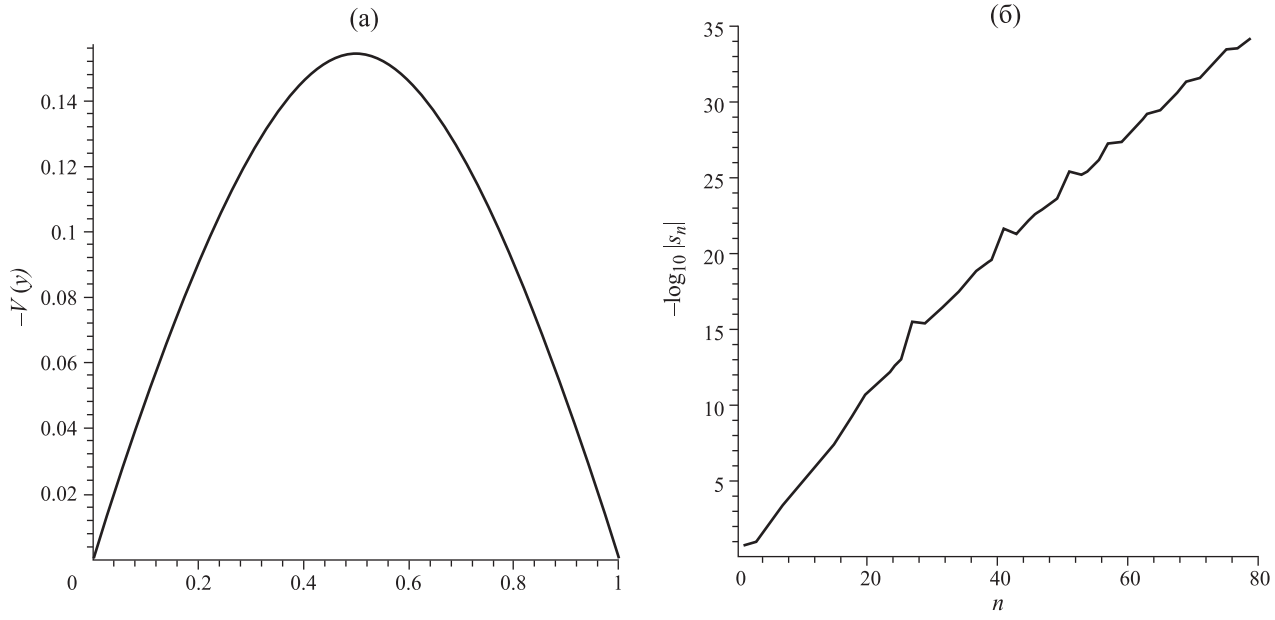
Фиг. 1б содержит информацию о достигнутой равномерной погрешности полученного решения ($\approx 10^{-35}$), а также указывает на то, что функция $V(y)$ не может быть голоморфной в полной окрестности отрезка $[0,1]$. Иначе этот график приближался бы к некоторой прямой (экспоненциальное убывание коэффициентов Фурье). В то время как на рисунке видно, что это не так. Таким образом, (некоторые) аналитические свойства решений, полученных нашим методом, выводятся из анализа убывания их коэффициентов Фурье.

5. ФУНКЦИОНАЛЬНОЕ СУММИРОВАНИЕ РЯДОВ

Перейдем к задачам, которые возникают при функциональном суммировании рядов (см. [9]). Напомним некоторые формулы, которые нам понадобятся.

При суммировании числового ряда его частичные суммы

$$s(n) = \sum_{k=1}^{n} a(k) \tag{19}$$



Фиг. 1. Функция $V(y)$ и ее коэффициенты Фурье–Чебышёва.

удовлетворяют разностному уравнению

$$s(n) - s(n-1) = a(n), \tag{20}$$

где $a(x)$ (по предположению) – это некоторая аналитическая функция при $x \in \mathbb{R}$, $x > \text{const}$. При этом ряд может быть и знакопеременным.

Сделаем замену переменных в уравнении (20), $n = 1/x$, $S(x) = s(1/x)$, $A(x) = a(1/x)$. Тогда получим функциональное уравнение

$$S(x) - S\left(\frac{x}{1-x}\right) = A(x). \tag{21}$$

Наконец, сделаем замену $x \to x/(1+x)$ в уравнении (21) и получим функциональное уравнение на интервале [0, 1],

$$S\left(\frac{x}{1+x}\right) - S(x) = A\left(\frac{x}{1+x}\right). \tag{22}$$

При суммировании знакопеременных рядов вместо уравнения (22) получим (см. [9]) уравнение

$$U\left(\frac{x}{1+x}\right) + U(x) = A\left(\frac{x}{1+x}\right). \tag{23}$$

Пусть исходный ряд (19) сходится. Тогда его сумма

$$C = \lim_{n\to\infty} s(n)$$

однозначно восстанавливается условием согласования двух решений разностного уравнения (20), т.е. дискретного, $s(n)$, и непрерывного, $s(1/x) = S(x)$:

$$C = s(n) - S(1/n), \quad n \in \mathbb{N}, \tag{24}$$

где решение $S(x)$ уравнения (22) получается описанным в разд. 3 способом с учетом начального значения $S(0) = 0$.

Рассмотрим Базельскую проблему в качестве примера, т.е. $A(x) = x^2$ и $C = \pi^2/6$. Действуя описанным способом (и в QD-арифметике), для размерностей аппроксимации $N = 10,\ 20,\ 40$ получим, соответственно,

$$\left|C_{10} - \frac{\pi^2}{6}\right| \approx 4.5 \times 10^{-8}, \quad \left|C_{20} - \frac{\pi^2}{6}\right| \approx 3.4 \times 10^{-14}, \quad \left|C_{40} - \frac{\pi^2}{6}\right| \approx 9.4 \times 10^{-23},$$

где погрешности получаются по формуле (24) при $n = 1$. При $n = 2, 3, \ldots$ погрешности аналогичны, что также дает оценку равномерной аппроксимации функции $S(x)$, $x \in [0, 1]$.

Суммирование сходящихся рядов со сложно устроенной правой частью $A(x)$ ничем не отличается от суммирования ряда для Базельской проблемы, но только если функция $A(x/(1 + x))$ хорошо аппроксимируется полиномами Чебышёва на интервале $[0, 1]$. В случае, когда это не так, а это может быть только при достаточно сложной особенности функции $A(x)$ в нуле, необходимо сделать замены переменных в уравнении (22), которые учитывают эту сингулярность (см. ниже).

В случае если ряд (19) расходится в обычном смысле, формула (24) (вообще говоря) однозначно определяет его обобщенную сумму, так как асимптотика частичных сумм $s(n)$ и асимптотика решения $S(x)$ могут сократить друг друга как разрешение неопределенности $\infty - \infty$. В этом случае, очевидно, функция $S(x)$ неограниченна на интервале $[0, 1]$ и уравнение (22) решается по-другому.

Возьмем в качестве примера обычный ряд для ζ-функции Римана

$$\zeta(\sigma + 1) = \sum_{n=1}^{\infty} \frac{1}{n^{\sigma+1}}, \quad \mathrm{Re}(\sigma) > 0, \tag{25}$$

т.е. $A(x) = x^{\sigma+1}$ в уравнении (21).

Следуя [9], сделаем в уравнении (21) замену $S(x) = x^{\sigma} V(x)$, а затем подстановку $x \to x/(1+x)$. Тогда вместо уравнения (22) получим уравнение

$$V\left(\frac{x}{1+x}\right) - (1+x)^{\sigma} V(x) = \frac{x}{1+x}, \tag{26}$$

которое не имеет логарифмической особенности в нуле.

Уравнение (26) при $\sigma \neq 0$ имеет явное решение

$$V(x) = -\frac{1}{x^{\sigma}} \zeta\left(\sigma + 1, 1 + \frac{1}{x}\right),$$

где $\zeta()$ – это ζ-функция Гурвица.

Уравнение (26) теперь аппроксимируется без учета начального значения. Это объясняется тем, что однородное уравнение (26) имеет решение

$$V_0(x) = \mathrm{const}\, x^{-\sigma}, \tag{27}$$

которое является полиномом только при $\sigma = 0, -1, -2, \ldots$, т.е. в особой точке и там, где функция Римана определена явно и принимает рациональные значения. Там матрица аппроксимации левой части (26) будет вырождена.

В точке $x = 0$ при $\sigma \neq 0$ функция $V(x)$ имеет асимптотическое разложение в полуплоскости $\mathrm{Re}(x) > 0$,

$$V(x) = -\frac{1}{\sigma} + \frac{x}{2} - \frac{x^2(\sigma+1)}{12} + \cdots = -\frac{1}{\sigma} \sum_{n=0}^{\infty} B(n)\, C(n + \sigma - 1, n)\, x^n, \tag{28}$$

где $B(n)$ – это числа Бернулли, и $C()$ – биномиальный коэффициент. Этот ряд обрывается при $\sigma = -1, -2, \ldots$ и дает решение уравнения (26) в полиномах Бернулли, т.е.

$$V(x) = -\frac{(-1)^{-\sigma}}{\sigma} x^{-\sigma} B(-\sigma, -1/x), \quad \sigma = -1, -2, \ldots.$$

Таким образом, в результате несложных преобразований мы фактически осуществили аналитическое продолжение ζ-функции Римана, определенной рядом (25), на всю комплексную плоскость, кроме особой точки $\sigma = 0$.

Уравнение (26) позволяет дать характеризацию всего множества нулей ζ-функции Римана (как тривиальных, так и нетривиальных) в терминах свойств аналитического решения функционального уравнения, зависящего от параметра.

Предложение 3. *Функция* $\zeta(\sigma + 1) = 0$ *для тех и только тех* $\sigma \in \mathbb{C}$*, для которых решение уравнения* (26) *удовлетворяет набору тождеств*

$$x^{\sigma} V(x) = \sum_{k=1}^{n} \frac{1}{k^{\sigma+1}}, \quad x = 1/n, \quad n \in \mathbb{N},$$

т.е. когда функция $x^\sigma V(x)$ интерполирует данную сумму на всем интервале $[0,1]$ при условии, что $V(x)$ имеет асимптотику (28).

Доказательство очевидно, так как это всего лишь переформулировка формулы суммирования (24). Условие на асимптотику нельзя опустить, так как при $\mathrm{Re}(\sigma) < 0$ всегда существует решение уравнения (26), такое, что $V(1) = 1$ и $V(+0) = -1/\sigma$, в силу (27). Что требовалось доказать.

Например, для $\sigma = -1, -2, \ldots$ имеем согласно (28), $V(0) = -1/\sigma$, и

$$V(1) = 1 - \zeta(\sigma+1) \in \left\{ \frac{3}{2}, \frac{13}{12}, 1, \frac{119}{120}, 1, \frac{253}{252}, 1, \frac{239}{240}, 1, \frac{133}{132}, 1, \frac{32069}{32760}, 1, \frac{13}{12}, \ldots \right\},$$

где единицы соответствуют тривиальным нулям ζ-функции Римана.

При этих значениях σ функция $V(x)$, будучи полиномом, находится точно нашим методом (при $N \geqslant -\sigma$), если выбрать соответствующее краевое значение $V(1)$. Термин «точно» здесь понимается в обозначенном ранее смысле «сколь угодно точно», так как вычисления проводятся в плавающей арифметике, а коэффициенты решения являются рациональными числами.

Здесь мы обращаем внимание на тот факт, что иррациональности, которые использовались в конструировании матриц аппроксимации нашим методом, в результате сокращают друг друга. Это же справедливо для вычислений с полиномами Лежандра.

Если для $\sigma = -1, -2, \ldots$ назначить краевое значение $V(1) = 1$, то функция $V(x)$ по-прежнему находится точно в виде полинома, и с тем же значением $V(0) = -1/\sigma$, так как отличие от предыдущего полинома только в мономе $\mathrm{const}\, x^{-\sigma}$. Однако асимптотика (28) будет нарушена.

Для σ, соответствующих нетривиальным нулям ζ-функции Римана, краевые условия функции $V(x)$ не могут быть назначены, так как моном $\mathrm{const}\, x^{-\sigma}$ уже не будет полиномом, т.е. формально существующее решение $V(x) \in \mathcal{H}$ уравнения (26) с этими свойствами не может быть найдено нашим методом (что скорее преимущество, чем недостаток).

Напомним, что аппроксимация уравнения (26) в пространстве $\mathcal{A}_N$ согласно формулам разд. 3 имеет вид

$$A.s = F.\left([p_{k-1}(f_j)]_{1\leqslant j,k\leqslant N} - \mathrm{Diag}\left[(1+x_1)^\sigma, \ldots, (1+x_N)^\sigma\right].F^{-1} \right).s = r,$$

где вектор $f \in \mathcal{H}_N$ имеет вид

$$f = \left\langle \frac{x_1}{1+x_1}, \ldots, \frac{x_N}{1+x_N} \right\rangle^t,$$

а вектор $r = F.f = X.(E+X)^{-1}.e \in \mathcal{A}_N$.

Тогда вектор коэффициентов Фурье–Чебышёва решения получается как $s = A^{-1}.r$, причем никакой модификации матрицы A и вектора r для учета краевых условий не требуется (так как мы уже разобрались с тривиальными нулями).

Альтернативный (и более эффективный) способ вычислить то же самое – это использовать коллокационное представление функций и операторов. Но это работает только потому, что краевые условия здесь можно не учитывать.

В пространстве $\mathcal{H}_N$ эта задача аппроксимируется как

$$B.v = \left([p_{k-1}(f_j)]_{1\leqslant j,k\leqslant N}.F - Diag\left[(1+x_1)^\sigma, \ldots, (1+x_N)^\sigma\right] \right).v = f,$$

где вектор f определен выше, а неизвестный вектор $v = \langle V(x_1), \ldots, V(x_N)\rangle^t$. Затем находим $v = B^{-1}.f$, а затем вычисляем вектор коэффициентов Фурье решения, $s = F.v$.

Можно проверить, что для любой размерности аппроксимации N результаты, полученные в $\mathcal{A}_N$ или в $\mathcal{H}_N$, всегда различаются лишь в пределах машинной точности. Это является следствием хорошей обусловленности матриц F, F^{-1}.

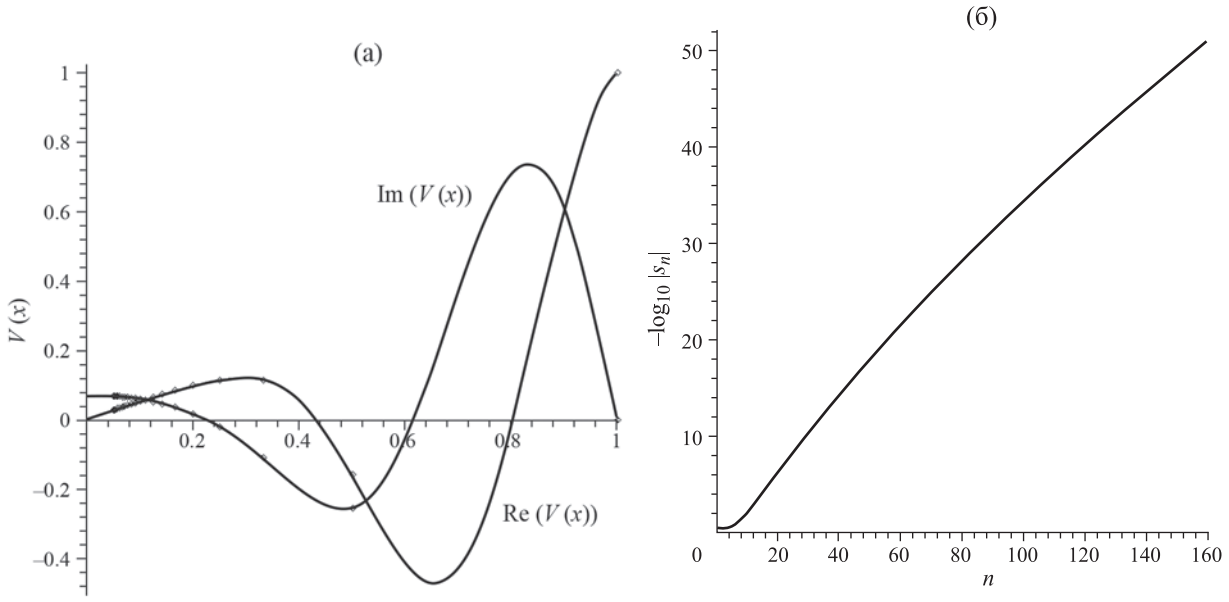
Выберем в качестве примера первый нетривиальный нуль ζ-функции Римана, $\zeta(\sigma+1) = 0$, который в QD-арифметике (в CAS Maple) имеет вид

$$\sigma = -0.5 + i\,14.134725141734693790457251983562470270784257115699243175685567 46,$$

и найдем численное решение задачи для $N = 160$. Результат представлен на фиг. 2.

На фиг. 2а показаны графики вещественной и мнимой части функции $V(x)$, посчитанные по ее интерполяционному полиному Чебышёва

$$V(x) = \sum_{n=0}^{N-1} s_n\, p_n(x), \tag{29}$$



Фиг. 2. Функция $V(x)$ и ее коэффициенты Фурье–Чебышёва.

а также дискретные точки для $n = 1, 2, \ldots, 20$, в которых функция $V(x)$ известна явно, согласно предложению 3.

На фиг. 2б приведены абсолютные величины коэффициентов s_n в логарифмической шкале. Фиг. 2б содержит информацию о достигнутой равномерной погрешности полученного решения ($\approx 10^{-50}$), что также подтверждается краевыми значениями, вычисленными по формуле (29), т.е. $V(1) = 1 - \zeta(\sigma+1) = 1$ и $V(0) = -1/\sigma$ с данной точностью.

Расчеты при $N = 80$ дают точность $\approx 7 \times 10^{-29}$, что также считывается с фиг. 2.

Заметим, что точно так же и с той же точностью вычисляются значения ζ-функции Римана $\zeta(\sigma+1) \approx 1 - V(1)$ для всех близких значений σ. Для других σ размерность аппроксимации и величину разрядной сетки, возможно, придется изменить с учетом требуемой точности.

Таким образом, мы имеем альтернативный способ вычисления ζ-функции Римана с контролируемой точностью (хотя и не самый эффективный).

Также на фиг. 2 видно, что экспоненциального убывания коэффициентов Фурье здесь ожидать не следует, что вполне аналогично фиг. 1. Отличие состоит в том, что здесь коэффициенты Фурье функции $V(x)$ ведут себя весьма регулярно.

Ну и, наконец, заметим, что если бы расчеты велись, например, в DD-арифметике (≈ 32 десятичных разряда), то фиг. 2б оказался бы обрезанным по высоте ординаты ≈ 32, что также является полезной информацией, указывающей на бесполезность (в данном примере и этой арифметике) увеличения размерности аппроксимации больше $N \approx 100$.

В завершение этого раздела дадим пример суммирования расходящегося знакопеременного ряда нашим методом. Рассмотрим классическую задачу вычисления константы Эйлера–Гомперца $\delta = \exp(1)\,\mathrm{Ei}(1,1)$ как суммы всюду расходящегося ряда (8) при $x = 1$.

Мы показали в [9], что этот ряд суммируется точно методом функционального суммирования, но здесь нас интересует численное решение задачи.

Действуя описанными в [9] способом, получим функциональное уравнение

$$W\left(\frac{x}{1+x}\right) + x\,W(x) = 1, \quad x \in [0,1], \tag{30}$$

а также набор тождеств

$$\delta = \sum_{n=1}^{m} (-1)^{n-1}\,(n-1)! + (-1)^m\,(m-1)!\,W\left(\frac{1}{m}\right),\ m \in \mathbb{N}. \tag{31}$$

Уравнение (30) имеет точное решение

$$W(x) = 1 - \exp(1)\,\mathrm{Ei}\left(\frac{1}{x}, 1\right), \tag{32}$$

что легко проверяется. Так что численное решение уравнения (30) – это по сути задача аппроксимации этой функции полиномами Чебышёва.

Здесь следует отметить одно обстоятельство, относящееся к отличию математически точного решения от численного, полученного даже, возможно, с очень большой точностью.

Дело в том, что уравнение (31) выполняется точно для функции $W(x)$ в (32) для $m \in \mathbb{N}$. Однако подстановка в (31) численной аппроксимации функции $W(x)$ приводит к умножению (даже очень малой) погрешности на $(m-1)!$, так что процесс быстро расходится. Но для небольших m в (31) это несущественно.

В результате для размерностей аппроксимации $N = 10,\ 20,\ 40$ получим, соответственно,

$$|\delta_{10} - \delta| \approx 1.8 \times 10^{-5}, \quad |\delta_{20} - \delta| \approx 6.9 \times 10^{-9}, \quad |\delta_{40} - \delta| \approx 1.1 \times 10^{-13},$$

где δ_N вычисляется по формуле (31) при $m = 1$.

Заметим, что диагональная $[N, N]$ Паде–аппроксимация ряда (8) дает приближение для δ на 2–3 порядка хуже.

6. РЕШЕНИЕ НЕЛИНЕЙНЫХ И НЕСТАНДАРТНЫХ ЗАДАЧ

Здесь мы дадим два примера решения нелинейных задач нашим методом, которые примечательны только тем, что стандартные давно известные численные методы (реализованные в CAS) здесь либо неприменимы, либо неэффективны и весьма трудоемки.

До сих пор мы рассматривали только линейные задачи, так что, казалось бы, решение нелинейных задач является существенным обобщением. Однако это не так.

Решение нелинейной задачи в смысле, определенном в начале разд. 3, – это всегда итерационный процесс, в котором участвуют только линейные операторы. Например, метод Ньютона можно охарактеризовать как метод последовательных линеаризаций.

Рассмотрим задачу Коши $y(0) = 0$ для уравнения

$$y'(x) = y^3(x) + x. \tag{33}$$

Это уравнение примечательно тем, что Maple зависает при попытке его проинтегрировать в явном виде (что крайне необычно).

Поставим задачу определить значение $y(1)$ (если оно существует) с большой (скажем, $\approx 10^{-50}$) и контролируемой точностью. Заметим, что стандартная процедура численного интегрирования в Maple (без дополнительных настроек) даст точность $\approx 1.3 \times 10^{-6}$, причем без указания, что точность именно такова.

В Maple существует способ проинтегрировать это уравнение методом тейлоровских разложений с данной точностью. Однако эта процедура для пользователя является «черным ящиком», и нет никаких подтверждений достигнутой точности.

Решение нелинейных задач итерационными методами возможно, в принципе, в пространстве коэффициентов Фурье $\mathcal{A}_N$. Но коллокационный подход, т.е. когда функции представлены таблицами их значений в узлах (12), здесь значительно удобнее. Однако и спектральное представление функций по-прежнему необходимо для учета краевых и иных условий.

Самый простой способ решить уравнение (33) – это применить инерации Пикара, которые здесь, очевидно, сходятся.

Выберем QD-арифметику и $N = 100$ (чтобы технические моменты не влияли на точность аппроксимации).

Возьмем матрицу D и заменим ее последнюю строку на вектор i_0 для учета начального значения. Тогда получим обратимый оператор дифференцирования $D_0 \colon \mathcal{A}_N \to \mathcal{A}_N$. Поэтому матрицы

$$D_p = F^{-1}.\,D_0\,.\,F \colon\ \mathcal{H}_N \to \mathcal{H}_N, \quad \text{и} \quad I_p = D_p^{-1} = F^{-1}.\,D_0^{-1}\,.\,F \colon\ \mathcal{H}_N \to \mathcal{H}_N$$

представляют, соответственно, операторы дифференцирования и интегрирования в пространстве $\mathcal{H}_N$ с учетом начального значения.

Иными словами, применение оператора I_p к вектору $y(x) \in \mathcal{H}_N$ дает вектор $z(x) = \langle z(x_1), \ldots, z(x_N)\rangle^t$, где

$$z(x) = \int\limits_0^x y(t)dt.$$

Поэтому итерации Пикара в пространстве $\mathcal{H}_N$ имеют вид

$$y_{n+1} = I_p \,.\, \langle y_n^3(x_k) + x_k \rangle_{k=1,\dots,N}^t, \quad n \in \mathbb{N}_0,$$

где $y_0 = 0$ – это начальное приближение решения.

В данном случае (т.е. $y(0) = 0$) 35 итераций Пикара дают решение с машинной точностью, т.е. норма вектора $y_n - y_{n-1}$ близка к машинному нулю. Однако это лишь косвенная оценка достигнутой точности. Пробразование Фурье $F.y$ дает коэффициенты разложения функции $y(x)$ по полиномам Чебышёва, по которым видно (как на фиг. 1, 2), что достигнута точность $y(1) \approx 0.5190566558$ порядка 10^{-63}.

Итерации Пикара имеют линейную скорость сходимости, но весьма малозатратны в вычислительном плане и обладают двумя важными преимуществами.

Во-первых, они всегда дают правильный результат, если он есть. Это утверждение – всего лишь переформулировка стандартной теоремы о существовании и единственности решения ОДУ.

Второе преимущество состоит в том, что если нужное решение не существует, то итерации Пикара просто не будут сходиться и укажут тем самым на этот факт.

Например, решение задачи Коши $y(0) = a$ для уравнения (33) имеет вид

$$y_{n+1} = \langle a \rangle_{k=1,\dots,N}^t + I_p \,.\, \langle y_n^3(x_k) + x_k \rangle_{k=1,\dots,N}^t, \quad n \in \mathbb{N}_0,$$

т.е. операторы дифференцирования и интегрирования в $\mathcal{H}_N$ всегда строятся для однородных краевых (и иных) условий. Неоднородности учитываются в самом уравнении или в его аппроксимации.

В данном случае легко проверить, что решение задачи Коши $y(0) = 1/2$ для уравнения (33) существует на интервале $[0, 1]$, но требует большего количества инераций, а решение задачи Коши $y(0) = 1$ не продолжается до $x = 1$.

Метод Ньютона обладает, как известно, квадратичной сходимостью. Однако он весьма чувствителен к выбору начального приближения. Это всегда является отдельной задачей, требующей индивидуального подхода.

В данном случае начальное приближение можно выбрать как отрезок ряда Тейлора решения задачи Коши (33), $y_0(x) = x^2/2$.

Уравнение в вариациях уравнения (33) имеет вид

$$v'(x) = 3\,y^2(x)\,v(x), \tag{34}$$

поэтому итерации метода Ньютона здесь имеют вид $y_{n+1} = y_n - \Delta_n$, $n \in \mathbb{N}_0$,

$$\Delta_n = \left(D_p - 3\,\mathrm{Diag}\left[y_n^2(x_1), \dots, y_n^2(x_N)\right]\right)^{-1} . \Big(D_p . y_n - \langle y_n^3(x_k) + x_k \rangle_{k=1,\dots,N}^t \Big).$$

Здесь мы обращаем внимание на тот факт, что линейный оператор в методе Ньютона вычисляется для конечномерной аппроксимации задачи, а не для исходного уравнения, т.е. уравнение в вариациях (34) играет лишь вспомогательную роль. Иными словами, оператор D_p здесь не может быть заменен на другой оператор дифференцирования.

В данном случае 6 итераций метода Ньютона дают то же, что ранее мы получили с помощью 35 итераций Пикара.

В завершение этого раздела (и статьи) приведем несколько искусственный пример, который иллюстрирует ограничения, существующие в стандартных численных методах. Рассмотрим краевую задачу для уравнения

$$(1 + y(x))\,y''(x) + x = 0$$

на интервале $[0, 1]$ с нестандартными краевыми условиями, $y'(0) = 2\,y(1)$, $2\,y'(1) = -y(1/2)$.

Если бы краевые условия задавались только по краям интервала, то, например, Maple (вообще говоря) генерирует численную процедуру для решения такой задачи. Иными словами, подобные нелинейные краевые задачи являются стандартными. Однако при выбранных нами краевых условиях Maple выдает ошибку типа «слишком много краевых условий».

Тем не менее метод Ньютона в нашей аппроксимации этой задачи дает нужное решение за несколько итераций. Приведем начальные данные этого решения:

$$y(0) \approx -0.035949204045792, \quad y'(0) \approx 0.370656406646618,$$

которые можно проверить численно.

СПИСОК ЛИТЕРАТУРЫ

1. *Варин В.П.* Аппроксимация дифференциальных операторов с учетом граничных условий // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 8. С. 1251–1271.
2. *Варин В.П.* Аппроксимация дифференциальных операторов с учетом граничных условий // Препринты ИПМ им. М.В. Келдыша. 2022. № 77.
3. *Wilf H.S.* Mathematics for the physical sciences. NewYork. Wiley. 1962.
4. *Gantmacher F.R.* Application of the Theory of Matrices. New-York. Chelsea Press. 1960.
5. *Boyd J.P., Petschek R.* The Relationships Between Chebyshev, Legendre and Jacobi Polynomials: The Generic Superiority of Chebyshev Polynomials and Three Important Exceptions // J. of Scientific Computing. 2014. V. 59. P. 1–27.
6. *Варин В.П.* Факториальное преобразование некоторых классических комбинаторных последовательностей // Ж. вычисл. матем. и матем. физ. 2018. Т. 59. № 6. С. 1747–1770.
7. *Pashkovskii S.* Computational Application of Chebyshev Polynomials and Series Moscow. Nauka. 1983. [in Russsian].
8. *Варин В.П.* Инвариантные кривые некоторых дискретных динамических систем // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 2. С. 199–216.
9. *Варин В.П.* Функциональное суммирование рядов // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 1. С. 3–17.

SPECTRAL METHODS FOR SOLVING DIFFERENTIAL AND FUNCTIONAL EQUATIONS

V. P. Varin*

Keldysh Institute of Applied Mathematics RAS, Miusskaya Sq. 4, Moscow, 125047, Russia
**e-mail: varin@keldysh.ru*
Received 16 October, 2023
Revised 16 October, 2023
Accepted 14 January, 2024

Abstract. The operator approach previously developed for the spectral method using Legendre polynomials is generalized here to any systems of basis functions (not necessarily orthogonal) that satisfy two conditions: the result of the operation of multiplication by x or differentiation with respect to x is expressed in the same functions. All systems of classical orthogonal polynomials meet these conditions. In particular, a spectral method utilizing Chebyshev polynomials is constructed, which is most efficient for numerical calculations. This method is applied for the numerical solution of linear functional equations that arise in generalized series summation problems, as well as in problems of analytic continuation of discrete mappings. It is also shown how these methods solve non-standard and nonlinear boundary value problems for which conventional algorithms are not applicable.

Keywords: spectral methods, Chebyshev polynomials, boundary value problems, functional equations, high-precision computations.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 729–731

УДК 512.643

ЕЩЕ РАЗ ОБ ОДНОВРЕМЕННОМ ПРИВЕДЕНИИ ЮНИТОИДОВ К ДИАГОНАЛЬНОМУ ВИДУ

© 2024 г. Х. Д. Икрамов[1,*]

[1]119992 Москва, Ленинские горы, МГУ, ВМК, Россия
**e-mail: ikramov@cs.msu.su*

Поступила в редакцию 01.09.2023 г.
Переработанный вариант 01.09.2023 г.
Принята к публикации 06.02.2024 г.

Настоящая заметка представляет собой дополнение к статье, опубликованной автором на ту же тему ранее. Ее назначение в том, чтобы точнее охарактеризовать пары юнитоидов, допускающих одновременное приведение к диагональному виду. Библ. 15.

Ключевые слова: конгруэнция, юнитоид, коквадрат, канонические углы.

DOI: 10.31857/S0044466924050035, **EDN**: YDMURE

1. Квадратная матрица A называется *юнитоидной*, или попросту *юнитоидом*, если она может быть приведена к диагональному виду *конгруэнцией*, которая здесь понимается как матричное преобразование вида

$$A \to P^*AP,$$

где P — невырожденная матрица.

Если и сама матрица A не вырождена, то все диагональные элементы в ее диагональной форме отличны от нуля. Их аргументы, отсчитываемые в интервале $[0, 2\pi)$, однозначно определены матрицей A и называются ее *каноническими углами*.

Невырожденной матрице A можно сопоставить произведение

$$\mathcal{C}_A = A^{-*}A,$$

называемое *коквадратом* этой матрицы. Если A — юнитоид, то ее коквадрат диагонализуем подобием и имеет унимодулярный спектр. При этом аргументы собственных значений коквадрата равны удвоенным каноническим углам матрицы A.

Настоящая заметка представляет собой дополнение к статье [1]. Главным результатом последней является следующее утверждение.

Предложение 1. *Пусть A и B — невырожденные юнитоиды одинакового порядка n, причем канонические углы матрицы B попарно различны по модулю π. Предположим, что матрицы $C = A^{-1}B$ и $D = A^{-1}B^*$ диагонализуемы одним и тем же подобием, т.е. найдется невырожденная матрица R такая, что*

$$R^{-1}CR = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \tag{1}$$

и

$$R^{-1}DR = M = \operatorname{diag}(\mu_1, \ldots, \mu_n). \tag{2}$$

Тогда A и B могут быть приведены к диагональному виду посредством одной и той же конгруэнции.

Наша цель здесь состоит в том, чтобы точнее охарактеризовать пары юнитоидов, допускающих одновременное приведение к диагональному виду.

2. Наряду с матрицами $C = A^{-1}B$ и $D = A^{-1}B^*$ будем рассматривать коквадрат $\mathcal{C}_B = B^{-*}B$.

Теорема 1. *Всякая матрица R, диагонализующая подобием любые две из матриц C, D и $\mathcal{C}_B$, диагонализует и третью из этих матриц.*

Доказательство. Рассмотрим три возможных ситуации:

а) R диагонализует C и D, а значит, диагонализует и D^{-1}. Так как

$$B^{-*}B = (B^{-*}A)(A^{-1}B) = (A^{-1}B^*)^{-1}(A^{-1}B) = D^{-1}C, \tag{3}$$

то R диагонализует и $\mathcal{C}_B$.

б) R диагонализует C и $\mathcal{C}_B$, а значит, и $\mathcal{C}_B^{-1}$. Имеем

$$A^{-1}B^* = (A^{-1}B)(B^{-1}B^*) = (A^{-1}B)(B^{-*}B)^{-1} = C(\mathcal{C}_B)^{-1}.$$

Следовательно, R диагонализует и D.

в) R диагонализует D и $\mathcal{C}_B$, а потому и $C = A^{-1}B$, поскольку

$$A^{-1}B = (A^{-1}B^*)(B^{-*}B) = D\mathcal{C}_B.$$

(Это же сразу следует из (3).)

3. Покажем теперь, что в условиях предложения 1 обе матрицы A и B переводят подобием матрицу $F = BB^{-*}$ в $\mathcal{C}_B$. Будучи приводимы одновременным подобием, C и D должны быть перестановочны:

$$(A^{-1}B)(A^{-1}B^*) = (A^{-1}B^*)(A^{-1}B).$$

Отсюда выводим

$$BA^{-1}B^* = B^*A^{-1}B$$

и

$$A^{-1}(BB^{-*}) = (B^{-*}B)A^{-1},$$

$$A^{-1}FA = \mathcal{C}_B. \tag{4}$$

Аналогичное соотношение

$$B^{-1}FB = \mathcal{C}_B \tag{5}$$

очевидно, если подставить в него выражение для F.

4. Выражая F из (4) и подставляя полученное выражение в (5), получаем

$$(B^{-1}A)\mathcal{C}_B(A^{-1}B) = \mathcal{C}_B$$

или

$$\mathcal{C}_B(A^{-1}B) = (A^{-1}B)\mathcal{C}_B.$$

Напомним, что множество матриц, перестановочных с данной квадратной матрицей M, называется централизатором этой матрицы.

Итак, матрицы A и B в предложении 1 должны быть таковы, чтобы произведение $C = A^{-1}B$ принадлежало централизатору матрицы $\mathcal{C}_B = B^{-*}B$.

Пусть теперь невырожденная матрица G является элементом централизатора матрицы $\mathcal{C}_B$. Определим матрицу A соотношением

$$G = A^{-1}B,$$

т.е. $A = BG^{-1}$. Тогда

$$A^{-1}B^* = GB^{-1}B^* = G(B^{-*}B)^{-1} = G\mathcal{C}_B^{-1}.$$

Отсюда следует, что

$$(A^{-1}B)(A^{-1}B^*) = G^2\mathcal{C}_B^{-1}$$

и

$$(A^{-1}B^*)(A^{-1}B) = G\mathcal{C}_B^{-1}G.$$

Поскольку G коммутирует с $\mathcal{C}_B$ и $\mathcal{C}_B^{-1}$, то оба произведения $(A^{-1}B)(A^{-1}B^*)$ и $(A^{-1}B^*)(A^{-1}B)$ равны, т.е. матрицы C и D перестановочны.

Тем самым из пп. 3 и 4 можно извлечь следующий вывод.

Теорема 2. *Пусть $n \times n$-матрицы A и B не вырождены. Матрицы $C = A^{-1}B$ и $D = A^{-1}B^*$ коммутируют тогда и только тогда, когда C принадлежит централизатору матрицы $\mathcal{C}_B = B^{-*}B$.*

5. Вернемся к предложению 1. Оно требует от матриц A и B большего, чем просто перестановочность C и D, а именно, эти последние должны быть диагонализуемы подобием и иметь общий базис из собственных

векторов. В то же время в силу теоремы 1 этот базис состоит из собственных векторов матрицы $\mathcal{C}_B = B^{-*}B$. Линии действия этих векторов определены однозначно условием попарного различия канонических углов матрицы B. Поэтому централизатор $\mathcal{C}_B$ представляет собой n-мерную (комплексную) алгебру. В соответствии с теоремой 2 для выбора матриц A годятся только невырожденные матрицы из этого централизатора. Таким образом, мощность множества матриц A, подходящих для заданного юнитоида B, не слишком велика. Этому не стоит удивляться, ведь похожая ситуация имеет место в классической теореме об одновременной приводимости к диагональному виду пары перестановочных эрмитовых матриц (A, B). Если одна из этих матриц, скажем B, имеет простой спектр, то множество подходящих для нее матриц A образует как раз алгебру размерности n. Есть, правда, и отличие: от A и B не требуется невырожденности.

6. Воспользуемся этой возможностью, чтобы прокомментировать доказательство утверждения б) в теореме 2 из [1]. Недостаток приведенного там рассуждения в том, что не показана отчетливо роль предположения о попарном различии канонических углов матрицы B. Между тем это предположение сильно упрощает доказательство. Объясним, почему.

Из равенств (1) и (2) прежним образом выводятся соотношения

$$R^*BR = R^*AR\Lambda \tag{6}$$

и

$$(R^*BR)\Lambda^{-1} = (R^*B^*R)M^{-1}. \tag{7}$$

Положим $S = R^*BR$. Будучи конгруэнтна B, матрица S является юнитоидом. Теперь соотношение (7) можно переписать в виде

$$S^{-*}S = M^{-1}\Lambda \equiv \Gamma.$$

Тем самым диагональная матрица Γ есть коквадрат юнитоида S, а потому все ее диагональные элементы унимодулярны. Более того, их аргументы попарно различны по модулю 2π. Как следствие, диагональна и матрица S. (Аргументы ее диагональных элементов также попарно различны, но по модулю π, как это требуется предложением 1 от канонических углов матрицы B.) Из равенства (6) теперь следует, что

$$R^*AR = S\Lambda^{-1},$$

т.е. R приводит (конгруэнцией) к диагональному виду не только B, но и матрицу A. Этим доказательство завершается.

СПИСОК ЛИТЕРАТУРЫ

1. *Икрамов Х.Д.* Об одновременном приведении к диагональному виду пары юнитоидных матриц // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 2. С. 227–229.

ON THE SIMULTANEOUS DIAGONALIZATION OF UNITOIDS

H. D. Ikramov*

Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Moscow, 119992, Russia
**e-mail: ikramov@cs.msu.su*

Received 01 September, 2023
Revised 01 September, 2023
Accepted 06 February, 2024

Abstract. This note serves as a supplement to a previous article by the author on the same topic. Its purpose is to more precisely characterize pairs of unitoids that allow simultaneous diagonalization.

Keywords: congruence, unitoid, cosquare, canonical angles.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 732–744

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

УДК 517.977

АСИМПТОТИКА РЕШЕНИЯ БИСИНГУЛЯРНОЙ ЗАДАЧИ ОПТИМАЛЬНОГО РАСПРЕДЕЛЕННОГО УПРАВЛЕНИЯ В ВЫПУКЛОЙ ОБЛАСТИ С МАЛЫМ ПАРАМЕТРОМ ПРИ ОДНОЙ ИЗ СТАРШИХ ПРОИЗВОДНЫХ

© 2024 г. А. Р. Данилин[1,*]

[1]620990 Екатеринбург, ул. С. Ковалевской, 16, Институт математики и механики им. Н.Н. Красовского УрО РАН, Россия

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

Поступила в редакцию 27.11.2023 г.

Переработанный вариант 13.01.2024 г.

Принята к публикации 06.02.2024 г.

Рассматривается задача оптимального распределенного управления в плоской строго выпуклой области с гладкой границей и малым параметром при одной из старших производных эллиптического оператора. На границе области в этой задаче задано нулевое условие Дирихле, а управление аддитивно входит в неоднородность. В качестве множества допустимых управлений используется единичный шар в соответствующем пространстве функций, суммируемых с квадратом. Решения получающихся краевых задач рассматриваются в обобщенном смысле как элементы некоторого гильбертова пространства. В качестве критерия оптимальности выступает сумма квадрата нормы отклонения состояния от заданного и квадрата нормы управления с некоторым коэффициентом. Такая структура критерия оптимальности позволяет при необходимости усилить роль либо первого, либо второго слагаемого в этом критерии. В первом случае более важным является достижение заданного состояния, а во втором случае – минимизация ресурсных затрат. Подробно изучена асимптотика задачи, порожденная дифференциальным оператором второго порядка с малым коэффициентом при одной из старших производных, к которому прибавлен дифференциальный оператор нулевого порядка. Библ. 15.

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

DOI: 10.31857/S0044466924050043, **EDN**: YDMIFK

ВВЕДЕНИЕ

Статья посвящена исследованию асимптотики решения задачи оптимального распределенного управления (см. [1]) в плоской строго выпуклой области с гладкой границей и малым параметром при одной из старших производных эллиптического оператора. Такие операторы характерны для установившихся процессов теплопроводности и диффузии в слоистых средах, когда распространение тепла (диффузия) имеют существенно различные коэффициенты по перпендикулярным направлениям (в слое и при переходе в новый слой) (см. [2, гл. III, § 1, п. 3]).

Асимптотика решения задачи Дирихле для подобных эллиптических уравнений в подобных областях была исследована в [3], [4].

Исследование задач оптимального управления, определяемых уравнениями в частных производных, не теряет своей актуальности (см., например, [5]–[7] и библиографию в них).

Асимптотика распределенного управления для оператора с малым коэффициентом при операторе Лапласа и в существенно другой области рассматривалась в [8], [9], а в аналогичной области – в [10]. Регулярный случай подобной задачи рассмотрен в [11].

1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ И УСЛОВИЯ ОПТИМАЛЬНОСТИ

Пусть $\Omega \subset \mathbb{R}^2$ — ограниченная строго выпуклая область с гладкой границей $\Gamma := \partial\Omega$ (Ω — многообразие класса C^∞ с краем).

Рассматривается следующая задача распределенного управления (см. [1, гл. 2, § 2, (2.8)–(2.9)]):

$$\mathcal{L}_\varepsilon z_\varepsilon := -\varepsilon^6\frac{\partial^2 z_\varepsilon}{\partial x^2} - \frac{\partial^2 z_\varepsilon}{\partial y^2} + a(x,y)z_\varepsilon = f(x,y) - u_\varepsilon(x,y), \qquad (x,y)\in\Omega,\ z_\varepsilon\in H_0^1(\Omega), \tag{1.1}$$

$$J(u) := \|z_\varepsilon - z_d\|^2 + \beta^{-1}\|u\|^2 \longrightarrow \inf,\ u\in\mathcal{U}, \tag{1.2}$$

$$\mathcal{U} = \mathcal{U}(1),\ \text{где}\ \mathcal{U}(r) := \{u\in L_2(\Omega) : \|u\|\leqslant r\}. \tag{1.3}$$

Здесь $\beta>0$, $H_0^1(\Omega)$ — соболевское пространство дифференцируемых функций, равных нулю на границе $\partial\Omega$ (см., например, [14]), $\|\cdot\|$ – норма в пространстве $L_2(\Omega)$,

$$f, z_d, a, \in C^\infty(\overline{\Omega_{\widetilde{\delta}}}),\ a(x,y)\geqslant\alpha^2>0 \text{ при } (x,y)\in\Omega, \tag{1.4}$$

где $\widetilde{\delta}>0$, а $\Omega_{\widetilde{\delta}}$ — $\widetilde{\delta}$-окрестность области Ω.

Скалярное произведение в $L_2(\Omega)$ будем обозначать $\langle\cdot,\cdot\rangle$.

Отметим, что степень шесть малого параметра взята для технического удобства, чтобы не писать в дальнейшем дробных степеней этого параметра.

Решение уравнения (1.1) понимается в слабом смысле: для любого $v\in H_0^1(\Omega)$ справедливо равенство

$$\varepsilon^6\left(\frac{\partial z}{\partial x},\frac{\partial v}{\partial x}\right) + \left(\frac{\partial z}{\partial y},\frac{\partial v}{\partial y}\right) + (a(x,y)z,v) = (f+u,v). \tag{1.5}$$

В силу (1.4) и (1.5) при всех малых $\varepsilon>0$ справедливо соотношение

$$(\mathcal{L}_\varepsilon v,v) = \varepsilon^6\left\|\frac{\partial v}{\partial x}\right\|^2 + \left\|\frac{\partial v}{\partial y}\right\|^2 + (a(x,y)v,v) \geqslant \varepsilon^6\left\|\frac{\partial v}{\partial x}\right\|^2 + \left\|\frac{\partial v}{\partial y}\right\|^2 + \alpha^2\|v\|^2 \geqslant \varepsilon^6\|v\|^2_{H_0^1(\Omega)}.$$

В этом случае единственное оптимальное управление $u_\varepsilon(\cdot)$ и соответствующее ему $z_\varepsilon(\cdot)$ характеризуются следующим образом: существует $p_\varepsilon\in H_0^1(\Omega)$ такое, что (см. [1, Гл. 2, § 2, (2.10)])

$$\mathcal{L}_\varepsilon z_\varepsilon = f(x,y) + u_\varepsilon, \quad \mathcal{L}_\varepsilon p_\varepsilon - z_\varepsilon = -z_d(x,y),\ \ (x,y)\in\Omega,\ z_\varepsilon,\ p_\varepsilon\in H_0^1(\Omega),$$

$$\forall\,\widetilde{v}\in\mathcal{U} \qquad \big(p+\beta^{-1}u_\varepsilon, (\widetilde{v}-u_\varepsilon)\big)\geqslant 0. \tag{1.6}$$

Как показано в [15, лемма 1] в этом случае условие (1.6) эквивалентно следующему: существует $\lambda_\varepsilon>0$ такое, что

$$\big(u_\varepsilon = -\lambda_\varepsilon p_\varepsilon\big)\wedge\big(\lambda_\varepsilon\in(0;\beta]\big)\wedge\big(\lambda_\varepsilon\|p_\varepsilon\|\leqslant 1\big)\wedge\Big((\beta-\lambda_\varepsilon)\cdot\big(1-\lambda_\varepsilon\|p_\varepsilon\|\big)=0\Big). \tag{1.7}$$

Таким образом, исходная задача свелась к системе уравнений

$$\mathcal{L}_\varepsilon z_\varepsilon + \lambda_\varepsilon p_\varepsilon = f(x,y), \quad \mathcal{L}_\varepsilon p_\varepsilon - z_\varepsilon = -z_d(x,y),\ \ (x,y)\in\Omega,\ z_\varepsilon,\ p_\varepsilon\in H_0^1(\Omega), \tag{1.8}$$

зависящей от положительного скалярного параметра λ_ε с дополнительным условием (1.7).

Отметим, что в силу (1.4) при любом $\lambda_\varepsilon>0$ решения системы (1.8) бесконечно дифференцируемы в Ω.

Цель работы — изучить поведение z_ε, p_ε и λ_ε при $\varepsilon\to 0$ и найти полное асимптотические разложение указанных величин при $\varepsilon\to 0$.

Обратим внимание, несмотря на то что порядок уравнения $\mathcal{L}_\varepsilon z_\varepsilon = f$ в задаче Дирихле при $\varepsilon=0$ не вырождается, эта задача при некоторых условиях на границу области $\partial\Omega$ может быть бисингулярной. В работах [3] и [4] показано, что это связано с порядком касания границы области Ω некоторых линейных многообразий. Для рассматриваемого оператора – это прямые, параллельные оси Oy. При порядке касания, равном 1, обычный ряд теории возмущений (внешнее разложение) является пригодным во всей области Ω и, тем самым, дает асимптотическое разложение решения z_ε.

Для рассматриваемой задачи оптимального управления ситуация усложняется за счет появления задачи Дирихле для системы уравнений (1.8) с подобными операторами и появлением дополнительного скалярного параметра $\lambda_\varepsilon>0$, связанного с решением системы (1.8) соотношением (1.7).

Если ограничения на управления не по существу, то $\lambda_\varepsilon=\beta$ — известная константа, и нахождение асимптотического разложения z_ε и p_ε сводится к построению либо только внешнего разложения этой системы (регулярный случай — порядок касания указанных прямых равен 1), либо к методу согласования асимптотических разложений (см. [12], [13]), адаптированному для системы уравнений (сингулярный случай — порядок касания хотя бы одной прямой больше 1).

Однако даже в регулярном случае, когда ограничения на управления по существу, само внешнее разложение зависит от асимптотического разложения параметра λ_ε, и коэффициенты всех разложений z_ε, p_ε и λ_ε необходимо находить одновременно.

В работе [11] доказаны общие теоремы о предельной задаче, об априорных оценках и теоремы аппроксимации, показывающие, что для нахождения асимптотических разложений z_ε, p_ε и λ_ε достаточно построить формальное асимптотическое решение (ФАР) (см., например, [12, гл. I, § 1], [13, § 27, (27.3)]) системы (1.8) с соотношением (1.7).

Кроме того, там был рассмотрен регулярный случай (когда порядок касания обеих указанных прямых равен 1).

В настоящей работе рассматривается сингулярный случай, когда порядок касания одной из прямых равен 3, а порядок касания второй (для упрощения выкладок) равен 1.

2. ПРЕДЕЛЬНЫЕ СООТНОШЕНИЯ

Замечание 1. Разрешимость указанных в этом разделе задач и приведенные факты об их решениях доказаны в [11].

Наряду с (1.8) рассмотрим также систему вида

$$\begin{aligned}&\mathcal{L}_\varepsilon z_{\varepsilon,\lambda}+\lambda p_{\varepsilon,\lambda}=f_{\varepsilon,1}(x,y),\quad \mathcal{L}_\varepsilon p_{\varepsilon,\lambda}-z_{\varepsilon,\lambda}=f_{\varepsilon,2}(x,y),\quad (x,y)\in\Omega,\\&z=0, p=0, (x,y)\in\Gamma,\end{aligned}\tag{2.1}$$

где $\lambda>0$.

При $f_{\varepsilon,1}=f$ и $f_{\varepsilon,2}=-z_d$ решение системы (2.1) будем обозначать: $z_{\varepsilon,\lambda,d}$, $p_{\varepsilon,\lambda,d}$.

Замечание 2. Отметим, что если $\beta\|p_{\varepsilon,\beta,d}\|\leqslant 1$, то $z_{\varepsilon,\beta,d}=z_\varepsilon$ и $p_{\varepsilon,\beta,d}=p_\varepsilon$, а ограничения на управление не по существу.

Предельной для (2.1) будет задача

$$\begin{aligned}&\mathcal{L}_0 z_{0,\lambda}+\lambda p_{0,\lambda}=f_{0,1}(x,y),\quad \mathcal{L}_0 p_{0,\lambda}-z_{0,\lambda}=f_{0,2}(x,y),\quad (x,y)\in\Omega,\\&z=0, p=0, (x,y)\in\Gamma,\end{aligned}\tag{2.2}$$

где оператор $\mathcal{L}_0$ получается из $\mathcal{L}_\varepsilon$, если положить формально $\varepsilon=0$:

$$\mathcal{L}_0 v:=-\frac{\partial^2 v}{\partial y^2}+a(x,y)v.$$

Поскольку область Ω *строго выпукла*, то существуют точки $M_i=(x_i,y_i)\in\Gamma$, $i=1,2$, в которых уравнение касательной к Γ имееет вид $x=x_i$ соответственно. Точки M_i разбивают границу Γ на две части Γ_j: нижнюю ($j=1$) и верхнюю ($j=2$). Обе эти части являются графиками функций $\varphi_j(x)$, $x\in[x_1;x_2]$. При этом

$$\varphi_j(x)\in C([x_1;x_2])\cap C^\infty(x_1;x_2),\qquad \varphi_j(x_i)=y_i,\qquad \varphi_j'(x_i-(-1)^i 0)=\infty.\tag{2.3}$$

В окрестностях точек M_i существует еще одна параметризация границы Γ: $x=\psi_i(y)$ соответственно. Отметим, что ψ_1 — выпуклая ($\psi_1''\geqslant 0$), а ψ_2 — вогнутая ($\psi_1''\leqslant 0$) функции и $\psi_i(x_i)=0$. При выполнении условий (1.4), (2.3) и $f_{\varepsilon,1}, f_{\varepsilon,2}f_{0,1}, f_{0,2}\in C^\infty(\overline{\Omega_\delta})$ задачи (2.1) и (2.2) разрешимы единственным образом, их решения бесконечно дифференцируемы в $\overline{\Omega}\setminus\{M_1,M_2\}$ и

$$\|z_{\varepsilon,\lambda,d}-z_{0,\lambda,d}\|\to 0,\quad \|p_{\varepsilon,\lambda,d}-p_{0,\lambda,d}\|\to 0\quad\text{при }\varepsilon\to 0.$$

Поэтому, если

$$\beta\|p_{0,\beta,d}\|<1,\tag{2.4}$$

то $\lambda_\varepsilon=\beta$ при всех малых $\varepsilon>0$, т.е. ограничения на управление в задаче (1.1)–(1.3) не по существу, и $\|z_\varepsilon-z_{0,\beta,d}\|\to 0$, $\|p_\varepsilon-p_{0,\beta,d}\|\to 0$ при $\varepsilon\to 0$, если $\beta\|p_{0,\beta,d}\|>1$, то при всех малых $\varepsilon>0$ ограничения на управление в задаче (1.1)–(1.3) по существу и

$$\lambda_\varepsilon\|p_\varepsilon\|=1\tag{2.5}$$

при всех таких ε.

При выполнении условий

$$\mathcal{L}_0 z_d\neq f\ \text{ и }\ \beta\|p_{0,\beta,d}\|>1\tag{2.6}$$

существует единственное $\lambda_0\in(0,\beta)$ такое, что

$$\lambda_0\|p_{0,\lambda_0,d}\|=1\text{ и }\lambda_\varepsilon\longrightarrow\lambda_0,\quad \|z_\varepsilon-z_{0,\lambda_0,d}\|\to 0,\quad \|p_\varepsilon-p_{0,\lambda_0,d}\|\to 0\quad\text{при }\varepsilon\to 0.\tag{2.7}$$

3. ВНЕШНЕЕ АСИМПТОТИЧЕСКОЕ РАЗЛОЖЕНИЕ

В отличие от [10], поскольку при $\varepsilon = 0$ система (1.8) остается системой обыкновенных дифференциальных уравнений второго порядка, гладко зависящей от параметра x, то с помощью внешнего разложения удается удовлетворить граничным условиям без пограничного экспоненциально убывающего пограничного слоя.

Внешнее разложение для z_ε и p_ε и разложение для λ_ε ищем в виде

$$z_{\text{out}} := \sum_{k=0}^{\infty} \varepsilon^k z_k(x,y), \quad p_{\text{out}} := \sum_{k=0}^{\infty} \varepsilon^k p_k(x,y), \quad \Lambda := \sum_{k=0}^{\infty} \varepsilon^k \lambda_k. \tag{3.1}$$

Подставим ряды (3.1) в систему (1.8) и приравняем слагаемые одинакового порядка малости. В результате получим уравнения, связывающие между собой z_k, p_k и λ_k:

$$\begin{aligned} &\mathcal{L}_0 z_0 + \lambda_0 p_0 = f(x,y), && \mathcal{L}_0 p_0 - z_0 = -z_d(x,y), && \\ &\mathcal{L}_0 z_k + \lambda_0 p_k + \lambda_k p_0 = F_{1,k}, && \mathcal{L}_0 p_k - z_k = F_{2,k}, && k \geqslant 1, \\ &z_k|_\Gamma = 0 = p_k|_\Gamma, && && k \geqslant 0, \end{aligned} \tag{3.2}$$

где

$$F_{1,k} = b_k \frac{\partial^2 z_k}{\partial x^2} - \sum_{l=1}^{k-1} \lambda_l p_{k-l}, \quad F_{2,k} = b_k \frac{\partial^2 p_{k-1}}{\partial x^2},$$

и $b_k = 0$ при k не кратном 6.

Отметим, что система (3.2) имеет единственное решение при любом заданном наборе $\{\lambda_k\}$. Однако в системе (3.2) никак не учтено дополнительное условие (1.7).

Итак, внешнее разложение (при заданном наборе $\{\lambda_k\}$) построено. Оно по построению является ФАР задачи (1.8) (с заданным рядом Λ из (3.1)) в тех подобластях области Ω, где ряды для z_ε и p_ε из (3.1) не теряют своей асимптотичности.

В отличие от [11] в дальнейшем будем предполагать, что

$$x_1 = y_1 = 0, \quad \psi_1(y) = y^4 \text{ при } 0 < x < \delta_0 \text{ и некотором малом } \delta_0, \text{ а } \psi_2''(y_2) < 0. \tag{3.3}$$

Отметим, что эти ряды пригодны во всей области Ω, за исключением малой окрестности точки M_1 .

Покажем, что эти ряды в малой окрестности точки M_1 теряют свой асимптотический характер. Для этого исследуем асимптотику функций z_k и p_k при $(x,y) \to M_1$.

Нахождение асимптотики коэффициентов внешнего разложения будем вести по схеме работы [3] с необходимыми изменениями, связанными с решением системы уравнений (в [3] исследовалась асимптотика коэффициентов внешнего разложения для одного уравнения). Кроме этого для дальнейшего нам потребуется уточнение вида разложений из [3].

При выполнении условия (3.3) функции φ_j, определяющие Γ_j, при $0 < x < \delta_0$ имеют следующий вид:

$$\varphi_j(x) = (-)^j \sqrt[4]{x}. \tag{3.4}$$

Как и в [3] рассмотрим функции вида $x^{m/4} P(\omega(x,y))$, $(x,y) \in \Omega$, где $m \in \mathbb{Z}$, $\omega(x,y) := y/\sqrt[4]{x}$, а P — полином. Функции такого вида будем обозначать $R_m(x,y)$ (возможно с дополнительными индексами). Число m естественно назвать *порядком* такой функции. Если полином P есть четная (нечетная функция) т.е. все степени ω в P четные (нечетные), то такой полином будем обозначать $P(\omega;1)$ ($P(\omega;-1)$). Соответственно будем использовать обозначения

$$R_m(x,y;i) := x^{m/4} P(\omega;i) \;\text{ и }\; \mathcal{R}_m^+(x,y;i) := \sum_{s=0}^{\infty} R_{m+s}(x,y;i\cdot(-1)^s).$$

Замечание 3. При использовании обозначения $\mathcal{R}_m^+(x,y;i)$ не исключается случай, что некоторые слагаемые в сумме равны нулю (включая и начальные, или даже все). Тем самым, содержательно запись $f \overset{\text{as}}{=} \mathcal{R}_m^+(x,y;i)$ говорит о том, что у функции f асимптотика в нуле не хуже, чем $x^{m/4}$. В частности, функцию, тождественно равную нулю, при необходимости можно считать имеющей асимптотику вида $\mathcal{R}_m^+(x,y;i)$ для нужного нам порядка m.

Отметим простейшие свойства функций $R_m(x,y;i)$:

$$\begin{gathered} x^m y^n = x^{(4m-n)/4}\Big(\frac{y}{\sqrt[4]{x}}\Big)^n = R_{4m-n}(x,y;(-1)^n), \\ R_m(x,y;i)\cdot R_n(x,y;j) = R_{m+n}(x,y;i\cdot j), \\ \frac{\partial^2}{\partial x^2}R_m(x,y;i) = R_{m-8}(x,y;i), \quad \frac{\partial^2}{\partial y^2}R_m(x,y;i) = R_{m-2}(x,y;i), \\ \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} R_m(x,y;i)dy = \gamma_{m,i}x^{(m+1)/4}, \text{ причем } \gamma_{m,-1}=0. \end{gathered} \tag{3.5}$$

Кроме того, если при $(x+|y|)\to 0$

$$f(x,y)\overset{\text{as}}{=}\sum_{n,m=0}^{\infty} a_{n,m}x^n y^m, \text{ то } f(x,y)\overset{\text{as}}{=}\mathcal{R}_0^+(x,y;1). \tag{3.6}$$

Как показано в [3, следствие 1], единственное решение задачи

$$\mathcal{L}_{0,0}Z := -\frac{\partial^2}{\partial y^2}Z = R_m(x,y;i), \quad Z(-\sqrt[4]{x}) = 0 = Z(\sqrt[4]{x}) \tag{3.7}$$

имеет вид $Z = R_{m+2}(x,y;i)$.

В частности, если $R_0(x,y;1)=1$, то $Z = 1/2(y^2-\sqrt{x}) = R_2(x,y;1)$.

Лемма 1. *Пусть выполнены условия* (1.4) *и* (3.4) *и* $\lambda>0$ *— некоторое число. Тогда решение системы*

$$\begin{aligned} &\mathcal{L}_0 z + \lambda p = f_1(x,y)\overset{\text{as}}{=}\mathcal{R}_{m,1}(x,y,i), \quad \mathcal{L}_0 p - z = f_1(x,y)\overset{\text{as}}{=}\mathcal{R}_{m,2}(x,y,i), \\ &z|_\Gamma = 0 = p|_\Gamma, \end{aligned} \tag{3.8}$$

имеет при $(x+|y|)\to 0$ *асимптотику вида*

$$z(x,y)\overset{\text{as}}{=}\mathcal{R}^+_{m+2,z}(x,y;i), \quad p(x,y)\overset{\text{as}}{=}\mathcal{R}^+_{m+2,p}(x,y;i).$$

Это асимптотическое представление можно дифференцировать сколько угодно раз.

Доказательство. Отметим прежде всего, что в силу (3.6)

$$a(x,y)\overset{\text{as}}{=}\mathcal{R}^+_{0,a}(x,y;1).$$

Будем искать асимптотику функций z и p в виде

$$z(x,y)\overset{\text{as}}{=}\sum_{s=0}^{\infty} R_{m(s),1}(x,y;i_{m(s)}), \quad p(x,y)\overset{\text{as}}{=}\sum_{s=0}^{\infty} R_{n(s),2}(x,y;i_{n(s)})$$

с помощью стандартной процедуры: подставив это представление в систему (3.8) и приравнивая слагаемые с наименьшим порядком:

$$\begin{aligned} &\sum_{s=0}^{\infty}\mathcal{L}_0\left(R_{m(s),1}(x,y;i_{m(s)})\right) + \lambda\sum_{s=0}^{\infty}R_{n(s),2}(x,y;i_{n(s)}) = \sum_{s=0}^{\infty}R_{m+s,1}(x,y;i(-1)^s), \\ &\sum_{s=0}^{\infty}\mathcal{L}_0\left(R_{n(s),1}(x,y;i_{n(s)})\right) - \sum_{s=0}^{\infty}R_{m(s),1}(x,y;i_{m(s)}) = \sum_{s=0}^{\infty}R_{m+s,2}(x,y;i(-1)^s). \end{aligned} \tag{3.9}$$

Поскольку в силу (3.7) и (3.5)

$$\mathcal{L}_0\left(R_m(x,y;i)\right) = \mathcal{L}_{0,0}R_m(x,y;i) + R_m(x,y;i)\mathcal{R}^+_{0,a}(x,y;1) = R_{m-2}(x,y;i) + \mathcal{R}^+_{m,a}(x,y;i),$$

то в (3.9) наименьшие порядки равны $m(0)-2$, $n(0)-2$ и m. Поэтому $R_{m(0),1}(x,y;i_{m(0)})$ и $R_{n(0),2}(x,y;i_{n(0)})$ есть решения задач вида (3.7):

$$\mathcal{L}_{0,0}R_{m(0),1}(x,y;i_{m(0)}) = R_{m,1}(x,y;i), \quad \mathcal{L}_{0,0}R_{n(0),2}(x,y;i_{n(0)}) = R_{m,2}(x,y;i).$$

Тем самым, в силу (3.7) $m(0) = m+2 = n(0)$, $i_{m(0)} = i = i_{n(0)}$ и $R_{m(0),1}(x,y;i_{m(0)}) = R_{m+2,z}(x,y;i)$, а $R_{n(0),1}(x,y;i_{n(0)}) = R_{m+2,p}(x,y;i)$.

Подставив найденные слагаемые в (3.9), получим равенства

$$\sum_{s=1}^{\infty}\mathcal{L}_0\left(R_{m(s),1}(x,y;i_{m(s)})\right)+\lambda\sum_{s=1}^{\infty}R_{m+2,p}(x,y;i)+\lambda\sum_{s=1}^{\infty}R_{n(s),2}(x,y;i_{n(s)})=$$

$$=\sum_{s=0}^{\infty}R_{m+s,1}(x,y;i(-1)^s),$$

$$\sum_{s=0}^{\infty}\mathcal{L}_0\left(R_{n(s),1}(x,y;i_{n(s)})\right)-R_{m+2,z}(x,y;i)-\sum_{s=0}^{\infty}R_{m(s),1}(x,y;i_{m(s)})=$$

$$=\sum_{s=1}^{\infty}R_{m+s,2}(x,y;i(-1)^s).$$

Теперь равенство элементов наименьшего порядка выглядит следующим образом:

$$\mathcal{L}_{0,0}R_{m(1),1}(x,y;i_{m(1)})=R_{m+1,1}(x,y;-i),\quad \mathcal{L}_{0,0}R_{n(1),2}(x,y;i_{n(1)})=R_{m,2}(x,y;-i),$$

откуда получаем

$R_{m(1),1}(x,y;i_{m(1)})=R_{m+3,z}(x,y;-i)$, а $R_{n(1),1}(x,y;i_{n(1)})=R_{m+3,p}(x,y;-i)$.

Далее, применяя метод математической индукции, получим асимптотические ряды указанного в лемме вида для z и p, дающие асимптотическое разложение при $(x+|y|)\to 0$ функций z и p, что следует из теорем об аппроксимации (см. [11]).

Возможность почленного дифференцирования этих асимптотических разложений доказывается стандартно, так как вид построенных рядов не меняется после почленного дифференцирования (см., например, [12, гл. IV, §3, лемма 3.1]). Лемма 1 доказана.

Теорема 1. *Пусть выполнены условия* (1.4) *и* (3.3). *Тогда для любого набора* $\{\lambda_k\}_{k=0}^{\infty}$ *с* $\lambda_0>0$ *решение системы* (3.2) *имеет при* $(x+|y|)\to 0$ *асимптотику вида*

$$\begin{aligned} z_{6m+r}(x,y)&\overset{\text{as}}{=}\begin{cases}\mathcal{R}^{+}_{2-6m,z}(x,y;1),\ r=0,\\ \mathcal{R}^{+}_{4-6m,z}(x,y;1),\ r=\overline{1,5},\end{cases}\\ p_{6m+r}(x,y)&\overset{\text{as}}{=}\begin{cases}\mathcal{R}^{+}_{2-6m,p}(x,y;1),\ r=0,\\ \mathcal{R}^{+}_{4-6m,p}(x,y;1),\ r=\overline{1,5}.\end{cases}\end{aligned}\tag{3.10}$$

Это асимптотическое представление можно дифференцировать сколько угодно раз.

Доказательство. Отметим прежде всего, что в силу (3.6)

$$f(x,y)\overset{\text{as}}{=}\mathcal{R}^{+}_{0,f}(x,y;1),\quad -z_d(x,y)\overset{\text{as}}{=}\mathcal{R}^{+}_{0,d}(x,y;1).$$

По лемме 1 получим $z_0\overset{\text{as}}{=}\mathcal{R}^{+}_{2,z,0}(x,y;1)$, $p_0\overset{\text{as}}{=}\mathcal{R}^{+}_{2,p,0}(x,y;1)$.

Для z_1 и p_1 получим уравнения с правыми частями

$$F_{1,1}(x,y)=-\lambda_1 p_0\overset{\text{as}}{=}\mathcal{R}^{+}_{2,1,z}(x,y;1),\quad F_{1,2}(x,y)=0\overset{\text{as}}{=}\mathcal{R}^{+}_{2,1,p}(x,y;1).$$

Применив лемму 1, получим $z_1\overset{\text{as}}{=}\mathcal{R}^{+}_{4,z,1}(x,y;1)$, $p_1\overset{\text{as}}{=}\mathcal{R}^{+}_{4,p,1}(x,y;1)$.

Для z_2 и p_2 правые части имеют вид

$$F_{2,1}(x,y)=-\lambda_2 p_0-\lambda_1 p_1\overset{\text{as}}{=}\mathcal{R}^{+}_{2,1,z}(x,y;1),\quad F_{2,2}(x,y)=0\overset{\text{as}}{=}\mathcal{R}^{+}_{2,1,p}(x,y;1).$$

И, тем самым, опять $z_2\overset{\text{as}}{=}\mathcal{R}^{+}_{4,z,2}(x,y;1)$, $p_1\overset{\text{as}}{=}\mathcal{R}^{+}_{4,p,2}(x,y;1)$.

Для z_6 и p_6 ситуация меняется, поскольку

$$F_{6,1}(x,y)=\frac{\partial^2}{\partial x^2}z_0-\sum_{s=1}^{6}\lambda_s p_{6-s}\overset{\text{as}}{=}\mathcal{R}^{+}_{-6,1,z}(x,y;1),\quad F_{6,2}(x,y)=\frac{\partial^2}{\partial x^2}p_0\overset{\text{as}}{=}\mathcal{R}^{+}_{-6,1,p}(x,y;1).$$

Поэтому $z_6\overset{\text{as}}{=}\mathcal{R}^{+}_{-4,z,6}(x,y;1)$, $p_6\overset{\text{as}}{=}\mathcal{R}^{+}_{-4,p,6}(x,y;1)$.

Далее применяем метод математической индукции. Теорема 1 доказана.

Для одного уравнения аналогичный результат получен в [3, теорема 2].

Замечание 4. Отметим, что порядки особенностей асимптотик, указанные в теореме, могут быть сильно завышены. Так, например, если $\lambda_1 = 0$, то $z_1 = 0$ и $p_1 = 0$. С другой стороны, если $z_{d,0,0} \neq 0$, то p_6 действительно имеет в нуле особенность порядка x^{-1}.

Таким образом, внешнее разложение имеет при $(x + |y|) \to 0$ нарастающие особенности в асимптотике коэффициентов разложения, что говорит о бисингулярности рассматриваемой задачи.

Отметим, что внешнее разложения не теряет асимптотичности (и, тем самым, “вполне пригодно”) в области $\Omega_{\varepsilon,\gamma} := \{(x, y) \in \Omega : x > \varepsilon^\gamma\}$ при $\gamma \in (0; 4)$.

4. ВНУТРЕННЕЕ РАЗЛОЖЕНИЕ

В связи с тем, что внешнее разложение непригодно в малой окрестности точки M_1, необходимо в окрестностях этой точки рассмотреть новое — *внутреннее* — разложение в растянутых переменных.

В окрестности точки M_1 введем новые растянутые переменные, подобно тому, как это было сделано в [3, п. 1.2]: $x = \varepsilon^4\xi$, $y = \varepsilon\eta$.

В этих переменных функции $V_\varepsilon(\xi, \eta) := z_\varepsilon\big(\varepsilon^4\xi, \varepsilon\eta\big)$, $W_\varepsilon(\xi, \eta) := p_\varepsilon\big(\varepsilon^4\xi, \varepsilon\eta\big)$ будут удовлетворять системе

$$\begin{aligned} &-\frac{\partial^2}{\partial\xi^2}V_\varepsilon - \frac{\partial^2}{\partial\eta^2}V_\varepsilon + \varepsilon^2 a\big(\varepsilon^4\xi, \varepsilon\eta\big)V_\varepsilon + \varepsilon^2\lambda_\varepsilon W_\varepsilon = \varepsilon^2 f\big(\varepsilon^4\xi, \varepsilon\eta\big), \\ &-\frac{\partial^2}{\partial\xi^2}W_\varepsilon - \frac{\partial^2}{\partial\eta^2}W_\varepsilon + \varepsilon^2 a\big(\varepsilon^4\xi, \varepsilon\eta\big)W_\varepsilon - \varepsilon^2 V_\varepsilon = -\varepsilon^2 z_d\big(\varepsilon^4\xi, \varepsilon\eta\big) \end{aligned} \tag{4.1}$$

в области (см. (3.3)) $0 < \xi < \delta_1\varepsilon^{-4}$, $|\eta| < \sqrt[4]{\xi}$, где δ_1 — некоторая константа $\delta_1 < \delta_0$, и граничным условиям

$$V_\varepsilon\big(\xi, -\sqrt[4]{\xi}\big) = 0 = V_\varepsilon\big(\xi, \sqrt[4]{\xi}\big), \quad W_\varepsilon\big(\xi, -\sqrt[4]{\xi}\big) = 0 = W_\varepsilon\big(\xi, \sqrt[4]{\xi}\big). \tag{4.2}$$

Теперь надо построить внутреннее разложение при $(\xi^2 + \eta^2) \to +\infty$ для функций V_ε и W_ε, согласованное с внешним разложением при $(x + |y|) \to 0$ для z_ε и p_ε (см., например, [12, (0.9)], [13, § 28, (28.21)]). Для получения вида такого внутреннего разложения надо переразложить внешнее разложение через новые переменные.

Отметим, что если $x = \varepsilon^4\xi$, а $y = \varepsilon\eta$, то

$$R_m(x, y; i) = x^{m/4}P(\omega(x, y); i) = \varepsilon\xi^{m/4}P(\widetilde{\omega}(\xi, \eta); i) = \varepsilon^m R_m(\xi, \eta; i), \tag{4.3}$$

где $\widetilde{\omega}(\xi, \eta) = \xi/\sqrt[4]{\eta}$.

Для асимптотических разложений при $(\xi^2 + \eta^2) \to +\infty$ будем использовать обозначение

$$\mathcal{R}^-_m(\xi, \eta; i) := \sum_{s=0}^{\infty} R_{m-s}(\xi, \eta; i \cdot (-1)^s).$$

Перераскладывая внешнее разложение (см. (3.10)) через новые переменные, получим

$$z_\varepsilon \overset{\text{as}}{=} \sum_{m=0}^{\infty} \varepsilon^{6m} \left(\sum_{s=0}^{\infty} R_{2-6m+s,z}(x, y; (-1)^s) + \sum_{j=1}^{5}\sum_{s=0}^{\infty} \varepsilon^j R_{4-6m+s,z}(x, y; (-1)^s) \right) \overset{(4.3)}{=}$$

$$\overset{(4.3)}{=} \sum_{m=0}^{\infty} \varepsilon^{6m} \left(\sum_{s=0}^{\infty} \varepsilon^{2-6m+s} R_{2-6m+s,z}(\xi, \eta; (-1)^s) + \sum_{j=1}^{5}\sum_{s=0}^{\infty} \varepsilon^{j+4-6m+s} R_{4-6m+s,z}(\xi, \eta; (-1)^s) \right) =$$

$$= \varepsilon^2 \sum_{s=0}^{\infty} \varepsilon^s \left(\sum_{m=0}^{\infty} R_{2-6m+s,z}(\xi, \eta; (-1)^s) + \sum_{j=1}^{5}\sum_{s=0}^{\infty} \varepsilon^{j+2} R_{4-6m+s,z}(\xi, \eta; (-1)^s) \right) =$$

$$= \varepsilon^2 \sum_{n=0}^{\infty} \varepsilon^n \mathcal{R}^-_{n+2,z}(\xi, \eta; (-1)^n).$$

Аналогично

$$p_\varepsilon \overset{\text{as}}{=} \varepsilon^2 \sum_{n=0}^{\infty} \varepsilon^n \mathcal{R}^-_{n+2,p}(\xi, \eta; (-1)^n).$$

Замечание 5. Отметим, что порядки особенностей асимптотик, указанные в предыдущих формулах, могут быть сильно завышены. Так, в частности,

$$\mathcal{R}^-_{2,z}(\xi,\eta;1)=\sum_{\sigma=0}^{\infty}R_{2-6\sigma}(\xi,\eta;1).$$

Таким образом, внутреннее разложение функций будем искать в виде

$$z_{in}:=V(\xi,\eta)=\varepsilon^2\sum_{n=0}^{\infty}\varepsilon^nV_n(\xi,\eta),\quad p_{in}:=W(\xi,\eta)=\varepsilon^2\sum_{n=0}^{\infty}\varepsilon^nW_n(\xi,\eta) \tag{4.4}$$

с дополнительным условием; при $\xi^2+\eta^2\to+\infty$

$$V_n(\xi,\eta)\overset{\text{as}}{=}\mathcal{R}^-_{n+2,z}(\xi,\eta;(-1)^n),\quad W_n(\xi,\eta)\overset{\text{as}}{=}\mathcal{R}^-_{n+2,p}(\xi,\eta;(-1)^n), \tag{4.5}$$

где $\mathcal{R}^-_{n+2,z}(\xi,\eta;(-1)^n)$ и $\mathcal{R}^-_{n+2,p}(\xi,\eta;(-1)^n)$ – известные, построенные по внешнему разложению асимптотические ряды.

Подставим ряды (3.1) в систему (4.1) и приравняем слагаемые одинакового порядка малости. В результате для определения функций V_n и W_n получим систему уравнений в неограниченной области $\widetilde{\Omega}:=\{(\xi,\eta):\xi>0,\ |\eta|<\sqrt[4]{\xi}\}$:

$$\begin{aligned}&\Delta V_0=-f_{0,0},\quad &&\Delta W_0=-z_{d,0,0},\\ &\Delta V_1=-f_{0,1}\eta,\quad &&\Delta W_1=-z_{d,0,1}\eta,\\ &\Delta V_n=H_{n,1},\quad &&\Delta W_n=H_{n,2},\qquad n>1,\end{aligned} \tag{4.6}$$

где $H_{n,1}$ и $H_{n,2}$ вычисляются рекуррентно по предыдущим V_j, W_j, λ_k и коэффициентам асимптотических разложений в нуле известных функций a, f и z_d, в частности,

$$H_{2,1}=-f_{0,2}\eta^2-a_{0,0}V_0-\lambda_0W_0,\quad H_{2,2}=-z_{d,0,2}\eta^2-a_{0,0}W_0+V_0.$$

При этом в силу (4.2) нам нужны решения системы, удовлетворяющие граничным условиям

$$V_n(\xi,-\sqrt[4]{\xi})=0=V_n(\xi,\sqrt[4]{\xi}),\quad W_n(\xi,-\sqrt[4]{\xi})=0=W_n(\xi,\sqrt[4]{\xi}),\ n\geqslant 0, \tag{4.7}$$

и условиям поведения на бесконечности (4.5).

Теорема 2. *Пусть выполнено условие* (1.4) *и* $\{\lambda_k\}$ *— заданная последовательность. Тогда существуют функции* $V_n(\xi,\eta)$, $W_n(\xi,\eta)\in C^\infty(\widetilde{\Omega})$ *такие, что они являются решениями системы* (4.6) *в области* $\widetilde{\Omega}$ *и удовлетворяют условиям* (4.7) *и* (4.5).

Доказательство. Поскольку при каждом n рассматриваемая система распадается на два независимых уравнения одинакового вида, то существование нужных решений можно получить, следуя доказательству теоремы 3.1 из [12, гл. IV, § 3]. Теорема 2 доказана.

Отметим, что построенные ряды не теряют своей асимптотичности при $x<\varepsilon^\gamma$, где $\gamma>0$, т.е. области асимптотичности внешнего и внутреннего разложений пересекаются. Отсюда следует (см., например, доказательство теоремы 1.4 из [12, гл. IV, § 1]), что внешнее разложение есть асимптотическое разложение функций $z_{\varepsilon,\Lambda}$ и $p_{\varepsilon,\Lambda}$ (решение системы (1.8) при разложении λ_ε в ряд Λ из (3.1) с заданными $\{\lambda_k\}$) в области $\Omega_{1,\mu}$, а внутреннее — асимптотическое разложение этих же функций в области $\Omega_{2,\mu}$, где

$$\Omega_{1,\mu}:=\{(x,y)\in\Omega,0<x<\mu\},\quad \Omega_{2,\mu}:=\{(x,y)\in\Omega,x>\mu\},\quad \mu=\varepsilon^\gamma,\ \gamma\in(0;4). \tag{4.8}$$

Замечание 6. При рассмотрении внутреннего разложения в области $\Omega_{1,\mu}$ необходимо вернуться к переменным x и y:

$$z_{\text{in}}=V\Big(\frac{x}{\varepsilon^4},\frac{y}{\varepsilon}\Big)=\varepsilon^2\sum_{n=0}^{\infty}\varepsilon^nV_n\Big(\frac{x}{\varepsilon^4},\frac{y}{\varepsilon}\Big),$$

$$p_{\text{in}}=W\Big(\frac{x}{\varepsilon^4},\frac{y}{\varepsilon}\Big)=\varepsilon^2\sum_{n=0}^{\infty}\varepsilon^nW_n\Big(\frac{x}{\varepsilon^4},\frac{y}{\varepsilon}\Big).$$

5. ПОЛНАЯ АСИМПТОТИКА РЕШЕНИЯ

Итак, при фиксированном наборе $\{\lambda_k\}$ построены согласованные внешние и внутренние ФАР системы (1.8). Эти разложения аппроксимируют решения системы в областях $\Omega_{1,\mu}$ и $\Omega_{1,\mu}$ соответственно.

1. Пусть выполнено условие (2.4). В этом случае $\lambda_\varepsilon = \beta$ и, тем самым, $\lambda_0 = \beta$, $\lambda_k = 0$ при $k > 0$. Поэтому построенные для таких λ_k согласованные внешнее и внутреннее разложения будут ФАР всей задачи, и тем самым справедлива теорема.

Теорема 3. *Пусть выполнены условия* (1.4), (3.3) *и* (2.4). *Тогда внешнее разложение* (3.1) *с* $\lambda_0 = \beta$, $\lambda_k = 0$ *при* $k > 0$ *и согласованное с ним внутреннее разложение* (4.4) *есть асимптотическое разложение решения задачи* (1.8) *с* $\lambda_\varepsilon = \beta$ *при* $\varepsilon \to +0$ *в областях* (4.8) *соответственно.*

2. Пусть выполнено условие (2.6). В этом случае условие (2.5) с разложением λ_ε в силу (3.1) для произвольного набора $\{\lambda_k\}$, вообще говоря, не выполнено. В силу (2.7) известно только λ_0. В этом случае построение внешнего и внутреннего разложений необходимо вести совместно с определением величин $\{\lambda_k\}$, $k > 0$. Эти числа можно найти из асимптотического равенства, порожденного равенством (2.5) и разложениями (3.1) и (4.4). Зафиксируем начальный набор $\{\lambda_k\}$, $k > 0$, и найдем соответствующую ему асимптотику величины $\lambda_\varepsilon^2\|p_\varepsilon\|^2$ при $\varepsilon \to 0$ через p_k и W_n.

Для нахождения такой асимптотики воспользуемся методом *вспомогательного параметра* (см., например, [8], [13, § 30, II, теорема 30.1]).

Пусть μ удовлетворяет условию из (4.8). Тогда

$$\lambda_\varepsilon^2\|p_\varepsilon\|^2 = \lambda_\varepsilon^2 \int\limits_0^\mu dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} p_\varepsilon^2(x,y)dy + \lambda_\varepsilon^2 \int\limits_\mu^{x_2} dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} p_\varepsilon^2(x,y)dy \overset{\text{as}}{=}$$

$$\overset{\text{as}}{=} \left(\sum_{k=0}^\infty \varepsilon^k\lambda_k\right)^2 \int\limits_0^\mu dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} \varepsilon^4 \left(\sum_{n=0}^\infty \varepsilon^n W_n\Big(\frac{x}{\varepsilon^4}, \frac{y}{\varepsilon}\Big)\right)^2 dy +$$

$$+ \left(\sum_{k=0}^\infty \varepsilon^k\lambda_k\right)^2 \int\limits_\mu^{x_2} dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} \left(\sum_{k=0}^\infty \varepsilon^k p_k(x,y)\right)^2 dy.$$

Сделав в первом интеграле замену $x = \varepsilon^4\xi$, $y = \varepsilon\eta$, получим асимптотическое равенство для нахождения коэффициентов λ_k:

$$\begin{aligned} 1 \overset{\text{as}}{=} \varepsilon^7 \left(\sum_{k=0}^\infty \varepsilon^k\lambda_k\right)^2 \int\limits_0^{\mu/\varepsilon^4} d\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} \left(\sum_{n=0}^\infty \varepsilon^n W_n(\xi,\eta)\right)^2 d\eta + \\ + \left(\sum_{k=0}^\infty \varepsilon^k\lambda_k\right)^2 \int\limits_\mu^{x_2} dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} \left(\sum_{k=0}^\infty \varepsilon^k p_k(x,y)\right)^2 dy. \end{aligned} \tag{5.1}$$

Поясним нахождение асимптотики получившихся слагаемых.

Во внешнем разложении надо из $p_k(x,y)p_l(x,y)$ вычесть неинтегрируемую по Ω часть асимптотики (слагаемые вида $R_m(x,y;j_m)$ с $m < 0$), а интеграл по $\Omega_{2,\mu}$ от получившейся части представить в виде разности интегралов по Ω и по $\Omega_{1,\mu}$. Это даст нам слагаемое вида $\varepsilon^{(k+l)}A_{k,l}$ и степенной (по $\mu^{1/4}$) асимптотический ряд $\varepsilon^{(k+l)}\mathcal{F}_1(\mu)$. Наконец, проинтегрировав вычтенную часть асимптотики, состоящую из слагаемых вида $x^{(k+l)/4}P(\omega(x,y); j_k \cdot j_l)$, по y получим сумму выражений

$$\varepsilon^{(k+l)} \int\limits_\mu^{x_2} \gamma_{k,l} x^{(k+l+1)/4} dx,$$

которая в итоге дает сумму $\varepsilon^{(k+l)}A_{k,l,1} + \varepsilon^{(k+l)}\mathcal{F}_2(\mu)$.

Аналогичная процедура — вычитание неинтегрируемой по $\widetilde{\Omega}$ части асимптотики (слагаемые вида $R_m(\xi,\eta;j_m)$ с $m > -2$). Замена интеграла от регуляризованной части по $\widetilde{\Omega}_{1,\varepsilon,\mu}$ на разность интегралов по $\widetilde{\Omega}$ и по $\widetilde{\Omega}_{2,\varepsilon,\mu}$ с последующим интегрированием вычтенной части асимптотики по $\widetilde{\Omega}_{1,\varepsilon,\mu}$ даст слагаемые вида $\varepsilon^m B_{k,l,1} + \varepsilon^m\mathcal{F}_3(\mu,\varepsilon)$.

Здесь

$$\widetilde{\Omega}_{1,\varepsilon,\mu}:=\{(\xi,\eta):\xi\in(0;\mu/(\varepsilon^4),\ |\eta|<\sqrt[4]{\xi}\},\quad \widetilde{\Omega}_{2,\varepsilon,\mu}:=\{(\xi,\eta):\xi>\mu/(\varepsilon^4),\ |\eta|<\sqrt[4]{\xi}\}.$$

При этом слагаемые, зависящие только от μ или от μ и ε вместе, при нахождении асимптотики можно не учитывать (они взаимно уничтожаются).

Покажем, что

$$\int\limits_0^{\mu/(\varepsilon^4)} d\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} W_n(\xi,\eta)W_m(\xi,\eta)d\eta = D_{n,m}+\mathcal{F}(\varepsilon,\mu), \tag{5.2}$$

где $D_{n,m}$ — константа.

Обозначим через $W^{\mathrm{sing}}_{n,m}(\xi,\eta)$ неинтегрируемые в $\widetilde{\Omega}$ особенности $W_n(\xi,\eta)W_m(\xi,\eta)$, а через $W^{\mathrm{reg}}_{n,m}(\xi,\eta)$ "регуляризованное" вычитанием этих особенностей рассматриваемое произведение, и

$$W^{\mathrm{sing}}_{n,m}(\xi):=\int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} W^{\mathrm{sing}}_{n,m}(\xi,\eta)\,d\eta,\quad W^{\mathrm{reg}}_{n,m}(\xi):=\int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} W^{\mathrm{reg}}_{n,m}(\xi,\eta)\,d\eta.$$

Отметим, что $W^{\mathrm{sing}}_{n,m}(\xi,\eta)$ есть линейная комбинация слагаемых вида

$$\xi^{(4+n+m-s-\sigma)/4}P_{n,m,s,\sigma}(\widetilde{\omega};(-1)^{4+n+m-s-\sigma}),\quad 4+n+m-s-\sigma<-5, \tag{5.3}$$

поэтому асимптотика на $+\infty$ у функции $W^{\mathrm{reg}}_{n,m}(\xi)$ есть ряд по степеням $\xi^{-1/4}$, начинающийся с $\xi^{-3/2}$.

Тогда

$$\int\limits_0^{\mu/\varepsilon^4} d\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} W_n(\xi,\eta)W_m(\xi,\eta)d\eta = \int\limits_0^1 W^{\mathrm{reg}}_{n,m}(\xi)\,d\xi+\int\limits_1^{+\infty} W^{\mathrm{reg}}_{n,m}(\xi)\,d\,\xi+\int\limits_{+\infty}^{\mu/\varepsilon^4} W^{\mathrm{reg}}_{n,m}(\xi)\,d\,\xi+$$

$$+\int\limits_1^{\mu/\varepsilon^4} W^{\mathrm{sing}}_{n,m}(\xi)\,d\,\xi = Dn,m,1+Dn,m,2+\int\limits_{+\infty}^{\mu/\varepsilon^4} W^{\mathrm{reg}}_{n,m}(\xi)\,d\,\xi+\int\limits_1^{\mu/\varepsilon^4} W^{\mathrm{sing}}_{n,m}(\xi)\,d\,\xi.$$

Но оставшийся интеграл от $W^{\mathrm{reg}}_{n,m}$ есть асимптотический ряд по степеням $(\varepsilon^4/\mu)^{-l}$ ($l>5$), поэтому он имеет вид $\mathcal{F}_1(\varepsilon,\mu)$. Интеграл от $W^{\mathrm{sing}}_{n,m}(\xi)$ есть конечная сумма слагаемых вида (см. (5.3))

$$\int\limits_1^{\mu/\varepsilon^4} d\,\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} \xi^{(4+n+m-s-\sigma)/4}P_{n,m,s,\sigma}(\widetilde{\omega};(-1)^{4+n+m-s-\sigma})\,d\,\eta = \int\limits_1^{\mu/\varepsilon^4} \xi^{(5+n+m-s-\sigma)/4}\gamma_{n,m,s,\sigma}\,d\,\xi.$$

При этом, если $4+n+m-s-\sigma=-4$, то $(n+m-s-\sigma)$ — нечетно и, значит, $\gamma_{n,m,s,\sigma}=0$. Окончательно последний интеграл есть конечная сумма слагаемых $(\mu/\varepsilon^4)^{(9+n+m-s-\sigma)/4}$ и 1, т.е. равен $D_{n,m,2}+\mathcal{F}_2(\varepsilon,\mu)$, и, тем самым, справедливо равенство (5.2).

Отметим, что

$$\left(\sum_{n=0}^{\infty}a_n\right)\left(\sum_{n=0}^{\infty}b_n\right)=a_0b_0+\left(\sum_{n=1}^{\infty}(a_0b_n+a_nb_0+c_{n,[0;n-1]}\right),$$

где индекс $[0;n-1]$ говорит о том, что данное слагаемое зависит только от слагаемых перемножаемых рядов,

имеющих номера меньше n. Тем самым, асимптотическое равенство (5.1) имеет вид

$$
\begin{aligned}
1 \stackrel{\text{as}}{=} \varepsilon^7\lambda_0^2 &\left[\int\limits_0^{\mu/\varepsilon^4} d\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} W_0^2(\xi,\eta)d\eta\right]_\varepsilon + \\
+\varepsilon^7\sum_{k=1}^{\infty} &\left[\varepsilon^k \int\limits_0^{\mu/\varepsilon^4} d\xi \int\limits_{-\sqrt[4]{\xi}}^{\sqrt[4]{\xi}} \Big(2\lambda_0^2 W_0(\xi,\eta)W_k(\xi,\eta)+\right. \\
+\ 2\lambda_0\lambda_k V_0^2(\xi,\eta) &\left.+ G_{k,[0;k-1],1}(\xi,\eta)\Big)\,d\eta\right]_\varepsilon + \lambda_0^2\left[\int\limits_\mu^{x_2} dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} p_0^2(x,y)dy\right]_\varepsilon + \\
+\sum_{k=1}^{\infty}\varepsilon^k &\left[\int\limits_\mu^{x_2} dx \int\limits_{-\sqrt[4]{x}}^{\sqrt[4]{x}} \Big(2\lambda_0^2 p_0(x,y)p_k(x,y) + \right. \\
&\left.+ 2\lambda_0\lambda_k p_0^2(x,y) + G_{k,[0;k-1],2}(x,y)\Big)\,dy\right]_\varepsilon .
\end{aligned}
\tag{5.4}
$$

Здесь через $[\cdot]_\varepsilon$ обозначены слагаемые, полученные указанной процедурой получения асимптотики методом вспомогательного параметра, зависящие только от ε. Отметим, что в данном случае это будут только константы.

Из последнего асимптотического равенства видно, что λ_k впервые появляется во внешнем разложении при ε^k, а во внутреннем — только при ε^{k+7}.

Как и в [10], удобно при $k>0$ представить z_k и p_k в виде $z_k = \widetilde{z}_k + \lambda_k\widetilde{z}$, $p_k = \widetilde{p}_k + \lambda_k\widetilde{p}$, где $\widetilde{z}_k$, $\widetilde{p}_k$ — решение задачи

$$
\begin{aligned}
&\mathcal{L}_0\widetilde{z}_k + \lambda_0\widetilde{p}_k = \widetilde{F}_{1,k,[0;k-1]}, \quad \mathcal{L}_0\widetilde{p}_k - \widetilde{z}_k = \widetilde{F}_{2,k,[0;k-1]}, \quad k\geqslant 1,\\
&\widetilde{z}_k|_\Gamma = 0 = \widetilde{p}_k|_\Gamma, \qquad\qquad\qquad\qquad\qquad\quad k\geqslant 0,
\end{aligned}
\tag{5.5}
$$

где функции $\widetilde{F}_{1,k,[0;k-1]}$ и $\widetilde{F}_{2,k,[0;k-1]}$ зависят только от предыдущих z_l, p_l и λ_l, $0\leqslant l<k$, а $\widetilde{z}$, $\widetilde{p}$ — решение задачи

$$
\begin{aligned}
&\mathcal{L}_0\widetilde{z} + \lambda_0\widetilde{p} + p_0 = 0, \quad \mathcal{L}_0\widetilde{p} - \widetilde{z} = 0, \quad k\geqslant 1,\\
&\widetilde{z}_k|_\Gamma = 0 = \widetilde{p}_k|_\Gamma, \qquad\qquad\qquad\quad k\geqslant 0.
\end{aligned}
\tag{5.6}
$$

В силу леммы 3.1 и теоремы 3.1 при $(x+|y|)\to 0$

$$
\widetilde{z}(x,y) \stackrel{\text{as}}{=} \mathcal{R}^+_{4,\widetilde{z}}(x,y;1), \quad \widetilde{p}(x,y) \stackrel{\text{as}}{=} \mathcal{R}^+_{4,\widetilde{p}}(x,y;1),
\tag{5.7}
$$

и, тем самым, как z_0, так и p_0, интегрируемы с квадратом по Ω.

При таком представлении уравнения для нахождения (правильных) λ_k принимают вид

$$
2\lambda_k\lambda_0\Big(\|p_0\|^2 + \lambda_0\langle p_0,\widetilde{p}\rangle\Big) = \widetilde{h}_{k,[0;k-1]}, \quad k\in\mathbb{N},
\tag{5.8}
$$

где $\widetilde{h}_{k,[0;k-1]}$ — константы, уже известные к моменту решения этого уравнения.

В силу леммы 3 из [10] при выполнении условия (2.6) справедливо соотношение

$$
\|p_0\|^2 + \lambda_0(p_0,\widetilde{p}) > 0.
$$

Опишем процедуру одновременного построения рядов (3.1) и согласованных с ними рядов (4.4).

Алгоритм согласования:

S1. Возьмем λ_0 из (2.7), а остальные λ_k положим равными нулю. Построим ряды (3.1) и согласованные с ними ряды (4.4), соответствующие данным λ_k. При этом однозначно определятся z_0 и p_0.

S2. Из уравнения (5.8) для $k=1$, в котором $\widetilde{h}_{1,[0;0]}$ определяется только через z_0 и p_0, найдем λ_1 и соответствующие этому λ_1 функции $z_{1,1}$ и $p_{1,1}$. После чего получим новые наборы $\{z_{k,1}\}$ и $\{p_{k,1}\}$ ($k>1$). При этом уравнения для z_0 и p_0 (как и сами эти функции) не изменятся.

Новые функции $z_{1,1}$ и $p_{1,1}$ будут отличаться от старых $z_{1,0}$ и $p_{1,0}$ в силу (5.7) на слагаемые вида $\mathcal{R}_4^+(x,y;1)$ При этом изменится и асимптотика при $(x+|y|)\to 0$ функций $\{z_{k,1}\}$ и $\{p_{k,1}\}$ ($k\geqslant 1$) и в силу условий согласования (4.5) изменится асимптотика при $(\xi^2+\eta^2)\to+\infty$ и некоторых функций из внутреннего разложения.

Рассмотрим подробнее ситуацию изменения асимптотик коэффициентов внутреннего разложения при изменении старых $\varepsilon^k z_k$ и $\varepsilon^k p_k$ на величину $\varepsilon^k\mathcal{R}_4^+(x,y;1)$. Все предыдущие функции z_l и p_l ($l<k$) останутся неизменными, поскольку уравнения, их определяющие, не изменятся. Уравнения для V_l и W_l ($l<k$) тоже не изменятся. Покажем, что у этих функций V_l и W_l не изменятся и асимптотики при $(\xi^2+\eta^2)\to+\infty$, продиктованные условиями согласования (4.5).

Поскольку при $(x+|y|)\to 0$ асимптотика добавленного слагаемого имеет вид $\mathcal{R}_4^+(x,y;1)$, то при переходе к ξ и η ряд $\varepsilon^k\mathcal{R}_4^+(x,y;1)$ даст слагаемые вида $\varepsilon^{k+4+s}\xi^{(k+4+s)/4}P_s$. Тем самым, изменятся асимптотики (4.5) лишь у функций V_n и W_n при $n\geqslant 4+k+s\geqslant 4+k$. Поэтому функции V_l и W_l, $l<n+2$, не изменяются.

S3. Последовательно находя λ_k и корректируя уравнения и условия поведения при $(x+|y|)\to 0$ и $(\xi^2+\eta^2)\to$ $\to+\infty$, получаем окончательно коэффициенты всех рядов из (3.1) и (4.4). При этом внутреннее разложение будет согласовано с внешним, и будет выполняться (5.4).

Тем самым, и в этом случае справедлива итоговая теорема.

Теорема 4. *Пусть выполнены условия* (1.4), (3.3) *и* (2.6), *а* (3.1) *и* (4.4) *— ряды, построенные по алгоритму согласования. Тогда внешнее и внутреннее разложения есть асимптотические разложение решения задачи* (1.8) *с дополнительным условием* $\lambda_\varepsilon\|p_\varepsilon\|=1$ *при* $\varepsilon\to+0$ *в областях* (4.8) *соответственно.*

СПИСОК ЛИТЕРАТУРЫ

1. *Лионс Ж.-Л.* Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972, 414 с.
2. *Тихонов А.Н., Самарский А.А.* Уравнения математической физики. М.: Наука, 1972, 736 с.
3. *Леликова Е.Ф.* Об асимптотике решения эллиптического уравнения второго порядка с малым параметром при одной из старших производных // Тр. ИММ УрО РАН. 2003. Т. 9. № 1. С. 107–120.
4. *Ильин А.М., Леликова Е.Ф.* Об асимптотике решения одного уравнения с малым параметром // Алгебра и анализ. 2010. Т. 22. № 6. С. 109–126.
5. *Casas Eduardo.* A review on sparse solutions in optimal control of partial differential equations // SeMA J. 2017. V. 74. P. 319–344.
6. *Lou H., Yong J.* Second-order necessary conditions for optimal control of semilinear elliptic equations with leading term containing controls // Math. Control Relat. Field. 2018. V. 8. № 1. P. 57–88.
7. *Betz Livia M.* Second-order sufficient optimality conditions for optimal control of nonsmooth, semilinear parabolic equations // SIAM J. Control Optim. 2019. V. 57. № 6. P. 4033–4062.
8. *Данилин А.Р.* Аппроксимация сингулярно возмущенной эллиптической задачи оптимального управления // Матем. сб. 2000. Т. 191. №10. С. 3–12.
9. *Данилин А.Р.* Асимптотика решений системы сингулярных эллиптических уравнений в прямоугольнике // Матем. сб. 2003. Т. 194. №1. С. 31–60.
10. *Данилин А.Р.* Асимптотика решения сингулярной задачи оптимального распределённого управления с существенными ограничениями в выпуклой области // Дифференц. ур-ния. 2020. Т. 56. № 2. С. 256–268.
11. *Данилин А.Р.* Асимптотика решения задачи оптимального распределенного управления в выпуклой области с малым параметром при одной из старших производных // Уфимский матем. ж. 2023. Т. 15. № 2. С. 42–54.
12. *Ильин А.М.* Согласование асимтотических разложений решений краевых задач. М.: Наука, 1989. 336 с.
13. *Ильин А.М., Данилин А.Р.* Асимптотические методы в анализе. М.: Наука, 1989. 248 с.
14. *Лионс Ж.-Л., Мадженес Э.* Неоднородные граничные задачи и их приложения. М.: Мир, 1971. 371 с.
15. *Данилин А.Р., Зорин А.П.* Асимптотика решения задачи оптимального граничного управления // Тр. ИММ УрО РАН. 2009. Т. 15. № 4. С. 95–107.

ASYMPTOTICS OF THE SOLUTION TO A BISINGULAR PROBLEM OF OPTIMAL DISTRIBUTED CONTROL IN A CONVEX DOMAIN WITH A SMALL PARAMETER IN ONE OF THE HIGHER DERIVATIVES

A. R. Danilin*

N.N. Krasovsky Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences, S. Kovalevskaya St. 16, Yekaterinburg, 620990, Russia

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

Received 27 November, 2023
Revised 13 January, 2024
Accepted 06 February, 2024

Abstract. The paper considers the problem of optimal distributed control in a strictly convex planar domain with a smooth boundary and a small parameter in one of the higher derivatives of the elliptic operator. In this problem, a zero Dirichlet boundary condition is imposed, and the control enters additively into the inhomogeneity. The set of admissible controls is the unit ball in the corresponding space of square-integrable functions. The solutions to the resulting boundary value problems are treated in the generalized sense as elements of a certain Hilbert space. The optimality criterion is the sum of the square of the norm of the state deviation from a given state and the square of the norm of the control, with a weighting coefficient. This structure of the optimality criterion allows either the first or the second term to be emphasized, depending on the need. In the first case, achieving the desired state is prioritized, while in the second case, minimizing resource costs becomes more important. The asymptotics of the problem are studied in detail, arising from a second-order differential operator with a small coefficient in one of the higher derivatives, to which a zero-order differential operator is added.

Keywords: singular problems, optimal control, boundary value problems for systems of partial differential equations, asymptotic expansions.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 745–765

УДК 517.977.56

О СУЩЕСТВОВАНИИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ПОЛУЛИНЕЙНЫМ ЭВОЛЮЦИОННЫМ УРАВНЕНИЕМ С НЕОГРАНИЧЕННЫМ ОПЕРАТОРОМ

© 2024 г. А.В. Чернов[1,*]

[1]603950 Нижний Новгород, пр-т Гагарина, 23, Нижегородский государственный университет им. Н.И. Лобачевского, Россия

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

Поступила в редакцию 28.11.2023 г.

Переработанный вариант 16.01.2024 г.

Принята к публикации 31.01.2024 г.

Исследуется задача оптимального управления абстрактным полулинейным дифференциальным уравнением первого порядка по времени в гильбертовом пространстве, с неограниченным оператором и линейно входящим в правую часть управлением. Целевой функционал предполагается аддитивно разделенным по состоянию и управлению, при зависимости от состояния достаточно общего вида. Для этой задачи доказывается теорема существования оптимального управления, а также устанавливаются свойства множества оптимальных управлений. В связи с нелинейностью изучаемого уравнения, развиваются ранее полученные автором результаты о тотальном сохранении однозначной глобальной разрешимости (о тотально глобальной разрешимости) и об оценке решений для подобных уравнений. Указанная оценка оказывается существенной при проведении исследования. В качестве примеров рассматриваются нелинейное уравнение теплопроводности и нелинейное волновое уравнение. Библ. 22.

Ключевые слова: полулинейное эволюционное уравнение с неограниченным оператором в гильбертовом пространстве, существование оптимального управления, нелинейное уравнение теплопроводности, нелинейное волновое уравнение.

DOI: 10.31857/S0044466924050055, **EDN**: YDLDOA

ВВЕДЕНИЕ

Краткий обзор основных подходов к исследованию проблемы существования оптимального управления в распределенных задачах оптимизации автором уже был представлен в его предыдущей работе по данной теме [1]. Чтобы не повторяться, кратко напомним лишь основные моменты, повлиявшие на мотивацию автора. Достаточно часто рассматривается ситуация, когда динамика управляемой системы допускает описание в виде операторного дифференциального уравнения, в частности, первого порядка по времени:

$$\frac{d\varphi}{dt} + (G\varphi)(t) = z(t), \quad t \in (0;T]; \quad \varphi(0) = a, \tag{1}$$

где на оператор G накладываются специальные условия (типа параболичности, монотонности, коэрцитивности, липшицевости, либо, скажем, линейности и ограниченности, и т.д. и т.п.), обеспечивающие разрешимость этого уравнения для любых z из заданного пространства, см., например, [2], [3, гл. 3], [4, chapter 5], [5]. При этом либо сама правая часть z трактуется как управление, либо она заменяется на выражение, линейно зависящее от управления (но не зависящее от переменной состояния). При замене z на управляемую нелинейность вида $f(.,\varphi,u)$ (не удовлетворяющую специальным требованиям типа тех, которые были отнесены выше к оператору G), с управлением $u \in U$ разрешимость управляемой системы $\forall\, u \in U$ становится негарантированной. В этом случае, как правило, переходят к рассмотрению пар «управление–состояние» (см., например, [6, 7]), а управляемую систему рассматривают как ограничение особого типа. Эта схема предполагает лишь целевые функционалы того или иного специального вида (как правило, суммы q-степеней L_q-норм), см., например, [3, 7] и непустоту множества допустимых пар; при этом обычно доказывается лишь сам по себе факт существования оптимального управления (о свойствах множества оптимальных управлений речь не идет) [7, гл. 1, § 7, § 8; гл. 4, § 1], [3, гл. 3, § 15; гл. 4, § 10], [4, §§ 4.4, 5.3], [5]. Другой стандартный путь — наложение условий (глобальной) липшицевости по переменной состояния на нелинейность $f(.,\varphi,u)$ и/или ее равномерной ограниченности,

см., например, [2]. В работе [1] было предложено при исследовании проблемы существования оптимального управления использовать достаточные условия тотально глобальной разрешимости управляемой эволюционной системы. *Тотальное сохранение глобальной разрешимости* (ТСГР), или *тотально глобальная разрешимость* (ТГР) — это свойство управляемой системы сохранять глобальную[1)] разрешимость для всех допустимых управлений. При наличии теорем о ТСГР и единственности глобального решения и выполнении их условий функционалы оптимизационной задачи выступают как функции, зависящие только от управлений, что позволяет опираться на соответствующие теоремы функционального анализа или их обобщения и тем самым, упрощать исследование и/или получать достаточно сильные результаты.

Собственно, в [1] рассматривалось операторное дифференциальное уравнение в банаховом пространстве псевдопараболического типа, допускающее сведение к уравнению

$$\frac{d\varphi}{dt} + (G\varphi)(t) = \Phi\big(t, \varphi(t)\big) + b\big(t, u(t), \varphi(t)\big), \quad t \in (0;T]; \quad \varphi(0) = a, \tag{2}$$

с ограниченным линейным оператором G и функцией $b(t,u,\varphi)$, билинейной по (u,φ), при естественных предположениях относительно функции $\Phi(t,\varphi)$ (локальная липшицевость и локальная оценка роста по переменной состояния φ). В данной статье результаты [1] распространяются на случай неограниченного оператора G (но на этот раз действующего в гильбертовом пространстве). При этом слагаемое $\Phi\big(t,\varphi(t)\big)$ рассматривается в более общей форме $\Phi\big(t,\varphi(.)\big)$ как оператор от φ, вольтерровый в том смысле, что значения $\Phi\big(t,\varphi(.)\big)$ зависят лишь от значений $\varphi(s)$ при $s \in [0;t]$ и не зависят от значений $\varphi(s)$ при $s \in (t;T]$ для п.в. $t \in [0;T]$. Слагаемое $b(t,u,\varphi)$ допускает нелинейный характер зависимости от φ. Так же, как и в [1], доказывается теорема существования оптимального управления и устанавливаются свойства множества оптимальных управлений U_*. А именно, при сделанных предположениях доказывается, что U_* непусто и слабо компактно в пространстве управлений, и что всякая минимизирующая последовательность слабо сходится к этому множеству. В качестве примеров рассматриваются нелинейное уравнение теплопроводности и нелинейное волновое уравнение. Отметим наконец, что для понимания доказательств основных утверждений данной статьи необходимо иметь под рукой работу [1].

1. ПРЕДВАРИТЕЛЬНЫЕ ПОСТРОЕНИЯ И СОГЛАШЕНИЯ

Пусть X — вещественное гильбертово пространство со скалярным произведением $[.,.]_X$, $G : X \to X$ — инфинитезимальный генератор (производящий оператор) сильно непрерывной полугруппы $S(t)$, $t \in [0;T]$, с областью определения $D(G) \subset X$, $z \in Z = L_2\big(0,T;X\big)$, $x_0 \in X$. Следуя [8, § 4.8], рассмотрим задачу Коши для эволюционного уравнения (абстрактного дифференциального уравнения в пространстве X):

$$x'(t) = Gx(t) + z(t), \quad t \in [0;T]; \quad x(0) = x_0. \tag{3}$$

Справедливы следующие утверждения.

Лемма 1.1. (см. [8, теорема 4.8.3]). *Для любых $z \in Z$, $x_0 \in X$ существует единственная функция $x : [0;T] \to X$ такая, что для всех $y \in D(G^*)$ функция $\big[x(t), y\big]_X$ абсолютно непрерывна на $[0;T]$,*

$$\frac{d}{dt}\big[x(t), y\big]_X = \big[x(t), G^*y\big]_X + \big[z(t), y\big]_X \quad \textit{п.в. } t \in [0;T],$$

$$\lim_{t\to 0}\big[x(t), y\big]_X = [x_0, y] \quad \forall\, y \in D(G^*).$$

Более того, справедлива формула

$$x(t) = S(t)x_0 + \int_0^t S(t-s)z(s)\,ds, \quad t \in [0;T]. \tag{4}$$

Лемма 1.2. (см. [8, следствие 4.8.1]). *Для любых $z \in Z$, $x_0 \in X$ существует единственная слабонепрерывная функция $x : [0;T] \to X$ такая, что для всех $y \in D(G^*)$ имеем*

$$\big[x(t), y\big]_X = [x_0, y]_X + \int_0^t \big[x(s), G^*y\big]_X\,ds + \int_0^t \big[z(s), y\big]_X\,ds,$$

[1)] Глобальную разрешимость эволюционной системы мы понимаем как разрешимость на произвольно фиксированном промежутке времени.

и более того, эта функция представляется формулой (4).

Замечание 1.1. В [8] интеграл от функции со значениями в пространстве X понимается в смысле Петтиса; $L_2(0,T;X)$ определяется в [8, гл. 3, п.3.5] и понимается как пространство слабо измеримых функций $z:[0;T]\to X$, для которых функция $\|z(t)\|_X^2$ интегрируема по Лебегу; в [8, гл. 3, п.3.5] (чтобы не перегружать изложение, как там сказано) выкладки проводятся для случая сепарабельного пространства X — в этом случае измеримость функции $\|z(t)\|_X^2$ заведомо гарантируется. Но можно понимать пространство $L_2(0,T;X)$ и в обычном смысле — как пространство измеримых по Бохнеру (а следовательно, сильно измеримых) функций $z:[0;T]\to X$, для которых функция $\|z(t)\|_X^2$ интегрируема по Лебегу, поскольку это просто сужение пространства по сравнению с его трактовкой в [8, глава 3, п.3.5]. Функция $z:[0;T]\to X$ сильно измерима тогда и только тогда, когда она слабо измерима и почти сепарабельнозначна [9, глава 3, §1, теорема 3.5.3, с. 86]. Для сепарабельного пространства X понятия слабой измеримости и измеримости по Бохнеру совпадают [10, гл. 4, теорема 1.4]. Отметим, кстати, что из интегрируемости по Бохнеру следует интегрируемость по Петтису, с совпадением интегралов в том и другом смысле, [9, с. 94]. Условимся понимать обозначение $L_2(0,T;X)$ и ему подобные в обычном смысле.

Напомним, см., например, [9, глава III, § 1, п.3.2, с.72], [11, с. 96], что функция $x:[0;T]\to X$ (для, вообще говоря, линейного нормированного пространства X) называется *слабо непрерывной* (иногда говорят *деминепрерывной*), если для любого $y\in X^*$ функция $y\big[x(t)\big]$ непрерывна на $[0;T]$. Множество всех слабо непрерывных функций $x:[0;T]\to X$ будем обозначать $\mathbb{C}_w\big(0,T;X\big)$. Для дальнейшего важно, что норма $\big\|x(t)\big\|_X$ всякой функции $x\in\mathbb{C}_w\big(0,T;X\big)$ ограничена на $[0;T]$. С другой стороны, см., например, [10, гл. IV, теорема 1.9, с. 154], всякая функция $x\in\mathbb{C}_w\big(0,T;X\big)$ интегрируема по Бохнеру, следовательно, измерима по Бохнеру. Стало быть, $\mathbb{C}_w\big(0,T;X\big)\subset L_\infty\big(0,T;X\big)$.

Функцию $x(t)$, существование и единственность которой в множестве $\mathbb{C}_w\big(0,T;X\big)$ утверждается в леммах 1.1, 1.2, будем называть *слабым решением* задачи (3). Далее будем предполагать, что оператор G удовлетворяет следующему условию.

Условие $\mathbf{G}_1$. Полугруппа $S(.)$ равномерно ограничена на промежутке $[0;T]$, т.е. $\big\|S(t)\big\|\leqslant M$ для всех $t\in[0;T]$.

Замечание 1.2. Для любой сильно непрерывной полугруппы $S(t)$ существуют константы $\omega\geqslant 0$, $R\geqslant 1$ такие, что выполняется условие на порядок роста: $\big\|S(t)\big\|\leqslant Re^{\omega t}\ \forall\, t\geqslant 0$, см. [12, § 1.2, theorem 2.2]. Поэтому условие $\mathbf{G}_1$ заведомо выполнено на любом конечном промежутке $[0;T]$ с константой $M=M(T)=Re^{\omega T}$.

Лемма 1.3. *Пусть* $A[z](t)=\int\limits_0^t S(t-s)z(s)\,ds$, $t\in[0;T]$, *— слабое решение задачи* (3) *при* $x_0=0$, $z\in Z$. *Тогда для п.в.* $t\in[0;T]$ *справедлива оценка* $\big\|A[z](t)\big\|_X\leqslant M\int\limits_0^t\|z(s)\|_X\,ds$.

Доказательство. По свойствам интеграла Петтиса [8, с. 176] и в силу условия $\mathbf{G}_1$

$$\big\|A[z](t)\big\|_X\leqslant\int\limits_0^t\big\|S(t-s)\,z(s)\big\|_X\,ds\leqslant M\int\limits_0^t\|z(s)\|_X\,ds.$$

Лемма доказана.

Будем предполагать заданными числа $T>0$, $p\in[2;+\infty)$. Чтобы определить управляемую систему, сделаем следующие предположения.

Условие $\mathbf{F}_1$. Для всех $z\in E(T)=L_\infty(0,T;X)$ отображение $[0;T]\ni t\to\Phi\big(t,z(.)\big)$ принадлежит классу $L_2\big(0,T;X\big)$ и является вольтерровым в том смысле, что значения $\Phi\big(t,z(.)\big)$ зависят лишь от значений $z(s)$ при $s\in[0;t]$ для п.в. $t\in[0;T]$.

Условие $\mathbf{F}_2$. Существует функция $\mathcal{N}=\mathcal{N}(t,r):[0;T]\times\mathbb{R}_+\to\mathbb{R}_+$, суммируемая с квадратом по $t\in[0;T]$ и неубывающая по $r\in\mathbb{R}_+$ такая, что для всех $\xi,\eta\in E=E(T)$, $\|\xi\|_E$, $\|\eta\|_E\leqslant r$, п.в. $t\in[0;T]$ имеем

$$\big\|\Phi(t,\xi)-\Phi(t,\eta)\big\|_X\leqslant\mathcal{N}(t,r)\,\big\|\xi-\eta\big\|_{L_\infty(0,t;X)}.$$

Условие $\mathbf{F}_3$. Существует функция $\mathcal{N}_1=\mathcal{N}_1(t,r):[0;T]\times\mathbb{R}_+\to\mathbb{R}_+$, неубывающая по r и суммируемая по Лебегу с квадратом по t такая, что $\Big\|\Phi(t,\xi)\Big\|_X\leqslant\mathcal{N}_1(t,r)$ для всех $r>0$, $\xi\in E(t)=L_\infty(0,t;X)$, $\|\xi\|_{E(t)}\leqslant r$, п.в. $t\in[0;T]$.

Замечание 1.3. По поводу условия $\mathbf{F}_2$ заметим, что функция $\psi(t) = \big\|\xi - \eta\big\|_{L_\infty(0,t;X)}$ является неубывающей и ограниченной. Согласно [13, глава VIII, § 1, теорема 1, с. 192], множество точек разрыва такой функции не более, чем счетно. Иначе говоря, функция $\psi(t)$ непрерывна почти всюду. Стало быть, она интегрируема по Риману [13, глава V, § 4, теорема 2, с. 125], а тем самым, и по Лебегу [13, глава V, § 4, теорема 3, с. 125], и во всяком случае измерима. А в силу ограниченности, $\psi \in L_\infty[0;T]$.

В качестве множества допустимых управлений далее будет выступать выпуклое, замкнутое и ограниченное множество U в пространстве $L_p(0,T;Y)$, где Y — сепарабельное рефлексивное банахово пространство, $p \geqslant 2$. Пусть, наконец, задана функция $b(t,v,x) : [0;T] \times Y \times X \to X$, измеримая по $t \in [0;T]$, линейная по $v \in Y$ и удовлетворяющая следующим условиям.

Условие $\mathbf{B}_1$. При каждом $u \in U$ оператор суперпозиции $b\big(.,u(.),x(.)\big)$ осуществляет отображение $L_\infty(0,T;X) \to L_2(0,T;X)$.

Условие $\mathbf{B}_2$. Существует функция $\mathcal{N}_2(t,r) : [0;T] \times \mathbb{R}_+ \to \mathbb{R}_+$, неубывающая по r и суммируемая по Лебегу со степенью $\widetilde{p} = \dfrac{2p}{p-2}$, такая, что при при п.в. $t \in [0;T]$, $x \in X$, $\|x\|_X \leqslant r$, выполняется оценка

$$\big\|b(t,v,x)\big\|_X \leqslant \mathcal{N}_2(t,r)\,\|v\|_Y \quad \forall\, v \in Y.$$

Условие $\mathbf{B}_3$. Существует функция $\mathcal{N}_3 = \mathcal{N}_3(t,r) : [0;T] \times \mathbb{R}_+ \to \mathbb{R}_+$, суммируемая с квадратом по $t \in [0;T]$ и неубывающая по $r \in \mathbb{R}_+$ такая, что для всех $u \in U$, $\xi, \eta \in E = E(T)$, $\|\xi\|_E$, $\|\eta\|_E \leqslant r$, п.в. $t \in [0;T]$ имеем

$$\big\|b\big(t,u(t),\xi(t)\big) - b\big(t,u(t),\eta(t)\big)\big\|_X \leqslant \mathcal{N}_3(t,r)\,\big\|\xi - \eta\big\|_{L_\infty(0,t;X)}.$$

Для $u \in U$ будем рассматривать управляемую задачу вида

$$\varphi'(t) = G\varphi(t) + \Phi\big(t,\varphi(.)\big) + b\big(t,u(t),\varphi(t)\big),\ t \in [0;T];\ \varphi(0) = \varphi_0. \tag{5}$$

Слабое решение задачи (5) на $[0;T]$ понимаем как решение операторного уравнения типа Гаммерштейна:

$$\varphi(t) = \theta(t) + A\Big[\Phi\big(.,\varphi(.)\big) + b\big(.,u(.),\varphi(.)\big)\Big](t), \quad t \in [0;T]; \quad \varphi \in E, \tag{6}$$

при $\theta(t) = S(t)\varphi_0$, $E = E(T) = L_\infty(0,T;X)$, $\theta \in \mathbb{C}_w\big(0,T;X\big) \subset E$. Ясно, что всякое решение $\varphi \in E$ в соответствии со свойствами оператора правой части принадлежит также и пространству $\mathbb{C}_w\big(0,T;X\big)$.

2. ПОСТАНОВКА ОПТИМИЗАЦИОННОЙ ЗАДАЧИ И ФОРМУЛИРОВКА ПРЕДВАРИТЕЛЬНЫХ РЕЗУЛЬТАТОВ

Пусть $\gamma(t) = \sup\limits_{u \in U}\big\|u(t)\big\|_Y$, $\widetilde{\mathcal{N}}_1(t,r) = M\big\{\mathcal{N}_1(t,r) + \mathcal{N}_2(t,r)\gamma(t)\big\}$, см. условия $\mathbf{F}_3$, $\mathbf{B}_2$; $\nu(t) \geqslant \big\|S(t)x_0\big\|_{L_\infty(0,t;X)}$.

Теорема 2.1. *Пусть выполнены сделанные выше предположения,* $\gamma(.) \in L_p[0;T]$, $\nu(.) \in L_\infty[0;T]$, *и кроме того, функции* $\mathcal{N}_1(t,r)$, $\mathcal{N}_2(t,r)$ *локально липшицевы по* $r \in \mathbb{R}_+$. *Тогда справедливы следующие утверждения:*

а) *существует* $T_0 \in (0;+\infty]$ *такое, что при всех* $T \in (0;T_0)$ *задача Коши*

$$\frac{d\beta}{dt} = \widetilde{\mathcal{N}}_1\big(t,\nu(t) + \beta(t)\big), \quad t \in (0;T]; \quad \beta(0) = 0, \tag{7}$$

имеет абсолютно непрерывное решение $\beta \in \mathbf{AC}[0;T]$, *понимаемое в смысле п.в.;*

б) *для любых* $T \in (0;T_0)$, $u \in U$ *задача* (5) *имеет решение* $\varphi[u] \in \mathbf{C}_w(0,T;X)$, *удовлетворяющее оценке* $\big\|\varphi[u](t)\big\|_X \leqslant \big\|S(t)x_0\big\|_X + \beta(t)$ *для п.в.* $t \in [0;T]$, *и это решение единственно.*

Доказательство теоремы 2.1 см. в разд. 3.

Теорема 2.1 гарантирует, что управляемая задача (5) разрешима для всех допустимых управлений на *одном и том же* отрезке $[0;T]$, т.е. на данном отрезке имеет смысл постановка задачи оптимального управления. В этом и есть значение теоремы 2.1. При отсутствии теоремы 2.1 горизонт существования решения будет зависеть от выбора конкретного управления, и будет непонятно, как выбирать общий для всех управлений отрезок существования решения $[0;T]$.

Замечание 2.1. Если просто постулировать разрешимость задачи (7) для любого $T \in (0;T_0)$, то требование локальной липшицевости функций $\mathcal{N}_1(t,r)$, $\mathcal{N}_2(t,r)$ по $r \in \mathbb{R}_+$ в условиях теоремы можно опустить (для корректности суперпозиции достаточно непрерывности). Указанную разрешимость можно устанавливать различными способами. Подробные пояснения по этому вопросу см. в [1].

Далее везде будем считать условия теоремы 2.1 выполненными. Поставим задачу оптимизации. Пусть задан непрерывный функционал $I_0 : L_\infty(0,T;X) \to \mathbb{R}$, ограниченный на ограниченных множествах, $J_0[u] = I_0\big(\varphi[u]\big)$, $u \in U$. Для произвольно заданного $\alpha \geqslant 0$ будем исследовать задачу

$$J_\alpha[u] = J_0[u] + \frac{1}{2}\alpha\|u\|^2_{L_p(0,T;Y)} \to \min_{u\in U}.$$

Пусть $J^*_\alpha = \inf\limits_{u\in U} J_\alpha[u]$, $U_* = \big\{u \in U :\ J_\alpha[u] = J^*_\alpha\big\}$.

Теорема 2.2. *Множество U слабо компактно в $L_p(0,T;Y)$.*

Доказательство теоремы 2.2 см. в [1, теорема 1.2].

Теорема 2.3. *Функционал $J_0[u]$, а следовательно, и $J_\alpha[u]$, ограничен на множестве U.*

Доказательство теоремы 2.3 вытекает непосредственно из теоремы 2.1 и ограниченности функционала I_0 на ограниченных множествах в пространстве $L_\infty(0,T;X)$.

Вопрос существования оптимального управления в поставленной задаче оптимизации будем исследовать отдельно для двух основных различных случаев в разд. 4 и 6. Там же будут сформулированы и доказаны соответствующие теоремы.

3. ДОКАЗАТЕЛЬСТВО ТОТАЛЬНО ГЛОБАЛЬНОЙ РАЗРЕШИМОСТИ

Общие замечания по вопросу тотально глобальной разрешимости управляемых операторных уравнений (и распределенных систем, в частности) см. в работе [1]. Следующее утверждение известно как лемма Гронуолла.

Лемма 3.1. *Пусть на $[0;T]$ заданы: $f \in L^+_\infty[0;T]$ и $g \geqslant 0$ — вещественная неубывающая функция. Если имеет место неравенство*

$$f(t) \leqslant g(t) + c\int\limits_0^t f(s)\,ds \quad \textit{для п.в. } t \in [0;T], \quad c \geqslant 0,$$

то $f(t) \leqslant e^{ct}g(t)$ для п.в. $t \in [0;T]$.

Обозначим для краткости $f(t,x,u) = \Phi(t,x)+b(t,u,x)$. Непосредственно из предположений $\mathbf{F}_1$–$\mathbf{F}_3$, $\mathbf{B}_1$–$\mathbf{B}_3$ получаем, что функция f удовлетворяет следующим условиям.

Условие $\mathbf{R}_1$. Для всех $u \in U$, $z \in E(T) = L_\infty(0,T;X)$ отображение $[0;T] \ni t \to f\big(t,z(.),u(t)\big)$ принадлежит классу $L_2\big(0,T;X\big)$ и является вольтерровым по z в том смысле, что значения $f\big(t,z(.),u(t)\big)$ зависят лишь от значений $z(s)$ при $s \in [0;t]$ для п.в. $t \in [0;T]$.

Условие $\mathbf{R}_2$. Существует функция $\widehat{\mathcal{N}}_1(t,r) = \mathcal{N}_1(t,r) + \mathcal{N}_2(t,r)\gamma(t) : [0;T]\times\mathbb{R}_+ \to \mathbb{R}_+$, неубывающая по r и суммируемая по Лебегу с квадратом по t такая, что $\Big\|f\big(t,\xi,u(t)\big)\Big\|_X \leqslant \widehat{\mathcal{N}}_1(t,r)\ \forall\, r>0, \xi \in E(t) = L_\infty(0,t;X)$, $\|\xi\|_{E(t)} \leqslant r$, $u \in U$, п.в. $t \in [0;T]$.

Условие $\mathbf{R}_3$. Существует функция $\widehat{\mathcal{N}}_2(t,r) = \mathcal{N}(t,r) + \mathcal{N}_3(t,r) : [0;T]\times\mathbb{R}_+ \to \mathbb{R}_+$, суммируемая с квадратом по $t \in [0;T]$ и неубывающая по $r \in \mathbb{R}_+$ такая, что $\forall\, u \in U$, $\xi,\eta \in E = E(T)$, $\|\xi\|_E$, $\|\eta\|_E \leqslant r$, п.в. $t \in [0;T]$ имеем $\big\|f\big(t,\xi,u(t)\big) - f\big(t,\eta,u(t)\big)\big\|_X \leqslant \widehat{\mathcal{N}}_2(t,r)\,\big\|\xi-\eta\big\|_{L_\infty(0,t;X)}$.

Лемма 3.2. *Каковы бы ни были $T>0$, $u \in U$, задача* (5) *не может иметь более одного решения.*

Доказательство. Предположим, от противного, что управлению $u \in U$ отвечают два решения φ_1, φ_2 задачи (5). Это означает, что

$$\varphi_i(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)f\big(s,\varphi_i,u(s)\big)\,ds, \quad i=1,2, \quad t\in[0;T].$$

Пусть $\psi = \varphi_1 - \varphi_2$, $r = \max\limits_{i=1,2}\|\varphi_i\|_E$, $\sigma = \big\|\widehat{\mathcal{N}}_2(.,r)\big\|_{L_2}$. Пользуясь условиями $\mathbf{G}_1$, $\mathbf{R}_3$, получаем

$$M^{-1}\big\|\psi(t)\big\|_X \leqslant \int\limits_0^t \Big\|f\big(s,\varphi_1,u(s)\big) - f\big(s,\varphi_2,u(s)\big)\Big\|_X\,ds \leqslant \int\limits_0^t \widehat{\mathcal{N}}_2(s,r)\,\big\|\psi\big\|_{L_\infty(0,s;X)}\,ds \leqslant \sigma\sqrt{\int\limits_0^t \|\psi\|^2_{L_\infty(0,s;X)}\,ds}.$$

Поскольку правая часть не убывает по $t \in [0;T]$, то ясно, что

$$\|\psi\|^2_{L_\infty(0,t;X)} \leqslant M^2\sigma^2 \int\limits_0^t \|\psi\|^2_{L_\infty(0,s;X)}\, ds.$$

По лемме 3.1 (Гронуолла), $\|\psi\|^2_{L_\infty(0,T;X)} = 0$, т.е. $\varphi_1 = \varphi_2$. Лемма доказана.

Лемма 3.3. *Пусть при* $\nu \equiv 0$ *задача Коши* (7) *имеет решение* $\beta(t)$*;* $\varphi_0 = 0$*. Тогда* $\forall\, u \in U$ *задача* (5) *имеет решение, удовлетворяющее оценке* $\big\|\varphi(t)\big\|_X \leqslant \beta(t)$*,* $t \in [0;T]$.

Доказательство. Пусть $\sigma = \|\beta\|_{\mathbf{C}[0;T]}$. Зафиксируем произвольно $u \in U$. Нам достаточно доказать разрешимость уравнения

$$\varphi(t) = \int\limits_0^t S(t-s) f\big(s, \varphi, u(s)\big)\, ds, \quad t \in [0;T]; \quad \varphi \in L_\infty(0,T;X). \tag{8}$$

Выберем разбиение $0 = t_0 < t_1 < \ldots < t_k = T$, $\max\limits_{i=\overline{1,k}} |t_i - t_{i-1}| < \delta$, где число $\delta > 0$ определяется исходя из условия (с учетом абсолютной непрерывности интеграла Лебега):

$$M \int\limits_h \widehat{\mathcal{N}}_2(s,\sigma)\, ds < \frac{1}{2} \quad \text{для измеримых } h \subset [0;T], \quad \operatorname{mes} h \leqslant \delta.$$

Докажем разрешимость уравнения (8) на $[0;t_1]$. Для $i = \overline{1,k}$ определим множества $W_i = \big\{\varphi \in E(t_i) = L_\infty(0,t_i;X) :\ \big\|\varphi(t)\big\|_X \leqslant \beta(t),\ t \in [0;t_i]\big\}$ и операторы $\mathcal{F}_i : W_i \to L_\infty(0,t_i;X)$ формулой

$$\mathcal{F}_i[\varphi](t) = \int\limits_0^t S(t-s) f\big(s, \varphi, u(s)\big)\, ds, \quad t \in [0;t_i].$$

Заметим, что функция β неубывающая, так как $\beta' \geqslant 0$. Стало быть,

$$\big\|\varphi(s)\big\|_X \leqslant \beta(s) \leqslant \beta(t) \quad \text{п.в. } s \in [0;t],$$

и таким образом, $\big\|\varphi\big\|_{L_\infty(0,t;X)} \leqslant \beta(t)$ п.в. $t \in [0;t_i]$, $\forall\, \varphi \in W_i$; $i = \overline{1,k}$. Поэтому для любого $\varphi \in W_1$, пользуясь условиями $\mathbf{R}_1$, $\mathbf{R}_2$, получаем

$$\big\|\mathcal{F}_1[\varphi](t)\big\|_X \leqslant M \int\limits_0^t \Big\|f\big(s, \varphi, u(s)\big)\Big\|_X\, ds \leqslant$$

$$\leqslant \int\limits_0^t M\widehat{\mathcal{N}}_1\big(s, \beta(s)\big)\, ds = \int\limits_0^t \widetilde{\mathcal{N}}_1\big(s, \beta(s)\big)\, ds = \beta(t) \quad \text{п.в. } t \in [0;t_1],$$

согласно определению функции $\beta(.)$ как решения задачи (7). Стало быть, $\mathcal{F}_1 : W_1 \to W_1$. Докажем сжимаемость оператора $\mathcal{F}_1$ на множестве W_1. Действуя совершенно аналогично тому, как это было при доказательстве леммы 3.2, для любых $\varphi_1, \varphi_2 \in W_1$ получаем оценку

$$\big\|\mathcal{F}_1[\varphi_1](t) - \mathcal{F}_1[\varphi_2](t)\big\|_X \leqslant M \int\limits_0^{t_1} \widehat{\mathcal{N}}_2(s,\sigma)\, ds\, \big\|\varphi_1 - \varphi_2\big\|_{L_\infty(0,t_1;X)}.$$

Используя условие на мелкость разбиения, заключаем, что

$$\big\|\mathcal{F}_1[\varphi_1] - \mathcal{F}_1[\varphi_2]\big\|_{L_\infty(0,t_1;X)} \leqslant \frac{1}{2}\big\|\varphi_1 - \varphi_2\big\|_{L_\infty(0,t_1;X)}.$$

Это означает, что оператор $\mathcal{F}_1$ является сжимающим на множестве W_1. И согласно принципу сжимающих отображений существует единственное $\varphi = \varphi_1 \in W_1$: $\varphi_1 = \mathcal{F}_1[\varphi_1]$, т.е. уравнение (8) разрешимо на $[0;t_1]$.

Действуя по индукции, предположим, что разрешимость уравнения (8) на $[0;t_{i-1}]$ уже доказана, и соответствующее решение $\varphi_{i-1} \in W_{i-1}$. Определим множество

$$W_{i-1,i} = \big\{\varphi \in E(t_i) :\ \varphi\,\big|_{[0;t_{i-1}]} = \varphi_{i-1};\ \big\|\varphi(t)\big\|_X \leqslant \beta(t),\ t \in (t_{i-1};t_i]\big\}.$$

Очевидно, что $W_{i-1,i} \subset W_i$. Поэтому совершенно аналогично тому, как это было сделано для $[0;t_1]$, устанавливается, что $\mathcal{F}_i : W_{i-1,i} \to W_i$. При этом очевидно, что $\mathcal{F}_i[\varphi]\,\big|_{[0;t_{i-1}]} = \mathcal{F}_{i-1}[\varphi_{i-1}] = \varphi_{i-1}$ для всех $\varphi \in W_{i-1,i}$. Стало быть, $\mathcal{F}_i : W_{i-1,i} \to W_{i-1,i}$. Докажем сжимаемость оператора $\mathcal{F}_i$ на $W_{i-1,i}$. Для произвольных $\psi_1, \psi_2 \in W_{i-1,i}$ имеем $\mathcal{F}_i[\psi_j](t) = \varphi_{i-1}(t)$ при $t \in [0;t_{i-1}]$, $j = 1,2$. Если же $t \in (t_{i-1};t_i]$, то

$$\big\|\mathcal{F}_i[\psi_1](t) - \mathcal{F}_i[\psi_2](t)\big\|_X \leqslant M \int\limits_0^{t_{i-1}} \Big\|f\big(s,\varphi_{i-1},u(s)\big) - f\big(s,\varphi_{i-1},u(s)\big)\Big\|_{E(t_{i-1})} +$$

$$+M \int\limits_{t_{i-1}}^{t} \Big\|f\big(s,\psi_1,u(s)\big) - f\big(s,\psi_2,u(s)\big)\Big\|_{E(t_i)} \leqslant$$

$$\leqslant M \int\limits_{t_{i-1}}^{t_i} \widehat{\mathcal{N}}_2(s,\sigma)\,ds\,\big\|\psi_1 - \psi_2\big\|_{L_\infty(0,t_i;X)} \leqslant \frac{1}{2}\,\big\|\psi_1 - \psi_2\big\|_{L_\infty(0,t_i;X)}.$$

Это означает, что оператор $\mathcal{F}_i$ является сжимающим на множестве $W_{i-1,i}$. И согласно принципу сжимающих отображений существует единственное $\varphi = \varphi_i \in W_{i-1,i} \subset W_i$: $\varphi_i = \mathcal{F}_i[\varphi_i]$, т.е. уравнение (8) разрешимо на $[0;t_i]$. По индукции делаем вывод, что уравнение (8) разрешимо на множестве W_k. Лемма доказана.

Доказательство теоремы 2.1. Что касается утверждения а), отметим, что при сделанных предположениях относительно функции $\widehat{\mathcal{N}}_1(t,r)$ задача (7) разрешима локально (это хорошо известный классический факт; в [14, теорема 5] он доказывается даже в более общей постановке, когда речь идет об абсолютно непрерывных функциях со значениями в банахом пространстве). Займемся далее доказательством утверждения б) при произвольно фиксированном $T \in (0;T_0)$. Нам требуется доказать разрешимость уравнения

$$\varphi(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)f\big(s,\varphi,u(s)\big)\,ds, \quad t \in [0;T]; \quad \varphi \in L_\infty(0,T;X). \tag{9}$$

Заменой $\psi(t) = \varphi(t) - S(t)\varphi_0$ это уравнение сводится к виду (8). Тогда по лемме 3.3 уравнение (9) имеет решение вида $\varphi(t) = S(t)\varphi_0 + \psi(t)$, где $\psi(t)$ — решение аналога (8). Учитывая представление

$$f\big(t,\varphi,u(t)\big) = f\big(t,S(.)\varphi_0 + \psi,u(t)\big) = \widetilde{f}(t,\psi,u(t)\big),$$

для функции $\widetilde{f}$ выполняется аналог условия $\mathbf{R}_2$ при замене функции $\widehat{\mathcal{N}}_1(t,r)$ на функцию $\widehat{\mathcal{N}}_1'(t,r) = \widehat{\mathcal{N}}_1\big(t,\nu(t)+ +r\big)$, $\nu(t) \geqslant \big\|S(.)\varphi_0\big\|_{L_\infty(0,t;X)}$. Это означает, что $\big\|\psi(t)\big\|_X \leqslant \beta(t)$, где $\beta(.)$ — решение задачи (7). И соответственно, $\big\|\varphi(t)\big\|_X \leqslant \big\|S(t)\varphi_0\big\|_X + \beta(t)$. Единственность решения уравнения (9) доказана в лемме 3.2. Теорема 2.1 доказана.

4. СЛУЧАЙ АППРОКСИМАТИВНОЙ КОМПАКТНОСТИ МНОЖЕСТВА РЕШЕНИЙ ЛИНЕЙНОЙ ЗАДАЧИ

В этом разделе будем считать выполненным следующее предположение.

Условие $\mathbf{W}_1$. Существуют $\widetilde{x}_0 \in X$, всюду плотное множество $\widetilde{Z} \subset Z = L_2(0,T;X)$ и пространство W, компактно вложенное в $L_q(0,T;H)$, $H \supset X$ непрерывно и плотно, $q \in (1,\infty)$, такие, что для всех $z \in \widetilde{Z}$ и $x_0 = \widetilde{x}_0$ задача (3) имеет решение $x = x[z] \in W$; и более того, $\big\|x[z]\big\|_W \leqslant \mathcal{N}_0(\sigma)$ для всех $z \in \widetilde{Z}$, $\|z\|_Z \leqslant \sigma$, где $\mathcal{N}_0 : \mathbb{R}_+ \to \mathbb{R}_+$ — некоторая неубывающая функция.

Условие $\mathbf{W}_1$ естественно назвать *условием аппроксимативной компактности* множества решений задачи (3). В той или иной конкретной ситуации зачастую оказывается, что уже существуют известные результаты (в том числе классические), обеспечивающие выполнение предположения $\mathbf{W}_1$. См. на этот счет примеры в разд. 7. Кроме того, в следующем разделе мы обсудим общие достаточные условия выполнения предположения $\mathbf{W}_1$. Сделаем также следующие предположения.

Условие $\mathbf{F}_0$. Оператор $\Phi\big(.,z(.)\big) : L_q(0,T;H) \to L_1(0,T;H)$ определен и непрерывен; и имеет место отображение
$\Phi\big(.,z(.)\big) : L_\infty(0,T;H) \cap L_q(0,T;X) \to L_2(0,T;X)$.

Условие $\mathbf{B}_0$. Существует неубывающая функция $\mathcal{N}_3 = \mathcal{N}_3(r) : \mathbb{R}_+ \to \mathbb{R}_+$, такая, что для всех $\xi, \eta \in \widetilde{E}(T) \cap$ $\cap L_q(0,T;X)$, $\widetilde{E}(T) = L_\infty(0,T;H)$, $\|\xi\|_{\widetilde{E}}$, $\|\eta\|_{\widetilde{E}} \leqslant r$, $u \in U$, имеем:

$$\big\|b(t,u,\xi) - b(t,u,\eta)\big\|_{L_1(0,T;H)} \leqslant \mathcal{N}_3(r)\,\big\|\xi - \eta\big\|_{L_q(0,T;H)}.$$

Условие $\mathbf{B}_2'$. Условие $\mathbf{B}_2$ выполняется в следующей усиленной форме. При каждом $u \in U$ оператор суперпозиции $b\big(.,u(.),x(.)\big)$ осуществляет отображение $L_\infty(0,T;H) \cap L_q(0,T;X) \to L_2(0,T;X)$. Более того, существует функция $\mathcal{N}_2'(t,r) : [0;T] \times \mathbb{R}_+ \to \mathbb{R}_+$, неубывающая по r и суммируемая по Лебегу со степенью $\widetilde{p} = \dfrac{2p}{p-2}$, такая, что при при п.в. $t \in [0;T]$, $x \in X$, $\|x\|_H \leqslant r$, выполняется оценка

$$\big\|b(t,v,x)\big\|_X \leqslant \mathcal{N}_2'(t,r)\,\|v\|_Y \quad \forall\, v \in Y.$$

Лемма 4.1. *Пусть $\{z_m\}$ — ограниченная последовательность в пространстве Z, $\{x_m\}$ — последовательность соответствующих решений задачи* (3) *при $z = z_m$. Тогда существует подпоследовательность $\{x_{m_k}\}$ сходящаяся в пространстве $L_q(0,T;H)$.*

Доказательство. Выберем последовательность $\varepsilon_m \to 0$ при $m \to \infty$. Поскольку множество $\widetilde{Z}$ всюду плотно в Z, для каждого $m \in \mathbb{N}$ найдется $\widetilde{z}_m \in \widetilde{Z}$ такое, что $\|z_m - \widetilde{z}_m\|_Z < \varepsilon_m$. Сравним соответствующие решения задачи (3) при $x_0 = \widetilde{x}_0$: $\overline{x}_m(t) = S(t)\widetilde{x}_0 + A[z_m](t)$, $\widetilde{x}_m(t) = S(t)\widetilde{x}_0 + A[\widetilde{z}_m](t)$, $A[z](t) = \int\limits_0^t S(t-s)z(s)\,ds$. В силу условия $\mathbf{G}_1$ имеем

$$\big\|\overline{x}_m(t) - \widetilde{x}_m(t)\big\|_X \leqslant M\|z_m - \widetilde{z}_m\|_{L_1} \leqslant M\sqrt{T}\|z_m - \widetilde{z}_m\|_Z \leqslant M\sqrt{T}\varepsilon_m \equiv \gamma_m;$$

$$\|\overline{x}_m - \widetilde{x}_m\|_{L_\infty(0,T;X)} \leqslant \gamma_m \quad \Rightarrow \quad \|\overline{x}_m - \widetilde{x}_m\|_{L_\infty(0,T;H)} \leqslant c\gamma_m \equiv \delta_m,$$

где c — константа непрерывного вложения $X \subset H$. Согласно условию $\mathbf{W}_1$, $\widetilde{x}_m \in W$ для всех $m \in \mathbb{N}$. По условию, $\|z_m\|_Z \leqslant \sigma_1\ \forall\, m \in \mathbb{N}$. Без ограничения общности рассуждений, считаем, что $\varepsilon_m \leqslant 1\ \forall\, m \in \mathbb{N}$. Тогда $\|\widetilde{z}_m\|_Z \leqslant$ $\leqslant 1 + \sigma_1 = \sigma\ \forall\, m \in \mathbb{N}$. Вновь по условию $\mathbf{W}_1$, $\|\widetilde{x}_m\|_W \leqslant \mathcal{N}_0(\sigma)$. Тогда, учитывая компактное вложение $W \subset$ $\subset L_q(0,T;H)$, существует сходящаяся подпоследовательность $\widetilde{x}_{m_k} \to \widetilde{x}$ в $L_q(0,T;H)$. Рассмотрим

$$\|\overline{x}_{m_k} - \widetilde{x}\|_{L_q} \leqslant \|\overline{x}_{m_k} - \widetilde{x}_{m_k}\|_{L_q} + \|\widetilde{x}_{m_k} - \widetilde{x}\|_{L_q} \leqslant T^{1/q}\delta_{m_k} + \|\widetilde{x}_{m_k} - \widetilde{x}\|_{L_q} \to 0.$$

Итак, $\overline{x}_{m_k} \to \widetilde{x}$ в $L_q(0,T;H)$. Для произвольного $x_0 \in X$ обозначим через

$$x_m(t) = S(t)x_0 + \int\limits_0^t S(t-s)z_m(s)\,ds = S(t)x_0 - S(t)\widetilde{x}_0 + \overline{x}_m(t)$$

решение $x[z_m]$ задачи (3). По доказанному, очевидно, что $x_{m_k} \to \overline{x}$ в $L_q(0,T;H)$, где $\overline{x}(t) = S(t)x_0 - S(t)\widetilde{x}_0 + \widetilde{x}(t)$. Лемма доказана.

Лемма 4.2. *Пусть $u_m \rightharpoonup u$, $\{u_m\} \subset U$, и по теореме* 2.2, *$u \in U$; $\varphi_m = \varphi[u_m]$, причем последовательность $\{\varphi_m\}$ ограничена в $L_\infty(0,T;X)$. Тогда существует подпоследовательность $\varphi_{m_k} \to \varphi[u]$ в $L_q(0,T;H)$.*

Доказательство. Будем считать, что

$$\|\varphi_m\|_{L_\infty(0,T;X)} \leqslant \sigma, \quad \|u_m\|_{L_p(0,T;Y)} \leqslant \sigma \quad \forall\, m \in \mathbb{N}.$$

Положим $z_m = \Phi\big(.,\varphi_m(.)\big)$, $\zeta_m = b\big(.,u_m(.),\varphi_m(.)\big) + z_m(.)$. По условиям $\mathbf{F}_1$, $\mathbf{B}_1$, $z_m \in L_2(0,T;X)$, $\zeta_m \in$ $\in L_2(0,T;X)$. Согласно условию $\mathbf{F}_3$ имеем

$$\big\|z_m(s)\big\|_X \leqslant \mathcal{N}_1(s,\sigma), \quad \text{п.в. } s \in [0;T]; \quad \mathcal{N}_1(.,\sigma) \in L_2[0;T].$$

Таким образом, $\|z_m\|_{L_2} \leqslant \sigma_1 = \big\|\mathcal{N}_1(.,\sigma)\big\|_{L_2}$. И аналогично, по условию $\mathbf{B}_2$ имеем $\Big\|b\big(s,u_m(s),\varphi_m(s)\big)\Big\|_X \leqslant$ $\leqslant \mathcal{N}_2(s,\sigma)\,\big\|u_m(s)\big\|_Y$. Следовательно,

$$\Big\|b\big(.,u_m(.),\varphi_m(.)\big)\Big\|_{L_2} \leqslant \big\|\mathcal{N}_2(.,\sigma)\big\|_{L_{\widetilde{p}}[0;T]}\|u_m\|_{L_p(0,T;Y)} \leqslant \sigma_2,$$

где $\sigma_2 = \sigma\big\|\mathcal{N}_2(.,\sigma)\big\|_{L_{\widetilde{p}}[0;T]}$. Таким образом, $\|\zeta_m\|_{L_2} \leqslant \sigma_1 + \sigma_2$.

Так как $L_q(0,T;X)$ — рефлексивное банахово пространство (в случае $q=2$ гильбертово) [10, гл. IV, § 1, теоремы 1.13, 1.14, замечания 1.10, 1.11, с. 159–163]), замкнутый шар в $L_q(0,T;X)$ слабо компактен [15, с. 51, теорема 4]. Стало быть, существует подпоследовательность $\varphi_{m_k} \rightharpoonup \overline{\varphi}$ в $L_q(0,T;X)$. Без ограничения общности рассуждений, примем, что $\varphi_m \rightharpoonup \overline{\varphi}$ в $L_q(0,T;X)$. Поскольку $X \subset H$ непрерывно и плотно, то $H^* \subset X^* = X$, и соответственно, $\big(L_q(0,T;H)\big)^* = L_{q'}(0,T;H^*) \subset L_{q'}(0,T;X) = \big(L_q(0,T;X)\big)^*$. Отсюда ясно, что $\varphi_m \rightharpoonup \overline{\varphi}$ в $L_q(0,T;H)$.

По лемме 4.1, найдется подпоследовательность $\varphi_{m_k} \to \varphi$ в $L_q(0,T;H)$. Без ограничения общности рассуждений, будем считать, что $\varphi_m \to \varphi$ в $L_q(0,T;H)$. Поскольку из сильной сходимости следует слабая, а слабый предел определяется однозначно [16, утверждение 2.22, с. 17], заключаем, что $\overline{\varphi} = \varphi$. Таким образом, $\varphi_m \rightharpoonup \varphi$ в $L_q(0,T;X)$, $\varphi_m \to \varphi$ в $L_q(0,T;H)$.

Поскольку из сходимости в $L_q[0;T]$ следует сходимость по мере [17, теорема VIII.5.1, с. 210], а из нее по теореме Ф. Рисса [17, теорема VI.5.3, с. 142] — существование подпоследовательности, сходящейся п.в., можем считать, с точностью до перехода к подпоследовательности, что

$$\big\|\varphi(t)\big\|_H \leqslant \lim_{m\to\infty}\big\|\varphi(t)-\varphi_m(t)\big\|_H + \big\|\varphi_m(t)\big\|_H \leqslant c\sigma,$$

где c — константа непрерывного вложения $X \subset H$.

Итак, $\varphi \in L_\infty(0,T;H) \cap L_q(0,T;X)$. По условию $\mathbf{F}_0$ получаем, что $z_m \to z = \Phi\big(.,\varphi(.)\big)$ в $L_1(0,T;H)$, и при этом $z \in L_2(0,T;X)$. Рассмотрим

$$\varphi_m(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)z_m(s)\,ds + \int\limits_0^t S(t-s)b\big(s,u_m(s),\varphi_m(s)\big)\,ds.$$

при фиксированном $t \in [0;T]$. По условиям $\mathbf{B}_1$, $\mathbf{B}_2$ $\forall\,\omega \in X$ функционал

$$g_t[u] = \big[\xi[u](t),\omega\big], \quad \xi[u](t) = \int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds$$

является линейным и непрерывным в $L_p(0,T;Y)$. Поэтому, переходя к пределу, получаем $\lim\limits_{m\to\infty}\big[\xi[u_m](t),\omega\big] = \big[\xi[u](t),\omega\big]$. Иначе говоря,

$$\big[\xi[u_m](t)-\xi[u](t),\omega(t)\big] \to 0 \quad \forall\omega \in L_{q'}(0,T;X), \quad t \in [0;T].$$

Это предел в смысле п.в., а следовательно, и предел сходимости по мере. Тогда, учитывая равномерную поточечную ограниченность функций

$$\Big|\big[\xi[u_m](t)-\xi[u](t),\omega(t)\big]\Big| \leqslant 2M\sigma\sqrt{T}\big\|\mathcal{N}_2'(.,c\sigma)\big\|_{L_{\tilde{p}}[0;T]}\big\|\omega(t)\big\|_X,$$

см. условие $\mathbf{B}_2'$, по теореме Лебега о предельном переходе под знаком интеграла [17, теорема VII.3.1, с. 166], получаем

$$\int\limits_0^T \big[\xi[u_m](t)-\xi[u](t),\omega(t)\big]\,dt \to 0 \quad \forall\omega \in L_{q'}(0,T;X).$$

Иначе говоря, $\xi[u_m] \rightharpoonup \xi[u]$ в $L_q(0,T;X)$. Следовательно, $\xi[u_m] \rightharpoonup \xi[u]$ в $L_q(0,T;H)$. Рассмотрим

$$P_m(t) = \int\limits_0^t S(t-s)b\big(s,u_m(s),\varphi_m(s)\big)\,ds - \int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds.$$

Ясно, что $P_m(t) = f_m(t) + \xi[u_m](t) - \xi[u](t)$, где

$$f_m(t) = \int\limits_0^t S(t-s)\Big\{b\big(s,u_m(s),\varphi_m(s)\big) - b\big(s,u_m(s),\varphi(s)\big)\Big\}\,ds.$$

Непосредственно из условия $\mathbf{B}_0$ получаем

$$\big\|f_m(t)\big\|_H \leqslant M\big\|b(t,u_m,\varphi_m) - b(t,u_m,\varphi)\big\|_{L_1(0,T;H)} \leqslant M\mathcal{N}_3(c\sigma)\,\big\|\varphi_m-\varphi\big\|_{L_q(0,T;H)} \to 0.$$

Таким образом, $\|f_m\|_{L_\infty(0,T;H)} \leqslant M\mathcal{N}_3(c\sigma)\,\|\varphi_m - \varphi\|_{L_q(0,T;H)} \to 0$. Следовательно, $f_m \to 0$ в $L_q(0,T;H)$, а значит, $f_m \rightharpoonup 0$ в $L_q(0,T;H)$. Соответственно, $P_m \rightharpoonup 0$ в $L_q(0,T;H)$. Аналогично, в силу условия $\mathbf{F}_0$ получаем $Q_m \to 0$ в $L_q(0,T;H)$, а значит, $Q_m \rightharpoonup 0$ в $L_q(0,T;H)$, где

$$Q_m(t) = \int\limits_0^t S(t-s)\Phi\big(s,\varphi_m(s)\big)\,ds - \int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds.$$

Из полученных соотношений вытекает, что $\varphi_m \rightharpoonup \widetilde{\varphi}$ в $L_q(0,T;H)$, где

$$\widetilde{\varphi}(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds + \int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds.$$

Однако выше уже было показано, что $\varphi_m \rightharpoonup \varphi$ в $L_q(0,T;H)$. И поскольку слабый предел существует только один, заключаем, что $\widetilde{\varphi} = \varphi$. А это, в свою очередь означает, что выполняется тождество

$$\varphi(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds + \int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds,$$

причем функция $\zeta(s) = \Phi\big(s,\varphi(s)\big) + b\big(s,u(s),\varphi(s)\big)$ принадлежит пространству $L_2(0,T;X)$, откуда вытекает, что $\varphi = \widetilde{\varphi} \in \mathbf{C}_w(0,T;X)$. Стало быть, $\varphi = \varphi[u]$. Итак, с точностью до перехода к подпоследовательности, $\varphi_m \rightharpoonup \varphi[u]$ в $L_q(0,T;X)$, $\varphi_m \to \varphi[u]$ в $L_q(0,T;H)$. Лемма доказана.

Далее будем предполагать, что функционал I_0 непрерывен на пространстве $L_q(0,T;H)$. Отсюда и из леммы 4.2 вытекает, что при $u_m \rightharpoonup u$, $\{u_m\} \subset U$ существует подпоследовательность такая, что $J_0[u_{m_k}] \to J_0[u]$.

Теорема 4.1. *Функционал $J_0[u]$ слабо непрерывен на множестве U. Функционал $J_\alpha[u]$ слабо полунепрерывен снизу на множестве U.*

Доказательство практически дословно воспроизводит доказательство [1, теорема 1.4], с тем лишь отличием, что вместо ссылки на [1, лемма 4.1] следует использовать ссылку на лемму 4.2.

Определение [15, гл. 1, § 1, определение 6, с. 49]. Пусть E — банахово пространство. Говорят, что последовательность $\{u_m\} \subset E$ *сходится к множеству U слабо в E*, если $\{u_m\}$ имеет хотя бы одну слабо сходящуюся подпоследовательность, причем все точки v, являющиеся слабым пределом какой-либо подпоследовательности последовательности $\{u_m\}$, принадлежат U.

Непосредственно из теорем 2.1–2.3, 4.1 и [15, гл. 1, § 1, теорема 2, с. 49] вытекает

Следствие 1. Множество U_* непусто и слабо компактно в пространстве $L_p(0,T;Y)$, и более того, всякая минимизирующая последовательность слабо сходится к множеству U_* в $L_p(0,T;Y)$.

Отдельно рассмотрим следующий, практически достаточно интересный случай. Пусть $X = X^+ \times X^-$, где X^+, X^- — гильбертовы пространства. Соответственно, всякий элемент $x \in X$ будем представлять в виде $x = (x^+, x^-)$. Известно, что скалярное произведение

$$[x,y] = [x^+,y^+]_+ + [x^-,y^-]_-,$$

где $[.,.]_+$, $[.,.]_-$ — скалярные произведения соответственно в X^+, X^-. Для $\varphi \in L_q(0,T;X)$ будем использовать аналогичное представление: $\varphi(t) = \big(\varphi^+(t), \varphi^-(t)\big)$, где $\varphi^+ \in L_q(0,T;X^+)$, $\varphi^- \in L_q(0,T;X^-)$. Соответственно,

$$\big(L_q(0,T;X)\big)^* = L_{q'}(0,T;X^*) = L_{q'}(0,T;(X^+)^*) \times L_{q'}(0,T;(X^-)^*) =$$

$$= L_{q'}(0,T;X^+) \times L_{q'}(0,T;X^-).$$

Предположим, что функции Φ и b таковы, что

$$\Phi(.,\varphi) = \Phi(.,\varphi^+), \quad b(.,u,x) = b(.,u,x^+).$$

Соответственно, и правая часть в задаче (3) представляется в виде $z = (z^+, z^-) \in L_2(0,T;X^+) \times L_2(0,T;X^-)$. Мы будем предполагать для простоты, что $\Phi^+ \equiv 0$, $b^+ \equiv 0$ (можно рассмотреть более общий случай, когда они просто не зависят от переменной состояния). Соответственно, при рассмотрении задачи (3) в качестве вспомогательной можно считать, что $z^+ = 0$. В этой ситуации предположение $\mathbf{W}_1$ можно, оставаясь в достаточно содержательных рамках, ослабить следующим образом.

Условие $\mathbf{V}_1$. Существуют элемент $\widetilde{x}_0 \in X$, всюду плотное множество $\widetilde{Z} \subset Z^- = L_2(0,T;X^-)$, банахово пространство $H \supset X^+$ непрерывно и плотно, $q \in (1,\infty)$ такие, что множество первых («положительных») компонент решений задачи (3) $\{x^+[(0,z^-)] : z^- \in \widetilde{Z},\ \|z^-\|_{Z^-} \leqslant \sigma\}$ при $x_0 = \widetilde{x}_0$ для любого $\sigma > 0$ предкомпактно в $L_q(0,T;H)$.

Соответствующая модификация условий на функции Φ и b производится тривиальным образом. Лемма 4.1 заменяется следующей.

Лемма 4.3. *Пусть $\{z_m = (0, z_m^-)\}$ — ограниченная последовательность в пространстве Z, $\{x_m\}$ — последовательность соответствующих решений задачи* (3) *при $z = z_m$. Тогда существует подпоследовательность $\{x^+_{m_k}\}$ сходящаяся в пространстве $L_q(0,T;H)$.*

Доказательство получается очевидной компиляцией из доказательства леммы 4.1.

Лемма 4.2 заменяется следующей.

Лемма 4.4. *Пусть $u_m \rightharpoonup u$, $\{u_m\} \subset U$, и по теореме* 2.2, *$u \in U$; $\varphi_m = \varphi[u_m]$, причем последовательность $\{\varphi_m\}$ ограничена в пространстве $L_\infty(0,T;X)$. Тогда существует подпоследовательность φ_{m_k} такая, что $\varphi^+_{m_k} \to \varphi^+[u]$ в $L_q(0,T;H)$, $\varphi_{m_k} \rightharpoonup \varphi[u]$ в $L_q(0,T;X)$.*

Доказательство получается очевидной компиляцией из доказательства леммы 4.2.

Если теперь предположить, что $J_0[u] = I_0\big[\varphi^+[u]\big]$ и что функционал I_0 непрерывен на пространстве $L_q(0,T;H)$, то теорема 4.1 и ее следствие останутся справедливыми.

5. ДОСТАТОЧНЫЕ УСЛОВИЯ ВЫПОЛНЕНИЯ ТРЕБОВАНИЯ АППРОКСИМАТИВНОЙ КОМПАКТНОСТИ

Пусть V — произвольное банахово пространство. Напомним следующее определение [18, chapter 7, comment 1, p. 197]. Неограниченный линейный оператор $B : D(B) \subset V \to V$ называется m-аккретивным (m-accretive), если $\overline{D(B)} = V$ и $\forall \lambda > 0$ оператор $I + \lambda B$ осуществляет биекцию $D(B) \to V$, причем $\big\|(I+\lambda B)^{-1}\big\| \leqslant 1$.

Лемма 5.1. *Пусть $S(t)$ — сильно непрерывная полугруппа сжатий в V, $z \in \mathbf{C}^1(0,T;V)$. Тогда существует плотное подмножество $V' \subset V$, $V' \ni 0$, такое, что для всех $x_0 \in V'$ функция*

$$x(t) = S(t)x_0 + \int_0^t S(t-s)z(s)\,ds \tag{10}$$

такова, что

$$x \in \mathbf{C}^1(0,T;V) \cap \mathbf{C}(0,T;V'). \tag{11}$$

Доказательство. Как указано в [18, chapter 7, comment 1, p. 197], для полугруппы $S(t)$ существует единственный m-аккретивный оператор B такой, что $S(t) = S_B(t)$. Здесь $S_B(t)$ — это отображение $D(B) \ni x_0 \to x[x_0]$, распространенное по непрерывности на все V, где $x[x_0]$ — сильное решение однородной задачи

$$\frac{dx}{dt} + Bx(t) = 0,\ t \in [0;T]; \quad x(0) = x_0,$$

существующее в силу теоремы Хилле–Иосиды (Hille–Yosida) [18, chapter 7, theorem 7.8]. В соответствии с теоремой Хилле–Иосиды в неоднородном случае [18, chapter 7, theorem 7.10, p. 198], для оператора B существует единственное решение неоднородной задачи (при условиях леммы, $V' = D(B)$):

$$\frac{dx}{dt} + Bx(t) = z(t),\ t \in [0;T]; \quad x(0) = x_0,$$

удовлетворяющее условию (11), и это решение определяется формулой (10). В силу линейности оператора B, $0 \in D(B)$. Лемма доказана.

Если заметить, что (10) — это слабое решение задачи (3), то непосредственно из леммы 5.1 вытекает

Лемма 5.2. *Пусть G — инфинитезимальный генератор сильно непрерывной полугруппы сжатий, $z \in \mathbf{C}^1(0,T;X)$. Тогда существует всюду плотное в X подмножество $X' \ni 0$ такое, что $\forall\, x_0 \in X'$ слабое решение задачи* (3) *обладает свойством* (11)*, и, стало быть, является сильным решением.*

Далее мы везде считаем, что G — генератор сильно непрерывной полугруппы сжатий (достаточно, чтобы $(-G)$ был максимальным монотонным оператором). Сделаем следующее предположение.

Условие $\mathbf{W}_2$. Существует подмножество $Z' \subset \mathbf{C}^1(0,T;X)$, плотное по норме пространства $L_2(0,T;X)$ такое, что для всех решений $x = x[z]$ задачи (3) при $x_0 = 0$, $z \in Z'$ (а по лемме 5.2 это, фактически, сильные решения) имеем

$$x \in \mathbf{C}^1\big(0,T;D(G)\big), \quad G\left(\frac{dx}{dt}\right) = \frac{d}{dt}\,Gx \tag{12}$$

(тем самым, предполагается, что производная справа существует и непрерывна).

Замечание 5.1. Если G — некоторый дифференциальный оператор, то условие (12) означает всего лишь равенство смешанных производных (для обобщенных производных это, очевидно, имеет место).

Напомним следующее определение [18, § 7.4, p. 193]. Неограниченный линейный оператор $G : D(G) \subset X \to X$, $\overline{D(G)} = X$, называется *самосопряженным*, если $D(G^*) = D(G)$, $G^* = G$.

Отметим, что всякий самосопряженный оператор является симметричным: $[Gx, y] = [x, Gy]$ для всех $x, y \in D(G)$. Если оператор $(-G)$ — максимальный монотонный, то для самосопряженности достаточно его симметричности [18, proposition 7.6]. Сделаем еще одно предположение.

Условие $\mathbf{W}_3$. Оператор G самосопряженный; оператор $(-G)$ — максимальный монотонный.

Лемма 5.3. *Пусть выполнены предположения $\mathbf{W}_2$, $\mathbf{W}_3$; $z \in Z'$, $x = x[z]$ — решение задачи* (3) *при $x_0 = 0$. Тогда справедливы следующие оценки:*

$$\|Gx\|_{L_2(0,T;X)} \leqslant \|z\|_{L_2(0,T;X)}, \quad \left\|\frac{dx}{dt}\right\|_{L_2(0,T;X)} \leqslant 2\|z\|_{L_2(0,T;X)}.$$

Доказательство. В соответствии с леммой 5.2 x имеет непрерывную сильную производную, причем выполнены соотношения (11), (12). Поэтому справедливо равенство

$$\frac{d}{dt}\,[x,y] = [x, G^*y] + [z,y] = [Gx,y] + [z,y] \quad \forall\, y \in D(G^*) = D(G).$$

Поскольку $\overline{D(G)} = X$, это равенство справедливо и для всех $y \in X$, откуда вытекает, что [19, доказательство леммы 2.2]

$$\frac{dx}{dt} = Gx + z. \tag{13}$$

Домножая (13) скалярно на Gx, получаем

$$[x', Gx] = [Gx, Gx] + [z, Gx], \quad t \in [0;T]. \tag{14}$$

Для сильной производной справедливо тождество [19, доказательство лемм 2.2, 2.3]:

$$\frac{d}{dt}\,[x, Gx] = [x', Gx] + \left[x, \frac{d}{dt}Gx\right],$$

и с учетом предположений $\mathbf{W}_2$, $\mathbf{W}_3$ имеем

$$\frac{d}{dt}\,[x, Gx] = [x', Gx] + [x, Gx'] = 2[x', Gx].$$

Подставляя в (14), получаем

$$\frac{1}{2}\,\frac{d}{dt}\,[x, Gx] = \|Gx\|_X^2 + [z, Gx].$$

Интегрируя полученное тождество на $[0;T]$, имеем

$$\int\limits_0^T \|Gx\|_X^2\,dt = -\int\limits_0^T [z, Gx]\,dt + \frac{1}{2}\big[x(T), Gx(T)\big],$$

учитывая, что $x(0) = x_0 = 0$. В силу монотонности оператора $(-G)$, получим

$$\big[x(T), Gx(T)\big] = -\big[-Gx(T), x(T)\big] \leqslant 0.$$

Таким образом, по неравенствам Коши–Буняковского и Гельдера, имеем

$$\|Gx\|^2_{L_2(0,T;X)} = \int\limits_0^T \|Gx\|^2_X\,dt \leqslant \int\limits_0^T \big|[z,Gx]\big|\,dt \leqslant$$

$$\leqslant \int\limits_0^T \|z\|_X\,\|Gx\|_X\,dt \leqslant \|z\|_{L_2(0,T;X)}\|Gx\|_{L_2(0,T;X)}.$$

Стало быть, $\|Gx\|_{L_2(0,T;X)} \leqslant \|z\|_{L_2(0,T;X)}$. Теперь, в силу (13), можем оценить

$$\left\|\frac{dx}{dt}\right\|_{L_2(0,T;X)} \leqslant \|Gx\|_{L_2(0,T;X)} + \|z\|_{L_2(0,T;X)} \leqslant 2\|z\|_{L_2(0,T;X)}.$$

Лемма доказана.

Замечание 5.2. Мы рассмотрели случай $x_0 = 0$. Если же $x_0 \neq 0$, то можно сделать замену: $y = x - x_0$. Тогда получим

$$\frac{dy}{dt} = \frac{dx}{dt} = Gx + z = Gy + z + Gx_0, \quad y(0) = 0.$$

Тем самым, справедлива оценка

$$\left\|\frac{dx}{dt}\right\|_{L_2(0,T;X)} = \left\|\frac{dy}{dt}\right\|_{L_2(0,T;X)} \leqslant 2\|z + Gx_0\|_{L_2(0,T;X)}.$$

Следующее утверждение — это известная теорема Лионса–Темама (J.L.Lions–R.Temam), см., например, [20, гл. 1, теорема 5.1, с. 70, включая ее доказательство].

Лемма 5.4. *Пусть V, V' — рефлексивные банаховы пространства, H — банахово пространство, $V \subset H$ компактно, $H \subset V'$ непрерывно, $p, q \in (1;+\infty)$. Тогда пространство*

$$W = \big\{z \in L_q(0,T;V) :\ z' \in L_p(0,T;V')\big\}$$

с нормой $\|z\|_W = \|z\|_{L_q(0,T;V)} + \|z'\|_{L_p(0,T;V')}$ является рефлексивным банаховым пространством, непрерывно вложенным в $\mathbf{C}(0,T;V')$ и компактно вложенным в $L_q(0,T;H)$.

Считая далее выполненными предположения $\mathbf{W}_2$, $\mathbf{W}_3$, положим $W = W[0;T]$ — множество решений $x = x[z]$ задачи (3), отвечающих всевозможным $z \in Z'$ при $x_0 = 0$. В соответствии с леммами 5.2, 5.3 справедливо вложение

$$W \subset W' = \big\{x \in \mathbf{C}^1(0,T;X) :\ x' \in L_2(0,T;X)\big\}.$$

Предположим, что $X \subset H$ компактно, где H — рефлексивное банахово пространство. Выберем произвольно $q \in (1;\infty)$. Тогда, очевидно, имеет место непрерывное вложение

$$W \subset W'_q = \big\{x \in L_q(0,T;X) :\ x' \in L_2(0,T;H)\big\}.$$

Применяя лемму 5.4 при $V = X$, $V' = H$, получаем, что $W'_q \subset L_q(0,T;H)$ компактно. Более того, согласно лемме 5.3, при $\|z\|_{L_2(0,T;X)} \leqslant \sigma$ справедлива оценка

$$\big\|x[z]\big\|_{W'_q} = \|x\|_{L_q(0,T;X)} + \|x'\|_{L_2(0,T;H)} \leqslant \|x\|_{L_q(0,T;X)} + c\|x'\|_{L_2(0,T;X)} \leqslant$$

$$\leqslant (T^{1/q}M\sqrt{T} + 2c)\sigma \equiv \mathcal{N}_0(\sigma).$$

Теперь для того, чтобы доказать выполнение предположения $\mathbf{W}_1$, осталось лишь установить плотность вложения $\mathbf{C}^1(0,T;X)$ в пространство $L_2(0,T;X)$, и потребовать также плотность вложения $X \subset H$.

Лемма 5.5. *Пространство $\mathbf{C}^1(0,T;X)$ плотно в $L_p(0,T;X)$ для любого $p \in [1;\infty)$.*

Доказательство. В соответствии с [10, гл. IV, § 1.3, лемма 1.3, с. 156], множество ступенчатых функций плотно в $L_p(0,T;X)$. Поэтому нам достаточно доказать, что любую ступенчатую функцию можно сколь угодно точно в метрике $L_p(0,T;X)$ приблизить функцией из $\mathbf{C}^1(0,T;X)$. Выберем произвольную ступенчатую функцию:

$$y(t) = x_i \in X,\ t \in [t_{i-1};t_i),\ i = \overline{1,k},$$

где $0 = t_0 < t_1 < \ldots < t_k = T$, а также произвольное число $\varepsilon > 0$ и (пока неопределенное) число $\delta > 0$. Для каждого $i = \overline{1, k-1}$ выберем непрерывно дифференцируемую функцию $\lambda_i(t)$ на $[t_i - \delta; t_i + \delta]$ со значениями в $[0; 1]$, исходя из условий

$$\lambda_i(t_i - \delta) = \lambda_i'(t_i - \delta) = 0,\ \lambda_i(t_i + \delta) = 1,\ \lambda_i'(t_i + \delta) = 0.$$

Можно взять, например,

$$\lambda_i(t) = \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{2\delta}\,(t - t_i - \delta)\right)\right).$$

Определим функцию

$$z_\delta(t) = \begin{cases} x_i + \lambda_i(t)(x_{i+1} - x_i), & t \in [t_i - \delta; t_i + \delta),\ i = \overline{1, k-1}, \\ x_i, & t \in [t_{i-1} + \delta; t_i - \delta),\ i = \overline{2, k-1}, \\ x_1, & t \in [0; t_1 - \delta), \\ x_k, & t \in [t_{k-1} + \delta; T], \end{cases}$$

т.е.

$$z_\delta(t) = \begin{cases} x_i + \lambda_i(t)(x_{i+1} - x_i), & t \in [t_i - \delta; t_i + \delta),\ i = \overline{1, k-1}, \\ y(t) & \text{иначе.} \end{cases}$$

Очевидно, что существует непрерывная сильная производная

$$z_\delta'(t) = \begin{cases} \lambda_i'(t)(x_{i+1} - x_i), & t \in [t_i - \delta; t_i + \delta),\ i = \overline{1, k-1}, \\ 0 & \text{иначе.} \end{cases}$$

Таким образом, $z_\delta \in \mathbf{C}^1(0, T; X)$. Оценим

$$\|y - z_\delta\|^p_{L_p(0,T;X)} = \sum_{i=1}^{k-1} \int\limits_{t_i-\delta}^{t_i+\delta} \|y(t) - z_\delta(t)\|^p_X\, dt \leqslant \sum_{i=1}^{k-1} \int\limits_{t_i-\delta}^{t_i+\delta} \|x_{i+1} - x_i\|^p_X\, dt.$$

Выбирая число $\delta = \delta(\varepsilon) > 0$ из условия $\sum\limits_{i=1}^{k-1} \int\limits_{t_i-\delta}^{t_i+\delta} \|x_{i+1} - x_i\|^p_X\, dt < \varepsilon^p$, получаем $\|y - z_\delta\|_{L_p(0,T;X)} < \varepsilon$. Лемма доказана.

Замечание 5.3. При соответствующем выборе вещественных функций $\lambda_i(t)$, $i = \overline{1, k-1}$, точно так же доказывается плотность вложения $\mathbf{C}^m(0, T; X)$ в $L_p(0, T; X)$, $m \in \mathbb{N}$.

Таким образом, из лемм 5.2–5.5 вытекает

Теорема 5.1. *Пусть H — рефлексивное банахово пространство, $X \subset H$ компактно и плотно, $q \in (1, \infty)$, и выполнены предположения $\mathbf{W}_2$, $\mathbf{W}_3$. Тогда выполнено предположение $\mathbf{W}_1$.*

При проверке рефлексивности, сепарабельности и компактного вложения конкретных пространств полезны следующие две леммы. Первая — это один из вариантов теоремы Реллиха–Кондрашова, [21, § I.11.5, с. 106].

Лемма 5.6. *Если $1 < p < \infty$, $n \geqslant \ell p$, $q < \dfrac{np}{n - \ell p}$, область $\Omega \subset \mathbb{R}^n$ представляет объединение конечного числа ограниченных областей, каждая из которых звездна относительно своего шара, то вложение $W^\ell_p(\Omega) \subset L_q(\Omega)$ компактно.*

О других вариантах см., например, [10, гл. II, § 1, лемма 1.28], [22, § 3.5, с. 82]. Во втором собраны известные свойства пространств Соболева [22, § 2.2, теорема 2.4, с. 30].

Лемма 5.7. *Пространство $W^\ell_p(\Omega)$ банахово при $1 \leqslant p \leqslant \infty$, рефлексивно при $1 < p < \infty$, сепарабельно при $1 \leqslant p < \infty$ и гильбертово при $p = 2$ относительно скалярного произведения $(x, y) = \sum\limits_{0 \leqslant |\alpha| \leqslant \ell} (D^\alpha x, D^\alpha y)$.*

6. СЛУЧАЙ НАРУШЕНИЯ ТРЕБОВАНИЯ АППРОКСИМАТИВНОЙ КОМПАКТНОСТИ

В данном случае приходится накладывать аналог требования аппроксимативной компактности на множество допустимых управлений. А именно, в этом разделе будем предполагать, что заданы: сепарабельное рефлексивное банахово пространство Y, компактно вложенное в рефлексивное банахово пространство H; пространство

$$W = \left\{u \in L_p(0, T; Y) : u' \in L_q(0, T; H)\right\}, \quad q \in (1, \infty),$$

$$\|u\|_W = \|u\|_{L_p(0,T;Y)} + \|u'\|_{L_q(0,T;H)},$$

с производной u' (по $t \in [0;T]$), понимаемой в смысле распределений; ограниченное и выпуклое подмножество $\widetilde{U} \subset W$. В качестве множества допустимых управлений U будет выступать замыкание множества $\widetilde{U}$ по норме $L_p(0,T;Y)$. По сути дела, это означает, что (для существования оптимального управления) допустимые управления должны быть хотя бы аппроксимативно достаточно гладкими (первого порядка). Учитывая, что оператор G у нас неограничен, а уравнение будет исследоваться нелинейное и при этом ничего не известно о компактных свойствах множества решений, вряд ли стоит сильно удивляться этому обстоятельству. Очевидно, что множество U будет выпуклым, замкнутым и ограниченным в пространстве $L_p(0,T;Y)$.

Далее будем считать выполненными следующие предположения.

Условие $\widetilde{\mathbf{B}}_2$. Условие $\mathbf{B}_2$ выполняется в следующей усиленной форме. Существует функция $\mathcal{N}_2'(t,r) : [0;T] \times \mathbb{R}_+ \to \mathbb{R}_+$, неубывающая по r и суммируемая по Лебегу со степенью $\widetilde{p} = \dfrac{2p}{p-2}$ такая, что при при п.в. $t \in [0;T]$, $x \in X$, $\|x\|_X \leqslant r$, выполняется оценка

$$\big\|b(t,v,x)\big\|_X \leqslant \mathcal{N}_2'(t,r)\,\|v\|_H \quad \forall\, v \in Y.$$

Лемма 6.1. *Пусть $\varphi_m = \varphi[u_m]$, $\{u_m\} \subset U$. Тогда существует подпоследовательность $\varphi_{m_k} \to \varphi$ в $L_\infty(0,T;X)$.*

Доказательство. Выберем числовую последовательность $\varepsilon_m \to 0$. Для каждого $m \in \mathbb{N}$, пользуясь плотностью вложения $\widetilde{U} \subset U$ в пространстве $L_p(0,T;Y)$, найдем $\widetilde{u}_m \in \widetilde{U}$ такое, что $\|u_m - \widetilde{u}_m\|_{L_p(0,T;Y)} < \varepsilon_m$. Положим $\widetilde{\varphi}_m = \varphi[\widetilde{u}_m]$. Поскольку $\widetilde{U} \subset U$, то согласно теореме 2.1, последовательности $\{\varphi_m\}$, $\{\widetilde{\varphi}_m\}$ ограничены в $L_\infty(0,T;X)$. Поэтому можно считать, что

$$\|u_m\|_{L_p(0,T;Y)} \leqslant \sigma, \quad \|\widetilde{u}_m\|_{L_p(0,T;Y)} \leqslant \sigma,$$

$$\|\varphi_m\|_{L_\infty(0,T;X)} \leqslant \sigma, \quad \|\widetilde{\varphi}_m\|_{L_\infty(0,T;X)} \leqslant \sigma \quad \forall\, m \in \mathbb{N}.$$

Рассмотрим

$$(\varphi_m - \widetilde{\varphi}_m)(t) = \int\limits_0^t S(t-s)\Big[\Phi\big(s,\varphi_m(s)\big) - \Phi\big(s,\widetilde{\varphi}_m(s)\big)\Big]\,ds +$$

$$+ \int\limits_0^t S(t-s)\Big[b\big(s,u_m(s),\varphi_m(s)\big) - b\big(s,\widetilde{u}_m(s),\widetilde{\varphi}_m(s)\big)\Big]\,ds.$$

Добавляя и вычитая под вторым интегралом $b\big(s,\widetilde{u}_m(s),\varphi_m(s)\big)$ и пользуясь предположениями $\mathbf{F}_2$, $\mathbf{B}_2$, $\mathbf{B}_3$, получаем

$$M^{-1}\big\|(\varphi_m - \widetilde{\varphi}_m)(t)\big\|_X \leqslant \int\limits_0^t \Big\|\Phi\big(s,\varphi_m(s)\big) - \Phi\big(s,\widetilde{\varphi}_m(s)\big)\Big\|_X\,ds +$$

$$+ \int\limits_0^t \Big\|b\big(s,u_m(s) - \widetilde{u}_m(s),\varphi_m(s)\big)\Big\|_X\,ds +$$

$$+ \int\limits_0^t \Big\|b\big(s,\widetilde{u}_m(s),\varphi_m(s)\big) - b\big(s,\widetilde{u}_m(s),\widetilde{\varphi}_m(s)\big)\Big\|\,ds \leqslant$$

$$\leqslant \big\|\mathcal{N}(.,\sigma)\big\|_{L_2}\sqrt{\int\limits_0^t \|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,s;X)}\,ds} + \big\|\mathcal{N}_2(.,\sigma)\big\|_{L_{p'}}\|u_m - \widetilde{u}_m\|_{L_p(0,T;Y)} +$$

$$+ \big\|\mathcal{N}_3(.,\sigma)\big\|_{L_2}\sqrt{\int\limits_0^t \|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,s;X)}\,ds}.$$

Пусть

$$\gamma_1 = \big\|\mathcal{N}(.,\sigma)\big\|_{L_2} + \big\|\mathcal{N}_3(.,\sigma)\big\|_{L_2},\ \gamma_2 = \big\|\mathcal{N}_2(.,\sigma)\big\|_{L_{p'}},\ \gamma = M\,\max\{\gamma_1,\gamma_2\}.$$

Используя очевидное неравенство $(a+b)^2 \leqslant 2(a^2+b^2)$, получаем

$$\big\|(\varphi_m - \widetilde{\varphi}_m)(t)\big\|_X^2 \leqslant 2\gamma^2 \left\{ \int\limits_0^t \|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,s;X)}\, ds + \varepsilon_m^2 \right\}.$$

Поскольку выражение справа не убывает по $t \in [0;T]$, то ясно, что для функции $f_m(t) = \|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,t;X)}$ справедлива оценка

$$f_m(t) \leqslant 2\gamma^2\varepsilon_m^2 + 2\gamma^2 \int\limits_0^t f_m(s)\, ds.$$

Функция $f_m \in L_\infty[0;T]$, см. замечание 1.3. Тогда по лемме 3.1 получим

$$f_m(t) = \|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,t;X)} \leqslant 2e^{2\gamma^2 t}\gamma^2\varepsilon_m^2 \leqslant 2e^{2\gamma^2 T}\gamma^2\varepsilon_m^2 \equiv \delta_m^2.$$

В частности, $\|\varphi_m - \widetilde{\varphi}_m\|^2_{L_\infty(0,T;X)} \leqslant \delta_m^2 \to 0$ при $m \to \infty$. Поскольку согласно нашим исходным предположениям в этом разделе множество $\widetilde{U}$ содержится и ограничено в W, а пространство W по лемме 5.4 компактно вложено в $L_p(0,T;H)$, ясно, что у последовательности $\{\widetilde{u}_m\} \subset \widetilde{U}$ существует подпоследовательность, сходящаяся в $L_p(0,T;H)$. Без ограничения общности рассуждений, будем считать, что $\widetilde{u}_m \to \widetilde{u}$ в $L_p(0,T;H)$. Следовательно, $\|\widetilde{u}_m - \widetilde{u}_n\|_{L_p(0,T;H)} \to 0$ при $m,n \to \infty$. Повторяя почти дословно проведенные выше рассуждения, с тем лишь отличием, что вместо предположения $\mathbf{B}_2$, используется на этот раз предположение $\widetilde{\mathbf{B}}_2$, получаем оценку вида $\|\widetilde{\varphi}_m - \widetilde{\varphi}_n\|_{L_\infty(0,T;X)} \leqslant \gamma_3\|\widetilde{u}_m - \widetilde{u}_n\|_{L_p(0,T;H)}$. Таким образом, при $E = E(T) = L_\infty(0,T;X)$ имеем

$$\|\varphi_m - \varphi_n\|_E = \big\|(\varphi_m - \widetilde{\varphi}_m) + (\widetilde{\varphi}_m - \widetilde{\varphi}_n) + (\widetilde{\varphi}_n - \varphi_n)\big\|_E \leqslant$$

$$\leqslant \delta_m + \gamma_3\|\widetilde{u}_m - \widetilde{u}_n\|_{L_p(0,T;H)} + \delta_n \to 0 \quad \text{при} \quad m,n \to \infty.$$

Это означает, что последовательность $\{\varphi_m\}$ фундаментальна в банаховом пространстве $L_\infty(0,T;X)$. Стало быть, сходится. Лемма доказана.

Лемма 6.2. *Пусть $u_m \rightharpoonup u$, $\{u_m\} \subset U$, и по теореме* 2.2, *$u \in U$; $\varphi_m = \varphi[u_m]$, причем по теореме* 2.1, *последовательность $\{\varphi_m\}$ ограничена в пространстве $L_\infty(0,T;X)$. Тогда существует подпоследовательность $\varphi_{m_k} \to \varphi[u]$ в $L_\infty(0,T;X)$.*

Доказательство. Зафиксируем произвольно $\varkappa \in (1;\infty)$. Будем считать, что $\|\varphi_m\|_{L_\infty(0,T;X)} \leqslant \sigma$, $\|u_m\|_{L_p(0,T;Y)} \leqslant \sigma\ \forall\, m \in \mathbb{N}$. Положим

$$z_m(.) = \Phi\big(., \varphi_m(.)\big) \in L_2(0,T;X),$$

$$\zeta_m(.) = b\big(., u_m(.), \varphi_m(.)\big) + z_m(.) \in L_2(0,T;X)$$

согласно условиям $\mathbf{F}_1$, $\mathbf{B}_1$. По лемме 6.1, существует подпоследовательность $\varphi_{m_k} \to \varphi$ в $L_\infty(0,T;X)$. Далее, без ограничения общности рассуждений, будем считать, что $\varphi_m \to \varphi$ в $L_\infty(0,T;X)$. А стало быть, $\varphi_m \to \varphi$ в $L_\varkappa(0,T;X)$. Поскольку из сильной сходимости следует слабая, а слабый предел определяется однозначно [16, утверждение 2.22, с. 17], заключаем, что $\varphi_m \rightharpoonup \varphi$ в $L_\varkappa(0,T;X)$. Заметим, что

$$\big\|\varphi(t)\big\|_X \leqslant \varliminf_{m\to\infty}\big\|\varphi(t) - \varphi_m(t)\big\|_X + \big\|\varphi_m(t)\big\|_X \leqslant \sigma \quad \Rightarrow \quad \|\varphi\|_{L_\infty(0,T;X)} \leqslant \sigma.$$

По условию $\mathbf{F}_2$, $z_m \to z = \Phi\big(., \varphi(.)\big)$ в $L_2(0,T;X)$. Рассмотрим

$$\varphi_m(t) = S(t)\varphi_0 + \int\limits_0^t S(t-s)z_m(s)\, ds + \int\limits_0^t S(t-s)b\big(s, u_m(s), \varphi_m(s)\big)\, ds$$

при фиксированном $t \in [0;T]$. По условиям $\mathbf{B}_1$, $\mathbf{B}_2\ \forall\, \omega \in X$ функционал

$$g_t[u] = \big[\xi[u](t), \omega\big], \quad \xi[u](t) = \int\limits_0^t S(t-s)b\big(s, u(s), \varphi(s)\big)\, ds$$

является линейным и непрерывным в $L_p(0,T;Y)$. Поэтому, переходя к пределу, получаем $\lim\limits_{m\to\infty}\big[\xi[u_m](t),\omega\big] = \big[\xi[u](t),\omega\big]$. Иначе говоря,

$$\big[\xi[u_m](t)-\xi[u](t),\omega(t)\big]\to 0\quad\forall\omega\in L_{\kappa'}(0,T;X),\quad t\in[0;T].$$

Это можно понимать как предел в смысле п.в., а следовательно, и как предел сходимости по мере. Тогда, учитывая равномерную поточечную ограниченность функций

$$\Big|\big[\xi[u_m](t)-\xi[u](t),\omega(t)\big]\Big|\leqslant 2M\sigma\sqrt{T}\big\|\mathcal{N}_2(.,\sigma)\big\|_{L_{\tilde p}[0;T]}\big\|\omega(t)\big\|_X,$$

см. условие $\mathbf{B}_2$, по теореме Лебега о предельном переходе под знаком интеграла [17, теорема VII.3.1, с. 166], получаем

$$\int\limits_0^T\big[\xi[u_m](t)-\xi[u](t),\omega(t)\big]\,dt\to 0\quad\forall\omega\in L_{\kappa'}(0,T;X).$$

Иначе говоря, $\xi[u_m]\rightharpoonup\xi[u]$ в $L_\kappa(0,T;X)$. Рассмотрим

$$P_m(t)=\int\limits_0^t S(t-s)b\big(s,u_m(s),\varphi_m(s)\big)\,ds-\int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds.$$

Ясно, что $P_m(t)=f_m(t)+\xi[u_m](t)-\xi[u](t)$, где

$$f_m(t)=\int\limits_0^t S(t-s)\Big\{b\big(s,u_m(s),\varphi_m(s)\big)-b\big(s,u_m(s),\varphi(s)\big)\Big\}\,ds.$$

Непосредственно из условия $\mathbf{B}_3$ получаем

$$\big\|f_m(t)\big\|_X\leqslant M\big\|b(t,u_m,\varphi_m)-b(t,u_m,\varphi)\big\|_{L_1(0,T;X)}\leqslant$$

$$\leqslant M\big\|\mathcal{N}_3(.,\sigma)\big\|_{L_1}\big\|\varphi_m-\varphi\big\|_{L_\infty(0,T;X)}\to 0.$$

Таким образом, $\big\|f_m\big\|_{L_\infty(0,T;X)}\leqslant M\big\|\mathcal{N}_3(.,\sigma)\big\|_{L_1}\big\|\varphi_m-\varphi\big\|_{L_\infty(0,T;X)}\to 0$. Следовательно, $f_m\to 0$ в $L_\kappa(0,T;X)$, а значит, $f_m\rightharpoonup 0$ в $L_\kappa(0,T;X)$. Соответственно, $P_m\rightharpoonup 0$ в $L_\kappa(0,T;X)$.

Аналогично, в силу условия $\mathbf{F}_2$ получаем $Q_m\to 0$ в $L_\kappa(0,T;X)$, а значит, $Q_m\rightharpoonup 0$ в $L_\kappa(0,T;X)$, где

$$Q_m(t)=\int\limits_0^t S(t-s)\Phi\big(s,\varphi_m(s)\big)\,ds-\int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds.$$

Из полученных соотношений вытекает, что $\varphi_m\rightharpoonup\widetilde\varphi$ в $L_\kappa(0,T;X)$, где

$$\widetilde\varphi(t)=S(t)\varphi_0+\int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds+\int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds.$$

Однако выше уже было показано, что $\varphi_m\rightharpoonup\varphi$ в $L_\kappa(0,T;X)$. И поскольку слабый предел существует только один, заключаем, что $\widetilde\varphi=\varphi$. А это, в свою очередь означает, что выполняется тождество

$$\varphi(t)=S(t)\varphi_0+\int\limits_0^t S(t-s)\Phi\big(s,\varphi(s)\big)\,ds+\int\limits_0^t S(t-s)b\big(s,u(s),\varphi(s)\big)\,ds,$$

причем функция $\zeta(s)=\Phi\big(s,\varphi(s)\big)+b\big(s,u(s),\varphi(s)\big)$ принадлежит пространству $L_2(0,T;X)$, откуда вытекает, что $\varphi=\widetilde\varphi\in\mathbf{C}_w(0,T;X)$. Стало быть, $\varphi=\varphi[u]$. Итак, с точностью до перехода к подпоследовательности, $\varphi_m\rightharpoonup\varphi[u]$ в $L_\kappa(0,T;X)$, $\varphi_m\to\varphi[u]$ в $L_\infty(0,T;X)$. Лемма доказана.

Далее будем предполагать, что функционал I_0 непрерывен на пространстве $L_\infty(0,T;X)$. Отсюда и из леммы 6.2 вытекает, что при $u_m\rightharpoonup u$, $\{u_m\}\subset U$ существует подпоследовательность такая, что $J_0[u_{m_k}]\to J_0[u]$.

Теорема 6.1. *Функционал $J_0[u]$ слабо непрерывен на множестве U. Функционал $J_\alpha[u]$ слабо полунепрерывен снизу на множестве U.*

Доказательство практически дословно воспроизводит доказательство [1, теорема 1.4], с тем лишь отличием, что вместо ссылки на [1, лемма 4.1] следует использовать ссылку на лемму 6.2.

Непосредственно из теорем 2.1–2.3, 6.1 и [15, гл. 1, § 1, теорема 2, с. 49] вытекает

Следствие 2. Множество U_* непусто и слабо компактно в пространстве $L_p(0,T;Y)$, и более того, всякая минимизирующая последовательность слабо сходится к множеству U_* в $L_p(0,T;Y)$.

7. ПРИМЕРЫ

Пусть $T>0$; $1\leqslant n\leqslant 3$; $\Omega\subset\mathbb{R}^n$ — открытое ограниченное множество. Следуя [6, (1.2),(1.3), с. 22–23], мы предполагаем, что его граница Γ регулярна (дважды непрерывно дифференцируема), причем Ω расположено локально с одной стороны от Γ. Положим $Q=\Omega\times(0;T]$, $\Sigma=\Gamma\times(0;T]$. Следуя [18], мы несколько усилим сделанные выше предположения, считая дополнительно, что Ω — это область класса $\mathbf{C}^\infty$ с ограниченной границей Γ. Отметим, что это соответствует также и предположениям [2].

7.1. Уравнение теплопроводности

Рассмотрим задачу об отыскании функции $\varphi(x,t):\overline{\Omega}\times[0;T]\to\mathbb{R}$ такой, что

$$\frac{\partial\varphi}{\partial t}-\Delta\varphi=f(t,\varphi,u)=\Phi(t,\varphi)+b(t,\varphi,u),\quad (x,t)\in Q; \tag{15}$$

$$\varphi\,\big|_\Sigma=0;\quad \varphi(x,0)=\varphi_0(x),\quad x\in\Omega, \tag{16}$$

где $\Delta=\sum\limits_{i=1}^{n}\dfrac{\partial^2}{\partial x_i^2}$ — оператор Лапласа.

Прежде всего, чтобы указать и обосновать выбор функциональных пространств, рассмотрим линейную неуправляемую задачу: $f(t,\varphi,u)=z(t)$. Только после этого можно будет сформулировать (соответствующим образом согласованные) условия на выбор функций Φ, b, u.

В [18, sec.10.1] рассматривалась аналогичная задача при $f\equiv 0$. Следуя [18, sec.10.1], возьмем

$$\varphi(t)=\varphi(.,t),\quad X=L_2(\Omega),\quad G\varphi=\Delta\varphi,\quad D(G)=\mathbb{H}^2(\Omega)\cap\mathbb{H}_0^1(\Omega).$$

Таким образом, краевое условие встраивается в область определения $D(G)$. В итоге задача (15), (16) при $f=z=0$ переписывается в виде абстрактного дифференциального уравнения (3). Как показано в [18, sec.10.1], оператор $(-G)$ является максимальным монотонным, т.е. G есть инфинитезимальный генератор сильно непрерывной полугруппы сжатий — условие $\mathbf{G}_1$ выполнено при $M=1$. Далее будем считать, что $n\in\overline{1,3}$. Если $f=z\in L_2(Q)=L_2(0,T;X)$, $\varphi_0=0$, то, как показано в [6], существует единственное решение задачи

$$\varphi=\varphi[z]\in L_2\big(0,T;\mathbb{H}^2(\Omega)\big),\quad \frac{\partial\varphi}{\partial t}\in L_2(Q),$$

причем, см. [6, (1.31)], имеет место оценка

$$\left\|\frac{\partial\varphi}{\partial t}\right\|_{L_2(Q)}\leqslant c\,\|z\|_{L_2(Q)}.$$

В соответствии с [22, § 3.5], $X'=\mathbb{H}^2(\Omega)\subset L_q(\Omega)$ компактно при $q\in(1;\infty)$. Таким образом, имеет место вложение множества

$$\big\{\varphi[z]:\ z\in\mathbf{C}^1(0,T;X)\big\}\subset W=\big\{\varphi\in L_2(0,T;X'):\varphi'\in L_2(0,T;X)\big\},$$

где, согласно лемме 5.4, W компактно вложено в $L_2(0,T;X)$. Впрочем, за счет повышения гладкости функции z, см. замечание 5.3, можно взять

$$\big\{\varphi[z]:\ z\in\mathbf{C}^k(0,T;X')\big\}\subset W_q=\big\{\varphi\in L_q(0,T;X'):\varphi'\in L_2(0,T;X)\big\},$$

с компактным вложением $W_q\subset L_q(0,T;X)$, $q\in(2;\infty)$. В любом случае, оказывается, что условие $\mathbf{W}_1$ выполнено. Значит, можно пользоваться результатами разд. 4 при соответствующем выборе условий на функции Φ, b, u. А именно, будем считать, что заданы $T>0$, $p\in[2;+\infty)$, а также выпуклое, замкнутое, ограниченное множество U в пространстве $L_p(0,T;Y)$, где Y — сепарабельное рефлексивное банахово пространство; $u\in U$;

$X = L_2(\Omega)$. Соответственно, будем предполагать, что функция Φ удовлетворяет условиям $\mathbf{F}_1$–$\mathbf{F}_3$, $\mathbf{F}_0$; функция b — условиям $\mathbf{B}_1$, $\mathbf{B}_2'$, $\mathbf{B}_3$, $\mathbf{B}_0$. По схеме, описанной в разд. 1, управляемая задача (16) может быть представлена в виде абстрактного дифференциального уравнения (5). Тем самым, применима теорема 2.1. А это, в свою очередь, позволяет нам утверждать существование числа $T_0 > 0$ такого, что для всех $T \in (0; T_0]$ и $u \in U$ управляемая задача (16) имеет единственное решение $\varphi = \varphi[u]$. Далее число $T \in (0; T_0]$ будем считать произвольно фиксированным.

Пусть задан непрерывный функционал $I_0 : L_\infty(0, T; X) \to \mathbb{R}$, ограниченный на ограниченных множествах, $J_0[u] = I_0\big(\varphi[u]\big)$, $u \in U$, где $\varphi[u]$ — решение задачи (16), отвечающее управлению u. Для произвольно заданного $\alpha \geqslant 0$ рассмотрим задачу оптимизации

$$J_\alpha[u] = J_0[u] + \frac{1}{2}\alpha\|u\|^2_{L_p(0,T;Y)} \to \min_{u\in U}. \tag{17}$$

Пусть $J^*_\alpha = \inf\limits_{u\in U} J_\alpha[u]$, $U_* = \big\{u \in U : \ \ J_\alpha[u] = J^*_\alpha\big\}$. Применяя теорему 4.1 и ее следствие, получаем, что справедлива

Теорема 7.1. *При сделанных предположениях множество U_* в задаче* (17) *непусто и слабо компактно в пространстве $L_p(0, T; Y)$, и более того, всякая минимизирующая последовательность слабо сходится к множеству U_* в $L_p(0, T; Y)$.*

7.2. Волновое уравнение

Рассмотрим задачу об отыскании функции $\varphi(x, t) : \overline{\Omega} \times [0; T] \to \mathbb{R}$ такой, что

$$\frac{\partial^2\varphi}{\partial t^2} - \Delta\varphi = g(t, \varphi, u) = \Phi_1(t, \varphi) + b_1(t, u, \varphi), \quad (x, t) \in Q; \tag{18}$$

$$\varphi\,\big|_\Sigma = 0; \tag{19}$$

$$\varphi(x, 0) = \varphi_0(x), \quad x \in \Omega, \tag{20}$$

$$\frac{\partial\varphi}{\partial t}(x, 0) = \psi_0(x), \quad x \in \Omega. \tag{21}$$

Так же, как и в предыдущем примере, условия на выбор функций Φ, b, u укажем позже. А предварительно рассмотрим случай $g(t, \varphi, u) = z^-(t)$. В [18, sec.10.3] рассматривалась аналогичная задача при $g \equiv 0$. Следуя [18, sec.10.3], перепишем уравнение (18) в виде системы уравнений первого порядка

$$\frac{\partial\varphi}{\partial t} = \psi, \ \frac{\partial\psi}{\partial t} = \Delta\varphi + g(t, \varphi, u), \quad (x, t) \in Q. \tag{22}$$

Пусть $\eta = (\varphi, \psi)^*$ (здесь $*$ — знак транспонирования); $\eta(t) = \eta(., t)$. Тогда систему (22) можно переписать в виде:

$$\frac{d\eta}{dt} = G\eta + f(., \eta, u), \tag{23}$$

$$G = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix}, \ \eta = \begin{pmatrix} \varphi \\ \psi \end{pmatrix}, \ f(t, \eta, u) = \begin{pmatrix} 0 \\ g(t, \varphi, u) \end{pmatrix}.$$

Опять же следуя [18, sec.10.3], возьмем $X = X^+ \times X^- = \mathbb{H}^1_0(\Omega) \times L_2(\Omega)$;

$$D(G) = \big\{\mathbb{H}^2(\Omega) \cap \mathbb{H}^1_0(\Omega)\big\} \times \mathbb{H}^1_0(\Omega).$$

Таким образом, краевое условие (19) встраивается в область определения $D(G)$. Как указано в [18, sec.10.3, remark 7, p. 338] со ссылкой на [18, corollary 9.19], в случае, когда множество Ω ограничено, можно использовать на $\mathbb{H}^1_0(\Omega)$ скалярное произведение $\int\limits_\Omega \nabla\varphi_1 \cdot \nabla\varphi_2\, dx$, а на $X = \mathbb{H}^1_0(\Omega) \times L_2(\Omega)$ скалярное произведение

$$[\eta_1, \eta_2]_X = \int\limits_\Omega \nabla\varphi_1 \cdot \nabla\varphi_2\, dx + \int\limits_\Omega \psi_1\psi_2\, dx.$$

При этом оказывается, что оба оператора G и $(-G)$ являются максимальными монотонными. В частности, отсюда следует, что G — инфинитезимальный генератор сильно непрерывной полугруппы сжатий. Следовательно, условие $\mathbf{G}_1$ выполнено при $M = 1$. Тем самым, можем использовать результаты разд. 1 для задачи

$$\eta'(t) = G\eta(t) + z(t), \quad t \in [0; T]; \quad \eta(0) = \eta_0, \tag{24}$$

вида (3) при указанном выше выборе пространства X и оператора G, а также $z \in L_2(0,T;X)$. При $g(t,\varphi,u) = z^-(t)$ задача (18)–(21) переписывается в виде (24) при $z = (0,z^-)$, $z^- \in L_2(Q)$.

Будем считать, что заданы $T > 0$, $p \in [2;+\infty)$, а также выпуклое, замкнутое, ограниченное множество U в пространстве $L_p(0,T;Y)$, где Y — сепарабельное рефлексивное банахово пространство; $u \in U$. Соответственно, будем предполагать, что функция Φ удовлетворяет условиям $\mathbf{F}_1$–$\mathbf{F}_3$, $\mathbf{F}_0$; функция b — условиям $\mathbf{B}_1$, $\mathbf{B}_2'$, $\mathbf{B}_3$, $\mathbf{B}_0$, где $\eta = (\eta^+,\eta^-)^* = (\varphi,\psi)^*$,

$$\Phi(.,\eta) = \Phi(.,\eta^+) = \begin{pmatrix} 0 \\ \Phi_1(.,\eta^+) \end{pmatrix}, b(.,u,\eta) = b(.,u,\eta^+) = \begin{pmatrix} 0 \\ b_1(.,u,\eta^+) \end{pmatrix}.$$

По схеме, описанной в разд. 1, управляемая задача (18)–(21) может быть представлена в виде абстрактного дифференциального уравнения

$$\eta'(t) = G\eta(t) + \Phi\big(t,\eta(.)\big) + b\big(t,u(t),\eta(t)\big),\ t \in [0;T];\ \eta(0) = \eta_0 \tag{25}$$

вида (5). Тем самым, применима теорема 2.1. А это, в свою очередь, позволяет нам утверждать существование числа $T_0 > 0$ такого, что для всех $T \in (0;T_0]$ и $u \in U$ управляемая задача (25) (а тем самым, и задача (18)–(21)) имеет единственное решение $\eta = \eta[u]$. Далее число $T \in (0;T_0]$ будем считать произвольно фиксированным.

Заметим, что $\|f\|_X = \sqrt{[f,f]_X} = \big\|g(t,\varphi,u)\big\|_{L_2(\Omega)}$; $\eta^+ = \varphi \in X^+ = \mathbb{H}_0^1(\Omega) \subset L_6(\Omega) = H$ компактно (при $n = 1,2,3$); $Z^- = L_2(0,T;X^-) = L_2(Q)$. Для решений $\eta = \eta[z]$ при $f \equiv z$, $\eta_0 = (0,0)^*$ имеет место оценка, см., например, [2]:

$$\|\eta^+ = \varphi\|_{\mathbb{H}_0^1(\Omega)} + \left\|\eta^- = \psi = \frac{\partial\varphi}{\partial t}\right\|_{L_2(\Omega)} \leqslant c\,\|z^-\|_{L_2(Q)}, \quad t \in [0;T].$$

Поэтому в соответствии с леммой 5.4 множество первых компонент решений вспомогательной задачи (при $f \equiv z$, $\eta_0 = (0,0)^*$)

$$\big\{\eta^+[(0,z^-)]:\ z^- \in Z^-,\ \|z^-\|_{Z^-} \leqslant \sigma\big\}$$

предкомпактно в $L_q(0,T;H)$ при любом $q \in (1;\infty)$. Таким образом, выполнено условие $\mathbf{V}_1$, и мы здесь находимся в ситуации, описанной в завершающей части разд. 4.

Пусть задан непрерывный функционал $I_0 : L_q(0,T;H) \to \mathbb{R}$, ограниченный на ограниченных множествах, $J_0[u] = I_0\big(\eta^+[u]\big) = I_0\big(\varphi[u]\big)$, $u \in U$, где $\eta[u]$ — решение задачи (25), отвечающее управлению u; соответственно, $\varphi[u]$ — решение задачи (18)–(21), отвечающее управлению u. Для произвольно заданного $\alpha \geqslant 0$ рассмотрим задачу оптимизации

$$J_\alpha[u] = J_0[u] + \frac{1}{2}\alpha\|u\|^2_{L_p(0,T;Y)} \to \min_{u\in U}. \tag{26}$$

Пусть $J_\alpha^* = \inf\limits_{u\in U} J_\alpha[u]$, $U_* = \big\{u \in U:\ J_\alpha[u] = J_\alpha^*\big\}$.

В соответствии с замечаниями из завершающей части разд. 4, справедлива

Теорема 7.2. *При сделанных предположениях множество U_* в задаче* (26) *непусто и слабо компактно в пространстве $L_p(0,T;Y)$, и более того, всякая минимизирующая последовательность слабо сходится к множеству U_* в $L_p(0,T;Y)$.*

СПИСОК ЛИТЕРАТУРЫ

1. Чернов А.В. О существовании оптимального управления в задаче оптимизации младшего коэффициента полулинейного эволюционного уравнения // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 7. С. 1084–1099.

2. Ismayilova G.G. The problem of the optimal control with a lower coefficient for weakly nonlinear wave equation in the mixed problem // European journal of pure and applied mathematics 2020. Vol. 13. № 2. P. 314–322.

3. Лионс Ж.-Л. Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972. 415 с.

4. Tröltzsch F. Optimal control of partial differential equations. Theory, methods and applications. Graduate Studies in Mathematics. V. 112. Providence, RI: American Mathematical Society (AMS), 2010. xv+399 p.

5. Bewley T., Temam R., Ziane M. Existence and uniqueness of optimal control to the Navier-Stokes equations // C. R. Acad. Sci., Paris, Ser. I, Math. 2000. V. 330. № 11. P. 1007–1011.

6. Лионс Ж.-Л. Управление сингулярными распределенными системами. М.: Наука, 1987. 368 с.

7. Фурсиков А.В. Оптимальное управление распределенными системами. Теория и приложения. Новосибирск: Научная книга, 1999. xii+352 с.

8. Балакришнан А.В. Прикладной функциональный анализ. М.: Наука, 1980. 383 с.

9. Хилле Э., Филлипс Р. Функциональный анализ и полугруппы. М.: Изд-во иностр. лит., 1962. 830 с.

10. Гаевский Х., Грёгер К., Захариас К. Нелинейные операторные уравнения и операторные дифференциальные уравнения. М.: Мир, 1978. 336 с.

11. Функциональный анализ / под ред. С.Г. Крейна. М.: Наука, 1979. 418 с.

12. Pazy A. Semigroups of Linear Operators and Applications to Partial Differential Equations. New York etc.: Springer-Verlag, 1983. viii+279 p.

13. Натансон И.П. Теория функций вещественной переменной. М.: Наука, 1974. 480 с.

14. Чернов А.В. Операторные уравнения II рода: теоремы о существовании и единственности решения и о сохранении разрешимости // Дифференц. ур-ния. 2022. Т. 58. № 5. С. 656–668.

15. Васильев Ф.П. Методы решения экстремальных задач. М.: Наука, 1981. 400 с.

16. Рыжиков В.В. Курс лекций по функциональному анализу. М.: МГУ, 2004. 24 с.

17. Вулих Б.З. Краткий курс теории функций вещественной переменной. М.: Наука, 1973. 352 с.

18. Brezis H. Functional analysis, Sobolev spaces and partial differential equations. N.Y., Dordrecht, Heidelberg, London: Springer, 2011. xiv+600 p.

19. Чернов А.В. О дифференцировании функционала в задаче параметрической оптимизации коэффициента уравнения глобальной электрической цепи // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 9. С. 1586–1601.

20. Лионс Ж.-Л. Некоторые методы решения нелинейных краевых задач. М.: Мир, 1972. 588 с.

21. Соболев С.Л. Некоторые применения функционального анализа в математической физике. М.: Наука, 1988. 336 с.

22. Павлова М.Ф., Тимербаев М.Р. Пространства Соболева (теоремы вложения). Казань: КГУ, 2010. 123 с.

ON THE EXISTENCE OF OPTIMAL CONTROL FOR A SEMILINEAR EVOLUTION EQUATION WITH AN UNBOUNDED OPERATOR

A. V. Chernov*

N.I. Lobachevsky State University of Nizhny Novgorod, Gagarin Ave. 23, Nizhny Novgorod, 603950, Russia
**e-mail: chavnn@mail.ru*

Received 28 November, 2023
Revised 16 January, 2024
Accepted 31 January, 2024

Abstract. The paper studies the problem of optimal control for an abstract first-order semilinear differential equation in a Hilbert space, with an unbounded operator and a control linearly entering the right-hand side. The objective functional is assumed to be additively separable with respect to the state and control, with a fairly general dependence on the state. A theorem on the existence of an optimal control is proved for this problem, and properties of the set of optimal controls are established. Due to the nonlinearity of the equation under study, the author further develops previous results on total preservation of unique global solvability and solution estimates for similar equations. This estimate proves essential for the investigation. As examples, a nonlinear heat conduction equation and a nonlinear wave equation are considered.

Keywords: semilinear evolution equation with an unbounded operator in a Hilbert space, existence of optimal control, nonlinear heat conduction equation, nonlinear wave equation.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 766–779

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 519.62

СКОРОСТЬ СХОДИМОСТИ АЛГОРИТМОВ РЕШЕНИЯ ЛИНЕЙНОГО УРАВНЕНИЯ МЕТОДОМ КВАНТОВОГО ОТЖИГА[1)]

© 2024 г. С. Б. Тихомиров[1],*, В. С. Шалгин[2],**

[1] Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio, Rua Marquês de São Vicente, 225, Gávea – Rio de Janeiro, RJ – Brasil Cep: 22451-900 – Cx. Postal: 38097

[2] 199034 Санкт-Петербург, Университетская наб., 7/9, Санкт-Петербургский государственный университет, Россия

**e-mail: sergey.tikhomirov@gmail.com*

***e-mail: st086496@student.spbu.ru*

Поступила в редакцию 06.11.2023 г.

Переработанный вариант 26.12.2023 г.

Принята к публикации 06.02.2024 г.

Рассмотрены различные итеративные алгоритмы решения линейного уравнения $ax = b$ с помощью квантового вычислительного устройства, работающего по принципу квантового отжига. В предположении, что результат работы компьютера описывается распределением Больцмана, показано, при каких условиях алгоритмы решения уравнения сходятся, и дана оценка скорости их сходимости. Рассмотрено применение данного подхода для алгоритмов, использующих как бесконечное количество кубитов, так и малое количество кубитов. Библ. 31. Фиг. 2.

Ключевые слова: адиабатические квантовые вычисления, квантовый отжиг, линейное уравнение, распределение Больцмана, усеченное нормальное распределение.

DOI: 10.31857/S0044466924050061, **EDN**: YDHDHX

1. ВВЕДЕНИЕ

Квантовые вычисления представляют собой новую парадигму выполнения вычислений, предложенную Ю.И. Маниным (см. [1]) и Р. Фейнманом (см. [2]). Функционирование таких вычислительных устройств основано на квантовой механике. В основе вычислений лежат квантовые биты (кубиты), которые могут находиться не только в состоянии “0” или “1”, но и в их суперпозиции. Что более важно, квантовые биты могут находиться в запутанном состоянии. Таким образом, система из n кубитов описывается 2^n комплексными числами, более того операция над одним кубитом “меняет состояние” всех запутанных кубитов, что является основой квантового параллелизма — одновременного проведения вплоть до $O(2^n)$ операций над числами, описывающими состояние системы (см. [3], [4]). В квантовых вычислениях появляются и ограничения, не свойственные классическим вычислениям, например, невозможность копирования состояния и считывания состояния без его изменения. Для множества задач разработаны алгоритмы для квантовых компьютеров, работающие быстрее, чем их классические аналоги, вплоть до экспоненциального ускорения: например, алгоритм поиска (алгоритм Гровера) (см. [5]), разложение на множители (алгоритм Шора) (см. [6]), приближенное решение систем линейных уравнений (алгоритм HHL) (см. [7]).

Есть две основные модели квантовых вычислений: универсальная схемная модель (circuit based) (см. [3], [4]) и адиабатическая модель (см. [8]). В схемной модели операции выполняются одна за другой, как в классических вычислениях. Операции — квантовые вентили — представляют собой унитарные операторы, действующие на состояние системы кубитов. На основе этой модели функционируют квантовые компьютеры таких компаний,

[1)] Авторы благодарят Михаила Скопенкова за внимание к работе и полезные замечания. Исследование разд. 1, 2 и п. 3.1 выполнено при финансовой поддержке гранта РНФ № 21-11-00047. Исследование пп. 3.3 и 3.4 выполнено в Санкт-Петербургском международном математическом институте имени Леонарда Эйлера при финансовой поддержке Минобрнауки РФ (соглашение № 075–15–2022–287 от 06.04.2022). Исследование п. 3.2 выполнено при поддержке Projeto Paz и Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Finance Code 001.

как IBM, Google, Intel. Основными ограничениями для практического использования таких квантовых компьютеров в настоящее время является небольшой размер (порядка 100 кубитов) и низкая точность выполнения операций.

Принцип работы адиабатических квантовых компьютеров основан не на последовательном выполнении операций, а на адиабатической теореме (см. [8]). Если изначально система находилась в состоянии минимальной энергии для гамильтониана H_1, то при достаточно медленной эволюции гамильтониана:

$$H(t) = (1 - vt)H_1 + vtH_2, \quad t \in [0, 1/v],$$

в конечный момент времени система будет находиться в состоянии минимальной энергии для гамильтониана H_2. Гамильтониан H_2 строится таким образом, что состояние минимальной энергии для него будет решением некоторой задачи. Такой подход позволяет эффективно решать задачи дискретной оптимизации, например, задачу коммивояжера (см. [9]) и задачу разрешимости булевых функций (см. [10]). Известно, что адиабатическая модель эквивалентна универсальной схемной модели квантовых вычислений (см. [11]).

С адиабатическими квантовыми вычислениями тесно связан квантовый отжиг (quantum annealing) — квантовый аналог алгоритма имитации отжига (см. [12]). Устройство, работающее по принципу квантового отжига, носит название “quantum annealer” (QA). Процесс поиска также начинается с состояния минимальной энергии системы для гамильтониана H_1. Однако структура целевого гамильтониана H_2 более ограничена по сравнению с той, что фигурирует в общей адиабатической модели. А именно, квантовый отжиг нацелен на поиск точки минимума целевой функции модели Изинга (см. [12], [13]). В упрощенном виде ее можно представить так:

$$F(\sigma) = \sum_i h_i \sigma_i + \sum_{i<j} J_{ij} \sigma_i \sigma_j, \tag{1}$$

где $\sigma_i \in \{-1, 1\}$ представляют собой спины кубитов, а h_i и J_{ij} — коэффициенты линейных и квадратичных слагаемых соответственно. Результат работы квантового отжига — набор спинов кубитов $\{\sigma_i\}$, которые доставляют минимум функции $F(\sigma)$.

Точный результат может быть получен только в случае нулевой абсолютной температуры у QA, что в текущих реализациях является недостижимым. На практике такое устройство будет выдавать ”сэмпл” (sample) из распределения Больцмана (см. [14]). Вероятность получить состояние σ зависит от значения функции F и параметра β:

$$P(\sigma) \propto e^{-\beta^2 F(\sigma)},$$

параметр $\beta \to \infty$ при стремлении температуры устройства к нулю. Ввиду этого квантовый отжиг является неточным, эвристическим алгоритмом. Кроме того, наличие шума также может внести помехи в работу компьютера.

Высокий интерес к модели квантового отжига обусловлен наличием реализации устройства, работающего по данной модели с большим количеством кубитов. Соответствующая реализация представлена устройствами компании D-Wave Systems (см. [15]), количество кубитов в которых достигает 5000. Дополнительными ограничениями в работе компьютера являются граф связности между кубитами и неточность выполнения операций (квантовый шум). Перед тем, как решить задачу с помощью компьютера D-Wave, ее необходимо перевести в термины модели Изинга. Согласованность распределения Больцмана и результата работы D-Wave достаточно хорошо продемонстрирована в [13], [16]–[18] (для обсуждения вопроса о наличии превосходства QA над классическими компьютерами см., например, [19], [20]).

В нашей работе мы будем опираться на модель квантовых вычислений, работающую по принципу квантового отжига. Одной из важных для приложений задач является решение систем линейных алгебраических уравнений. Задача решения системы $Ax = b$ эквивалентна задаче минимизации функции $||Ax - b||^2$, часто называемой в литературе “linear least squares problem” (LLS). Данная задача может быть решена с помощью QA путем ее перевода в целевую функцию (1) модели Изинга. Заметим, что линейность по переменной x является необходимым условием, так как в противном случае целевая функция не будет иметь форму (1).

Во многих работах исследуется решение задачи LLS с помощью QA. В [21] предлагается подход к переформулировке задачи LLS в форму, эквивалентную модели Изинга. Также в работе авторы предположили, что QA лучше всего подходит для решения задачи LLS в случае, если матрица A разреженная, или когда компоненты вектора x бинарные. Этот подход получил дальнейшее развитие. В [22] предлагается подход к решению произвольной системы линейных уравнений, а также приводятся условия, при которых возможно получить ускорение по сравнению с лучшим из известных классических алгоритмов решения произвольных систем линейных уравнений. В [23] рассматривается задача решения одного уравнения с одной неизвестной и задача LLS, подробно излагается процесс переформулировки исходной задачи в форму модели Изинга и встраивания полного графа задачи в граф компьютера D-Wave. В [24] исследуется вопрос о целесообразности использования

QA для решения систем уравнений и предлагается гибридный алгоритм решения линейных систем. Подход решения линейных систем с помощью QA нашел применение во многих задачах, в которых возникает необходимость решения системы уравнений: оценке линейной регрессии (см. [25]), для задач сейсмической томографии (см. [26]), в задаче определения преобразования из точечного множества (см. [27]), решения краевой задачи для эллиптических уравнений (см. [28]).

Во всех упомянутых работах были предложены алгоритмы решения линейных уравнений и систем, включая и итеративные алгоритмы (см. [23], [26]–[28]). Однако рассматривалась только экспериментальная постановка задачи, теоретические вопросы сходимости подобных итеративных алгоритмов не исследовались.

В настоящей работе рассматривается подход к решению задачи LLS, аналогичный [23], [26]–[28], для случая одного уравнения с одной неизвестной:

$$ax = b.$$

В рамках данного подхода мы учитываем подверженность квантового компьютера ошибкам и вероятностную природу QA. Нас будет интересовать вопрос устойчивости итеративного алгоритма к погрешностям, связанным с вероятностной обусловленностью результата работы компьютера. Ввиду этого простейшее уравнение является подходящей моделью для анализа сходимости и устойчивости итеративных алгоритмов. Так как результат работы QA подчиняется распределению Больцмана, это позволяет проводить оценку работы алгоритмов на основе нормальных распределений и их модификаций.

Мы рассматриваем различные итеративные алгоритмы, которые работают как для большого (стремящегося к бесконечности) количества кубитов, так и для малого числа кубитов. Характерной чертой рассмотренных алгоритмов является адаптация размера поправки на каждом шаге в зависимости от текущего значения невязки, что аналогично подходам из [23], [26]–[28] и расширяет предложенные ранее подходы из [21], [22], [25]. Мы доказываем, что предложенные алгоритмы сходятся к точному значению при достаточно малых ошибках в квантовом компьютере, и оцениваем скорость сходимости (см. теоремы 1, 2). Для реализации алгоритмов адаптации используем сложение, умножение и домножение на целые степени двойки, что соответствует сдвигу битов, и не используем деление на произвольное число.

В разд. 2 мы даем предварительные определения, переводим задачу решения уравнения в термины модели, эквивалентной модели Изинга, и устанавливаем вероятностную модель вычислений. В разд. 3 рассматриваем итеративные алгоритмы, основанные на последовательном улучшении приближенных решений уравнения. В п. 3.1 мы рассматриваем идеализированный случай, при котором результат работы QA подчиняется закону нормального распределения. В п. 3.2 рассматриваем общий подход к изучению скорости сходимости итеративных алгоритмов. Случай усеченного нормального распределения решений, соответствующий бесконечному количеству кубитов, представлен в п. 3.3, случай распределения Больцмана, соответствующий конечному количеству кубитов, — в п. 3.4.

2. ПРЕДВАРИТЕЛЬНЫЕ ПОСТРОЕНИЯ

Модель Изинга эквивалентна так называемой модели QUBO (quadratic unconstrained binary optimization) (см. [29]):

$$H(q_1, \dots, q_n) = \sum_{i=1}^{n} \sum_{j=i}^{n} Q_{ij} q_i q_j, \tag{2}$$

где $(Q_{ij})_{i,j=1}^{n}$ — верхнетреугольная квадратная матрица порядка n, $q_i \in \{0, 1\}$. Для переформулировки задачи решения уравнения $ax = b$ в задачу QUBO мы используем функцию

$$H(x) = (ax - b)^2. \tag{3}$$

Переменную x представляем с конечной точностью в виде

$$x = \vartheta q_p + \sum_{i=r}^{p-1} 2^i q_i, \tag{4}$$

где $p, r \in \mathbb{Z}$, $r < p$, $\vartheta = -2^p + 2^r$, $q_i \in \{0, 1\}$. Бит q_p отвечает за знак переменной x. Роль константы ϑ заключается в том, что набор битов $q_r, \dots, q_{p-1}$ для отрицательных значений x представляет собой дополнительный код к аналогичному набору битов для положительных значений x. Отметим, что можно использовать и другие представления переменной (см. [22], [23], [28]). Обозначим через $\Omega_{r,p}$ множество чисел вида (4). Тогда

$$\Omega_{r,p} = \left\{ \pm \sum_{i=r}^{p-1} q_i 2^i \ : \ q_i \in \{0, 1\} \right\}.$$

После подстановки (4) в (3) и отбрасывания постоянного слагаемого получаем целевую функцию вида (2):

$$H(q_r,\dots,q_p)=\sum_{i=r}^{p}\sum_{j=i}^{p}Q_{ij}q_iq_j,$$

где

$$Q_{ii}=\begin{cases}a^2\vartheta^2-2ab\vartheta, & i=p,\\ 2^{2i}a^2-2^{i+1}ab,\ r\leqslant i\leqslant p-1,\end{cases}\qquad Q_{ij}=\begin{cases}2^{i+1}a^2\vartheta,\ \ j=p, r\leqslant i\leqslant p-1,\\ 2^{i+j+1}a^2,\ r\leqslant i<j\leqslant p-1.\end{cases}$$

Как было сказано ранее, ошибки в работе QA имеют вероятностный характер, и мы считаем, что они подчиняются распределению Больцмана. Будем использовать вариацию этого распределения, определенную ниже.

Определение 1. Пусть Ω — конечное подмножество вещественных чисел, $H:\mathbb{R}\to[0,+\infty)$, $\beta>0$. *Распределением Больцмана* $\mathrm{B}(\beta,\Omega,H(x))$ на множестве Ω с параметром β и целевой функцией $H(x)$ назовем вероятностное распределение на Ω, в котором вероятность элемента $x\in\Omega$ определяется как

$$P(x)=\frac{1}{Z}e^{-\beta^2H(x)},\quad \text{где } Z=\textstyle\sum_{y\in\Omega}e^{-\beta^2H(y)}.$$

Чем меньше значение $H(x)$, тем больше вероятность получить x в качестве результата работы компьютера. Параметр β отражает точность работы компьютера. Чем больше значение β, тем более вероятно результат работы компьютера будет близок к точке минимума функции $H(x)$ на множестве Ω. Если $\beta\to+\infty$, то компьютер будет работать без ошибок.

Вернемся к целевой функции (3). Если предположить, что количество кубитов в QA стремится к бесконечности, и для двоичного представления (4) переменной x мы используем как все положительные степени двойки, так и все отрицательные, то распределение решений будет стремиться к нормальному, что показывает следующее очевидное предложение.

Предложение 1. *Пусть* $r,p\in\mathbb{Z}$, $r<p$, $\Omega_{r,p}=\left\{\pm\sum\limits_{i=r}^{p-1}q_i\,2^i\ :\ q_i\in\{0,1\}\right\}$, $\beta>0$, $a,b\in\mathbb{R}$, $a\neq0$. *Тогда*

$$\lim_{\substack{p\to+\infty\\ r\to-\infty}}\mathrm{B}\left(\beta,\Omega_{r,p},(ax-b)^2\right)=\mathcal{N}\left(\frac{b}{a},\frac{1}{2a^2\beta^2}\right)$$

по распределению.

Таким образом, в качестве приближения к распределению Больцмана мы можем использовать нормальное распределение.

3. УЛУЧШЕНИЕ РЕШЕНИЯ УРАВНЕНИЯ

3.1. Модель улучшения решения, основанная на нормальном распределении

В этом пункте мы будем предполагать, что результат работы QA по решению уравнения $ax=b$ имеет распределение $\mathcal{N}\left(\frac{b}{a},\frac{1}{2a^2\beta^2}\right)$ согласно предложению 1. Ниже рассмотрим, как будут распределены ошибки приближения к решению, если мы будем итерировать алгоритм. Пусть x_n — n-ое фиксированное приближение к решению уравнения $ax=b$. Точное решение x мы можем представить как сумму x_n и поправки: $x=x_n+\Delta_n$. Подставляя ее в исходное уравнение, мы получаем уравнение относительно поправки Δ_n:

$$a\Delta_n=b-ax_n. \tag{5}$$

Пусть целое число l_n такое, что

$$\frac{1}{2^{l_n+1}}<|b-ax_n|\leqslant\frac{1}{2^{l_n}}. \tag{6}$$

Вместо уравнения (5) будем решать на QA уравнение

$$a\widetilde{\Delta}_n=2^{l_n}(b-ax_n) \tag{7}$$

с неизвестной $\widetilde{\Delta}_n$. По предположению имеем, что

$$\widetilde{\Delta}_n\sim\mathcal{N}\left(\frac{2^{l_n}(b-ax_n)}{a},\frac{1}{2a^2\beta^2}\right). \tag{8}$$

Так как $\Delta_n = \frac{1}{2^{l_n}}\widetilde{\Delta}_n$, то по свойствам нормального распределения получаем, что

$\Delta_n \sim \mathcal{N}\left(\frac{b}{a} - x_n, \frac{1}{2^{2l_n}} \cdot \frac{1}{2a^2\beta^2}\right)$.

Пусть $\xi_n \sim \mathcal{N}(0,1)$, тогда

$$\Delta_n \stackrel{d}{=} \frac{b}{a} - x_n + \frac{\xi_n}{2^{l_n}\sqrt{2}a\beta}.$$

Следующее приближение к решению будем вычислять по формуле

$$x_{n+1} = x_n + \Delta_n.$$

Заметим, что следующая поправка Δ_{n+1} и число l_{n+1} зависят от предыдущей поправки Δ_n. Следующее предложение позволяет построить последовательность приближений x_n, учитывая эти зависимости.

Предложение 2. *Пусть $a, b \in \mathbb{R}\backslash\{0\}$, $\sigma > 0$. Пусть $\xi_n \sim \mathcal{N}(0,1)$, $n \geqslant 0$, — независимые в совокупности случайные величины. Построим последовательности случайных величин x_n, x'_n l_n, l'_n, Δ_n, Δ'_n при $n \geqslant 0$. Положим $x_0 = x'_0 = = l_0 = l'_0 = 0$. Дадим дальнейшие определения при $n \geqslant 0$. Пусть условное распределение случайной величины Δ_n при условии x_n равно $\mathcal{N}\left(\frac{b}{a} - x_n, \frac{\sigma^2}{2^{2l_n}}\right)$. Положим*

$$\Delta'_n = \frac{b}{a} - x'_n + \frac{\sigma\xi_n}{2^{l'_n}},$$

$$x_{n+1} = x_n + \Delta_n,$$

$$x'_{n+1} = x'_n + \Delta'_n.$$

Определим l_{n+1} и l'_{n+1} следующим образом: $l_{n+1}, l'_{n+1} \in \mathbb{Z}$ и

$$2^{l_{n+1}}|b - ax_{n+1}|,\ 2^{l'_{n+1}}|b - ax'_{n+1}| \in (1/2, 1]\,.$$

Тогда $(\Delta_0, \ldots, \Delta_n) \stackrel{d}{=} (\Delta'_0, \ldots, \Delta'_n)$ для любого $n \geqslant 0$.

Доказательство. Доказательство будем вести индукцией по n. При $n = 0$ имеем $\Delta_0 \sim \mathcal{N}\left(\frac{b}{a}, \sigma^2\right)$ и $\Delta'_0 = = \frac{b}{a} + \sigma\xi_0 \sim \mathcal{N}\left(\frac{b}{a}, \sigma^2\right)$. Значит, $\Delta_0 \stackrel{d}{=} \Delta'_0$.

Пусть $(\Delta_0, \ldots, \Delta_{n-1}) \stackrel{d}{=} \left(\Delta'_0, \ldots, \Delta'_{n-1}\right)$, то есть для любого борелевского множества $A \subset \mathbb{R}^n$ выполнено $P\left((\Delta_0, \ldots, \Delta_{n-1}) \in A\right) = P\left(\left(\Delta'_0, \ldots, \Delta'_{n-1}\right) \in A\right)$. Покажем, что тогда для любого борелевского множества $A \subset \mathbb{R}^{n+1}$ будет выполнено $P\left((\Delta_0, \ldots, \Delta_n) \in A\right) = P\left((\Delta'_0, \ldots, \Delta'_n) \in A\right)$. Достаточно доказать, что

$$P\left((\Delta_0, \ldots, \Delta_n) \in A_0 \times \cdots \times A_n\right) = P\left((\Delta'_0, \ldots, \Delta'_n) \in A_0 \times \cdots \times A_n\right)$$

для любых борелевских $A_i \subset \mathbb{R}$, $i = 0, 1, \ldots, n$. По формуле полной вероятности получаем, что вероятность $P\left((\Delta'_0, \ldots, \Delta'_n) \in A_0 \times \cdots \times A_n\right)$ равна

$$\int_{\mathbb{R}^n} P\left((\Delta'_0, \ldots, \Delta'_n) \in A_0 \times \cdots \times A_n \,|\, \left(\Delta'_0, \ldots, \Delta'_{n-1}\right) = (r_0, \ldots, r_{n-1})\right) dP'_{n-1},$$

где P'_{n-1} — распределение вектора $\left(\Delta'_0, \ldots, \Delta'_{n-1}\right)$ и интегрирование ведется по переменным $r_0, \ldots, r_{n-1}$. Если $(r_0, \ldots, r_{n-1}) \notin A_0 \times \cdots \times A_{n-1}$, то $(\Delta'_0, \ldots, \Delta'_n) \notin A_0 \times \cdots \times A_n$, и вероятность под знаком интеграла равна нулю. Поэтому последний интеграл равен

$$\int\limits_{A_0 \times \cdots \times A_{n-1}} P\left(\Delta'_n \in A_n \,|\, \Delta'_0 = r_0, \ldots, \Delta'_{n-1} = r_{n-1}\right) dP'_{n-1}. \tag{9}$$

Зафиксируем числа $r_0, \ldots, r_{n-1}$. Положим $S = r_0 + \cdots + r_{n-1}$ и возьмем целое l таким, что $2^l|b - aS| \in (1/2, 1]$. Покажем, что условное распределение случайной величины Δ'_n при условии $\Delta'_0 = r_0, \ldots, \Delta'_{n-1} = r_{n-1}$ равно условному распределению случайной величины Δ_n при условии $\Delta_0 = r_0, \ldots, \Delta_{n-1} = r_{n-1}$.

При этих условиях $x_n = S$ и $x'_n = S$. Тогда по определению l'_n получим, что $l'_n = l$. По определению Δ'_n имеем

$$\Delta'_n = \frac{b}{a} - S + \frac{\sigma\xi_n}{2^l} \sim \mathcal{N}\left(\frac{b}{a} - S, \frac{\sigma^2}{2^{2l}}\right),$$

что есть условное распределение случайной величины Δ_n при условии $x_n = S$.

По индукционному предположению мы имеем $P'_{n-1} = P_{n-1}$, где P_{n-1} — распределение вектора $(\Delta_0, \ldots, \Delta_{n-1})$. Значит, интеграл (9) равен

$$\int\limits_{A_0 \times \cdots \times A_{n-1}} P\left(\Delta_n \in A_n \,|\, \Delta_0 = r_0, \ldots, \Delta_{n-1} = r_{n-1}\right) dP_{n-1} =$$

$$= \int\limits_{\mathbb{R}^n} P\left((\Delta_0, \ldots, \Delta_n) \in A_0 \times \cdots \times A_n \,|\, (\Delta_0, \ldots, \Delta_{n-1}) = (r_0, \ldots, r_{n-1})\right) dP_{n-1}.$$

Последний интеграл равен вероятности $P\left((\Delta_0, \ldots, \Delta_n) \in A_0 \times \cdots \times A_n\right)$, что и требовалось показать.

Полагая, что $\sigma^2 = \frac{1}{2a^2\beta^2}$ в предложении 2, получаем, что $x'_{n+1} = \frac{b}{a} + \frac{\xi_n}{2^{l'_n}\sqrt{2}a\beta}$.

Следующая теорема показывает, что эта последовательность при определенных условиях сходится к решению уравнения $ax = b$.

Теорема 1. *Пусть $a, b \in \mathbb{R}$, $a \neq 0$, $\beta > 0$, γ — постоянная Эйлера–Маскерони. Зафиксируем последовательность независимых в совокупности случайных величин $\xi_n \sim \mathcal{N}(0,1)$, $n \geqslant 0$. Построим последовательности случайных величин l_n и x_n по правилу $l_0 = 0$, $x_0 = 0$, и при $n \geqslant 0$ положим $x_{n+1} = \frac{b}{a} + \frac{\xi_n}{2^{l_n}\sqrt{2}a\beta}$, целое число l_{n+1} таково, что $2^{l_{n+1}}|b - ax_{n+1}| \in \left(\frac{1}{2}, 1\right]$. Тогда*

1) *если $s \in \left[1, \beta e^{\gamma/2}\right)$, то $s^n\left(x_n - \frac{b}{a}\right) \xrightarrow[n\to\infty]{\text{п.н.}} 0$,*

2) *если $s > 2\beta e^{\gamma/2}$, то $s^n\left(x_n - \frac{b}{a}\right) \xrightarrow[n\to\infty]{\text{п.н.}} \infty$,*

3) *если $\beta < \frac{1}{2}e^{-\gamma/2}$, то $x_n \xrightarrow[n\to\infty]{\text{п.н.}} \infty$.*

Замечание 1. Процесс, описанный в теореме, соответствует процессу работы квантового компьютера. Процесс последовательного улучшения решения описывался в статьях [23], [26]–[28], однако теоретические вопросы сходимости подобных итеративных алгоритмов не исследовались.

Из п. 1) теоремы 1 следует, что если $\beta > e^{-\gamma/2} \approx \frac{3}{4}$, то последовательность x_n будет сходиться к решению уравнения $ax = b$ почти наверное и притом с экспоненциальной скоростью. Пункт 2) устанавливает верхнюю границу скорости сходимости. Пункт 3) устанавливает достаточное условие расходимости последовательности x_n: если β мало (точность работы QA слишком плоха), то последовательность x_n будет расходиться.

Для доказательства теоремы 1 нам понадобится следующая лемма, непосредственно следующая из усиленного закона больших чисел Колмогорова (см. [30]).

Лемма 1. *Пусть $\delta > 0$, и пусть $(X_i)_{i=1}^{\infty}$ — независимые в совокупности случайные величины такие, что для любого натурального i существуют $\mathbb{E}\ln|X_i|$ и $\mathbb{E}\ln^2|X_i|$. Пусть дисперсии $\operatorname{Var}(\ln|X_i|)$ ограничены в совокупности. Тогда*

1) *если $\mathbb{E}\ln|X_i| < -\delta$ для любого i, то $X_1 \cdots X_n \xrightarrow[n\to\infty]{\text{п.н.}} 0$,*

2) *если $\mathbb{E}\ln|X_i| > \delta$ для любого i, то $X_1 \cdots X_n \xrightarrow[n\to\infty]{\text{п.н.}} \infty$.*

Доказательство теоремы 1. Проведем предварительные построения. Рассмотрим случайную величину $z_n = \frac{b}{a} - x_n$. Оценим сверху $|z_{n+1}|$, используя определение случайных величин x_n и l_n:

$$|z_{n+1}| = \left|\frac{\xi_n}{2^{l_n}\sqrt{2}a\beta}\right| = \left|\frac{(b - ax_n)\xi_n}{2^{l_n}(b - ax_n)\sqrt{2}a\beta}\right| < \left|\frac{\sqrt{2}(b - ax_n)\xi_n}{a\beta}\right| = \frac{\sqrt{2}}{\beta}|z_n\xi_n|.$$

Так как $|z_1| = \frac{|\xi_0|}{\sqrt{2}a\beta}$, то

$$|z_{n+1}| < \left|\frac{1}{2a}\right|\left(\frac{\sqrt{2}}{\beta}\right)^{n+1}|\xi_0 \cdots \xi_n|. \tag{10}$$

Аналогично получаем оценку $|z_{n+1}|$ снизу:

$$|z_{n+1}| \geqslant \left|\frac{1}{a}\right| \left(\frac{1}{\sqrt{2}\beta}\right)^{n+1} |\xi_0 \cdots \xi_n|. \tag{11}$$

Найдем математическое ожидание случайной величины $\ln|\xi_0|$:

$$\mathbb{E}\ln|\xi_0| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \ln|x|\, dx = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} \ln\left(\sqrt{2}x\right) dx = \ln\frac{1}{\sqrt{2}e^{\gamma/2}},$$

где мы воспользовались соотношением для γ из [31]:

$$\int_0^{\infty} e^{-x^2} \ln x\, dx = -\frac{\sqrt{\pi}}{4}(\gamma + \ln 4).$$

Тогда

$$\mathbb{E}\ln\frac{\sqrt{2}|\xi_0|}{\beta} = \ln\frac{1}{\beta e^{\gamma/2}}, \quad \mathbb{E}\ln\frac{|\xi_0|}{\sqrt{2}\beta} = \ln\frac{1}{2\beta e^{\gamma/2}}. \tag{12}$$

Докажем п. 1). Достаточно показать, что $\left(\frac{s\sqrt{2}}{\beta}\right)^n |\xi_0 \cdots \xi_{n-1}| \xrightarrow[n\to\infty]{\text{п.н.}} 0$, исходя из неравенства (10). Ввиду (12) и того, что $s \in \left[1, \beta e^{\gamma/2}\right)$, имеем

$$\mathbb{E}\ln\frac{s\sqrt{2}|\xi_0|}{\beta} = \ln\frac{1}{\beta e^{\gamma/2}} + \ln s < 0.$$

Значит, по лемме 1 произведение $\left(\frac{s\sqrt{2}}{\beta}\right)^n |\xi_0 \cdots \xi_{n-1}|$ сходится к нулю почти наверное.

Докажем п. 2). Достаточно показать, что $\left(\frac{s}{\sqrt{2}\beta}\right)^n |\xi_0 \cdots \xi_{n-1}| \xrightarrow[n\to\infty]{\text{п.н.}} \infty$, исходя из неравенства (11). Ввиду (12) и того, что $s > 2\beta e^{\gamma/2}$, имеем

$$\mathbb{E}\ln\frac{s|\xi_0|}{\sqrt{2}\beta} = \ln\frac{1}{2\beta e^{\gamma/2}} + \ln s > 0.$$

Значит, по лемме 1 произведение $\left(\frac{s}{\sqrt{2}\beta}\right)^n |\xi_0 \cdots \xi_{n-1}|$ сходится к бесконечности почти наверное.

Докажем п. 3). Достаточно показать, что $\left(\frac{1}{\sqrt{2}\beta}\right)^n |\xi_0 \cdots \xi_{n-1}| \xrightarrow[n\to\infty]{\text{п.н.}} \infty$, исходя из неравенства (11). Ввиду (12) и того, что $\beta < \frac{1}{2}e^{-\gamma/2}$, имеем $\mathbb{E}\ln\frac{|\xi_0|}{\sqrt{2}\beta} = \ln\frac{1}{2\beta e^{\gamma/2}} > 0$. Значит, по лемме 1 произведение $\left(\frac{1}{\sqrt{2}\beta}\right)^n |\xi_0 \cdots \xi_{n-1}|$ сходится к бесконечности почти наверное.

3.2. Общий подход к исследованию сходимости алгоритмов решения линейного уравнения

В этом пункте рассмотрим общий подход к построению и исследованию сходимости алгоритма, решающего уравнение $ax = b$. Мы рассмотрим общую схему последовательных приближений и докажем теорему 2, позволяющую оценить скорость сходимости подобных алгоритмов. Домножая a и b на одинаковую степень двойки, можно добиться выполнения неравенства

$$1/2 \leqslant a < 1. \tag{13}$$

В дальнейшем в статье будем предполагать, что выполнено неравенство (13).

Вернемся к построению последовательности приближений к решению уравнения $ax = b$. Имея очередное фиксированное приближение x_n, решаем на QA уравнение (7) относительно $\widetilde{\Delta}_n$, где целое l_n выбирается в соответствии с (6). Следующее приближение вычисляется как

$$x_{n+1} = x_n + 2^{-l_n}\widetilde{\Delta}_n. \tag{14}$$

Распределение случайной величины $\widetilde{\Delta}_n$ зависит от выбранного нами алгоритма и определяется значениями c_n, $\operatorname{sign}(b - ax_n)$, a, β, где

$$c_n = \frac{1}{2^{l_n}|b - ax_n|}. \tag{15}$$

Заметим, что $c_n \in [1, 2)$. В этом пункте мы не фиксируем конкретный алгоритм и соответствующее распределение $\widetilde{\Delta}_n$. В дальнейшем будем предполагать, что зависимость от $\operatorname{sign}(b - ax_n)$ имеет специальный вид, а именно,

существует такая функция $q(u, c, a, \beta)$, что если η — случайная величина, равномерно распределенная на $[0, 1]$, то

$$\widetilde{\Delta}_n \,|\, x_n \stackrel{d}{=} \operatorname{sign}(b - ax_n) q(\eta, c_n, a, \beta) \,|\, x_n. \tag{16}$$

Условие (16) означает, что распределение поправок $\widetilde{\Delta}_n$ в случае положительных и отрицательных невязок отличается лишь знаком.

Заметим, что если функция $q(\cdot, c_n, a, \beta)$ равна обратной функции распределения закона $\mathcal{N}\left(\frac{1}{ac_n}, \frac{1}{2a^2\beta^2}\right)$, то $q(\eta, c_n, a, \beta) \sim \mathcal{N}\left(\frac{1}{ac_n}, \frac{1}{2a^2\beta^2}\right)$, и поправка $\widetilde{\Delta}_n$ распределена, как в (8).

По аналогии с предложением 2 верно следующее

Предложение 3. *Пусть $b \neq 0$, $\beta > 0$. Пусть η_n, $n \geqslant 0$, — независимые в совокупности случайные величины, равномерно распределенные на промежутке $[0, 1]$. Зафиксируем функцию $q(\eta, c, a, \beta)$, определяемую выбранным алгоритмом последовательных приближений.*

Введем функции

$$G_1(u_0) = q(u_0, 1/b, a, \beta) - \frac{b}{a},$$

$$G_{n+1}(u_0, \ldots, u_n) = G_n(u_0, \ldots, u_{n-1}) \left(1 - ac'_n q(u_n, c'_n, a, \beta)\right), \; n \geqslant 1, \tag{17}$$

где $u_i \in [0, 1]$, $c'_n = \frac{1}{2^{l'_n}|aG_n|}$, и целое l'_n выбрано так, что $2^{l'_n}|aG_n| \in \left(\frac{1}{2}, 1\right]$.

Пусть $x_0 = l_0 = 0$. Пусть $\widetilde{\Delta}_0 \stackrel{d}{=} q(\eta_0, 1/b, a, \beta)$ и при $n \geqslant 1$ выполнено (16) *для $\eta = \eta_n$, где l_n и c_n определяются в* (6) *и* (15). *При $n \geqslant 0$ положим*

$$x_{n+1} = x_n + 2^{-l_n} \widetilde{\Delta}_n,$$

$$x'_{n+1} = \frac{b}{a} + G_{n+1}(\eta_0, \ldots, \eta_n). \tag{18}$$

Тогда $(x_1, \ldots, x_n) \stackrel{d}{=} (x'_1, \ldots, x'_n)$ для любого $n \geqslant 1$.

Здесь и далее будем рассматривать последовательность x_n, заданную равенством (18). Последовательность x_n определяется выбором функции q. В следующей теореме мы будем использовать обозначения и определения из предложения 3.

Теорема 2. *Введем функцию*

$$r(u, a, \beta) = \max_{c \in [1,2]} |1 - c \cdot a \cdot q(u, c, a, \beta)|. \tag{19}$$

Пусть $E(a, \beta)$ — математическое ожидание случайной величины $\ln r(\eta, a, \beta)$, где η — случайная величина, равномерно распределенная на $[0, 1]$:

$$E(a, \beta) = \int_0^1 \ln r(u, a, \beta) du.$$

Тогда

1) *если $E(a, \beta) < 0$, то $x_n \xrightarrow[n\to\infty]{\text{п.н.}} \frac{b}{a}$,*

2) *если $\ln s + E(a, \beta) < 0$, то $s^n \left(x_n - \frac{b}{a}\right) \xrightarrow[n\to\infty]{\text{п.н.}} 0$.*

Доказательство. Докажем п. 1). Исходя из (18), достаточно показать, что $G_{n+1}(\eta_0, \ldots, \eta_n) \xrightarrow{\text{п.н.}} 0$. Используя (17), получаем

$$|G_{n+1}(\eta_0, \ldots, \eta_n)| = |G_1(\eta_0)| \prod_{i=1}^{n} |1 - a\, c_i\, q(\eta_i, c_i, a, \beta)| \leqslant |G_1(\eta_0)| \prod_{i=1}^{n} r(\eta_i, a, \beta).$$

Так как математические ожидания случайных величин $\ln r(\eta_i, a, \beta)$ меньше нуля, то по лемме 1 получаем требуемое.

Докажем п. 2). Достаточно показать, что $s^{n+1} G_{n+1}(\eta_0, \ldots, \eta_n) \xrightarrow{\text{п.н.}} 0$. Аналогично получаем

$$s^{n+1} |G_{n+1}(\eta_0, \ldots, \eta_n)| \leqslant s\, |G_1(\eta_0)| \prod_{i=1}^{n} \left(s\, r(\eta_i, a, \beta)\right).$$

Из неравенства $\mathbb{E}\ln(s\,r(\eta_i,a,\beta))<0$ следует требуемое.

Замечание 2. В теореме 2 условие сходимости алгоритма зависит от a и β, в то время как в теореме 1 условие сходимости зависит только от β. Это порождает вопрос: от чего может зависеть скорость сходимости аналогичных итеративных алгоритмов для систем линейных уравнений – размер матрицы, число обусловленности и др.? Рассмотрение систем линейных уравнений выходит за рамки работы и нуждается в отдельном исследовании.

3.3. Модели вычислений, основанные на усеченном нормальном распределении

В разд. 2 и п. 3.1 мы рассматривали представление (4) переменной x по положительным и отрицательным степеням двойки. В текущем пункте мы рассмотрим более "гибкое" представление:

$$x=(d_2-d_1)\sum_{i=1}^{r}q_i2^{-i}+d_1, \tag{20}$$

где $q_i\in\{0,1\}$, $d_1<d_2$, $r\in\mathbb{N}$. В таком представлении x принимает значения в промежутке $[d_1,d_2)$. Коэффициенты Q_{ij} в модели QUBO (2) находятся из подстановки представления (20) в функцию (3).

Мы будем рассматривать несколько различных алгоритмов поиска поправки, отличающихся значениями d_1, d_2. На каждом шаге будем искать поправку $\widetilde{\Delta}_n$ в шаге (14) алгоритма как решение уравнения (7) на промежутке $\operatorname{sign}(b-ax_n)[d_1,d_2]$, представляя $\widetilde{\Delta}_n$ по аналогии с формулой (20):

$$\widetilde{\Delta}_n=\operatorname{sign}(b-ax_n)\left((d_2-d_1)\sum_{i=1}^{r}q_i2^{-i}+d_1\right). \tag{21}$$

Таким образом, числа d_1,d_2 определяют алгоритм и могут влиять на характер его сходимости. В дальнейшем рассмотрим четыре алгоритма с различными значениями d_1,d_2.

Предположим, что количество кубитов в QA стремится к бесконечности, т.е. $r\to\infty$. Посмотрим, как при этом ведет себя распределение Больцмана на множестве, состоящем из чисел вида (20), с целевой функцией $(ax-b)^2$. Для начала введем следующее определение.

Определение 2. Пусть $\sigma>0$, $\mu,d_1,d_2\in\mathbb{R}$, $d_1<d_2$. *Усеченное нормальное распределение* $\mathcal{N}(\mu,\sigma^2,d_1,d_2)$ — это распределение с функцией плотности

$$f(t)\propto e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathbf{1}_{(d_1,d_2)}(t).$$

Если $d_1>d_2$, то обозначение $\mathcal{N}(\mu,\sigma^2,d_1,d_2)$ будет пониматься как усеченное нормальное распределение $\mathcal{N}(\mu,\sigma^2,d_2,d_1)$, а обозначение (d_1,d_2) будет пониматься как интервал (d_2,d_1).

Предложение 4. *Пусть* $\beta>0$, $a,b\in\mathbb{R}$, $a\neq0$, $d_1<d_2$, $r\in\mathbb{N}$. *Тогда*

$$\mathrm{B}\left(\beta,\left\{(d_2-d_1)\sum_{i=1}^{r}q_i2^{-i}+d_1\ :\ q_i\in\{0,1\}\right\},(ax-b)^2\right)\xrightarrow[r\to\infty]{d}\mathcal{N}\left(\frac{b}{a},\frac{1}{2a^2\beta^2},d_1,d_2\right).$$

Таким образом, мы можем использовать усеченное нормальное распределение в качестве приближения к распределению Больцмана.

Согласно представлению (21) и предложению 4 будем считать, что

$$\widetilde{\Delta}_n\sim\operatorname{sign}(b-ax_n)\mathcal{N}\left(\frac{1}{ac_n},\frac{1}{2a^2\beta^2},d_1,d_2\right).$$

Обозначим через $q(u,c,a,\beta)$ такие функции, что если c,a,β фиксированы и η равномерно распределена на промежутке $[0,1]$, то

$$q(\eta,c,a,\beta)\sim\mathcal{N}\left(\frac{1}{ac},\frac{1}{2a^2\beta^2},d_1,d_2\right).$$

Функция q представляет собой обратную функцию распределения соответствующего закона распределения и в явном виде записывается следующим образом:

$$q(u,c,a,\beta)=\frac{1}{ac}+\frac{1}{a\beta}\operatorname{erf}^{-1}\left((1-u)\operatorname{erf}\left(d_1a\beta+\frac{\beta}{c}\right)+u\operatorname{erf}\left(d_2a\beta-\frac{\beta}{c}\right)\right),$$

где $\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ — функция ошибок.

Из неравенств (6),(13) следует, что точное решение уравнения (7) удовлетворяет неравенствам

$$|\widetilde{\Delta}_n| \leqslant 2, \quad \operatorname{sign}(b - ax_n)\widetilde{\Delta}_n \in [1/2, 2]. \tag{22}$$

Числа d_1, d_2 определяют функцию q и соответственно задают алгоритм улучшения решения. Ниже приводятся четыре алгоритма для соответствующих значений d_1, d_2, три из них учитывают правильный знак поправки на каждом шаге. Ввиду (22) мы ограничиваемся значениями d_1, d_2 такими, что $|d_1|, |d_2| \leqslant 2$.

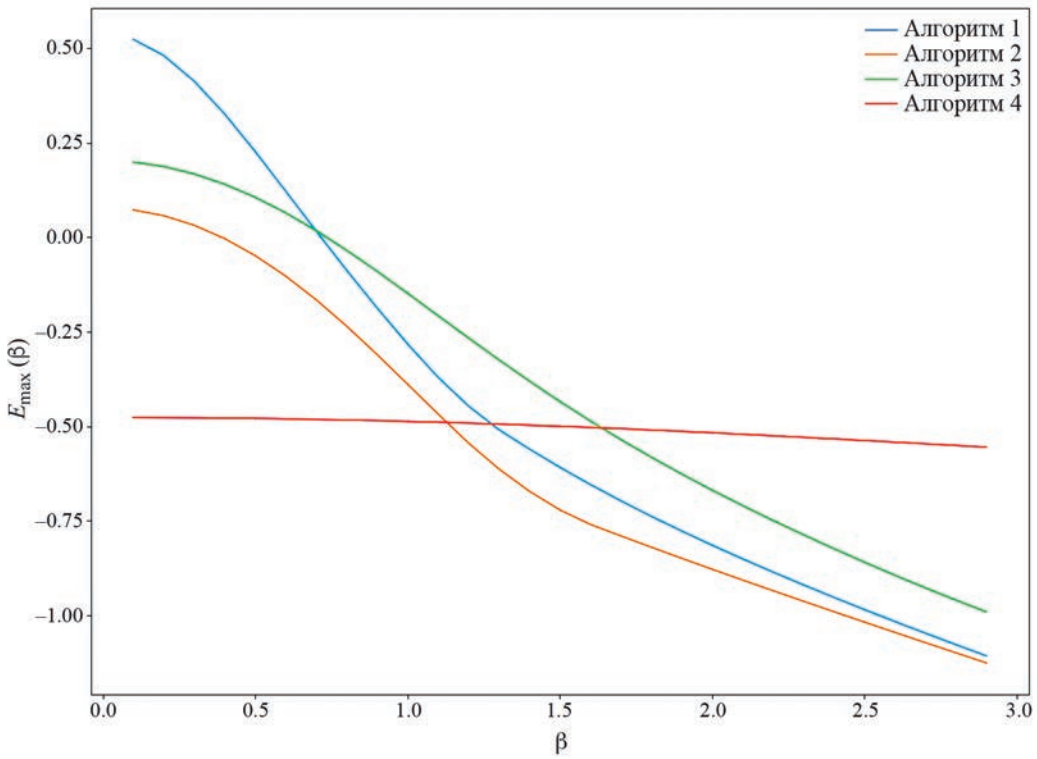
Алгоритм 1: $d_1 = -2$, $d_2 = 2$. Не учитывает знак поправки, а лишь наибольшее значение модуля $|\widetilde{\Delta}_n|$.

Алгоритм 2: $d_1 = 0$, $d_2 = 2$. Учитывает знак поправки и наибольшее значение модуля $|\widetilde{\Delta}_n|$.

Алгоритм 3: $d_1 = 1/2$, $d_2 = 2$. Учитывает знак поправки и наибольшее и наименьшее значения модуля $|\widetilde{\Delta}_n|$.

Алгоритм 4: $d_1 = 1/2$, $d_2 = 1$. Консервативный алгоритм, при котором гарантированно выполняется неравенство $r(u, a, \beta) \leq 1$, где r определено в (19). По теореме 2 такой алгоритм сходится для любых a, β. Отметим, что при этом точное значение $\frac{1}{ac_n}$ поправки $\widetilde{\Delta}_n$ не всегда лежит в интервале $[d_1, d_2]$.

На фиг. 1 приведены сравнительные графики для соответствующих функций $E_{\max}(\beta) = \max\limits_{a\in[1/2,1]} E(a, \beta)$, дающие пессимистичную оценку на скорость сходимости алгоритма. Если $E_{\max}(\beta) < 0$, то алгоритм сходится при любых $a \in [0.5, 1)$, $b \in \mathbb{R}$, чем меньше значение $E_{\max}(\beta)$, тем сходимость быстрее.



Фиг. 1. Графики функций $E_{\max}(\beta)$ для различных алгоритмов.

Поскольку алгоритм 4 при любом выборе $\widetilde{\Delta}_n$ уменьшает значение невязки, то он сходится в любом случае, что отражено в отрицательности функции $E_{\max}(\beta)$ для всех значений β. При этом, поскольку включение $\frac{1}{ac} \in [0.5, 1]$ не всегда выполнено, алгоритм показывает не быструю скорость сходимости даже при больших значениях β. Алгоритмы 1–3 ведут себя приблизительно одинаково при больших значениях β, это объясняется высокой вероятностью получить значение $\widetilde{\Delta}_n$, близкое к точному решению уравнения (7). Лучшие показатели сходимости наблюдаются у алгоритма 2, учитывающего знак, но разрешающего малые значения поправки на каждом шаге. Его преимущество над алгоритмом 3 вероятно объясняется уменьшением веса хвоста усеченного нормального распределения при котором $r(u, a, \beta) > 1$.

3.4. Модели вычислений, основанные на распределении Больцмана

В пп. 3.1, 3.3 мы рассматривали непрерывные распределения, приближающие распределение Больцмана. Здесь мы непосредственно рассмотрим модель вычислений, основанную на распределении Больцмана, и будем считать, что количество кубитов в QA конечно.

Будем рассматривать несколько различных алгоритмов поиска поправки $\widetilde{\Delta}_n$ в шаге (14) алгоритма, отличающихся представлениями переменной $\widetilde{\Delta}_n$ в уравнении (7). Используемое представление переменной определяет алгоритм и может влиять на характер его сходимости.

В одной группе алгоритмов мы не будем учитывать правильный знак поправки и будем искать решение уравнения (7), представляя $\widetilde{\Delta}_n$ как в формуле (4):

$$\widetilde{\Delta}_n = \operatorname{sign}(b - ax_n)\left(\vartheta q_p + \sum_{i=r}^{p-1} 2^i q_i\right), \tag{23}$$

где $r, p \in \mathbb{Z}$, $r < p$, $\vartheta = -2^p + 2^r$.

В другой группе алгоритмов будем учитывать знак поправки и искать $\widetilde{\Delta}_n$, представляя его как

$$\widetilde{\Delta}_n = \operatorname{sign}(b - ax_n)\sum_{i=r}^{p-1} 2^i q_i. \tag{24}$$

Заметим, что количество n_q участвующих в представлении $\widetilde{\Delta}_n$ кубитов в группе алгоритмов, не учитывающих знак, равно $p - r + 1$, а в группе, учитывающих знак, равно $p - r$.

Обозначим

$$\Omega^{\pm}_{r,p} = \left\{ \pm \sum_{i=r}^{p-1} q_i 2^i \ : \ q_i \in \{0,1\} \right\}, \ \ \Omega^{+}_{r,p} = \Omega_{r,p} \cap [0, \infty).$$

Таким образом, алгоритм определяется выбором $\Omega^{\pm}_{r,p}$ или $\Omega^{+}_{r,p}$ в качестве множества, на котором ищется поправка $\widetilde{\Delta}_n$. Ввиду (22) мы будем ограничиваться значениями $p \leqslant 1$.

Распределение поправки на n-м шаге задается соотношением

$$\widetilde{\Delta}_n \sim \operatorname{sign}(b - ax_n)\mathrm{B}\left(\beta, \Omega_{r,p}, \left(ax - \frac{1}{c_n}\right)^2\right),$$

где $\Omega_{r,p}$ равно либо $\Omega^{\pm}_{r,p}$, либо $\Omega^{+}_{r,p}$.

Обозначим через $q_{r,p}(u, c, a, \beta)$, такие функции, что если c, a, β фиксированы и η равномерно распределена на промежутке $[0, 1]$, то

$$q_{r,p}(\eta, c, a, \beta) \sim \mathrm{B}\left(\beta, \Omega_{r,p}, \left(ax - \frac{1}{c}\right)^2\right),$$

где $\Omega_{r,p}$ определяется в соответствии с выбранным алгоритмом. Функции $q_{r,p}$ представляют собой обратные функции распределения соответствующих законов распределения и определяются как

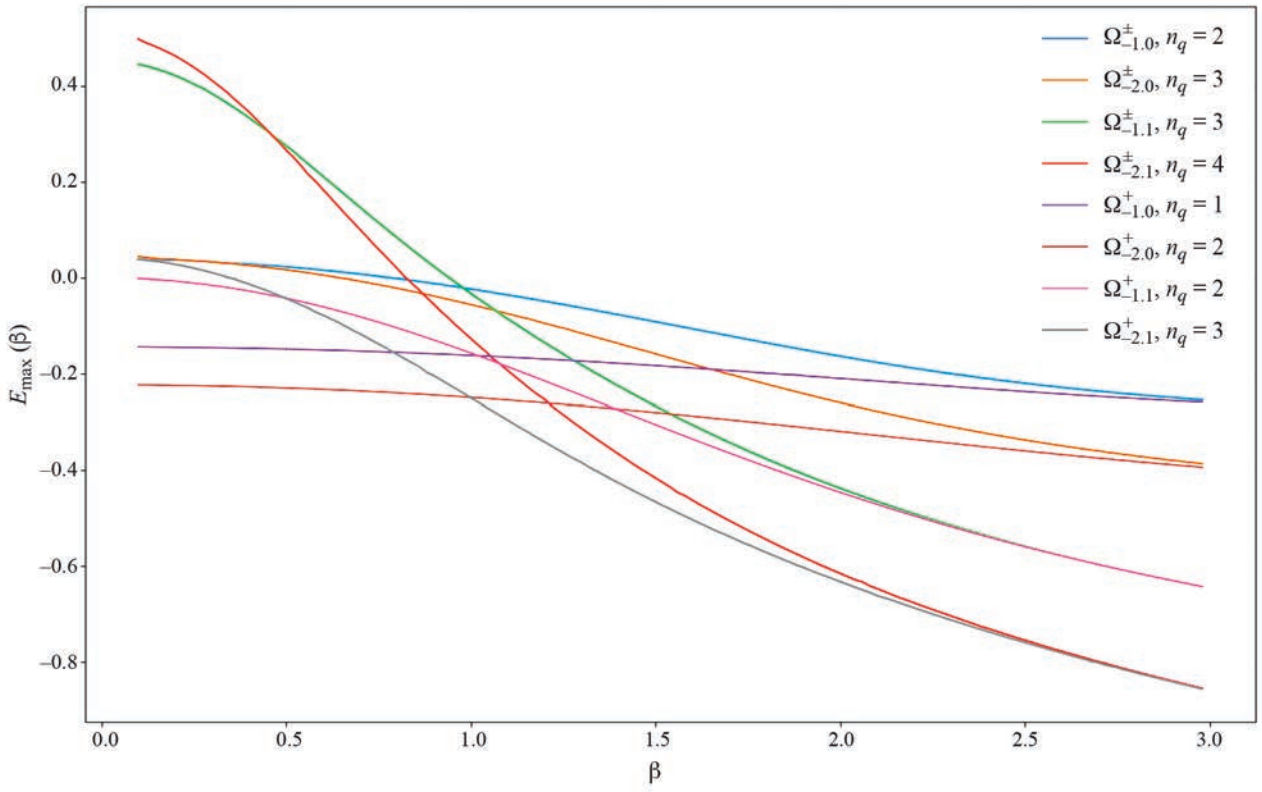
$$q_{r,p}(u, c, a, \beta) = \inf\left\{t \,|\, F_{r,p}(t) \geqslant u\right\}, \ u \in (0, 1],$$

где $F_{r,p}(t)$ — функция распределения закона $\mathrm{B}\left(\beta, \Omega_{r,p}, \left(ax - \frac{1}{c}\right)^2\right)$. Отметим, что функция $q_{r,p}(\cdot, c, a, \beta)$ кусочно-постоянная. Дополнительно определим

$$q_{r,p}(0, c, a, \beta) = \lim_{u \to 0+} q_{r,p}(u, c, a, \beta).$$

На фиг. 2 приведены сравнительные графики для соответствующих функций $E_{\max}(\beta) = \max\limits_{a \in [1/2, 1]} E(a, \beta)$, дающие пессимистичную оценку на скорость сходимости алгоритма. Если $E_{\max}(\beta) < 0$, то алгоритм сходится при любых $a \in [0.5, 1)$, $b \in \mathbb{R}$, чем меньше значение $E_{\max}(\beta)$, тем сходимость быстрее.

Из графиков следует, что при достаточно больших β сходятся все методы, включая основанные на одном кубите. Методы, учитывающие знак поправки, в которых $\Omega_{r,p} = \Omega^{+}_{r,p}$, сходятся быстрее, чем не учитывающие знак поправки, в которых $\Omega_{r,p} = \Omega^{\pm}_{r,p}$. При этом методы, учитывающие знак поправки используют меньшее



Фиг. 2. Графики функций $E_{\max}(\beta)$ для различных $\Omega_{r,p}$. Число n_q — количество кубитов, кодирующих поправку.

количество кубитов. Ожидаемо, с увеличением количества используемых кубитов скорость сходимости возрастает, но остается ниже, чем предельная скорость, соответствующая усеченному нормальному распределению на фиг. 1. По аналогии со сравнением алгоритмов 2 и 3 из п. 3.3 отметим, что методы с $p = 1$ включают больше значений поправок, при которых точность может ухудшиться, но при этом гарантированно содержат наилучшую возможную поправку, в то время как методы с $p = 0$ гарантированно не ухудшают точность приближения на каждом шаге, но при этом имеют меньшую вероятность для оптимальной поправки. Как и в случае бесконечного количества кубитов, при больших β методы с $p = 1$ оказываются более эффективными, чем при $p = 0$.

4. ВЫВОДЫ

В статье рассмотрены несколько адаптивных итеративных методов для поиска корня линейного уравнения $ax = b$ при помощи устройства, работающего по принципу квантового отжига. Результат работы QA моделируется распределением Больцмана. Для широкого класса алгоритмов предложен метод доказательства их сходимости и оценки скорости сходимости. Рассмотрены алгоритмы с бесконечным количеством кубитов и с малым количеством кубитов. Показано, что при достаточно малом шуме скорость сходимости экспоненциальная. При этом алгоритмы, учитывающие знак поправки сходятся быстрее, чем не учитывающие знак.

СПИСОК ЛИТЕРАТУРЫ

1. *Манин Ю.И.* Вычислимое и невычислимое. М.: Советское радио, 1980.
2. *Feynman R.P.* Simulating physics with computers // Int. J. Theor. Phys. 1982. V. 21. № 6. P. 467–488. https://doi.org/10.1007/BF02650179.
3. *Williams C.P.* Explorations in quantum computing. New York: Springer, 1998. https://doi.org/10.1007/978-1-84628-887-6.
4. *Nielsen M.A., Chuang I.L.* Quantum Computation and Quantum Information. Cambridge: Cambridge Univ. Press, 2010. https://doi.org/10.1017/CBO9780511976667.

5. *Grover L.K.* A fast quantum mechanical algorithm for database search // Proceed. of the twenty eighth Ann. ACM Symp. on Theory of Computing, Philadelphia, Pennsylvania, USA: Association for Computing Machinery. 1996. P. 212–219. https://doi.org/10.1145/237814.237866.

6. *Shor P.W.* Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer // SIAM J. on Comp. 1997. V. 26. № 5. P. 1484–1509. https://doi.org/10.1137/s0097539795293172.

7. *Harrow A.W., Hassidim A., Lloyd S.* Quantum algorithm for linear systems of equations // Phys. Rev. Lett. 2009. V. 103. № 15. P. 150502. https://doi.org/10.1103/physrevlett.103.150502.

8. *Albash T., Lidar D.A.* Adiabatic quantum computation // Rev. Mod. Phys. 2018. V. 90. № 1. P. 015002. https://link.aps.org/doi/10.1103/RevModPhys.90.015002.

9. *Kieu T.D.* The travelling salesman problem and adiabatic quantum computation: an algorithm // Quant. Inform. Proces. 2019. V. 18. № 3. P. 1–19. https://doi.org/10.1007/s11128-019-2206-9.

10. *Farhi E., Goldstone J., Gutmann S., Sipser M.* Quantum Computation by Adiabatic Evolution // arXiv preprint quant-ph/0001106. 2000. https://doi.org/10.48550/arXiv.quant-ph/0001106.

11. *Aharonov D., van Dam W., Kempe J., Landau Z., Lloyd S., Regev O.* Adiabatic quantum computation is equivalent to standard quantum computation // SIAM Rev. 2008. V. 50. № 4. P. 755–787. https://doi.org/10.1137/080734479.

12. *Kadowaki T., Nishimori H.* Quantum annealing in the transverse Ising model // Phys. Rev. E. 1998. V. 58. № 5. P. 5355–5363. https://doi.org/10.1103/physreve.58.5355.

13. *Bian Z., Chudak F., Macready W.G., Rose G.* The Ising model: teaching an old problem new tricks // D-Wave Systems. 2010.

14. *Albash T., Martin-Mayor V., Hen I.* Temperature scaling law for quantum annealing optimizers // Phys. Rev. Lett. 2017. V. 119. № 11. P. 110502. https://doi.org/10.1103/physrevlett.119.110502.

15. D-Wave Systems. QPU Solver Datasheet. https://docs.dwavesys.com/docs/latest/doc_qpu.html, accessed 24 Oct 2023.

16. *Vinci W., Buffoni L., Sadeghi H., Khoshaman A., Andriyash E., Amin M.H.* A path towards quantum advantage in training deep generative models with quantum annealers // Machine Learning: Science and Technology. 2020. V. 1. № 4. P. 045028. https://doi.org/10.1088/2632-2153/aba220.

17. *Korenkevych D., Xue Y., Bian Z., Chudak F., Macready W., Rolfe J., Andriyash E.* Benchmarking quantum hardware for training of fully visible boltzmann machines // arXiv preprint arXiv:1611.04528. 2016. https://doi.org/10.48550/arXiv.1611.04528.

18. *Denil M., de Freitas N.* Toward the implementation of a quantum RBM // NIPS Deep Learning and Unsupervised Feature Learning Workshop. 2011.

19. *Albash T., Lidar D.A.* Demonstration of a scaling advantage for a quantum annealer over simulated annealing // Phys. Rev. X. 2018. V. 8. № 3. P. 031016. https://doi.org/10.1103/physrevx.8.031016.

20. *King A.D., Raymond J., Lanting T., Harris R., Zucca A., Altomare F., Berkley A.J., Boothby K., Ejtemaee S., Enderud C., Hoskinson E., Huang S., Ladizinsky E., MacDonald A.J.R., Marsden G., Molavi R., Oh T., Poulin-Lamarre G., Reis M., Rich C., Sato Y., Tsai N., Volkmann M., Whittaker J.D., Yao J., Sandvik A.W., Amin M.H.* Quantum critical dynamics in a 5000-qubit programmable spin glass // Nature. 2023. V. 617. № 7959. P. 61–66. https://doi.org/10.1038/s41586-023-05867-2.

21. *O'Malley D., Vesselinov V.V.* ToQ.jl: A high-level programming language for D-Wave machines based on Julia // 2016 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA. 2016. P. 1–7. 10.1109/HPEC.2016.7761616.

22. *Borle A., Lomonaco S.J.* Analyzing the quantum annealing approach for solving linear least squares problems // Lect. Notes Comp. Sci. 2018. P. 289–301. https://doi.org/10.1007/978-3-030-10564-8_23.

23. *Rogers M.L., Singleton R.L.* Floating-point calculations on a quantum annealer: Division and matrix inversion // Front. Phys. 2020. V. 8. https://doi.org/10.3389/fphy.2020.00265.

24. *Borle A., Lomonaco S.J.* How viable is quantum annealing for solving linear algebra problems? // arXiv preprint arXiv:2206.10576, 2022. https://doi.org/10.48550/arXiv.2206.10576.

25. *Date P., Potok T.* Adiabatic quantum linear regression // Sci. Rep. 2021. V. 11. № 1. https://doi.org/10.1038/s41598-021-01445-6.

26. *Souza A.M., Martins E.O., Roditi I., Sa N., Sarthour R.S., Oliveira I.S.* An application of quantum annealing computing to seismic inversion // Front. Phys. 2022. V. 9. https://doi.org/10.3389/fphy.2021.748285.

27. *Meli N.K., Mannel F., Lellmann J.* An Iterative Quantum Approach for Transformation Estimation from Point Sets // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA. 2022. P. 519–527. https://doi.org/10.1109/CVPR52688.2022.00061.

28. *Conley R., Choi D., Medwig G., Mroczko E., Wan D., Castillo P., Yu K.* Quantum optimization algorithm for solving elliptic boundary value problems on D-Wave quantum annealing device // Proc. SPIE 12446, Quantum Computing, Communication, and Simulation III, 124460A. 2023. https://doi.org/10.1117/12.2649076.

29. *Lewis M., Glover F.* Quadratic Unconstrained Binary Optimization Problem Preprocessing: Theory and Empirical Analysis // Networks. 2017. V. 70. № 2. P. 79–97. https://doi.org/10.1002/net.21751.

30. *Ширяев А.Н.* Вероятность. М.: Наука, 1980.

31. *Lagarias J.C.* Euler's constant: Euler's work and modern developments // Bull. Am. Math. Soc. 2013. V. 50. № 4. P. 527–628. https://doi.org/10.1090/s0273-0979-2013-01423-x.

CONVERGENCE RATE OF ALGORITHM FOR SOLVING LINEAR EQUATIONS BY QUANTUM ANNEALING

S. B. Tikhomirov[a,]*, V. S. Shalgin[b,]**

[a] *Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio, Rua Marquês de São Vicente, 225, Gávea – Rio de Janeiro, RJ - Brazil Zip: 22451-900 - P.O. Box: 38097*

[b] *St. Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia*

**e-mail: sergey.tikhomirov@gmail.com*

***e-mail: st086496@student.spbu.ru*

Received 06 November, 2023

Revised 26 December, 2023

Accepted 06 February, 2024

Abstract. Various iterative algorithms for solving the linear equation $ax = b$ using a quantum computer operating on the principle of quantum annealing are studied. Assuming that the result produced by the computer is described by the Boltzmann distribution, conditions under which these algorithms converge are obtained and an estimate of their convergence rate is provided. Application of this approach for algorithms that use an infinite number of qubits and a small number of qubits is considered.

Keywords: adiabatic quantum computing, quantum annealing, linear equation, Boltzmann distribution, truncated normal distribution.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 780–790

УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

УДК 519.634

О СТРУКТУРЕ ВИНТОВЫХ ОСЕСИММЕТРИЧНЫХ РЕШЕНИЙ СИСТЕМЫ НАВЬЕ–СТОКСА ДЛЯ НЕСЖИМАЕМОЙ ЖИДКОСТИ[1)]

© 2024 г. В.А. Галкин[1,2,*]

[1] *628408 ХМАО-Югра, Сургут, ул. Энергетиков, 4, Сургутский филиал ФГУ ФНЦ НИИСИ РАН, Россия*
[2] *628400 ХМАО-Югра, Сургут, пр-т Ленина, 1, Сургутский государственный университет, Россия*
**e-mail: val-gal@yandex.ru*

Поступила в редакцию 13.11.2023 г.
Переработанный вариант 29.12.2023 г.
Принята к публикации 14.01.2024 г.

Получен класс точных решений уравнений Навье–Стокса для осесимметричного вихревого течения несжимаемой жидкости. Выделены инвариантные многообразия течений, обладающих вращательной симметрией относительно заданной оси в трехмерном координатном пространстве, приведено описание структуры решений. Установлено, что типичными инвариантными областями таких течений являются фигуры вращения, гомеоморфные тору, образующие структуру топологического расслоения, например, в шаре, цилиндре и в общих комплексах, составленных из таких фигур. Полученные результаты распространяются на подобные решения системы уравнений МГД, уравнения электродинамики Максвелла, обладающие в $\mathbb{R}_3$ аналогичными свойствами. Приведены примеры осесимметричных вихревых векторных полей и порожденных ими топологических расслоений на многообразиях в $\mathbb{R}_3$, инвариантных относительно динамических систем, задаваемых этими полями. Библ. 23. Фиг. 3.

Ключевые слова: уравнения несжимаемой жидкости, точные решения, точные решения системы Навье–Стокса, МГД, уравнения Максвелла, инвариантные многообразия, топологическое расслоение.

DOI: 10.31857/S0044466924050076, **EDN**: YDGAQB

1. ВВЕДЕНИЕ

Настоящая работа является развитием круга идей о построении поля с универсальной динамикой, появившихся у автора настоящей статьи во время участия в международных проектах по физике высоких энергий в 2000–2010 г.г.: HERA-B (DESY, Hamburg, Germany), OPERA (INFN, Gran-Sasso, Italy) в составе российской группы от Обнинского государственного технического университета атомной энергетики (IATE), (см. [1–6]).

В [7–23] даны содержательные результаты, относящиеся к теме настоящей работы.

В координатном пространстве $\mathbb{R}_3 = \{x = (x_1, x_2, x_3)\}$ рассматривается динамика во времени $t \in \mathbb{R}$ гладкого поля скоростей $V : \mathbb{R}_3 \times \mathbb{R} \to \mathbb{R}_3$ ($V \in C^2$), удовлетворяющего системе Навье–Стокса для несжимаемой жидкости

$$\frac{\partial V}{\partial t} + (V \cdot \nabla)V + \frac{1}{\overline{\rho}}\nabla P(x,t) = G(x,t) + \varepsilon^2 \Delta V, \tag{1.1}$$

$$\mathrm{div} V = 0, \tag{1.2}$$

где $P : \mathbb{R}_3 \times \mathbb{R} \to \mathbb{R}$ — давление жидкости, $\overline{\rho}$ и ε^2 — постоянные, характеризующие плотность и вязкость жидкости, $G(x,t)$ — плотность объемных сил. В настоящей работе рассматривается класс решений $\{\mathbf{V}, P\}$, обладающих свойством симметрии относительно некоторой оси $\mathcal{L}$, которую без потери общности направим вдоль вектора $\overline{e}_3 = (0,0,1)$, ($\{\overline{e}_i\}_{i=1}^3$ — ортонормированный базис в $\mathbb{R}_3$). Проекцию вектора x на ось $\mathcal{L}$ обозначим

[1)] Работа выполнена при финансовой поддержке в рамках государственного задания ФГУ ФНЦ НИИСИ РАН (НИЦ «Курчатовский институт») — Выполнение фундаментальных научных исследований ГП 47) по теме No 0580-2021-0007 «Развитие методов математического моделирования распределенных систем и соответствующих методов вычисления».

в дальнейшем $z = (x, \bar{e}_3)$, положим $x_i \stackrel{\text{def}}{=} (x, \bar{e}_i)$. Предположение об осевой симметрии функции u относительно оси $\mathcal{L}$ означает, что ее зависимость от пространственных аргументов x_1, x_2 осуществляется через величину $\rho(x_1, x_2) \stackrel{\text{def}}{=} \sqrt{(x_1^2 + x_2^2)}$. Для этого класса задач наряду с декартовыми координатами удобно использовать цилиндрическую систему координат (ρ, φ, z). Для функции u, обладающей осевой симметрией, положим $u(x) \equiv \overline{u}(\rho, z)$, и в этом случае оператор Лапласа имеет вид

$$\Delta u(x) = \Delta_\rho \overline{u} + \overline{u}_{zz}, \quad \Delta_\rho \overline{u} = \frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho \overline{u}_\rho).$$

Осесимметричные однородно-винтовые решения системы Навье–Стокса уже изучались в научной литературе. Одними из первых были работы [7, 8]. Анализ свойств этого класса решений проводился и другими авторами, см., например, [9–11].

2. ВИХРЕВЫЕ ОСЕСИММЕТРИЧНЫЕ ПОЛЯ

Пусть множество $M = \mathbb{R}_3 \backslash \mathcal{L}$ и функция $F \in C^4(M)$ при некоторой постоянной $\lambda \in \mathbb{R}$ удовлетворяет соотношению

$$\Delta F(x) + \lambda^2 F(x) = 0, \quad x \in M. \tag{2.1}$$

Определение 2.1. Предположим, что функция F удовлетворяет соотношению (2.1) при некотором $\lambda \neq 0$ и обладает на M осевой симметрией: $F(x) \equiv \overline{F}(\rho, z)$. Назовем осесимметричным вихревым полем, порожденным функцией F на множестве M, отображение $U : M \to \mathbb{R}_3$, заданное соотношением

$$U(x) = \lambda^{-2}\rho^{-1}\Big[\big(\overline{F}_{\rho,z}x_1 + \lambda\overline{F}_\rho x_2\big)\bar{e}_1 + \big(-\lambda\overline{F}_\rho x_1 + \overline{F}_{\rho,z}x_2\big)\bar{e}_2 - \big(\rho\Delta_\rho\overline{F}\big)\bar{e}_3\Big], \quad \rho = \rho(x_1, x_2), \quad x \in M. \tag{2.2}$$

Замечание 2.1. Отметим, что произвольные линейные комбинации осесимметричных полей вида (2.1), (2.2) сохраняют это свойство для результирующего поля.

С векторным полем (2.2) свяжем динамическую систему в цилиндрических координатах, определенную при $\lambda \neq 0$ в области $\rho > 0$:

$$\begin{aligned} \dot{\rho} &= \lambda^{-2}\overline{F}_{\rho,z}, \\ \dot{\varphi} &= -\lambda^{-1}\rho^{-1}\overline{F}_\rho, \\ \dot{z} &= -\lambda^{-2}\Delta_\rho\overline{F}. \end{aligned} \tag{2.3}$$

Будем полагать, что система (2.3) порождает однопараметрическую группу преобразований $T_t : M \to M$, $t \in \mathbb{R}$.

Лемма 2.1. *Пусть векторное поле U задано соотношениями* (2.1), (2.2). *Тогда на M справедливы тождества*

$$\operatorname{div} U = 0, \tag{2.4}$$

$$\Delta U + \lambda^2 U = 0, \tag{2.5}$$

$$(U \cdot \nabla)U = \frac{1}{2}\nabla(U, U). \tag{2.6}$$

Доказательство. Соотношения (2.4)–(2.6) являются прямым следствием тождества

$$\nabla \times U = \lambda U, \quad x \in M, \tag{2.7}$$

которое получается из формул (2.1) и (2.2).

Лемма доказана.

Замечание 2.2. Решения уравнения Бельтрами (2.7) вида (2.2) могут быть построены из решений уравнения Гельмгольца (2.1) методом Чандрасекара–Кендала [12, 13].

Лемма 2.2. *Динамическая система* (2.3) *имеет инвариант движения*

$$\overline{\Phi}(\rho, z) = \rho\overline{F}_\rho(\rho, z), \quad \rho > 0, \quad z \in \mathbb{R}. \tag{2.8}$$

Доказательство. Вычисление полной производной по времени функции $\overline{\Phi}\big(\rho(t), z(t)\big)$ с учетом соотношений (2.3) приводит к утверждению леммы.

Лемма доказана.

Следствие 2.1. Область M является расслоением с базой $\mathbb{R}$ и слоями $\Phi^{-1}(c) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^{\Phi}_3(x) = c\}\ \forall c \in \mathbb{R}$, где $\Phi(x) \equiv \overline{\Phi}\big(\rho(x_1, x_2), z\big)$:

$$M = \bigcup_{c\in\mathbb{R}} \Phi^{-1}(c), \tag{2.9}$$

при этом в силу утверждения леммы 2.2 группа преобразований $T_t : M \to M$ динамической системы

$$\dot{x} = U(x) \tag{2.10}$$

оставляет инвариантными слои $\Phi^{-1}(c)$:

$$T_t : \Phi^{-1}(c) \to \Phi^{-1}(c)\ \ \forall c, t \in \mathbb{R}. \tag{2.11}$$

Следствие 2.2. Каждая компонента связности открытого множества $M\backslash\big(\Phi^{-1}(0)\bigcup\mathcal{L}\big)$ является фигурой вращения вокруг оси $\mathcal{L}$.

Действительно, поскольку множество $\big(\Phi^{-1}(0)\bigcup\mathcal{L}\big)$ является замкнутым, а M — открытое подмножество в $\mathbb{R}_3$, то $M\backslash\big(\Phi^{-1}(0)\bigcup\mathcal{L}\big)$ открытое. Таким образом, $M\backslash\big(\Phi^{-1}(0)\bigcup\mathcal{L}\big)$ является объединением непересекающихся открытых компонент связности, границы которых принадлежат многообразию $\big(\Phi^{-1}(0)\bigcup\mathcal{L}\big)$. На многообразии $\Phi^{-1}(0)$ выполнено соотношение $\overline{F}_\rho(\rho, z) = 0$ при $\rho > 0$, и, значит, для динамической системы (2.3), являющейся представлением системы (2.10) в цилиндрической системе координат при $\rho > 0$ на $\Phi^{-1}(0)$ выполняется тождество $\dot{\varphi} \equiv 0$, т.е. линии тока систем (2.3), (2.10) на многообразии $\Phi^{-1}(0)$ инвариантны относительно поворотов относительно оси $\mathcal{L}$. Поскольку эти линии тока образуют $\Phi^{-1}(0)$, то границы компонент связности инвариантны относительно поворотов вокруг оси $\mathcal{L}$.

Ниже в качестве примеров таких областей рассматриваются комбинации цилиндров, шаров, торообразных и воронкообразных фигур.

Следствие 2.3. Группа T_t оставляет инвариантной каждую компоненту связности множества

$$M_{c,d} = \{x \in M : c < \Phi(x) < d\}\ \ \forall c, d \in \mathbb{R}. \tag{2.12}$$

Лемма 2.3. *Пусть $Q_{c,d}$ является открытой компонентой связности множества $M_{c,d}$, $c < d$, определенного в* (2.12), *$\partial Q_{c,d}$ — граница $Q_{c,d}$. Тогда группа T_t оставляет инвариантными $Q_{c,d}$ и $\partial Q_{c,d}$. В точках $q \in \partial Q_{c,d}$, где определен вектор $\nabla\Phi(q) \neq 0$, выполнено соотношение*

$$\big(U(q), n(q)\big) = 0, \quad n(q) = \|\nabla\Phi(q)\|^{-1}\nabla\Phi(q). \tag{2.13}$$

Доказательство. Инвариантность $Q_{c,d}$ и $\partial Q_{c,d}$ следует из соотношения (2.11). Поскольку $U(q) = \frac{d}{d\tau}T_t(q)|_{\tau=0}$, то $\frac{d}{d\tau}\Phi\big(T_t(q)\big)|_{\tau=0} = \big(U(q), \nabla\Phi(q)\big) = 0$, так как значения $\Phi\big(T_t(q)\big)$ на траектории постоянны. Тем самым устанавливается справедливость соотношения (2.13). Лемма доказана.

Замечание 2.3. На подмножествах границы компонент связности, где нормаль (2.13) не определена, условие непротекания (скольжения) по определению означает, что группа T_t оставляет инвариантной эту часть границы

Замечание 2.4. Отмеченный в лемме случай, когда вектор нормали не определен, является типичным для цилиндрических областей [17] на участках границы, состоящих из пересечения боковых образующих цилиндра и его торцов.

Замечание 2.5. Пусть векторные поля U_α определены формулами (2.1), (2.2) при фиксированном значении параметра $\lambda \neq 0$. Если U_α являются достаточно гладкими на M, то произвольные конечные линейные комбинации осесимметричных полей U_α

$$\tilde{U} = \sum_\alpha a_\alpha U_\alpha$$

удовлетворяют соотношениям (2.4)–(2.6) на M. При надлежащих требованиях на слагаемые в правой части этого выражения это утверждение распространяется на бесконечные наборы U_α посредством интегрирования по параметру α.

3. РЕШЕНИЯ СИСТЕМЫ (1.1), (1.2) НА КОМПОНЕНТАХ СВЯЗНОСТИ

Свойство инвариантности подобластей в $D \subset \mathbb{R}_3$ относительно действия группы $T_t : D \to D$, выделенных в лемме 2.3, позволяет рассматривать решения системы Навье–Стокса (1.1), (1.2) только на этих множествах с дополнительным условием непротекания (скольжения):

$$\big(V(q,t), \mathbf{n}(q)\big)|_{q\in\partial D} = 0, \tag{3.1}$$

где $\mathbf{n}(q)$ — поле единичных нормалей к кусочно-гладкой границе ∂D области D.

Теорема 3.1. *Пусть $D = Q_{c,d}$ является открытой компонентой связности множества $M_{c,d}$, $c < d$, определенного в* (2.12), *$\partial Q_{c,d}$ — граница $Q_{c,d}$, в точках $\forall a \in \partial Q_{c,d}$ которой определено поле единичных нормалей $\mathbf{n}(q)$. Предположим, что объемные силы $G(x,t) = \nabla g(x,t)$ с гладким потенциалом в области $M = \mathbb{R}\backslash\mathcal{L}$. Тогда система уравнений Навье–Стокса* (1.1), (1.2) *имеет в области $Q_{c,d}$ решение, удовлетворяющее условию непротекания* (*скольжения*) (3.1):

$$V(x,t) = U(x)\exp(-\lambda^2\varepsilon^2 t), \tag{3.2}$$

$$P(x,t) = -\frac{\overline{\rho}}{2}\big(V(x,t), V(x,t)\big) + g(x,t) + \beta(t), \tag{3.3}$$

где $\beta(t)$ — произвольная функция, зависящая от времени $t \in \mathbb{R}$. Если в формулах (3.2), (3.3) *заменить поле U на $\tilde{U}$, определенное в замечании* (2.5), *то эти выражения определяют гладкое решение системы уравнений Навье-Стокса* (1.1), (1.2) *на множестве аргументов $x \in \mathbb{R}_3$, $t \in \mathbb{R}$.*

Доказательство. Утверждение теоремы является прямым следствием соотношений (2.4)–(2.6) из леммы 2.1 и свойства инвариантности границы области относительно группы T_t, заданной соотношением (2.10).

Теорема доказана.

Теорема 3.2. *Пусть рассматривается произвольная линейная комбинация*

$$U(x) = \sum_k \alpha_k U_k(x) \tag{3.4}$$

осесимметричных полей $U_k(x)$, заданных формулами (2.1), (2.2) *при помощи функций $F_k = \overline{F}_k(\rho, z)$, соответствующих в уравнении* (2.1) *фиксированному значению параметра. Тогда инвариантные многообразия* (2.8) *для динамической системы T_t, порожденной полем* (3.4), *удовлетворяют соотношению*

$$\overline{\Phi}(\rho, z) \equiv \rho \sum_k \alpha_k \frac{\partial}{\partial\rho}\overline{F}_k(\rho, z) = c \quad \forall c \in \mathbb{R}. \tag{3.5}$$

Доказательство. Рассмотрим динамическую систему (2.3) с полем (3.4) и сопутствующую ей группу преобразований T_t. Производная по времени на траекториях этой системы удовлетворяет тождествам

$$\frac{d}{dt}\overline{\Phi}\big(\rho(t), z(t)\big) = \overline{\Phi}_\rho\big(\rho(t), z(t)\big)\dot{\rho}(t) + \overline{\Phi}_z\big(\rho(t), z(t)\big)\dot{z}(t),$$

$$\dot{\rho}(t) = \lambda^{-2}\overline{F}_{\rho,z},$$

$$\dot{z} = -\lambda^{-2}\Delta_\rho\overline{F},$$

где функция $\overline{F} \equiv \sum\limits_k \alpha_k\overline{F}_k$, и для каждой $\overline{F}_k$ в силу определения полей U_k посредством соотношений (2.1), (2.2) справедливы соотношения

$$\Delta_\rho\overline{F}_k + \frac{\partial^2}{\partial z^2}\overline{F}_k + \lambda^2\overline{F}_k = 0, \quad \rho > 0,\ z \in \mathbb{R}.$$

Таким образом, для функции $\overline{F}$ имеем

$$\Delta_\rho\overline{F} + \frac{\partial^2}{\partial z^2}\overline{F} + \lambda^2\overline{F} = 0, \quad \rho > 0,\ z \in \mathbb{R}.$$

Поскольку

$$\frac{d}{dt}\overline{\Phi}\big(\rho(t), z(t)\big) \equiv \dot{\rho}\big(\overline{F}_\rho + \rho\overline{F}_{\rho\rho}\big) + \rho\overline{F}_{\rho z}\dot{z},$$

то

$$\frac{d}{dt}\overline{\Phi}\big(\rho(t),z(t)\big)\equiv\lambda^2\rho\overline{F}_{\rho,z}\Big[\overline{F}_{\rho,\rho}+\frac{1}{\rho}\overline{F}_{\rho}-\Delta_{\rho}\overline{F}\Big].$$

Так как последний сомножитель в правой части этого тождества равен нулю, то функция $\overline{\Phi}$ является инвариантом группы преобразований T_t. Теорема доказана.

Представление области M в виде расслоения (2.9) и инвариантность открытых компонентов связности области M относительно группы T_t позволяет рассматривать их как независимые структуры гидродинамического течения (1.1), (1.2). Это дает возможность расширить класс задач для системы уравнений Навье–Стокса на случай многокомпонентных систем, рассматривая величину $\overline{\rho}$ как кусочно-постоянную функцию в M с участками постоянства на открытых компонентах связности, инвариантных относительно группы T_t. Очевидно, что поле скоростей $V(x,t)$ на M определяется формулой (3.2), а давление в этом случае является кусочно-непрерывной функцией пространственных переменных, удовлетворяющей на каждой компоненте соотношению (3.3). Тем самым появляется возможность рассмотрения многофазных нереагирующих жидкостей, обладающих различными плотностями. На компонентах, где $\overline{\rho}=0$, течение считаем не определенным, хотя формально поле $V(x,t)$ имеет гладкое продолжение на эту область.

Замечание 3.1. Решения (3.2), (3.3) при различных значениях параметра $\lambda\neq 0$ топологически эквивалентны случаю $\lambda=1$, к которому приводит масштабирование пространственно-временных координат

$$(x',t')=(\lambda x,\lambda^2 t). \tag{3.6}$$

Замечание 3.2. Все результаты настоящей работы переносятся на специальный класс решений системы МГД (магнитной гидродинамики)

$$\frac{\partial V}{\partial t}+(V\cdot\nabla)V+\frac{1}{\overline{\rho}}\nabla P(x,t)=-\frac{1}{4\pi\overline{\rho}}[\mathrm{H},\operatorname{rot}\mathrm{H}]+\nabla g+\varepsilon^2\Delta V, \tag{3.7}$$

$$\frac{\partial H}{\partial t}=\operatorname{rot}[V,\mathrm{H}]+\mu^2\Delta\mathrm{H}, \tag{3.8}$$

$$\operatorname{div}V=0, \tag{3.9}$$

$$\operatorname{div}\mathrm{H}=0, \tag{3.10}$$

где $\mathrm{H}(x,t)$ — напряженность магнитного поля, постоянная величина μ — магнитная вязкость жидкости. Указанный класс решений системы МГД (3.7)–(3.10) определяется следующими соотношениями:

$$V(x,t)=V_0\exp(-\lambda^2\varepsilon^2 t)U(x),\quad \mathrm{H}(x,t)=\mathrm{H}_0\exp(-\lambda^2\mu^2 t)U(x), \tag{3.11}$$

$$P(x,t)=-\frac{\overline{\rho}}{2}\big(V(x,t),V(x,t)\big)+g(x,t)+\beta(t), \tag{3.12}$$

с векторным полем U на M, заданным формулами (2.1), (2.2), V_0 и H_0 — произвольные постоянные. Особо следует выделить случай нетривиальных стационарных магнитных полей $\mathrm{H}(x,t)\equiv\mathrm{H}_0U(x)$ при $\mu=0$, указывающий на возможность существования решений системы МГД типа “геодинамо”. В частности, в работе [16] рассмотрено векторное поле $U:\mathbb{R}_3\to\mathbb{R}_3$ в классе вещественно-аналитических функций на $\mathbb{R}_3$ с инвариантными многообразиями на шарах и шаровых слоях:

$$U(x)=\frac{\overline{u}'(\lambda r)}{r}\begin{bmatrix}x_2\\-x_1\\-2\lambda^{-1}\end{bmatrix}+\frac{1}{r^2}\left(\overline{u}''(\lambda r)-\frac{\overline{u}'(\lambda r)}{\lambda r}\right)\begin{bmatrix}x_1 z\\x_2 z\\-(x_1^2+x_2^2)\end{bmatrix},\quad \lambda\neq 0, \tag{3.13}$$

где

$$\overline{u}(r)=\begin{cases}r^{-1}\sin(r). & r>0,\\ 1, & r=0.\end{cases}\qquad r(x)\equiv\sqrt{x_1^2+x_2^2+z^2}>0. \tag{3.14}$$

Это поле удовлетворяет условию регулярности

$$U(x)\sim\mathcal{O}\left(\frac{1}{r(x)}\right)\to 0,\quad r(x)\to\infty. \tag{3.15}$$

В силу теоремы 3.2 такими же свойствами обладают линейные комбинации полей (3.13), (3.14) вида

$$\tilde{U}(x) = \sum_k A_k U(x + z_k \bar{e}_3), \tag{3.16}$$

определяющие симметричные решения относительно оси $\mathcal{L}$. Более общий класс решений системы (3.7)–(3.10) порождают поля $\tilde{U}$ из замечания 2.5.

Примеры решений системы МГД (3.7)–(3.10) с инвариантными многообразиями, имеющих форму цилиндров, симметричных относительно оси $\mathcal{L}$, дают векторные поля (2.2), соответствующие функциям F в соотношении (2.1), определенных формулами

$$F = \overline{F}(\rho, z) \equiv \sum_k \Big\{ A_k J_0(a_k \rho) \sin\big[b_k(z - z'_k)\big] + B_k J_0(c_k \rho) \cos\big[d_k(z - z''_k)\big] \Big\}, \tag{3.17}$$

где J_0 — функция Бесселя, постоянные A_k, B_k, a_k, b_k, c_k, d_k, z'_k, z''_k подчинены условию достаточно быстрой сходимости ряда (3.15), обеспечивающей его гладкость в классе $C^4(M)$ для выполнения дифференциальных операций в формулах (2.2), (2.6), и выполняются соотношения

$$a_k^2 + b_k^2 = \lambda^2, \quad c_k^2 + d_k^2 = \lambda^2, \quad k \in \mathbb{N}, \quad \lambda \neq 0. \tag{3.18}$$

В силу замечания 2.1 линейные комбинации вышеупомянутых осесимметричных полей, определяемых формулами (3.16) и (3.17) позволяют рассматривать широкий класс бесконечно гладких решений системы МГД (3.7)—(3.10) вида (3.11), (3.12), имеющих структуру расслоения (2.9) на $\mathbb{R}_3$.

Множество векторных полей, порожденных функциями (3.17), (3.18), которые естественно называть цилиндрическими, можно расширить, добавив к ним линейные комбинации полей, порожденных функциями с особенностями на оси $\mathcal{L}$:

$$F = \overline{F}(\rho, z) \equiv \sum_k \Big\{ A'_k N_0(a'_k \rho) \sin\big[b'_k(z - \tilde{z}'_k)\big] + B'_k N_0(c'_k \rho) \cos\big[d'_k(z - \tilde{z}''_k)\big] \Big\}, \tag{3.19}$$

$$(a'_k)^2 + (b'_k)^2 = \lambda^2, \quad (c'_k)^2 + (d'_k)^2 = \lambda^2, \quad k \in \mathbb{N}, \quad \lambda \neq 0, \tag{3.20}$$

где N_0 — функция Неймана. Требования к сходимости рядов (3.19), (3.20) аналогичны вышеупомянутому случаю рядов (3.17).

В случае стационарных решений системы уравнений Эйлера для уравнений гидродинамики, когда в системе (1.1), (2.1) параметр вязкости $\varepsilon = 0$, решения аналогичные классам вида (3.17)–(3.20), иными методами получены в [11].

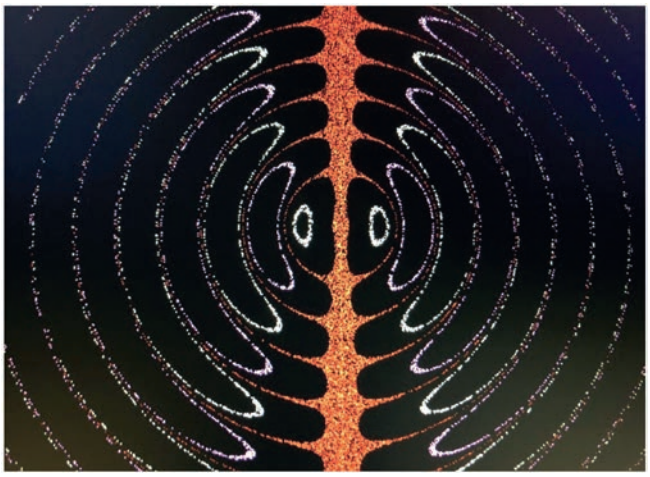
Расслоение $\mathbb{R}_3$ на инвариантные многообразия для решений системы МГД, порожденных приведенными выше функциями $F = \overline{F}(\rho, z)$, получается на основании формулы (2.8) вращением вокруг оси $\mathcal{L}$ семейства многообразий

$$\rho \sum_k \Big\{ A_k J_1(a_k \rho) \sin\big[b_k(z - z'_k)\big] + B_k J_1(c_k \rho) \cos\big[d_k(z - z''_k)\big] + \\ + A'_k N_1(a'_k \rho) \sin\big[b'_k(z - \tilde{z}'_k)\big] + B'_k N_1(c'_k \rho) \cos\big[d'_k(z - \tilde{z}''_k)\big] \Big\} = c. \tag{3.21}$$

с произвольными постоянными $c \in \mathbb{R}$. Эти семейства естественным образом можно расширить за счет использования аналогичных комбинаций функций Бесселя и Неймана мнимого аргумента в сочетании с гиперболическими синусами и гиперболическими косинусами.

Замечание 3.3. Наличие вращательной симметрии относительно оси $\mathcal{L}$ для рассматриваемых гидродинамических течений позволяет выделить инвариантные компоненты, диффеоморфные торам, для решений вида (3.1), (3.2), (3.3) в шаре и цилиндре [16, 17]. Более того, таким свойством обладают произвольные линейные комбинации этих решений при одинаковых значениях параметра $\lambda \neq 0$ в областях, составленных из шаров и цилиндров, для которых оси симметрии течений совпадают с $\mathcal{L}$. Пример осевого сечения семейства инвариантных многообразий, диффеоморфных шарам и торам, для решения из [16, 17] приведен на фиг. 1 (структура сечения инвариантных многообразий, приведенная на рисунке, соответствует формулам (3.13)–(3.18)). Инвариантные слои течения в этом случае удовлетворяют в $\mathbb{R}$ уравнению

$$\left[\cos(\sqrt{\rho^2 + z^2}) - \frac{\sin(\sqrt{\rho^2 + z^2})}{\sqrt{\rho^2 + z^2}} \right] \frac{\rho^2}{\rho^2 + z^2} = c, \quad \rho(x_1, x_2) = \sqrt{x_1^2 + x_2^2},$$



Фиг. 1. Осевое сечение семейства вложенных инвариантных сфер $\Phi^{-1}(0)$ (красный цвет) и торов $\Phi^{-1}(c)$, $c \neq 0$ (белый цвет) для гладкого решения в шаре (3.13), (3.14). Инвариант Φ определяется формулой (2.8).

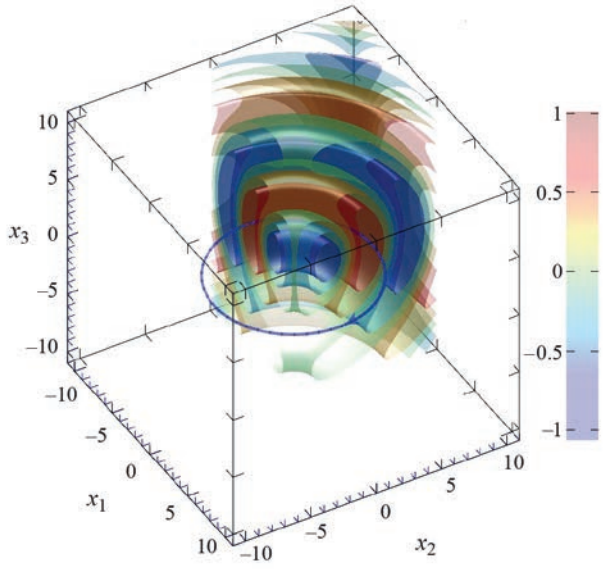
где c — произвольная постоянная. 3-D визуализация структуры этого решения в шаре приведена ниже на фиг. 2.

Течение, двойственное к (3.13)–(3.15), с аналогичной сферической структурой, но с особенностью в начале координат, порождается по формулам (2.1), (2.2) функцией

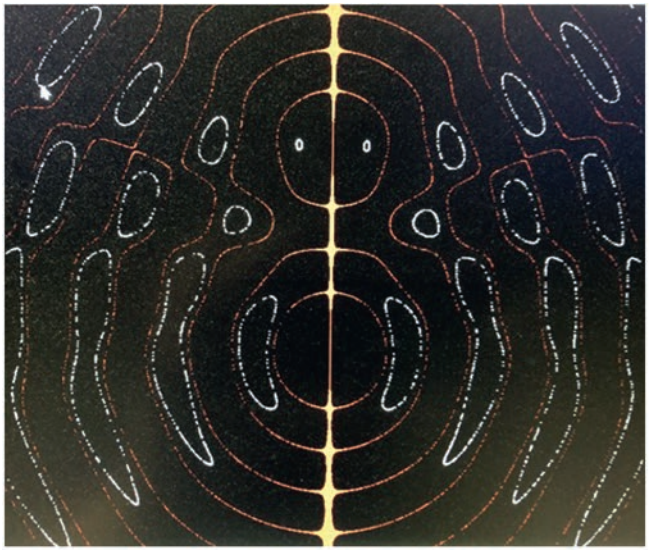
$$F = \overline{F}(\rho, z) \equiv \frac{\cos(\sqrt{\rho^2 + z^2})}{\sqrt{\rho^2 + z^2}}, \quad \rho(x_1, x_2) = \sqrt{x_1^2 + x_2^2}, \quad \lambda = 1.$$

Пример визуализации осевого сечения течения, порожденного линейной комбинацией вида (3.16) $\tilde{U}(x) = A_1 U(x + z_1 \overline{e}_3) + A_2 U(x + z_2 \overline{e}_3)$ для двух сферических течений (3.13), (3.14) с центрами, расположенными на оси $\mathcal{L}$ в точках $z_k \overline{e}_3$, где $z_1 = -8$, $z_2 = 8$, и коэффициентами “смешивания” $A_1 = 1$, $A_2 = -2$, приведен ниже на фиг. 3.

Отметим, что сечения инвариантных многообразий, выделенные на фиг. 1, 2 красным цветом, соответствуют траекториям динамической системы (2.3), движущимся вдоль оси симметрии $\mathcal{L}$, и для них выполняются соотношения $\dot{\varphi}(t) \equiv 0$, $\Phi\big(x(t)\big) \equiv 0$. Соответственно, на этих сечениях в силу тождеств (2.3) справедливы соот-



Фиг. 2. 3-D визуализация структуры гладкого решения в шаре (3.13), (3.14).



Фиг. 3. Осевое сечение инвариантных многообразий для течения (3.16), возникающих при “смешивании” двух сферических течений (3.13)–(3.14), смещенных симметрично относительно начала координат вдоль оси $\mathcal{L}$, окрашенной желтым цветом. Многообразия $\Phi^{-1}(0)$ выделены красным цветом, белый цвет соответствует многообразию $\Phi^{-1}(1)$, где инвариант Φ определяется формулой (3.5).

ношения

$$\dot{\rho} = \lambda^{-2}\overline{F}_{\rho,z}(\rho, z), \quad \dot{z} = -\lambda^{-2}\Delta_\rho \overline{F}(\rho, z), \quad \overline{\Phi}\big(\rho(0), z(0)\big) = 0, \quad \varphi(t) \equiv \varphi(0).$$

Многообразия, сечения которых окрашены белым цветом, диффеоморфны двумерным торам $T^{(c)} \stackrel{\text{def}}{=} \{x \in \in \mathbb{R}_3 : \Phi(x) = c\}$, $c \neq 0$, и траектории динамической системы $T_t\big(x(0)\big) \equiv x(t)$ образуют их “обмотку”. Трехмерные фигуры $D^{(c)}$, ограниченные торами $T^{(c)}$, получающиеся вращением этих сечений вместе с их внутренней частью вокруг оси симметрии $\mathcal{L}$, вложены друг в друга, а их пересечение $S = \bigcap\limits_{c\neq 0} D^{(c)}|_{D^{(c)}\neq\varnothing}$ является набором окружностей, лежащих в плоскостях, ортогональных оси $\mathcal{L}$. На этих окружностях располагаются периодические траектории динамической системы (2.3). Пример такой круговой орбиты в экваториальной плоскости выделен синим цветом в 3-D визуализации динамики в шаре на фиг. 2. Координаты таких окружностей в плоских сечениях, параллельных оси симметрии $\mathcal{L}$, выделятся как решения (ρ_*, z_*) системы уравнений

$$\overline{F}_{\rho,z}(\rho_*, z_*) = 0,$$
$$\Delta_\rho \overline{F}(\rho_*, z_*) = 0.$$

Каждая точка $(\rho_*, z_*) \in \mathbb{R}$ задает вышеуказанную окружность, на которой динамика системы (2.3) определяется следующими тождествами:

$$\rho(t) = \rho_*, \varphi(t) = \varphi(0) - \lambda^{-1}\rho_*^{-1}\overline{F}_\rho(\rho_*, z_*)t,$$
$$z = z_*.$$

В работе [16] приведены координаты таких окружностей, лежащих в плоскости $z_* = 0$, для течения (3.13), (3.14) при значении $\lambda = 1$. В этом случае их радиусы ρ_* являются счетным набором положительных корней уравнения $\mathrm{tg}(\rho_*) = \frac{\rho_*}{1-\rho_*^2}$.

Пересечение $\mathcal{L} \bigcap \Phi^{-1}(0)$ состоит из стационарных точек динамической системы (2.3).

Экспериментальные данные показывают, что аналогичная топологическая структура наблюдается в высокоинтенсивных газодинамических потоках.

Замечание 3.4. Осесимметричные решения системы МГД, порождающее расслоение на $\mathbb{R}_3$ со слоями в форме вложенных друг в друга “воронок”, ориентированных вдоль оси $\mathcal{L}$, порождаются векторным полем

$$U(x) = J_0(\rho)\begin{bmatrix}0\\0\\z\end{bmatrix} - J_1(\rho)\rho^{-1}\begin{bmatrix}x_1 + x_2 z\\ x_2 - x_1 z\\ 0\end{bmatrix}, \quad \lambda = 1. \tag{3.22}$$

Структура инвариантных слоев для поля (3.22) в осевом сечении в координатах (ρ, z) задается формулой

$$z = \frac{c}{\rho J_1(\rho)}, \quad \rho > 0, \tag{3.23}$$

где c — произвольная постоянная, задающая инвариантный слой в $\mathbb{R}_3$, который получается вращением кривой (3.23) вокруг оси $\mathcal{L}$.

Замечание 3.5. Отметим, что классы вихревых векторных полей $U(x)$, задаваемых формулами (2.1), (2.2), позволяют построить в $\mathbb{R}_3$ решения системы уравнений электродинамики Дж.К. Максвелла в случае отсутствия свободных зарядов и токов:

$$\begin{gathered} \frac{\partial \mathrm{E}}{\partial t} = \operatorname{rot} \mathrm{B}, \quad \frac{\partial \mathrm{B}}{\partial t} = -\operatorname{rot} \mathrm{E}, \\ \operatorname{div} \mathrm{E} = 0, \quad \operatorname{div} \mathrm{B} = 0, \\ x \in \mathbb{R}_3, \quad t \in \mathbb{R}, \end{gathered} \tag{3.24}$$

где $\mathrm{E}(x,t)$ — напряженность электрического поля, $\mathrm{B}(x,t)$ — вектор магнитной индукции. Положим

$$\mathrm{E}(x,t) = \sin(\lambda t)U(x), \quad \mathrm{B}(x,t) = \cos(\lambda t)U(x), \tag{3.25}$$

с векторным полем $U(x)$, которое определяем по формулам (2.1), (2.2), либо из замечания 2.5. При заданном значении параметра $\lambda \neq 0$, устанавливаем, что формулы (3.25) являются решением системы (3.24).

Учитывая линейность системы (3.24), инвариантность операций rot, div относительно группы поворотов SO_3 и сдвигов, формулы (3.25) позволяют конструировать решения системы (3.24) в виде линейных комбинаций (рядов)

$$\mathrm{E}(x,t) = \sum_k A_k \sin\bigl(\lambda_k(t-t_k)\bigr)U_k(x), \quad \mathrm{B}(x,t) = \sum_k A_k \cos\bigl(\lambda_k(t-t_k)\bigr)U_k(x)$$

с произвольными A_k, $t_k \in \mathbb{R}$, параметрами $\lambda_k \neq 0$, и полями U_k, построенными по формулам (2.1), (2.2) для осей симметрии $\mathcal{L}_k \in \mathbb{R}_3$, которые получатся произвольными поворотами в $\mathbb{R}_3$ выделенной оси $\mathcal{L}$ и соответствующего поля U.

Широкий набор примеров $U(x)$ приведен выше в следствиях 3.2–3.4 и замечании 2.5.

4. ЗАКЛЮЧЕНИЕ

В работе исследована структура топологического расслоения для одного класса вихревых нестационарных осесимметричных точных решений уравнений Навье–Стокса в случае несжимаемой жидкости. На основе единого подхода результаты распространены на специальные случаи системы МГД и уравнения электродинамики Дж.К. Максвелла.

Автор выражает благодарность за обсуждение работы и участие в исследованиях по визуализации построенных решений сотрудникам Сургутского филиала ФГУ ФНЦ НИИСИ РАН Т.В. Гавриленко, Д.А. Моргуну, А.О. Дубовику, А.Д. Смородинову, Т.Н. Садыкову.

СПИСОК ЛИТЕРАТУРЫ

1. *Galkin V. A. et al.* The detection of neutrino interactions in the emulsion/lead target of the OPERA experiment // J. of Instrumentation. 2009. V. 4. № 1.

2. *Galkin V. A. et al.* The OPERA experiment in the CERN to Gran Sasso neutrino beam // J. of Instrumentation. 2009. V. 4. № 1.

3. *Галкин В. А.* Анализ математических моделей: системы законов сохранения, уравнения Больцмана и Смолуховского. М.: БИНОМ, 2009. 408 с.

4. *Галкин В. А., Савельев В. И.* Энциклопедия низкотемпературной плазмы (Серия «Б», том VII-1 «Математическое моделирование в низкотемпературной плазме». Гл. 6. Математическое моделирование переходного излучения в средах с быстропеременными электромагнитными свойствами (Ч. III). М.: Янус-К, 2009. С. 348–364.

5. *Galkin V. A. et al.* Study of the effects induced by lead on the emulsion films of the OPERA experiment // J. of Instrumentation. 2008. V. 3. № 1.

6. *Galkin V. A. et al.* Emulsion sheet doublets as interface trackers for the OPERA experiment // J. of Instrumentation. 2008. V. 3. № 1.

7. *Caldonazzo D.* Moti helicoidali, simmetrici ad un asse, di liquidi viscosi // Ist. Lombardo Accad. Sci. Lett. Rend A. 1926. Vol. 59. P. 657—665.

8. *Mattei G.* Sui moti di Beltrami- Caldonazzo in magnetofluidodinamica // Rendiconti del Seminaro Matematico della Universita di Padova. 1982. Vol. 68. P. 11—15.

9. *Богоявленский О. И.* О задаче Кельвина 1880 года и точных решениях уравнений Навье—Стокса // Общеинститутский семинар «Математика и ее приложения» Математического института им. В.А. Стеклова Российской академии наук 21 мая 2015 г. Москва, конференц-зал МИАН (ул. Губкина, 8). Электронный ресурс: www.mathnet.ru

10. *Bogoyavlenskij O.* New exact axisymmetric solutions to the Navier–Stokes equations // Zeitschricht Naturforschung A. 2020. Vol. 75. № 1. P. 29–42.

11. *Ковалев В.П., Сизых Г.Б.* Осесимметричные винтовые течения идеальной жидкости // Тр. МФТИ. 2016. Т. 8. № 3. С. 171–178.

12. *Chandrasekhar, Subrahmanyan.* On force-free magnetic fields // Proc. of the National Academy of Sciences. 1956. 42 (1): 1–5.

13. *Chandrasekhar, Subrahmanyan; Kendall, P. C.* On Force-Free Magnetic Fields // The Astrophysical Journal. September 1957. 126 (1): 1–5.

14. *Бетелин В.Б., Галкин В.А.* Управление параметрами несжимаемой жидкости при изменении во времени геометрии течения // Докл. АН. 2015. Т. 463. № 2. С. 149–51.

15. *Бетелин В.Б., Галкин В.А., Дубовик А.О.* Точные решения системы Навье–Стокса для несжимаемой жидкости в случае задач, связанных с нефтегазовой отраслью // Докл. АН. Математика, информатика, процессы управления. 2020. Т. 495. № 1. С. 13–6.

16. *Галкин В. А.* Об одном классе точных решений системы Навье–Стокса для несжимаемой жидкости в шаре и сферическом слое // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 6. С.1000–005.

17. *Галкин В. А., Дубовик А. О.* Об одном классе точных решений системы уравнений Навье–Стокса для несжимаемой жидкости // Матем. моделирование. 2023. Т. 35. № 8. С. 3–3.

18. *Галкин В. А., Смородинов А. Д., Моргун Д. А.* Решение уравнения Навье–Стокса для сталкивающихся потоков // Успехи кибернетики. 2023. Т. 4. № 2. С. 8–5.

19. *Галкин В. А., Дубовик А. О.* Моделирование трехмерного потенциального течения жидкости в области, изменяющейся во времени // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 7. С. 1180–1186.

20. *Trkal V.* A note on the hydrodynamics of viscous fluids // Czechoslovak Journal of Physics. 1994. Vol. 44. № 2. P. 97–106.

21. *Шеретов Ю.В.* О решениях задачи Коши для квазигидродинамической системы // Вестник ТвГУ. Серия: Прикладная математика. 2020. № 1. С. 84–96. https://doi.org/10.26456/vtpmk557

22. *Arnold V.I.* Sur la topologie des ecoulements stationnaires des fluides parfaits. C. R. Acad. Sci. Paris, 1965. 261:17–20.

23. *Галкин В.А., Дубовик А.О.* О моделировании слоистого течения вязкой проводящей жидкости в области, изменяющейся во времени // Матем. моделирование. 2020. Т. 32. № 4. С. 31–42.

ON THE STRUCTURE OF HELICAL AXISYMMETRIC SOLUTIONS OF THE NAVIER-STOKES SYSTEM FOR INCOMPRESSIBLE FLUIDS[1)]

V. A. Galkin[a,b,*]

[a] *Surgut Branch of the Federal Research Center "NIISI" of the Russian Academy of Sciences, Energetikov St. 4, Surgut, Khanty-Mansi Autonomous Okrug-Yugra, 628408, Russia*
[b] *Surgut State University, Lenin Ave. 1, Surgut, Khanty-Mansi Autonomous Okrug-Yugra, 628400, Russia*
**e-mail: val-gal@yandex.ru*

Received 13 November, 2023
Revised 29 December, 2023
Accepted 14 January, 2024

Abstract. A class of exact solutions to the Navier-Stokes equations for axisymmetric vortex flows of incompressible fluids is obtained. Invariant manifolds of flows with rotational symmetry relative to a given axis in three-dimensional coordinate space are identified, and the structure of the solutions is described. It is established that typical invariant regions of such flows are rotational figures homeomorphic to a torus, forming a structure of topological fibration, such as in a sphere, cylinder, and more complex configurations composed of such figures. The results are extended to similar solutions of the magnetohydrodynamics (MHD) system and Maxwell's electrodynamics equations, which possess $\mathbb{R}_3$ analogous properties. Examples of axisymmetric vortex vector fields and the topological fibrations they generate on manifolds invariant $\mathbb{R}_3$ under the dynamical systems defined by these fields are provided.

Keywords: incompressible fluid equations, exact solutions, Navier-Stokes system exact solutions, MHD, Maxwell's equations, invariant manifolds, topological fibration.

[1)] The work was financially supported under the state assignment of the Federal Research Center "NIISI" RAS (Kurchatov Institute Research Center) — Implementation of fundamental scientific research GP 47) on the topic No 0580-2021-0007 "Development of methods for mathematical modeling of distributed systems and corresponding computational methods".

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 791–803

УДК 517.956.223+517.575

ФУНКЦИЯ ГРИНА ЗАДАЧИ РИКЬЕ–НЕЙМАНА ДЛЯ ПОЛИГАРМОНИЧЕСКОГО УРАВНЕНИЯ В ЕДИНИЧНОМ ШАРЕ

© 2024 г. В. В. Карачик[1,*]

[1]454080 Челябинск, пр-т Ленина, 76, ЮУрГУ (НИУ), Россия
**e-mail: karachik@susu.ru*

Поступила в редакцию 10.01.2023 г.
Переработанный вариант 10.01.2023 г.
Принята к публикации 06.02.2024 г.

Строится функция Грина задачи Рикье–Неймана для полигармонического уравнения в единичном шаре и приводится интегральное представление решений задачи Рикье–Неймана. Приведены два примера. Библ. 26.

Ключевые слова: полигармоническое уравнение, задача Рикье–Неймана, функция Грина.

DOI: 10.31857/S0044466924050089, **EDN**: YDFLBA

1. ВВЕДЕНИЕ

Много работ посвящено построению функции Грина в явном виде для различных классических краевых задач. Функции Грина бигармонических задач Дирихле, Неймана, Робина и др. в двумерном диске построены в [1] с помощью гармонических функций Грина задачи Дирихле, а в [2], [3] найдено явное представление гармонической функции Робина. Явная форма функции Грина в секторе для бигармонического и тригармонического уравнений приведена в работах [4], [5]. Статьи [6], [7] посвящены построению функции Грина задачи Дирихле для полигармонического уравнения в единичном шаре. В [8] дано явное представление функции Грина задачи Робина для уравнения Пуассона, а в [9] приведен явный вид функции Грина для 3-гармонического уравнения в единичном шаре.

Условия разрешимости некоторых вариантов задач для бигармонического уравнения в шаре были получены также в работах [10], [11]. В [12] исследована фредгольмовость и индекс обобщенной задачи Неймана, содержащей степени нормальных производных в граничных условиях. В [13] приводятся функции Грина задач Навье [14] и Рикье–Неймана для бигармонического уравнения в шаре, а в [15] исследована разрешимость четырех нелокальных задач для бигармонического уравнения с инволюцией.

Хорошо известно, что функция Грина задачи Дирихле для уравнения Пуассона в шаре $S = \{x \in \mathbb{R}^n : |x| < 1\}$ при $n \geqslant 2$ имеет вид

$$G_2(x,\xi) = E(x,\xi) - E\Big(\frac{x}{|x|}, |x|\xi\Big), \tag{1}$$

где $E(x,\xi)$ – элементарное решение уравнения Лапласа (см. [16]). В работах [9], [17], [18] были определены элементарные решения бигармонического и тригармонического уравнений $E_4(x,\xi)$, $E_6(x,\xi)$ и найдены функции Грина соответствующих задач Дирихле в S. В [19] была построена функция Грина задачи Неймана для уравнения Пуассона в S

$$\mathcal{N}_2(x,\xi) = E_2(x,\xi) - E_0(x,\xi), \tag{2}$$

где гармоническая по $x, \xi \in S$ функция $E_0(x,\xi)$ записывается в форме

$$E_0(x,\xi) = \int_0^1 \Big(\hat{E}_2\Big(\frac{x}{|x|}, t|x|\xi\Big) + 1\Big)\frac{dt}{t}$$

и $\hat{E}_2(x,\xi) = \Lambda_x E_2(x,\xi)$, где обозначено $\Lambda u = \sum_{i=1}^n x_i u_{x_i}$, а индекс x указывает, что оператор Λ применяется по переменным x. Нетрудно заметить, что $\Lambda u = \partial u/\partial \nu$ на ∂S. Поскольку $\hat{E}_2(x,\xi) = -(|x|^2 - x\cdot\xi)/|x-\xi|^n$, то функция

$$\hat{E}_2\Big(\frac{x}{|x|}, t|x|\xi\Big) = -\frac{1-(x\cdot\xi)t}{(1-2t(x\cdot\xi)+|x|^2|\xi|^2t^2)^{n/2}}$$

симметрична, и значит, функция $E_0(x,\xi)$, а следовательно, и функция $\mathcal{N}_2(x,\xi)$ тоже симметричны. Функция $\mathcal{N}_2(x,\xi)$ обладает свойствами (см. [8, теорема 3.1] и [13, теорема 3])

$$\Lambda_x\mathcal{N}_2(x,\xi)=\Lambda_xE_2(x,\xi)-(\Lambda_xE_2)\Big(\frac{x}{|x|},|x|\xi\Big)-1,\quad x,\xi\in S,\ x\neq\xi, \tag{3}$$

$$\Lambda_x\mathcal{N}_2(x,\xi)\big|_{\xi\in\partial S}=-\frac{\partial G_2(x,\xi)}{\partial\nu_\xi}-1,\quad x\in S,$$

а поэтому верны равенства

$$\begin{gathered}\int_S\frac{\partial\mathcal{N}_2(x,\xi)}{\partial\nu_x}f(\xi)\,d\xi\big|_{x\in\partial S}=-\int_S f(\xi)\,d\xi,\\ \frac{1}{\omega_n}\int_{\partial S}\frac{\partial\mathcal{N}_2(x,\xi)}{\partial\nu_x}\psi(\xi)\,ds_\xi\big|_{x\in\partial S}=\psi(x)\big|_{\partial S}-\frac{1}{\omega_n}\int_{\partial S}\psi(\xi)\,ds_\xi.\end{gathered} \tag{4}$$

В [13, теорема 3] показано, что решение задачи Неймана для уравнения Пуассона

$$\Delta u(x)=f(x),\ x\in S;\quad \frac{\partial u(x)}{\partial\nu}\big|_{\partial S}=\psi(x),\ x\in\partial S,$$

при выполнении известного условия $\int_{\partial S}\psi(\xi)\,ds_\xi=\int_S f(\xi)\,d\xi$ записывается в виде

$$u(x)=\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}_2(x,\xi)\psi(\xi)\,ds_\xi-\frac{1}{\omega_n}\int_S\mathcal{N}_2(x,\xi)f(\xi)\,d\xi+C.$$

Задача Неймана для полигармонического уравнения исследована в работах [20], [21], а в [22] приведено решение этой задачи.

В настоящей работе определяется элементарное решение полигармонического уравнения и в леммах 1 и 2 приводятся его свойства. В теореме 1 дается интегральное представление функций класса $u\in C^{2m}(D)\cap\cap C^{2m-1}(\bar D)$ в ограниченной области с гладкой границей. Далее исследуется задача Рикье–Неймана (см. [23]). В теореме 2 из разд. 3 определяется функция Грина задачи Рикье–Неймана, а в теореме 3 из разд. 4 находится интегральное представление решения этой задачи. В теореме 4 доказывается, что функция, найденная в теореме 3, действительно представляет собой решение задачи Рикье–Неймана. В теореме 5 из разд. 5 рассматривается частный случай задачи Рикье–Неймана для однородного уравнения и дается пример решения задачи при простых граничных данных.

2. ЭЛЕМЕНТАРНОЕ РЕШЕНИЕ И ИНТЕГРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ

Пусть $m\in\mathbb{N}$. Тогда множество $\mathbb{N}\setminus\{1\}$ можно разбить на два непересекающихся множества $\mathbb{N}_m=\{n\in\in\mathbb{N}:n>2m>1\}\cup(2\mathbb{N}+1)$ и дополнение к нему $\mathbb{N}_m^c=\{2,4,\dots,2m\}$. Поскольку множество $\mathbb{N}_m^c$ – конечное, то $\mathbb{N}_m$ — бесконечное. Ясно, что $\mathbb{N}_{m-1}^c\subset\mathbb{N}_m^c$, а поэтому $\mathbb{N}_m\subset\mathbb{N}_{m-1}$. Определим элементарное решение m-гармонического уравнения $\Delta^m u=0$ в виде

$$E_{2m}(x,\xi)=\begin{cases}\dfrac{(-1)^m|x-\xi|^{2m-n}}{(2-n,2)_m(2,2)_{m-1}}, & n\in\mathbb{N}_m,\\ \dfrac{(-1)^m|x-\xi|^{2m-n}}{(2-n,2)_m^*(2,2)_{m-1}}\Big(\ln|x-\xi|-\sum\limits_{k=1}^{m-n/2}\frac{1}{2k}-\sum\limits_{k=n/2}^{m-1}\frac{1}{2k}\Big), & n\in\mathbb{N}_m^c,\end{cases} \tag{5}$$

где $(a,b)_k=a(a+b)\cdots(a+kb-b)$ – обобщенный символ Похгаммера с соглашением $(a,b)_0=1$, а символ $(a,b)_k^*$ означает, что если среди сомножителей $a,(a+b),\dots,(a+kb-b)$, входящих в $(a,b)_k$, есть 0, то его следует заменить на 1, например, $(-2,2)_3^*=(-2)\cdot1\cdot2=-4$. Кроме того, если в суммах, входящих в (5), верхний индекс становится меньше нижнего, то сумма считается равной нулю. Заметим, что $(2-n,2)_m=(2-n)(4-n)\cdots(2m-n)\neq0$ при $n\in\mathbb{N}_m$, и значит, первая часть формулы (5) определена корректно.

Справедливы следующие простые утверждения.

Лемма 1. *Функция $E_{2m}(x,\xi)$ совпадает с элементарными функциями $E(x,\xi)$,$E_4(x,\xi)$ и $E_6(x,\xi)$ при $m=1$, $m=2$ и $m=3$ соответственно.*

Лемма 2. *Симметричная функция $E_{2m}(x,\xi)$, определенная при $x\neq\xi$, удовлетворяет равенствам*

$$\Delta_\xi E_{2m}(x,\xi)=-E_{2(m-1)}(x,\xi),\quad \Delta_\xi E_2(x,\xi)=0.$$

Приведем интегральное представление функции класса $u \in C^{2m}(D) \cap C^{2m-1}(\bar{D})$, где $D \subset \mathbb{R}^n$ – ограниченная область с гладкой границей ∂D с помощью $E_{2m}(x, \xi)$.

Теорема 1. *Для функции $u \in C^{2m}(D) \cap C^{2m-1}(\bar{D})$ имеет место следующее интегральное представление:*

$$u(x) = \frac{1}{\omega_n} \int_{\partial D} \sum_{k=0}^{m-1} (-1)^k \Big(E_{2k+2}(x,\xi) \frac{\partial \Delta^k u}{\partial \nu} - \frac{\partial E_{2k+2}(x,\xi)}{\partial \nu} \Delta^k u \Big)\, ds_\xi + \frac{(-1)^m}{\omega_n} \int_D E_{2m}(x,\xi) \Delta^m u(\xi)\, d\xi, \quad (6)$$

где $\omega_n = |\partial S|$ – площадь единичной сферы в $\mathbb{R}^n$, ν – внешняя единичная нормаль к ∂D.

Доказательства этих утверждений опустим. Пусть $n \geqslant 3$. В рассуждениях, приводимых ниже, необходима также следующая функция, задаваемая рекуррентно:

$$E^r_{2k}(x,\xi) = \frac{1}{\omega_n} \int_S E^r_{2k-2}(x,y) E_2(y,\xi)\, dy, \quad k \geqslant 2, \quad (7)$$

где $E^r_2(x,\xi) = E_2(x,\xi)$. Например,

$$E^r_4(x,\xi) = \frac{1}{\omega_n} \int_S E_2(x,y) E_2(y,\xi)\, dy, \quad E^r_6(x,\xi) = \frac{1}{\omega_n^2} \int_S E_2(x,\eta) \int_S E_2(\eta,y) E_2(y,\xi)\, dy\, d\eta.$$

Лемма 3. *Функция $E^r_{2m}(x,\xi)$ $(m > 1)$ определена при $\xi, x \in S$, $\xi \neq x$, и имеет, быть может, особенность при $\xi = x$ такую, что $E^r_{2m}(x,\xi) \leqslant C|x-\xi|^{3-n}$, где C – некоторая положительная константа. При $\xi \neq x$ справедливо равенство $\Delta E^r_{2m}(x,\xi) = -E^r_{2m-2}(x,\xi)$.*

По теореме о стирании особенностей (см. [24]) функция $h_{2k}(x,\xi) = E_{2k}(x,\xi) - E^r_{2k}(x,\xi)$ является k-гармонической в S по ξ.

3. ФУНКЦИЯ ГРИНА ЗАДАЧИ РИКЬЕ–НЕЙМАНА

Задача Рикье–Неймана (см. [23]) (в [1] она называется также задачей Неймана-2) заключается в нахождении функции $u \in C^{2m}(S) \cap C^{2m-1}(\bar{S})$, являющейся решением следующей граничной задачи для неоднородного полигармонического уравнения:

$$\Delta^m u(x) = f(x), \quad x \in S,$$
$$\frac{\partial u}{\partial \nu}\Big|_{\partial S} = \varphi_0(\xi), \ \frac{\partial \Delta u}{\partial \nu}\Big|_{\partial S} = \varphi_1(\xi), \dots, \frac{\partial \Delta^{m-1} u}{\partial \nu}\Big|_{\partial S} = \varphi_{m-1}(\xi), \quad \xi \in \partial S. \quad (8)$$

Определение 1. Функцию вида

$$\mathcal{N}_{2m}(x,\xi) = E_{2m}(x,\xi) + g^n_{2m}(x,\xi), \quad m \geqslant 1, \quad (9)$$

где $g^n_{2m}(x,\xi)$ – есть m-гармоническая функция по переменным $x, \xi \in S$ такая, что

$$\frac{\partial \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi \in \partial S} = \dots = \frac{\partial \Delta^{m-2}_\xi \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi \in \partial S} = 0, \ \frac{\partial \Delta^{m-1}_\xi \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi \in \partial S} = (-1)^m, \quad (10)$$

где $x \in S$ назовем *функцией Грина задачи Рикье–Неймана* (8).

Теорема 2. *Функция $\mathcal{N}_{2m}(x,\xi)$ при $\xi, x \in S$ и $m > 1$, определяемая рекуррентно равенством*

$$\mathcal{N}_{2m}(x,\xi) = \frac{1}{\omega_n} \Big(\int_S \mathcal{N}_{2m-2}(x,y) \mathcal{N}_2(y,\xi)\, dy - \frac{1}{\tau_n} \int_S \mathcal{N}_{2m-2}(x,y)\, dy \cdot \int_S \mathcal{N}_2(y,\xi)\, dy \Big), \quad (11)$$

где $\tau_n = |S|$, а $\mathcal{N}_2(x,y)$ – функция Грина задачи Неймана из (2)*, является функцией Грина задачи Рикье–Неймана* (8)*. Функция $\mathcal{N}_{2m}(x,\xi)$ обладает свойством*

$$\Lambda_x \mathcal{N}_{2k}(x,\xi)\big|_{x \in \partial S} = 0, \ k > 1, \quad \Lambda_x \mathcal{N}_2(x,\xi)\big|_{x \in \partial S} = -1, \quad \xi \in S, \quad (12)$$

а $g^n_{2m}(x,\xi)$ из представления (9) *имеет непрерывные производные по ξ в $\bar{S}$ при $x \in S$.*

Доказательство. Проведем доказательство теоремы методом математической индукции по m. При $m = 2$ утверждение теоремы доказано в [13, Теорема 4]. В этом случае из представления

$$\mathcal{N}_4(x,\xi) = \frac{1}{\omega_n}\Big(\int_S \mathcal{N}_2(x,y)\mathcal{N}_2(y,\xi)\,dy - \frac{1}{\tau_n}\int_S \mathcal{N}_2(x,y)\,dy\cdot\int_S \mathcal{N}_2(y,\xi)\,dy\Big),$$

используя (2), нетрудно получить равенство

$$\mathcal{N}_4(x,\xi) = E_4^r(x,\xi) + \hat{g}_4^n(x,\xi),$$

где бигармоническая функция $\hat{g}_4^n(x,\xi)$ определяется как

$$\hat{g}_4^n(x,\xi) = -\frac{1}{\omega_n}\Big(\int_S E_2(x,y)E_0(y,\xi)\,dy + \int_S E_0(x,y)E_2(y,\xi)\,dy - \\ - \int_S E_0(x,y)E_0(y,\xi)\,dy + \frac{1}{\tau_n}\int_S \mathcal{N}_2(x,y)\,dy\cdot\int_S \mathcal{N}_2(y,\xi)\,dy\Big).$$

Действительно, первый и третий интегралы в полученной формуле – гармонические функции по ξ, поскольку особенность в первом интеграле интегрируемая, а дифференцирование эту особенность не увеличивает. Второй и четвертый интегралы, по свойству объемного потенциала, – бигармонические функции по ξ, поскольку гармоническая функция $E_0(x,\xi)$ имеет ограниченные производные по ξ в $\bar{S}$ при $x \in S$. Кроме того, $\hat{g}_4^n(x,\xi)$ в силу особенностей порядка $|\xi - y|^{2-n}$ под интегралами имеет непрерывные производные по ξ в $\bar{S}$ при $x \in S$ (см. [24]). Если вспомнить, что $E_4(x,\xi) = E_4^r(x,\xi) + h_4(x,\xi)$, то будем иметь

$$\mathcal{N}_4(x,\xi) = E_4(x,\xi) + \hat{g}_4^n(x,\xi) - h_4(x,\xi) \equiv E_4(x,\xi) + g_4^n(x,\xi),$$

где $g_4^n(x,\xi)$ в силу свойств $\hat{g}_4^n$ и h_4 имеет ограниченные производные по ξ в $\bar{S}$ при $x \in S$.

Пусть утверждение теоремы верно для некоторого $m \geqslant 2$. Тогда согласно определению для функции Грина $\mathcal{N}_{2m-2}(x,\xi)$ имеет место представление (9):

$$\mathcal{N}_{2m-2}(x,\xi) = E_{2m-2}(x,\xi) + g_{2m-2}^n(x,\xi),$$

где $g_{2m-2}^n(x,\xi)$ – некоторая $(m-1)$-гармоническая функция в S по ξ при $x \in S$, имеющая ограниченные производные по ξ в $\bar{S}$. Если вспомнить функцию $E_{2k}^r(x,\xi)$ из (7) и k-гармоническую функцию $h_{2k}(x,\xi) = E_{2k}(x,\xi) - E_{2k}^r(x,\xi)$, то можем записать

$$\mathcal{N}_{2m-2}(x,\xi) = E_{2m-2}^r(x,\xi) + g_{2m-2}^n(x,\xi) + h_{2m-2}(x,\xi) \equiv E_{2m-2}^r(x,\xi) + \hat{g}_{2m-2}^n(x,\xi). \tag{13}$$

Покажем, что функция

$$\mathcal{N}_{2m}(x,\xi) = \frac{1}{\omega_n}\Big(\int_S \mathcal{N}_{2m-2}(x,y)\mathcal{N}_2(y,\xi)\,dy - \frac{1}{\tau_n}\int_S \mathcal{N}_{2m-2}(x,y)\,dy\cdot\int_S \mathcal{N}_2(y,\xi)\,dy\Big)$$

при $m \geqslant 2$ может быть представлена в виде (9). Используя (13) и имея в виду (2), нетрудно получить, что

$$\mathcal{N}_{2m}(x,\xi) = E_{2m}^r(x,\xi) - \frac{1}{\omega_n}\Big(\int_S E_{2m-2}^r(x,y)E_0(y,\xi)\,dy - \int_S \hat{g}_{2m-2}^n(x,y)E_2(y,\xi)\,dy + \\ + \int_S \hat{g}_{2m-2}^n(x,y)E_0(y,\xi)\,dy + \frac{1}{\tau_n}\int_S \mathcal{N}_2(x,y)\,dy\cdot\int_S \mathcal{N}_2(y,\xi)\,dy\Big),$$

где функция $E_{2k}^r(x,\xi)$ определена в (7). Оценим интегральные члены в полученном равенстве. Первый и третий интегралы – гармонические функции по ξ (особенность в первом интеграле интегрируемая, а дифференцирование эту особенность не увеличивает). Второй интеграл, по свойству объемного потенциала, является m-гармонической функцией по ξ, поскольку $(m-1)$-гармоническая функция $\hat{g}_{2m-2}^n(x,\xi)$ имеет ограниченные производные по ξ в $\bar{S}$ при $x \in S$. Четвертый интеграл, содержащий переменную ξ, является бигармонической функцией в S. Если сумму всех интегральных членов в этом равенстве обозначить через $\hat{g}_{2m}^n(x,\xi)$, то в силу того, что особенности под интегралами имеют порядок не выше $|\xi - y|^{3-n}$ (см. лемму 3), функция $\hat{g}_{2m}^n(x,\xi)$ имеет непрерывные производные по ξ в $\bar{S}$ при $x \in S$. Наконец, если вспомнить, что $h_{2m}(x,\xi) = E_{2m}(x,\xi) - E_{2m}^r(x,\xi)$, то будем иметь (9) при $g_{2m}^n(x,\xi) = \hat{g}_{2m}^n(x,\xi) - h_{2m}(x,\xi)$. Равенство (9) доказано.

Проверим граничные условия для функции $\mathcal{N}_{2m}(x,\xi)$ из определения. В силу симметрии функции $\mathcal{N}_2(x,\xi)$ имеем $\Lambda_\xi \mathcal{N}_2(x,\xi)|_{\xi\in\partial S} = -1$. Поэтому

$$\frac{\partial \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi\in\partial S} = \Lambda_\xi \mathcal{N}_{2m}(x,\xi)\big|_{\xi\in\partial S} = \frac{1}{\omega_n}\Big(- \int_S \mathcal{N}_{2m-2}(x,y)\,dy + \frac{1}{\tau_n}\int_S \mathcal{N}_{2m-2}(x,y)\,dy \cdot \tau_n\Big) = 0.$$

При $x \in S$, по свойству объемного потенциала и в силу непрерывной дифференцируемости функции $\mathcal{N}_{2m-2}(x,\xi)$ по $\xi \in \bar{S}$, но $\xi \neq x$, найдем

$$\Delta_\xi \mathcal{N}_{2m}(x,\xi) = -\mathcal{N}_{2m-2}(x,\xi) + \frac{1}{\tau_n}\int_S \mathcal{N}_{2m-2}(x,y)\,dy. \tag{14}$$

Поэтому по предположению индукции при $x \in S$ имеем

$$\frac{\partial \Delta_\xi^k \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi\in\partial S} = -\frac{\partial \Delta_\xi^{k-1} \mathcal{N}_{2m-2}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi\in\partial S} = 0, \quad k = 1,\dots,m-2,$$

и

$$\frac{\partial \Delta_\xi^{m-1} \mathcal{N}_{2m}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi\in\partial S} = -\frac{\partial \Delta_\xi^{m-2} \mathcal{N}_{2m-2}(x,\xi)}{\partial \nu_\xi}\Big|_{\xi\in\partial S} = -(-1)^{m-1} = (-1)^m,$$

что доказывает равенства (10).

Докажем формулу (12) при $k > 1$, поскольку при $k = 1$ она следует из (3). Пусть $k = 2$, тогда в силу слабой особенности функции $\mathcal{N}_2(x,\xi)$, при $x \in \partial S$ и $\xi \in S$ запишем

$$\Lambda_x \mathcal{N}_4(x,\xi) = \frac{1}{\omega_n}\int_S \Lambda_x \mathcal{N}_2(x,y)\mathcal{N}_2(y,\xi)\,dy - \frac{1}{\tau_n}\int_S \Lambda_x \mathcal{N}_2(x,y)\,dy \cdot \frac{1}{\omega_n}\int_S \mathcal{N}_2(y,\xi)\,dy =$$
$$= -\frac{1}{\omega_n}\int_S \mathcal{N}_2(y,\xi)\,dy + \frac{1}{\omega_n}\int_S \mathcal{N}_2(y,\xi)\,dy = 0.$$

Поэтому при $k > 2$ и $x \in \partial S$, $\xi \in S$, используя формулу (11), будем иметь

$$\Lambda_x \mathcal{N}_{2k}(x,\xi) = \frac{1}{\omega_n}\int_S \Lambda_x \mathcal{N}_{2k-2}(x,y)\mathcal{N}_2(y,\xi)\,dy - \frac{1}{\tau_n}\int_S \Lambda_x \mathcal{N}_{2k-2}(x,y)\,dy \cdot \frac{1}{\omega_n}\int_S \mathcal{N}_2(y,\xi)\,dy = 0,$$

откуда и следует (12). Теорема доказана.

Пример 1. Для нахождения решения задачи Рикье–Неймана с многочленами в граничных условиях или в правой части уравнения необходима следующая формула:

$$\frac{1}{\omega_n}\int_S \mathcal{N}_2(x,\xi)|\xi|^{2l}\,d\xi = -\frac{|x|^{2l+2}}{(2l+2)(2l+n)} + \frac{1}{(2l+2)(n-2)}, \tag{15}$$

где $l \in \mathbb{N}_0$. Докажем ее. В работе [19, замечание 2] была получена аналогичная формула

$$\frac{1}{\omega_n}\int_S \mathcal{N}_2(x,\xi)|\xi|^{2l} H_k(\xi)\,d\xi = -\frac{|x|^{2l+2} - (2l+2+k)/k}{(2l+2)(2l+2k+n)} H_k(x), \tag{16}$$

где $H_k(x)$ – однородный гармонический полином степени k, которая не работает в рассматриваемом случае, так как правая часть в ней не определена при $k = 0$.

Пусть $\{H_k^{(i)}(x) : i = 1,\dots,h_k,\ k \in \mathbb{N}_0\}$ – полная система однородных степени $k \in \mathbb{N}_0$ ортогональных на ∂S сферических гармоник такая, что $\int_{\partial S}(H_k^{(i)}(\xi))^2\,ds_\xi = \omega_n$ (см. [25]) и $h_k = \frac{2k+n-2}{n-2}\binom{k+n-3}{n-3}$ при $n > 2$ ($h_k = 2$ при $n = 2$) – размерность базиса однородных гармонических полиномов степени k. В [8, теорема 1] установлено, что имеет место равенство

$$\frac{1}{\omega_n}\int_S E_2(x,\xi)|\xi|^{2l} H_k(\xi)\,d\xi = -\frac{|x|^{2l+2} H_k(x)}{(2l+2)(2l+2k+n)} + \frac{H_k(x)}{(2l+2)(2k+n-2)},$$

где $k \in \mathbb{N}_0$ и $l \in \mathbb{N}_0$. Отсюда получаем

$$\frac{1}{\omega_n}\int_S E_2(x,\xi)|\xi|^{2l}\,d\xi = -\frac{|x|^{2l+2}}{(2l+2)(2l+n)} + \frac{1}{(2l+2)(n-2)}.$$

В [19, теорема 1] доказано, что при $x \in S$ и $\xi \in \bar{S}$

$$E_0(x,\xi) = -\sum_{k=1}^{\infty} \frac{k+n-2}{k(2k+n-2)} \sum_{i=1}^{h_k} H_k^{(i)}(x) H_k^{(i)}(\xi),$$

причем приведенный ряд сходится равномерно по ξ. Поэтому имеем

$$\int_S E_0(x,\xi)|\xi|^{2l}\,d\xi = -\sum_{k=1}^{\infty} \frac{k+n-2}{k(2k+n-2)} \sum_{i=1}^{h_k} H_k^{(i)}(x) \int_S H_k^{(i)}(\xi)|\xi|^{2l}\,d\xi = 0,$$

и, значит, учитывая (2), получаем доказываемую формулу.

4. РЕШЕНИЕ ЗАДАЧИ РИКЬЕ-НЕЙМАНА

Найдем интегральное представление решения задачи Рикье–Неймана.

Теорема 3. *Пусть функция $u \in C^{2m}(S) \cap C^{2m-1}(\bar{S})$ является решением задачи Рикье–Неймана* (8)*, тогда она может быть представлена в виде*

$$u(x) = \frac{(-1)^{m-1}}{\omega_n} \int_{\partial S} \sum_{k=0}^{m-1} \Big(\Delta_\xi^{m-k-1} \mathcal{N}_{2m}(x,\xi)\Big) \varphi_k(\xi)\,ds_\xi + \frac{(-1)^m}{\omega_n} \int_S \mathcal{N}_{2m}(x,\xi) f(\xi)\,d\xi + C. \tag{17}$$

Доказательство. Пусть $u \in C^{2m}(S) \cap C^{2m-1}(\bar{S})$ – решение задачи Рикье–Неймана (8). Воспользуемся формулой

$$\int_D \big(v\Delta^m u - u\Delta^m v\big)\,d\xi = \int_{\partial D} \sum_{k=0}^{m-1} \Big(\Delta^k v \frac{\partial \Delta^{m-k-1} u}{\partial \nu} - \frac{\partial \Delta^k v}{\partial \nu} \Delta^{m-k-1} u\Big)\,ds_\xi$$

при $D = \{\xi : |\xi| < 1 - \varepsilon\}$, где $\varepsilon > 0$ – достаточно мало ($x \in D$) и $v(\xi) = E_{2m}(x,\xi)$. Аналогично доказательству (6) найдем

$$\int_D E_{2m}(x,\xi) f(\xi)\,d\xi = \int_{\partial D} \sum_{k=0}^{m-1} \Big(\Delta_\xi^k E_{2m}(x,\xi) \frac{\partial \Delta^{m-k-1} u}{\partial \nu} - \frac{\partial \Delta_\xi^k E_{2m}(x,\xi)}{\partial \nu} \Delta^{m-k-1} u\Big)\,ds_\xi + (-1)^m \omega_n u(x).$$

Если теперь опять воспользоваться предыдущей формулой для такой же области D, но при $v(\xi) = g_{2m}^n(x,\xi)$ (m-гармоническая в S функция из (9)), то получим аналогичное равенство с $g_{2m}^n(x,\xi)$ вместо $E_{2m}(x,\xi)$ и без последнего члена справа. Складывая эти равенства, группируя интегральные члены и меняя $k \to m-k-1$, будем иметь

$$u(x) = \frac{(-1)^m}{\omega_n} \int_{|\xi|=1-\varepsilon} \sum_{k=0}^{m-1} \Big(\frac{\partial \Delta_\xi^{m-k-1} \mathcal{N}_{2m}(x,\xi)}{\partial \nu} \Delta^k u - \Delta_\xi^{m-k-1} \mathcal{N}_{2m}(x,\xi) \frac{\partial \Delta^k u}{\partial \nu}\Big)\,ds_\xi +$$
$$+ \frac{(-1)^m}{\omega_n} \int_{|\xi|<1-\varepsilon} \mathcal{N}_{2m}(x,\xi) f(\xi)\,d\xi. \tag{18}$$

Перейдем к пределу при $\varepsilon \to +0$. При этом учтем, что функции $g_{2k}^n(x,\xi)$, а значит, и $\mathcal{N}_{2m}(x,\xi)$ и ее производные по ξ, непрерывны по ξ в $\{\xi : 1-\varepsilon < |\xi| \leqslant 1\}$ при фиксированном $x \in S$. Поэтому в силу свойств функции Грина (10) поверхностные интегралы по $\partial S = \{\xi : |\xi| = 1\}$, содержащие функции $\frac{\partial}{\partial \nu_\xi}(\Delta^{m-k-1}\mathcal{N}_{2m}(x,\xi))$ при $k = 1, \ldots, m-1$ обратятся в 0, а функция $\frac{\partial}{\partial \nu_\xi}(\Delta^{m-1}\mathcal{N}_{2m}(x,\xi))$ обратится в $(-1)^m$. Кроме того, в силу граничных условий задачи Рикье–Неймана будем также иметь $\frac{\partial}{\partial \nu}\Delta^k u(\xi) \to \varphi_k(\xi)$, $\varepsilon \to +0$ при $k = 0, \ldots, m-1$. Таким образом, из (18) в пределе при $\varepsilon \to +0$ получим представление (17) при $C = \frac{1}{\omega_n} \int_{\partial S} u(\xi)\,ds_\xi$.

Лемма 4. 1. *Пусть $f \in C^1(\bar{S})$, тогда функция*

$$u_f(x) = \frac{(-1)^m}{\omega_n} \int_S \mathcal{N}_{2m}(x,\xi) f(\xi)\,ds_\xi \tag{19}$$

является решением следующей задачи Рикье–Неймана:

$$\Delta^m u(x) = f(x),\ x \in S,$$
$$\frac{\partial \Delta^k u}{\partial \nu}\Big|_{\partial S} = 0,\ k = 0, \dots, m-2,\ \frac{\partial \Delta^{m-1} u}{\partial \nu}\Big|_{\partial S} = \frac{1}{\omega_n}\int_S f(\xi)\, d\xi.$$

2. *Пусть* $\psi \in C^{1+\varepsilon}(\partial S)$ $(\varepsilon > 0)$*, тогда функция*

$$v_\psi(x) = \frac{(-1)^{m-1}}{\omega_n}\int_{\partial S} \mathcal{N}_{2m}(x,\xi)\psi(\xi)\, ds_\xi \tag{20}$$

является решением следующей задачи Рикье–Неймана:

$$\Delta^m v(x) = 0, \quad x \in S,$$
$$\frac{\partial \Delta^k v}{\partial \nu}\Big|_{\partial S} = 0,\ k = 0, \dots, m-2,\ \frac{\partial \Delta^{m-1} v}{\partial \nu}\Big|_{\partial S} = \psi(x) - \frac{1}{\omega_n}\int_{\partial S} \psi(\xi)\, ds_\xi.$$

Доказательство. 1. Исследуем функцию $u_f(x)$ без конкретизации гладкости функции $f(x)$. Для этого введем функции

$$\hat{\mathcal{N}}_2(x,\xi) = \mathcal{N}_2(x,\xi) - \frac{1}{\tau_n}\int_S \mathcal{N}_2(y,\xi)\, dy,\ \hat{\mathcal{N}}_{2k}(x,\xi) = \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_{2k-2}(x,y)\hat{\mathcal{N}}_2(y,\xi)\, dy,\ k > 1.$$

Функция $\hat{\mathcal{N}}_2(x,\xi)$ обладает свойством

$$\int_S \hat{\mathcal{N}}_2(x,\xi)\, dx = \int_S \mathcal{N}_2(x,\xi)\, dx - \frac{\tau_n}{\tau_n}\int_S \mathcal{N}_2(y,\xi)\, dy = 0.$$

Нетрудно видеть, что верно равенство

$$\mathcal{N}_{2m}(x,\xi) = \frac{1}{\omega_n}\int_S \mathcal{N}_{2m-2}(x,y)\Big(\mathcal{N}_2(y,\xi) - \frac{1}{\tau_n}\int_S \mathcal{N}_2(y_1,\xi)\, dy_1\Big)\, dy = \frac{1}{\omega_n}\int_S \mathcal{N}_{2m-2}(x,y)\hat{\mathcal{N}}_2(y,\xi)\, dy. \tag{21}$$

Поэтому в силу свойств объемного потенциала, переставляя интегралы, будем иметь

$$\Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_4(x,\xi) f(\xi)\, d\xi = \Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_2(x,y)\, dy \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(y,\xi) f(\xi)\, d\xi = -\frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(x,\xi) f(\xi)\, d\xi,$$

откуда аналогично следует

$$\Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_6(x,\xi) f(\xi)\, d\xi = \Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_4(x,y)\, dy \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(y,\xi) f(\xi)\, d\xi =$$
$$= -\frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(x,y)\, dy \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(y,\xi) f(\xi)\, d\xi = -\frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_4(y,\xi) f(\xi)\, d\xi.$$

Используя предыдущие формулы, по индукции, меняя порядок интегрирования будем иметь

$$\Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_{2m}(x,\xi) f(\xi)\, d\xi = \Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_{2m-2}(x,y)\, dy \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(y,\xi) f(\xi)\, d\xi =$$
$$= -\frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_{2m-4}(x,y)\, dy \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(y,\xi) f(\xi)\, d\xi = -\frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_{2m-2}(y,\xi) f(\xi)\, d\xi. \tag{22}$$

Из последней формулы, используя свойство объемного потенциала, нетрудно получить

$$\Delta_x^m u_f(x) = \Delta_x^m \frac{(-1)^m}{\omega_n}\int_S \mathcal{N}_{2m}(x,\xi) f(\xi)\, d\xi = -\Delta_x \frac{1}{\omega_n}\int_S \hat{\mathcal{N}}_2(x,\xi) f(\xi)\, d\xi =$$
$$= -\Delta_x \frac{1}{\omega_n}\int_S \mathcal{N}_2(x,\xi) f(\xi)\, d\xi = f(x). \tag{23}$$

Далее, в силу слабой особенности функции $\mathcal{N}_{2m}(x,\xi)$ (особенность такая же, как у элементарного решения $E_{2m}(x,\xi)$ (9)) ее можно дифференцировать под знаком интеграла, и поэтому из (12) следует, что при $m>1$

$$\frac{\partial u_f}{\partial \nu}\Big|_{\partial S}=\Lambda_x u_f(x)\big|_{\partial S}=\frac{(-1)^m}{\omega_n}\int_S \Lambda_x \mathcal{N}_{2m}(x,\xi)\big|_{x\in\partial S} f(\xi)\,d\xi=0.$$

Кроме того, в силу (22) при $k=1,\dots,m-2$ получим

$$\frac{\partial \Delta^k u_f}{\partial \nu}\Big|_{\partial S}=\Lambda_x\Delta_x^k u_f(x)\big|_{\partial S}=\frac{(-1)^{m-k}}{\omega_n}\int_S \Lambda_x \hat{\mathcal{N}}_{2m-2k}(x,\xi)\big|_{x\in\partial S} f(\xi)\,d\xi=$$
$$=\frac{(-1)^{m-k}}{\omega_n}\int_S \Lambda_x \mathcal{N}_{2m-2k}(x,\xi)\big|_{x\in\partial S} f(\xi)\,d\xi=0.$$

Аналогично сделанному выше, в соответствии с (22) и (12) найдем

$$\frac{\partial \Delta^{m-1} u_f}{\partial \nu}\Big|_{\partial S}=-\frac{1}{\omega_n}\int_S \Lambda_x \hat{\mathcal{N}}_2(x,\xi)\big|_{x\in\partial S} f(\xi)\,d\xi=\frac{1}{\omega_n}\int_S f(\xi)\,d\xi.$$

Это значит, что функция $u_f(x)$ из (19) является решением задачи Рикье–Неймана (8) при $\varphi_k=0$, $k=0,\dots,m-2$ и $\varphi_{m-1}=\frac{1}{\omega_n}\int_S f(\xi)\,d\xi$.

Какую гладкость достаточно наложить на функцию $f(x)$, чтобы рассуждения сделанные относительно $u_f(x)$, были справедливы? Для выполнения последнего равенства из (23) для объемного потенциала достаточно, чтобы $f\in C^1(\bar S)$ (см. [16]). При этом объемный потенциал с плотностью $f(x)$ будет также из $C^1(\bar S)$ [24], и значит, вся цепочка равенств из (23) верна.

2. Рассмотрим функцию $v_\psi(x)$ из (20). Подставляя значение $\mathcal{N}_{2m}(x,\xi)$ из (21) и меняя порядок итегрирования, будем иметь

$$v_\psi(x)=\frac{(-1)^{m-1}}{\omega_n}\int_S \mathcal{N}_{2m-2}(x,y)\,dy\,\frac{1}{\omega_n}\int_{\partial S}\hat{\mathcal{N}}_2(y,\xi)\psi(\xi)\,ds_\xi.$$

В силу полученных выше свойств функций $u_f(x)$ и $\hat{\mathcal{N}}_2(x,\xi)$ функция $v_\psi(x)$ является решением задачи Рикье–Неймана (8) при $m=m-1$, $\varphi_k=0$, $k=0,\dots,m-3$,

$$\varphi_{m-2}=\frac{1}{\omega_n}\int_S\frac{1}{\omega_n}\int_{\partial S}\hat{\mathcal{N}}_2(y,\xi)\psi(\xi)\,ds_\xi\,dy=\frac{1}{\omega_n}\int_{\partial S}\psi(\xi)\,ds_\xi\frac{1}{\omega_n}\int_S\hat{\mathcal{N}}_2(y,\xi)\,dy=0$$

и

$$f(x)=\Delta^{m-1}v(x)=\frac{1}{\omega_n}\int_{\partial S}\hat{\mathcal{N}}_2(x,\xi)\psi(\xi)\,ds_\xi.$$

Поэтому по свойству потенциала простого слоя будем иметь

$$\Delta^m v_\psi(x)=\Delta\Delta^{m-1}v_\psi(x)=\Delta\frac{1}{\omega_n}\int_{\partial S}\hat{\mathcal{N}}_2(x,\xi)\psi(\xi)\,ds_\xi=0,$$

и, кроме того, по свойству (4) функции Грина $\mathcal{N}_2(x,\xi)$ запишем

$$\frac{\partial \Delta^{m-1}v_\psi}{\partial\nu}\Big|_{\partial S}=\psi(x)-\frac{1}{\omega_n}\int_{\partial S}\psi(\xi)\,ds_\xi.$$

Таким образом, функция $v_\psi(x)$ является решением задачи Рикье–Неймана (8) при $f=0$, $\varphi_k=0$, $k=0,\dots,m-2$ и $\varphi_{m-1}(x)=\psi(x)-\frac{1}{\omega_n}\int_{\partial S}\psi(\xi)\,ds_\xi$, что и утверждалось.

Какая гладкость функции $\psi(x)$ достаточна для правомерности сделанных выше рассуждений? Поскольку функция $v_\psi(x)$ была представлена в виде, аналогичном виду функции $u_f(x)$ с плотностью $\frac{1}{\omega_n}\int_S\hat{\mathcal{N}}_2(y,\xi)\psi(\xi)\,ds_\xi$, то также, как и в предыдущем случае, достаточно, чтобы эта плотность была из $C^1(\bar S)$, а это выполнено, если $\psi\in C^{1+\varepsilon}(\partial S)$ (см. [26]). Лемма доказана.

Замечание 1. При $m=1$ для функции ψ достаточно потребовать, чтобы $\psi\in C(\partial S)$.

Теорема 4. *Пусть* $\varphi_0\in C(\partial S)$, $\varphi_k\in C^{1+\varepsilon}(\partial S)$ *при* $k=1,\dots,m-1$ *и* $f\in C^1(\bar S)$. *Тогда функция*

$$u(x)=\frac{1}{\omega_n}\int_{\partial S}\sum_{k=0}^{m-1}(-1)^k\Big(\mathcal{N}_{2k+2}(x,\xi)-\frac{1}{\tau_n}\int_S\mathcal{N}_{2k+2}(x,y)\,dy\Big)\varphi_k(\xi)\,ds_\xi+\frac{(-1)^m}{\omega_n}\int_S\mathcal{N}_{2m}(x,\xi)f(\xi)\,d\xi+C \quad (24)$$

является решением задачи Рикье–Неймана (8) *при условии, что*

$$\int_{\partial S}\varphi_{m-1}(\xi)\,ds_\xi=\int_S f(\xi)\,d\xi. \tag{25}$$

Доказательство. Докажем, что функция $u(x)$, определяемая по формуле (17), является решением задачи

$$\begin{gathered}\Delta^m u(x)=f(x),\quad x\in S,\\ \frac{\partial\Delta^k u}{\partial\nu}\Big|_{\partial S}=\varphi_k,\quad k=0,\dots,m-2,\\ \frac{\partial\Delta^{m-1}u}{\partial\nu}\Big|_{\partial S}=\varphi_{m-1}(x)-\frac{1}{\omega_n}\int_{\partial S}\varphi_{m-1}(\xi)\,ds_\xi+\frac{1}{\omega_n}\int_S f(\xi)\,d\xi.\end{gathered} \tag{26}$$

Нетрудно видеть, что в этом случае при выполнении условия (25) следует утверждение теоремы. Доказательство этого утверждения проведем индукцией по m. При $m=2$ в [13, теорема 5] доказано, что для функции вида

$$u(x)=-\frac{1}{\omega_n}\int_{\partial S}\Delta_\xi\mathcal{N}_4(x,\xi)\varphi_0(\xi)\,ds_\xi-\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}_4(x,\xi)\varphi_1(\xi)\,ds_\xi+$$
$$+\frac{1}{\omega_n}\int_S\mathcal{N}_4(x,\xi)f(\xi)\,d\xi+C\equiv v_1(x)+v_2(x)+v_3(x)+C$$

верны равенства

$$\begin{gathered}\Delta^2u(x)=0+0+f(x)=f(x),\\ \frac{\partial u}{\partial\nu}\big|_{x\in\partial S}=\varphi_0(x)+0+0=\varphi_0(x),\\ \frac{\partial\Delta u}{\partial\nu}\big|_{x\in\partial S}=0+\Big(\varphi_1(x)-\frac{1}{\omega_n}\int_{\partial S}\varphi_1(\xi)\,d\xi\Big)+\frac{1}{\omega_n}\int_S f(\xi)\,d\xi.\end{gathered}$$

Здесь, в формулах справа, указан результат применения соответствующих операторов к функциям v_1, v_2 и v_3. Это соответствует доказываемому утверждению при $m=2$. Предположим верность утверждения (26) при $m=m-1$. Представим $u(x)$ из (17) в виде

$$u(x)=u_m(x)+u_f^{(m)}(x)+C, \tag{27}$$

где функция $u_f^{(m)}(x)$ определена в (19) (здесь дополнительно введен верхний индекс m, чтобы избежать путаницы) и

$$u_m(x)=\frac{(-1)^{m-1}}{\omega_n}\int_{\partial S}\sum_{k=0}^{m-1}\Big(\Delta_\xi^{m-k-1}\mathcal{N}_{2m}(x,\xi)\Big)\varphi_k(\xi)\,ds_\xi.$$

Заметим, что функция $u_f(x)$ при $f\in C^1(\bar S)$ по лемме 4 является решением задачи (26) при $\varphi_k=0$, $k=0,\dots,m-1$. Поэтому достаточно доказать, что функция $u_m(x)=u(x)-u_f^{(m)}(x)+C$ является решением задачи (26) при $f=0$. Преобразуем функцию $u_m(x)$. Из равенства (14) при $k>1$ нетрудно получить

$$\Delta_\xi^k\mathcal{N}_{2m}(x,\xi)=-\Delta_\xi^{k-1}\mathcal{N}_{2m-2}(x,\xi),$$

а поэтому

$$\begin{aligned}u_m(x)&=\frac{(-1)^{m-1}}{\omega_n}\Big(\int_{\partial S}\mathcal{N}_{2m}(x,\xi)\varphi_{m-1}(\xi)\,ds_\xi+\int_{\partial S}\Delta_\xi\mathcal{N}_{2m}(x,\xi)\varphi_{m-2}(\xi)\,ds_\xi\Big)+\\&+\frac{(-1)^{m-2}}{\omega_n}\Big(\int_{\partial S}\sum_{k=0}^{m-2}\Big(\Delta_\xi^{m-k-2}\mathcal{N}_{2m-2}(x,\xi)\Big)\varphi_k(\xi)\,ds_\xi-\int_{\partial S}\mathcal{N}_{2m-2}(x,\xi)\varphi_{m-2}(\xi)\,ds_\xi\Big)=\\&=u_{m-1}(x)+\frac{(-1)^{m-1}}{\omega_n}\Big(\int_{\partial S}\mathcal{N}_{2m}(x,\xi)\varphi_{m-1}(\xi)\,ds_\xi+\int_{\partial S}\Delta_\xi\mathcal{N}_{2m}(x,\xi)\varphi_{m-2}(\xi)\,ds_\xi+\\&\qquad+\int_{\partial S}\mathcal{N}_{2m-2}(x,\xi)\varphi_{m-2}(\xi)\,ds_\xi\Big).\end{aligned}$$

Используя равенство (14) и обозначения из (19) и (20), можем записать

$$u_m(x) = u_{m-1}(x) + v_{\varphi_{m-1}}(x) + \frac{(-1)^{m-1}}{\omega_n}\int_S \mathcal{N}_{2m-2}(x,y)\,dy \cdot \frac{1}{\tau_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi = \\ = u_{m-1}(x) + v_{\varphi_{m-1}}(x) + u_1^{(m-1)}(x)\cdot\frac{1}{\tau_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi. \quad (28)$$

Для законности использования функции $v_{\varphi_{m-1}}(x)$ по лемме 4 достаточно потребовать $\varphi_0 \in C(\partial S)$, $\varphi_k \in C^{1+\varepsilon}(\partial S)$ при $k = 1,\dots,m-1$. Поскольку по предположению индукции функция $u_{m-1}(x)$ является $(m-1)$-гармонической, то согласно лемме 4 можно записать

$$\Delta^m u_m(x) = 0 + 0 + \Delta 1 \cdot \frac{1}{\tau_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi = 0.$$

Кроме того, из (28) также следует, что на ∂S в силу леммы 4 и предположения индукции верны равенства

$$\frac{\partial \Delta^k u_m}{\partial \nu} = \frac{\partial \Delta^k u_{m-1}}{\partial \nu} + 0 + 0 = \varphi_k(x), \quad k = 0,\dots,m-3,$$
$$\frac{\partial \Delta^{m-2} u_m}{\partial \nu} = \frac{\partial \Delta^{m-2} u_{m-1}}{\partial \nu} + 0 + \frac{1}{\omega_n}\int_S \frac{1}{\tau_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi\,dy =$$
$$= \varphi_{m-2}(x) - \frac{1}{\omega_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi + \frac{1}{\omega_n}\int_{\partial S}\varphi_{m-2}(\xi)\,ds_\xi = \varphi_{m-2}(x),$$
$$\frac{\partial \Delta^{m-1} u_m}{\partial \nu} = 0 + \Big(\varphi_{m-1}(x) - \frac{1}{\omega_n}\int_{\partial S}\varphi_{m-1}(\xi)\,ds_\xi\Big) + 0 = \varphi_{m-1}(x) - \frac{1}{\omega_n}\int_{\partial S}\varphi_{m-1}(\xi)\,ds_\xi,$$

а значит, функция $u_m(x)$ является решением задачи (26) при $f = 0$. Шаг индукции доказан, и, значит, функция из (17) является решением задачи (8).

Из формулы (14) нетрудно получить равенство

$$\Delta_\xi^k \mathcal{N}_{2m}(x,\xi) = (-1)^k\Big(\mathcal{N}_{2m-2k}(x,\xi) - \frac{1}{\tau_n}\int_S \mathcal{N}_{2m-2k}(x,y)\,dy\Big),$$

с помощью которого формула (17) преобразуется к виду (24). Теорема доказана.

Замечание 2. Необходимое и достаточное условие разрешимости задачи Рикье–Неймана (25) для полигармонического уравнения ранее было получено в [23]. Из доказательства теоремы 4 следует, что если условие (25) не выполнено, то функция $u(x)$ из (24) является решением задачи Рикье–Неймана (26).

5. ЧАСТНЫЙ СЛУЧАЙ

Рассмотрим один частный случай граничных условий задачи Рикье–Неймана.

Теорема 5. *Пусть* $\varphi_0 \in C(\partial S)$, $\varphi_k \in C^{1+\varepsilon}(\partial S)$ *и* $\int_{\partial S}\varphi_k(\xi)\,ds_\xi = 0$ *при* $k = 1,\dots,m-1$, *а* $f = 0$. *Тогда решение задачи Рикье–Неймана* (8) *существует и его можно записать в виде*

$$u(x) = \sum_{k=0}^{m-1} u_k[\varphi_k](x) + C, \quad (29)$$

где обозначено

$$u_k[\varphi](x) = \frac{(-1)^k}{\omega_n}\int_{\partial S}\mathcal{N}^0_{2k+2}(x,\xi)\varphi(\xi)\,ds_\xi,$$
$$\mathcal{N}^0_{2k+2}(x,\xi) = \frac{1}{\omega_n}\int_S \mathcal{N}_2(x,y)\mathcal{N}^0_{2k}(y,\xi)\,dy,\ k > 0; \quad \mathcal{N}^0_2(x,\xi) = \mathcal{N}_2(x,\xi).$$

Для $(k+1)$*-гармонических функций* $u_k[\varphi_k](x)$ *выполнены условия*

$$\Delta u_k[\varphi](x) = u_{k-1}[\varphi](x),\ \frac{\partial u_k[\varphi]}{\partial \nu}\Big|_{\partial S} = 0, k \in \mathbb{N}; \quad \Delta u_0[\varphi](x) = 0,\ \frac{\partial u_0[\varphi]}{\partial \nu}\Big|_{\partial S} = \varphi(x). \quad (30)$$

Доказательство. Нетрудно видеть, что при требуемых в теореме условиях все условия теоремы 4 тоже выполнены, а значит, решение задачи Рикье–Неймана существует. Докажем, что формула (29) задает это решение. Для этого достаточно убедиться в справедливости равенств (30). Согласно определению функции $\mathcal{N}^0_{2k}(x,\xi)$ и по свойству о бъемного потенциала при $k>0$ запишем

$$\Delta u_k[\varphi](x)=\Delta_x\frac{(-1)^k}{\omega_n}\int_{\partial S}\mathcal{N}^0_{2k+2}(x,\xi)\varphi(\xi)\,ds_\xi=\frac{(-1)^{k-1}}{\omega_n}\int_{\partial S}\mathcal{N}^0_{2k}(x,\xi)\varphi(\xi)\,ds_\xi=u_{k-1}[\varphi](x),$$

а также найдем $\Delta u_0[\varphi](x)=0$. Кроме того, при $k>0$, используя (4), получим

$$\frac{\partial u_k[\varphi]}{\partial\nu}\Big|_{\partial S}=\frac{(-1)^k}{\omega_n}\int_S\Lambda_x\mathcal{N}_2(x,y)\big|_{x\in\partial S}\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}^0_{2k}(y,\xi)\varphi(\xi)\,ds_\xi\,dy=$$
$$=\frac{(-1)^{k-1}}{\omega_n}\int_{\partial S}\Big(\int_S\mathcal{N}^0_{2k}(y,\xi)\,dy\Big)\varphi(\xi)\,ds_\xi\equiv\int_{\partial S}F(\xi)\varphi(\xi)\,ds_\xi.$$

Воспользовавшись формулой

$$\frac{1}{\omega_n}\int_S\mathcal{N}^0_{2k}(y_1,\xi)\,dy_1=\frac{1}{\omega_n}\int_S\mathcal{N}_2(y_1,y_2)\,dy_1\cdots\frac{1}{\omega_n}\int_S\mathcal{N}_2(y_k,\xi)\,dy_k,$$

симметричностью функция $\mathcal{N}_2(x,\xi)$ и формулой (15) убеждаемся, что функция

$$F(\xi)=\frac{(-1)^{k-1}}{\omega_n}\int_S\mathcal{N}^0_{2k}(y,\xi)\,dy$$

является многочленом степени k по $|\xi|^2$, и, значит, $F(\xi)|_{\partial S}=C$, а поэтому по условию теоремы

$$\frac{\partial u_k[\varphi_k]}{\partial\nu}\Big|_{\partial S}=\int_{\partial S}F(\xi)\varphi_k(\xi)\,ds_\xi=C\int_{\partial S}\varphi_k(\xi)\,ds_\xi=0.$$

Если $k=0$, то согласно (4)

$$\frac{\partial u_0[\varphi_0]}{\partial\nu}\Big|_{\partial S}=\frac{1}{\omega_n}\int_{\partial S}\frac{\partial\mathcal{N}_2(x,\xi)}{\partial\nu_x}\big|_{\partial S}\varphi_0(\xi)\,ds_\xi=\varphi_0(x)\big|_{\partial S}-\frac{1}{\omega_n}\int_{\partial S}\varphi_0(\xi)\,ds_\xi=\varphi_0(x).$$

Теорема доказана.

Пример 2. Вычислим функции $u_p[H_k](x)$ из формулы (29) при $p\in\mathbb{N}_0$, где $H_k(x)$ – однородный гармонический полином степени $k\in\mathbb{N}$. В этом случае условия теоремы 5 выполнены, поскольку справедливо равенство $\int_{\partial S}H_k(\xi)\,ds_\xi=0$.

В работе [19] было установлено, что при $x\in S$ и $\xi\in\partial S$ верны равенства

$$\mathcal{N}_2(x,\xi)=E(x,\xi)-E_0(x,\xi)=\frac{1}{n-2}+$$
$$+\sum_{k=1}^{\infty}\Big(\frac{1}{2k+n-2}+\frac{k+n-2}{k(2k+n-2)}\Big)\sum_{i=1}^{h_k}H_k^{(i)}(x)H_k^{(i)}(\xi)=\frac{1}{n-2}+\sum_{k=1}^{\infty}\frac{1}{k}\sum_{i=1}^{h_k}H_k^{(i)}(x)H_k^{(i)}(\xi),$$

где гармонические полиномы $H_k^{(i)}(x)$ определены в примере 1. Поэтому в силу ортонормируемости полиномов $H_k^{(i)}(x)$ на ∂S и равномерной сходимости ряда по $\xi\in\partial S$ имеем

$$u_0[H_k](x)=\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}_2(x,\xi)H_k(\xi)\,ds_\xi=\frac{1}{(n-2)\omega_n}\int_{\partial S}H_k(\xi)\,ds_\xi+$$
$$+\sum_{m=1}^{\infty}\frac{1}{m}\sum_{i=1}^{h_m}H_m^{(i)}(x)\frac{1}{\omega_n}\int_{\partial S}H_m^{(i)}(\xi)H_k(\xi)\,ds_\xi=\frac{1}{k}H_k(x).$$

Вычислим $u_1[H_k](x)$. С помощью (16) при $l=0$ и предыдущих вычислений найдем

$$u_1[H_k](x)=-\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}^0_4(x,\xi)H_k(\xi)\,ds_\xi=-\frac{1}{\omega_n}\int_S\mathcal{N}_2(x,y)\frac{1}{\omega_n}\int_{\partial S}\mathcal{N}_2(y,\xi)H_k(\xi)\,ds_\xi\,dy=$$
$$-\frac{1}{\omega_n}\int_S\mathcal{N}_2(x,y)u_0[H_k](y)\,dy=-\frac{1}{k}\frac{1}{\omega_n}\int_S\mathcal{N}_2(x,y)H_k(y)\,dy=\frac{1}{k}\frac{|x|^2-1-2/k}{2(2k+n)}H_k(x).$$

Аналогично в общем случае для $u_p[H_k](x)$ при $p \in \mathbb{N}$ имеем

$$u_p[H_k](x) = \frac{1}{\omega_n}\int_{\partial S}\mathcal{N}^0_{2p+2}(x,\xi)H_k(\xi)\,ds_\xi = -\frac{1}{\omega_n}\int_S \mathcal{N}_2(x,y)u_{p-1}[H_k](y)\,dy. \tag{31}$$

Отсюда, используя найденную выше функцию $u_1[H_k](x)$ и формулу (16), получим

$$u_2[H_k](x) = \Big(\frac{|x|^4 - 1 - 4/k}{8k(2k+n)(2k+2+n)} - (k+2)\frac{|x|^2 - 1 - 2/k}{4k^2(2k+n)^2}\Big)H_k(x).$$

Нетрудно убедиться, что 3-гармоническая функция $u_2[H_k](x)$ удовлетворяет условию

$$\frac{\partial u_2[H_k]}{\partial \nu}\Big|_{\partial S} = \Big(\frac{(k+4)|x|^4 - k - 4}{8k(2k+n)(2k+2+n)} - (k+2)\frac{(k+2)|x|^2 - k - 2}{4k^2(2k+n)^2}\Big)H_k(x)\Big|_{\partial S} = 0,$$

и поскольку

$$\Delta u_2[H_k] = \Big(\frac{|x|^2}{2k(2k+n)} - \frac{(k+2)}{2k^2(2k+n)}\Big)H_k(x) = u_1[H_k],$$

то

$$\frac{\partial \Delta u_2[H_k]}{\partial \nu}\Big|_{\partial S} = \frac{(k+2)|x|^2 - k - 2}{2k(2k+n)}H_k(x)\Big|_{\partial S} = 0.$$

Кроме того, верны равенства

$$\Delta^2 u_2[H_k] = \frac{1}{k}H_k(x) = u_0[H_k], \quad \frac{\partial \Delta^2 u_2[H_k]}{\partial \nu}\Big|_{\partial S} = H_k(x).$$

Используя (31) и (16), можно последовательно найти любую функцию $u_p[H_k](x)$.

СПИСОК ЛИТЕРАТУРЫ

1. *Begehr H.* Biharmonic Green functions // Le Matematiche. 2006. V. LXI. P. 395–405.
2. *Begehr H., Vaitekhovich T.* Modified harmonic Robin function // Complex Var. and Elliptic Equat. 2013. V. 58. № 4. P. 483–496.
3. *Sadybekov M.A., Torebek B.T., Turmetov B.Kh.* On an explicit form of the Green function of the Robin problem for the Laplace operator in a circle // Adv. Pure Appl. Math. 2015. V. 6. № 3. P. 163–172.
4. *Ying Wang, Liuqing Ye.* Biharmonic Green function and biharmonic Neumann function in a sector // Complex Var. Elliptic. Equat. 2013. V. 58. № 1. P. 7–22.
5. *Ying Wang* Tri-harmonic boundary value problems in a sector // Complex Var. Elliptic. Equat. 2014. V. 59. № 5. P. 732–749.
6. *Boggio T.* Sulle funzioni di Green d'ordine m // Palermo Rend. 1905. V. 20. P. 97–135.
7. *Kalmenov T.Sh., Koshanov B.D., Nemchenko M.Y.* Green function representation for the Dirichlet problem of the polyharmonic equation in a sphere // Complex Var. Elliptic. Equat. 2008. V. 53. P. 177–183.
8. *Karachik V.V., Turmetov B.Kh.* On Green's function of the Robin problem for the Poisson equation // Adv. in Pure and Appl. Math. 2019. V. 10. № 3. C. 203–214.
9. *Карачик В.В.* Функция Грина задачи Дирихле для 3-гармонического уравнения в шаре // Матем. заметки. 2020. V. 107. № 1. С. 87–105.
10. Карачик В.В., Торебек Б.Т. О задаче Дирихле—Рикье для бигармонического уравнения // Матем. заметки. 2017. Т. 102. № 1. С. 39–51.
11. *Карачик В.В.* Об одной задаче типа Неймана для бигармонического уравнения // Матем. тр. 2016. Т. 19. № 2. С. 86–108.
12. *Солдатов А.П.* О фредгольмовости и индексе обобщённой задачи Неймана // Дифференц. ур-ния. 2020. Т. 56. № 2. С. 217–225.

13. *Карачик В.В.* Функции Грина задач Навье и Рикье–Неймана для бигармонического уравнения в шаре // Дифференц. ур-ния. 2021. Т. 57. № 5. Р. 673–686.

14. *Sweers G.* A survey on boundary conditions for the biharmonic // Complex Var. and Ellipt. Equat. 2009. V. 54. P. 79–93.

15. *Karachik V., Turmetov B., Yuan H.* Four Boundary Value Problems for a Nonlocal Biharmonic Equation in the Unit Ball // Mathematics. 2022. V. 10. № 7. P. 1–21.

16. *Бицадзе А.В.* Уравнения математической физики. М.: Наука, 1982.

17. *Karachik V.V.* Greens function of Dirichlet problem for biharmonic equation in the ball // Complex Var. and Ellipt. Equat. 2019. V. 64. № 9. P. 1500–1521.

18. *Карачик В.В.* О функции Грина задачи Дирихле для бигармонического уравнения в шаре // Ж. вычисл. матем. и матем. физ. 2019. Т. 59. № 1. С. 71–86.

19. *Карачик В.В., Турметов Б.Х.* О функции Грина третьей краевой задачи для уравнения Пуассона // Матем. тр. 2018. Т. 21. № 1. С. 17–34.

20. *Бицадзе А.В.* К задаче Неймана для гармонических функций // Докл. АН СССР. 1990. Т. 311. № 1. С. 11–13.

21. *Карачик В.В.* Об арифметическом треугольнике, возникающем из условий разрешимости задачи Неймана // Матем. заметки. 2014. Т. 96. № 2. С. 228–238.

22. *Karachik V.V.* Dirichlet and Neumann boundary value problems for the polyharmonic equation in the unit ball // Mathematics. 2021. V. 9. № 16. 1907.

23. *Карачик В.В.* Задача Рикье–Неймана для полигармонического уравнения в шаре // Дифференц. ур-ния. 2018. Т. 54. № 5. С. 653–662.

24. *Владимиров В.С.* Уравнения математической физики М.: Наука, 1981

25. *Karachik V.V.* On one set of orthogonal harmonic polynomials // Proc. of the Am. Math Soc. 1998. V. 126. № 12. P. 3513–3519.

26. *Алимов Ш.А.* Об одной задаче с наклонной производной // Дифференц. ур-ния 1981. Т. 17. № 10. С. 1738–1751.

GREEN'S FUNCTION FOR THE RIEMANN–NEUMANN PROBLEM FOR A POLYHARMONIC EQUATION IN THE UNIT SPHERE

V. V. Karachik*

South Ural State University (NIU), Lenina Ave. 76, Chelyabinsk, 454080, Russia
**e-mail: karachik@susu.ru*

Received 10 January, 2023
Revised 10 January, 2023
Accepted 06 February, 2024

Abstract. The Green's function for the Riemann–Neumann problem for a polyharmonic equation in the unit sphere is constructed, and an integral representation of the solutions to the Riemann–Neumann problem is provided. Two examples are presented.

Keywords: polyharmonic equation, Riemann–Neumann problem, Green's function.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 804–818

УДК 519.622

ОБ ОДНОМ МЕТОДЕ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ КОШИ ДЛЯ СИНГУЛЯРНО ВОЗМУЩЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ[1)]

© 2024 г. Д. А. Маслов[1,*]

[1] 111250 Москва, ул. Красноказарменная, 14, НИУ МЭИ, Россия
**e-mail: maslovdma@mpei.ru*

Поступила в редакцию 10.11.2023 г.
Переработанный вариант 10.11.2023 г.
Принята к публикации 14.01.2024 г.

В работе предлагается новый способ численного решения нелинейных жестких задач, основанный на численной реализации метода голоморфной регуляризации задачи Коши для сингулярно возмущенных нелинейных дифференциальных уравнений. Библ. 31. Фиг. 6. Табл. 4.

Ключевые слова: сингулярно возмущенное дифференциальное уравнение, задача Коши, жесткость, нелинейность, метод голоморфной регуляризации, численное решение.

DOI: 10.31857/S0044466924050099, **EDN**: YDEVYQ

ВВЕДЕНИЕ

Существует много важных прикладных задач, описываемых задачей Коши для жестких систем дифференциальных уравнений: задачи химической кинетики, задачи моделирования нестационарных процессов в электрических цепях и др. [1]. Примером жесткой задачи является задача Коши для сингулярно возмущенного дифференциального уравнения [2]. К такой задаче применимы методы численного решения жестких задач, а также численные реализации асимптотических методов [1]. Подробному обзору методов численного решения жестких задач посвящена обширная литература, см., например, [1–4]. Многие методы и рекомендации по решению жестких задач хорошо подходят для линейных и слабонелинейных жестких задач, однако, для существенно нелинейных сверхжестких задач они становятся ненадежными, требуют сильного уменьшения шага в некоторые критические моменты, причем для определения этих моментов не разработаны достаточно надежные алгоритмы [5]. Применение явных методов типа Рунге–Кутты и Адамса требует выбора неприемлемо малого шага, гарантирующего устойчивость численного решения, что приводит к возрастанию трудоемкости и делает данные методы непригодными для решения жестких задач, и тем более неприемлемыми для нелинейных сверхжестких задач. Поэтому для решения жестких задач используют так называемые A-устойчивые методы, которые не накладывают ограничений на шаг. Среди явных многошаговых методов не существует A-устойчивых, а порядок неявных A-устойчивых многошаговых методов не может быть выше второго (барьер Далквиста [1]). Для решения сверхжестких задач требуется также Lp-устойчивость, что дополнительно сужает класс допустимых методов численного решения жестких задач. Многошаговые методы требовательны к выбору стартовых приближений, чувствительны к адаптивному изменению шага и уступают в устойчивости неявным схемам и схемам типа Розенброка [6]. Неявные методы Рунге-Кутты применяются достаточно редко, так как требуют итерационного процесса на каждом шаге с вытекающей из этого проблемой сходимости и увеличением трудоемкости. Наиболее широкое распространение получили методы типа Розенброка, в которых для решения нелинейной системы алгебраических уравнений используется одна итерация метода Ньютона. Методы типа Розенброка относительно просты в реализации и обладают достаточно хорошими свойствами точности и устойчивости. В [5] проведено тестирование на нелинейной задаче методов, которые считаются наиболее надежными и превосходно справляются со сверхжесткими линейными задачами, и сделан вывод, что ни одна из известных схем не является гарантированно надежной для существенно нелинейных сверхжестких задач. Заметим, что жесткой может быть не только задача Коши для системы уравнений, но и для одного скалярного уравнения, что используется для тестирования и разработки методов решения жестких задач [5]. Перспективными

[1)] Работа выполнена при финансовой поддержке РНФ (код проекта 23-21-00496).

методами решения нелинейных сингулярно возмущенных задач являются численно-аналитические методы, в связи с чем стоит отметить активное развитие теории численных методов для задач с переходными слоями на основе асимптотических методов решения сингулярно возмущенных задач [7–10], в том числе развитие самих асимптотических методов решения задач с контрастными структурами [10–14].

В данной работе предлагается численная реализация метода голоморфной регуляризации нелинейных сингулярно возмущенных задач [15–23]. Данный метод позволяет строить приближение к решению в виде ряда по степеням малого параметра, сходящегося не только асимптотически, но и в обычном смысле. Учитывая трудоемкость вычислений, предлагается численное приближение функций, составляющих первые два члена ряда по малому параметру, имеющие погрешность, соответственно, первого и второго порядка по малому параметру. Такой подход к решению позволяет снять проблему нарушения устойчивости при увеличении шага по времени, однако точность метода зависит от малого параметра. Поэтому данный метод ориентирован на уравнения высокой жесткости, а также он применим к задачам с существенными нелинейностями.

1. ЗАДАЧА КОШИ ДЛЯ СКАЛЯРНОГО СИНГУЛЯРНО ВОЗМУЩЕННОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ

1.1. Метод голоморфной регуляризации

Рассмотрим задачу Коши

$$\begin{aligned} &\varepsilon\frac{dy}{dt} = f(t,y), \quad t \in (t_0, T], \\ &y(t_0) = y_0, \end{aligned} \tag{1}$$

где $\varepsilon > 0$ — малый параметр.

Потребуем выполнение условий теоремы Тихонова о предельном переходе [24], адаптированных для уравнений, содержащих только быстрые переменные [25].

I. Пусть функция $f(t,y)$ непрерывна и удовлетворяет условию Липшица по y в некоторой области

$$\bar{\Omega} = \{(t,y) : |y| \le H, t_0 \le t \le T\}.$$

II. Пусть вырожденная задача, полученная из (1) при $\varepsilon = 0$,

$$f(t,\bar{y}) = 0, \quad t \in [t_0, T],$$

имеет непрерывный изолированный корень $\bar{y} = \psi(t)$ на отрезке $[t_0, T]$, т.е. существует такое $\delta > 0$, что в некоторой окрестности $U = \{y : \|y(t) - \psi(t)\| < \delta,\ t_0 \le t \le T\}$ нет других корней: $f(t,y) \ne 0,\ (t,y) \in U$.

III. Уравнение

$$\frac{d\tilde{y}}{ds} = f(t,\tilde{y}),$$

где t выступает в роли параметра, называется присоединенным. Пусть точка покоя $\tilde{y} = \psi(t)$ присоединенного уравнения является асимптотически устойчивой по Ляпунову равномерно по $t \in [t_0, T]$. В таком случае корень вырожденной задачи $\bar{y} = \psi(t)$ называют устойчивым.

IV. Пусть точка y_0 такова, что решение $\tilde{y}(s)$ начальной задачи для присоединенного уравнения

$$\begin{aligned} &\frac{d\tilde{y}}{ds} = f(\tilde{y}, t_0), \\ &\tilde{y}(t_0) = y_0, \end{aligned}$$

существует при всех $s \ge t_0$ и $\tilde{y}(s) \underset{s\to\infty}{\to} \psi(t_0)$. В этом случае говорят, что y_0 принадлежит области влияния точки покоя $\tilde{y} = \psi(t_0)$.

При выполнении условий I–IV справедлива теорема (Тихонова) [24]: найдется постоянная $\varepsilon_0 > 0$ такая, что при $0 < \varepsilon < \varepsilon_0$ решение $y(t,\varepsilon)$ задачи (1) существует на $[t_0, T]$, единственно и справедлив предельный переход

$$\lim_{\varepsilon\to+0} y(t,\varepsilon) = \bar{y}(t) = \psi(t), \quad t \in (t_0, T].$$

Теорема А.Н. Тихонова устанавливает предельный переход, что имеет большое значение в следующем смысле: какими бы методами ни решалась сингулярно возмущенная задача, построенные с их помощью асимптотические приближения должны сходиться при $\varepsilon \to +0$ к предельному решению, указанному в теореме. Могут применяться методы Васильевой–Бутузова–Нефедова, Крылова–Боголюбова–Митропольского и др. Мы

применим метод голоморфной регуляризации, который является логичным продолжением метода регуляризации С.А. Ломова [26], [27]. В рамках метода регуляризации С.А. Ломова было доказано существование сходящихся в обычном смысле рядов по степеням малого параметра, представляющих решения линейных сингулярно возмущенных задач. Трудности переноса теории на нелинейные задачи привели к поиску новых подходов к проблеме аналитической зависимости от малого параметра решений сингулярно возмущенных уравнений, и таким подходом стало введение понятия псевдоголоморфного решения сингулярно возмущенной задачи [28]. Решение $y(t,\varepsilon)$ задачи (1) называется псевдоголоморфным в точке $\varepsilon = 0$, если при представлении $y = Y\left(t, \frac{\varphi(t)}{\varepsilon}, \varepsilon\right)$ функция $Y(t,\eta,\varepsilon)$ голоморфна по третьей переменной в точке $\varepsilon = 0$ равномерно по $t \in [t_0, T]$ при каждом η из некоторого неограниченного множества.

В [15] доказана теорема о псевдоголоморфности решения задачи Коши (1). Пусть функция $f(t,y)$ является голоморфной в некоторой замкнутой области $\bar{\Omega}_{t,y} \in \mathbb{R}^2$ и не обращается в ноль в $\Omega_{t,y}$, а отрезок $[t_0, T]$ и начальная точка (t_0, y_0) принадлежат области $\Omega_{t,y}$. Если голоморфная на $[t_0, T]$ функция $\varphi(t)$, такая что $\varphi(t_0) = 0, \quad \varphi'(t) < 0$ для $\forall t \in [t_0, t_0 + \Delta]$, и уравнение

$$\varphi'(t) \int\limits_{y_0}^{y} \frac{ds}{f(t,s)} = \frac{\varphi(t)}{\varepsilon} \tag{2}$$

имеет решение $y = Y_0\left(t, \frac{\varphi(t)}{\varepsilon}\right)$, равномерно ограниченное при $\varepsilon \to +0$ на отрезке $[t_0, T]$, то решение $y(t,\varepsilon)$ задачи Коши (1) является псевдоголоморфным в точке $\varepsilon = 0$ и определяется общим интегралом [15]:

$$U(t,y,\varepsilon) = 0,$$
$$U(t,y,\varepsilon) = \varphi(t) - \varepsilon \int\limits_{y_0}^{y} \frac{\varphi'(t)ds}{f(t,s)} + \varepsilon^2 \int\limits_{y_0}^{y} \left(\int\limits_{y_0}^{s} \frac{\partial}{\partial t} \left(\frac{\varphi'(t)}{f(t,\xi)} \right) d\xi \right) \frac{ds}{f(t,s)} - \ldots. \tag{3}$$

Замечание 1. Из условия III для теоремы Тихонова (асимптотической устойчивости точки покоя присоединенного уравнения) следует равномерная ограниченность решения $Y_0(t, \varphi/\varepsilon)$ при $\varepsilon \to +0$ на отрезке $[t_0, T]$.

Замечание 2. Если $f(t_0, y_0) = 0$, то введением новой неизвестной функции $v = y + t$ задача (1) может быть сведена к задаче, для которой применим метод голоморфной регуляризации. Если нет цели применения метода голоморфной регуляризации, то в случае согласованности начального условия с правой частью, $f(t_0, y_0) = 0$, приближение порядка $O(\varepsilon)$ может быть получено по теореме Тихонова [24] решением алгебраического уравнения $f(t,y) = 0$, $t \in (t_0, T]$.

Замечание 3. Коэффициенты ряда

$$y(t,\varepsilon) = \sum_{n=0}^{\infty} Y_n\left(t, \frac{\varphi(t)}{\varepsilon}\right) \varepsilon^n \tag{4}$$

могут быть определены по теореме о неявной функции из соотношения [15]:

$$V(t,y,\varepsilon) = \frac{\varphi(t)}{\varepsilon},$$

где

$$V(t,y,\varepsilon) = \int\limits_{y_0}^{y} \frac{\varphi'(t)ds}{f(t,s)} - \varepsilon \int\limits_{y_0}^{y} \left(\int\limits_{y_0}^{s} \frac{\partial}{\partial t} \left(\frac{\varphi'(t)}{f(t,\xi)} \right) d\xi \right) \frac{ds}{f(t,s)} + \ldots,$$

например,

$$Y_1 = -\left. \frac{V'_\varepsilon}{V'_y} \right|_{\varepsilon=0, y=Y_0(t,\varphi(t)/\varepsilon)}, \tag{5}$$

где

$$V'_y\big|_{\varepsilon=0} = \frac{\varphi'(t)}{f(t,y)}, \quad V'_\varepsilon\big|_{\varepsilon=0} = -\int\limits_{y_0}^{y} \left(\int\limits_{y_0}^{s} \frac{\partial}{\partial t} \left(\frac{\varphi'(t)}{f(t,\xi)} \right) d\xi \right) \frac{ds}{f(t,s)}.$$

Замечание 4. Ряд (4) сходится к решению задачи (1) на $[t_0, T]$, и остаточный член имеет вид

$$r_n(t,\varepsilon) = \tilde{Y}_{n+1}\left(t, \frac{\varphi(t)}{\tilde{\varepsilon}(t)}\right) \varepsilon^{n+1}, \tag{6}$$

где $0 < \tilde{\varepsilon}(t) < \varepsilon$ и

$$\lim_{n\to\infty} r_n(t,\varepsilon) = 0$$

равномерно по $t \in [t_0, T]$.

Из (6) и сходимости ряда (4) следует оценка, удобная для вычислений:

$$|r_n(t,\varepsilon)| < |Y_n(t)| \cdot \varepsilon^{n+1}, \quad t \in [t_0, T]\,. \tag{7}$$

1.2. Численная реализация

Введем разбиение отрезка $[t_0, T]$ на n частей с равномерным шагом $h_t = (T - t_0)/n$, и сетку $\{t_i\}_{i=0}^{n}, t_i = t_0 + i \cdot h_t, i = 0, 1, \ldots, n$. Выбираем голоморфную функцию $\varphi(t)$, такую что $\varphi(t_0) = 0, \quad \varphi'(t) < 0 \quad \forall t \in [t_0, T]$. Положим $\varphi(t) = -\operatorname{sh}(t - t_0)$. Далее для каждого t_i, $i = 1, \ldots, n$, решаем нелинейное алгебраическое уравнение (2) и получаем сеточную функцию $\{Y_0^{(i)}\}_{i=0}^{n}$. Для численного интегрирования с контролем погрешности применяются формулы Гаусса-Кронрода [29]: сначала пара формул (G7,K15), если погрешность превышает заданную, то делается перерасчет по паре формул (G15,K31). Для уточнения корней при решении нелинейных алгебраических уравнений используется гибридный гарантированно сходящийся алгоритм Деккера-Брента [30], [31]. По формулам (5) определяется сеточная функция $\{Y_1^{(i)}\}_{i=0}^{n}$ с применением для вычисления повторного интеграла формул Гаусса-Кронрода.

Погрешности значений сеточной функции $\{Y_0^{(i)}\}_{i=0}^{n}$ равны погрешностям решения нелинейного алгебраического уравнения (2) при каждом t_i, $i = 1, \ldots, n$. Однако в данных уравнениях, заданных в виде интеграла с переменным верхним пределом, интервалом неопределённости корня будет

$$\bar{\Delta} Y_0^{(i)} \approx \Delta I_i \cdot |f(t_i, y)|\,, \tag{8}$$

точнее которого не может быть определено значение $Y_0^{(i)}$, $i = 1, \ldots, n$. Из формулы (8) следует, что минимально возможная погрешность определения $Y_0^{(i)}$, $i = 1, \ldots, n$, прямо пропорциональна погрешности ΔI_i вычисления интеграла из (2). Из оценки (8) следует, что погрешность численного приближения не зависит от шага по времени h_t и длины отрезка $[t_0, T]$, на котором ведётся решение задачи Коши.

Введем относительную погрешность для численного решения через евклидову норму сеточных функций:

$$\delta Y = \frac{\|Y - y\|}{\|y\|}, \tag{9}$$

$$\|Y - y\| = \sqrt{\sum_{i=0}^{n} \left(Y^{(i)} - y(t_i)\right)^2}, \quad \|y\| = \sqrt{\sum_{i=0}^{n} \left(y(t_i)\right)^2},$$

где $Y = \left\{Y^{(i)}\right\}_{i=0}^{n}$ — сеточная функция численного решения, $\{y(t_i)\}_{i=0}^{n}$ – точные значения решения задачи Коши на сетке $\{t_i\}_{i=0}^{n}$.

Выведем оценки относительных погрешностей приближений δY_0 и δY_ε для сеточных функций $Y_0 = \left\{Y_0^{(i)}\right\}_{i=0}^{n}$ и $Y_\varepsilon = \left\{Y_0^{(i)} + \varepsilon Y_1^{(i)}\right\}_{i=0}^{n}$. δ_m будем обозначать погрешность, которая следует из заданной в алгоритме точности r определения Y_0, δ_n — погрешность метода голоморфной регуляризации, которая неустранима с точки зрения численной реализации, и согласно (7) имеет порядок $O(\varepsilon)$ для Y_0 и $O\left(\varepsilon^2\right)$ для Y_ε. Из (7) следует, что

$$\delta_n Y_0 \le \frac{\|Y_0\| \cdot \varepsilon}{\|y\|} \approx \varepsilon,$$

и при выборе $r = 0.05\varepsilon$ получим оценку погрешности

$$\delta Y_0 \le \delta_n Y_0 + \delta_m Y_0 = 1.05\varepsilon. \tag{10}$$

При оценке δY_ε заметим, что погрешность определения Y_0 приводит к изменению всего уточненного выражения $Y_\varepsilon = Y_0 + \varepsilon Y_1$ на $\tilde{Y}_\varepsilon = \tilde{Y}_0 + \tilde{\varepsilon}\tilde{Y}_1$: пусть найдено приближение $\tilde{Y}_0^i$, $Y_0^i \in [\tilde{Y}_0^i(1 - r),\ \tilde{Y}_0^i(1 + r)]$, которое удовлетворяет (2) в пределах погрешности, но тогда точное равенство будет справедливо при некотором новом значении малого параметра $\tilde{\varepsilon}$:

$$\int_{y_0}^{\tilde{Y}_0} \frac{ds}{f(t,s)} = \frac{\varphi(t)}{\tilde{\varepsilon}\varphi'(t)},$$

что соответствует применению метода голоморфной регуляризации к задаче

$$\begin{aligned} &\tilde{\varepsilon}\frac{d\tilde{y}}{dt} = f(t,\tilde{y}), \quad t \in (t_0, T], \\ &\tilde{y}(t_0) = y_0, \end{aligned} \tag{11}$$

где

$$\tilde{\varepsilon} = \frac{\varphi(t)}{\varphi'(t)} \left(\int\limits_{y_0}^{\tilde{Y}_0} \frac{ds}{f(t,s)} \right)^{-1}.$$

Очевидно, что, если для задачи (1) выполняется условие

$$\frac{\partial f(t,y)}{\partial y} \leq -\sigma < 0, \quad (t,y) \in \bar{\Omega}_{t,y}, \quad \sigma = \text{const},$$

то можно вывести

$$(y(t) - \tilde{y}(t))' = \frac{f(t,y)}{\varepsilon} - \frac{f(t,\tilde{y})}{\tilde{\varepsilon}},$$

$$(y(t) - \tilde{y}(t))' = \frac{1}{\varepsilon}\frac{\partial f(t,\widehat{y})}{\partial y}(y(t) - \tilde{y}(t)) + f(t,\tilde{y}) \cdot \left(\frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}}\right), \quad \widehat{y} \in (y,\tilde{y}),$$

и с учетом начального условия

$$y(t_0) - \tilde{y}(t_0) = 0,$$

получить оценку

$$|y(t) - \tilde{y}(t)| = \left| \left(\frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}}\right) \cdot \int_{t_0}^{t} e^{\frac{1}{\varepsilon}\frac{\partial f(t,\widehat{y})}{\partial y}(t-s)} f(s,\tilde{y}(s))ds \right| \leq \left| \frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}} \right| \cdot \int_{t_0}^{t} e^{-\frac{\sigma}{\varepsilon}(t-s)} |f(s,\tilde{y}(s))|ds.$$

Поскольку точное решение $\tilde{y}$ неизвестно, для вычисления приближенной оценки будем использовать интерполяцию сплайнами по сеточной функции $\tilde{Y}_\varepsilon$:

$$|y(t) - \tilde{y}(t)| \leq \left| \frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}} \right| \cdot \int\limits_{t_0}^{t} e^{-\frac{\sigma}{\varepsilon}(t-s)} \left| f\left(s, \tilde{Y}_\varepsilon\right) \right| ds, \quad t \in (t_0, T]. \tag{12}$$

Из (7) следует, что

$$\delta_n Y_\varepsilon \leq \frac{\|Y_1\|}{\|y\|}\varepsilon^2,$$

также имеется погрешность численного интегрирования при вычислении $\tilde{Y}_1$, которую оцениваем по методу Гаусса-Кронрода [29]:

$$\Delta\tilde{Y}_1 = \left|\tilde{Y}_1 - \tilde{Y}_1^*\right| \approx \left(200\left|\tilde{Y}_1^{(G)} - \tilde{Y}_1^{(K)}\right|\right)^{1.5}, \tag{13}$$

где $\tilde{Y}_1^{(G)}$ — значение, рассчитанное по квадратурной формуле Гаусса с n узлами, $\tilde{Y}_1^{(K)}$ — значение, рассчитанное по квадратурной формуле Кронрода с $2n+1$ узлами.

Таким образом, получим итоговую оценку

$$\delta Y_\varepsilon = \frac{\left\|y - \tilde{Y}_\varepsilon^*\right\|}{\|y\|} \leq \frac{\|y - \tilde{y}\|}{\|y\|} + \frac{\left\|\tilde{y} - \tilde{Y}_\varepsilon\right\|}{\|y\|} + \frac{\left\|\tilde{Y}_\varepsilon - \tilde{Y}_\varepsilon^*\right\|}{\|y\|} \approx \frac{\|y - \tilde{y}\|}{\|\tilde{Y}_\varepsilon^*\|} + \frac{\left\|\tilde{Y}_1\right\|}{\|\tilde{Y}_\varepsilon^*\|}\varepsilon^2 + \frac{\|\Delta\tilde{Y}_1\|}{\|\tilde{Y}_\varepsilon^*\|}, \tag{14}$$

где должны использоваться оценки (12) и (13).

Компьютерная программа для разработанного численного метода написана на языке программирования C++.

Основной проблемой численного решения по методу голоморфной регуляризации является значительное возрастание вычислительной трудоемкости, если возрастает $|y(t) - y_0|$ при возрастании t, особенно, если велика длина отрезка $[t_0, T]$. Предлагается рассматривать задачи, про которые известна ограниченность решения: $|y(t)| \leqslant C\ \forall t \geqslant t_0$, $C = \text{const}$. Также программа выдает предупреждение о возможности прекратить вычисления при существенном возрастании значения $|y(t) - y_0|$, влияющего на трудоемкость и время вычислений.

К важнейшим преимуществам численной реализации метода голоморфной регуляризации следует отнести отсутствие ограничения на шаг по времени; отсутствие зависимости точности полученных приближений от шага по времени; возможность вычисления на больших временных отрезках, если нет сильного возрастания решения, в том числе возможность прохождения с большим шагом по времени длительного промежутка с последующими вычислениями без потери точности; наличие оценки погрешности приближенного решения; возможность эффективного применения параллельного программирования для ускорения программы (распараллеливание алгоритма вычисления интегралов).

1.3. Вычислительные эксперименты

Рассмотрим сначала задачи с известными точными решениями: кубический тест на отрезке $[0,1]$ [5] и линейную сингулярно возмущенную задачу на большом временном промежутке $[0,T]$, $T \gg 1$. Вычисления проводились на процессоре Intel Core i5 10210U (базовая частота 1.6 ГГц, максимальная частота 4.2 ГГц, уровни кэша: 256 КБ, 1 МБ, 6МБ). Проводилось измерение времени вычислений встроенной функцией времени C++, имеющей точность 15 миллисекунд.

Пример 1. Рассмотрим кубический тест, который хорошо иллюстрирует характерные особенности задач химической кинетики [5]:

$$\begin{aligned} &\varepsilon\frac{dy}{dt} = -y\left(y^2 - a^2\right), \quad t \in (0,1], \\ &y(0) = y_0, \end{aligned} \tag{15}$$

где $\varepsilon > 0$ — малый параметр, параметр $a > 0$, рекомендуется брать $a \approx 1$, функция $y(t)$ имеет смысл концентрации, поэтому осмысленным будет начальное значение $0 \leqslant y_0 \leqslant 1$. Данная задача является нелинейной с уравнением с разделяющимися переменными, поэтому известно точное решение:

$$y(t) = \frac{ay_0}{\sqrt{y_0^2 + (a^2 - y_0^2)\cdot e^{-2a^2 t/\varepsilon}}}.$$

Положим $y_0 = 0.5$, $a = 1$, шаг сетки по времени $h_t = 0.1$, $n = 10$. В табл. 1 приведены зависимости от ε погрешностей δY_0 и δY_ε, рассчитанных по формуле (9). Значения малого параметра ε уменьшаем на порядок с 10^{-1} до 10^{-13}, меньшее значение не может рассматриваться из-за машинной точности $\sim 10^{-16}$ при стандартных вычислениях с типом вещественных данных “double”.

Таблица 1

ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
δY_0	$1.03\cdot 10^{-3}$	$9.7\cdot 10^{-10}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$
δY_ε	$6.7\cdot 10^{-4}$	$9.6\cdot 10^{-10}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$

ε	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}	10^{-13}
δY_0	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$
δY_ε	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$	$1.1\cdot 10^{-16}$

Из полученных значений погрешности следует, что кубический тест успешно пройден. Кроме того, полученные погрешности для данного примера гораздо меньше оценочных и с уменьшением малого параметра быстро достигли машинной точности вычислений. Простота прохождения кубического теста объясняется спецификой численной реализации метода голоморфной регуляризации, которая кардинально отличает его от пошаговых реализаций традиционных методов численного решения задачи Коши. Аналогично не представляет трудности и тест Далквиста для линейной задачи.

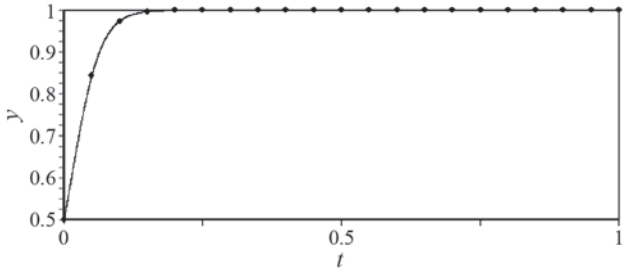
Для лучшего визуального восприятия профиль точного $y(t)$ и приближенного Y_0 решений построены на фиг. 1 при $h_t = 0.05$, $n = 20$, $\varepsilon = 0.05$. В данном случае получены погрешности $\delta Y_0 = 3.4\cdot 10^{-4}$, $\delta Y_\varepsilon = 2.2\cdot 10^{-4}$.

Пример 2. В качестве линейной задачи на большом временном промежутке рассмотрим задачу Коши

$$\begin{aligned} &\varepsilon\frac{dy}{dt} = -y + \sin t, \quad t \in (0,T], \\ &y(0) = 1, \end{aligned} \tag{16}$$

где $\varepsilon > 0$ — малый параметр. Точное решение задачи Коши с линейным дифференциальным уравнением

$$y(t) = \left(1 + \frac{\varepsilon}{1+\varepsilon^2}\right)e^{-t/\varepsilon} + \frac{\varepsilon}{1+\varepsilon^2}\left(\frac{\sin t}{\varepsilon} - \cos t\right) \tag{17}$$



Фиг. 1. График точного решения задачи (15) $y(t)$ — сплошная линия, график численного решения Y_0 — набор точек, $n = 20$.

будем использовать для оценки погрешности численного решения. Также из (17) видно, что при $\varepsilon \to 0$ вблизи $t = 0$ имеется пограничный слой и при малых t можно приближенно записать $y(t) \approx e^{-t/\varepsilon}$, а при увеличении t первое слагаемое в (17) становится малым и $y(t) \approx \sin t$, что делает (16) классическим примером жесткой задачи.

Положим $y_0 = 1$, $T = 100$, шаг сетки по времени $h_t = 5$, $n = 20$. В табл. 2 приведены зависимости от малого параметра ε погрешностей δY_0 и δY_ε, рассчитанных по формуле (9), их оценок $\delta\bar{Y}_0$ и $\delta\bar{Y}_\varepsilon$, рассчитанных по формулам (10), (14), и соответствующего времени вычислений time_0 и time_ε (в миллисекундах).

Таблица 2

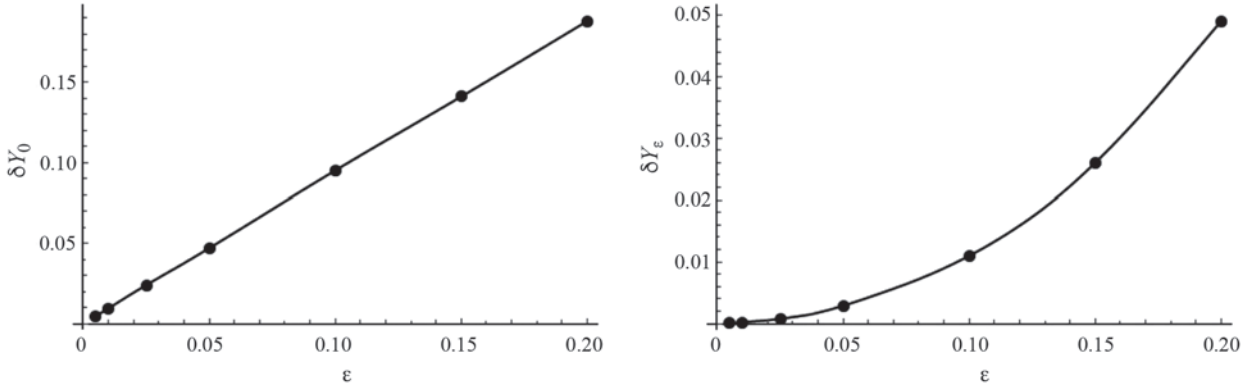
ε	0.2	0.15	0.1	0.05	0.025	10^{-2}
δY_0	0.188	0.141	0.095	0.047	0.024	$9.5 \cdot 10^{-3}$
$\delta\bar{Y}_0$	0.21	0.16	0.11	0.053	0.027	$1.1 \cdot 10^{-2}$
time_0	< 15	< 15	< 15	< 15	< 15	< 15
δY_ε	0.049	0.026	0.011	$2.9 \cdot 10^{-3}$	$7.5 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$
$\delta\bar{Y}_\varepsilon$	0.054	0.031	0.014	$3.9 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$3.9 \cdot 10^{-4}$
time_ε	15	15	31	31	47	94

ε	$5 \cdot 10^{-3}$	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
δY_0	$4.7 \cdot 10^{-3}$	$9.5 \cdot 10^{-4}$	$9.5 \cdot 10^{-5}$	$9.5 \cdot 10^{-6}$	$9.5 \cdot 10^{-7}$	$9.5 \cdot 10^{-8}$
$\delta\bar{Y}_0$	$5.3 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-5}$	$1.1 \cdot 10^{-6}$	$1.1 \cdot 10^{-7}$
time_0	< 15	< 15	< 15	< 15	< 15	< 15
δY_ε	$1 \cdot 10^{-4}$	$2.5 \cdot 10^{-5}$	$2.3 \cdot 10^{-6}$	$2.5 \cdot 10^{-7}$	$3.2 \cdot 10^{-8}$	$4.2 \cdot 10^{-9}$
$\delta\bar{Y}_\varepsilon$	$1.7 \cdot 10^{-4}$	$3.3 \cdot 10^{-5}$	$2.9 \cdot 10^{-6}$	$3.1 \cdot 10^{-7}$	$3.9 \cdot 10^{-8}$	$4.6 \cdot 10^{-9}$
time_ε	141	359	1703	8781	42828	227046

ε	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}	10^{-13}
δY_0	$9.5 \cdot 10^{-9}$	$9.5 \cdot 10^{-10}$	$9.5 \cdot 10^{-11}$	$9.5 \cdot 10^{-12}$	$9.5 \cdot 10^{-13}$	$9.5 \cdot 10^{-14}$
$\delta\bar{Y}_0$	$1.1 \cdot 10^{-8}$	$1.1 \cdot 10^{-9}$	$1.1 \cdot 10^{-10}$	$1.1 \cdot 10^{-11}$	$1.1 \cdot 10^{-12}$	$1.1 \cdot 10^{-13}$
time_0	< 15	< 15	< 15	< 15	< 15	< 15

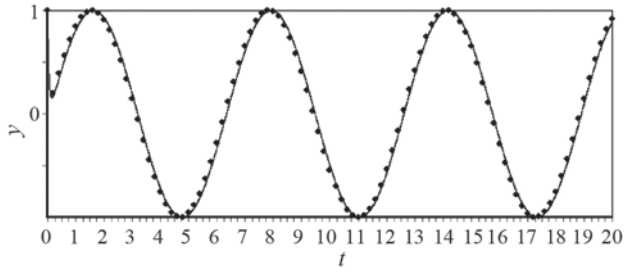
Для рассмотренного примера выявлена прямо пропорциональная зависимость δY_0 от малого параметра ε (фиг. 2), что полностью соответствует теоретической оценке (10). Также получена зависимость δY_ε от ε, но она имеет более сложный характер: квадратичная зависимость имеет место при относительно больших значениях малого параметра, $10^{-2} < \varepsilon < 0.2$ (фиг. 2), а при меньших значениях ε наблюдается почти линейная зависимость, что в целом соответствует оценке (14), по которой рассчитаны значения $\delta\bar{Y}_\varepsilon$. Также из значений времени вычислений time_ε следует, что при уменьшении ε значительно вырастает трудоемкость расчета Y_ε и делает его

нецелесообразным при $\varepsilon < 10^{-5}$. Из полученных данных следует, что наилучшие соотношения точности и трудоемкости ориентируют предложенный метод на вычисление приближений Y_0 для задач средней жесткости и сверхжестких задач при экстремально малых значениях малого параметра, $10^{-13} < \varepsilon < 10^{-4}$. Расчет уточненного приближения Y_ε имеет смысл только для умеренно жестких задач, $10^{-4} < \varepsilon < 10^{-1}$.



Фиг. 2. Графики зависимости погрешностей δY_0 и δY_ε от малого параметра ε.

Для лучшего визуального восприятия профиль точного $y(t)$ и приближенного Y_0 решений построены на фиг. 3 при $T = 20$, $h_t = 0.2$, $n = 100$, $\varepsilon = 0.05$. В данном случае получены погрешности $\delta Y_0 = 4.8 \cdot 10^{-2}$, $\delta Y_\varepsilon = 3.1 \cdot 10^{-3}$. Построенные графики точного и приближенного решений визуально почти неотличимы (фиг. 3).



Фиг. 3. График точного решения (17) $y(t)$ – сплошная линия, график численного решения Y_0 – набор точек, $n = 100$.

Пример 3. Рассмотрим нелинейную задачу Коши на большом временном промежутке

$$\begin{aligned} &\varepsilon \frac{dy}{dt} = -y(y^2+1) + 4\cos^2 t, \quad t \in (0, T], \\ &y(0) = 0, \end{aligned} \tag{18}$$

где $\varepsilon > 0$ — малый параметр. Положим $T = 500$, $n = 100$, шаг сетки по времени $h_t = 5$. В табл. 3 приведены зависимости от малого параметра ε оценок $\delta\bar{Y}_0$ и $\delta\bar{Y}_\varepsilon$, рассчитанных по формулам (10), (14), и соответствующего времени вычислений time_0 и time_ε (в миллисекундах).

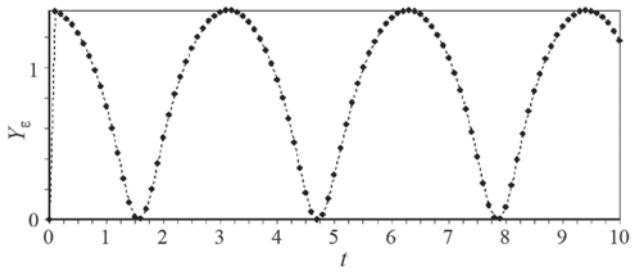
Для лучшего визуального восприятия профиль точного $y(t)$ и приближенного Y_ε решений построены на фиг. 4 при $T = 10$, $h_t = 0.1$, $n = 100$, $\varepsilon = 0.01$. В данном случае получены оценки погрешностей $\delta\bar{Y}_0 = 1.05 \cdot 10^{-2}$, $\delta\bar{Y}_\varepsilon = 1.03 \cdot 10^{-3}$. Построен график приближенного решения Y_ε (фиг. 4).

Из приведенных примеров видно, что предложенный в статье метод позволяет строить численные решения задач Коши с нелинейными сингулярно возмущенными дифференциальными уравнениями и дает гарантированные оценки погрешностей построенных приближенных решений.

Таблица 3

ε	0.1	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$\delta\bar{Y}_0$	0.11	$1.1 \cdot 10^{-2}$	$1.1 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-5}$	$1.1 \cdot 10^{-6}$
time_0	< 15	< 15	15	15	23	23
$\delta\bar{Y}_\varepsilon$	0.017	$1 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$	$1 \cdot 10^{-5}$	$9.2 \cdot 10^{-7}$	$9.8 \cdot 10^{-8}$
time_ε	110	437	2032	10110	50578	267500

ε	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}
$\delta\bar{Y}_0$	$1.1 \cdot 10^{-7}$	$1.1 \cdot 10^{-8}$	$1.1 \cdot 10^{-9}$	$1.1 \cdot 10^{-10}$	$1.1 \cdot 10^{-11}$	$1.1 \cdot 10^{-12}$
time_0	23	31	31	31	39	39



Фиг. 4. График численного решения Y_ε – набор точек, $n = 100$, соединенных штриховой линией.

2. ЗАДАЧА КОШИ ДЛЯ СИНГУЛЯРНО ВОЗМУЩЕННОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ ВТОРОГО ПОРЯДКА

2.1. Метод голоморфной регуляризации

Рассмотрим задачу Коши для дифференциального уравнения второго порядка

$$\begin{aligned} &\varepsilon y'' = f\left(t, y, y'\right), \quad t \in \left(t_0, T\right], \\ &y\left(t_0\right) = y_0, \quad y'\left(t_0\right) = v_0, \end{aligned} \tag{19}$$

где $\varepsilon > 0$ — малый параметр, и перепишем ее для системы дифференциальных уравнений

$$\begin{aligned} &\frac{dy}{dt} = v, \\ &\varepsilon\frac{dv}{dt} = f(t, y, v), \\ &y\left(t_0\right) = y_0, \quad v\left(t_0\right) = v_0. \end{aligned} \tag{20}$$

Пусть для задачи (20) выполнены условия теоремы Тихонова о предельном переходе [17], [24]. Пусть функция $f(t, y, v)$ является голоморфной в некоторой замкнутой области $\bar{\Omega}_{t,y,v} \in \mathbb{R}^3$ и не обращается в ноль в $\Omega_{t,y,v}$, а отрезок $[t_0, T]$ и начальная точка (t_0, y_0, v_0) принадлежат области $\Omega_{t,y,v}$. Пусть $\bar{V}(t, y)$ — изолированный корень уравнения $f(t, y, v) = 0$, и соответствующая (20) вырожденная задача

$$\begin{aligned} &\frac{d\bar{y}}{dt} = \bar{V}(t, \bar{y}), \quad t \in \left(t_0, T\right], \\ &\bar{y}\left(t_0\right) = y_0, \end{aligned} \tag{21}$$

имеет единственное голоморфное на $[t_0, T]$ решение $\bar{y}(t)$.

При наложенных условиях, из доказанных в [16] теорем, следует голоморфность первого интеграла

$$\varphi(t, y) - \varepsilon\int\limits_{v_0}^{v}\frac{L\varphi(t, y)ds}{f(t, y, s)} + \varepsilon^2\int\limits_{v_0}^{v}\left(L\int\limits_{v_0}^{s}\left(\frac{L\varphi(t, y)}{f(t, y, \xi)}\right)d\xi\right)\frac{ds}{f(t, y, s)} - \ldots = 0, \tag{22}$$

где введен оператор $L = \frac{\partial}{\partial t} + v\frac{\partial}{\partial y}$.

Построим два независимых первых интеграла, применяя в (22) регуляризирующие функции $\varphi_1(t,y) = \varphi(t)$ и $\varphi_2(t,y) = y - \bar{y}(t)$:

$$\begin{aligned} &\int\limits_{v_0}^{v} \frac{\varphi'(t)ds}{f(t,y,s)} - \varepsilon \int\limits_{v_0}^{v} \left(L \int\limits_{v_0}^{s} \left(\frac{\varphi'(t)}{f(t,y,\xi)} \right) d\xi \right) \frac{ds}{f(t,y,s)} + \ldots = \frac{\varphi(t)}{\varepsilon}, \\ &y - \bar{y}(t) - \varepsilon \int\limits_{v_0}^{v} \frac{(s - \bar{y}'(t))\, ds}{f(t,y,s)} + \varepsilon^2 \int\limits_{v_0}^{v} \left(L \int\limits_{v_0}^{s} \left(\frac{\xi - \bar{y}'(t)}{f(t,y,\xi)} \right) d\xi \right) \frac{ds}{f(t,y,s)} - \ldots = 0. \end{aligned} \tag{23}$$

Пусть функция $\varphi(t)$ голоморфна на $[t_0, T]$ и $\varphi\,(t_0) = 0$, $\varphi'(t) < 0$ для $\forall t \in [t_0, T]$. Если уравнение

$$\varphi'(t) \int\limits_{v_0}^{v} \frac{ds}{f(t,\bar{y},s)} = \frac{\varphi(t)}{\varepsilon} \tag{24}$$

имеет решение $v = V_0\left(t, \frac{\varphi(t)}{\varepsilon}\right)$, равномерно ограниченное при $\varepsilon \to +0$ на отрезке $[t_0, T]$, то решение $y(t,\varepsilon)$ задачи (20) является псевдоголоморфным в точке $\varepsilon = 0$ [16].

Из (23) определяются коэффициенты разложений

$$y(t,\varepsilon) = \sum_{k=0}^{\infty} Y_k\left(t, \frac{\varphi(t)}{\varepsilon}\right) \cdot \varepsilon^k, \quad v(t,\varepsilon) = \sum_{k=0}^{\infty} V_k\left(t, \frac{\varphi(t)}{\varepsilon}\right) \cdot \varepsilon^k.$$

Заметим, что $Y_0(t) = \bar{y}(t)$ и

$$Y_1\left(t, \frac{\varphi(t)}{\varepsilon}\right) = \left.\frac{\partial y}{\partial \varepsilon}\right|_{\varepsilon=0,\ y=\bar{y}(t),\ v=V_0(t,\varphi(t)/\varepsilon)} = \int\limits_{v_0}^{V_0} \frac{(s - \bar{y}'(t))\,\mathrm{d}s}{f(t,\bar{y},s)}. \tag{25}$$

2.2. Численная реализация и вычислительные эксперименты

Первым этапом является нахождение решения вырожденной задачи (21). Предварительно должен быть найден и указан изолированный корень $\bar{V}(t,\bar{y}(t))$ уравнения $f(t,y,v) = 0$. Тогда вырожденная задача (21) может быть решена любым удобным численным методом, например, Рунге-Кутты четвертого порядка точности, так как не является жесткой задачей. Шаг численного интегрирования вырожденной задачи (21) должен быть согласован (укладываться целое число раз) с шагом по времени, который будет использоваться для получения приближений по методу голоморфной регуляризации. Либо, если возможно, может быть указано непосредственно решение вырожденной задачи (21) в виде функции $\bar{y}(t)$.

Второй этап представляет численную реализацию метода голоморфной регуляризации, во многом схожую с реализацией метода для уравнения первого порядка. Из (24) определяется $\{V_0^{(i)}\}_{i=0}^n$, затем по формуле (25) определяется $\{Y_1^{(i)}\}_{i=0}^n$, для численного интегрирования используются формулы Гаусса-Кронрода [29].

Выведем оценку относительной погрешности δY_ε для сеточной функции $Y_\varepsilon = \{\bar{y}(t_i) + \varepsilon Y_1^{(i)}\}_{i=0}^n$. Решая (24), мы определяем некоторое приближение $\tilde{V}_0^{(i)}$, $V_0^i \in [\tilde{V}_0^{(i)}(1-r),\ \tilde{V}_0^{(i)}(1+r)]$, на которое, в том числе, влияет и погрешность нахождения решения вырожденной задачи $\bar{y}(t)$. Тогда, полученное приближение $\tilde{V}_0^{(i)}$ соответствует решению задачи (20) при некотором новом значении малого параметра $\tilde{\varepsilon}$:

$$\begin{aligned} &\frac{d\tilde{y}}{dt} = \tilde{v}, \\ &\tilde{\varepsilon}\frac{d\tilde{v}}{dt} = f(t,\tilde{y},\tilde{v}), \\ &\tilde{y}\,(t_0) = y_0, \quad \tilde{v}\,(t_0) = v_0, \end{aligned} \tag{26}$$

где

$$\tilde{\varepsilon} = \frac{\varphi(t)}{\varphi'(t)} \left(\int\limits_{y_0}^{\tilde{V}_0} \frac{ds}{f(t,\bar{Y}(t),s)} \right)^{-1}.$$

Пусть

$$\mathbf{z}(t) = \begin{pmatrix} y(t) - \tilde{y}(t) \\ v(t) - \tilde{v}(t) \end{pmatrix}.$$

Тогда

$$\begin{aligned} &\frac{d\mathbf{z}}{dt} = A(t)\mathbf{z} + \mathbf{g}(t), \quad t \in (t_0, T], \\ &\mathbf{z}(t_0) = \mathbf{0}, \end{aligned} \tag{27}$$

где

$$A(t) = \begin{pmatrix} 0 & 1 \\ a(t) & b(t) \end{pmatrix}, \quad \mathbf{g}(t) = \begin{pmatrix} 0 \\ g(t) \end{pmatrix}, \quad g(t) = f(t, \tilde{y}(t), \tilde{v}(t)) \cdot \left(\frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}}\right),$$
$$a(t) = \frac{1}{\varepsilon}\frac{\partial f(t, \widehat{y}(t), \tilde{v}(t))}{\partial y}, \quad b(t) = \frac{1}{\varepsilon}\frac{\partial f(t, y(t), \widehat{v}(t))}{\partial v}, \quad \widehat{y} \in (y, \tilde{y}), \quad \widehat{v} \in (y, \tilde{y}),$$

Выведем оценку $|y(t) - \tilde{y}(t)|$ при следующих ограничениях:

$$-\beta \le \frac{\partial f(t, y, v)}{\partial v} \le -\sigma < 0, \quad (t, y, v) \in \bar{\Omega}_{t,y,v}, \quad \beta, \sigma = \text{ const},$$
$$\frac{\partial f(t, y, v)}{\partial y} \le -\frac{\beta^2}{2} < 0, \quad (t, y, v) \in \bar{\Omega}_{t,y,v},$$

когда $4a + b^2 \le -\beta^2 < 0$ и собственные значения матрицы $A(t)$ комплексные

$$\lambda_{1,2}(t) = \frac{1}{2}\left(b(t) \pm \sqrt{4a(t) + b^2(t)}\right), \quad \operatorname{Re}\lambda_{1,2}(t) = \frac{b(t)}{2} \le -\frac{\sigma}{2\varepsilon} < 0, \quad t \in (t_0, T].$$

По формуле Коши с учетом начального условия, имеем решение задачи (27):

$$\mathbf{z}(t) = \Phi(t)\int_{t_0}^{t} \Phi^{-1}(s)\mathbf{g}(s)ds, \tag{28}$$

где $\Phi(t)$ — фундаментальная матрица решений однородной системы дифференциальных уравнений из (27),

$$\Phi(t) = \begin{pmatrix} \dfrac{1}{\lambda_1(t)}e^{\lambda_1(t)t} & \dfrac{1}{\lambda_2(t)}e^{\lambda_2(t)t} \\ e^{\lambda_1(t)t} & e^{\lambda_2(t)t} \end{pmatrix}, \qquad \Phi^{-1}(s)\mathbf{g}(s) = \begin{pmatrix} \dfrac{\lambda_1(s)e^{\lambda_1(s)s}}{\lambda_1(s) - \lambda_2(s)}g(s) \\ \dfrac{\lambda_2(s)e^{\lambda_2(s)s}}{\lambda_2(s) - \lambda_1(s)}g(s) \end{pmatrix},$$

$$\left|\int_{t_0}^{t} \frac{\lambda_1(s)e^{\lambda_1(s)s}}{\lambda_1(s) - \lambda_2(s)}g(s)ds\right| \le \int_{t_0}^{t} \left|\frac{\left(b(s) - \sqrt{4a(s) + b^2(s)}\right)e^{b(s)s/2}}{2\sqrt{4a(s) + b^2(s)}}\right| \cdot |g(s)|ds \le$$
$$\le \frac{1}{2}\int_{t_0}^{t} \sqrt{1 + \frac{b^2(s)}{|4a(s) + b^2(s)|}} \cdot e^{-\sigma\cdot s/2\varepsilon}|g(s)|ds \le \frac{\sqrt{2}}{2}\int_{t_0}^{t} e^{-\sigma\cdot s/2\varepsilon}|g(s)|ds.$$

Аналогично

$$\left|\int_{t_0}^{t} \frac{\lambda_2(s)e^{\lambda_2(s)s}}{\lambda_2(s) - \lambda_1(s)}g(s)ds\right| \le \frac{\sqrt{2}}{2}\int_{t_0}^{t} e^{-\sigma\cdot s/2\varepsilon}|g(s)|ds.$$

Таким образом, из (28) следует, что

$$\begin{aligned} |y(t) - \tilde{y}(t)| &\le \frac{\sqrt{2}}{2} \cdot \left|\frac{1}{\lambda_1(t)}e^{\lambda_1(t)t} + \frac{1}{\lambda_2(t)}e^{\lambda_2(t)t}\right| \cdot \int_{t_0}^{t} e^{-\sigma\cdot s/2}|g(s)|ds \le \\ &\le \frac{\sqrt{2}}{2}\left|2\frac{2}{2\sigma/\varepsilon}\right|\int_{t_0}^{t} e^{-\sigma\cdot(t-s)/2\varepsilon}|g(s)|ds = \frac{\sqrt{2}\varepsilon}{\sigma}\left|\frac{1}{\varepsilon} - \frac{1}{\tilde{\varepsilon}}\right|\int_{t_0}^{t} e^{-\sigma\cdot(t-s)/2\varepsilon}|f(s, \tilde{y}(s), \tilde{v}(s))|ds. \end{aligned} \tag{29}$$

Оценка (29) используется для вычисления итоговой оценки

$$\delta Y_\varepsilon \le \frac{\|y-\tilde{y}\|}{\|y\|} + \frac{\left\|\tilde{Y}_1\right\|}{\|y\|}\varepsilon^2 + \frac{\|\Delta\tilde{Y}_1\|}{\|y\|}, \tag{30}$$

где также должна использоваться оценка (13) для вычислительной погрешности интегрирования при расчете $\tilde{Y}_1$ по формуле (25).

Пример 4. Рассмотрим задачу Коши для сингулярно возмущенного уравнения с кубической нелинейностью

$$\begin{aligned} &\varepsilon y'' + y' + y + y^3 = \frac{\cos(t)}{4}, \quad t \in (0,T]\,, \\ &y\,(0) = 1, \quad y'\,(0) = 0, \end{aligned} \tag{31}$$

где $\varepsilon > 0$ — малый параметр, который соответствует, например, малой индуктивности при описании электрических колебаний. Перепишем задачу в виде

$$\begin{aligned} &\frac{dy}{dt} = v, \\ &\varepsilon\frac{dv}{dt} = -y' - y - y^3 + \frac{\cos(t)}{4}, \\ &y(0) = 1, \quad v(0) = 0. \end{aligned} \tag{32}$$

Вырожденная задача примет вид

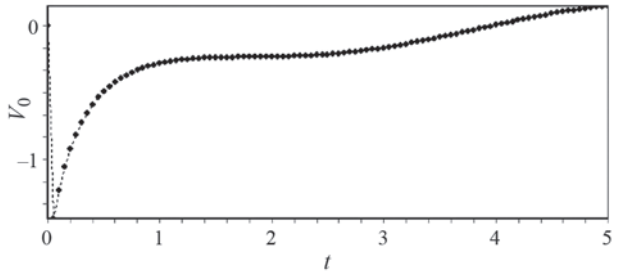
$$\begin{aligned} &\frac{d\bar{y}}{dt} = -\bar{y} - \bar{y}^3 + \frac{\cos(t)}{4}, \quad t \in (0,T], \\ &\bar{y}(0) = 1. \end{aligned}$$

Положим $T = 10$, $n = 50$, шаг сетки по времени $h_t = 0.2$. В табл. 4 приведены зависимости от малого параметра ε оценки $\delta\bar{Y}_\varepsilon$, рассчитанной по формуле (30) и соответствующего времени вычислений time_ε (в миллисекундах).

Таблица 4

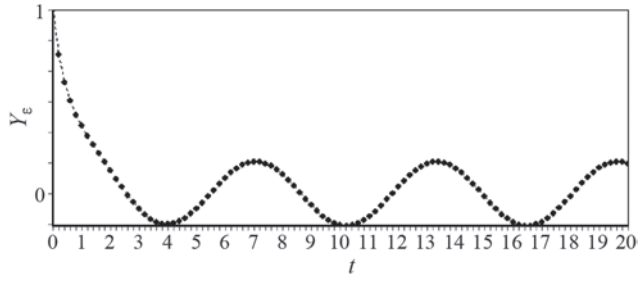
ε	0.1	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
$\delta\bar{Y}_\varepsilon$	0.011	$3.1\cdot10^{-3}$	$7.1\cdot10^{-5}$	$8.4\cdot10^{-7}$	$1.6\cdot10^{-9}$	$1.3\cdot10^{-11}$	$1.1\cdot10^{-13}$
time_ε	31	39	43	51	55	62	66

На фиг. 5 представлен график сеточной функции V_0, которая является приближением для производной от функции $y(t)$. Из графика (фиг. 5) следует наличие пограничного слоя, что соответствует второму уравнению в (32) с сингулярно входящим малым параметром ε. На фиг. 6 представлен график сеточной функции Y_ε. Для лучшего визуального восприятия наличия пограничного слоя, на фиг. 5 и фиг. 6 взяты разные временные промежутки T.



Фиг. 5. График сеточной функции V_0 задачи (31) при $\varepsilon = 0.01$, $T = 5$, – набор точек, $n = 100$, соединенных штриховой линией.

Заметим, что для задачи Коши с уравнением второго порядка предложена численная реализация метода голоморфной регуляризации с определением приближения только в виде Y_ε, которое имеет второй порядок точности по малому параметру ε. При этом трудоемкость вычислений меньше, чем для построения приближения



Фиг. 6. График численного решения Y_ε задачи (31) при $\varepsilon = 0.01$, $T = 20$, – набор точек, $n = 100$, соединенных штриховой линией.

второго порядка описанной выше задачи Коши для скалярного уравнения, так как здесь рассчитывается интеграл только по одной переменной. Зависимости погрешности от малого параметра и сравнения трудоемкостей подтверждаются вычислительными экспериментами.

СПИСОК ЛИТЕРАТУРЫ

1. *Хайрер Э., Ваннер Г.* Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. Пер. с англ. М.: Мир, 1999.

2. *Ракитский Ю.В., Устинов С.М., Черноруцкий И.Г.* Численные методы решения жестких систем. М.: Наука, 1979.

3. *Lambert J.D.* Numerical methods for ordinary differential systems: the initial value problem. New York: Wiley-Sons, 1991.

4. *Новиков Е.А., Шорников Ю.В.* Компьютерное моделирование жестких гибридных систем. Новосибирск: Изд-во НГТУ, 2012.

5. *Белов А.А., Калиткин Н.Н.* Проблема нелинейности при численном решении сверхжестких задач Коши // Матем. моделирование. 2016. Т. 28. № 4. С. 16–32.

6. *Калиткин Н.Н.* Численные методы решения жестких систем // Матем. моделирование. 1995. Т. 7. № 5. С. 8–11.

7. *Нефедов Н.Н.* Развитие методов асимптотического анализа переходных слоев в уравнениях реакции–диффузии–адвекции: теория и применение // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 12. С. 2074–2094.

8. *Kopteva N., Stynes M.* Stabilised approximation of interior-layer solutions of a singularly perturbed semilinear reaction diffusion problem // Numerische Mathematik. 2011. V. 119. № 2. P. 787–810.

9. *Quinn J.* A numerical method for a nonlinear singularly perturbed interior layer problem using an approximate layer location // Comput. and Appl. Math. 2015. V. 290. № 15. P. 500–515.

10. *Нефедов Н.Н., Никулин Е.И., Орлов А.О.* О периодическом внутреннем слое в задаче реакция-диффузия с источником модульно-кубичного типа // Ж. вычисл. матем. и матем. физ. 2020. Т. 60. № 9. С. 1513–1532.

11. *Нефедов Н.Н., Орлов А.О.* О неустойчивых контрастных структурах в одномерных задачах реакция-диффузия-адвекция с разрывными источниками // Теор. и матем. физ. 2023. Т. 215. № 2. С. 297–310.

12. *Нефедов Н.Н.* Периодические контрастные структуры в задаче реакция-диффузия с быстрой реакцией и малой диффузией // Матем. заметки. 2022. Т. 112. № 4. С. 601–612.

13. *Волков В.Т., Нефедов Н.Н.* Асимптотическое решение задачи граничного управления для уравнения типа Бюргерса с модульной адвекцией и линейным усилением // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 11. С. 1851–1860.

14. *Нефедов Н.Н., Руденко О.В.* О движении, усилении и разрушении фронтов в уравнениях типа Бюргерса с квадратичной и модульной нелинейностью // Докл. АН. Матем., информ., проц. упр. 2020. Т. 493. С. 26–31.

15. *Качалов В.И.* Голоморфная регуляризация сингулярно возмущенных задач // Вестник МЭИ. 2010. № 6. С. 54–62.

16. *Качалов В.И.* Голоморфная регуляризация сингулярно возмущенного уравнения второго порядка // Вестник МЭИ. 2013. № 6. С. 95–103.

17. *Качалов В.И.* Теорема Тихонова о предельном переходе и псевдоголоморфные решения сингулярно возмущенных задач // Докл. АН. 2014. Т. 458. № 6. С. 630–632.

18. *Качалов В.И.* О методе голоморфной регуляризации сингулярно возмущенных задач // Изв. вузов. матем. 2017. № 6. С. 52–59.

19. *Качалов В.И.* О голоморфной регуляризации сингулярно возмущенных систем дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 4. С. 654–661.

20. *Качалов В.И.* О голоморфной регуляризации сильно нелинейных сингулярно возмущенных задач // Уфимск. матем. ж. 2018. Т. 10. № 3. С. 35–43.

21. *Качалов В.И.* Голоморфная регуляризация сингулярных возмущений в банаховом пространстве // Дифференц. ура-ния. 2018. Т. 54. № 6. С. 794–802.

22. *Bobodzhanov A.A., Safonov V.F., Kachalov V.I.* Asymptotic and Pseudoholomorphic Solutions of Singularly Perturbed Differential and Integral Equations in the Lomov's Regularization Method // Axioms. 2019. 8(1), 27.

23. *Besova M.I., Kachalov V.I.* Axiomatic Approach in the Analytic Theory of Singular Perturbations // Axioms. 2020. 9(1), 9.

24. *Васильева А.Б., Бутузов В.Ф.* Асимптотические разложения решений сингулярно возмущенных задач. М.: Наука, 1973.

25. *Сафонов В.Ф., Бободжанов А.А.* Сингулярно возмущенные задачи и метод регуляризации. М.: Изд-во МЭИ, 2010.

26. *Ломов С.А.* Введение в общую теорию сингулярных возмущений. М.: Наука, 1981.

27. *Ломов С.А., Ломов И.С.* Основы математической теории пограничного слоя. М.: Изд-во МГУ, 2011.

28. *Качалов В.И., Ломов С.А.* Псевдоаналитические решения сингулярно возмущенных задач. Докл. АН. 1994. Т. 334. № 6. С. 694–695.

29. *Кронрод А.С.* Узлы и веса квадратурных формул: шестнадцатизначные таблицы. М.: Наука, 1964.

30. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. Пер. с англ. М.: Мир, 1980.

31. *Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P.* Numerical Recipes in C: The Art of Scientific Computing. Second Edition. 2002.

ON A NUMERICAL METHOD FOR SOLVING THE CAUCHY PROBLEM FOR SINGULARLY PERTURBED DIFFERENTIAL EQUATIONS

D. A. Maslov*

National Research University "Moscow Power Engineering Institute" (NRU MPEI), Krasnokazarmennaya st., 14, Moscow, 111250, Russia
**e-mail: maslovdma@mpei.ru*
Received 10 November, 2023
Revised 10 November, 2023
Accepted 14 January, 2024

Abstract. The paper proposes a new method for numerically solving nonlinear stiff problems based on the numerical implementation of the holomorphic regularization method of the Cauchy problem for singularly perturbed nonlinear differential equations.

Keywords: singularly perturbed differential equation, Cauchy problem, stiffness, nonlinearity, holomorphic regularization method, numerical solution.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 819–834

УДК 519.63

ТОЖДЕСТВА ДЛЯ МЕР ОТКЛОНЕНИЙ ОТ РЕШЕНИЙ ПАРАБОЛО-ГИПЕРБОЛИЧЕСКИХ УРАВНЕНИЙ

© 2024 г. С.И. Репин[1,2,*]

[1] 191023 С-Петербург, Фонтанка, 27, СПб отд. Математического ин-та им. В.А. Стеклова РАН, Россия
[2] 195251 С.-Петербург, Политехническая, 29, СПб Политехнический ун-т Петра Великого, Россия
**e-mail: repin@pdmi.ras.ru*

Поступила в редакцию 19.11.2023 г.
Переработанный вариант 19.11.2023 г.
Принята к публикации 14.01.2024 г.

В статье получены интегральные тождества, которые выполняются для разности между точным решением начально-краевой задачи для парабологиперболического уравнения и любой функцией из соответствующего энергетического класса. Тождества позволяют получать двусторонние апостериорные оценки для приближенных решений соответствующей задачи Коши. Левая часть оценки представляет собой естественную меру отклонения от решения, а правая зависит только от данных задачи и самого приближенного решения и поэтому может быть явно вычислена. Полученные оценки используются для сравнения решений задач Коши для параболического уравнения и парабологиперболического уравнения с малым параметром при второй производной по времени. Также оценки позволяют количественно оценить эффекты, возникающие из-за неточности начальных данных и коэффициентов уравнения. Библ. 16. Фиг. 5.

Ключевые слова: задача Коши для параболо-гиперболических уравнений, контроль точности приближенных решений, апостериорные оценки функционального типа.

DOI: 10.31857/S0044466924050101, **EDN**: YDBSXC

1. ВВЕДЕНИЕ

Пусть Ω является открытой, ограниченной областью в $\mathbb{R}^d$ с липшицевой границей Γ, $T > 0$, $I := (0,T)$, а $Q_T := \Omega \times I$ обозначает пространственно-временной цилиндр с боковой поверхностью $S_T := \Gamma \times I$. Мы рассматриваем задачу нахождения функции $u(x,t)$, удовлетворяющей уравнению

$$\delta u_{tt} + u_t - \operatorname{div} A\nabla u + \nu u = f \qquad \text{в } Q_T \tag{1.1}$$

с условиями

$$u = 0 \quad \text{на } S_T, \tag{1.2}$$
$$u(x,0) = \phi(x), \tag{1.3}$$
$$u_t(x,0) = \psi(x). \tag{1.4}$$

Предполагается, что δ и ν это вещественные функции, зависящие только от x и такие, что

$$0 < \delta_1 \le \delta(x) \le \delta_2 \le 1, \qquad 0 \le \nu_1 \le \nu(x) \le \nu_2. \tag{1.5}$$

Далее, $A = A(x)$ это симметричная матрица, удовлетворяющая условию

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

где $c_1 > 0$, точка обозначает скалярное произведение векторов, а $|\xi|$ обозначает евклидову норму вектора ξ. Также предполагается, что

$$f,\ f_t \in L_{2,1}(Q_T), \tag{1.7}$$
$$\phi \in \overset{\circ}{H}{}^1(\Omega) \cap H^2(\Omega), \qquad \psi \in \overset{\circ}{H}{}^1(\Omega). \tag{1.8}$$

Здесь $L_{2,1}(Q_T)$ обозначает пространство функций $g \in L_1(Q_T)$ с конечной нормой $\|g\|_{2,1,Q_T} := \int_0^T \|g(\cdot,t)\|_\Omega\, dt$, а $\overset{\circ}{H}{}^1(\Omega)$ является подпространством $H^1(\Omega)$, состоящим из функций, обращающихся в ноль на границе. В статье используются стандартные обозначения пространств Лебега и Соболева $L^p(\Omega)$ и $W_p^l(\Omega)$, $l, p \geq 1$, $H^k(\Omega) := W_2^k(\Omega)$, $k \geq 1$, а символ $\circ$ означает, что функции обращаются в ноль на S_T. L^2 — нормы функций в Ω и Q_T обозначаются $\|\cdot\|_\Omega$ и $\|\cdot\|_{Q_T}$ соответственно. Также используются нормы

$$\|\nabla w\|^2_{A,\Omega} := \int\limits_\Omega A\nabla w \cdot \nabla w dx, \quad \|\nabla w\|^2_{A,Q_T} := \int\limits_0^T \|\nabla\|^2_{A,\Omega} dt$$

и

$$\|y^*\|^2_{A^{-1},\Omega} := \int\limits_\Omega A^{-1} y^* \cdot y^* dx, \quad \|y^*\|^2_{A^{-1},Q_T} := \int\limits_0^T \|y^*\|^2_{A^{-1},\Omega} dt.$$

Для разности значений функции при $t = t_1$ и $t = t_2 \geq t_1$ используется обозначение

$$\Big[\!\Big[g(t) \Big]\!\Big]_{t_1}^{t_2} := g(t_2) - g(t_1),$$

а $z_+ := \max\{0, z\}$. Производные функции v по пространственным переменным x_i обозначаются $v_{,i}$, а v_t обозначает производную по времени. Операторы ∇ и div означают градиент и дивергенцию по пространственным переменным. Замыкание (в норме $H^1(Q_T)$) множества гладких функций, которые обращаются в ноль вблизи S_T обозначается $\mathcal{H}(Q_T)$. Это гильбертово пространство со скалярным произведением

$$(u,v)_{\mathcal{H}} := \int\limits_{Q_T} (uv + u_t v_t + \nabla u \cdot \nabla v)\, dxdt$$

и нормой $\|v\|_{\mathcal{H}} := (v,v)_{\mathcal{H}}^{1/2}$. Также мы будем использовать пространство векторно-значных функций $\Sigma(Q_T) := L^2(Q_T, \mathbb{R}^d)$ и его подпространства

$$Y^*(Q_T) := \{y^* \in \Sigma(Q_T),\, y_t^* \in \Sigma(Q_T)\} \;\text{ и }\; Y^*_{\mathrm{div}}(Q_T) := \{y^* \in \Sigma(Q_T),\, \mathrm{div} y^* \in L^2(Q_T)\},$$

которые снабжены нормами

$$\|y^*\|^2_{Y^*} := \|y^*\|^2 + \|y_t^*\|^2 \quad \text{и} \quad \|y^*\|^2_{\mathrm{div}} := \|y^*\|^2 + \|\mathrm{div} y^*\|^2$$

соответственно.

Математические свойства уравнений типа (1.1) изучались многими авторами (см., например, [1–4]). Обобщенное решение определяется как функция $u \in V_0(Q_T)$, удовлетворяющая интегральному тождеству

$$\int\limits_{Q_T} (\delta u_{tt} w + u_t w + A\nabla u \cdot \nabla w + \nu u w)\, dxdt = \int\limits_{Q_T} f w dxdt \qquad \forall w \in V_0(Q_T) := \mathcal{H}(Q_T) \cap H^2(Q_T) \tag{1.9}$$

и условиям (1.3) и (1.4).

Если $d = 1$, то (1.1) совпадает с хорошо известным телеграфным уравнением, которое описывает распространение электрического сигнала. При $d > 1$ рассмотрение (1.1)–(1.4) мотивировано рядом физических соображений [5]. При моделировании процессов переноса в теории жидкости и газа уравнения подобного вида имеют ряд преимуществ по сравнению с соответствующими параболическими уравнениями [6], поскольку первый член уравнения играет роль регуляризатора и улучшает свойства конечномерных задач, возникающих при аппроксимации системы.

Точное решение задачи (1.1)–(1.4), как правило, построить не удается и оно замещается некоторым приближением v. Неизбежно возникает вопрос: насколько v отличается от u? Нас интересует ответ на этот вопрос не в асимптотическом смысле, т.е. не априорные оценки погрешности (как, например, те что показывают скорость, с которой галеркинская аппроксимация u_h на сетке с характерным размером ячейки h стремится к u в некоторой метрике при $h \to 0$). Асимптотические оценки устанавливают принципиальную правильность избранного метода аппроксимации при выполнении ряда дополнительных условий (обычно это регулярность решения,

ограничения на структуру сетки и требование абсолютной точности вычислений). Кроме того, эти оценки содержат константы, зависящие от точного решения. Значения констант как правило неизвестны, а возможные их оценки часто являются весьма грубыми и могут сильно переоценивать истинную величину.

Априорные оценки важны в теоретическом плане, но малопригодны для оценки точности конкретного приближенного решения, полученного в реальных вычислениях. Для этой цели используются так называемые апостериорные оценки. Существует несколько подходов к построению таких оценок (см., например, монографии [7–10]). Часто апостериорный анализ сводится к получению просто вычисляемого индикатора погрешности, главная цель которого состоит в предоставлении информации, необходимой для адаптации конечномерного подпространства используемого для аппроксимации решения. В частности, индикаторы основанные на усреднении градиента (gradient averaging) или на анализе невязки уравнения (residual based), широко используются для адаптации сетки в методе конечных элементов. Работоспособность этих методов существенно зависит от ряда условий, которые предъявляются к приближенному решению (галеркинская ортогональность), точному решению (регулярность) и вычислительной сетке. Аналогичные условия возникают при использовании других индикаторов. Как правило, индикаторы дают некоторое представление о распределении ошибки в области, но не могут гарантированно оценить величину отклонения приближенного решения от точного. Для получения полностью гарантированных и вычисляемых оценок необходимо несколько иначе сформулировать саму задачу контроля точности и использовать для ее решения другие (более общие) методы. Это привело к возникновению класса так называемых функциональных апостериорных оценок (a posteriori estimates of the functional type), которые выполняются для любых функций из функционального класса, содержащего обобщенное решение задачи. Подробное изложение соответствующей теории можно найти в монографии [9].

Основой для получения апостериорных оценок функционального типа являются тождества, которые устанавливают равенство между мерами отклонения произвольной функции от точного решения и некоторым функционалом, зависящим от этой функции и данных задачи. Для эллиптических и параболических уравнений такие тождества и вытекающие из них оценки были ранее исследованы в [11–16] и ряде других работ, ссылки на которые даны в указанных публикациях.

В данной статье тождества для мер отклонений произвольных функций от точного решения получены для начально–краевой задачи (1.1)–(1.4) (Теорема 1). Их левые части представляют собой некоторые меры отклонения функций от решения задачи, а правые содержат комплексы, которые можно трактовать как невязки в соотношениях, которые определяют уравнение и начальные условия рассматриваемой задачи. Важно отметить, что эти меры имеют вполне определенный вид. Они автоматически возникают при преобразовании соотношений, определяющих обобщенное решение задачи. Фактически меры индуцируются самим дифференциальным уравнением и в этом смысле являются наиболее естественными характеристиками точности приближенных решений. Правые части тождеств также содержат неизвестные функции e и e^*, которые, однако, можно оценить и подчинить упомянутым мерам. Это можно сделать различными способами. Один из вариантов подробно рассмотрен в § 3. Здесь же показано как оценки можно использовать для учета ошибок, связанных с неточностью в данных задачи. Заключительный § 4 содержит несколько примеров, целью которых является численная проверка тождеств (2.4) и (2.5), а также работоспособности оценок, полученных в § 3.

2. ТОЖДЕСТВА ДЛЯ ОТКЛОНЕНИЙ ОТ РЕШЕНИЯ ЗАДАЧИ (1.1)–(1.4)

Уравнение (1.1) удобно записать в виде

$$\delta u_{tt} + u_t - \operatorname{div} p^* + \nu u = f \qquad \text{в } Q_T, \tag{2.1}$$

$$p^* = A\nabla u, \tag{2.2}$$

введя в рассмотрение вектор-функцию p^* (поток). Пусть $v(x,t)$ является приближением функции $u(x,t)$, а вектор функция $y^*(x,t)$ рассматривается в качестве приближения точного потока $p^*(x,t)$ и

$$v \in V_0(Q_T), \quad y^* \in \mathcal{Q}^*(Q_T) := Y^*(Q_T) \cap Y^*_{\mathrm{div}}(Q_T). \tag{2.3}$$

Функции $\mathsf{e} := v - u$ и $\mathsf{e}^* := y^* - p^*$ являются отклонениями от решения u и соответствующего потока p^*. Определим функционалы

$$\boldsymbol{\mu}_1(\mathsf{e}, T) := \quad 2\int\limits_0^T \Big(\|\nabla \mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}\|^2_{\nu,\Omega} + \|\mathsf{e}_t\|^2_{1-\delta,\Omega} \Big) dt + \mathcal{E}(T)$$

и

$$\mathbf{\mu}_2(\mathsf{e}, \mathsf{e}^*, T) := \int\limits_0^T \Big(\|\nabla \mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + 2\|\mathsf{e}\|^2_{\nu,\Omega} + 2\|\mathsf{e}_t\|^2_{1-\delta,\Omega} \Big) dt + \mathcal{E}(T),$$

где

$$\mathcal{E}(t) := \int\limits_\Omega \Big((1 + \nu - \delta)|\mathsf{e}|^2 + \delta|\mathsf{e} + \mathsf{e}_t|^2 + A^{-1}\mathsf{e}^* \cdot \mathsf{e}^* \Big) dx.$$

Нетрудно видеть, что $\mathbf{\mu}_i \geq 0$, $i = 1, 2$, $\mathbf{\mu}_i = 0$ в том и только том случае, когда v совпадает с u, а y^* с p^*. Таким образом, $\mathbf{\mu}_i(\mathsf{e}, \mathsf{e}^*)$ можно рассматривать как меры отклонения v и y^* от u и p^*. Фактически $\mathbf{\mu}_1(\mathsf{e}, T)$ образована квадратом энергетической нормы, естественной для задачи (2.1), (2.2), а $\mathbf{\mu}_2(\mathsf{e}, \mathsf{e}^*, T)$ отличается от нее введением слагаемого, контролирующего норму отклонения в терминах потока.

Определим функции

$$\mathcal{S}(v, y^*) := A\nabla v - y^* \quad \text{и} \quad \mathcal{R}(v, y^*) := f - \delta v_{tt} - v_t + \operatorname{div} y^* - \nu v,$$

которые можно рассматривать как невязки в соотношениях (2.1) и (2.2). Следующая теорема устанавливает интегральное тождество, связывающее эти функции с мерами $\mathbf{\mu}_1$ и $\mathbf{\mu}_2$.

Теорема 1. *При выполнении условий* (1.5), (1.6) *и* (2.3) *имеют место тождества*

$$\mathbf{\mu}_1(\mathsf{e}, T) = \mathcal{E}(0) + \Big[\!\Big[\|\mathcal{S}(v, y^*)\|^2_{A^{-1},\Omega} \Big]\!\Big]_0^T + 2\int\limits_0^T \int\limits_\Omega \Big((\mathcal{S}(v, y^*) - \mathcal{S}_t(v, y^*)) \cdot \nabla \mathsf{e} - \mathcal{R}(v, y^*)(\mathsf{e} + \mathsf{e}_t) \Big) dxdt \tag{2.4}$$

и

$$\mathbf{\mu}_2(\mathsf{e}, \mathsf{e}^*, T) = \mathcal{E}(0) + \|\mathcal{S}(v, y^*)\|^2_{A^{-1},Q_T} - 2\int\limits_{Q_T} \Big(\mathcal{S}_t(v, y^*) \cdot A^{-1}\mathsf{e}^* + \mathcal{R}(v, y^*)(\mathsf{e} + \mathsf{e}_t) \Big) dxdt. \tag{2.5}$$

Доказательство. Преобразуем интегральное тождество (1.9) к виду

$$\int\limits_{Q_T} (\delta(u - v)_{tt} w + (u - v)_t w + A\nabla(u - v) \cdot \nabla w + \nu(u - v)w) dxdt =$$

$$= \int\limits_{Q_T} \Big((f - \delta v_{tt} - v_t - \nu v)w - A\nabla v \cdot \nabla w \Big) dxdt \quad \forall w \in V_0(Q_T) \tag{2.6}$$

и используем тот факт, что

$$\int\limits_\Omega (w \operatorname{div} y^* + \nabla w \cdot y^*) dx = \int\limits_\Gamma w(y^* \cdot n) ds = 0. \tag{2.7}$$

Тогда для $w = u - v$ тождество (2.6) можно записать в виде

$$\int\limits_{Q_T} (\delta \mathsf{e}_{tt}\mathsf{e} + \mathsf{e}_t\mathsf{e} + A\nabla \mathsf{e} \cdot \nabla \mathsf{e} + \nu \mathsf{e}^2)\, dxdt = \int\limits_{Q_T} \mathcal{S}(v, y^*) \cdot \nabla \mathsf{e}\, dxdt - \int\limits_{Q_T} \mathcal{R}(v, y^*)\, \mathsf{e} dxdt. \tag{2.8}$$

Поскольку

$$\int\limits_{Q_T} \mathsf{e}_t \mathsf{e} dxdt = \frac{1}{2}\Big[\!\Big[\|\mathsf{e}\|^2_\Omega \Big]\!\Big]_0^T$$

и

$$\int\limits_{Q_T} \delta \mathsf{e}_{tt}\mathsf{e} = \int\limits_{Q_T} \delta(\frac{1}{2}(\mathsf{e}^2)_{tt} - \mathsf{e}_t^2) dxdt = \Big[\!\Big[\int_\Omega \delta \mathsf{e}_t\, \mathsf{e} \Big]\!\Big]_0^T - \|\mathsf{e}_t\|^2_{\delta,Q_T},$$

где $\|\mathsf{e}_t\|^2_{\delta,Q_T} := \int\limits_{Q_T} \delta|\mathsf{e}_t|^2 dxdt$, то из (2.8) следует что для любых v и y^* имеет место равенство

$$\int\limits_0^T \|\nabla \mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}\|^2_{\nu,Q_T} + \frac{1}{2}\Big[\!\Big[\, \|\mathsf{e}\|^2_{\Omega} \,\Big]\!\Big]_0^T + \Big[\!\Big[\int\limits_\Omega \delta \mathsf{e}_t\, \mathsf{e} \Big]\!\Big]_0^T - \|\mathsf{e}_t\|^2_{\delta,Q_T} = \int\limits_{Q_T} \mathcal{S}(v,y^*)\cdot\nabla \mathsf{e} dxdt - \int\limits_{Q_T} \mathcal{R}(v,y^*)\mathsf{e} dxdt. \quad (2.9)$$

Положим в (2.6) $w = -\mathsf{e}_t$ и используем тождество, аналогичное (2.7). В результате имеем тождество

$$\int\limits_{Q_T} (\delta e_{tt}\mathsf{e}_t + \mathsf{e}_t\mathsf{e}_t + A\nabla \mathsf{e}\cdot\nabla \mathsf{e}_t + \nu \mathsf{e}_t \mathsf{e})\, dxdt = \int\limits_{Q_T} \mathcal{S}(v,y^*)\cdot\nabla \mathsf{e}_t dxdt - \int\limits_{Q_T} \mathcal{R}(v,y^*)\mathsf{e}_t dxdt. \quad (2.10)$$

Первое слагаемое в правой части (2.10) преобразуем следующим образом:

$$\int\limits_{Q_T} (A\nabla v - y^*)\cdot\nabla \mathsf{e}_t\, dxdt = \int\limits_{Q_T} A^{-1}(A\nabla v - y^*)\cdot A\nabla(v_t - u_t)\, dxdt =$$

$$= \int\limits_{Q_T} A^{-1}(A\nabla v - y^*)\cdot(A\nabla v - y^*)_t\, dxdt + \int\limits_{Q_T} (\nabla v - A^{-1}y^*)\cdot(y^*_t - p^*_t)\, dxdt.$$

Поскольку

$$\int\limits_0^t \int\limits_\Omega A^{-1}\mathcal{S}(v,y^*)\cdot\mathcal{S}_t(v,y^*) dxdt = \frac{1}{2}\int\limits_0^t \int\limits_\Omega \frac{\partial}{\partial t} A^{-1}\mathcal{S}(v,y^*)\cdot\mathcal{S}(v,y^*) dxdt = \frac{1}{2}\Big[\!\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]\!\Big]_0^t, \quad (2.11)$$

приходим к равенству

$$\int\limits_{Q_T} (A\nabla v - y^*)\cdot\nabla \mathsf{e}_t\, dxdt = \frac{1}{2}\Big[\!\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]\!\Big]_0^T +$$

$$+ \int\limits_{Q_T} (\nabla v - A^{-1}p^*)\cdot(y^*_t - p^*_t)\, dxdt - \int\limits_{Q_T} A^{-1}(y^* - p^*)\cdot(y^*_t - p^*_t)\, dxdt. \quad (2.12)$$

С учетом того, что

$$\int\limits_{Q_T} (\nabla v - A^{-1}p^*)\cdot(y^*_t - p^*_t)\, dxdt = \int\limits_{Q_T} \nabla \mathsf{e}\cdot(y^*_t - A\nabla v_t) dxdt + \int\limits_{Q_T} \nabla \mathsf{e}\cdot(A\nabla v_t - p^*_t)\, dxdt =$$

$$= \int\limits_{Q_T} \nabla \mathsf{e}\cdot(y^*_t - A\nabla v_t) dxdt + \int\limits_{Q_T} A\nabla \mathsf{e}\cdot\nabla \mathsf{e}_t\, dxdt =$$

$$= -\int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot\nabla \mathsf{e}\, dxdt + \int\limits_{Q_T} A\nabla \mathsf{e}\cdot\nabla \mathsf{e}_t\, dxdt, \quad (2.13)$$

и

$$\int\limits_{Q_T} A^{-1}(y^* - p^*)\cdot(y^*_t - p^*_t)\, dxdt = \int\limits_{Q_T} A^{-1}\mathsf{e}^*\cdot\mathsf{e}^*_t dxdt = \frac{1}{2}\Big[\!\Big[\, \|\mathsf{e}^*\|^2_{A^{-1},\Omega} \,\Big]\!\Big]_0^T, \quad (2.14)$$

представляем (2.12) в виде

$$\int\limits_{Q_T} (A\nabla v - y^*)\cdot\nabla \mathsf{e}_t\, dxdt = \frac{1}{2}\Big[\!\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]\!\Big]_0^T -$$

$$- \int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot\nabla \mathsf{e}\, dxdt + \int\limits_{Q_T} A\nabla \mathsf{e}\cdot\nabla \mathsf{e}_t\, dxdt - \frac{1}{2}\Big[\!\Big[\, \|\mathsf{e}^*\|^2_{A^{-1},\Omega} \,\Big]\!\Big]_0^T.$$

Учитывая, что

$$\int\limits_{Q_T} \delta e_{tt} \mathsf{e}_t dxdt = \frac{1}{2}\Big[\, \|\mathsf{e}_t\|^2_{\delta,Q_T} \,\Big]_0^T \quad \text{и} \quad \int\limits_{Q_T} \nu \mathsf{e}_t e dxdt = \frac{1}{2}\Big[\, \|\mathsf{e}\|^2_{\nu,\Omega} \,\Big]_0^T,$$

мы теперь можем записать (2.10) в виде

$$\frac{1}{2}\Big[\, \|\mathsf{e}_t\|^2_{\delta,Q_T} + \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + \|\mathsf{e}\|^2_{\nu,\Omega} \,\Big]_0^T + \|\mathsf{e}_t\|^2_{Q_T} = \\ = \frac{1}{2}\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},Q_T} \,\Big]_0^T - \int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot\nabla \mathsf{e} dxdt - \int\limits_{Q_T} \mathcal{R}(v,y^*)\mathsf{e}_t dxdt. \quad (2.15)$$

Суммируя (2.9) и (2.15), приходим к тождеству

$$\|\nabla \mathsf{e}\|^2_{A,Q_T} + \|\mathsf{e}\|^2_{\nu,Q_T} + \|\mathsf{e}_t\|^2_{1-\delta,Q_T} + \frac{1}{2}\Big[\, \|\mathsf{e}\|^2_{\Omega} + \int_\Omega \delta \mathsf{e}_t\, dx + \|\mathsf{e}_t\|^2_{\delta,Q_T} + \|\mathsf{e}\|^2_{\nu,\Omega} + \|\mathsf{e}^*\|^2_{A^{-1},\Omega} \,\Big]_0^T = \\ = \int\limits_{Q_T} (\mathcal{S}(v,y^*) - \mathcal{S}_t(v,y^*))\cdot\nabla \mathsf{e} dxdt - \int\limits_{Q_T} \mathcal{R}(v,y^*)(\mathsf{e}+\mathsf{e}_t) dxdt + \frac{1}{2}\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},Q_T} \,\Big]_0^T. \quad (2.16)$$

Поскольку

$$\frac{1}{2}\left(\Big[\, \|\mathsf{e}\|^2_{\delta,\Omega} + 2\int_\Omega \delta \mathsf{e}_t\, e dx + \|\mathsf{e}_t\|^2_{\delta,Q_T} \,\Big]_0^T\right) = \frac{1}{2}\Big[\, \|\mathsf{e}+\mathsf{e}_t\|^2_{\delta,\Omega} \,\Big]_0^T,$$

левая часть (2.16) совпадает с $\frac{1}{2}\boldsymbol{\mu}_1(\mathsf{e},T)$. Умножив обе части на 2, получаем (2.4). Для доказательства тождества (2.5) заметим, что

$$2\int\limits_{\Omega} \mathcal{S}(v,y^*)\cdot\nabla \mathsf{e} dxdt = \|\nabla \mathsf{e}\|^2_{A,\Omega} - \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega}. \quad (2.17)$$

Кроме того (см. (2.11)),

$$\int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot(\nabla \mathsf{e} - A^{-1}\mathsf{e}^*) dxdt = \int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot(\nabla(v-u) - A^{-1}(y^*-p^*)) dxdt = \\ = \int\limits_{Q_T} \mathcal{S}_t(v,y^*)\cdot A^{-1}\mathcal{S}(v,y^*) dxdt = \frac{1}{2}\Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]_0^t,$$

Поэтому

$$2\int\limits_0^t\int\limits_\Omega \mathcal{S}_t(v,y^*)\cdot\nabla \mathsf{e} dxdt = 2\int\limits_0^t\int\limits_\Omega \mathcal{S}_t(v,y^*)\cdot A^{-1}\mathsf{e}^* dxdt + \Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]_0^t. \quad (2.18)$$

С учетом (2.17) и (2.18) получаем

$$2\int\limits_{Q_t} (\mathcal{S}(v,y^*) - \mathcal{S}_t(v,y^*))\cdot\nabla \mathsf{e} dxdt = \\ = \int\limits_0^t \Big(\|\nabla \mathsf{e}\|^2_{A,\Omega} - \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega}\Big) dt - 2\int\limits_{Q_t} \mathcal{S}_t(v,y^*)\cdot A^{-1}\mathsf{e}^* dxdt - \Big[\, \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \,\Big]_0^t$$

и тождество (2.4) преобразуется в (2.5).

Замечание 1. Если $\delta = 0$, то $\mathcal{E}(T) := \|\mathsf{e}(x,T)\|^2_{1+\nu,\Omega} + \|\mathsf{e}^*(x,T)\|^2_{A^{-1},\Omega}$ и тождества (2.4) и (2.5) принимают вид

$$\int\limits_0^T \Big(2\|\nabla \mathsf{e}\|^2_{A,\Omega} + 2\|\mathsf{e}\|^2_{\nu,\Omega} + 2\|\mathsf{e}_t\|^2_{\Omega}\Big)dt + \mathcal{E}(T) =$$

$$= \mathcal{E}(0) + \Big[\!\Big[\|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega} \Big]\!\Big]_0^T + 2\int\limits_0^T\int\limits_\Omega \Big(\mathcal{S}(v,y^*) - \mathcal{S}_t(v,y^*))\cdot\nabla \mathsf{e} - \mathcal{R}(v,y^*)(\mathsf{e}+\mathsf{e}_t)\Big)dxdt$$

и

$$\int\limits_0^T \Big(\|\nabla \mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + 2\|\mathsf{e}\|^2_{\nu,\Omega} + 2\|\mathsf{e}_t\|^2_{\Omega}\Big)dt + \mathcal{E}(T) =$$

$$= \mathcal{E}(0) + \int\limits_0^T \|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega}dt - 2\int\limits_0^T\int\limits_\Omega \Big(\mathcal{S}_t(v,y^*)\cdot A^{-1}\mathsf{e}^* + \mathcal{R}(v,y^*)(\mathsf{e}+\mathsf{e}_t)\Big)dxdt,$$

где $\mathcal{R}(v,y^*) = \mathrm{div} y^* + f - v_t - \nu v$. Для $\nu = 0$ и единичной матрицы A такие тождества были ранее получены в [16].

Замечание 2. Подстановка $\widehat{u}(x,t) = e^{-\frac{t}{2}}u(x,t)$ преобразует гиперболическое уравнение

$$u_{tt} - \mathrm{div} A\nabla u + \nu u = f \tag{2.19}$$

с условиями (1.2)–(1.4) в уравнение

$$\widehat{u}_{tt} + \widehat{u}_t - \mathrm{div} A\nabla \widehat{u} + \widehat{\nu}\widehat{u} = \widehat{f}, \quad \widehat{\nu} = \frac{1}{4} + \nu, \quad \widehat{f} = fe^{-\frac{t}{2}} \tag{2.20}$$

с начальными условиями $\widehat{u}(x,0) = \phi(x)$, $\widehat{u}_t(x,0) = -\frac{1}{2}\phi(x) + \psi(x)$ и краевым условием $\widehat{u}(x,t) = 0$ на S_T. Пусть $v(x,t)$ и $y^*(x,t)$ являются приближениями функций $u(x,t)$ и $p^*(x,t) := A\nabla u$. Функции $\widehat{v}(x,t) := e^{-\frac{t}{2}}v(x,t)$ и соответствующий поток $\widehat{y^*}(x,t) := e^{-\frac{t}{2}}y^*(x,t)$ можно рассматривать как приближенные решения задачи Коши для уравнения (2.20), для которых выполняются тождества (2.4) и (2.5). При этом $\widehat{\delta} = 1$,

$$\widehat{\mathsf{e}}(x,t) = e^{-\frac{t}{2}}\mathsf{e}(x,t),\ \widehat{\mathsf{e}^*}(x,t) = e^{-\frac{t}{2}}\mathsf{e}^*(x,t),\ \widehat{\mathsf{e}}_t = e^{-\frac{t}{2}}\left(\mathsf{e}_t - \frac{1}{2}\mathsf{e}\right),\ \widehat{\mathsf{e}} + \widehat{\mathsf{e}}_t = e^{-\frac{t}{2}}\left(\frac{1}{2}\mathsf{e} + \mathsf{e}_t\right),$$

а меры $\widehat{\boldsymbol{\mu}}_1(\mathsf{e},T)$ и $\widehat{\boldsymbol{\mu}}_2(\mathsf{e},\mathsf{e}^*,T)$ для уравнения (2.20) имеют вид

$$\widehat{\boldsymbol{\mu}}_1(\widehat{\mathsf{e}},T) := \quad 2\int\limits_0^T (\|\nabla\widehat{\mathsf{e}}\|^2_{A,\Omega} + \|\widehat{\mathsf{e}}\|^2_{\widehat{\nu},\Omega})dt + \widehat{\mathcal{E}}(T),$$

$$\widehat{\boldsymbol{\mu}}_1(\widehat{\mathsf{e}},\widehat{\mathsf{e}^*},T) := \quad \int\limits_0^T (\|\nabla\widehat{\mathsf{e}}\|^2_{A,\Omega} + \|\widehat{\mathsf{e}^*}\|_{A^{-1},\Omega} + 2\|\widehat{\mathsf{e}}\|^2_{\widehat{\nu},\Omega})dt + \widehat{\mathcal{E}}(T),$$

где

$$\widehat{\mathcal{E}}(t) := \int\limits_\Omega \Big(|\widehat{\mathsf{e}} + \widehat{\mathsf{e}}_t|^2 + A^{-1}\widehat{\mathsf{e}^*}\cdot\widehat{\mathsf{e}^*} + \widehat{\nu}|\widehat{\mathsf{e}}|^2\Big)dx.$$

Таким образом, для (2.19) меры преобразуются в

$$\boldsymbol{\mu}_1(\mathsf{e},T) := \quad 2\int\limits_0^T e^{-t}\Big(\|\nabla\mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}\|^2_{\widehat{\nu},Q_T}\Big) + \mathcal{E}(T),$$

$$\boldsymbol{\mu}_2(\mathsf{e},\mathsf{e}^*,T) := \quad \int\limits_0^T e^{-t}\Big(\|\nabla\mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}^*\|_{A^{-1},\Omega} + 2\|\mathsf{e}\|^2_{\widehat{\nu},Q_T}\Big) + \mathcal{E}(T),$$

где

$$\mathcal{E}(t) := e^{-t}\int\limits_{\Omega}\Big(|\frac{1}{2}\mathsf{e}+\mathsf{e}_t|^2 + A^{-1}\mathsf{e}^*\cdot\mathsf{e}^* + \widehat{\mathrm{v}}|\mathsf{e}|^2\Big)dx.$$

С учетом того, что

$$\widehat{\mathcal{R}}(\widehat{v},\widehat{y^*}) = e^{-\frac{t}{2}}\mathcal{R}(v,y^*), \qquad \widehat{\mathcal{S}}(\widehat{v},\widehat{y^*}) = e^{-\frac{t}{2}}\mathcal{S}(v,y^*),$$
$$\widehat{\mathcal{S}}_t(\widehat{v},\widehat{y^*}) = e^{-\frac{t}{2}}\left(\mathcal{S}_t(v,y^*) - \frac{1}{2}\mathcal{S}(v,y^*)\right),$$

получаем тождества для задачи Коши, связанной с гиперболическим уравнением (2.19):

$$\mathbf{\mu}_1(\mathsf{e},T) = \mathcal{E}(0) + \Big[\!\Big[\, e^{-t}\|\mathcal{S}(v,y^*)\|^2_{A^{-1},\Omega}\,\Big]\!\Big]_0^T + \\ + \int\limits_0^T\int\limits_{\Omega} e^{-t}\Big(\big(3\mathcal{S}(v,y^*) - 2\mathcal{S}_t(v,y^*)\big)\cdot\nabla\mathsf{e} - \mathcal{R}(v,y^*)(\mathsf{e}+2\mathsf{e}_t)\Big)dxdt \quad (2.21)$$

и

$$\mathbf{\mu}_2(\mathsf{e},\mathsf{e}^*,T) = \mathcal{E}(0) + \|e^{-t/2}\mathcal{S}(v,y^*)\|^2_{A^{-1},Q_T} + \\ + \int\limits_{Q_T} e^{-t}\Big(\big(\mathcal{S}(v,y^*) - 2\mathcal{S}_t(v,y^*)\big)\cdot A^{-1}\mathsf{e}^* - \mathcal{R}(v,y^*)(\mathsf{e}+2\mathsf{e}_t)\Big)dxdt. \quad (2.22)$$

3. ОЦЕНКИ, ВЫТЕКАЮЩИЕ ИЗ ТОЖДЕСТВ (2.4) И (2.5)

Тождества установленные в разд. 2 позволяют получать апостериорные оценки точности приближенных решений. Также они позволяют сравнивать решения различных задач, которые могут отличаться коэффициентами или начальными условиями. Эти два случая рассматриваются ниже.

3.1. Апостериорные оценки

Для того чтобы получить полностью вычисляемые оценки для мер отклонения от решения необходимо в тождествах (2.4) и (2.5) оценить интегралы, которые стоят в правых частях и содержат неизвестные величины e, e_t и e^*. Это можно сделать разными способами. Рассмотрим наиболее простой, основанный на использовании интегральных неравенств Юнга. Определим вектор $\boldsymbol{\alpha} = \{\alpha_1,\alpha_2,\alpha_3,\alpha_4\}$, где

$$\alpha_i \in L^\infty_+(0,T) := \big\{\,\alpha(t)\in L^\infty(0,T),\ \alpha(t)>0,\ \forall t\in[0,T]\,\big\}, \quad i=1,2,3,4.$$

Имеют место следующие неравенства:

$$2\left|\int\limits_{Q_T}(\mathcal{S}(v,y^*)-\mathcal{S}_t(v,y^*))\cdot\nabla\mathsf{e}dxdt\right| \le \int\limits_0^T\Big(\tfrac{1}{\alpha_1}\|\mathcal{S}(v,y^*)-\mathcal{S}_t(v,y^*)\|^2_{A^{-1},\Omega} + \alpha_1\|\nabla\mathsf{e}\|^2_{A,\Omega}\Big)dt, \quad (3.1)$$

$$2\left|\int\limits_{Q_T}\mathcal{S}_t(v,y^*)\cdot A^{-1}\mathsf{e}^*dxdt\right| \le \int\limits_0^T\Big(\tfrac{1}{\alpha_2}\|\mathcal{S}_t(v,y^*)\|^2_{A^{-1},\Omega} + \alpha_2\|\mathsf{e}^*\|^2_{A^{-1},\Omega}\Big)dt, \quad (3.2)$$

$$2\left|\int\limits_{Q_T}\mathcal{R}(v,y^*)(\mathsf{e}+\mathsf{e}_t)dxdt\right| \le \int\limits_0^T\Big(\varkappa\|\mathcal{R}(v,y^*)\|^2_{\Omega} + \alpha_3\|\mathsf{e}\|^2_{\Omega} + \alpha_4\|\mathsf{e}_t\|^2_{\Omega}\Big)dt, \quad (3.3)$$

где $\varkappa = \dfrac{\alpha_3+\alpha_4}{\alpha_3\alpha_4}$. Также заметим, что

$$\int\limits_0^T\alpha_3\|\mathsf{e}\|^2_{\Omega}dt \le \int\limits_0^T(\alpha_3-2\mathrm{v})_+\|\mathsf{e}\|^2_{\Omega}dt + \int\limits_0^T 2\mathrm{v}\|\mathsf{e}\|^2_{\Omega}dt \le \int\limits_0^T(\alpha_3-2\mathrm{v})_+C_\Omega^2c_1^{-1}\|\nabla\mathsf{e}\|^2_{A,\Omega}dt + \int\limits_0^T 2\mathrm{v}\|\mathsf{e}\|^2_{\Omega}dt. \quad (3.4)$$

Из тождеств (2.4) и (2.5) с учетом (3.1)–(3.4) следуют оценки

$$\mathbb{[\![}\, \mathsf{e} \,\mathbb{]\!]}_1^2 \;\le\; \mathrm{M}_1(v, y^*), \tag{3.5}$$

$$\mathbb{[\![}\, \mathsf{e}; \mathsf{e}^* \,\mathbb{]\!]}_2^2 \;\le\; \mathrm{M}_2(v, y^*). \tag{3.6}$$

Функционалы, стоящие в правых частях (3.5) и (3.6), задаются соотношениями

$$\mathrm{M}_1(v, y^*) \;:=\; \mathcal{E}(0) \;+\; \Big[\!\Big[\|\mathcal{S}(v, y^*)\|^2_{A^{-1},\Omega} \Big]\!\Big]_0^T \;+\; \int\limits_0^T \Big(\tfrac{1}{\alpha_1} \|\mathcal{S}(v, y^*) \;-\; \mathcal{S}_t(v, y^*)\|^2_{A^{-1},\Omega} \;+\; \varkappa \|\mathcal{R}(v, y^*)\|^2_{\Omega} \Big) dt$$

и

$$\mathrm{M}_2(v, y^*) := \mathcal{E}(0) + \|\mathcal{S}(v, y^*)\|^2_{A^{-1},Q_T} + \int\limits_0^T \Big(\tfrac{1}{\alpha_2} \|\mathcal{S}_t(v, y^*)\|^2_{A^{-1},\Omega} + \varkappa \|\mathcal{R}(v, y^*)\|^2_{\Omega} \Big) dt.$$

Они зависят только от известных функций и легко вычисляются. Левые части задаются комбинированными нормами

$$\mathbb{[\![}\, \mathsf{e} \,\mathbb{]\!]}_1 := \left(\mathcal{E}(T) + \int\limits_0^T \Big(\lambda_1 \|\nabla \mathsf{e}\|^2_{A,\Omega} + \lambda_3 \|\mathsf{e}_t\|^2_{\Omega} \Big) dt \right)^{1/2}$$

и

$$\mathbb{[\![}\, \mathsf{e}; \mathsf{e}^* \,\mathbb{]\!]}_2 := \left(\mathcal{E}(T) + \int\limits_0^T \Big(\widetilde{\lambda}_1 \|\nabla \mathsf{e}\|^2_{A,\Omega} + \lambda_2 \|\mathsf{e}^*\|^2_{A^{-1},\Omega} + \lambda_3 \|\mathsf{e}_t\|^2_{\Omega} \Big) dt \right)^{1/2},$$

где

$$\lambda_1 := 2 - \alpha_1 - (\alpha_3 - 2\nu)_+ C_F^2 c_1^{-1},\ \widetilde{\lambda}_1 := \lambda_1 - 1,\ \lambda_2 := 1 - \alpha_2,\ \lambda_3 := 2(1 - \delta) - \alpha_4, \tag{3.7}$$

а параметры выбираются так, чтобы весовые функции $\lambda_i(t)$ были неотрицательны.

Замечание 3. Если $\delta > 1$, то сделав замену переменной $\widetilde{t} = \frac{1}{\varrho} t$, с $\varrho > \delta$ для $\widetilde{u}(x, \widetilde{t}) = u(x, \frac{1}{\rho} t)$ получаем уравнение

$$\widetilde{\delta} \widetilde{u}_{\widetilde{t}\widetilde{t}} + \widetilde{u}_{\widetilde{t}} - \rho \mathrm{div} A \widetilde{\nabla} \widetilde{u}_{xx} + \nu \widetilde{u} = f,$$

где $\widetilde{\delta} := \frac{\delta}{\rho} < 1$, $\widetilde{t} \in [0, \widetilde{T}]$ и $\widetilde{T} := \frac{1}{\varrho} T$. Для начально-краевой задачи с условиями $\widetilde{u}(x, 0) = \phi$, $\widetilde{u}_{\widetilde{t}}(x, 0) = \varrho\psi$ полученные выше оценки применимы.

3.2. Оценки ошибок, связанных с данными задачи

Тождества, установленные в теореме 1, позволяют сравнить решение задачи (1.1)–(1.4) с решением $\widehat{u}$, которое удовлетворяет такому же уравнению, но с другими начальными условиями:

$$\widehat{u}(x, 0) = \widehat{\phi}(x), \qquad \widehat{u}_t(x, 0) = \widehat{\psi}(x).$$

Функции $\mathsf{e}_\phi := \widehat{\phi} - \phi$ и $\mathsf{e}_\psi := \widehat{\psi} - \psi$ характеризуют разницу в начальных данных.

Нетрудно видеть, что для $v = \widehat{u}$ и $y^* = A\nabla \widehat{u}$ имеют место равенства $\mathcal{S}(v, y^*) = 0$ и $\mathcal{R}(v, y^*) = 0$. Поэтому из (2.4) следует, что при любом T выполняется тождество

$$\int\limits_0^T \Big(\|\nabla \mathsf{e}\|^2_{A,\Omega} + \|\mathsf{e}\|^2_{\nu,\Omega} + \|\mathsf{e}_t\|^2_{1-\delta,\Omega} \Big) dt + \frac{1}{2} \mathcal{E}(T) = \tag{3.8}$$

$$= \frac{1}{2} \int\limits_\Omega \left((1 + \nu - \delta) |\mathsf{e}_\phi|^2 + \delta |\mathsf{e}_\phi + \mathsf{e}_\psi|^2 + A \nabla \mathsf{e}_\phi \cdot \nabla \mathsf{e}_\phi \right) dx = \mathcal{E}(0).$$

Поскольку с ростом T интеграл в левой части (3.8) может только возрастать, а $\mathcal{E}(0)$ не зависит от T, это тождество показывает, что $\mathcal{E}(T)$ монотонно убывает. При этом $\mathcal{E}(T)$ не может стремиться к положительной величине (так как в этом случае интеграл в (3.8) будет неограниченно возрастать с ростом T). Таким образом, $\mathcal{E}(T)$ стремится к нулю. В частности, если $f = 0$, то при любых начальных данных решение затухает и стремится к нулю при $T \to +\infty$. Из (2.5) следует аналогичная оценка, которая отличается от (3.8) заменой первого слагаемого в левой части на сумму $\frac{1}{2}\|\nabla \mathsf{e}\|^2_{A,\Omega} + \frac{1}{2}\|\mathsf{e}^*\|^2_{A^{-1}}$.

С помощью (3.5) и (3.6) можно оценить разность между решениями двух задач с разными коэффициентами. В частности, такая ситуация возникает, если надо оценить насколько серьезно изменится решение при упрощении данных, например при сглаживании незначительных осцилляций коэффициентов. Для широкого класса линейных и нелинейных эллиптических уравнений этот вопрос подробно рассматривается в монографии [12]. Аналогичная проблема возникает при анализе погрешностей индуцированных неточной (неполной) информацией о данных (так называемая Uncertain Input Data Problem). Поскольку оценки (3.5) и (3.6) не требуют каких-либо особых свойств (ортогональности относительно некоторого множества, дополнительной регулярности и т.п.) их нетрудно использовать и для решения таких вопросов.

Пусть $\widehat{u}$ и $\widehat{y^*}$ определяют решение с матрицей $\widehat{A}$ вместо A, $\widehat{\nu}$ вместо ν и $\widehat{\delta}$ вместо δ. В этом случае

$$\mathcal{R}(\widehat{u}, \widehat{y^*}) = \varepsilon_\delta \widehat{u}_{tt} + \varepsilon_\nu \widehat{u}, \quad \|\mathcal{S}(v, y^*)\|^2_{A^{-1},\Omega} = \int\limits_\Omega B\nabla\widehat{u}\cdot\nabla\widehat{u}dx,$$

где $B := A^{-1}(\widehat{A} - A)(\widehat{A} - A)$, $\varepsilon_\nu := \widehat{\nu} - \nu$ и $\varepsilon_\delta := \widehat{\delta} - \delta$. Поскольку

$$\mathcal{S}(v, y^*) - \mathcal{S}_t(v, y^*) = (A - \widehat{A})\nabla(\widehat{u} - \widehat{u}_t) \quad \text{и} \quad \mathcal{R}(\widehat{u}, \widehat{p^*}) = \varepsilon_\delta \widehat{u}_{tt} + \varepsilon_\nu \widehat{u},$$

то для $\mathsf{e} = \widehat{u} - u$ и $\mathsf{e}^* = \widehat{p^*} - p^*$ с помощью (3.5) и (3.6) получаем оценки

$$[\![\, \mathsf{e} \,]\!]^2_1 \le \mathcal{E}(0) + \Big[\|\nabla\widehat{u}\|^2_{B,\Omega}\Big]_0^T + \int\limits_0^T \Big(\tfrac{1}{\alpha_1}\|\nabla(\widehat{u} - \widehat{u}_t)\|^2_{B,\Omega} + \varkappa\|\varepsilon_\delta\widehat{u}_{tt} + \varepsilon_\nu\widehat{u}\|^2_\Omega\Big)dt \tag{3.9}$$

и

$$[\![\, \mathsf{e}, \mathsf{e}^* \,]\!]^2_2 \le \mathcal{E}(0) + \int\limits_0^T \Big(\|\nabla\widehat{u}\|^2_{B,\Omega} + \tfrac{1}{\alpha_2}\|\nabla\widehat{u}_t\|^2_{B,\Omega} + \varkappa\|\varepsilon_\delta\widehat{u}_{tt} + \varepsilon_\nu\widehat{u}\|^2_\Omega\Big)dt, \tag{3.10}$$

которые позволяют оценить насколько может измениться решение при изменении коэффициентов.

4. ПРИМЕРЫ

Выполнение тождеств (2.4) и (2.5), а также эффективность оценок (3.5) и (3.6) проверялось численно в серии тестов. В качестве (1.1), (1.2) рассматривалась задача

$$\delta u_{tt} + u_t - \rho u_{xx} = 0, \qquad u(0) = u(\pi) = 0 \tag{4.1}$$

в области $Q_T = [0, T] \times (0, \pi)$ при различных положительных δ и ρ. Решение этой задачи можно представить в виде

$$u(x, t) = e^{-\dfrac{t}{2\delta}} \sum_{j=1}^{\infty} \sin(jx) T_j(t), \tag{4.2}$$

где

$$T_j(t) = \begin{cases} a_j e^{\omega_j t} + b_j e^{-\omega_j t}, & j \le N_\delta, \\ a_j \sin(\omega_j t) + b_j \cos(\omega_j t), & j > N_\delta, \end{cases}$$

$$\omega_j := \frac{1}{2\delta}\sqrt{|1 - 4\rho\delta j^2|},$$

а N_δ это целая часть $\frac{1}{2\sqrt{\delta\rho}}$. Начальные условия для (4.1) задаются соотношениями

$$\phi(x) = \sum_{j=1}^{\infty} q_j \sin(jx), \qquad q_j = \begin{cases} a_j + b_j, & j \le N_\delta, \\ b_j, & j > N_\delta \end{cases}$$
$$\psi(x) = \sum_{j=1}^{\infty} \sin(jx)(r_j - \tfrac{1}{2\delta} q_j), \qquad r_j = \begin{cases} \omega_j(a_j - b_j), & j \le N_\delta, \\ \omega_j a_j, & j > N_\delta. \end{cases}$$

Левые части оценок (3.5) и (3.6) приобретают вид

$$\mathbb{[\![}\, \mathsf{e} \,\mathbb{]\!]}_1^2 = \int_0^T \int_0^\pi \Big(\lambda_1 \rho |e_x|^2 + \lambda_3 |\mathsf{e}_t|^2 \Big) dxdt + \mathcal{E}(T),$$
$$\mathbb{[\![}\, \mathsf{e}, \mathsf{e}^* \,\mathbb{]\!]}_2^2 = \int_0^T \int_0^\pi \Big(\widetilde{\lambda}_1 \rho |e_x|^2 + \frac{\lambda_2}{\rho} |\mathsf{e}^*|^2 + \lambda_3 |\mathsf{e}_t|^2 \Big) dxdt + \mathcal{E}(T),$$

где в соответствии с (3.7)

$$\lambda_1 = 2 - \alpha_1 - \alpha_3 C_F^2(\Omega),\ \widetilde{\lambda}_1 := \lambda_1 - 1,\ \lambda_2 := 1 - \alpha_2,\ \ \lambda_3 := 2 - \alpha_4,$$

а параметры удовлетворяют условиям $\alpha_2 \le 1$, $\alpha_1 + \alpha_3 C_F^2(\Omega) \le 1$, $\alpha_4 \le 2$.

Функции v выбираются в виде

$$v(x,t) = u(x,t) + \sum_{j=1}^{m} \lambda_j \chi_j(x) \tau_j(t) + e_0(x) + t g_0(x), \tag{4.3}$$

где e_0, g_0, λ_j и χ_i – некоторые заданные функции такие, что $\chi_j(0) = \chi_j(\pi) = 0$ и $\tau(0) = 0$. В зависимости от выбора e_0 и g_0 функция v может удовлетворять или не удовлетворять условиям (1.3) и (1.4). Выбирая величину параметров λ_j можно получать как весьма близкие аппроксимации решения, так и функции, которые сильно от него отклоняются. Функции $y^*(x,t)$ выбирались в виде

$$y^* = \rho v_x - \varphi(x,t). \tag{4.4}$$

В этом случае

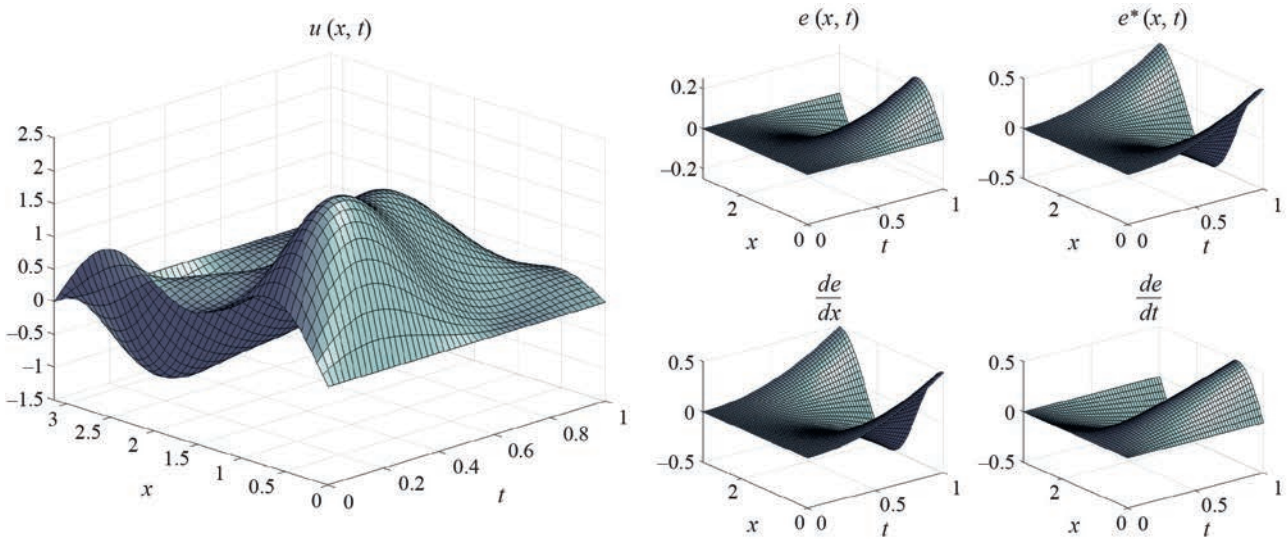
$$\mathcal{S}(v, y^*) = \varphi(x,t), \qquad \mathcal{S}_t(v, y^*) = \varphi_t(x,t),$$
$$\mathcal{R}(v, y^*) = \mathcal{R}(v) - \varphi_x(x,t), \quad \mathcal{R}(v) = -\delta v_{tt} - v_t + \rho v_{xx}.$$

В частности, если $\varphi(x,t) = 0$, то y^* определяется с помощью равенства (2.2) (что соответствует нередко применяемому в вычислительной практике варианту, когда поток реконструируется с помощью определяющего физического закона и градиента приближенного решения). Функция φ позволяет рассмотреть и другие варианты, например она может рассматриваться как корректор. Заметим, что левая часть оценки (3.5) не зависит от y^*, поэтому оценку можно улучшить за счет минимизации мажоранты по φ. Оптимальная реконструкция y^* на основе приближенного решения v достигается, если φ минимизирует функционал

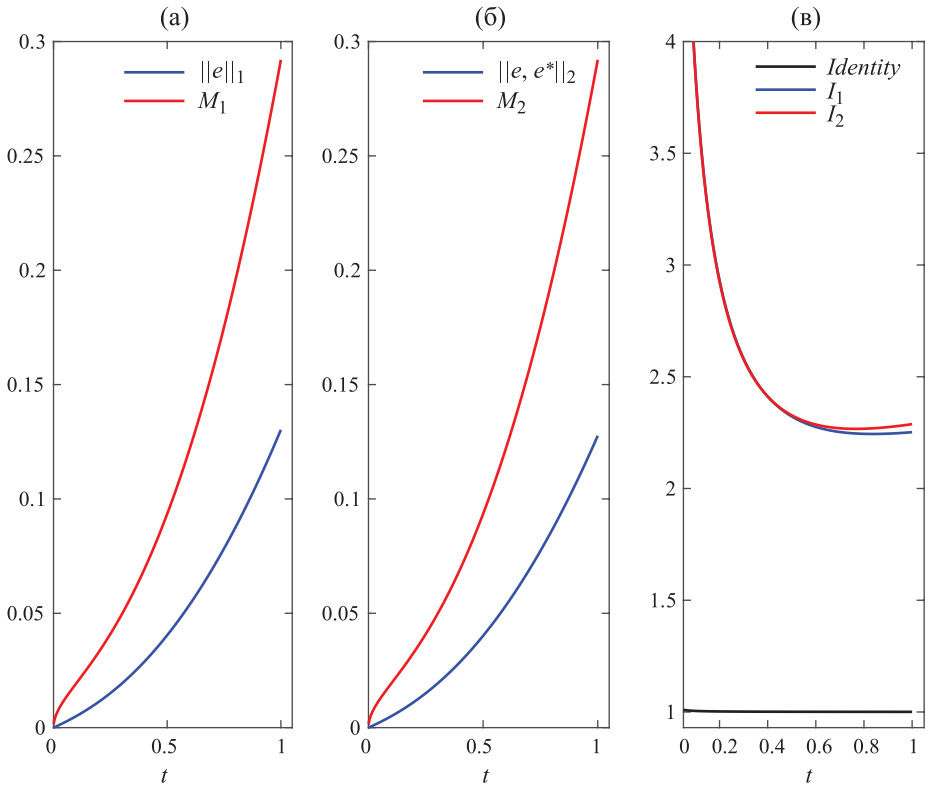
$$J_{\mathrm{corr}}(\varphi) := \int_0^T \int_0^\pi \Big(\frac{1}{\rho} |\varphi|^2 + \frac{1}{\alpha_2 \rho} |\varphi_t|^2 + \varkappa |\varphi_x|^2 - 2\varkappa \mathcal{R}(v) \varphi_x + \varkappa |\mathcal{R}(v)|^2 \Big) dxdt,$$

на некотором множестве функций (в качестве последнего можно взять например функции вида $\varphi(x,t) = \sum_{j=0}^{m} \sum_{k=0}^{n} d_{ij} (x - \frac{\pi}{2})^j t^k$). В большом количестве тестов с различными функциями u вида (4.2) и функциями v и y^* вида (4.3) (4.4) было проверено, что тождества (2.4) и (2.5) для мер отклонений от точных решений выполняются, а оценки (3.5) и (3.6) дают правильное представление о величине отклонения. Некоторые результаты обсуждаются ниже.

Пример 1. В первом примере $a_1 = 2$, $a_2 = 3$, $a_4 = 1$, $b_1 = 1$, $b_3 = 0.5$, $\delta = 0.3$, $\rho = 1$. Функции e_0 и g_0 выбраны так, что начальные условия выполняются. Решение задачи в области $(0,\pi) \times (0,1)$ и функции e,



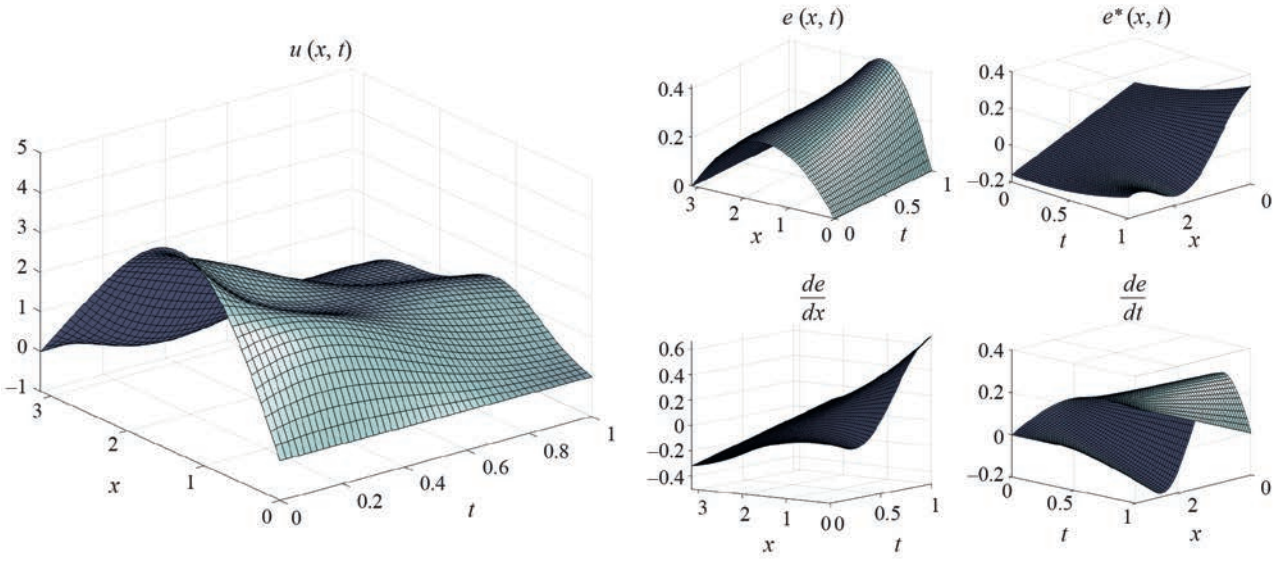
Фиг. 1. Решение u и функции e и e^*.



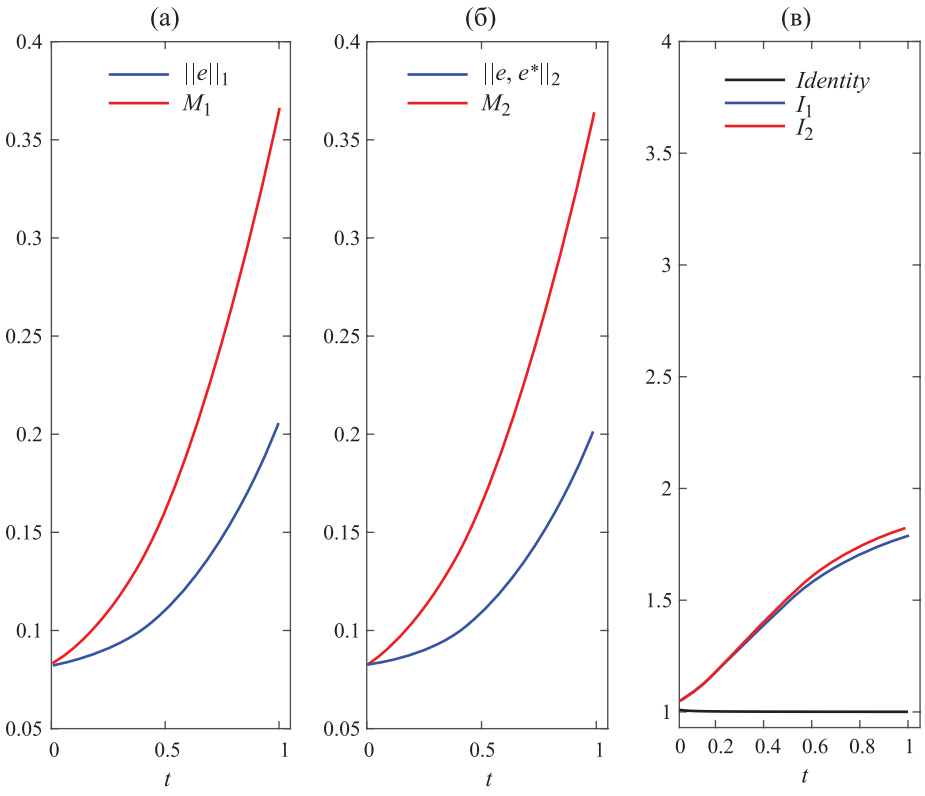
Фиг. 2. Пример 1. Меры $\|\mathsf{e}\|$ и $\|\mathsf{e}, \mathsf{e}^*\|$ и их оценки.

e_t, e_x и e^* представлены на фиг. 1. Функция ϕ равна нулю. Этот пример соответствует достаточно хорошей аппроксимации, которая удовлетворяет начальным и краевым условиям.

Тождества (2.4) и (2.5) выполняются при всех $t \in (0, 1)$. На фиг. 2в горизонтальная прямая показывает отношение правой части (2.4) к левой. Для (2.5) такое отношение также равно 1 при всех t. На фиг. 2а и 2б показано как изменяются нормы $\|\mathsf{e}\|_1$ и $\|\mathsf{e}, \mathsf{e}^*\|_2$ с ростом t (нижние кривые). Также показаны значения соответствующих мажорант M_1 и M_2 (верхние кривые). Поскольку абсолютные значения отклонений недостаточно инфор-



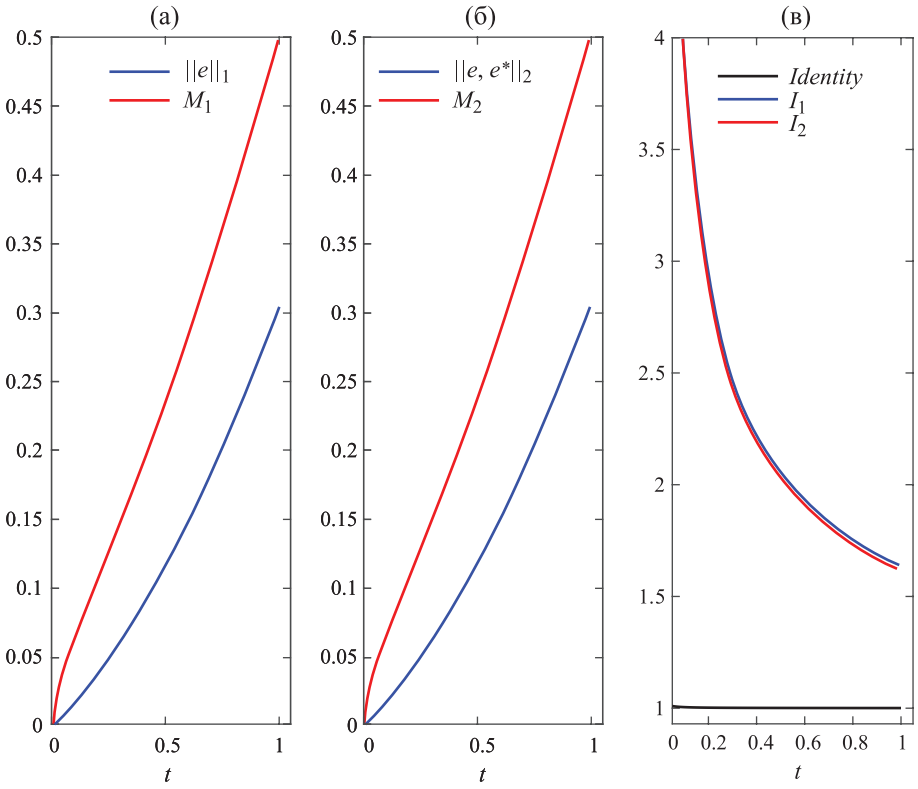
Фиг. 3. Решение u и функции e и e*.



Фиг. 4. Пример 2. Меры $[\![\mathrm{e}]\!]$ и $[\![\mathrm{e}, \mathrm{e}^*]\!]$ и их оценки.

мативны, на графиках представлены нормализованные значения, которые соотносят все данные величиной

$$[\![u]\!] := \left(\int_0^T \int_0^\pi \left(\rho |u_x|^2 + \frac{1}{\rho} |p^*|^2 + |u_t|^2 \right) dx dt \right)^{1/2},$$



Фиг. 5. Пример 3. Меры $[\![\mathsf{e}]\!]$ и $[\![\mathsf{e}, \mathsf{e}^*]\!]$ и их оценки.

которая характеризует само решение. Графики показывают, что для таких (достаточно близких к решению) аппроксимаций оценки работают хорошо. Величины $I_1 = \frac{M_1}{[\![\mathsf{e}]\!]_1}$ и $I_2 = \frac{M_2}{[\![\mathsf{e},\mathsf{e}^*]\!]_1}$ (так называемые индексы эффективности) характеризуют точность оценок. Чем ближе эти значения к 1 тем лучше оценка. Значения индексов эффективности для мажорант M_1 и M_2 представлены двумя кривыми на фиг. 2в. Видно что они переоценивают истинную меру отклонения не более чем в 2–3 раза.

Пример 2. В этом примере $a_1 = 2$, $a_2 = 1$, $a_4 = 1$, $b_1 = 1$, $b_2 = 0.5$, $\delta = 0.3$, $\rho = 0.5$ а функции v и y^* представляют собой весьма грубые аппроксимации, которые не удовлетворяют начальным условиям задачи, так что $\mathcal{E}(0) > 0$. Соответствующие функции e и e^* изображены на фиг. 3. Фиг. 4 показывает поведение мер отклонений и соответствующих мажорант. Поскольку функция v нарушает начальные условия, соответствующие кривые не равны нулю при $t = 0$. Мажоранты правильно отражают это обстоятельство и в целом качество оценок очень хорошее.

Пример 3. Данные в этом примере отличаются от примера 1 тем, что здесь $\rho = 0.01$. Тождества (2.4) и (2.5) выполняются при любых $\rho > 0$, однако эффективность оценок (3.5) и (3.6) может зависеть от величины этого коэффициента. Данный пример показывает, что даже при малых значениях ρ оценки могут оставаться работоспособными (см. фиг. 5).

В заключение отметим, что оценки (3.5) и (3.6) были получены из тождеств весьма простым способом. Они обеспечивают гарантированную мажоранту меры отклонения, но можно ожидать, что в определенных ситуациях эта мажоранта будет сильно переоценивать истинное значение. Эта трудность преодолима поскольку тождества (2.4) и (2.5) позволяют получать и более точные оценки. В контексте уравнений эллиптического типа этот вопрос подробно рассмотрен в работах [13] и [15].

СПИСОК ЛИТЕРАТУРЫ

1. *Ладыженская О.А.* О нестационарных операторных уравнениях и их приложениях к линейным задачам математической физики// Матем. сб. 1958. Т. 87. № 2. Р. 123–158.

2. *Ладыженская О.А., Солонников В.А., Уральцева Н.Н.*. Линейные и квазилинейные уравнения параболического типа, М.: Наука, 1967.

3. *Бубнов Б. А.* Смешанная задача для некоторых параболо-гиперболических уравнений // Дифференц. урния 1976. Т. 12. № 3. Р. 494–501.

4. *Ларькин Н. А.* Краевые задачи в целом для одного класса гиперболических уравнений // Сиб. Матем. ж. 1977. Т. XVIII. № 6. Р. 1414–1419.

5. *Четверушкин Б. Н.* Пределы детализации и формулировка моделей уравнений сплошных сред // Матем. моделирование. 2012. Т. 24. № 11. Р. 33–52.

6. *Давыдов А. А., Б. Н. Четверушкин Б. Н., Шильников Е. В.* Моделирование течений несжимаемой жидкости и слабосжимаемого газа на многоядерных гибридных вычислительных системах // Ж. вычисл. матем. и матем. физ. 2010. N. 50. № 12. Р. 2275–2284.

7. *Ainthworth M.,Oden J. T.* A posteriori error estimation in finite element analysis, Wiley, New York, 2000.

8. *Babuška I., Strouboulis T.* The finite element method and its reliability. Claderon Press, Oxford, 2001.

9. *Repin S.* A posteriori estimates for partial differential equations, volume 4 of Radon Series on Computational and Applied Mathematics. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

10. *Verfurth R.* A review of a posteriori error estimation and adaptive mesh-refinement techniques Wiley, Teubner, New-York, 1996.

11. *Repin S.* Estimates of deviations from exact solutions initial-boundary value problem for the heat equation // Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 2002. V. 13. № 2. P. 121–133.

12. *Repin S., Sauter S.* Accuracy of Mathematical Models. Dimension Reduction, Homogenization, and Simplification, volume 33 of EMS Tracts Math. European Mathematical Society (EMS), Berlin, 2020.

13. *Репин С.И.* Тождество для отклонений от точного решения задачи $\Lambda^* \mathcal{A} \Lambda u + \ell = 0$ и его следствия // Ж. вычисл. матем. и матем. физ. 2021 Т. 61. № 12 Р. 1986–2009.

14. *Репин С. И.* Апостериорные тождества для мер отклонений от точных решений нелинейных краевых задач // Ж. вычисл. матем. и матем. физ. 2023. Т. 63. № 6. Р. 896–919.

15. *Репин С. И.* Контроль точности приближенных решений одного класса сингулярно возмущенных краевых задач // Ж. вычисл. матем. и матем. физ. 2022. Т. 62. № 11. Р. 1822–1839.

16. *Repin S.* Error identities for parabolic initial boundary value problems // Zap. Nauchn. Sem. POMI. 2021. V. 508. P. 147–172.

IDENTITIES FOR MEASURES OF DEVIATIONS FROM SOLUTIONS OF PARABOLO-HYPERBOLIC EQUATIONS

S.I. Repin[a,b,*]

[a] *St. Petersburg Branch of the V.A. Steklov Mathematical Institute of the Russian Academy of Sciences, St. Petersburg, Fontanka 27, 191023, Russia*
[b] *St. Petersburg Polytechnic University, Polytechnicheskaya 29, St. Petersburg, 195251, Russia*
**e-mail: repin@pdmi.ras.ru*
Received 19 November, 2023
Revised 19 November, 2023
Accepted 14 January, 2024

Abstract. The article presents integral identities that hold for the difference between the exact solution of the initial-boundary value problem for a parabolo-hyperbolic equation and any function from the corresponding energy class. These identities allow for the derivation of two-sided a posteriori estimates for approximate solutions to the corresponding Cauchy problem. The left side of the estimate provides a natural measure of deviation from the solution, while the right side depends only on the problem data and the approximate solution itself, making it computable. The obtained estimates are utilized to compare solutions of Cauchy problems for both the parabolic equation and the parabolo-hyperbolic equation with a small parameter in the second time derivative. Additionally, the estimates enable a quantitative assessment of the effects arising from inaccuracies in initial data and coefficients of the equation.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 835–841

МАТЕМАТИЧЕСКАЯ ФИЗИКА

УДК 519.6

ОБ УСТОЙЧИВОСТИ СХЕМЫ СТАБИЛИЗИРУЮЩЕЙ ПОПРАВКИ С ЦЕНТРАЛЬНЫМИ РАЗНОСТЯМИ ПО ПРОСТРАНСТВЕННЫМ ПЕРЕМЕННЫМ ДЛЯ 3-МЕРНОГО УРАВНЕНИЯ ПЕРЕНОСА

© 2024 г. В.П. Жуков[1,*]

[1]630090 Новосибирск, пр-т Акад.Лаврентьева, 6, Федеральный исследовательский центр информационных и вычислительных технологий, Россия

**e-mail: zukov@ict.nsc.ru*

Поступила в редакцию 05.12.2023 г.

Переработанный вариант 25.12.2023 г.

Принята к публикации 14.01.2024 г.

Принято считать, что схема стабилизирующей поправки с центральными разностями по пространственным переменным для уравнения переноса в 3-мерном случае является условно устойчивой. В настоящей работе показано, что, строго говоря, эта схема абсолютно неустойчива. Однако область неустойчивых гармоник в пространстве волновых векторов и величина их инкрементов быстро стремятся к нулю при стремлении параметра Куранта к нулю, что позволяет успешно использовать эту схему. Поэтому правильнее говорить о практически условной устойчивости данной схемы. Библ. 3. Фиг. 4.

Ключевые слова: устойчивость конечно-разностных схем в многомерном случае, метод дробных шагов, схема стабилизирующей поправки, гиперболические уравнения.

DOI: 10.31857/S0044466924050118, **EDN**: YCZKUW

1. ВВЕДЕНИЕ

Оператор, соответствующий уравнению переноса $\frac{\partial}{\partial t} + V_1\frac{\partial}{\partial x_1} + V_2\frac{\partial}{\partial x_2} + V_3\frac{\partial}{\partial x_3}$, входит во многие уравнения, описывающие сплошные среды (гидро-, газо-, гео-, плазмодинамика). Поэтому выбор конечно-разностных схем для успешного решения уравнения переноса

$$\frac{\partial u}{\partial t} + V_1\frac{\partial u}{\partial x_1} + V_2\frac{\partial u}{\partial x_2} + V_3\frac{\partial u}{\partial x_3} = 0 \tag{1}$$

и знание свойств этих схем очень важен. Уравнения, в которые входит оператор переноса, могут быть очень сложными и нелинейными, и связаны с другими уравнениями. Поэтому на практике часто ограничиваются первым порядком аппроксимации по времени $\partial u/\partial t \approx (u^{n+1} - u^n)/\tau$ (обозначения стандартные). Однако по пространственным переменным первого порядка обычно недостаточно. Так использование простейшей условно-устойчивой явной схемы с учетом знака скорости первого порядка по пространственной переменной зачастую дает плохие результаты, так как эта схема обладает большой схемной вязкостью и конечно-разностное решение подвержено сильной нефизической диффузии. Явная схема с аппроксимацией пространственных производных центральной разностью с порядком аппроксимации $O(\tau, h^2)$

$$\frac{u^{n+1}_{i,j,k} - u^n_{i,j,k}}{\tau} + V_1\frac{u^n_{i+1,j,k} - u^n_{i-1,j,k}}{2h_1} + V_2\frac{u^n_{i,j+1,k} - u^n_{i,j-1,k}}{2h_2} + V_3\frac{u^n_{i,j,k+1} - u^n_{i,j,k-1}}{2h_3} = 0 \tag{2}$$

абсолютно неустойчива. Использование абсолютно устойчивой полностью неявной схемы наталкивается на известные сложности реализации, которые преодолеваются с помощью схем с дробными шагами [3]. При этом, если в двумерном случае абсолютной устойчивостью обладают различные варианты метода дробных шагов, то в трехмерном случае проблемы сохраняются [3]. Потеря безусловной устойчивости схем в трехмерном случае

имеет место для всех схем, приводимых к канонической форме: схемы приближенной факторизации, стабилизирующей поправки и предиктор-корректор [1, 2]. Более того, например, трехмерный аналог схемы продольно-поперечной прогонки для (1) оказывается абсолютно неустойчивым. В классических работах [1, 2] утверждается, что класс схем приближенной факторизации

$$(I+\alpha\tau V_3\Lambda_3)(I+\alpha\tau V_2\Lambda_2)(I+\alpha\tau V_1\Lambda_1)\frac{u^{n+1}-u^n}{\tau}=-(V_1\Lambda_1+V_2\Lambda_2+V_3\Lambda_3)u^n, \tag{3}$$
$$\Lambda_1 u^n=(u^n_{i+1,j,k}-u^n_{i-1,j,k})/(2h_1)\approx\partial u/\partial x_1,\quad \Lambda_2 u^n=(u^n_{i,j+1,k}-u^n_{i,j-1,k})/(2h_2)\approx\partial u/\partial x_2,$$
$$\Lambda_3 u^n=(u^n_{i,j,k+1}-u^n_{i,j,k-1})/(2h_3)\approx\partial u/\partial x_3.$$

при параметре $\alpha>1/2$ обладает условной устойчивостью. Ниже будет показано, что это неправильно и схема (3) абсолютно неустойчива. В (2), (3) индексы i,j,k и пространственные шаги h_1,h_2,h_3 относятся к переменным x_1,x_2,x_3 соответственно.

Для выяснения условий устойчивости схемы (3) ищем ее решение в виде ($\bar{i}$ — мнимая единица, k_m — волновые числа)

$$u^n_{i,j,k}=u_0\lambda^n\exp\bigl(\bar{i}(k_1h_1i+k_2h_2j+k_3h_3k)\bigr). \tag{4}$$

Подставляя (4) в (3), в согласии с [1], имеем

$$(1+\bar{i}\alpha d_1)(1+\bar{i}\alpha d_2)(1+\bar{i}\alpha d_3)(\lambda-1)+\bar{i}d=0,$$
$$d_m=\tau h_m^{-1}V_m\sin(k_mh_m),\quad m=1,2,3,\quad d=d_1+d_2+d_3.$$

Это дает

$$\lambda=\frac{A+\bar{i}(B-d)}{A+\bar{i}B}, \tag{5}$$
$$A=1-\alpha^2(d_1d_2+d_2d_3+d_1d_3),\quad B=\alpha(d-\alpha^2d_1d_2d_3).$$

Условие устойчивости $|\lambda|\le1$ дает $(B-d)^2\le B^2$. Это неравенство можно переписать в нескольких эквивалентных формах

$$\bigl((1-\alpha)d+\alpha^3d_1d_2d_3\bigr)^2\le(\alpha d-\alpha^3d_1d_2d_3)^2, \tag{6}$$
$$\bigl(\alpha^3d_1d_2d_3+(1-\alpha)d\bigr)^2\le(\alpha^3d_1d_2d_3-\alpha d)^2, \tag{7}$$
$$2\alpha^3d_1d_2d_3d\le(2\alpha-1)d^2. \tag{8}$$

В [1, 2] рассмотрен только случай $d_1d_2d_3\ge0$ и отмечены следующие свойства схемы, вытекающие из (6)–(8).

1. Если хотя бы одно из d_m равно нулю, то эти условия выполняются. В частности, в 2-мерном случае схема абсолютно устойчива.

2. Из (6) видно, что для устойчивости при малых d_m необходимо, чтобы α было больше $1/2$.

3. При положительных d_1, d_2, d_3 и $d_3\gg d_1,d_2$ условие (8) дает

$$d_1d_2\le(2\alpha-1)/(2\alpha^3). \tag{9}$$

   То есть в этом случае устойчивость определяется меньшими d_m. Условие (9) верно и для разных знаков при $|d_3|\gg|d_1|,|d_2|$. Это свойство важно, например, при получении решения в окрестности оси при использовании цилиндрической и сферической систем координат.

4. При $d_1=d_2=d_3=d_0$ формулы (6)–(8) дают

$$d_0^2\le\frac{3(2\alpha-1)}{2\alpha^3}. \tag{10}$$

   При $\alpha>1/2$ условие (10) может быть удовлетворено. Причем максимальное $d_0=4/3$ достигается при $\alpha=3/4$. При $\alpha=1$ имеем $d_0\le\sqrt{3/2}\approx1.225$.

На основании пп. 1–4, изложенных выше, в [1, 2] предполагается условная устойчивость (3) в 3-мерном случае при $\alpha > 1/2$. Но в [1, 2] есть принципиальная ошибка. Она состоит в том, что величины d_1, d_2, d_3 могут иметь разные знаки. Поэтому их сумма может быть малой, а произведение — значительным. Если знаки суммы и произведения разные, то правая часть (7) будет заведомо меньше левой и условие устойчивости выполняться не будет. Покажем это.

Без ограничения общности можно положить, что $d_1, d_2 > 0$, $d_3 < 0$. В случае $|d_3| \le d_1 + d_2$ величина $d \ge 0$ и условие устойчивости (7) выполняется. При $|d_3| > d_1 + d_2$ удобно ввести обозначение $d_3 = -(d_1 + d_2)(1 + \varepsilon)$, $\varepsilon > 0$. В этом случае $d = -(d_1 + d_2)\varepsilon < 0$ и (7) дает $2\alpha^3 d_1 d_2 (d_1 + d_2)^2 \varepsilon(1 + \varepsilon) \le (2\alpha - 1)(d_1 + d_2)^2\varepsilon^2$ или

$$\big((2\alpha - 1) - 2\alpha^3 d_1 d_2\big)\varepsilon \ge 2\alpha^3 d_1 d_2.$$

Это условие может быть выполнено только если выполняется условие (9), т.е. d_1, d_2 достаточно малы. При этом для устойчивости необходимо, чтобы ε было достаточно велико. В диапазоне

$$\varepsilon \in \left(0 : \frac{2\alpha^3 d_1 d_2}{(2\alpha - 1) - 2\alpha^3 d_1 d_2}\right)$$

условие устойчивости не выполняется. При $d_1, d_2 \to 0$ ширина диапазона неустойчивости по ε уменьшается пропорционально $d_1 d_2$, а соответственно, по d_3 — как $d_1 d_2 (d_1 + d_2)$. Таким образом, неустойчивыми являются возмущения с волновыми векторами близкими к вектору $\big(d_1, d_2, -(d_1 + d_2)\big)$, который направлен поперек вектора скорости. Для превышения квадрата модуля множителя перехода (5) над единицей имеем

$$|\lambda|^2 - 1 = \frac{d(d - 2B)}{A^2 + B^2} = \frac{(d_1 + d_2)\varepsilon\Big((d_1 + d_2)\varepsilon - 2\Big(\alpha\big((d_1 + d_2)\varepsilon + \alpha^2 d_1 d_2 d_3\big)\Big)\Big)}{A^2 + B^2} =$$
$$= \frac{(d_1 + d_2)^2\varepsilon\big(2\alpha^3 d_1 d_2 - (2\alpha - 1 - 2\alpha^3 d_1 d_2)\varepsilon\big)}{A^2 + B^2}.$$

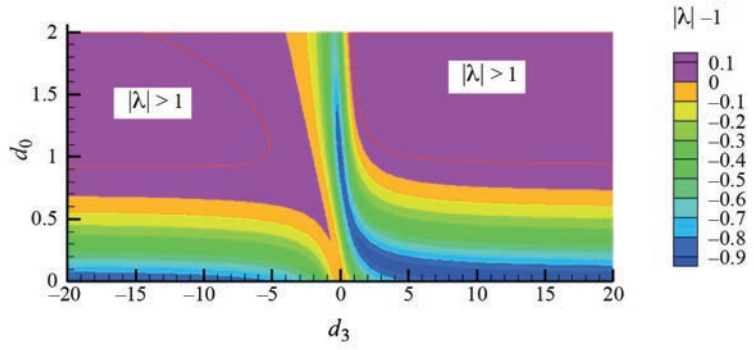
Так как $A^2 + B^2 \approx 1$ при $d_1, d_2 \to 0$, то в этом пределе максимум $|\lambda|^2 - 1$ достигается в точке $\varepsilon = \frac{\alpha^3 d_1 d_2}{(2\alpha - 1) - 2\alpha^3 d_1 d_2}$ и равен

$$\big(|\lambda|^2 - 1\big)_{\max} = \frac{\big(\alpha^3 d_1 d_2 (d_1 + d_2)\big)^2}{2\alpha - 1}.$$

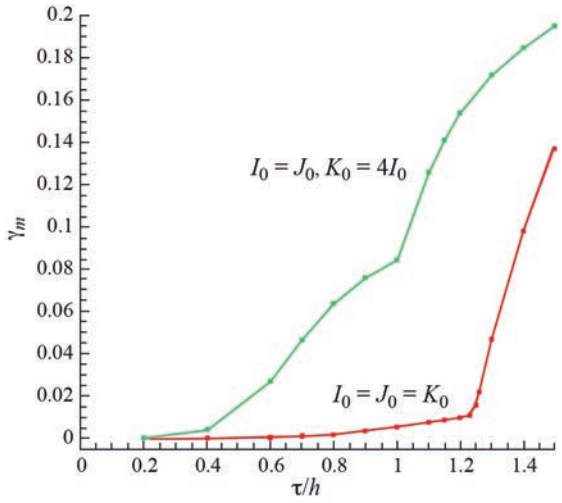
Заметим, что при уменьшении α в диапазоне $\alpha \in (1/2 : 1)$, в отличие от случая положительных d_1, d_2, d_3, инкремент неустойчивости монотонно возрастает при уменьшении α. Таким образом, область неустойчивости существует всегда (абсолютная неустойчивость), но при уменьшении d_1, d_2 ее ширина по d_3 уменьшается как куб, а превышение инкремента над 1 — как шестая степень d_1, d_2. Сказанное иллюстрирует фиг. 1, где показано распределение превышения модуля множителя перехода над единицей $|\lambda| - 1$ при $d_1 = d_2 = d_0$ в зависимости от d_0 и d_3.

Для еще одной иллюстрации рассмотрим решения разностного аналога (3) уравнения (1) при $\alpha = 1$ и $V_1 = V_2 = V_3 = 1$ в единичном кубе с периодическими граничными условиями на равномерной сетке в виде плоской волны

$$u^n_{i,j,k} = \lambda^n_{N,M,L} \exp\big(\bar{i}(2\pi Ni/I_0 + 2\pi Mj/J_0 + 2\pi Lk/K_0)\big).$$



Фиг. 1. $|\lambda| - 1$ как функция d_3 и d_0 ($d_1 = d_2 = d_0$) при $\alpha = 1$. Область неустойчивости при $d_3 < 0$ простирается до нуля.



Фиг. 2. Зависимость инкремента γ_m от τ/h в случае расчетной области в виде единичного куба и равномерной сетки с одинаковыми шагами по всем направлениям (красная кривая) и в случае когда по одному из направлений число узлов в 4 раза больше.

Величины d_m в рассматриваемом случае имеют вид

$$d_1 = (\tau/h_1)\sin(2\pi N/I_0), \quad d_2 = (\tau/h_2)\sin(2\pi M/J_0), \quad d_3 = (\tau/h_3)\sin(2\pi L/K_0).$$

Изучим зависимость максимального инкремента $\gamma_m = \max\limits_{N,\,M,\,L} |\lambda_{N,\,M,\,L}| - 1$ от шагов сетки (фиг. 2). Величина γ_m определялась численно перебором всех значений N, M и L. Рассмотрим случай $I_0 = J_0 = K_0$, $h_1 = h_2 = h_3 = h = 1/I_0$. При $\tau/h \geq 1.2$ инкремент γ_m уменьшается с уменьшением τ/h в соответствии с формулами работ [1, 2]. При этом γ_m соответствует гармоникам с $d_1 = d_2 = d_3$. Но при $\tau/h < 1.2$, в отличие от [1, 2], γ_m не равен нулю, а изменяется в соответствии с развитыми выше представлениями. Он больше нуля, но резко падает при уменьшении τ/h. При этом максимальный инкремент соответствует гармоникам с $d_3 \approx -(d_1 + d_2)$.

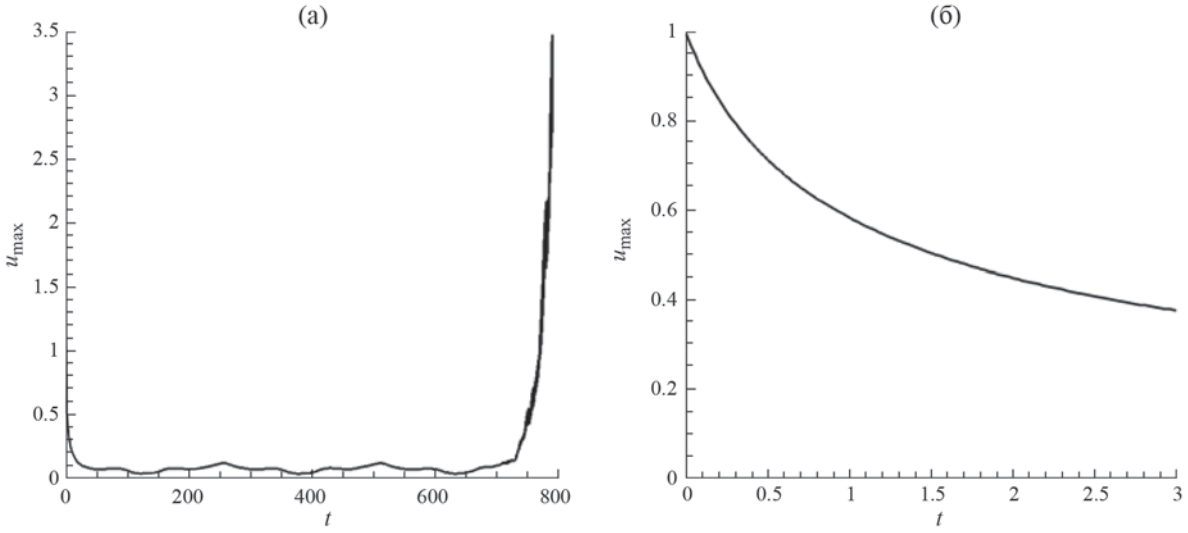
Если в одном из направлений величины $\tau V/h$ намного превышают эти величины в других направлениях, то γ_m больше, чем в рассмотренном случае. На фиг. 2 зеленая кривая соответствует $I_0 = J_0$, $K_0 = 4I_0$, $h = 1/I_0$. Заметим, что дальнейшее увеличение отношения K_0/I_0 не оказывает существенного влияния на γ_m.

2. ЗАДАЧА С ПЕРИОДИЧЕСКИМИ ГРАНИЧНЫМИ УСЛОВИЯМИ

Рассмотрим решение уравнения (1) с помощью схемы (3) с $\alpha = 1$ в единичном кубе при периодических граничных условиях и $V_1 = V_2 = V_3 = 1$. Начальное условие соответствовало локализованному вблизи центра расчетной области возмущению с гауссовым распределением $u = \exp\Big(-\big((x_1-0.5)^2+(x_2-0.5)^2+(x_3-0.5)^2\big)/R^2\Big)$, единичной амплитудой и $R = 1/8$. Если отношение $\tau/h < 1$ (условие устойчивости [1, 2] выполнено), то наблюдается ожидаемая картина: возмущение переносится вдоль главной диагонали куба с постоянной скоростью, исчезая на одной границе и появляясь на другой. В отличие от точного решения амплитуда возмущения $u_{\max} = \max\limits_{i,j,k} u_{i,j,k}$ уменьшается со временем (фиг. 3). Это ожидаемо, так как большинство гармоник имеет $|\lambda| < 1$. Со временем также происходит изменение формы решения. На временах, когда решение конечно-разностной задачи совсем не соответствует решению дифференциальной задачи $u_{\max}$ испытывает колебания. Затем происходит бурный рост наиболее неустойчивых гармоник и $u_{\max}$ устремляется к бесконечности.

Характерное время, в течении которого конечно-разностное решение еще примерно соответствует решению дифференциальной задачи (в качестве этого времени может выступать, например, момент времени, когда амплитуда возмущения уменьшается на 20% от начальной) увеличивается при уменьшении шагов сетки в соответствии с порядком аппроксимации схемы $o(\tau, h^2)$. Если по одному из направлений взять намного более мелкую сетку, например, $K_0 = 4I_0$, то эти результаты изменяться слабо, что ожидаемо.

Время яркого проявления неустойчивости в сотни раз превышает время, в течении которого конечно-разностное решение еще приблизительно соответствует точному даже при $\tau/h \approx 1$. Поскольку максимальное $|\lambda|$ зависит от τ/h, а необходимое для достижения определенного времени число шагов по времени пропорционально τ^{-1}, то время развития неустойчивости уменьшается с уменьшением шагов по пространству при постоянном τ/h. При уменьшении τ/h время развития неустойчивости резко увеличивается. Например, в рассмат-



Фиг. 3. Зависимость $u_{\max}$ от t, б — более подробно при малых временах. Вариант $I_0 = J_0 = K_0 = 100$, $\tau/h = 0.5$.

риваемой задаче при изменении τ/h от 1 до 0.5 оно увеличивается более чем в 10 раз. При различающихся в различных направлениях шагах сетки, например при $J_0 = I_0$, $K_0 = 4I_0$ ($h = 1/I_0$), время развития неустойчивости уменьшается в несколько раз, но по-прежнему намного превышает время соответствия конечно-разностного решения точному решению. Заметим, что время проявления неустойчивости зависит не только от инкрементов гармоник, но и от амплитуд этих гармоник, присутствующих в начальный момент времени.

Из представленных результатов можно сделать вывод, что для рассмотренной задачи можно считать схему (3) практически условно устойчивой, но с условием более жестким, чем предсказывает [1, 2]. Рассматриваемая неустойчивость присутствует и при малых τ/h, но она не успевает развиться: намного раньше решение теряет смысл из-за потери точности.

3. ЗАДАЧА НА УСТАНОВЛЕНИЕ

Рассмотрим задачу, отличающуюся от предыдущей задачи граничными условиями по x_1, которые имели вид

$$i = 1: \quad u = \exp\Big(-\big((x_2 - 0.5)^2 + (x_3 - 0.5)^2\big)/R^2\Big), \quad i = I_0: \quad \frac{u^{n+1}_{I_0,j,k} - u^n_{I_0,j,k}}{\tau} + V_1 \frac{u^{n+1}_{I_0,j,k} - u^{n+1}_{I_0-1,j,k}}{h_1} = 0,$$

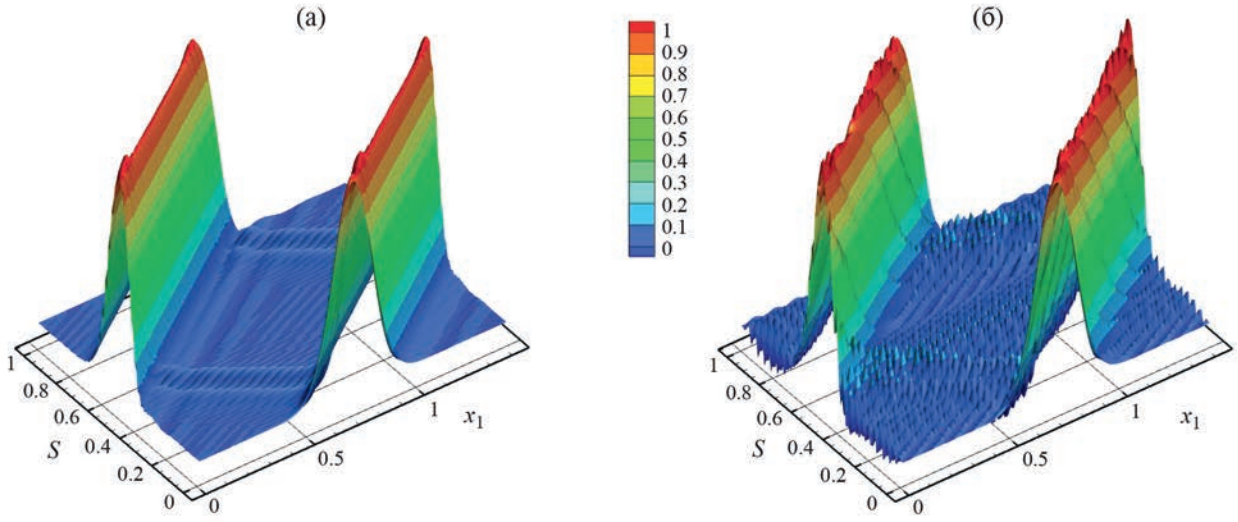
и нулевыми начальными условиями. Для более правильной передачи граничных условий дробный шаг с поправкой вдоль координаты x_1 при реализации (3) производился в последнюю очередь.

Расчеты говорят о том, что, например, при $I_0 = J_0 = K_0 = 100$ и 200 и $\tau/h = 1$ решение неустойчиво. Причина этой неустойчивости, которая возникает у границы $i = I_0$, связана с несовершенным граничным условием на этой границе, целью которого является беспрепятственный вынос возмущения из расчетной области.

При $\tau/h = 0.5$ на временах больше 1 устанавливается практически стационарное ожидаемое распределение u (фиг. 4). Наблюдаемая небольшая рябь связана с несовершенством граничных условий при $i = I_0$. Она появляется по достижении возмущением этой границы и уменьшается при уменьшении шагов сетки. Решение, формирующееся на временах порядка 1, долгое время сохраняет свою форму без изменений. Но ко времени $t \approx 456$ для $I_0 = 100$ и $t \approx 270$ для $I_0 = 200$ постепенно возникает описанная выше неустойчивость, приводящая к развалу решения, что соответствует развитым выше представлением.

Подчеркнем, что в реальных задачах расчет до таких больших времен бессмыслен: стационар достигается намного раньше.

Заметим следующее: 1) введение вязкого члена $\chi\Delta u$ в правую часть (1) при малом $\chi = 10^{-4}$ сильно уменьшает мелкомасштабную рябь и не дает развиться неустойчивостям, не влияя на главные гармоники; 2) применение явной схемы первого порядка с учетом знака скорости при тех же шагах сетки и при выполнении условия устойчивости явной схемы ($\tau/h < 0.3$) приводит к сильному нефизическому затуханию решения; 3) замена в (3) даже одного оператора Λ_m на оператор направленной с учетом знака скорости разности первого порядка сильно повышает устойчивость (3), но приводит к сильной, только что упомянутой, потере точности решения; 4) немонотонность конечно-разностного решения в рассмотренных задачах (при $\alpha = 1$) не является сильной. Напомним, что рябь на фиг. 4 связана с граничным условием.



Фиг. 4. Распределение u в сечении $x_2 = x_3$ в стационаре ($t < 400$ — (а)) и в начале яркого проявления неустойчивости ($t = 500$ — (б)). Величина s соответствует расстоянию по x_2, x_3: $x_2 = s/\sqrt{2}$, $x_3 = s/\sqrt{2}$. Вариант $I_0 = J_0 = K_0 = 100$, $\tau/h = 0.5$.

4. ЗАКЛЮЧЕНИЕ

Таким образом, схема (3) строго говоря абсолютно неустойчива, но область неустойчивых гармоник в пространстве d_1, d_2, d_3 узкая и сужается при $d_1, d_2, d_3 \to 0$ (при уменьшении τ/h) как $(\tau/h)^3$, а инкремент уменьшается пропорционально $\big(d_1 d_2 (d_1 + d_2)\big)^2 \sim (\tau/h)^6$. Поэтому неустойчивость зачастую не успевает развиваться за время расчета. Существуют другие обстоятельства, которые при решении более сложных задач подавляют обсуждаемую неустойчивость.

1. Наличие даже небольшой физической вязкости, диффузии и т.п.
2. Схемы для решения более сложных задач могут иметь запас устойчивости, связанный с аппроксимацией других членов, компенсирующий обсуждаемую неустойчивость.
3. Во многих случаях скорость не является постоянной величиной во времени и пространстве.
4. Наличие градиентов часто приводит к возникновению скорости вдоль этих градиентов и прекращению неустойчивости, так как в обсуждаемой неустойчивости возбуждаются возмущения с волновыми векторами, направленными поперек вектора скорости.
5. Для развития неустойчивости все 3 компоненты скорости должны быть отличны от нуля. При решении физических задач как правило координаты выбирают так, чтобы скорость была направлена преимущественно вдоль одного из направлений.

В итоге можно заключить, что схему стабилизирующей поправки можно считать “практически условно устойчивой” с условием устойчивости вида $V\tau/h < C$ или (в случае сильно отличающихся скоростей) (9), но величина C может быть ощутимо меньше, чем предсказывает [1, 2]. Также необходимо соблюдать осторожность при расчетах на установление. При возникновении проблем рекомендуется уменьшить шаг по времени или ввести очень малую вязкость.

Автор выражает благодарность В. М. Ковеня за полезные обсуждения.

СПИСОК ЛИТЕРАТУРЫ

1. *Ковеня В. М., Тарнавский Г. А., Яненко Н. Н.,* Неявная разностная схема для численного решения пространственных уравнений газовой динамики // Ж. вычисл. матем. и матем. физ. 1980. Т. 20. № 6. С. 1465–1482.
2. *Ковеня В. М., Яненко Н. Н.* Метод расщепления в задачах газовой динамики. Новосибирск: Наука, 1981.
3. *Яненко Н. Н.* Метод дробных шагов решения многомерных задач математической физики. Новосибирск: Наука, Сибирское отделение, 1967.

ON THE STABILITY OF A STABILIZING CORRECTION SCHEME WITH CENTRAL DIFFERENCES FOR SPATIAL VARIABLES IN THE 3D TRANSPORT EQUATION

V. P. Zhukov*

Federal Research Center for Information and Computational Technologies, Acad. Lavrentiev Ave., 6, Novosibirsk, 630090, Russia

**e-mail: zukov@ict.nsc.ru*

Received 05 December, 2023
Revised 25 December, 2023
Accepted 14 January, 2024

Abstract. It is generally accepted that the stabilizing correction scheme with central differences for spatial variables in the 3D transport equation is conditionally stable. The work shows that, strictly speaking, this scheme is absolutely unstable. However, the region of unstable harmonics in the wave vector space and the magnitude of their increments rapidly approach zero as the Courant parameter tends to zero, allowing successful use of this scheme. Therefore, it is more accurate to refer to this scheme as practically conditionally stable.

Keywords: stability of finite-difference schemes in multidimensional cases, fractional step method, stabilizing correction scheme, hyperbolic equations.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 842–851

УДК 519.634

ИССЛЕДОВАНИЕ И ОПТИМИЗАЦИЯ N-ЧАСТИЧНОГО ЧИСЛЕННОГО СТАТИСТИЧЕСКОГО АЛГОРИТМА РЕШЕНИЯ УРАВНЕНИЯ БОЛЬЦМАНА[1)]

© 2024 г. Г. З. Лотова[1,2,*], Г. А. Михайлов[1,2], С. В. Рогазинский[1,2]

[1] *630090 Новосибирск, пр-т Акад. Лаврентьева, 6, Институт вычислительной математики и математической геофизики СО РАН, Россия*

[2] *630090 Новосибирск, ул. Пирогова, 2, Новосибирский государственный университет, Россия*

**e-mail: lot@osmf.sscc.ru*

Поступила в редакцию 27.11.2023 г.

Переработанный вариант 27.11.2023 г.

Принята к публикации 14.01.2024 г.

Основной целью работы является проверка гипотезы о том, что известный N-частичный статистический алгоритм дает оценку решения нелинейного уравнения Больцмана с погрешностью порядка $O(1/N)$. Для этого определяются практически важные оптимальные соотношения между значением N и числом n выборочных значений оценки. Численные результаты для задачи с известным решением подтверждают удовлетворительность сформулированных оценок и выводов. Библ.14. Табл.3.

Ключевые слова: метод Монте-Карло, статистическое моделирование, уравнение Больцмана, N-частичная цепь Маркова, молекулярный хаос, метод мажорантной частоты.

DOI: 10.31857/S0044466924050121, **EDN**: YCXUYI

ВВЕДЕНИЕ

Для построения и обоснования метода прямого статистического моделирования с целью нахождения приближенного решения однородного по пространству нелинейного кинетического уравнения Больцмана может быть использовано линейное интегральное уравнение, которое эквивалентно N-частичному уравнению Каца [1]. Однако использовать это уравнение непосредственно для построения стандартных весовых модификаций моделирования [2] невозможно, так как его ядро представляет собой сумму взаимно сингулярных слагаемых [3]. Для определенности далее будет рассматриваться задача об однородной релаксации простого однокомпонентного газа, однако все построения весовой схемы носят общий характер и без труда переносятся на более общие случаи. Итак, рассматривается физический процесс однородной релаксации простого однокомпонентного газа, который описывается следующим нелинейным кинетическим уравнением Больцмана:

$$\frac{\partial}{\partial t} f(\mathbf{v},t) = \int [f(\mathbf{v}',t) f(\mathbf{v}_1',t) - f(\mathbf{v},t) f(\mathbf{v}_1,t)] \times w(\mathbf{v}', \mathbf{v}_1' \,|\mathbf{v}, \mathbf{v}_1) d\mathbf{v}' d\mathbf{v}_1' d\mathbf{v}_1. \tag{0.1}$$

Здесь $f(\mathbf{v},t)$ – "одночастичная" функция плотности распределения по скорости $\mathbf{v}$ в момент времени t. В данном случае уравнение Больцмана записано с использованием условной плотности вероятности w перехода пары скоростей частиц от $(\mathbf{v}', \mathbf{v}_1')$ к $(\mathbf{v}, \mathbf{v}_1)$. Удобство этой формы заключается в том, что она позволяет произвести все построения в максимально общей форме и легко перейти к рассмотрению конкретного взаимодействия пары частиц.

Плотность $w(\mathbf{v}', \mathbf{v}_1'|\mathbf{v}, \mathbf{v}_1)$ и дифференциальное сечение рассеяния частиц связаны следующим соотношением (см., например, [4]):

$$w(\mathbf{v}', \mathbf{v}'_1 \,|\mathbf{v}, \mathbf{v}_1) = \sigma(h,\Omega) \delta_1 \left(\frac{(\mathbf{v}-\mathbf{v}_1)^2 - (\mathbf{v}'-\mathbf{v}'_1)^2}{2} \right) \times \delta_3 \left(\frac{\mathbf{v}+\mathbf{v}_1-\mathbf{v}'-\mathbf{v}'_1}{2} \right),$$

которое следует из того, что скорости $(\mathbf{v}', \mathbf{v}'_1)$ и $(\mathbf{v}, \mathbf{v}_1)$ удовлетворяют законам сохранения импульса и энергии при столкновении:

$$\mathbf{v} + \mathbf{v}_1 = \mathbf{v}' + \mathbf{v}'_1, \quad \mathbf{v}^2 + \mathbf{v}_1^2 = \mathbf{v}'^2 + \mathbf{v}'^2_1.$$

[1)] Работа выполнена при финансовой поддержке государственного задания ИВМиМГ СО РАН (проект 0251-2022-0002).

Здесь $h = |\mathbf{v} - \mathbf{v}_1|$ – модуль относительной скорости сталкивающихся частиц, Ω – телесный угол поворота относительной скорости при столкновении, δ_1 и δ_3 – одно- и трехмерная дельта-функции, соответственно:

$$\int \delta_1(h)dh = 1, \quad \int \delta_3(\mathbf{v} + \mathbf{v}_1)d(\mathbf{v} + \mathbf{v}_1) = 1.$$

Функция $f(\mathbf{v}, t)$ удовлетворяет условию нормировки

$$\int f(\mathbf{v}, t)d\mathbf{v} = 1, \quad t \geqslant 0.$$

Присоединяя к (0.1) начальные данные

$$f(\mathbf{v}, t)|_{t=0} = f_0(\mathbf{v}), \quad t \in (0, T\,], \quad \mathbf{v} \in R^3, \tag{0.2}$$

получим задачу Коши для нелинейного уравнения Больцмана.

Численное решение задачи Коши (0.1), (0.2) мы будем понимать в смысле оценки линейных функционалов от функции $f(\mathbf{v}, t)$ (в среднеквадратичной метрике).

1. N-ЧАСТИЧНЫЙ АЛГОРИТМ, ОЦЕНКА ПАРАМЕТРА ПОГРЕШНОСТИ

Перепишем уравнение (0.1) в виде

$$\begin{aligned} \frac{\partial}{\partial t}f(\mathbf{v}, t) \quad &= \quad -\int f(\mathbf{v}, t)f(\mathbf{v}_1, t)w(\mathbf{v}', \mathbf{v}'_1|\mathbf{v}, \mathbf{v}_1)d\mathbf{v}'d\mathbf{v}'_1d\mathbf{v}_1 + \\ &\quad + \int f(\mathbf{v}', t)f(\mathbf{v}'_1, t)w(\mathbf{v}', \mathbf{v}'_1|\mathbf{v}, \mathbf{v}_1)d\mathbf{v}'d\mathbf{v}'_1d\mathbf{v}_1. \end{aligned} \tag{1.1}$$

Полученное таким образом уравнение можно рассматривать как итерационное соотношение баланса при условии, что плотность $f(\mathbf{v}_1, t)$ известна — это соответствует предположению о “переносе хаоса” (chaos propagation) при выводе уравнения (0.1) (см., например, [1]).

Соотношение (1.1) формально можно записать в виде линейного кинетического уравнения:

$$\frac{\partial}{\partial t}f(\mathbf{v}, t) = -\nu(\mathbf{v})f(\mathbf{v}, t) + \int f(\mathbf{v}', t)\left[f(\mathbf{v}'_1, t)w(\mathbf{v}', \mathbf{v}'_1|\mathbf{v}, \mathbf{v}_1)d\mathbf{v}'d\mathbf{v}_1\right]d\mathbf{v}', \tag{1.2}$$

где $\nu(\mathbf{v}) = \int f(\mathbf{v}_1, t)w(\mathbf{v}', \mathbf{v}'_1|\mathbf{v}, \mathbf{v}_1)d\mathbf{v}'d\mathbf{v}'_1d\mathbf{v}_1$.

По аналогии с линейным кинетическим уравнением соотношение (1.2) формально задает “нелинейный” марковский процесс [5], в котором изменения системы за время dt определяется ее состоянием в момент времени t. На этой основе строятся “пошаговые” дискретные вычислительные схемы, которые принято называть методом Берда [6, 7]. Соотношение (1.2) фактически также позволило построить рассматриваемую математическую N-частичную модель.

В связи с этим для физического процесса однородной релаксации простого однокомпонентного газа хорошо известна математическая модель [5], в основу которой положено представление о газе как об ансамбле конечного числа взаимодействующих частиц. При выполнении определенных требований, накладываемых на характеристики этого ансамбля и на стохастический процесс его эволюции во времени можно исследовать вопрос о степени аппроксимации данной математической моделью рассматриваемого физического процесса. Однородная релаксация простого газа описывается нелинейным кинетическим уравнением Больцмана (0.1) и его решением является одночастичная функция распределения по скоростям. Для его приближенного решения используется линейное интегродифференциальное уравнение, описывающее эволюцию ансамбля N частиц во времени, так называемое “основное” кинетическое уравнение [4].

Используя условную плотность вероятности $w(\mathbf{v}'_i, \mathbf{v}'_j\,|\mathbf{v}_i, \mathbf{v}_j\,) = w(\mathbf{v}_i, \mathbf{v}_j\,\big|\mathbf{v}'_i, \mathbf{v}'_j\,)$ перехода пары скоростей частиц от $(\mathbf{v}'_i, \mathbf{v}'_j)$ к $(\mathbf{v}_i, \mathbf{v}_j)$, уравнение типа (0.1) для N-частичной модели можно записать в виде [4]:

$$\frac{\partial}{\partial t}f_N(t, V) = \frac{1}{N-1}\sum_{i<j}\int \left\{f_N(t, V'_{ij}) - f_N(t, V)\right\}w(\mathbf{v}'_i, \mathbf{v}'_j\,|\mathbf{v}_i, \mathbf{v}_j\,)d\mathbf{v}'_id\mathbf{v}'_j, \tag{1.3}$$

где $V = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N)$ есть $3N$-мерный вектор,

$$V'_{ij} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}'_i, ..., \mathbf{v}'_j, ..., \mathbf{v}_N),$$

$(\mathbf{v}'_i, \mathbf{v}'_j)$ и $(\mathbf{v}_i, \mathbf{v}_j)$ – скорости пары частиц до и после столкновения соответственно, удовлетворяющие законам сохранения момента и энергии, $\int f_N(t,V)dV = 1$. Плотность w и дифференциальное сечение $\sigma(h_{ij}, \chi_{ij})$ связаны следующим соотношением:

$$w(\mathbf{v}'_i, \mathbf{v}'_j|\mathbf{v}_i, \mathbf{v}_j) = \rho\sigma(h_{ij}, \chi_{ij})\, \delta_1\left(\frac{(\mathbf{v}_i - \mathbf{v}_j)^2 - (\mathbf{v}'_i - \mathbf{v}'_j)^2}{2}\right) \times \delta_3\left(\frac{\mathbf{v}_i + \mathbf{v}_j - \mathbf{v}'_i - \mathbf{v}'_j}{2}\right),$$

где ρ – плотность среды; $h_{ij} = |\mathbf{v}_i - \mathbf{v}_j|$, χ_{ij} – угол рассеяния.

Как указано в [1], модельный процесс стохастической кинетики системы из N частиц представляет собой однородную цепь Маркова, переходы в которой осуществляются в результате элементарных парных взаимодействий. Плотность распределения времени между элементарными взаимодействиями в системе определяется состоянием системы и является экспоненциальным. Участок N-частичной траектории, соответствующей прямолинейному движению всех частиц между двумя последовательными элементарными взаимодействиями, называется *свободным пробегом системы*. Вероятность элементарного взаимодействия в системе N частиц за время dt равна $\nu(V)dt$, где

$$\nu(V) = \frac{1}{N-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\int w(\mathbf{v}_i, \mathbf{v}_j|\mathbf{v}'_i, \mathbf{v}'_j)\, d\mathbf{v}'_i\, d\mathbf{v}'_j = \frac{1}{N-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} a(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{N-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \rho\, h_{ij}\, \sigma_t(h_{ij}),$$

а

$$\sigma_t(h) = \int \sigma(h, \Omega)\, d\Omega.$$

Вероятность того, что столкновение в системе N частиц реализует пара частиц с номерами i и j равна $a(\mathbf{v}_i, \mathbf{v}_j)/\nu(V)$, причем распределение новых скоростей частиц $(\mathbf{v}'_i, \mathbf{v}'_j)$ определяется дифференциальным сечением столкновения пары

$$k(\mathbf{v}_i, \mathbf{v}_j \to \mathbf{v}'_i, \mathbf{v}'_j) = w(\mathbf{v}'_i, \mathbf{v}'_j\,|\mathbf{v}_i, \mathbf{v}_j)\,[\rho\, h_{ij}\, \sigma_t(h_{ij})]^{-1}, \tag{1.4}$$

а скорости остальных частиц не изменяются. Время свободного пробега системы распределено с экспоненциальной плотностью $p(t) = \nu(V)\exp(-t\,\nu(V))$.

Этот алгоритм прямого моделирования хорошо известен. Он был эвристически сформулирован на основе марковского характера стохастического процесса эволюции N-частичной системы [1].

Относительно основного кинетического уравнения (1.3) известно (см. [1]), что оно асимптотически при $N \to \infty$ эквивалентно уравнению Больцмана в предположении молекулярного хаоса, которое фактически определяет форму интеграла столкновений в (0.1). Однако в [1] не приведена оценка скорости сходимости по N решения “основного” кинетического уравнения, т. е. плотности соответствующего одночастичного распределения к решению уравнения Больцмана. По-видимому, напрямую она не может быть получена из приведенного там доказательства. В связи с этим актуальна задача исследования зависимости решения N-частичного уравнения (1.3) от числа модельных частиц N, решению которой, как и работа [8], посвящена настоящая статья.

В [8] для простой модельной задачи были впервые получены результаты, позволяющие предположить, что погрешность N-частичного алгоритма имеет порядок $O(N^{-1})$(хотя эвристически можно было ожидать порядок $O(N^{-1/2})$). На этой основе рассмотрим какой-либо функционал G_N от решения $f_N(V,t)$ уравнения (1.3). Предполагая аналитическую зависимость G_N от $1/N$, представим его в следующем виде:

$$G_N = G_\infty + \frac{\gamma}{N} + O\left(\frac{1}{N^2}\right). \tag{1.5}$$

Отсюда получаем оценку точного значения G_∞ по двум уравнениям с $N = N_1, N_2$:

$$G_\infty \approx G_{N_2} + \frac{N_1}{N_2 - N_1}(G_{N_2} - G_{N_1}). \tag{1.6}$$

Значение разности $G_{N_2} - G_{N_1}$ может оказаться малым по сравнению со значениями G_{N_2}, G_{N_1}, поэтому для ее вычисления, при сравнительно небольших N_1, N_2, целесообразно применять метод коррелированной выборки (см. далее разд. 2).

Из (1.5) получаем также оценку величины γ:

$$\gamma = N_1 N_2 \frac{G_{N_2} - G_{N_1}}{N_1 - N_2}, \tag{1.7}$$

которая в [8] не вычислялась, так как проводились расчеты для сравнительно малых значений N_1, N_2.

2. N-ЧАСТИЧНЫЙ АЛГОРИТМ С ИСПОЛЬЗОВАНИЕМ КОРРЕЛИРОВАННОЙ ВЫБОРКИ И "МАЖОРАНТНОЙ ЧАСТОТЫ"

В работе [8] была использована так называемая "локальная оценка", которая дает вклады в некоторый искомый функционал $z(t)$ от всех столкновений. Исследования, проведенные в [9] для решения задач переноса "со временем", показали, что алгоритм прямого подсчета числа частиц в заданные моменты времени может обеспечивать не меньшую точность и тем самым быть менее трудоемким. Далее предварительно детально излагается такой алгоритм для псевдомаксвелловских молекул. Моделирование N-частичной цепи Маркова производится согласно описанию, данному в разд. 1.

1. Предполагается, что система из N частиц стартует в момент времени $t = 0$, причем начальные скорости частиц $\mathbf{v}_i^{(0)}$, $i = 1, 2, ..., N$, выбираются соответственно заданной плотности $f_0(\mathbf{v})$.

Для реализации используемой далее коррелированной выборки рассматриваются два варианта системы: с $N = N_1$ и $N = N_2$, $N_1 > N_2$, причем N_1 начальных значений скорости частиц получаются дополнением предварительно выбранных N_2 значений.

2. Поскольку для псевдомаксвелловских молекул $a(\mathbf{v}_i, \mathbf{v}_j) \equiv a_0$, то время τ свободного пробега соответствующей системы выбирается соответственно плотности

$$p_1(t) = (Na_0/2)\exp[-(Na_0/2)t].$$

В случае, если рассматриваются не псевдомаксвелловские молекулы, то используется метод мажорантной частоты [4], в котором разыгрывается альтернатива: с вероятностью $a(\mathbf{v}_i, \mathbf{v}_j)/M$ – столкновение реальное, где M удовлетворяет неравенству $a(\mathbf{v}_i, \mathbf{v}_j) \leqslant M$, т.е. происходит изменение скоростей; с дополнительной вероятностью – фиктивное столкновение, т.е. скорости частиц не изменяются.

3. Для реализации коррелированной выборки длины пробега в статье [8] предложен следующий алгоритм.

Временной шаг τ моделируется соответственно плотности $\sigma\exp(-\sigma\tau)$, причем $a_0N_2/2 \leqslant \sigma \leqslant a_0N_1/2$. После выбора τ веса Q_1, Q_2, соответствующие ансамблям объемов N_1, N_2 пересчитываются по формулам:

$$Q_1 = Q_1' \frac{a_0N_1}{2\sigma}\ \exp\bigl(-(a_0N_1/2 - \sigma)\tau\bigr), \quad Q_2 = Q_2' \frac{a_0N_2}{2\sigma}\ \exp\bigl(-(a_0N_2/2 - \sigma)\tau\bigr).$$

Здесь Q_1', Q_2' – веса на предыдущем шаге. Если на очередном свободном пробеге происходит пересечение заданной временной границы T, то веса домножаются на отношения вероятностей этого события, т.е. на величины

$$\exp\bigl(-(a_0N_1/2 - \sigma)(T - t')\bigr), \quad \exp\bigl(-(a_0N_2/2 - \sigma)(T - t')\bigr),$$

где t' – момент столкновения, непосредственно предшествующего указанному пересечению. С учетом равенства $\mathbf{E}Q_1 = \mathbf{E}Q_2 = 1$, очевидно, что целесообразно подобрать σ так, чтобы достигался следующий минимакс: $\min\limits_{\sigma} \max\bigl\{\mathbf{E}Q_1^2(T), \mathbf{E}Q_2^2(T)\bigr\}$. Простые вычисления дают равенства

$$\mathbf{E}Q_i^2(T) = \exp\bigl((a_0N_i/2 - \sigma)^2T/\sigma\bigr), \qquad i = 1, 2.$$

Ясно, что рассматриваемый минимакс достигается при $\mathbf{E}Q_1^2(T) = \mathbf{E}Q_2^2(T)$. Отсюда получаем требуемое минимаксное значение $\sigma = \sigma^* = a_0(N_1 + N_2)/4$.

В данном случае "вес" является глобальным [3] и определяется разностью сильно растущих при $T \to \infty$ величин для больших значений N. Поэтому здесь весовой алгоритм целесообразно применять, как в [8], лишь для малых значений N_1, $N_2 \approx 10$ с целью определения начальных тестовых значений N (см. далее разд. 5).

4. Коррелированный выбор номеров $\pi_1 = (i_1, j_1)$, $\pi_2 = (i_2, j_2)$ пар частиц, взаимодействующих в рассматриваемых системах в момент времени t, реализуется с помощью следующего алгоритма.

Далее в тексте через $\{\alpha_i\}$ будут обозначаться независимые случайные числа, равномерно распределенные в интервале $(0, 1)$. Сначала выбираем $\pi_1 = \pi_2$:

$$i_1 = i_2 = [\alpha_1N_2] + 1, \quad j' = [\alpha_2(N_2 - 1)] + 1,$$

$$\text{если } j' < i_2, \text{ то } j_1 = j_2 = j', \text{ иначе } j_1 = j_2 = j' + 1.$$

Если $\alpha_3 > N_2(N_2 - 1)/(N_1(N_1 - 1))$, то i_1 принимает другое значение:

$$i_1 = [\alpha_1(N_1 - N_2)] + N_2 + 1.$$

Если же при этом $\alpha_3 > 1 - (N_1 - N_2)(N_1 - N_2 - 1)/(N_1(N_1 - 1))$, то меняется и j_1:

$$j' = [\alpha_2(N_1 - N_2 - 1)] + N_2 + 1, \text{ если } j' < i_1, \text{ то } j_1 = j', \text{ иначе } j_1 = j' + 1.$$

Здесь квадратные скобки [ ] означают взятие целой части числа.

5. Соответственно заданной индикатрисе рассеяния (1.4) (см. также разд. 4) моделируются новые скорости $\mathbf{v}_i$, $\mathbf{v}_j$, остальные остаются без изменений.

6. Значение линейного функционала $G(T) = \int f(\mathbf{v}, T)q(\mathbf{v})d\mathbf{v}$ оценивается средним

$$\tilde{G}_N = \frac{1}{N}\sum_{i=1}^{N} Q(T)q(\mathbf{v}_i),$$

где $\{\mathbf{v}_i\}$ — скорости частиц в момент времени T. С целью уменьшения трудоемкости вычислений можно получать эту сумму пересчетом предыдущей суммы для $T' < T$ путем добавления к ней вкладов $\{Q(T)q(\mathbf{v}_i)\}$ “новых” частиц (скорости которых изменились за это время), предварительно удалив вклады “старых” частиц с теми же номерами. Для упрощения алгоритма в настоящей работе такой пересчет оценки осуществлялся после каждого столкновения.

3. ОПТИМИЗАЦИЯ N-ЧАСТИЧНОЙ ОЦЕНКИ

Обозначим через $\tilde{G}_N$ случайное значение N-частичной оценки G_N функционала $G = G(t) = \int f(\mathbf{v}, t)q(\mathbf{v})d\mathbf{v}$. Для построения достаточно точной статистической оценки величины G_N реализуется выборка значений $\{\tilde{G}_N^{(i)}\}$, $i = 1, ..., n$, и вычисляется величина

$$\tilde{G}_{N,n} = \frac{1}{n}\sum_{i=1}^{n} \tilde{G}_N^{(i)} \approx G_N.$$

Трудоемкость такой оценки можно минимизировать, используя подходящую связь между величинами N и n. Полагая, что $|G - G_N| = \gamma/N$, получаем соотношение:

$$\mathrm{E}(\tilde{G}_{N,n} - G)^2 = \mathrm{D}\tilde{G}_{N,n} + \frac{\gamma^2}{N^2} \approx \frac{d}{nN} + \frac{\gamma^2}{N^2}.$$

Следовательно, оптимизация трудоемкости здесь достигается в результате минимизации величины $S = nNs_1$ при условии

$$\frac{d}{nN} + \frac{\gamma^2}{N^2} = \varepsilon^2, \tag{3.1}$$

где ε — требуемая среднеквадратическая погрешность оценки $\tilde{G}_{N,n}$, $s_1 = \nu T s_0$, s_0 — трудоемкость розыгрыша столкновения. Из (3.1) в предположении, что $\gamma^2/N^2 \le \varepsilon^2$, $N \ge 1$, $n \ge 1$, получаем

$$nN = \frac{d}{\varepsilon^2 - \gamma^2/N^2}. \tag{3.2}$$

Минимальное значение этой величины достигается при $N = N_0 = +\infty$, причем надо полагать $n = n_0 = 1$. Однако вследствие ограниченности компьютерной памяти имеет место ограничение $N \le N_m$, которое в наилучшем случае позволяет реализовать оценку лишь с

$$\varepsilon^2 = \frac{d}{nN_m} + \frac{\gamma^2}{N_m^2}$$

при

$$n = n_m = \frac{d}{\varepsilon^2 N_m - \gamma^2/N_m}.$$

Можно также уменьшать трудоемкость оценки, уравнивая, как это иногда делается в математической статистике, слагаемые в (3.1). При этом $n = n_2 = dN_2/\gamma^2$, $N = N_2 = \gamma\sqrt{2}/\varepsilon$ и $n_2 N_2 = 2d/\varepsilon^2$.

Неравенство

$$n_m N_m = d/(\varepsilon^2 - \gamma^2/N_m^2) < 2d/\varepsilon^2 = n_2 N_2$$

выполняется, если $\varepsilon^2 - \gamma^2/N_m^2 \ge \varepsilon^2/2$, т.е. при $N_m > 2\gamma/\varepsilon$, причем выигрыш в трудоемкости при оптимальных значениях n_m и N_m получается не более, чем в два раза. Учитывая универсальность способа уравнивания можно рекомендовать его для практического использования.

4. АЛГОРИТМ ОЦЕНКИ РЕШЕНИЯ МОДЕЛЬНОГО УРАВНЕНИЯ БОЛЬЦМАНА С ПРОСТРАНСТВЕННО-ОДНОРОДНЫМИ НАЧАЛЬНЫМИ УСЛОВИЯМИ

В работах [10, 11] представляется точное решение пространственно-однородного уравнения Больцмана для максвелловских молекул. В этом случае уравнение Больцмана можно записать в виде

$$\frac{\partial f(\mathbf{v},t)}{\partial t} = \int g\left(\frac{(\mathbf{u},\mathbf{n})}{|\mathbf{u}|}\right)[f(\mathbf{v}',t)f(\mathbf{v}_1',t) - f(\mathbf{v},t)f(\mathbf{v}_1,t)]\,d\mathbf{v}_1\,d\mathbf{n}, \tag{4.1}$$

где $\mathbf{u} = \mathbf{v} - \mathbf{v}_1$; $g(\cos\theta) = g\left(\frac{(\mathbf{u},\mathbf{n})}{|\mathbf{u}|}\right) = \sigma(|\mathbf{u}|,\Omega)|\mathbf{u}|$ – произведение дифференциального сечения рассеяния на модуль относительной скорости; $\mathbf{n}$ – единичный вектор направления относительной скорости частиц после столкновения; $d\mathbf{n} = \sin\theta\,d\theta\,d\varphi$.

Скорости частиц после столкновения имеют вид

$$\mathbf{v}' = \frac{\mathbf{v}+\mathbf{v}_1}{2} + \frac{|\mathbf{v}-\mathbf{v}_1|}{2}\mathbf{n}; \quad \mathbf{v}_1' = \frac{\mathbf{v}+\mathbf{v}_1}{2} - \frac{|\mathbf{v}-\mathbf{v}_1|}{2}\mathbf{n}. \tag{4.2}$$

В [10] рассматривается задача Коши для уравнения (4.1) с начальным условием

$$f(\mathbf{v},0) = f_0(\mathbf{v}) = \frac{\exp\left(-\frac{\mathbf{v}^2}{2(1-\beta)}\right)}{(2\pi(1-\beta))^{3/2}}\left[\frac{\beta}{(1-\beta)}\left(\frac{\mathbf{v}^2}{2(1-\beta)} - \frac{3}{2}\right) + 1\right], \tag{4.3}$$

которое удовлетворяет следующим условиям нормировки:

$$\int f(\mathbf{v},0)d\mathbf{v} = 1, \quad \int f(\mathbf{v},0)\mathbf{v}d\mathbf{v} = 0, \quad \int f(\mathbf{v},0)\mathbf{v}^2 d\mathbf{v} = 3.$$

Решение этого уравнения Больцмана, которое получено в [10], имеет вид

$$f(\mathbf{v},t) = (2\pi\tau)^{-\frac{3}{2}}\exp\left(-\frac{\mathbf{v}^2}{2\tau}\right)\left[\frac{1-\tau}{\tau}\left(\frac{\mathbf{v}^2}{2\tau} - \frac{3}{2}\right) + 1\right], \tag{4.4}$$

где

$$\tau = \tau(t) = 1 - \beta e^{-\lambda t}, \quad \lambda = \frac{\pi}{2}\int_0^\pi g(\cos\theta)\sin^3\theta d\theta, \quad t \geqslant 0, \quad 0 \leqslant \beta \leqslant 0.4.$$

Моделирование цепи Маркова для оценки решения задачи Коши (4.1),(4.3) производится согласно алгоритму, описанному в разд. 2.

1. Предполагается, что система из N частиц стартует в момент времени $t = 0$, начальные скорости частиц $\mathbf{v}_i$, $i = 1, ..., N$, выбираются соответственно плотности $f_0(\mathbf{v})$.

Моделирование согласно плотности $f_0(\mathbf{v})$ осуществляется методом суперпозиции (см., например, [12]) на основе представления:

$$f_0(\mathbf{v}) = p_1 f_1(\mathbf{v}) + p_2 f_2(\mathbf{v}),$$

где

$$p_1 = 1 - \frac{3\beta}{2(1-\beta)}; \quad p_2 = \frac{3\beta}{2(1-\beta)}; \quad \sigma = \sqrt{1-\beta};$$

$$f_1(\mathbf{v}) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{\mathbf{v}^2}{2\sigma^2}\right); \qquad f_2(\mathbf{v}) = \frac{\mathbf{v}^2}{3(2\pi)^{3/2}\sigma^5}\exp\left(-\frac{\mathbf{v}^2}{2\sigma^2}\right).$$

Для вывода формулы моделирования соответственно плотности $f_2(\mathbf{v})$, сделаем замену переменных $\mathbf{v} = v\omega$, где $v = |\mathbf{v}|$. Плотность распределения вектора (ω, v) выражается формулой:

$$\tilde{f}_2(\omega, v) = \frac{1}{4\pi}\frac{4\pi v^4}{3(2\pi)^{3/2}\sigma^5}\exp\left(-\frac{v^2}{2\sigma^2}\right) = \frac{1}{4\pi}p_{\chi_5(\sigma)}(v),$$

где $p_{\chi_5(\sigma)}(v)$ – плотность распределения случайной величины $\chi_5(\sigma) = \left(\sum_{i=1}^5 \xi_i^2\right)^{1/2}$. Величины ξ_i, $i = 1, ..., 5$, имеют нормальное распределение $N(0,\sigma^2)$ и независимы в совокупности. Соответственно выражению плотности

$\tilde{f}_2(\omega, v)$, следует выбрать единичный изотропный вектор ω и значение модуля вектора по формуле (см. [12], п. 1.9.1)

$$v = \sigma\sqrt{-2\ln(\alpha_1\alpha_2) - 2\sin^2(2\pi\alpha_3)\ln\alpha_4}.$$

2. В задаче Коши (4.1),(4.3) рассматривается случай максвелловских молекул, т.е. $a(\mathbf{v}_i, \mathbf{v}_j) = \text{const} = a_0$, $i, j = 1, \ldots, N$. Следовательно, как указано в разд. 2, время свободного пробега τ выбирается соответственно плотности $p_1(t) = (Na_0/2)\exp(-(Na_0/2)t)$, т.е. временная координата нового столкновения вычисляется по формуле: $t = t' - 2\ln\alpha/(Na_0)$.

3. Номера пары частиц $\pi = (i, j)$, моделируются по формулам:

$$i = [\alpha_1 N] + 1, \quad j' = [\alpha_2(N-1)] + 1, \text{ если } j' < i, \text{ то } j = j', \text{ иначе } j = j' + 1.$$

Если необходимо коррелировать оценки для $N = N_1, N_2$, то используется алгоритм (3) из разд. 2.

4. Вычисляются направляющие косинусы a, b, c вектора $\mathbf{n}$ по стандартным формулам (см. [12]), причем можно полагать $a' = b' = 0$, $c' = 1$, т.к. $\mathbf{n}$ – изотропный вектор. Для вариантов с $N = N_1$, $N = N_2$ этот вектор повторяется.

Величина $\mu = \cos\theta$ моделируется соответственно плотности, пропорциональной функции $g(\cos\theta)$ из уравнения (4.1), $\varphi = 2\pi\alpha$, т.е. имеет равномерное на промежутке $(0, 2\pi)$ распределение. Приведенный алгоритм вытекает из того, что углы φ и θ соответствуют сферической системе координат с осью (0,0,1). Новые скорости $\mathbf{v}_i'$, $\mathbf{v}_j'$, которые частицы приобретают после столкновения, вычисляются согласно формуле (4.2). Скорости частиц, не участвующих в столкновении, не меняются.

5. Вычисляется результативная оценка (как в п.6 из разд. 2):

$$\tilde{G_N}(T) = \frac{1}{N}\sum_{i=1}^{N} Q(T)q(\mathbf{v}_i).$$

5. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

В качестве функционалов от решения задачи Коши (4.1), (4.3) рассматриваются четные моменты скоростей частиц

$$z_m(t) = \frac{1}{(2m+1)!!}\int f(\mathbf{v}, t)|\mathbf{v}|^{2m}d\mathbf{v}, \quad m = 0, 1\ldots,$$

где $f(\mathbf{v}, t)$ – точное решение этой задачи.

В [11] приведены формулы для записанных выше функционалов. Они имеют вид

$$z_0 = z_1 = 1, \quad z_m(t) = (1 - \beta e^{-\lambda t})^{m-1}\left[1 + (m-1)\beta e^{-\lambda t}\right].$$

В численных экспериментах использовалась модель молекул, соответствующая случаю псевдомаксвелловских молекул [11], т.е. произведение дифференциального сечения рассеяния на модуль относительной скорости частиц было постоянным, а угловое распределение относительной скорости частиц после столкновения определялась индикатрисой: $g(\cos\theta) = 1/(4\pi)$, $-1 \leqslant \cos\theta \leqslant 1$, т.е. рассеяние изотропно.

Для исследования зависимости функционала $z_5(t)$ рассмотрим задачу Коши для уравнения (4.1) с начальным условием (4.3) и параметрами $\beta = 0.4$, $\lambda = 1/6$. Решение этой задачи $f(\mathbf{v},t)$ имеет вид (4.4).

Рассмотрим теперь результаты расчетов для сформулированной тестовой задачи об оценке функционала $z_5(t)$. В табл. 1 даны статистические оценки $\tilde{z}_5(t)$ этого функционала для различных значений N и n при $t = 12$, а также соответствующие оценки значения γ по формуле $\tilde{\gamma} = N(z_5 - \tilde{z}_5)$, где z_5 – точное значение. Полная погрешность оценки здесь и в табл. 2 вычислена соответственно формуле $\Delta = \sqrt{\mathrm{D}\tilde{G}_{N,n} + \gamma^2/N^2}$. Видно, что $\gamma \approx 2.32$. Отклонение от этого значения для $N = 10, 11$ связано с полной погрешностью оценок.

Для уточнения реальных расчетов с помощью выбора подходящих значений N, n и особенно с помощью экстраполяции по формуле (1.6), целесообразно использовать тестовые значения N, в какой-то степени близкие к $N = N^{(\varepsilon)}$, такому, что $|z_5 - \tilde{z}_5(N^{(\varepsilon)})| \leq \varepsilon$, где ε – требуемая погрешность. Такие значения N можно предварительно определять при $t = t_{\max}$ с помощью представленных в разд. 2 малотрудоемких зависимых испытаний, например, для $N_1 = 11$, $N_2 = 10$, как в [8]. При этом тестовое значение N определяется соотношением $\tilde{\gamma}/N = \varepsilon$, т.е. $N = \tilde{\gamma}/\varepsilon$, где $\tilde{\gamma}$ вычисляется по формуле (1.7) для $t = t_{\max}$. Расчеты, проведенные при $n = 10^7$, $N_1 = 11$, $N_2 = 10$ в рассматриваемой задаче для $\varepsilon = 0.001$ дали $\tilde{\gamma} = 1.56$ и $N = 1560$. Поэтому для реализации экстраполяционной оценки (1.6) были выбраны значения $N_1 = 2000$ и $N_2 = 1000$. Соответствующие независимые оценки функционала $z_5(t)$, их статистическая погрешность и значения γ даны в табл. 3. Видно, что экстраполяция уменьшает полную погрешность примерно на порядок.

Таблица 1. Значения оценки и полной погрешности при $t = 12$

N	n	$\tilde{z}_5 \pm \Delta$	$\tilde{\gamma}$
10	10^8	0.7864 ± 0.1874	1.873
11	10^8	0.7999 ± 0.1739	1.912
100	10^8	0.9512 ± 0.0226	2.251
200	10^8	0.9622 ± 0.0116	2.308
300	10^8	0.9660 ± 0.0078	2.322
500	10^7	0.9690 ± 0.0047	2.290
1000	10^7	0.9714 ± 0.0023	2.320
2000	10^7	0.9726 ± 0.0012	2.320

Таблица 2. Значения оптимизированной (при $t = 12$) оценки функционала и полной погрешности

t	При $\varepsilon = 0.001$, $N = 3268$, $n = 65500$	При $\varepsilon = 0.0001$, $N = 32680$, $n = 655000$	Точное решение
0	0.33697 ± 0.00016	0.33695 ± 0.00002	0.33696
1	0.45054 ± 0.00023	0.45055 ± 0.00002	0.45055
2	0.55571 ± 0.00032	0.55595 ± 0.00003	0.55593
3	0.64815 ± 0.00041	0.64838 ± 0.00004	0.64839
4	0.72603 ± 0.00049	0.72617 ± 0.00005	0.72625
5	0.78880 ± 0.00061	0.78971 ± 0.00006	0.78980
6	0.83967 ± 0.00065	0.84035 ± 0.00007	0.84043
7	0.87928 ± 0.00074	0.87997 ± 0.00007	0.88000
8	0.90956 ± 0.00082	0.91043 ± 0.00008	0.91046
9	0.93257 ± 0.00086	0.93362 ± 0.00009	0.93363
10	0.95072 ± 0.00083	0.95108 ± 0.00008	0.95106
11	0.96332 ± 0.00095	0.96416 ± 0.00010	0.96408
12	0.97341 ± 0.00100	0.97378 ± 0.00010	0.97374

Далее будет дано полуэвристическое объяснение этого эффекта, фактически связанного почти с занулением детерминированной погрешности. В связи с этим в табл. 3 даны оценки статистической погрешности. В табл. 2 даны результаты расчетов для оптимальных значений $N = N_2$, $n = n_2$ (см. разд. 3), вычисленных при $t = 12$ для $\varepsilon = 0.001$ и $\varepsilon = 0.0001$.

Теперь дадим полуэвристическое объяснение эффекта экстраполяции. Для этого формулу (1.6) перепишем в виде

$$\tilde{G}_\infty \approx -\frac{N_2}{N_1 - N_2}\tilde{G}_{N_2} + \frac{N_1}{N_1 - N_2}\tilde{G}_{N_1}.$$

Подстановка сюда выражений $\tilde{G}_{N_i} \approx G_\infty + \gamma/N_i$ дает равенство $\tilde{G}_\infty = G_\infty$, т.е. в линейном (по $1/N$) приближении детерминированная погрешность уменьшается до нуля. В то же время дисперсия экстраполяционной оценки σ_e^2 может не сильно возрасти. Это можно получить из соотношения

$$\sigma_e^2 = \mathrm{D}(\tilde{G}_\infty) = \frac{N_2^2}{(N_1 - N_2)^2}\mathrm{D}\tilde{G}_{N_2} + \frac{N_1^2}{(N_1 - N_2)^2}\mathrm{D}\tilde{G}_{N_1}$$

особенно в случае $N_2 \ll N_1$. Величина $\mathrm{D}\tilde{G}_\infty$ оценивается численно и определяет статистическую оценку погрешности, приведенную в табл. 3. Простые вычисления здесь дают соотношение $\sigma_e \approx \sqrt{6}\sigma_1 \approx 2.4\sigma_1$, где

Таблица 3. Значения независимых оценок функционала $z_5(t)$ и статистические погрешности (одна σ)

t	$N_2 = 1000$, $n = 10^7$	$N_1 = 2000$, $n = 10^7$	Точное решение	Экстраполяция	γ	d
0	0.33697 ± 0.00002	0.33696 ± 0.00001	0.33696	0.33694	0.0320	5
1	0.45046 ± 0.00003	0.45051 ± 0.00002	0.45055	0.45057	0.1140	12
2	0.55572 ± 0.00004	0.55583 ± 0.00003	0.55593	0.55595	0.2333	20
3	0.64795 ± 0.00005	0.64817 ± 0.00003	0.64839	0.64840	0.4437	31
4	0.72548 ± 0.00006	0.72582 ± 0.00004	0.72625	0.72616	0.6847	43
5	0.78870 ± 0.00007	0.78927 ± 0.00005	0.78980	0.78984	1.1399	55
6	0.83915 ± 0.00008	0.83971 ± 0.00005	0.84043	0.84027	1.1205	66
7	0.87847 ± 0.00008	0.87919 ± 0.00006	0.88000	0.87991	1.4359	76
8	0.90864 ± 0.00009	0.90952 ± 0.00006	0.91046	0.91040	1.7584	85
9	0.93170 ± 0.00009	0.93262 ± 0.00006	0.93363	0.93355	1.8499	92
10	0.94913 ± 0.00009	0.94994 ± 0.00007	0.95106	0.95075	1.6241	98
11	0.96185 ± 0.00010	0.96292 ± 0.00007	0.96408	0.96399	2.1405	103
12	0.97142 ± 0.00010	0.97258 ± 0.00007	0.97374	0.97374	2.3108	107

$\sigma_1 = \sqrt{\mathrm{D}\tilde{G}_{N_1}}$. Погрешность экстраполяционной оценки $\tilde{z}_5^{(e)}(t)$ лишь для значений $t = 6, 10$ несколько превышает σ_e:

$$|\tilde{z}_5^{(e)}(6) - z_5(6)| \approx 1.3\sigma_e, \quad |\tilde{z}_5^{(e)}(10) - z_5(10)| \approx 1.8\sigma_e.$$

Анализ оценок из таблиц 2, 3 с учетом выражения $S = CnN$ среднего числа арифметических операций (см. разд. 3) показывает, что оптимизационные оценки являются менее трудоемкими.

В заключение заметим, что полученные результаты убедительно подтверждают соотношение

$$G_N(t) = z_5(t; N) \approx z_5(t) + \gamma(t)/N.$$

Ясно, что это соотношение связано с квадратичным порядком $O(N^2)$ числа столкновений частиц вследствие chaos propogation, однако строгое доказательство здесь пока не получено. Можно также отметить, что многочастичный алгоритм является важной составной частью методики mean field game, возникнув ранее появления этого термина.

Отметим, что в [13] проведен сравнительный анализ дифференциальной и соответствующей стохастической пуассоновской SEIR моделей [14] для тестовой задачи моделирования пандемии COVID-19 в Новосибирске с 23 марта по 21 июня 2020 г. с начальной численностью $N = 2\,798\,170$. Путем варьирования начальной численности вида $N_k = kN$ при $k \geq 2$ было показано, что среднестатистические числа случаев выявления заболевших меньше (начиная с 7 апреля) соответствующих дифференциальных значений на величину, статистически неотличимую от $C(t)/k$, причем для 21 июня значение $C \approx 27.3$; это соотношение позволяет использовать стохастическую модель для больших значений N. Исследовалось также влияние на прогноз введения запаздывания, т.е. инкубационного периода, соответствующего пуассоновской модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Кац М.* Вероятность и смежные вопросы в физике. М.: Мир, 1965. 408 с.

2. *Михайлов Г.А.* Весовые методы Монте-Карло. Новосибирск: Изд-во СО РАН, 2000. 248 с.

3. *Михайлов Г.А., Рогазинский С.В.* Весовые методы Монте-Карло для приближённого решения нелинейного уравнения Больцмана // Сиб. матем. журнал. 2002. Т. 48. № 3. С. 620–621.

4. *Ivanov H.S., Rogasinsky S.V.* Analysis of numerical techniques of the direct simulation Monte Carlo method in the rarefied gas dynamics // Sov. J. Numer. Anal. Math. Modeling. 1988. Vol. 3. № 6. P. 453-465.

5. *Денисик С.А., Лебедев С.Н., Малама Ю.Г.* Об одной проверке нелинейной схемы метода Монте-Карло // Ж. вычисл. матем. и матем. физ. 1971. Т.11. № 3. С. 783–785.

6. *Бёрд Г.* Молекулярная газовая динамика. М.: Мир, 1981.

7. *Королев А.Е., Яницкий В.Е.* Прямое статистическое моделирование столкновительной релаксации в смесях газов с большим различием в концентрациях // Ж. вычисл. матем. и матем. физ. 1983. Т. 23. № 3. С. 674–680.

8. *Иванов М.С., Коротченко М.А., Михайлов Г.А., Рогазинский С.В.* Глобально-весовой метод Монте-Карло для нелинейного уравнения Больцмана // Ж. вычисл. матем. и матем. физ. 2005. Т.45. № 10. С. 1860–1870.

9. *Лотова Г.З., Михайлов Г.А.* Исследование сверхэкспоненциального роста среднего потока частиц в случайной размножающей среде // Сиб. ж. вычисл. матем. 2023. Т. 26. № 4. С. 401—413.

10. *Бобылев А.В.* О точных решениях уравнения Больцмана // Докл. АН СССР. 1975. Т. 225. № 6. С. 1296–1299.

11. *Бобылев А.В.* Точные решения нелинейного уравнения Больцмана и теория релаксации максвелловского газа // Теор. и матем. физ. 1984. Т. 60. № 2. С. 280–310.

12. *Михайлов Г.А., Войтишек А.В.* Численное статистическое моделирование, методы Монте-Карло. М.: Академия, 2006. 367 с.

13. *Lotova G.Z., Lukinov V.L., Marchenko M.A., Mikhailov G.A., and Smirnov D.D.* Numerical-statistical study of the prognostic efficiency of the SEIR model // Rus. J. Numer. Analysis Math. Modelling. 2021. Vol. 36. № 6. P. 337–345.

14. *Pertsev N.V., Loginov K.K., Topchii V.A.* Analysis of a stage-dependent epidemic model based on a non-Markov random process // J. Appl. Industr. Math. 2020. V. 14. № 3. P. 566–580.

INVESTIGATION AND OPTIMIZATION OF THE N-PARTIAL NUMERICAL STATISTICAL ALGORITHM FOR SOLVING THE BOLTZMANN EQUATION

G. Z. Lotova[a,b,*], G. A. Mikhailov[a,b], S. V. Rogazinsky[a,b]

[a] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Academician Lavrentyev Ave., 6, Novosibirsk, 630090, Russia*

[b] *Novosibirsk State University, Pirogov st., 2, Novosibirsk, 630090, Russia*

**e-mail: lot@osmf.sscc.ru*

Received 27 November, 2023

Revised 27 November, 2023

Accepted 14 January, 2024

Abstract. The primary goal of the study is to test the hypothesis that the known N-partial statistical algorithm provides an estimate of the solution to the nonlinear Boltzmann equation with an error of order $O(1/N)$. To achieve this, practically important optimal relationships between the value of N and the number n of sample estimates are determined. Numerical results for a problem with a known solution confirm the adequacy of the formulated estimates and conclusions.

Keywords: Monte Carlo method, statistical modeling, Boltzmann equation, N-partial Markov chain, molecular chaos, majorant frequency method.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 852–866

УДК 519.63

ПРИМЕНЕНИЕ СХЕМ CABARET И WENO ДЛЯ РЕШЕНИЯ НЕЛИНЕЙНОГО УРАВНЕНИЯ ПЕРЕНОСА В ЗАДАЧЕ МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ ВОЛНЫ ЗВУКОВОГО УДАРА В АТМОСФЕРЕ[1)]

© 2024 г. П. А. Мищенко[1,*], Т. А. Гимон[1], В. А. Колотилов[1,2], А. Н. Кудрявцев[1]

[1]630090 Новосибирск, ул. Институтская, 4/1, Институт теоретической и прикладной механики им. С.А. Христиановича СО РАН, Россия

[2]630090 Новосибирск, пр-т Акад. Лаврентьева, 15, Институт гидродинамики им. М. А. Лаврентьева СО РАН

**e-mail: mischenko.polina.16@gmail.com*

Поступила в редакцию 21.09.2023 г.

Переработанный вариант 20.12.2023 г.

Принята к публикации 06.02.2024 г.

Наиболее удобной моделью описания явления распространения волн звукового удара в атмосфере является расширенное уравнение Бюргерса. В настоящей работе исследовалось влияние численной схемы на результат решения уравнения, учитывающего нелинейный характер распространения в атмосфере волн звукового удара. Это уравнение является ключевым компонентом расширенного уравнения Бюргерса и определяет характер трансформации профиля возмущенного давления при его распространении. Для решения применялись две численные схемы: CABARET и WENO, квазимонотонные сквозные счетные схемы, позволяющие получить решение без значительных численных осцилляций. Проводился анализ применимости данных схем для решения рассматриваемой задачи. Библ. 19. Фиг. 12.

Ключевые слова: звуковой удар, нелинейное уравнение переноса, распространение волн малой амплитуды, схема CABARET, схема WENO.

DOI: 10.31857/S0044466924050136, **EDN**: YCUYOH

ВВЕДЕНИЕ

Звуковой удар, возникающий при достижении поверхности земли возмущений, генерируемых при крейсерском полете сверхзвукового самолета, может приводить к ряду нежелательных последствий. Минимизация звукового удара — одна из актуальных проблем в области конструирования сверхзвуковых пассажирских самолетов нового поколения (см. [1]). Точное численное моделирование сверхзвукового обтекания геометрии самолета и распространения возникающих возмущений параметров среды до поверхности земли необходимо для успешного прогнозирования звукового удара.

Наиболее эффективным подходом к моделированию распространения в атмосфере на дальние расстояния волн звукового удара является решение расширенного уравнение Бюргерса (см. [2]). Одна из составляющих этого уравнения представляет собой одномерное квазилинейное уравнение переноса и описывает перенос в пространстве волны малой амплитуды со скоростью распространения, определяемой параметрами среды и интенсивностью самих возмущений. При распространении начального профиля волны на дальнее расстояние, как например, с высоты полета до поверхности земли, профиль волны может значительно трансформироваться, могут образовываться разрывы решения.

Распространенные способы решения рассматриваемого уравнения основываются на его неявном аналитическом решении (решение Пуассона) (см. [2]–[4], [9]), а также метода параметров формы волны (см. [10]), разработанный из теории геометрической акустики Ч. Томасом.

При возникающей многозначности решения для локализации разрывов применяются условие Ренкина–Гюгонио (см. [8]) и следствие из него “правило равенства площадей” (см. [5], [7]). Также в работах [6], [9] используется метод Бургера–Хейса, заключающийся в выявлении узлов с минимальными значениями потенциала численного решения. Также во избежание возникновения областей многозначности решения в [2], [4] ограничивается шаг дискретизации, не позволяющий достичь момента формирования разрывов решения. Данные

[1)] Работа выполнена при финансовой поддержке Минобрнауки РФ (проект №121030900260-6).

аналитические подходы имеют свои недостатки. Некоторые из методов сложно реализуемы для решения практических задач, могут в зависимости от параметров задачи повышать требуемые вычислительные мощности численного алгоритма, а также не удовлетворять условию консервативности.

Схемы сквозного счета кажутся более универсальным подходом. Однако они нашли меньшее распространение при решении рассматриваемого в работе нелинейного уравнения переноса, несмотря на множество схем сквозного счета, разработанных для решения уравнений гиперболического типа (см. [11]–[13]). Частой проблемой данных схем является возникающие нефизические осцилляции в областях особенностей решения. Схемы со значительной численной вязкостью, помогающей бороться с ними, могут наоборот гасить физические локальные экстремумы.

Важными параметрами при моделировании распространения акустических волн являются амплитуда возмущений и время нарастания фронта волны, которые определяют интенсивность воздействия звукового удара. Сложность рассматриваемой задачи, включая большую протяженность области моделирования и сложную структуру исходных данных с волнами разрежения, скачками разной интенсивности и разрывами, требует от используемых численных методов не только высокой точности, но также низкой диссипации и дисперсии.

В настоящей работе рассматриваемое нелинейное уравнение переноса решалось с помощью численных схем сквозного счета CABARET и WENO. Также для оценки влияния выбора численной схемы на точность конечного численного решения применялась явная схема Годунова. Сравнение точности применяемых схем и определение основных особенностей полученных численных решений проводились на ряде тестовых примеров с распространением плоских волн различной сложности. Решение данными схемами тестового примера для уравнения Хопфа, подобного рассматриваемому в работе уравнению переноса, позволило верифицировать численные результаты расчетов. Программная реализация проводилась на языке программирования Python.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим задачу Коши для одномерного квазилинейного уравнения переноса в следующей постановке:

$$\frac{\partial p}{\partial x} + C\frac{\partial p^2}{\partial t'} = 0, \quad p(t',0) = p_0(t'), \tag{1}$$

где $p(t',x)$ — акустическое давление (переносимая величина), x — криволинейная координата вдоль траектории распространения волны (пространственная переменная распространения), $t' = t - \int_0^x \frac{1}{c_0}dx$ — временная переменная длительности сигнала, коэффициент C определяет скорость переноса $2Cp$, $p_0(t')$ — начальный профиль распределения давления. В дальнейшем в работе для простоты будем t' обозначать t.

В работе рассматривалось уравнение с коэффициентом C, принимающим постоянное значение $1/2$ (уравнение Хопфа), так и с C, определяемым параметрами среды

$$C = \frac{\beta}{2\rho_0 c_0^3}, \tag{2}$$

где коэффициент нелинейности β определяется через показатель адиабаты γ, как $\beta = (\gamma+1)/2$, ρ_0 — плотность окружающей среды, и c_0 — скорость звука на высоте распространения.

В настоящей работе задача Коши (1) для уравнения переноса решалась с применением численных схем CABARET, WENO и Годунова. Сравнение точности расчетов по данным схемам проводилось на ряде тестовых примеров с распространением плоских волн при двух отличающихся скоростях переноса C.

2. ЧИСЛЕННЫЙ МЕТОД РЕШЕНИЯ

В работе для численного решения уравнения (1) с начальными данными (2) применялись три численные схемы: CABARET, WENO и Годунова.

Схема CABARET (см. [14], [15]) имеет второй порядок точности по времени и по пространству и характеризуется использованием дробных узлов сетки, помимо основных. Для нивелирования нефизических осцилляций проводится нелинейная коррекция потоковых членов на основе принципа максимума. В этой схеме используются потоковые и консервативные переменные, заданные соответственно в полуцелых и целых пространственных узлах разностной сетки. Именно в консервативных переменных обеспечивается выполнение законов сохранения. В качестве начальных данных консервативных переменных использовались дискретные значения сеточных начальных функций. Учитывая разрывность начальных данных, начальные потоковые переменные в полуцелых узлах согласовывались с консервативными с применением оператора Римана (см. [15]). Данный оператор представляет собой решение задачи о распаде разрыва в текущем узле.

Конечно-объемная схема WENO (Weighted essentially non-oscillatory scheme) (см. [16]) — схема пятого порядка точности по пространству и третьего по времени на гладких решениях. Данная схема основывается на

применении линейной комбинации реконструкций по трем различным шаблонам искомой функции для обоих сторон грани ячейки и интегрирования по времени по схеме Рунге–Кутты. Значения потоковых переменных вычисляются из решения задачи о распаде разрыва.

Схема Годунова — консервативная схема первого порядка точности, также использующая для вычисления потоковых переменных решение задачи о распаде разрыва, но с состояниями, соответствующими значениям искомой функции в соседних ячейках сетки. Данная схема является широко распространенной численной схемой решения уравнений гиперболического типа, характеризующаяся высокой схематической вязкостью. В настоящей работе она была реализована для оценки точности схем CABARET и WENO при решении тестовых примеров.

Аппроксимируем задачу (1) схемами CABARET, WENO и Годунова, заданными на прямоугольной равномерной разностной сетке:

$$\{t_j, x_n\} : t_j = j\Delta\tau, \quad x_{n+1} = x_n + \Delta x_n, \quad x_0 = 0, \tag{3}$$

в котором $\Delta\tau$ — постоянный шаг сетки по времени, а Δx_n — шаг сетки по пространству, определяемый из условия устойчивости

$$\Delta x_n = \frac{z\Delta\tau}{\max\limits_j |a^n_{j+1/2}|},$$

где $z \in (0,1)$ — коэффициент запаса, $a^n_{j+1/2} = 2Cp_{j+1/2}$, а значения $p_{j+1/2}$ заданы в полуцелых временных узлах разностной сетки $t_{j+1/2} = t_j + \Delta\tau/2$. Данное условие устойчивости можно считать аналогом условия устойчивости Куранта (см. [17]) для рассматриваемого уравнения переноса, тогда коэффициент запаса z соответствует числу Куранта. Далее в работе для простоты вместо коэффициента запаса z будет часто применяться понятие числа Куранта и обозначение CFL. Далее в работе будет показано, что число Куранта является важным параметром, оказывающим влияние на точность расчетов для схем CABARET и Годунова.

3. РЕШЕНИЕ ТЕСТОВЫХ ЗАДАЧ

3.1. Решение задачи Коши для уравнения переноса (1) с постоянным коэффициентом $C = 1/2$

В качестве тестового примера рассмотрим задачу Коши для уравнения Хопфа (нелинейного уравнения переноса с фиксированным коэффициентом скорости переноса $= 1/2$) следующего вида:

$$\frac{\partial p}{\partial x} + \frac{1}{2}\frac{\partial p^2}{\partial t} = 0, \quad p(t,0) = p_0(t), \tag{4}$$

На фиг. 1а приведены точное решение и результаты численного решения (4) с кусочно-постоянными начальными данными, представляющими собой прямоугольное возмущение “ступенька”:

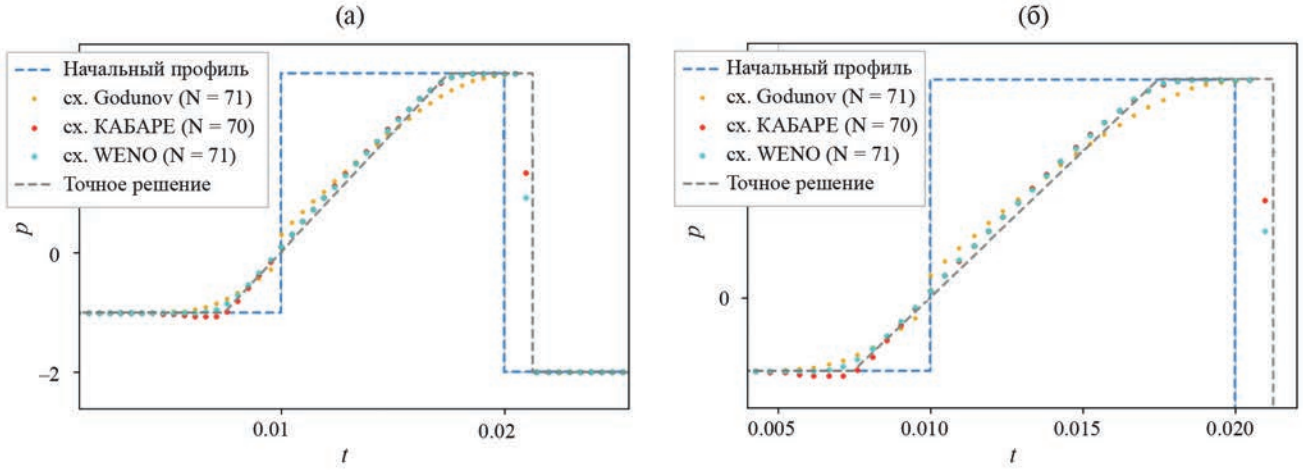
$$p_0(t) = \begin{cases} -1, & t < 0.01, \\ 3, & 0.01 \le t \le 0.02, \\ -2, & t > 0.02. \end{cases} \tag{5}$$

Также на фиг. 2 приведены результаты численного решения задачи с гладкими синусоидальными начальными данными:

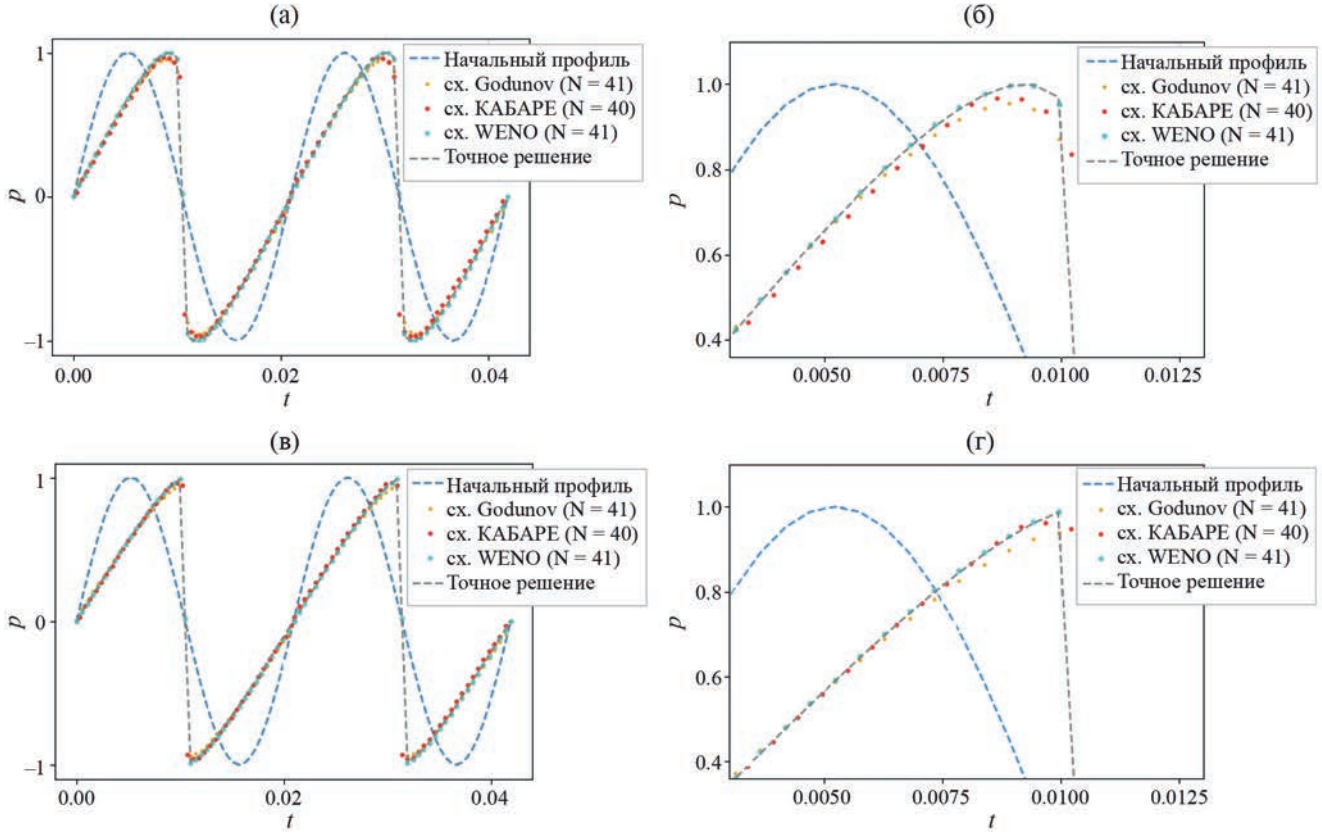
$$p_0(t) = \sin(t). \tag{6}$$

Начальные данные и точные решения этих задач показаны пунктирными линиями, а точками — численные результаты, полученные с применением схем CABARET, WENO и Годунова. Точное решение находилось из аналитического решения для данного уравнения, а расположение и амплитуда скачков профиля решения из условия Ренкина–Гюгонио (см. [18]). Расчеты проводились на прямоугольной разностной сетке (3) с размерностью в 70 узлов и с временным шагом $\Delta\tau = 0.0004$ (фиг. 1), а также на сетке в 80 узлов с временным шагом $\Delta\tau = 0.0005$ (фиг. 2) при коэффициенте запаса $z = 0.5$ в условии устойчивости.

Результаты вычислений на фиг. 1 и 2 демонстрируют хорошее согласование с точным решением задачи Коши. Данные схемы сохраняют высокую точность при локализации сильных и слабых разрывов и воспроизводят волны разрежения. Численные решения отличаются монотонностью, не генерируются нефизические осцилляции в областях разрывов профиля решения.



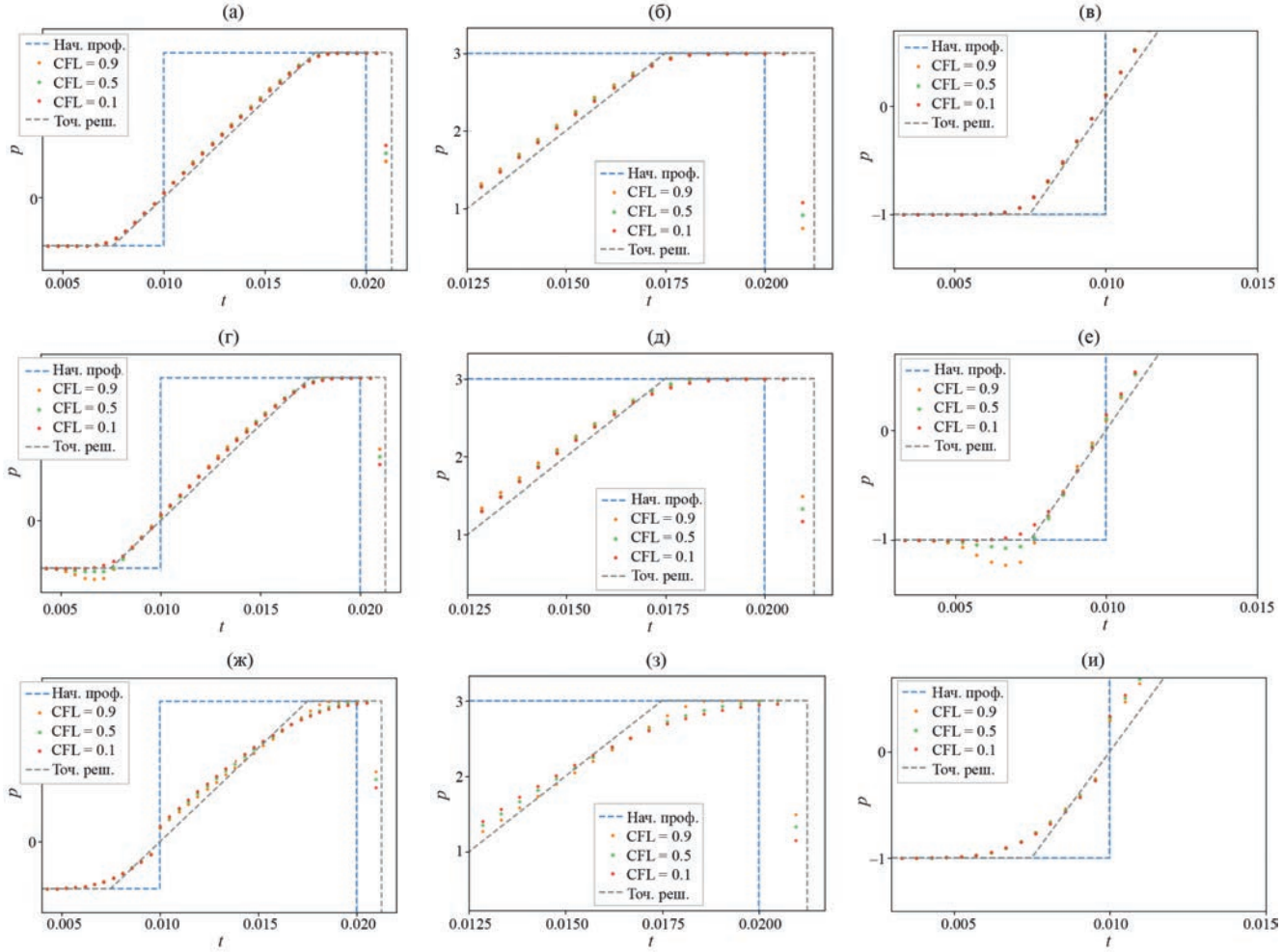
Фиг. 1. Сравнение точного и численных решений задачи Коши (4), (5) при распространении на $x = 225$ м (а), область ударных волн профиля решения (б).



Фиг. 2. Сравнение аналитического и численных решений задачи Коши (4), (6) при распространении на $x = 360$ м (а), (б) — область головной ударной волны профиля решения, (в) — сравнение решений при распространении на $x = 480$ м, (г) — область головной ударной волны профиля решения.

На масштабированных фиг. 1б и 1г видно, что результаты WENO схемы относительно результатов схем CABARET и Годунова характеризуются большей крутизной фронта, большей интенсивностью скачков и большим совпадением с точным решением. В областях с ударными волнами схема CABARET сглаживает неоднородности решения сильнее, чем WENO, а схема Годунова – еще сильнее. Это соответствует упомянутому ранее свойству схемы Годунова. Также наблюдается разрыв численного решения схемы Годунова в области волны разрежения.

На фиг. 3 представлены численные результаты распространения профиля (5) “ступеньки” на расстояние 225 м при числе Куранта, принимающего значения 0.1, 0.5 и 0.9. Представлены численные решения в общем



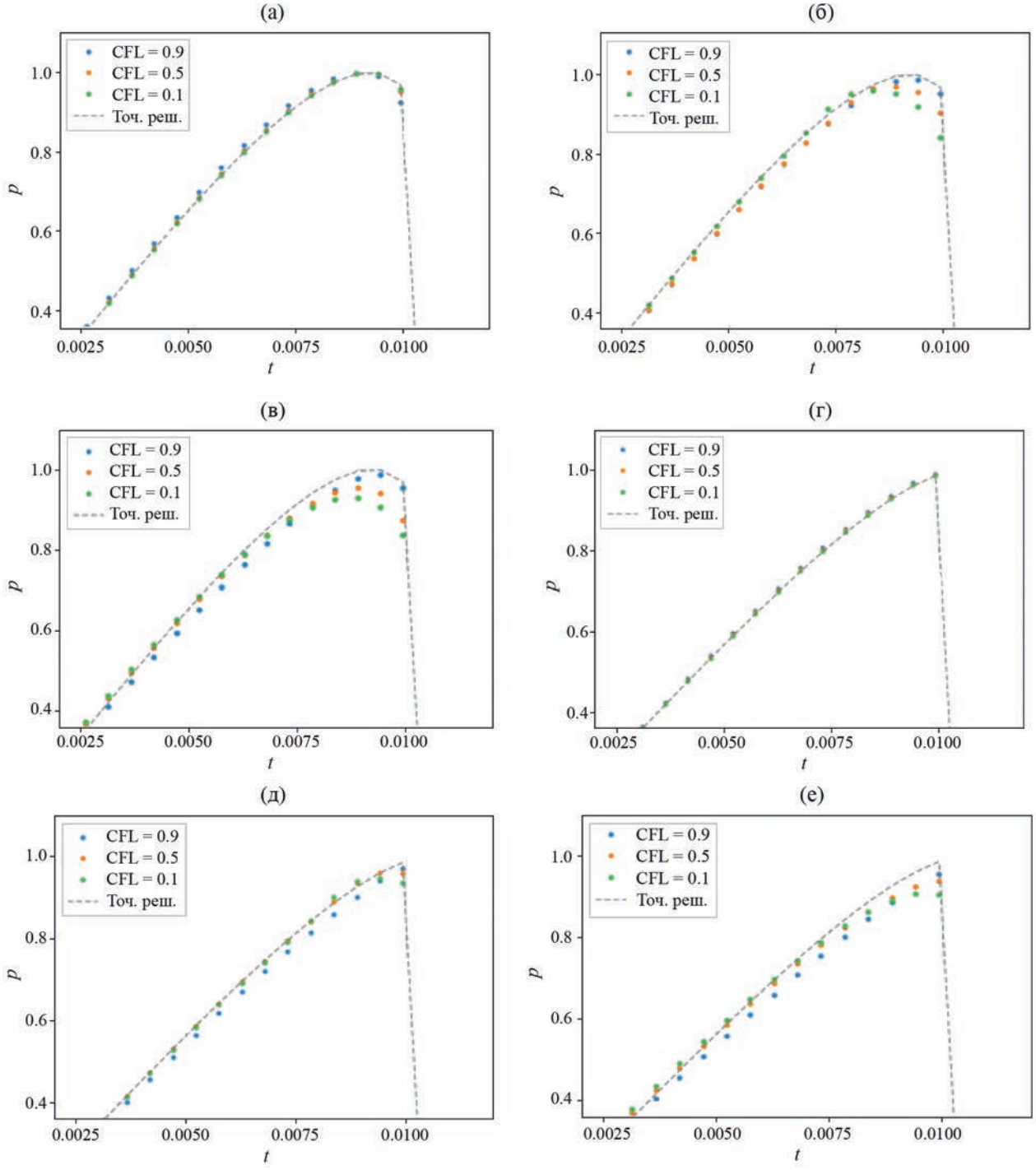
Фиг. 3. Влияние числа Куранта на вид численного решения задачи Коши (4), (5), полученного по схемам WENO (а)–(в), CABARET (г)–(е) и Годунова (ж)–(и).

виде, а также области решений с разрывами более детально. Численные решения на фиг. 3 представлены точками, а точное решение — пунктирной линией. Для схемы WENO изменение числа Куранта не приводит к значительным изменениям профиля решения. Число Куранта для схем Годунова и CABARET оказывает большее влияние на точность решения. При числах Куранта 0.1 и 0.5 численные результаты WENO и CABARET показывают более точное разрешение слабых разрывов решения и волны разрежения. Как и следует из свойств схемы CABARET, при приближении числа Куранта к 0.5 точность численного решения растет. Для числа Куранта 0.9 схема Годунова показывает более точное разрешение разрывов и скачков решения.

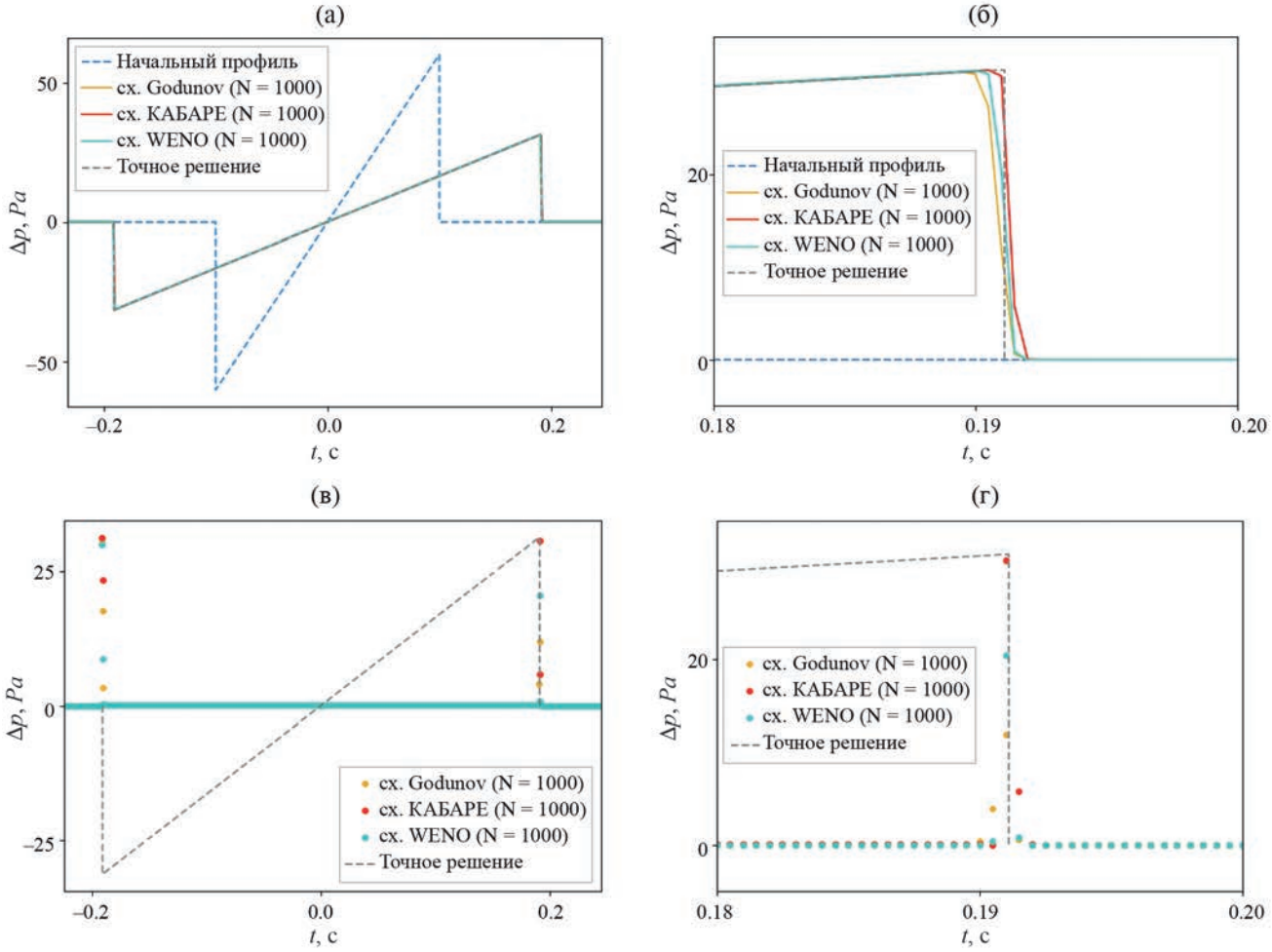
На фиг. 4 представлены области численных решений со скачками, полученные для случая распространении начального синусоидального профиля (6) на расстояние $x = 360$ м (фиг. 4а–в) и 480 м (фиг. 4г–е) при числах Куранта 0.1, 0.5 и 0.9. Численные решения на фиг. 4 представлены точками, а точное решение — пунктирной линией. Также, как и для результатов пересчета “ступеньки” (фиг. 3), для схем Годунова и CABARET наблюдается большее влияние числа Куранта на точность решения по сравнению со схемой WENO. Для двух рассмотренных случаев при числе Куранта 0.5 численные решения WENO и CABARET лучше разрешают слабые разрывы решения и волны разрежения.

3.2. Решение задачи Коши для уравнения переноса (1) с коэффициентом C, определяемым параметрами среды

Для второго тестового примера рассмотрим задачу о нелинейном распространении волны возмущенного давления в однородной атмосфере. В качестве начального профиля возмущенного давления p_0 брались волны с сигнатурой N-образной формы и с сигнатурой более сложной формы с ударными волнами и волнами разрежения.



Фиг. 4. Область головной ударной волны профиля численного решения задачи Коши (4), (6) при распространении на $x = 360$ м, полученного по схемам WENO (а), CABARET (б) и Годунова (в) при различных числах Куранта и при распространении на $x = 480$ м (г)–(е) аналогично.



Фиг. 5. Сравнение аналитического и численных решений задачи о распространении с высоты 15 000 м до 264 м начального профиля (8) с шагом $\Delta\tau = 0.0005$ (а), (б) — область головной ударной волны профиля решения, (в) — абсолютная погрешность численных решений, (г) — абсолютная погрешность численных решений в области головной ударной волны.

Рассмотрим численное решение задачи Коши для уравнения переноса (1):

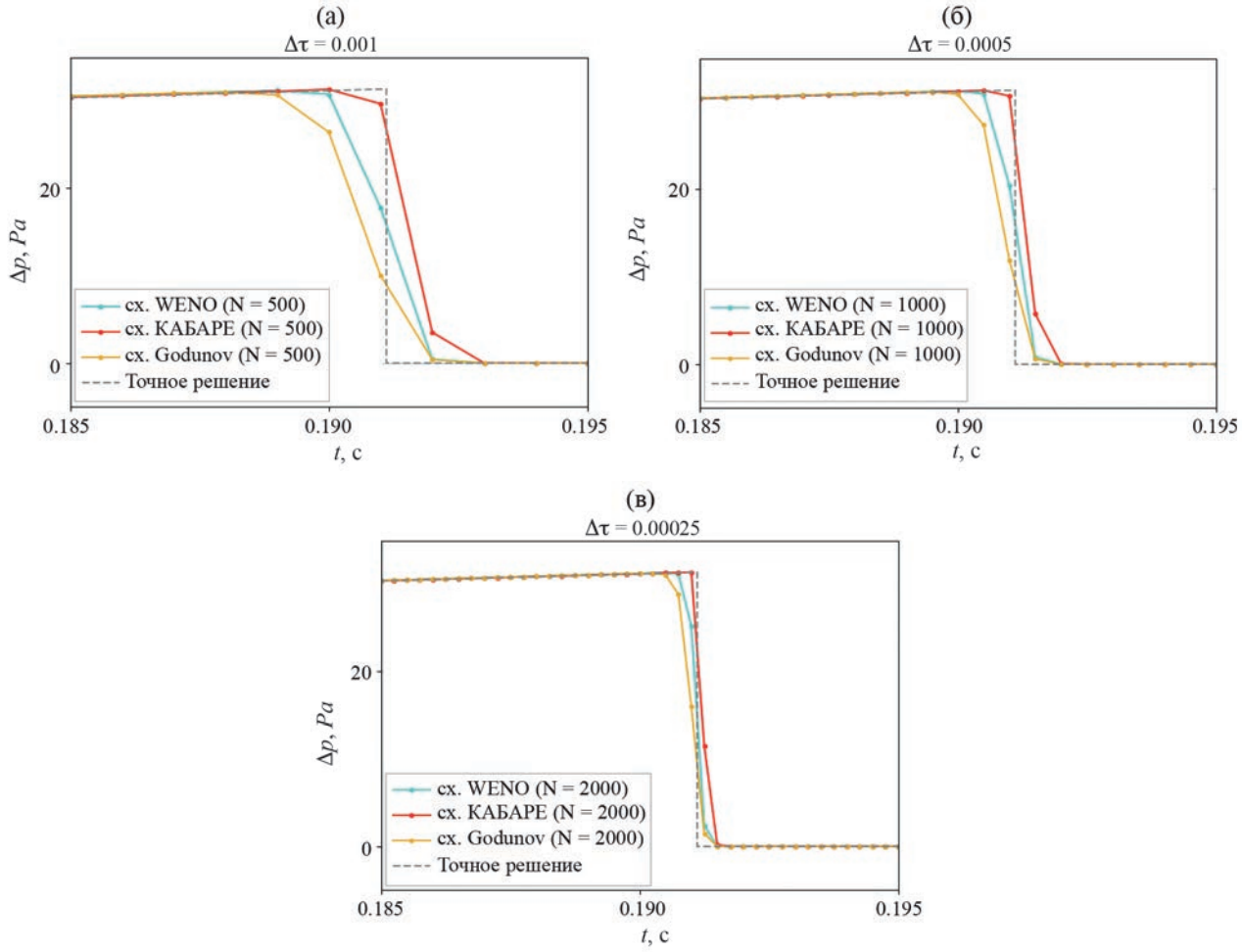
$$\frac{\partial p}{\partial x} + C\frac{\partial p^2}{\partial t} = 0, \quad p(t,0) = p_0(t), \tag{7}$$

где коэффициент скорости распространения C определяется плотностью окружающей среды ρ_0 и скоростью звука c_0 из соотношения (2). Данные параметры среды принимают значения $\rho_0 = 0.1515$ кг/м^3 и $c_0 = 297.8$ м/с на начальной высоте распространения 15 000 м из данных о стандартной атмосфере США (1976 г.) (см. [19]). При таких значениях параметров среды коэффициент C принимает значение $C = 1.5 \times 10^{-7}$. Распространение сигнала происходит с начальной высоты 15 000 м и до высоты в 264 м в перпендикулярном направлении к поверхности земли. Длина траектории распространения начального профиля волны составляет 14 736 м.

3.2.1. Распространение N-волны. Рассмотрим задачу Коши для уравнения переноса (1), полагая в качестве профиля начальных данных N-волну. N-волна — это простая, но реалистичная модель волны звукового удара. Ее большим преимуществом является возможность нахождения аналитического решения задачи Коши из условия Ренкина-Гюгонио для локализации скачков решения.

На фиг. 5а приведены аналитическое и численные решения уравнения переноса (7) с начальным профилем N-волны следующего вида:

$$p_0(t) = \begin{cases} \widehat{p_0}\dfrac{t}{t_0}, & -t_0 < t < t_0, \\ 0 \quad \text{иначе.} \end{cases} \tag{8}$$

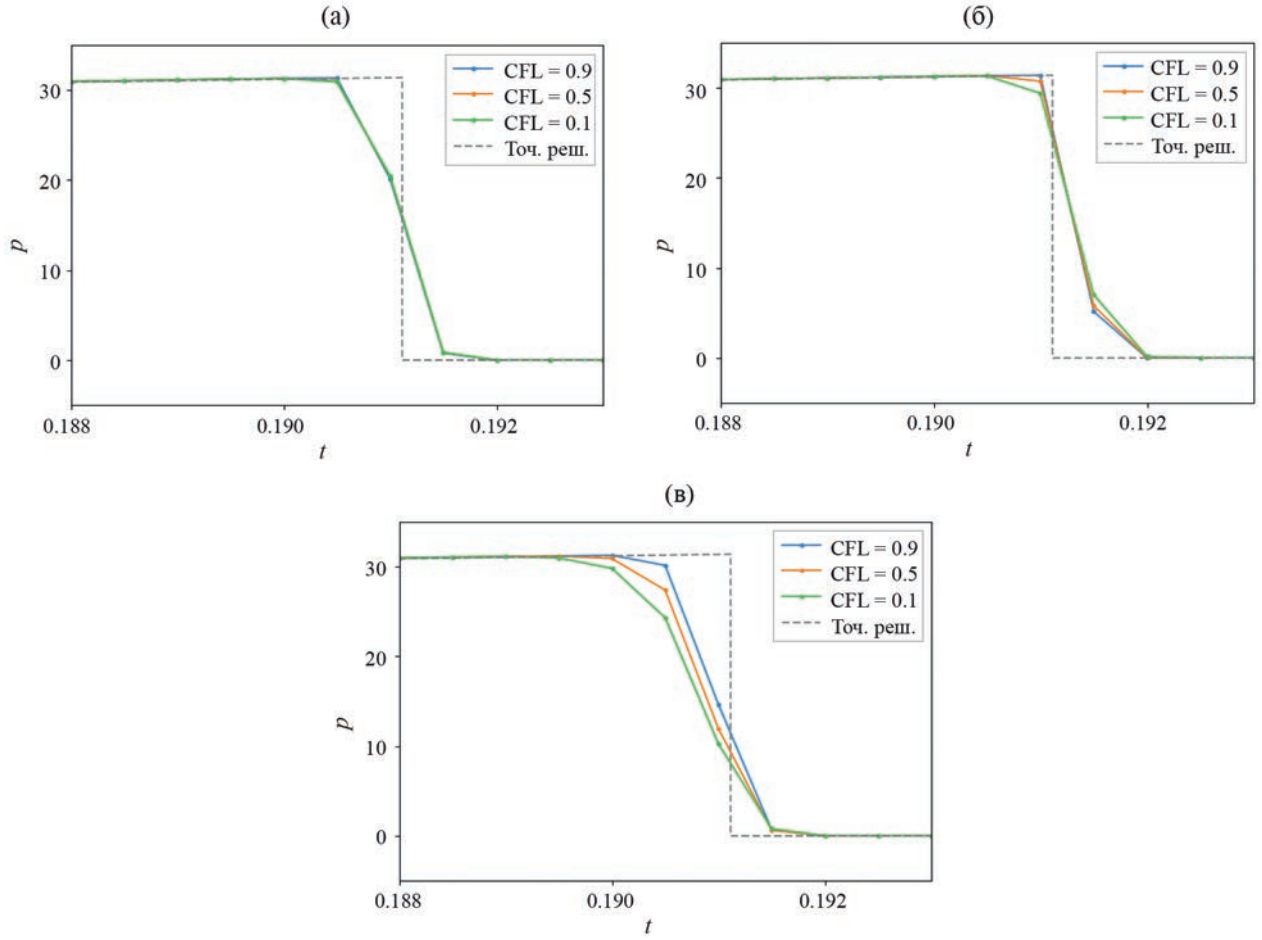


Фиг. 6. Область головной ударной волны аналитического и численного решений при распространении начального профиля (8) с высоты 15 000 м до 264 м при шаге $\Delta\tau = 0.001$ (а), $\Delta\tau = 0.0005$ (б) и $\Delta\tau = 0.00025$ (в).

Начальный профиль N-волны с $t_0 = 0.1$ с и аналитическое решение задачи о распространении начального профиля показаны пунктирной линией, а сплошными линиями — численные результаты, полученные с помощью схем CABARET, WENO и Годунова. Расчеты проводились на разностной сетке (3) из 1000 узлов с временным шагом $\Delta\tau = 0.0005$ и при коэффициенте запаса $z = 0.5$ в условии устойчивости. В ходе распространения вдоль траектории начального профиля N-волны его протяженность увеличивается, а интенсивность скачков уменьшается.

На фиг. 5а наблюдается хорошее согласование результатов вычислений с аналитическим решением задачи Коши (7), (8). Численные схемы показывают хорошую точность локализации ударных волн и сохраняют монотонность численного решения. Абсолютные погрешности численных результатов представлены на фиг. 5в,г точками. Значения погрешностей для всех схем малы и близки к нулю на гладких участках решения. В областях разрыва погрешности принимают большие значения и отличаются для различных схем.

На фиг. 6 представлены аналитические и численные решения задачи Коши (7), (8) при сильном приближении к области с головной ударной волной. Аналитическое решение задачи показано пунктирной линией, а сплошными линиями с точками — численные результаты, полученные с помощью схем CABARET, WENO и Годунова. Сравнивается точность разрешения разрыва численными схемами при расчете с различным шагом дискретизации $\Delta\tau = 0.001$, $\Delta\tau = 0.0005$ и $\Delta\tau = 0.00025$ (сетки из 500, 1000 и 2000 узлов соответственно) при коэффициенте запаса $z = 0.5$ в условии устойчивости. Наблюдается слабое разрешение ударных волн схемой Годунова по сравнению с схемой WENO и CABARET, как и в ранее рассмотренных случаях. Для задач, связанных с распространением волн малой амплитуды на дальние расстояния, данная потеря в точности и размазывание скачка на большее количество узлов сетки может значительно повлиять на достоверность предсказания конечного результата распространения волны. Численные результаты, полученные с помощью схем CABARET и



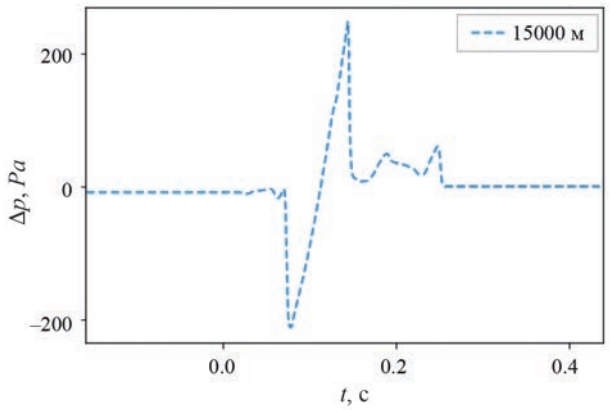
Фиг. 7. Область головной ударной волны численного решения задачи Коши (7), (8), полученного по схемам WENO (а), CABARET (б) и Годунова (в) при различных числах Куранта.

WENO, показывают незначительные различия. Точность разрешения разрыва и время нарастания скачка имеют достаточную степень соответствия. Результаты схемы CABARET по сравнению с результатами схемы WENO имеют большую крутизну фронта на скачках решения и большую их интенсивность.

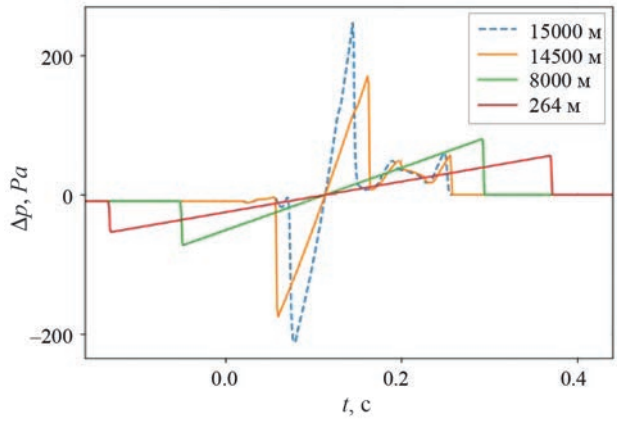
На фиг. 7 представлены численные результаты пересчета (головная ударная волна) при распространении начального профиля (8) N-волны с высоты 15 000 м до 264 м с шагом сетки $\Delta\tau = 0.0005$ и при числах Куранта 0.1, 0.5 и 0.9. Численные решения на фиг. 7 представлены точками, а точное решение — пунктирной линией. Также, как и для результатов пересчета в первом тестовом примере, наблюдается большее влияние числа Куранта на точность решения схемами Годунова и CABARET по сравнению со схемой WENO. Численные решения, полученные по схеме CABARET при числах Куранта 0.5 и 0.9, показывают более хорошее согласование с точным решением и более крутой фронт разрыва решения. Как и ранее наблюдалось в результатах первого тестового примера, для числа Куранта 0.9 схема Годунова показывает более точное разрешение разрывов и скачков решения.

3.2.2. Распространение профиля сложной структуры Рассмотрим задачу Коши для уравнения переноса (7), полагая в качестве профиля начальных данных профиль сложной структуры с ударными волнами и волнами разрежения (фиг. 8). Данный профиль соответствует значениям возмущенного давления, зафиксированного в ближней зоне возмущенного течения у модели самолета при ее обтекании набегающим сверхзвуковым потоком воздуха с числом Маха $M = 2$. Модель представляет собой компоновку “корпус-крыло” сверхзвукового пассажирского самолета с массой $G = 70$ т и длиной $L = 40$ м.

На фиг. 9 представлены результаты пересчета профиля избыточного давления на различное удаление от начальной высоты. Данные результаты были получены решением задачи Коши (7) с помощью схемы WENO. Расчеты проводились на разностной сетке (3) в 677 узлов с временным шагом $\Delta\tau = 0.001$ и при коэффициенте



Фиг. 8. Начальный профиль избыточного давления.

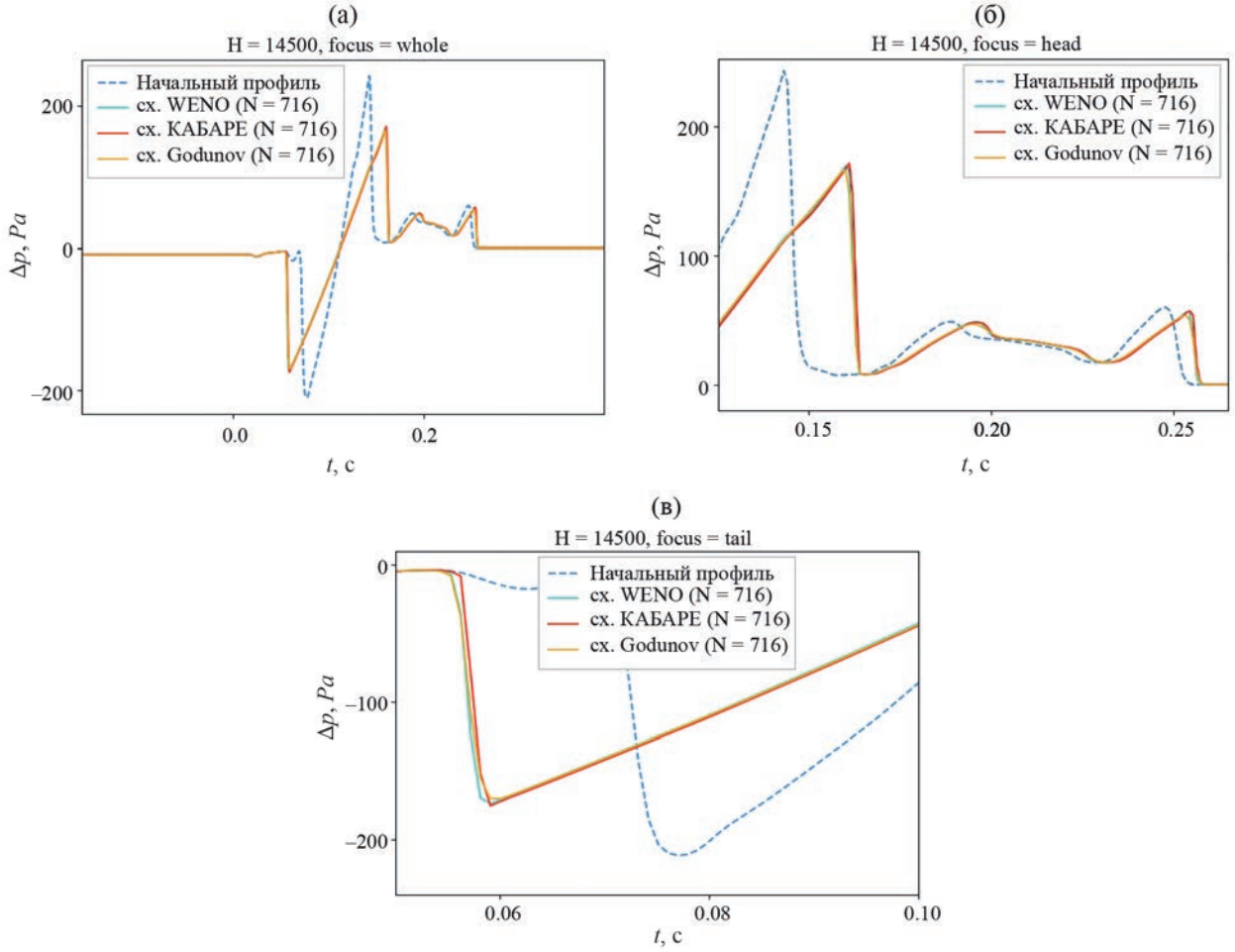


Фиг. 9. Трансформация профиля сложной структуры при его распространении с начальной высоты 15 000 м до 264 м.

запаса $z = 0.5$ в условии устойчивости. Под действием нелинейных эффектов при распространении профиля возмущенного давления скачки большей интенсивности движутся с большей скоростью относительно скачков меньшей интенсивности. В результате “догоняет” его и, не достигая высоты 8000 метров, профиль избыточного давления трансформируется в N-волну, которая далее растягивается в стороны с потерей интенсивности скачков.

На фиг. 10–12 представлены результаты расчета распространения начального профиля с высоты 15 000 м до высоты 14 500 м, 8000 м и 264 м соответственно. Начальный профиль показан пунктирной линией, сплошными линиями — численные результаты решения схемами WENO, CABARET и Годунова рассматриваемой задачи Коши (7). Результаты вычислений по схемам WENO и CABARET демонстрируют хорошее согласование друг с другом. В случае рассматриваемого начального профиля более сложной структуры также сохраняется монотонность численных решений. В результатах схемы Годунова для рассматриваемых случаев наблюдается более сглаженное разрешение разрывов и меньшая интенсивность скачков по сравнению с результатами схем WENO и CABARET. Результаты CABARET и WENO слабо отличаются друг от друга. Результаты схемы CABARET по сравнению с результатами схемы WENO, как и в ранее рассмотренном примере, имеют большую крутизну фронта на скачках решения и большую их интенсивность: на фиг. 12б, фиксирующем головной скачок, результат CABARET ушел вперед относительно результатов WENO.

Вычислительное время для решения рассматриваемого тестового примера по схеме WENO заняло 23.89 с, 0.55 с – по схеме CABARET и 0.85 с – по схеме Годунова. Использовался ноутбук HP Pavilion 13 with Core-i5-8265U. Расчет по схеме CABARET занимает значительно меньшее время по сравнению с временем расчета по схеме WENO. Данное соотношение между расчетными временами наблюдается и для других рассмотренных в работе тестовых примеров. Для данного тестового примера расчетное время WENO превышает расчетное время CABARET более, чем в 43 раза. Учитываяблизкие по точности численные результаты схем WENO и CABARET



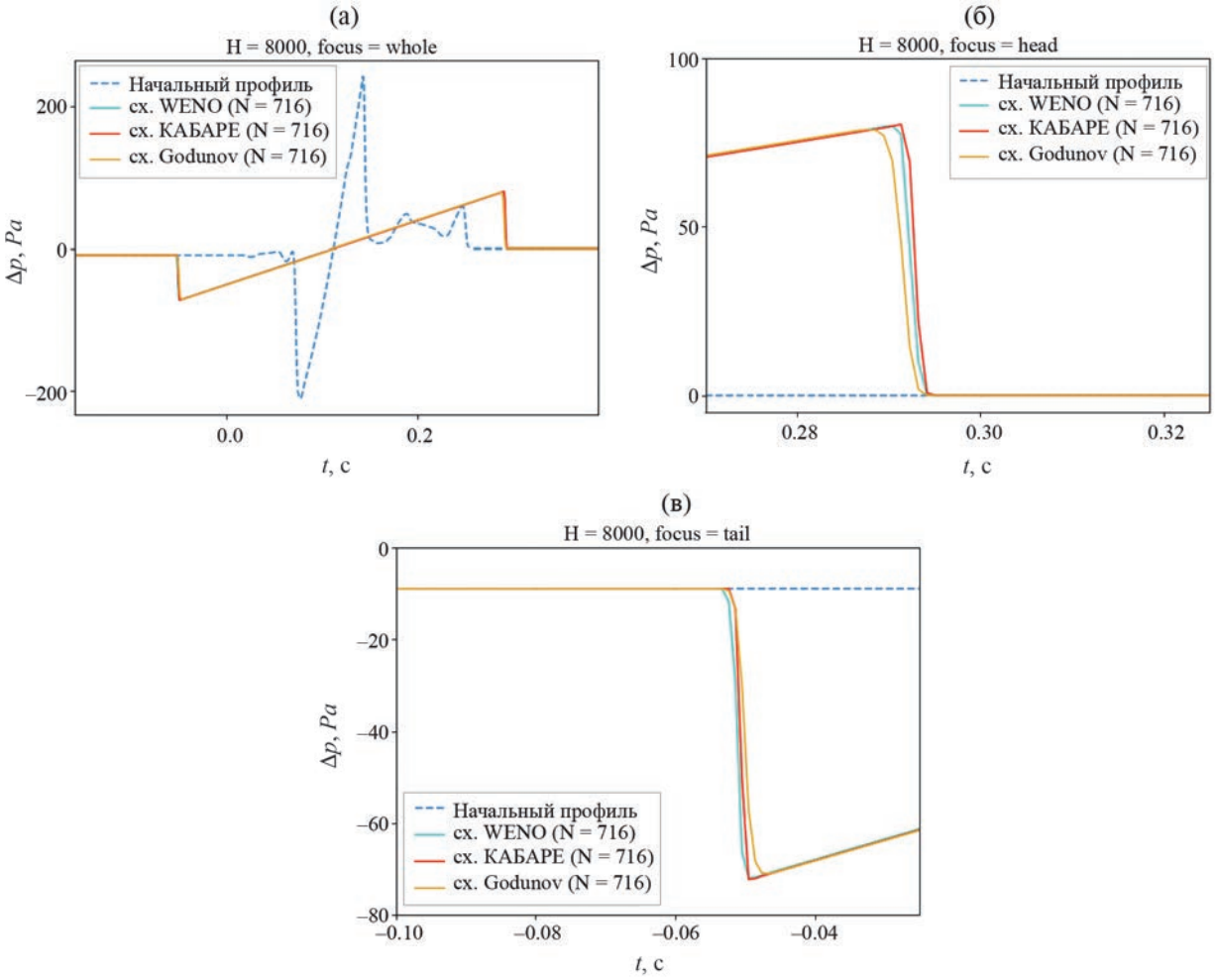
Фиг. 10. Сравнение численных решений задачи Коши (7) при распространении начального профиля сложной структуры с высоты 15 000 м до 14 500 м (а), (б) и (в) — области численных решений с головными скачками и хвостовым скачком соответственно.

для тестовых примеров, рассмотренных в п. 3.2.1 и 3.2.2, схема CABARET – более оптимальный выбор численной схемы по сравнению со схемой WENO для решения задачи распространения волны возмущенного давления на дальние расстояния в атмосфере. Однако нужно отметить необходимость предварительной подготовки согласованных наборов начальных данных для консервативных и потоковых переменных схемы CABARET.

ЗАКЛЮЧЕНИЕ

В настоящей работе рассматривались численные схемы сквозного счета CABARET и WENO для решения нелинейного уравнение переноса, моделирующего перенос в пространстве волны возмущенного давления со скоростью распространения, определяемой параметрами среды и интенсивностью самих возмущений. Также для оценки влияния выбора численной схемы для рассматриваемой задачи применялась явная схема Годунова. Таким образом, задача Коши для рассматриваемого уравнения переноса решалась с применением трех численных схем: WENO, CABARET и Годунова.

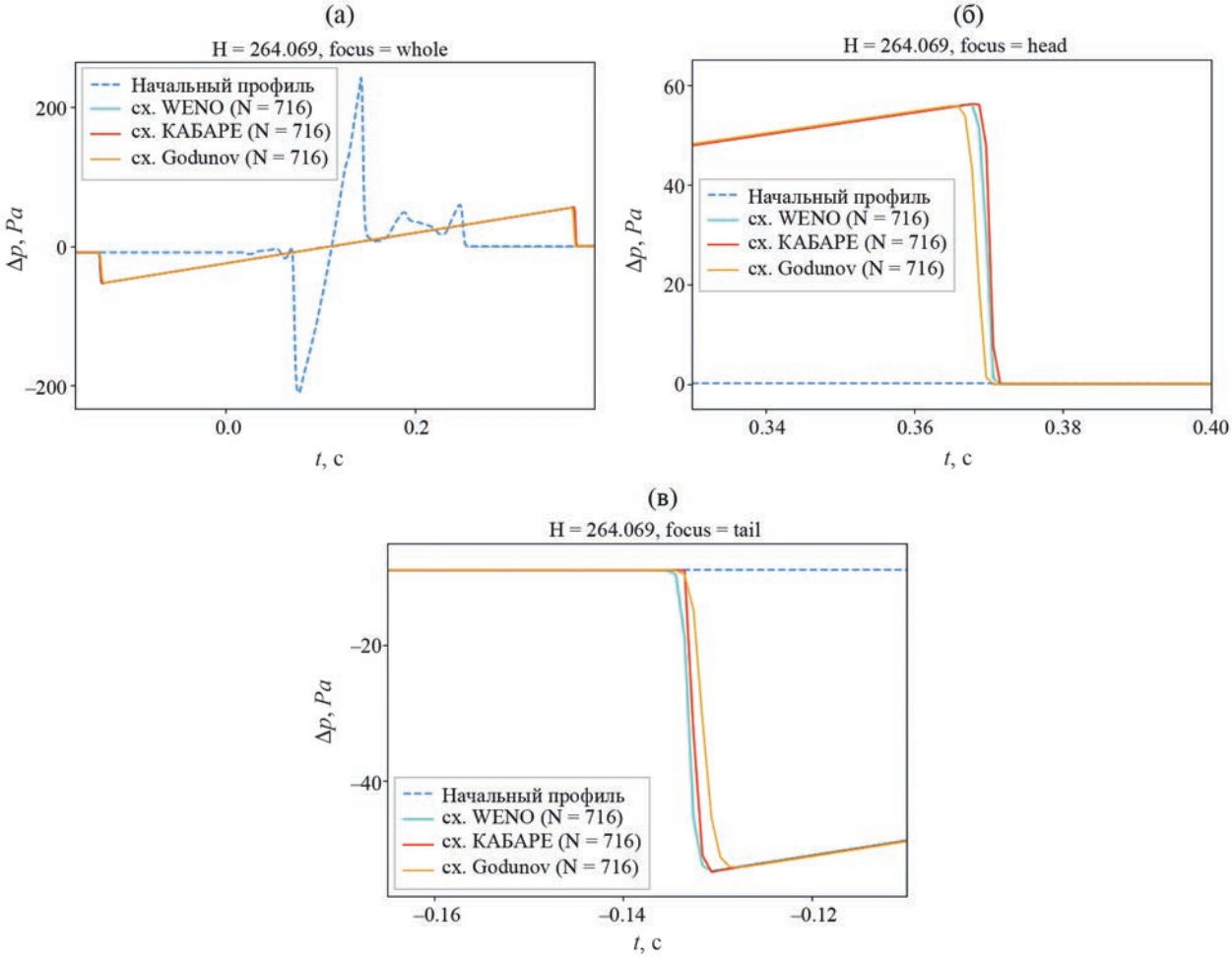
Решались тестовые примеры с распространением плоских волн различной сложности. Тестовый пример с фиксированным коэффициентом скорости распространения (уравнение Хопфа) позволяет верифицировать результаты расчетов реализованных численных схем. По результатам тестовых примеров для распространения начального профиля с фиксированным и нефиксированным коэффициентами скорости отмечалась потеря интенсивности скачков и значительное их сглаживание схемой Годунова. Схемы WENO и CABARET показали более точные результаты, сохранение монотонности решений и достоверность разрешения разрывов. В случае распространения начального профиля с фиксированным коэффициентом скорости схема WENO по срав-



Фиг. 11. Сравнение численных решений задачи Коши (7) при распространении начального профиля сложной структуры с высоты 15 000 м до 8000 м (а), (б) и (в) – области численных решений с головным и хвостовым скачками соответственно.

нению со схемой CABARET показывает более точные результаты и меньшее размазывание скачков. В случае распространения начального профиля с фиксированным коэффициентом скорости, схема WENO демонстрирует более высокую точность результатов и меньшее размазывание скачков по сравнению со схемой CABARET. Однако для задачи решения в протяженной области моделирования нелинейного уравнения переноса с коэффициентом скорости, зависящим от параметров среды, численные результаты, полученные с использованием схем CABARET и WENO, демонстрируют незначительные различия, что вполне ожидаемо, поскольку обе эти схемы относятся к методам с нелинейной коррекцией переменных потока. По сравнению с результатами схемы WENO результаты схемы CABARET характеризуются более крутым фронтом разрывов решения и более высокой их интенсивностью. Расчетное время численного метода со схемой CABARET значительно меньше, чем у метода со схемой WENO. Схема CABARET более предпочтительна для решения задачи распространения волны возмущенного давления на дальние расстояния в атмосфере.

Рассматриваемое в работе уравнение переноса описывает нелинейное распространение в стационарной атмосфере волн малой амплитуды и применяется в моделях распространения волн звукового удара, генерируемых при полете сверхзвукового пассажирского самолета в атмосфере. Поиск более точных и эффективных методов решения рассматриваемого в работе уравнения переноса необходим для разработки более достоверных методик прогнозирования звукового удара. Что позволяет более эффективно решать задачу минимизации звукового удара и многопараметрической задачи разработки компоновки “тихого” сверхзвукового пассажирского самолета нового поколения.



Фиг. 12. Сравнение численных решений задачи Коши (7) при распространении начального профиля сложной структуры с высоты 15 000 м до 264 м (а), (б) и (в) — области численных решений с головным и хвостовым скачками соответственно.

СПИСОК ЛИТЕРАТУРЫ

1. *Чернышев С.Л.* Звуковой удар. М.: Наука, 2011.
2. *Cleveland R.O.* Propagation of sonic booms through a real, stratified atmosphere : PhD thesis. Univer. Texas at Austin, 1995.
3. *Blackstock D.T.* Nonlinear acoustics (theoretical) // Am. Inst. Phys. Handbook. 1972. V. 3.
4. *Rallabhandi S.K.* Advanced sonic boom prediction using the augmented Burgers equation // J. Aircraft. 2011. V. 48. № 4. P. 1245–1253.
5. *Qiao J.L. et al.* Development of sonic boom prediction code for supersonic transports Based on augmented Burgers equation // AIAA Aviation 2019 Forum. 2019. P. 3571.
6. *Kanamori M. et al.* Comparison of simulated sonic boom in stratified atmosphere with flight test measurements // AIAA J. 2018. V. 56. № 7. P. 2743–2755.
7. *Lonzaga J.B.* Recent Enhancements to NASA PCBoom Sonic Boom Propagation Code // AIAA Aviation 2019 Forum. 2019. P. 3386.
8. *Pilon A.R.* Spectrally accurate prediction of sonic boom signals // AIAA J. 2007. V. 45. № 9. P. 2149–2156.
9. *Jianling Q. et al.* Far-field sonic boom prediction considering atmospheric turbulence effects: An improved approach // Chin. J. Aeronaut. 2022. V. 35. № 9. P. 208–225.

10. *Thomas C.L.* Extrapolation of wind-tunnel sonic boom signatures without use of a Whitham F-function // NASA SP-255. 1970. P. 205–217.

11. *Холодов А.С.* Численные методы решения уравнений и систем гиперболического типа // Энциклопедия низкотемпературной плазмы (сер. Б). 2008. Т. 1. Ч. 2. С. 141–174.

12. *Куликовский А.Г., Погорелов Н.В., Семёнов А.Ю.* Математические вопросы численного решения гиперболических систем уравнений. М.: Физматлит, 2012.

13. *Самарский А.А., Гулин А.В.* Численные методы математической физики: Учеб. пособие по прикл. математике. Науч. мир, 2003.

14. *Зюзина Н.А., Ковыркина О.А., Остапенко В.В.* О монотонности схемы CABARET, аппроксимирующей скалярный закон сохранения со знакопеременным характеристическим полем и выпуклой функцией потоков // Матем. моделирование. 2018. Т. 30. № 5. С. 76–98.

15. *Головизнин В.М. и др.* Новые алгоритмы вычислительной гидродинамики для многопроцессорных вычислительных комплексов // М.: Изд-во Моск. ун-та, 2013.

16. *Jiang G.S., Shu C.W.* Efficient implementation of weighted ENO schemes // J. Comput. Phys. 1996. V. 126. № 1. P. 202–228.

17. *Courant R., Friedrichs K., Lewy H.* On the partial difference equations of mathematical physics // IBM J. Res. and Development. 1967. V. 11. № 2. P. 215–234.

18. *Pierce A.D., Acoustics A.* Introduction to its physical principles and applications // Acoustic. Soc. Am. and Am. Inst. Phys. 1981. P. 122.

19. United States Committee on Extension to the Standard Atmosphere et al. US standard atmosphere. — National Oceanic and Amospheric [sic] Administration, National Aeronautics and Space Administration, US Air Force, 1962.

APPLICATION OF CABARET AND WENO SCHEMES FOR SOLVING THE NONLINEAR TRANSPORT EQUATION IN THE MODELING OF SOUND WAVE PROPAGATION IN THE ATMOSPHERE

P. A. Mishchenko[a,*], T. A. Himon[a], V. A. Kolotilov[a,b], A. N. Kudryavtseva[a]

[a] *Khristianovich Institute of Theoretical and Applied Mechanics SB RAS, Institutskaya st., 4/1, Novosibirsk, 1630090, Russia*

[b] *M. A. Lavrentiev Institute of Hydrodynamics SB RAS, Akademik Lavrentiev Ave., 15, Novosibirsk, 630090, Russia*

**e-mail: mischenko.polina.16@gmail.com*

Received 21 September, 2023
Revised 20 December, 2023
Accepted 06 February, 2024

Abstract. The most convenient model for describing the phenomenon of shock wave propagation in the atmosphere is the extended Burgers equation. This work investigates the influence of the numerical scheme on the results of solving the equation, which accounts for the nonlinear nature of shock wave propagation in the atmosphere. This equation is a key component of the extended Burgers equation and defines the transformation of the perturbed pressure profile during its propagation. Two numerical schemes were applied for the solution: CABARET and WENO, which are quasi-monotonic finite difference schemes that allow for solutions without significant numerical oscillations. An analysis of the applicability of these schemes for solving the considered problem was conducted.

Keywords: shock wave, nonlinear transport equation, small amplitude wave propagation, CABARET scheme, WENO scheme.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 867–880

УДК 519.6

К ВОПРОСУ ОБ ОДНОВРЕМЕННОМ ОПРЕДЕЛЕНИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ЭКВИВАЛЕНТНЫХ ПО ВНЕШНЕМУ ПОЛЮ ИСТОЧНИКОВ И СПЕКТРА ПОЛЕЗНОГО СИГНАЛА[1)]

© 2024 г. И. Э. Степанова[1,*], Д. В. Лукьяненко[2], И. И. Колотов[2], А. В. Шепетилов[2], А. Г. Ягола[2], И. А. Керимов[1], А. Н. Левашов[2]

[1] 123242 Москва, ул. Б.Грузинская, 10, ГБУ Ин-т физики Земли РАН, Россия
[2] 119991 Москва, Ленинские горы, МГУ, Россия
**e-mail: tet@ifz.ru*

Поступила в редакцию 28.06.2023 г.
Переработанный вариант 28.06.2023 г.
Принята к публикации 14.01.2024 г.

В статье исследуется возможность одновременного восстановления эквивалентных по внешнему полю источников и спектральных характеристик полезного сигнала. Приводятся примеры вариационных постановок для различных версий метода линейных интегральных представлений, а также формулируется задача о нахождении плотности распределения гравитирующих или магнитных масс на нескольких горизонтальных плоскостях и преобразования Фурье элемента аномального поля по известным в точках некоторой сети наблюдений значениям сигнала, осложненного помехой. Библ.17.

Ключевые слова: системы линейных алгебраических уравнений, интегральные представления, формула суммирования Пуассона.

DOI: 10.31857/S0044466924050147, **EDN**: YCQUUV

ВВЕДЕНИЕ

В рамках аппроксимационного подхода к решению обратных линейных и нелинейных задач геофизики, геодезии и геоморфологии [1, 2] практически все постановки по определению параметров геологической среды можно редуцировать к решению систем линейных (в некоторых случаях — и нелинейных) систем алгебраических уравнений. Как было отмечено в работе [3], основным методом, позволяющим реализовать такой подход, является метод интегральных представлений.

В статье исследуется возможность одновременного определения плотностей эквивалентных по внешнему гравитационному или магнитному полю источников и спектров полезных сигналов. Приводится описание методики нахождения численного решения обратной задачи по поиску распределений эквивалентных по внешнему полю носителей масс как в обычном (двумерном или трехмерном пространстве), так и в двойственном ему пространстве частот спектра полезного сигнала — некоторой компоненты гравитационного или магнитного поля.

Основы теории локальных и региональных S-аппроксимаций, а также локальных F- и R-аппроксимаций как примеров применения метода линейных интегральных представлений изложены в целой серии работ авторов [4–8].

В рамках трехмерного метода S-аппроксимаций известная компонента гравитационного поля аппроксимируется суммой простого и двойного слоев, распределенных на некоторой совокупности областей (в локальном случае ими являются горизонтальные плоскости, в региональном — сферы или сфероиды). Но, как уже подчеркивалось в работе [3], источники поля масс могут иметь любую форму (конечно, при условии, что выполнены требования гладкости для границы области, занятой массами) и любую размерность, меньшую или равную размерности рассматриваемого пространства. Плоскости в локальном варианте и сферы (или эллипсоиды вращения) — в региональном и глобальном) выбирались нами в качестве носителей масс исключительно

[1)] Работа выполнена при финансовой поддержке РНФ (грант № 23-41-00002).

из-за простоты выражений для элементов матрицы системы линейных алгебраических уравнений, к которой редуцировалась обратная задача.

В методе F-аппроксимаций элементы аномальных потенциальных полей представляются интегралом Фурье, а R-аппроксимации получаются при так называемом лучевом преобразовании.

Методы F-, R- и S-аппроксимаций позволяют получить решение, с помощью которого можно эффективно строить линейные трансформанты поля, а также использовать их в качестве нулевого приближения для решения нелинейной обратной задачи по локализации источников. Для того чтобы изложить основные моменты новой методики одновременного определения распределений масс в различных пространствах, напомним читателю, как строятся R- F- и S-аппроксимации элементов потенциальных полей (а также функции, описывающей рельеф поверхности планеты) в трехмерном декартовом пространстве. Поскольку все три типа аппроксимаций связаны между собой [3], то возникает естественный вопрос: а нельзя ли по решению вариационной задачи для одного типа представлений определить другие важнейшие характеристики сигнала: например, по восстановленной плотности распределения эквивалентных по внешнему гравитационному или магнитному полю источников масс найти спектр элемента поля или интегральную плотность масс вдоль фиксированного направления (луча)? Оказывается, можно. И при этом еще раз решать системы линейных алгебраических уравнений не нужно: элементы матрицы системы для постановки одного типа аппроксимаций в точности совпадают с элементами матрицы для другой.

1. R-АППРОКСИМАЦИЯ ЭЛЕМЕНТОВ АНОМАЛЬНЫХ ПОТЕНЦИАЛЬНЫХ ПОЛЕЙ

В [8] показано, что для функции $f(x) \in S(\mathbb{R}^n)$, где $\mathbb{R}^n$ — пространство быстро убывающих на бесконечности непрерывно дифференцируемых функций (точнее говоря, для непрерывно дифференцируемых функций, имеющих порядок убывания $O\big(1+\sum_{i=1}^n x_i^2\big)^{-1}$) — пространство Шварца — существует преобразование Радона:

$$\hat{f}(\omega, p) = \int\limits_{(\omega,x)=p} f(x)dm(x), \tag{1}$$

где ω — единичный вектор, $dm(x)$ — мера на прямой $(\omega, x) = p$.

В двумерном случае формула (1) принимает вид:

$$\hat{f}(\omega, p) = \int\limits_{-\infty}^{\infty} f(-t\sin s + x_1\cos\varphi, t\cos s + x_2\sin\varphi)ds, \quad \omega = (\cos\omega, \sin\omega), \quad x = (x_1, x_2). \tag{2}$$

Запишем формулу обращения преобразования Радона

$$f(x) = \gamma L_x^{(n-1/2)} \left(\int\limits_{S^{n-1}} \hat{f}\big(\omega, (x,\omega)\big)d\omega \right),$$

где постоянная $\gamma = (2\pi i)^{1-n}/2$. Отметим, что функция $f_\omega(x) = \hat{f}\big(\omega, (x,\omega)\big)d\omega$ есть плоская волна в направлении ω, другими словами, она постоянна на каждой гиперплоскости, перпендикулярной ω. Здесь $L_n^{(n-1)/2}$ — оператор Лапласа порядка $(n-1)/2$. В случае нечетных n она принимает вид

$$f(x) = \frac{1}{2}(2\pi)^{-n}(-i)^{n-1} \int\limits_{S^{n-1}} \left\{ \frac{d^{n-1}}{dp^{n-1}} \hat{f}(\omega, p) \right\}_{p=(x,\omega)} d\omega.$$

Покажем, как связаны преобразование Радона и n-мерное преобразование Фурье:

$$\tilde{f}(u) = \int\limits_{R^n} f(x)e^{-i(x,u)}dx, \quad u \in R^n. \tag{3}$$

Действительно, если $s \in R$ и ω — единичный вектор, то

$$\tilde{f}(s\omega) = \int\limits_{-\infty}^{\infty} dr \int\limits_{(x,\omega)=r} f(x)e^{-is(x,\omega)}dm(x), \tag{4}$$

следовательно,

$$\tilde{f}(s\omega) = \int\limits_{-\infty}^{\infty} \hat{f}(\omega, r)e^{-isr}dr. \tag{5}$$

Из (4) и (5) следует, что n-мерное преобразование Фурье есть композиция одномерного преобразования Фурье и преобразования Радона. Введем среднее функции f по сферам с центром в фиксированной точке :

$$F(x, r) = \frac{1}{4\pi} \int\limits_{|\omega|=1} f(x + r\omega)d\omega. \tag{6}$$

Пусть

$$\hat{F}(x, p) = \frac{1}{4\pi} \int\limits_{|\omega|=1} \Re f(\omega, p + \langle \omega, x\rangle)d\omega$$

среднее $\Re f$ по плоскостям, равноотстоящим от точки x, т.е. по плоскостям, касающимся сферы радиуса p с центром в точке x. Тогда

$$f(x) = F(x, 0) = -\frac{1}{2\pi}\hat{F}''_p(\omega, 0). \tag{7}$$

Аналогичным образом определяется среднее функции $\Re f$ по прямым, касающимся окружности с центром в точке $x = (x_1, x_2)$ радиуса p на плоскости:

$$\hat{F}(x, p) = -\frac{1}{2\pi} \int\limits_{0}^{2\pi} \Re f(\varphi, p + x_1 \cos\varphi + x_2 \sin\varphi)d\varphi. \tag{8}$$

Теперь функцию f можно восстановить, пользуясь следующим представлением:

$$f(x) = -\frac{1}{\pi} \int\limits_{0}^{\infty} \frac{\hat{F}'_p(x, p)}{p}dp, \tag{9}$$

где последний интеграл понимается в смысле главного значения.

Напомним основную формулу теории гармонических функций для полупространства, ограниченного плоскостью $x_3 = 0$ (далее упоминаемой как плоскость "*П*") [9]:

$$V(M) = \iint\limits_{-\infty}^{+\infty} \frac{\rho_1(\xi_1, \xi_2)d\xi_1 d\xi_2}{\sqrt{(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + x_3^2}} + \iint\limits_{-\infty}^{+\infty} \frac{\rho_2(\xi_1, \xi_2)x_3 d\xi_1 d\xi_2}{\left[\sqrt{(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + x_3^2}\right]^3}, \tag{10}$$

$$M = (x_1, x_2, x_3), \quad \xi = (\xi_1, \xi_2, \xi_3).$$

Мы выбрали систему координат так, чтобы плоскость простого и двойного слоев задавалась уравнением $x_3 = 0$. Тогда производная по x_3 потенциала V, взятая с обратным знаком, будет иметь вид

$$-\frac{\partial V}{\partial x_3}(M) = \iint\limits_{-\infty}^{+\infty} \frac{\rho_1(\hat{\xi})x_3 d\hat{\xi}}{\left[\sqrt{(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + x_3^2}\right]^3} + \iint\limits_{-\infty}^{+\infty} \frac{\rho_2(\hat{\xi})\left(2x_3^2 - (x_1 - \xi_1)^2 - (x_2 - \xi_2)^2\right)^2 d\hat{\xi}}{\left[\sqrt{(x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + x_3^2}\right]^5}, \tag{11}$$

$$M = (x_1, x_2, x_3), \quad \xi = (\xi_1, \xi_2, \xi_3).$$

Функции ρ_1, ρ_2 неизвестны. Пусть компоненты поля заданы в конечном множестве точек $M_i = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$, $i = 1, 2, \ldots, N$. Обозначим подынтегральную функцию в первом слагаемом в (11) в точке M_i через $Q_1^{(i)}$, а во втором слагаемом — через $Q_2^{(i)}$. Тогда получим

$$-\frac{\partial V(M_i)}{\partial x_3} \equiv f_i = \iint\limits_{-\infty}^{+\infty} \left(\rho_1(\hat{\xi})Q_1^{(i)}(\hat{\xi}) + \rho_2(\hat{\xi})Q_2^{(i)}(\hat{\xi})\right)d\hat{\xi}, \quad i = 1, 2, \ldots, N. \tag{12}$$

Здесь необходимо отметить, что формулы (10)–(12) являются основными при построении S-аппроксимаций искомого элемента аномального потенциального поля в локальном варианте, когда сферичностью Земли или другой планеты можно пренебречь [1].

Применим к обеим частям равенства (12) преобразование Радона:

$$\hat{V}_{x_3}(\omega,p)=\int\limits_{-\infty}^{+\infty}\left[\hat{\rho}_1(\omega,q)\cdot\hat{Q}_1^{(i)}(\omega,p-q)+\hat{\rho}_2(\omega,q)\cdot\hat{Q}_2^{(i)}(\omega,p-q)\right]dq. \tag{13}$$

Функции $\hat{Q}_1^{(i)}$, $\hat{Q}_2^{(i)}$ можно найти аналитически:

$$\begin{aligned}
&\int\limits_{-\infty}^{+\infty}\frac{x_3}{\left[\sqrt{(x_1+t\sin\varphi-p\cos\varphi)^2+(x_2+t\cos\varphi-p\sin\varphi)^2+x_3^2}\right]^3}dt=\\
&=\frac{2x_3}{x_3^2+p^2-2px_1\cos\varphi-2px_2\sin\varphi+(x_1\cos\varphi+x_2\sin\varphi)^2},\quad \omega=(\cos\varphi,\sin\varphi),\\
&\int\limits_{-\infty}^{+\infty}\frac{2x_3^2-\left((x_1+t\sin\varphi-p\cos\varphi)^2+(x_2+t\cos\varphi-p\sin\varphi)^2\right)}{\left[\sqrt{(x_1+t\sin\varphi-p\cos\varphi)^2+(x_2+t\cos\varphi-p\sin\varphi)^2+x_3^2}\right]^5}dt=\\
&=\frac{\partial}{\partial x_3}\left[\frac{2x_3}{x_3^2+p^2-2px_1\cos\varphi-2px_2\sin\varphi+(x_1\cos\varphi+x_2\sin\varphi)^2}\right],\quad \omega=(\cos\varphi,\sin\varphi),
\end{aligned} \tag{14}$$

где $\omega_1\xi+\omega_2\eta=p$ — прямая, по которой производится интегрирование. Если теперь записать для $-\frac{\partial V}{\partial x_3}(M_i)$ его выражение с помощью формулы обращения преобразования Радона, мы получим:

$$\begin{aligned}
-\frac{\partial V}{\partial x_3}(M_i)&=-\frac{1}{\pi}\int\limits_0^{+\infty}dp\int\limits_0^{2\pi}d\varphi\frac{1}{p}\left[\int\limits_{-\infty}^{+\infty}\left[\hat{\rho}_1(\omega,q)(\hat{Q}_1^{(i)})'_p(\omega,p-q)+\hat{\rho}_2(\omega,q)(\hat{Q}_2^{(i)})'_p(\omega,p-q)\right]dq\right]=\\
&=-\frac{1}{2\pi}\int\limits_0^{2\pi}d\varphi\int\limits_{-\infty}^{+\infty}\left\{\left[\int\limits_{-\infty}^{+\infty}dp\frac{(\hat{Q}_1^{(i)})'_p(\omega,p-q)}{p}\right]\rho_1(\omega,q)+\left[\int\limits_{-\infty}^{+\infty}dp\frac{(\hat{Q}_2^{(i)})'_p(\omega,p-q)}{p}\right]\rho_2(\omega,q)\right\}dq.
\end{aligned} \tag{15}$$

В (15) интеграл по переменной p понимается в смысле главного значения. Его можно вычислить явно:

$$\begin{aligned}
I_1(\omega,q)&=\int\limits_{-\infty}^{+\infty}\frac{(Q_1)'_p(\omega,p-q)dp}{p}=2\cdot\frac{x_3^2-(q-r\cos\varphi)^2}{\left(x_3^2+(q-r\cos\varphi)^2\right)^2},\\
I_2(\omega,q)&=\int\limits_{-\infty}^{+\infty}\frac{(Q_2)'_p(\omega,p-q)dp}{p}=8\cdot\frac{x_3(q-r\cos\varphi)^2}{\left(x_3^2+(q-r\cos\varphi)^2\right)^2}.
\end{aligned} \tag{16}$$

На практике компоненты поля бывают заданы с некоторой погрешностью, поэтому входной информацией являются значения $f_{i,\delta}$. С помощью решения вариационной задачи:

$$\begin{aligned}
\Omega(\rho)&=\int\limits_0^{+\infty}\int\limits_{-\infty}^{+\infty}dq\int\limits_0^{2\pi}\left(\hat{\rho}_1^2(\omega,q)+\hat{\rho}_2^2(\omega,q)\right)dpd\varphi=\min_{\rho},\\
f_{i,\delta}&=-\frac{1}{2\pi}\int\limits_0^{2\pi}d\varphi\int\limits_{-\infty}^{+\infty}\left\{\left[\int\limits_{-\infty}^{+\infty}\frac{(\hat{Q}_1^{(i)})'_p(\omega,p-q)dp}{p}\right]\rho_1(\omega,q)+\left[\int\limits_{-\infty}^{+\infty}\frac{(\hat{Q}_2^{(i)})'_p(\omega,p-q)dp}{p}\right]\rho_2(\omega,q)\right\}dq
\end{aligned} \tag{17}$$

получим, что искомые функции должны иметь вид [3, 8]:

$$\begin{gathered}
\hat{\rho}_1^{(a)}(\omega,q)=\tilde{\hat{\rho}}_1(\omega,q,\lambda),\quad \hat{\rho}_2^{(a)}(\omega,q)=\tilde{\hat{\rho}}_2(\omega,q,\lambda),\\
\tilde{\hat{\rho}}_1(\omega,q,\lambda)=-\frac{1}{2\pi}\sum_{i=1}^{N}\lambda_i\int\limits_{-\infty}^{+\infty}\frac{(\hat{Q}_1^{(i)})'_p(\omega,p-q)}{p}dp,\quad \tilde{\hat{\rho}}_2(\omega,q,\lambda)=-\frac{1}{2\pi}\sum_{i=1}^{N}\lambda_i\int\limits_{-\infty}^{+\infty}\frac{(\hat{Q}_2^{(i)})'_p(\omega,p-q)}{p}dp.
\end{gathered} \tag{18}$$

Таким образом, мы приходим к следующей системе линейных алгебраических уравнений (СЛАУ):

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{f}_\delta, \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N), \quad \mathbf{f}_\delta = (f_{1,\delta}, \ldots, f_{N,\delta}), \tag{19}$$

элементы матрицы которой в нашем случае имеют вид

$$a_{ij} = \frac{1}{4\pi^2} \int\limits_0^{2\pi} \int\limits_{-\infty}^{+\infty} \big\{ I_{1,i}(\omega, q) I_{1,j}(\omega, q) + I_{2,i}(\omega, q) I_{2,j}(\omega, q) \big\} d\varphi dq, \quad 1 \le i \le N, \quad 1 \le j \le N. \tag{20}$$

Элементы матрицы системы (19), (20) имеют абсолютно такой же вид, как и элементы матрицы системы в методе локальных S-аппроксимаций [1] при условии, что элементы поля представляются в виде потенциала простого слоя.

Как было показано выше, преобразование Фурье и лучевое преобразование тесно связаны друг с другом:

$$\begin{aligned} \bar{f}(p, \varphi) &= \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} F(u, \nu) e^{-i|\omega|p} d(|\bar{\omega}|), \quad \bar{\omega} = (\cos\varphi, \sin\varphi), \\ F(u, \nu) &= \int\limits_{-\infty}^{+\infty} \bar{f}(p, \varphi) e^{-i|\omega|p} dp, \end{aligned} \tag{21}$$

где через $\bar{f}(p, \varphi)$ обозначено преобразование Радона от соответствующих переменных, а через $F(u, \nu)$ — преобразование Фурье. Если ввести обозначения

$$\frac{ux + \nu y}{\sqrt{u^2 + \nu^2}} = p, \quad \varphi = \arccos \frac{u}{\sqrt{u^2 + \nu^2}}$$

и принять во внимание тот факт, что

$$\int\limits_{-\infty}^{+\infty} e^{i\omega x} d\omega = 2\pi\delta(x)$$

(здесь $\delta(x)$ — дельта-функция Дирака), то можно показать, что

$$\begin{aligned} \int\limits_{-\infty}^{+\infty} e^{-i(p-q)\rho_1} \cdot e^{-iq\rho_2} dq = e^{-ip\rho_1} \int\limits_{-\infty}^{+\infty} e^{-i(\rho_1-\rho_2)q} dq = 2\pi\delta(\rho_1 - \rho_2) e^{ip\rho_1} \cdot 2\pi \int\limits_{-\infty}^{+\infty} e^{ip\rho_1} \cdot e^{-\rho_1(x_{3i}+H)} \times \\ \times e^{i\rho_1(x_i\cos\varphi + y_i\sin\varphi)} d\rho_1 \cdot \int\limits_{-\infty}^{+\infty} e^{-\rho_2(x_{3j}+H)} \cdot e^{i\rho_2(x_j\cos\varphi + y_j\sin\varphi)} \delta(\rho_1 - \rho_1) d\rho_2 = \\ = \frac{2\pi}{(x_{3i} + 2H + x_{3j})^2 + \Big(p - \big((x_{3i} + x_{3j})\cos\varphi + (y_{3i} + y_{3j})\sin\varphi\big)\Big)^2}. \end{aligned} \tag{22}$$

Если затем осуществить обратное преобразование Радона, отнесенное к i-й точке, (см. формулы (6)–(9)), то мы убедимся в том, что элементы матрицы системы линейных алгебраических уравнений (20) — это вторые производные по i-й координате функции, преобразование Радона которой, деленное на $x_{3i} + 2H + x_{3j}$, есть как раз (22). Данное утверждение отражает очень важный факт: интегральные представления аномальных потенциальных полей (т.е. гармонических в некоторых областях пространства истокообразно представимых функций) весьма тесно связаны друг с другом. Если вспомнить выражения для элементов матрицы в методе S-аппроксимаций [1]:

$$\begin{aligned} a_{ij} &= 2\pi \left\{ \frac{z_i + z_j}{\rho_{i,j}^3} + \frac{(z_i + z_j)\big(9\rho_{i,j}^2 - 6(z_i + z_j)^2\big)}{\rho_{i,j}^7} \right\}, \\ \rho_{i,j}^2 &= (z_i + z_j)^2 + (x_i - x_j)^2 + (y_i - y_j)^2, \quad 1 \le i, j \le N, \end{aligned} \tag{23}$$

то можно сделать вывод, что преобразование Радона приводит к точно такой же системе линейных алгебраических уравнений, как и S-аппроксимация в локальном варианте, но с представлением искомого элемента поля

в виде потенциала простого слоя. Таким образом, интегральная плотность масс, описываемая выражениями вида (18), может быть восстановлена из решения той же самой системы линейных алгебраических уравнений, которая записывается для определения носителей простого и двойного слоев (11). Опишем этот процесс более подробно.

Если поставить вариационную задачу:

$$\Omega(\rho) = \sum_{l=1}^{L} \iint_{-\infty}^{+\infty} \big(\rho_{1,l}^2(\hat{\xi}) + \rho_{2,l}^2(\hat{\xi})\big) d\hat{\xi} = \min_{\rho}, \tag{24}$$

$$f_{i,\delta} - \sum_{l=1}^{L} \iint_{-\infty}^{+\infty} \big(\rho_{1,l}(\hat{\xi}) Q_{1,l}^{(i)}(\hat{\xi}) + \rho_{2,l}(\hat{\xi}) Q_{2,l}^{(i)}(\hat{\xi})\big) d\hat{\xi} = 0, \quad i = 1, 2, \ldots, N, \tag{25}$$

то для ее решения нам нужно будет определить вектор $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_N)$, а затем найти приближения к искомым неизвестным функциям по формулам

$$\rho_{1,l}^{(a)}(\hat{\xi}) = \tilde{\rho}_{1,l}(\hat{\xi}, \lambda), \quad \rho_{2,l}^{(a)}(\hat{\xi}) = \tilde{\rho}_{2,l}(\hat{\xi}, \lambda), \quad \tilde{\rho}_{1,l}(\hat{\xi}, \lambda) = \sum_{i=1}^{N} \lambda_i Q_{1,l}^{(i)}(\hat{\xi}), \quad \tilde{\rho}_{2,l}(\hat{\xi}, \lambda) = \sum_{i=1}^{N} \lambda_i Q_{2,l}^{(i)}(\hat{\xi}), \quad l = 1, 2, \ldots, L. \tag{26}$$

Компоненты вектора $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_N)$ находятся из решения системы линейных алгебраических уравнений, матрица которой имеет элементы, выражающиеся по формулам (23) [1].

2. ОДНОВРЕМЕННОЕ ОПРЕДЕЛЕНИЕ ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЯ ЭКВИВАЛЕНТНЫХ ПО ВНЕШНЕМУ ПОЛЮ ИСТОЧНИКОВ И СПЕКТРОВ ПОЛЕЙ

Спектральный анализ элементов аномальных потенциальных полей был широко распространен в 50–80-е годы прошлого века [3, 7].

При определении двумерных спектров различных сигналов предполагалось, что элемент V_x, $x = (x_1, x_2, x_3)$ аномального поля непрерывно задан на всей бесконечной плоскости $x_3 = 0$ и что однозначно восстанавливается преобразование Фурье $F(u, \nu)$ элемента $V(x)|_{x_3=0}$:

$$F(u, \nu) = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} V(x)|_{x_3=0} \exp\big(i(ux_1 + \nu x_2)\big) dx_1 dx_2. \tag{27}$$

Обратное к (27) преобразование можно записать следующим образом:

$$T\big\{V(x)\big\} = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} K(u, \nu) F(u, \nu) \exp\big(-i(ux_1 + \nu x_2)\big) du d\nu, \tag{28}$$

где $T\big\{V(x)\big\} = W(x)$ есть некоторая линейная трансформанта функции $V(x)$, которой в спектральной области соответствует умножение спектра $F(u\nu)$ на частотную характеристику $K(u, \nu)$. Метод линейных интегральных представлений позволяет принципиально по-новому подойти к использованию метода анализа Фурье в задачах гравиметрии и магнитометрии [3, 7]. Подчеркнем, что формулы (27) и (28) имеют абсолютно аналогичный вид и для четырехмерного пространства (вариационные постановки для которого также рассматривались нами ранее, в [3]), частотная характеристика и спектр будут иметь еще одну компоненту: $K(\omega) = K(u, \nu, \omega)$, $F = F(u, \nu, \omega)$.

Рассмотрим одну из основных постановок задач на нахождение спектров Фурье элементов аномальных потенциальных полей по данным экспериментальных исследований этих полей. В данной работе ограничимся случаем гравитационного поля и задания значений одного элемента:

$$\Delta g(x) = -\frac{\partial V_a(x)}{\partial x_3}, \tag{29}$$

где $V_a(x)$, $x = (x_1, x_2, x_3)$ — потенциал аномального гравитационного поля, а ось Ox_3 направлена вверх, в силу чего в (29) фигурирует знак минус.

Напомним, как решается задача определения спектра сигнала в рамках одного из вариантов метода F-аппроксимаций [7], который состоит в том, что вводится спектральное представление функции $\frac{\partial V_a(x)}{\partial x_3}$, гармонической в полупространстве $x_3 > -H$, через спектр Фурье $F(u,\nu)$ потенциала $V_a(x)$:

$$\frac{\partial V_a(x)}{\partial x_3} = \mathrm{Re}\left\{\frac{1}{2\pi}\iint\limits_{-\infty}^{+\infty} K(u,\nu;x_3+H)F(u,\nu)\exp\bigl(-i(ux_1+\nu x_2)\bigr)dud\nu\right\} =$$

$$= \frac{1}{2\pi}\iint\limits_{-\infty}^{+\infty} K(u,\nu;x_3+H)\bigl(A(u,\nu)\cos(ux_1+\nu x_2)+B(u,\nu)\sin(ux_1+\nu x_2)\bigr)dud\nu. \quad (30)$$

В (30) положено

$$K(u,\nu;x_3+H) = \exp\bigl(-(x_3+H)\sqrt{u^2+\nu^2}\bigr) \quad (31)$$

и

$$F(u,\nu) = A(u,\nu) + iB(u,\nu). \quad (32)$$

Вариационная постановка на нахождение функций $A(u,\nu)$ и $B(u,\nu)$ (действительной и мнимой частей комплексного спектра Фурье) и имеет следующий вид:

$$\iint\limits_{-\infty}^{+\infty} |F(u,\nu)|^2 dud\nu = \iint\limits_{-\infty}^{+\infty}\bigl(A^2(u,\nu)+B^2(u,\nu)\bigr)dud\nu = \min_{\substack{A(u,\nu)\\B(u,\nu)}}. \quad (33)$$

при линейных условиях

$$f_{i,\delta} - \frac{1}{2\pi}\iint\limits_{-\infty}^{+\infty} K(u,\nu;x_3^{(i)}+H)\times\bigl[A(u,\nu)\cos(ux_1^{(i)}+\nu x_2^{(i)})+B(u,\nu)\sin(ux_1^{(i)}+\nu x_2^{(i)})\bigr]dud\nu = 0, \quad i=1,2,\ldots,N, \quad (34)$$

где

$$f_{i,\delta} = \frac{\partial V(x^{(i)})}{\partial x_3} + \delta\left(\frac{\partial V(x^{(i)})}{\partial x_3}\right). \quad (35)$$

Аналогичная вариационная постановка возникает и в том случае, когда требуется найти спектр при частотной характеристике $K(u,\nu;x_3+H) = \sqrt{u^2+\nu^2}\exp\bigl(-(x_3+H)\sqrt{u^2+\nu^2}\bigr)$, соответствующей двойному слою. Компоненты спектра будем обозначать тогда через

$$G(u,\nu) = C(u,\nu) + iD(u,\nu). \quad (36)$$

Задача (33)–(35) решается методом множителей Лагранжа [10]. Имеют место представления

$$\begin{aligned} A(u,\nu) &= \sum_{i=1}^{N}\lambda_i P_i(u,\nu),\\ B(u,\nu) &= \sum_{i=1}^{N}\lambda_i P_i(u,\nu), \end{aligned} \quad (37)$$

где положено

$$\begin{aligned} P_i(u,\nu) &= \frac{1}{2\pi}K(u,\nu;x_3^{(i)}+H)\cos(ux_1^{(i)}+\nu x_2^{(i)}),\\ S_i(u,\nu) &= \frac{1}{2\pi}K(u,\nu;x_3^{(i)}+H)\sin(ux_1^{(i)}+\nu x_2^{(i)}), \quad i=1,2,\ldots,N. \end{aligned} \quad (38)$$

С учетом (31) получаем

$$\begin{aligned} A(u,\nu) &= \frac{1}{2\pi}\sum_{i=1}^{N}\lambda_i e^{-(x_3^{(i)}+H)\sqrt{u^2+\nu^2}}\cos(ux_1^{(i)}+\nu x_2^{(i)}) = \sum_{i=1}^{N}\lambda_i P_i(u,\nu),\\ B(u,\nu) &= \frac{1}{2\pi}\sum_{i=1}^{N}\lambda_i e^{-(x_3^{(i)}+H)\sqrt{u^2+\nu^2}}\sin(ux_1^{(i)}+\nu x_2^{(i)}) = \sum_{i=1}^{N}\lambda_i S_i(u,\nu), \end{aligned} \quad (39)$$

где

$$P_i(u,\nu) = \frac{1}{2\pi} e^{-(x_3^{(i)}+H)\sqrt{u^2+\nu^2}} \cos(ux_1^{(i)} + \nu x_2^{(i)}),$$
$$S_i(u,\nu) = \frac{1}{2\pi} e^{-(x_3^{(i)}+H)\sqrt{u^2+\nu^2}} \sin(ux_1^{(i)} + \nu x_2^{(i)}). \tag{40}$$

Значения параметров λ_i (множителей Лагранжа) находятся из решения (СЛАУ) вида (19), в которой $\mathbf{A}$ есть $(N \times N)$ — матрица со свойством

$$\mathbf{A} = \mathbf{A}^{\mathrm{T}} \geq 0 \tag{41}$$

и элементами a_{pq}, $p = 1, 2, \ldots, N$ $q = 1, 2, \ldots, N$:

$$a_{pq} = \iint\limits_{-\infty}^{+\infty} \big[P_p(u,\nu)P_q(u,\nu) + S_p(u,\nu)S_q(u,\nu)\big]\,du d\nu, \tag{42}$$

С учетом условных обозначений (37) и (38) получим окончательное выражение для элементов искомой матрицы:

$$a_{p,q} = \frac{x_3^{(p)} + x_3^{(q)} + 2H}{2\pi\big[(x_3^{(p)} + x_3^{(q)} + 2H)^2 + (x_1^{(p)} - x_1^{(q)})^2 + (x_2^{(p)} - x_2^{(q)})^2\big]^{\frac{3}{2}}}. \tag{43}$$

Обратим внимание на тот факт, что условная экстремальная задача (33)–(35) не содержит априорной информации о свойствах погрешностей δV_i в экспериментальных данных — в значениях $f_{i,\delta}$. Однако эта информация может быть учтена при нахождении устойчивых приближенных решений СЛАУ (19).

Теперь можно перейти от вариационных постановок вида (24), (25) или (33)–(35) к более сложной задаче. Поясним, как это сделать, на примере объединения вариационных постановок в рамках F- и S-аппроксимаций.

Сформулируем следующую вариационную постановку.

Пусть компоненты поля заданы в конечном множестве точек $M_i = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$, $i = 1, 2, \ldots, N$. Рассмотрим интегральные представления вида (12) элемента аномального поля $\frac{\partial V_a(x)}{\partial x_3}$ на этом множестве. Функции ρ_1, ρ_2 неизвестны. Как и выше, обозначим подынтегральную функцию в первом слагаемом в (12) в точке M_i через $Q_1^{(i)}$, а во втором слагаемом — через $Q_2^{(i)}$. Кроме того, попытаемся одновременно с функциями ρ_1, ρ_2 определить спектральные характеристики указанной компоненты аномального поля $A(u,\nu)$, $B(u,\nu)$, $C(u,\nu)$, $D(u,\nu)$ (см. (32) и (36)). В отличие от варианта представления элемента поля формулой (12), рассмотрим случай L плоскостей, на которых распределены простой и двойной слои. Тогда нам потребуется определить две вектор-функции:

$$P = (\rho_1^{(l)}, \rho_2^{(l)}), \quad A = \big(A^{(l)}(u,\nu); B^{(l)}(u,\nu); C^{(l)}(u,\nu); D^{(l)}(u,\nu)\big), \quad l = 1, 2, \ldots, L, \tag{45}$$

причем компоненты этих функций заданы на плоскостях H_l, $l = 1, 2, \ldots, L$.

Получим следующую задачу на поиск экстремумов двух функционалов при различных линейных ограничениях на искомые функции [11, 12]:

$$\Omega(\rho) = \sum_{l=1}^{L} \iint\limits_{-\infty}^{+\infty} \big(\rho_{1,l}^2(\hat{\xi}) + \rho_{2,l}^2(\hat{\xi})\big)\,d\hat{\xi} = \min_{\rho}, \tag{46}$$

$$f_{i,\delta} - \sum_{l=1}^{L} \iint\limits_{-\infty}^{+\infty} \big(\rho_{1,l}(\hat{\xi})Q_{1,l}^{(i)}(\hat{\xi}) + \rho_{2,l}(\hat{\xi})Q_{2,l}^{(i)}(\hat{\xi})\big)\,d\hat{\xi} = 0, \quad i = 1, 2, \ldots, N, \tag{47}$$

$$\int\limits_{-\infty}^{+\infty} |F^{(l)}(u,\nu)|^2 du d\nu + \int\limits_{-\infty}^{+\infty} |G^{(l)}(u,\nu)|^2 du d\nu = \sum_{l=1}^{L} \iint\limits_{-\infty}^{+\infty} \Big(\big(A^{(l)}(u,\nu)\big)^2 + \big(B^{(l)}(u,\nu)\big)^2 +$$
$$+ \big(C^{(l)}(u,\nu)\big)^2 + \big(D^{(l)}(u,\nu)\big)^2\Big) du d\nu = \min\big\{A^{(l)}(u,\nu), B^{(l)}(u,\nu), C^{(l)}(u,\nu), D^{(l)}(u,\nu)\big\}, \tag{48}$$

$$f_{i,\delta} - \frac{1}{2\pi}\sum_{l=1}^{L}\iint\limits_{-\infty}^{+\infty} K_1(u,\nu;x_3^{(i)}+H_l)\times\left[A^{(l)}(u,\nu)\cos(ux_1^{(i)}+\nu x_2^{(i)})+B^{(l)}(u,\nu)\sin(ux_1^{(i)}+\nu x_2^{(i)})\right]dud\nu+$$

$$+\frac{1}{2\pi}\sum_{l=1}^{L}\iint\limits_{-\infty}^{+\infty} K_2(u,\nu;x_3^{(i)}+H_l)\times\left[C^{(l)}(u,\nu)\cos(ux_1^{(i)}+\nu x_2^{(i)})+D^{(l)}(u,\nu)\sin(ux_1^{(i)}+\nu x_2^{(i)})\right]dud\nu=0,$$

$$i=1,2,\ldots,N, \quad (49)$$

где положено

$$f_{i,\delta}=\frac{\partial V_a(x^{(i)})}{\partial z}+\delta f_i. \quad (50)$$

Решение задачи (46)–(50) будет иметь вид

$$\rho_{1,l}^{(a)}(\hat{\xi})=\tilde{\rho}_{1,l}(\hat{\xi},\lambda), \quad \rho_{2,l}^{(a)}(\hat{\xi})=\tilde{\rho}_{2,l}(\hat{\xi},\lambda), \quad \tilde{\rho}_{1,l}(\hat{\xi},\lambda)=\sum_{i=1}^{N}\lambda_i Q_{1,l}^{(i)}(\hat{\xi}), \quad \tilde{\rho}_{2,l}(\hat{\xi},\lambda)=\sum_{i=1}^{N}\lambda_i Q_{2,l}^{(i)}(\hat{\xi}), \quad l=1,2,\ldots,L,$$

$$A^{(l)}(u,\nu)=\sum_{i=1}^{N}\lambda_i P_{i,1}^{(l)}(u,\nu), \quad B^{(l)}(u,\nu)=\sum_{i=1}^{N}\lambda_i S_{i,1}^{(l)}(u,\nu), \quad (51)$$

$$C^{(l)}(u,\nu)=\sum_{i=1}^{N}\lambda_i P_{i,2}^{(l)}(u,\nu), \quad D^{(l)}(u,\nu)=\sum_{i=1}^{N}\lambda_i S_{i,2}^{(l)}(u,\nu),$$

где положено

$$\begin{aligned}
&P_{i,m}^{(l)}(u,\nu)=\frac{1}{2\pi}K_m(u,\nu;x_3^{(i)}+H_l)\cos(ux_1^{(i)}+\nu x_2^{(i)}),\\
&S_{i,m}^{(l)}(u,\nu)=\frac{1}{2\pi}K_m(u,\nu;x_3^{(i)}+H_l)\sin(ux_1^{(i)}+\nu x_2^{(i)}), \quad m=1,2,\\
&K_1(u,\nu;x_3^{(i)}+H_l)=\exp\bigl(-(x_3^{(i)}+H_l)\sqrt{u^2+\nu^2}\bigr), \quad i=1,\ldots,N, \quad l=1,\ldots,L,\\
&K_2(u,\nu;x_3^{(i)}+H_l)=\sqrt{u^2+\nu^2}\exp\bigl(-(x_3^{(i)}+H_l)\sqrt{u^2+\nu^2}\bigr), \quad i=1,\ldots,N, \quad l=1,\ldots,L.
\end{aligned} \quad (52)$$

Таким образом, приходим к системе линейных уравнений (аналогичной (19)):

$$\mathbf{A}\boldsymbol{\lambda}=\mathbf{f}_\delta, \quad (53)$$

элементы матрицы которой в нашем случае имеют вид

$$a_{ij}=\sum_{l=1}^{L}\iint\limits_{-\infty}^{+\infty}\bigl(Q_{1,l}^{(i)}(\hat{\xi})Q_{1,l}^{(j)}(\hat{\xi})+Q_{2,l}^{(i)}(\hat{\xi})Q_{2,l}^{(j)}(\hat{\xi})\bigr)d\hat{\xi}, \quad 1\le i\le N, \quad 1\le j\le N. \quad (54)$$

Элементы a_{ij} матрицы $\mathbf{A}$ при использовании интегральных представлений (11) и (34) могут быть вычислены явно с помощью интеграла Пуассона. Например, в случае представления вертикальной компоненты гравитационного поля получим выражения

$$a_{ij}=2\pi\sum_{l=1}^{L}\left\{\frac{z_i+z_j-2H_l}{\bigl(\sqrt{(z_i+z_j-2H_l)^2+(x_i-x_j)^2+(y_i-y_j)^2}\bigr)^3}-\right.$$

$$\left.-\frac{(z_i+z_j-2H_l)\Bigl(9\bigl[(x_i-x_j)^2+(y_i-y_j)^2\bigr]-6(z_i+z_j-2H_l)^2\Bigr)}{\bigl(\sqrt{(z_i+z_j-2H_l)^2+(x_i-x_j)^2+(y_i-y_j)^2}\bigr)^7}\right\}, \quad 1\le i\le N, \quad 1\le j\le N. \quad (55)$$

Необходимо подчеркнуть, что компоненты вектор-функции (44) (плотности простого и двойного слоев, а также спектральные плотности соответствующих распределений) находятся из решения одной и той же системы уравнений (53). Если частотная характеристика имеет вид

$$K_2(u,\nu;x_3+H)=\sqrt{u^2+\nu^2}\exp\bigl(-(x_3+H)\sqrt{u^2+\nu^2}\bigr), \quad (56)$$

то элементы матрицы системы для определения спектральных плотностей представляют собой второе слагаемое в формуле (55):

$$a_{p,q} = \frac{6(x_3^{(p)} + x_3^{(q)} + 2H)^3 - 9\big((x_1^{(p)} - x_1^{(q)})^2 + (x_2^{(p)} - x_2^{(q)})^2\big) \cdot (x_3^{(p)} + x_3^{(q)} + 2H)}{2\pi\big[(x_3^{(p)} + x_3^{(q)} + 2H)^2 + (x_1^{(p)} - x_1^{(q)})^2 + (x_2^{(p)} - x_2^{(q)})^2\big]^{\frac{7}{2}}}. \tag{57}$$

Фактически, нами доказана следующая

Теорема 1. *Плотности простых слоев, распределенных на нескольких горизонтальных плоскостях, и спектральные плотности источников при частотной характеристике вида* (31) *определяются в рамках метода линейных интегральных представлений из решения одной и той же системы линейных алгебраических уравнений* (СЛАУ).

Замечание 1. СЛАУ (19) и аналогичные ей необходимо решать с помощью одного из методов регуляризации, см., например, [3, 5, 6].

Как хорошо известно [13], для финитных функций справедлива формула суммирования Пуассона:

$$\sum_{\mathbf{g}\in\mathbb{Z}^n} F(\mathbf{x} + a\mathbf{g}) = \left(\frac{k}{a}\right)^n \sum_{\mathbf{g}\in\mathbb{Z}^n} f\left(\frac{k\mathbf{g}}{a}\right) \exp\big(i(\mathbf{x},\mathbf{g})k/a\big). \tag{58}$$

В (58), как и далее по тексту статьи, приняты следующие обозначения: $\mathbb{Z}^n$ есть n-мерное пространство векторов с целочисленными координатами, $\mathbb{R}^n$ — вещественное n-мерное пространство; $\mathbf{g} = (g_1, g_2 < \ldots, g_n) \in \mathbb{Z}^n$; $\mathbf{x} = (x_1, x_2 < \ldots, x_n) \in \mathbb{R}^n$, a — положительное вещественное число (период функции F). В правой части (58) суммируются значения преобразования Фурье исходной функции, заданные в узлах дополнительной решетки (в фазовом пространстве).

Формула (58) может быть получена различными способами, в частности, путем определения так называемых условно периодических функций.

А именно, для быстро убывающей вместе со всеми производными на бесконечности функции $F(x)$ на вещественной прямой $\mathbb{R}$ положим

$$F_t(x) = \sum_{n=-\infty}^{+\infty} F(x+n)e^{\mathrm{int}}. \tag{59}$$

Функция (59) является условно периодической по x с параметром t.

Поэтому ее можно разложить в ряд Фурье:

$$F_t(x) = \sum_{n=-\infty}^{+\infty} a_n(t)e^{i2\pi nx}, \quad a_n(t) = \sum_{m=-\infty}^{+\infty} \int_0^1 F(x+m)e^{i(mt+2\pi nx)}dx = f(t+2\pi n), \tag{60}$$

где через $f(t+2\pi n)$ обозначено преобразование Фурье функции $F(x)$. Отсюда получаем при $t = 1$ искомую формулу Пуассона в одномерном случае. Если период условно периодической функции равен a, то формула Пуассона приобретает вид:

$$\sum_{g\in\mathbb{Z}^n} f(x+ga) = \left(\frac{2\pi}{a}\right)^n \sum_{g\in\mathbb{Z}^n} F\left(\frac{2\pi g\mathbf{g}}{a}\right) \exp\left(i\frac{2\pi}{a}\mathbf{xg}\right), \tag{61}$$
$$\mathbf{xg} = g_1x_1 + \cdots + g_nx_n, \quad g_i \in \mathbb{Z} \quad x = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n.$$

В (61) через $\mathbb{Z}$, $\mathbb{Z}^n$, $\mathbb{R}^n$ обозначены множество целых чисел, множество n-мерных векторов с целыми и вещественными координатами, соответственно.

Ввиду справедливости формулы (58), можно ввести дополнительный критерий качества решения задачи (46)–(50). А именно, среди всех решений, удовлетворяющих условиям (46)–(50) отберем те, которые дают минимальное отклонение левых и правых частей формулы (58) друг от друга. При этом важно обратить внимание на следующий факт: сеть наблюдений всегда конечна и расположена в ограниченной области пространства [1, 2] (неважно, является ли эта сеть в полном смысле трехмерной или мы интерпретируем результаты измерений вдоль кривых, на кусках поверхностей и т.п.). Поэтому в формуле (58) левая часть рассматривается лишь в конечном числе точек. Что же касается правой части (58), то суммирование значений фурье-преобразования сигнала выполняется во всем пространстве частот. Но частотный спектр сигнала также может быть ограничен. Фактически, мы оказываемся в ситуации, когда более “узкому” спектру сигнала в фазовом пространстве соответствует более широкий спектр в реальном трехмерном пространстве, и наоборот. Это становится понятным, если вспомнить вид преобразования фурье-компонент гравитационного поля согласно методу

F-аппроксимаций (см. (34)). Одна точка наблюдения “дает” одну косинусоиду с экспотенциально убывающей амплитудой:

$$A^{(i)}(u,\nu)=\lambda_i e^{-(x_3^{(i)}+H)\sqrt{u^2+\nu^2}}\cos(ux_1^{(i)}+\nu x_2^{(i)}). \tag{62}$$

При суммировании правой и левой частей формулы (58) можно применять различные варианты формулы Эйлера-Маклорена для многомерного случая [14]. Напомним читателю формулу Эйлера-Маклорена в одномерном случае [15]:

$$\sum_{j=m}^{n}\varphi(j)=\int_m^n\varphi(x)dx+\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!}\{\varphi^{(\nu-1)}(n)-\varphi^{(\nu-1)}(m)\}=$$
$$=\int_m^n\varphi(x)dx+\sum_{\nu=1}^{k-1}\frac{B_\nu}{\nu!}\{\varphi^{(\nu-1)}(n)-\varphi^{(\nu-1)}(m)\}-$$
$$-\frac{1}{k!}\int_0^1\bigl(B_k(x)-B_k\bigr)\sum_{j=m}^{n-1}\varphi^k(j-x+1)dx. \tag{63}$$

В двумерном случае формула (63) приобретает вид

$$\sum_{m=m_1}^{m_2}\sum_{n=n_1}^{n_2}\varphi(m.n)=\int_{m1}^{m2}\int_{n1}^{n2}\varphi(x,y)dxdy+\sum_{\nu}\frac{B_\nu}{\nu!}\varphi^{(\nu-1)}(x,y)|_{m_1,m_2}^{n_1,n_2},\quad \nu=(\nu_1,\nu_2),\quad B_\nu=B_{\nu_1}B_{\nu_2},\quad \nu!=\nu_1!\nu_2!;$$
$$\varphi^{(\nu-1)}=\frac{\partial^{\nu_1-1}\partial^{\nu_2-1}\varphi}{\partial x^{\nu_1-1}\partial y^{\nu_2-1}},\quad \varphi^{(\nu-1)}(x,y)|_{m_1,m_2}^{n_1,n_2}=\varphi^{(\nu-1)}(m_2,n_2)-\varphi^{(\nu-1)}(m_1,n_2)- \tag{64}$$
$$-\varphi^{(\nu-1)}(m_2,n_1)+\varphi^{(\nu-1)}(m_1,n_1).$$

При вычислении суммы значений некоторой функции, зависящей от двух переменных, согласно выражению (64) целесообразно пользоваться формулой Бруно [15]:

$$D^nf\bigl[g(t)\bigr]=n!\sum_{k=1}^{n}f_k\sum_{\substack{\sum jk_j=n\\ \sum k_j=k}}\prod_{j=1}^{n}\frac{g_j^{k_j}}{(j!)^{k_j}k_j!},\quad f_k\equiv D_s^nf(s)|_{s=g(t)},\quad g^k=D_t^kg(t),\quad D_t^kg(t)\equiv\frac{d^kg(t)}{dt^k}. \tag{65}$$

В нашем случае формулу (65) приходится применять дважды, поскольку при суммировании значений экспоненты от функции двух переменных, представляющей собой корень квадратный из суммы квадратов этих переменных, лучше применить (65) к сложной функции вида $f\bigl(g(u,\nu)\bigr)=\exp\bigl(-(z_i+H_l)\sqrt{u^2+\nu^2}\bigr)$ для переменных u и ν в отдельности, но считая функцией $g(w)$ сначала $g_1(w)=-aw$, а затем $w=g_2\bigl(\eta(u,\nu)\bigr)=\sqrt{\eta}$, $\eta=u^2+\nu^2$.

На втором этапе в роли функции f выступает $w(\eta)=\sqrt{\eta}$, а в роли g — функция $g(u,\nu)=u^2+\nu^2$.

Таким образом, мы приходим к выражению

$$f\Bigl(g_1\bigl(g_2(u,\nu)\bigr)\Bigr)\equiv F(u,\nu). \tag{66}$$

Для функции $w(\eta)=\sqrt{\eta}$ верны, как легко установить, следующие выражения для производных k-го порядка:

$$\frac{d^kw}{d\eta^k}=\frac{(-1)^{k+1}(2k-3)!!}{2^k(\eta)^{(2k-1)/2}},\quad k=1,2,\ldots;\quad (2k-3)!!=1,k=1;$$
$$(2k-3)!!=1\cdot3\cdot5\cdot\ldots,k=2,\ldots.$$

Аналогичные соображения верны и для производной по второй переменной ν.

В результате получим следующие соотношения:

$$F(u,\nu)=g_1\bigl(g_2(u,\nu)\bigr),\qquad \frac{\partial F\Bigl(g_1\bigl(g_2(u,\nu)\bigr)\Bigr)}{\partial u}=\frac{dF}{dg}\Big|_{g=g_1\circ g_2(u,\nu)}\cdot\frac{dg_1}{d\tau}\Big|_{\tau=g_2(u,\nu)}\cdot\frac{\partial g_2}{\partial u}. \tag{67}$$

Описанный выше алгоритм действий при выполнении операций суммирования значений функций, заданных в узлах двумерной целочисленной решетки, безусловно, не является единственно возможным, но двухэтапное применение формулы Бруно позволит существенно упростить выражения для производных сложной функции, поскольку у $g(u) = u^2 + \nu^2$ отличны от нуля лишь производные первого и второго порядка (см. формулу (65)), а производные функции $f\big(g_1(w)\big) = \exp(-aw)$ легко вычисляются:

$$\frac{\partial^k g(u)}{\partial u^k} \equiv 0, \quad \frac{\partial^k g(\nu)}{\partial \nu^k} \equiv 0, \quad k \geq 3, \quad \frac{d^k\big(\exp(-aw)\big)}{dw^k} = (-1)^k a^k \exp(-aw).$$

Как было отмечено в предыдущих работах авторов (см. [16–17]), матрицы систем линейных алгебраических уравнений вида (19) бывают невырожденными в тех случаях, когда координаты точек наблюдений располагаются в точках оптимальной в некотором смысле сети. Такую оптимальную сеть построить бывает непросто, но опыт численных расчетов при интерпретации реальных геофизических данных показывает, что в случае равномерной сети точек матрицы СЛАУ, как правило, не вырождена, если имеет место диагональное преобладание (оно наблюдается при задании всех пунктов наблюдений на одной и той же горизонтальной плоскости, например, но не только). Можно ввести дополнительный функционал качества приближенного решения вариационной постановки (46)–(50):

$$X[P,A] = \left\{ \left(\frac{2\pi}{a}\right)^2 \sum_{i=1}^{N} \sum_{n\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} \lambda_i e^{-(x_3^{(i)}+H)\frac{2\pi}{a}\sqrt{n^2+m^2}} \cos\left(\frac{2\pi}{a} n x_1^{(i)} + \frac{2\pi}{a} m x_2^{(i)}\right) - \sum_{i=1}^{N} \sum_{n\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} f_{i,\delta(ma,na)} \right\}^2 = \min_{P,A}. \quad (68)$$

В (68) используются те же обозначения, что и ранее определенные в статье. Если предположить, что координаты сети точек наблюдений удовлетворяют условиям $x_1^{(i)} = ia$, $x_2^{(i)} = ia$, то (68) принимает вид

$$X[P,A] = \left\{ \left(\frac{2\pi}{a}\right)^2 \sum_{i=1}^{N} \sum_{n\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} \lambda_i e^{-(x_3^{(i)}+H)\frac{2\pi}{a}\sqrt{n^2+m^2}} - \sum_{i=1}^{N} \sum_{n\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} f_{i,\delta(ma,na)} \right\}^2 = \min_{P,A}. \quad (69)$$

Если значения полезного сигнала, осложненные случайной помехой заданы в ограниченной области трехмерного пространства (а именно так и бывает на практике), то суммирование в (69) происходит не по всему множеству пар целых чисел (m,n), а только по некоторому конечному подмножеству: $(m,n) \in \mathbb{Z}^2_{\lim} = \big\{m \in \mathbb{Z}, n \in \mathbb{Z} : m_1 \leq m \leq m_2; n_1 \leq n \leq n_2\big\}$.

Множество ограничений на области определения функций, фигурирующих в (69), можно задавать и иными способами. С точки зрения простоты формул для практического использования, наиболее удобным является следующее условие:

$$(m,n) \in \mathbb{Z}^2_{\lim} = \big\{m \in \mathbb{Z}, n \in \mathbb{Z} : m_1^2 + n_1^2 = r_1^2 \leq m^2 + n^2 \leq r_2^2 = m_2^2 + n_2^2\big\}. \quad (70)$$

Применяя формулу приближенного суммирования значений функции вида (64), получим тогда следующее выражение для интегрального слагаемого:

$$\left(\frac{2\pi}{a}\right)^2 \sum_{i=1}^{N} \sum_{(m,n)\in\mathbb{Z}^2_{\lim}} \lambda_i e^{-(x_3^{(i)}+H)\frac{2\pi}{a}\sqrt{m^2+n^2}} \approx -2\pi \sum_{i=1}^{N} \lambda_i \frac{1}{(x_3^{(i)}+H)^2} \left\{ e^{-(x_3^{(i)}+H)\frac{2\pi}{a}\sqrt{m^2+n^2}} \right\}_{r_1}^{r_2} -$$
$$- \left(\frac{2\pi}{a}\right) \cdot 2\pi \sum_{i=1}^{N} \lambda_i \frac{\sqrt{m^2+n^2}}{(x_3^{(i)}+H)} \left\{ e^{-(x_3^{(i)}+H)\frac{2\pi}{a}\sqrt{m^2+n^2}} \right\}_{r_1}^{r_2}, \quad (71)$$

$$r(m,n) = \sqrt{m^2+n^2},$$
$$(m,n) \in \mathbb{Z}^2_{\lim} = \big\{m \in \mathbb{Z}, n \in \mathbb{Z} : m_1^2 + n_1^2 = r_1^2 \leq m^2 + n^2 \leq r_2^2 = m_2^2 + n_2^2\big\}.$$

В (64) мы интегрируем по мере $rdrd\vartheta$, при том $r(m,n) = \sqrt{m^2+n^2}$, $0 \leq \vartheta \leq 2\pi$, $r_1 \leq r \leq r_2$. Формулы (69), (70) позволяет сузить множество приближенных решений вариационной задачи (46)–(50) и, тем самым, повысить надежность предлагаемой методики решения некорректных геофизических задач. Кроме того, предложенный в статье подход к одновременной интерпретации пространственных и спектральных данных (под “пространственными” данными подразумеваются значения поля в точках некоторой сети наблюдений — сенсорах) открывает, на наш взгляд, ряд дополнительных возможностей для уточнения геологического строения Земли и планет: методика чувствительна к погрешностям в определении спектра и, следовательно, мелкие неоднородности можно выявить при ее применении.

3. ЗАКЛЮЧЕНИЕ

1. В статье показана связь вариационных постановок в рамках методов S-, F- и R-аппроксимаций в трехмерном декартовом пространстве. Доказано, что элементы матрицы СЛАУ во всех трех случаях совпадают, при условии если в качестве эквивалентных носителей выбираются простые и двойные слои, распределенные на некотором множестве горизонтальных плоскостей. По одному набору множителей Лагранжа (который представляет собой решение СЛАУ) можно находить различные характеристики аномальных геофизических полей, включая спектр и интегральную плотность носителя вдоль фиксированного направления (при построении F- и R-аппроксимаций соответственно).

2. В статье формулируются условия для повышения качества совместной интерпретации различных данных об изучаемом аномальном поле: вводится дополнительный стабилизатор при восстановлении плотности распределений источников на горизонтальных плоскостях одновременно со спектральной плотностью сигнала.

3. В статье делается акцент на возможность дальнейшего улучшения свойств решений некорректных задач геофизики путем выбора оптимальной сети точек наблюдений. Если из большого и сверхбольшого массива данных об аномальных полях Земли и рельефе «вычленить» информацию о значениях физических величин в точках трехмерного пространства, координаты которых соответствуют пунктам оптимальной сети и одновременно близки к целым числам, то точность результатов совместной интерпретации значительно возрастет. Такой эффект наблюдается благодаря специфическим особенностям различных версий метода интегральных представлений и свойствам преобразования Фурье интегрируемых с квадратом функций во всем пространстве $\mathbb{R}^3$.

СПИСОК ЛИТЕРАТУРЫ

1. *Страхов В.Н., Степанова И.Э.* Метод S-аппроксимаций и его использование при решении задач гравиметрии (локальный вариант) // Физика Земли. 2002. № 2. С. 3–19.
2. *Страхов В.Н., Степанова И.Э.* Метод S-аппроксимаций и его использование при решении задач гравиметрии (региональный вариант) // Физика Земли. 2002. № 7. С. 3–12.
3. *Stepanova I.E., Kerimov I.A., Yagola A.G.* Approximation approach in various modifications of the method of linear integral representations// Izvestiya. Physics of the Solid Earth. 2019. Vol. 55. No 2. P. 218–231.
4. *Страхов В.Н., Керимов И.А., Степанова И.Э.* Разработка теории и компьютерной технологии построения линейных аналитических аппроксимаций гравитационных и магнитных полей. М.: ИФЗ РАН. 2009. 254 с.
5. *Раевский Д.Н., Степанова И.Э.* О решении обратных задач гравиметрии с помощью модифицированного метода S-аппроксимаций // Физика Земли. 2015. № 2. С. 44–54.
6. *Раевский Д.Н., Степанова И.Э.* Модифицированный метод S-аппроксимаций. Региональный вариант // Физика Земли. 2015. № 2. С. 55–66.
7. *Керимов И.А.* Метод F-аппроксимаций при решении задач гравиметрии и магнитометрии. М.:Физматлит, 2011. 262 с.
8. *Степанова И.Э.* Метод R-аппроксимаций при интерпретации данных детальной гравиметрической и магнитометрической съемок // Физика Земли. 2009. № 4. С. 17–30.
9. *Кошляков Н.С., Глинер Э.Б., Смирнов М.М.* Основные дифференциальные уравнения математической физики. М.: Физматгиз, 1962. 767 с.
10. *Лаврентьев М.А., Люстерник Л.А.* Курс вариационного исчисления. М.-Л.: Гостоптехиздат, 1950. 296 с.
11. *Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г.* Численные методы решения некорректных задач. М.: Наука, 1990. 230 с.
12. *Ягола А.Г., Степанова И.Э., Ван Янфей, Титаренко В.Н.* Обратные задачи и методы их решения. Приложения к геофизике. М.: Бином. Лаборатория знаний. 2014. 214 с.
13. *Виленкин Н.Я.* Специальные функции и теория представлений групп. М.: Наука, 1965. 588 с.
14. *Лейнартас Е.К., Петроченко М.Е.* Многомерные аналоги формулы суммирования Эйлера—Маклорена и преобразование Бореля степенных рядов// Сиб. электрон. матем. изв. 2022. Т. 19. Вып. 1. С. 91–100. DOI: 10.33048/semi.2022.19.008

15. *Сачков В.Н.* Комбинаторные методы дискретной математики. М.: Наука, 1977. 320 с.

16. *Kolotov I.I., Lukyanenko D.V., Stepanova I.E., Shchepetilov A.V., Yagola A.G.* On the uniqueness of solution to systems of linear algebraic equations to which the inverse problems of gravimetry and magnetometry are reduced: a regional variant// Comput. Math. and Math.Physics. 2023. V. 63. № 9. P. 1588–1599.

17. *Kolotov I.I., Lukyanenko D.V., Stepanova I.E., Yagola A.G.* On the uniqueness of solutions to systems of linear algebraic equations resulting from the reduction of linear inverse problems of gravimetry and magnetometry: a local case// Comput. Math. and Math.Physics. 2023. V. 63. № 8. P. 1452–1465.

ON THE SIMULTANEOUS DETERMINATION OF THE DENSITY DISTRIBUTION OF EQUIVALENT SOURCES UNDER AN EXTERNAL FIELD AND THE SPECTRUM OF USEFUL SIGNAL

I. Stepanova[a,*], D. Lukyanenko[b], I. Kolotov[b], A. Shepetilov[b], A. Yagola[b], I. Kerimov[a], A. Levashov[b]

[a] *Institute of Physics of the Earth RAS, B. Grusinskaya St. 10, Moscow, 123242, Russia*
[b] *Lomonosov Moscow State University, Leninskiye gory, Moscow, 119991, Russia*
**e-mail: tet@ifz.ru*

Received 28 June, 2023
Revised 28 June, 2023
Accepted 14 January, 2024

Abstract. The article investigates the possibility of simultaneously recovering equivalent sources under an external field and the spectral characteristics of a useful signal. Examples of variational formulations for different versions of the method of linear integral representations are presented, and the problem of finding the density distribution of gravitating or magnetic masses on several horizontal planes is formulated. Additionally, the Fourier transform of the anomalous field element based on the known signal values at certain observation points, complicated by noise, is discussed.

Keywords: systems of linear algebraic equations, integral representations, Poisson summation formula.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 5, с. 881–892

УДК 519.634

РАСЧЕТ НАГРЕВА ПЛАЗМЫ ЗАРЯЖЕННЫМИ ПРОДУКТАМИ ТЕРМОЯДЕРНЫХ РЕАКЦИЙ НА ОСНОВЕ УПРОЩЕННОГО УРАВНЕНИЯ ФОККЕРА–ПЛАНКА

© 2024 г. К. В. Хищенко[1,*], А. А. Чарахчьян[2,]**

[1] 125412 Москва, ул. Ижорская, 13, стр. 2, ОИВТ РАН, Россия
[2] 119333 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия
**e-mail: konst@ihed.ras.ru*
***e-mail: chara@ccas.ru*

Поступила в редакцию 05.12.2023 г.
Переработанный вариант 20.12.2023 г.
Принята к публикации 14.01.2024 г.

Создана двухслойная по времени схема расчета упрощенного кинетического уравнения Фоккера–Планка применительно к переносу заряженных продуктов термоядерной реакции, которая включает в себя интерполяционную процедуру в 4-мерном сеточном пространстве. Обнаружена неустойчивость схемы при малых значениях скорости частицы и специальном выборе скорости торможения частицы в поле иона, которая входит в кинетическое уравнение в качестве параметра. Показано, что условие термализации, которое запрещает расчет кинетического уравнения для частицы с энергией меньше средней энергии иона, существенно ограничивает число термоядерных реакций, где неустойчивость может проявиться. Схема тестирована на задаче релаксации к стационарному состоянию и на задаче с заданной зависимостью от времени скорости термоядерной реакции, для которой можно найти точное решение кинетического уравнения. Библ. 16. Фиг. 7.

Ключевые слова: термоядерная реакция, уравнение Фоккера–Планка, конечно-разностная схема, неустойчивость.

DOI: 10.31857/S0044466924050157, **EDN**: YCLPUZ

ВВЕДЕНИЕ

Инерционный управляемый термоядерный синтез основан на зажигании с помощью внешнего драйвера небольшой части топлива с последующим распространением волны термоядерного горения на его остальную часть [1]. Механизмом горения термоядерного топлива является нагрев плазмы при торможении в ней надтепловых заряженных частиц, которые возникают в результате термоядерных реакций. Простейшая модель локального нагрева, когда частица отдает свою энергию в той точке, где она родилась, имеет очень узкую область применимости.

Перенос надтепловых заряженных частиц в плазме описывается кинетическим уравнением Фоккера–Планка [1], которое можно получить из уравнения Больцмана (см. [2]). Помимо термоядерных реакций это уравнение описывает различные процессы, в частности в астрофизике [3].

Применительно к заряженным продуктам термоядерных реакций имеется замечательная возможность отбросить в уравнении Фоккера–Планка диффузию функции распределения в скоростном пространстве (см. [4], где приведено обоснование малости отброшенного слагаемого). Далее такое упрощенное уравнение Фоккера–Планка будем для краткости называть *кинетическим уравнением*. Решение этого уравнения сводится к решению обыкновенного дифференциального уравнения вдоль характеристики, а уравнения характеристики описывают процесс торможения частицы электронами и ионами плазмы.

Во многих работах для расчета переноса надтепловых заряженных частиц используется трековый метод (см. [1, 5–7]), построенный на простых физических соображениях, применяемых к дискретной среде. Как показано в [7], трековый метод является методом расчета стационарного кинетического уравнения. Что касается нестационарного кинетического уравнения, то, насколько известно авторам, в настоящее время в качестве модели используется не само уравнение, а его диффузионное приближение по телесному углу в одногрупповом [8] или многогрупповом [1] приближении по скорости частицы.

Настоящая работа является продолжением работы [7], в которой рассматривается стационарное кинетическое уравнение. Любопытным результатом этой работы является следующее утверждение. Пусть внутри расчетной области имеется подобласть, где коэффициенты уравнения постоянны. Тогда во всех точках подобласти, удаленных от ее границы на расстояние, превышающее длину пробега частицы, решение уравнения совпадает с упомянутой выше моделью локального нагрева.

В работе [7] применительно к двумерным осесимметричным течениям развит обратный трековый метод, когда пробные частицы влетают в центр ячейки пространственной сетки, а не вылетают из них, как в обычном прямом трековом методе. Для одной и той же сетки по пространственным и угловым переменным объем вычислений для обратного метода заметно больше, чем для прямого. Однако прямой метод, в отличие от обратного, дает значительную потерю точности вблизи оси симметрии для не слишком подробных сеток по телесному углу, что связано с известным “эффектом луча”. В [7] приводятся рекомендации по значительному снижению объема вычислений обратного метода без существенного снижения точности расчета.

Метод [7] применялся для расчета горения оболочечной цилиндрической дейтерий-тритиевой мишени в [9, 10].

В работе [9] впервые было показано, что квазистационарная быстрая безударная волна горения с механизмом распространения в виде нелокального нагрева альфа-частицами, перенос которых описывался стационарным кинетическим уравнением, может возникать из нестационарной детонационной волны по мере увеличения ее интенсивности. Настоящая работа открывает возможность изучения структуры квазистационарной волны горения на основе нестационарного кинетического уравнения.

Параметры зажигающего драйвера и мишени в [9] выбираются так, чтобы по возможности уменьшить энергию зажигания. В частности, произведение начальной плотности и радиуса цилиндра с DT смесью равно примерно 0.5 г/см^2, что примерно соответствует длине торможения альфа-частицы. Учет нестационарности кинетического уравнения возможно приведет к увеличению этого параметра. Размер мишени вдоль оси симметрии выбирается достаточно большим, чтобы увидеть процесс превращения детонационной волны в быструю безударную волну.

Далее мы ограничимся дейтерий-тритиевой реакцией, заряженным продуктом которой является α-частица. Будем также предполагать, что рождающиеся частицы имеют одну и ту же скорость $v_{\max}$, которая определяется энергией частицы 3.5 МэВ, и изотропное распределение по телесному углу. Это предположение основано на том, что для типичной температуры 10^8 K средняя скорость ионов примерно в 10 раз меньше $v_{\max}$ (см. [11]).

Продуктом DT-реакции являются также высокоэнергичные нейтроны. Длина их свободного пробега намного больше характерных размеров области горения. Поэтому они часто не учитываются, как, например, в [9]. В будущем авторы предполагают использовать модель нейтронного нагрева в оптически тонком пределе [12], где учитывается только первый акт рассеяния нейтрона.

1. КИНЕТИЧЕСКОЕ УРАВНЕНИЕ

Помимо начальной скорости частицы $v_{\max}$ в модель переноса α-частиц входят скорости торможения (отрицательные по величине ускорения) частицы на электронах a_e и ионах a_i, скорость рождения частиц в единице объема F и скорость частицы в конце траектории $v_{\rm th}$, при которой она перестает нагревать плазму. Эти функции зависят от термодинамических функций плазмы (плотности и температур электронов и ионов), а скорости торможения a_e и a_i еще и от скорости частицы v. Зависимость термодинамических функций плазмы от пространственной координаты $\mathbf{r}$ и времени t будем считать известной, и тем самым полагать известными функции $a_{e,i}(\mathbf{r}, t, v)$, $F(\mathbf{r}, t)$, $v_{\rm th}(\mathbf{r}, t)$. Задача рассматривается на некотором интервале времени в некоторой пространственной области. Предполагается отсутствие частиц, влетающих в область извне.

Рассматриваемое кинетическое уравнение имеет вид

$$\frac{\partial f}{\partial t} + v(\mathbf{\Omega}\nabla)f + \frac{\partial af}{\partial v} = \tilde{F}\delta(v - v_{\max}),\ a = a_e + a_i,\ \tilde{F}{=}F/4\pi, \tag{1}$$

где $\mathbf{\Omega}$ — единичный вектор, который задает направление полета частицы; $f(\mathbf{r}, t, v, \mathbf{\Omega})$ — функция распределения, которая определяет $f(\mathbf{r}, t, v, \mathbf{\Omega})dvd\Omega$ как число частиц в единице объема в точке $(\mathbf{r}, t)$, имеющих модуль скорости в интервале dv вблизи v и направление в интервале телесного угла $d\Omega$ вблизи $\mathbf{\Omega}$; δ есть δ-функция Дирака.

При $0 \leqslant v < v_{\max}$ правая часть уравнения (1) равна нулю. Следуя [4], в точке $v = v_{\max}$ поставим граничное условие

$$f(\mathbf{r}, t, v_{\max}, \mathbf{\Omega}) = -\frac{\tilde{F}(\mathbf{r}, t)}{a(\mathbf{r}, t, v_{\max})}, \tag{2}$$

которое получается интегрированием уравнения (1) по v от $v_{\max} - \Delta v$ до $v_{\max} + \Delta v$, $\Delta v > 0$ при $\Delta v \to 0$.

Вместо пространственной координаты $\mathbf{r}$ введем в рассмотрение постоянный вектор $\mathbf{r}_0$ и координату ξ вдоль луча $\mathbf{\Omega}$, $\mathbf{r} = \mathbf{r}_0 + \xi\mathbf{\Omega}$, а также переменную $\varepsilon = v^2/2$, $\varepsilon_{\max} = v_{\max}^2/2$. Тогда уравнение (1) при $v < v_{\max}$ (или при $\varepsilon < \varepsilon_{\max}$) принимает вид линейного гиперболического уравнения

$$\frac{1}{v}\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \xi} + a\frac{\partial f}{\partial \varepsilon} = -f a_\varepsilon,\ a = a(\xi, t, \varepsilon). \tag{3}$$

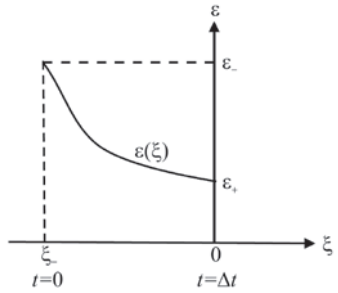
Семейство характеристик уравнения (3) определяется уравнениями

$$\frac{d\xi}{dt} = v,\ \frac{d\varepsilon}{d\xi} = a, \tag{4}$$

где дифференцирование выполняется вдоль характеристики. Уравнения (4) описывают торможение частицы при увеличении ξ под действием отрицательного ускорения a.

При построении численного метода значение $\varepsilon = \varepsilon_+$ задается при $t = \Delta t > 0$, а функция распределения известна при $t = 0$. Кроме того, функция распределения известна при $\varepsilon = \varepsilon_{\max}$ (2) и на границе области (т.е. при некотором значении $\xi < 0$ в точке пересечения луча $-\mathbf{\Omega}$ с границей), где $f = 0$.

Для нахождения нужного участка характеристики надо интегрировать уравнения (4), начиная от точки $t = \Delta t$, в сторону уменьшения t (и, следовательно, уменьшения ξ, так как $v > 0$) до ближайшей точки, где задана функция распределения. На этом участке уравнения (4) описывают движение частицы с положительным ускорением $-a$, как это показано на фиг. 1 применительно к случаю, когда ближайшей точкой с заданной функцией распределения оказывается точка с $t = 0$. В двух других случаях также определяется точка на характеристике, координаты которой будем обозначать ξ_-, t_- и ε_-.



Фиг. 1. Характеристика $\varepsilon(\xi)$ на интервале от $t = \Delta t > 0$, где $\xi = 0$ и задано ε_+, до $t = 0$, где определяются ξ_- и ε_-.

Уравнение (3) вдоль характеристики удобно взять в виде

$$\frac{df}{d\xi} = -f a_\varepsilon\,. \tag{5}$$

Далее с помощью равенства

$$a_\varepsilon = \frac{da}{d\varepsilon} - a_\xi\frac{d\xi}{d\varepsilon} - a_t\frac{dt}{d\varepsilon} = \frac{1}{a}\left(\frac{da}{d\xi} - a_{\xi t}\right),\ a_{\xi t} = a_\xi + a_t/v, \tag{6}$$

уравнение (5) приводится к виду

$$\frac{daf}{d\xi} = f a_{\xi t}, \tag{7}$$

откуда получаем связь между функциями распределения $f_- = f(\xi_-, t_-, \varepsilon_-)$ и $f_+ = f(0, \Delta t, \varepsilon_+)$

$$f_+ a_+ = f_- a_- \exp\left(\int\limits_{\xi_-}^{0} \frac{a_{\xi t}}{a} d\xi\right), \tag{8}$$

где $a_- = a(\xi_-, t_-, \varepsilon_-)$, $a_+ = a(0, \Delta t, \varepsilon_+)$, интегрирование выполняется вдоль характеристики, $a_{\xi t}$ определено в (6).

Для построения численного метода нужна процедура интерполяции термодинамических функций, включая скорость рождения частиц в единице объема F, определенных в серединах ячеек пространственной сетки, на луч, пересекающий эти ячейки. В работе [7] была реализована простейшая интерполяция в виде кусочно постоянной аппроксимации вдоль луча, который разбивался на интервалы, принадлежащие определенным ячейкам сетки. Термодинамические функции на интервале полагались равными своим значениям в центре соответствующей ячейки. Поэтому ниже будет рассмотрен частный случай формулы (8), когда отрезок $[\xi_-, 0]$ разбит на интервалы, в каждом из которых $a(\xi, t, \varepsilon)$ не зависит от ξ и t, а на границах интервалов функция a разрывна.

Пусть отрезок $[\xi_-, 0]$ разбит на N интервалов $[\xi_j, \xi_{j+1}]$, $j = 1, \ldots, N$. На каждом интервале j функция a задана в виде функции $a_j(\varepsilon)$. Функции $f(\xi)$ и $\varepsilon(\xi)$ непрерывны. Формула (8), которая справедлива на любом интервале, дает на интервале $[\xi_j, \xi_{j+1}]$

$$f_j a_j(\varepsilon_j) = f_{j+1} a_j(\varepsilon_{j+1}),\ \varepsilon_j = \varepsilon(\xi_j),\ f_j = f(\xi_j),$$

откуда получаем

$$f_+ = f_- \prod_{j=1}^{N} \frac{a_j(\varepsilon_j)}{a_j(\varepsilon_{j+1})}\,. \tag{9}$$

Мощность нагрева единицы объема электронной W_e и ионной W_i компоненты плазмы задается формулой

$$W_{e,i}(\mathbf{r}, t) = -m_p \int\limits_{(4\pi)} \int\limits_{\varepsilon_{\rm th}}^{\varepsilon_{\max}} a_{e,i}(\mathbf{r}, t, \varepsilon) f(\mathbf{r}, t, \varepsilon, \mathbf{\Omega}) d\varepsilon d\Omega\,, \tag{10}$$

где $\varepsilon_{\rm th} = v_{\rm th}^2/2$, m_p — масса частицы.

В работе [7] применительно к стационарному кинетическому уравнению во внутреннем интеграле формулы (10) делается замена переменных $\xi_- = \xi_-(\varepsilon)$ (см. фиг. 1, где переменная ε из (10) обозначена через ε_+). В результате формула (10) превращается в известную интегральную формулу трекового метода, которая и используется в расчетах. В расчетах настоящей работы используется формула (10), в частности из-за того, что связь между дифференциалами $d\xi_-$ и $d\varepsilon$ зависит от типа точки с заданной функцией распределения ($t = 0$ или $\varepsilon = \varepsilon_{\max}$).

2. ЧИСЛЕННЫЙ МЕТОД ДЛЯ ДВУМЕРНЫХ ОСЕСИММЕТРИЧНЫХ ТЕЧЕНИЙ

Как и в работе [7], будем рассматривать двумерные осесимметричные течения, когда искомая функция распределения и коэффициенты модели зависят только от двух пространственных цилиндрических координат (координаты вдоль оси симметрии z и расстояния до оси r) и не зависят от угловой координаты ψ. В плоскости (z, r) имеется регулярная сетка из четырехугольных ячеек. Все упомянутые выше функции определены при $t = 0$ в серединах ячеек, которые будем называть узлами сетки. Рассматривается двухслойная схема, которая позволяет определять функцию распределения в тех же узлах в следующий момент времени $t = \Delta t$. Шаг Δt предполагается достаточно малым, чтобы можно было пренебречь зависимостью от времени коэффициентов модели.

Из каждого узла пространственной сетки выпускаются лучи. Единичный вектор вдоль луча определяется двумя угловыми координатами следующим образом. Возьмем точку в трехмерном пространстве с цилиндрическими координатами (z_0, r_0, ψ_0), в которой задан единичный вектор $\mathbf{\Omega}$. На фиг. 2 показана связанная с этой точкой локальная декартова система координат, ось z' которой параллельна оси z, ось x' лежит в одной плоскости с осью z и ортогональна ей, а ось y' выбрана так, чтобы оси x', y', z' образовывали правую систему координат. Направление единичного вектора $\mathbf{\Omega}$ задается углами θ и φ, как это показано на фиг. 2.

Обозначим через n_z, n_r и n_ψ проекции вектора $\mathbf{\Omega}$ на оси z', x' и y' соответственно. Как видно из фиг. 2,

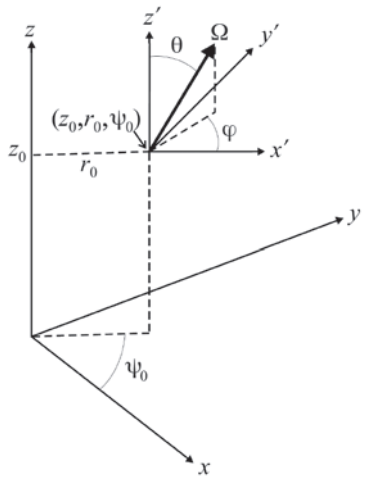
$$n_z = \cos(\theta),\ \ n_r = \sin(\theta)\cos(\varphi),\ \ n_\psi = \sin(\theta)\sin(\varphi). \tag{11}$$

Проекции вектора $\mathbf{\Omega}$ на оси x и y декартовой системы координат

$$n_x = n_r \cos(\psi_0) - n_\psi \sin(\psi_0),\ \ n_y = n_r \sin(\psi_0) + n_\psi \cos(\psi_0). \tag{12}$$

Введем координату ξ вдоль луча, задаваемого вектором $\mathbf{\Omega}$, и найдем зависимость цилиндрических координат $z(\xi)$ и $r(\xi)$ вдоль луча. В дальнейшем понадобятся также функции $\psi(\xi)$ ($\psi(0) = \psi_0$), $n_r(\xi)$ и $n_\psi(\xi)$. Используя зависимость декартовых координат от ξ (см. фиг. 2)

$$\begin{aligned} x &= r(\xi)\cos(\psi(\xi)) = r_0\cos(\psi_0) + n_x\xi, \\ y &= r(\xi)\sin(\psi(\xi)) = r_0\sin(\psi_0) + n_y\xi, \end{aligned} \tag{13}$$



Фиг. 2. Угловые координаты θ и φ единичного вектора $\mathbf{\Omega}$ в локальной системе координат (x', y', z') относительно точки с цилиндрическими координатами (z_0, r_0, ψ_0).

формулу (12) и связь $r^2{=}x^2{+}y^2$, получим формулы

$$z(\xi) = z_0 + n_z\xi, \;\; r(\xi) = \left[(r_0 + n_r\xi)^2 + (n_\psi\xi)^2\right]^{1/2}, \tag{14}$$

которые являются параметрической формой уравнения гиперболы в плоскости z, r. Неравенство $\xi{\leqslant}0$ выделяет часть этой гиперболы.

Как и следовало ожидать, функции (14) не зависят от ψ_0. Отметим также инвариантность этих функций при замене φ на $-\varphi$, что позволяет ограничиться рассмотрением диапазона $0 \leqslant \varphi \leqslant \pi$.

Для нахождения функции распределения в конечной точке характеристики при $t = 0$ необходимо знать значение угла φ, который зависит от ξ. Из второй формулы (11), которая остается справедливой после замены переменных φ и n_r на соответствующие функции, получаем

$$\varphi(\xi) = \arccos\left(n_r(\xi)/\sin(\theta)\right). \tag{15}$$

Формулы (12) остаются в силе после замены переменных ψ_0, n_r и n_ψ соответствующими функциями. Из этих формул можно получить связь

$$n_r(\xi) = n_x\cos(\psi(\xi)) + n_y\sin(\psi(\xi)).$$

После подстановки n_x, n_y из исходных формул (12) и $\cos(\psi(\xi))$, $\sin(\psi(\xi))$ из (13), получим простую формулу

$$n_r(\xi) = r'_\xi(\xi), \tag{16}$$

где $r(\xi)$ определено в (14).

Дифференциал телесного угла

$$d\Omega = d\mu d\varphi, \;\; -1 \leqslant \mu = \cos(\theta) \leqslant 1, \;\; -\pi \leqslant \varphi \leqslant \pi.$$

Пусть имеется для простоты равномерная сетка по угловым переменным

$$\begin{aligned} \tilde{\mu}_m &= -1 + (m-1)\Delta\mu, \;\; \Delta\mu = 2/N_\mu, \;\; m = 1, \ldots, N_\mu + 1, \\ \tilde{\varphi}_k &= (k-1)\Delta\varphi, \;\; \Delta\varphi = \pi/N_\varphi, \;\; k = 1, \ldots, N_\varphi + 1, \end{aligned} \tag{17}$$

которая делит половину полного телесного угла $-1{\leqslant}\mu{\leqslant}1$, $0{\leqslant}\varphi{\leqslant}\pi$ на $N_\mu N_\varphi$ угловых групп с одинаковым телесным углом $\Delta\mu\Delta\varphi$. Каждой угловой группе соответствует единичный вектор с координатами

$$\theta_m = \arccos((\tilde{\mu}_m + \tilde{\mu}_{m+1})/2), \;\; \varphi_k = (\tilde{\varphi}_k + \tilde{\varphi}_{k+1})/2, \;\; 1 \leqslant m \leqslant N_\mu, \;\; 1 \leqslant k \leqslant N_\varphi.$$

Для приближенного представления функции распределения нужна также сетка в скоростном пространстве ε_i, $i = 0, \ldots, M$, $\varepsilon_0 = 0$, $\varepsilon_M = \varepsilon_{\max}$.

При $t = 0$ в узлах всех рассмотренных выше сеток заданы соответствующие значения функции распределения. Для нахождения значений при $t = \Delta t$ из каждого узла пространственной сетки выпускаются лучи, вдоль которых для всех значений ε_i (кроме $i = M$ и узлов с $\varepsilon_i \leqslant \varepsilon_{\text{th}}$, где функция распределения задана формулой (2) и нулевым значением соответственно) находится представленное в разд. 1 решение кинетического уравнения при $\varepsilon_+ = \varepsilon_i$. Вначале находится ближайшая вдоль характеристики точка с заданной функцией распределения (при $t = 0$, на границе области или при $\varepsilon = \varepsilon_{\max}$), а затем по формуле (9) находится искомое значение функции распределения.

Для реализации схемы необходимо находить функцию распределения f_- при $t = 0$, $\varepsilon = \varepsilon_-$, $\mathbf{r} = \mathbf{r}_0 + \xi_-\mathbf{\Omega}$ ($\mathbf{r}_0$ — заданный узел пространственной сетки, $\mathbf{\Omega}$ определяется заданными значениями угловой сетки θ_m, φ_k), $\varphi = \varphi(\xi_-)$ (см. (15), (16)) и $\theta = \theta_m$, по заданным значениям функции распределения в узлах пространственной, угловой и скоростной сеток. Соответствующая процедура содержит интерполяцию высокого порядка точности на 16-ти точечном шаблоне пространственной сетки из [13] и билинейную интерполяцию на объединении скоростной ε_i и угловой φ_k сеток.

Описанная выше конечно-разностная схема при $\Delta t = \infty$ переходит в схему для уравнения (1) с отброшенной производной по времени, которую далее будем называть *стационарной схемой*.

3. ТЕСТОВЫЕ РАСЧЕТЫ

Параметры тестовых задач примерно соответствуют параметрам горения цилиндрических мишеней для инерциального термоядерного синтеза из [9, 10].

Рассмотрены две задачи. В первой задаче все параметры кинетического уравнения не зависят от пространственных координат и времени. Во второй задаче скорость рождения α-частиц в единице объема F зависит заданным образом от времени, а коэффициенты a_e и a_i по-прежнему зависят только от скорости частицы, что позволяет точно решать задачу вдоль характеристики (так как приближенная формула (9) превращается в точную формулу (8) в силу равенства $a_{\xi t} = 0$).

3.1. Задача о релаксации к стационарному состоянию

Рассматривается цилиндр радиусом $R = 0.1$ мм и длиной $H = 1$ мм, состоящий из полностью ионизованной эквимолярной смеси дейтерия и трития, параметры которой не зависят от пространственных координат и времени. Число ячеек пространственной сетки вдоль оси z $N_z = 120$ и вдоль оси r $N_r = 50$. Угловая сетка $N_\mu = 8$, $N_\varphi = 4$.

Температура ионов (T_i) и электронов (T_e) $T_i = T_e = 10^8$ K, плотность смеси $\rho = 100\rho_a$, $\rho_a \approx 0.22$ г/см^3 — плотность жидкой смеси при атмосферном давлении. Скорость рождения α-частиц в единице объема $F = n_{\text{D}} n_{\text{T}} \langle \sigma v \rangle_{\text{DT}}(T_i)$, где n_{D} и $n_{\text{T}} = n_{\text{D}}$ — концентрации ядер дейтерия и трития, которые определяются плотностью плазмы ρ, $\langle \sigma v \rangle_{\text{DT}}(T_i)$ — взятая из [5] скорость реакции.

Для расчета скоростей торможения α-частицы на электронах (a_e) и ионах (a_i), которые зависят от термодинамических функций и скорости α-частицы, как правило используются различные варианты теории парных столкновений в плазме. В настоящей работе использовались два варианта формул. Первый вариант взят из работы [14]. Эта работа посвящена торможению тяжелых ионов, степень ионизации которых также подлежит определению, но после очевидных упрощений можно получить формулы и для полностью ионизованных частиц.

Второй вариант этой формулы из работы [15] для электронов и из работы [16] для ионов. Заметим, что в [16] используется последовательная теория парных столкновений, которая допускает наличие небольшой области скоростной координаты вблизи нуля с положительными значениями a_i, что означает передачу энергии от ионов среды к продуктам термоядерной реакции, а не наоборот.

Наличие такой области связано с функцией

$$G(x, \beta) = \int_0^x \exp(-y)dy - x(1 + \beta)\exp(-x^2),$$

которая входит в качестве множителя в формулу для a_i. При $\beta > 0$ на некотором интервале вблизи $x = 0$ $G(x, \beta) < 0$ (см. [16]). Здесь $x = v(m_i/2k_{\text{B}}T_i)^{1/2}$, m_i — масса иона, k_{B} — постоянная Больцмана, T_i — температура ионов, $\beta = m_i/m_p$. При расчете a_i по формулам из [16] мы полагали $a_i = 0$, если формулы давали $a_i > 0$.

В работе [14] использовался другой подход. Функция $G(x, \beta)$ заменялась положительной при $x > 0$ функцией $G(x, 0)$. Такой подход кажется менее предпочтительным, так как вносит изменения в теорию парных столкновений для случая $a_i < 0$, что нуждается в обосновании. Тем не менее ниже будет показано, что подход с

занулением положительных значений a_i имеет существенный дефект, который может приводить к неустойчивости разностной схемы.

В работе [16] приведена таблица значений функции $x_{cr}(\beta)$, которая является решением уравнения

$$G(x_{cr}(\beta), \beta) = 0.$$

Рассмотрим теперь как влияет на возможность появления положительных значений a_i условие термализации частицы $f(v) = 0$ при $v \leqslant v_{th} = (3k_{\mathrm{B}}T_i/m_i)^{1/2}$. Введем в рассмотрение скорость $v_{cr} = x_{cr}(2k_{\mathrm{B}}T_i/m_i)^{1/2}$. Условие отсутствия положительных значений a_i принимает вид

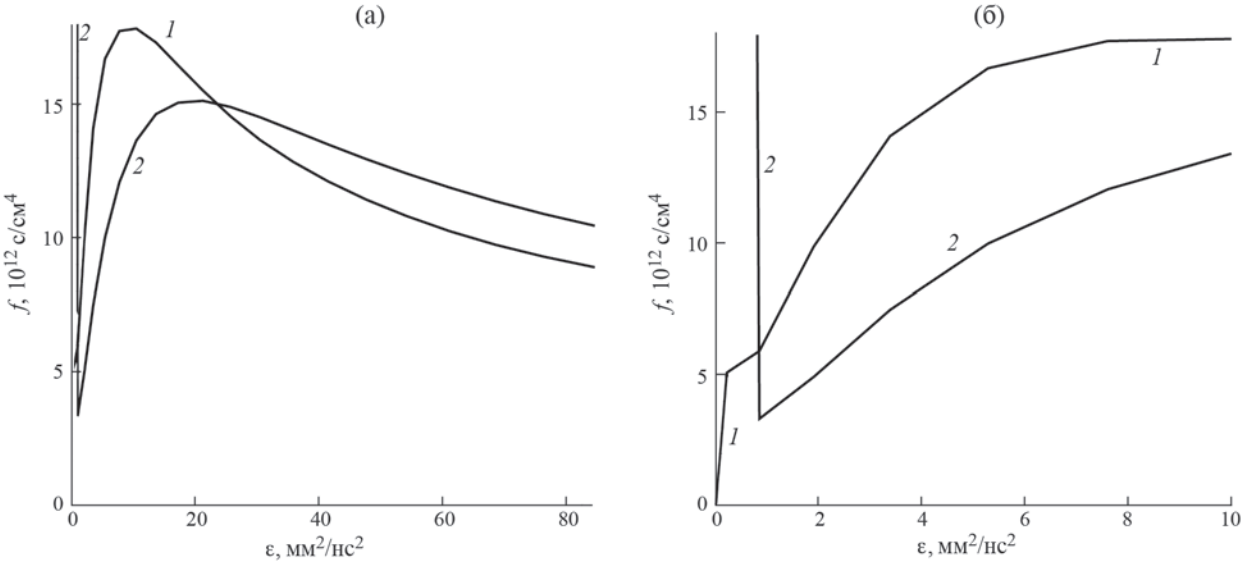
$$\frac{v_{cr}}{v_{th}} = \sqrt{\frac{2}{3}}x_{cr}(\beta) \leqslant 1. \tag{18}$$

Одна из точек упомянутой выше таблицы, ($\beta = 2$, $x_{cr} = 1.23$), с хорошей точностью совпадает с решением уравнения

$$x_{cr}(\beta) = \sqrt{\frac{3}{2}} \approx 1.225\,.$$

Так как функция $x_{cr}(\beta)$ возрастающая, из неравенства (18) получим $\beta \lessapprox 2$. Для DT-реакции $\beta = 2.5/4 \approx 0.6$, что гарантирует отсутствие положительных значений a_i.

Для других термоядерных реакций, в частности протонного типа (см. [1]), положительные значения a_i могут возникать в расчетах. Ниже мы приводим результаты расчетов для DT-реакции, но без условия термализации ($v_{th} = 0$), чтобы продемонстрировать возможное аномальное поведение вычислительной схемы при наличии точек скоростной сетки с положительными значениями a_i, которые искусственно заменяются нулевыми значениями.

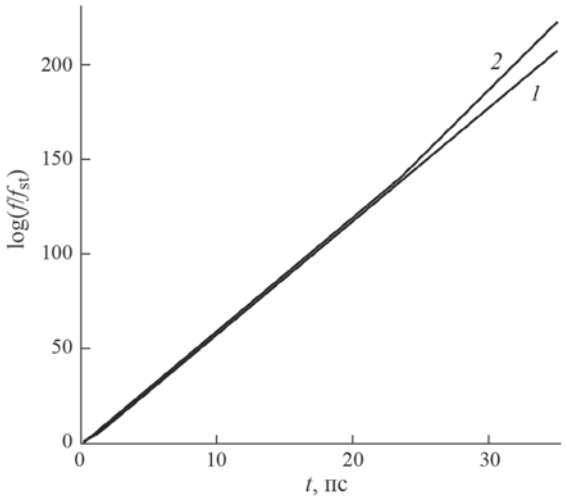


Фиг. 3. Функция распределения (стационарная схема) в некотором узле пространственной сетки вдоль некоторого направления для всего скоростного диапазона (а) и вблизи точки $\varepsilon = 0$ (б) для разных формул вычисления a_e и a_i: *1* – расчет по формулам из [12], *2* — из [13] для a_e и из [14] для ai (с занулением положительных значений a_i); $\varepsilon_{\mathrm{th}} = 0$, $M = 20$, сетка ε_i с равномерным распределением по $v = \sqrt{2\varepsilon}$.

На фиг. 3 приведена функция $f(\varepsilon)$, полученная по стационарной схеме в некоторых узлах пространственной и угловой сетки для двух рассмотренных выше вариантов вычисления a_e и a_i.

В отсутствие условия термализации узел скоростной сетки ε_1 входит в область положительности a_i для второго варианта вычислений, что дает аномально высокое значение f (примерно в 50 раз больше, чем в соседнем узле ε_2), указывающее на наличие множителей заметно больше 1 под знаком произведения в формуле (9). В первом варианте такой аномалии нет.

Рассмотрим теперь численное решение нестационарной задачи о релаксации к стационарному состоянию для той же сетки. Функция распределения в начальный момент времени полагается равной результату расчета по стационарной схеме, умноженной на 20. Сетка и другие параметры те же, что и во втором варианте расчета на фиг. 3, который дает аномально высокое значение функции распределения в узле ε_1.



Фиг. 4. Неустойчивость при релаксации к стационарному состоянию. Функция распределения в узле с аномально высоким стационарным решением f_{st} на фиг. 3 (*1*) и максимальное значение функции распределения (*2*) от времени; расчет с занулением положительных значений a_i, $\varepsilon_{\text{th}} = 0$, сетка та же, что и для стационарной схемы.

Как видно на фиг. 4, численное решение в том же узле ε_1, что и на фиг. 3, как и максимальное значение по узлам всех сеток, стремится не к стационарному значению, а к бесконечности.

Напомним, что для DT-реакции такая неустойчивость возможна только в отсутствие условия термализации ($\varepsilon_{\text{th}} = 0$). Учет этого условия делает неустойчивость возможной для термоядерных реакций с $\beta = m_i/m_p \gtrapprox 2$. Заметим, что для таких реакций неустойчивости можно избежать, если к условию термализации добавить условие $f(v) = 0$ при $x \leqslant x_{cr}(\beta)$.

Заметим, что кривые на фиг. 4 слабо зависят от начальных данных. Например, если в качестве начальных данных взять результат расчета по стационарной схеме, деленный на 20, то кривые на фиг. 4 почти не изменятся.

Далее будем использовать только первый вариант расчета коэффициентов a_e и a_i [14]. Число узлов скоростной сетки $M = 20$. Вместо сетки с равномерной расстановкой узлов по переменной $v = \sqrt{2\varepsilon}$ будет использоваться равномерная по ε сетка, более подробная вблизи $\varepsilon = \varepsilon_{\max}$ и, соответственно, менее подробная вблизи $\varepsilon = 0$.

На фиг. 5 показан результат расчета упомянутой выше задачи о релаксации. Шаг по времени выбирался по формуле

$$\Delta t = \frac{R}{v_{\max}} \frac{5}{N_t},$$

где $N_t = 200$ — число шагов по времени. Видно, что функция распределения сходится к некоторой функции, которую назовем *стационарным состоянием*. Эта функция близка к решению по стационарной схеме. Небольшие отличия связаны с тем, что стационарное состояние зависит от шага по времени и от упомянутой выше интерполяционной процедуры.

На фиг. 6 показано сравнение результатов расчетов с шагом по времени для $N_t = 200$ и в два раза большим шагом для $N_t = 100$ в один и тот же момент времени. Видно, что результаты с хорошей точностью близки друг к другу.

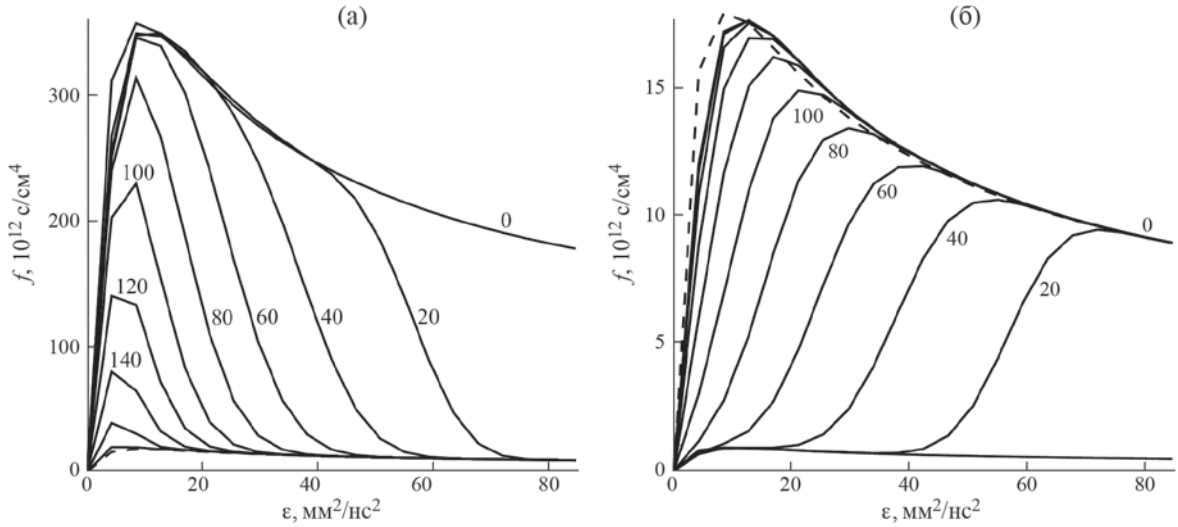
3.2. Задача с заданной временной зависимостью скорости рождения частиц

Рассматривается задача, в которой скорость рождения α-частиц в единице объема

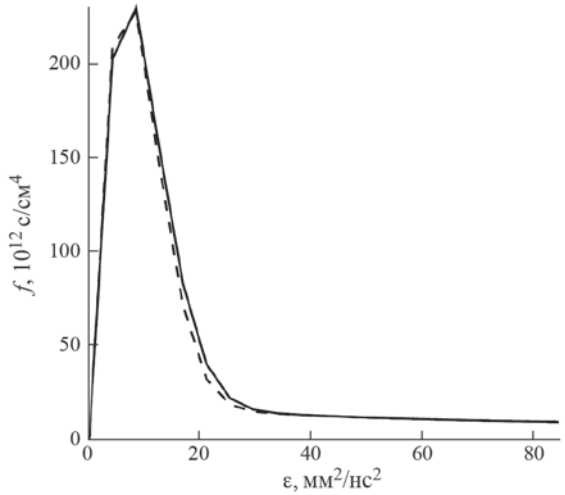
$$F = n_{\text{D}} n_{\text{T}} \langle \sigma \upsilon \rangle_{\text{DT}}(T_i(t)), \tag{19}$$

где $n_{\text{D}} = n_{\text{T}} = \text{const}$, а функция $T_i(t) = (tT_i^{(1)} + (\tau - t)T_i^{(0)})/\tau$ имитирует быстрый рост температуры при зажигании мишени. Здесь $T_i^{(1)} = 10^8$K, $T_i^{(0)} = 6000$ K, $\tau = 20$ пс.

Параметры мишени и сеток: $R = 0.07$ мм, $H = 0.1$ мм, $\rho = 500\rho_a$, $N_z = 100$, $N_r = 40$, $N_\mu = 8$, $N_\varphi = 4$. Температура электронов и ионов, от которых зависят коэффициенты a_e и a_i, полагается независящей от пространственных координат и времени, $T_e^{(2)} = T_i^{(2)} = (T_i^{(1)} + T_i^{(0)})/2$. Начальные данные при $t = 0$: $f = 0$ для всех значений пространственных, угловых и скоростной переменных.



Фиг. 5. Релаксация к стационарному состоянию (коэффициенты a_e и a_i из работы [14]) для начальных данных в 20 раз больше (а) и в 20 раз меньше (б) решения по стационарной схеме (штриховая линия), цифры у сплошных линий — число шагов по времени (0 — начальная функция распределения).



Фиг. 6. Функция распределения после 100 шагов по времени с $N_t = 200$ (сплошная линия) и для того же момента времени с $N_t = 100$ (штриховая линия); задача о релаксации для начальных данных в 20 раз больше решения по стационарной схеме.

Точное решение кинетического уравнения вдоль характеристики в момент времени t легко находится. Для этого надо в расчетных формулах схемы заменить Δt на t, а скорость рождения частиц в единице объема вычислять по формуле (19) с $t = \Delta t - t_-$, где t_- определяется характеристикой.

Двухслойная конечно-разностная схема для кинетического уравнения вычисляет значения функции распределения в момент $t + \Delta t$ по значениям в момент t. Шаг Δt полагается достаточно малым, чтобы можно было пренебречь зависимостью от времени термодинамических функций, в частности скорости рождения частиц в единице объема, которая предварительно вычисляется в момент t в серединах ячеек пространственной сетки и аппроксимируется вдоль луча в виде кусочно-постоянной функции.

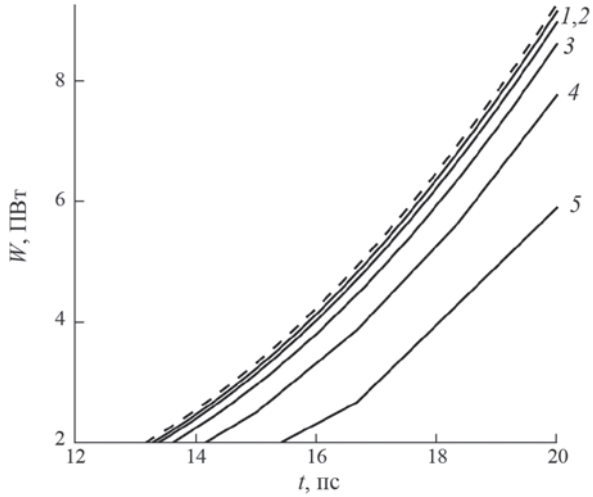
Полная мощность нагрева плазмы

$$W(t) = \int (W_e + W_i) dv,$$

где W_e, W_i определены формулой (10), интегрирование ведется по объему области.

Для вычисления W надо вначале решить кинетическое уравнение, приближенно или точно, а затем воспользоваться квадратурными формулами. Мы не будем проверять точность квадратурных формул. Ясно, что их точность возрастает с ростом числа значений по переменной интегрирования.

Функция $W(t)$ используется для контроля точности двухслойной конечно-разностной схемы относительно точного решения кинетического уравнения. Для каждого интеграла выбирается некоторая квадратурная формула, одна и та же для точного и приближенного расчета кинетического уравнения.



Фиг. 7. Задача с заданной временной зависимостью скорости рождения частиц: точное решение (штриховая линия) и двухслойная схема (сплошные линии) для числа шагов по времени 100 (*1*), 50 (*2*), 25 (*3*), 12 (*4*) и 6 (*5*).

На фиг. 7 представлены результаты расчетов функции $W(t)$ для точного решения и для двухслойной конечно-разностной схемы с разным числом шагов по времени на интервале $0 \leqslant t \leqslant 20$ пс. Видно, что с уменьшением шага по времени погрешность схемы уменьшается.

ЗАКЛЮЧЕНИЕ

Создана двухслойная схема расчета упрощенного нестационарного кинетического уравнения переноса заряженных продуктов термоядерной реакции. По сравнению с развитым ранее обратным трековым методом для упрощенного стационарного кинетического уравнения, появляется новая независимая переменная, что меняет логическую схему вычислений.

Создана необходимая для расчета нестационарного уравнения интерполяционная процедура в 4-мерном сеточном пространстве (интерполяция высокого порядка точности для двумерных криволинейных пространственных сеток применительно к осесимметричным течениям и билинейная интерполяция для угловой и скоростной сеток).

Обнаружена неустойчивость схемы при малых значениях скорости частицы и специальном выборе скорости торможения частицы в поле иона a_i, которая входит в кинетическое уравнение в качестве параметра. Неустойчивость проявляется в бесконечном росте функции распределения с ростом числа шагов по времени и связана с наличием области в пространстве термодинамических функций и скорости частицы (далее аргументов функции a_i), где теория парных столкновений дает $a_i > 0$, что означает ускорение (а не торможение) частицы в поле иона. Чтобы исключить эффект ускорения частицы, не затрагивая область аргументов функции a_i, где $a_i < 0$, положительные значения a_i полагались равными нулю.

Изучена роль условия термализации, которое запрещает расчет кинетического уравнения для частицы с энергией меньше средней энергии иона. Показано, что область аргументов функции a_i, где теория парных столкновений дает $a_i > 0$, может возникать только при $m_i/m_p > 2$, где m_i — масса иона, m_p — масса частицы, что существенно ограничивает число термоядерных реакций, где неустойчивость может проявиться.

Схема тестирована на задаче релаксации к стационарному состоянию и на задаче с заданной зависимостью от времени скорости термоядерной реакции, для которой можно найти точное решение кинетического уравнения.

СПИСОК ЛИТЕРАТУРЫ

1. *Дюдерштадт Дж., Мозес Г.* Инерционный термоядерный синтез. М.: Энергоатомиздат, 1984.

2. *Лифшиц Е.М., Питаевский Л.П.* Физическая кинетика. М.: Наука, 1979.

3. *Aksenov A.G., Ruffini R., Vereshchagin G.V.* Comptonization of photons near the photosphere of relativistic outflows // Monthly Notices of the Royal Astronomical Society: Letters. 2013. Vol. 436. Issue 1. P. L54–L58. https://doi.org/10.1093/mnrasl/slt112

4. *Гуськов С.И., Крохин О.Н., Розанов В.Б.* Перенос энергии заряженными частицами в лазерной плазме // Квантовая электроника. 1974. Т. 1. № 7. С. 1617–1623.

5. *Бракнер К., Джорна С.* Управляемый лазерный синтез. М.: Атомиздат, 1977.

6. *Charakhch'yan A.A., Khishchenko K.V.* Plane thermonuclear detonation waves initiated by proton beams and quasi-one-dimensional model of fast ignition // Laser and Particle Beams. 2015. V. 33. Issue 1. P. 65–80. https://doi.org/10.1017/S0263034614000780

7. *Фролова А.А., Хищенко К.В., Чарахчьян А.А.* Трековый метод расчета нагрева плазмы заряженными продуктами термоядерных реакций для осесимметричных течений // Ж. вычисл. матем. и матем. физ. 2016. Т. 36. № 3. С. 443–454. https://doi.org/10.7868/S0044466916030054

8. *Баско М.М.* Диффузионное описание переноса энергии заряженными продуктами термоядерных реакций // Физика плазмы. 1987. Т. 13. № 8. С. 967–973.

9. *Фролова А.А., Хищенко К.В., Чарахчьян А.А.* Быстрое зажигание пучком протонов и горение цилиндрической оболочечной DT-мишени // Физика плазмы. 2019. Т. 45. № 9. С. 804–824. https://doi.org/10.1134/S0367292119080043

10. *Хищенко К.В., Чарахчьян А.А.* Отражение детонационной волны от плоскости симметрии внутри цилиндрической мишени для управляемого термоядерного синтеза // Ж. вычисл. матем. и матем. физ. 2021. Т. 61. № 10. С. 1715–1733. https://doi.org/10.31857/S0044466921100069

11. *Гуськов С.Ю., Розанов В.Б.* Кинетика термоядерных частиц в лазерной плазме // Труды ФИАН. 1982. Т. 134. С. 115–152.

12. *Баско М.М.* Физические основы инерциального термоядерного синтеза. М.: МИФИ, 2009.

13. *Чарахчьян А.А.* Расчет сжатия дейтерия в конической мишени в рамках уравнений Навье—Стокса для двухтемпературной магнитной гидродинамики // Ж. вычисл. матем. и матем. физ. 1993. Т. 33. № 5. С. 766–784.

14. *Баско М.М.* Торможение быстрых ионов в плотной плазме // Физика плазмы. 1984. Т. 10. № 6. С. 1195–1203.

15. *Выговский О.Б., Ильин Д.А., Левковский А.А. и др.* Торможение быстрых заряженных частиц в идеальной плазме с произвольной степенью вырождения: Препринт № 72. М.: ФИАН, 1990.

16. *Сивухин Д.В.* Кулоновские столкновения в полностью ионизованной плазме // Вопросы теории плазмы. М.: Атомиздат, 1964. Вып. 4. С. 81–187.

CALCULATION OF PLASMA HEATING BY CHARGED PRODUCTS OF THERMONUCLEAR REACTIONS BASED ON A SIMPLIFIED FOKKER–PLANCK EQUATION

K. V. Khishchenko[a,*], A. A. Charakhchyan[b,]**

[a] *Joint Institute for High Temperatures of the Russian Academy of Sciences (JIHT RAS), Izhorskaya St. 13, Bldg. 2, Moscow, 125412, Russia*

[b] *Federal Research Center for Computer Science and Control of the Russian Academy of Sciences (FRC CSC RAS), Vavilov St. 44, Moscow, 119333, Russia*

**e-mail: konst@ihed.ras.ru*

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

Received 05 December, 2023

Revised 20 December, 2023

Accepted 14 January, 2024

Abstract. A two-time-layer scheme has been developed for solving the simplified kinetic Fokker–Planck equation related to the transport of charged products of thermonuclear reactions, which includes an interpolation procedure in four-dimensional grid space. Instabilities in the scheme were detected at low particle velocities and for a specific choice of particle deceleration in the ion field, which enters the kinetic equation as a parameter. It was shown that the thermalization condition, which prohibits solving the kinetic equation for particles with energy lower than the average ion energy, significantly limits the number of thermonuclear reactions where instability can manifest. The scheme was tested on the problem of relaxation to a stationary state and on a problem with a prescribed time-dependent thermonuclear reaction rate, for which an exact solution to the kinetic equation can be found.

Keywords: thermonuclear reaction, Fokker–Planck equation, finite-difference scheme, instability.