СОДЕРЖАНИЕ

Том 47, номер 3, 2007 год

Contents

Vol. 47, No. 3, 2007

Simultaneous English language translation of the journal is available from Pleiades Publishing, Ltd.
Distributed worldwide by Springer. *Computational Mathematics and Mathematical Physics* ISSN 0965-5425.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 355–363

УДК 519.642.8

МЕТОД ЛОКАЛЬНЫХ ВЫПУКЛЫХ МАЖОРАНТ ДЛЯ РЕШЕНИЯ ВАРИАЦИОННО-ПОДОБНЫХ НЕРАВЕНСТВ[1)]

© 2007 г. Е. А. Нурминский*, Н. Б. Шамрай**

(690041 Владивосток, ул. Радио, 5, Ин-т автоматики и процессов управления ДВО РАН;*
*** 690039 Владивосток, ул. Суханова, 8, Дальневосточный гос. ун-т)*
e-mail: nurmi@dvo.ru; nb_shamray@mail.ru

Поступила в редакцию 09.03.2006 г.
Переработанный вариант 15.09.2006 г.

Для вариационно-подобных неравенств предложен численный метод решения, основанный на выпуклых аппроксимациях, локально мажорирующих оценочную функцию. Дано теоретическое обоснование алгоритма, и приведены результаты сравнения вычислительной эффективности предлагаемого подхода с традиционными. Библ. 23. Фиг. 1.

Ключевые слова: вариационные и вариационно-подобные неравенства, оценочные функции, выпуклая оптимизация, метод локальных выпуклых мажорант.

ВВЕДЕНИЕ

В настоящее время для изучения многих задач из разных областей знаний, например таких, как математическая физика, исследование операций, математическая экономика и др., активно используется аппарат вариационных неравенств и их обобщений. Задача решения вариационного неравенства, обозначаемая в дальнейшем через VI(G, X), состоит в нахождении точки $x \in X$ такой, что

$$G^{\mathrm{T}}(x)(y - x) \geq 0 \quad \forall y \in X, \tag{1}$$

где $G : X \longrightarrow \mathbb{R}^n$ – некоторое заданное отображение, $X \subseteq \mathbb{R}^n$ – непустое выпуклое замкнутое множество (см. [1]–[3]). Используемые обозначения будут объяснены далее.

Аппарат решения VI(G, X) уже достаточно хорошо разработан (см., например, монографии [4], [5] и ссылки в них), однако для обобщений (1) условия применимости существующих методов остаются весьма ограничительными. Для сходимости обычно требуются сильная монотонность и липшицевость входящих в обобщенное вариационное неравенство отображений или их производных (см. [6], [7]).

Одним из широко используемых подходов к исследованию и решению VI(G, X) является построение эквивалентной оптимизационной задачи и дальнейшее применение методов математического программирования для ее решения. Важную роль в таком преобразовании играют оценочные функции неравенства (см., например, [8]).

В настоящей работе рассматривается вариационно-подобное неравенство, являющееся одним из непосредственных обобщений задачи (1), и предлагается метод его решения, также основанный на условной оптимизации оценочной функции. Дается теоретическое обоснование метода, и приводятся результаты сравнения вычислительной эффективности предлагаемого подхода с традиционными. Основная идея алгоритма состоит в комбинации метода доверительных окрестностей (trust-region, box-step) (см. [9]) с построением выпуклой аппроксимирующей мажоранты оценочной функции. При достаточно необременительных условиях на отображения, задающие вариационно-подобное неравенство, может быть построена слабо выпуклая (см. [10]) оценочная функция. Построение выпуклых мажорант, эквивалентных таким оценочным функциям с точки зрения локальной оптимизации, является задачей, по сути дела совпадающей с вычислением самой оценочной функции. Хотя алгоритм является локальным и нуждается в достаточно хорошем начальном приближении, он не требует от отображений свойств типа монотонности.

[1)]Работа выполнена при государственной поддержке ведущих научных школ (номер гранта НШ-9004.2006.1) и гранта ДВО РАН (01-052).

1. ИСПОЛЬЗУЕМЫЕ ОБОЗНАЧЕНИЯ И ОПРЕДЕЛЕНИЯ

В работе используются следующие обозначения: $\mathbb{R}^n$ есть n-мерное евклидово пространство, элементы которого считаются вектор-столбцами; $^{\mathrm{т}}$ – символ транспонирования; $a_+ = \max\{a, 0\}$ – положительная срезка числа a; $\|x\| = \sqrt{x^{\mathrm{т}}x}$ – евклидова норма вектора x; $\nabla F(x)$ – матрица Якоби отображения $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ в точке x, при $m = 1$ – градиент $F(x)$ в точке x; $U_\delta(\bar{x}) = \{x : \|x - \bar{x}\| \le \delta\}$ есть δ-окрестность точки $\bar{x}$; $f'(x, d) = \lim\limits_{\tau \to +\infty} [f(x + \tau d) - f(x)]/\tau$ – производная функции f по направлению d.

В дальнейшем потребуются некоторые свойства слабо выпуклых функций.

Определение 1 (см. [10]). Функция $f : X \longrightarrow \mathbb{R}$ называется *слабо выпуклой на* X, если для любого $x \in X$ существует непустое множество $f(x)$ векторов g таких, что

$$f(y) - f(x) \ge g^{\mathrm{т}}(y - x) + r(x, y) \quad \forall y \in X,$$

где $|r(x, y)|/\|x - y\| \longrightarrow 0$ при $x \longrightarrow y$ равномерно по x в каждом компактном подмножестве X.

Для настоящей работы наибольший интерес представляют следующие свойства слабо выпуклых функций:

1) непрерывно дифференцируемая функция является слабо выпуклой;

2) если $f(x)$ – слабо выпуклая функция, то существует ее производная по направлению $f'(x, d)$;

3) пусть $f(x, y)$ – слабо выпуклая по x функция для любого фиксированного $y \in X$, $Y(x)$ – множество тех $y \in Y$, на которых достижим $\sup\limits_{y \in Y} f(x, y) = w(x)$, тогда $w(x)$ – слабо выпуклая функция и $w'(x, d) = \sup\limits_{y \in Y(x)} f'(x, y, d)$, где $f'(x, y, d)$ – производная $f(x, y)$ по x в направлении d.

2. ВАРИАЦИОННО-ПОДОБНЫЕ НЕРАВЕНСТВА

Пусть заданы непустое выпуклое замкнутое подмножество X пространства $\mathbb{R}^n$ и однозначные непрерывные отображения $G, F : X \longrightarrow \mathbb{R}^m$. Задача решения вариационно-подобного неравенства, обозначаемая в дальнейшем через VLI(G, F, X), состоит в нахождении точки $x \in X$ такой, что

$$G^{\mathrm{т}}(x)[F(y) - F(x)] \ge 0 \quad \forall y \in X. \tag{2}$$

Термин “вариационно-подобное неравенство” (variational-like inequality) впервые введен в [11] для задачи нахождения точки $x \in X$, удовлетворяющей условию

$$G^{\mathrm{т}}(x)\eta(y, x) \ge 0 \quad \forall y \in X, \tag{3}$$

где $G : X \longrightarrow \mathbb{R}^n$, $\eta : X \times X \longrightarrow \mathbb{R}^n$ – заданные непрерывные отображения, $X \subseteq \mathbb{R}^n$ – непустое выпуклое замкнутое множество. Так как при $\eta(y, x) = F(y) - F(x)$ и $m = n$ условия (2) и (3) совпадают, то неравенство (2) также будем называть вариационно-подобным.

Как показано в [11], VLI(G, F, X) имеет решение, если выполнено одно из следующих условий:

1) X – ограниченное множество и при любом фиксированном $x \in X$ функция $G^{\mathrm{т}}(x)F(y)$ квазивыпукла по $y \in X$;

2) при любом фиксированном $x \in X$ функция $G^{\mathrm{т}}(x)F(y)$ выпукла по $y \in X$ и существует компактное подмножество $\Omega \in X$ такое, что для любого $x \in X\backslash\Omega$ справедливо неравенство

$$G^{\mathrm{т}}(x)(F(y) - F(x)) < 0 \quad \forall y \in X.$$

К вариационно-подобным неравенствам (2) приводит, например, замена переменных в (1) с целью упрощения допустимой области X. Кроме того, в терминах VLI(G, F, X) могут быть записаны задачи определения обобщенного решения системы неравенств (см. [12]), задачи транспортного и экономического равновесий (см. [13]–[15]).

Один из широко применяемых способов решения задачи (1) состоит в использовании оценочных функций. Это понятие легко переносится и на вариационно-подобные неравенства (2).

Определение 2. *Оценочной функцией* для вариационного (1) (вариационно-подобного (2)) неравенства назовем функцию $\varphi : X \longrightarrow R \cup \{+\infty\}$, обладающую следующими свойствами:

(i) $\varphi(x) \geq 0 \quad \forall x \in X$;

(ii) $x^* \in X$ является решением вариационного (1) (вариационно-подобного (2)) неравенства тогда и только тогда, когда $\varphi(x^*) = 0$ и $x^* \in X$.

Очевидно, что при этом VLI(G, F, X) эквивалентна задаче условной оптимизации

$$\min_{x \in X} \varphi(x). \tag{4}$$

Для вариационного неравенства (1) существует много работ, посвященных различным типам оценочных функций. Наиболее общий вид оценочной функции предложен в [16]:

$$\nu(x) = \sup_{y \in X} \{ f(x) - f(y) + [G(x) - \nabla f(x)]^{\mathrm{T}}(x - y) \}, \tag{5}$$

где $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ – выпуклая, полунепрерывная снизу дифференцируемая на X функция. Отметим, что (5) является задачей выпуклого программирования, но трудно гарантировать свойства выпуклости и дифференцируемости для $\nu(x)$.

Наряду с прямой оценочной функцией (5) в [16] рассматривается ее дуальная формулировка:

$$\tilde{\nu}(y) = \inf_{x \in X} \{ f(x) - f(y) + [G(x) - \nabla f(x)]^{\mathrm{T}}(x - y) \}, \tag{6}$$

и эквивалентная оптимизационная задача для (1) будет состоять в максимизации $\tilde{\nu}(y)$ по $y \in X$. Здесь (6), вообще говоря, не является задачей выпуклого программирования, однако $\tilde{\nu}(y)$ вогнута.

Частные случаи функций (5) и (6) с $f \equiv 0$ были предложены в [17] для нахождения равновесия в задачах теории игр, далее изучены в [18] и последующих работах. При этом основные результаты получены для случая, когда супремум в (5) достижим в единственной точке и, следовательно, функция $\nu(x)$ является дифференцируемой. Другим способом достижения дифференцируемости $\nu(x)$ является использование сильно выпуклых функций $f(x)$, например $f(x) = (1/2)\|x\|^2$ (см. [19]). Дифференцируемые функции исследованы также в [18], [20] и др.

В данной работе для VLI(G, F, X) рассмотрена оценочная функция

$$\varphi(x) = \sup_{y \in X} G^{\mathrm{T}}(x)[F(x) - F(y)]. \tag{7}$$

Легко видеть, что эта функция удовлетворяет свойствам (i) и (ii) определения 2.

В самом деле, для любого $x \in X$

$$\varphi(x) = \sup_{y \in X} G^{\mathrm{T}}(x)[F(x) - F(y)] \geq G^{\mathrm{T}}(x)[F(x) - F(x)] = 0. \tag{8}$$

Если $x^* \in X$ – решение задачи VLI(G, F, X), то

$$G^{\mathrm{T}}(x^*)[F(x^*) - F(y)] \leq 0 \quad \forall y \in X$$

и, следовательно,

$$\varphi(x^*) = \sup_{y \in X} G^{\mathrm{T}}(x^*)[F(x^*) - F(y)] \leq 0.$$

Однако, в силу (8), $\varphi(x^*) \geq 0$, следовательно, $\varphi(x^*) = 0$.

Обратно: пусть для некоторого $x^* \in X$ справедливо соотношение

$$0 = \varphi(x^*) = \sup_{y \in X} G^{\mathrm{T}}(x^*)[F(x^*) - F(y)] \geq G^{\mathrm{T}}(x)[F(x^*) - F(y)] \quad \forall y \in X,$$

следовательно, x^* – решение задачи VLI(G, F, X).

При всей своей концептуальной простоте функция (7) обладает рядом недостатков: это, вообще говоря, негладкая функция и, кроме того, трудно гарантировать ее выпуклость для нелинейных G и F. Однако если G и F являются непрерывно дифференцируемыми и для любого $x \in X$ супремум в (7) достижим, то $\varphi(x)$ – слабо выпуклая функция (см. [10]).

В настоящей работе для решения задачи (4) предлагается построить выпуклую аппроксимацию оценочной функции (7), в значительной степени эквивалентную $\varphi(x)$ с точки зрения ее оптимизации в окрестности приближенного решения.

3. ЛОКАЛЬНАЯ ВЫПУКЛАЯ МАЖОРАНТА ОЦЕНОЧНОЙ ФУНКЦИИ

Далее будем предполагать, что отображения G и F являются дважды непрерывно дифференцируемыми на X и для любого $x \in X$ супремум в (7) достижим.

Зафиксируем точку $\bar{x} \in X$ и выделим ее δ-окрестность $U_\delta(\bar{x})$, где $\delta > 0$ достаточно мало. Окрестность $U_\delta(\bar{x})$ выбирается так, чтобы для некоторого $R > 0$ и любых $y \in X$, $x_\delta \in U_\delta(\bar{x})$, $z \in \mathbb{R}^n$ выполнялись неравенства

$$\left|z^{\mathrm{T}} H_1(x_\delta) z\right| \le R\|z\|^2, \quad \left|z^{\mathrm{T}} H_2(x_\delta, y) z\right| \le R\|z\|^2,$$

где $H_1(x_\delta)$ и $H_2(x_\delta, y)$ – матрицы вторых производных в точке x_δ функций $G^{\mathrm{T}}(x)F(x)$ и $-G^{\mathrm{T}}(x)F(y)$ соответственно.

Введем обозначение

$$\begin{aligned} h(x, y) &= G^{\mathrm{T}}(x)[F(x) - F(y)], \\ c_0(\bar{x}) &= G^{\mathrm{T}}(\bar{x})F(\bar{x}), \\ C(\bar{x}) &= F^{\mathrm{T}}(\bar{x})\nabla G(\bar{x}) + G^{\mathrm{T}}(\bar{x})\nabla F(\bar{x}), \\ A(\bar{x}, z) &= \nabla G(\bar{x})z + G(\bar{x}), \end{aligned} \tag{9}$$

и пусть $Z(x) = \{z : \|z\| \le \delta,\ x + z \in X\}$ – множество допустимых смещений из точки x, по норме не превышающих δ.

Пусть $z \in Z(\bar{x})$, $x = \bar{x} + z$. Для $h(x, y)$ справедлива оценка

$$\begin{aligned} h(x, y) &= h(\bar{x} + z, y) = G^{\mathrm{T}}(\bar{x} + z)[F(\bar{x} + z) - F(y)] = c_0(\bar{x}) + C(\bar{x})z - F^{\mathrm{T}}(y)A(\bar{x}, z) + \\ &+ (1/2)z^{\mathrm{T}}H_1(x_\delta)z + (1/2)z^{\mathrm{T}}H_2(x_\delta)z \le c_0(\bar{x}) + C(\bar{x})z + R\|z\|^2 - F^{\mathrm{T}}(y)A(\bar{x}, z) = \tilde{h}(\bar{x}, y, z). \end{aligned} \tag{10}$$

Рассмотрим функцию

$$\psi(\bar{x}, z) = \sup_{y \in X} \tilde{h}(\bar{x}, y, z) = c_0(\bar{x}) + C(\bar{x})z + R\|z\|^2 - \inf_{y \in X} F^{\mathrm{T}}(y)A(\bar{x}, z), \tag{11}$$

где будем предполагать, что супремум достижим для всех $\bar{x} \in X$ и $z \in Z(\bar{x})$. Заметим, что $\psi(\bar{x}, z)$ выпукла по z.

В силу оценки (10) выполнено следующее неравенство:

$$\sup_{y \in X} h(\bar{x} + z, y) = \varphi(\bar{x} + z) \le \psi(\bar{x}, z),$$

причем при $z = 0$ имеет место равенство.

Взаимосвязь между вариационно-подобным неравенством (2) и функцией $\psi(\bar{x}, z)$ устанавливает

Теорема 1. *Точка x^* является решением задачи* VLI(G, F, X) *тогда и только тогда, когда $z = 0$ есть решение задачи*

$$\min_{z \in Z(x^*)} \psi(x^*, z) \tag{12}$$

и $\psi(x^*, 0) = 0$.

Доказательство. Пусть x^* – решение задачи VLI(G, F, X) и, следовательно, x^* – решение задачи (4) и $\varphi(x^*) = 0$; тогда для любого $z \in Z(x^*)$ имеем

$$\psi(x^*, 0) = \varphi(x^*) \le \varphi(x) = \varphi(x^* + z) \le \psi(x^*, z).$$

Таким образом, $z = 0$ – решение задачи (12) и $\psi(x^*, 0) = 0$.

Обратно: пусть $z = 0$ является решением задачи (12) и $\psi(x^*, 0) = 0$. Так как $x^* \in X$ и верна оценка

$$0 = \psi(x^*, 0) = \varphi(x^*),$$

то, в силу свойства (ii) определения 2, точка x^* – решение VLI(G, F, X).

Обозначим через $\varphi'(x, d)$ и $\psi'(\bar{x}, z, d)$ производные по направлениям функций (7) и (11) по переменным x и z соответственно, т.е.

$$\varphi'(x, d) = \lim_{\tau \to +0} [\varphi(x + \tau d) - \varphi(x)]/\tau, \tag{13}$$

$$\psi'(\bar{x}, z, d) = \lim_{\tau \to +0} [\psi(\bar{x}, z + \tau d) - \psi(\bar{x}, z)]/\tau. \tag{14}$$

Пределы (13) и (14) существуют в силу предположения о достижимости соответствующих супремумов в (7) и (11), слабой выпуклости $\varphi(\cdot)$ и выпуклости $\psi(\bar{x}, \cdot)$ (см. [10]).

Как показывает следующее утверждение, $\psi(\bar{x}, z)$ достаточно хорошо аппроксимирует поведение $\varphi(x)$ в окрестности точки $\bar{x}$.

Лемма. *Пусть $\bar{x} \in X$, тогда для любого $d \in Z(\bar{x})$ имеет место соотношение*

$$\varphi'(\bar{x}, d) = \psi'(\bar{x}, 0, d).$$

Доказательство. Пусть $Y(x)$ – множество тех $y \in X$, на которых достигается супремум в (7). В силу слабой выпуклости $\varphi(x)$,

$$\varphi'(x, d) = \sup_{y \in Y(x)} h'(x, y, d) = \sup_{y \in Y(x)} h'_x(x, y)d = C(x)d - \inf_{y \in Y(x)} [F^{\mathrm{T}}(y)\nabla G(x)d].$$

Пусть $\tilde{Y}(\bar{x}, z)$ – множество тех $y \in X$, на которых достигается супремум в (11). Аналогично, в силу выпуклости $\psi(\bar{x}, z)$ по z,

$$\psi'(\bar{x}, z, d) = \sup_{y \in Y(\bar{x}, z)} \tilde{h}'(\bar{x}, y, z, d) = \sup_{y \in Y(\bar{x}, z)} \tilde{h}'_z(\bar{x}, y, z)d = C(\bar{x})d + 2Rz^{\mathrm{T}}d - \inf_{y \in \tilde{Y}(\bar{x}, z)} [F^{\mathrm{T}}(y)\nabla G(x)d].$$

При $x = \bar{x}$ и $z = 0$ множества $Y(x)$ и $\tilde{Y}(\bar{x}, z)$ совпадают, т.е. $Y(\bar{x}) = \tilde{Y}(\bar{x}, 0)$, следовательно,

$$\varphi'(\bar{x}, d) = \psi'(\bar{x}, 0, d).$$

Из приведенной леммы, в частности, следует, что точка $\bar{x}$ является стационарной для $\varphi(x)$ тогда и только тогда, когда $z = 0$ – стационарная точка функции $\psi(\bar{x}, z)$.

4. МЕТОД ЛОКАЛЬНОЙ ВЫПУКЛОЙ МАЖОРАНТЫ

Используя локальную аппроксимируемость $\varphi(x)$ при $x \in U_\delta(\bar{x})$ выпуклой мажорантой $\psi(\bar{x}, z)$ при $z \in Z(\bar{x})$, описываемую леммой, предлагаем следующий алгоритм решения вариационно-подобного неравенства (2).

Инициализация алгоритма

Выберем точку $x^0 \in X$ и рассмотрим множество $X^0 = \{x : \varphi(x) \le \varphi(x^0)\}$. Определим $\delta > 0$ такое, что для любого $\bar{x} \in X^0$ и $z \in Z(\bar{x})$ выполняется условие

$$\varphi(\bar{x} + z) \le \psi(\bar{x}, z).$$

В качестве начальной возьмем точку $\bar{x}^0 \in X^0$. Положим $k = 0$.

Итерация алгоритма

Шаг 1. Решить задачу

$$\min_{z \in Z(\bar{x}^k)} \psi(\bar{x}^k, z) = \psi(\bar{x}^k, z^k). \tag{15}$$

Шаг 2. Если $z^k = 0$ и $\psi(\bar{x}^k, 0) = 0$, то $\bar{x}^k$ – решение задачи VLI(G, F, X). Если $z^k = 0$, но $\psi(\bar{x}^k, 0) \ne 0$, то VLI(G, F, X) либо решения не имеет, либо мы нашли локальный минимум оценочной функции φ. Алгоритм заканчивает работу.

Шаг 3. Положить $\bar{x}^{k+1} = \bar{x}^k + z^k$, $k = k + 1$ и перейти на шаг 1.

Обоснование сходимости метода дает

Теорема 2. *Пусть вектор* x^0 *таков, что множество* X^0 *не содержит локальных минимумов и стационарных точек оценочной функции* $\varphi(x)$, *отличных от глобальных. Тогда последовательность* $\{\bar{x}^k\}$, *генерируемая алгоритмом, сходится к решению задачи* VLI(G, F, X).

Доказательство. Пусть z^k – решение задачи (15), тогда

$$\varphi(\bar{x}^k) = \psi(\bar{x}^k, 0) \geq \psi(\bar{x}^k, z^k) \geq \varphi(\bar{x}^k + z^k) = \varphi(\bar{x}^{k+1});$$

таким образом, последовательность $\{\varphi(\bar{x}^k)\}$ монотонно убывает и существует предел $\varphi^* = \lim\limits_{k\to\infty}\varphi(\bar{x}^k)$.

Зафиксируем некоторое $z \in Z(\bar{x}^k)$. По построению,

$$\psi(\bar{x}^k, z^k) \leq \min_{\theta\in[0,1]}\psi(\bar{x}^k, \theta z) \leq \psi(\bar{x}^k, 0) + \min_{\theta\in[0,1]}[\theta\psi'(\bar{x}^k, 0, z) + o(\theta)].$$

Если предположить, что для сколь угодно большого K существуют $k > K$, $\gamma > 0$ и $\bar{z}^k \in Z(\bar{x}^k)$ такие, что $\psi(\bar{x}^k, 0, \bar{z}^k) \leq -\gamma < 0$, то

$$\psi(\bar{x}^k, z^k) \leq \psi(\bar{x}^k, 0) + \theta[-\gamma + o(\theta)/\theta] \leq \psi(\bar{x}^k, 0) - \theta\gamma/2$$

для достаточно малых $\theta > 0$. Выбрав такое θ и зафиксировав его, получим

$$\varphi(\bar{x}^{k+1}) \leq \psi(\bar{x}^k, 0, z^k) \leq \psi(\bar{x}^k, 0) - \varepsilon = \varphi(\bar{x}^k) - \varepsilon, \tag{16}$$

где $\varepsilon = \theta\gamma > 0$. Переходя к пределу в (16) по $K \longrightarrow \infty$, получаем

$$\varphi^* \leq \varphi^* - \varepsilon,$$

что невозможно. Отсюда, в частности следует, что

$$\lim_{k\to\infty}\inf_{z\in Z(\bar{x}^k)}\psi'(\bar{x}^k, 0, z) = \lim_{k\to\infty}\inf_{z\in Z(\bar{x}^k)}\varphi'(\bar{x}^k, z) \geq 0,$$

и в силу полунепрерывности сверху $Y(x)$ получаем для любого $z \in Z(\bar{x})$ неравенство

$$\varphi'(\bar{x}, z) \geq \inf_{z\in Z(\bar{x})}\varphi'(\bar{x}, z) \geq \lim_{k\to\infty}\inf_{z\in Z(\bar{x}^k)}\varphi'(x^k, z) \geq 0$$

для любой предельной точки $\bar{x}$ последовательности $\{\bar{x}^k\}$. Таким образом, в $\bar{x}$ удовлетворяются необходимые условия экстремума для $\varphi(x)$, $x \in X$, и, следовательно, $\bar{x}$ является решением задачи (2).

5. ЧИСЛОВОЙ ПРИМЕР

Чтобы подчеркнуть некоторые преимущества переформулировки исходной задачи в виде вариационно-подобного неравенства и применения метода локальных выпуклых мажорант, сравним предлагаемый подход с традиционным, основанным на минимизации регуляризованной оценочной функции. В качестве примера рассмотрим вариационное и вариационно-подобное неравенства, порождаемые оптимизационной задачей

$$\min_{u\in U} g(u), \tag{17}$$

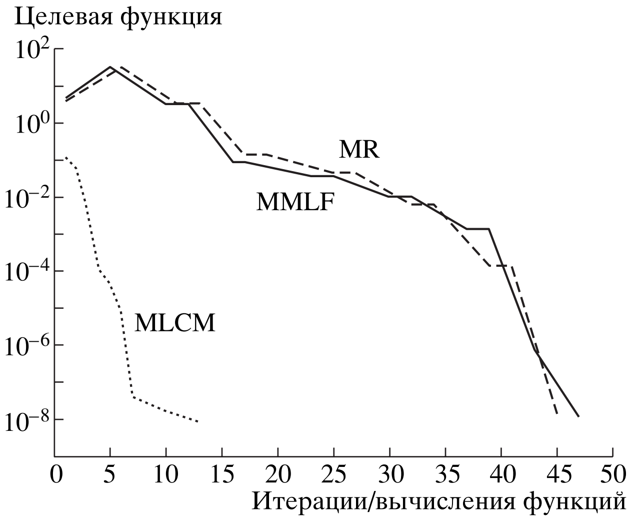
где $g(u) = (1/4)(u_1 - u_2)^2 - (1/2)(u_1 + u_2)$, $U = \{u : \|u\|^2 \leq 1, u \geq 0\}$, решением которой является точка $(\sqrt{2}/2, \sqrt{2}/2)$.

Метод регуляризации (MR)

Задача (17) эквивалентна вариационному неравенству

$$Q^{\mathrm{T}}(u)(v - u) \geq 0 \quad \forall v \in U, \tag{18}$$

где $Q^{\mathrm{T}}(u) = (1/2)(u_1 - u_2 - 1, u_2 - u_1 - 1) = \nabla g(u)$.



Фигура.

В формуле (5) положим $f(u) = (1/2)\|u\|^2$. При этом регуляризованная оценочная функция для (18) имеет вид

$$\nu(u) = \|u - a(u)\|^2 - [\|a(u)\| - 1]_+^2, \tag{19}$$

где $a^{\mathrm{T}}(u) = (1/2)(1 + u_1 + u_2, 1 + u_1 + u_2)$. Подобная регуляризация наиболее часто используется, и $\nu(u)$ называется оценочной функцией Фукушимы (см. [19]). Отметим, что $\nu(u)$ является в данном случае невыпуклой.

Для минимизации $\nu(u)$ при $u \in U$ был применен метод модифицированной функции Лагранжа, реализованный в пакете MINOS (см. [21]). Уменьшение значения $\nu(u)$ в процессе работы алгоритма представлено на фигуре (штриховая линия MR).

Можно предположить, что именно разрыв вторых производных оценочной функции (19) не позволяет достичь более чем линейной скорости сходимости на большей части процесса оптимизации. Ускорение в конце счета связано с выходом в достаточно малую окрестность решения, где $\nu(u)$ имеет уже непрерывные вторые производные.

Для сравнения тем же методом модифицированной функции Лагранжа была решена и исходная задача (17), уменьшение значения целевой функции $g(u)$ также приведено на фигуре (сплошная линия MMLF). Видно, что оба подхода (MR и MMLF) к решению (17) демонстрируют очень близкую вычислительную эффективность, несмотря на существенные отличия в свойствах решаемых задач. По всей видимости, это связано с тем, что в обеих постановках сохранен нелинейный характер ограничений допустимой области U и именно это представляет собой основную вычислительную сложность.

Метод локальных выпуклых мажорант (MLCM)

Поскольку предлагаемый в разд. 4 метод может быть применен к более общему классу вариационных неравенств, чем (18), то возможно упрощение допустимой области U. Замена переменных $u_1^2 = x_1$, $u_2^2 = x_2$ превращает неравенство (18) в вариационно-подобное (2) с основными отображениями

$$G^{\mathrm{T}}(x) = (1/2)(\sqrt{x_1} - \sqrt{x_2} - 1, \sqrt{x_2} - \sqrt{x_1} - 1), \quad F^{\mathrm{T}}(x) = (\sqrt{x_1}, \sqrt{x_2}), \tag{20}$$

и допустимой областью

$$X = \{x : x_1 + x_2 \le 1, x_1 \ge 0, x_2 \ge 0\}. \tag{21}$$

Заметим, что X описывается теперь только линейными ограничениями (21).

Оценочная функция (7) для (2) при (20), (21) имеет вид

$$\varphi(x) = G^{\mathrm{т}}(x)F(x) + \begin{cases} \|G(x)\|, & \text{если } G(x) > 0, \\ \mu(x) & \text{в противном случае,} \end{cases} \tag{22}$$

где $\mu(x) = (1/2)\max\{0, 1 - \sqrt{x_1} + \sqrt{x_2}, 1 + \sqrt{x_1} - \sqrt{x_2}\}$.

Для $\bar{x} \in X$, $\delta > 0$ локальная выпуклая мажоранта имеет вид

$$\psi(\bar{x}, z) = c_0(\bar{x}) + C(\bar{x})z + R\|x\|^2 + \begin{cases} \|A(\bar{x}, z)\|, & \text{если } A(\bar{x}, z) < 0, \\ \mu(\bar{x}, z) & \text{в противном случае,} \end{cases}$$

где $z \in Z(\bar{x})$, $\mu(\bar{x}, z) = (1/2)\max\{0, a_1(\bar{x}, z), a_2(\bar{x}, z)\}$, $c_0(\bar{x})$, $C(\bar{x})$, $A(\bar{x}, z)$ определены в (9), $a_1(\bar{x}, z)$, $a_2(\bar{x}, z)$ – компоненты вектор-функции $A(\bar{x}, z)$ для вариационно-подобного неравенства (2) при (20), (21).

В качестве начальной была выбрана точка $x^0 = [0.2, 0.4]^{\mathrm{т}}$, параметры алгоритма определены как $\delta = 0.1$, $R = 0.5$. Алгоритм реализован на свободно распространяемом MATLAB-подобном языке octave (см. [22]). Для решения задачи выпуклой оптимизации (15) использовался метод отделяющих плоскостей из [23].

На фигуре пунктирной линией MLCM показано убывание оценочной функции $\varphi(x)$ в зависимости от итераций алгоритма. Результат численных экспериментов наглядно демонстрирует существенно более быструю сходимость MLCM по сравнению с MR и MMLF. Следует отметить, что итерации MLCM представляют собой достаточно сложную операцию, поскольку на шаге 1 необходимо решать оптимизационную задачу (15), однако трудоемкость ее решения может быть существенно снижена за счет перезапуска алгоритма с предыдущего оптимума и других способов “горячего старта”. В этих условиях можно надеяться на то, что и общая вычислительная эффективность MLCM окажется достаточно высокой.

СПИСОК ЛИТЕРАТУРЫ

1. *Киндерлелер Д., Стампаккья Г.* Введение в вариационные неравенства и их приложения. М.: Мир, 1983.
2. *Коннов И.В.* Методы решения конечномерных вариационных неравенств: Курс лекций. Казань: Изд-во “ДАС”, 1998.
3. *Harker P.T., Pang J.S.* Finite-dimensional variational inequalities and nonlinear complementarity problems: a survey of theory, algorithms and applications // Math. Program. Ser. B. 1990. V. 48. № 2. P. 161–220.
4. *Попов Л.Д.* Введение в теорию, методы и экономические приложения задач о дополнительности: Учебное пособие. Екатеринбург: Изд-во Уральского ун-та, 2001.
5. *Konnov I.V.* Combined relaxation methods for variational inequalities. Berlin: Springer, 2001.
6. *Коннов И.В.* Комбинированный релаксационный метод для обобщенных вариационных неравенств // Изв. вузов. Математика. 2001. № 12. С. 46–54.
7. *Коннов И.В., Пинягина О.В.* Метод спуска по интервальной функции для негладких задач равновесия // Изв. вузов. Математика. 2003. № 12. С. 71–77.
8. *Patriksson M.* A class of gap functions for variational inequalities // Math. Program. 1994. V. 64. № 1–3. P. 53–79.
9. *Conn A.R., Gould N.I.M., Toint Ph.L.* Trust region methods. Philadelphia: SIAM, 2000.
10. *Нурминский Е.А.* Численные методы решения детерминированных и стохастических минимаксных задач. Киев: Наук. думка, 1979.
11. *Parida J., Sahoo M., Kumar A.* A variational-like inequality problem // Bull. Austral. Math. Soc. 1989. V. 39. P. 225–231.
12. *Булавский В.А.* Методы релаксации для систем неравенств. Учебное пособие. Новосибирск: НГУ, 1981.
13. *Шамрай Н.Б.* О двух подходах к решению систем неравенств // Омский науч. вестн. 2003. Т. 24. № 3. С. 55–57.
14. *Шамрай Н.Б.* Проективный метод для систем неравенств // Докл. АН ВШ РФ. 2005. Т. 4. № 3. С. 36–43.
15. *Шамрай Н.Б.* Применение вариационно-подобных неравенств для решения задач транспортного ценового равновесия // Информатика и системы управления. 2006. Т. 11. № 1. С. 62–72.

16. *Auchmuty G.* Variational principles for variational inequalities // Numer. Funct. Analys. and Optimizat. 1989. V. 10. № 9–10. P. 863–874.
17. *Зуховицкий С.И., Поляк Р.А., Примак М.Е.* Два метода отыскания точек равновесия вогнутых игр n лиц // Докл. АН СССР. 1969. Т. 185. № 1. С. 24–27.
18. *Демьянов В.Ф., Певный А.Б.* Численные методы отыскания седловых точек // Ж. вычисл. матем. и матем. физ. 1972. Т. 12. № 5. С. 1099–1127.
19. *Fukushima M.* Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems // Math. Program. 1992. V. 53. № 1. P. 99–110.
20. *Пшеничный Б.Н., Калжанов М.У.* Метод решения вариационных неравенств // Кибернетика и систем. анализ. 1992. № 6. С. 48–55.
21. Stanford Business Software Inc. [Электронный ресурс] http://www.sbsi-sol-optimize.com/asp/sol_product_minos.htm.
22. Octave Home Page [Электронный ресурс] http://www.octave.org.
23. *Нурминский Е.А.* О методе отделяющих плоскостей с ограниченной памятью // Вычисл. методы и программирование. 2006. Т. 7. С. 133–137.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 364–375

УДК 519.626.2

НЕОБХОДИМЫЕ УСЛОВИЯ МИНИМУМА В АНОРМАЛЬНЫХ ЗАДАЧАХ С ГЕОМЕТРИЧЕСКИМИ ОГРАНИЧЕНИЯМИ[1)]

© 2007 г. А. В. Арутюнов*, Д. Ю. Карамзин**

(* 119198 Москва, ул. Миклухо-Маклая, 6, РУДН;

** 119991 Москва, ул. Вавилова, 40, ВЦ РАН)

e-mail: arutun@orc.ru; dmitry_karamzin@mail.ru

Поступила в редакцию 06.10.2006 г.

Рассматривается анормальная задача минимизации с конечномерным образом и геометрическими ограничениями, которые, в частности, включают в себя ограничения типа неравенств. Для этой задачи получены необходимые условия второго порядка, усиливающие ранее известные результаты. Библ. 7.

Ключевые слова: анормальные задачи, необходимые условия второго порядка, геометрические ограничения, индекс квадратичной формы.

Пусть X – линейное пространство, $Y = \mathbb{R}^k$ есть k-мерное арифметическое пространство. Заданы замкнутое выпуклое множество $C \subseteq Y$, отображение $F : X \longrightarrow Y$ и функция $f : X \longrightarrow \mathbb{R}^1$. Рассмотрим экстремальную задачу

$$f(x) \longrightarrow \min, \quad F(x) \in C. \tag{P}$$

Целью настоящей работы является вывод для этой задачи необходимых условий локального минимума второго порядка при естественных предположениях гладкости F и f в случае, когда рассматриваемая точка экстремума анормальна в определяемом ниже смысле.

Поясним сказанное на примере частного случая задачи (P), а именно задачи с ограничениями типа равенств

$$f(x) \longrightarrow \min, \quad F(x) = 0, \tag{1}$$

т.е. при $C = 0$, и когда пространство X конечномерно, а f и F дважды непрерывно дифференцируемы. Пусть x_0 является локальным минимумом в задаче (1). Могут представиться две возможности. Пусть вначале точка x_0 нормальна. Последнее означает, что $\operatorname{im}F'(x_0) = Y$, где $\operatorname{im}F'(x_0)$ – образ оператора $F'(x_0)$. В этом случае необходимые условия первого и второго порядков хорошо известны (см. [1]). Они заключаются в существовании такого множителя Лагранжа λ, что вторая производная функции Лагранжа $\dfrac{\partial^2\mathcal{L}}{\partial x^2}(x_0, \lambda)$ неотрицательно определена на ядре $\ker F'(x_0)$ оператора $F'(x_0)$, которое является касательным подпространством ко множеству $\{x : F(x) = 0\}$ в точке x_0.

Откажемся теперь от предположения нормальности, допустив тем самым, что точка x_0 может быть анормальной, т.е. что $\operatorname{im}F'(x_0) \neq Y$. Тогда приведенные выше необходимые условия второго порядка, как известно (см. [2]), вообще говоря, могут не выполняться. Тем не менее в [2] получены необходимые условия второго порядка, которые остаются содержательными и без априорных предположений нормальности точки x_0. Они заключаются в том, что для любого вектора $h \in \ker F'(x_0)$ существует такой (зависящий от h) множитель Лагранжа $\lambda \in \Lambda_k(x_0)$, что $\dfrac{\partial^2\mathcal{L}}{\partial x^2}(x_0, \lambda)[h, h] \geq 0$. Здесь $\Lambda_k(x_0)$ – множество тех нормированных множителей Лагранжа λ, для каждого из ко-

[1)] Работа выполнена при финансовой поддержке РФФИ (коды проектов 05-01-00193, 06-01-00530, 06-01-81004), Фонда содействия отечественной науке и грантов Президента РФ (проекты МК-5591.2006.1, НШ-5344.2006.1).

торых найдется такое линейное подпространство $\Pi \subseteq X$, что

$$\Pi \subseteq \ker F'(x_0), \quad \operatorname{codim} \Pi \le k, \quad \frac{\partial^2 \mathcal{L}}{\partial x^2}(x_0, \lambda)[x, x] \ge 0 \quad \forall x \in \Pi,$$

где codim обозначает коразмерность линейного подпространства. Эти необходимые условия естественным образом обобщают классические результаты на анормальный случай. При этом непустота множества $\Lambda_k(x_0)$ уже сама является важным необходимым условием минимума.

Приведенные результаты из [2] были затем обобщены в [3] на более общую задачу (P), в которой множество C предполагается всего лишь замкнутым. После этого в [4] для задачи (1) при дополнительном предположении, что точка x_0 анормальна, необходимые условия из [2] были усилены. А именно, в [4] доказано, что если точка x_0 анормальна, то в приведенной выше формулировке необходимых условий множество $\Lambda_k(x_0)$ можно заменить на, вообще говоря, меньшее множество $\Lambda_{k-1}(x_0)$. В настоящей работе этот результат обобщается на более общую, чем (1), задачу (P).

Перейдем к точным формулировкам. Вначале определим на пространстве X так называемую конечную топологию. Для этого обозначим через $\mathcal{M}$ множество всех линейных конечномерных подпространств $M \subseteq X$. Открытыми в конечной топологии являются те и только те множества, для которых пересечение с любым подпространством $M \in \mathcal{M}$ открыто в единственной отделимой векторной топологии пространства M. Локальный минимум относительно конечной топологии является слабейшим среди обычно рассматриваемых типов минимума (подробности см. в [2]). В дальнейшем под локальным минимумом будем подразумевать локальный минимум относительно конечной топологии.

Пусть $x_0 \in X$, причем $y_0 = F(x_0) \in C$. Предположим, что x_0 является локальным минимумом задачи (P). Будем также предполагать, что отображения F, f дважды непрерывно дифференцируемы в окрестности точки x_0 относительно конечной топологии. Это означает, что для произвольного $M \in \mathcal{M}$, содержащего точку x_0, сужения f и F на M дважды непрерывно дифференцируемы в некоторой (зависящей от M) окрестности точки x_0.

Рассмотрим функцию Лагранжа $\mathcal{L} : X \times \mathbb{R}^{k+1} \longrightarrow \mathbb{R}^1$:

$$\mathcal{L}(x, \lambda) = \lambda_0 f(x) + \langle y^*, F(x) \rangle, \quad \lambda = (\lambda_0, y^*), \quad \lambda_0 \in \mathbb{R}^1, \quad y^* \in Y.$$

(Здесь и ниже пространство, сопряженное с арифметическим пространством, отождествляется с ним самим.)

Обозначим через $\Lambda(x_0)$ множество векторов $\lambda = (\lambda_0, y^*)$, которые отвечают принципу Лагранжа в точке x_0:

$$\frac{\partial \mathcal{L}}{\partial x}(x_0, \lambda) = 0, \quad \lambda_0 \ge 0, \quad y^* \in N_C(y_0), \quad |\lambda| = 1, \tag{2}$$

где $N_C(y)$ – нормальный конус к выпуклому множеству C в точке $y \in C$. В силу принципа Лагранжа (см. [5], [3]), множество $\Lambda(x_0)$ не пусто. Элементы λ этого множества называются множителями Лагранжа.

Линейное подпространство $I \subseteq Y$ называется инвариантным относительно множества C, если $I + C \subseteq C$ (см. [3, определение 2.1]). Заметим, что если I_1, I_2 – два инвариантных подпространства, то $I = I_1 + I_2$ – также инвариантное подпространство относительно C, так как

$$I + C = (I_1 + I_2) + C = I_1 + (I_2 + C) \subseteq I_1 + C \subseteq C.$$

Поэтому в силу конечномерности пространства Y существует максимальное по включению среди всех инвариантных подпространств I, которое обозначим через $\mathcal{I}$.

Положим $s = \dim \mathcal{I}^{\perp}$, где $^{\perp}$ обозначает ортогональное дополнение. Отметим (это будет использовано ниже), что по построению $y^* \in \mathcal{I}^{\perp}$ $\forall (\lambda_0, y^*) \in \Lambda(x_0)$.

Для целого неотрицательного числа r обозначим через $\Lambda_r(x_0)$ множество векторов $\lambda \in \Lambda(x_0)$, для каждого из которых существует линейное подпространство $\Pi = \Pi(\lambda) \subseteq F'(x_0)^{-1}(\mathcal{I})$ такое, что

$$\operatorname{codim} \Pi \le r, \quad \frac{\partial^2 \mathcal{L}}{\partial x^2}(x_0, \lambda)[x, x] \ge 0 \quad \forall x \in \Pi.$$

Будем говорить, что точка x_0 *анормальна*, если $\operatorname{im} F'(x_0) + \mathcal{I} \neq Y$.

Рассмотрим множества

$$T_C(y_0) = \{d \in Y : \operatorname{dist}(y_0 + \varepsilon d, C) = o(\varepsilon), \varepsilon \geq 0\},$$

$$T_C^2(y_0, d) = \left\{w \in Y : \operatorname{dist}\left(y_0 + \varepsilon d + \frac{1}{2}\varepsilon^2 w, C\right) = o(\varepsilon^2), \varepsilon \geq 0\right\},$$

которые называются, соответственно, *касательным конусом* и *внутренним касательным множеством второго порядка* к C в точке y_0, где dist – расстояние от точки до множества. Отметим, что оба эти множества замкнуты и выпуклы.

Рассмотрим конус критических направлений в точке x_0:

$$\mathcal{K}(x_0) = \left\{x \in X : \left\langle \frac{\partial f}{\partial x}(x_0), x\right\rangle \leq 0, \frac{\partial F}{\partial x}(x_0)x \in T_C(F(x_0))\right\}.$$

Теорема 1. *Пусть точка x_0 является локальным минимумом в задаче* (P) *и анормальна. Тогда $\Lambda_{s-1}(x_0) \neq \varnothing$ и для любых векторов $h \in \mathcal{K}(x_0)$, $w \in T_C^2\left(F(x_0), \frac{\partial F}{\partial x}(x_0)h\right)$ имеет место неравенство*

$$\max_{\lambda \in \Lambda_{s-1}(x_0)} \left\{\frac{\partial^2 \mathcal{L}}{\partial x^2}(x_0, \lambda)[h, h] - \langle y^*, w\rangle\right\} \geq 0. \tag{3}$$

Доказательство. Не ограничивая общности, будем считать, что $x_0 = 0$, $F(x_0) = 0$, $f(x_0) = 0$. Положим $\mathcal{K} = \mathcal{K}(0)$, $\Lambda = \Lambda(0)$. Доказательство теоремы разобьем на четыре этапа. При этом будем использовать три вспомогательных утверждения (леммы 1, 2, 3), которые приведены в Приложении в конце статьи.

Этап I. Докажем утверждение теоремы при следующих дополнительных предположениях (от которых будем последовательно избавляться на последующих этапах):

а) пространство X конечномерно и $X = \mathbb{R}^n$;

б) $\mathcal{I} = \{0\}$ (и поэтому $s = k$);

в) множество C является конечнопорожденным конусом, т.е. существуют такие векторы $a_i \in Y$, $i = 1, 2, \ldots, i_1$, что $C = \{y \in Y : \exists \alpha_i \geq 0 \lfloor y = \sum_{i=1}^{i_1} \alpha_i a_i\ \}$;

г) $w = 0$.

Заметим, что предположение г) корректно, так как в силу предположения в) имеет место $0 \in T_C^2(0, d)$ $\forall d \in T_C(0)$.

Рассмотрим два случая.

Случай 1. Пусть вначале

$$f'(0) \in \operatorname{im}(F'(0)^*). \tag{4}$$

Возьмем произвольную точку $y \in C$. Положим $Z(y) = T_C(y) \cap (-T_C(y))$. Заметим, что $Z(0) = \mathcal{I} = \{0\}$, но $Z(y) \neq \{0\}$ при $y \neq 0$, так как $y \in Z(y)$. При этом если $y \in \operatorname{int} C$, то $Z(y) = Y$, а если $y \in \partial C$, то $Z(y)$ – это линейная оболочка наименьшей грани, содержащей y. Обозначим через $P^{\perp}(y)$ оператор ортогонального проектирования Y на подпространство $(Z(y))^{\perp}$.

Как известно, конечно-порожденный конус имеет конечное число граней. Поэтому когда y пробегает все значения из конуса C, то $Z(y)$ принимает лишь конечное множество значений (подпространств). Среди этих значений выберем только те подпространства, для которых $f'(0) \in \operatorname{im}(P^{\perp}(y)F'(0)^*)$. Эти подпространства обозначим через $Z_1, \ldots, Z_m$. Оператор ортогонального проектирования на Z_i обозначим через $P_i^{\perp}$. Отметим, что $m \geq 1$, так как, в силу (4), в качестве Z_1 можно взять $Z(0)$.

Положим $L_i = \ker(P_i^{\perp}F'(0))$, $N_i = \ker(P_i^{\perp}F'(0))^*$, $i = 1\ 2, \ldots, m$. По определению, для каждого $i = 1, 2, \ldots, m$ существует вектор $\zeta_i \in Y$ такой, что $f'(0) + (P_i^{\perp}F'(0))^*\zeta_i = 0$.

Определим билинейные отображения Q_i и квадратичные формы $Q_{i,0}$ по формулам

$$Q_i[x, \xi] = P_i^{\perp}F''(0)[x, \xi], \quad Q_{0,i}[x, \xi] = f''(0)[x, \xi] + \zeta_i^* Q_i[x, \xi], \quad x, \xi \in X.$$

Для каждого номера $i = 1, 2, \ldots, m$ по подпространствам L_i, N_i, квадратичному отображению Q_i и квадратичной форме $Q_{i,0}$ в соответствии с приведенными ниже в приложении формулами (30) и (31) определим множества M_1^i, M_2^i.

По лемме 1, для всех достаточно малых $\delta > 0$ функция $\phi_i(x) := Q_{0,i}[x, x]$ не принимает значение $(-\delta)$ на множествах M_1^i и M_2^i при всех $i = 1, 2, \ldots, m$. Зафиксируем такое $\delta > 0$ и положим $f_\delta(x) = f(x) + \frac{1}{2}\delta|x|^2$.

Вначале докажем, что для любого $h \in \mathcal{K}$ существуют (зависящие от δ и, естественно, от h) линейное подпространство $\Pi(\delta)$ и вектор λ такие, что

$$\Pi(\delta) \subseteq \ker F'(0), \quad \operatorname{codim}\Pi(\delta) \le k - 1, \quad \lambda \in \Lambda, \tag{5}$$

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda)[\xi, \xi] + \lambda_0\delta|\xi|^2 \ge 0 \quad \forall \xi \in \Pi(\delta), \tag{6}$$

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda)[h, h] + \lambda_0\delta|h|^2 \ge 0. \tag{7}$$

В силу принципа Лагранжа, $\Lambda \ne \emptyset$. Положим

$$H = H(\delta) = \{h \in \mathcal{K} : \exists l = l(h) | \langle f'(0), l\rangle - f_\delta''(0)[h, h] \le 0, F'(0)l - F''(0)[h, h] \in C\}.$$

Очевидно, что если $h \in H$, то

$$\lambda_0 f_\delta''(0)[h, h] + \langle y^*, F''(0)[h, h]\rangle \ge 0 \quad \forall \lambda = (\lambda_0, y^*) \in \Lambda. \tag{8}$$

Могут представиться две возможности. Пусть вначале $H = \mathcal{K}$. Возьмем произвольное $\lambda \in \Lambda$. Тогда в силу (8) имеем

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda)[h, h] + \lambda_0\delta|h|^2 \ge 0 \quad \forall h \in \mathcal{K}. \tag{9}$$

Положим $\Pi(\delta) = \ker F'(0)$. Тогда $\operatorname{codim}\Pi(\delta) \le k - 1$, так как $\operatorname{im}F'(0) \ne Y$ в силу анормальности точки $x_0 = 0$. Покажем, что $\Pi(\delta) \subseteq \mathcal{K}$. Действительно, если $x \in \Pi(\delta)$, то $F'(0)x = 0$, откуда, в силу (4), $\langle f'(0), x\rangle = 0 \Rightarrow x \in \mathcal{K}$. Таким образом, в силу (9) получаем, что соотношения (5)–(7) выполняются для произвольных $h \in \mathcal{K}$.

Рассмотрим вторую возможность. Пусть $H \ne \mathcal{K}$. Зафиксируем произвольный вектор $h \in \mathcal{K}$. Предположим вначале, что $h \notin H$. Для $\varepsilon = 1/i$, $i = 1, 2, \ldots$, рассмотрим так называемую ε-задачу:

$$f_{\varepsilon,\delta}(x, \chi) := f_\delta(x) - \chi f_\delta(\varepsilon h) + \gamma(\chi) \longrightarrow \min,$$
$$F(x) - \chi(F(\varepsilon h) - \varepsilon F'(0)h) \in C,$$
$$\chi \ge 0, \quad |x| \le c.$$

Здесь $\gamma(\chi) = (\max\{0, \chi - 1\})^4$, а число $c > 0$ выбрано так, чтобы точка $x_0 = 0$ являлась глобальным минимумом исходной задачи (P) в c-окрестности нуля. Будем рассматривать только те значения ε, для которых $\varepsilon \le c/|h|$. Очевидно, точка $(\varepsilon h, 1)$ удовлетворяет всем ограничениям ε-задачи, так как, в силу предположения в), $F'(0)h \in C$.

Повторяя рассуждения, приведенные в [6], несложно получить, что каждая из ε-задач имеет решение $(x_\varepsilon, \chi_\varepsilon)$, для которого $\chi_\varepsilon > 0$ и, кроме того, $x_\varepsilon \longrightarrow 0$ при $\varepsilon \longrightarrow 0$.

Покажем, что $x_\varepsilon \neq 0$ $\forall\varepsilon$. Очевидно,

$$f_{\varepsilon,\delta}(x_\varepsilon, \chi_\varepsilon) \le f_{\varepsilon,\delta}(0, 0) = 0 \Rightarrow f_\delta(x_\varepsilon) - \chi_\varepsilon f_\delta(\varepsilon h) \le 0, \tag{10}$$

$$y_\varepsilon := F(x_\varepsilon) - \chi_\varepsilon(F(\varepsilon h) - \varepsilon F'(0)h) \in C. \tag{11}$$

Предположим, что $x_\varepsilon = 0$. Тогда $f_\delta(x_\varepsilon) = 0$, $F(x_\varepsilon) = 0$, и, раскладывая f и F в окрестности нуля в ряд Тейлора до членов второго порядка, из (10), (11) получаем

$$-\chi_\varepsilon(\varepsilon^2 f''_\delta(0)[h, h] + o(\varepsilon^2)) \le 0, \quad -\chi_\varepsilon(\varepsilon^2 F''(0)[h, h] + o(\varepsilon^2)) \in C.$$

Здесь использовалось то, что $h \in \mathcal{K} \Rightarrow \langle f'(0), h\rangle \le 0$. Поскольку $\chi_\varepsilon > 0$, то из полученных соотношений, деля их на $\varepsilon^2\chi_\varepsilon > 0$, имеем $h \in H$ (при $l = 0$). Пришли к противоречию, так как, по сделанному предположению, $h \notin H$, а значит, $x_\varepsilon \neq 0$ $\forall\varepsilon$.

Положим $l_\varepsilon = x_\varepsilon/|x_\varepsilon|$, $\nu_\varepsilon = \chi_\varepsilon\varepsilon^2/2|x_\varepsilon| > 0$. Переходя к подпоследовательности, имеем $l_\varepsilon \longrightarrow l$ и $|l| = 1$. Покажем, что $\nu_\varepsilon \longrightarrow 0$.

Из формул (10), (11), деля на $|x_\varepsilon|$ и учитывая, что $\langle f'(0), h\rangle \le 0$, получаем

$$\langle f'(0), l_\varepsilon\rangle + O(|x_\varepsilon|) - \nu_\varepsilon f''_\delta(0)[h, h] + \nu_\varepsilon o(\varepsilon^2)/\varepsilon^2 \le 0,$$

$$F'(0)l_\varepsilon + O(|x_\varepsilon|) - \nu_\varepsilon F''(0)[h, h] + \nu_\varepsilon o(\varepsilon^2)/\varepsilon^2 \in C.$$

Если ν_ε не стремится к нулю, то, переходя к подпоследовательности, получаем, что либо $\nu_\varepsilon \longrightarrow \nu > 0$, либо $\nu_\varepsilon \longrightarrow +\infty$. Разделив полученные соотношения на ν_ε и устремив ε к нулю, в обоих случаях получим, что $h \in H$. Но это противоречит сделанному предположению $h \notin H$. Следовательно, $\nu_\varepsilon \longrightarrow 0$.

Применим к каждой из ε-задач необходимые условия минимума, полученные в [3]. Положим $s_\varepsilon = \dim(Z(y_\varepsilon))^\perp$. В силу необходимых условий второго порядка из [3], существуют множитель Лагранжа $\lambda_\varepsilon = (\lambda_{0,\varepsilon}, y_\varepsilon^*)$, $y_\varepsilon^* \in N_C(y_\varepsilon)$, $|\lambda_\varepsilon| = 1$, и линейное подпространство Π_ε такие, что

$$\Pi_\varepsilon \subseteq \ker(P^\perp(y_\varepsilon)F'(x_\varepsilon)), \quad \Pi_\varepsilon \subseteq \ker f'_\delta(x_\varepsilon), \quad \operatorname{codim}\Pi_\varepsilon \le s_\varepsilon, \tag{12}$$

$$\lambda_{0,\varepsilon} f'_\delta(x_\varepsilon) + F'(x_\varepsilon)^* y_\varepsilon^* = 0, \tag{13}$$

$$\lambda_{0,\varepsilon} f_\delta(\varepsilon h) + \langle y_\varepsilon^*, (F(\varepsilon h) - \varepsilon F'(0)h)\rangle \ge 0, \tag{14}$$

$$\lambda_{0,\varepsilon} f''_\delta(\varepsilon h)[\xi, \xi] + \langle y_\varepsilon^*, F''(x_\varepsilon)[\xi, \xi]\rangle \ge 0 \quad \forall\xi \in \Pi_\varepsilon. \tag{15}$$

Переходя к подпоследовательности, имеем $\lambda_\varepsilon \longrightarrow \lambda = (\lambda_0, y^*)$, $s_\varepsilon = s_0$ при $\varepsilon \longrightarrow 0$. По теореме 4.4 из [2], существует линейное подпространство $\Pi \subseteq \operatorname{Ls}\{\Pi_\varepsilon\}$: $\operatorname{codim}\Pi \le s_0$. Здесь $\operatorname{Ls}\{\Pi_\varepsilon\}$ обозначает верхний топологический предел последовательности множеств $\{\Pi_\varepsilon\}$. Он состоит из тех векторов $\xi \in X$, для каждого из которых существует такая последовательность точек $\{\xi_\varepsilon\}$, что $\xi_\varepsilon \in \Pi_\varepsilon$ $\forall\varepsilon$ и ξ является предельной точкой последовательности $\{\xi_\varepsilon\}$. (Подробности см. в [2].) Учитывая конечность числа граней конуса C и переходя к подпоследовательности, будем не ограничивая общности считать, что $Z(y_\varepsilon) = Z$, $P^\perp(y_\varepsilon) = P_Z^\perp$ $\forall\varepsilon$, где Z – некоторое подпространство в Y (а именно, это линейная оболочка какой-то из граней конуса C), а $P_Z^\perp$ – оператор ортогонального проектирования на $Z^\perp$. Ясно, что при этом $s_0 = \dim Z^\perp$ и

$$y_\varepsilon = F(x_\varepsilon) - \chi_\varepsilon(F(\varepsilon h) - \varepsilon F'(0)h) \in Z \quad \forall\varepsilon. \tag{16}$$

Предположим вначале, что $f'(0) \notin \operatorname{im}(P_Z^\perp F'(0))^*$. В этом случае $s_0 < k$ в силу (4). В силу (4) и того, что $f'(0) \notin \operatorname{im}(P_Z^\perp F'(0))^*$ получим выражение

$$\exists e_1 \in Z : e_1 \neq 0, \quad F'(0)^* e_1 \in \operatorname{Lin}\{f'(0)\} + \operatorname{im}(P_Z^\perp F'(0))^*. \tag{17}$$

Дополним вектор e_1 до базиса $e_1, \ldots, e_{k-s_0}$ подпространства Z. Из (12) при $\varepsilon \longrightarrow 0$ имеем

$$\operatorname{Ls}\{\Pi_\varepsilon\} \subseteq \ker(P_Z^\perp F'(0))^* \cap \ker f'(0) \Rightarrow \Pi \subseteq \ker(P_Z^\perp F'(0))^* \cap \ker f'(0),$$

откуда, в силу (17), $\Pi \subseteq \ker(F'(0)^* e_1)$.

Положим

$$\Pi(\delta) = \bigcap_{i=2}^{k-s_0} \ker(F'(0)^* e_i) \cap \Pi.$$

Тогда, в силу доказанного, $\Pi(\delta) \subseteq \ker F'(0)$ и $\operatorname{codim}\Pi(\delta) \le k-1$. Поэтому, переходя в соотношениях (13)–(15) к пределу при $\varepsilon \longrightarrow 0$, получаем (5)–(7).

Пусть теперь $f'(0) \in \operatorname{im}(P_Z^\perp F'(0))^*$. По построению $P_Z = P_i$ для некоторого номера $i \in \{1, 2, \ldots, m\}$, и пусть для удобства $i = 1$. Переобозначая, для удобства положим $Q := Q_1$, $Q_0 := Q_{0,1}$, $L := L_1$, $N := N_1$, $\zeta := \xi_1$, $\phi := \phi_1$, $M_j := M_j^1$, $j = 1, 2$.

Деля обе части включения (16) на $|x_\varepsilon|$, после чего переходя к пределу при $\varepsilon \longrightarrow 0$ с учетом того, что $\nu_\varepsilon \longrightarrow 0$, получаем $F'(0)l \in Z \Rightarrow l \in L$.

Раскладывая f и F в (13) в ряд Тейлора в нуле и используя то, что $Z \perp N_C(y_\varepsilon) \Rightarrow P_Z^\perp y_\varepsilon^* = y_\varepsilon^*$, получаем[2)]

$$\lambda_{0,\varepsilon}(f_\delta'(0) + f_\delta''(0)x_\varepsilon) + (F'(0) + F''(0)x_\varepsilon)^* P_Z^\perp y_\varepsilon^* + o(|x_\varepsilon|) = 0. \tag{18}$$

По сделанному предположению, $f'(0) \in \operatorname{im}(P_Z^\perp F'(0))^*$, откуда $\langle f_\delta'(0), \xi\rangle = 0$ $\forall \xi \in L$. Поэтому, деля равенство (18) на $|x_\varepsilon|$ и переходя к пределу при $\varepsilon \longrightarrow 0$, получаем

$$\lambda_0(Q_0[l, \xi] + \delta\langle l, \xi\rangle) + \langle y^*, Q[l, \xi]\rangle = 0 \quad \forall \xi \in L. \tag{19}$$

Переходя в (15) к пределу при $i \longrightarrow \infty$ и используя при этом определение верхнего топологического предела, получаем

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda)[\xi, \xi] + \delta|\xi|^2 \ge 0 \quad \forall \xi \in \Pi. \tag{20}$$

Покажем теперь, что $Q[l, \xi] \in N^\perp$ $\forall \xi \in \Pi$. В самом деле, возьмем произвольное $\xi \in \Pi$. По определению верхнего топологического предела существует такая последовательность $\{\xi_\varepsilon\}$, что $\xi_\varepsilon \in \Pi_\varepsilon$ $\forall \varepsilon$, и после перехода к подпоследовательности имеет место $\xi_\varepsilon \longrightarrow \xi$. Поэтому, в силу (12), $P_Z^\perp F'(x_\varepsilon)\xi_\varepsilon = 0$, откуда

$$\langle y, P_Z^\perp(F'(0)\xi_\varepsilon + F''(0)[x_\varepsilon, \xi_\varepsilon])\rangle + o(|x_\varepsilon|) = 0 \quad \forall y \in N.$$

Разделив это соотношение на $|x_\varepsilon|$, учтя, что $y \in N$, и перейдя к пределу при $\varepsilon \longrightarrow 0$, получим $\langle y, Q[l, \xi]\rangle = 0$ $\forall y \in N$. Это равенство доказано для произвольного $\xi \in \Pi$. Точно так же (с помощью (12)) доказывается, что $Q_0[l, \xi] = -\delta\langle l, \xi\rangle$ $\forall \xi \in \Pi$. Таким образом, доказано, что

$$Q[l, \xi] \in N^\perp, \quad Q_0[l, \xi] = -\delta\langle l, \xi\rangle \quad \forall \xi \in \Pi. \tag{21}$$

Покажем, что существует такое линейное подпространство $\tilde{\Pi} \subseteq \ker(P_Z^\perp F'(0))$, что

$$\operatorname{codim}\tilde{\Pi} \le s_0 - 1, \quad \frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda)[\xi, \xi] + \delta|\xi|^2 \ge 0 \quad \forall \xi \in \tilde{\Pi}. \tag{22}$$

Если $l \notin \Pi$, то из (19) и (20) легко вытекает, что подпространство $\tilde{\Pi} = \Pi + \{\alpha l, \alpha \in \mathbb{R}^1\}$ является искомым. Пусть теперь $l \in \Pi$. Тогда, в силу (21), $\phi(l) = -\delta$, $Q[l, l] \in N^\perp$. Поэтому $\lambda^0 = 0$, так как в противном случае, в силу (19), мы имели бы $(l, y^*/\lambda^0 - \zeta, \delta) \in M_1$, что привело бы к противоречию, так как, в силу выбора числа $\delta > 0$, должно быть $\phi(l) \ne -\delta$.

Докажем, что

$$\exists \hat{l} \in L,\ \hat{\mu} \in N: \quad Q[l, \hat{l}] \in N^\perp,\ \hat{\mu}Ql + y^*Q\hat{l} \in L^\perp,\ \langle (Q_0 l + \delta l), \hat{l}\rangle \ne 0. \tag{23}$$

[2)] $F''(0)x_\varepsilon$ – линейный оператор, действующий из X в Y по формуле $(F''(0)x_\varepsilon)x = F''(0)[x_\varepsilon, x]$.

Действительно, учитывая (19) и (21), имеем $\langle l, l\rangle = 1$, $y^*Ql \in L^\perp$, $Q[l, l] \in N^\perp$. Предположим, что (23) нарушается. Тогда

$$\langle (Q_0 l + \delta l), x\rangle = 0 \quad \forall x \in L,\ \mu \in N:\quad Q[l, x] \in N^\perp,\ \mu Ql + y^*Qx \in L^\perp.$$

(Обозначения см. в Приложении.) Отсюда вытекает[3)] существование таких $\bar\mu$, $\bar\xi$, что $(l, y^*, \delta, \bar\mu, \bar\xi) \in M_2$. С другой стороны, $\phi(l) = -\delta$ и, значит, в силу выбора $\delta > 0$, имеем $(l, y^*, \delta, \mu, \xi) \notin M_2$ $\forall \mu, \xi$. Полученное противоречие доказывает (23).

Положим $\tilde\Pi = \Pi + \{\alpha\hat l, \alpha \in \mathbb{R}^1\}$. В силу (21), (23), $\hat l \notin \Pi$ и, значит, $\operatorname{codim}\tilde\Pi \le s_0 - 1$. В то же время, в силу (23), $y^*Q[\hat l, \hat l] = 0$, $y^*Q[\hat l, \xi] = 0$ $\forall \xi \in \Pi$. Поэтому $y^*Q[\xi + \alpha\hat l, \xi + \alpha\hat l] = y^*Q[\xi, \xi]$ $\forall \xi \in \Pi$, $\alpha \in \mathbb{R}^1$, и, следовательно, в силу (20), имеет место (22). Легко видеть, что подпространство

$$\Pi(\delta) = \bigcap_{i=1}^{k-s_0} \ker(F'(0)^* e_i) \cap \tilde\Pi,$$

где $\{e_i\}$ – произвольный базис подпространства Z, удовлетворяет условиям (5)–(7).

Пусть теперь $h \in H$. Из сделанного предположения, что $\mathcal{K}\backslash H \ne \emptyset$, в силу доказанного выше, вытекает существование вектора λ и подпространства $\Pi(\delta)$, которые удовлетворяют условиям (5), (6). Поэтому, в силу (8), для них имеет место (7).

Таким образом, доказано, что в случае 1 для всех достаточно малых $\delta > 0$ для произвольного $h \in \mathcal{K}$ существуют множитель Лагранжа λ и линейное подпространство $\Pi(\delta)$, для которых выполняются соотношения (5)–(7). Осуществляя в этих соотношениях предельный переход при $\delta \longrightarrow 0$, завершаем рассмотрение первого случая.

Случай 2. Предположим, что

$$f'(0) \notin \operatorname{im}(F'(0)^*). \tag{24}$$

Рассмотрим множество

$$\tilde H = \{h \in \mathcal{K} : \exists l = l(h),\ |F'(0)l - F''(0)[h, h] \in C\}.$$

Очевидно, что для любых $\lambda = (\lambda_0, y^*) \in \Lambda$ имеет место

$$\langle y^*, F''(0)[x, x]\rangle \ge 0 \quad \forall x \in \tilde H, \tag{25}$$

поскольку, в силу (24) и (2), $\lambda_0 = 0$ $\forall \lambda = (\lambda_0, y^*) \in \Lambda$.

Пусть вначале $\tilde H = \mathcal{K}$. Положим $\Pi = \ker F'(0)$. В силу анормальности точки $x_0 = 0$ имеем $\operatorname{im}F'(0) \ne$ $\ne Y \Rightarrow \operatorname{codim}\Pi \le k - 1$. Возьмем произвольный $\lambda \in \Lambda$ ($\Lambda \ne \emptyset$ в силу принципа Лагранжа). Заметим, что если $x \in \Pi$, но $x \notin \mathcal{K}$, то $-x \in \mathcal{K}$. Отсюда в силу (25) и того, что квадратичная форма является четной функцией, вытекает, что (25) выполняется для всех $x \in \Pi$ и, следовательно, $\lambda \in \Lambda_{k-1}(0)$, $\frac{\partial^2\mathcal{L}}{\partial x^2}(0, \lambda)[h, h] \ge 0$ $\forall h \in \mathcal{K}$, что завершает рассмотрение случая $\tilde H = \mathcal{K}$.

Пусть теперь $\tilde H \ne \mathcal{K}$. Возьмем произвольный вектор $h \in \mathcal{K}\backslash\tilde H$, для которого $\langle f'(0), h\rangle < 0$.

Зафиксируем $\delta > 0$ такое, что $\delta < \langle f'(0), h\rangle^2/|h|^2$, и рассмотрим задачу

$$\Xi_\delta(x) = -(\min\{0, \langle f'(0), x\rangle\})^2 + \delta|x|^2 \longrightarrow \min,\quad F(x) \in C. \tag{26}$$

Покажем, что точка $x_0 = 0$ является локальным минимумом в задаче (26). Действительно, предположим противное. Тогда существует такая последовательность $\{x_i\}$, что $x_i \longrightarrow 0$, $i \longrightarrow \infty$, $F(x_i) \in C$, $\Xi_\delta(x_i) < 0$ $\forall i$. Несложно показать, что тогда найдется такой касательный вектор l ко множеству $D = \{x \in X : F(x) \in C\}$, что после перехода к подпоследовательности выполняется

[3)] При этом мы используем следующее утверждение из линейной алгебры. Пусть X, Y – линейные конечномерные подпространства, $X_1 \subseteq X$ и $Y_1 \subseteq Y$, $A : X \longrightarrow Y$ – линейный оператор и a – заданный вектор из X. Тогда если $\langle a, x\rangle = 0$ $\forall x \in X_1 : Ax \in Y_1^\perp$, то $\exists \xi \in Y_1 : a - A^*\xi \in X_1^\perp$.

$x_i = |x_i|l + o(|x_i|)$. Но так как $x_0 = 0$ является точкой минимума в исходной задаче (Р), то, как известно (см. [2, с. 74]), имеет место $\langle f'(0), l\rangle \le 0$. Отсюда, в силу определения функции Ξ_δ и неравенства $\delta > 0$, непосредственно вытекает, что $\Xi_\delta(x_i) > 0$ при всех больших i. Полученное противоречие доказывает, что $x_0 = 0$ – локальный минимум в задаче (26).

Минимизируемая функция Ξ_δ непрерывна на $\mathbb{R}^n$ и является дважды непрерывно дифференцируемой на множестве $\{x : \langle f'(0), x\rangle \ne 0\}$. Покажем, что тем не менее к задаче (26) применимы все рассуждения, проведенные для случая 1.

Рассмотрим определенную выше ε-задачу при $f_\delta := \Xi_\delta$. Пусть $(x_\varepsilon, \chi_\varepsilon)$ – ее решение, и $\chi_\varepsilon > 0$. Тогда $\langle f'(0), x_\varepsilon\rangle < 0$. Действительно, в силу определения Ξ_δ, если $\langle f'(0), x_\varepsilon\rangle \ge 0$, то $\Xi_\delta(x_\varepsilon) = \delta|x_\varepsilon|^2 \ge 0$. Отсюда, в силу выбора числа δ, и того, что $\langle f'(0), x_\varepsilon\rangle < 0$, имеем $f_{\varepsilon,\delta}(x_\varepsilon, \chi_\varepsilon) > 0$. В то же время точка $(\varepsilon h, 1)$ удовлетворяет всем ограничениям ε-задачи и $f_{\varepsilon,\delta}(\varepsilon h, 1) = 0$. Полученное противоречие доказывает, что $\langle f'(0), x_\varepsilon\rangle < 0$. Следовательно, функция Ξ_δ дважды непрерывно дифференцируема в окрестности каждой из точек x_ε. Поэтому к ε-задаче применимы все рассуждения, проведенные при рассмотрении случая 1, но в которых множество H заменено на $\tilde{H}$. Проводя эти рассуждения и осуществляя затем предельный переход при $\delta \longrightarrow 0$, получаем, что существуют вектор $\lambda = (\lambda_0, y^*) \in \Lambda$ и подпространство $\Pi \subseteq \ker F'(0)$, $\operatorname{codim}\Pi \le k - 1$ такие, что

$$-\lambda_0\langle f'(0), h\rangle^2 + \langle y^*, F''(0)[h, h]\rangle \ge 0, \quad \langle y^*, F''(0)[\xi, \xi]\rangle \ge 0 \quad \forall\xi \in \Pi.$$

Отсюда следует, что $y^* \ne 0$, $(0, y^*) \in \Lambda_{k-1}(0)$, а также $\langle y^*, F''(0)[h, h]\rangle \ge 0$ и, значит, для рассматриваемого h имеет место (3).

Пусть теперь $h \in \mathcal{K}\backslash\tilde{H}$, $\langle f'(0), h\rangle = 0$. В силу (24), $\exists\xi \in \ker F'(0) : \langle f'(0), \xi\rangle < 0$. Поэтому $\xi \in \mathcal{K} \Rightarrow$ $\Rightarrow h + i^{-1}\xi \in \mathcal{K}$ для всех натуральных i. В то же время, в силу замкнутости множества $\tilde{H}$, при всех больших i имеем $(h + i^{-1}\xi) \in \mathcal{K}\backslash\tilde{H}$. Применяя доказанное выше утверждение к векторам $(h + i^{-1}\xi)$ и переходя к пределу при $i \longrightarrow \infty$, получаем (3) для векторов $h \in \mathcal{K}\backslash\tilde{H}$. Остается отметить, что для векторов $h \in \tilde{H}$ неравенство (3) вытекает из только что доказанной непустоты множества $\Lambda_{k-1}(0)$ и неравенства (25). Это завершает рассмотрение случая 2. Таким образом, в предположениях этапа I теорема доказана.

Этап II. Пусть выполняются предположения а) и б) этапа I, а C – произвольное выпуклое множество. Пусть $h \in \mathcal{K}$, $w \in O_C^2(0, F'(0)h)$. Выберем произвольный вектор r, принадлежащий $\operatorname{ri}C$ – относительной внутренности выпуклого множества C ($\operatorname{ri}C \ne \emptyset$, так как Y конечномерно). По лемме 2 (см. ниже), кривая

$$\varphi(\tau) = \tau F'(0)h + \tau^2(w + r)/2 \tag{27}$$

целиком принадлежит множеству C при всех малых $\tau \ge 0$. Пусть B_1 – единичный шар с центром в нуле в пространстве Y. Зафиксируем произвольное $\varepsilon > 0$. Выберем во множестве $C \cap B_1$ конечную ε-сеть $\{a_i\}_{i=1}^N$, $N = N(\varepsilon)$, т.е.

$$a_i \in C \cap B_1,\ i = 1, 2, \ldots, N,\ \forall y \in C \cap B_1 \quad \exists j \in \{1, 2, \ldots, N\} : |y - a_j| \le \varepsilon.$$

Возьмем произвольное $\alpha \in (0, 1/N)$ и рассмотрим задачу минимизации

$$\begin{gathered} f(x) \longrightarrow \min, \\ F(x) = \alpha N\sum_{i=1}^{N} a_i\theta_i/N + (1 - \alpha N)\varphi(\tau), \\ 0 \le \theta_i, \quad i = 1, 2, \ldots, N, \quad 0 \le \tau, \end{gathered} \tag{P_α}$$

относительно неизвестных x, τ, $\theta = (\theta_1, \ldots, \theta_N)$. По построению, в силу выпуклости множества C точка $x = 0$, $\tau = 0$, $\theta = 0$ является локальным минимумом в рассматриваемой задаче. Применим к задаче (P_α) необходимые условия минимума, полученные на этапе I доказательства. Для нее конус критических направлений в нуле имеет вид

$$\mathcal{K}_\alpha = \{(\bar{x}, \bar{\tau}, \bar{\theta}) \in \mathbb{R}^n \times \mathbb{R}^1 \times \mathbb{R}^N : \bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_N), \bar{\theta}_i \ge 0, \bar{\tau} \ge 0,$$
$$\langle f'(0), \bar{x} \rangle \le 0, F'(0)\bar{x} = \alpha(a_1\bar{\theta}_1 + \ldots + a_N\bar{\theta}_N) + (1 - \alpha N)\varphi'(0)\bar{\tau}\}.$$

В силу (27), $\left(h, \frac{1}{1-\alpha N}, 0\right) \in \mathcal{K}_\alpha$. Подставляя этот вектор в неравенство (3) (это необходимое условие минимума для задачи типа (P_α) было доказано на этапе I), после несложной расшифровки получаем, что существуют вектор $\lambda_\alpha = (\lambda_{0,\alpha}, y_\alpha^*)$, $|\lambda_\alpha| = 1$, и подпространство $\Pi_\alpha \subseteq \ker F'(0) : \operatorname{codim}\Pi_\alpha \le$ $\le k - 1$ такие, что

$$\frac{\partial \mathcal{L}}{\partial x}(0, \lambda_\alpha) = 0, \quad \lambda_{0,\alpha} \ge 0, \quad \langle y_\alpha^*, a_i \rangle \le 0, \quad i = 1, 2, \ldots, N,$$

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda_\alpha)[\xi, \xi] \ge 0 \quad \forall \xi \in \Pi_\alpha,$$

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(0, \lambda_\alpha)[h, h] - \frac{1}{1-\alpha N}\langle y_\alpha^*, \varphi''(0) \rangle \ge 0.$$

Но, в силу (27), $\varphi''(0) = w + r$. Поэтому, переходя стандартным образом в полученных условиях последовательно к пределу вначале при $\alpha \longrightarrow 0$, затем при $\varepsilon \longrightarrow 0$ и, наконец, при $r \longrightarrow 0$, $r \in \operatorname{ri} C$ (см. подробности в [2]), находим вектор $\lambda \in \Lambda_{k-1}(0)$, для которого

$$\frac{\partial^2 \mathcal{L}}{\partial x^2}(x_0, \lambda)[h, h] - \langle y^*, w \rangle \ge 0. \tag{28}$$

Это завершает доказательство неравенства (3) в предположениях, сделанных на этапе II.

Этап III. Пусть выполняется предположение а). На этом этапе доказательства избавимся от предположения $\mathcal{I} = \{0\}$. Итак, пусть $\mathcal{I} \ne \{0\}$. Обозначим через $P_\mathcal{I}$ и $P_\mathcal{I}^\perp$ операторы ортогонального проектирования на подпространства $\mathcal{I}$ и $\mathcal{I}^\perp$ соответственно. Рассмотрим задачу

$$f(x) \longrightarrow \min, \quad P_\mathcal{I}^\perp F(x) \in P_\mathcal{I}^\perp C. \tag{29}$$

Покажем, что точка $x = 0$ является локальным минимумом в задаче (29).

Для этого, очевидно, достаточно показать, что если точка x удовлетворяет ограничениям задачи (29), то она удовлетворяет и ограничениям исходной задачи (P). Действительно, пусть x удовлетворяет ограничениям задачи (29). Тогда

$$P_\mathcal{I} F(x) \in \mathcal{I}, \quad P_\mathcal{I}^\perp F(x) \in P_\mathcal{I}^\perp C.$$

Складывая левые части этих включений, имеем $F(x) \in \mathcal{I} + P_\mathcal{I}^\perp C$. По лемме 3, $P_\mathcal{I}^\perp C = C \cap \mathcal{I}^\perp \subseteq C$. Значит, $F(x) \in \mathcal{I} + C$. Поэтому, в силу инвариантности подпространства $\mathcal{I}$, имеем $F(x) \in C$, что и требовалось доказать.

Пусть $h \in \mathcal{K}$, $w \in T_C^2(0, F'(0)h)$. Рассмотрим для задачи (29) конус критических направлений в нуле. Он имеет вид

$$\mathcal{K}_\mathcal{I} = \left\{\xi \in X : \left\langle \frac{\partial f}{\partial x}(x_0), \xi \right\rangle \le 0, P_\mathcal{I}^\perp F'(0)\xi \in T_{P_\mathcal{I}^\perp C}(0)\right\}.$$

Легко заметить, что $h \in \mathcal{K}_\mathcal{I}$ (поскольку $AT_C = T_{AC}$, где A – линейный оператор) и, в силу определения внутреннего касательного множества второго порядка, $P_\mathcal{I}^\perp w \in T^2_{P_\mathcal{I}^\perp C}(0, P_\mathcal{I}^\perp F'(0)h)$.

Применяя в задаче (29) к векторам h и $P_{\mathcal{J}}^{\perp}w$ результаты, полученные на этапе II, используя то, что $P_{\mathcal{J}}^{\perp}y^* = y^*\ \forall(\lambda_0, y^*) \in \Lambda \Rightarrow \langle P_{\mathcal{J}}^{\perp}w, y^*\rangle = \langle w, y^*\rangle$, находим вектор $\lambda \in \Lambda_{s-1}(0)$, для которого имеет место (28). Это завершает рассмотрения этапа III.

Этап IV. Чтобы доказать теорему в полной общности, нам осталось избавиться от предположения $X = \mathbb{R}^n$. А это делается дословным повторением рассуждений, проведенных на последнем этапе доказательства теоремы 4.1 из [2, с. 37–38].

Теорема доказана.

Замечание. Утверждение теоремы остается в силе, если в определении множества $\Lambda_r(x_0)$ инвариантное подпространство $\mathcal{J}$ заменить на локально-инвариантное подпространство, а именно: пусть для удобства $F(x_0) = 0$. Линейное подпространство I называется локально-инвартантным относительно множества C, если $\exists\varepsilon_1, \varepsilon_2 > 0 : I \cap B_{\varepsilon_1} + C \cap B_{\varepsilon_1} \subseteq C \cap B_{\varepsilon_2}$, где B_ε – шар радиуса ε с центром в нуле. Возможность замены инвариантного подпространства на локально-инвартантное вытекает из того, что рассматриваемый минимум на этапе III доказательства является локальным.

Для $h \in X$ через $\sigma(\cdot, h)$ обозначим опорную функцию множества

$$T_C^2\left(F(x_0), \frac{\partial F}{\partial x}(x_0)h\right), \quad \text{т.е.} \quad \sigma(y^*, h) = \sup\left\{\langle y, y^*\rangle, y \in T_C^2\left(F(x_0), \frac{\partial F}{\partial x}(x_0)h\right)\right\}, \quad y^* \in Y.$$

Теорема 2. *Пусть точка x_0 является локальным минимумом в задаче* (P) *и анормальна. Тогда $\Lambda_{s-1}(x_0) \neq \varnothing$ и для любого $h \in \mathcal{K}(x_0)$ имеет место*

$$\max_{\lambda \in \operatorname{conv}\Lambda_{s-1}(x_0)}\left\{\frac{\partial^2\mathcal{L}}{\partial x^2}(x_0, \lambda)[h, h] - \sigma(y^*, h)\right\} \geq 0, \quad \lambda = (\lambda_0, y^*),$$

где conv *– выпуклая оболочка множества.*

Доказательство теоремы 2 выводится из теоремы 1 дословным повторением рассуждений, приведенных на этапе IV доказательства теоремы 2.1 из [3].

ПРИЛОЖЕНИЕ

Пусть заданы линейные подпространства $L \subseteq \mathbb{R}^n$, $N \subseteq \mathbb{R}^k$ и вещественные симметричные $n \times n$-матрицы Q_i, $i = \overline{0, k}$. Определим билинейное отображение $Q : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^k$ по формуле

$$Q[x_1, x_2] = (\langle Q_1x_1, x_2\rangle, \ldots, \langle Q_kx_1, x_2\rangle), \quad x_1, x_2 \in \mathbb{R}^n.$$

Для $y = (y^1, \ldots, y^k) \in \mathbb{R}^k$ положим $yQ = \sum_{i=1}^k y^iQ_i$. Введем в рассмотрение множества

$$M_1 = \{(x, y, y^0) \in L \times N \times \mathbb{R}^1 : \langle x, x\rangle = 1, Q[x, x] \in N^\perp, (Q_0 + yQ)x + y^0x \in L^\perp\}, \tag{30}$$

$$\begin{gathered} M_2 = \{(x, y, y^0, \mu, \xi) \in L \times N \times \mathbb{R}^1 \times N \times L : \langle x, x\rangle = 1, Q[x, x] \in N^\perp, \\ yQx \in L^\perp, (Q_0x + y^0x) + \mu Qx + yQ\xi \in L^\perp, Q[x, \xi] \in N^\perp\}. \end{gathered} \tag{31}$$

Положим $\phi(x, y, y^0, \mu, \xi) = \phi(x) = \langle Q_0x, x\rangle$.

Лемма 1. *На каждом из множеств M_1 и M_2 функция ϕ принимает лишь конечное число значений.*

Доказательство. Нам понадобятся некоторые дополнительные сведения из вещественной алгебраической геометрии. Далее будем использовать некоторые обозначения и результаты из [7]. Множество $A \subset \mathbb{R}^n$ называется полуалгебраическим, если оно может быть представлено в виде конечного числа объединений или пересечений множеств вида $Z = \{x : p(x) = 0\}$ или $S = \{x : p(x) < 0\}$, где $p(x)$ – полиномы (с вещественными коэффициентами) от n переменных $x_1, \ldots, x_n$. В предложении 2.2.4 из [7] доказано, что любое множество, записанное в виде формулы, содержащей конечное число операций объединения, пересечения, отрицания над множествами вида Z и S, а также кванторы существования и единственности по переменным $x_1, \ldots, x_n$, будет полуалгебраическим. Полуалгебраическое множество может быть, очевидно, не замкнутым, однако замыкание и внутренность полуалгебраического множества всегда есть полуалгебраическое множество. Для полуалгебраического множества A отображение $f : A \longrightarrow \mathbb{R}^k$ называется полуалгебраическим, если его график является полуалгебраическим множеством в $\mathbb{R}^{n+k}$. Если A – интервал, отрезок или полуинтервал, то $\gamma = f$ называется полуалгебраической кривой.

Кривой Нэша называется заданная на интервале (0, 1) полуалгебраическая кривая, координаты которой суть алгебраические аналитические функции. При этом функция $\phi : U \longrightarrow \mathbb{R}^1$, где U – открытое полуалгебраическое подмножество $\mathbb{R}^n$, называется алгебраической аналитической (ее еще называют функцией Нэша), если она является решением уравнения

$$p_m(x)[\phi(x)]^m + \ldots + p_1(x)\phi(x) + p_0(x) = 0, \quad x \in U,$$

где $p_0, p_1, \ldots, p_m$ – полиномы, причем $p_m \neq 0$. Оказывается, что функции Нэша – это в точности бесконечно дифференцируемые полуалгебраические функции (см. [7, гл. 8]). Наконец, нам понадобится понятие нэш-диффеоморфизма между двумя полуалгебраическими множествами. Последнее означает, что отображение, задающее диффеоморфизм, а также его обратное являются бесконечно дифференцируемыми полуалгебраическими отображениями. Подмногообразием Нэша называется множество M, которое нэш-диффеоморфно открытому шару в $\mathbb{R}^{\dim(M)}$, где $\dim(M)$ – размерность подмногообразия. При этом одноточечное множество по определению является нэш-подмногообразием размерности 0. Приведем предложение 2.9.10 из [7].

Предложение. *Пусть $S \subset \mathbb{R}^n$ – полуалгебраическое множество. Тогда S представимо в виде объединения конечного числа попарно непересекающихся подмногообразий Нэша.*

Доказательство леммы 1. Множества M_1, M_2 являются вещественными алгебраическими многообразиями. Нам нужно показать, что функция $\phi(x) = \langle Q_0x, x\rangle$ на каждом из M_1 и M_2 принимает лишь конечное число значений. Сделаем это, используя элементы вещественной алгебраической геометрии (см. [7]). Действительно, в силу предложения, всякое алгебраическое многообразие представимо в виде дизъюнктного объединения конечного числа подмногообразий Нэша. Пусть $\overline{M}$ – одно из них, т.е. оно нэш-диффеоморфно открытому шару в некотором подпространстве. Тогда оно связно и, более того, любые две точки из $\overline{M}$ можно соединить кривой Нэша, лежащей в $\overline{M}$ (в силу нэш-диффеоморфизма).

Итак, пусть M_1 и M_2 представимы в виде дизъюнктного объединения конечного числа подмногообразий Нэша $M_{1,j}$ и $M_{2,j}$. Докажем, что на каждом из них функция ϕ постоянна. В силу сказанного выше, для этого достаточно показать, что если $z(t)$, $t \in [0, 1]$ – гладкая кривая (т.е. $z(t)$ непрерывна на [0, 1] и бесконечно дифференцируема на (0, 1)), лежащая в одном из подмногообразий $M_{1,j}$ и $M_{2,j}$, то производная функции ϕ вдоль этой кривой тождественно равна нулю. Действительно, рассмотрим вначале подмногообразие $M_{1,j}$. Пусть дана гладкая кривая $z(t) = (x(t), y(t), y^0(t)) \in M_{1,j}$, $t \in [0, 1]$. Положим $\phi(t) = \phi(x(t))$. Имеем $\dot{\phi}(t)/2 = \langle Q_0x(t), \dot{x}(t)\rangle$. Дифференцируя по t тождество $\langle x(t), x(t)\rangle \equiv 1$ и включение $Q[x(t), x(t)] \in N^\perp$, имеем

$$\langle x(t), \dot{x}(t)\rangle \equiv 0, \quad Q[x(t), \dot{x}(t)] \in N^\perp. \tag{32}$$

Из определения M_1 получаем, что $Q_0x(t) \in -[y(t)Qx(t) + y^0(t)x(t)] + L^\perp$. Умножая последнее тождество скалярно на $\dot{x}(t)$, используя (32), получаем $\langle Q_0x(t), \dot{x}(t)\rangle \equiv 0 \Rightarrow \dot{\phi}(t) \equiv 0$.

Подмногообразия $M_{2,j}$ рассматриваются аналогично, а именно: пусть дана гладкая кривая

$$z(t) = (x(t), y(t), y^0(t), \mu(t), \xi(t)) \in M_{2,j}, \quad t \in [0, 1].$$

В силу определения M_2 имеем $Q[x(t), \xi(t)] \in N^\perp\ \forall t \in [0, 1]$. Отсюда, дифференцируя по t, получаем

$$Q[\dot{x}(t), \xi(t)] + Q[x(t), \dot{\xi}(t)] \in N^\perp \quad \forall t \in [0, 1].$$

Умножая левую часть полученного включения скалярно на $y(t)$ и учитывая, что в силу определения множества M_2 имеет место $y(t)Qx(t) \in L^\perp$, получаем

$$y(t)Q[\xi(t), \dot{x}(t)] \in L^\perp \quad \forall t \in [0, 1]. \tag{33}$$

Умножая левую часть включения $[Q_0x(t) + y^0(t)x(t)] + \mu(t)Qx(t) + y(t)Q\xi(t) \in L^\perp\ \forall t \in [0, 1]$ скалярно на $\dot{x}(t)$, в силу (32) и (33) получаем $\langle Q_0x(t), \dot{x}(t)\rangle \equiv 0 \Rightarrow \dot{\phi}(t) \equiv 0$. Лемма доказана.

Лемма 2. *Пусть C – выпуклое замкнутое множество и даны векторы $d, w \in Y$. Предположим, что $y(t) = td + t^2w + o(t^2) \in C$ для всех достаточно малых $t \ge 0$. Тогда для любого вектора r, принадлежащего относительной внутренности* $\mathrm{ri}\,C$ *множества C, существует $\varepsilon(r) > 0$ такое, что*

$$\hat{y}(t) = td + t^2(w + r) \in C \quad \forall t \in [0, \varepsilon(r)].$$

Доказательство. Для малых $t \ge 0$ обозначим через $\tau = \tau(t)$ наименьшее положительное решение уравнения $t = (1 - \tau^2)\tau$. Подставляя это выражение в формулу для $\hat{y}(t)$, имеем

$$\hat{y}(t) = td + t^2(w + r) = (1 - \tau^2)\tau d + (1 - \tau^2)^2\tau^2(w + r) =$$
$$= (1 - \tau^2)(\tau d + \tau^2 w) + \tau^2[(1 - \tau^2)^2 r - \tau^2(1 - \tau^2)w].$$

Здесь и ниже $\tau = \tau(t)$. Положим $\Delta(t) = y(t) - td - t^2w$. Тогда $\Delta(t) = o(t^2)$. Прибавляя к предыдущему равенству выражение $(1 - \tau^2)\Delta(\tau)$ и одновременно его вычитая, после очевидных преобразований получаем

$$\begin{gathered}\hat{y}(t) = (1 - \tau^2)(\tau d + \tau^2 w + \Delta(\tau)) + \\ + \tau^2[(1 - \tau^2)^2 r - \tau^2(1 - \tau^2)w - (1 - \tau^2)o(\tau^2)/\tau^2].\end{gathered} \tag{34}$$

При этом для малых $\tau \ge 0$ имеет место $y(\tau) = \tau d + \tau^2 w + o(\tau^2) \in C$, а также $(1 - \tau^2)^2 r - \tau^2(1 - \tau^2)w - (1 - \tau^2)\Delta(\tau)/\tau^2 \in \mathrm{ri}\,C$, так как $r \in \mathrm{ri}\,C$ и $\Delta(\tau)/\tau^2 \longrightarrow 0$ при $\tau \longrightarrow 0$. Поэтому из (34) в силу выпуклости множества C вытекает, что $\hat{y}(t) \in C$ для всех достаточно малых $t \ge 0$. Лемма доказана.

Лемма 3. *Пусть $C \subseteq Y$ – выпуклое замкнутое подмножество и $0 \in C$, а M – линейное подпространство, для которого $M \subseteq C$, и P – оператор ортогонального проектирования на $M^\perp$. Тогда $PC = C \cap M^\perp$.*

Доказательство. Включение $C \cap M^\perp \subseteq PC$ очевидно. Докажем включение $C \cap M^\perp \supseteq PC$. Пусть $y \in PC$. Покажем, что $y \in C$. Для этого, в силу замкнутости C, достаточно показать, что $\alpha y \in C$ $\forall \alpha \in (0, 1)$. Действительно, $y \in PC \Rightarrow \exists z \in C$: $y = Pz$. Далее имеем очевидное равенство

$$\alpha y = (1 - \alpha)\left(\frac{\alpha}{1 - \alpha}(y - z)\right) + \alpha z, \quad \alpha \in (0, 1). \tag{35}$$

Но $y - z \in M$ по определению. Поэтому в правой части (35) стоит выпуклая комбинация двух элементов из C. В силу выпуклости C из (35), $\alpha y \in C$, что и требовалось доказать.

СПИСОК ЛИТЕРАТУРЫ

1. *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. М.: Наука, 1979.
2. *Арутюнов А.В.* Условия экстремума. М.: Факториал, 1997.
3. *Arutyunov A., Pereira F.L.* Second-order necessary optimality conditions for problems without a priori normality assumptions // Math. Operat. Res. 2006. V. 31. № 1. P. 1–12.
4. *Арутюнов А.В., Карамзин Д.Ю.* Необходимые условия экстремума в анормальной задаче с ограничениями типа равенств // Ж. вычисл. матем. и матем. физ. 2006. Т. 46. № 8. С. 1363–1368.
5. *Иоффе А.Д., Тихомиров В.М.* Теория экстремальных задач. М.: Наука, 1974.
6. *Arutyunov A.V.* Optimality conditions: Abnormal and degenerate problems. Dordrecht: Kluwer Acad. Publs, 2000.
7. *Bochnak J., Coste M., Roy M.F.* Real Algebraic Geometry // Ser. Mod. Math. Springer, 1998.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 376–396

УДК 519.626.2

О НЕКОТОРЫХ ЗАДАЧАХ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ И ИХ РАЗНОСТНЫХ АППРОКСИМАЦИЯХ И РЕГУЛЯРИЗАЦИИ ДЛЯ КВАЗИЛИНЕЙНЫХ ЭЛЛИПТИЧЕСКИХ УРАВНЕНИЙ С УПРАВЛЕНИЯМИ В КОЭФФИЦИЕНТАХ

© 2007 г. Ф. В. Лубышев, А. Р. Манапова

(450074 Уфа, ул. Фрунзе, 32, Башкирский гос. ун-т)

e-mail: LubyshevFV@bsu.bashedu.ru

Поступила в редакцию 18.07.2006 г.

Рассматриваются и исследуются математические постановки нелинейных задач оптимального управления для квазилинейных эллиптических уравнений с управлениями в переменных коэффициентах уравнения состояния. Рассмотрены как локальные, так и интегральные ограничения на управления. Функционалы цели соответствуют оптимизации по некоторому конечному числу критериев качества. Построены разностные аппроксимации экстремальных задач, и установлены оценки точности аппроксимаций по состоянию и функционалу, доказана слабая сходимость по управлению. Проведена регуляризация аппроксимаций по Тихонову. Рассмотрены содержательные примеры некоторых прикладных оптимизационных задач, математические постановки которых естественным образом приводят к поставленным и исследованным в настоящей работе нелинейным задачам оптимального управления. Библ. 23.

Ключевые слова: задача оптимального управления, квазилинейные эллиптические уравнения, разностный метод решения, метод регуляризации.

ВВЕДЕНИЕ

Характер конкретных постановок задач оптимального управления для распределенных систем существенно зависит от того, куда входят управления: в свободные члены уравнений состояния или в коэффициенты уравнений, а также линейными или нелинейными дифференциальными уравнениями математической физики (УМФ) описываются состояния систем (см. [1]–[5]). В настоящее время наиболее полно исследован случай, когда управления достаточно простым образом входят в линейные уравнения состояния и линейные предельные условия (в правые части линейных уравнений, граничных и начальных условий). Менее изучены задачи оптимального управления нелинейного типа, характерной особенностью которых является то, что отображение $g \longrightarrow u(g)$ из множества допустимых управлений U в пространство состояний W является нелинейным. Особый интерес представляют задачи оптимального управления, когда нелинейность обусловлена вхождением управлений в коэффициенты уравнений для состояний, в том числе в коэффициенты при старших производных. Такие задачи весьма существенно отличаются от задач, где управление осуществляется путем внешних воздействий на систему и которые наименее изучены, хотя развитие теории и методов их решения вызвано потребностями математического моделирования нелинейных оптимальных процессов, большой прикладной важностью таких задач при оптимизации процессов теплофизики, диффузии, фильтрации, теории упругости и др., а также при решении обратных задач для УМФ, рассматриваемых в вариационной постановке. Задачи оптимального управления с управлениями в коэффициентах и в первую очередь с управляющими параметрами, содержащимися в главной части дифференциального оператора (в старших коэффициентах) являются “сильно нелинейными” оптимизационными задачами. Нелинейность еще более усугубляется, если, кроме того, и состояния процессов описываются нелинейными уравнениями.

Проблема численного решения задач оптимального управления приводит к необходимости их аппроксимации задачами более простой природы – “конечномерными задачами” (см. [1]). Правильно построенная аппроксимация позволяет получить содержательные результаты качественного и численного характера об изучаемом процессе. Обзор работ, посвященных основам общей теории и методов устойчивости и аппроксимации экстремальных задач, в том числе аппроксимаций задач оптимального управления и результатов в данной области, представлен, например, в [1], [6]–[9]. Одним из наиболее удобных, универсальных и широко распространенных методов конечномерных аппроксимаций задач оптимального управления является метод сеток (см. [10], [11]). Центральными в проблеме аппроксимации являются вопросы “конструирования” аппроксимаций, сходимости аппроксимаций по состоянию, функционалу, управлению, регуляризации аппроксимаций. Для систем с распределенными параметрами построения и исследования аппроксимаций задач оптимального управления проводились в основном также для линейных задач оптимального управления. Поэтому особенно актуальным является рассмотрение этих вопросов для нелинейных задач оптимального управления (в том числе для задач, когда нелинейность обусловлена вхождением управлений в переменные коэффициенты уравнений состояния и/или нелинейностью самих уравнений состояния).

В настоящей работе, по тематике примыкающей к [9], [12]–[17], рассмотрены и исследованы математические постановки задач оптимального управления для квазилинейных уравнений эллиптического типа с управлениями в коэффициентах уравнения состояния, отвечающих различным видам “управляющих воздействий”: управления в переменных коэффициентах при старших производных, младших производных, в переменном коэффициенте нелинейного члена уравнения, зависящего от функции состояния и различными вариантами критериев оптимальности (функционалов цели). Рассмотрены как локальные, так и интегральные ограничения на управления. Построены разностные аппроксимации исходных экстремальных задач, и установлены оценки скорости сходимости аппроксимаций по состоянию и функционалу, слабая сходимость по управлению. Никакие дополнительные априорные требования на гладкость обобщенных решений для состояния при этом не накладываются. Причем для аппроксимации уравнения состояния предложена некоторая “модифицированная” разностная схема, отличная от традиционных схем другим способом вычисления переменных коэффициентов в главной части сеточного оператора. Проведена регуляризация аппроксимаций. Рассмотренные постановки задач включают в себя в качестве частных вариантов постановок большой круг конкретных прикладных оптимизационных задач теории теплопроводности, конвекции-диффузии-реакции, теории упругости и др. (при соответствующей конкретизации уравнений состояния, управляющих воздействий, ограничений на управления и функционалов цели, соответствующих оптимизации по некоторому конечному числу критериев качества).

Так, например, в теории упругости возникает множество задач, которые естественно формируются как проблемы оптимального управления. При этом роль управляющих факторов в этих задачах могут выполнять, например, функции, входящие в главную часть основного дифференциального оператора, задающие внутреннюю структуру конструкций, т.е. описывающие распределение упругих характеристик материала. Рассмотрим, например, следующий содержательный частный вариант постановок экстремальных **задач** $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$ (см. разд. 1), когда в уравнении состояния (1.1) $b_\alpha(x) = 0$, $\alpha = 1, 2$, $d(x) = 0$, $f(x) = 2\theta$, а коэффициент $k(x)$ – управление; множества допустимых управлений для $g(x) = k(x)$ имеют вид либо U_0, либо $\hat{U}_0$, либо $\hat{U}_0(p)$, $p \geq 1$, а функционал цели $J(g) = \sum_{k=0}^{3} \alpha_k J_k(g)$, соответствующий оптимизации по заданному конечному числу критериев качества, имеет вид (1.3), где $\rho(x) = 2/\theta$. Подобная модель для функции состояния $u(x)$, $x \in \Omega$, возникает, например, при математическом моделировании важной в теории упругости задачи о кручении упругого изотропного неоднородного призматического стержня с заданным односвязным поперечным сечением $\Omega \subset \mathbb{R}^2$ (см. [4]). Предполагается, что боковая поверхность стержня свободна от напряжений и стержень нагружен только при помощи закручивающего момента $M_{x_3} = \theta J_0(g)$, приложенного на его торцах и направленного по оси стержня x_3. Здесь θ – угол закручивания, приходящийся на единицу длины стержня, вызванный крутящим моментом M_{x_3}. Этим углом характеризуется деформированное состояние стержня. Напряженное состояние в стержне, деформированном при помощи закручивающего момента M_{x_3} относительно оси x_3, определяется только двумя компонентами напряжения $\sigma_{13}(x)$ и $\sigma_{23}(x)$ (так как возникает только два касательных напряжения). В приложениях большой интерес представляет не только определение функции напряжений Прандтля $u(x) = u(x_1, x_2)$, $x \in \Omega$, удовлетворяющей прямой задаче для состояния, но и такие важные характеристики, как касательные напряжения $\sigma_{13}(x)$, $\sigma_{23}(x)$, а также жесткость стержня на кручение $J_0(g) = J_0(k)$. Для данного частного варианта экстремальные задачи $\mathbf{A}^{(\alpha)}$ можно трактовать как оптимальные задачи теории упругости об экстремуме функционала, соответствующего оптимизации по конечному числу критериев качества $J(g) = \sum_{k=0}^{3} \alpha_k J_k(g)$. Здесь $J_0(g) = J_0(k)$ – один из основных функционалов решения задачи о кручении. Этот функционал характеризует жесткость кручения неоднородного упругого изотропного призматического стержня. Второй функционал $J_1(k)$ характеризует меру отклонения в $L_2(\Omega)$-норме функции напряжений Прандтля $u(x)$, характеризующей состояние системы, от заданного (желаемого) распределения $u_0(x)$. Третий и четвертый функционалы $J_2(k)$ и $J_3(k)$ характеризуют среднеквадратичное отклонение действующих в стержне касательных (сдвиговых) напряжений $\partial u/\partial x_1 = \sigma_{23}(x)$ и $\partial u/\partial x_2 = \sigma_{13}(x)$ от заданных (желаемых) касательных напряжений $\psi_1(x)$ и $\psi_2(x)$ соответственно. Роль "управляющего" фактора в этих экстремальных задачах $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, выполняет функция $g(x) = k(x)$ – упругая податливость материала призматического стержня (величина, обратная $\mu(x)$ – модулю упругого сдвига материала, из которого состоит стержень), описывающая распределение упругих характеристик материала. Локальные ограничения на управление $g(x) = k(x)$ характеризуют границы допустимого изменения упругой податливости материала стержня, а также учитывают недопустимость резких перепадов в изменении упругих характеристик материала (ограничение на градиент функции – управления $k(x)$). Интегральные соотношения в ограничениях на управление $k(x)$ можно трактовать как условные оценки стоимости применяемого материала (см. [4]). Величина функционала $J(g) = \sum_{k=0}^{3} \alpha_k J_k(g)$ зависит от распределения податливости $k(x)$ по сечению Ω призматического стержня. Целью в каждой из задач $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, является отыскание такого распределения податливости $k(x) \in$ $\in U^{(\alpha)}$, где $U^{(\alpha)} = U_0$, либо $U^{(\alpha)} = \hat{U}_0$, либо $U^{(\alpha)} = \hat{U}_0(p)$, $p \ge 1$ (см. разд. 1), чтобы на этом управлении функционал $J(g) = J(k)$ достигал минимума. В зависимости от выбора параметров α_k, $k = 0, 1, 2, 3$, получаем различные частные критерии качества. Таким образом, поставленные в настоящей работе экстремальные задачи $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, в рассмотренном выше частном содержательном варианте можно трактовать как задачи оптимального распределения модуля сдвига материала упругого неоднородного изотропного призматического стержня, находящегося в условиях кручения. Заметим, что задача о максимизации жесткости на кручение упругого призматического стержня состоит в отыскании $\inf \tilde{J}_0(k)$, где $\tilde{J}_0(k) = -J_0(k)$. Заметим также, что в [4] рассмотрены более частные постановки экстремальных задач о кручении. Кроме того, вопросы, связанные с аппроксимацией экстремальных задач о кручении, там не затрагивались.

Рассмотрим другой содержательный частный вариант постановок экстремальных задач $\mathbf{A}^{(\alpha)}$, когда в уравнении состояния (1.1) $b_\alpha(x) = 0$, $\alpha = 1, 2$, $d(x)q(u) = d(x)u(x)$ (т.е. принято $q(u) = u$), а коэффициенты $k(x)$ и $d(x)$ – управления; множества допустимых управлений для вектор-функции управления $g = (k(x), d(x))$ имеют вид либо $U_0 \times U_3$, либо $\hat{U}_0 \times U_3$, либо $\hat{U}_0(p) \times U_3$, $p \ge 1$, а функционал цели $J(g)$, соответствующий оптимизации по заданному конечному числу критериев качества, имеет вид (1.3), где $\rho(x) = f(x)$, $\alpha_2 = \alpha_3 = 0$. В данном частном варианте задача для состояния $u(x)$ может трактоваться как задача о положении равновесия жестко закрепленной на границе Γ неоднородной упругой мембраны, находящейся под действием системы сил – внешней нормальной стационарной нагрузки с непрерывной поверхностной плотностью $f(x)$ и силы сопротивления упругой среды, плотность которой $d(x)u(x)$ пропорциональна смещению $u(x)$ точек мембраны и обратна этому смещению по знаку (предполагается, что окружающая мембрану среда оказывает сопротивление перемещению мембраны, пропорциональное смещению точек мембраны). Причем в спокойном, ненагруженном состоянии мембрана расположена в горизонтальной плоскости (x_1, x_2) и занимает область $\Omega \subset \mathbb{R}^2$, ограниченную границей Γ. Здесь $u(x) = u(x_1, x_2)$, $x \in \Omega$, – функция, определяющая прогиб мембраны (т.е. нормальное перемещение мембраны в точке $x = (x_1, x_2)$). Для данного частного варианта экстремальные задачи $\mathbf{A}^{(\alpha)}$ можно трактовать как оптимальные задачи об экстремуме функционала $J(g) = \alpha_0 J_0(g) + \alpha_1 J_1(g)$, в котором функционал $J_0(g)$ выражает значение работы внешней нормальной силы плотности $f(x)$ на возможных перемещениях $u(x)$ точек мембраны, а второй функционал $J_1(g)$ характеризует величину среднеквадратичного отклонения положения (формы) мембраны от требуемого положения (формы) $u_0(x)$. Роль "управляющих" факторов в этих экстремальных задачах выполняют функции $k(x)$ и $d(x)$, $x \in \Omega$, где $k(x) \ge \nu >$ > 0 – коэффициент натяжения мембраны, характеризующий внутренние упругие свойства материала мембраны, а $d(x) \ge 0$ – коэффициент упругости окружающей среды, характеризующий упругую восстанавливающую силу окружающей мембрану упругой среды. Величина функционала $J(g) = = J(k, d)$ зависит от распределения величин $k(x)$ и $d(x)$. Цель задач оптимизации $\mathbf{A}^{(\alpha)}$ состоит в том, что требуется отыскать допустимые функции $k(x)$ и $d(x)$ такие, чтобы функционал

$J(g) = \alpha_0 J_0(g) + + \alpha_1 J_1(g)$ принимал минимальное возможное значение. В зависимости от выбора параметров α_0 и α_1 получаем различные частные критерии качества. Например, минимизация функционала $J_1(g)$ отвечает требованию, чтобы прогиб нагруженной мембраны, закрепленной по краю Γ, был близок в $L_2(\Omega)$-норме к заданному прогибу $u_0(x)$. Таким образом, поставленные в настоящей работе экстремальные задачи в рассмотренном выше частном содержательном варианте можно трактовать так же, как задачи оптимального распределения коэффициента натяжения упругой мембраны и коэффициента упругости окружающей мембрану среды, когда упругая неоднородная мембрана находится под действием системы сил (нагрузок).

Экстремальные задачи $\mathbf{A}^{(\alpha)}$, поставленные в разд. 1, в других содержательных частных вариантах можно трактовать, например, и как оптимальные задачи экологического прогнозирования, как проблемы, связанные с оптимизацией параметров математических моделей экологического прогнозирования, в основе которых лежит уравнение конвекции-диффузии-реакции, описывающее процесс распространения вещества в некоторой области экологического прогнозирования Ω (например, переноса в атмосфере или в реке неоднородной загрязняющей субстанции, см. [18], [19]).

1. ПОСТАНОВКА ЗАДАЧ И ИХ КОРРЕКТНОСТЬ

Пусть $\Omega = \{\xi = (\xi_1, \xi_2) \in \mathbb{R}^2 : 0 < \xi_\alpha < l_\alpha, \alpha = 1, 2\}$ – прямоугольник в $\mathbb{R}^2$ с границей Γ. Пусть управляемый процесс описывается в Ω следующей задачей Дирихле для квазилинейного уравнения эллиптического типа:

$$L(g)u = -\sum_{\alpha=1}^{2} \frac{\partial}{\partial \xi_\alpha}\left(k(\xi)\frac{\partial u}{\partial \xi_\alpha}\right) + \sum_{\alpha=1}^{2} b_\alpha(\xi)\frac{\partial u}{\partial \xi_\alpha} + d(\xi)q(u) = f(\xi), \quad \xi \in \Omega, \tag{1.1а}$$

$$u(\xi) = 0, \quad \xi \in \Gamma, \tag{1.1б}$$

где $q(u)$, $f(\xi)$ – известные функции; $g(\xi) = (g_0(\xi), g_1(\xi), g_2(\xi), g_3(\xi)) = (k(\xi), b_1(\xi), b_2(\xi), d(\xi))$ – управление. Относительно заданных функций будем предполагать следующее: $f(\xi) \in L_2(\Omega)$, функция $q(s)$ определена на R со значениями в R и удовлетворяет условиям $q(0) = 0$, $0 \le q_0 \le [q(s_1) - q(s_2)]/(s_1 - s_2) \le$ $\le L_q < \infty$, для всех $s_1, s_2 \in R$, $s_1 \ne s_2$.

Введем множество допустимых управлений

$$U^{(0)} = \prod_{k=0}^{3} U_k \subset B = W_\infty^1(\Omega) \times (L_\infty(\Omega))^3,$$

$$U_0 = \left\{k(\xi) \in W_\infty^1(\Omega) : 0 < \nu \le k(\xi) \le \bar{\nu} < \infty, \left|\frac{\partial k}{\partial \xi_\alpha}\right| \le R_\alpha, \alpha = 1, 2, \text{ п.в. на } \Omega\right\},$$
$$U_\alpha = \{b_\alpha(\xi) \in L_\infty(\Omega) : \zeta_\alpha \le b_\alpha(\xi) \le \bar{\zeta}_\alpha, \text{ п.в. на } \Omega\}, \quad \alpha = 1, 2,$$
$$U_3 = \{d(\xi) \in L_\infty(\Omega) : \zeta_3 \le d(\xi) \le \bar{\zeta}_3, \text{ п.в. на } \Omega\}; \tag{1.2}$$

здесь п.в. – почти всюду. Предполагается выполнение следующих условий: $-m \le \zeta_1 \le \bar{\zeta}_1 \le m$, $-p \le$ $\le \zeta_2 \le \bar{\zeta}_2 \le p$, $\bar{\nu}$, R_α, m, $p = \text{Const} > 0$; $\delta_0 = \max\limits_{\substack{\varepsilon_1, \varepsilon_2 > 0 \\ \varepsilon_1 + \varepsilon_2 \le \nu}} \left\{\frac{\nu - (\varepsilon_1 + \varepsilon_2)}{C_\Omega^2} + \lambda - \frac{m^2}{4\varepsilon_1} - \frac{p^2}{4\varepsilon_2}\right\} > 0$, $C_\Omega^2 = \left(\frac{8}{l_1^2} + \frac{8}{l_2^2}\right)^{-1}$; ζ_3, $\bar{\zeta}_3$ – некоторые постоянные; здесь λ – любая из следующих констант: а) $\lambda = q_0\zeta_3$, $\zeta_3 \ge 0$; б) $\lambda = \zeta_3$ – любая константа, когда $q(u) = u$; в) $\lambda = -L_q\zeta_0$, где $\zeta_0 = \max\{|\zeta_3|, |\bar{\zeta}_3|\}$.

Зададим функционал цели $J : U^{(0)} \longrightarrow R^1$ в виде

$$g \longrightarrow J(g) = \sum_{k=0}^{3} \alpha_k J_k(g) = I(u(\xi; g)), \tag{1.3}$$

где отображение $I : V \longrightarrow R^1$, задаваемое на классе V решений задачи (1.1) (см. определение V ниже), определяется выражением

$$u \longrightarrow I(u) = \alpha_0\int_\Omega \rho(\xi)ud\Omega + \alpha_1\int_\Omega |u-u_0(\xi)|^2 d\Omega + \int_\Omega \sum_{k=1}^{2}\alpha_{k+1}\left|\frac{\partial u}{\partial \xi_k} - \psi_k(\xi)\right|^2 d\Omega, \tag{1.4}$$

где $u_0(\xi)$, $\psi_k(\xi)$ – заданные функции из $W_2^1(\Omega)$, $k = 1, 2$, а $\rho(\xi)$ – заданная функция из $L_2(\Omega)$, $\alpha_m =$ $=$ Const ≥ 0, $m = 0, 1, 2, 3$, $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 > 0$. Под решением прямой задачи (1.1) при фиксированном управлении $g \in U^{(0)}$ понимается функция $u(\xi) \equiv u(\xi; g) \in \overset{0}{W}{}_2^1(\Omega) = V$, удовлетворяющая для всех $\eta \in \overset{0}{W}{}_2^1(\Omega) = V$ тождеству

$$Q(g; u, \eta) = \int_\Omega \left\{\sum_{\alpha=1}^{2} k(\xi)\frac{\partial u}{\partial \xi_\alpha}\frac{\partial \eta}{\partial \xi_\alpha} + \sum_{\alpha=1}^{2} b_\alpha(\xi)\frac{\partial u}{\partial \xi_\alpha}\eta + d(\xi)q(u)\eta\right\}d\Omega = \int_\Omega f(\xi)\eta d\Omega = l(\eta). \tag{1.5}$$

Поставим задачу оптимального управления.

Задача A$^{(0)}$. Найти управление $g_* \in U^{(0)}$ такое, что $J(g_*) = I(u(\xi; g_*)) = \inf\limits_{g \in U^{(0)}} I(u(\xi; g)) = \inf\limits_{g \in U^{(0)}} J(g) =$ $= J_*^{(0)}$, где функционал цели $J : U^{(0)} \longrightarrow R^1$ имеет вид (1.3), $g \longrightarrow J(g) = I(u(\xi; g))$, причем $I : V =$ $= \overset{0}{W}{}_2^1(\Omega) \longrightarrow R^1$ – отображение, определяемое выражением (1.4), $u \longrightarrow I(u)$, где $u = u(\xi; g) \in V$ – решение задачи состояния (1.5).

Используя теорию монотонных операторов (см. [20]), а также результаты из [21], можно показать, что при любом $g \in U^{(0)}$ существует единственное обобщенное решение $u \equiv u(\xi; g) \in V$ задачи (1.5) и справедлива оценка

$$\|u(\xi; g)\|_V \leq C\|f(\xi)\|_{L_2(\Omega)}, \tag{1.6}$$

где $C =$ Const > 0, не зависящая от управления g; более того, обобщенное (из $V = \overset{0}{W}{}_2^1(\Omega)$) решение краевой задачи (1.1) принадлежит также пространству $W_{2,0}^2(\Omega) = W_2^2(\Omega) \cap \overset{0}{W}{}_2^1(\Omega)$ и при каждом фиксированном управлении $g \in U$ справедлива оценка

$$\|u(\xi; g)\|_{W_2^2(\Omega)} \leq C\|f(\xi)\|_{L_2(\Omega)}, \tag{1.7}$$

где $C =$ Const > 0, не зависящая от управления g.

Нетрудно видеть, что множество U_0 слабо компактно в пространстве $W_2^1(\Omega)$, так как U_0 ограничено по норме в $W_2^1(\Omega)$. Точно так же множество U_0 слабо компактно в $W_\infty^1(\Omega)$.

Сузим теперь множество допустимых управлений U_0 для компоненты $k(\xi)$ вектор-функции управления g до множества $\hat{U}_0$ (см. ниже), присоединяя к множеству ограничений U_0 дополнительное ограничение (равенство изопериметрического типа)

$$\int_\Omega k(\xi)d\xi = M, \text{ где } M = \text{Const} > 0. \tag{1.8}$$

В результате получим следующее определение подмножества $\hat{U}_0$ функций из U_0, удовлетворяющих дополнительному условию (1.8):

$$\hat{U}_0 = \left\{ k(\xi) \in W^1_\infty(\Omega) : 0 < \nu \le k(\xi) \le \bar{\nu} < \infty, \left|\frac{\partial k}{\partial \xi_\alpha}\right| \le R_\alpha, \alpha = 1, 2, \text{ п.в. на } \Omega; \int\limits_\Omega k(\xi)d\xi = M \right\} =$$
$$= \left\{ k(\xi) \in U_0 : \int\limits_\Omega k(\xi)d\xi = M \right\}. \tag{1.9}$$

Множество допустимых управлений $\hat{U}_0$ учитывает наряду с локальными ограничениями U_0 и изопериметрическое ограничение на управление $k(\xi)$. При этом предполагаем, что положительные постоянные ν, $\bar{\nu}$, R_1, R_2, M таковы, что множество (1.9) не пусто. В частности, естественно предположить, что положительная постоянная M такова, что выполняется условие $\nu|\Omega| < M < \bar{\nu}|\Omega|$, где $|\Omega| = \operatorname{mes}\Omega$ – мера множества Ω. Определяя теперь множество допустимых управлений $U^{(1)}$ для вектор-функции управления $g(\xi)$ в виде

$$U^{(1)} = \hat{U}_0 \times U_1 \times U_2 \times U_3 \subset U^{(0)} \subset B, \tag{1.10}$$

поставим следующую задачу оптимального управления.

Задача $\mathbf{A}^{(1)}$. Найти управление $g_* \in U^{(1)}$ такое, что $J(g_*) = I(u(\xi; g_*)) = \inf\limits_{g \in U^{(1)}} I(u(\xi; g)) = \inf\limits_{g \in U^{(1)}} J(g) = J_*^{(1)}$, где функционал цели $J : U^{(1)} \longrightarrow R^1$ имеет вид (1.3), $g \longrightarrow J(g) = I(u(\xi; g))$, причем $I : V \longrightarrow R^1$ – отображение, определяемое выражением (1.4), $u \longrightarrow I(u)$, где $u = u(\xi; g) \in V$ – решение задачи (1.5).

Сузим теперь множество допустимых управлений U_0 для компоненты $k(\xi)$ вектор-функции управления g, присоединяя к множеству ограничений U_0 дополнительное ограничение (в отличие от (1.8)) вида

$$\int\limits_\Omega k^p(\xi)d\xi \le M^p|\Omega|, \quad \nu \le M \le \bar{\nu}, \tag{1.11}$$

где $p \ge 1$ – целое число. В результате получим определение подмножества $\hat{U}_0(p)$ функций из U_0, удовлетворяющих дополнительному условию (1.11):

$$\hat{U}_0(p) = \left\{ k(\xi) \in W^1_\infty(\Omega) : 0 < \nu \le k(\xi) \le \bar{\nu} < \infty, \left|\frac{\partial k}{\partial \xi_\alpha}\right| \le R_\alpha, \alpha = 1, 2, \text{ п.в. на } \Omega; \right.$$
$$\left. \int\limits_\Omega k^p(\xi)d\xi \le M^p|\Omega|, \nu < M < \bar{\nu} \right\} = \left\{ k(\xi) \in U_0 : \int\limits_\Omega k^p(\xi)d\xi \le M^p|\Omega| \right\}, \tag{1.12}$$

где $|\Omega| = \operatorname{mes}\Omega$ – мера множества Ω, $p \ge 1$ – целое число. При этом предполагаем, что положительные постоянные ν, $\bar{\nu}$, R_1, R_2 таковы, что множество (1.12) не пусто. Определяя теперь для вектор-функции управления $g(\xi)$ множество допустимых управлений в виде

$$U^{(2)} \equiv U^{(2)}(p) = \hat{U}_0(p) \times U_1 \times U_2 \times U_3 \subset U^{(0)} \subset B, \quad p \ge 1, \tag{1.13}$$

поставим следующую задачу оптимального управления.

Задача $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$. Найти управление $g_* \in U^{(2)} \equiv U^{(2)}(p)$ такое, что $J(g_*) = I(u(\xi; g_*)) = \inf\limits_{g \in U^{(2)}(p)} I(u(\xi; g)) = \inf\limits_{g \in U^{(2)}(p)} J(g) = J_*^{(p)}$, где $p \ge 1$ – целое число, функционал цели $J : U^{(2)}(p) \longrightarrow R^1$ имеет вид (1.3), $g \longrightarrow J(g) = I(u(\xi; g))$, причем $I : V \longrightarrow R^1$ – отображение, определяемое выражением (1.4), $u \longrightarrow I(u)$, где $u = u(\xi; g) \in V$ – решение задачи (1.5).

Справедлива теорема о разрешимости экстремальных задач $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$.

Теорема 1. *Пусть* $U^{(0)}$, $U^{(1)}$, $U^{(2)} \equiv U^{(2)}(p)$ – *множества допустимых управлений экстремальных задач* $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ *соответственно. Существует, по крайней мере, одно опти-*

мальное управление $g_* \in U^{(\alpha)}$ *задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, т.е. $J_*^{(0)} = \inf\{J(g): g \in U^{(0)}\} > -\infty$, $U_*^{(0)} = \{g_* \in U^{(0)} : J(g_*) = J_*^{(0)}\} \neq \varnothing$; $J_*^{(1)} = \inf\{J(g) : g \in U^{(1)}\} > -\infty$, $U_*^{(1)} = \{g_* \in U^{(1)} : J(g_*) = J_*^{(1)}\} \neq \varnothing$; $J_*^{(2)} = J_*^{(2)}(p) = \inf\{J(g) : g \in U^{(2)} = U^{(2)}(p)\} > -\infty$, $U_*^{(2)} = U_*^{(2)}(p) = \{g_* \in U^{(2)} = U^{(2)}(p) : J(g_*) = J_*^{(2)} = J_*^{(2)}(p)\} \neq \varnothing$. *Множества точек минимума* $U_*^{(0)}$, $U_*^{(1)}$, $U_*^{(2)} = U_*^{(2)}(p)$ *функционала цели* $J(g)$ *в задачах* $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ *соответственно слабо компактны в* $H = W_2^1(\Omega) \times (L_2(\Omega))^3$. *Любая минимизирующая последовательность* $\{g^{(n)}\} = \{(g_0^{(n)}, g_1^{(n)}, g_2^{(n)}, g_3^{(n)})\}_{n=1}^{\infty} \subset U^{(\alpha)}$ *функционала* $J(g)$ *соответствующих экстремальных задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *слабо в* H *сходится к соответствующим множествам* $U_*^{(\alpha)}$, $\alpha = 0, 1, 2$, *точек минимума функционала* $J(g)$.

Доказательство. Можно показать, что множества $U^{(0)}$, $U^{(1)}$ и $U^{(2)} \equiv U^{(2)}(p)$ при любом целом $p \geq 1$ являются слабо бикомпактными в пространстве $H = W_2^1(\Omega) \times (L_2(\Omega))^3$. При этом при доказательстве используется, в частности, тот факт, что так как $\Omega \subset \mathbb{R}^2$, то оператор вложения пространства $W_2^1(\Omega)$ в $L_p(\Omega)$ непрерывен (и даже компактен) для любого конечного $p \geq 1$. Дальнейшее доказательство утверждений теоремы 1 будет опираться на следующее свойство отображения $g \longrightarrow u(\xi, g)$.

Лемма 1. *Пусть* $U^{(0)}$, $U^{(1)}$, $U^{(2)} \equiv U^{(2)}(p)$ – *множества допустимых управлений экстремальных задач* $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ *соответственно. Пусть, далее,* $\{g^{(n)}\}_{n=1}^{\infty} = \{(g_0^{(n)}, g_1^{(n)}, g_2^{(n)}, g_3^{(n)}) = (k^{(n)}(\xi), b_1^{(n)}(\xi), b_2^{(n)}(\xi), d^{(n)}(\xi))\}_{n=1}^{\infty} \subset U^{(\alpha)}$ – *произвольная последовательность, а* $u^{(n)} \equiv u(\xi, g^{(n)})$ – *решение задачи* (1.5) *при* $g = g^{(n)}$. *Тогда из условия* $g^{(n)} \longrightarrow \bar{g} = (\bar{g}_0, \bar{g}_1, \bar{g}_2, \bar{g}_3) = (\bar{k}(\xi), \bar{b}_1(\xi), \bar{b}_2(\xi), \bar{d}(\xi))$ *слабо в* $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ *следует, что* $u^{(n)}(\xi) = u(\xi, g) \longrightarrow u(\xi, \bar{g}) = \bar{u}(\xi)$ *слабо в* $W_2^2(\Omega)$, *где* $\bar{u}(\xi) = u(\xi, \bar{g})$ – *решение задачи* (1.5) *при* $g = \bar{g}$, *т.е. отображение* $g \longrightarrow u(\xi, g)$ *является слабо непрерывным из* $U^{(\alpha)}$ *в* $W_2^2(\Omega)$, *переводящим слабую сходимость в* $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ – *пространстве управлений на множестве* $U^{(\alpha)}$ *в слабую сходимость в* $W_2^2(\Omega)$ – *пространстве состояний.*

Доказательство. Пусть $\{g^{(n)}\}_{n=1}^{\infty} = \{(g_0^{(n)}, g_1^{(n)}, g_2^{(n)}, g_3^{(n)}) = (k^{(n)}(\xi), b_1^{(n)}(\xi), b_2^{(n)}(\xi), d^{(n)}(\xi))\}_{n=1}^{\infty} \subset U^{(\alpha)}$ – произвольная последовательность такая, что $g^{(n)} \longrightarrow \bar{g} = (\bar{g}_0, \bar{g}_1, \bar{g}_2, \bar{g}_3) = (\bar{k}, \bar{b}_1, \bar{b}_2, \bar{d})$ слабо в $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ к некоторому элементу $g \in H$, т.е.

$$k^{(n)}(\xi) \longrightarrow \bar{k}(\xi) \text{ слабо в } W_2^1(\Omega),\quad b_\alpha^{(n)}(\xi) \longrightarrow \bar{b}_\alpha(\xi) \text{ слабо в } L_2(\Omega),\quad \alpha = 1, 2,\quad d^{(n)}(\xi) \longrightarrow \bar{d}(\xi) \text{ слабо в } L_2(\Omega). \tag{1.14}$$

Так как $U^{(\alpha)} \subset H = W_2^1(\Omega) \times (L_2(\Omega))^3$ – выпуклое и замкнутое в H множество, то оно и слабо замкнуто в рефлексивном пространстве H. Поэтому $\bar{g} = (\bar{k}(\xi), \bar{b}_1(\xi), \bar{b}_2(\xi), \bar{d}(\xi)) \in U^{(\alpha)}$. В силу однозначной разрешимости задачи (1.5), каждому элементу $g^{(n)} \in U^{(\alpha)}$ сопоставляется $u^{(n)}(\xi) = u(\xi; g^{(n)})$ – единственное решение задачи (1.5) при $g = g^{(n)}$; кроме того,

$$\|u^{(n)}(\xi)\|_{W_2^2(\Omega)} = \|u(\xi; g^{(n)})\|_{W_2^2(\Omega)} \leq \mathrm{Const} \quad \forall n, \tag{1.15}$$

т.е. последовательность $\{u^{(n)}(\xi)\}_{n=1}^{\infty}$ равномерно ограничена в норме пространства $W_2^2(\Omega)$. Из соотношений (1.14), (1.15) следует, что можно извлечь подпоследовательность $\{g^{(n_k)}, u^{(n_k)}\}_{k=1}^{\infty} = \{g^{(n_k)}, u(\xi; g^{(n_k)})\}_{k=1}^{\infty}$ такую, что

$$k^{(n_k)}(\xi) \longrightarrow \bar{k}(\xi) \text{ сильно в } L_2(\Omega) \text{ и п.в. на } \Omega,\quad b_\alpha^{(n_k)}(\xi) \longrightarrow \bar{b}_\alpha(\xi) \text{ слабо в } L_2(\Omega),\quad \alpha = 1, 2,$$

$$d^{(n_k)}(\xi) \longrightarrow \bar{d}(\xi) \text{ слабо в } L_2(\Omega), \quad u^{(n_k)}(\xi) \longrightarrow \bar{u} \text{ слабо в } W_2^2(\Omega), \tag{1.16}$$
$$u^{(n_k)}(\xi) \longrightarrow \bar{u} \text{ сильно в } W_2^1(\Omega), \quad u^{(n_k)}(\xi) \longrightarrow \bar{u} \text{ сильно в } L_2(\Omega) \text{ и п.в. в } \Omega,$$

где $\bar{u} = \bar{u}(\xi)$ – некоторый элемент из $W_2^2(\Omega)$. Далее, на основе соотношений (1.5), (1.14)–(1.16) и ограничений на входные данные и управление g можно показать, что $\bar{u} = \bar{u}(\xi)$ – решение исходной задачи (1.5) при $g = \bar{g}(\xi) = (\bar{k}(\xi), \bar{b}_1(\xi), \bar{b}_2(\xi), \bar{d}(\xi))$, т.е. $Q(\bar{g}; \bar{u}; \eta) = l(\eta)$ $\forall \eta \in W_2^1(\Omega)$, где $\bar{u} = u(\xi; \bar{g})$. Таким образом, установлено, что при выполнении соотношений (1.14) из последовательности $\{u^{(n)} \equiv u(\xi, g^{(n)})\}_{n=1}^{\infty}$ можно выделить подпоследовательность $\{u^{(n_k)} \equiv u(\xi, g^{(n_k)})\}_{k=1}^{\infty}$, для которой справедлива сходимость

$$u^{(n_k)}(\xi) = u(\xi, g^{(n_k)}) \longrightarrow \bar{u}(\xi) = u(\xi, \bar{g}) \text{ слабо в } W_2^2(\Omega). \tag{1.17}$$

Можно показать, что соотношение (1.17) справедливо не только для подпоследовательности $\{u^{(n_k)}\}_{k=1}^{\infty}$, но и для всей последовательности $\{u^{(n)} \equiv u(\xi, g^{(n)})\}_{n=1}^{\infty}$, т.е.

$$u^{(n)}(\xi) = u(\xi, g^{(n)}) \longrightarrow \bar{u}(\xi) = u(\xi, \bar{g}) \text{ слабо в } W_2^2(\Omega). \tag{1.18}$$

Тем самым установлено, что при выполнении соотношения $g^{(n)} \longrightarrow \bar{g}$ слабо в $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ следует (1.18), т.е. $g \longrightarrow u(\xi, g)$ – слабо непрерывное отображение $U^{(\alpha)} \longrightarrow W_2^2(\Omega)$. Лемма 1 доказана.

Следствие 1. Отображение $g \longrightarrow u(\xi, g)$ является усиленно непрерывным из $U^{(\alpha)}$ в $W_2^1(\Omega)$ в том смысле, что оно переводит слабую сходимость в H на множестве $U^{(\alpha)}$ в сильную сходимость в $W_2^1(\Omega)$.

Продолжение доказательства теоремы 1. Рассмотрим теперь функционал цели $g \longrightarrow J(g)$, определенный формулой (1.3) в задачах $\mathbf{A}^{(\alpha)}$. Так как решение $u(\xi) = u(\xi, g)$ задачи (1.5) принадлежит пространству $W_2^2(\Omega)$, то определение функционала стоимости $J(g)$ в виде (1.3) корректно. Покажем, что функционал $J(g)$ слабо в $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ непрерывен на множестве $U^{(\alpha)}$. Пусть $\{g^{(n)}\}_{n=1}^{\infty} = \{(g_0^{(n)}, g_1^{(n)}, g_2^{(n)}, g_3^{(n)}) = (k^{(n)}(\xi), b_1^{(n)}(\xi), b_2^{(n)}(\xi), d^{(n)}(\xi))\}_{n=1}^{\infty} \subset U^{(\alpha)}$ – произвольная последовательность управлений, слабо сходящаяся к некоторому элементу $g = (g_0, g_1, g_2, g_3) = (k(\xi), b_1(\xi), b_2(\xi), d(\xi)) \in U^{(\alpha)}$. Покажем, что тогда $J(g^{(n)}) \longrightarrow J(g)$ при $n \longrightarrow \infty$. Действительно, нетрудно убедиться, что справедлива оценка

$$\begin{aligned} |J(g^{(n)}) - J(g)| &\le \alpha_0 \|\rho(\xi)\|_{L_2(\Omega)} \|u(\xi, g^{(n)}) - u(\xi, g)\|_{L_2(\Omega)} + \\ &+ \alpha_1 (\|u(\xi, g^{(n)})\|_{L_2(\Omega)} + \|u(\xi, g)\|_{L_2(\Omega)} + 2\|u_0(\xi)\|_{L_2(\Omega)}) \|u(\xi, g^{(n)}) - u(\xi, g)\|_{L_2(\Omega)} + \\ &+ \alpha_2 \left(\left\| \frac{\partial u(\xi, g^{(n)})}{\partial \xi_1} \right\|_{L_2(\Omega)} + \left\| \frac{\partial u(\xi, g)}{\partial \xi_1} \right\|_{L_2(\Omega)} + 2\|\psi_1(\xi)\|_{L_2(\Omega)} \right) \left\| \frac{\partial u(\xi, g^{(n)})}{\partial \xi_1} - \frac{\partial u(\xi, g)}{\partial \xi_1} \right\|_{L_2(\Omega)} + \\ &+ \alpha_3 \left(\left\| \frac{\partial u(\xi, g^{(n)})}{\partial \xi_2} \right\|_{L_2(\Omega)} + \left\| \frac{\partial u(\xi, g)}{\partial \xi_2} \right\|_{L_2(\Omega)} + 2\|\psi_2(\xi)\|_{L_2(\Omega)} \right) \left\| \frac{\partial u(\xi, g^{(n)})}{\partial \xi_2} - \frac{\partial u(\xi, g)}{\partial \xi_2} \right\|_{L_2(\Omega)}. \end{aligned} \tag{1.19}$$

Далее, в силу следствия 1 имеем $\|u(\xi, g^{(n)}) - u(\xi, g)\|_{W_2^1(\Omega)} \longrightarrow 0$ при $n \longrightarrow \infty$. Кроме того, справедливы оценки $\|u(\xi, g^{(n)})\|_{W_2^1(\Omega)} \le \mathrm{Const}$ $\forall n$. Поэтому из (1.19) получаем, что $J(g^{(n)}) \longrightarrow J(g)$ при $n \longrightarrow \infty$, т.е. функционал цели $J(g)$ непрерывен на $U^{(\alpha)}$ в слабой топологии пространства $H = W_2^1(\Omega) \times (L_2(\Omega))^3$. Кроме того, как было показано выше, множество $U^{(\alpha)}$ является слабо бикомпактным в H, откуда, учитывая, что слабо непрерывный функционал слабо полунепрерывен снизу, и применяя результат из [1], получаем, что экстремальные задачи $\mathbf{A}^{(\alpha)}$ корректно поставлены в слабой топологии пространства H, т.е. справедливы все утверждения теоремы 1. Теорема 1 доказана.

Замечания. 1. Из теоремы 1 следует существование решения экстремальных задач $\mathbf{A}^{(\alpha)}$, но решения этих задач, вообще говоря, могут быть не единственными.

2. Если при сужении класса допустимых управлений U_0 до множества $\hat{U}_0(p)$ (см. постановку задачи $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$) вместо ограничения (1.11) взять более сильное ограничение

$$\int_{\Omega} k^p(\xi)d\Omega = M^p|\Omega|, \quad \nu \le M \le \bar{\nu}, \quad p \ge 1, \tag{1.20}$$

то приведенное выше доказательство теоремы 1 не проходит в случае $p > 1$ в силу того, что при таком сужении множества U_0 до $\hat{U}_0(p)$ соответствующее множество $\hat{U}_0(p)$ не будет обладать свойством слабой бикомпактности в $W_2^1(\Omega)$. В то же время, как следует из теоремы 1, существование решения задачи $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ при $p > 1$ гарантировано, когда множество $\hat{U}_0(p)$ учитывает более слабое ограничение (1.11), чем (1.20). В связи с отмеченным, более естественным при постановке задачи $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ был переход в ограничении (1.20) от знака равенства к знаку меньше или равно. В результате чего мы и пришли к постановке задачи $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$.

2. РАЗНОСТНАЯ АППРОКСИМАЦИЯ ЗАДАЧ УПРАВЛЕНИЯ. КОРРЕКТНОСТЬ АППРОКСИМАЦИЙ

В связи с численным решением задач оптимального управления существенный интерес представляет вопрос об аппроксимации бесконечномерных задач $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ последовательностями конечномерных задач оптимального управления. Ниже построим и изучим аппроксимации задач на основе метода сеток (см. [10], [11]) и исследуем сходимость этих аппроксимаций при неограниченном измельчении шага h сетки дискретизации. Для аппроксимации задач $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ нам понадобятся сетки на $[0, l_\alpha]$, $\alpha = 1, 2$, и в $\bar{\Omega}$: $\bar{\omega}_\alpha = \{x_\alpha^{(i_\alpha)} = i_\alpha h_\alpha \in [0, l_\alpha] : i_\alpha = 0, 1, \ldots, N_\alpha; N_\alpha h_\alpha = l_\alpha\}$, $\omega_\alpha = \bar{\omega}_\alpha \cap (0, l_\alpha)$, $\omega_\alpha^{+} = \bar{\omega}_\alpha \cap (0, l_\alpha]$, $\omega_\alpha^{-} = \bar{\omega}_\alpha \cap [0, l_\alpha)$, $\alpha = 1, 2$; $\bar{\omega} = \bar{\omega}_1 \times \bar{\omega}_2$, $\omega = \omega_1 \times \omega_2$; $\gamma = \bar{\omega}/\omega$, $\omega^{(\pm 1)} = \omega_1^{\pm} \times \omega_2$, $\omega^{(\pm 2)} = \omega_1 \times \omega_2^{\pm}$, $\bar{\omega}^{(\pm 1)} = \omega_1^{\pm} \times \bar{\omega}_2$, $\bar{\omega}^{(\pm 2)} = \bar{\omega}_1 \times \omega_2^{\pm}$. Пусть $|h|^2 = h_1^2 + h_2^2$. Введем также средний шаг сетки $\bar{\omega}_\alpha$: $\hbar_\alpha = \hbar_\alpha(x_\alpha) = h_\alpha$, если $x \in \omega_\alpha$, и $\hbar_\alpha = 0.5h_\alpha$, если $h_\alpha = 0, l_\alpha$, $\alpha = 1, 2$. При исследовании свойств разностных аппроксимаций нам понадобятся скалярные произведения, нормы и полунормы сеточных функций, заданных на сетке $\bar{\omega}$ или на ее частях:

$$(u, \nu)_{\bar{\omega}} = \sum_{\bar{\omega}} \hbar_1 \hbar_2 u\nu, \quad (u, \nu)_{\omega} = \sum_{\omega} h_1 h_2 u\nu, \quad (u, \nu)_{\omega^{(\pm\alpha)}} = \sum_{\omega^{(\pm\alpha)}} h_1 h_2 u\nu, \quad (u, \nu)_{\omega_1^{\pm} \times \bar{\omega}_2} = \sum_{\omega_1^{\pm} \times \bar{\omega}_2} h_1 \hbar_2 u\nu,$$

$$(u, \nu)_{\bar{\omega}_1 \times \omega_2^{\pm}} = \sum_{\bar{\omega}_1 \times \omega_2^{\pm}} \hbar_1 h_2 u\nu, \quad \|u\|_{L_2(\bar{\omega})} = (u, u)_{\bar{\omega}}^{1/2}, \quad \|u\|_{L_2(\omega)} = (u, u)_{\omega}^{1/2}, \quad \|u\|_{L_2(\omega^{(\pm\alpha)})} = (u, u)_{\omega^{(\pm\alpha)}}^{1/2},$$

$$\|u\|_{L_2(\omega_1^{\pm} \times \bar{\omega}_2)} = (u, u)_{\omega_1^{\pm} \times \bar{\omega}_2}^{1/2}, \quad \|u\|_{L_2(\bar{\omega}_1 \times \omega_2^{\pm})} = (u, u)_{\bar{\omega}_1 \times \omega_2^{\pm}}^{1/2}, \quad \|u\|_{L_\infty(\bar{\omega})} = \|u\|_{C(\bar{\omega})} = \max_{x \in \bar{\omega}} |u(x)|,$$

$$\|u\|^2_{\overset{0}{W}{}_2^1(\bar{\omega})} = \sum_{\omega_1^{+} \times \omega_2} h_1 h_2 u_{\bar{x}_1}^2 + \sum_{\omega_1 \times \omega_2^{+}} h_1 h_2 u_{\bar{x}_2}^2 = \sum_{\alpha = 1}^{2} \sum_{\omega^{(+\alpha)}} h_1 h_2 u_{\bar{x}_\alpha}^2,$$

$$\|u\|^2_{W_2^1(\bar{\omega})} = \sum_{\omega_1^{-} \times \bar{\omega}_2} h_1 \hbar_2 u_{x_1}^2 + \sum_{\bar{\omega}_1 \times \omega_2^{-}} \hbar_1 h_2 u_{x_2}^2 = \sum_{\alpha = 1}^{2} \sum_{\bar{\omega}^{(-\alpha)}} h_\alpha \hbar_\beta u_{x_\alpha}^2, \quad \|u\|^p_{L_p(\bar{\omega})} = \sum_{\bar{\omega}} \hbar_1 \hbar_2 |u|^p, \quad p \ge 1,$$

$$\|u\|^2_{W_2^1(\bar{\omega})} = \|u\|^2_{L_2(\bar{\omega})} + |u|^2_{W_2^1(\bar{\omega})}, \quad \|u\|_{L_\infty(\omega^{(+\alpha)})} = \max_{\omega^{(+\alpha)}} |u(x)|, \quad \alpha = 1, 2,$$

$$\|u\|_{W^1_\infty(\bar\omega)} = \max_{x\in\bar\omega}|u(x)| + \max_{\omega_1^-\times\bar\omega_2}|u_{x_1}(x)| + \max_{\bar\omega_1\times\omega_2^-}|u_{x_2}(x)|.$$

Пусть $\tilde\omega \subset \bar\Omega$ – разные сетки в области $\bar\Omega$. Через $L_p(\tilde\omega)$, $p \ge 1$, $L_\infty(\tilde\omega)$, $W^1_\infty(\tilde\omega)$, $W^1_2(\tilde\omega)$ будем обозначать пространства сеточных функций, определенных на соответствующих сетках $\tilde\omega$ с нормами $\|\cdot\|_{L_p(\tilde\omega)}$, $p \ge 1$, $\|\cdot\|_{L_\infty(\tilde\omega)}$, $\|\cdot\|_{W^1_\infty(\tilde\omega)}$, $\|\cdot\|_{W^1_2(\tilde\omega)}$. Определим для функций $\nu(x)$ усредняющие операторы по Стеклову S^{x_α} по переменным x_α, $\alpha = 1, 2$:

$$S^{x_\alpha}(\nu)(x) = \frac{1}{\hbar_\alpha}\int\limits_{e_\alpha(x_\alpha)}\nu(x(\xi_\alpha))d\xi_\alpha, \quad x_\alpha\in\bar\omega_\alpha,$$

где $x(\xi_\alpha) = (\xi_1, x_2)$, если $\alpha = 1$, и $x(\xi_\alpha) = (x_1, \xi_2)$, если $\alpha = 2$; $e_\alpha(x_\alpha)$ – элементарные ячейки отрезка $[0, l_\alpha]$, $\alpha = 1, 2$: $e_\alpha(x_\alpha) = \{\xi_\alpha : x_\alpha - 0.5h_\alpha \le \xi_\alpha \le x_\alpha + 0.5h_\alpha\}$, $x_\alpha \in \omega_\alpha$, $e_\alpha(0) = \{\xi_\alpha : 0 \le \xi_\alpha \le 0.5h_\alpha\}$, $e_\alpha(l_\alpha) = \{\xi_\alpha : l_\alpha - 0.5h_\alpha \le \xi_\alpha \le l_\alpha\}$, $\alpha = 1, 2$. С помощью одномерных операторов S^{x_α}, действующих по направлению x_α, $\alpha = 1, 2$, определим усредняющий оператор $S^x = S^{x_1}S^{x_2}$ как произведение одномерных усредняющих операторов. Введем также элементарные ячейки области $\bar\Omega$: $e(x) = e(x_1) \times e(x_2)$, $x \in \bar\omega$, $e^1(x) = \bar e_1(x_1) \times e_2(x_2)$, $x \in \omega_1^- \times \bar\omega_2$, $e^2(x) = e_1(x_1) \times \bar e_2(x_2)$, $x \in \bar\omega_1 \times \omega_2^-$, $\bar e_\alpha(x_\alpha) = \{\xi_\alpha : x_\alpha \le \xi_\alpha \le x_\alpha + h_\alpha\}$, $x_\alpha \in \omega_\alpha^-$, $\alpha = 1, 2$. Задаче $\mathbf{A^{(0)}}$ оптимального управления поставим в соответствие следующее семейство конечномерных сеточных задач оптимального управления.

Задача $\mathbf{A_h^{(0)}}$. Найти сеточное управление $\Phi_{h*} \in U_h^{(0)}$ такое, что $J_h(\Phi_{h*}) = I_h(y(x, \Phi_{h*})) = \inf\limits_{\Phi_h\in U_h^{(0)}} I_h(y(x, \Phi_h)) = \inf\limits_{\Phi_h\in U_h^{(0)}} J_h(\Phi_h) = J_{h*}^{(0)}$, где сеточный функционал цели $\Phi_h \longrightarrow J_h(\Phi_h) = I_h(y(x, \Phi_h))$ задается формулой

$$\begin{aligned} J_h(\Phi_h) \equiv I_h(y(x, \Phi_h)) = {} & \alpha_0\sum_{\bar\omega}\rho^h(x)y(x,\Phi_h)\hbar_1\hbar_2 + \alpha_1\sum_{\bar\omega_1\times\bar\omega_2}\left|y(x;\Phi_h) - u_0^h(x)\right|^2\hbar_1\hbar_2 + \\ & + \alpha_2\sum_{\omega_1^-\times\bar\omega_2}\left|y_{x_1}(x;\Phi_h) - \Psi_1^h(x)\right|^2 h_1\hbar_2 + \alpha_3\sum_{\bar\omega_1\times\omega_2^-}\left|y_{x_2}(x;\Phi_h) - \Psi_2^h(x)\right|^2\hbar_1 h_2; \end{aligned} \tag{2.1}$$

$\nu_h = \nu(x; \Phi_h)$ – решение следующей сеточной задачи состояния, являющейся разностной аппроксимацией прямой задачи (1.1) на сетке $\bar\omega = \bar\omega_h$: найти $\nu_h(x) = \nu(x; \Phi_h) \in \overset{0}{W}{}^1_2(\bar\omega)$ такую, что для любой сеточной функции $\vartheta(x) \in \overset{0}{W}{}^1_2(\bar\omega)$ справедливо сумматорное тождество

$$\begin{aligned} Q_h(\Phi_h; y, \vartheta) = {} & \sum_{\omega_1^+\times\omega_2}\frac{\Phi_{0h}^{(-1_2)}(x) + \Phi_{0h}^{(-1_1,-1_2)}(x) + \Phi_{0h}^{(+1_2)}(x) + \Phi_{0h}^{(-1_1,+1_2)}(x)}{4}y_{\bar x_1}\vartheta_{\bar x_1}h_1h_2 + \\ & + \sum_{\omega_1\times\omega_2^+}\frac{\Phi_{0h}(x) + \Phi_{0h}^{(-1_2)}(x)}{2}y_{\bar x_2}\vartheta_{\bar x_2}h_1h_2 + \sum_{\alpha=1}^{2}\sum_{\omega}\Phi_{\alpha h}(x)y_{0_{x_\alpha}}\vartheta h_1h_2 + \sum_{\omega}\Phi_{3h}(x)q(y)\vartheta h_1h_2 = \\ & = \sum_{\omega}f^h(x)\vartheta h_1h_2 = l_h(\vartheta), \end{aligned} \tag{2.2}$$

а сеточные управления $\Phi_h(x)$ таковы, что

$$\Phi_h(x) = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in \prod_{\alpha=0}^{3}U_{\alpha h} = U_h^{(0)} \subset B_h = W^1_\infty(\bar\omega)\times(L_2(\omega))^3, \tag{2.3a}$$

$$U_{0h} = \{\Phi_{0h}(x) \in W^1_\infty(\bar{\omega}) : 0 < \nu \le \Phi_{0h}(x) \le \bar{\nu} < \infty, x \in \bar{\omega},$$
$$\left|\Phi_{0hx_1}(x)\right| \le R_1, x \in \omega_1^- \times \bar{\omega}_2, \left|\Phi_{0hx_2}(x)\right| \le R_2, x \in \bar{\omega}_1 \times \omega_2^-\}, \tag{2.3б}$$
$$U_{\alpha h} = \{\Phi_{\alpha h}(x) \in L_\infty(\omega) : \zeta_\alpha \le \Phi_{\alpha h}(x) \le \bar{\zeta}_\alpha, x \in \omega\}, \quad \alpha = 1, 2, 3.$$

Здесь $\Phi_{0h}^{(-1_1, -1_2)}(x) = \Phi_{0h}(x_1 - h_1, x_2 - h_2)$, $\Phi_{0h}^{(-1_2)}(x) = \Phi_{0h}(x_1, x_2 - h_2)$, $\Phi_{0h}^{(-1_1, +1_2)}(x) = \Phi_{0h}(x_1 - h_1, x_2 + h_2)$, $\Phi_{0h}^{(+1_2)}(x) = \Phi_{0h}(x_1, x_2 + h_2)$, а $f^h(x)$, $u_0^h(x)$, $\psi_\alpha^h(x)$, $\rho^h(x)$, $\alpha = 1, 2$ – сеточные аппроксимации функций $f(\xi)$, $u_0(\xi)$, $\psi_\alpha(\xi)$, $\alpha = 1, 2$, $\rho(\xi)$, определяемые через операцию усреднения по Стеклову:

$$f^h(x) = S^{x_1}S^{x_2}(f)(x) = \frac{1}{\hbar_1 h_2}\int\limits_{e(x)} f(\xi_1, \xi_2)d\xi_1\xi_2, \quad x \in \omega,$$
$$u_0^h(x) = S^{x_1}S^{x_2}(u_0)(x) = \frac{1}{\hbar_1\hbar_2}\int\limits_{e(x)} u_0(\xi_1, \xi_2)d\xi_1\xi_2, \quad \rho^h(x) = S^x\rho(x), \quad x \in \bar{\omega}, \tag{2.4}$$
$$\psi_1^h(x) = \frac{1}{h_1\hbar_2}\int\limits_{e^1(x)} \psi_1(\xi_1, \xi_2)d\xi_1\xi_2, \quad x \in \omega_1^- \times \bar{\omega}_2, \quad \psi_2^h = \frac{1}{\hbar_1 h_2}\int\limits_{e^2(x)} \psi_2(\xi_1, \xi_2)d\xi_1\xi_2, \quad x \in \bar{\omega}_1 \times \omega_2^-.$$

На основе энергетического метода, используя теорию монотонных операторов и разностные аналоги теорем вложения Соболева, можно показать, что справедлива следующая

Теорема 2. *Сеточная задача* (*разностная схема*) *для состояния* (2.2) *однозначно разрешима для любого сеточного управления* $\Phi_h \in U_h^{(0)}$, *причем справедлива априорная оценка*

$$\|y(x; \Phi_h)\|_{\overset{0}{W}{}_2^1(\bar{\omega})} \le C\|f^h(x)\|_{L_2(\omega)} \le C\|f\|_{L_2(\Omega)}. \tag{2.5}$$

Замечание 3. Разностная схема (2.2) – аппроксимация уравнения состояния – является некоторой “модификацией” традиционных схем в том смысле, что она отличается способом вычисления переменных коэффициентов в главной части сеточного оператора.

Рассмотрим теперь экстремальные задачи $\mathbf{A}^{(1)}$ и $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$. Задачам $\mathbf{A}^{(1)}$ и $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ поставим в соответствие при $|h| \longrightarrow 0$ следующие семейства конечномерных сеточных задач оптимального управления, которые будем называть задачами $\mathbf{A}_\mathbf{h}^{(1)}$ и $\mathbf{A}_\mathbf{h}^{(2)} = \mathbf{A}_\mathbf{h}^{(2)}(\mathbf{p})$ соответственно.

Задача $\mathbf{A}_\mathbf{h}^{(1)}$. Найти сеточное управление $\Phi_{h*} \in U_h^{(1)}$ такое, что $J_h(\Phi_{h*}) = I_h(y(x, \Phi_{h*})) = \inf\limits_{\Phi_h \in U_h^{(1)}} I_h(y(x, \Phi_h)) = \inf\limits_{\Phi_h \in U_h^{(1)}} J_h(\Phi_h) = J_{h*}^{(1)}$, где сеточный функционал цели $\Phi_h \longrightarrow J_h(\Phi_h) = I_h(y(x, \Phi_h))$ задается формулой (2.1), $y_h = y(x; \Phi_h)$ – решение сеточной задачи состояния (2.2), а сеточные управления $\Phi_h(x)$ таковы, что

$$\Phi_h(x) = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in \hat{U}_{0h} \times U_{1h} \times U_{2h} \times U_{3h} = U_h^{(1)} \subset B_h,$$
$$\hat{U}_{0h} = \Bigg\{\Phi_{0h}(x) \in W^1_\infty(\bar{\omega}) : 0 < \nu \le \Phi_{0h}(x) \le \bar{\nu} < \infty, x \in \bar{\omega}, \left|\Phi_{0hx_1}(x)\right| \le R_1, x \in \omega_1^- \times \bar{\omega}_2,$$
$$\left|\Phi_{0hx_2}(x)\right| \le R_2, x \in \bar{\omega}_1 \times \omega_2^-; \sum_{\bar{\omega}} \Phi_{0h}(x)\hbar_1\hbar_2 = M\Bigg\};$$

$U_{\alpha h}$, $\alpha = \overline{1, 3}$, – множества допустимых управлений, имеющие вид (2.3б).

Задача $\mathbf{A}_\mathbf{h}^{(2)} = \mathbf{A}_\mathbf{h}^{(2)}(\mathbf{p})$. Найти сеточное управление $\Phi_{h*} \in U_h^{(2)} = U_h^{(2)}(p)$ такое, что $J_h(\Phi_{h*}) = I_h(y(x, \Phi_{h*})) = \inf\limits_{\Phi_h \in U_h^{(2)}(p)} I_h(y(x, \Phi_h)) = \inf\limits_{\Phi_h \in U_h^{(p)}(p)} J_h(\Phi_h) = J_{h*}^{(p)}$, где сеточный функционал цели $\Phi_h \longrightarrow$

$\longrightarrow J_h(\Phi_h) = I_h(y(x, \Phi_h))$ задается формулой (2.1), $y_h = y(x; \Phi_h)$ – решение сеточной задачи состояния (2.2), а сеточные управления $\Phi_h(x)$ таковы, что

$$\Phi_h(x) = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in \hat{U}_{0h}(p) \times U_{1h} \times U_{2h} \times U_{3h} = U_h^{(2)}(p) \equiv U_h^{(2)} \subset B_h,$$

$$\hat{U}_{0h}(p) = \Big\{\Phi_{0h}(x) \in W_\infty^1(\bar{\omega}) : 0 < \nu \le \Phi_{0h}(x) \le \bar{\nu} < \infty,\ x \in \bar{\omega},\ \left|\Phi_{0hx_1}(x)\right| \le R_1,\ x \in \omega_1^- \times \bar{\omega}_2,$$

$$\left|\Phi_{0hx_2}(x)\right| \le R_2,\ x \in \bar{\omega}_1 \times \omega_2^-;\ \sum_{\bar{\omega}} \Phi_{0h}^p(x)\hbar_1\hbar_2 \le M^p|\Omega|\Big\};$$

$U_{\alpha h}$, $\alpha = \overline{1, 3}$, – множества допустимых управлений, имеющие вид (2.3б).

Справедлива

Теорема 3. *Пусть* $U_h^{(0)}$, $U_h^{(1)}$, $U_h^{(2)} \equiv U_h^{(2)}(p)$ *– множества допустимых управлений в сеточных экстремальных задачах* $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ *соответственно. Для каждого* $h > 0$ *существует по крайней мере одно оптимальное управление* $\Phi_{h*} \in U_h^{(\alpha)}$ *задач* $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$, т.е. $J_{h*}^{(0)} = \inf\{J_h(\Phi_h) : \Phi_h \in U_h^{(0)}\} > -\infty$, $U_{h*}^{(0)} = \{\Phi_{h*} \in U_h^{(0)} : J(\Phi_{h*}) = J_{h*}^{(0)}\} \ne \varnothing$, $J_{h*}^{(1)} = \inf\{J_h(\Phi_h) : \Phi_h \in U_h^{(1)}\} > -\infty$, $U_{h*}^{(1)} = \{\Phi_{h*} \in U_h^{(1)} : J(\Phi_{h*}) = J_{h*}^{(1)}\} \ne \varnothing$; $J_{h*}^{(2)} = J_{h*}^{(2)}(p) = \inf(J_h(\Phi_h) : \Phi_h \in U_h^{(2)} = U_h^{(2)}(p)\} > -\infty$, $U_{h*}^{(2)} = U_{h*}^{(2)}(p) = \{\Phi_{h*} \in U_h^{(2)} = U_h^{(2)}(p) : J(\Phi_{h*}) = J_{h*}^{(2)} = J_{h*}^{(2)}(p)\} \ne \varnothing$.

3. АПРИОРНЫЕ ОЦЕНКИ ПОГРЕШНОСТИ И СКОРОСТИ СХОДИМОСТИ СЕТОЧНЫХ ЗАДАЧ $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ ПО СОСТОЯНИЮ

Установим связь между $u(\xi, g)$ – решением прямой задачи (1.1) и $y(x, \Phi_h)$ – решением аппроксимирующей ее разностной задачи состояния (2.2) при $h \longrightarrow 0$ для любых фиксированных управлений $g \in U^{(\alpha)}$ и $\Phi_h \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, где $U^{(\alpha)}$ – множества допустимых управлений задач $\mathbf{A}^{(\alpha)}$, а $U_h^{(\alpha)}$ – множества допустимых управлений задач $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$. Пусть $u(\xi, g) \in W_2^2(\Omega)$ – решение задачи (1.1), отвечающее управлению $g \in U^{(\alpha)}$, а $y(x, \Phi_h) \in W_2^1(\bar{\omega})$ – решение задачи (2.2), отвечающее сеточному управлению $\Phi_h \in U_h^{(\alpha)}$. Обозначим через $z(x; \Phi_h, g) = y(x; \Phi_h) - u(x; g)$ погрешность метода по состоянию.

Априорную оценку погрешности метода по состоянию устанавливает

Теорема 4. *Пусть* $g = (g_0(\xi), g_1(\xi), g_2(\xi), g_3(\xi)) = (k(\xi), b_1(\xi), b_2(\xi), d(\xi)) \in U^{(\alpha)}$ *и* $\Phi_h(x) = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in U_h^{(\alpha)}$ *– произвольные управления, а* $u(\xi, g)$ *и* $y(x, \Phi_h)$ *– соответствующие им решения задач состояния в экстремальных задачах* $\mathbf{A}^{(\alpha)}$ *и* $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$. *Тогда для любых* $h > 0$ *справедлива оценка*

$$\|y(x; \Phi_h) - u(x; g)\|_{W_2^1(\bar{\omega})} \le$$

$$\le C\left[\sum_{\alpha, k = 1}^{2} \|\eta_\alpha^{(k)}\|_{L_2(\omega^{(+\alpha)})} + \sum_{\alpha = 1}^{2}\left(\|B_\alpha^-\|_{L_2(\omega)} + \|B_\alpha^+\|_{L_2(\omega)} + \|\hat{B}_\alpha^-\|_{L_2(\omega)} + \|\hat{B}_\alpha^+\|_{L_2(\omega)}\right) + \sum_{\alpha = 1}^{2} \|\eta_3^{(\alpha)}\|_{L_2(\omega)}\right], \tag{3.1}$$

$$\eta_\alpha^{(1)}(x) = [S^{x_\beta}k_\alpha^{(-0.5\alpha)}(x)]u_{\bar{x}_\alpha}(x) - S^{x_\beta}\left(k_\alpha\frac{\partial u}{\partial \xi_\alpha}\right)^{(-0.5\alpha)}(x), \quad x \in \omega^{(+\alpha)}, \quad \beta = 3 - \alpha, \tag{3.2а}$$

$$\eta_1^{(2)}(x) = \left[\frac{\Phi_{0h}^{(-1_2)}(x) + \Phi_{0h}^{(-1_1, -1_2)}(x) + \Phi_{0h}^{(+1_2)}(x) + \Phi_{0h}^{(-1_1, +1_2)}(x)}{4} - S^{x_2}k^{(-0.5_1)}(x)\right]u_{\bar{x}_1}(x), \quad x \in \omega^{(+1)}, \tag{3.2б}$$

$$\eta_2^{(2)}(x) = \left[\frac{\Phi_{0h}(x)+\Phi_{0h}^{(-1_2)}(x)}{2} - S^{x_1}k^{(-0.5_2)}(x)\right]u_{\bar{x}_2}(x), \quad x \in \omega^{(+2)}, \tag{3.2в}$$

$$B_\alpha^-(x) = [S^x b_\alpha(x)]u_{x_\alpha}(x) - S^x\left(b_\alpha\frac{\partial u}{\partial \xi_\alpha}\right)(x), \quad B_\alpha^+(x) = [S^x b_\alpha(x)]u_{\bar{x}_\alpha}(x) - S^x\left(b_\alpha\frac{\partial u}{\partial \xi_\alpha}\right)(x), \quad x \in \omega, \tag{3.2г}$$

$$\hat{B}_\alpha^-(x) = (\Phi_{\alpha h}(x) - S^x b_\alpha(x))u_{x_\alpha}(x), \quad \hat{B}_\alpha^+(x) = (\Phi_{\alpha h}(x) - S^x b_\alpha(x))u_{\bar{x}_\alpha}(x), \quad x \in \omega, \quad \alpha = 1, 2, \tag{3.2д}$$

$$\eta_3^{(1)}(x) = [S^x d(x)]q(u(x)) - S^x[d(\xi)q(u(\xi))](x), \quad \eta_3^{(2)}(x) = (\Phi_{3h}(x) - S^x d(x))q(u(x)), \quad x \in \omega, \tag{3.2е}$$

ξ – *переменная, по которой производится интегрирование в* S^{x_α}, $\alpha = 1, 2$, x – *узел сетки* $\bar{\omega}$.

Для оценки левой части неравенства (3.1а) через параметр h – шаг сетки $\bar{\omega}$ достаточно установить оценки величин (3.2). Справедливы следующие леммы.

Лемма 2. *Пусть* $u(\xi) \in W_2^2(\Omega)$. *Тогда справедливы оценки*

$$\{\|u_{x_\alpha}\|_{L_2(\bar{\omega}^{(-\alpha)})}\} \le M_1\left\lfloor |h||u(\xi)|_{W_2^2(\Omega)} + |u(\xi)|_{W_2^1(\Omega)}\right\rfloor, \quad \alpha = 1, 2, \tag{3.3а}$$

$$\|u\|_{L_2(\bar{\omega})} \le M_1\left\lfloor |h|^2|u(\xi)|_{W_2^2(\Omega)} + |h||u(\xi)|_{W_2^1(\Omega)} + \|u(\xi)\|_{L_2(\Omega)}\right\rfloor, \tag{3.3б}$$

где через $|\cdot|_{W_2^k(\Omega)}$ *обозначены полунормы в пространствах* W_2^k, $k = 1, 2$.

Лемма 3. *Пусть* $k(\xi)$, $b_\alpha(\xi)$, $d(\xi) \in L_\infty(\Omega)$, $\alpha = 1, 2$; *решение* $u(\xi) = u(\xi, g)$ *задачи* (1.1) *принадлежит классу* $W_{2,0}^2(\Omega)$. *Тогда справедливы оценки*

$$\left\|\eta_\alpha^{(1)}(x)\right\|_{L_2(\omega^{(+\alpha)})} \le |h|\,\|k(\xi)\|_{L_\infty(\Omega)}\|u(\xi, g)\|_{W_{2,0}^2(\Omega)}, \quad \alpha = 1, 2,$$

$$\left\|\eta_1^{(2)}(x)\right\|_{L_2(\omega^{(+1)})} \le M_1\left\|\frac{\Phi_{0h}^{(-1_2)}(x) + \Phi_{0h}^{(-1_1,-1_2)}(x) + \Phi_{0h}^{(+1_2)}(x) + \Phi_{0h}^{(-1_1,+1_2)}(x)}{4} - S^{x_2}k^{(-0.5_1)}(x)\right\|_{L_\infty(\omega^{(+1)})}\|u(\xi)\|_{W_{2,0}^2(\Omega)},$$

$$\left\|\eta_2^{(2)}(x)\right\|_{L_2(\omega^{(+2)})} \le M_1\left\|\frac{\Phi_{0h}(x) + \Phi_{0h}^{(-1_2)}(x)}{2} - S^{x_1}k^{(-0.5_2)}(x)\right\|_{L_\infty(\omega^{(+2)})}\|u(\xi)\|_{W_{2,0}^2(\Omega)},$$

$$\left\|B_\alpha^\pm(x)\right\|_{L_2(\omega)} \le M_1|h|\,\|b_\alpha(\xi)\|_{L_\infty(\Omega)}\|u(\xi)\|_{W_{2,0}^2(\Omega)},$$

$$\left\|\hat{B}_\alpha^\pm(x)\right\|_{L_2(\omega)} \le M_1\left\|\Phi_{\alpha h}(x) - S^x b_\alpha(x)\right\|_{L_\infty(\Omega)}\|u(\xi)\|_{W_{2,0}^2(\Omega)}, \quad \alpha = 1, 2,$$

$$\left\|\eta_3^{(1)}(x)\right\|_{L_2(\omega)} \le M_1|h|L_q\|d(\xi)\|_{L_\infty(\Omega)}\|u(\xi)\|_{W_{2,0}^2(\Omega)},$$

$$\left\|\eta_3^{(2)}(x)\right\|_{L_2(\omega)} \le M_1L_q\left\|\Phi_{3h}(x) - S^x d(x)\right\|_{L_\infty(\Omega)}\|u(\xi)\|_{W_{2,0}^2(\Omega)}.$$

Доказательства теоремы 4 и лемм 2, 3 проводятся на основе методики из [9].

Теорема 4 и леммы 2, 3 позволяют установить справедливость следующей теоремы.

Теорема 5. *Пусть* $g(\xi) = (g_0(\xi), g_1(\xi), g_2(\xi), g_3(\xi)) = (k(\xi), b_1(\xi), b_2(\xi), d(\xi)) \in U^{(\alpha)}$ *и* $\Phi_h(x) = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x)) \in U_h^{(\alpha)}$ – *произвольные управления, а* $u = u(\xi, g)$ *и* $y = y(x, \Phi_h)$ – *соответствующие им решения задач состояния в экстремальных задачах* $\mathbf{A}^{(\alpha)}$ *и* $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$. *Тогда для любых* $h > 0$ *справедлива следующая оценка скорости сходимости метода сеток по состоянию для экстремальных задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$:

$$\|y(x;\Phi_h) - u(x;g)\|_{W_2^1(\bar{\omega})} \le C\left[|h|\left(\|k\|_{L_\infty(\Omega)} + \sum_{\alpha=1}^{2}\|b_\alpha\|_{L_\infty(\Omega)} + L_q\|d\|_{L_\infty(\Omega)}\right) + \right.$$

$$+\left\|S^{x_1}k^{(-0.5_2)}(x)-\frac{\Phi_{0h}(x)+\Phi_{0h}^{(-1_2)}(x)}{2}\right\|_{L_\infty(\omega^{(+2)})}+$$
$$+\left\|S^{x_2}k^{(-0.5_1)}(x)-\frac{\Phi_{0h}^{(-1_2)}(x)+\Phi_{0h}^{(-1_1,-1_2)}(x)+\Phi_{0h}^{(+1_2)}(x)+\Phi_{0h}^{(-1_1,+1_2)}(x)}{4}\right\|_{L_\infty(\omega^{(+1)})}+ \tag{3.4}$$
$$\left.+\sum_{\alpha=1}^{2}\left\|\Phi_{\alpha h}(x)-S^x b_\alpha(x)\right\|_{L_\infty(\omega)}+\left\|\Phi_{3h}(x)-S^x d(x)\right\|_{L_\infty(\omega)}\right]\|u(\xi;g)\|_{W_{2,0}^2(\Omega)}.$$

4. ОЦЕНКИ ПОГРЕШНОСТИ СЕТОЧНОГО ФУНКЦИОНАЛА И СКОРОСТИ СХОДИМОСТИ СЕТОЧНЫХ ЗАДАЧ $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}=\mathbf{A}^{(2)}(\mathbf{p})$ ПО ФУНКЦИОНАЛУ, СХОДИМОСТЬ ПО УПРАВЛЕНИЮ. РЕГУЛЯРИЗАЦИЯ АППРОКСИМАЦИЙ

Для ответа на вопрос о сходимости сеточных задач оптимального управления $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}=\mathbf{A}^{(2)}(\mathbf{p})$ по функционалу и управлению необходимо прежде всего установить связь между функционалами $J_h(\Phi_h)$ и $J(g)$ экстремальных задач $\mathbf{A}_\mathbf{h}^{(\alpha)}$ и $\mathbf{A}^{(\alpha)}$, $\alpha=0,1,2$, для любых фиксированных управлений $\Phi_h\in U_h^{(\alpha)}$ и $g\in U^{(\alpha)}$, $\alpha=0,1,2$, и любых $h>0$. Для этого нам понадобятся некоторые кусочно-линейные и кусочно-постоянные восполнения сеточных функций, некоторые дискретизации функций непрерывного аргумента, а также их свойства.

Пусть $\vartheta(x)$, $x\in\bar\omega$, – сеточная функция, заданная на двумерной сетке $\bar\omega=\bar\omega_1\times\bar\omega_2\subset\bar\Omega$. Через $\bar\vartheta(\xi)=P_h\vartheta(\xi)$, $\xi\in\Omega$, будем обозначать кусочно-постоянное восполнение на Ω сеточной функции $\vartheta(x)$, $x\in\bar\omega$, определяемое по формуле

$$\bar\vartheta_h(\xi)=P_h\vartheta(\xi)=\vartheta(x),\quad \xi\in e(x),\quad x\in\bar\omega. \tag{4.1}$$

Для сеточной функции $\vartheta_1(x)$, заданной на двумерной сетке $\omega_1^-\times\bar\omega_2\in\bar\Omega$, а также для сеточной функции $\vartheta_2(x)$, заданной на двумерной сетке $\bar\omega_1\times\omega_2^-\in\bar\Omega$, введем кусочно-постоянные восполнения на Ω, соответственно, по формулам

$$\bar\vartheta_{1h}(\xi)=P_{1h}\vartheta_1(\xi)=\vartheta_1(x),\quad \xi\in e^1(x),\quad x\in\omega_1^-\times\bar\omega_2, \tag{4.2}$$

$$\bar\vartheta_{2h}(\xi)=P_{2h}\vartheta_2(\xi)=\vartheta_2(x),\quad \xi\in e^2(x),\quad x\in\bar\omega_1\times\omega_2^-. \tag{4.3}$$

Справедливы следующие леммы.

Лемма 4. *Пусть* $u(\xi)\in W_2^2(\Omega)$, $u_0(\xi)$, $\psi_\alpha(\xi)\in W_2^1(\Omega)$, $\alpha=1,2$, $\rho(\xi)\in W_2^1(\Omega)$; $u_0^h(x)$, $\rho^h(x)$, $x\in\bar\omega$ *и* $\psi_1^h(x)$, $x\in\omega_1^-\times\bar\omega_2$, $\psi_2^h(x)$, $x\in\bar\omega_1\times\omega_2^-$ *– дискретизации функций* $u_0(\xi)$, $\rho(\xi)$ *на сетке* $\bar\omega$ *и функций* $\psi_1(\xi)$, $\psi_2(\xi)$ *на сетках* $\omega_1^-\times\bar\omega_2$, $\bar\omega_1\times\omega_2^-$ *соответственно, определяемые по формулам* (2.4). *Пусть* $P_hu(\xi)$, $P_hu_0^h(\xi)$, $P_h\rho^h(\xi)$ *– кусочно-постоянные восполнения на* Ω *сеточных функций* $u(x)$, $u_0^h(x)$, $\rho^h(x)$, $x\in\bar\omega\subset\bar\Omega$, *определяемые по формулам, аналогичным* (4.1); $P_{1h}u_{x_1}(\xi)$, $P_{1h}\psi_1^h(\xi)$ *и* $P_{2h}u_{x_2}(\xi)$, $P_{2h}\psi_2^h(\xi)$ *– кусочно-постоянные восполнения на* Ω *сеточных функций* $u_{x_1}(x)$, $\psi_1^h(x)$, $x\in\omega_1^-\times\bar\omega_2\subset\bar\Omega$ *и* $u_{x_2}(x)$, $\psi_2^h(x)$, $x\in\bar\omega_1\times\omega_2^-\subset\bar\Omega$, *определяемые по формулам, аналогичным* (4.2) *и* (4.3) *соответственно. Тогда справедливы оценки*

$$\|u(\xi)-P_hu(\xi)\|_{L_2(\Omega)}\le M|h||u(\xi)|_{W_2^2(\Omega)},\quad \left\|\frac{\partial u(\xi)}{\partial\xi_\alpha}-P_{\alpha h}u_{x_\alpha}(\xi)\right\|_{L_2(\Omega)}\le M|h||u(\xi)|_{W_2^2(\Omega)},\quad \alpha=1,2,$$
$$\|u_0(\xi)-P_hu_0^h(\xi)\|_{L_2(\Omega)}\le 2^{1/2}|h|\|u_0(\xi)\|_{W_2^1(\Omega)},\quad \|\rho(\xi)-P_h\rho^h(\xi)\|_{L_2(\Omega)}\le 2^{1/2}|h|\|\rho(\xi)\|_{W_2^1(\Omega)}, \tag{4.4}$$
$$\|\psi_\alpha(\xi)-P_{\alpha h}\psi_\alpha^h(\xi)\|_{L_2(\Omega)}\le 2^{1/2}|h|\|\psi_\alpha(\xi)\|_{W_2^1(\Omega)},\quad \alpha=1,2.$$

Лемма 5. *Пусть выполнены условия леммы* 4. *Тогда справедливы оценки*

$$\|P_h u_0^h(\xi)\|_{L_2(\Omega)} \le \|u_0(\xi)\|_{L_2(\Omega)}, \quad \|u_0^h(x)\|_{L_2(\bar\omega)} \le \|u_0(\xi)\|_{L_2(\Omega)}, \quad \|\psi_1^h(\xi)\|_{L_2(\omega_1^-\times\bar\omega_2)} \le \|\psi_1(\xi)\|_{L_2(\Omega)},$$

$$\|P_{\alpha h}\psi_\alpha(\xi)\|_{L_2(\Omega)} \le \|\psi_\alpha(\xi)\|_{L_2(\Omega)}, \quad \|\rho^h(x)\|_{L_2(\bar\omega)} \le \|\rho(\xi)\|_{L_2(\Omega)}, \quad \|\psi_2^h(\xi)\|_{L_2(\bar\omega_1\times\omega_2^-)} \le \|\psi_2(\xi)\|_{L_2(\Omega)},$$

$$\|P_h\rho^h(\xi)\|_{L_2(\Omega)} \le \|\rho(\xi)\|_{L_2(\Omega)}, \quad \|P_h u(\xi)\|_{L_2(\Omega)} = \|u(x)\|_{L_2(\bar\omega)} \le M\|u(\xi)\|_{W_2^2(\Omega)}, \tag{4.5}$$

$$\|P_{\alpha h}u_{x_\alpha}(\xi)\|_{L_2(\Omega)} = \|u_{x_\alpha}(x)\|_{L_2(\bar\omega^{(-\alpha)})} \le M\|u(\xi)\|_{W_2^2(\Omega)}, \quad \alpha = 1, 2.$$

Доказательства лемм 4, 5 опираются на определение отображений $P_{\alpha h}$, $\alpha = 1, 2$, P_h, определение оператора усреднения по Стеклову S^x, довольно громоздки, и мы их опускаем из-за ограниченности объема статьи.

Оценку погрешности сеточного функционала $J_h(\Phi_h)$ экстремальных задач $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$ устанавливает следующая

Теорема 6. *Для любых управлений* $g \in U^{(\alpha)}$ *и* $\Phi_h \in U_h^{(\alpha)}$ *экстремальных задач* $\mathbf{A}^{(\alpha)}$ *и* $\mathbf{A}_\mathbf{h}^{(\alpha)}$ *соответственно и любых* $h > 0$ *для погрешности сеточного функционала* $J_h(\Phi_h)$ *экстремальных задач* $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$, *справедлива оценка*

$$|J(g) - J_h(\Phi_h)| = |I(u(\xi, g)) - I_h(y(x, \Phi_h))| \le$$

$$\le M(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3)\Bigg[|h| + \left\|S^{x_1}k^{(-0.5_2)}(x) - \frac{\Phi_{0h}(x) + \Phi_{0h}^{(-1_2)}(x)}{2}\right\|_{L_\infty(\omega^{(+2)})} +$$

$$+ \left\|S^{x_2}k^{(-0.5_1)}(x) - \frac{\Phi_{0h}^{(-1_2)}(x) + \Phi_{0h}^{(-1_1,-1_2)}(x) + \Phi_{0h}^{(+1_2)}(x) + \Phi_{0h}^{(-1_1,+1_2)}(x)}{4}\right\|_{L_\infty(\omega^{(+1)})} + \tag{4.6}$$

$$+ \|S^x d(x) - \Phi_{3h}(x)\|_{L_\infty(\omega)} + \sum_{\alpha=1}^{2}\|S^x b_\alpha(x) - \Phi_{\alpha h}(x)\|_{L_\infty(\omega)}\Bigg],$$

где $M = \mathrm{Const} > 0$, *не зависящая от* h, y, u, Φ_h, g; $\alpha_k = \mathrm{Const} \ge 0$, $k = 0, 1, 2, 3$, $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 > 0$.

Доказательство. Пусть $g \in U^{(\alpha)}$ и $\Phi_h \in U_h^{(\alpha)}$ – произвольные управления, а $u = u(\xi, g)$ и $y = y(x, \Phi_h)$ – соответствующие им решения задач состояния в экстремальных задачах $\mathbf{A}^{(\alpha)}$ и $\mathbf{A}_\mathbf{h}^{(\alpha)}$, $\alpha = 0, 1, 2$. Принимая во внимание (1.3), (2.1) и проводя некоторые преобразования, приведем погрешность $\theta_h(g, \Phi_h) = J(g) - J_h(\Phi_h)$ к специальному виду:

$$\theta_h(g; \Phi_h) = J(g) - J_h(\Phi_h) = \alpha_0\sum_{\alpha=1}^{2}A_h^{(k)} + \alpha_1\sum_{\alpha=1}^{2}B_h^{(k)} + \alpha_2\sum_{\alpha=1}^{2}C_h^{(k)} + \alpha_3\sum_{\alpha=3}^{4}C_h^{(k)}, \tag{4.7}$$

$$A_h^{(1)} = (\rho(\xi), u(\xi, g))_{L_2(\Omega)} - (P_h\rho^h(\xi), P_h u(\xi, g))_{L_2(\Omega)},$$

$$A_h^{(2)} = (P_h\rho^h(\xi), P_h u(\xi, g))_{L_2(\Omega)} - (\rho^h(x), y(x, \Phi_h))_{L_2(\bar\omega)},$$

$$B_h^{(1)} = \|u(\xi; g) - u_0(\xi)\|_{L_2(\Omega)}^2 - \|P_h u(\xi; g) - P_h u_0^h(\xi)\|_{L_2(\Omega)}^2,$$

$$B_h^{(2)} = \|P_h u(\xi; g) - P_h u_0^h(\xi)\|_{L_2(\Omega)}^2 - \|y(x; \Phi_h) - u_0^h(x)\|_{L_2(\bar\omega)}^2,$$

$$C_h^{(2\alpha-1)} = \left\|\frac{\partial u(\xi; g)}{\partial\xi_\alpha} - \psi_\alpha(\xi)\right\|_{L_2(\Omega)}^2 - \|P_{\alpha h}u_{x_\alpha}(\xi; g) - P_{\alpha h}\psi_\alpha^h(\xi)\|_{L_2(\Omega)}^2,$$

$$C_h^{(2\alpha)} = \|P_{\alpha h}u_{x_\alpha}(\xi; g) - P_{\alpha h}\psi_\alpha^h(\xi)\|_{L_2(\Omega)}^2 - \|y_{x_\alpha}(x; \Phi_h) - \psi_\alpha^h(x)\|_{L_2(\bar\omega^{(-\alpha)})}^2, \quad \alpha = 1, 2.$$

Можно показать, что для величин $A_h^{(k)}$, $B_h^{(k)}$, $k = 1, 2$, $C_h^{(m)}$, $m = 1, 2, 3, 4$, справедливы следующие оценки:

$$\left|A_h^{(1)}\right| \le [\|\rho(\xi) - P_h\rho^h(\xi)\|_{L_2(\Omega)}\|u(\xi, g)\|_{L_2(\Omega)} + \|u(\xi; g) - P_h u(\xi; g)\|_{L_2(\Omega)}\|P_h\rho^h(\xi)\|_{L_2(\Omega)}],$$

$$\left|A_h^{(2)}\right| \le \|\rho^h(x)\|_{L_2(\bar\omega)}\|y(x; \Phi_h) - u(x, g)\|_{L_2(\bar\omega)},$$

$$\left|B_h^{(1)}\right| \le [\|u(\xi; g) - P_h u(\xi; g)\|_{L_2(\Omega)} + \|u_0(\xi) - P_h u_0^h(\xi)\|_{L_2(\Omega)}] \times$$

$$\times [\|u(\xi; g)\|_{L_2(\Omega)} + \|P_h u(\xi; g)\|_{L_2(\Omega)} + \|u_0(\xi)\|_{L_2(\Omega)} + \|P_h u_0^h(\xi)\|_{L_2(\Omega)}],$$

$$\left|B_h^{(2)}\right| \le \|y(x; \Phi_h) - u(\xi; g)\|_{L_2(\bar\omega)}[\|u(x; g)\|_{L_2(\bar\omega)} + \|y(x; \Phi_h)\|_{L_2(\bar\omega)} + 2\|u_0^h(x)\|_{L_2(\Omega)}], \tag{4.8}$$

$$\left|C_h^{(2\alpha-1)}\right| = \left[\left\|\frac{\partial u(\xi; g)}{\partial \xi_\alpha} - P_{\alpha h}u_{x_\alpha}(\xi; g)\right\|_{L_2(\Omega)} + \|\psi_\alpha(\xi) - P_{\alpha h}\psi_\alpha^h(\xi)\|_{L_2(\Omega)}\right] \times$$

$$\times \left[\left\|\frac{\partial u(\xi; g)}{\partial \xi_\alpha}\right\|_{L_2(\Omega)} + \|P_{\alpha h}u_{x_\alpha}(\xi; g)\|_{L_2(\Omega)} + \|\psi_\alpha(\xi)\|_{L_2(\Omega)} + \|P_{\alpha h}\psi_\alpha^h(\xi)\|_{L_2(\Omega)}\right],$$

$$\left|C_h^{(2\alpha)}\right| \le \|y_{x_\alpha}(x; \Phi_h) - u_{x_\alpha}(x; g)\|_{L_2(\bar\omega^{(-\alpha)})}[\|u_{x_\alpha}(x; g)\|_{L_2(\bar\omega^{(-\alpha)})} + \|y_{x_\alpha}(x; \Phi_h)\|_{L_2(\bar\omega^{(-\alpha)})} + 2\|\psi_\alpha^h(x)\|_{L_2(\bar\omega^{(-\alpha)})}],$$

$$\alpha = 1, 2.$$

Далее, проводя оценки правых частей неравенств (4.8) на основе оценок (4.4), (4.5), (3.3), (1.7), (2.5), (3.4), теорем вложения Соболева [22] и их разностных аналогов [10], [11] и принимая во внимание представление (4.7), устанавливаем оценку (4.6).

Определим теперь некоторые восполнения сеточных управлений $\Phi_{0h}(x)$, $x \in \bar\omega$, $\Phi_{\alpha h}(x)$, $x \in \omega$, $\alpha = 1, 2, 3$, на Ω. Рассмотрим сеточное управление $\Phi_{0h}(x)$, $x \in \bar\omega$. Для построения кусочно-линейного восполнения сеточного управления $\Phi_{0h}(x)$ на Ω каждую элементарную ячейку $e^+(x) \subset \bar\Omega$ области $\bar\Omega$, где $e^+(x) = e^+(x_1, x_2) = \{\xi = (\xi_1, \xi_2) : x_\alpha \le \xi_\alpha \le x_\alpha + h_\alpha, \alpha = 1, 2\} = \bar e_1'(x_1) \times \bar e_2'(x_2)$, $x \in \omega_1^- \times \omega_2^-$, разобьем диагональю, образующей тупой угол с осью ξ_1, на два треугольника. Треугольники, расположенные выше диагонали, будем обозначать через $\Delta_-(x)$, а нижние – через $\Delta_+(x)$. Здесь $x = (x_1, x_2)$ – узел, соответствующий вершине прямого угла треугольника $\Delta_\pm(x)$. Пусть r – управляющий параметр, принимающий значения ± 1. Если через $\Phi_{0hr}'(\xi)$ обозначить функцию, определенную в $\Delta_r(x)$, линейную по ξ_1 и ξ_2 и совпадающую в вершинах треугольника с сеточной функцией $\Phi_{0h}(x)$, то кусочно-линейное восполнение на $\bar\Omega$ определится в виде $F_h^{(1)}\Phi_{0h}(\xi) = \{\Phi_{0hr}'(\xi), \xi \in \Delta_r(x), r = -1, +1, \Delta_r(x) \in \bar\Omega\}$.

Нетрудно убедиться, что явный вид восполнения в каждом из треугольников таков:

$$F_h^{(1)}\Phi_{0h}(\xi) = \begin{cases} \Phi_{0h}(x) + \Phi_{0hx_1}(x)(\xi_1 - x_1) + \Phi_{0hx_2}(x)(\xi_2 - x_2), & \xi \in \Delta_+(x), \quad x \in \omega_1^- \times \omega_2^-, \\ \Phi_{0h}(x) + \Phi_{0h\bar x_1}(x)(\xi_1 - x_1) + \Phi_{0h\bar x_2}(x)(\xi_2 - x_2), & \xi \in \Delta_-(x), \quad x \in \omega_1^+ \times \omega_2^+. \end{cases}$$

Затем каждую элементарную ячейку $e^+(x) \subset \bar\Omega$ разобьем диагональю, образующей острый угол с осью ξ_1, на два треугольника. Треугольники, расположенные выше диагонали, будем обозначать через $\Delta_-'(x)$, а нижние – через $\Delta_+'(x)$. Здесь $x = (x_1, x_2)$ – узел, соответствующий вершине прямого угла треугольника $\Delta_\pm'(x)$. Пусть r – управляющий параметр, принимающий значения ± 1. Через $\Phi_{0hr}''(\xi)$ обозначим функцию, определенную в $\Delta_r'(x)$, линейную по ξ_1 и ξ_2 и совпадающую в вершинах треугольника с сеточной функцией $\Phi_{0h}(x)$. Тогда кусочно-линейное восполнение на $\bar\Omega$, соответствующее данному разбиению, определится в виде $F_h^{(2)}\Phi_{0h}(\xi) = \{\Phi_{0hr}''(\xi), \xi \in \Delta_r'(x), r = -1,$

$+1,\ \Delta'_r(x) \in \overline{\Omega}\}$. Нетрудно убедиться, что явный вид восполнения в каждом из треугольников таков:

$$F_h^{(2)}\Phi_{0h}(\xi) = \begin{cases} \Phi_{0h}(x) + \Phi_{0h\bar{x}_1}(x)(\xi_1 - x_1) + \Phi_{0hx_2}(x)(\xi_2 - x_2), & \xi \in \Delta'_+(x), \quad x \in \omega_1^+ \times \omega_2^-, \\ \Phi_{0h}(x) + \Phi_{0hx_1}(x)(\xi_1 - x_1) + \Phi_{0h\bar{x}_2}(x)(\xi_2 - x_2), & \xi \in \Delta'_-(x), \quad x \in \omega_1^- \times \omega_2^+. \end{cases}$$

Теперь положим

$$F_h\Phi_{0h}(\xi) = \frac{1}{2}F_h^{(1)}\Phi_{0h}(\xi) + \frac{1}{2}F_h^{(2)}\Phi_{0h}(\xi), \quad \xi \in \overline{\Omega}. \tag{4.9}$$

Рассмотрим сеточные управления $\Phi_{\alpha h}(x)$, $x \in \omega$, $\alpha = 1, 2, 3$, и определим кусочно-постоянные восполнения на Ω сеточных управлений $\Phi_{\alpha h}(x)$, $x \in \omega$, $\alpha = 1, 2, 3$, по формуле

$$\hat{g}_\alpha(\xi) = \hat{P}_h\Phi_{\alpha h}(\xi) = \Phi_{\alpha h}(x), \quad \xi \in \hat{e}(x), \quad x \in \omega, \quad \alpha = 1, 2, 3, \tag{4.10}$$

где $\hat{e}(x) \subset \overline{\Omega}$ – элементарные ячейки области $\overline{\Omega}$:

$$\hat{e}(x_1, x_2) = e(x_1, x_2) = e_1(x_1) \times e_2(x_2), \quad x_\alpha = 2h_\alpha, 3h_\alpha, \ldots, l_\alpha - 2h_\alpha, \quad \alpha = 1, 2,$$

$$\hat{e}(h_1, h_2) = \{0 \le \xi_1 \le 1.5h_1, 0 \le \xi_2 \le 1.5h_2\};$$

$$\hat{e}(h_1, x_2) = \{0 \le \xi_1 \le 1.5h_1, x_2 - 0.5h_2 \le \xi_2 \le x_2 + 0.5h_2\}, \quad x_2 = 2h_2, 3h_2, \ldots, l_2 - 2h_2,$$

$$\hat{e}(h_1, l_2 - h_2) = \{0 \le \xi_1 \le 1.5h_1, l_2 - 1.5h_2 \le \xi_2 \le l_2\},$$

$$\hat{e}(l_1 - h_1, h_2) = \{l_1 - 1.5h_1 \le \xi_1 \le l_1, 0 \le \xi_2 \le 1.5h_2\},$$

$$\hat{e}(l_1 - h_1, x_2) = \{l_1 - 1.5h_1 \le \xi_1 \le l_1, x_2 - 0.5h_2 \le \xi_2 \le x_2 + 0.5h_2\}, \quad x_2 = 2h_2, 3h_2, \ldots, l_2 - 2h_2,$$

$$\hat{e}(l_1 - h_1, l_2 - h_2) = \{l_1 - 1.5h_1 \le \xi_1 \le l_1, l_2 - 1.5h_2 \le \xi_2 \le l_2\};$$

$$\hat{e}(x_1, h_2) = \{x_1 - 0.5h_1 \le \xi_1 \le x_1 + 0.5h_1, 0 \le \xi_2 \le 1.5h_2\}, \quad x_1 = 2h_1, 3h_1, \ldots, l_1 - 2h_1,$$

$$\hat{e}(x_1, l_2 - h_2) = \{x_1 - 0.5h_1 \le \xi_1 \le x_1 + 0.5h_1, l_2 - 1.5h_2 \le \xi_2 \le l_2\}, \quad x_1 = 2h_1, 3h_1, \ldots, l_1 - 2h_1.$$

Справедливы следующие леммы.

Лемма 6. *Пусть* $F_h\Phi_{0h}(\xi)$ *– кусочно-линейное восполнение на* Ω *сеточного управления* $\Phi_{0h}(x)$, $x \in \overline{\omega}$, *а* $\hat{P}_h\Phi_{\alpha h}(\xi)$ *– кусочно-постоянные восполнения на* Ω *сеточных управлений* $\Phi_{\alpha h}(x)$, $x \in \omega$, $\alpha = 1, 2, 3$, *определяемые формулами* (4.9) *и* (4.10) *соответственно. Пусть* $R_hg_0(x) \equiv R_hk(x) = S^xk(x)$, $x \in \overline{\omega}$, *и* $R_hg_\alpha(x) \equiv R_hb_\alpha(x) = S^xb_\alpha(x)$, $x \in \omega$, $\alpha = 1, 2$, $R_hg_3(x) \equiv R_hd(x) = S^xd(x)$, $x \in \omega$, *– дискретизации на сетках* $\overline{\omega}$ *и* ω *управлений* $g_0(\xi) \equiv k(\xi)$ *и* $g_\alpha(\xi) \equiv b_\alpha(\xi)$, $\alpha = 1, 2$, $g_3(\xi) \equiv d(\xi)$ *соответственно, где* $S^x = S^{x_1}S^{x_2}$ *– оператор усреднения по Стеклову. Тогда справедливы оценки*

$$\|F_h\Phi_{0h}(\xi)\|_{L_1(\Omega)} = \|\Phi_{0h}(x)\|_{L_1(\overline{\omega})}, \quad \|F_h\Phi_{0h}(\xi)\|_{L_2(\Omega)} \le \|\Phi_{0h}(x)\|_{L_2(\overline{\omega})}, \quad \|R_hk(x)\|_{L_1(\overline{\omega})} = \|k(\xi)\|_{L_1(\Omega)},$$

$$\left\|\frac{\partial F_h\Phi_{0h}(\xi)}{\partial \xi_\alpha}\right\|_{L_2(\Omega)} \le \|\Phi_{0hx_\alpha}(x)\|_{L_2(\overline{\omega}^{(-\alpha)})}, \quad \alpha = 1, 2, \quad \|R_hk(x)\|_{L_p(\overline{\omega})} \le \|k(\xi)\|_{L_p(\Omega)}, \quad p \ge 1,$$

$$\|R_hk(x)\|^2_{W_2^1(\overline{\omega})} \le \|k(\xi)\|^2_{W_2^1(\Omega)} + \gamma_h^{(0)}, \quad \|\hat{P}_h\Phi_{\alpha h}(\xi)\|^2_{L_2(\Omega)} \le \|\Phi_{\alpha h}(x)\|^2_{L_2(\omega)} + \gamma_h^{(\alpha)},$$

$$\|R_hg_\alpha(x)\|_{L_2(\omega)} \le \|g_\alpha(\xi)\|_{L_2(\Omega)}, \quad \alpha = 1, 2, 3, \quad \text{где } 0 < \gamma_h^{(m)} = O(h) \text{ при } h \longrightarrow 0, \quad m = 0, 1, 2, 3.$$

Лемма 7. *В предположениях леммы* 6 *справедливы оценки*

$$\left\|S^{x_1}k^{(-0.5_2)}(x) - \frac{R_hk(x) + (R_hk(x))^{(-1_2)}}{2}\right\|_{L_\infty(\omega^{(+2)})} \le M|h|\left\|\frac{\partial k(\xi)}{\partial \xi_2}\right\|_{L_\infty(\Omega)},$$

$$\left\| S^{x_2} k^{(-0.5_1)}(x) - \frac{(R_h k(x))^{(-1_2)} + (R_h k(x))^{(-1_1,-1_2)} + (R_h k(x))^{(+1_2)} + (R_h k(x))^{(-1_1,+1_2)}}{4} \right\|_{L_\infty(\omega^{(+1)})} \le$$

$$\le M|h| \sum_{\alpha=1}^{2} \left\| \frac{\partial k(\xi)}{\partial \xi_\alpha} \right\|_{L_\infty(\Omega)},$$

$$\left\| S^{x_1} [F_h \Phi_{0h}]^{(-0.5_2)}(x) - \frac{\Phi_{0h}(x) + \Phi_{0h}^{(-1_2)}(x)}{2} \right\|_{L_\infty(\omega^{(+2)})} \le M|h| \|\Phi_{0hx_1}(x)\|_{L_\infty(\omega_1^- \times \bar\omega_2)},$$

$$\left\| S^{x_2} [F_h \Phi_{0h}]^{(-0.5_1)}(x) - \frac{\Phi_{0h}^{(-1_2)}(x) + \Phi_{0h}^{(-1_1,-1_2)}(x) + \Phi_{0h}^{(+1_2)}(x) + \Phi_{0h}^{(-1_1,+1_2)}(x)}{4} \right\|_{L_\infty(\omega^{(+1)})} \le M|h| \|\Phi_{1hx_2}(x)\|_{L_\infty(\bar\omega_1 \times \omega_2^-)},$$

$$\|S^x b_\alpha(x) - R_h b_\alpha(x)\|_{L_\infty(\omega)} = 0, \quad \alpha = 1, 2, \quad \|S^x d(x) - R_h d(x)\|_{L_\infty(\omega)} = 0,$$

$$\|S^x [\hat{P}_h \Phi_{\alpha h}](x) - \Phi_{\alpha h}(x)\|_{L_\infty(\omega)} = 0, \quad \alpha = 1, 2, 3.$$

Доказательства лемм 6, 7 опираются на определения отображений F_h, $\hat{P}_h$, P_h и определение оператора усреднения по Стеклову $S^x = S^{x_1}S^{x_2}$. Доказательства довольно громоздки, и мы их опускаем из-за ограниченности объема статьи (см. также [9]).

Для исследования сходимости сеточных экстремальных задач $\mathbf{A^{(0)}}$, $\mathbf{A^{(1)}}$, $\mathbf{A^{(2)}} = \mathbf{A^{(2)}(p)}$ по функционалу и управлению рассмотрим последовательность разностных задач $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, зависящих от шага $h = (h_1, h_2)$ сетки $\bar\omega = \bar\omega_h = \bar\omega_1 \times \bar\omega_2 \subset \bar\Omega$ при $|h| \longrightarrow 0$. Причем всюду ниже в задачах $\mathbf{A^{(2)}} = \mathbf{A^{(2)}(p)}$ и $\mathbf{A_h^{(2)}} = \mathbf{A_h^{(2)}(p)}$ будем считать, что параметр p принимает значения $p = 1, 2$. Для изучения этих вопросов нам понадобятся результаты, полученные выше. Кроме того, для исследования связи между экстремальными задачами $\mathbf{A_h^{(\alpha)}}$ и $\mathbf{A^{(\alpha)}}$, $\alpha = 0, 1, 2$, введем отображения

$$\tilde{R}_h : H \longrightarrow H_h \text{ и } N_h : H_h \longrightarrow H, \tag{4.11}$$

которые определим следующим образом: $\tilde{R}_h g = \Phi_h$, где $g = (g_0, g_1, g_2, g_3) = (k(\xi), b_1(\xi), b_2(\xi), d(\xi))$, $\Phi_h = (R_h k(x), R_h b_1(x), R_h b_2(x), R_h d(x))$; $N_h \Phi_h = g$, где $\Phi_h = (\Phi_{0h}(x), \Phi_{1h}(x), \Phi_{2h}(x), \Phi_{3h}(x))$, $g = (F_h \Phi_{0h}(\xi)$, $\hat{P}_h \Phi_{1h}(\xi)$, $\hat{P}_h \Phi_{2h}(\xi)$, $\hat{P}_h \Phi_{3h}(\xi))$. Отображения R_h, F_h, $\hat{P}_h$ определены в лемме 6. Нетрудно показать, что для произвольных управлений $g \in U^{(\alpha)}$, $\Phi_h \in U_h^{(\alpha)}$ экстремальных задач $\mathbf{A^{(\alpha)}}$ и $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, соответственно справедливы включения $\tilde{R}_h g \in U_h^{(\alpha)}$ и $N_h \Phi_h \in U^{(\alpha)}$, $\alpha = 0, 1, 2$.

Лемма 8. *Пусть $g \in U^{(\alpha)}$ и $\Phi_h \in U_h^{(\alpha)}$ – произвольные управления в экстремальных задачах $\mathbf{A^{(\alpha)}}$ и $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$. Тогда для любого $h > 0$ справедливы оценки*

$$\begin{aligned} |J(g) - J_h(\tilde{R}_h g)| &= |I(u(\xi, g)) - I_h(y(x, \tilde{R}_h g))| \le C|h| \sum_{k=0}^{3} \alpha_k, \\ |J(N_h \Phi_h) - J_h(\Phi_h)| &= |I(u(\xi, N_h \Phi_h)) - I_h(y(x, \Phi_h))| \le C|h| \sum_{k=0}^{3} \alpha_k. \end{aligned} \tag{4.12}$$

Доказательство леммы 8 опирается на определения отображений $\tilde{R}_h g$ и $N_h \Phi_h$ и следует из теоремы 6 и лемм 6, 7.

Теорема 7. *Пусть $J_*^{(\alpha)}$ и $J_{h*}^{(\alpha)}$ – нижние грани функционалов $J(g)$ и $J_h(\Phi_h)$ в экстремальных задачах $\mathbf{A^{(\alpha)}}$ и $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, соответственно. Семейства сеточных задач $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, зависящих от шага $h = (h_1, h_2)$ сетки $\bar\omega \subset \bar\Omega$ при $|h| \longrightarrow 0$, аппроксимируют исходные экстремаль-*

ные задачи $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *по функционалу, т.е.* $\lim J_{h*}^{(\alpha)} = J_*^{(\alpha)}$ *при* $|h| \longrightarrow 0$, *и справедливы оценки скорости сходимости*

$$\left|J_{h*}^{(\alpha)} - J_*^{(\alpha)}\right| \le M(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3)|h|, \quad \alpha = 0, 1, 2. \tag{4.13}$$

Доказательство теоремы следует из оценок (4.12) и методики из [1].

Предположим теперь, что при каждом $h = (h_1, h_2)$ и при соответствующей сетке $\overline{\omega} = \overline{\omega}_h$ с помощью какого-либо метода минимизации получены приближенные значения $J_{h*}^{(\alpha)} + \varepsilon_h^{(\alpha)}$ нижних граней $J_{h*}^{(\alpha)}$ функционала $J_h(\Phi_h)$ на $U_h^{(\alpha)}$ в экстремальных задачах $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, и найдены сеточные управления $\Phi_{h\varepsilon_h}^{(\alpha)}(x) \equiv \hat{\Phi}_h^{(\alpha)}(x) \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, дающие приближенные решения соответствующих сеточных экстремальных задач $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$, в следующем смысле:

$$J_{h*}^{(\alpha)} \le J_h(\hat{\Phi}_h^{(\alpha)}) \le J_{h*}^{(\alpha)} + \varepsilon_h^{(\alpha)}, \quad \hat{\Phi}_h^{(\alpha)} \in U_h^{(\alpha)}, \quad \alpha = 0, 1, 2, \tag{4.14}$$

где последовательности $\{\varepsilon_h^{(\alpha)}\}$, $\alpha = 0, 1, 2$, таковы, что $\varepsilon_h^{(\alpha)} \ge 0$ и $\varepsilon_h^{(\alpha)} \longrightarrow 0$ при $|h| \longrightarrow 0$, $\alpha = 0, 1, 2$.

Справедлива следующая

Теорема 8. *Пусть последовательности сеточных управлений* $\{\hat{\Phi}_h^{(\alpha)}(x)\} \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, *определены из условий* (4.14). *Тогда последовательности управлений* $\{N_h\hat{\Phi}_h^{(\alpha)}(\xi)\}$, $\alpha = 0, 1, 2$, *где отображение* $N_h : H_h \longrightarrow H$ *определено в* (4.11), *являются минимизирующими для функционалов* $J(g)$ *исходных задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *оптимального управления, т.е.* $\lim J(N_h\hat{\Phi}_h^{(\alpha)}) = J_*^{(\alpha)}$, $\alpha = 0, 1, 2$, *при* $|h| \longrightarrow 0$ *и справедливы оценки скорости сходимости*

$$0 \le J(N_h\hat{\Phi}_h^{(\alpha)}) - J_*^{(\alpha)} \le C[(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3)|h| + \varepsilon_h^{(\alpha)}], \quad \alpha = 0, 1, 2;$$

последовательности $\{N_h\hat{\Phi}_h^{(\alpha)}\}$, $\alpha = 0, 1, 2$, *слабо в* $H = W_2^1(\Omega) \times (L_2(\Omega))^3$ *сходятся к соответствующим множествам* $U_*^{(\alpha)} \ne \varnothing$, $\alpha = 0, 1, 2$, *оптимальных управлений исходных задач* $\mathbf{A}^{(\alpha)}$, $\alpha =$ $= 0, 1, 2$.

Доказательство теоремы проводится на основе методики из [1] и опирается на полученные выше результаты. Слабая сходимость минимизирующих последовательностей $\{N_h\hat{\Phi}_h^{(\alpha)}\}$, $\alpha = 0, 1, 2$, к соответствующим множествам $U_*^{(\alpha)} \ne \varnothing$, $\alpha = 0, 1, 2$, следует из теоремы 1.

Рассмотрим теперь вопрос о сильной сходимости в H по аргументу (управлению) разностных аппроксимаций $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$. Для регуляризации семейств сеточных экстремальных задач $\mathbf{A^{(0)}}$, $\mathbf{A^{(1)}}$, $\mathbf{A^{(2)}} = \mathbf{A^{(2)}}(\mathbf{p})$ введем на $U^{(\alpha)}$, $\alpha = 0, 1, 2$, функционал-стабилизатор $\Omega(g) = \|g\|_H^2$, $g \in U^{(\alpha)}$, $\alpha =$ $= 0, 1, 2$, и его сеточный аналог $\Omega_h(\Phi_h) = \|\Phi_h\|_{H_h}^2$, $\Phi_h \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$. При каждом $h = (h_1, h_2)$ рассмотрим на $U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, сеточный функционал Тихонова экстремальных задач $\mathbf{A_h^{(\alpha)}}$, $\alpha = 0, 1, 2$: $T_{h\alpha_h}(\Phi_h) = J_h(\Phi_h) + \alpha_h\Omega_h(\Phi_h)$, $\Phi_h \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, где $\{\alpha_h\}$ – произвольная последовательность положительных чисел, сходящаяся к нулю при $|h| \longrightarrow 0$. Рассмотрим теперь задачи минимизации функционала $T_{h\alpha_h}(\Phi_h)$ на $U_h^{(\alpha)}$, $\alpha = 0, 1, 2$: при каждом $h = (h_1, h_2)$ определим сеточные управления $\hat{\Phi}_h^{(\alpha)} = \Phi_{h\alpha_h\nu_h}^{(\alpha)}(x) \in U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, удовлетворяющие условиям

$$T_{h\alpha_h*}^{(\alpha)} = \inf\{T_{h\alpha_h}(\Phi_h) : \Phi_h \in U_h^{(\alpha)}\} \le T_{h\alpha_h*}^{(\alpha)} + \nu_h^{(\alpha)}, \tag{4.15}$$

где $\nu_h^{(\alpha)} \ge 0$ и $\nu_h^{(\alpha)} \longrightarrow 0$ при $|h| \longrightarrow 0$, $\alpha = 0, 1, 2$.

Теорема 9. *Пусть последовательности сеточных управлений* $\{\hat{\Phi}_h^{(\alpha)}\} \subset U_h^{(\alpha)}$, $\alpha = 0, 1, 2$, *определены из условий* (4.15). *Тогда последовательности управлений* $\{N_h\hat{\Phi}_h^{(\alpha)}(\xi)\}$, $\alpha = 0, 1, 2$, *где отображение* $N_h : H_h \longrightarrow H$ *определено в* (4.11), *являются минимизирующими для функционалов* $J(g)$ *исходных задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *оптимального управления, т.е.* $\lim J(N_h\hat{\Phi}_h^{(\alpha)}) = J_*^{(\alpha)}$, $\alpha = 0, 1, 2$, *при* $|h| \longrightarrow 0$ *и справедливы оценки скорости сходимости*

$$0 \le J(N_h\hat{\Phi}_h^{(\alpha)}) - J_*^{(\alpha)} \le M[(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3)|h| + \nu_h^{(\alpha)} + \alpha_h], \quad \alpha = 0, 1, 2.$$

Если, кроме того, параметры $\nu_h^{(\alpha)}$, α_h *согласованы с* $|h|$ *так, что* $\nu_h^{(\alpha)}, \alpha_h \longrightarrow +0$ *при* $|h| \longrightarrow 0$ *и* $(|h| + \nu_h^{(\alpha)})/\alpha_h \longrightarrow 0$, *то последовательности* $\{N_h\hat{\Phi}_h^{(\alpha)}\}$, $\alpha = 0, 1, 2$, *сильно в* H *сходятся к множествам* Ω*-нормальных (в смысле минимальной нормы) оптимальных управлений* $U_{**}^{(\alpha)}$, $\alpha = 0, 1, 2$, *задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *т.е.*

$$\lim\rho(N_h\hat{\Phi}_h^{(\alpha)}, U_{**}^{(\alpha)}) = \lim \inf\{\|N_h\hat{\Phi}_h^{(\alpha)} - g_{**}\|_H : g_{**} \in U_{**}^{(\alpha)}\} = 0 \ \ \text{при} \ \ |h| \longrightarrow 0, \quad \alpha = 0, 1, 2,$$

$$\lim\Omega(N_h\hat{\Phi}_h^{(\alpha)}) = \lim\|N_h\hat{\Phi}_h^{(\alpha)}\|_H^2 = \Omega_*^{(\alpha)} = \inf_{g_* \in U_*^{(\alpha)}} \Omega(g_*) \ \ \text{при} \ \ |h| \longrightarrow 0,$$

где $U_{**}^{(\alpha)} = \{g_{**} \in U_*^{(\alpha)} : \Omega(g_{**}) = \inf\{\Omega(g_*) : g_* \in U_*^{(\alpha)}\} = \Omega_*^{(\alpha)}\}$ *– множества* Ω*-нормальных решений задач* $\mathbf{A}^{(\alpha)}$, $\alpha = 0, 1, 2$, *оптимального управления*; $\Omega(g) = \|g\|_H^2$.

Доказательство теоремы 9 проводится на основе методики из [1], [23] и опирается на полученные выше результаты.

Полученные результаты не зависят от способа решения разностных задач оптимального управления $\mathbf{A}^{(0)}$, $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)} = \mathbf{A}^{(2)}(\mathbf{p})$.

СПИСОК ЛИТЕРАТУРЫ

1. *Васильев Ф.П.* Методы оптимизации. М.: Факториал Пресс, 2002.
2. *Лионс Ж.-Л.* Оптимальное управление системами, описываемыми уравнениями с частными производными. М.: Мир, 1972.
3. *Литвинов В.Г.* Оптимизация в эллиптических граничных задачах с приложениями к механике. М.: Наука, 1987.
4. *Лурье К.А.* Оптимальное управление в задачах математической физики. М.: Наука, 1975.
5. *Райтум У.Е.* Задачи оптимального управления для эллиптических уравнений. Рига: Зинатне, 1989.
6. *Ишмухаметов А.З.* Вопросы устойчивости и аппроксимации задач оптимального управления. М.: ВЦ РАН, 1999.
7. *Ишмухаметов А.З.* Вопросы устойчивости и аппроксимации задач оптимального управления системами с распределенными параметрами. М.: ВЦ РАН, 2001.
8. *Потапов М.М.* Аппроксимация экстремальных задач в математической физике (гиперболические уравнения). М.: Изд-во МГУ, 1985.
9. *Лубышев Ф.В.* Разностные аппроксимации задач оптимального управления системами, описываемыми уравнениями в частных производных. Уфа: БашГУ, 1999.
10. *Самарский А.А., Андреев В.Б.* Разностные методы для эллиптических уравнений. М.: Наука, 1976.
11. *Самарский А.А., Лазаров Р.Д., Макаров В.Л.* Разностные схемы для дифференциальных уравнений с обобщенными решениями. М.: Высш. школа, 1987.
12. *Лубышев Ф.В.* Точность разностных аппроксимаций и регуляризация задач оптимального управления для эллиптического уравнения с управлениями в коэффициентах // Ж. вычисл. матем. и матем. физ. 1989. Т. 29. № 9. С. 1431–1444.
13. *Лубышев Ф.В.* Аппроксимация и регуляризация задач оптимального управления для несамосопряженного эллиптического уравнения с переменными коэффициентами // Ж. вычисл. матем. и матем. физ. 1991. Т. 31. № 1. С. 17–30.
14. *Лубышев Ф.В.* Аппроксимация и регуляризация задач оптимального управления коэффициентами параболических уравнений // Ж. вычисл. матем. и матем. физ. 1993. Т. 33. № 8. С. 1166–1183.

15. *Лубышев Ф.В.* Аппроксимация и регуляризация задач оптимального управления для параболических уравнений с управлениями в коэффициентах // Докл. РАН. 1996. Т. 349. № 5. С. 98–602.
16. *Лубышев Ф.В., Файрузов М.Э.* Аппроксимация и регуляризация задач оптимального управления для квазилинейных эллиптических уравнений // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. № 8. С. 1148–1164.
17. *Лубышев Ф.В., Манапова А.Р.* О некоторых аппроксимациях задач оптимального управления коэффициентами квазилинейных эллиптических уравнений // Тр. СВМО. М., 2006. Т. 8. № 1. С. 250–259.
18. *Марчук Г.И.* Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982.
19. *Марчук Г.И.* Сопряженные уравнения и анализ сложных систем. М.: Наука, 1992.
20. *Гаевский Х., Грегер К., Захариас К.* Нелинейные операторные уравнения и операторные дифференциальные уравнения. М.: Мир, 1978.
21. *Ладыженская О.А.* Краевые задачи математической физики. М.: Наука, 1973.
22. *Соболев С.Л.* Некоторые применения функционального анализа в математической физике. Новосибирск: СО АН СССР, 1962.
23. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М.: Наука, 1986.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 397–413

УДК 519.852.6

ЛОКАЛЬНЫЙ ПОИСК В ЗАДАЧАХ С НЕВЫПУКЛЫМИ ОГРАНИЧЕНИЯМИ[1)]

© 2007 г. Т. В. Груздева, А. С. Стрекаловский

(664033 Иркутск, ул. Лермонтова, 134, Ин-т динамики систем и теории управления СО РАН)

e-mail: yak@icc.ru; strekal@icc.ru

Поступила в редакцию 07.08.2006 г.

Рассмотрены невыпуклые задачи оптимизации с ограничением-неравенством, заданным разностью двух выпуклых функций (d.c.-функцией). Предложены два метода поиска локальных решений в этой задаче, сочетающие решение частично линеаризованных задач и спуск на поверхность уровня d.c.-функции. Исследована сходимость этих методов, и предложены критерии останова. Приведены результаты вычислительного эксперимента по сравнительному тестированию предложенных методов локального поиска. Библ. 15. Табл. 4.

Ключевые слова: разность двух выпуклых функций, локальный поиск, линеаризованная задача, поверхность уровня, критическая точка.

1. ВВЕДЕНИЕ

Как известно (см. [1], [2]), в невыпуклых задачах может существовать большое количество локальных и седловых решений, а также стационарных (ККТ) точек, весьма далеких от глобального оптимума (многоэкстремальность, см. [1]–[6]). И, как следствие, классические методы выпуклой оптимизации чаще всего демонстрируют свою неэффективность при решении невыпуклых задач.

Поэтому развитие “глобальной” оптимизации наряду с методами ветвей и границ и отсечений в настоящий момент включает в себя построение и исследование методов решения специальных классов невыпуклых задач (см. [4]–[12]). Весьма привлекательными в этом плане выглядят задачи с d.c.-функциями, представимыми в виде разности двух выпуклых функций (d.c.-оптимизация).

Известно (см. [4]–[6]), что любую непрерывную задачу оптимизации можно с любой точностью приблизить некоторой d.c.-задачей. В частности, задачи с d.c.-ограничениями типа неравенства оказываются весьма обширным классом задач оптимизации. Некоторые авторы (см., например, [5]) сводят к этому классу почти все другие невыпуклые задачи. В частности, d.c.-ограничениями описывается объединение конечного семейства многогранников (см. [3]). К тому же конечное число d.c.-ограничений нетрудно свести к одному (см. [5], [6]). Поэтому простой вид задачи

$$\begin{gathered} f(x) \downarrow \min, \quad x \in S, \\ F(x) \triangleq g(x) - h(x) \geq 0, \end{gathered} \tag{P}$$

где $f(\cdot), g(\cdot), h(\cdot)$ – выпуклые функции, а S – выпуклое множество из $\mathbb{R}^n$, не должен обманывать исследователя. Это очень трудная задача при условии, что ограничение $F(x) \geq 0$ активно (см. [4]–[6]). Иначе, решая релаксированную задачу

$$f(x) \downarrow \min, \quad x \in S, \tag{PW}$$

можно отыскать решения и для задачи (P). Поэтому мы будем предполагать в дальнейшем выполнение одного из условий регулярности (см. [4]–[6]), например

$$\exists v \in S, \quad F(v) < 0 : f(v) < f_* \triangleq V(P), \tag{H_0}$$

где $V(P) \triangleq \inf\limits_x \{ f(x) \mid x \in S, F(x) \geq 0 \}$ – (оптимальное) значение задачи (P).

1) Работа выполнена при финансовой поддержке РФФИ (код проекта 05-01-00110) и Фонда содействия отечественной науке.

Одним из следствий условия регулярности (H_0) является то, что, решая релаксированную задачу (PW) одним из современных методов выпуклой оптимизации (см. [1]–[3]) и исходя из произвольной стартовой точки $x^0 \in S$, $F(x^0) > 0$, мы непременно "сядем" на ограничение $F(x) = 0$ или найдем точку v, удовлетворяющую условию (H_0). Этот факт используется различными авторами для глобального поиска в задаче (P) (см. [4]–[6]).

Сама схема глобального поиска чаще всего носит двухступенчатый характер: сначала локальный поиск, а затем процедуры глобального поиска (имеющие порой даже нематематическую природу), например процедуры ветвей и границ, отсечений (см. [4], [5]), стратегии глобального поиска (см. [6]), процедуры смены окрестностей (генетические, табу и т.д.). При этом специальным методам локального поиска придается весьма важное значение (см. [4]–[6]). Это объясняется, в частности, тем, что применение напрямую методов выпуклой оптимизации к многоэкстремальным задачам может привести (и чаще всего приводит) к неожиданным эффектам (см. [1], [2], [6]), когда, например, метод для задачи минимизации попадает в окрестность локального максимума и не может оттуда "выскочить" (см. [2, с. 167]).

Наш вычислительный опыт (см. [6]) свидетельствует о том, что для каждого класса невыпуклых задач нужно разрабатывать специальные методы локального поиска, включая в них известные методы выпуклой оптимизации как составную часть.

Настоящая работа как раз и посвящена этой проблематике. Так, в разд. 2 предлагается специальный метод локального поиска, порожденный последовательной линеаризацией ассоциированной (см. [4], [5]) с (P) задачи d.c.-максимизации

$$F(x) \uparrow \max, \quad x \in S, \quad f(x) \le \beta. \qquad (Q_\beta)$$

При этом исследуется сходимость самого метода и составляющих его процедур.

Далее, в разд. 3 предлагается другая схема локального поиска, развивающая идею Розена (см. [13], [14]) для обратно-выпуклой задачи на задачу с d.c.-ограничением. Также исследуется сходимость разработанных процедур.

В заключение приводятся результаты численного решения предложенными алгоритмами серий тестовых задач, а также анализируются результаты численного эксперимента.

2. СПЕЦИАЛЬНЫЙ МЕТОД ЛОКАЛЬНОГО ПОИСКА

Предположим, что в задаче (P) функции $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ выпуклы, а $g(\cdot)$ еще и дифференцируема на некотором выпуклом открытом множестве Ω из $\mathbb{R}^n$, S – выпуклое множество, $S \in \Omega$, $\{x \in \mathbb{R}^n \mid F(x) = 0\} \subset \Omega$.

Предлагается один метод локального поиска, состоящий из двух процедур. Первая процедура начинается с некоторой допустимой точки $y \in S$, $F(y) \ge 0$ и строит точку $x(y) \in S$ со свойствами $F(x(y)) = 0$, $f(x(y)) \le f(y)$.

Вторая процедура заключается в последовательном решении линеаризованных задач (которые, как нетрудно видеть, выпуклы)

$$(LQ(u, \beta)): \quad h(x) - \langle \nabla g(u), x \rangle \downarrow \min, \quad x \in S, \quad f(x) \le \beta, \qquad (2.1)$$

где параметры u и β будут определены ниже. Задача $(LQ(u, \beta))$ является частичной линеаризацией задачи d.c.-максимизации (Q_β), ассоциированной с задачей (P) (β – некоторый параметр).

Перейдем к более детальному описанию процесса вычислений.

Процедура 1. Пусть задана допустимая точка $y \in S$, $F(y) \ge 0$. Если $F(y) = 0$, то полагаем $x(y) = y$. Если же $F(y) > 0$, то, поскольку $0 > F(v)$ согласно (H_0), по непрерывности функции $F(\cdot)$ найдется $\lambda \in \,]0, 1[$ такое, что $F(x_\lambda) = 0$, где $x_\lambda = \lambda v + (1 - \lambda)y$. В этом случае полагаем $x(y) := x_\lambda$.

Далее, в силу выпуклости $f(\cdot)$ справедлива следующая цепочка:

$$f(x_\lambda) \le \lambda f(v) + (1 - \lambda) f(y) < \lambda f_* + (1 - \lambda) f(y) \le f(y), \qquad (2.2)$$

так что $f(x(y)) < f(y)$, что и требовалось.

Процедуру 1 иногда называют свободным (от ограничения $F(x) \ge 0$) спуском.

Процедура 2 стартует из допустимой точки $\tilde{x} \in S$, $F(\tilde{x}) = 0$ и строит последовательность $\{u^r\}$ такую, что $(r = 0, 1, \dots)$

$$u^r \in S, \quad F(u^r) \ge 0, \quad f(u^r) \le \beta, \qquad (2.3)$$

где $u^0 := \tilde{x}$, $\beta := f(\tilde{x})$.

Последовательность $\{u^r\}$ строится по следующему правилу.

Если известна точка u^r, $r \geq 0$, удовлетворяющая условию (2.3), то следующая точка u^{r+1} строится как приближенное решение линеаризованной выпуклой задачи

$$(\mathrm{LQ}(u^r, \beta)):\ h(x) - \langle \nabla g(u^r), x \rangle \downarrow \min, \quad x \in S, \quad f(x) \leq \beta. \tag{2.4}$$

Другими словами, справедливо неравенство

$$h(u^{r+1}) - \langle \nabla g(u^r), u^{r+1} \rangle - \delta_r \leq \inf_x \{h(x) - \langle \nabla g(u^r), x \rangle \mid x \in S,\ f(x) \leq \beta\}, \tag{2.5}$$

где числовая последовательность $\{\delta_r\}$ такова, что

$$\delta_r > 0, \quad r = 0, 1, 2, \ldots, \quad \sum_{r=0}^{\infty} \delta_r < +\infty. \tag{2.6}$$

При этом будем предполагать, что поиск u^{r+1}, удовлетворяющего неравенству (2.5), начинается с точки u^r, так что

$$h(u^{r+1}) - \langle \nabla g(u^r), u^{r+1} \rangle \leq h(u^r) - \langle \nabla g(u^r), u^r \rangle. \tag{2.7}$$

Естественно, это никак не ограничивает общности рассмотрений.

Предложение 1. *Предположим, что значение ассоциированной с* (P) *задачи* (Q_ζ) *конечно для некоторого* $\zeta \geq \beta$:

$$V(Q_\zeta) := \sup_x \{F(x) \mid x \in S,\ f(x) \leq \zeta\} < +\infty, \tag{2.8}$$

функции $g(\cdot)$ *и* $h(\cdot)$ *являются выпуклыми, а* $g(\cdot)$ *еще и непрерывно дифференцируемой на открытой области* Ω, *содержащей множество* $S \cup \{x \in \mathbb{R}^n \mid F(x) = 0\}$.

(i) *Тогда последовательность* $\{u^r\}$, *генерируемая процедурой* 2, *удовлетворяет условию*

$$\lim_{r \to \infty} [\inf_x \{h(x) - h(u^r) - \langle \nabla g(u^r), x - u^r \rangle \mid x \in S,\ f(x) \leq \beta\}] = 0. \tag{2.9}$$

(ii) *При этом для всякой предельной точки* u_* *последовательности* $\{u^r\}$ *справедливы неравенства*

$$h(x) - \langle \nabla g(u_*), x \rangle \geq h(u_*) - \langle \nabla g(u_*), u_* \rangle \quad \forall x \in S : f(x) \leq \beta, \tag{2.10}$$

$$F(u_*) \triangleq g(u_*) - h(u_*) \geq 0. \tag{2.11}$$

(iii) *В случае замкнутости S эта предельная точка* u_* *является решением линеаризованной* (*в этой точке* u_*) *задачи* $(\mathrm{LQ}(u_*, \beta))$ *и нормально стационарной в ассоциированной задаче* (Q_β).

Доказательство. В силу выпуклости $g(\cdot)$, из (2.5) вытекает

$$\begin{gathered} 0 \geq \inf_x \{h(x) - h(u^r) - \langle \nabla g(u^r), x - u^r \rangle \mid x \in S,\ f(x) \leq \beta\} \geq \\ \geq h(u^{r+1}) - h(u^r) - \langle \nabla g(u^r), u^{r+1} - u^r \rangle - \delta_r \geq \\ \geq h(u^{r+1}) - h(u^r) - g(u^{r+1}) + g(u^r) - \delta_r = F(u^r) - F(u^{r+1}) - \delta_r. \end{gathered} \tag{2.12}$$

Отсюда получаем

$$F(u^r) \leq F(u^{r+1}) + \delta_r,$$

так что последовательность $\{F(u^r)\}$ является почти монотонно возрастающей. Поэтому, принимая во внимание (2.6), по лемме 3 из [1, §2.6, с. 87] заключаем, что существует конечный предел этой последовательности, поскольку по формуле (2.8) функция $F(\cdot)$ ограничена сверху на множестве $\mathcal{F}_\beta := \{x \in S \mid f(x) \leq \beta\}$. А тогда из (2.12) следует (2.9).

Далее, при переходе из u^r в u^{r+1} не происходит "ухудшения" целевой функции задачи (2.4) (см. (2.7)). А тогда из выпуклости функции $g(\cdot)$ получаем

$$0 \geq h(u^{r+1}) - h(u^r) - \langle \nabla g(u^r), u^{r+1} - u^r \rangle \geq F(u^r) - F(u^{r+1}),$$

так что $F(u^{r+1}) \ge F(u^r)$. Поэтому из $F(u^0) = 0$ вытекает $F(u^r) \ge 0$ $\forall r = 0, 1, \ldots$, откуда и следует справедливость (2.11).

Наконец, если $u^r \longrightarrow u_*$, то (2.10) следует по непрерывности $\nabla g(\cdot)$ и $h(\cdot)$ из (2.9), а также из очевидного неравенства

$$h(x) - h(u^r) - \langle \nabla g(u^r), x - u^r \rangle \ge$$
$$\ge \inf_u \{ h(u) - h(u^r) - \langle \nabla g(u^r), u - u^r \rangle \mid u \in S,\ f(u) \le \beta \}$$

при произвольном, но фиксированном $x \in S$, $f(x) \le \beta$.

Условие (2.10) означает, что если $u_* \in S$, то точка u_* является решением линеаризованной задачи ($LQ(u_*, \beta)$). Поэтому в случае дифференцируемости в точке u_* функции $h(\cdot)$ найдутся множители λ_0, $\lambda \ge 0$, $\lambda_0 + \lambda > 0$, при которых справедливы соотношения

$$\langle \lambda(\nabla h(u_*) - \nabla g(u_*)) + \lambda_0 \nabla f(u_*), x - u_* \rangle \ge 0 \quad \forall x \in S,$$
$$\lambda_0 [f(u_*) - \beta] = 0.$$

При этом из условия (H_0) следует, что

$$\exists v \in S, \quad f(v) < f_* \le f(u_*) \le \beta.$$

Это означает, что в задаче ($LQ(u_*, \beta)$) выполнено условие Слейтера. Поэтому $\lambda > 0$, что и завершает доказательство.

Теперь перейдем к вопросу о том, как с помощью последовательности $\{u^r\}$, где $u^0 := \tilde{x}$, $\beta = f(\tilde{x})$, определить точку $y(\tilde{x})$. Для этого зададимся некоторым числом $\varepsilon > 0$, и пусть $r \ge 0$. Тогда если $\delta_r \le \varepsilon/2$ и справедливо неравенство

$$h(u^{r+1}) - h(u^r) - \langle \nabla g(u^r), u^{r+1} - u^r \rangle \le \varepsilon/2, \tag{2.13}$$

то полагаем $y = y(\tilde{x}, \varepsilon) := u^r$. При этом вместо (2.13) можно использовать неравенство

$$F(u^{r+1}) - F(u^r) \le \varepsilon/2.$$

В обоих случаях из (2.12) вытекает, что точка $y = y(\tilde{x}, \varepsilon)$ удовлетворяет условию

$$\inf_x \{ h(x) - h(y) - \langle \nabla g(y), x - y \rangle \mid x \in S,\ f(x) \le \beta \} \ge -\varepsilon,$$

т.е. является ε-приближенным решением задачи ($LQ(y, \beta)$) – (2.4), где $\beta = f(\tilde{x})$.

Теперь объединим процедуры 1 и 2 в один метод и запишем его в более алгоритмической форме. Пусть задана допустимая точка $x_0 \in S$, $F(x_0) \ge 0$, а также две числовые последовательности: $\{\delta_r\}$, удовлетворяющая условию (2.6), и $\{\varepsilon_s\}$ такая, что

$$\varepsilon_s > 0, \quad s = 0, 1, 2, \ldots, \quad \varepsilon_s \downarrow 0 \ \ (s \longrightarrow +\infty). \tag{2.14}$$

Специальный метод локального поиска (СМЛП)

Шаг 0 (инициализация). Положить $s := 0$, $x^s := x_0$, $\beta_s := f(x^s)$.

Шаг 1 (процедура 2). Начиная из точки x^s, построить точку $y^s = y(x^s, \varepsilon_s)$:

$$y^s \in S, \quad F(y^s) \ge 0, \quad f(y^s) \le \beta_s,$$

являющуюся ε_s-решением линеаризованной задачи ($LQ(y^s, \beta_s)$), т.е.

$$h(x) - h(y^s) - \langle \nabla g(y^s), x - y^s \rangle \ge -\varepsilon_s \ \forall x \in S : \ f(x) \le \beta_s.$$

Шаг 2 (критерий останова). Если $F(y^s) \le 0$, то STOP.

Шаг 3 (процедура 1). С помощью y^s построить точку $u := x(y^s)$:

$$u \in S, \quad F(u) = 0, \quad f(u) < f(y^s) \le \beta_s.$$

Шаг 4. Положить $s := s + 1$, $x^s := u$, $\beta_s := f(u)$ и вернуться на шаг 1.

Лемма 1. *Метод локального спуска, описанный выше, обладает такими свойствами:*

(а) *либо конечен, причем* $F(y^N) = 0$, *где* N – *итерация останова,*

(б) *либо генерирует две последовательности* $\{x^s\}$ *и* $\{y^s\}$ *со свойствами*

$$x^s \in S, \quad F(x^s) = 0, \quad y^s \in S, \quad F(y^s) > 0, \tag{2.15}$$

$$\beta_{s+1} := f(x^{s+1}) < f(y^s) \le \beta_s := f(x^s). \tag{2.16}$$

При этом справедливо равенство

$$\beta_* := \lim_{s\to\infty} \beta_s = \lim_{s\to\infty} f(y^s). \tag{2.17}$$

Доказательство. (а) Согласно критерию останова (шаг 2), если N – итерация останова счета, то $F(y^N) \le 0$. А поскольку из доказательства предложения 1 видно, что $F(u^r) \ge 0$, $r = 1, 2, \dots$, то и $F(y^s) \ge 0$, $s = 0, 1, \dots$. Поэтому $F(y^N) = 0$.

(б) Если процесс счета по СМЛП бесконечен, согласно описаниям метода и процедуры 1, то $F(y^s) \ge 0$ и благодаря выпуклости $f(\cdot)$ имеем

$$\begin{gathered}\beta_{s+1} := f(x^{s+1}) = f(\lambda_s v + (1-\lambda_s)y^s) \le \lambda_s f(v) + (1-\lambda_s)f(y^s) < \\ < \lambda_s f_* + (1-\lambda_s)f(y^s) \le f(y^s) \le \beta_s,\end{gathered} \tag{2.18}$$

что и доказывает (2.16). А поскольку

$$-\infty < f_* < \beta_{s+1} < f(y^s) \le \beta_s, \quad s = 0, 1, \dots, \tag{2.19}$$

то существует предел $\lim \beta_s = \beta_* \ge f_*$, откуда с помощью (2.19) получаем (2.17).

Теорема 1. *Пусть функция* $f(\cdot)$ *и множество* S *выпуклы, множество* $\mathcal{F}_0 = \{x \in S \mid f(x) \le f(x_0)\}$ *ограничено и выполнено условие* (H_0).

Тогда специальный метод локального поиска имеет особенности:

а) *при конечном числе итераций получает точку* $y^N \in S$, $F(y^N) = 0$, *которая является* ε_N-*решением линеаризованной задачи* $(\mathrm{LQ}(y^N, \beta_N)$, *где* N – *номер итерации останова;*

б) *при бесконечном числе итераций для последовательностей* $\{x^s\}$ *и* $\{y^s\}$ *в дополнение к свойствам* (2.15)–(2.17) *выполнены условия*

$$0 = F(x^s) = \lim_{s\to\infty} F(y^s), \tag{2.20}$$

$$x_* = \lim_{s\to\infty} x^s = \lim_{s\to\infty} y^s \tag{2.21}$$

для некоторого $x_* \in \mathbb{R}^n$, $F(x_*) = 0$.

Кроме того, в случае замкнутости множества S *точка* x_* *является решением линеаризованной задачи* $(\mathrm{LQ}(x_*, \beta_*))$, *которое характеризуется условием*

$$\langle \nabla h(x_*) - \nabla g(x_*), x - x_* \rangle \ge 0 \;\; \forall x \in S : f(x) \le \beta_*, \tag{2.22}$$

а x_* *оказывается также нормально стационарной в ассоциированной задаче* (Q_{β_*}).

Доказательство. Утверждение а) вытекает из леммы 1, утверждение б) – из (2.16)–(2.18), откуда следует $\beta_{s+1} \le \lambda_s[f(v) - f(y^s)] + f(y^s)$, или

$$0 \le \lambda_s[f_* - f(v)] \le \lambda_s[f(y^s) - f(v)] \le f(y^s) - \beta_{s+1}.$$

Поэтому из (2.17) получаем

$$\lim_{s\to\infty} \lambda_s = 0, \quad \lambda_s \in \left]0, 1\right[, \tag{2.23}$$

поскольку $f(y^s) \ge f_* > f(v)$.

По построению имеем $f(x^0) = \beta_0$, $\{x^s\} \subset \mathcal{F}_0$, а $\mathcal{F}_0$ ограничено. Поэтому с точностью до подпоследовательности можно считать, что $x^s \longrightarrow x_*$.

Далее, поскольку $x^{s+1} = y^s + \lambda_s(v - y^s)$, то

$$\begin{aligned} \|y^s - x_*\| &= \|x^{s+1} + \lambda_s(y^s - v) - x_*\| \le \\ &\le \|x^{s+1} - x_*\| + \lambda_s\|y^s - v\| \le \|x^{s+1} - x_*\| + \lambda_s K \end{aligned} \tag{2.24}$$

для некоторого $K > 0$ такого, что $\|y^s - v\| \le K$.

Последняя оценка справедлива, поскольку по построению $y^s \in \mathcal{F}_0$, $s = 0, 1, 2, \ldots$. При этом равенство $\lim y^s = x_*$ следует из (2.23) и (2.24). Поэтому в силу равенств $F(x^s) = 0$, $s = 0, 1, \ldots$, и непрерывности $F(\cdot)$ имеем

$$0 = F(x^s) = F(x_*) = \lim_{s \to \infty} F(y^s),$$

так что (2.20) и (2.21) доказаны.

Далее напомним, что по построению для каждого $y^s = y(x^s, \varepsilon_s)$ при любом $s = 0, 1, \ldots$ справедливо неравенство

$$h(x) - h(y^s) - \langle \nabla g(y^s), x - y^s \rangle \ge -\varepsilon_s \ \forall x \in S, \quad f(x) \le \beta_s. \tag{2.25}$$

А поскольку последовательность $\{\beta_s\}$ строго убывающая: $\beta_{s+1} < \beta_s$, то очевидно включение $X_* \subset X_s$, где $X_* \triangleq \{x \in S \,|\, f(s) \le \beta_*\}$, $X_s \triangleq \{x \in S \,|\, f(x) \le \beta_s\}$. Поэтому неравенство (2.25) справедливо при любом фиксированном $x \in X_*$. Переходя в (2.25) к пределу ($s \longrightarrow \infty$) при таком $x \in X_*$, имеем

$$h(x) - \langle \nabla g(x_*), x \rangle \ge h(x_*) - \langle \nabla g(x_*), x_* \rangle.$$

Это и означает, что $x_* \in \mathrm{Sol}(\mathrm{LQ}(x_*, \beta_*))$ в случае замкнутости множества S.

Нормальная стационарность точки x_* доказывается так же, как в предложении 1.

Замечание 1. Если критерий останова СМЛП $F(y^s) \le 0$ заменить на более реалистичный и заключающийся, скажем, в одновременном выполнении трех следующих неравенств:

$$F(y^s) \le \rho, \quad \varepsilon_s \le \rho, \quad \beta_{s-1} - \beta_s \le \rho, \tag{2.26}$$

где ρ – заданная точность, то нетрудно видеть, что СМЛП окажется конечным.

При этом в качестве выхода он будет выдавать точку y^N со свойствами

$$F(y^N) \le \rho, \quad f(y^N) \le \beta_N,$$
$$\langle \nabla g(y^N), x - y^N \rangle - h(x) + h(y^N) \le \rho \ \forall x \in S : f(x) \le \beta_N,$$

что вполне приемлемо для локального поиска.

3. МОДИФИЦИРОВАННЫЙ МЕТОД РОЗЕНА

Следующий метод локального поиска комбинирует предложенную в [13] для частного случая задачи (P) – обратно-выпуклой задачи – идею линеаризации по базовой невыпуклости в ограничении (т.е. по функции $g(\cdot)$) и процедуру свободного спуска.

Рассмотрим линеаризованные задачи вида

$$(\mathrm{PL}(u)): \ f(x) \downarrow \min, \quad x \in S, \quad h(x) - \langle \nabla g(u), x - u \rangle - g(u) \le 0. \tag{3.1}$$

Здесь, в отличие от задачи ($\mathrm{LQ}(u, \beta)$), целевая функция задачи (P) остается на месте (не переходит в ограничение, как в задаче ($\mathrm{LQ}(u, \beta)$), а линеаризуется ограничение $F(x) = g(x) - h(x) \ge 0$, причем лишь частично по функции $g(\cdot)$).

Предположим, что в задаче (P) точка $u \in S$, в которой производится линеаризация функции $g(\cdot)$, является допустимой: $u \in D = \{x \in S \,|\, F(x) \ge 0\} \ne \emptyset$.

При этом задача (3.1) оказывается выпуклой и ее допустимое множество непусто:

$$D(u) \triangleq \{x \in S \,|\, h(x) - \langle \nabla g(u), x - u \rangle - g(u) \le 0\} \ne \emptyset, \tag{3.2}$$

поскольку $u \in D(u)$. С другой стороны, нетрудно видеть, что $D(u) \subset D$, поскольку из выпуклости $g(\cdot)$ следует неравенство

$$0 \le \langle \nabla g(u), x-u\rangle + g(u) - h(x) \le F(x),$$

которое и обеспечивает требуемое включение.

Далее будет рассмотрено следующее условие (см. [6], [7], [9]):

$$\begin{gathered} \forall y \in S \quad F(y) = 0 \ \exists d = d(y) \in S: \\ h(d) - h(y) < \langle \nabla g(y), d - y\rangle, \end{gathered} \tag{3.3}$$

а также условие регулярности (H_0).

Метод локального поиска состоит из процедуры 1 из СМЛП и следующей процедуры последовательного решения задачи ($\mathrm{PL}(u)$).

Процедура 3 стартует из допустимой точки $\tilde{x} \in S$, $F(\tilde{x}) = 0$ и строит последовательность $\{u^r\}$ такую, что

$$u^r \in S, \quad F(u^r) \ge 0, \quad f(u^r) \le f(\tilde{x}), \quad r = 0, 1, \ldots, \tag{3.4}$$

причем $u^0 := \tilde{x}$. Последовательность $\{u^r\}$ строится по следующему правилу.

Если известна точка u^r, $r \ge 0$, удовлетворяющая условию (3.4), то следующая точка u^{r+1} строится как приближенное решение выпуклой задачи

$$(\mathrm{PL}_r):\ f(x) \downarrow \min, \quad x \in S, \quad h(x) - \langle \nabla g(u^r), x - u^r\rangle - g(u^r) \le 0. \tag{3.5}$$

Точнее говоря, справедливо неравенство

$$f(u^{r+1}) \le \inf_x \{ f(x) \mid x \in S, h(x) - \langle \nabla g(u^r), x - u^r\rangle - g(u^r) \le 0\} + \delta_r, \tag{3.6}$$

где $\{\delta_r\}$ – числовая последовательность, удовлетворяющая условию (2.6).

Обозначим допустимое множество задачи (PL_r), а именно (3.5), через D_r:

$$D_r := D(u^r) \triangleq \{x \in S \mid h(x) - \langle \nabla g(u^r), x - u^r\rangle - g(u^r) \le 0\}.$$

Как и выше, можно показать, что $u^r \in D_r \subset D$, в частности $F(u^r) \ge 0$, $r = 0, 1, \ldots$. При этом $u^{r+1} \in D_r \cap D_{r+1}$ по построению, поэтому $u^{r+1} \in D = \{x \in S \mid F(x) \ge 0\}$.

Предложение 2. *Предположим, что значение задачи* (P) *конечно, а функции* $f(\cdot)$, $g(\cdot)$ *и* $h(\cdot)$ *выпуклы и выполнено условие* (3.3).

Пусть, кроме того, функция $g(\cdot)$ *непрерывно дифференцируема на открытой области* Ω, *содержащей множество* S.

(i) *Тогда последовательность* $\{u^r\}$, *генерируемая процедурой* 3, *удовлетворяет условию*

$$\lim_{r\to\infty} [\inf_x \{ f(x) - f(u^{r+1}) \mid x \in D_r\}] = 0. \tag{3.7}$$

(ii) *При этом если множество* $\mathcal{F} = \{x \in S \mid f(x) \le f(u^0) + \sum_{r=0}^{\infty} \delta_r\}$ *ограничено, то всякая предельная точка* $\hat{u}$ *последовательности* $\{u^r\}$ *удовлетворяет неравенству*

$$f(\hat{u}) \le f(x) \ \forall x \in D(\hat{u}). \tag{3.8}$$

(iii) *В случае замкнутости множества* S *предельная точка* $\hat{u}$ *является решением задачи*

$$f(x) \downarrow \min, \ \ x \in D(\hat{u}). \tag{PL($\hat{u}$)}$$

(iv) *В случае дифференцируемости* $f(\cdot)$ *в точке* $\hat{u}$ *эта предельная точка* $\hat{u}$ *оказывается нормально стационарной в задаче* (P).

Доказательство. (i) Из неравенства (3.6) с помощью включения $u^r \in D_r$ имеем

$$\begin{gathered} -\delta_r \le \inf_x \{ f(x) \mid x \in D_r\} - f(u^{r+1}) = \\ = \inf_x \{ f(x) - f(u^r) \mid x \in D_r\} + f(u^r) - f(u^{r+1}) \le f(u^r) - f(u^{r+1}). \end{gathered} \tag{3.9}$$

Отсюда получаем

$$f(u^{r+1}) - \delta_r \le f(u^r),$$

так что числовая последовательность $\{f(u^r)\}$ оказывается почти монотонно убывающей. Принимая во внимание ограниченность снизу функции $f(\cdot)$ на допустимом множестве задачи (P), по лемме 2 из [1, §2.6, с. 87] заключаем, что существует конечный предел $\hat{f}$ числовой последовательности $\{f(u^r)\}$:

$$\lim_{r\to\infty} f(u^r) = \hat{f}.$$

Переходя к пределу в соотношениях (3.9), получаем (3.7).

(ii) По построению (3.6), $u^{r+1} \in D_r \subset S$ и $f(u^{r+1}) \le f(u^r) + \delta_r \le f(u^0) + \sum_{k=0}^{r} \delta_k$, так что $\{u^r\} \subset \mathcal{F}$, а $\mathcal{F}$ ограничено. Поэтому с точностью до подпоследовательности можно считать, что $u^r \longrightarrow \hat{u}$ для некоторого $\hat{u} \in \mathbb{R}^n$.

Введем обозначение $V_r := \inf_x \{ f(x) \mid x \in D_r\}$. Тогда из (3.9) следует

$$\lim_{r\to\infty} V_r = \lim_{r\to\infty} f(u^r) = f(\hat{u}) = \hat{f}. \qquad (3.10)$$

Кроме того, поскольку $F(u^r) \ge 0$, то по непрерывности $F(\cdot) = g(\cdot) - h(\cdot)$ получаем $F(\hat{u}) = \lim_{r\to\infty} F(u^r) \ge 0$. Теперь положим

$$\hat{V} := \inf_x \{ f(x) \mid x \in D(\hat{u})\} \qquad (3.11)$$

и покажем, что

$$\hat{V} = \hat{f} = f(\hat{u}). \qquad (3.12)$$

С одной стороны, по (3.11) и согласно включению $\hat{u} \in D(\hat{u})$ справедливо неравенство $\hat{V} \le f(\hat{u})$. Для доказательства противоположного неравенства покажем прежде всего, что для любой точки $w \in D(\hat{u})$ и любого числа $\eta > 0$ существуют число $p = p(w, \eta) \in Z_+ \triangleq \{0, 1, \ldots\}$ и вектор $w^p \in D_p$ такие, что $\|w^p - w\| \le \eta$.

1) Если $w \in S$ таково, что $\Psi(w) < 0$, где

$$\Psi(x) \triangleq h(x) - \langle \nabla g(\hat{u}), x - \hat{u}\rangle - g(\hat{u}),$$

$$\Psi_r(x) \triangleq h(x) - \langle \nabla g(u^r), x - u^r\rangle - g(u^r),$$

то, поскольку $\Psi_r(w) \longrightarrow \Psi(w)$ $(r \longrightarrow +\infty)$, существует число $p = p(w) : \Psi_p(w) \le 0$. Поэтому полагаем $w^p := w$.

2) Пусть теперь $\Psi(w) = 0$, $w \in S$. Покажем, что непременно найдется элемент

$$\bar{w} \in S \ \text{ такой, что } \ \Psi(\bar{w}) < 0. \qquad (3.13)$$

Действительно, если $F(\hat{u}) > 0$, то имеем $\Psi(\hat{u}) = h(\hat{u}) - g(\hat{u}) = -F(\hat{u}) < 0$, так что $\bar{w} = \hat{u}$. Если же $F(\hat{u}) = 0$, то, по условию (3.3), при $y = \hat{u}$ получаем, что существует $d = d(\hat{u})$:

$$\Psi(d) \triangleq h(d) - \langle \nabla g(\hat{u}), d - \hat{u}\rangle - g(\hat{u}) < h(\hat{u}) - g(\hat{u}) = 0.$$

Итак, при $\bar{w} = d(\hat{u})$ условие (3.13) также выполнено.

Далее, в силу выпуклости $\Psi(\cdot)$ при $w_\lambda = \lambda\bar{w} + (1 - \lambda)w$, $\lambda \in]0, 1[$, имеем

$$\Psi(w_\lambda) \le \lambda\Psi(\bar{w}) + (1-\lambda)\Psi(w) = \lambda\Psi(\bar{w}) < 0.$$

При этом очевидно, что $w_\lambda \in S$, $w_\lambda \longrightarrow w$ $(\lambda \downarrow 0)$. Поэтому полагаем $w^p := w_\lambda$, где λ таково, что $\|w_\lambda - w\| \le \eta$, а для номера $p = p(w, \eta)$ справедливо неравенство $\Psi_p(w_\lambda) \le 0$.

3) Далее, из (3.11) следует, что $\forall \varepsilon > 0 \ \exists v_\varepsilon \in D(\hat{u})$, для которого имеем

$$f(v_\varepsilon) - \varepsilon/2 \le \hat{V}. \qquad (3.14)$$

По непрерывности $f(\cdot)$ найдется $\eta = \eta(\varepsilon) > 0$ такое, что

$$|f(z) - f(v_\varepsilon)| \le \varepsilon/2 \quad \forall z : \|z - v_\varepsilon\| \le \eta,$$

откуда, в частности, следует неравенство

$$f(v_\varepsilon) \ge f(z) - \varepsilon/2 \quad \forall z : \|z - v_\varepsilon\| \le \eta. \tag{3.15}$$

Согласно доказанному выше (см. 1) и 2)), для любого $\varepsilon > 0$ существуют $\eta = \eta(\varepsilon)$ и $p = p(\varepsilon) \in Z_+$, а также $w^p \in D_p$ такие, что $\|w^p - v_\varepsilon\| \le \eta$. А тогда с помощью (3.15) получаем $f(v_\varepsilon) \ge f(w^p) - \varepsilon/2$, откуда в силу (3.14) имеем $f(w^p) - \varepsilon \le \hat{V}$.

Поскольку $w^p \in D_p$, то из последнего неравенства вытекает неравенство

$$V_p - \varepsilon \le \hat{V}. \tag{3.16}$$

Теперь рассмотрим некоторую последовательность $\{\varepsilon_k\}$, $\varepsilon_k > 0$, $k \in Z_+$, $\varepsilon_k \downarrow 0$ $(k \longrightarrow \infty)$. По доказанному выше, $\forall k \in Z_+ \; \exists p = p(k) \in Z_+$ такое, что, согласно (3.16), верно неравенство

$$V_{p(k)} - \varepsilon_k \le \hat{V}.$$

Отсюда при $k \longrightarrow \infty$ с учетом (3.10) получаем $\hat{f} \le \hat{V}$, что и требовалось.

Итак, (3.12) доказано, а вместе с ним и неравенство (3.8).

(iii) Из (3.8) в случае замкнутости множества S следует, что $\hat{u} \in \mathrm{Sol}(\mathrm{PL}(\hat{u}))$.

(iv) Из (iii) вытекает, что найдутся множители λ_0, $\lambda \ge 0$, $\lambda_0 + \lambda > 0$, при которых справедливы условия

$$\begin{gathered} \langle \lambda_0 \nabla f(\hat{u}) - \lambda \nabla F(\hat{u}), x - \hat{u} \rangle \ge 0 \;\; \forall x \in S, \\ \lambda F(\hat{u}) \;=\; 0. \end{gathered} \tag{3.17}$$

Обратим внимание на то, что (3.17) является ККТ-системой оптимальности и для задачи (P). Пусть $\lambda_0 = 0$. Тогда $\lambda > 0$ и из системы (3.17) получаем $F(\hat{u}) = 0$,

$$\langle \nabla g(\hat{u}), x - \hat{u} \rangle - \langle \nabla h(\hat{u}), x - \hat{u} \rangle \;=\; \langle \nabla F(\hat{u}), x - \hat{u} \rangle \le 0 \;\; \forall x \in S.$$

Отсюда в силу выпуклости функции $h(\cdot)$ вытекает

$$F(\hat{u}) \;=\; 0, \quad h(x) - h(\hat{u}) \ge \langle \nabla g(\hat{u}), x - \hat{u} \rangle \;\; \forall x \in S,$$

что противоречит условию регулярности (3.3). Значит, $\lambda_0 > 0$.

Замечание 2. Если в (3.17) $\lambda = 0$, то $\langle \nabla f(\hat{u}), x - \hat{u} \rangle \ge 0$ или $f(\hat{u}) \le f(x)$ $\forall x \in S$. Последнее, очевидно, противоречит условию (H_0), так как $F(\hat{u}) \ge 0$. Поэтому $\lambda > 0$, $F(\hat{u}) = 0$.

Теперь рассмотрим вопрос о том, как с помощью последовательности $\{u^r\}$, где $u_0 = \tilde{x}$, определить точку $y(\tilde{x})$. Для этого зададим число $\varepsilon > 0$. Тогда если $\delta_r \le \varepsilon$ и справедливо неравенство

$$f(u^r) - f(u^{r+1}) \le \varepsilon, \tag{3.18}$$

то полагаем $y = y(\tilde{x}, \varepsilon) = u^{r+1}$, поскольку в этом случае из (3.9) и (3.18) вытекает $y \in D_r \subset D$,

$$f(y) - \varepsilon \;=\; f(u^{r+1}) - \varepsilon \le \inf_x \{ f(x) \mid x \in D_r \},$$

так что $y = u^{r+1}$ является ε-решением линеаризованной в точке y задачи (PL_r).

Объединим процедуры 1 и 3 в один метод и запишем его в более алгоритмической форме. Пусть заданы точка $x_0 \in S$, $F(x_0) \ge 0$, и числовые последовательности $\{\delta_r\}$ и $\{\varepsilon_s\}$, удовлетворяющие условиям (2.6) и (2.14) соответственно.

Модифицированный метод Розена

Шаг 0 (инициализация). Положить $s := 0$, $x^s := x_0$.

Шаг 1 (процедура 3). Начиная с точки x^s, построить точку $y^s = y(x^s, \varepsilon_s)$, являющуюся ε_s-решением линеаризованной задачи (PL_s) – (3.5), т.е.

$$f(x) + \varepsilon_s \ge f(y^s) \;\; \forall x \in S : h(x) - \langle \nabla g(y^s), x - y^s \rangle - g(y^s) \le 0.$$

Шаг 2 (критерий останова). Если $F(y^s) \le 0$, то Stop.

Шаг 3 (процедура 1). С помощью y^s построить точку $u := x(y^s)$:

$$u \in S, \quad F(u) = 0, \quad f(u) < f(y^s) \le f(x^s).$$

Шаг 4. Положить $x^{s+1} := u$, $s := s + 1$ и вернуться на шаг 1.

Лемма 2. *Модифицированный метод Розена обладает следующими свойствами:*

(а) *либо конечен, причем* $F(y^N) = 0$, *где* N *– итерация останова;*

(б) *либо генерирует две последовательности* $\{x^s\}$ *и* $\{y^s\}$ *со свойствами*

$$x^s \in S, \quad F(x^s) = 0, \quad y^s \in S, \quad F(y^s) > 0, \tag{3.19}$$

$$f(x^{s+1}) < f(y^s) \le f(x^s). \tag{3.20}$$

При этом справедливо предельное соотношение

$$\lim_{s \to \infty} f(x^s) = \lim_{s \to \infty} f(y^s). \tag{3.21}$$

Доказательство. (а) Согласно критерию останова (шаг 2), если N – итерация остановки счета, то $F(y^N) \le 0$. А поскольку каждая точка u^{r+1}, $r = 0, 1, \ldots$, допустима в линеаризованной задаче $(\mathrm{PL}(u^r))$, то и $F(y^s) \ge 0$, $s = 0, 1, \ldots$. Поэтому $F(y^N) = 0$.

(б) Если процесс счета бесконечен, то, согласно описанию метода, благодаря выпуклости $f(\cdot)$ и способу построения точки $x^{s+1} = \lambda_s v + (1 - \lambda_s) y^s$, в силу (H_0) получаем (3.20). Поскольку $f(\cdot)$ ограничена снизу в задаче (P), то существует предел $\lim\limits_{s \to \infty} f(x^s) > f(v)$, откуда с помощью (3.20) получаем (3.21).

Теорема 2. *Пусть функция* $f(\cdot)$ *и множество* S *выпуклы, множество* $\mathcal{F}_0 = \{x \in S \mid f(x) \le f(x_0) + \sum_{r=0}^{\infty} \delta_r\}$ *ограничено и справедливы условия* (H_0) *и* (3.3).

Тогда модифицированный метод Розена обладает следующими свойствами:

а) *при конечном числе итераций получает точку* $y^N \in S$, $F(y^N) = 0$, *которая является* ε_N-*решением линеаризованной задачи* (PL_N) – (3.5), *где* N *– номер итерации останова;*

б) *при бесконечном числе итераций для последовательностей* $\{y^s\}$ *и* $\{x^s\}$ *в дополнение к свойствам* (3.19)–(3.21) *справедливы также условия*

$$0 = F(x^s) = \lim_{s \to \infty} F(y^s), \tag{3.22}$$

$$x_* = \lim_{s \to \infty} x^s = \lim_{s \to \infty} y^s \tag{3.23}$$

для некоторого $x_* \in \mathbb{R}^n$, $F(x_*) = 0$.

Кроме того, точка x_* *является решением линеаризованной задачи*

$$(\mathrm{PL}_*): \quad f(x) \downarrow \min, \quad x \in S, \quad g(x_*) - h(x) + \langle \nabla g(x_*), x - x_* \rangle \ge 0, \tag{3.24}$$

и нормально стационарной в задаче (P).

Доказательство. Утверждение а) вытекает из леммы 2 и (3.19). Для б) из (3.20), (3.21) и выпуклости функции $f(\cdot)$ следует

$$f(x^{s+1}) \le \lambda_s [f(v) - f(y^s)] + f(y^s),$$

или

$$0 \le \lambda_s [f_* - f(v)] \le \lambda_s [f(y^s) - f(v)] \le f(y^s) - f(x^{s+1}),$$

где $f_* \triangleq V(P)$. Поэтому из (3.21), поскольку $f(y^s) \ge f_* > f(v)$, получаем

$$\lim_{s \to \infty} \lambda_s = 0, \quad \lambda_s \in \,]0, 1[. \tag{3.25}$$

По построению, $\{x^s\} \subset \mathcal{F}_0$, а $\mathcal{F}_0$ ограничено. Поэтому с точностью до подпоследовательности можно считать, что $x^s \longrightarrow x_*$.

Далее, поскольку $x^{s+1} = y^s + \lambda_s(v - y^s)$, то

$$\|y^s - x_*\| = \|x^{s+1} + \lambda_s(y^s - v) - x_*\| \le \|x^{s+1} - x_*\| + \lambda_s\|y^s - v\| \le \|x^{s+1} - x_*\| + \lambda_s K$$

для некоторого $K > 0$ такого, что $\|y^s - v\| \le K$. Эта оценка справедлива, поскольку $y^s \in \mathcal{F}_0$, $s = 0, 1, \ldots$. Поэтому в силу равенства $F(x^s) = 0$, $s = 0, 1, \ldots$, и непрерывности $F(\cdot)$ имеем $0 = F(x^s) = \lim_{s \to \infty} F(y^s) = 0 = F(x_*)$, тем самым (3.22) и (3.23) доказаны.

Наконец, по построению (см. (3.20)), для каждого $y^s = y(x^s, \varepsilon_s)$ согласно (3.9) и (3.18) справедливо неравенство $\inf_x\{f(x) - f(y^s) \mid x \in D_s\} \le \varepsilon_s$, откуда, аналогично п. (ii) доказательства предложения 2, при $s \longrightarrow \infty$ получаем

$$f(x) \ge f(x_*) \;\; \forall x \in S : \langle \nabla g(x_*), x - x_* \rangle + g(x_*) - h(x) \ge 0,$$

т.е. точка x_* является решением линеаризованной задачи (3.24).

Нормальная стационарность точки x_* доказывается так же, как в предложении 2.

4. ТЕСТИРОВАНИЕ МЕТОДОВ ЛОКАЛЬНОГО ПОИСКА

Сравнительное тестирование предложенных методов локального поиска – специального метода (СМЛП) и модифицированного метода Розена (ММР) – было проведено на двух сериях задач.

Первая состояла из квадратичных задач вида

$$\begin{aligned} f(x) &\triangleq \|x - p\|^2 \downarrow \min, \quad x \in S \triangleq \Pi = [-1; 5]^n, \\ F(x) &\triangleq \langle x, Cx \rangle - \langle x, Dx \rangle - \gamma \ge 0, \end{aligned} \tag{P$_1$}$$

где C и D симметричные ($C = C^{\text{т}}$, $D = D^{\text{т}}$), положительно-определенные ($C > 0$, $D > 0$) матрицы $(n \times n)$; вектор $p \in \mathbb{R}^n$ и число $\gamma \in \mathbb{R}$ согласованы с размерами параллелепипеда Π так, чтобы глобальный минимум функции $f(\cdot)$ на $\mathbb{R}^n$ лежал в Π, но был недопустим по ограничению $F(x) \ge 0$ ($v = p = \operatorname{argmin}(f, \mathbb{R}^n)$).

Отметим, что для построения задач (P_1) случайным образом генерировалась знаконеопределенная матрица $A = C - D$, а затем известным способом (см. [6, с. 270]) строилось d.c.-представление квадратичной функции.

Поскольку линеаризованные задачи ($LQ(u, \beta)$) и ($PL(u)$) в СМЛП и ММР имеют достаточно простую структуру, то их решение проводилось в два этапа. Вначале приближенное решение отыскивалось комбинацией аналитического подхода и метода последовательных проекций на параллелепипед Π. При этом учитывалось, что точка $x_c = D^{-1}Cu$ является точкой абсолютного минимума выпуклой функции $\psi(x, u) \triangleq \langle x, Dx \rangle - 2\langle Cu, x \rangle + \langle u, x \rangle$. Затем найденная точка выбиралась в качестве стартовой для квазиньютоновского метода из [1], которым производился поиск минимума модифицированной функции Лагранжа (см. [1]) задач ($LQ(u, \beta)$) и ($PL(u)$).

Далее, согласно теореме 1 о сходимости СМЛП, в качестве критерия останова можно выбирать одновременное выполнение неравенств (2.26), однако вычислительные эксперименты показали, что второе неравенство $\beta_{s-1} - \beta_s \le \rho$ является следствием первого неравенства $F(y^s) \le \rho$. В то же время, согласно теореме 2 о сходимости ММР, за критерий останова выбирается неравенство $F(y^s) \le \rho$. Поэтому при проведении численного тестирования в обоих методах останов производился при выполнении $F(y^s) \le \rho$, где $\rho = 10^{-4}$.

В табл. 1 и 2 приведены результаты численного тестирования методов локального поиска на серии задач (P_1), где в качестве начальных приближений были выбраны три различные точки (x^1, x^2, x^3), причем x^1 выбиралась самой далекой по значению целевой функции (для каждой тестовой

Таблица 1

n	f_0	СМЛП			ММР		
		f_*	PL	Time	f_*	PL	Time
	144,75	11,61	31	0.00	11,87	9	0.00
10	43,95	11,31	57	0.00	11,77	42	0.00
	46,94	11,61	49	0.00	11,77	45	0.00
	300,01	23,73	32	0.00	23,75	36	0.00
20	78,77	23,72	30	0.00	23,75	32	0.00
	114,16	23,73	39	0.00	23,16	15	0.00
	474,79	42,82	87	0.00	44,16	74	0.00
30	113,94	42,84	54	0.00	42,81	51	0.00
	150,42	42,84	45	0.00	42,81	40	0.00
	561,93	61,06	37	0.00	59,64	56	0.00
40	99,89	61,07	37	0.00	59,66	40	0.00
	216,75	61,49	45	0.00	61,27	43	0.00
	714,48	89,81	62	0.00	88,67	20	0.00
50	128,62	89,82	86	0.00	88,79	66	0.00
	283,51	89,74	116	0.00	88,49	68	0.00
	905,31	112,15	74	0.00	109,42	97	0.02
60	245,05	111,76	58	0.00	109,07	55	0.04
	332,84	112,11	57	0.00	109,42	96	0.02
	1155,38	137,27	70	0.00	133,34	58	0.01
70	241,41	137,26	79	0.00	133,76	69	0.01
	336,04	137,38	79	0.00	133,41	84	0.01
	1334,67	149,83	43	0.00	148,62	73	0.02
80	319,59	149,99	100	0.01	148,55	80	0.02
	372,81	149,85	133	0.03	148,51	99	0.03
	1378,19	221,87	206	0.05	215,96	301	0.13
90	410,57	221,83	64	0.02	213,35	61	0.03
	429,39	222,01	182	0.03	214,14	111	0.05

задачи – это первая строка в таблице). Чтобы найти такую точку, решалась релаксированная задача с противоположной целью (на максимум):

$$f(x) \uparrow \max, \quad x \in S,$$

а затем полученная точка проверялась на допустимость по ограничению $F(x) \geq 0$.

В таблицах использованы следующие обозначения: n – размерность тестовой задачи, f_0 – значение целевой функции в стартовой точке. Далее для каждого метода приведены f_* – значение целевой функции в полученной критической точке; PL – количество решенных линеаризованных задач; Time – время решения в секундах.

Результаты численного эксперимента показывают, что методы во всех тестовых задачах вида (P_1) нашли разные критические точки. По качеству счета лучше себя зарекомендовал ММР, который во всех задачах нашел лучшую по значению целевой функции точку, а по времени счета в основном преимущество сохраняется за СМЛП (максимальное время не превысило 1 мин 15 с). Причем в задаче при $n = 600$ (для x_2) ММР работал в 11 раз медленнее, а в задаче при $n = 300$ (для

Таблица 2

n	f_0	СМЛП			ММР		
		f_*	PL	Time	f_*	PL	Time
100	1714,94	240,39	57	0.00	236,75	40	0.03
	339,35	233,69	102	0.02	232,01	110	0.06
	517,78	233,67	104	0.02	231,17	58	0.03
200	2997,07	680,09	253	0.22	654,93	377	0.73
	1059,33	679,97	416	0.36	654,93	390	0.77
	1255,27	680,02	864	0.73	654,93	634	1.14
300	4414,71	1259,24	618	1.83	1199,15	278	1.78
	1980,58	1259,41	460	1.36	1199,21	231	1.53
	1888,77	1256,31	192	0.58	1199,14	1468	8.34
400	5549,62	1962,91	90	0.47	1850,72	336	4.64
	2925,71	1963,65	432	2.25	1850,72	688	8.52
	3216,72	1963,98	462	2.42	1850,72	519	6.66
500	6762,64	2853,51	551	4.49	2576,89	238	6.03
	3920,82	2857,28	1228	10.05	2576,72	1192	22.36
	4067,11	2856,78	2154	17.63	2576,72	1223	22.94
600	8153,71	3886,08	495	5.81	3495,74	171	7.64
	5333,28	3902,14	357	4.17	3491,73	1784	47.16
	5786,36	3828,41	739	8.67	3516,54	543	16.77
700	9448,22	4904,27	371	5.91	4346,63	311	15.80
	6605,56	4901,76	2130	33.83	4346,48	823	32.80
	6687,46	4929,57	364	5.80	4346,46	655	28.13
800	10758,15	6130,58	300	6.22	5349,19	663	36.94
	8489,67	6069,38	1839	38.13	5344,53	456	27.92
	8246,92	6068,98	888	18.41	5346,07	1473	72.52
900	11895,77	7397,63	2854	74.92	6340,01	606	44.92
	9967,62	7481,64	974	25.53	6341,24	2047	124.17
	10110,39	7372,61	1048	27.49	6341,14	2206	134.42
1000	13154,19	8889,71	780	25.22	7503,44	665	61.08
	11439,69	8989,36	1804	58.33	7503,85	1733	133.38
	11638,25	8970,38	1512	49.02	7503,88	2190	167.42
1100	14455,83	10272,76	462	18.19	8521,44	415	55.25
	12837,48	10409,45	1556	61.02	8521,68	769	84.36
	13048,76	10458,61	792	31.02	8538,52	715	79.86
1200	15877,89	11860,39	909	42.34	9804,54	1141	139.14
	14448,15	11791,89	1234	57.45	9804,66	972	124.67
	14616,48	12314,38	823	38.28	9804,73	1298	154.49

x_3) – в 14 раз медленнее СМЛП. В то же время есть примеры, когда ММР и работал быстрее, и нашел лучшую точку (см. $n = 700$, 800 (для x_2), $n = 900$ (для x_1)).

Итак, СМЛП работает, грубо говоря, быстрее, но ММР находит лучшую критическую точку, поэтому можно предложить использовать при решении задач с квадратичной целевой функцией

Таблица 3

Name	n	f_0	СМЛП			ММР		
			f_*	PL	Time	f_*	PL	Time
c-fat200-1	200	0	67,087	31	0.03	67,051	43	0.03
		2,546	66,259	36	0.03	66,510	29	0.02
		3,472	66,259	62	0.06	68,249	29	0.02
c-fat200-2	200	0	2,435	3	0.00	2,435	3	0.00
		2,426	66,372	40	0.05	66,372	20	0.02
		1,755	66,372	38	0.04	66,372	21	0.02
c-fat200-5	200	0	1,523	2	0.00	1,523	2	0.00
		1,228	11,848	319	0.28	50,173	36	0.03
		1,083	11,694	502	0.45	48,737	40	0.05
san200_0.7_1	200	0	12,947	196	0.17	37,405	74	0.08
		1,048	15,232	276	0.27	36,100	55	0.05
		1,086	13,492	298	0.28	36,411	61	0.06
san200_0.7_2	200	0	18,732	351	0.31	1,210	5	0.00
		1,051	20,245	168	0.16	29,565	49	0.05
		1,082	19,253	256	0.25	33,641	56	0.05
san200_0.9_1	200	0	8,171	430	0.39	31,429	93	0.09
		1,041	7,275	325	0.30	28,705	77	0.08
		0,771	6,729	343	0.31	26,902	81	0.08
san200_0.9_2	200	0	7,733	317	0.30	28,458	95	0.09
		1,027	8,561	437	0.39	25,986	80	0.08
		0,775	8,719	321	0.30	29,291	75	0.06
p_hat300-1	300	0	73,085	78	0.23	116,932	21	0.06
		1,606	73,229	85	0.27	117,437	23	0.08
		1,763	72,464	87	0.27	117,183	31	0.09
p_hat300-2	300	0	64,281	132	0.39	112,262	40	0.13
		1,198	65,524	100	0.30	111,750	32	0.09
		1,213	63,583	120	0.36	111,775	28	0.08
p_hat300-3	300	0	21,849	256	0.76	57,796	46	0.14
		0,992	19,485	295	0.88	58,151	44	0.14
		1,158	21,359	200	0.59	58,447	42	0.14
san400_0.5_1	400	0	58,157	65	0.36	96,940	66	0.34
		1,345	57,611	52	0.28	106,408	51	0.27
		1,201	57,291	113	0.59	112,228	97	0.52
san400_0.7_1	400	0	18,393	443	2.33	62,860	109	0.58
		1,13	18,634	325	1.70	57,170	80	0.42
		1,177	17,993	372	1.95	60,826	88	0.47
san400_0.9_1	400	0	8,578	907	4.76	45,897	150	0.78
		0,931	9,133	684	3.59	45,117	113	0.59
		0,929	7,885	601	3.16	44,459	121	0.64
sanr400_0.5	400	0	28,703	232	1.20	80,849	65	0.34
		1,326	26,524	482	2.53	80,099	56	0.30
		1,191	27,882	288	1.52	82,197	61	0.33
sanr400_0.7	400	0	20,326	366	1.92	60,361	96	0.50
		1,084	19,494	329	1.73	59,979	73	0.38
		1,1667	16,342	391	2.05	58,196	77	0.41

Таблица 4

Name	n	f_0	СМЛП			ММР		
			f_*	PL	Time	f_*	PL	Time
		0	1,631	1	0.00	1,631	1	0.00
c-fat500-10	500	1,541	184,248	63	0.53	184,248	23	0.19
		1,469	184,248	65	0.53	184,248	25	0.22
		0	172,463	119	0.98	176,058	41	0.34
c-fat500-2	500	3,233	189,088	44	0.38	189,088	21	0.17
		2,764	189,088	44	0.36	189,088	24	0.20
		0	2,304	2	0.02	2,304	2	0.02
c-fat500-5	500	1,925	185,473	52	0.44	185,473	22	0.19
		1,939	185,473	55	0.47	185,473	25	0.20
		0	115,631	112	0.91	184,601	35	0.30
p_hat500-1	500	1,808	113,164	100	0.81	184,159	28	0.24
		1,667	113,749	118	0.97	183,352	30	0.25
		0	104,777	104	0.86	179,628	29	0.23
p_hat500-2	500	1,198	105,821	132	1.08	179,628	28	0.23
		1,272	106,517	179	1.47	179,795	28	0.23
		0	23,289	355	2.91	85,419	55	0.45
p_hat500-3	500	0,998	20,431	396	3.23	85,435	49	0.41
		1,001	22,374	419	3.44	84,543	49	0.41
		0	158,368	143	2.30	259,655	30	0.48
p_hat700-1	700	1,874	157,008	165	2.64	259,666	30	0.48
		1,758	158,355	128	2.05	259,933	31	0.50
		0	151,693	119	1.91	252,518	34	0.55
p_hat700-2	700	1,475	152,269	122	1.94	253,038	30	0.48
		1,363	152,914	149	2.38	252,101	37	0.58
		0	26,554	624	9.92	118,456	61	0.98
p_hat700-3	700	1,079	25,124	410	6.52	118,924	49	0.78
		1,151	27,238	583	9.30	117,649	52	0.84
		0	23,865	962	19.94	74,711	180	3.74
brock800_1	800	1,221	23,333	1182	24.55	73,594	135	2.81
		1,144	21,489	1051	21.84	73,178	164	3.41
		0	24,087	1196	24.84	77,866	175	3.64
brock800_2	800	1,855	21,559	971	20.13	74,208	143	2.97
		1,131	20,707	1201	24.95	76,679	145	3.02
		0	25,621	1204	24.98	76,171	165	3.42
brock800_3	800	1,191	20,466	976	20.23	74,514	141	2.94
		1,312	23,899	1007	20.91	73,771	144	3.00
		0	22,096	1073	22.28	74,113	164	3.41
brock800_4	800	1,229	19,337	1036	21.52	72,489	136	2.84
		. 1,151	23,573	1037	21.52	75,489	148	3.09

и квадратичным d.c.-ограничением в качестве метода локального поиска некую комбинацию этих двух методов.

Переходя к описанию второй части вычислительного эксперимента, заметим, что частным случаем задачи (Q_β), ассоциированной с задачей (P), является максимизация знаконеопределенной квадратичной формы на каноническом симплексе (см. [6, гл. 7]). Эта экстремальная задача описывает известную дискретную задачу нахождения максимальной клики в неориентированном графе в непрерывной постановке Моцкина–Штрауса (см. [15]). Принимая во внимание этот факт, знаконеопределенные матрицы A второй серии тестовых задач

$$\begin{aligned} f(x) &\triangleq \langle e, x\rangle \uparrow \max, \quad x \in S \triangleq \Pi = [0; 1]^n, \\ F(x) &\triangleq \langle x, Ax\rangle - \gamma_* \le 0 \end{aligned} \qquad (P_2)$$

брали из библиотеки DIMACS (Discrete Mathematics and Computer Science). Они расположены в Интернете на сайте ftp://dimacs.rutgers.edu/pub/challenge/graph/benchmarks. Здесь A – матрица смежности графа $(n \times n)$, $e = (1, \ldots, 1) \in \mathbb{R}^n$, число γ_* связано с размерностью максимальной клики K_*, найденной в графе: $\gamma_* = 1 - 1/K_*$.

Было решено 45 задач размерности от 200 до 800. В табл. 3 и 4 приведены некоторые наиболее интересные результаты численного тестирования методов локального поиска на этой тестовой серии. К обозначениям, введенным в табл. 1 и 2, добавлено name – название тестового примера.

Аналогично предыдущей серии тестовых задач, выбирались три различные стартовые точки (x^1, x^2, x^3) и в качестве критерия останова использовалось неравенство $F(y^s) \le \rho$, где $\rho = 10^{-4}$.

Результаты численного тестирования, приведенные в табл. 3 и 4, показывают, что есть задачи, в которых оба метода локального поиска сработали примерно одинаково и по времени, и по качеству решения (см. c-fat200-1, c-fat200-2). В задачах c-fat размерности $n = 500$ значения f_* целевой функции в полученных алгоритмами критических точках оказались равными (за исключением задачи c-fat500-2 для x^1), однако ММР нашел их быстрее.

В целом преимущество и по качеству решения, и по времени счета в этой серии тестов сохраняется за ММР. Например, для задач san_200_0.9_1 и san_400_0.9_1 решение, полученное ММР, в 4–6 раз “лучше”, чем полученное СМЛП, а для задач brock размерности $n = 800$ ММР работал в 7–8 раз быстрее СМЛП (максимальное время не превысило 3 с). Однако в одной задаче (san_200_0.7_2 для x^1) СМЛП нашел лучшую по значению целевой функции точку ($f_* = 18{,}732$, в то время как ММР – точку $f_* = 1.21$), но для этого потребовалось решить 351 линеаризованную задачу (LQ(u, β)). Итак, тестированием выявлено некоторое преимущество ММР над СМЛП при решении задач с линейной целевой функцией и квадратичным d.c.-ограничением, что уже наблюдалось при тестировании методов локального поиска, разработанных для обратно-выпуклой задачи [6], [10], [11] – частного случая задачи с d.c.-ограничением.

В заключение отметим, что, несмотря на трудность задач с d.c.-ограничениями, удалось разработать два варианта методов локального поиска, один из которых (СМЛП) связан с условиями глобальной оптимальности [6], [10], а второй (ММР) является развитием идеи Розена (см. [13]). Проведено сравнительное тестирование предложенных методов на задачах с d.c.-ограничением, целевая функция в которых задавалась как квадратичной формой, так и линейной.

Интересным, на наш взгляд, оказался тот факт, что эти методы ведут себя в тестовых задачах по-разному: СМЛП “лучше показал себя” на задачах (P_1) с квадратичной целевой функцией, а ММР – на задачах (P_2), целевая функция которых линейна. Поэтому в дальнейшем при построении алгоритма глобального поиска предполагается использовать комбинацию представленных здесь локальных методов.

СПИСОК ЛИТЕРАТУРЫ

1. *Васильев Ф.П.* Методы оптимизации. М.: Факториал Пресс, 2002.
2. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983.
3. *Еремин И.И.* Теория линейной оптимизации. Екатеринбург: Изд-во Ин-та матем. и механ. УрО РАН, 1998.
4. Handbook of global optimization / Eds by R. Horst, P. Pardalos. Dordrecht: Kluwer Acad. Publs., 1995.
5. *Horst R., Tuy H.* Global optimization (deterministic approaches). Berlin: Springer, 1993.
6. *Стрекаловский А.С.* Элементы невыпуклой оптимизации. Новосибирск: Наука, 2003.

7. *Стрекаловский А.С.* Об экстремальных задачах с d.c.-ограничениями // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. № 12. С. 1833–1843.
8. *Стрекаловский А.С.* О минимизации разности двух выпуклых функций на допустимом множестве // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 3. С. 399–409.
9. *Стрекаловский А.С.* Минимизирующие последовательности в задачах с d.c.-ограничениями // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. № 3. С. 435–447.
10. *Стрекаловский А.С., Яковлева Т.В.* О локальном и глобальном поиске в невыпуклых задачах оптимизации // Автоматика и телемехан. 2004. № 3. С. 23–34.
11. *Стрекаловский А.С., Яковлева Т.В.* Модификация метода Розена в обратно-выпуклой задаче // Изв. вузов. Матем. 2005. № 12. С. 70–75.
12. *Заботин В.И., Черняев Ю.А.* Обобщение метода проекции градиента на экстремальные задачи с предвыпуклыми ограничениями // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. № 3. С. 367–373.
13. *Rosen J.B.* Iterative solution of nonlinear optimal control problems // J. SIAM Control. 1966. V. 4. № 1. P. 223–244.
14. *Meyer R.R.* The validity of a family of optimization methods // J. SIAM Control. 1970. V. 8. № 1. P. 41–54.
15. *Motzkin T.S., Straus E.G.* Maxima for graphs and a new proof of a theorem of Turán // Canad. J. Math. 1965. V. 17. P. 533–540.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 414–423

УДК 519.624.2+517.589

ВЫЧИСЛЕНИЕ РЕШЕНИЙ УРАВНЕНИЯ МАТЬЕ И СВЯЗАННЫХ С НИМИ ВЕЛИЧИН[1)]

© 2007 г. А. А. Абрамов, С. В. Курочкин

(119991 Москва, ул. Вавилова, 40, ВЦ РАН)

e-mail: alalabr@ccas.ru; kuroch@ccas.ru

Поступила в редакцию 03.10.2006 г.

Для уравнения Матье рассматриваются следующие вопросы: нахождение собственных значений с нужным номером (с использованием осцилляционных теорем для возникающих разностных уравнений); устойчивость решений разностных уравнений; корректное определение и вычисление собственных значений и функций Матье с нецелым номером; корректное определение и вычисление характеристического показателя Матье; вычисление значений решений уравнения Матье для больших значений аргумента. Для численного решения указанных проблем предложены вычислительные алгоритмы. Библ. 14. Фиг. 5. Табл. 1.

Ключевые слова: функции Матье, осцилляционные теоремы, рекуррентные формулы, численная устойчивость.

1. ВВЕДЕНИЕ

Уравнение Матье

$$y'' + (p - 2q\cos 2t)y = 0, \tag{1.1}$$

где $-\infty < t < \infty$, p, q – вещественные числа, играет важную роль во многих разделах математической физики и в течение долгого времени является объектом систематического исследования. Основные факты, касающиеся уравнения (1.1), представлены в классических руководствах [1], [2]. Современные работы сосредоточиваются на конкретных вопросах: получение асимптотических представлений для собственных значений (СЗ) при малых и больших значениях q (см. [3], [4]), аппроксимация, нормировка и автомодельное поведение для СЗ (см. [5]), а также на различных вычислительных аспектах (см. [6]–[8]), включая вычисление СЗ в случае невещественного q (см. [9]). В данной работе рассматривается ряд вопросов как аналитического, так и вычислительного характера, до настоящего времени остававшихся не полностью исследованными.

Содержание статьи следующее. В разд. 2 дается корректное определение СЗ и функций Матье с произвольным (неотрицательным) вещественным номером. Предложен метод вычисления СЗ и собственных функций (СФ) с заданным номером. Отыскание хорошего приближения к СЗ для последующего применения методов типа Ньютона является трудной самостоятельной задачей (это отмечено, в частности, в [5]). Предложенный в настоящей статье метод не требует предварительного поиска приближенного значения. Метод основан на одной осцилляционной теореме для возникающих разностных уравнений, которая, как и соответствующая возможность непосредственного нахождения СЗ и СФ, является новой уже для случая классической периодической задачи Матье. В разд. 3 исследуется аналитическая зависимость характеристического показателя задачи (1.1) от параметра p, в результате чего получается другая (эквивалентная предыдущему определению) характеризация СЗ с произвольным вещественным номером, а также метод его вычисления. Далее, проведенный анализ приводит к корректному определению характеристического показателя Матье, основанному на конструкции аналитического продолжения. При этом устраняется (якобы имеющаяся) его неоднозначность с точностью до слагаемого вида $2k$, где k – произвольное целое. В этом вопросе в литературе распространены ошибочные представления, которые также воспроизводятся в некоторых известных современных математических пакетах. В разд. 4 предложен метод устойчивого вычисления значений решения уравнения (1.1) для больших значений аргумента t. В разд. 5 кратко описана численная реализация предложенных методов и приведены примеры расчетов.

[1)] Работа выполнена при финансовой поддержке РФФИ (код проекта 05-01-00257).

2. СОБСТВЕННЫЕ ЗНАЧЕНИЯ И ФУНКЦИИ МАТЬЕ С ПРОИЗВОЛЬНЫМ (НЕОТРИЦАТЕЛЬНЫМ) ВЕЩЕСТВЕННЫМ НОМЕРОМ

В классической постановке задачи на СЗ для уравнения (1.1), где p – спектральный параметр, ставятся условия периодичности или антипериодичности на промежутке $[0, \pi]$. Соответствующие СЗ образуют возрастающие последовательности a_n, $n = 0, 1, \ldots$, b_n, $n = 1, 2, \ldots$, соответственно, для четных и нечетных решений уравнения (1.1). Получающиеся при этом СФ – четные ce_{2m} (периода π), ce_{2m+1} (периода 2π) и нечетные se_{2m} (периода π), se_{2m+1} (периода 2π) – имеют по m нулей на интервале $(0, \pi/2)$. Обобщение этих понятий на нецелые рациональные номера СЗ (см. [1]) достигается с использованием кратных периодов, при этом СЗ оказываются двукратными ($a_\mu = b_\mu$ при $\mu \in \mathbb{Q}\backslash\mathbb{Z}$; далее для таких СЗ будем использовать обозначение p_μ). Для обобщения на случай произвольных (иррациональных) μ далее предложена конструкция, существенно уточняющая известное определение, основанное на теореме Флоке (см., например, [10]).

Этот вопрос может быть поставлен также следующим образом. При $q = 0$, $0 < p = \mu^2$ функции $y_1(t) = \cos\mu t$ и $y_2(t) = (\sin\mu t)/\mu$ являются решениями уравнения Матье. Как продолжить эти решения на значения $q \neq 0$? Отметим, что при целом положительном μ в качестве таких функций возникают функции Матье $ce_\mu(t)$ и $se_\mu(t)$. Важно, что в этой задаче при $q \neq 0$ число p уже не является заданным, здесь возникает задача на СЗ.

В данном разделе с целью технических упрощений сделаем в (1.1) замену $q \longrightarrow -q$, что приводит к следующей форме уравнения Матье:

$$y'' + (p + 2q\cos 2t)y = 0. \tag{2.1}$$

Достаточно рассмотреть случай $q \geq 0$. Случай $q < 0$ приводится к рассматриваемому заменой $t \longrightarrow t + \pi/2$.

Пусть $0 \leq p = \mu^2$ и $q = 0$ в (2.1). Уравнение (2.1) имеет решение вида

$$y(t) = e^{i\mu t}. \tag{2.2}$$

При фиксированном μ продолжением по q, $q \neq 0$, решения вида (2.2) естественно считать функцию, полученную следующим образом (см., например, [11]). Надо найти такое $p = p(\mu)$, при котором уравнение (2.1) имеет решение Флоке вида $F(t) = e^{i\mu t}P(t)$, где $P(t)$ является π-периодичной, $P(t) \not\equiv 0$.

Исследуем полученную задачу. Представив $P(t) = \Sigma_{k=-\infty}^{+\infty} c_k e^{i2kt}$, получим

$$(\mu + 2k)^2 c_k - q(c_{k-1} + c_{k+1}) = pc_k, \quad k = 0, \pm 1, \pm 2, \ldots. \tag{2.3}$$

Из (2.3) видно, что временно можно положить $0 \leq \mu < 2$, так как изменение μ на четное целое число может быть компенсировано сдвигом в последовательности индексов и не меняет свойств возникшей задачи. Однако (ср. [10]) имеется единственный естественный выбор этого слагаемого. Далее однозначность будет восстановлена.

Задача (2.3) – спектральная задача в гильбертовом пространстве двусторонних последовательностей $\{c_k\}$ таких, что $\Sigma_{k=-\infty}^{+\infty} |c_k|^2 < \infty$, со скалярным произведением $(\{c_k\}, \{d_k\}) = \Sigma_{k=-\infty}^{+\infty} c_k \bar{d}_k$. Исследуем некоторые ее свойства. Возьмем какое-либо вещественное Q большее, чем $2q$. Приведем (2.3) к виду

$$((\mu + 2k)^2 + Q)c_k - q(c_{k-1} + c_{k+1}) = \lambda c_k, \quad k = 0, \pm 1, \pm 2, \ldots, \quad \lambda = p + Q. \tag{2.4}$$

Оператор в левой части имеет вид $L + S$, где L эрмитов и положительно определен, L^{-1} вполне непрерывен, $|L^{-1}| \leq 1/Q$, S эрмитов, $|S| = 2q$. Тогда $(L + S)^{-1} = L^{-1/2}(I + L^{-1/2}SL^{-1/2})^{-1}L^{-1/2}$, $|L^{1/2}SL^{-1/2}| \leq$ $\leq \frac{1}{\sqrt{Q}} 2q \frac{1}{\sqrt{Q}} < 1$. Поэтому $(I + L^{-1/2}SL^{-1/2})^{-1}$ существует и ограничен. Следовательно, $(L + S)^{-1}$ существует и вполне непрерывен, а спектр $L + S$ дискретен и собственные элементы оператора $L + S$ образуют базис. Так как все нужные p суть СЗ эрмитова оператора, то они вещественны; следовательно, можно ограничиться вещественными $\{c_k\}$ (если $\{c_k\}$ – решение уравнения (2.3), то $\{c_k + \bar{c}_k\}$ и $i\{c_k - \bar{c}_k\}$ – решение уравнения (2.3)). Далее все величины считаются вещественными.

Из (2.3) имеем

$$\sum_{k=-\infty}^{\infty}(\mu+2k)^2c_k^2+\sum_{k=-\infty}^{\infty}q(c_{k+1}-c_k)^2 = (p+2q)\sum_{k=-\infty}^{\infty}c_k^2.$$

Для самого левого СЗ задачи (2.3) (обозначим его через p_1) получаем

$$p_1 = -2q+\min\left\{\left[\sum_{k=-\infty}^{\infty}(\mu+2k)^2c_k^2+\sum_{k=-\infty}^{\infty}(c_{k+1}-c_k)^2\right]\Big/\sum_{k=-\infty}^{\infty}c_k^2\right\}.$$

Отсюда сразу следует, что $p_1 \ge -2q$. Так как $(c_{k+1}-c_k)^2 \ge (|c_{k+1}|-|c_k|)^2$, то нужные для минимума величины c_k все либо неотрицательны, либо неположительны. Подряд двух c_k и c_{k+1}, равных нулю, быть не может, так как формула (2.3) трехчленная и все c_k стали бы нулями. Ситуации $c_k = 0$, $c_{k+1}c_{k-1} > 0$ также быть не может (см. (2.3)). Поэтому для p_1 все c_k отличны от нуля и одного знака (без ограничения будем считать их положительными). Отсюда следует некратность p_1 (далее мы увидим, что каждое СЗ задачи (2.3) некратное).

Достаточно полное исследование нужных нам свойств задачи (2.3) может быть проведено теми же методами, которыми исследовалась аналогичная задача в [12]. Поэтому мы ограничимся формулировками нужных нам выводов и фиксацией возникающих численных алгоритмов.

Каждое СЗ задачи (2.3) некратное. Через p_m обозначим m-е (при перечислении слева направо) СЗ задачи (2.3). Тогда соответствующая p_m последовательность c_k имеет $m-1$ перемен знака (т.е. ровно в $m-1$ местах имеет место $c_kc_{k-1} < 0$ или $c_k = 0$, $c_{k-1}c_{k+1} < 0$), ситуация $c_k = 0$, $c_{k-1}c_{k+1} > 0$ невозможна. Из (2.3) получаем

$$\frac{c_k}{c_{k-1}} = \frac{1}{M_k-\dfrac{c_{k+1}}{c_k}}, \tag{2.5}$$

где $M_k = [(\mu+2k)^2-p]/q$. Пусть $\alpha \le p \le \beta$. Возьмем такое k_f, что $[(\mu+2k)^2-\beta]/q \ge 2$ при $k \ge k_f$. Тогда при $\alpha \le p \le \beta \wedge k \ge k_f$ непрерывная дробь

$$\cfrac{1}{M_k-\cfrac{1}{M_{k+1}-\cfrac{1}{M_{k+2}-\dots}}}$$

сходится, ее значение – непрерывная положительная неубывающая функция p, обозначим ее через $\varphi(p)$; $c_k/c_{k-1} = \varphi(p)$ при $k = k_f$. Аналогично получаем $c_{k-1}/c_k = \psi(p)$ при $k = -k_f-1$, $\psi(p)$ – непрерывная положительная неубывающая функция. Поэтому задачу (2.3) можно решать не при $k = 0, \pm1, \pm2, \dots$, а при $-k_f \le k \le k_f-1$, используя "граничные условия"

$$c_{k_f}/c_{k_f-1} = \varphi(p), \quad c_{-k_f-1}/c_{-k_f} = \psi(p). \tag{2.6}$$

Используя приведенные утверждения, получаем следующий алгоритм решения задачи (2.3). В соответствии с пояснениями к формуле (2.3) берем $0 \le \mu < 2$. Фиксируем m. Пусть из каких-то соображений мы знаем оценку $\alpha \le p_m \le \beta$. Для этого β определяем k_f. Берем какое-либо p из указанного диапазона. Вычисляем (приблизительно, оборвав ценные дроби) значения $\varphi(p)$ и $\psi(p)$. Вычисляем по (2.5) значения c_{k+1}/c_k до k_0, близкого к нулю. Аналогично слева направо вычисляем c_k/c_{k+1}. Значение k_0 берем таким, чтобы $(c_{k_0+1}/c_{k_0})_r$ – значение, вычисленное при счете справа налево, и $(c_{k_0}/c_{k_0+1})_l$ – значение, вычисленное при счете слева направо, были бы положительными. Тогда верно следующее:

– суммарное (от $-k_f-1$ до k_f) число перемен знака в $\{c_k\}$ больше $m-1 \Rightarrow p > p_m$;

– суммарное число перемен знака в $\{c_k\}$ меньше $m-1 \Rightarrow p < p_m$;

– суммарное число перемен знака равно $m-1$ и $(c_{k_0+1}/c_{k_0})_r(c_{k_0}/c_{k_0+1})_l > 1 \Rightarrow p > p_m$;

– суммарное число перемен знака равно $m-1$ и $(c_{k_0+1}/c_{k_0})_r(c_{k_0}/c_{k_0+1})_l < 1 \Rightarrow p < p_m$;

– суммарное число перемен знака равно $m-1$ и $(c_{k_0+1}/c_{k_0})_r(c_{k_0}/c_{k_0+1})_l = 1 \Rightarrow p = p_m$.

В качестве алгоритма вычисления нужного значения p_m можно взять алгоритм, предложенный в [12].

Вычислив достаточно точно p_m, переходим к вычислению нужной последовательности $\{c_k\}$. Для этого, используя "граничные условия" (2.6), вычисляем отношения c_k/c_{k+1}, идя от $-k_f - 1$ слева направо, и отношения c_{k+1}/c_k, идя от k_f справа налево. Так как с достаточной точностью имеет место равенство

$$\left(\frac{c_{k_0+1}}{c_{k_0}}\right)_r = \left(\frac{c_{k_0+1}}{c_{k_0}}\right)_l,$$

то фиксируем какое-либо ненулевое значение c_{k_0} и, используя найденное отношение соседних членов, последовательно вычисляем c_{k_0+1}, c_{k_0+2}, …, c_{k_0-1}, c_{k_0-2}, … (о нормировке нужных решений исходного уравнения (2.1) сказано ниже). Возможное осложнение в вычислениях – обращение какого-либо из c_k в 0 и тем самым невозможность использовать c_{k+1}/c_k (или c_{k-1}/c_k). Простой способ преодоления этого осложнения – использование формулы (2.5) не для получения частного c_k/c_{k-1}, а для получения "проективного отношения" $c_k : c_{k-1}$, т.е. введение на каждом таком шаге какой-либо пары чисел, отношение которых равно $c_k : c_{k-1}$; если $c_k \approx 0$, то отношение $c_{k+1} : c_{k-1}$ определяется использованием (2.3).

Характеристическое уравнение на k-м шаге для (2.3) имеет вид

$$\omega^2 + \frac{p - (\mu + 2k)^2}{q}\omega + 1 = 0.$$

Корни его при больших k^2 – величины порядка $q/(4k^2)$ и $4k^2/q$, один корень очень маленький и один очень большой. Нужные нам последовательности быстро убывают в каждом из двух направлений: $\lim\limits_{k\to\infty} k^2 c_{k+1}/c_k = \lim\limits_{k\to-\infty} k^2 c_{k-1}/c_k = q/4$; ряд для $P(t)$ быстро сходится. Вследствие наличия большого корня, вычислив c_{k_0} и c_{k_0+1}, нецелесообразно вычислять c_{k_0+2}, c_{k_0+3}, …, используя формулу (2.3) (аналогично при счете налево); при таком вычислении влияние даже малых погрешностей в исходных данных (например, малая погрешность в p_m) и погрешностей арифметических операций катастрофически нарастает. Возможно, зафиксировав c_{k_0}, решать две возникшие "краевые задачи" (рекуррентная формула (2.3), значение c_{k_0}, одно из условий (2.6)), используя метод, предложенный в [13].

Как указано в пояснении к формуле (2.3), изменение μ на четное число не меняет спектр задачи. Но такое изменение меняет номер m нужного p_m (см. выше). Приведем окончательный рецепт получения тех решений задачи, которые являются продолжением функций $\cos\mu t$ и $(\sin\mu t)/\mu$, возникших при $q = 0$. Далее все формулируем для $q > 0$, $\mu \geq 0$.

Тот номер m, который нужен для продолжения $\cos\mu t$ (обозначим его через $m_{\cos}$), и тот номер m, который нужен для продолжения $(\sin\mu t)/\mu$ (обозначим его через $m_{\sin}$), вычисляются по следующим формулам:

– для μ нецелого $m_{\cos}(\mu) = m_{\sin}(\mu) = \mathrm{entier}(\mu) + 1$,

– для μ целого $m_{\cos}(\mu) = \mu + 1$ при $\mu \geq 0$, $m_{\sin}(\mu) = \mu$ при $\mu \geq 1$.

Для нецелого μ берем указанное выше m, находим соответствующее p_m. Вычислив $P(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2kt}$, получим $y(t) = e^{i\mu t}P(t)$. Берем $[y(t) + y(-t)]/[2y(0)]$ в качестве продолжения $\cos\mu t$ и $[y(t) - y(-t)]/[2y'(0)]$ в качестве продолжения $(\sin\mu t)/\mu$. Если μ – рациональное число, $\mu = 2k/l$, где k, l – целые, то так полученные функции являются $l\pi$-периодическими.

Для целого четного μ, беря $m_{\cos}(\mu)$ и вычисляя p_m, получаем четное π-периодическое решение, а беря $m_{\sin}(\mu)$ и вычисляя p_m, получаем нечетное π-периодическое решение. Это функции $ce_\mu(t)$ и $se_\mu(t)$.

Для целого нечетного μ получаем, соответственно, 2π-периодические решения $ce_\mu(t)$ и $se_\mu(t)$, не являющиеся π-периодическими.

3. ХАРАКТЕРИСТИЧЕСКИЙ ПОКАЗАТЕЛЬ МАТЬЕ

Характеристический показатель (ХП) Матье определяется из соотношения

$$F_\mu(t) = e^{i\mu t}P(t), \tag{3.1}$$

где F_μ – решение Флоке уравнения (1.1), $P(t)$ есть π-периодическая функция (см. [1], [2]). Соотношение (3.1) не определяет μ однозначным образом. Проведенный далее анализ позволяет уточнить этот вопрос.

Пусть $y_1(t)$ и $y_2(t)$ – фундаментальная система решений уравнения (1.1), определяемая начальными условиями

$$y_1(0) = 1, \quad y_1'(0) = 0, \tag{3.2}$$

$$y_2(0) = 0, \quad y_2'(0) = 1. \tag{3.3}$$

При этом, в силу четности коэффициентов уравнения (1.1), $y_1(t)$ – четная функция, $y_2(t)$ – нечетная функция. Введем величину

$$S(p) = y_1(\pi) + y_2'(\pi)$$

(описанная далее процедура применима к любому уравнению с периодическим потенциалом и не использует того факта, что для уравнения Матье слагаемые в правой части равны).

Далее значение параметра q предполагается зафиксированным. Тогда (см., например, [14], гл. I, § 4]) верно следующее:

1) p является СЗ периодической задачи (в обозначениях [1], a_n и b_n с четным n) $\Leftrightarrow S(p) = 2$;

2) p является СЗ антипериодической задачи (a_n и b_n с нечетным n) $\Leftrightarrow S(p) = -2$.

Далее, из асимптотик для решений (1.1) следует, что при $p \longrightarrow -\infty$ имеет место $S(p) \sim$ $\sim 2\cosh(\pi\sqrt{|p|}) \longrightarrow +\infty$ и при $p \longrightarrow +\infty$ значение $S(p)$ колеблется от 2 к –2 и обратно.

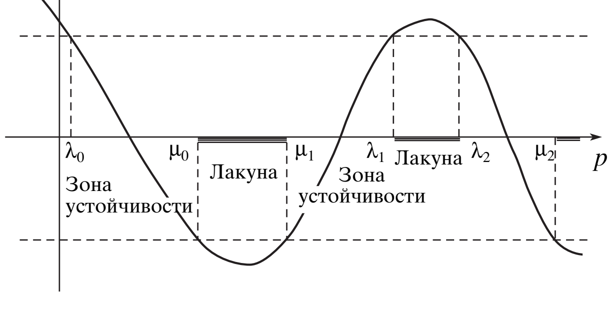
Пусть $\lambda_0, \lambda_1, \ldots$ – корни уравнения $S(p) = 2$, $\mu_0, \mu_1, \ldots$ – корни уравнения $S(p) = -2$. В вырожденных случаях эти уравнения могут иметь кратные корни (т.е. $\lambda_k = \lambda_{k+1}$ или $\mu_k = \mu_{k+1}$), однако (см. [1]) для уравнения Матье так будет лишь в тривиальном случае $q = 0$. Промежутки $[\mu_0, \mu_1]$, (λ_1, λ_2), $[\mu_2, \mu_3]$, (λ_3, λ_4) называются лакунами. График функции $S(p)$ для уравнения (1.1) при $q \neq 0$ схематично изображен на фиг. 1. Из свойств симметрии коэффициентов уравнения (1.1) и осцилляционных теорем Штурма вытекает следующее:

1) $a_0 = \lambda_0$;

2) при $q > 0$ имеет место $b_1 = \mu_0$, $a_1 = \mu_1$, $b_2 = \lambda_1$, $a_2 = \lambda_2, \ldots$;

3) при $q < 0$ имеет место $a_1 = \mu_0$, $b_1 = \mu_1$, $b_2 = \lambda_1$, $a_2 = \lambda_2, \ldots$;

4) при $q = 0$ имеет место $\mu_0 = \mu_1 = a_1 = b_1$, $\lambda_1 = \lambda_2 = a_2 = b_2, \ldots$.



Фиг. 1.

Далее, из периодичности коэффициентов уравнения (1.1) следует, что для любого его решения $y(t)$ имеет место соотношение

$$\begin{pmatrix} y(\pi) \\ y'(\pi) \end{pmatrix} = \begin{pmatrix} y_1(\pi) & y_2(\pi) \\ y_1'(\pi) & y_2'(\pi) \end{pmatrix} \begin{pmatrix} y(0) \\ y'(0) \end{pmatrix}. \tag{3.4}$$

СЗ матрицы, стоящей в правой части, равны

$$\kappa_\pm = \left[S(p) \mp \sqrt{S^2(p) - 4}\right]/2.$$

Для $p \in (-\infty, \lambda_0)$ и в лакунах $\kappa_\pm$ вещественны, $|\kappa_+| < 1$, $|\kappa_-| > 1$ (области неустойчивости). Для p вне лакун $|\kappa_\pm| = 1$ и комплексно сопряжены (области устойчивости). В [14, гл. II, § 6] указано, что $\kappa_\pm$ при различных p являются аналитическими продолжениями друг друга. Рассмотрим более подробно аналитическую зависимость $S(p)$ и $\kappa_\pm$ от p. Пусть p проходит вещественную ось в положительном направлении, обходя СЗ a_k и b_k (точки ветвления $\kappa_\pm$) сверху. Учтем, что на промежутках $[\lambda_0, \mu_0]$, $[\lambda_2, \mu_2]$, … будет $S'(p) < 0$, а на промежутках $[\mu_1, \lambda_1]$, $[\mu_3, \lambda_3]$, … будет $S'(p) > 0$.

Тогда $S(p)$ будет двигаться по "гантеле" от –2 к 2, обходя эти точки по часовой стрелке. Движение κ_+ будет следующим (см. фиг. 2): из нуля $\kappa = +0$ ($p = -\infty$) вправо до $\kappa = 1$ ($p = \lambda_0$), затем по окружности $|\kappa| = 1$ против часовой стрелки до $\kappa = -1$ ($p = \mu_0$), затем (в лакуне $[\mu_0, \mu_1]$) по вещественной оси вправо (не доходя до $\kappa = 0$) и обратно до $\kappa = -1$ ($p = \mu_1$), затем по окружности против часовой стрелки до $\kappa = 1$ ($p = \lambda_1$), и т.д.

Проведенный анализ позволяет дать еще одно определение СЗ с произвольным вещественным номером: при произвольном $r \in \mathbb{R}^+$ значение p_r однозначно определяется следующими условиями:

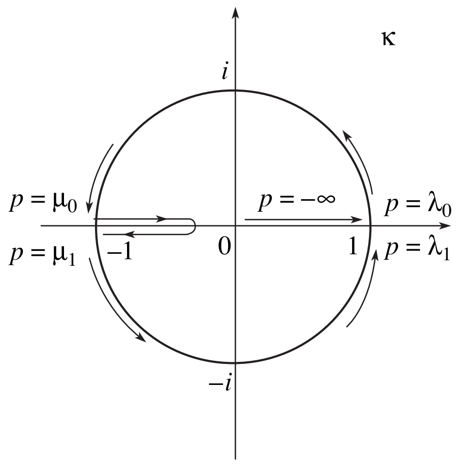
1)

$$\kappa_+(p_r) = e^{i\pi r}; \tag{3.5}$$

2) число полных оборотов, которое сделало $\kappa_+(p)$ при изменении p от $-\infty$ до p_r равно целой части от $r/2$.

Замечание. Для иррациональных r обозначения a_r, b_r вместо p_r, по-видимому, были бы неудачными, так как соответствующие решения уравнения Матье не могут быть охарактеризованы как собственные функции периодической краевой задачи на кратном промежутке.

Описанная конструкция позволяет однозначно и корректно определить вещественную часть ХП Матье, а именно избежать неопределенного слагаемого вида $2k$, где k целое. Для этого нужно сформулированные выше условия 1), 2) использовать в обратную сторону – от p к r: нужно взять аналитическое продолжение логарифма вдоль пути $(-\infty, p]$. В частности, в лакунах веще-



Фиг. 2.

ственная часть ХП равна удвоенному числу оборотов, которые к этому моменту сделало $\kappa_+(p)$. Второй ХП определяется как комплексно-сопряженный к первому.

По обсуждаемому вопросу в [1, 20.3.5] и в [2, 16.2] утверждается, что ХП определяется в принципе неоднозначно. В некоторых других текстах ошибочно утверждается, что ХП однозначно определяется уже из известного соотношения $\cos\pi\mu = y_1(\pi) = y_2'(\pi)$. В программных реализациях встречаются также различные произвольные выборы вещественной части и знака мнимой части с самопроизвольными скачками. Определенный же описанным выше образом ХП Матье, помимо определяющего соотношения (3.2), удовлетворяет всем естественным требованиям непрерывности.

4. ВЫЧИСЛЕНИЕ ЗНАЧЕНИЙ РЕШЕНИЙ УРАВНЕНИЯ МАТЬЕ ПРИ БОЛЬШИХ ЗНАЧЕНИЯХ АРГУМЕНТА

Пусть имеется какой-либо способ вычисления решений уравнения Матье в форме (2.1) с условиями (3.1) и (3.2) и их производных при $0 < t \le \pi/2$. Пусть $|t|$ велико; как устойчиво вычислить $y_1(t)$, $y_2(t)$, $y_1'(t)$, $y_2'(t)$ для такого t? Прежде всего отметим, что выбранный при $0 < t \le \pi/2$ способ вычисления этих функций дает возможность легко вычислить их при $\pi/2 < t \le \pi$. Действительно, рассмотрим также уравнение в форме (1.1):

$$\tilde{y}'' + (p - 2q\cos 2t)\tilde{y} = 0 \tag{4.1}$$

и два его решения: $\tilde{y}_1(t)$ и $\tilde{y}_2(t)$, где

$$\tilde{y}_1(0) = 1, \quad \tilde{y}_1'(0) = 0, \tag{4.2}$$

$$\tilde{y}_2(0) = 0, \quad \tilde{y}_2'(0) = 1. \tag{4.3}$$

Очевидно, что $\tilde{y}_1(t - \pi/2)$ и $\tilde{y}_2(t - \pi/2)$ – решения уравнения (1.1) и так как $\tilde{y}_1\tilde{y}_2'(t) - \tilde{y}_2\tilde{y}_1' \equiv 1$, то

$$\begin{aligned} y_1(t) &= \tilde{y}_2'\left(\frac{\pi}{2}\right)\tilde{y}_1\left(t - \frac{\pi}{2}\right) + \tilde{y}_1'\left(\frac{\pi}{2}\right)\tilde{y}_2\left(t - \frac{\pi}{2}\right), \\ y_2(t) &= \tilde{y}_2\left(\frac{\pi}{2}\right)\tilde{y}_1\left(t - \frac{\pi}{2}\right) + \tilde{y}_1\left(\frac{\pi}{2}\right)\tilde{y}_2\left(t - \frac{\pi}{2}\right). \end{aligned} \tag{4.4}$$

Поэтому, используя упомянутый выше способ решения уравнения (2.1) при $0 < t \le \pi/2$ и применяя его к уравнению (4.1), получаем распространение этого способа на диапазон $(0, \pi]$.

Рассмотрим задачу вычисления $y_1(t + k\pi)$ и $y_2(t + k\pi)$, где k целое, возможно большое. Эти две функции также являются решениями уравнения (2.1). Поэтому имеем

$$\begin{aligned} y_1(t + k\pi) &= c_{11}y_1(t) + c_{12}y_2(t), \\ y_2(t + k\pi) &= c_{21}y_1(t) + c_{22}y_2(t), \end{aligned} \tag{4.5}$$

где константы c_{rs} зависят от k, их несложно вычислить. Далее мы используем факт существования для уравнения (2.1) решения Флоке

$$F(t) = e^{i\mu t}P(t), \tag{4.6}$$

где $P(t)$ есть π-периодическая функция, $P(t) \not\equiv 0$; μ – характеристический показатель. В окончательный алгоритм функция $P(t)$ не входит, используется только формула для определения μ:

$$\cos\mu\pi = y_1(\pi) = y_2'(\pi) = 1 + 2y_1'\left(\frac{\pi}{2}\right)y_2\left(\frac{\pi}{2}\right); \tag{4.7}$$

пользуясь упомянутым выше способом вычисления $y_1(t)$, $y_1'(t)$, $y_2(t)$, $y_2'(t)$ в диапазоне $(0, \pi]$, вычисляем какое-либо (возможно, невещественное) значение μ.

Выводя формулы для c_{rs}, временно предположим, что $F(0) \ne 0$ и $F'(0) \ne 0$. Тогда имеем

$$\begin{aligned} y_1(t) &= \frac{F(t)+F(-t)}{2F(0)}, \\ y_2(t) &= \frac{F(t)-F(-t)}{2F'(0)}. \end{aligned} \tag{4.8}$$

Отсюда, учитывая (4.6), получаем

$$\begin{aligned} y_1(t+k\pi) &= \frac{e^{i\mu k\pi}F(t)+e^{-i\mu k\pi}F(-t)}{2F(0)}, \\ y_2(t+k\pi) &= \frac{e^{i\mu k\pi}F(t)-e^{-i\mu k\pi}F(-t)}{2F'(0)}. \end{aligned} \tag{4.9}$$

Исключив из (4.8), (4.9) величины $F(t)$ и $F(-t)$, получим

$$y_1(t+k\pi) = (\cos k\pi\mu)y_1(t) + i\frac{F'(0)}{F(0)}(\sin k\pi\mu)y_2(t), \tag{4.10}$$

$$y_2(t+k\pi) = i\frac{F(0)}{F'(0)}(\sin k\pi\mu)y_1(t) + (\cos k\pi\mu)y_2(t). \tag{4.11}$$

Из (4.10) имеем

$$y_1'(t+k\pi) = (\cos k\pi\mu)y_1'(t) + i\frac{F'(0)}{F(0)}(\sin k\pi\mu)y_2'(t). \tag{4.12}$$

Положив $t = 0$, $k = 1$ в (4.11), (4.12), получим

$$y_2(\pi) = i\frac{F(0)}{F'(0)}\sin\pi\mu, \quad y_1'(\pi) = i\frac{F'(0)}{F(0)}\sin\pi\mu.$$

Дополнительно предположив, что μ не является целым, получим

$$\begin{aligned} i\frac{F(0)}{F'(0)}\sin k\pi\mu &= y_2(\pi)\frac{\sin k\pi\mu}{\sin\pi\mu}, \\ i\frac{F'(0)}{F(0)}\sin k\pi\mu &= y_1'(\pi)\frac{\sin k\pi\mu}{\sin\pi\mu}. \end{aligned}$$

Подставив эти значения в (4.10) и (4.11), окончательно получим

$$\begin{aligned} y_1(t+k\pi) &= (\cos k\pi\mu)y_1(t) + y_1'(\pi)\frac{\sin k\pi\mu}{\sin\pi\mu}y_2(t), \\ y_2(t+k\pi) &= y_2(\pi)\frac{\sin k\pi\mu}{\sin\pi\mu}y_1(t) + (\cos k\pi\mu)y_2(t). \end{aligned} \tag{4.13}$$

Так как $(\sin k\pi\mu)/\sin\pi\mu$ и $\cos k\pi\mu$ суть многочлены от $\cos\pi\mu$, а $\cos\pi\mu = y_1(\pi) = y_2'(\pi)$ (см. (4.7)), то коэффициенты c_{rs} в (4.5) суть многочлены от $y_1(\pi)$, $y_1'(\pi)$, $y_2(\pi)$, $y_2'(\pi)$. Использовать это обстоятельство для непосредственных вычислений, по-видимому, неразумно, но для обоснования рассматриваемого далее предельного перехода оно существенно. Именно формула (4.13) была доказана при дополнительном предположении, что $F(0) \neq 0$, $F'(0) \neq 0$ и μ не является целым. Случай несоблюдения этого предположения получаем как предельный. Поэтому формула (4.13) верна и в общем случае.

5. О ЧИСЛЕННОЙ РЕАЛИЗАЦИИ

Описанные методы были реализованы в системе вычислений с произвольной точностью. Для определения внутренних параметров, таких как число взятых членов ряда, место обрыва цепных дробей и др., потребовалось разработать специальные методы контроля точности. Ниже приведены некоторые результаты расчетов.

В таблице, являющейся “продолжением” табл. 4.1 из [6], представлены СЗ b_{2n} при $q = 25$. Результаты расчетов с 40 значащими цифрами подтверждаются (воспроизводятся) при расчетах с

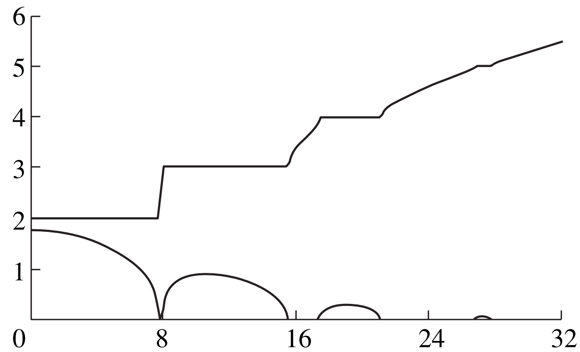
Таблица

СЗ	[1]	[8]	[6]	Расчеты авторов
b_2	–21.31486062	–21.314860622	–21.3148606222499	–21.31486062224985085431466497257381226977
b_4	–	12.986489953	12.9864899527425	12.98648995274245978696086926962446752855
b_6	–	41.801071292	41.8010712918115	41.80107129181058013238706064957626657798
b_8	–	69.057988351	69.0579883512758	69.05798835128618256012392585342334608242
b_{10}	103.22568004	103.225680042	103.225680042418	103.2256800423734700047997305444380455190
b_{12}	–	146.207674647	146.207674647279	146.2076746474580792325359615455730525781
b_{14}	–	197.611164916	197.611164920589	197.6111649156508603480957728194897503429
b_{16}	–	257.229284862	257.22928528627	257.2292848625012979647682267875409801588

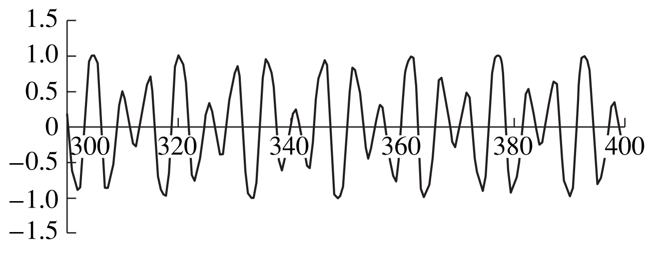
бóльшим числом цифр (60, 100 и т.д.). Отметим, что наши результаты несколько отличаются от результатов из [6] в b_6, …, b_{12}, а в b_{14}, b_{16} отличаются более существенно. Наоборот, получено полное подтверждение результатов из [8] со всеми приведенными там значащими цифрами.

На фиг. 3 представлен график характеристического показателя (верхняя кривая – вещественная часть, нижняя кривая – мнимая часть) для значений параметров $q = 10$, $0 < p < 32$. Как видно, обе кривые непрерывны, а вещественная часть монотонно возрастает в промежутках вне лакун.

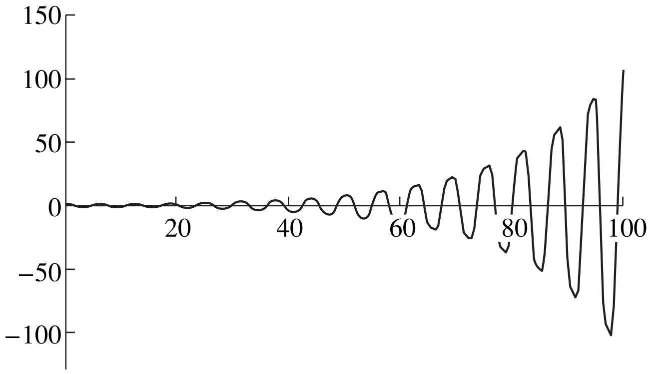
На фиг. 4, 5 приведены графики решения уравнения Матье на достаточно больших промежутках изменения аргумента. Использовалась обычная (ср. [1, рис. 20.2 и 20.3]) нормировка условиями вида (3.2), (3.3), соответственно, для четных и нечетных решений с соответствующей пере-



Фиг. 3.



Фиг. 4.



Фиг. 5.

нормировкой коэффициентов c_k. Фиг. 4 соответствует параметрам $p = 2.0$, $q = 1.0$, $300 < t < 400$ (зона устойчивости $a_1 < p < b_2$), фиг. 5 – параметрам $p = 1.85$, $q = 1.0$, $0 < t < 100$ (лакуна $b_1 < p < a_1$).

СПИСОК ЛИТЕРАТУРЫ

1. *Абрамовиц М., Стиган И.* Справочник по специальным функциям с формулами, графиками и математическими таблицами. М.: Наука, 1979.
2. *Бейтман Г., Эрдейи А.* Высшие трансцендентные функции. Т. 1–3. М.: Наука, 1967.
3. *Frenkel D., Portugal T.* Algebraic methods to compute Mathieu functions: Laboratório Nacional de Computação Cientifica. Abstract. Petrópolis, Brazil, 2002. P. 1–14.
4. *Kokkorakis G.C., Roumeliotis J.A.* Power series expansions for Mathieu functions with small arguments // Math. Comput. 2000. V. 70. № 235. P. 1221–1235.
5. *Algarhan F.A.* A complete method for the computations of Mathieu characteristic numbers of integer orders // SIAM Rev. 1996. V. 38. 2. P. 239–255.
6. *Gutiérrez Vega J.C.* Theory and numerical analysis of the Mathieu functions. Monterrey, NL, México, 2003. http://homepages.mty.itesm.mx/jgutierr/Mathieu/Mathieu.pdf.
7. *Власов В.К., Глухова М.Н., Королев Л.Н. и др.* О вычислении функций Матье // Вестн. МГУ. Сер. 15. Вычисл. матем. и кибернетика. 1992. № 1. С. 65–69.
8. *Leeb W.* Algorithm 537. Characteristic values of Mathieu's differential equation // ACM Trans. Math. Software. 1979. V. 5. № 1. P. 112–117.
9. *Березман А.М., Керимов М.К., Скороходов С.Л., Шадрин Г.А.* О вычислении собственных значений уравнения Матье с комплексным параметром // Ж. вычисл. матем. и матем. физ. 1986. Т. 26. № 9. С. 1350–1361.
10. *Shirts R.B.* The computation of eigenvalues and solutions of Mathieu's differential equation for noninteger order // ACM Trans. Math. Software. 1993. V. 19. № 3. P. 377–390.
11. *Meixner J., Schäfke F.W., Wolf G.* Mathieu functions and spheroidal functions and their mathematical foundations. Further studies // Lect. Notes Math. 837. Berlin: Springer, 1981.
12. *Абрамов А.А., Курочкин С.В.* Высокоточное вычисление угловых сфероидальных функций // Ж. вычисл. матем. и матем. физ. 2006. Т. 46. № 1. С. 12–17.
13. *Абрамов А.А.* Выделение медленно растущих последовательностей, члены которых удовлетворяют заданным рекуррентным соотношениям // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. № 4. С. 661–668.
14. *Левитан Б.М., Саргсян И.С.* Операторы Штурма–Лиувилля и Дирака. М.: Наука, 1988.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 424–437

УДК 519.624.2

РЕШЕНИЯ СИНГУЛЯРНО ВОЗМУЩЕННОГО ПАРАБОЛИЧЕСКОГО УРАВНЕНИЯ С ВНУТРЕННИМИ И ПОГРАНИЧНЫМИ СЛОЯМИ, ЗАВИСЯЩИМИ ОТ РАСТЯНУТЫХ ПЕРЕМЕННЫХ РАЗНОГО ПОРЯДКА[1)]

Е. Е. Букжалёв, А. Б. Васильева

(119992 Москва, Ленинские горы, МГУ, физ. ф-т)

e-mail: bukzhalev@mail.ru

Поступила в редакцию 29.12.2005 г.
Переработанный вариант 06.09.2006 г.

Рассматривается решение сингулярно возмущенного параболического уравнения, имеющее внутренние и пограничные слои, растянутые переменные которых могут зависеть от различных степеней параметра возмущения. Проводится построение и обоснование асимптотики этого решения, и дается доказательство его устойчивости. Библ. 12.

Ключевые слова: сингулярные возмущения, контрастные структуры типа ступеньки, метод пограничных функций, метод дифференциальных неравенств, устойчивость решений параболических уравнений.

1. ПОСТАНОВКА ЗАДАЧИ

Настоящая работа посвящена исследованию краевой задачи для следующего параболического уравнения с условием периодичности по временно́му аргументу:

$$L[y] \equiv \varepsilon^4\left(\frac{\partial^2 y}{\partial x^2} - \frac{\partial y}{\partial t}\right) - \varepsilon\frac{\partial y}{\partial x}A(y, x, t) - B(y, x, t) = 0,$$

$$y(x, t + T, \varepsilon) = y(x, t, \varepsilon), \quad (x, t) \in \Omega \equiv (0, 1) \times \mathbb{R}, \tag{1.1}$$

$$y(0, t, \varepsilon) = y^0(t), \quad y(1, t, \varepsilon) = y^1(t), \quad t \in \mathbb{R},$$

где $\varepsilon > 0$ – малый параметр, A, B, y^0 и y^1 суть T-периодические по t функции. Постановка (1.1) является естественным обобщением задачи, рассмотренной ранее в работе [1]:

$$\varepsilon^4\frac{d^2 y}{dx^2} = \varepsilon\frac{dy}{dx}A(y, x) + B(y, x), \quad x \in (0, 1),$$

$$y(0, \varepsilon) = y^0, \quad y(1, \varepsilon) = y^1. \tag{1.2}$$

Отметим, что (1.1) представляет собой (1.2) в случае, когда все входящие в нее величины, включая само решение, не зависят от t.

В [1] с помощью метода пограничных функций (см. [2]) была построена асимптотика решений задачи (1.2), обладающих двумя пограничными и одним внутренним слоями, и с помощью метода дифференциальных неравенств (см. [3]) проведено ее обоснование. При этом внутренний слой описывал переход между некоторыми двумя решениями вырожденного уравнения. Решения с внутренними переходными слоями носят название контрастных структур типа ступеньки (КСТС).

Характерной особенностью задачи (1.2) является то, что как погранслойные, так и внутренние переходные составляющие ее решений могут быть функциями растянутых переменных разного порядка относительно параметра возмущения (по этому поводу см. также [4]). В соответствии с этим выделяют пограничные и переходные слои двух видов: резкого (растянутая переменная $\sim 1/\varepsilon^3$) и плавного (растянутая переменная $\sim 1/\varepsilon$).

[1)] Работа выполнена при финансовой поддержке РФФИ (код проекта 04-01-00710).

Решение задачи (1.1), так же как и решение задачи (1.2), может обладать как пограничными, так и внутренними переходными слоями. Ниже строится и обосновывается асимптотика решения, обладающего двумя пограничными (в окрестностях граничных прямых $x = 0$ и $x = 1$) и одним внутренним (в окрестности некоторой внутренней кривой $x = x^*(t, \varepsilon)$) слоями. При этом внутренний переходный слой описывается функцией, зависящей от растянутой переменной $\sim 1/\varepsilon$. Кроме того, будет проведено исследование на устойчивость указанного решения уравнения (1.1), рассматриваемого как стационарное решение следующей начально-краевой задачи:

$$\begin{gathered} L[y] = 0, \quad (x, t) \in \Omega^+ \equiv (0, 1) \times (0, +\infty), \\ y(x, 0, \varepsilon) = y_0(x), \quad x \in [0, 1], \\ y(0, t, \varepsilon) = y^0(t), \quad y(1, t, \varepsilon) = y^1(t), \quad t \in \mathbb{R}^+ \equiv [0, +\infty), \end{gathered} \tag{1.3}$$

где функция $y_0(x)$ удовлетворяет условиям согласования порядка нуль: $y_0(0) = y^0(0)$, $y_0(1) = y^1(0)$.

2. ПОСТРОЕНИЕ АСИМПТОТИКИ

Пусть уравнение $B(y, x, t) = 0$ имеет три корня $y = \varphi_i(x, t)$, $i = 1, 2, 3$, причем

$$\varphi_1(x, t) < \varphi_2(x, t) < \varphi_3(x, t),$$

$$B_y(\varphi_{1,3}(x, t), x, t) > 0, \quad B_y(\varphi_2(x, t), x, t) < 0,$$

$$A(\varphi_1(0, t), 0, t) \neq 0, \quad A(\varphi_3(1, t), 1, t) \neq 0.$$

Тогда при определенных условиях может возникнуть КСТС с переходом с $\varphi_1(x, t)$ на $\varphi_3(x, t)$ в окрестности некоторой кривой $x = x^*(t, \varepsilon)$. В зависимости от знаков A на φ_1 и φ_3 при $x = x^*(t, 0) = x_0(t)$ внутренний слой является функцией следующих растянутых переменных:

1) если $A(\varphi_1(x_0, t), x_0, t) > 0$, $A(\varphi_3(x_0, t), x_0, t) < 0$, то $\eta = (x - x^*(t, \varepsilon))/\varepsilon^3$, – КСТС резкого вида,

2) если $A(\varphi_1(x_0, t), x_0, t) < 0$, $A(\varphi_3(x_0, t), x_0, t) > 0$, то $\eta = (x - x^*(t, \varepsilon))/\varepsilon$, – КСТС плавного вида.

Как уже было сказано, мы ограничимся рассмотрением случая 2), когда переходная переменная η имеет порядок $1/\varepsilon$.

Значение функции $y(x, t, \varepsilon)$ на переходной кривой $x = x^*(t, \varepsilon)$ обозначим через $y(x^*, t, \varepsilon) = \gamma(t, \varepsilon)$ ($\varphi_1(x_0, t) < \gamma(t, 0) < \varphi_3(x_0, t)$).

Итак, ищем решение вида

$$y(x, t, \varepsilon) = \tilde{y}(x, t, \varepsilon) + \Pi(\tau, t, \varepsilon) + Q(\rho, t, \varepsilon), \quad (x, t) \in \bar{\Omega},$$

где

$$\tilde{y}(x, t, \varepsilon) = \begin{cases} \bar{y}^m(x, t, \varepsilon) + T^{(-)}(\eta, t, \varepsilon), & (x, t) \in \bar{\Omega}_l, \\ \bar{y}^p(x, t, \varepsilon) + T^{(+)}(\eta, t, \varepsilon), & (x, t) \in \bar{\Omega}_r; \end{cases}$$

здесь, в свою очередь,

$$\Omega_l = (0, x^*(t, \varepsilon)) \times \mathbb{R},$$

$$\Omega_r = (x^*(t, \varepsilon), 1) \times \mathbb{R},$$

а порядок погранслойных переменных определяется знаком A на решениях вырожденного уравнения в соответствующих точках: $\tau = x/\varepsilon^3$ при $A(\varphi_1(0, t), 0, t) < 0$ и $\tau = x/\varepsilon$ при $A(\varphi_1(0, t), 0, t) > 0$, $\rho = (x - 1)/\varepsilon$ при $A(\varphi_3(1, t), 1, t) < 0$ и $\rho = (x - 1)/\varepsilon^3$ при $A(\varphi_3(1, t), 1, t) > 0$.

Ниже строим только внутренние переходные слои, ибо регулярные и пограничные части по существу уже построены в [5].

Представляя x^*, γ^*, $\bar{y}^m$, $\bar{y}^p$, $T^{(-)}$, $T^{(+)}$ в виде рядов по малому параметру, а именно

$$x^*(t, \varepsilon) = x_0(t) + \varepsilon x_1(t) + \ldots,$$

$$\gamma^*(t, \varepsilon) = \gamma_0(t) + \varepsilon \gamma_1(t) + \ldots,$$

$$\bar{y}^m(x^* + \eta\varepsilon, t, \varepsilon) = \bar{y}_0^m(x^* + \eta\varepsilon, t) + \varepsilon \bar{y}_1^m(x^* + \eta\varepsilon, t) + \ldots = y_0^{(-)}(t) + \varepsilon y_1^{(-)}(t, \eta) + \ldots,$$

$$\bar{y}^p(x^* + \eta\varepsilon, t, \varepsilon) = \bar{y}_0^p(x^* + \eta\varepsilon, t) + \varepsilon\bar{y}_1^p(x^* + \eta\varepsilon, t) + \ldots = y_0^{(+)}(t) + \varepsilon y_1^{(+)}(t, \eta) + \ldots,$$

$$T^{(-)}(\eta, t, \varepsilon) = T_0^{(-)}(\eta, t) + \varepsilon T_1^{(-)}(\eta, t) + \ldots,$$

$$T^{(+)}(\eta, t, \varepsilon) = T_0^{(+)}(\eta, t) + \varepsilon T_1^{(+)}(\eta, t) + \ldots,$$

согласно методу пограничных функций получаем ряд задач для определения переходных членов асимптотики. Поскольку структуры задач для $T_n^{(-)}$ и $T_n^{(+)}$ полностью идентичны, ниже ограничиваемся построением первых.

Отметим, что члены переходного ряда будут определяться из уравнений первого порядка, так что их гладкость в переходной точке будет следствием их непрерывности. Поэтому для определения составляющих $x^*(t, \varepsilon)$ придется задействовать не связанные с гладкостью (как это наиболее часто бывает) соображения. Эти соображения, как будет показано ниже, связаны с возможностью разрешимости возникающих для членов переходного ряда уравнений.

В *нулевом приближении* имеем

$$\frac{\partial T_0^{(-)}}{\partial\eta} = P^{(-)}(T_0^{(-)}, t) \equiv -\frac{B(\varphi_1 + T_0^{(-)}, x_0, t)}{A(\varphi_1 + T_0^{(-)}, x_0, t)},$$

$$T_0^{(-)}(0, t) = \gamma_0(t) - \varphi_1(x_0(t), t), \tag{2.1}$$

$$T_0^{(-)}(-\infty, t) = 0.$$

Вспоминая, что $B_y(\varphi_1(x_0, t), x_0, t) > 0$, а $A(\varphi_1(x_0, t), x_0, t) < 0$, легко убеждаемся в том, что $T_0^{(-)} = 0$ – устойчивая точка покоя при $\eta \longrightarrow -\infty$. Значит, если $\gamma_0 - \varphi_1$ принадлежит ее области влияния, то $T_0^{(-)}(-\infty, t) = 0$. Аналогичное требование возникает и при рассмотрении соответствующей задачи для $T_0^{(+)}$.

Для выполнения условия данной принадлежности необходимо, чтобы $P^{(-)}(T_0^{(-)}, t)$ и $P^{(+)}(T_0^{(+)}, t) \equiv$ $\equiv -B(\varphi_3 + T_0^{(+)}, x_0, t)/A(\varphi_3 + T_0^{(+)}, x_0, t)$ не обращались в нуль на $(0, \gamma_0 - \varphi_1] \times \mathbb{R}$ и $[\gamma_0 - \varphi_3, 0) \times \mathbb{R}$ соответственно. Поэтому приходим к выводу, что

$$A(\varphi_2(x_0, t), x_0, t) \equiv I(x_0, t) = 0, \tag{2.2}$$

иначе (в силу $B(\varphi_2(x_0(t), t), x_0(t), t) = 0$) $P^{(-)}(T_0^{(-)}, t)$ при $T_0^{(-)} = \varphi_2 - \varphi_1$ либо $P^{(+)}(T_0^{(+)}, t)$ при $T_0^{(+)} =$ $= \varphi_2 - \varphi_3$ будет равно нулю. Причем для того, чтобы $P^{(\mp)}(T_0^{(\mp)}, t)$ при этом действительно не обращалось в нуль (или бесконечность), потребуем выполнения условий

$$A_y(\varphi_2(x_0(t), t), x_0(t), t) \neq 0, \quad B_y(\varphi_2(x_0(t), t), x_0(t), t) \neq 0 \tag{2.3}$$

(в этом случае B_y будет меньше нуля, а A_y – больше).

Кроме того, $P^{(-)}(T_0^{(-)}, t)$ и $P^{(+)}(T_0^{(+)}, t)$ не должны обращаться в бесконечность всюду на $[0, \gamma_0 - \varphi_1]$ и $[\gamma_0 - \varphi_3, 0]$ соответственно. Отсюда приходим к еще одному требованию на $x_0(t)$:

$$A(y, x_0(t), t) \neq 0 \ \text{ при } \ y \in [\varphi_1(x_0(t), t), \varphi_3(x_0(t), t)] \backslash \varphi_2(x_0(t), t) \tag{2.4}$$

(при его выполнении на первом полуинтервале A, естественно, будет иметь отрицательный знак, а на втором – положительный).

Выражения (2.2), (2.3), (2.4) можно рассматривать как систему соотношений для определения $x_0(t)$. Будем считать, что существует ее решение $x = x_0(t)$ такое, что

$$0 < x_0(t) < 1, \quad t \in \mathbb{R},$$
$$I_x(x, t)\big|_{x = x_0(t)} = \left.\frac{\tilde{B}_y\tilde{A}_x - \tilde{B}_x\tilde{A}_y}{\tilde{B}_y}\right|_{x_0 = x_0(t)} < 0, \quad t \in \mathbb{R}. \tag{2.5}$$

Здесь, а также для описания старших приближений использованы обозначения

$$\bar{y}_n^{(-)}(t) = \bar{y}_n^m(x_0(t), t), \quad \bar{y}_n^{(+)}(t) = \bar{y}_n^p(x_0(t), t),$$

$$\bar{E}^{(-)}(t) = E(\bar{y}_0^{(-)}(t), x_0(t), t), \quad \bar{E}^{(+)}(t) = E(\bar{y}_0^{(+)}(t), x_0(t), t),$$

$$\tilde{E}^{(-)}(\eta, t) = E(\bar{y}_0^{(-)}(t) + T_0^{(-)}(\eta, t), x_0(t), t),$$

$$\tilde{E}(t) = E(\varphi_2(x_0(t), t), x_0(t), t), \quad \tilde{P}^{(-)}(\eta, t) = P^{(-)}(T_0^{(-)}(\eta, t), t),$$

$$\tilde{P}_0^{(-)}(\eta, t) = \tilde{P}^{(-)}(\eta, t)/\tilde{P}^{(-)}(0, t),$$

$$\bar{\varphi}_1 = \bar{\varphi}_1(t) \equiv \varphi_1(x_0(t), t) = \bar{y}_0^{(-)}(t),$$

$$\bar{\varphi}_3 = \bar{\varphi}_3(t) \equiv \varphi_3(x_0(t), t) = \bar{y}_0^{(+)}(t),$$

где в качестве функции E могут выступать A и B, а также их частные производные.

Кроме того, обратим внимание на тот факт, что, начиная с $n = 1$, $y_n^{(-)}$ может быть представлено в виде

$$y_n^{(-)} = \frac{\partial \bar{y}_0^{(-)}}{\partial x}(t)x_n + \Delta y_n^{(-)},$$

где $\Delta y_n^{(-)}$ является функцией $\bar{y}_0^m(x_0, t), \ldots, \bar{y}_n^m(x_0, t)$ и их производных в точке $(x_0(t), t)$, а также $x_1(t) + \eta$, $x_2(t), \ldots, x_{n-1}(t)$ и не зависит от $x_n(t)$. Ниже под $\Delta y_n^{(-)}$ понимаем его значение при $\eta = 0$. При постановке задач для $T_n^{(+)}$ используется обозначение $\Delta y_n^{(+)}$, имеющее аналогичный смысл.

Первое приближение:

$$\begin{gathered}\tilde{A}^{(-)}\frac{\partial T_1^{(-)}}{\partial \eta} + (\tilde{B}_y^{(-)} + \tilde{P}^{(-)}\tilde{A}_y^{(-)})T_1^{(-)} + \tilde{G}_1^{(-)} = 0,\\ T_1^{(-)}(0, t) = \gamma_1(t) - \frac{\partial \bar{y}_0^{(-)}}{\partial x}(t)x_1(t) - \bar{y}_1^{(-)}(t).\end{gathered} \tag{2.6}$$

Решением уравнений (2.6) будет

$$T_1^{(-)} = \left(\gamma_1(t) - \frac{\partial \bar{y}_0^{(-)}}{\partial x}(t)x_1(t) - \bar{y}_1^{(-)}(t)\right)\tilde{P}_0^{(-)}(\eta, t) - \int_0^\eta \frac{\tilde{P}^{(-)}(\eta, t)}{\tilde{P}^{(-)}(s, t)}\frac{\tilde{G}_1^{(-)}(s, t)}{\tilde{A}^{(-)}(s, t)}ds. \tag{2.7}$$

Пусть для определенности $\bar{\varphi}_2(t) \le \gamma_0(t) < \bar{\varphi}_3(t)$. Тогда обратим внимание на то обстоятельство, что подынтегральное выражение из (2.7) имеет смысл, вообще говоря, не при всех значениях аргумента, а именно при $\eta = \eta^*$: $T_0^{(-)}(\eta^*, t) = \bar{\varphi}_2(t) - \bar{\varphi}_1(t)$, стоящее в знаменателе $\tilde{A}^{(-)}$, обращается в нуль. Поэтому для обеспечения корректности определения (2.7) (а вместе с тем и самой постановки (2.6)) потребуем, чтобы $\tilde{G}_1^{(-)}(\eta^*, t) = 0$:

$$\begin{gathered}\tilde{G}_1^{(-)}(\eta^*, t) = \frac{\partial \bar{y}_0^{(-)}}{\partial x}\tilde{A} + \left(\frac{\partial \bar{y}_0^{(-)}}{\partial x}(x_1 + \eta^*) + \bar{y}_1^{(-)}\right)\left(-\frac{\tilde{B}}{\tilde{A}}\right)\tilde{A}_y + (x_1 + \eta^*)\left(-\frac{\tilde{B}}{\tilde{A}}\right)\tilde{A}_x + \\ + \left(\frac{\partial \bar{y}_0^{(-)}}{\partial x}(x_1 + \eta^*) + \bar{y}_1^{(-)}\right)\tilde{B}_y + (x_1 + \eta^*)\tilde{B}_x = (x_1 + \eta^*)\frac{-\tilde{B}_y\tilde{A}_x + \tilde{B}_x\tilde{A}_y}{\tilde{A}_y} = 0.\end{gathered}$$

В силу (2.5) и (2.3), последнее уравнение разрешимо и $x_1(t)$ можно считать определенным: $x_1(t) = -\eta^*(\gamma_0, t)$. Подчеркнем, что x_1 определено в том смысле, что мы нашли связь между функциями $x_1(t)$ и $\gamma_0(t)$. Функция же $\gamma_0(t)$ может быть выбрана произвольно (считаем только, что $\bar{\varphi}_1 < \gamma_0 < \bar{\varphi}_3$, и что $\gamma_0(t)$ бесконечно дифференцируема).

Докажем взаимно однозначную зависимость между η^* и γ_0 (а следовательно, и между x_1 и γ_0). Так как $\frac{\partial}{\partial\eta}T_0^{(-)} \neq 0$ (считаем, что $T_0^{(-)} = T_0^{(-)}(\eta, t, \gamma_0)$), то η может быть однозначно выражено через $T_0^{(-)}$, t и γ_0 : $\eta = \eta(T_0^{(-)}, t, \gamma_0)$. Кроме того, из (2.1) получаем

$$\eta = -\int_{\gamma_0}^{\bar{\varphi}_1 + T_0^{(-)}} \frac{A(s, x_0, t)}{B(s, x_0, t)} ds,$$

$$\frac{\partial\eta}{\partial\gamma_0} = \frac{A(\gamma_0, x_0, t)}{B(\gamma_0, x_0, t)} < 0.$$

Значение $\partial\eta/\partial\gamma_0$, таким образом, не зависит от $T_0^{(-)}$ и при всех $T_0^{(-)}$ (а значит, и при $T_0^{(-)} = \bar{\varphi}_2 - \bar{\varphi}_1$) не равно нулю, что и доказывает высказанное утверждение.

Для *всех приближений*, начиная со второго, можно получить

$$\begin{gathered} T_n^{(-)} = \left(\gamma_n(t) - \frac{\partial\bar{y}_0^{(-)}}{\partial x}(t)x_n(t) - \Delta y_n^{(-)}(t)\right)\tilde{P}_0^{(-)}(\eta, t) - \int_0^{\eta} \frac{\tilde{P}^{(-)}(\eta, t)}{\tilde{P}^{(-)}(s, t)} \frac{\tilde{G}_n^{(-)}(s, t)}{\tilde{A}^{(-)}(s, t)} ds, \\ x_n(t) = \frac{\tilde{B}_y(t)}{\tilde{A}_y(t)} \frac{\tilde{A}^{(-)}(0, t)}{\tilde{B}^{(-)}(0, t)} \gamma_{n-1}(t) + I_{kn}(t), \end{gathered} \tag{2.8}$$

где I_{kn} – некоторое известное в данном контексте выражение.

Для переходных членов асимптотики могут быть установлены следующие экспоненциальные оценки (по поводу метода их получения см. [5]):

$$\left|\frac{\partial^{k+l} T_n^{(\mp)}}{\partial\tau^k \partial t^l}\right| < C_n^{k+l}(1 + |\eta|^{2n+l}) \exp\left\{-\frac{\bar{B}_y^{(\mp)}(t)}{\bar{A}^{(\mp)}(t)}\eta\right\}.$$

3. ОБОСНОВАНИЕ АСИМПТОТИКИ

При обосновании асимптотики используем следующие обозначения:

$$\bar{Y}_n^{(-)}(x, t, \varepsilon) = \bar{y}_0^m(x, t) + \ldots + \varepsilon^n \bar{y}_n^m(x, t),$$

$$\bar{Y}_n^{(+)}(x, t, \varepsilon) = \bar{y}_0^p(x, t) + \ldots + \varepsilon^n \bar{y}_n^p(x, t),$$

$$X_n(t, \varepsilon) = x_0(t) + \ldots + \varepsilon^n x_n(t), \quad \eta_n = \frac{x - X_{n+1}(t, \varepsilon)}{\varepsilon},$$

$$T_n^{(-)} Y(\eta_n, t, \varepsilon) = T_0^{(-)}(\eta_n, t) + \ldots + \varepsilon^n T_n^{(-)}(\eta_n, t),$$

$$T_n^{(+)} Y(\eta_n, t, \varepsilon) = T_0^{(+)}(\eta_n, t) + \ldots + \varepsilon^n T_n^{(+)}(\eta_n, t),$$

$$\Pi_n Y(\tau, t, \varepsilon) = \Pi_0(\tau, t) + \ldots + \varepsilon^n \Pi_n(\tau, t),$$

$$Q_n Y(\rho, t, \varepsilon) = Q_0(\rho, t) + \ldots + \varepsilon^n Q_n(\rho, t);$$

C и c – соответственно, достаточно большие и достаточно малые положительные (от ε не зависящие) постоянные.

Доказательство существования КСТС проводится методом дифференциальных неравенств с применением соответствующей теоремы для параболического уравнения, доказанной Аманном в [6]. Изложим необходимые элементы этого метода.

Рассмотрим задачу

$$
\begin{gathered}
Lu \equiv ku_{xx} - u_t - f(u_x, u, x, t) = 0, \\
u(x, t+T) = u(x, t), \quad (x, t) \in \Omega \equiv (0, 1) \times \mathbb{R}, \\
u(0, t) = u^0(t), \quad u(1, t) = u^1(t), \quad t \in \mathbb{R},
\end{gathered} \tag{3.1}
$$

где $k = \text{const} > 0$, f, u^0 и u^1 суть T-периодические по t функции.

Определение 1. *Нижним* и *верхним* (барьерными) решениями задачи (3.1) называются, соответственно, T-периодические по t функции $\alpha = \alpha(x, t)$ и $\beta = \beta(x, t)$, принадлежащие $C^2(\bar{\Omega})$ и такие, что

$$
\begin{gathered}
\alpha(0, t) \le u^0(t) \le \beta(0, t), \quad \alpha(1, t) \le u^1(t) \le \beta(1, t), \ \ t \in \mathbb{R}, \\
L\alpha(x, t) = -\psi_\alpha(x, t) < 0, \quad L\beta(x, t) = +\psi_\beta(x, t) > 0, \quad (x, t) \in \Omega.
\end{gathered}
$$

Обозначим через B область $(\alpha, \beta) \times \Omega$, а через B^* – область $\mathbb{R} \times B$.

Определение 2. Будем говорить, что функция $f(z, u, x, t)$ *принадлежит классу A* в области $B(f \in A(B))$, если $f(z, u, x, t) = O(z^2)$ при $|z| \longrightarrow \infty$, $(u, x, t) \in \bar{B}$.

В [6] была доказана следующая теорема (точнее – ее более общий вариант, где, в частности, в качестве x бралась точка n-мерного пространства).

Теорема 1. *Пусть существуют нижнее* $\alpha(x, t)$ *и верхнее* $\beta(x, t)$ *решения задачи* (3.1), *образующие упорядоченную на* $\bar{\Omega}$ *пару* (α, β). *Далее, пусть функция* $f(z, u, x, t) \in C^1(\bar{B}^*) \cap A(B)$, $u^0(t)$, $u^1(t) \in C^2(\mathbb{R})$.

Тогда существует решение задачи (3.1) *такое, что*

$$
\alpha(x, t) \le u(x, t) \le \beta(x, t).
$$

Верхнее $\beta = \beta_{n+1}(x, t, \varepsilon)$ и нижнее $\alpha = \alpha_{n+1}(x, t, \varepsilon)$ решения уравнений (1.1) формируются на основе построенной асимптотики. Ниже исследуем лишь нижнее решение и левую окрестность линии перехода $[x_l(t), X_{n+2}(t, \varepsilon)]$, $0 < x_l(t) < x_0(t)$:

$$
\alpha_{n+1}(x, t, \varepsilon) = \bar{\alpha} + T\alpha + \Pi\alpha + Q\alpha,
$$

где

$$
\bar{\alpha}(x, t, \varepsilon) = \begin{cases} \bar{Y}^{(-)}_{n+1}(x, t, \varepsilon) + \varepsilon^{n+2}(\bar{y}^m_{n+2}(x, t) - \Upsilon), & (x, t) \in [0, X_{n+2}(t, \varepsilon)] \times \mathbb{R}, \\ \bar{Y}^{(+)}_{n+1}(x, t, \varepsilon) + \varepsilon^{n+2}(\bar{y}^p_{n+2}(x, t) - \Upsilon), & (x, t) \in [X_{n+2}(t, \varepsilon), 1] \times \mathbb{R}, \end{cases}
$$

$$
T_n\alpha(\eta_{n+1}, t, \varepsilon) = \begin{cases} T^{(-)}_n Y(\eta_{n+1}, t, \varepsilon) + \varepsilon^{n+1} T^{(-)}_{(n+1)\alpha}(\eta_{n+1}, t) + \varepsilon^{n+2} T^{(-)}_{(n+2)\alpha}(\eta_{n+1}, t) + \varepsilon^{n+3}\Psi_\alpha(\eta_{n+1}, t, \varepsilon), \\ (\eta_{n+1}, t) \in (-\infty, 0] \times \mathbb{R}, \\ T^{(+)}_n Y(\eta_{n+1}, t, \varepsilon) + \varepsilon^{n+1} T^{(+)}_{(n+1)\alpha}(\eta_{n+1}, t) + \varepsilon^{n+2} T^{(+)}_{(n+2)\alpha}(\eta_{n+1}, t), \\ (\eta_{n+1}, t) \in [0, +\infty) \times \mathbb{R}, \end{cases}
$$

$$
\Pi\alpha(\tau, t, \varepsilon) = \Pi_n Y(\tau, t, \varepsilon) + \varepsilon^{n+1}\Pi_{(n+1)\alpha}(\tau, t),
$$

$$
Q\alpha(\rho, t, \varepsilon) = Q_n Y(\rho, t, \varepsilon) + \varepsilon^{n+1} Q_{(n+1)\alpha}(\rho, t),
$$

где, в свою очередь, $\Upsilon = \text{const} > 0$, а $T^{(-)}_{(n+2)\alpha}$, $T^{(+)}_{(n+2)\alpha}$, $\Pi_{(n+1)\alpha}$, $Q_{(n+1)\alpha}$ определяются из задач, аналогичных задачам для соответствующих членов формальной асимптотики, и отличаются от них добавлением в уравнениях экспоненциально убывающих невязок и заменой $\bar{y}^{m,p}_{n+2}(x_0, t)$ на $\bar{y}^{m,p}_{n+2}(x_0, t) - \Upsilon$.

В частности,

$$\tilde{A}^{(-)}\frac{\partial T^{(-)}_{(n+2)\alpha}}{\partial\eta} + (\tilde{B}^{(-)}_y + \tilde{P}^{(-)}\tilde{A}^{(-)}_y)T^{(-)}_{(n+2)\alpha} + \tilde{G}^{(-)}_{(n+2)\alpha} + \psi_l = 0,$$

$$T^{(-)}_{(n+2)\alpha}(0, t) = \gamma_{n+2}(t) - y^{(-)}_{n+2}(t);$$

здесь $\psi_l(\eta, t)$ – некоторая экспоненциально убывающая функция, опредленная ниже.

Невязка $\psi_l(\eta, t)$ помимо всего прочего определяется из соображений гладкости решения при $\eta_{n+1} = 0$. Она берется в виде

$$\psi_l(\eta, t) = (C_\psi + \Upsilon\bar{B}^{(+)})\exp\left\{-\frac{\bar{B}^{(-)}_y}{\bar{A}^{(-)}}\eta\right\} - C_\psi\exp\{\chi_l\eta\},$$

где $C_\psi > 0$ – некоторая свободная положительная постоянная, а

$$\chi_l(t) = \left(1 + \frac{\Upsilon\bar{B}^{(+)}_y}{C_\psi}\right)\left(-\frac{\bar{B}^{(-)}_y}{\bar{A}^{(-)}}\right) > -\frac{\bar{B}^{(-)}_y}{\bar{A}^{(-)}}.$$

Аналогично, при постановке задачи для $T^{(+)}_{(n+1)\alpha}$ используется невязка

$$\psi_r(\eta, t) = (C_\psi + \Upsilon\bar{B}^{(-)})\exp\left\{-\frac{\bar{B}^{(+)}_y}{\bar{A}^{(+)}}\eta\right\} - C_\psi\exp\{\chi_r\eta\},$$

$$\chi_r(t) = \left(1 + \frac{\Upsilon\bar{B}^{(-)}_y}{C_\psi}\right)\left(-\frac{\bar{B}^{(+)}_y}{\bar{A}^{(+)}}\right) > -\frac{\bar{B}^{(+)}_y}{\bar{A}^{(+)}}.$$

Подобный выбор ψ_l и ψ_r связан с тем, что

$$\left.\frac{\partial}{\partial\eta}(T^{(-)}_{(n+2)\alpha} + y^{(-)}_{n+2}) - \frac{\partial}{\partial\eta}(T^{(+)}_{(n+2)\alpha} + y^{(+)}_{n+2})\right|_{\eta=0} = \left.\frac{\Upsilon\bar{B}^{(+)}_y + \psi_r}{\tilde{A}^{(+)}} - \frac{\Upsilon\bar{B}^{(-)}_y + \psi_l}{\tilde{A}^{(-)}}\right|_{\eta=0},$$

$$\left.\frac{\partial^2}{\partial\eta^2}(T^{(-)}_{(n+2)\alpha} + y^{(-)}_{n+2}) - \frac{\partial^2}{\partial\eta^2}(T^{(+)}_{(n+2)\alpha} + y^{(+)}_{n+2})\right|_{\eta=0} = \left.\frac{1}{\tilde{A}^{(+)}}\frac{\partial\psi_r}{\partial\eta} - \frac{1}{\tilde{A}^{(-)}}\frac{\partial\psi_l}{\partial\eta}\right|_{\eta=0},$$

а следовательно, с осуществлением требования, чтобы выполнялись условия

$$\Upsilon\bar{B}^{(-)}_y + \psi_l = \Upsilon\bar{B}^{(+)}_y + \psi_r,$$

$$\frac{\partial\psi_l}{\partial\eta}(0, t) = \frac{\partial\psi_r}{\partial\eta}(0, t).$$

Индекс α у $T^{(\mp)}_{(n+1)\alpha}$ связан с тем, что значение γ_{n+1} зависит от того, какое барьерное решение рассматривается (для нижнего обозначим его через $\gamma_{(n+1)\alpha}$, для верхнего – через $\gamma_{(n+1)\beta}$). Линию перехода $X_{n+2}(t, \varepsilon)$ считаем фиксированной.

Слагаемое $\Psi_\alpha(\eta_{n+1}, t, \varepsilon)$ в представлении $T\alpha$ необходимо для обеспечения гладкости в переходной точке. Оно имеет вид

$$\Psi_\alpha(\eta, t, \varepsilon) = C_1(t, \varepsilon)e^{\chi_1\eta} + C_2(t, \varepsilon)e^{\chi_2\eta} + C_3(t, \varepsilon)e^{\chi_3\eta},$$

где $-\bar{B}^{(-)}_y/\bar{A}^{(-)} < \chi_1 < \chi_2 < \chi_3$, а $C_1(t, \varepsilon)$, $C_2(t, \varepsilon)$, $C_3(t, \varepsilon)$ находятся из системы уравнений

$$C_1 + C_2 + C_3 = \alpha_{n+1}(X_{n+2}(t, \varepsilon) + 0, t, \varepsilon) - \alpha_{n+1}(X_{n+2}(t, \varepsilon) - 0, t, \varepsilon),$$

$$\chi_1 C_1 + \chi_2 C_2 + \chi_3 C_3 = \frac{\partial \alpha_{n+1}}{\partial x}(X_{n+2}(t, \varepsilon) + 0, t, \varepsilon) - \frac{\partial \alpha_{n+1}}{\partial x}(X_{n+2}(t, \varepsilon) - 0, t, \varepsilon),$$

$$\chi_1^2 C_1 + \chi_2^2 C_2 + \chi_3^2 C_3 = \frac{\partial^2 \alpha_{n+1}}{\partial x^2}(X_{n+2}(t, \varepsilon) + 0, t, \varepsilon) - \frac{\partial^2 \alpha_{n+1}}{\partial x^2}(X_{n+2}(t, \varepsilon) - 0, t, \varepsilon).$$

В силу того, что ее определитель $\chi_1^2(\chi_3 - \chi_2) + \chi_2^2(\chi_1 - \chi_3) + \chi_3^2(\chi_2 - \chi_1) > 0$ (так как $\chi_1 \neq \chi_2 \neq \chi_3$), она однозначно разрешима.

Пусть

$$L_\varepsilon y(x, t, \varepsilon) = L_\varepsilon[y(x, t, \varepsilon)] = \varepsilon^4 \frac{\partial^2 y}{\partial x^2} - \varepsilon \frac{\partial y}{\partial t} - \varepsilon \frac{\partial y}{\partial x} A(y, x, t) - B(y, x, t).$$

Покажем положительность $L_\varepsilon \alpha$.

С точностью до экспоненциально бесконечно малых имеем $L_\varepsilon \alpha_{n+1} = \overline{L\alpha} + TL\alpha$, где

$$\overline{L\alpha}(x, t, \varepsilon) \equiv L_\varepsilon \bar{\alpha}(x, t, \varepsilon),$$

$$TL\alpha(\eta_{n+1}, t, \varepsilon) \equiv L_\varepsilon(\bar{\alpha}(X_{n+2}(t, \varepsilon) + \eta_{n+1}\varepsilon, t, \varepsilon) + T\alpha(\eta_{n+1}, t, \varepsilon)) - L_\varepsilon \bar{\alpha}(X_{n+2}(t, \varepsilon) + \eta_{n+1}\varepsilon, t, \varepsilon).$$

Несложно убедиться, что справедливо неравенство

$$\overline{L\alpha}(x, t, \varepsilon) = \varepsilon^{n+2} \Upsilon B_y(\bar{y}_0^m(x, t), x, t) + O(\varepsilon^{n+3}) > 0.$$

Далее $TL\alpha$ представим в виде

$$\begin{gathered} TL\alpha(\eta_{n+1}, t, \varepsilon) = T_0 L\alpha(\eta_{n+1}, t) + \ldots + \varepsilon^{n+2} T_{n+2} L\alpha(\eta_{n+1}, t) + \varepsilon^{n+3} R(\eta_{n+1}, t, \varepsilon) = \\ = \varepsilon^{n+2} \psi_l(\eta_{n+1}, t) + \varepsilon^{n+3} R(\eta_{n+1}, t, \varepsilon). \end{gathered} \tag{3.2}$$

Для $R(\eta_{n+1}, t, \varepsilon)$ может быть получена оценка

$$|R| < C(1 + |\eta_{n+1}|^{3n+7}) \exp\{-(\bar{B}_y^{(-)}/\bar{A}^{(-)})\eta_{n+1}\}.$$

Отсюда и из (3.2) видно, что $TL\alpha > 0$ при $|\eta_{n+1}| < c/\sqrt[3n+7]{\varepsilon}$. При бо́льших значениях $|\eta_{n+1}|$ величина $TL\alpha$ пренебрежимо мала. Значит, $L_\varepsilon \alpha_{n+1} > 0$.

Покажем положительность разности верхнего и нижнего решений:

$$\begin{gathered} \beta_{n+1}(x, t, \varepsilon) - \alpha_{n+1}(x, t, \varepsilon) = \\ = 2\Upsilon\varepsilon^{n+2} + (T^{(-)}_{(n+1)\beta} - T^{(-)}_{(n+1)\alpha})\varepsilon^{n+1} + (T^{(-)}_{(n+2)\beta} - T^{(-)}_{(n+2)\alpha})\varepsilon^{n+2} + (\Psi_\beta - \Psi_\alpha)\varepsilon^{n+3}. \end{gathered} \tag{3.3}$$

Рассмотрим первую скобку:

$$T^{(-)}_{(n+1)\beta} - T^{(-)}_{(n+1)\alpha} = (\gamma_{(n+1)\beta} - \gamma_{(n+1)\alpha})\tilde{P}_0^{(-)}(\eta_{n+1}).$$

Для возможности построения барьерных решений необходимо выполнение условия (2.8) (с учетом вариаций, вносимых в барьерные решения), из которого получаем

$$\gamma_{(n+1)\beta} - \gamma_{(n+1)\alpha} = \frac{2\Delta(\eta^*_{n+1}, t)}{\tilde{P}_0^{(-)}(\eta^*_{n+1}, t)},$$

где

$$\Delta(\eta) = \begin{cases} -[\gamma \bar{B}_y^{(-)} + \psi_l(\eta, t)]\left[\tilde{P}^{(-)}(\eta, t)\dfrac{d}{dx_0}\tilde{A}\right]^{-1}, & \eta \le 0, \\ -[\gamma \bar{B}_y^{(+)} + \psi_r(\eta, t)]\left[\tilde{P}^{(+)}(\eta, t)\dfrac{d}{dx_0}\tilde{A}\right]^{-1}, & \eta \ge 0, \end{cases}$$

а η^*_{n+1} означает величину η_{n+1}, при которой $T_0^{(-)} = \varphi_2 - \varphi_1$.

Таким образом, для первой разности из (3.3) получаем

$$T^{(-)}_{(n+1)\beta} - T^{(-)}_{(n+1)\alpha} = \frac{2\Delta(\eta^*_{n+1}, t)}{\tilde{P}^{(-)}(\eta^*_{n+1}, t)}\tilde{P}^{(-)}(\eta_{n+1}, t) > c\exp\{-(\bar{B}_y^{(-)}/\bar{A}^{(-)})\eta_{n+1}\}.$$

Оставшиеся в (3.3) разности по абсолютной величине не превосходят $\varepsilon^{n+2}C(1 + |\eta_{n+1}|^{2n+4})\exp\{-(\bar{B}_y^{(-)}/\bar{A}^{(-)})\eta_{n+1}\}$, так что при $|\eta_{n+1}| < c/\sqrt[2n+4]{\varepsilon}$ они по модулю меньше первой скобки. При бо́льших $|\eta_{n+1}|$ все скобки перекрываются первым слагаемым. Значит, $\beta - \alpha > 0$.

Теперь, используя теорему Аманна, приходим к теореме существования и асимптотического представления решения задачи (1.1)

Теорема 2. *Пусть выполнен ряд требований*:

1) *вырожденное уравнение* $B(y, x, t) = 0$ *имеет корни* $y = \varphi_1(x, t)$, $(x, t) \in [0, x_r(t)] \times \mathbb{R}$, $y = \varphi_2(x, t)$, $(x, t) \in [x_l(t), x_r(t)] \times \mathbb{R}$, $y = \varphi_3(x, t)$, $(x, t) \in [x_l(t), 1] \times \mathbb{R}$, $0 \le x_l(t) < x_r(t) \le 1$, *удовлетворяющие условиям*

$$\varphi_1(x, t) < \varphi_2(x, t) < \varphi_3(x, t), \quad (x, t) \in [x_l(t), x_r(t)] \times \mathbb{R},$$

$$B_y(\varphi_1(x, t), x, t) > 0, \quad (x, t) \in [0, x_r(t)] \times \mathbb{R},$$

$$B_y(\varphi_2(x, t), x, t) < 0, \quad (x, t) \in [x_l(t), x_r(t)] \times \mathbb{R},$$

$$B_y(\varphi_3(x, t), x, t) > 0, \quad (x, t) \in [x_l(t), 1] \times \mathbb{R},$$

$$A(\varphi_1(0, t), 0, t) < 0, \quad A(\varphi_3(1, t), 1, t) < 0, \quad t \in \mathbb{R};$$

2) *переходное уравнение* (2.2) *имеет корень* $x_0 = x_0(t)$, *удовлетворяющий условиям* (2.3), (2.4) *и* (2.5); *кроме того*,

$$(x_0, t) \in (x_l(t), x_r(t)) \times \mathbb{R},$$

$$A(\bar{\varphi}_1(t), x_0(t), t) < 0, \quad A(\bar{\varphi}_3(t), x_0(t), t) > 0;$$

3) *A и B являются достаточно гладкими функциями своих аргументов*

$$A(y, x, t) \in C^{n+6}(\mathbb{R} \times \Omega), \quad B(y, x, t) \in C^{n+4}(\mathbb{R} \times \Omega);$$

4) *граничные функции* $y^0(t)$ *и* $y^1(t)$ *принадлежат областям влияния корней вырожденного уравнения* $y = \varphi_1(x, t)$ *и* $y = \varphi_3(x, t)$ *соответственно*.

Тогда при достаточно малых ε $(0 < \varepsilon \le \varepsilon_0)$ *существует решение* $y = y(x, \varepsilon)$ *задачи* (1.1), *для которого справедливо следующее асимптотическое представление*:

$$y = \begin{cases} \bar{Y}_n^{(-)}(x, t, \varepsilon) + \Pi_n Y\left(\dfrac{x}{\varepsilon^3}, t, \varepsilon\right) + T_n^{(-)}Y\left(\dfrac{x - X_{n+1}}{\varepsilon}, t, \varepsilon\right) + O(\varepsilon^{n+1}), & 0 \le x \le X_{n+1}(t, \varepsilon), \\ \bar{Y}_n^{(+)}(x, t, \varepsilon) + Q_n Y\left(\dfrac{x - 1}{\varepsilon}, t, \varepsilon\right) + T_n^{(+)}Y\left(\dfrac{x - X_{n+1}}{\varepsilon}, t, \varepsilon\right) + O(\varepsilon^{n+1}), & X_{n+1}(t, \varepsilon) \le x \le 1. \end{cases} \tag{3.4}$$

4. ОЦЕНКА ПРОИЗВОДНОЙ ПО ПРОСТРАНСТВЕННОМУ АРГУМЕНТУ

Вновь рассмотрим задачу (3.1).

Обозначим через B^M область $(\alpha_x - M, \beta_x + M) \times B$, где $\alpha(x, t)$ и $\beta(x, t)$ – соответственно, нижнее и верхнее решения (3.1), и потребуем, чтобы выполнялись неравенства

$$\|\beta - \alpha\|_{C(\bar{\Omega})} < l, \quad \|\psi_\alpha\|_{C(\bar{\Omega})} < \Delta_\alpha, \quad \|\psi_\beta\|_{C(\bar{\Omega})} < \Delta_\beta, \quad \|f_z\|_{C(\bar{B}^M)} < f_1(M), \quad \|f_y\|_{C(\bar{B}^M)} < f_2(M).$$

Введем

Определение 3. Будем говорить, что функция $f(z, y, x, t)$ *принадлежит классу* A1 в области B^M $(f \in \mathrm{A1}(B^M))$, если существует такое положительное число M, что верны неравенства

$$M > 4l/(b - a),$$

$$M > 2lf_1 + \sqrt{4(lf_1)^2 + 4l(f_2 l + \Delta_\alpha + \Delta_\beta) + 2l^2}.$$

Для оценки производной $\partial y/\partial x$ используется (см. [7]) следующая

Теорема 3. *Пусть выполнены требования теоремы* 1 *с тем изменением, что функция* $f(z, y, x, t) \in C^1(\bar{B}^M) \cap \mathrm{A}1(B^M)$.

Тогда существует решение задачи (3.1) *такое, что*

$$\alpha(x, t) \le y(x, t) \le \beta(x, t),$$

причем для каждого такого решения справедливо неравенство

$$\alpha_x - M < y_x < \beta_x + M.$$

Применяя теорему 3 к задаче (1.1), несложно убедиться в том, что для производной по x любого решения, заключенного между построенными в предыдущем разделе барьерами α_{n+1} и β_{n+1}, справедливо следующее представление:

$$\frac{\partial y}{\partial x} = \begin{cases} \dfrac{\partial \bar{Y}^{(-)}_{n-3}}{\partial x}(x, t, \varepsilon) + \varepsilon^{-3}\dfrac{\partial \Pi_n Y}{\partial \tau}\left(\dfrac{x}{\varepsilon^3}, t, \varepsilon\right) + \varepsilon^{-1}\dfrac{\partial T^{(-)}_{n-2}Y}{\partial \eta}\left(\dfrac{x - X_{n-1}}{\varepsilon}, t, \varepsilon\right) + O(\varepsilon^{n-2}), \ 0 \le x \le X_{n+1}(t, \varepsilon), \\ \dfrac{\partial \bar{Y}^{(+)}_{n-3}}{\partial x}(x, t, \varepsilon) + \varepsilon^{-1}\dfrac{\partial Q_{n-2} Y}{\partial \rho}\left(\dfrac{x-1}{\varepsilon}, t, \varepsilon\right) + \varepsilon^{-1}\dfrac{\partial T^{(+)}_{n-2}Y}{\partial \eta}\left(\dfrac{x - X_{n-1}}{\varepsilon}, t, \varepsilon\right) + O(\varepsilon^{n-2}), \ X_{n+1}(t, \varepsilon) \le x \le 1. \end{cases} \tag{4.1}$$

5. ИССЛЕДОВАНИЕ НА УСТОЙЧИВОСТЬ

Любое решение задачи (3.1) может рассматриваться как стационарное решение задачи вида

$$\begin{gathered} Lu = 0, \quad (x, t) \in \Omega^+, \\ u(0, t) = u^0(t), \quad u(1, t) = u^1(t), \quad t \in \mathbb{R}^+, \\ u(x, 0) = u_0(x), \quad x \in [0, 1]. \end{gathered} \tag{5.1}$$

Пусть $D \equiv \{u_0(x) \mid 0 < x < 1\}$ – произвольное подмножество функций из $C^2[0, 1]$, удовлетворяющих краевым условиям

$$u_0(0) = u^0(0), \quad u_0(1) = u^1(0). \tag{5.2}$$

Введем следующие определения (см. также [8]).

Определение 4. Решение $u = u_s(x, t)$ задачи (3.1) называется *устойчивым по Ляпунову на множестве* D, если для любого $\varepsilon > 0$ существует $\delta = \delta(\varepsilon)$ такое, что для каждой функции $u_0(x) \in D$: $\max\limits_{x \in [0, 1]} |u_s(x, 0) - u_0(x)| \le \delta(\varepsilon)$ существует решение $u = u(x, t)$ задачи (5.1), причем для каждого такого решения

$$\max_{(x, t) \in \bar{\Omega}^+} |u(x, t) - u_s(x, t)| \le \varepsilon.$$

Определение 5. Решение $u = u_s(x, t)$ задачи (3.1) называется *асимптотически устойчивым на* D, если оно устойчиво на D и если для любой функции $u_0(x) \in D$ существует решение $u = u(x, t)$ задачи (5.1), причем для каждого такого решения

$$\lim_{t \to +\infty} \max_{x \in [0, 1]} |u(x, t) - u_s(x, t)| = 0.$$

При этом множество D называется областью влияния решения $u_s(x, t)$.

Определение 6. Решение $u = u_s(x, t)$ задачи (3.1) называется *экспоненциально устойчивым на* D, если оно асимптотически устойчиво на D и если существуют постоянные $C > 0$ и $\tau > 0$ такие, что для любой функции $u_0(x) \in D$ отвечающие ей решения задачи (5.1) удовлетворяют неравенству

$$\max_{x \in [0, 1]} |u(x, t) - u_s(x, t)| \le C\exp(-t/\tau). \tag{5.3}$$

Точная нижняя грань значений τ, при которых справедлива оценка (5.3) (при этом, вообще говоря, $C = C(\tau)$), называется *временем выхода на стационарный режим*.

Определение 7. *Нижним и верхним* (барьерными) *решениями задачи* (5.1) называются, соответственно, функции $\alpha = \alpha(x, t)$ и $\beta = \beta(x, t)$, принадлежащие $C^2(\overline{\Omega}^+)$ и такие, что выполняется следующее:

$$1)\ \alpha(0, t) \le u^0(t) \le \beta(0, t), \quad \alpha(1, t) \le u^1(t) \le \beta(1, t), \quad t \in \mathbb{R}^+;$$

$$2)\ L\alpha(x, t) \ge 0, \quad L\beta(x, t) \le 0, \quad (x, t) \in \Omega^+.$$

Докажем экспоненциальную устойчивость решения задачи (1.1), рассмотренного в предыдущих пунктах, опираясь на следующую теорему (более общий ее вариант см. в [6]).

Теорема 4. *Пусть выполнены все условия теоремы* 1 *с тем изменением, что* α *и* β *являются, соответственно, нижним и верхним решениями задачи* (5.1), *причем достаточно, чтобы* α *и* β *образовывали упорядоченную пару лишь при* $t = 0$.

Тогда для любой начальной функции $u_0(x) \in C^2[0, 1]$, *удовлетворяющей условиям* (5.2) *и такой, что* $\alpha(x, 0) \le u_0(x) \le \beta(x, 0)$, *существует единственное решение задачи* (5.1), *причем* $\alpha(x, t) \le \le u(x, t) \le \beta(x, t)$.

Обозначим решение (3.4) задачи (1.1) через $y_s(x, t, \varepsilon)$. Будем строить барьеры задачи (1.3) путем специальной вариации (вычитания и прибавления некоторых функций) стационарного решения $y_s(x, t, \varepsilon)$, а именно выберем нижнее и верхнее решения, соответственно, в виде

$$\alpha(x, t, \varepsilon) = y_s(x, t, \varepsilon) - \frac{N}{\varepsilon} r_\alpha(x, t, \varepsilon) E_K(t, \varepsilon),$$

$$\beta(x, t, \varepsilon) = y_s(x, t, \varepsilon) + \frac{N}{\varepsilon} r_\beta(x, t, \varepsilon) E_K(t, \varepsilon),$$

где

$$r_\alpha(x, t, \varepsilon) = y_s(x, t, \varepsilon) - \alpha_4(x, t, \varepsilon),$$

$$r_\beta(x, t, \varepsilon) = \beta_4(x, t, \varepsilon) - y_s(x, t, \varepsilon),$$

$$E_K(t, \varepsilon) = \exp(-Kt/\varepsilon^3),$$

N, K – некоторые положительные постоянные.

Поскольку выполнение п. 1) определения 7 не вызывает сомнений, остановимся на доказательстве п. 2). При этом ограничимся рассмотрением нижнего решения:

$$L_\varepsilon\alpha = L_\varepsilon\Big[y_s - \frac{N}{\varepsilon} r_\alpha E_K\Big] = \varepsilon^4\Big\{(y_s)_{xx} - (y_s)_t - \frac{N}{\varepsilon}[(r_\alpha)_{xx} - (r_\alpha)_t]E_K - \frac{N}{\varepsilon} r_\alpha \frac{K}{\varepsilon^3} E_K\Big\} -$$

$$- \varepsilon\frac{\partial}{\partial x}\Big[y_s - \frac{N}{\varepsilon} r_\alpha E_K\Big] A\Big(y_s - \frac{N}{\varepsilon} r_\alpha E_K, x, t\Big) - B\Big(y_s - \frac{N}{\varepsilon} r_\alpha E_K, x, t\Big) =$$

$$= \varepsilon^4\Big\{(y_s)_{xx} - (y_s)_t - \frac{N}{\varepsilon}[(r_\alpha)_{xx} - (r_\alpha)_t]E_K - \frac{NK}{\varepsilon^4} r_\alpha E_K\Big\} -$$

$$- \varepsilon\Big[(y_s)_x - \frac{N}{\varepsilon}(r_\alpha)_x E_K\Big]\Big\{A(y_s, x, t) - A_y(y_s, x, t)\frac{N}{\varepsilon} r_\alpha E_K + \frac{1}{2} A_{yy}(y_s^*, x, t)\frac{N^2}{\varepsilon^2} r_\alpha^2 E_{2K}\Big\} -$$

$$- \Big\{B(y_s, x, t) - B_y(y_s, x, t)\frac{N}{\varepsilon} r_\alpha E_K + \frac{1}{2} B_{yy}(y_s^*, x, t)\frac{N^2}{\varepsilon^2} r_\alpha^2 E_{2K}\Big\},$$

где y_s^* означает функцию $y_s^*(x, t, \varepsilon)$ такую, что $\alpha \le y_s^* \le y_s$.

Вспоминая, что $Ly_s \equiv 0$, получаем

$$L_\varepsilon\alpha = \frac{N}{\varepsilon}E_K\Big\{-\varepsilon^4[(r_\alpha)_{xx}-(r_\alpha)_t]-\varepsilon K r_\alpha+\varepsilon(r_\alpha)_x A(y_s, x, t)+$$

$$+\varepsilon\left[(y_s)_x-\frac{N}{\varepsilon}(r_\alpha)_x E_K\right]\left[A_y(y_s, x, t)r_\alpha-\frac{1}{2}A_{yy}(y_s^*, x, t)\frac{N}{\varepsilon}r_\alpha^2 E_K\right]+B_y(y_s, x, t)r_\alpha-\frac{1}{2}B_{yy}(y_s^*, x, t)\frac{N}{\varepsilon}r_\alpha^2 E_K\Big\}.$$

Далее, учитывая определение r_α, для первого слагаемого в фигурной скобке имеем

$$-\varepsilon^4[(r_\alpha)_{xx}-(r_\alpha)_t] = L_\varepsilon\alpha_4-\varepsilon(y_s)_x A(y_s, x, t)+\varepsilon(\alpha_4)_x A(\alpha_4, x, t)-B(y_s, x, t)+B(\alpha_4, x, t) = L_\varepsilon\alpha_4-$$

$$-\varepsilon(\alpha_4)_x A_y(y_s, x, t)r_\alpha+\frac{1}{2}\varepsilon(\alpha_4)_x A_{yy}(y_s^{**}, x, t)r_\alpha^2-\varepsilon(r_\alpha)_x A(y_s, x, t)-B_y(y_s, x, t)r_\alpha+\frac{1}{2}B_{yy}(y_s^{**}, x, t)r_\alpha^2,$$

где y_s^{**} обладает тем же смыслом, что и y_s^*.

Подставляя последнее выражение в соотношение для $L_\varepsilon\alpha$ и приводя подобные члены, приходим к следующему представлению:

$$L_\varepsilon\alpha = \frac{N}{\varepsilon}E_K\Big\{-\varepsilon K r_\alpha+L_\varepsilon\alpha_4+\varepsilon(r_\alpha)_x A_y(y_s, x, t)r_\alpha-N(r_\alpha)_x A_y(y_s, x, t)r_\alpha E_K-\frac{1}{2}N(y_s)_x A_{yy}(y_s^*, x, t)r_\alpha^2 E_K+$$

$$+\frac{1}{2}\frac{N^2}{\varepsilon}(r_\alpha)_x A_{yy}(y_s^*, x, t)r_\alpha^2 E_{2K}-\frac{1}{2}\frac{N}{\varepsilon}B_{yy}(y_s^*, x, t)r_\alpha^2 E_K+\frac{1}{2}\varepsilon(\alpha_4)_x A_{yy}(y_s^{**}, x, t)r_\alpha^2+\frac{1}{2}B_{yy}(y_s^{**}, x, t)r_\alpha^2\Big\}.$$

Обратим внимание на то, что частные производные коэффициентов A и B (при достаточной гладкости этих функций) в точках (y_s^*, x, t), (y_s^{**}, x, t) суть ограниченные функции. Кроме того, при $(x, t) \in \Omega_1 \equiv [\Delta, X_7(t, \varepsilon)] \times \mathbb{R}^+$ (Δ – некоторая постоянная: $0 < \Delta < x_0(t)$) справедливо

$$(\alpha_4)_x = O\left(\frac{1}{\varepsilon}\right), \quad r_\alpha = O(\varepsilon^4), \quad L_\varepsilon\alpha_4 \ge c(\varepsilon^5) \text{ по построению,}$$

$$(r_\alpha)_x = O(\varepsilon) \text{ в силу представления (4.1),}$$

$$(y_s)_x = (\alpha_4)_x+(r_\alpha)_x = O\left(\frac{1}{\varepsilon}\right).$$

Аналогичные оценки имеют место и при остальных значениях (x, t). Таким образом, при $x \in \Omega_1$ для $L_\varepsilon\alpha$ окончательно получим

$$L_\varepsilon\alpha \ge \frac{N}{\varepsilon}E_K\{KO(\varepsilon^5)+cO(\varepsilon^5)+O(\varepsilon^6)+NO(\varepsilon^5)+NO(\varepsilon^7)+N^2O(\varepsilon^8)+NO(\varepsilon^7)+O(\varepsilon^8)+O(\varepsilon^8)\}.$$

Теперь, выбирая K и N достаточно малыми, несложно убедиться в том, что $L_\varepsilon\alpha > 0$. Доказательство положительности $L_\varepsilon\alpha$ при всех оставшихся значениях переменных x и t проводится полностью аналогично.

Следовательно, опираясь на теорему 4, делаем вывод, что для любой начальной функции $y_0(x) \in C^2[0, 1]$, удовлетворяющей краевым условиям задачи (1.3) и неравенству $\alpha(x, 0, \varepsilon) \le y_0(x) \le \beta(x, 0, \varepsilon)$, задача (1.3) имеет единственное решение $y = y(x, t, \varepsilon)$, причем

$$\alpha(x, t, \varepsilon) \le y(x) \le \beta(x, t, \varepsilon).$$

Так как $\beta-\alpha = \frac{N}{\varepsilon}(r_\beta-r_\alpha)\exp(-Kt/\varepsilon^3)$, то решение $y_s(x, t, \varepsilon)$ задачи (1.1) экспоненциально устойчиво. При этом (учитывая также, что $\beta(x, 0, \varepsilon)-\alpha(x, 0, \varepsilon) = \frac{N}{\varepsilon}[\beta_4(x, 0, \varepsilon)-\alpha_4(x, 0, \varepsilon)] \ge c\varepsilon^4$) ширина области влияния имеет порядок ε^4, а время выхода – порядок ε^3.

Напомним, что из асимптотической устойчивости стационарного решения (в том виде, как это понятие было сформулировано выше) с областью влияния D непосредственно следует его единственность в этой области.

Таким образом, справедлива

Теорема 5. *Пусть выполнены все условия теоремы* 2 *при* $n = 3$. *Тогда при достаточно малых* ε *решение* $y_s(x, t, \varepsilon)$ *задачи* (1.1), *заключенное между барьерами* α_4 *и* β_4, *единственно и экспоненциально устойчиво на множестве*

$$D = \{y_0(x) \mid y_0(x) \in C^2[0, 1], \max_{x \in [0, 1]} |y_0(x) - y_s(x, 0, \varepsilon)| \le c\varepsilon^4, y_0(0) = y^0(0), y_0(1) = y^1(0)\}.$$

При этом время выхода на стационарное решение не превосходит $C\varepsilon^3$.

Замечания. 1. Отметим без доказательства, что в случае, когда отсутствуют внутренние переходные слои, область влияния может быть оценена снизу конечной (не зависящей от ε) постоянной. Верхняя же оценка времени выхода остается прежней.

2. Полученные оценки ширины области влияния и времени выхода являются характерными для рассматриваемого типа задач при условии использования метода барьерных функций в том виде, в котором он был впервые предложен в [9]. Это отнюдь не означает, что они не могут быть улучшены, и вопрос об их уточнении остается открытым (см. также [10], [11]).

Например, в [12] рассматривалось уравнение

$$\mu(y_{xx} - y_t) = y_x A(y, x, t) + B(y, x, t)$$

при дополнительных условиях задачи (1.1). Домножая обе части этого уравнения на μ, получаем

$$\mu^2(y_{xx} - y_t) = \mu y_x A(y, x, t) + \mu B(y, x, t). \tag{5.4}$$

В [12] была построена и обоснована асимптотика решения с внутренним переходным слоем. Если применить по отношению к (5.4) приведенные выше рассуждения, то можно показать экспоненциальную устойчивость построенной контрастной структуры. Причем область влияния будет иметь порядок μ^2, а время выхода – порядок μ.

Вернемся к уравнению (1.1). Домножая обе его части на ε^2 и производя замену $\varepsilon^3 = \mu$, придем к уравнению

$$\mu^2(y_{xx} - y_t) = \mu y_x A(y, x, t) + \mu^{2/3} B(y, x, t).$$

Таким образом, указанная в теореме 5 оценка времени выхода совпадает с таковой для случая (5.4), если принять во внимание, что характерным параметром задачи (1.1) является $\mu = \varepsilon^3$. Что касается ширины области влияния, то ее оценка даже несколько увеличивается, хотя главная особенность – то, что ширина области влияния есть $o(\mu)$, – остается.

СПИСОК ЛИТЕРАТУРЫ

1. *Букжалёв Е.Е.* Контрастные структуры типа ступеньки, растянутые переменные которых зависят от различных степеней параметра возмущения // Ж. вычисл. матем. и матем. физ. 2004. Т. 44. № 4. С. 662–675.

2. *Васильева А.Б., Бутузов В.Ф.* Асимптотические методы в теории сингулярных возмущений. М.: Высш. школа, 1990.

3. *Васильева А.Б., Бутузов В.Ф., Нефёдов Н.Н.* Контрастные структуры в сингулярно возмущенных задачах // Фундамент. и прикл. матем. 1998. Т. 4. № 3. С. 799–851.

4. *Васильева А.Б., Давыдова М.А.* Сингулярно возмущенное уравнение второго порядка с малыми параметрами при первой и второй производных // Ж. вычисл. матем. и матем. физ. 1999. Т. 39. № 9. С. 1504–1512.

5. *Букжалёв Е.Е.* Сингулярно возмущенное уравнение с погранслойным решением, растянутые переменные которого зависят от различных степеней параметра возмущения // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 12. С. 1775–1785.

6. *Amann H.* Periodic solutions of semi-linear parabolic equations // Nonlinear Analysis: A Collection of Papers in Honor of Erich Rothe. New York: Acad. Press, 1978. P. 1–29.

7. *Букжалёв Е.Е.* О применении метода дифференциальных неравенств к уравнениям параболического типа, правая часть которых растет по пространственному градиенту более чем квадратично // Дифференц. ур-ния. 2005. Т. 41. № 3. С. 356–365.
8. *Колесов Ю.С.* О некоторых признаках существования устойчивых периодических решений у квазилинейных параболических уравнений // Докл. АН СССР. 1964. Т. 157. № 6. С. 1288–1290.
9. *Бутузов В.Ф., Неделько И.В.* Устойчивость контрастных структур типа ступеньки в двумерном случае // Докл. РАН. 1999. Т. 366. № 3. С. 1–4.
10. *Бутузов В.Ф., Неделько И.В.* О глобальной области влияния решений с внутренними слоями // Докл. РАН. 2000. Т. 373. № 2. С. 155–156.
11. *Бутузов В.Ф., Неделько И.В.* О глобальной области влияния устойчивых решений с внутренними слоями в двумерном случае // Изв. РАН. Сер. матем. 2002. Т. 66. № 1. С. 3–42.
12. *Васильева А.Б., Омельченко О.Е.* Периодические контрастные структуры типа ступеньки для сингулярно возмущенного параболического уравнения // Дифференц. ур-ния. 2000. Т. 36. № 2. С. 198–208.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 438–444

УДК 519.63

АСИМПТОТИКА РЕШЕНИЯ СИНГУЛЯРНО ВОЗМУЩЕННОЙ СИСТЕМЫ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ПЕРВОГО ПОРЯДКА В ЧАСТНЫХ ПРОИЗВОДНЫХ С МАЛОЙ НЕЛИНЕЙНОСТЬЮ В КРИТИЧЕСКОМ СЛУЧАЕ

© 2007 г. А. В. Нестеров, О. В. Шулико

(249040 Обнинск, ул. Студгородок, 1, ИАТЭ)

e-mail: andrenesterov@yandex.ru, olga-shuliko@yandex.ru

Поступила в редакцию 28.12.2005 г.

Построено формальное асимптотическое представление решения начально-краевой задачи для сингулярно возмущенной системы дифференциальных уравнений первого порядка в частных производных с малой нелинейностью. Библ. 2.

Ключевые слова: малый параметр, сингулярные возмущения, дифференциальные уравнения в частных производных, асимптотическое представление решений.

1. ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Строится формальное асимптотическое представление решения начально-краевой задачи для сингулярно возмущенной системы дифференциальных уравнений первого порядка в частных производных с малой нелинейностью вида

$$\varepsilon^2(U_t + DU_x) = AU + \varepsilon F(U), \quad x, t \subset \Omega = \{0 < t < T, 0 < x < \infty\}, \tag{1а}$$

$$U(0, x) = U^0(x), \quad U(t, 0) = \Phi^0(t), \tag{1б}$$

где $U(x, t) = \{u_i(x, t)\}$, $1 \le i \le n$, – вектор решений, D, A – квадратные n-мерные матрицы, матрица A вырожденная, $F(U) = \{f_i(U), 1 \le i \le n\}$ – вектор-функция, определенная для всех $|u_i| < \infty$, $1 \le i \le n$.

Целью настоящей работы является построение формального асимптотического представления решения начально-краевой задачи с точностью $O(\varepsilon)$ вне малой окрестности начала координат.

Аналогичная задача для линейных уравнений (что эквивалентно равенству $F(U) = 0$) рассматривалась ранее в [1].

В настоящей статье рассматривается случай, когда элементы матриц A, D постоянны и вещественны, матрица $D = \mathrm{diag}\{d_{ii}\}$, $d_{ii} > 0$, диагональная.

Введем следующие обозначения: h_0 – собственный вектор матрицы A, соответствующий нулевому собственному значению $\lambda = 0$; h_0^* – собственный вектор транспонированной матрицы A^*, соответствующий собственному значению $\lambda^* = 0$; скалярное произведение векторов a и b будем обозначать стандартно через (a, b). Будем считать, что выполнены следующие условия.

1. $U^0(x)$, $\Phi^0(t)$, а также элементы n-мерной вектор-функции $F(U)$ дважды непрерывно дифференцируемы в областях своего определения.

2. $(F, h_0^*) = 0$.

3. $\lambda = 0$ есть однократное собственное значение матрицы A ($\mathrm{rang}A = n - 1$).

4. Остальные собственные значения λ матрицы A удовлетворяют условию $\mathrm{Re}\lambda < 0$.

Отметим, что $\tilde{\lambda} = 0$ есть также собственное значение матрицы $D^{-1}A$ и при наложенных на матрицы D, A условиях – однократное.

5. Остальные собственные значения $\tilde{\lambda}$ матрицы $D^{-1}A$ удовлетворяют условию $\mathrm{Re}\tilde{\lambda} < 0$.

Представим функцию $F(U)$ в виде

$$F(U) = \tilde{F} + SF + \Pi F + QF + RF, \tag{2}$$

где

$$\tilde{F} = F(U^+)\omega + F(U^-)(1-\omega),$$

$$SF = F(\tilde{U} + S) - [F(U^+)\omega + F(U^-)(1-\omega)],$$

$$\Pi F = F(\overline{U} + \Pi) - F(\overline{U}),$$

$$QF = F(\overline{U} + Q) - F(\overline{U}),$$

$$RF = F(U) - (\tilde{F} + SF + \Pi F + QF).$$

Асимптотическое представление (АП) решения задачи (1) с точностью $O(\varepsilon)$ строится в виде суммы сглаженной регулярной части $\tilde{U}(x, t, \zeta)$ пограничных функций $\Pi(x, \tau)$, $Q(\xi, t)$ и функции переходного слоя $S(\zeta, t)$:

$$U = \tilde{U}_0(x, t, \zeta) + \Pi_0(x, \tau) + Q_0(\xi, t) + S_0(\zeta, t) + r = U_0 + r, \tag{3}$$

где $U_0 = \tilde{U}_0(x, t, \zeta) + \Pi_0(x, \tau) + Q_0(\xi, t) + S_0(\zeta, t)$ – АП решения, r – остаточный член. Назначение и алгоритм построения функций, входящих в выражения (2), (3), а также аргументы этих функций описаны ниже.

Без ограничения общности, для сокращения выкладок при построении пограничных функций $\Pi(x, \tau)$, $Q(\xi, t)$ рассматривается случай однократных собственных значений матриц A и $D^{-1}A$.

2. ПОСТРОЕНИЕ АСИМПТОТИКИ РЕШЕНИЯ

2.1. Регулярная часть решения

Вначале, так же как и в [1], строится регулярная часть решения $\overline{U}$ в виде разложения по степеням ε:

$$\overline{U}(x, t) = \overline{U}_0 + \varepsilon\overline{U}_1 + \varepsilon^2\overline{U}_2 + \ldots; \tag{4}$$

здесь и далее вектор-функции с индексами 1 и 2 играют вспомогательную роль и в окончательный вид асимптотики не входят. Вектор-функцию $F(\overline{U})$ также разлагаем по степеням ε:

$$\varepsilon F(\overline{U}(x, t)) = \varepsilon F(\overline{U}_0 + \varepsilon\overline{U}_1 + \varepsilon^2\overline{U}_2 + \ldots) = \varepsilon F_1 + \varepsilon^2 F_2 + \ldots. \tag{5}$$

Отметим, что из подстановки разложения (5) в условие 2 и приравнивания коэффициентов при одинаковых степенях ε в этом выражении следует условие на компоненты функции $F(\overline{U}(x, t))$:

$$(F_i, h_0^*) = 0. \tag{6}$$

Подставляя разложения (4) и (5) в уравнение (1а) и приравнивая коэффициенты при ε^0, получаем уравнение $A\overline{U}_0 = 0$. Отсюда следует, что $\overline{U}_0 = g_0(x, t)h_0$, где $g_0(x, t)$ – пока не определенная функция. Из коэффициентов при ε^1 и ε^2 получаем уравнения для определения остальных членов разложения:

$$\begin{aligned} A\overline{U}_1 &= F_1, \\ A\overline{U}_2 &= -(\overline{U}_{0_t} + D\overline{U}_{0_x} - F_2). \end{aligned} \tag{7}$$

Для разрешимости этих уравнений должны выполняться условия

$$(F_1, h_0^*) = 0,$$

$$(\overline{U}_{0_t} + D\overline{U}_{0_x} - F_2, h_0^*) = 0.$$

Первое равенство выполняется в силу (6). Подставляя во второе условие функцию $\overline{U}_0$ в виде

$\overline{U}_0 = g_0(x, t)h_0$, получаем уравнение для определения функции $g_0(x, t)$:

$$g_{0_t} + Vg_{0_x} = 0, \text{ где } V = \frac{(Dh_0, h_0^*)}{(h_0, h_0^*)}. \tag{8}$$

Отметим, что векторные краевые и начальные условия (1б) не могут удовлетворятся одной функцией, поэтому краевые и начальные условия к уравнению (8) поставим ниже.

2.2. Пограничные П-*функции*

Для выполнения начальных условий строим пограничные вектор-функции в окрестности линии $t = 0$:

$$\Pi(x, \tau) = \Pi_0(x, \tau) + \varepsilon\Pi_1(x, \tau) + \varepsilon^2\Pi_2(x, \tau) + \ldots, \quad \tau = t/\varepsilon^2. \tag{9}$$

Пограничная функция П должна удовлетворять условию $\lim\limits_{\tau \to \infty} \Pi = 0$ и, кроме того, совместно с функцией $\overline{U}(x, t)$ формально удовлетворять уравнению

$$\varepsilon^2(\Pi_t + D\Pi_x) = A\Pi + \varepsilon\Pi F, \quad x > 0, \quad 0 < t < T, \tag{10}$$

и начальному условию (1б).

Подставляя разложения (9), (5) в уравнение (10), переходя к переменной τ и приравнивая коэффициенты при одинаковых степенях ε в обоих частях полученного равенства, получаем, что Π_0 является решением системы уравнений

$$\Pi_{0_\tau} = A\Pi_0$$

и имеет вид

$$\Pi_0 = C_0h_0 + \sum_{i=2}^{n} C_ih_i\exp(\lambda_i\tau),$$

где $C_i = C_i(x)$, $i = 0, 2, \ldots, n$. Из условия 4 и требования $\lim\limits_{\tau \to \infty} \Pi_0 = 0$ следует, что $C_0 = 0$. Подставляя в начальное условие сумму регулярной части и пограничной функции, получаем

$$g_0(x, 0)h_0 + \sum_{i=2}^{n} C_ih_i = U^0(x).$$

Так как векторы $(h_0, h_2, \ldots, h_n)$ линейно независимы, то $g(x, 0) \equiv g^-(x)$ и C_i однозначно определяются из этой системы уравнений. Тем самым начальные условия для регулярной части, или, что то же самое, для функции $g_0(x, t)$ и пограничная функция Π_0 однозначно определены. Отметим, что в силу условия 4 пограничная функция Π_0 удовлетворяет оценке

$$|\Pi_0| < C\exp(-\kappa\tau), \quad C > 0, \quad \kappa > 0.$$

2.3. Пограничные Q-функции

Для выполнения краевого условия строим краевые пограничные вектор-функции в окрестности линии $x = 0$:

$$Q(\xi, t) = Q_0(\xi, t) + \varepsilon Q_1(\xi, t) + \varepsilon^2 Q_2(\xi, t) + \ldots, \quad \xi = x/\varepsilon^2. \tag{11}$$

Пограничная функция Q должна удовлетворять условию $\lim\limits_{\xi \to \infty} Q = 0$, уравнению

$$\varepsilon^2(Q_t + DQ_x) = AQ + \varepsilon QF, \quad x > 0, \quad 0 < t < T, \tag{12}$$

и, кроме того, совместно с функцией $\overline{U}(x, t)$ формально удовлетворять краевому условию (1б).

Подставляя разложения (5) и (11) в систему (12), переходя к переменной ξ и приравнивая коэффициенты при одинаковых степенях ε в обеих частях полученного равенства, получаем, что Q_0 есть решение системы уравнений

$$Q_{0_\xi} = D^{-1}AQ_0$$

и имеет вид

$$Q_0 = E_0\tilde{h}_0 + \sum_{i=2}^{n} E_i\tilde{h}_i\exp(\tilde{\lambda}_i\xi),$$

где $\tilde{h}_i$, $i = 0, 2, \ldots, n$, $\tilde{h}_0 = h_0$ – собственные векторы, $\tilde{\lambda}_i$ – собственные значения матрицы $D^{-1}A$, $E_i = E_i(t)$, $i = 0, 2, \ldots, n$. Из условия 5 и требования $\lim\limits_{\xi\to\infty} Q_0 = 0$ следует, что $E_0 = 0$. Подставляя в краевое условие (5) сумму регулярной и погранслойной составляющих, получаем

$$g_0(0, t)\tilde{h}_0 + \sum_{i=2}^{n} E_i\tilde{h}_i = \Phi_0(t).$$

Поскольку векторы ($\tilde{h}_0, \tilde{h}_2, \ldots, \tilde{h}_n$) линейно независимы, то $g(0, t) \equiv g^+(t)$ и D_i также однозначно определяются из этой системы уравнений.

Тем самым краевые условия для регулярной части или, что то же самое, для функции $g_0(x, t)$ и пограничная функция Q_0 однозначно определены. Отметим, что в силу условия 5 пограничная функция Q_0 удовлетворяет оценке

$$|Q_0| < C\exp(-\kappa\xi), \quad C > 0, \quad \kappa > 0.$$

2.4. Внутренний пограничный слой

Функция g_0 является решением начально-краевой задачи для уравнения первого порядка в частных производных:

$$g_{0_t} + Vg_{0_x} = 0, \tag{13а}$$

$$g|_{t=0} = g^-(x), \quad g|_{x=0} = g^+(t). \tag{13б}$$

Отметим, что, вообще говоря, $g^-(0) \neq g^+(0)$.

Решение задачи (13) имеет разрыв на линии $l: t = \frac{1}{V}x$:

$$g(x, t) = \begin{cases} g^-(x - Vt) \equiv g^-(x, t), & x - Vt > 0, \\ g^+(t - x/V) \equiv g^+(x, t), & x - Vt < 0, \end{cases}$$

следовательно, регулярная часть $\bar{U}_0$ имеет вид

$$\bar{U}_0(x, t) = \begin{cases} g^-(x, t)h_0 = U_0^-(x, t), & x - Vt > 0, \\ g^+(x, t)h_0 = U_0^+(x, t), & x - Vt < 0, \end{cases} \tag{14}$$

и также имеет разрыв на линии l. В общем случае линия $l: t = \frac{1}{V}x$ не является характеристикой системы (1а) и, следовательно, решение системы должно быть гладким в окрестности линии l. Это говорит о том, что в окрестности указанной линии точное решение исходной задачи имеет переходный слой, вырождающийся при $\varepsilon \longrightarrow 0$ в разрыв.

Для описания переходного слоя в окрестности линии $l: t = \frac{1}{V}x$ применим метод сглаживания (см. [2]), для чего предварительно построим гладкую вектор-функцию $\tilde{U} = U^+(\omega(\zeta)) + U^-(1 - \omega(\zeta))$, где $\zeta = (t - x/V)/\varepsilon$, а $\omega(\zeta)$ – любая бесконечно дифференцируемая функция, удовлетворяющая условиям $\omega(\zeta) = 0$, $\zeta \le 0$, $\omega(\zeta) = 1$, $\zeta \ge 1$. Построим вектор-функцию переходного слоя S, исходя из требований

$$\varepsilon^2[(\tilde{U}_0 + S)_t + D(\tilde{U}_0 + S)_x] - A(\tilde{U}_0 + S) - \varepsilon F(\tilde{U}_0 + S) = O(\varepsilon^3), \tag{15а}$$

$$(\tilde{U}_0 + S + \Pi)\big|_{t=0} = U^0(x) + O(\varepsilon), \tag{15б}$$

$$(\tilde{U}_0 + S + Q)\big|_{x=0} = \Phi^0(t) + O(\varepsilon). \tag{15в}$$

Переходя к новым переменным (ζ, t), получаем уравнение для вектор-функции S:

$$\varepsilon^2 S_t + \varepsilon\Psi S_\zeta - AS = \varepsilon\Phi, \tag{16}$$

где

$$\begin{gathered}\Phi = SF - \Psi(U^+ - U^-)\omega_\zeta = F(\tilde{U} + S) - F(U^+)\omega - F(U^-)(1-\omega) - \Psi(U^+ - U^-)\omega_\zeta,\\ \Psi = \left(E - \frac{1}{V}D\right),\end{gathered} \tag{17}$$

E – единичная n-мерная матрица. При выводе системы (16) использовались (1а), (14), откуда получено

$$\varepsilon^2(U_t^- + DU_x^-) - AU^- = \varepsilon F(U^-),$$

$$\varepsilon^2(U_t^+ + DU_x^+) - AU^+ = \varepsilon F(U^+).$$

Для вывода функций переходного слоя будем считать, что

$$\tilde{U} = \tilde{U}_0 + \varepsilon\tilde{U}_1 + \varepsilon^2\tilde{U}_2 + \ldots,$$

где $\tilde{U}_1$ строятся в соответствии с уравнением (7), решение которого имеет вид

$$\tilde{U}_1 = \begin{cases} U_1^-(x,t) = g_1^- h_0 - GF(U_0^-), & x - Vt > 0,\\ U_1^+(x,t) = g_1^+ h_0 - GF(U_0^+), & x - Vt < 0;\end{cases} \tag{18}$$

здесь и далее матрица G псевдообратная к матрице A (общее решение системы $AY = F$ при условии $(F, h_0^*) = 0$ имеет вид $Y = GF + Ch_0$). Функции $\tilde{U}_1$, входящие в выражение (18), играют вспомогательную роль и в окончательный вид асимптотики не входят.

Вектор-функцию переходного слоя S ищем в виде разложения по степеням ε:

$$S = S_0 + \varepsilon S_1 + \varepsilon^2 S_2 + \ldots.$$

Подставляя в выражение (17) вектор-функцию S и U^+, U^- в виде разложений по степеням ε, разлагаем функцию Φ в ряд:

$$\varepsilon\Phi = \varepsilon[F(\tilde{U} + S) - F(U^+)\omega - F(U^-)(1-\omega)] - \varepsilon\omega_\zeta\Psi(U^+ - U^-) = \varepsilon\Phi_1 + \varepsilon^2\Phi_2 + \ldots.$$

Подставляя в уравнение (16) вектор-функции S и Φ в виде разложений по степеням ε и приравнивая коэффициенты при первых трех степенях ε, получаем уравнения для определения вектор-функций S_0, S_1, S_2 (вектор-функции S_1, S_2 играют вспомогательную роль и в окончательный вид АП не входят):

$$1)\ \varepsilon^0 : AS_0 = 0,$$

$$2)\ \varepsilon^1 : \Psi S_{0_\zeta} - AS_1 = \Phi_1,$$

$$3)\ \varepsilon^2 : S_{0_t} + \Psi S_{1_\zeta} - AS_2 = \Phi_2.$$

Из этих уравнений следует, что векторы S_0, S_1, S_2 находятся из равенств

$$\begin{aligned} S_0 &= s_0(\zeta, t)h_0, \\ S_1 &= s_1(\zeta, t)h_0 + G(\Psi S_{0_\zeta} - \Phi_1), \\ S_2 &= s_2(\zeta, t)h_0 + G(S_{0_t} + \Psi S_{1_\zeta} - \Phi_2), \end{aligned} \tag{19}$$

где h_0 – собственный вектор матрицы A, соответствующий нулевому собственному значению, s_0, s_1, s_2 – пока неизвестные функции.

Для разрешимости уравнений 2) и 3) должны выполняться следующие условия:

$$(\Psi S_{0_\zeta} - \Phi_1, h_0^*) = 0,$$

$$(S_{0_t} + \Psi S_{1_\zeta} - \Phi_2, h_0^*) = 0.$$

Переходя от вектор-функций S_0, S_1 к функциям s_0, s_1 в условии разрешимости для уравнения 3), получаем

$$s_{0_t}(h_0, h_0^*) + s_{1_\zeta}(\Psi h_0, h_0^*) + s_{0_{\zeta\zeta}}(\Psi G\Psi h_0, h_0^*) - (\Psi G\Phi_{1_\zeta}, h_0^*) - (\Phi_2, h_0^*) = 0.$$

Представим функцию

$$\Phi = \Phi_1 + \varepsilon\Phi_2 + \varepsilon^2\Phi_3 + \dots$$

в виде суммы $\Phi = \Phi^{\mathrm{I}} + \Phi^{\mathrm{II}}$, где

$$\Phi^{\mathrm{I}} = -\Psi(U^+ - U^-)\omega_\zeta = -\omega_\zeta\Psi(U_0^+ - U_0^-) - \omega_\zeta\Psi(U_1^+ - U_1^-) - \dots = \Phi_1^{\mathrm{I}} + \varepsilon\Phi_2^{\mathrm{I}} + \varepsilon^2\Phi_3^{\mathrm{I}} + \dots,$$

$$\Phi^{\mathrm{II}} = F(\tilde{U} + S) - F(U^+)\omega - F(U^-)(1 - \omega) = \Phi_1^{\mathrm{II}} + \varepsilon\Phi_2^{\mathrm{II}} + \varepsilon^2\Phi_3^{\mathrm{II}} + \dots.$$

Функции Φ^{I} и Φ^{II} зависят как от $\tilde{U}_0$, так и от $\tilde{U}_1$. Однако доказано, что функции $\tilde{U}_1$, имеющие вид (18), вклада в итоговое выражение не дают. (Доказательство весьма объемно, поэтому здесь не приводится.) В итоге слагаемые, содержащие функции $F(U)$, принимают вид

$$(\Psi GS\tilde{F}_{1_\zeta} + S\tilde{F}_2^{\mathrm{I}}, h_0^*) = -\omega_{\zeta\zeta}k(\Psi G\Psi h_0, h_0^*) + (\Psi GF((k\omega + g_0^-(0) + s_0)h_0), h_0^*)'_\zeta.$$

Подставляя в условия (15б), (15в) сумму ($\tilde{U}_0 + S_0$) (где S_0 находится из соотношений (19)), находим начальные и краевые условия для вектор-функции S_0 или, что то же самое, для функции s_0. Таким образом, получаем задачу на нахождение функций внутреннего переходного слоя:

$$s_{0_t} + Ms_{0_{\zeta\zeta}} + M\omega_{\zeta\zeta}\kappa - \frac{1}{(h_0, h_0^*)}(\Psi GF(k\omega + g_0^-(0) + s_0)h_0, h_0^*)_\zeta = 0, \tag{20а}$$

$$s_0\big|_{t = 0,\, \zeta > 0} = -k\omega, \tag{20б}$$

$$s_0\big|_{t = 0,\, \zeta < 0} = k(1 - \omega), \tag{20в}$$

где

$$k = g_0^+(0) - g_0^-(0), \quad M = \frac{(\Psi G\Psi h_0, h_0^*)}{(h_0, h_0^*)}.$$

Легко показать, что, хотя собственные векторы h_0, h_0^* и матрица G определены неоднозначно, коэффициент M определяется единственным образом.

В полученной системе уравнений можно сделать замену переменной $s_0 = -k\omega + \tilde{s}_0$. Тем самым для $\tilde{s}_0$ получается задача

$$\tilde{s}_{0_t} + M\tilde{s}_{0_{\zeta\zeta}} - \frac{(\Psi GF(g_0^-(0) + \tilde{s}_0)h_0, h_0^*)_\zeta}{(h_0, h_0^*)} = 0,$$

$$\tilde{s}_0\big|_{t = 0,\, \zeta > 0} = 0, \quad \tilde{s}_0\big|_{t = 0,\, \zeta < 0} = k.$$

При условии параболичности уравнения (20а) (т.е. $M < 0$) существует такое $T > 0$, что на сегменте $[0, T]$ существует единственное решение задачи (20а) и на $[0, T]$ справедлива оценка

$$|s_0| < C\exp\left(-\kappa\frac{\zeta^2}{t}\right), \quad C \geq 0, \quad \kappa > 0.$$

Необходимое и достаточное условие параболичности уравнения (20а) для матриц A общего вида, выраженное через элементы самой матрицы A, чрезвычайно громоздко и при размерности системы больше трех практически не поддается проверке.

Однако для матриц A, элементами которых являются заданные числа, вычисление коэффициента M и проверка условия $M < 0$ не представляет больших трудностей.

Условие параболичности задачи $M < 0$ доказано для некоторых матриц специального вида, в частности для симметричных матриц A, у которых все элементы, кроме главной диагонали, неотрицательны, а элементы главной диагонали равны суммам элементов по столбцам, взятых с отрицательными знаками. Такие матрицы встречаются в ряде физических приложений.

2.5. Асимптотическое представление решения

Представим решение задачи (1) в виде

$$U(x, t) = \tilde{U}_0(x, t) + \Pi_0(x, \tau) + Q_0(\xi, t) + S_0(\zeta, t) + r(x, t, \varepsilon),$$

где $\tilde{U}_0(x, t) + \Pi_0(x, \tau) + Q_0(\xi, t) + S_0(\zeta, t)$ – построенная асимптотика, $r(x, t, \varepsilon)$ – остаточный член.

Из алгоритма построения асимптотики следует, что $r(x, t, \varepsilon)$ удовлетворяет задаче

$$\varepsilon^2(r_t + Dr_x) = Ar + \varepsilon RF + O(\varepsilon^2), \quad x > 0, \quad 0 < t < T, \tag{21а}$$

$$r(0, x) = O(\varepsilon), \quad r(t, 0) = O(\varepsilon). \tag{21б}$$

Оценки правой части уравнения, а также начальных и краевых условий получаются непосредственным вычислением с учетом алгоритма построения асимптотики.

3. ЗАКЛЮЧЕНИЕ

Таким образом, построено формальное (с оценкой по невязке) асимптотическое представление решения начально-краевой задачи для сингулярно возмущенной системы дифференциальных уравнений первого порядка в частных производных с малой нелинейностью.

Особенностью задачи является наличие у точного решения параболического переходного слоя.

Построенное АП (21б) не является равномерным в окрестности начала координат, поскольку оно является гладким в области Ω при любых, в том числе и не согласованных, начальных и краевых условиях. Вместе с тем точное решение задачи (1) в этом случае будет иметь разрывы на характеристиках гиперболической системы (1а).

Структура АП в окрестности начала координат исследована в [2] (в случае $F = 0$).

Параболичность переходного слоя обусловлена порядком вырождения матрицы A: $\mathrm{rang}A = n - 1$. При ином вырождении матрицы A структура переходного слоя будет иной.

Условия однократности отличных от нуля собственных значений матриц A и $D^{-1}A$ не существенны и от них можно отказаться.

Алгоритм построения АП без труда распространяется и на случай диагональных матриц D, у которых часть диагональных элементов обращается в нуль (при соответствующем изменении краевых условий).

СПИСОК ЛИТЕРАТУРЫ

1. *Нестеров А.В.* О внутренних переходных параболических слоях // Тр. Междунар. конф. “Матем. идеи П.Л. Чебышёва и их приложение к современным проблемам естествознания”. Тезисы докл. Обнинск: ИАТЭ, 2002.
2. *Нестеров А.В.* Об асимптотике решения системы уравнений диффузия–сорбция при малых коэфициентах диффузии // Ж. вычисл. матем. и матем. физ. 1989. Т. 29. № 9. С. 1318–1330.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 445–459

УДК 519.63

О РАЗНОСТНЫХ АППРОКСИМАЦИЯХ ПЕРЕОПРЕДЕЛЕННЫХ ГИПЕРБОЛИЧЕСКИХ УРАВНЕНИЙ КЛАССИЧЕСКОЙ МАТЕМАТИЧЕСКОЙ ФИЗИКИ[1)]

© 2007 г. Д. П. Бабий*, С. К. Годунов*, В. Т. Жуков, О. Б. Феодоритова****

(630090 Новосибирск, пр-т Коптюга, 4, Ин-т матем. СО РАН;*

*** 125047 Москва, Миусская пл., 4, Ин-т прикл. матем. РАН)*

e-mail: godunov@math.nsc.ru; zhukov@kiam.ru; feodor@kiam.ru

Поступила в редакцию 17.10.2006 г.

Рассматриваются вопросы разностной аппроксимации переопределенных систем гиперболических уравнений. Для системы уравнений гидродинамики, магнитной гидродинамики, Максвелла и упругости приведены формулировки расширенных переопределенных систем. Обсуждаются подходы к построению разностных схем для таких систем. Библ. 9. Фиг. 5.

Ключевые слова: переопределенные гиперболические уравнения, разностные аппроксимации, уравнения газовой динамики, уравнения акустики, уравнения Максвелла.

1. ВВЕДЕНИЕ

Мы остановимся на трудностях, которые приходится преодолевать при разработках алгоритмов решения переопределенных систем уравнений, называемых *гиперболическими*. Вообще говоря, гиперболическими называют системы, состоящие из числа уравнений, совпадающего с числом неизвестных, и либо имеющими определенную структуру характеристического коноида (предложение И.Г. Петровского), либо записывающиеся с помощью коэффициентов, образующих симметрические (эрмитовы) матрицы (предложение Фридрихса). Одна из этих матриц должна быть положительно-определенной. Полезно отметить, что системы, гиперболические по Петровскому, совсем не обязательно допускают запись в виде симметрических гиперболических по Фридрихсу.

Пример, демонстрирующий это обстоятельство, был построен в 1984 году В.В. Ивановым, однако его нельзя считать общеизвестным, так как он опубликован только на русском языке в препринте (см. [1]).

Будем придерживаться определения Фридрихса и называть систему из N уравнений для N неизвестных функций, образующих N-мерный вектор u, гиперболической, если она записана в виде

$$A\frac{\partial u}{\partial t} + B_k\frac{\partial u}{\partial x_k} = f,$$

где квадратные матрицы $A = A^* > 0$, $B_k = B_k^*$ либо постоянны, либо являются гладкими функциями от u (случай квазилинейных уравнений).

Простейшая система уравнений газовой динамики в лагранжевых координатах, разностную схему для которой один из авторов разрабатывал зимой 1953–1954 гг. (она была опубликована лишь в 1959 г., см. [2]), а именно

$$\begin{aligned} &\partial u/\partial t + \partial p/\partial m = 0, \quad E = E(V, s),\\ &-\partial V/\partial t + \partial u/\partial m = 0, \quad p = -E_V(V, s), \\ &\partial s/\partial t = 0, \end{aligned} \tag{1.1}$$

[1)] Работа является расширенным изложением пленарного доклада на XI международной конференции по гиперболическим уравнениям (17 июля 2006 года, Франция, Лион) и выполнена при финансовой поддержке грантов 047.016.003 РФФИ–NWO (Нидерланды), “Ведущие научные школы” НШ-9019.2006.1 РФ, междисциплинарного проекта № 5 Президиума СО РАН, программы № 3.1 ОМН РАН.

после введения гамильтониана $H(u, p, s) = E + pv + u^2/2$, $dH = udu - vdp + Tds$ может быть записана в виде, предлагаемом Фридрихсом:

$$\begin{pmatrix} H_{uu} & H_{up} & H_{us} \\ H_{pu} & H_{pp} & H_{ps} \\ H_{su} & H_{sp} & H_{ss} \end{pmatrix} \frac{\partial}{\partial t} \begin{pmatrix} u \\ p \\ s \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{\partial}{\partial m} \begin{pmatrix} u \\ p \\ s \end{pmatrix} = 0.$$

Нетрудно видеть, что эта запись при обозначении $\hat{H} = pu$ (с учетом равенства $\hat{H}_s = 0$) эквивалентна следующей системе:

$$\begin{aligned} &u \quad \frac{\partial H_u}{\partial t} + \frac{\partial \hat{H}_u}{\partial m} = 0, \\ &p \quad \frac{\partial H_p}{\partial t} + \frac{\partial \hat{H}_p}{\partial m} = 0, \\ &T \quad \frac{\partial H_s}{\partial t} + \frac{\partial \hat{H}_s}{\partial m} = 0. \end{aligned} \tag{1.2}$$

Умножая последние уравнения на выписанные слева от них множители u, p, T и суммируя, мы приходим к дополнительному равенству, совместному с системой (1.2):

$$\begin{gathered} \frac{\partial\left(E + \frac{u^2}{2}\right)}{\partial t} + \frac{\partial pu}{\partial m} = 0, \\ E + \frac{u^2}{2} = uH_u + pH_p + sH_s - H. \end{gathered} \tag{1.3}$$

Таким образом, добавляя уравнение (1.3) к (1.1) или (1.2), получаем совместную переопределенную систему из четырех уравнений.

Интересно отметить, что из этой переопределенной системы симметрическую гиперболическую подсистему можно выделить иначе. Пусть эта подсистема составлена из уравнений

$$\begin{aligned} &-\frac{u}{T} \qquad \frac{\partial u}{\partial t} + \frac{\partial p}{\partial m} = 0, \\ &-\frac{p}{T} \qquad -\frac{\partial v}{\partial t} + \frac{\partial u}{\partial m} = 0, \\ &\frac{1}{T} \qquad \frac{\partial\left(E + \frac{u^2}{2}\right)}{\partial t} + \frac{\partial pu}{\partial m} = 0, \end{aligned} \tag{1.4}$$

слева от которых выписаны “интегрирующие множители”. Линейная комбинация уравнений с этими множителями в виде коэффициентов оказывается “законом сохранения энтропии” $\partial s/\partial t = 0$.

Если воспользоваться термодинамическим тождеством

$$ds = -\frac{u}{T}du - \frac{p}{T}d(-v) + \frac{1}{T}d\left(E + \frac{u^2}{2}\right),$$

$$d\left(\frac{E + pv - Ts - \frac{u^2}{2}}{T}\right) = ud\left(\frac{u}{T}\right) - vd\left(-\frac{p}{T}\right) + \left(E + \frac{u^2}{2}\right)d\frac{1}{T}$$

и ввести обозначения

$$q_1 = -\frac{u}{T}, \quad q_2 = -\frac{p}{T}, \quad q_0 = \frac{1}{T},$$
$$L = \frac{E + pV - Ts - \frac{u^2}{2}}{T} = L(q_0, q_1, q_2), \quad L^1 = \frac{q_1 q_2}{q_0}, \tag{1.5}$$

то уравнения (1.4) примут вид

$$\begin{aligned} q_0 \quad & \frac{\partial L_{q_0}}{\partial t} + \frac{\partial L^1_{q_0}}{\partial m} = 0, \\ q_1 \quad & \frac{\partial L_{q_1}}{\partial t} + \frac{\partial L^1_{q_1}}{\partial m} = 0, \\ q_2 \quad & \frac{\partial L_{q_2}}{\partial t} + \frac{\partial L^1_{q_2}}{\partial m} = 0. \end{aligned} \tag{1.6}$$

Так как

$$q_0 L_{q_0} + q_1 L_{q_1} + q_2 L_{q_2} - L = s,$$
$$q_0 L^1_{q_0} + q_1 L^1_{q_1} + q_2 L^1_{q_2} - L = 0,$$

то равенство $\partial s/\partial t = 0$ можно получить из уравнений (1.4) путем взятия их линейной комбинации с выписанными слева множителями q_0, q_1, q_2. Квазилинейная форма выписанных уравнений (1.6)

$$L_{q_i q_j}\frac{\partial q_i}{\partial t} + L^1_{q_i q_j}\frac{\partial q_j}{\partial m} = 0$$

при выпуклом производящем потенциале $L(q_0, q_1, q_2)$ свидетельствует об их гиперболичности.

Именно система (1.4), составленная из законов сохранения количества движения, объема и энергии, была положена в основу разработанной схемы, на решениях которой безусловно выполнялось требование неубывания энтропии. По этой причине схему можно было использовать при расчете разрывных решений – ударных волн. На ударных волнах энтропия растет и переопределенную систему, составленную из (1.1) и (1.3), уже нельзя считать совместной.

Размышления над описанными обстоятельствами подтолкнули к выделению класса "термодинамически согласованных" уравнений, образующих симметрические гиперболические системы, составленные из дивергенций (законов сохранения)

$$\frac{\partial L_{q_i}}{\partial t} + \frac{\partial M^j_{q_i}}{\partial x_j} = 0$$

и совместных (на гладких решениях) с дополнительным законом сохранения из [3], [4]

$$\frac{\partial(q_i L_q - L)}{\partial t} + \frac{\partial[q_i M^j_{q_i} - M^j]}{\partial x_j} = 0.$$

В дальнейшем, однако, выяснилось, что далеко не все уравнения классической математической физики к такому виду приводятся. Вот, например, аналогичная формализация уравнений магнитной гидродинамики (недивергентная):

$$\frac{\partial L_{q_0}}{\partial t} + \frac{\partial (u_k L) q_0}{\partial x_k} = 0,$$

$$\frac{\partial L_{h_i}}{\partial t} + \frac{\partial (u_k L)_{h_i}}{\partial x_k} - L_{h_k}\frac{\partial u_i}{\partial x_k} = 0, \tag{1.7}$$

$$\frac{\partial L_{u_i}}{\partial t} + \frac{\partial (u_k L)_{u_i}}{\partial x_k} - L_{h_k}\frac{\partial h_i}{\partial x_k} = 0.$$

Хотя система (1.7) без труда переписывается в симметрическом гиперболическом виде, она не содержит законов сохранения количества движения и энергии. Эти законы (дивергентные равенства)

$$\frac{\partial L_{u_i}}{\partial t} + \frac{\partial [(u_k L)_{u_i} - h_i L_{h_k}]}{\partial x_k} = 0,$$

$$\frac{\partial (q_0 L_{q_0} + h_i L_{h_i} + u_i L_{u_i} - L)}{\partial t} + \frac{\partial [u_k (q_0 L_{q_0} + h_i L_{h_i} + u_i L_{u_i}) - u_i h_i L_{h_k}]}{\partial x_k} = 0$$

справедливы не на всех решениях системы (1.7), а только на решениях, удовлетворяющих дополнительному уравнению

$$\partial L_{h_i}/\partial x_i = 0,$$

совместному с (1.7), как это вытекает из ее следствия

$$\frac{\partial}{\partial t}\left[\frac{1}{L_{q_0}}\ \frac{\partial L_{h_i}}{\partial x_i}\right] + u_k \frac{\partial}{\partial x_k}\left[\frac{1}{L_{q_0}}\ \frac{\partial L_{h_i}}{\partial x_i}\right] = 0.$$

Аналогично обстоит дело и в уравнениях нелинейной теории упругости в эйлеровых координатах.

Отмеченное обстоятельство вызывает трудности при численном решении уравнений магнитной гидродинамики. Обзор этих сложностей и описание всевозможных технологических приемов их преодоления составляет значительную часть в монографии [5], с которой мы знакомились после того, как профессор Ван-Леер обратил наше внимание на затруднения, причиной которых является накопление вычислительных погрешностей. Эти погрешности, по-видимому, накапливаются в основном при долговременном расчете больших гладких полей, тогда как в работах, реферируемых в [5], при построении расчетных схем упор делается на их адаптации к расчету зон, содержащих ударные переходы. Размышляя об этом, мы решили, что стоит разобраться в проблеме расчета гладких решений у совместных переопределенных систем, на некоторое время забыв о том, что у этих систем существуют и разрывные решения. К вопросу о расчете разрывных решений, на наш взгляд, следует вернуться лишь после того, как ситуация с гладкими решениями станет совершенно ясной. Так же как и 53 года тому назад, когда один из авторов готовил свой вариант расчета разрывных решений газовой динамики к пуску первой русской серийной ЭВМ "Стрела", мы решили сначала детально исследовать задачи, описываемые линейными уравнениями с постоянными коэффициентами. Вот об этих исследованиях пойдет речь ниже. Хотя они пока находятся на начальном этапе, мы в процессе их проведения натолкнулись на новые, неожиданные для нас приемы построения вычислительных алгоритмов.

2. СХЕМЫ С КОРРЕКТИРОВКОЙ ДЛЯ УРАВНЕНИЙ АКУСТИКИ

Начнем с разбора простейшего примера двумерных уравнений акустики:

$$\frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = a(x, y, t), \quad \frac{\partial v}{\partial t} + \frac{\partial p}{\partial y} = b(x, y, t),$$

$$\frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = c(x, y, t), \tag{2.1}$$

$$0 \le x \le \pi, \quad 0 \le y \le \pi, \quad t \ge 0.$$

Ограничимся для простоты изучением решений, удовлетворяющих граничному условию $p = 0$ на границе области и заданным при $t = 0$ начальным условиям.

Дополнительные уравнения, совместные с этой системой, имеют вид

$$\frac{\partial}{\partial t}\frac{u^2+v^2+p^2}{2}+\frac{\partial pu}{\partial x}+\frac{\partial pv}{\partial y} = au+bv+cp,$$
$$\frac{\partial}{\partial t}(u_y - v_x) = \frac{\partial a(x,y,t)}{\partial y}-\frac{\partial b(x,y,t)}{\partial x}. \tag{2.2}$$

Основная трудность, как известно, состоит в выборе разностной аппроксимации, обеспечивающей достаточно точное соблюдение уравнения (2.2). Расширим нашу систему, добавив в нее новые неизвестные потенциалы ω_1, ω_2, ω и новые уравнения

$$\frac{\partial\omega_1}{\partial t} = a(x,y,t),\quad \frac{\partial\omega_2}{\partial t} = b(x,y,t),\quad \frac{\partial\omega}{\partial t} = p(x,y,t),$$
$$\omega_1(x,y,0) = u(x,y,0),\quad \omega_2(x,y,0) = v(x,y,0),\quad \omega(x,y,0) = 0. \tag{2.3}$$

Легко убедиться, что ω удовлетворяет волновому уравнению

$$\frac{\partial^2\omega}{\partial t^2}-\frac{\partial^2\omega}{\partial x^2}-\frac{\partial^2\omega}{\partial y^2} = \frac{\partial}{\partial t}\left[\frac{\partial c}{\partial t}-\frac{\partial a}{\partial x}-\frac{\partial b}{\partial y}\right]$$

и что $u(x,y,t)$, $v(x,y,t)$ представимы в виде

$$u(x,y,t) = \omega_1(x,y,t)-\omega_x(x,y,t),$$
$$v(x,y,t) = \omega_2(x,y,t)-\omega_y(x,y,t). \tag{2.4}$$

Расчет одного временнóго шага $t \longrightarrow t+\tau$ разобьем на два этапа. На первом этапе используется аппроксимация симметрической гиперболической системы

$$\frac{\partial u}{\partial t}+\frac{\partial p}{\partial x} = a(x,y,t),\quad \frac{\partial v}{\partial t}+\frac{\partial p}{\partial y} = b(x,y,t),$$
$$\frac{\partial p}{\partial t}+\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y} = c(x,y,t), \tag{2.5}$$
$$\frac{\partial\omega_1}{\partial t} = a,\quad \frac{\partial\omega_2}{\partial t} = b,\quad \frac{\partial\omega}{\partial t} = p.$$

Несложный пример такой достаточно точной аппроксимации, основанный на симметрической гиперболичности, будет приведен ниже. После того как первый этап закончен, на втором этапе производим замену вычисленных $u(x, y, t+\tau)$, $v(x, y, t+\tau)$ на новые их значения, определяемые по представлениям (2.4), используя разностную аппроксимацию этих равенств на момент времени $t+\tau$. После этой замены расчет временного шага $t \longrightarrow t+\tau$ закончен и мы приступаем к расчету следующего шага по времени.

По существу описанная процедура может рассматриваться как реализация последовательного решения сначала волнового уравнения (первый этап)

$$\frac{\partial^2 p}{\partial t^2}-\frac{\partial^2 p}{\partial x^2}-\frac{\partial^2 p}{\partial y^2} = \frac{\partial c}{\partial t}-\frac{\partial a}{\partial x}-\frac{\partial b}{\partial y}$$

и уравнений для потенциалов (2.3) с переходом затем к расчету величин u, v с помощью соотношений (2.4) (второй этап). Значения u, v, используемые на первом этапе, рассматриваются лишь как промежуточные детали при решении волнового уравнения.

Изучим на примере уравнений акустики эффективность предложенной потенциальной коррекции при применении схем первого и второго порядка точности. Для краткого объяснения используемых схем запишем систему уравнений (2.1) в виде

$$\partial U/\partial t = DU+F(x,y,t),\quad (x,y)\in\Omega = [0,\pi]^2,\quad 0<t<T. \tag{2.6}$$

Дифференциальный оператор D действует на функцию $U(x, y, t) = (u, v, p)$ при фиксированном t

по правилу

$$DU = -\left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}, \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \tag{2.7}$$

Ограничимся простейшими аппроксимациями функций и дифференциального оператора D. Введем в области Ω квадратную сетку с шагом h и определим сеточные вектор-функции $\mathbf{u} = (u, v, p)$ в центрах ячеек (внутри ячейки каждая из компонент u, v, p является постоянной величиной). Построим разностные операторы D_1 и D_2, аппроксимирующие D с первым и вторым порядком по h. Оператор D_1 получается при аппроксимации симметрической гиперболической системы (2.6) по схеме С.К. Годунова (см. [9]); основные элементы этой схемы приведены в разд. 3. Оператор D_2 получается при замене входящих в определение (2.7) производных на центрально-разностные отношения. На границе области для каждого оператора D_1 и D_2 записываются краевые условия, обеспечивающие консервативность. Здесь и в дальнейшем, говоря о порядках точности, мы подразумеваем точность на гладких решениях.

Схему Годунова для системы уравнений (2.6) можно формально записать в виде

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} = D_1\mathbf{u}^n + \mathbf{f}^{n+1/2}, \tag{2.8}$$

где n – номер слоя по времени, $\mathbf{u}^n$, $n = 0, 1, \ldots,$ – искомая сеточная аппроксимация функции $U(x, y, t) = (u, v, p)$ на пространственно-временно́й сетке. Начальная функция $\mathbf{u}^0$ задана, а сеточная функция $\mathbf{f}^{n+1/2}$ есть приближенное среднее значение функции F на шаге интегрирования $t_n < t < t_{n+1} = t_n + \tau$; величина шага τ определяется известным условием устойчивости.

Схема (2.8), назовем ее **Схема 1**, используется нами для построения нескольких различных модификаций. На основе предложенного приема потенциальной коррекции нетрудно построить **Схему 1P**, которая состоит из упомянутых выше двух этапов: на первом этапе выполняется шаг интегрирования по времени гиперболической системы (2.5) по схеме 1 (с учетом входящих в эту систему уравнений для потенциалов (2.3)), на втором этапе делается коррекция функций u^{n+1}, v^{n+1} с помощью соотношений (2.4).

Построим для системы (2.6) **Схему 2** второго порядка точности. Предиктором в этой схеме является схема (2.8). Далее по известным значениям на двух временны́х слоях $\mathbf{u}^{n+1}$, $\mathbf{u}^n$ вычислим с использованием оператора D_2 невязку

$$\mathbf{r}^{n+\frac{1}{2}} = \frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\tau} - D_2\frac{\mathbf{u}^{n+1} + \mathbf{u}^n}{2} - \mathbf{f}^{n+1/2}.$$

Затем находим поправку δ^{n+1}, выполняя шаг по схеме (2.8)

$$(\delta^{n+1} - \delta^n)/\tau = D_1\delta^n - \mathbf{r}^{n+1/2}$$

с начальным условием $\delta^0 = 0$.

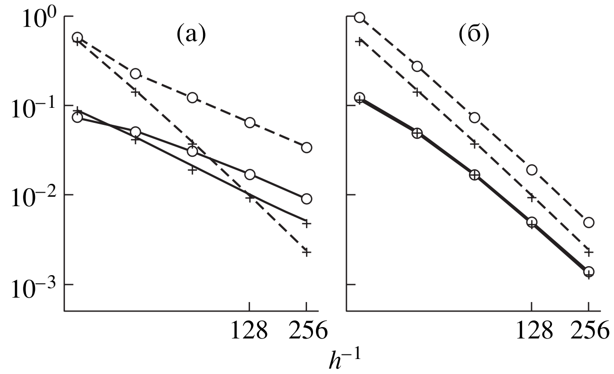
Аппроксимацию решения со вторым порядком дает коррекция $\mathbf{w}^n = \mathbf{u}^n + \delta^n$. Исключая промежуточные функции, записываем схему расчета пары функций $\{\mathbf{u}, \mathbf{w}\}$ в виде

$$\begin{gathered}(\mathbf{u}^{n+1} - \mathbf{u}^n)/\tau = D_1\mathbf{u}^n + \mathbf{f}^{n+1/2},\\ \frac{\mathbf{w}^{n+1} - \mathbf{w}^n}{\tau} = D_1\mathbf{w}^n + D_2\frac{\mathbf{u}^{n+1} + \mathbf{u}^n}{2} - D_1\mathbf{u}^n + \mathbf{f}^{n+1/2},\end{gathered} \tag{2.9}$$

$n = 0, 1, \ldots$; начальные условия $\mathbf{u}^0$ и $\mathbf{w}^0 = \mathbf{u}^0$. Функции $\mathbf{u}$, $\mathbf{w}$ аппроксимируют одну и ту же функцию U – решение системы (2.6). Детальное исследование схемы (2.9) будет дано в отдельной публикации. Идея схемы (2.9) возникла в процессе разработки и обсуждения алгоритма расчета гладких собственных функций симметрической гиперболической системы. Этот алгоритм приведен в [6], [7] и был предметом доклада на конференции по вычислительной гидродинамике (см. [8]).

Так же, как это сделано выше для схемы 1, построим **Схему 2P** второго порядка точности с потенциальной коррекцией: на первом этапе выполняется шаг по времени по схеме 2 для расширенной системы (2.5), на втором этапе делается коррекция компонент u^{n+1}, v^{n+1} вектора $\mathbf{w}$ с помощью соотношений (2.4).

Мы провели две серии экспериментальных расчетов. В первой серии метод тестировался на собственных функциях системы (2.1) при нулевом граничном значении для p и при нулевых пра-



Фиг. 1.

вых частях a, b, c. На этих разностных собственных функциях условия совместности $u_y - v_x = 0$ удовлетворяются с высокой точностью, и мы проверили, как они соблюдаются в наших расчетах с потенциальной корректировкой. Как и следовало ожидать, погрешность аппроксимации соотношения (2.2) оказалась практически нулевой, в отличие от приводимых ниже расчетов с ненулевыми a, b, c.

Вторая серия экспериментов состояла в расчете задачи с заранее указанным ответом, при котором правые части получаются при подстановке этого ответа в уравнения. Сравнивались решения с потенциальной корректировкой и без нее. Оказалось, что при применении корректировки, во-первых, точность аппроксимации функций u, v, p лучше, чем без нее. Во-вторых, погрешность аппроксимации дополнительного уравнения (2.2) существенно уменьшается и, что важно, не растет с увеличением времени. В качестве точного решения $U(x, y, t) = (u, v, p)$ берем

$$u(x, y, t) = \cos kt \cos mx \sin ly,$$
$$v(x, y, t) = \cos kt \sin mx \cos ly,$$
$$p(x, y, t) = \sin kt \sin mx \sin ly$$

при некоторых значениях k, m, l. В приводимых ниже результатах $k = 9$, $m = 6$, $l = 4$ и уравнения интегрировались на отрезке $0 \le t \le 0.7\pi$.

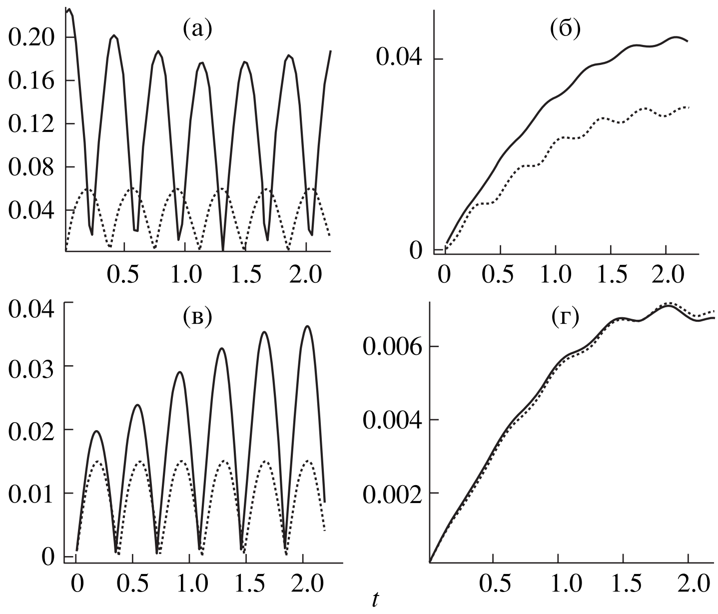
Напомним, что разностную аппроксимацию точного решения $U(x, y, t) = (u, v, p)$ на каждом шаге по времени t_n мы обозначаем через $\mathbf{u}^n = (u^n, v^n, p^n)$. Точность схем будем характеризовать сеточной L_2-нормой погрешности $E_U^n = \|\mathbf{u}^n - U(x, y, t^n)\|$ и нормой невязки дополнительного уравнения (2.2)

$$E_R^n = \left\| \frac{(u_y^{n+1} - v_x^{n+1}) - (u_y^n - v_x^n)}{\tau} - a_y^{n+1/2} - b_x^{n+1/2} \right\|,$$

а также их средними значениями E_U и E_R на отрезке интегрирования $0 \le t \le 0.7\pi$. В приведенной выше формуле для невязки $a_y^{n+1/2}$, $b_x^{n+1/2}$ являются средними значениями функций a_y, b_x на отрезке $(t, t + \tau)$, вычисленными с помощью квадратурной формулы Симпсона. Расчеты проведены на квадратных сетках с числом шагов в каждом направлении 16, 32, 64, 128, 256. На фиг. 1 показаны зависимости средних значений E_U (сплошная линия) и E_R (штриховая) от шага сетки h: (а) – схема 1 (кружочки) и схема 1Р (крестики); (б) – схема 2 (кружочки) и схема 2Р (крестики).

Видим, что точность расчета основных функций u, v, p не ухудшается после введения потенциальной корректировки, а точность выполнения дополнительного уравнения улучшается. Отметим, что для схемы первого порядка точности потенциальная корректировка повышает порядок аппроксимации уравнения (2.2) до второго. Для схемы второго порядка потенциальная корректировка не меняет порядка аппроксимации, но улучшает реальную точность.

На фиг. 2 показано поведение во времени t невязки E_R (фиг. 2 а, б) и погрешности E_U (фиг. 2 б, г) для сетки 128×128. Сплошными линиями показаны графики для схемы 1 (фиг. 2 а, б) и схемы 2 (фиг. 2 в, г), пунктирными – для аналогов этих схем с потенциальной корректировкой. Видно,



Фиг. 2.

что невязка E_R дополнительного уравнения (2.2) колеблется по времени. Потенциальная корректировка оказывает положительное действие – уменьшает амплитуду колебаний в несколько раз, а в случае схемы второго порядка – предотвращает рост амплитуды во времени (фиг. 2 в). Для схемы 1 и схемы 2 погрешность E_U (фиг. 2 б, г) вычисления основных функций u, v, p растет по времени (на данном участке интегрирования). Потенциальная корректировка не ухудшает точности расчета этих функций.

Отметим, что при дальнейшем увеличении времени графики невязки E_R и погрешности E_U для каждой из схем выходят на стационарный режим, т.е. рост, который наблюдается на графиках на начальном участке интегрирования, прекращается.

3. СХЕМЫ С КОРРЕКТИРОВКОЙ ДЛЯ УРАВНЕНИЙ МАКСВЕЛЛА

Попытаемся адаптировать приведенные выше приемы корректировки к решению систем уравнений Максвелла. Опишем один из достаточно широко употребляемых вариантов конструирования схем второго порядка точности, использующих в качестве предиктора стандартную схему Годунова для симметрических гиперболических систем. Алгоритмически она близка к схеме второго порядка точности, кратко описанной в разд. 2 и применяемой к решению уравнений акустики. Система симметрических гиперболических уравнений имеет вид

$$A\frac{\partial u}{\partial t} + B\frac{\partial u}{\partial x} + C\frac{\partial u}{\partial y} + D\frac{\partial u}{\partial z} = F, \tag{3.1}$$

где матрицы A, B, C и D могут зависеть от переменных x, y, z, t, а в нелинейном случае – и от решения u; $A^* = A > 0$, $B^* = B$, $C^* = C$, $D^* = D$.

Метод построения трехмерной разностной схемы состоит в использовании одномерных схем вдоль каждого из пространственных направлений. Поэтому построим сначала схему для одномерной задачи, рассматривая для простоты систему уравнений с нулевой правой частью и постоянными матрицами A и B:

$$A\frac{\partial u}{\partial t} + B\frac{\partial u}{\partial x} = 0. \tag{3.2}$$

Для дискретизации введем равномерную пространственную сетку с шагом h_x и составим для системы (3.2) схему Годунова, детали конструкции которой хорошо известны (см., например, [5],

[9]). Запишем эту схему в виде

$$\frac{v_i^{n+1} - v_i^n}{\tau} + \hat{B}\frac{V_{i+1/2} - V_{i-1/2}}{h_x} = 0, \tag{3.3}$$

где $\hat{B} = A^{-1/2}BA^{-1/2}$, v_i^n, $V_{i\pm 1/2}$ – значения сеточных функций. Верхний индекс $n = 0, 1, \dots$ есть номер слоя по времени, нижний целый индекс $i = 0, 1, \dots, I$ – это значения сеточной функции, отнесенные к центру ячейки сетки, нижние полуцелые индексы $i - 1/2$ и $i + 1/2$ – это, соответственно, значения функции на левой и правой границе ячейки с номером i. Значения $V_{i\pm 1/2}$ в схеме Годунова носят названия "больших величин" и получаются из решения задачи Римана на границе $i \pm 1/2$. Величина шага τ по времени определяется условием устойчивости Куранта–Фридрихса–Леви. Для простоты изложения ограничимся случаем периодических краевых условий, учет которых обеспечивается просто циклическим переносом значений. Схема (3.3) обладает первым порядком точности по пространству и времени и является "предиктором" для построения схемы второго порядка, описываемой ниже. Этап, уточняющий результат на верхнем временно́м слое, условно можно называть "корректором".

Для начала корректируются значения "больших величин" на границах ячеек:

$$\tilde{V}_{i+1/2} = \frac{v_{i-1}^n + v_{i+1}^n + v_{i-1}^{n+1} + v_{i+1}^{n+1}}{4} \tag{3.4}$$

(если индекс i выходит за пределы области определения, то значения функции определяются из условия периодичности).

Еще раз вычисляются значения на верхнем временно́м слое при помощи схемы (3.3):

$$\frac{v_i^{n+1} - v_i^n}{\tau} + \hat{B}\frac{\tilde{V}_{i+1/2} - \tilde{V}_{i-1/2}}{h_x} = 0. \tag{3.5}$$

Применяя схему (3.3)–(3.5) для каждой из трех одномерных систем, получаем результирующую схему для трехмерной системы (3.1), которую запишем в операторном виде:

$$\frac{v^{n+1} - v^n}{\tau} + R_1 v^n + R_2 v^n + R_3 v^n = F^{n+1/2}, \tag{3.6}$$

где операторы R_1, R_2, R_3 отвечают схемам вдоль направлений x, y, z соответственно.

Теперь опишем детали корректировок, применяемых при аппроксимации классических уравнений Максвелла

$$\begin{gathered} \partial E/\partial t = \operatorname{rot} H - j, \\ \partial H/\partial t = -\operatorname{rot} E, \\ \operatorname{div} H = 0, \end{gathered} \tag{3.7}$$

следствием которых является соотношение

$$\frac{\partial}{\partial t}(\operatorname{div} E) = -\operatorname{div} j. \tag{3.8}$$

Если определить плотность заряда ρ формулой $\rho = \operatorname{div} E$, то для этого заряда должно быть выполнено равенство (закон сохранения)

$$\partial\rho/\partial t + \operatorname{div} j = 0. \tag{3.9}$$

При аппроксимации системы (3.7) на первом этапе игнорируем это дополнительное соотношение и обращаем на него внимание только при корректировке. Последнему из уравнений $\operatorname{div} H = 0$, входящему в выписанную нами переопределенную, но совместную систему (3.7), мы уделим пристальное внимание. В уравнениях Максвелла роль потенциалов играют векторный и скалярный потенциалы A и Φ, удовлетворяющие, соответственно, векторному и скалярному волновым уравнениям

$$\partial^2 A/\partial t^2 - \Delta A = j, \quad \partial^2\Phi/\partial t^2 - \Delta\Phi = \rho;$$

здесь Δ – лапласиан. Сами эти потенциалы можно рассчитывать, используя симметрическую гиперболическую систему. Скалярный потенциал Φ определяется с некоторой степенью произвола (калибровочная инвариантность) и связан с A равенствами

$$\partial A/\partial t + \operatorname{grad}\Phi = -E, \quad \partial\Phi/\partial t + \operatorname{div}A = 0, \tag{3.10}$$

первое из которых, по существу, является определением A. Вектор H напряженности магнитного поля связан с вектор-потенциалом A с помощью равенства $H = \operatorname{rot}A$, обеспечивающего автоматическое выполнение нужного нам условия $\operatorname{div}H = 0$.

Итак, наше предложение сводится к тому, чтобы основываться на решении симметрической гиперболической системы, составленной как из уравнений для E, H, так и из уравнений для A, Φ:

$$\begin{aligned} \partial E/\partial t - \operatorname{rot}H &= -j, \quad \partial H/\partial t + \operatorname{rot}E = 0, \\ \partial A/\partial t + \operatorname{grad}\Phi &= -E, \quad \partial\Phi/\partial t + \operatorname{div}A = 0. \end{aligned} \tag{3.11}$$

Второе из равенств (3.10) приводит к выводу, что правые части волновых уравнений для Φ, A обязаны подчиняться закону сохранения (3.9). Это обстоятельство оправдывает отсутствие в правых частях системы (3.11) значений ρ – плотности заряда.

Вычисленные по разностной аппроксимации уравнений (3.11) значения E и H, так же как и в предыдущем примере уравнений акустики, подвергнем корректировке.

На первом этапе, начиная со значений E, H, A, Φ в момент t, вычисляем их на следующий момент времени $t + \tau$. Полученные на момент $t + \tau$ значения H подвергаются корректировке – они заменяются на $H = \operatorname{rot}A$ (A на момент времени $t + \tau$). При этой корректировке производные компонент A_k вектора A, входящие в выражение $\operatorname{rot}A$, заменяются на центральные разностные отношения вида

$$\frac{A_k(x_i + h) - A_k(x_i - h)}{2h};$$

здесь x_i – переменная по одному из пространственных направлений, $i = 1, 2, 3$. Подкорректированный таким образом вектор H не обязан точно совпадать с вычисленным на первом этапе решения системы (3.11), так как на этом этапе вектор A находился из других уравнений.

Представляется разумным включить в корректировку вычисление заряда ρ с помощью разностных равенств, аппроксимирующих определение $\rho = \operatorname{div}E$. Результаты интегрирования на некотором заданном интервале времени подлежат контролю, который состоит, во-первых, в вычислении разностных отношений до некоторого достаточно высокого порядка, т.е. в проверке того, что разностное решение обладает нужной гладкостью, а во-вторых – в вычислении (с помощью простейших центральных разностных отношений, приближающих производные) погрешностей аппроксимации как уравнений, входящих в систему (3.11), так и дополнительных уравнений

$$\begin{aligned} H - \operatorname{rot}A &= 0, \quad \operatorname{div}H = 0, \\ \operatorname{div}E &= \rho, \quad \frac{\partial\rho}{\partial t} + \operatorname{div}j = 0. \end{aligned} \tag{3.12}$$

Как известно, выполнение расширенной системы уравнений Максвелла, состоящей как из (3.11), так и из (3.12), влечет за собой в качестве формального алгебраического следствия еще и справедливость законов сохранения энергии и тензора максвелловских натяжений

$$\begin{aligned} \frac{\partial T}{\partial t} + \frac{\partial T_k}{\partial x_k} + j_k E_k &= 0, \\ \frac{\partial T_i}{\partial t} + \frac{\partial T_{ik}}{\partial x_k} + \rho E_i + [j \times H]_i &= 0, \end{aligned} \tag{3.13}$$

где

$$T = \frac{1}{2}[(E, E) + (H, H)], \quad T_i = [E \times H]_i,$$

$$T_{ii} = E_i E_i + H_i H_i - T, \quad T_{ik} = E_i E_k + H_i H_k \quad (i \neq k).$$

Эти законы сохранения можно записать в матричном виде. Обозначим через $\mathrm{D} = (\partial/\partial t\ \partial/\partial x_1\ \partial/\partial x_2\ \partial/\partial x_3\ 1)$ и $F = j_1E_1 + j_2E_2 + j_3E_3$, $F_1 = j_2H_3 - j_3H_2 - \rho E_1$, $F_2 = j_3H_1 - j_1H_3 - \rho E_2$, $F_3 = j_1H_2 - j_2H_1 - \rho E_3$. Тогда

$$\left[\mathrm{D}\begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & -H_3 & H_2 \\ E_2 & H_3 & 0 & -H_1 \\ E_3 & -H_2 & H_1 & 0 \\ -\rho & j_1 & j_2 & j_3 \end{pmatrix}\right]\begin{bmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & -H_3 & H_2 \\ E_2 & H_3 & 0 & -H_1 \\ E_3 & -H_2 & H_1 & 0 \end{bmatrix} + \left[\mathrm{D}\begin{pmatrix} 0 & H_1 & H_2 & H_3 \\ H_1 & 0 & E_3 & -E_2 \\ H_2 & -E_3 & 0 & E_1 \\ H_3 & E_2 & -E_1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\right]\begin{bmatrix} 0 & H_1 & H_2 & H_3 \\ H_1 & 0 & E_3 & -E_2 \\ H_2 & -E_3 & 0 & E_1 \\ H_3 & E_2 & -E_1 & 0 \end{bmatrix} =$$

$$= \mathrm{D}\begin{bmatrix} T & T_1 & T_2 & T_3 \\ T_1 & T_{11} & T_{12} & T_{13} \\ T_2 & T_{21} & T_{22} & T_{23} \\ T_3 & T_{31} & T_{32} & T_{33} \\ F & F_1 & F_2 & F_3 \end{bmatrix}.$$

Поэтому если с помощью корректировки удастся добиться достаточно точного выполнения всех уравнений расширенной системы (3.11), (3.12), то естественно ожидать от разностного решения хорошей точности выполнения законов сохранения (3.13).

Аналогичная корректировка может быть использована для решения уравнений Максвелла в материальной среде. Пусть матрицы $K = K(x_1, x_2, x_3) = K^* > 0$, $M = M(x_1, x_2, x_3) = M^* > 0$, описывающие диэлектрическую поляризацию и магнитную проницаемость, связывают E, H с D и B:

$$D = KE, \quad B = MH.$$

Уравнения Максвелла и связанные с ними уравнения для потенциалов

$$\frac{\partial KE}{\partial t} - \mathrm{rot}H = -j, \quad \frac{\partial MH}{\partial t} + \mathrm{rot}E = 0,$$

$$\frac{\partial A}{\partial t} + \mathrm{grad}\Phi = -E, \quad \frac{\partial \Phi}{\partial t} + \mathrm{div}A = 0,$$

$$\rho = \mathrm{div}(KE)$$

влекут в качестве следствий соотношения

$$\frac{\partial}{\partial t}(\mathrm{rot}A) + \mathrm{rot}E = 0, \quad \frac{\partial}{\partial t}\mathrm{div(KE)} = -\mathrm{div}j,$$

$$\frac{\partial}{\partial t}(MH - \mathrm{rot}A) = 0, \quad \frac{\partial \rho}{\partial t} + \mathrm{div}j = 0,$$

обеспечивающие совместность системы с дополнительными уравнениями

$$H = M^{-1}\mathrm{rot}A, \quad \mathrm{div}B = 0, \quad \partial\rho/\partial t + \mathrm{div}j = 0.$$

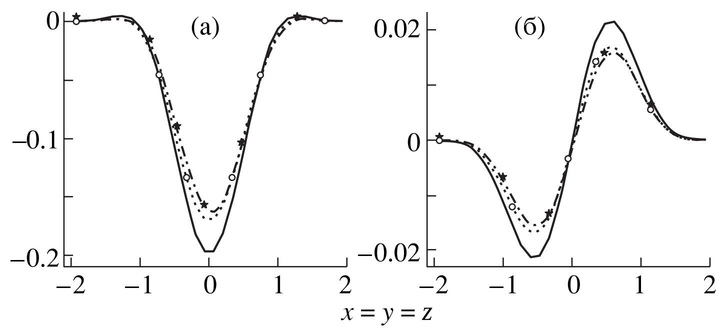
Первое из них позволяет перенести описанный выше прием корректировки и на случай электродинамики в материальных средах.

Для демонстрации описанного подхода рассмотрим решение уравнений Максвелла в трехмерном кубе $[-2; 2] \times [-2; 2] \times [-2; 2]$ с периодическими граничными условиями и нулевыми начальными условиями

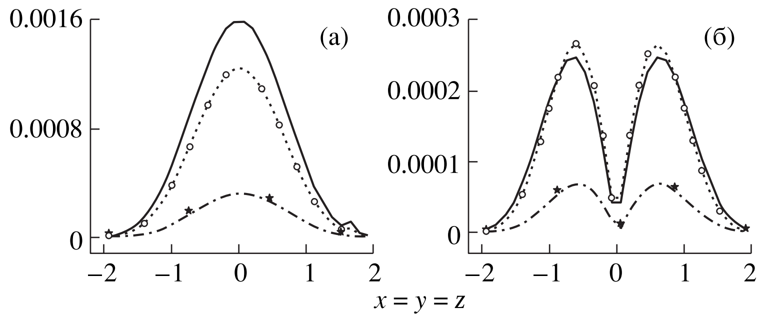
$$E|_{t=0} = H|_{t=0} = A|_{t=0} = 0, \quad \Phi|_{t=0} = 0.$$

Выберем значение j в виде

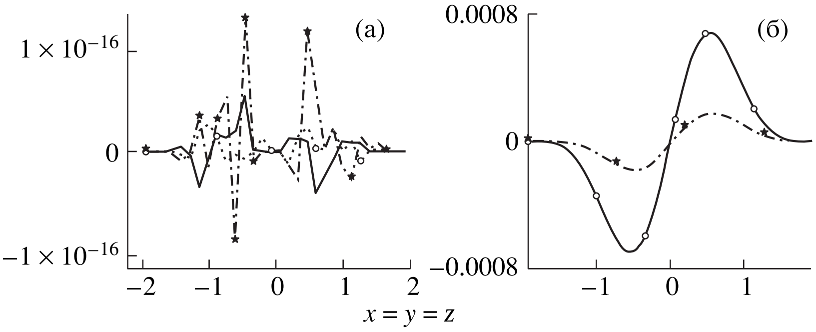
$$j = [j_x\ \ j_y\ \ j_z]^{\mathrm{T}} = t^2\left(\cos\frac{\pi x}{4}\cos\frac{\pi y}{4}\cos\frac{\pi z}{4}\right)^2[1\ \ 1/2\ \ 2]^{\mathrm{T}},$$



Фиг. 3.



Фиг. 4.



Фиг. 5.

а матрицы M и K зададим в диагональной форме: $M = \mathrm{diag}\{\mu, \mu, \mu\}$, $K = \mathrm{diag}\{\varepsilon, \varepsilon, \varepsilon\}$, взяв $\mu = 1.1$, $\varepsilon = 1.2$.

Расчеты проведены на кубических сетках $30 \times 30 \times 30$ и $60 \times 60 \times 60$ при 25 и 50 временны́х шагах соответственно. На фиг. 3–5 показаны профили сеточных функций вдоль диагонали куба $x = y = z$ для последнего момента времени. Результаты расчетов приведены для случаев $\mu = 1$, $\varepsilon = 1$ (тест 1) и $\mu = 1.1$, $\varepsilon = 1.2$ (тест 2). На фиг. 3 показаны профили компонент электрического (а) и магнитного (б) полей: сплошная линия соответствует решению для теста 1 на сетке $30 \times 30 \times 30$, символы $\circ$ и $*$ на пунктирных линиях отмечают графики для теста 2 на сетках $30 \times 30 \times 30$ и $60 \times 60 \times 60$ соответственно. Для оценки точности решаемых уравнений на последних трех временны́х слоях вычисляется невязка уравнений

$$\frac{\partial D}{\partial t} - \mathrm{rot} H = -j, \quad \frac{\partial B}{\partial t} + \mathrm{rot} E = 0,$$

где все входящие производные по пространству и времени аппроксимируются центральными разностными отношениями. На фиг. 4 показаны профили компонент невязки первого и второго из этих уравнений соответственно; обозначения такие же, как на фиг. 3.

Результаты проверки уравнения $\mathrm{div}H = 0$ и закона сохранения (3.9) приведены на фиг. 5а, б соответственно. Видно, что равенство $\mathrm{div}H = 0$ выполняется практически точно – невязка этого уравнения показана на фиг. 5а в масштабе 10^{-16}. Фиг. 5б подтверждает второй порядок аппроксимации, так как при уменьшении шага сетки вдвое невязка уравнения (3.9) уменьшается примерно в 4 раза.

По-видимому, аналогичную коррекцию можно применить и для уравнений магнитной гидродинамики (1.7):

$$\frac{\partial A}{\partial t} + \mathrm{grad}\Phi = \mathbf{u} \times L_{\mathbf{h}}, \quad \frac{\partial m\Phi}{\partial t} + \mathrm{div}A = 0,$$

$$\frac{\partial L_{q_0}}{\partial t} + \frac{\partial (u_k L) q_0}{\partial x_k} = 0,$$

$$\frac{\partial L_{\mathbf{h}}}{\partial t} + \frac{\partial (u_k L)_{\mathbf{h}}}{\partial x_k} - L_{\mathbf{h}}\frac{\partial \mathbf{u}}{\partial x_k} = 0,$$

$$\frac{\partial L_{\mathbf{u}}}{\partial t} + \frac{\partial (u_k L)_{u_i}}{\partial x_k} - L_{\mathbf{h}}\frac{\partial \mathbf{h}}{\partial x_k} = 0,$$

$$L_{\mathbf{h}} = \mathrm{rot}A.$$

При этом мы как бы отказались от упрощения – исключения потенциала электромагнитного поля Φ. Но в предлагаемой модели он носит чисто вспомогательный характер, обеспечивая выполнение дополнительного равенства $\mathrm{div}L_{\mathbf{h}} = 0$. В нашем распоряжении остается еще возможность выбора калибровочного множителя m, определяющего характеристический конус уравнений, добавляемых к системе (1.7).

4. О РЕШЕНИИ УРАВНЕНИЙ ТЕОРИИ УПРУГОСТИ

Теперь наметим, как уравнения теории упругости могут быть адаптированы к использованию описанного приема. Рассмотрим симметрическую гиперболическую систему

$$\begin{aligned} &\frac{\partial H_{u_i}}{\partial t} + \frac{\partial \sigma_{ik}}{\partial x_k} = 0, \\ &\frac{\partial H_{\sigma_{ik}}}{\partial t} + \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right) = 0, \end{aligned} \tag{4.1}$$

где

$$H = \rho_0 \frac{u_1^2 + u_2^2 + u_3^3}{2} + \frac{1}{\sigma(3\lambda + 2\mu)}(\sigma_{11} + \sigma_{22} + \sigma_{33})^2 +$$
$$+ \frac{1}{\mu}\left[\left(\sigma_{11} - \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{2}\right)^2 + \left(\sigma_{22} - \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{2}\right)^2 + \left(\sigma_{33} - \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{2}\right)^2 + 2(\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{31}^2)\right].$$

Ее обычно дополняют шестью совместными с ней уравнениями Сен-Венана. Два из них имеют вид

$$\begin{aligned} &\frac{\partial^2 \varepsilon_{11}}{\partial x_2^2} - \frac{\partial^2 \varepsilon_{12}}{\partial x_1 \partial x_2} + \frac{\partial^2 \varepsilon_{22}}{\partial x_1^2} = 0, \\ &\frac{\partial^2 \varepsilon_{12}}{\partial x_2^2} - \frac{\partial^2 \varepsilon_{13}}{\partial x_3 \partial x_2} - \frac{\partial^2 \varepsilon_{32}}{\partial x_1 \partial x_3} + 2\frac{\partial^2 \varepsilon_{33}}{\partial x_1 \partial x_2} = 0, \end{aligned} \tag{4.2}$$

а четыре других получаются из уравнений (4.2) циклической заменой индексов: (123) $\longrightarrow$ $\longrightarrow$ (231) $\longrightarrow$ (312). Элементы тензора напряжений σ_{ij} и тензора деформаций ε_{ij} связаны линейными соотношениями

$$\varepsilon_{ik} = H_{\sigma_{ik}}. \tag{4.3}$$

Пополняя систему (4.1) уравнениями (4.2), приходим к расширенной переопределенной системе. Если условия Сен-Венана выполняются в начальных данных, то их справедливость в последующие моменты времени обосновывается несложной выкладкой с помощью равенства (4.3) и второго уравнения системы (4.1). Чтобы при численном решении уравнений теории упругости обеспечить достаточно аккуратное выполнение условий Сен-Венана, удобно расширить систему уравнениями для смещений ζ_i:

$$\partial\zeta_i/\partial t = u_i. \tag{4.4}$$

Эти смещения играют роль потенциалов, через которые тензор деформации представляется в виде

$$\varepsilon_{ik} = \frac{1}{2}\left(\frac{\partial\zeta_i}{\partial\xi_k} + \frac{\partial\zeta_k}{\partial\xi_i}\right). \tag{4.5}$$

Описанная структура расширенной переопределенной системы уравнений линейной теории упругости позволяет применить при численном расчете прием корректировки. На первом этапе расчета временнóго шага $t \longrightarrow t + \tau$ решается основная система (4.1) и вычисляются смещения ζ_i путем решения уравнений (4.4). Из значений u_i, σ_{ik}, ζ_i, рассчитанных на момент $t + \tau$, удаляются все значения σ_{ik} и заменяются на σ_{ik}, вычисленные по значениям ζ_i (на этот момент времени $t + \tau$) с помощью формул (4.5) и соотношений (4.3).

5. ЗАКЛЮЧЕНИЕ

Приведенные предложения по конструированию разностных схем для переопределенных уравнений классической математической физики ни в коем случае не стоит считать окончательными. Мы лишь пытаемся наметить некоторые нестандартные направления исследований. Наше внимание к этим направлениям оказалось привлеченным во время размышлений над результатами многочисленных экспериментов, которые мы вели последние два с половиной года. Мы начали с задачи расчета частот колебаний, описываемых уравнениями акустики в двумерной области. Соответствующие собственные вектор-функции имеют нулевой вихрь скорости, если частота отлична от нуля. Нулевой же частоте отвечает бесконечномерное инвариантное подпространство, характеризуемое нулевым или постоянным давлением. При этом не зависящее от времени распределение вихря может быть совершенно произвольным. Занимаясь выделением низких ненулевых частот и соответствующих собственных функций путем анализа решений нестационарных уравнений акустики, возбуждаемых в процессе резонанса периодическим внешним воздействием, нам пришлось обратить особое внимание на удаление паразитических вихрей из результатов разностных расчетов. Эти расчеты проводились по простейшему варианту схемы Годунова первого порядка точности, которая паразитические вихри порождает.

Удаление вихрей в этих экспериментах удалось осуществить с помощью процедуры коррекции, основанной на методе Ритца. При этом производные, входящие в вариационный функционал для собственных функций, аппроксимировались простейшим образом центральными разностями, а в качестве допустимых функций рассматривались линейные комбинации разностных решений и отвечающих им остаточных членов. Эти остаточные члены получались после подстановки функций из грубого расчета в разностную схему второго порядка точности. Такая процедура приводила к выделению гладких собственных функций, несмотря на то, что использованная схема второго порядка для тех же собственных значений имела и другие собственные функции с паразитическими осцилляциями по пространству. Это неожиданное для нас явление было описано в [6]–[8] на примерах из акустики и теории упругости. По причине паразитических осцилляций упомянутая схема второго порядка точности обычно бракуется и на практике никогда не используется. У нас она сработала из-за того, что применялась к гладким решениям, полученным по другой схеме – схеме первого порядка точности.

Обдумывание явления, с которым мы в наших экспериментах столкнулись, повлекло за собой дальнейшую большую серию экспериментов, в которых мы пытались для одних и тех же величин использовать на разных этапах расчета различные схемы. На начальном этапе цель состоит в получении гладких решений разностных уравнений. На этапе коррекции мы позволяем себе эти гладкие решения “дифференцировать” с помощью простейших разностных отношений.

Именно так полученные разностные аппроксимации производных используются для коррекции и для проверки точности аппроксимации как основных уравнений, так и для дополнительных, входящих в расширенную систему. В настоящей работе мы рассказали о попытке, как нам кажется удачной, применить такой подход к решению переопределенных гиперболических уравнений.

Мы благодарны С.К. Селивановой за анализ критериев устойчивости некоторых используемых нами разностных схем.

СПИСОК ЛИТЕРАТУРЫ

1. *Иванов В.В.* Строго гиперболические полиномы, не допускающие симметризации: Препринт № 77. Новосибирск: Ин-т матем. СО АН СССР, 1984.
2. *Годунов С.К.* Разностный метод численного расчета разрывных решений уравнений гидродинамики // Матем. сб. 1959. Т. 47(89). № 3. С. 271–306.
3. *Годунов С.К., Гордиенко В.М.* Простейшие галилеево-инвариантные и термодинамически согласованные законы сохранения // Прикл. механ. и техн. физ. 2002. Т. 43. № 1. С. 1–12.
4. *Годунов С.К., Гордиенко В.М.* Усложненные структуры галилеево-инвариантных законов сохранения // Прикл. механ. и техн. физ. 2002. Т. 43. № 2. С. 175–189.
5. *Куликовский А.Г., Погорелов Н.В., Семенов А.Ю.* Математические вопросы численного решения гиперболических систем уравнений. М.: Физматлит, 2001.
6. *Годунов С.К., Селиванова С.В.* Эксперименты по использованию резонанса для спектрального анализа конечномерных кососимметрических операторов // Сибирский ж. вычисл. матем. 2006. Т. 9. № 2. С. 123–136.
7. *Годунов С.К., Жуков В.Т., Феодоритова О.Б.* Метод расчета инвариантных подпространств для симметрических гиперболических уравнений // Ж. вычисл. матем. и матем. физ. 2006. Т. 46. № 6. С. 1019–1031.
8. *Godunov S.K., Feodoritova O.B., Zhukov V.T.* Computation of eigenspaces of hyperbolic system // IV Internat. Conf. Comput. Fluid Dynamics. July 10–14. Ghent: Belguim, 2006.
9. Численное решение многомерных задач газовой динамики / Под ред. С.К. Годунова. М.: Наука, 1976. (French transl.: Resolution numerique des problemes miltidimensionnels de la dynamique des gaz. Moscou: Mir, 1979).

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 460–480

УДК 519.63

APPROXIMATION OF THE SOLUTION AND ITS DERIVATIVE FOR THE SINGULARLY PERTURBED BLACK–SCHOLES EQUATION WITH NONSMOOTH INITIAL DATA[1)]

© 2007 г. S. Li*, G. I. Shishkin, L. P. Shishkina****

(Department of Computational Science, National University of Singapore, Singapore 117543;*
*** Institute of Mathematics and Mechanics, Ural Division, Russian Academy of Sciences, Yekaterinburg 62019, Russia)*
e-mail: shishkin@imm.uran.ru
Received July 10, 2006

Аппроксимация решения и производной для сингулярно возмущенного уравнения Блэка–Шоулза с негладкими начальными данными. Ш. Ли, Г.И. Шишкин, Л.П. Шишкина. Задача для уравнения Блэка–Шоулза, возникающая в финансовой математике, преобразованием переменных приводится к задаче Коши для сингулярно возмущенного параболического уравнения в переменных x, t с возмущающим параметром ε, $\varepsilon \in (0, 1]$. Эта задача имеет такие особенности, как бесконечная область, ограниченная гладкость начальной функции (ее производная первого порядка по x терпит разрыв I рода в точке $x = 0$), переходный слой (движущийся во времени), порождаемый кусочно-гладкой начальной функцией при малых значениях параметра ε, и др. Рассматривается сеточная аппроксимация решения задачи и его первой производной на конечной области, содержащей переходный слой. На равномерной сетке с использованием метода аддитивного выделения особенности типа переходного слоя строится специальная разностная схема, аппроксимирующая ε-равномерно решение задачи и его первую производную по x с порядками скорости сходимости, близкими к 1 и 0.5 соответственно. Эффективность построенной схемы иллюстрируется численными экспериментами. Библ. 6. Фиг. 1. Табл. 7.

Keywords: Black–Scholes equation, singularly perturbed parabolic equation, nonsmooth initial data, interior layer, difference scheme, additive splitting of singularities, convergence.

1. INTRODUCTION

1.1. Mathematical modeling in financial mathematics leads to the Black–Scholes equation (that is backward parabolic) [1] with respect to the value $C = C(S, t')$, which is a European call option, where S and t' are the current values of the underlying asset and time,

$$\frac{\partial C}{\partial t'} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0, \quad (S, t') \in \mathbb{R}^+ \times [0, T), \tag{1.1a}$$

with the final condition

$$C(S, T) = \max(S - E, 0), \quad S \in \mathbb{R}^+, \tag{1.1b}$$

and the boundary conditions at $S = 0$ and at infinity $S = +\infty$

$$C(0, t') = 0; \quad C(S, t') \longrightarrow S \text{ for } S \longrightarrow \infty, \quad t' \in [0, T). \tag{1.1c}$$

Here σ, E, T and r are some financial parameters (the volatility, exercise price, expiry time and the interest rate, respectively).

For the problem (1.1), in addition to the solution itself, some of the partial derivatives of the solution are of interest (see [1, Ch. 3]).

When studying this problem, a standard approach is a transformation of the equation by the changes of variables.

1) Работа выполнена при финансовой поддержке Академического исследовательского фонда NUS (грант № R-151-000-025-112), РФФИ (коды проектов № 04-01-00578, 04-01-89007-НВО_а) и частично Нидерландской организации научных исследований NWO (проект № 047.016.008); и Булевского центра исследования по информатике Национального ун-та Ирландии в Корке.

By the transformations

$$S = E\exp(x), \quad t' = T - \tau r^{-1}, \quad C = Ev(x, \tau) \tag{1.1d}$$

and introducing the notation $k = 2\sigma^{-2}r$, $\tau^* = rT$, we come to the following problem for the dimensionless parabolic equation in the new variables x, τ:

$$Lv(x, \tau) \equiv \left\{\frac{\partial^2}{\partial x^2} + (k-1)\frac{\partial}{\partial x} - k - k\frac{\partial}{\partial \tau}\right\}v(x, \tau) = 0, \tag{1.2}$$
$$(x, \tau) \in \mathbb{R} \times (0, \tau^*]$$

with the initial condition

$$v(x, 0) = \varphi_v(x), \quad x \in \mathbb{R}, \tag{1.3a}$$

where

$$\varphi_v(x) = \max(\exp(x) - 1, 0), \quad x \in \mathbb{R},$$

and with the condition at infinity

$$\left.\begin{array}{l} v(x, \tau) \longrightarrow 0 \text{ for } x \longrightarrow -\infty \\ v(x, \tau) \longrightarrow \exp(x) \text{ for } x \longrightarrow \infty \end{array}\right\}, \quad \tau \in (0, \tau^*]. \tag{1.3b}$$

Under the condition T, $r = O(1)$ and for σ taking an arbitrary value from the half-open interval $(0, \sqrt{2r})$, we come to the Cauchy problem for the singularly perturbed parabolic equation

$$Lv(x, \tau) \equiv \left\{\varepsilon\frac{\partial^2}{\partial x^2} + (1-\varepsilon)\frac{\partial}{\partial x} - 1 - \frac{\partial}{\partial \tau}\right\}v(x, \tau) = 0, \quad (x, \tau) \in \mathbb{R} \times (0, \tau^*] \tag{1.4}$$

with conditions (1.3). Here $\varepsilon = 2^{-1}\sigma^2 r^{-1}$ is a dimensionless "perturbation" parameter, $\varepsilon \in (0, 1]$.

The initial function in condition (1.3a) is continuous; its first derivative in x has a discontinuity of the first kind at the point $x = 0$

$$\left[\frac{d}{dx}\varphi_v(0)\right] = 1,$$

where the jump of the derivative is defined by the relation

$$\left[\frac{d}{dx}\varphi_v(0)\right] = \lim_{x \searrow 0}\frac{d}{dx}\varphi_v(x) - \lim_{x \nearrow 0}\frac{d}{dx}\varphi_v(x).$$

The initial function and the solution itself for this problem grow (exponentially) without bound as $x \longrightarrow \infty$. If the parameter $\varepsilon = 1$ then the problem (1.4), (1.3) becomes the one of reaction-diffusion type, and for $\varepsilon < 1$, it is of convection-diffusion type. For small values of the parameter ε, an interior (moving in time) layer with the typical width of $\varepsilon^{1/2}$ appears in a neighbourhood of the characteristic (of the operator $L_1 \equiv (1-\varepsilon)\frac{\partial}{\partial x} - 1 - \frac{\partial}{\partial \tau}$) passing through the point (0, 0).

Thus, the Cauchy problem (1.4), (1.3) is a singularly perturbed problem with different types of singularities. In the present paper we are interested in approximations to both the solution and its first order derivative in a finite subdomain that contains the singularity of the interior layer type.

1.2. Boundary value problems in bounded domains for singularly perturbed parabolic reaction-diffusion equations with a discontinuous initial condition have been considered in [2]–[7]. To construct schemes that converge ε-uniformly, the method of condensing meshes (in a neighbourhood of boundary layers), and also either the fitted operator method [2]–[5] or the method of additive splitting of a singularity [6], [7] (in a neighbourhood of the points at which the initial function is discontinuous) were applied.

In [2]–[6], approximations to the normalized derivatives $\varepsilon(\partial/\partial x)u(x, t)$, i.e., the first order spatial derivative multiplied by the parameter ε, were considered. For this purpose, the method of additive splitting of the singularity generated by the discontinuity of the initial function was used; however, the approximation of the derivative $(\partial/\partial x)u(x, t)$ itself was not considered.

A boundary value problem on a segment for singularly perturbed parabolic convection-diffusion equations with a piecewise smooth initial condition has been considered in [8], [9]. In [9], by using the method of special meshes that condense in a neighbourhood of the boundary layer and the method of the additive splitting of a singularity of the interior layer type, special difference schemes are constructed that make it possible to approximate ε-uniformly the solution of the problem on the entire set under consideration, the normalized derivative on the entire set except for the discontinuity point (0, 0), and the first spatial derivative on the same set but outside a small neighbourhood of the boundary layer.

In the present paper, instead of the Cauchy problem (1.4), (1.3), we consider a singularly perturbed boundary value problem for equation (1.4) with a non-smooth initial condition similar to (1.3), namely, the problem (2.2), (2.1) (see the formulation of this problem in Section 2). The technique from [9] is used for studying the problem (2.2), (2.1). Note that in a problem of the type (2.2), (2.1) considered in a finite domain, except for the interior layer, an additional singularity appears, namely, a boundary layer with the typical width of ε. The singularity of the boundary layer is stronger than that of the interior layer, which makes it difficult to construct special numerical methods suitable for the adequate description of the singularity of the interior layer type. In contrast to [9], here conditions are defined that allow us to investigate each singularity of the problem separately. For the boundary value problem (2.2), (2.1), we construct a finite difference scheme that approximates the solution and its first order derivative in x. To construct ε-uniform approximations for the solution and its first derivative in a finite subdomain including only the interior-layer singularity, it suffices to use a uniform mesh and the method of the additive splitting of the singularity of the interior layer type. The efficiency of the scheme constructed in this paper is verified with numerical experiments.

The numerical method constructed for problem (2.2), (2.1), after the transformation to the original variables S, t' and the function C (see the change (1.1d)), allows us to approximate the solution of problem (1.1) and its first derivative $(\partial/\partial S)C(S, t')$ in a finite neighbourhood of the point (E, T) (the point of discontinuity of the derivative in condition (1.1b)), including the interior layer (appearing for small values of the dimensionless quantity $\sigma^2 r^{-1}$). Errors in the approximation of the solution and derivative (for $(S, t') \neq (E, T)$) are independent of the value of $\sigma^2 r^{-1}$; these errors (in the maximum norm) are defined only by the number of nodes in the mesh used for the numerical solution of the discrete problem.

About Contents. Formulation of the boundary value problem is given in Section 2. Difficulties involved in the approximation of the solution and derivatives on the basis of classical finite difference schemes are discussed in Section 3. *A priori* estimates used in the constructions are presented in Section 4. Classical difference approximations of the problem on uniform and piecewise uniform meshes are considered in Section 5. A difference scheme (approximating the solution and its first order derivative), which is constructed using the method of the additive splitting of the singular component of the solution generated by the discontinuity of the derivative of the initial function, is given in Section 6. In the same place, conditions are defined under which a certain singularity of the solution can be split off and investigated separately. Numerical experiments are analyzed in Section 7.

2. PROBLEM FORMULATION. AIM OF RESEARCH

2.1. On the set $\overline{G}$ with the boundary S,

$$\overline{G} = G \cup S, \quad G = D \times (0, T], \quad D = \{x : x \in (-d, d)\}, \tag{2.1}$$

we consider the Dirichlet problem for the singularly perturbed parabolic convection-diffusion equation[2)]

$$L_{(2.2a)}u(x, t) = f(x, t), \quad (x, t) \in G, \tag{2.2a}$$

$$u(x, t) = \varphi(x, t), \quad (x, t) \in S. \tag{2.2b}$$

Here

$$L_{(2.2a)} \equiv \varepsilon a\frac{\partial^2}{\partial x^2} + b\frac{\partial}{\partial x} - c - q\frac{\partial}{\partial t},$$

a, b, $q > 0$, $c \geq 0$, the right-hand side $f(x, t)$ is a sufficiently smooth function on $\overline{G}$; the parameter ε takes arbitrary values in the half-open interval (0, 1]. The boundary function $\varphi(x, t)$ is sufficiently smooth on the

2) The notation $L_{(j.k)}$, ($m_{(j.k)}$, $M_{(j.k)}$, $G_{h(j.k)}$)) means that these operators (constants, meshes) are introduced in formula ($j.k$).

sets $\bar{S}_0^-$, $\bar{S}_0^+$, $\bar{S}^L$ and is continuous on S; the first order derivative in x of the function $\varphi(x, t)$ has a discontinuity of the first kind on the set $S^{(*)} = \{(0, 0)\}$, i.e.,

$$\left[\frac{\partial}{\partial x}\varphi(x, t)\right] \neq 0, \quad (x, t) \in S^{(*)}. \tag{2.2c}$$

Here

$$S_0^- = \{(x, t) : x \in [-d, 0), t = 0\},$$

$$S_0^+ = \{(x, t) : x \in (0, d], t = 0\}, \quad S_0 = \bar{S}_0^- \cup \bar{S}_0^+,$$

S_0 and S^L are the lower and lateral parts of the boundary S, $S^L = \Gamma \times (0, T]$, $\Gamma = \bar{D} \backslash D$.

The solution of problem (2.2) is a function $u \in C(\bar{G}) \cap C^{2,1}(G)$ satisfying the differential equation on G and the boundary conditions on S.

For simplicity, we assume that compatibility conditions that ensure the smoothness of the solution for fixed values of ε [10] are fulfilled on the set $S_* = S_0 \cap \bar{S}^L$. Let $\bar{G}^\delta$ be the δ-neighbourhood of the set S_*, i.e.,

$$\bar{G}^\delta = \{(x, t) : r((x, t), S_*) \le \delta\},$$

where $r((x, t), S_*)$ is the distance from the point (x, t) to the set S_*. We suppose that the following inclusion holds on the set $\bar{G}^\delta$:

$$u \in C^{l+\alpha, (l+\alpha)/2}(\bar{G}^\delta), \quad l \ge 2, \quad \alpha \in (0, 1). \tag{2.3}$$

It follows from [10] that, under the condition (2.3), the solution of the problem (for sufficiently smooth functions $f(x, t)$ on $\bar{G}$ and $\varphi(x, t)$ on $\bar{S}_0^-$, $\bar{S}_0^+$, $\bar{S}^L$) is smooth on the set

$$\bar{G}^* = \bar{G} \backslash S^{(*)}, \tag{2.4}$$

i.e., $u \in C^{l+\alpha, (l+\alpha)/2}(\bar{G}^*)$. The derivative $(\partial/\partial x)u(x, t)$ is continuous on $\bar{G}^*$, bounded on $\bar{G}^*$ for fixed values of ε and has a discontinuity on the set $S^{(*)}_{(2.2c)}$.

Under the condition

$$a = c = p = 1, \quad b = 1 - \varepsilon, \quad f(x, t) = 0, \quad (x, t) \in \bar{G},$$

the equation (2.2a) becomes the equation (1.4).

We are interested in an approximation of the solution $u(x, t)$, $(x, t) \in \bar{G}$, and of the derivative $(\partial/\partial x)u(x, t)$, $(x, t) \in \bar{G}^*$. Let us describe the behaviour of the solution and derivatives more precisely.

Let $S^L = S^l \cup S^r$, S^l and S^r be the left and right parts of the boundary S^L, and let

$$S^\gamma = \{(x, t) : x = \gamma(t), (x, t) \in \bar{G}\}, \quad \gamma(t) = -bq^{-1}t, \quad t \ge 0,$$

be the characteristic of the reduced equation passing through the point (0, 0). When the parameter ε tends to zero, boundary and interior layers with the typical length scales ε and $\varepsilon^{1/2}$, respectively, appear in a neighbourhood of the sets S^l and S^γ; as opposed to the boundary layer, the interior layer is weak (the first order derivative in x of the interior-layer function is bounded ε-uniformly).

For simplicity, we assume that the characteristic S^γ does not meet the boundary S^l. The derivative $(\partial/\partial x)u(x, t)$ (denoted by $p(x, t)$) in a neighbourhood of the set S^l grows without bound as $\varepsilon \longrightarrow 0$. It is convenient to consider the quantity $P(x, t) = \varepsilon(\partial/\partial x)u(x, t)$, i.e., the normalized first derivative in x, in an m-neighbourhood of the set S^l, instead of the derivative $(\partial/\partial x)u(x, t)$, because this quantity $P(x, t)$ is bounded ε-uniformly. Outside a neighbourhood of the set S^l, the derivative $(\partial/\partial x)u(x, t)$ is bounded ε-uniformly. The quantity $P(x, t)$ will be called the diffusion flux (or briefly, the flux). Outside the neighbourhood of the set S^l, the derivative $p(x, t)$ is bounded ε-uniformly on $\bar{G}^*$. For small values of the parameter ε, the derivative $p(x, t)$ is more "informative" (on the set where it is bounded) than the flux $P(x, t)$.

2.2. It is well known (see, e.g., [11]) that even in the case of singularly perturbed problems with sufficiently smooth data, solutions of classical finite difference schemes do not converge ε-uniformly; for small values of the parameter ε, errors in the discrete solutions are commensurable with the actual solutions of the differential problem. The diffusion fluxes obtained on the basis of such schemes also do not converge ε-uniformly. It will be shown in Section 3 that for a boundary value problem whose solution is regular, classical difference schemes do not allow one to obtain ε-uniformly convergent approximations of the derivative in x.

Due to this it would be interesting to construct a difference scheme that allows us to approximate ε-uniformly both the solution on the whole domain $\overline{G}$ and diffusion fluxes in this domain excluding the discontinuity point $S^{(*)}$. Also, it would be interesting to determine conditions under which the boundary layer does not appear, and for such a problem, to find the ε-uniform approximation of the derivative in x on the set $\overline{G}*$.

Definition 1. Let

$$\overline{G}_0^* = \overline{G}_0^*(m) = \overline{G}* \cap \{x \geq -d+m\} \tag{2.5}$$

be the set $\overline{G}*$ excluding an m-neighbourhood[3)] of the set $\bar{S}^l$ (the m-neighbourhood of the boundary layer). If the interpolants constructed using the solution of some finite difference scheme converge on $\overline{G}$ ε-uniformly, we say that the discrete solution (the difference scheme) converges on $\overline{G}$ uniformly with respect to the parameter ε (or, briefly, ε-uniformly in $C(\overline{G})$). If, moreover, the interpolants of the diffusion fluxes (the first order derivatives in x) converge ε-uniformly on $\overline{G}$ (ε-uniformly on $\overline{G}_0^*$), we say that the difference scheme converges ε-uniformly in $C^{1(n)}(\overline{G}*)$ (ε-uniformly in $C^1(\overline{G}_0^*)$).

Thus, it is attractive to find numerical methods that converge ε-uniformly in $C^{1(n)}(\overline{G}*) \cap C^1(\overline{G}_0^*)$, where $\overline{G}_0^* = \overline{G}_0^*(m)$, moreover, it is required that the value m could be chosen sufficiently small.

Our aim is to construct a difference scheme for problem (2.2), (2.1) that converges ε-uniformly in $C^{1(n)}(\overline{G}*) \cap C^1(\overline{G}_0^*)$, and also to determine conditions under which the boundary layer does not appear, and in this case to construct a difference scheme that converges ε-uniformly in $C^1(\overline{G}*)$.

Some preliminary results related to this problem are given in [12], [13]. To investigate the problem, a technique similar to that developed in [9] is used. In the present paper, in contrast to [9], the main attention is given to the study of a singularity of the interior layer type, because the boundary layer does not arise in the original problem (1.1).

3. ON THE APPROXIMATION OF THE DERIVATIVE IN x

3.1. Let us discuss difficulties arising in the approximation of derivatives for the regular components of the problem solution, i.e., when the solution of a singularly perturbed problem is regular (not containing the singular component) and sufficiently smooth. In this case the solution of a classical finite difference scheme on a uniform mesh converges to the exact solution ε-uniformly. However, its difference derivatives are no longer convergent ε-uniformly; the error in the derivative of the solution of the grid problem may have the order of the derivative of the solution itself for the differential problem.

Consider the stationary problem

$$L_{(3.1)}u(x) \equiv \left\{\varepsilon\frac{d^2}{dx^2} + \frac{d}{dx}\right\}u(x) = f(x), \quad x \in D,$$
$$u(x) = \varphi(x), \quad x \in \Gamma. \tag{3.1a}$$

Here

$$\overline{D} = [0, 1]. \tag{3.1b}$$

Let $u(x) = x^2$, $x \in \overline{D}$, be a solution of problem (3.1); this solution has no singular component.

[3)] Throughout this paper, M, M_i (or m) denote sufficiently large (small) positive constants that do not depend on ε and on the discretization parameters.

To solve problem (3.1), we apply the classical difference scheme [14]. On the segment $\overline{D}$, we introduce the uniform mesh

$$\overline{D}_h = \overline{D}_{h(3.2)} \tag{3.2}$$

with the step-size $h = N^{-1}$. On the mesh $\overline{D}_h$, we consider the difference scheme

$$\begin{gathered}\Lambda_{(3.3)}z(x) \equiv \{\varepsilon\delta_{x\bar{x}} + \delta_x\}z(x) = f(x), \quad x \in D_h,\\ z(x) = \varphi(x), \quad x \in \Gamma_h.\end{gathered} \tag{3.3}$$

The solution of problem (3.3), (3.2) has the explicit form

$$z(x_i) = x_i^2 + N^{-1}\{1 - x_i - [(1 + \varepsilon^{-1}N^{-1})^{-i} - (1 + \varepsilon^{-1}N^{-1})^{-N}][1 - (1 + \varepsilon^{-1}N^{-1})^{-N}]^{-1}\},$$

$$x_i = iN^{-1} \in \overline{D}_h, \quad i \le N.$$

Thus, the function $z(x)$, $x \in \overline{D}_h$, converges to $u(x)$, $x \in \overline{D}$, ε-uniformly with the estimate

$$|u(x) - z(x)| \le N^{-1}, \quad x \in \overline{D}_h.$$

For the first discrete derivative

$$\delta_x z(x_i) = 2x_i + N^{-1}(\varepsilon + N^{-1})^{-1}(1 + \varepsilon^{-1}N^{-1})^{-i}[1 - (1 + \varepsilon^{-1}N^{-1})^{-N}]^{-1},$$

$$x_i = iN^{-1} \in \overline{D}_h, \quad i \le N - 1$$

we have the error

$$\left|\frac{d}{dx}u(x_i) - \delta_x z(x_i)\right| = N^{-1}(\varepsilon + N^{-1})^{-1}(1 + \varepsilon^{-1}N^{-1})^{-i}[1 - (1 + \varepsilon^{-1}N^{-1})^{-N}]^{-1}, \quad x_i \in \overline{D}_h, \quad x_i < 1;$$

$$\max_{\overline{D}_h,\, x < 1}\left|\frac{d}{dx}u(x) - \delta_x z(x)\right| \ge mN^{-1}(\varepsilon + N^{-1})^{-1}.$$

Thus, the discrete derivative does not converge ε-uniformly: when $\varepsilon \le N^{-1}$, the error in this discrete derivative is of the order of the derivative itself.

3.2. If the solution of the boundary value problem (2.2), (2.1) is regular, moreover, the regular component is of the order of unity in a neighbourhood of the boundary layer, then the error in the approximation of the derivative $(\partial/\partial x)u(x, t)$, in general, grows unboundedly under the condition $(\varepsilon + N^{-1})^{-1}N_0^{-1} \longrightarrow \infty$ as $N, N_0 \longrightarrow \infty$, where $N + 1$ and $N_0 + 1$ are the numbers of nodes with respect to x and t in the uniform mesh on $\overline{G}$.

4. A PRIORI ESTIMATES OF THE SOLUTION AND DERIVATIVES

For the solution of the boundary value problem (2.2), (2.1) and its derivatives, we give *a priori* estimates used in the constructions (the more detailed derivation can be found in [9]).

4.1. We represent the set $\overline{G}$ as the sum of overlapping sets

$$\overline{G} = \bigcup_j \overline{G}^j, \quad j = 1, 2, 3, \tag{4.1}$$

where

$$G^1 = G^1(m^1) = \{(x, t) : |x - \gamma(t)| < m^1, t \in (0, T]\},$$

$$G^2 = G^2(m^2) = \{(x, t) : x \in (-d, -d + m^2), t \in (0, T]\},$$

$$G^3 = G^3(m^3) = G \backslash \{G^1(m^3) \cup G^2(m^3)\}, \quad m^3 < m^1, m^2,$$

G^1 and G^2 are neighbourhoods of the interior and boundary layers, respectively. The solution of problem (2.2), (2.1) considered on the set $\overline{G}^j$ will be also denoted by the $u^j(x, t)$, $j = 1, 2, 3$.

The solution on the set $\overline{G}^3$ is smooth; we have the estimate

$$\left|\frac{\partial^{k+k_0}}{\partial x^k \partial t^{k_0}} u^3(x, t)\right| \le M, \quad (x, t) \in \overline{G}^3, \quad k + 2k_0 \le K. \tag{4.2}$$

The value K is defined by the problem data; and $K \ge 4$.

We represent the solution on the set $\overline{G}^2$ as the decomposition into two functions

$$u(x, t) = U(x, t) + V(x, t), \quad (x, t) \in \overline{G}^2, \tag{4.3}$$

where $U(x, t)$ and $V(x, t)$ are the regular and singular components of the solution. For the functions $U(x, t)$ and $V(x, t)$, the following estimates are valid:

$$\left|\frac{\partial^{k+k_0}}{\partial x^k \partial t^{k_0}} U(x, t)\right| \le M, \tag{4.4a}$$

$$\left|\frac{\partial^{k+k_0}}{\partial x^k \partial t^{k_0}} V(x, t)\right| \le M\varepsilon^{-k} \exp[-m\varepsilon^{-1} r((x, t), \bar{S}^l)], \tag{4.4b}$$

$$(x, t) \in \overline{G}^2, \quad k + 2k_0 \le K,$$

where $r((x, t), \bar{S}^l)$ is the distance from the point (x, t) to the set $\bar{S}^l$, m is an arbitrary constant from the interval $(0, m_0)$, $m_0 = a^{-1}b$.

4.2. On the set $\overline{G}^1$, we introduce the new variables

$$\tilde{u}(\xi, t) = u(x(\xi, t), t)\exp(\alpha t), \quad (\xi, t) \in \bar{\tilde{G}}^1, \quad \xi = x - \gamma(t), \quad (x, t) \in \overline{G}^1. \tag{4.5}$$

Here $\gamma(t) = -bq^{-1}t$, $\alpha = cq^{-1}$, and $\tilde{G}^1$ is the image of the set G^1. The function $\tilde{u}(\xi, t)$, $(\xi, t) \in \tilde{G}^1$, is the solution of the problem

$$\begin{gathered} L_{(4.6a)}\tilde{u}(\xi, t) \equiv \left\{\varepsilon a \frac{\partial^2}{\partial \xi^2} - q\frac{\partial}{\partial t}\right\}\tilde{u}(\xi, t) = \tilde{f}(\xi, t), \quad (\xi, t) \in \tilde{G}^1, \\ \tilde{u}(\xi, t) = \begin{cases} \tilde{u}^3(\xi, t), & (\xi, t) \in \tilde{S}^1 \backslash \tilde{S}, \\ \tilde{\varphi}(\xi, t), & (\xi, t) \in \tilde{S}^1 \cap \tilde{S}. \end{cases} \end{gathered} \tag{4.6a}$$

Here $\tilde{S}^1 = \bar{\tilde{G}}^1 \backslash \tilde{G}^1$, and $\tilde{v}(\xi, t)$ is the image of the function $v(x, t)$,

$$\tilde{v}(\xi, t) = v(x(\xi, t), t)\exp(\alpha t), \tag{4.6b}$$

where $v(x, t)$ is one of the functions $u(x, t)$, $f(x, t)$, $(x, t) \in \overline{G}^1$, $\varphi(x, t)$, $(x, t) \in S^1 \cap \{t = 0\}$, or $u^3(x, t)$, $(x, t) \in$ $\in \overline{G}^1 \cap \overline{G}^3$, $u^3(x, t) = u(x, t)$, $(x, t) \in \overline{G}^3$.

The function $\tilde{u}(\xi, t)$ is decomposed into the sum of functions

$$\tilde{u}(\xi, t) = \tilde{U}^1(\xi, t) + \tilde{W}^1(\xi, t), \quad (\xi, t) \in \bar{\tilde{G}}^1, \tag{4.7a}$$

where $\tilde{U}^1(\xi, t)$ and $\tilde{W}^1(\xi, t)$ are the regular ("sufficiently" smooth) and singular parts of the solution. The function $\tilde{W}^1(\xi, t)$ is the solution of the Cauchy problem

$$\begin{gathered} L_{(4.6)}\tilde{W}^1(\xi, t) = 0, \quad (\xi, t) \in \mathbb{R} \times (0, T], \\ \tilde{W}^1(\xi, t) = \tilde{\Phi}^1_W(\xi), \quad \xi \in \mathbb{R}, \quad t = 0. \end{gathered} \tag{4.8}$$

Here

$$\tilde{\Phi}_W^1(\xi) = 2^{-1}\left[\frac{\partial}{\partial\xi}\tilde{\varphi}(0, 0)\right]|\xi|, \quad \xi \in \mathbb{R},$$

$\left[\frac{\partial}{\partial\xi}\tilde{\varphi}(0, 0)\right] = \frac{\partial}{\partial\xi}\tilde{\varphi}(+0, 0) - \frac{\partial}{\partial\xi}\tilde{\varphi}(-0, 0)$ is the jump of the derivative $\left[\frac{\partial}{\partial\xi}\tilde{\varphi}(\xi, t)\right]$. The function $\tilde{W}^1(\xi, t)$ takes the form

$$\tilde{W}^1(\xi, t) = 2^{-1}\left[\frac{\partial}{\partial\xi}\tilde{\varphi}(0, 0)\right]\tilde{w}_1(\xi, t),$$

$$\begin{gathered}\tilde{w}_1(\xi, t) = \xi v(2^{-1}\varepsilon^{-1/2}a^{-1/2}q^{1/2}\xi t^{-1/2}) + 2\pi^{-1/2}\varepsilon^{1/2}a^{1/2}q^{-1/2}t^{1/2}\exp(-4^{-1}\varepsilon^{-1}a^{-1}q\xi^2t^{-1}),\\ (\xi, t) \in \mathbb{R}\times[0, T],\end{gathered} \tag{4.7b}$$

$$v(\xi) = \operatorname{erf}(\xi) = 2\pi^{-1/2}\int_0^{\xi}\exp(-\alpha^2)d\alpha, \quad \xi \in \mathbb{R}.$$

The function $\tilde{U}^1(\xi, t)$ is the solution of the problem

$$L_{(4.6)}\tilde{U}^1(\xi, t) = \tilde{f}(\xi, t), \quad (\xi, t) \in \tilde{G}^1,$$

$$\tilde{U}^1(\xi, t) = \begin{cases}\tilde{u}^3(\xi, t) - \tilde{W}^1(\xi, t), & (\xi, t) \in \tilde{S}^1, \quad t > 0,\\ \tilde{\varphi}(\xi, t) - \tilde{\Phi}_W^1(\xi, t), & (\xi, t) \in \tilde{S}^1, \quad t = 0.\end{cases}$$

Note that the first derivative of the function $\tilde{U}^1(\xi, t)$ in ξ is continuous at $t = 0$.

4.3. In the variables x, t, the following representation corresponds to representation (4.7a):

$$u(x, t) = U^1(x, t) + W^1(x, t), \quad (x, t) \in \bar{G}^1, \tag{4.9a}$$

where

$$\begin{gathered}W^1(x, t) = \tilde{W}^1(\xi(x, t), t)\exp(-\alpha t) = 2^{-1}\left[\frac{\partial}{\partial x}\varphi(0, 0)\right]\{(x-\gamma(t))v(2^{-1}\varepsilon^{-1/2}a^{-1/2}q^{1/2}[x-\gamma(t)]t^{-1/2}) +\\ + 2\pi^{-1/2}\varepsilon^{1/2}a^{1/2}q^{-1/2}t^{1/2}\exp(-4^{-1}\varepsilon^{-1}a^{-1}q[x-\gamma(t)]^2t^{-1})\}\exp(-\alpha t), \quad (x, t) \in \mathbb{R}\times[0, T].\end{gathered} \tag{4.9b}$$

For the components in representation (4.9), we have the estimates (similar to those in [9])

$$\begin{gathered}\left|\frac{\partial^{k+k_0}}{\partial x^k\partial t^{k_0}}U^1(x, t)\right| \le M[1 + \varepsilon^{(2-k-k_0)/2}\rho^{2-k-k_0} + \varepsilon^{(2-k)/2}\rho^{2-k-2k_0}], \quad (x, t) \in \bar{G}^1,\\
\left|\frac{\partial^{k+k_0}}{\partial x^k\partial t^{k_0}}W^1(x, t)\right| \le M[1 + \varepsilon^{(1-k-k_0)/2}\rho^{1-k-k_0} + \varepsilon^{(1-k)/2}\rho^{1-k-2k_0}]\exp(-m\varepsilon^{-1/2}|x-\gamma(t)|), \quad (x, t) \in \bar{G};\\
\left|\frac{\partial^{k+k_0}}{\partial \xi^k\partial t^{k_0}}\tilde{U}^1(\xi, t)\right| \le M[1 + \varepsilon^{(2-k)/2}\tilde{\rho}^{2-k-2k_0}], \quad (\xi, t) \in \bar{\tilde{G}}^1,\\
\left|\frac{\partial^{k+k_0}}{\partial \xi^k\partial t^{k_0}}\tilde{W}^1(\xi, t)\right| \le M[1 + \varepsilon^{(1-k)/2}\tilde{\rho}^{1-k-2k_0}]\exp(-m\varepsilon^{-1/2}|\xi|), \quad (\xi, t) \in \bar{\tilde{G}}, \quad k + 2k_0 \le K,\end{gathered} \tag{4.10}$$

where $\rho = \rho(x, t; \varepsilon) = \varepsilon^{-1/2}|x - \gamma(t)| + t^{1/2}$, $\tilde{\rho} = \tilde{\rho}(\xi, t; \varepsilon) = \varepsilon^{-1/2}|\xi| + t^{1/2}$, and m is an arbitrary constant.

Thus, the following theorem is valid.

Theorem 1. *Let the data of the boundary value problem* (2.2), (2.1) *satisfy the condition* $f \in C^{l,1/2}(\overline{G})$, $\varphi \in C^l(\overline{S}_0^-) \cap C^l(\overline{S}_0^+) \cap C^{l/2}(\overline{S}^L) \cap C(S)$, *and let the condition* (2.3) *holds for the solution of this problem, moreover*, $l = K$. *Then the solution of the boundary value problem and its components in representations* (4.3), (4.9) *satisfy the estimates* (4.2), (4.4), *and* (4.10).

Remark 1. Let us give the condition under which the boundary layer does not arise.

Let us define the set

$$G^4 = G^4(m) = \{(x,t);\ x > \gamma(t) - \gamma(T) - d + m\}, \quad \overline{G}^4 = G^4 \cup S^4, \tag{4.11a}$$

where $m < d + \gamma(T)$. Introduce the set

$$\overline{G}^5 = G^5 \cup S^5, \quad G^5 = G^5(m) = G \backslash \overline{G}^4(m). \tag{4.11b}$$

Let the functions $f(x, t)$ and $\varphi(x, t)$ satisfy the conditions

$$\begin{gathered} f(x,t) = 0, \quad (x,t) \in \overline{G}^5, \\ \varphi(x,t) = 0, \quad (x,t) \in S \cap \overline{G}^5. \end{gathered} \tag{4.12}$$

Then the boundary layer is absent, i.e., the singular component is absent in the representation (4.3):

$$V(x,t) = 0, \quad (x,t) \in \overline{G}^2, \tag{4.13}$$

and $u(x, t) = U(x, t)$ on the set $\overline{G}^2$. For the solution $u(x, t)$, the estimate (4.4a) takes place, moreover,

$$\left|\frac{\partial^{k+k_0}}{\partial x^k \partial t^{k_0}} u(x,t)\right| \le M\varepsilon^{K_1}, \quad (x,t) \in \overline{G}^2, \quad k + 2k_0 \le K, \tag{4.14}$$

where $\overline{G}^2 = \overline{G}^2_{(4.1)}(m^2)$, $m^2 < m_{(4.1)}$, and the constant K_1 in (4.14) can be chosen sufficiently large.

5. CLASSICAL GRID APPROXIMATIONS OF THE PROBLEM ON UNIFORM AND PIECEWISE UNIFORM MESHES

5.1. Let us consider a difference scheme based on classical approximations of problem (2.2), (2.1). On the set $\overline{G}_{(2.1)}$, we introduce the rectangular mesh

$$\overline{G}_h = \overline{D}_h \times \overline{\omega}_0 = \overline{\omega} \times \overline{\omega}_0, \tag{5.1}$$

where $\overline{\omega}$ and $\overline{\omega}_0$ are meshes on the segments $[-d, d]$ and $[0, T]$, respectively; the mesh $\overline{\omega}$ has an arbitrary distribution of nodes satisfying only the condition $h \le MN^{-1}$, where $h = \max\limits_i h^i$, $h^i = x^{i+1} - x^i$, $x^i, x^{i+1} \in \overline{\omega}$; the mesh $\overline{\omega}_0$ is uniform with the step-size $h_0 = TN_0^{-1}$. Here $N + 1$ and $N_0 + 1$ are the numbers of nodes in the meshes $\overline{\omega}$ and $\overline{\omega}_0$, respectively.

We approximate the boundary value problem (2.2) by the difference scheme

$$\begin{gathered} \Lambda_{(5.2)} z(x,t) = f(x,t), \quad (x,t) \in G_h, \\ z(x,t) = \varphi(x,t), \quad (x,t) \in S_h. \end{gathered} \tag{5.2}$$

Here

$$\Lambda_{(5.2)} \equiv \varepsilon a \delta_{\bar{x}\hat{x}} + b\delta_x - c - q\delta_{\bar{t}},$$

$\delta_{\bar{x}\hat{x}} z(x, t)$ is the second difference derivative on a nonuniform mesh. On the uniform mesh

$$\overline{G}_h = \overline{\omega} \times \overline{\omega}_0, \tag{5.3}$$

we obtain the estimate

$$|u(x,t) - z(x,t)| \le M[\varepsilon^{-1}N^{-1} + N^{-1/2} + N_0^{-1/2}], \quad (x,t) \in \overline{G}_h, \tag{5.4}$$

i.e., the scheme (5.2), (5.3) converges for fixed values of the parameter ε.

Under the condition (4.13), when $V(x, t) = 0$, $(x, t) \in \overline{G}^2$, we have the estimate

$$|u(x, t) - z(x, t)| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}_h, \tag{5.5}$$

i.e., the scheme (5.2), (5.3) converges ε-uniformly.

If, except (4.13), the following condition holds:

$$W^1(x, t) = 0, \quad (x, t) \in \overline{G}^1, \tag{5.6}$$

i.e.,

$$\left[\frac{\partial}{\partial x}\varphi(x, t)\right] = 0, \quad (x, t) \in S^{(*)},$$

then we have the ε-uniform estimate for the solution of difference scheme (5.2), (5.3):

$$|u(x, t) - z(x, t)| \le M[N^{-1} + N_0^{-1+\nu_0}], \quad (x, t) \in \overline{G}_h, \tag{5.7}$$

where ν^0 is an arbitrary constant in the interval (0, 1).

The following theorem holds.

Theorem 2. *Let the solution of problem* (2.2) *and its components satisfy the estimates* (4.2), (4.4), (4.8) *for* $K = 4$. *Then the difference scheme* (5.2), (5.3) *converges for fixed values of the parameter* ε. *In the case of the condition* (4.13), *the scheme* (5.2), (5.3) *converges* ε*-uniformly*. *The discrete solutions satisfy the estimate* (5.4); *and, in the case of the condition* (4.13), *the estimate* (5.5) *holds*; *and under conditions* (4.13) *and* (5.6), *the estimate* (5.7) *is valid*.

5.2. We now consider the case when the solution of the problem has a boundary layer. On the set $\overline{G}$, we construct the mesh condensing in a neighbourhood of the boundary layer (similar to that constructed in [9]–[16]):

$$\overline{G}_h = \overline{D}_h \times \overline{\omega}_0 = \overline{\omega}^* \times \overline{\omega}_0, \tag{5.8a}$$

where $\overline{\omega}_0 = \overline{\omega}_{0(5.1)}$, $\overline{\omega}^* = \overline{\omega}^*(\sigma)$ is a piecewise uniform mesh on $[-d, d]$, and σ is a parameter depending on ε and N. We choose the value σ satisfying the condition

$$\sigma = \sigma(\varepsilon, N) = \min[\beta, 2m^{-1}\varepsilon\ln N], \tag{5.8b}$$

where β is an arbitrary number in the half-open interval $(0, d]$, and $m = m_{(4.4)}$. The segment $[-d, d]$ is divided into two parts: $[-d, -d + \sigma]$ and $[-d + \sigma, d]$; on each part, the step-size is constant and is equal to $h^{(1)} = 2d\sigma\beta^{-1}N^{-1}$ on the segment $[-d, -d + \sigma]$ and to $h^{(2)} = 2d(2d - \sigma)(2d - \beta)^{-1}N^{-1}$ on the segment $[-d + \sigma, d]$, $\sigma \le d$.

On the mesh (5.8), we have the estimate

$$|u(x, t) - z(x, t)| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}_h, \tag{5.9}$$

i.e., the scheme (5.2), (5.8) converges ε-uniformly.

The following theorem holds [15].

Theorem 3. *Let the assumptions of Theorem* 2 *be fulfilled*. *Then the difference scheme* (5.2), (5.8) *converges* ε*-uniformly with the estimate* (5.9).

5.3. Let the conditions (4.13), (5.6) be satisfied (i.e., the interior and boundary layers are absent).

On the uniform mesh (5.3), the discrete derivative $p^h(x, t)$ is the solution of the equation

$$\Lambda_{(5.2)}p^h(x, t) = \delta_x f(x, t), \quad (x, t) \in G_h, \quad x \le d - 2h;$$

on the set S_{0h}, the following condition holds:

$$p^h(x, t) = \delta_x\varphi(x, t), \quad (x, t) \in S_{0h}, \quad x \ne d.$$

The following estimates are satisfied for the flux and for the discrete derivative:

$$\left|P(x, t) - P^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}_h, \quad x \ne d, \tag{5.10a}$$

$$\left|p(x, t) - p^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}][(\varepsilon + N^{-1})^{-1}(1 + 2a^{-1}bd\varepsilon^{-1}N^{-1})^{-i} + 1],$$
$$(x, t) \in \overline{G}_h, \quad x = x_i \ne d; \tag{5.10b}$$

$$\left|p(x, t) - p^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}_h, \quad x \in [-d + \beta_0, d). \tag{5.11}$$

Here $\beta_0 > 0$ is an arbitrary sufficiently small constant, and $M_{(5.11)} = M(\beta_0)$.

But if the condition (4.12) is fulfilled (which implies the condition (4.13)), and also the condition (5.6) holds, then for the discrete derivative $p(x, t)$, we obtain the estimate

$$\left|p(x, t) - p^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}_h, \quad x \ne d, \tag{5.12}$$

which is stronger than (5.11). Thus, in the case of scheme (5.2), (5.3), the condition (4.12) is sufficient for the ε-uniform convergence of the derivative $p^h(x, t)$ on the whole set $\overline{G}^h$, $x \ne d$.

5.4. We consider the approximation of the functions $u(x, t)$, $p(x, t)$, $P(x, t)$, $(x, t) \in \overline{G}$, using the interpolants constructed on the basis of the functions $z(x, t)$, $p^h(x, t)$, $P^h(x, t)$.

Let $z(x, t)$, $(x, t) \in \overline{G}_h$, be a solution of some scheme. For the function $z(x, t)$, we construct its extension $\bar{z}(x, t)$ to $\overline{G}$; $\bar{z}(x, t)$ is a bilinear interpolant on the elementary rectangles generated by the lines that pass through the nodes of the mesh $\overline{G}_h$ in parallel to the coordinate axes. Further, we construct the interpolant $\bar{p}^h(x, t)$, $(x, t) \in \overline{G}$, for the discrete derivative $p^h(x, t)$, $(x, t) \in \overline{G}_h$, $x \ne d$. At the interior points of the elementary rectangles, we assume $\bar{p}^h(x, t) = \bar{p}_z^h(x, t) = (\partial/\partial x)\bar{z}(x, t)$; the function $\bar{p}^h(x, t)$ is continuous on the upper and on the lower sides of the rectangles, and it is defined according to continuity on the left sides of the elementary rectangles. But if the rectangles are adjacent, by their right sides, to the set $\bar{S}^r$ (where S^r is the right side of the boundary S^L, $S^L = S^l \cup S^r$), then we also continuously define the function $\bar{p}^h(x, t)$ on these sides. Hence, we have constructed the function $\bar{p}^h(x, t)$, $(x, t) \in \overline{G}$. The interpolant $\bar{p}^h(x, t)$, in general, has discontinuities on the lines that are parallel to the t-axis and pass through the nodes of the mesh G_h. We define the interpolant of the diffusion flux by the $\overline{P}^h(x, t) = \overline{P}_z^h(x, t) = \varepsilon\bar{p}^h(x, t)$, $(x, t) \in \overline{G}$.

In the case of the difference scheme (5.2), (5.3) under conditions (4.13) and (5.6), we have the following estimates for the interpolants:

$$\left|p(x, t) - \bar{p}^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}, \quad x \ge -d + \beta_0; \tag{5.13a}$$

$$\left|P(x, t) - \overline{P}^h(x, t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x, t) \in \overline{G}, \tag{5.13b}$$

where $\beta_0 = \beta_{0(5.11)}$, $M_{(5.13a)} = M(\beta_0)$, and also the estimate similar to (5.7):

$$|u(x, t) - \bar{z}(x, t)| \le M[N^{-1} + N_0^{-1+\nu_0}], \quad (x, t) \in \overline{G}, \tag{5.13c}$$

where $\nu_0 = \nu_{0(5.7)}$.

Definition 2. In that case when the interpolants constructed on the basis of the solution of the difference scheme approximate the solution of the differential problem, its derivative and the diffusion flux ε-uniformly, we say that the difference scheme approximates the solution of the differential problem, its derivative and the diffusion flux ε-uniformly.

Theorem 4. *Let the assumptions of Theorem* 2 *be fulfilled for* $K = 6$. *Then*, *under the conditions* (4.13), (5.6), *the difference scheme* (5.2), (5.3) *approximates the solution of problem* (2.2), (2.1), *its derivative and the diffusion flux* ε*-uniformly with the estimates* (5.13); *for the diffusion flux and for the derivative*, *the estimates* (5.10), (5.11) *are also valid*.

6. SCHEME OF THE SOLUTION DECOMPOSITION METHOD APPROXIMATING THE DERIVATIVE $p(x, t)$

To construct a difference scheme that approximates the first order derivative $p(x, t) = (\partial/\partial x)u(x, t)$ on the set $\overline{G}^*$, we use the method of the additive splitting of a singularity such as the interior layer function [9] (or briefly, the singularity splitting method).

6.1. We represent the solution of problem (2.2), (2.1) as the sum of functions

$$u(x, t) = u_1(x, t) + u_2(x, t), \quad (x, t) \in \overline{G}. \tag{6.1a}$$

Here $u_1(x, t)$ and $u_2(x, t)$ are the components of the solution to the boundary value problem (2.2), (2.1) that include singularities of the boundary and interior layer types, respectively. We call the functions $u_1(x, t)$ and $u_2(x, t)$ the components containing the boundary and interior layers, respectively. We represent the function $u_2(x, t)$ as the sum of functions

$$u_2(x, t) = u_2^1(x, t) + u_2^2(x, t), \quad (x, t) \in \overline{G}, \tag{6.1b}$$

where $u_2^1(x, t)$ and $u_2^2(x, t)$ are the regular and singular parts of the component $u_2(x, t)$ containing the boundary layer;

$$u_2^2(x, t) = W^1_{(4.9b)}(x, t), \quad (x, t) \in \overline{G}.$$

The functions $u_1(x, t)$ and $u_2^1(x, t)$ are solutions of the following problems

$$\begin{aligned} Lu_2^1(x, t) &= f_2(x, t), \quad (x, t) \in G, \\ u_2^1(x, t) &= \varphi_2(x, t), \quad (x, t) \in S; \end{aligned} \tag{6.1c}$$

$$\begin{aligned} Lu_1(x, t) &= f_1(x, t), \quad (x, t) \in G, \\ u_1(x, t) &= \varphi_1(x, t), \quad (x, t) \in S. \end{aligned} \tag{6.1d}$$

The functions $f_i(x, t)$ and $\varphi_i(x, t)$, $i = 1, 2$, are defined by the relations

$$\begin{aligned} f_2(x, t) &= f(x, t)\eta(x, t), \\ f_1(x, t) &= f(x, t) - f_2(x, t), \quad (x, t) \in \overline{G}; \\ \varphi_2(x, t) &= [\varphi(x, t) - u_2^2(x, t)]\eta(x, t), \\ \varphi_1(x, t) &= \varphi(x, t) - \varphi_2(x, t) - u_2^2(x, t), \quad (x, t) \in S. \end{aligned} \tag{6.1e}$$

Here $\eta(x, t)$, $(x, t) \in \overline{G}$, is a sufficiently smooth function that vanishes in a neighbourhood of the boundary layer

$$\left.\begin{aligned} \eta(x, t) &= 0, \quad (x, t) \in \overline{G}^5_{(4.11)}(2^{-1}m_1) \\ \eta(x, t) &= 1, \quad (x, t) \in \overline{G}^4_{(4.11)}(m_1) \end{aligned}\right\}, \quad 0 \le \eta(x, t) \le 1, \quad (x, t) \in \overline{G},$$

where m_1 is an arbitrary number in the interval $(0, 2^{-1}(d + \gamma(T)))$.

Remark 2. The data of problem (6.1c) on the set $\overline{G}^5$ satisfy the condition similar to (4.12) ($f_2(x, t) = 0$, $\varphi_2(x, t) = 0$, $(x, t) \in \overline{G}^5$), moreover, for $t = 0$, the first order derivative in x of the function $\varphi_2(x, t)$ is continuous. For the singular components of the solution to problem (6.1c) in representations similar to (4.3), (4.9a), conditions of the type (4.13), (5.6) are satisfied. For this problem, a difference scheme, similar to (5.2), on the uniform mesh (5.3) ensures the ε-uniform convergence in $C^1(\overline{G})$.

The data of problem (6.1d) are sufficiently smooth, moreover, the functions $f_1(x, t)$ and $\varphi_1(x, t)$ vanish on the set $\overline{G}^4(m_1)$. For this problem, a difference scheme, similar to (5.2), on the piecewise uniform mesh (5.8) gives the ε-uniform convergence in $C^{1(n)}(\overline{G})$.

To solve problem (6.1d), we use the difference scheme

$$\Lambda_{(5.2)} z_1(x, t) = f_1(x, t), \quad (x, t) \in G_h,$$
$$z_1(x, t) = \varphi_1(x, t), \quad (x, t) \in S_h, \tag{6.2a}$$

where $\overline{G}_h$ is the piecewise uniform mesh (5.8).

To solve problem (6.1c), we use the difference scheme

$$\Lambda_{(5.2)} z_2^1(x, t) = f_2(x, t), \quad (x, t) \in G_h,$$
$$z_2^1(x, t) = \varphi_2(x, t), \quad (x, t) \in S_h, \tag{6.2b}$$

where $\overline{G}_h$ is the uniform mesh (5.3).

Further, we construct the special interpolants into which the singular component, i.e., the function of the interior layer type, enters in the explicit form

$$u_0^h(x, t) = \bar{z}_1(x, t) + u_2^h(x, t),$$
$$u_2^h(x, t) = \bar{z}_2^1(x, t) + u_2^2(x, t), \quad (x, t) \in \overline{G}; \tag{6.2c}$$

$$p_0^h(x, t) = \bar{p}_{z_1}^h(x, t) + p_2^h(x, t),$$
$$p_2^h(x, t) = \bar{p}_{z_2^1}^h(x, t) + \frac{\partial}{\partial x} u_2^2(x, t), \quad (x, t) \in \overline{G}^*; \tag{6.2d}$$

$$P_0^h(x, t) = \varepsilon p_0^h(x, t), \quad (x, t) \in \overline{G}^*, \tag{6.2e}$$

where $\bar{z}_1(x, t)$, $\bar{p}_{z_1}^h(x, t)$ and $\bar{z}_2(x, t)$, $\bar{p}_{z_2}^h(x, t)$ are bilinear interpolants constructed using the functions $z_1(x, t)$, $(x, t) \in \overline{G}_{h(5.8)}$ and $z_2(x, t)$, $(x, t) \in \overline{G}_{h(5.3)}$ (similarly to the construction of interpolants in Subsection 5.4). The use of these interpolants allows us to find the solution on the set $\overline{G}$, its first derivative in x and the diffusion flux on the set $\overline{G}^*$.

The function $u_0^h(x, t)$, $(x, t) \in \overline{G}$, is called the solution of the difference scheme (6.2), (5.3), (5.8), and the functions $p_0^h(x, t)$ and $P_0^h(x, t)$, $(x, t) \in \overline{G}^*$, are called the derivative and the diffusion flux, respectively, corresponding to this scheme. The scheme (6.2), (5.3), (5.8) is the scheme of the decomposition method for the solution in the case of the additive splitting of a singularity of the interior layer type (briefly, this scheme is called the scheme of the singularity splitting method).

6.2. If the condition (4.12) or the following (stronger) condition are fulfilled:

$$f(x, t) = 0, \quad (x, t) \in \overline{G};$$
$$\varphi(x, t) = 0, \quad (x, t) \in S, \quad x < 0, \tag{6.3}$$

then the scheme is simplified if we take

$$u_2^2(x, t) = W_{(4.9b)}^1(x, t) + 2^{-1}\left[\frac{\partial}{\partial x}\varphi(0, 0)\right][x - \gamma(t)]\exp(-\alpha t),$$
$$(x, t) \in \overline{G}, \quad \alpha = \alpha_{(4.5)}, \quad \gamma = \gamma_{(4.5)}; \tag{6.4}$$

$u_2^2(x, t) = 0$ for $x < 0$, $t = 0$. In this case, the component $u_1(x, t)$ that contains the boundary layer is absent in the representation (6.1a). Then the solution of problem (2.2), (2.1) takes the form

$$u(x, t) = u_2(x, t) = u_2^1(x, t) + u_{2(6.4)}^2(x, t), \quad (x, t) \in \overline{G}, \tag{6.5a}$$

where $u_2^1(x, t)$ is the solution of the problem

$$\begin{aligned} Lu_2^1(x,t) &= 0, \quad (x,t) \in G, \\ u_2^1(x,t) &= \varphi_2(x,t), \quad (x,t) \in S. \end{aligned} \tag{6.5b}$$

Here $\varphi_2(x, t) = \varphi(x, t) - u^2_{2(6.4)}(x, t)$, $(x, t) \in S$.

To solve problem (6.5), we use the difference scheme

$$\begin{aligned} \Lambda_{(5.2)} z_2^1(x,t) &= 0, \quad (x,t) \in G_h, \\ z_2^1(x,t) &= \varphi_2(x,t), \quad (x,t) \in S_h, \end{aligned} \tag{6.6a}$$

where $\overline{G}_h$ is the uniform mesh (5.3).

Further, we construct the following special interpolants (similar to (6.2c))

$$\begin{aligned} u_0^h(x,t) &= \bar{z}_2^1(x,t) + u^2_{2(6.4)}(x,t), \quad (x,t) \in \overline{G}, \\ p_0^h(x,t) &= \bar{p}^h_{z_2^1}(x,t) + \frac{\partial}{\partial x} u^2_{2(6.4)}(x,t), \quad (x,t) \in \overline{G}^*, \\ P_0^h(x,t) &= \varepsilon p_0^h(x,t), \quad (x,t) \in \overline{G}^*, \end{aligned} \tag{6.6b}$$

where $\bar{z}_2^1(x, t)$ is a bilinear interpolant constructed using the function $z_2^1(x, t) = z^1_{2(6.6a)}(x, t)$.

The scheme (6.6), (5.3) is the scheme of the singularity splitting method under condition (4.12) or (6.3).

Note that condition (6.3) is satisfied in the case of problem (1.2), (1.3).

6.3. Let us give estimates for the schemes constructed in this section.

In the case of scheme (6.2), (5.3), (5.8), we have the estimates

$$\left|u(x,t) - u_0^h(x,t)\right| \le M[N^{-1}\ln N + N_0^{-1+\nu_0}], \quad (x,t) \in \overline{G}, \tag{6.7a}$$

$$\left|P(x,t) - P_0^h(x,t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x,t) \in \overline{G}^*, \tag{6.7b}$$

$$\left|p(x,t) - p_0^h(x,t)\right| \le M[N^{-1/2} + N_0^{-1/2}], \quad (x,t) \in \overline{G}_0^*, \tag{6.7c}$$

where $\overline{G}_0^* = \overline{G}^*_{0(2.5)}(m)$, m is an arbitrary sufficiently small constant, $M_{(6.7c)} = M(m)$, and $\nu_0 = \nu_{0(5.7)}$.

In the case of scheme (6.6), (5.3) under condition (4.12) or (6.3), the following estimates are valid:

$$\begin{aligned} \left|u(x,t) - u_0^h(x,t)\right| &\le M[N^{-1} + N_0^{-1+\nu_0}], \quad (x,t) \in \overline{G}, \\ \left|p(x,t) - p_0^h(x,t)\right| &\le M[N^{-1/2} + N_0^{-1/2}], \quad (x,t) \in \overline{G}^*, \end{aligned} \tag{6.8}$$

where $\nu_0 = \nu_{0(5.7)}$.

Thus, scheme (6.2), (5.3), (5.8) converges ε-uniformly in $C^1(\overline{G}_0^*)$, and scheme (6.6), (5.3) under condition (4.12) or (6.3) converges ε-uniformly in $C^1(\overline{G}^*)$.

In the case of difference scheme (6.2), (5.3), (5.8), the component $u_{2(6.1a)}(x, t)$ containing the interior layer and its derivative in x satisfy the estimates

$$\begin{aligned} \left|u_2(x,t) - u_2^h(x,t)\right| &\le M[N^{-1} + N_0^{-1+\nu_0}], \quad (x,t) \in \overline{G}, \\ \left|p_2(x,t) - p_2^h(x,t)\right| &\le M[N^{-1/2} + N_0^{-1/2}], \quad (x,t) \in \overline{G}^*, \end{aligned} \tag{6.9}$$

where $p_2(x, t) = \frac{\partial}{\partial x} u_2(x, t)$, $p_2^h(x, t) = p^h_{2(6.2d)}(x, t)$. Thus, the component containing the interior layer converges ε-uniformly in $C^1(\overline{G}^*)$.

Theorem 5. *Let the assumptions of Theorem* 2 *be fulfilled for* $K = 6$. *Then the difference scheme* (6.2), (5.3), (5.8) (*the difference scheme* (6.6), (5.3) *under condition* (4.12) *or* (6.3)) *approximates the solution of* *the problem* (2.2), (2.1), *the derivative* $p(x, t)$ *and the diffusion flux* $P(x, t)$ (*the solution of the problem* (2.2), (2.1) *and the derivative* $p(x, t)$) ε-*uniformly with the estimates* (6.7) *and* (6.9), *respectively* (*with the estimates* (6.8)).

Note that the order of the ε-uniform convergence of schemes (6.2), (5.3), (5.8), and (6.6), (5.3) under condition (4.12) or (6.3) is essentially better than it is for the scheme (5.2), (5.8) (see the estimates (5.9), (6.7), (6.8)).

In the case of problem (1.1), i.e., the Cauchy problem for the Black–Scholes equation, the scheme of the singularity splitting method makes it possible to obtain the approximation of the solution $C(S, t')$ in a finite neighbourhood of the point (E, T) containing the interior layer, and also of its derivative $(\partial/\partial S)C(S, t')$ in this neighbourhood excluding the point (E, T), with errors independent of the dimensionless value $\sigma^2 r^{-1}$ for $\sigma^2 r^{-1} \le M$. The interpolants approximating the solution $C(S, t')$ and its derivative $(\partial/\partial S)C(S, t')$ converge in the maximum norm uniformly with respect to the value $\sigma^2 r^{-1}$ at a rate of convergence with the order close to 1 and 0.5, respectively.

7. NUMERICAL EXPERIMENTS

7.1. In this section, we present experimental results for the problem

$$L_{(7.1)}u(x, t) \equiv \left\{\varepsilon\frac{\partial^2}{\partial x^2} + (1-\varepsilon)\frac{\partial}{\partial x} - 1 - \frac{\partial}{\partial t}\right\}u(x, t) = 0, \quad (x, t) \in G, \tag{7.1a}$$

$$u(x, t) = \varphi(x, t), \quad (x, t) \in S,$$

that has the same singularity of the solution in a neighbourhood of the interior layer as problem (1.4), (1.3). Note that the problem (1.4), (1.3) is equivalent to problem (1.1). Here

$$G = D \times (0, T], \quad D = \{x : x \in (-d, d)\},$$

$$T = 1, \quad d = 2, \quad \varphi(x, 0) = \begin{cases} 0, & -d < x \le 0, \\ x + mx^2, & 0 < x < d; \end{cases} \tag{7.1b}$$

$$\varphi(-d, t) = 0, \quad \varphi(d, t) = e^{-t}\{md^2 + [2mt(1-\varepsilon) + 1]d + [2m\varepsilon + (1-\varepsilon)]t + m(1-\varepsilon)^2t^2\},$$

where $m = 4^{-1}$.

The jump of the first order derivative of the function $\varphi_{(7.1)}(x, t)$ at the point $(0, t)$ is the same as for the function $\varphi_{(1.3a)}(x)$ for $x = 0$ and is equal to 1 for both functions.

Note that $\varphi(x, t) = w(x, t)$ for $(x, t) \in S$, $x \ge 0$, where

$$w(x, t) = e^{-t}\{mx^2 + [2mt(1-\varepsilon) + 1]x + [2m\varepsilon + (1-\varepsilon)]t + m(1-\varepsilon)^2t^2\}, \quad (x, t) \in \mathbb{R} \times [0, T].$$

Here the function $w(x, t)$ is the solution of the Cauchy problem

$$\begin{gathered} L_{(7.1)}w(x, t) = 0, \quad (x, t) \in \mathbb{R} \times (0, T], \\ w(x, 0) = \varphi_w(x), \quad x \in \mathbb{R}, \end{gathered} \tag{7.2}$$

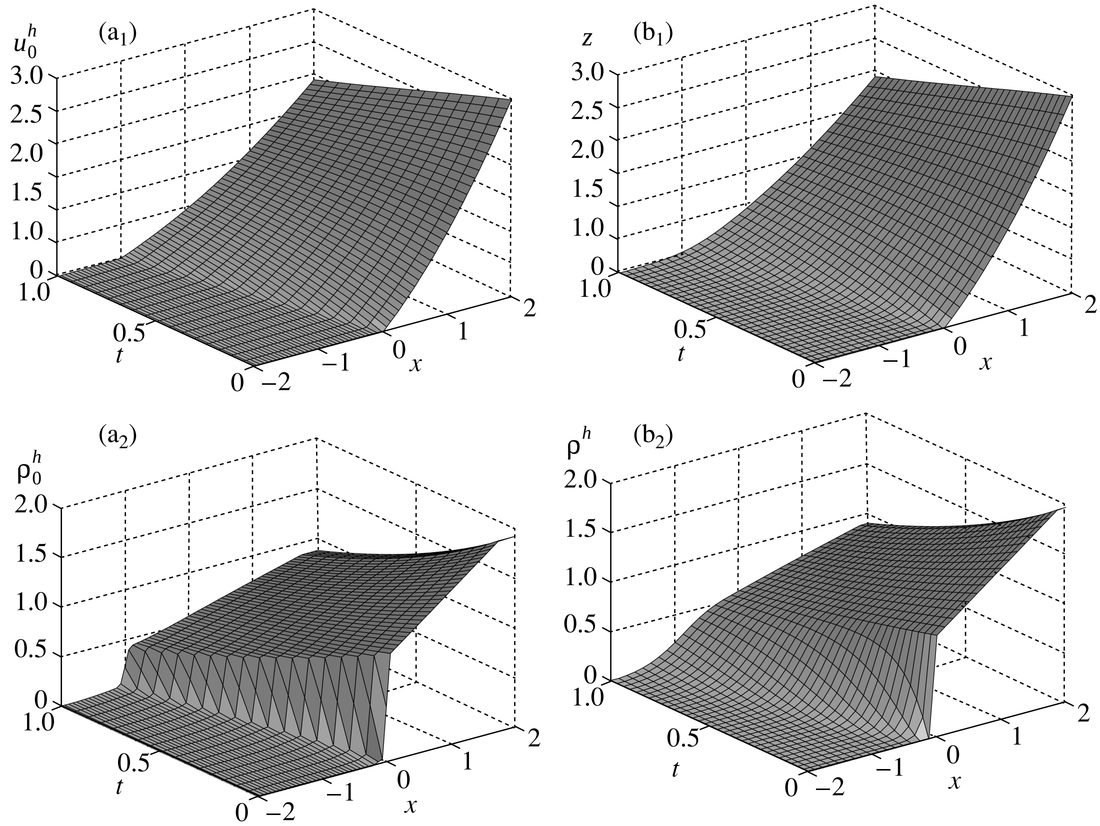
where $\varphi_w(x) = x + mx^2$, $x \in \mathbb{R}$.

The choice of boundary conditions for $x = -d, d$ ensures that compatibility conditions are fulfilled for the data of problem (7.1) and prevents the appearance of the boundary layer and of the interior layer in a neighbourhood of the characteristic passing through the point $(d, 0)$.

Under the chosen data of the problem, the singularity of the solution generated by the jump of the derivative of the initial function is not “polluted” by the other singularities, which allows us to study numerically the efficiency of the constructed difference scheme in a domain containing the interior layer and to compare this scheme with the classical finite difference scheme.

The data of problem (7.1) satisfy condition (6.3) (and condition (4.12)). Thus, for the numerical solution of this problem, it is possible to apply the simplified scheme (6.6), (5.3), i.e., the scheme of the singularity splitting method under condition (6.3) or (4.12) (we denote it briefly by Scheme A).

In order to estimate the efficiency of the developed method, we compare solutions generated using Scheme A in accuracy with the discrete solutions of problem (7.1) generated using the classical finite difference scheme (5.2), (5.3) (we denote it briefly by Scheme B).



Plots of the solutions (a_1), (b_1) and the derivatives (a_2), (b_2); plots of (a_1), (a_2) and (b_1), (b_2) are generated by Schemes A and B, respectively, for $\varepsilon = 2^{-10}$, $N = 16$ and $N_0 = 16$.

The plots of the solutions a_1, b_1 and the derivatives a_2, b_2 computed using Scheme A (see a_1, a_2) and Scheme B (see b_1, b_2) are presented on Fig. 1 for $\varepsilon = 2^{-10}$, $N = 16$ and $N_0 = 16$.

7.2. To analyze errors in the discrete solutions, a technique similar to that given in [11] is used, however, it is modified with regard to the singularity splitting method. Computations are made for the values of $\varepsilon = 2^{-j}$, $j = 0, \ldots, 34$ on grids with the number of nodes $N = N_0$ for $N = 2^i$, $i = 5, \ldots, 10$. The numerical solution $u_{0,\varepsilon}^{h,N^F}(x, t)$ generated by Scheme A on the finest mesh $\overline{G}_h^{N^F}$ with $N = N_0 = N^F = 2048$ for each value of ε is used as the exact solution of problem (7.1).

Errors in the numerical solutions in the maximum norm for each value of ε and N are computed by the formula

$$E_\varepsilon^N = E_\varepsilon^N(u_{0,\varepsilon}^{h,N}(\cdot)) = \left\|u_{0,\varepsilon}^{h,N^F}(x,t) - u_{0,\varepsilon}^{h,N}(x,t)\right\|_{\overline{G}_h^N} \tag{7.3}$$

for Scheme A and by the formula

$$E_\varepsilon^N = E_\varepsilon^N(z_\varepsilon^N(\cdot)) = \left\|u_{0,\varepsilon}^{h,N^F}(x,t) - z_\varepsilon^N(x,t)\right\|_{\overline{G}_h^N} \tag{7.4}$$

for Scheme B. Here the function $u_{0,\varepsilon}^{h,N}(x, t) = u_{0(6.6),\varepsilon}^{h,N}(x, t)$ in (7.3) and the function $z_\varepsilon^N(x, t)$ in (7.4) are the numerical solutions obtained, respectively, by Schemes A and B.

Tables 1 and 2 contain the values E_ε^N of errors in the solutions generated by Schemes A and B for various values of ε and N. The value E^N in the last rows of the tables is the maximal value of the errors E_ε^N with respect to ε, corresponding to the given value of N.

Table 1. Errors $E_\varepsilon^N = E_\varepsilon^N(u_{0,\varepsilon}^{h,N})$ and $E^N = E^N(u_{0,\varepsilon}^{h,N})$ for the solutions generated by Scheme A

ε	Number of intervals N					
	32	64	128	256	512	1024
2^0	0.1320–02	0.6633–03	0.3328–03	0.1666–03	0.8338–04	0.4170–04
2^{-1}	0.1140–02	0.5863–03	0.2972–03	0.1496–03	0.7507–04	0.3760–04
2^{-3}	0.2782–02	0.1426–02	0.7219–03	0.3632–03	0.1822–03	0.9123–04
2^{-3}	0.3904–02	0.2006–02	0.1017–02	0.5120–03	0.2569–03	0.1287–03
2^{-4}	0.4614–02	0.2384–02	0.1214–02	0.6131–03	0.3082–03	0.1545–03
2^{-5}	0.5034–02	0.2623–02	0.1345–02	0.6823–03	0.3440–03	0.1727–03
2^{-6}	0.5288–02	0.2769–02	0.1429–02	0.7289–03	0.3688–03	0.1856–03
…	…	…	…	…	…	…
2^{-15}	0.5557–02	0.2948–02	0.1544–02	0.8004–03	0.4116–03	0.2103–03
…	…	…	…	…	…	…
2^{-34}	0.5558–02	0.2948–02	0.1544–02	0.8006–03	0.4117–03	0.2104–03
E^N	0.5558–02	0.2948–02	0.1544–02	0.8006–03	0.4117–03	0.2104–03

Table 2. Errors $E_\varepsilon^N = E_\varepsilon^N(z_\varepsilon^N)$ and $E^N = E^N(z_\varepsilon^N)$ for the solutions generated by Scheme B

ε	Number of intervals N					
	32	64	128	256	512	1024
2^0	0.9868–02	0.5482–02	0.3296–02	0.2116–02	0.1419–02	0.9756–03
2^{-1}	0.8081–02	0.4425–02	0.2560–02	0.1584–02	0.1036–02	0.7016–03
2^{-3}	0.6600–02	0.3718–02	0.2093–02	0.1237–02	0.7772–03	0.5126–03
2^{-3}	0.9033–02	0.4972–02	0.2627–02	0.1354–02	0.6878–03	0.3850–03
2^{-4}	0.1242–01	0.7186–02	0.3933–02	0.2072–02	0.1065–02	0.5406–03
2^{-5}	0.1515–01	0.9293–02	0.5348–02	0.2919–02	0.1535–02	0.7887–03
2^{-6}	0.1706–01	0.1105–01	0.6736–02	0.3866–02	0.2107–02	0.1107–02
…	…	…	…	…	…	…
2^{-15}	0.1963–01	0.1403–01	0.9948–02	0.7029–02	0.4956–02	0.3486–02
…	…	…	…	…	…	…
2^{-34}	0.1964–01	0.1404–01	0.9960–02	0.7045–02	0.4978–02	0.3516–02
E^N	0.1964–01	0.1404–01	0.9960–02	0.7045–02	0.4978–02	0.3516–02

Tables 3 and 4, which are similar to Tables 1 and 2, demonstrate errors in the first derivatives computed by the formula

$$E_\varepsilon^N = E_\varepsilon^N(p_{0,\varepsilon}^{h,N}(\cdot)) = \left\| p_{0,\varepsilon}^{h,N^F}(x,t) - p_{0,\varepsilon}^{h,N}(x,t) \right\|_{\overline{G}_h^{N*}},$$
$$\overline{G}_h^{N*} = \overline{G}_h^N \backslash S^{(*)}, \quad S^{(*)} = S_{(2.2c)}^{(*)} \tag{7.5}$$

for Scheme A and by the formula

$$E_\varepsilon^N = E_\varepsilon^N(p_{z,\varepsilon}^N(\cdot)) = \left\| p_{0,\varepsilon}^{h,N^F}(x,t) - p_{z,\varepsilon}^N(x,t) \right\|_{\overline{G}_h^{N\{*\}}},$$
$$\overline{G}_h^{N\{*\}} = \overline{G}_h^N \backslash S^{\{*\}}, \quad S^{\{*\}} = \{(x,0) : x = x_{i-1}, x_i, x_{i+1};\ x_i = 0\} \tag{7.6}$$

Table 3. Errors $E_\varepsilon^N = E_\varepsilon^N(p_{0,\varepsilon}^{h,N})$ and $E^N = E^N(p_{0,\varepsilon}^{h,N})$ for the first discrete derivatives generated by Scheme A

ε	Number of intervals N					
	32	64	128	256	512	1024
2^0	0.1562–01	0.7931–02	0.4073–02	0.2106–02	0.1100–02	0.5835–03
2^{-1}	0.1562–01	0.7910–02	0.4016–02	0.2034–02	0.1048–02	0.5470–03
2^{-3}	0.1620–01	0.8583–02	0.4431–02	0.2253–02	0.1136–02	0.5705–03
2^{-3}	0.1701–01	0.9262–02	0.4858–02	0.2492–02	0.1263–02	0.6355–03
2^{-4}	0.1754–01	0.9768–02	0.5198–02	0.2690–02	0.1370–02	0.6915–03
2^{-5}	0.1788–01	0.1011–01	0.5446–02	0.2840–02	0.1454–02	0.7359–03
2^{-6}	0.1806–01	0.1034–01	0.5616–02	0.2947–02	0.1515–02	0.7688–03
…	…	…	…	…	…	…
2^{-15}	0.1831–01	0.1066–01	0.5915–02	0.3966–02	0.2697–02	0.1850–02
…	…	…	…	…	…	…
2^{-34}	0.1831–01	0.1066–01	0.5920–02	0.3974–02	0.2708–02	0.1865–02
E^N	0.1831–01	0.1066–01	0.5920–02	0.3974–02	0.2708–02	0.1865–02

Table 4. Errors $E_\varepsilon^N = E_\varepsilon^N(p_{z,\varepsilon}^{N})$ and $E^N = E^N(p_{z,\varepsilon}^{N})$ for the first discrete derivatives generated by Scheme B

ε	Number of intervals N					
	32	64	128	256	512	1024
2^0	0.1080+00	0.7864–01	0.6165–01	0.5118–01	0.4465–01	0.4000–01
2^{-1}	0.1296+00	0.1013+00	0.7555–01	0.5941–01	0.5017–01	0.4405–01
2^{-3}	0.1394+00	0.1219+00	0.9764–01	0.7388–01	0.5826–01	0.4964–01
2^{-3}	0.1302+00	0.1315+00	0.1178+00	0.9572–01	0.7301–01	0.5768–01
2^{-4}	0.1045+00	0.1228+00	0.1275+00	0.1157+00	0.9473–01	0.7256–01
2^{-5}	0.7901–01	0.9761–01	0.1189+00	0.1254+00	0.1147+00	0.9424–01
2^{-6}	0.8714–01	0.7835–01	0.9407–01	0.1170+00	0.1244+00	0.1141+00
…	…	…	…	…	…	…
2^{-15}	0.9914–01	0.9833–01	0.9985–01	0.1005+00	0.1006+00	0.1002+00
…	…	…	…	…	…	…
2^{-34}	0.9918–01	0.9839–01	0.1000+00	0.1008+00	0.1012+00	0.1014+00
E^N	0.1394+00	0.1315+00	0.1275+00	0.1254+00	0.1244+00	0.1238+00

for Scheme B. Here, $p_{z,\varepsilon}^N(x, t)$ in (7.6) is the first difference derivative

$$p_{z,\varepsilon}^N(x_i, t_j) = \frac{z_\varepsilon^N(x_{i+1}, t_j) - z_\varepsilon^N(x_i, t_j)}{x_{i+1} - x_i}, \quad i = 0, 1, \dots, N, \quad j = 0, 1, \dots, N_0. \tag{7.7}$$

The function $p_{0,\varepsilon}^{h,N^F}(x, t)$ in formulae (7.5) and (7.6) and the function $p_{0,\varepsilon}^{h,N}(x, t)$ in formula (7.5) are the special interpolants of the first order derivative of the solution computed by formula (6.6), respectively, on the finest mesh $\overline{G}_h^{N^F}$ and on the mesh $\overline{G}_h^N$ for fixed value of ε.

Analyzing the values of errors for the solutions in Tables 1 and 2, and for the first derivatives in Table 3, we observe the ε-uniform convergence, since, with decreasing ε, the errors are stabilized for each value of N approximately for one and the same values of ε, i.e., the errors are independent of the value of the parameter ε, moreover, the values of E^N (the last row) decrease as N increases. However, in Table 4 the first de-

Table 5. Convergence orders $q_\varepsilon^N = q_\varepsilon^N(u_{0,\varepsilon}^{h,N})$ and $q^N = q^N(u_{0,\varepsilon}^{h,N})$ for the solutions of Scheme A

ε	Number of intervals N				
	32	64	128	256	512
2^0	0.9928	0.9950	0.9983	0.9986	0.9997
2^{-1}	0.9593	0.9802	0.9903	0.9948	0.9975
2^{-3}	0.9641	0.9821	0.9910	0.9952	0.9979
2^{-3}	0.9606	0.9800	0.9901	0.9949	0.9972
2^{-4}	0.9526	0.9736	0.9856	0.9923	0.9963
2^{-5}	0.9405	0.9636	0.9791	0.9880	0.9941
2^{-6}	0.9334	0.9544	0.9712	0.9829	0.9906
…	…	…	…	…	…
2^{-15}	0.9146	0.9331	0.9479	0.9595	0.9688
…	…	…	…	…	…
2^{-34}	0.9148	0.9331	0.9475	0.9595	0.9685
q^N	0.9148	0.9331	0.9475	0.9595	0.9685

Table 6. Convergence orders $q_\varepsilon^N = q_\varepsilon^N(z_\varepsilon^N)$ and $q^N = q^N(z_\varepsilon^N)$ for the solutions of Scheme B

ε	Number of intervals N				
	32	64	128	256	512
2^0	0.8481	0.7340	0.6394	0.5765	0.5405
2^{-1}	0.8689	0.7895	0.6926	0.6125	0.5623
2^{-3}	0.8279	0.8290	0.7587	0.6705	0.6005
2^{-3}	0.8614	0.9204	0.9562	0.9772	0.8371
2^{-4}	0.7894	0.8696	0.9246	0.9602	0.9782
2^{-5}	0.7051	0.7971	0.8735	0.9272	0.9607
2^{-6}	0.6266	0.7141	0.8011	0.8757	0.9285
…	…	…	…	…	…
2^{-15}	0.4845	0.4960	0.5011	0.5041	0.5076
…	…	…	…	…	…
2^{-34}	0.4843	0.4953	0.4995	0.5010	0.5016
q^N	0.4843	0.4953	0.4995	0.5010	0.5016

rivative of the solution generated by Scheme B does not converge at all; the values of E^N practically do not change as N increases.

In Tables 5 and 6, the values of q_ε^N are shown that are the convergence orders for the solutions computed by Schemes A and B, respectively. In analogous Table 7, one can see the convergence orders for the first discrete derivatives generated by Scheme A for various values of ε and N. The value q^N in last rows of the tables is the minimal value of q_ε^N with respect to ε, corresponding to the given value of N.

The convergence order for the discrete solutions is defined by the formula

$$q_\varepsilon^N = \log_2\frac{E_\varepsilon^N}{E_\varepsilon^{2N}}. \tag{7.8}$$

The quantities E_ε^N, E_ε^{2N} are defined by formula (7.3) for Scheme A and by formula (7.4) for Scheme B.

Table 7. Convergence orders $q_\varepsilon^N = q_\varepsilon^N(p_{0,\varepsilon}^{h,N})$ and $q^N = q^N(p_{0,\varepsilon}^{h,N})$ for the first discrete derivatives generated by Scheme A

ε	Number of intervals N				
	32	64	128	256	512
2^0	0.9778	0.9614	0.9516	0.9370	0.9147
2^{-1}	0.9816	0.9779	0.9814	0.9567	0.9380
2^{-3}	0.9164	0.9538	0.9758	0.9879	0.9937
2^{-3}	0.8770	0.9310	0.9631	0.9804	0.9909
2^{-4}	0.8445	0.9101	0.9504	0.9734	0.9864
2^{-5}	0.8226	0.8925	0.9393	0.9659	0.9824
2^{-6}	0.8046	0.8806	0.9303	0.9599	0.9786
…	…	…	…	…	…
2^{-15}	0.7804	0.8498	0.5767	0.5563	0.5438
…	…	…	…	…	…
2^{-34}	0.7804	0.8485	0.5750	0.5534	0.5381
q^N	0.7804	0.8485	0.5750	0.5534	0.5381

The convergence order for the discrete derivatives is defined by (7.8), where the quantities E_ε^N, E_ε^{2N} are defined by formula (7.5) for Scheme A and by formula (7.6) for Scheme B.

Orders of the rate of ε-uniform convergence for the solutions generated by Schemes A and B (see Tables 5 and 6) are close, respectively, to 1 and 0.5; for the first derivative generated by Scheme A (see Table 7), the order of the rate of ε-uniform convergence is close to 0.5.

7.3. Thus, it follows from our numerical experiments that the solution and its first order derivative obtained by Scheme A (i.e., the scheme based on the method of the additive splitting of a singularity) converge ε-uniformly at the rate of convergence of the order close to 1 and 0.5, respectively. Whereas the convergence rate for the solutions of the classical finite difference Scheme B yields to that for scheme A, and the derivatives computed by Scheme B do not converge even for fixed values of the parameter ε. The numerical experiments, consistent with the theoretical results, illustrate the efficiency of the singularity splitting method for the approximation of the interior layer generated by the discontinuity of the first order derivative $\frac{\partial}{\partial x}\varphi(x, 0)$ in problem (7.1).

The numerical experiments showed that the special difference scheme constructed in this paper, i.e., the scheme of the method of the additive splitting of a singularity on uniform meshes, is effective both for small values of N and for its sufficiently large values, for which results of the theoretical study become apparent.

In the case of the Cauchy problem for the Black–Scholes equation, the interpolants constructed using solutions of the special difference scheme, which approximates the solution $C(S, t')$ of problem (1.1) and its derivative $(\partial/\partial S)C(S, t')$ (for $(S, t') \neq (E, T)$) in a neighbourhood of the interior layer, converge at the rate of $\sigma^2 r^{-1}$ – uniform convergence with orders close to 1 and 0.5, respectively.

REFERENCES

1. *P. Wilmott, S. Howison, and J. Dewynne*. The mathematics of financial derivatives. Cambridge: Cambridge Univ. Press, 2002.
2. *G.I. Shishkin*. A difference scheme for a singularly perturbed equation of parabolic type with a discontinuous initial condition, Soviet Math. Dokl., 1988, 37 (3), 792–796.
3. *G.I. Shishkin*. A difference scheme for a singularly perturbed equation of parabolic type with a discontinuous boundary condition, Comput. Math. and Math. Phys., 1988, 28 (6), 32–41.
4. *P.W. Hemker and G.I. Shishkin*. Discrete approximation of singularly perturbed parabolic PDEs with a discontinuous initial condition, Comput. Fluid Dynamics J., 1994, 2 (4), 375–392.
5. *G.I. Shishkin*. Approximation of solutions and diffusion flows in the case of singularly perturbed boundary value problems with discontinuous initial condition, Comput. Math. and Math. Phys., 1996, 36 (9), 1233–1250.

6. *G.I. Shishkin*. Singularly perturbed boundary value problems with concentrated sources and discontinuous initial conditions, Comput. Math. and Math. Phys., 1997, 37 (4), 417–434.
7. *V.L. Kolmogorov and G.I. Shishkin*. Numerical methods for singularly perturbed boundary value problems modelling diffusion processes, in: “Singular Perturbation Problems in Chemical Physics” (J.J.H. Miller ed.), Advances in Chemical Physics Series, vol. XCVII, J.Wiley & Sons, 1997, 181–362.
8. *G.I. Shishkin*. Grid approximations of parabolic convection-diffusion equations with piecewise smooth initial conditions, Dokl. Math., 2005, 72 (3), 850–853.
9. *G.I. Shishkin*. Grid approximation of singularly perturbed parabolic convection-diffusion equations subject to a piecewise smooth initial condition, Comput. Math. and Math. Phys., 2006, 46 (1), 49–72.
10. *O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. Ural’tseva*. Linear and quasilinear equations of parabolic type. Moscow: Nauka, 1967.
11. *P.A. Farrell, A.F. Hegarty, J.J.H. Miller, E. O’Riordan, and G.I. Shishkin*. Robust computational techniques for boundary layers. Boca Raton: Chapman and Hall/CRC, 2000.
12. *S. Li, D.B. Creamer, and G.I. Shishkin*. On numerical methods for a singularly perturbed Black-Scholes equation with nonsmooth initial data, In: Proceedings of the International Conference on Computational Mathematics ICCM’2004, Novosibirsk, June 2004, Novosibirsk: ICM& MG Publisher, 2004; Part II, pp. 896–900.
13. *S. Li, D.B. Creamer, and G.I. Shishkin*. Discrete approximation of a singularly perturbed Black–Scholes equation with nonsmooth initial data, in: An International Conference on Boundary and Interior Layers – Computational and Asymptotic Methods BAIL 2004, ONERA, Toulouse, 5th–9th July, 2004, pp. 441–446.
14. *A.A. Samarskii*. Theory of difference schemes. New York: Marcel Dekker, Inc., 2001.
15. *G.I. Shishkin*. Grid approximations of singularly perturbed elliptic and parabolic equations. Ekaterinburg: Ural Branch of Russian Acad. Sci., 1992.
16. *J.J.H. Miller, E. O’Riordan, and G.I. Shishkin*. Fitted numerical methods for singular perturbation problems. Singapore: World Scientific, 1996.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 481–489

УДК 519.634

ПРИМЕНЕНИЕ СОСТАВНЫХ РАЗНОСТНЫХ СХЕМ ДЛЯ РАСЧЕТА НЕСТАЦИОНАРНЫХ ТЕЧЕНИЙ С УЗКИМИ ТЕПЛОВЫМИ СЛОЯМИ[1)]

© 2007 г. Г. В. Беляков*, В. И. Грынь, А. А. Чарахчьян****

(119334 Москва, Ленинский пр-т, 38, ИДГ РАН;*

*** 119991 Москва, ул. Вавилова, 40, ВЦ РАН)*

e-mail: chara@ccas.ru

Поступила в редакцию 12.05.2006 г.

Описана методика расчета нестационарных гидродинамических течений с узкими тепловыми слоями, которая применяется к моделированию высокоскоростного нагрева трением при ударном сжатии пластины на клине. Библ. 21. Фиг. 6.

Ключевые слова: нестационарные гидродинамические течения, ударные волны, трение.

ВВЕДЕНИЕ

Для расчета гидродинамических течений с узкими погранслоями различного происхождения и ударными волнами в основной части течения необходимо использовать разностные сетки с сильным сгущением в области погранслоя. В случае стационарных течений такого типа высокую эффективность показали квазимонотонные неявные схемы переменных направлений (см., например, [1], [2]). Однако возможность эффективного применения таких схем к расчету нестационарных течений далеко не очевидна. В самом деле, в погранслое течение, как правило, дозвуковое, что в принципе дает возможность проводить расчет с шагом по времени

$$\tau \gg \tau_К = \min_{i,\,j}\tau_{i,\,j}^К,,$$

где $\tau_{i,\,j}^К$ определяется условием Куранта для ячейки сетки (i, j) и зависит от размеров ячейки и местной скорости звука. В настоящей работе мы ограничиваемся случаем двумерных течений. Далее сошлемся на работу [3], где изучалось применение схем переменных направлений к расчету нестационарной задачи конвекции, поставленной для уравнений Навье–Стокса сжимаемого газа. Было показано, что нарушение условия Куранта, по крайней мере для обычных двухслойных схем расщепления, приводит к значительной потере точности. Так как устойчивость расчета при этом сохраняется, диагностировать такую потерю точности можно только сравнением расчетов с существенно разными шагами по времени. Причиной потери точности, как было показано в [4], является рост той части погрешности аппроксимации, которая вносится в схему процедурой расщепления. Переход к неявной схеме без расщепления позволил в [4] значительно увеличить шаг по времени. Хотя в [3], [4] применялись равномерные прямоугольные сетки, а течение везде было дозвуковым, нет оснований надеяться, что эффект понижения точности схем переменных направлений исчезнет в случае неравномерных сеток.

Другая трудность создания эффективных методов расчета нестационарных течений с ударными волнами и узкими пограслоями заключается в том, что отказ от расщепления по направлениям в известных квазимонотонных схемах приводит к очень неэкономичным схемам, так как на каждом временном слое приходится решать систему алгебраических уравнений относительно значений нескольких разностных функций.

В [5], [6] был предложен новый принцип объединения двух разных схем в одну составную схему, свободную от рассмотренных выше недостатков. Такие схемы в погранслое переходят в неявную схему, в основной части течения – в явную квазимонотонную схему, а плавность перехода от одной схемы к другой определяется плавностью перехода от мелкой сетки к крупной. Экономичность составной схемы определяется, во-первых, тем, что отсутствие требования квазимоно-

[1)] Работа выполнена при финансовой поддержке РФФИ (код проекта 04-01-00051) и Российской академии наук (программа ОМН РАН “Соврем. вычисл. и информ. технологии решения больших задач”).

тонности позволяет построить неявную схему с системой алгебраических уравнений относительно значений только одной скалярной функции. Во-вторых, число уравнений примерно равно числу узлов сетки в погранслое, что значительно меньше общего числа узлов. Это позволяет построить эффективные составные схемы и для параболических уравнений, возникающих при использовании расщепления по физическим процессам на этапах учета вязкости и теплопроводности, объединяя схему переменных направлений, “работающую” в основной части течения, и чисто неявную схему, “работающую” в погранслое.

В [5], [6] составные схемы применяются для расчета вязкого нагрева дейтерия, сжимаемого внутри конической мишени. Настоящая работа посвящена применению составных схем для расчета нагрева трением, моделируемого заданием потока тепла на подвижной границе раздела сред. Так как скорость тепловой волны много меньше скорости вещества и скорости звука, тепловая волна имеет вид узкого погранслоя вблизи границы раздела.

1. ЧИСЛЕННЫЙ МЕТОД

Рассматриваются плоские течения сжимаемой теплопроводной жидкости, которые описываются уравнениями

$$\frac{d}{dt}\frac{1}{\rho} = \frac{1}{\rho}\operatorname{div}\mathbf{u}, \quad \operatorname{div}\mathbf{u} = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y},$$
$$\frac{\partial u_x}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \quad \frac{du_y}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial y}, \tag{1}$$
$$\frac{d\varepsilon}{dt} = -\frac{p}{\rho}\operatorname{div}\mathbf{u} + \operatorname{div}\kappa\nabla T,$$

где t – время, x, y – пространственные координаты, $\mathbf{u}$ – вектор скорости с координатами u_x, u_y, $d/dt = \partial/\partial t + u_x\partial/\partial x + u_y\partial/\partial y$ – лагранжева производная по времени, ρ – плотность, p – давление, ε – удельная внутренняя энергия, T – температура, $\kappa = \kappa(\rho, T)$ – коэффициент теплопроводности. Уравнения (1) замыкаются уравнениями состояния $p = p(\rho, T)$, $\varepsilon = \varepsilon(\rho, T)$.

В плоскости x, y имеется подвижная регулярная четырехугольная невырожденная сетка $G(t)$, некоторые линии которой являются лагранжевыми, т.е. движутся вместе с веществом, а внутренние узлы сетки, не принадлежащие выделенным линиям, вычисляются тем или иным методом построения сеток (см. [7]). Разностные функции определены в серединах ячеек сетки. Используется расщепление на лагранжев этап и этап пересчета поля течения с лагранжевой сетки $G^L(t)$ на сетку $G(t + \tau)$ (см. [8]).

Рассмотрим лагранжев этап. Множество всех ячеек сетки разделим на два подмножества, которые назовем множеством крупных Ω_L и мелких Ω_S ячеек. Например, если сетка сгущается к некоторой линии, к множеству Ω_S можно отнести некоторое количество слоев сетки вблизи этой границы, а к множеству Ω_L – остальные ячейки сетки. Шаг по времени определим формулой

$$\tau = \min_{i,\,j\,\in\,\Omega_L} \tau^{\mathrm{K}}_{i,\,j}. \tag{2}$$

Переход на верхний временно́й слой состоит из двух этапов. На первом этапе шаг по времени

$$\tau^{(1)}_{i,\,j} = \begin{cases} \tau, & i, j \in \Omega_L, \\ \tau^{\mathrm{K}}_{i,\,j}, & i, j \in \Omega_S, \end{cases} \tag{3}$$

свой для каждой ячейки сетки. На втором этапе шаг по времени

$$\tau^{(2)}_{i,\,j} = \begin{cases} 0, & i, j \in \Omega_L, \\ \tau - \tau^{\mathrm{K}}_{i,\,j}, & i, j \in \Omega_S, \end{cases} \tag{4}$$

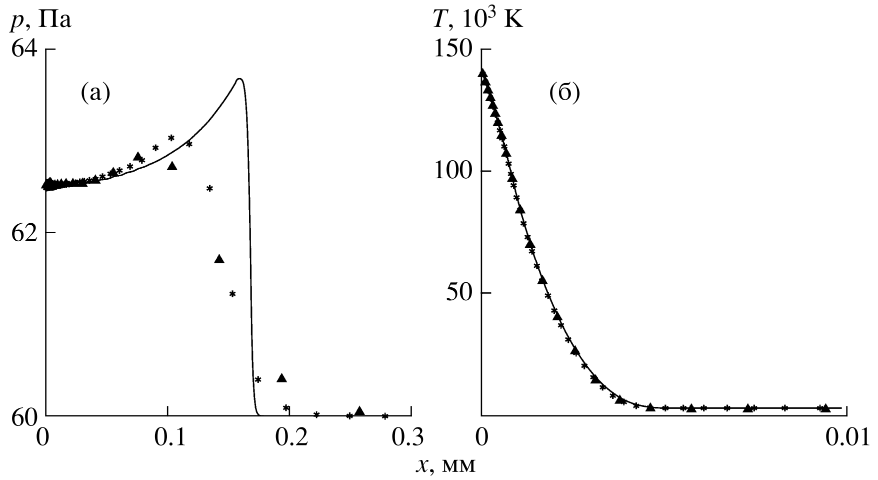
дополняет шаг (3) до полного шага τ. Шаг по времени первого этапа допускает использование явной схемы для уравнений гидродинамики. На втором этапе предполагается использование неявной схемы. Как видно из (3), (4), для крупных ячеек из множества Ω_L переход на верхний временно́й слой выполняется по схеме первого этапа, для очень мелких ячеек с $\tau^{\mathrm{K}}_{i,\,j} \ll \tau$ основной вклад в переход вносит неявная схема второго этапа, а плавность перехода от одной схемы к другой определяется плавностью перехода от мелкой сетки к крупной.

На первом этапе используется квазимонотонная схема типа С.К. Годунова второго порядка точности, которая применялась ранее для расчета различных течений с ударными волнами (см., например, [9]) в сочетании с неявной схемой для уравнения теплопроводности. Последняя схема, в свою очередь, строится как составная, объединяющая схему переменных направлений и схему без расщепления по направлениям (см. [5], [6]).

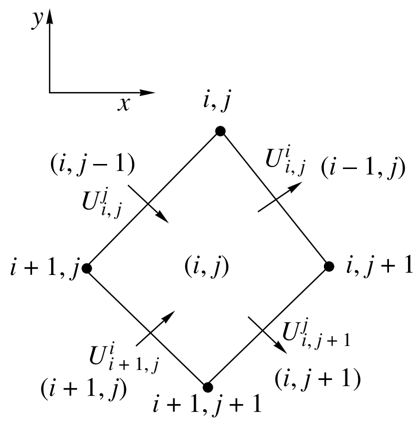
На втором этапе строится неявная схема без расщепления как по направлениям, так и по физическим процессам, обобщающая схему [10], [6] на уравнения (1). Применяется метод Ньютона, на каждой итерации которого возникает система линейных уравнений относительно значений двух разностных функций, которая, как и в [5], [6], решается прямым методом из [11], использующим разложение ленточной матрицы на треугольные множители. Так как множество Ω_S, на котором строится неявная схема, сравнительно невелико, вычислительные затраты на ее реализацию не влияют заметно на общую эффективность метода.

Схема лагранжевого этапа тестировалась на одномерной задаче о нагреве сжимаемой жидкости постоянным потоком тепла (см. [12]). В уравнениях (1) все неизвестные зависят только от x и t; $u_y = 0$. В начальный момент $t = 0$ область $x \geq 0$ заполнена неподвижным алюминием с давлением 60 ГПа и плотностью 3.5 г/см^3, что примерно соответствует параметрам алюминия за фронтом присоединенной ударной волны при бесструйном сжатии пластины со скоростью 4 км/с (см. [13]). Соответствующая двумерная задача будет рассмотрена в следующем разделе. На границе $x = 0$ ставится условие симметрии $u_x = 0$ и задается постоянный поток тепла $q = -\kappa\partial T/\partial x =$ $= 30$ ГВт/см^2. Используется широкодиапазонное уравнение состояния алюминия (см. [14]) в табличной форме. Для коэффициента теплопроводности используется та же модель, что и в [12].

После небольшого по времени начального этапа, когда формируется ударная волна, решение задачи имеет вид двух волн: ударной волны, после которой давление повышается до некоторого значения, и тепловой волны, в которой давление не меняется и скорость распространения которой много меньше скорости ударной волны. Расчет велся в переменных Лагранжа. При $t = 0$ задавалась сетка по x, которая являлась комбинацией сеток с постоянным и переменным шагом, меняющимся по геометрической или арифметической прогрессии. Результаты расчета на разных сетках в момент времени, когда плотность при $x = 0$ уменьшилась до некоторого заданного значения, показаны на фиг. 1. Для всех трех сеток шаг вблизи границы $x = 0$, $h_1 = 2 \times 10^{-5}$ мм достаточно мал, чтобы можно было разрешить структуру тепловой волны. Расчет на самой подробной сетке велся без применения составной схемы, т.е. множество ячеек Ω_s полагалось пустым. Расчет на грубых сетках велся по описанной выше составной схеме с шагом по времени $\tau \approx 100\tau_K$. Видно, что расчеты на грубых сетках правильно воспроизводят как структуру тепловой волны, так и положение ударной волны.



Фиг. 1. Давление вблизи ударной волны (график (а)) и температура вблизи тепловой волны (график (б)) в одномерной задаче о нагреве постоянным потоком тепла на разных сетках: сплошная линия – шаг перед ударной волной $h \approx 0.0005$, точки – шаг $h \approx 0.02$, треугольники – шаг $h \approx 0.05$ мм.



Фиг. 2. К построению схемы на этапе пересчета: узлы ячейки сетки (i, j) (точки) и соседние ячейки $(i-1, j)$, $(i+1, j)$, $(i, j-1)$ и $(i, j+1)$; стрелки – положительные направления к границам ячейки компонент скорости $U^i_{i,j}$, $U^j_{i,j}$.

Вернемся к описанию метода расчета двумерных течений. Расщепление на лагранжев этап и этап пересчета поля течения с сетки $G^L(t)$ на сетку $G(t+\tau)$ можно формально описать как расщепление интегродифференциальных уравнений, полученных интегрированием уравнений (1) по произвольной подвижной области $V(t)$ с заданной скоростью деформации границы $\mathbf{w}$ (см. [15]). Уравнения лагранжевого этапа получаются интегрированием по подвижной области со скоростью деформирования $\mathbf{u}$, а уравнение этапа пересчета имеет вид

$$\frac{d}{dt}\int \varphi dV + \oint \varphi(\mathbf{u} - \mathbf{w}, \mathbf{ds}) = 0, \tag{5}$$

где интегрирование ведется по области, деформируемой со скоростью $\mathbf{w} - \mathbf{u}$, вектор дифференциала границы области $\mathbf{ds}$ направлен по внешней нормали, а в качестве функций φ можно взять ρ, $\rho\mathbf{u}$ и $\rho\varepsilon$.

На фиг. 2 схематически изображена ячейка сетки i, j, состоящая из узлов (i, j), $(i+1, j)$, $(i+1, j+1)$, $(i, j+1)$. Нормальные к границам ячейки компоненты скорости $U^i_{i,j}$ и $U^j_{i,j}$ определяются как результат осреднения соответствующих величин, возникавших на каждом из двух лагранжевых этапов. Например,

$$U^i_{i,j} = \frac{\tau^{1i}_{i,j}U^{1i}_{i,j} + (\tau - \tau^{1i}_{i,j})U^{2i}_{i,j}}{\tau},$$

где $\tau^{1i}_{i,j}$ – шаг по времени на первом лагранжевом этапе для соответствующей границы, полученный интерполяцией значений $\tau^{(1)}_{i,j}$ в соседних ячейках, $U^{1i}_{i,j}$, $U^{2i}_{i,j}$ – значения нормальной компоненты скорости на первом и втором этапах.

Площадь четырехугольника с вершинами $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$, где $\mathbf{r} = (x, y)$, пронумерованных так, что их обход происходит против часовой стрелки, определяется известной формулой

$$V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \psi_{12} + \psi_{23} + \psi_{34} + \psi_{41}, \quad \psi_{mn} = (x_m y_n - x_n y_m)/2. \tag{6}$$

Площадь ячейки i, j сеток $G(t)$ и $G(t+\tau)$ определяется по формулам

$$V_{i,j} = V(\mathbf{r}_{i,j}, \mathbf{r}_{i+1,j}, \mathbf{r}_{i+1,j+1}, \mathbf{r}_{i,j+1}), \tag{7}$$
$$V^{i,j} = V(\mathbf{r}^{i,j}, \mathbf{r}^{i+1,j}, \mathbf{r}^{i+1,j+1}, \mathbf{r}^{i,j+1}),$$

где величины с верхними индексами относятся к сетке $G(t+\tau)$. Площади между границами ячеек сеток $G(t+\tau)$ и $G(t)$,

$$\Delta V^{iG}_{i,j} = V(\mathbf{r}_{i,j}, \mathbf{r}_{i,j+1}, \mathbf{r}^{i,j+1}, \mathbf{r}^{i,j}), \quad \Delta V^{jG}_{i,j} = V(\mathbf{r}_{i,j}, \mathbf{r}_{i+1,j}, \mathbf{r}^{i+1,j}, \mathbf{r}^{i,j}), \tag{8}$$

позволяют определить нормальные к границам скорости деформации ячеек сетки $G(t)$:

$$w^i_{i,j} = \Delta v^{iG}_{i,j}/(l^i_{i,j}\tau), \quad w^j_{i,j} = \Delta v^{jG}_{i,j}/(l^j_{i,j}\tau),$$

где $l^i_{i,j}$, $l^j_{i,j}$ – длины соответствующих границ сетки $G(t)$. Площади между границами ячеек сеток $G^L(t)$ и $G(t)$ определим формулами

$$\Delta v^{iL}_{i,j} = U^i_{i,j} l^i_{i,j}\tau, \quad \Delta v^{jL}_{i,j} = U^j_{i,j} l^j_{i,j}\tau.$$

Площадь ячейки i, j сетки $G^L(t)$ такова:

$$v^L_{i,j} = v_{i,j} + \Delta v^{iL}_{i,j} - \Delta v^{iL}_{i+1,j} + \Delta v^{jL}_{i,j+1} - \Delta v^{jL}_{i,j}. \tag{9}$$

Уравнение (5) аппроксимируется уравнением

$$\varphi^{i,j} v^{i,j} = \varphi_{i,j} v^L_{i,j} - \Delta v^i_{i,j}\varphi^i_{i,j} + \Delta v^i_{i+1,j}\varphi^i_{i+1,j} + \Delta v^j_{i,j+1}\varphi^j_{i,j+1} - \Delta v^j_{i,j}\varphi^j_{i,j}, \tag{10}$$

связывающим разностную функцию $\varphi^{i,j}$, определенную на сетке $G(t+\tau)$, и функцию $\varphi_{i,j}$, определенную на сетке $G^L(t)$. Здесь

$$\Delta v^i_{i,j} = \Delta v^{iL}_{i,j} - \Delta v^{iG}_{i,j}, \quad \Delta v^j_{i,j} = \Delta v^{jL}_{i,j} - \Delta v^{jG}_{i,j}, \tag{11}$$

а $\varphi^i_{i,j}$, $\varphi^j_{i,j}$ – некоторые выражения, аппроксимирующие функцию φ на границах ячеек. Как можно убедиться из (6)–(9) и (11), имеет место равенство

$$v^{i,j} = v^L_{i,j} - \Delta v^i_{i,j} + \Delta v^i_{i+1,j} + \Delta v^j_{i,j+1} - \Delta v^j_{i,j}, \tag{12}$$

в силу которого уравнение (10) обладает свойством интерполяционной формулы: если $\varphi_{i,j} = \varphi = \text{const}$, то $\varphi^{i,j} = \varphi$.

Рассмотрим схемы типа (10), в которых

$$\varphi^i_{i,j} = \begin{cases} \varphi^{i-}_{i,j}, & \Delta v^i_{i,j} > 0, \\ \varphi^{i+}_{i-1,j}, & \Delta v^i_{i,j} < 0, \end{cases} \qquad \varphi^j_{i,j} = \begin{cases} \varphi^{j+}_{i,j-1}, & \Delta v^j_{i,j} > 0, \\ \varphi^{j-}_{i,j}, & \Delta v^j_{i,j} < 0, \end{cases} \tag{13}$$

где величины $\varphi^{i-}_{i,j}$, $\varphi^{i+}_{i,j}$, $\varphi^{j-}_{i,j}$, $\varphi^{j+}_{i,j}$ относятся к ячейке (i, j). В случае $\varphi^{i-}_{i,j} = \varphi^{i+}_{i,j} = \varphi^{j-}_{i,j} = \varphi^{j+}_{i,j} = \varphi_{i,j}$ приходим к явной схеме первого порядка точности вида

$$\varphi^{i,j} = \sum_{m=-1}^{1}\sum_{n=-1}^{1} \alpha^{mn}_{i,j}\varphi_{i+m,j+n}, \tag{14}$$

коэффициенты которой удовлетворяют равенству

$$\sum_{m=-1}^{1}\sum_{n=-1}^{1} \alpha^{mn}_{i,j} = 1. \tag{15}$$

Как можно убедиться из (10), (13), $\alpha^{mn}_{i,j} \geq 0$ для всех m, n, кроме $m = n = 0$. В силу (15), неравенство

$$\alpha^{00}_{i,j} \geq 0 \tag{16}$$

является достаточным условием устойчивости схемы (14).

Явная квазимонотонная схема второго порядка точности (см. [16], [8]) строится с помощью разностных производных функции φ, которые принято называть наклонами и которые используются для вычисления величин $\varphi^{i-}_{i,j}$, $\varphi^{i+}_{i,j}$, $\varphi^{j-}_{i,j}$, $\varphi^{j+}_{i,j}$. Отсутствие заметных численных колебаний схемы [16], [8] обеспечивает так называемый принцип монотонных наклонов.

Непосредственное применение явной схемы на сетках с сильным сгущением и шагом по времени (2) невозможно из-за нарушения условия (16). Поэтому на этапе пересчета, как и на лагранжевом этапе, строится составная схема, объединяющая явную схему [16], [8] и неявную схему первого порядка точности $\varphi^{i-}_{i,j} = \varphi^{i+}_{i,j} = \varphi^{j-}_{i,j} = \varphi^{j+}_{i,j} = \varphi^{i,j}$. На первом этапе применяется явная схема, которая пересчитывает поле течения с сетки G^L на промежуточную сетку G_1, для которой условие (16) выполняется. При построении сетки G_1 она вначале совпадает с сеткой $G(t + \tau)$. Затем узлы тех ячеек, в которых условие (16) нарушено, по некоторому алгоритму двигаются по направлению к узлам сетки $G(t)$, пока условие (16) не выполнится для всех ячеек. Если коэффициенты в (10) на первом этапе обозначить через $\Delta v^{i1}_{i,j}$, $\Delta v^{j1}_{i,j}$, то коэффициенты второго этапа имеют вид $\Delta v^{i2}_{i,j} = \Delta v^{i}_{i,j} - \Delta v^{i1}_{i,j}$, $\Delta v^{j2}_{i,j} = \Delta v^{j}_{i,j} - \Delta v^{j1}_{i,j}$. При реализации неявной схемы вначале вычисляются обратная матрица, которая затем применяется для всех функций $\varphi = (\rho, \rho\mathbf{u}, \rho\varepsilon)$.

Некоторые линии сетки $G(t)$ объявляются лагранжевыми линиями, например граница раздела сред. Движение такой выделенной линии определяется решением задачи. На каждом шаге по времени узлы такой линии вначале перемещаются в соответствии с полем скорости, а затем перемещаются вдоль линии в соответствии с заданным законом их расстановки (см. [7]). Если линия сетки, отвечающая, например, значению $i = 1$, является лагранжевой, то переноса вещества через нее нет, т.е. в (10) следует положить $\Delta v^{i}_{1,j} = 0$. В то же время вычисленное по формуле (11) значение $\Delta v^{i}_{1,j} \neq 0$, а лишь стремится к нулю быстрее шага по времени и шага сетки вдоль лагранжевой линии. Можно сохранить формулу (11), положив $\Delta v^{iL}_{1,j} = \Delta v^{iG}_{1,j}$. Однако это приводит к неоднородности схемы на лагранжевом этапе, что может служить дополнительным источником колебаний численного решения вблизи границы. Поэтому было решено положить $\Delta v^{i}_{1,j} = 0$ в (10), не меняя формулу для $\Delta v^{iL}_{1,j}$. В этом случае для ячеек сетки, прилегающих к лагранжевой линии, значения $v^{1,j}$, вычисленные по формулам (7) и (12), уже не совпадают друг с другом. Чтобы схема (10) по-прежнему обладала свойством интерполяционной формулы, следует отказаться от (7) и вычислять $v^{1,j}$ по формуле (12). Для сеток без сильного сгущения к лагранжевой линии такой подход вполне приемлем. Однако для сеток с сильным сгущением, когда сеточный интервал вдоль лагранжевой линии много меньше интервала вдоль сеточных линий другого семейства, формула (12) может давать значение $v^{1,j} < 0$.

Заметим, что для внутренних ячеек сетки, ни одна из границ которых не принадлежит лагранжевой линии, $v^{i,j} > 0$, что следует из (7) и невырожденности сетки $G(t + \tau)$. По этой же причине, если $\Delta v^{i}_{1,j}$ вычислять по формуле (11), то и для ячеек, примыкающих к лагранжевой границе, формула (12) будет давать $v^{1,j} > 0$. Значение $v^{1,j}$ не изменится, если в (12) положить $\Delta v^{i}_{1,j} = 0$, $\Delta v^{i}_{2,j} = \Delta v^{i*}_{2,j} - \Delta v^{i*}_{1,j}$, где звездочкой отмечены величины, вычисляемые по формуле (11). Следующая модификация значений $\Delta v^{i}_{i,j}$ внутри области, т.е. при $i > 1$, позволила избежать появления отрицательных значений $v^{i,j}$:

$$\Delta v^{i}_{i,j} = \Delta v^{i*}_{i,j} - \Delta v^{i*}_{1,j}\xi(i), \quad i = 2, 3, \ldots, m,$$

где m – некоторое заданное число слоев, на которых выполняется модификация, $\xi(i)$ – некоторая монотонно убывающая функция, $\xi(1) = 1$, $\xi(m) = 0$.

2. МОДЕЛИРОВАНИЕ НАГРЕВА ТРЕНИЕМ ПРИ УДАРНОМ СЖАТИИ ПЛАСТИНЫ НА КЛИНЕ

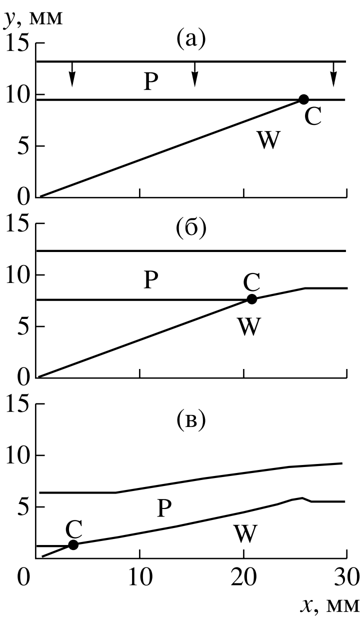
Рассматривается следующая задача (см. фиг. 3), моделирующая эксперимент из [17]. В алюминиевой пластине P толщиной 3 мм, лежащей на алюминиевом клине W, инициируется ударная волна. Угол между пластиной и клином $\alpha = 20°$. В эксперименте волна инициируется зарядом ВВ, а измеренная скорость свободной поверхности пластины после выхода на нее ударной волны примерно 4 км/с. В рассматриваемой задаче действие заряда ВВ моделируется алюминиевым ударником, налетающим на пластину со скоростью 4 км/с. В силу симметрии, скорость вещества за фронтом ударной волны в пластине 2 км/с, а правило удвоения скоростей (см. [18]) дает нужное значение скорости свободной поверхности 4 км/с. Точное решение соответствующей одномерной задачи с уравнением состояния алюминия из [14] отличается от этого приближенного результата на несколько процентов. Толщина ударника выбирается достаточно большой, чтобы волна разрежения от свободной поверхности ударника не успевала доходить до пластины за рассматриваемый интервал времени.

Как видно из фиг. 3, при выбранных параметрах задачи реализуется бесструйное сжатие, причиной которого являются присоединенные к точке контакта пластины и клина (точка С на фиг. 3) ударные волны в пластине и клине (см. [19], [20]). На фиг. 4 ударная волна в пластине видна в виде линии, на которой вектор скорости меняет направление.

Анализ экспериментальных данных, на котором мы здесь останавливаться не будем, позволяет предположить наличие в реальном течении сильного нагрева алюминия трением при скольжении пластины вдоль клина. Трение моделировалось заданием на границе раздела потока тепла

$$q = \nu p|u_{sl}|\chi(\rho), \quad \chi(\rho) = \begin{cases} 0, & \rho \le \rho_0, \\ (1-\rho_0/\rho)(1-\rho_0/\rho_1)^{-1}, & \rho_0 \le \rho \le \rho_1, \\ 1, & \rho \ge \rho_1, \end{cases}$$

где ν – коэффициент трения, u_{sl} – скорость скольжения, определенная как разность между проекциями на границу раздела скорости вещества в пластине и клине, множитель $\chi(\rho)$ обнуляет поток, если плотность становится меньше достаточно малого значения $\rho_0 \approx 1$ г/см^3 (нормальная плотность алюминия 2.7 г/см^3), и уменьшает поток в промежуточной области $\rho_0 \le \rho \le \rho_1 \approx 2$ г/см^3. Рассматривался нагрев только пластины. Основной вопрос, на который необходимо было дать ответ, заключался в том, может ли нагрев трением привести к нарушению бесструйного характера течения. Заметим, что угол между пластиной и клином $\alpha = 20°$ близок к критическому значению α_*, начиная с которого появляется кумулятивная струя. В частности, приближенная теория из [13] дает для скорости ударника 4 км/с величину $\alpha_* \approx 21°$.



Фиг. 3. Границы пластины (P) и клина (W) при $t = 0$ (график (а)), 0.85 (график (б)) и 2.4 мкс (график (в)).

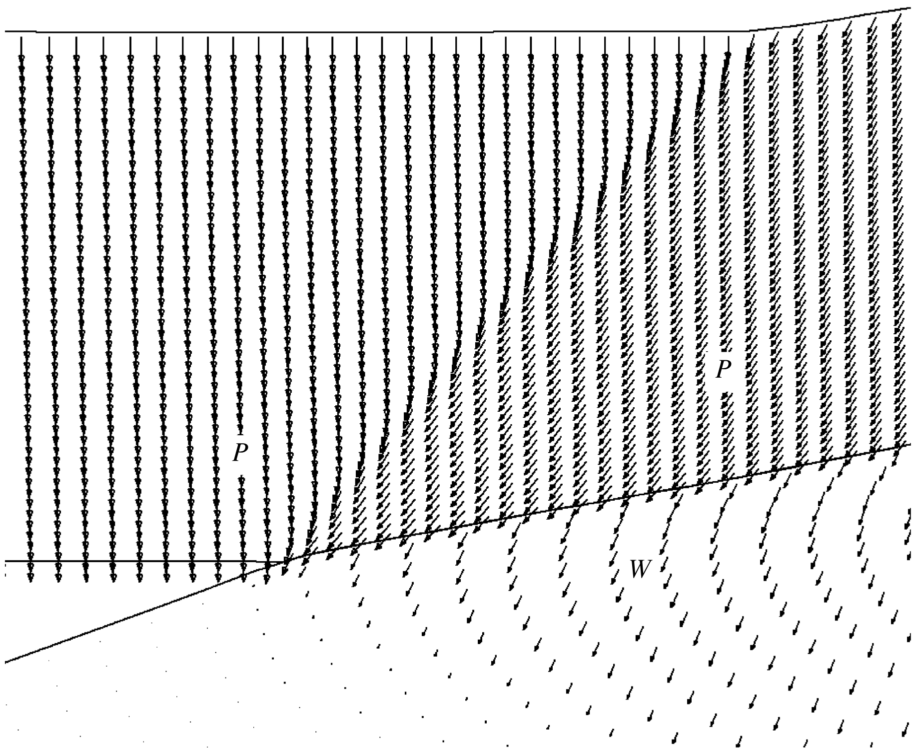
Тактика расчета следующая. В пластине и клине имеются два независимых сеточных блока. На границе между пластиной и клином сетки скользящие. Используется линейная переинтерполяция данных вдоль границы. Точка контакта С является фиксированным узлом сетки в пластине. До момента времени $t_0 = 0.85$ мкс, показанного на фиг. 3б, расчет велся по явной схеме на сетке без сгущения, что связано с желанием избежать использования немонотонной неявной схемы лагранжевого этапа при расчете отражения головной ударной волны от клина в области его начального контакта с пластиной. В момент времени t_0 поле течения пересчитывалось на сетку со сгущением, варианты которой показаны на фиг. 5, а затем использовалась описанная в предыдущем разделе составная схема.

Сетка строится с помощью модификации метода из [21], гарантирующего выпуклость всех ячеек сетки. Входными параметрами метода являются координаты узлов граничных линий и законы расстановки узлов вдоль линий одного семейства, в данном случае – линий, выходящих из нижней границы пластины. Эти линии отвечают значению индекса $j = \text{const}$, направление изменения которого показано на фиг. 5а стрелкой. В отличие от метода из [21], закон расстановки может быть своим для каждой линии.

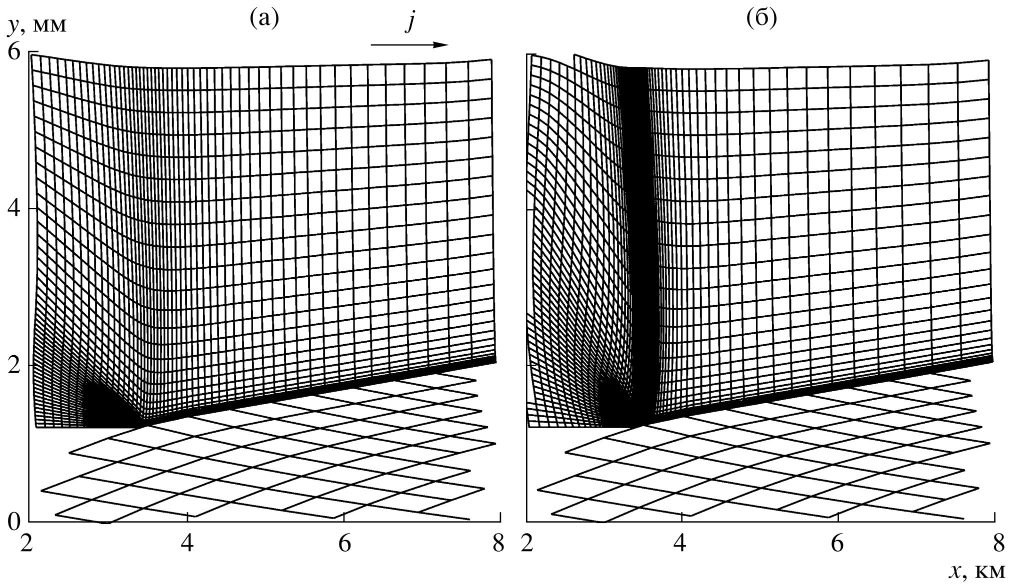
Под законом расстановки (см. [7]) понимается монотонно возрастающая последовательность

$$\eta_i, \quad i = 1, 2, \ldots, n, \quad \eta_1 = 0, \quad \eta_n = 1,$$

которая по заданному значению l определяет одномерную сетку $l_i = \eta_i l$. Для линий сетки с $j \ge j_C$, где индекс j_C отвечает линии, проходящей через точку контакта, строится фиктивная одномерная сетка $\{l_i\}$, по которой вычисляется закон расстановки. Сетка $\{l_i\}$ состоит из четырех участков с разным законом изменения шага сетки. На первом участке шаг сетки постоянен, затем идет участок с изменением шага по геометрической прогрессии, на третьем участке шаг изменяется по арифметической прогрессии, а на четвертом шаг опять постоянен. Параметрами сетки являются число интервалов на каждом участке, длина сетки l и первый шаг сетки l_1. В качестве l выбиралась длина линии на предыдущем шаге по времени, а значение $l_1 = 2 \times 10^{-5}$ мм выбира-



Фиг. 4. Поле скорости в окрестности точки контакта пластины и клина.



Фиг. 5. Сетки вблизи точки контакта с минимальным шагом вдоль границы раздела h_C = 0.05 (график (а)) и 0.01 мм (график (б)); t = 2.4 мкс.

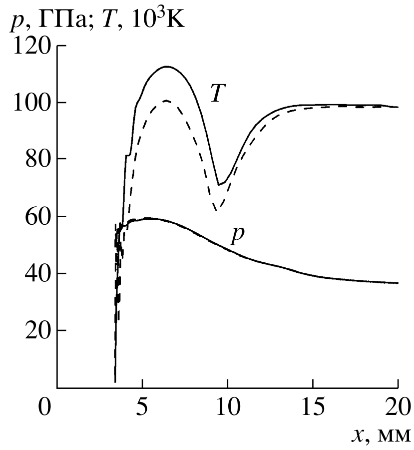
лось из анализа результатов расчета одномерной задачи, рассмотренной в предыдущем разделе. Для линий сетки вблизи левой границы пластины $1 \le j \le j_0$, где j_0 – параметр сетки, закон расстановки был равномерным: $\eta_i = = (i - 1)/n$, а в промежуточной области $j_0 < j < j_C$ закон расстановки вычислялся с помощью некоторой интерполяции законов при $j = j_0$ и j_C.

Для линии сетки, отвечающей нижней границе пластины, использовались свои законы расстановки слева и справа от точки контакта. Основным параметром был минимальный шаг h_C соответствующих фиктивных сеток вблизи точки контакта, изменение которого давало сетки с разным сгущением к точке контакта вдоль границы раздела. Две такие сетки показаны на фиг. 5.

Результаты расчета для коэффициента трения $\nu = 0.5$ на двух сетках показаны на фиг. 6. Как и в одномерной задаче, шаг по времени $\tau \approx 100\tau_K$. Распределение давления вдоль границы раздела с точностью до толщины линии на фиг. 6 не меняется при изменении сетки, а распределения температуры, полученные на разных сетках, близки.

Основной результат моделирования заключается в том, что трение не меняет бесструйного характера течения. Укажем также, что для выбранной модели трения температура алюминия на границе раздела, как видно из фиг. 6, может достигать значений порядка 10^5 K.

Авторы благодарят К.В. Хищенко из Института теплофизики экстремальных состояний Объединенного института высоких температур РАН за предоставленные в распоряжение одного из авторов таблицы уравнений состояния алюминия.



Фиг. 6. Температура и давление вдоль границы раздела для сеток с $h_C = 0.05$ (сплошная линия) и 0.01 мм (штриховая); $t = 2.4$ мкс.

СПИСОК ЛИТЕРАТУРЫ

1. *Rai M.M., Chakravarthy S.R.* An implicit form of the Osher upwing scheme // AIAA J. 1986. V. 24. № 5. P. 735–743.
2. *Иванов М.Я., Нигматуллин Р.З.* Неявная схема С.К. Годунова повышенной точности для численного интегрирования уравнений Эйлера // Ж. вычисл. матем. и матем. физ. 1987. Т. 27. № 11. С. 1725–1735.
3. *Махвидавзе Г.М., Щербак С.Б.* Разностная схема для численного исследования нестационарных двумерных движений сжимаемого газа: Препринт № 113. М.: ИПМехан. АН СССР, 1978.
4. *Гончаров В.А., Кривцов В.М., Чарахчьян А.А.* Численная схема моделирования дозвуковых течений вязкого сжимаемого газа // Ж. вычисл. матем. и матем. физ. 1988. Т. 28. № 12. С. 1858–1866.
5. *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.
6. *Charakhch'yan A.A.* Compound difference schemes for time-dependent equations on nonuniform nets // Communs. Numer. Meth. Eng. 1994. V. 10. № 2. P. 93–110.
7. *Годунов С.К., Забродин А.В., Иванов М.Я. и др.* Численное решение многомерных задач газовой динамики. М.: Наука, 1976.
8. *van Leer B.* A second-order sequel to Godunov's method // J. Comput. Phys. 1979. V. 32. № 1. P. 101–136.
9. *Чарахчьян А.А.* Об алгоритмах расчета распада разрыва для схемы С.К. Годунова // Ж. вычисл. матем. и матем. физ. 2000. Т. 40. № 5. С. 782–796.
10. *Чарахчьян А.А.* Расчет вязкого нагрева плазмы, сжимаемой ударником в конической мишени // Ж. вычисл. матем. и матем. физ. 1990. Т. 30. № 5. С. 767–779.
11. *Де Бор К.* Практическое руководство по сплайнам. М.: Радио и связь, 1985.
12. *Беляков Г.В., Чарахчьян А.А.* О нагреве сжимаемой жидкости постоянным потоком тепла // Прикл. механ. и техн. физ. 2003. Т. 44. № 2. С. 109–115.
13. *Чарахчьян А.А.* Ударное сжатие пластины на клине // Прикл. механ. и техн. физ. 2001. Т. 42. № 1. С. 17–24.
14. *Bushman A.V., Fortov V.E., Kanel G.I., Ni A.L.* Intense dynamic loading of condensed matter. Washington: Taylor & Francis, 1993.
15. *Чарахчьян А.А.* О симметричной и несимметричной схемах расщепления для уравнений газовой динамики // Ж. вычисл. матем. и матем. физ. 1991. Т. 31. № 11. С. 1692–1705.
16. *van Leer B.* Towards the altimate conservative difference scheme. IV. A new approach to numerical convection // J. Comput. Phys. 1977. V. 23. № 3. P. 276–299.
17. *Беляков Г.В., Чарахчьян А.А.* Высокотемпературные эффекты нагрева трением // VI Междунар. симп. по радиационной плазмодинамике. М.: НИИ "Инженер", 2003. С. 118–119.
18. *Зельдович Я.Б., Райзер Ю.П.* Физика ударных волн и высокотемпературных гидродинамических явлений. М.: Наука, 1966.
19. *Walsh J.M., Shreffler R.G., Willig F.J.* Limiting conditions for jet formation in high velocity collisions // J. Appl. Phys. 1953. V. 24. № 3. P. 349–359.
20. *Забабахин Е.И., Забабахин И.Е.* Явления неограниченной кумуляции. М.: Наука, 1988.
21. *Грынь В.И., Фролова А.А., Чарахчьян А.А.* Сеточный генератор барьерного типа и его применение для расчета течений с подвижными границами // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 6. С. 902–908.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 490–505

УДК 519.634

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ В ТРЕХМЕРНОЙ ПОСТАНОВКЕ СТРУИ ПЛАЗМЫ, ВЫХОДЯЩЕЙ В ОКРУЖАЮЩЕЕ ПРОСТРАНСТВО ИЗ СТАЦИОНАРНОГО ПЛАЗМЕННОГО ДВИГАТЕЛЯ

© 2007 г. А. С. Архипов, А. М. Бишаев

(125080 Москва, Ленинградское ш., 5, НИИ Прикл. механ. и электродинамики)

e-mail: bishaev@bk.ru

Поступила в редакцию 16.06.2006 г.

Предлагается численный метод, с помощью которого осуществляется моделирование струйного движения разреженной плазмы, возникающего от работы стационарного плазменного двигателя в трехмерной постановке. В отличие от работ, где эта задача рассматривалась в осесимметричном приближении, постановка задачи осуществляется так, чтобы было возможным определить влияние возникающих обратных ионных токов на области вверх по потоку и на корпус двигателя, который в рассматриваемом случае имеет конечный размер. Построенный численный метод является обобщением численных методов динамики разреженных газов на случай, когда движение происходит в незаданном аналитически силовом поле. Построение численного метода осуществляется так, чтобы учесть дельтообразность граничной функции распределения ионов и существенную разницу в масштабах скоростных пространств ионов и нейтралов, между которыми имеют место взаимные превращения. Приводятся результаты численных решений задачи, показывающие влияние некоторых факторов на возникающее течение. Библ. 10. Фиг. 12.

Ключевые слова: численное моделирование струи плазмы, модельные кинетические уравнения, численный метод конечных разностей.

ВВЕДЕНИЕ

Задача о струе плазмы, исходящей из стационарного плазменного двигателя (СПД), в осесимметричной постановке была рассмотрена в [1]. В [2] для моделирования упомянутого выше течения была использована специальная кинетическая модель, которая учитывала резонансную перезарядку – особый тип взаимодействия ионов и нейтралов, имеющих наибольшее сечение взаимодействия. С появлением ЭВМ типа “Пентиум” оказалось возможным увеличить размер счетной области и провести сравнение результатов расчетов с экспериментальными измерениями распределения плотности ионного тока. Методика проведения корректного сравнения экспериментальных измерений с расчетными данными и способ уточнения параметров в граничных условиях приведены в [3]. Проведение моделирования выявило наличие в струе ионных токов, направленных на срез двигателя. Чтобы определить влияние обратных токов на сам аппарат и увеличить возможности программного комплекса для проведения моделирования, не ограничиваясь спецификой геометрии течения, задача о струе была рассмотрена в трехмерной постановке.

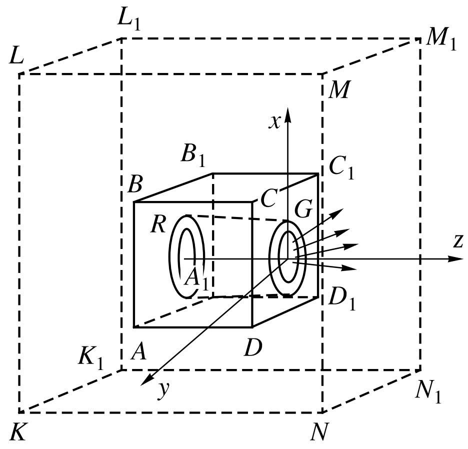
1. ПОСТАНОВКА ЗАДАЧИ О СТРУЕ

В основу моделирования струйного движения положена следующая система модельных кинетических уравнений для определения функции распределения ионов $f(\mathbf{x}, \boldsymbol{\xi})$ и нейтралов $g(\mathbf{x}, \mathbf{w})$, которая в безразмерных переменных имеет следующий вид:

$$\xi_j\frac{\partial f}{\partial x_j} + FE_k\frac{\partial f}{\partial \xi_k} = \nu_{\text{in}}(f_0 - f), \quad w_j\frac{\partial g}{\partial x_j} = \nu_{\text{ni}}(g_0 - g) + \nu_{\text{nn}}(g_M - g), \quad F = \frac{ekT_0^e}{2U_0}, \quad E_k = -\frac{\partial \Phi}{\partial x_k}, \tag{1.1}$$

$$\Phi = 5/2(n^i)^{2/3}, \quad j, k = 1, 2, 3.$$

Такое выражение для электрического поля получается, если использовать обобщение гипотезы “термолизованного” потенциала (см. [1], [2]), выдвинутой в [4], [5]. В (1.1) используются T_0^e – ха-



Фиг. 1.

рактерная температура электронов (≈3 эВ), U_0 – разрядное напряжение (300–600 В), k – постоянная Больцмана. В правой части (1.1) присутствуют частоты взаимодействия ион-нейтрал ν_{in}, нейтрал-ион ν_{ni} и нейтрал-нейтрал ν_{nn}. Они определены так же, как в [2]. Входящие в выражения для частот столкновений числа Кнудсена либо порядка единицы, либо больше единицы, f_0, g_0, g_M моделируют интегралы обратных столкновений. Они записываются в следующем виде:

$$
\begin{gathered}
f_0 = n^{\text{i}}(B_2/(\pi B_1 T^{\text{n}}))^{3/2} e^{-B_2 \mathbf{c}_\xi^2/T^{\text{n}}}, \quad \mathbf{c}_\xi = \boldsymbol{\xi} - \mathbf{u}^{\text{n}}/\sqrt{B_2}, \\
g_0 = n^{\text{n}}(B_1/(\pi B_2 T^{\text{i}}))^{3/2} e^{-B_1 \mathbf{c}_w^2/(B_2 T^{\text{i}})}, \quad \mathbf{c}_w = \mathbf{w} - \sqrt{B_2}\mathbf{u}^{\text{i}}, \\
g_M = n^{\text{n}}(\pi T^{\text{n}})^{-3/2} \exp\{-(\mathbf{w} - \mathbf{u})^2/T^{\text{n}}\},
\end{gathered}
\tag{1.2}
$$

где $B_1 = eU_0/(kT_0^{\text{i}})$, $B_2 = eU_0/(kT_0^{\text{n}})$, T_0^l, $l = \text{i, n}$, суть характерные значения температур ионов и нейтралов соответственно. Макропараметры в максвелловских функциях (1.2) определяются интегралами

$$
\begin{pmatrix} n \\ n\mathbf{u} \\ n\left(\dfrac{3}{2B_1}T + \mathbf{u}^2\right) \end{pmatrix}^i = B_1^{3/2}\int \begin{pmatrix} 1 \\ \boldsymbol{\xi} \\ \boldsymbol{\xi}^2 \end{pmatrix} f d\boldsymbol{\xi}, \quad \begin{pmatrix} n \\ nu \\ n(3T/2 + \mathbf{u}^2) \end{pmatrix}^n = \int \begin{pmatrix} 1 \\ \mathbf{w} \\ \mathbf{w}^2 \end{pmatrix} g d\mathbf{w}. \tag{1.3}
$$

Наличие в первом интеграле (1.3) множителя $B_1^{3/2}$ связано с введением двух скоростных масштабов для ионной компоненты $\xi_0 = (2eU_0/m)^{1/2}$, который есть масштаб скорости направленного движения иона, и $c_0 = (2kT_0^i/m)^{1/2}$ – соответственно, масштаб его теплового движения. Между этими масштабами имеет место соотношение $\xi_0 > c_0$.

Геометрию рассматриваемого струйного движения иллюстрирует фиг. 1. Параллелепипед $ABCDA_1B_1C_1D_1$ представляет корпус двигателя. Все грани являются твердыми поверхностями, кроме грани CC_1D_1D, в центре которой расположено кольцевое отверстие, откуда струя выходит в окружающее пространство. Внутри $ABCDA_1B_1C_1D_1$ штриховой линией показан ускорительный канал RG. Расчеты ведутся во внешности этого параллелепипеда, при этом граничные условия для (1.1) будут следующие. На CC_1D_1D ($z = 0$, $-1 \le x \le 1$, $-1 \le y \le 1$) для $\xi_z \ge 0$, $w_z \ge 0$ имеем

$$
f = \begin{cases} \bar{n}/\pi^{3/2}\exp\{-B_1(\boldsymbol{\xi} - \bar{\mathbf{u}})^2\}, & R_2 \le r \le R_1, \quad r = \sqrt{x^2 + y^2}, \\ 0, & (r < R_1) \cup (r > R_2); \end{cases} \tag{1.4}
$$

$$g = \begin{cases} \tilde{n}/\pi^{3/2}\exp\{-(\mathbf{w}-\tilde{\mathbf{u}})^2\}, & R_2 \le r \le R_1, \\ n_w(B_w/\pi)^{3/2}\exp\{-B_w\mathbf{w}^2\}, & (r < R_2)\cup(r > R_1); \end{cases} \tag{1.5}$$

здесь $\bar{n}$, $\tilde{n}$, $\bar{\mathbf{u}}$, $\tilde{\mathbf{u}}$ суть заданные функции x и y (отметим, что $x = x_1$, $y = x_2$, $z = x_3$). Они определялись так же, как в [3], кроме распределения азимутальной компоненты $\bar{\mathbf{u}}$, которая в отличие от цитируемой выше работы может быть не равной нулю. Величины R_1 и R_2 являются, соответственно, внутренним и внешним безразмерными радиусами отверстия, откуда выходит струя, $B_w = T_0^{\mathrm{n}}/T_w$, где T_w – температура корпуса двигателя. Фигурирующая во второй формуле (1.5) n_w находится из баланса потоков падающих и отраженных от поверхности нейтралов и ионов (см. [1]–[3]) и определяется следующей формулой:

$$n_w = -2\sqrt{\pi B_w}\left(\int\limits_{-\infty}^{0} w_z\left(\iint g(x, y, 0, \mathbf{w})dw_r dw_\varphi\right)dw_z + B_1^{3/2}\theta_1\sqrt{\frac{B_2}{B_1}}\int\limits_{-\infty}^{0}\xi_z\left(\iint f(x, y, 0, \boldsymbol{\xi})d\xi_r d\xi_\varphi\right)d\xi_z\right).$$

На остальных гранях $ABCDA_1B_1C_1D_1$ граничные условия задаются вторыми формулами (1.4) и (1.5) для $\xi_n = (\boldsymbol{\xi}\mathbf{n}) \ge 0$, $w_n = (\mathbf{wn}) \ge 0$, где $\mathbf{n}$ – внешняя нормаль к соответствующей грани.

Помимо граничных условий на твердых поверхностях, необходимо задать граничные условия на бесконечности. Они следующие:

$$f(\mathbf{x}, \boldsymbol{\xi}) = 0, \quad g(\mathbf{x}, \mathbf{w}) = \theta_2(B_\infty/\pi)^{3/2}\exp\{-B_\infty(\mathbf{w}-\mathbf{u}_\infty)^2\}\sqrt{x^2+y^2+z^2} \longrightarrow \infty, \tag{1.6}$$

где $\theta_2 = n_\infty/n_0^{\mathrm{n}}$, $B_\infty = T_0^{\mathrm{n}}/T_\infty$. Условия (1.6) при численном решении задачи сносятся на границу счетной области. Сама счетная область изображена на фиг. 1 штриховыми линиями и имеет форму параллелепипеда $KLMNK_1L_1M_1N_1$. В приведенной постановке существенно то, что расчеты проводятся как вниз, так и вверх по струе. Это позволило учесть влияние обратных токов на сам аппарат. Фактически рассматривается задача об обтекании аппарата исходящей из него струей. В [1]–[3] считалось, что срез двигателя простирался до бесконечности, поэтому граничная максвелловская функция нейтралов имела скорость, равную нулю, и температуру T_w. В рассматриваемой выше постановке соответствующая граничная функция может иметь произвольные значения $\mathbf{u}_\infty$ и T_∞. Это обстоятельство приводит к тому, что появляются дополнительные безразмерные параметры, определяющие рассматриваемое движение. Их наличие может позволить моделировать процесс откачки вакуумной камеры, куда обычно помещают СПД при проведении экспериментальных исследований.

2. ОБЩИЕ ВОПРОСЫ ПОСТРОЕНИЯ ЧИСЛЕННОЙ СХЕМЫ РЕШЕНИЯ

Содержащаяся в предыдущем разделе математическая задача формулируется так: найти решение уравнений (1.1)–(1.3) для функций f и g, удовлетворяющих граничным условиям (1.4), (1.5) и условиям на бесконечности (1.6). Эта задача решается численно с помощью метода, являющегося обобщением численных методов, которые применяются в динамике разреженных газов при решении стационарных кинетических уравнений (см. [6], [7]). Этот метод базируется на следующей итерационной процедуре:

$$\begin{gathered} \xi_j\frac{\partial f^k}{\partial x_j} + FE_j^{k-1}\frac{\partial f^k}{\partial \xi_j} = F_+ - \nu_1 f^k, \quad \nu_1 = \nu_{\mathrm{in}}^{k-1}, \quad F_+ = \nu_1 f_0^{k-1}, \quad w_j\frac{\partial g^k}{\partial x_j} = G_+^1 + G_+^2 - \nu_2 g^k, \\ \nu_2 = \nu_{\mathrm{ni}}^{k-1} + \nu_{\mathrm{nn}}^{k-1}, \quad G_+^1 = \nu_{\mathrm{ni}}^{k-1}g_0^{k-1}, \quad G_+^2 = \nu_{\mathrm{nn}}^{k-1}g_M^{k-1}. \end{gathered} \tag{2.1}$$

Индекс k в (2.1) означает номер итерации. Все выражения, отмеченные в (2.1) индексом $k-1$, известны, так как они выражаются через найденные в $(k-1)$-й итерации макропараметры ионов и нейтралов.

Основным методом, который применяется в разреженном газе, является метод характеристик. В большинстве задач разреженных газов силовое поле отсутствует (исключением здесь является работа [8]), поэтому характеристики дифференциальных частей уравнений – это прямые линии. В рассматриваемом случае уравнения характеристик (траекторий иона), приходящих в точку фазового пространства $(\mathbf{x}, \boldsymbol{\xi})$, будут следующие:

$$\frac{d\tilde{\mathbf{x}}}{d\tau} = -\boldsymbol{\xi}, \quad \frac{d\tilde{\boldsymbol{\xi}}}{d\tau} = -\mathbf{E}, \quad \tilde{\mathbf{x}}(0) = \mathbf{x}, \quad \tilde{\boldsymbol{\xi}}(0) = \boldsymbol{\xi}, \quad \mathbf{E} = F\mathbf{E}^{k-1}(\tilde{\mathbf{x}}(\tau)). \tag{2.2}$$

Зная $\tilde{\mathbf{x}}(\tau)$, $\tilde{\boldsymbol{\xi}}(\tau)$, решение первого уравнения (2.1) можно представить в виде

$$f^k(\mathbf{x}, \boldsymbol{\xi}) = f_b(\tilde{\mathbf{x}}(t_b), \tilde{\boldsymbol{\xi}}(t_b))\exp\left(-\int_0^{t_b} \nu_1(\tilde{\mathbf{x}}(s))ds\right) + \int_0^{t_b} F^+(\tilde{\mathbf{x}}(\tau), \tilde{\boldsymbol{\xi}}(\tau))\exp\left(-\int_0^{\tau} \nu_1(\tilde{\mathbf{x}}(s))ds\right)d\tau. \tag{2.3}$$

В (2.3) величина t_b есть значение параметра τ, когда характеристика пересекает счетную область. Это может быть грань $ABCDA_1B_1C_1D_1$ или грань $KLMNK_1L_1M_1N_1$. Фигурирующая в (2.3) величина f_b есть заданная на соответствующей границе функция распределения ионов.

Принципиальным отличием рассматриваемой задачи от задач разреженного газа является наличие электрического поля, определяемого численно и известного только в узлах или ячейках разбиения физического пространства. Определяемое в ходе численного решения электрическое поле может оказаться таким, что траектория иона никогда не будет выходить на границу счетной области. Ситуация здесь принципиально отличается от ситуации в [8], где такие характеристики выделялись аналитически.

Формула (2.3) (решение второго уравнения (2.1) записывается аналогично, см. [7]) позволяет проанализировать особенности решения уравнения (2.1). Видно, что f^k и g^k будут разрывными в пространстве скоростей функции, причем так как B_1 в (1.3) существенно больше единицы ($B_1 = 20$), то из кольцевого отверстия выходит дельтообразная функция. Площадь выходного отверстия много меньше площади грани CC_1D_1D, поэтому если не принять меры, то в точку $\mathbf{x}$ придет только малая часть характеристик, выходящих из отверстия, а это приведет к тому, что большая часть информации о струе будет утеряна. Из-за большой величины параметра B_1 функция распределения ионов будет иметь многогорбый характер в пространстве скоростей. Процессы перезарядки будут переносить указанную многогорбость на функцию распределения нейтралов. В разреженном газе многогорбость у функции распределения имеет место в гиперзвуковых течениях. Трудность расчета макропараметров в этом случае связана с тем, что из-за "горячих" частиц функция распределения будет иметь медленно затухающий хвост, что приведет к большому значению чисел, моделирующих бесконечные пределы в интегралах (1.3), а шаг разбиения скоростного пространства должен выбираться в соответствии с узким дельтообразным пиком у функции распределения, обусловленным "холодными" частицами. Это приводит к неприемлемо большому времени счета, которое требуется для решения задачи с таким большим количеством скоростных узлов.

Для преодоления указанной трудности предлагается следующая процедура. Из-за линейности (2.1) на k-й итерации их решения можно представить в виде $f^k = f_1 + f_2$, $g^k = g_1 + g_2 + g_3$, где введенные функции удовлетворяют следующим уравнениям:

$$\begin{gathered} \xi_i\frac{\partial f_1}{\partial x_i} + E_i\frac{\partial f_1}{\partial \xi_i} = -\nu_1 f_1, \quad \xi_i\frac{\partial f_2}{\partial x_i} + E_i\frac{\partial f_2}{\partial \xi_i} = \nu_1 F^+ - \nu_1 f_2, \\ w_i\frac{\partial g_1}{\partial x_i} = -\nu_2 g_1, \quad w_i\frac{\partial g_s}{\partial x_i} = \nu_s G^+_{s-1} - \nu_2 g_s, \quad s = 2, 3, \quad i = 1, 2, 3. \end{gathered} \tag{2.4}$$

Функции f_1 и g_1 подчинены граничным условиям (1.3), (1.4), а f_2, g_2, g_3 равны нулю на границе счетной области. Предполагается, что каждая из указанных выше функций ответственна за свой горб в функции распределения, к тому же эти функции, кроме f_1, непрерывны в пространстве скоростей, поэтому в каждой точке физического пространства можно специальным образом производить разбиение скоростного пространства и эффективно вычислять соответствующие вклады в макропараметры на небольшом количестве узлов скоростной сетки.

3. ЧИСЛЕННАЯ СХЕМА ДЛЯ ОПРЕДЕЛЕНИЯ ВЛИЯНИЯ f_1

Уравнение (2.4) для f_1 решается с граничным условием (1.3). Кроме выходного отверстия, на всех других границах и на бесконечности f_1 равна нулю, поэтому именно ионы, выходящие из отверстия, будут формировать струю ионов. Влияние этих частиц распространяется по характеристикам (2.2), выпущенным из отверстия и приходящим в точку $\mathbf{x}$ физического пространства. Так как $B_1 > 1$, то влияние отверстия будет передаваться теми характеристиками, скорость ионов вдоль которых мало отличается от $\mathbf{u} = \{\bar{u}_r(r), \bar{u}_z(r), \bar{u}_\varphi(r)\}$. Значение $\bar{u}_z(r)$ бралось из эксперимента (см. [3]), а оно таково, что по всему отверстию является величиной порядка единицы. Составляющая электрического поля E_z имеет порядок F. В струе F мaлó (≈ 0.01), поэтому электрическое поле слабо влияет на траектории ионов, определяющие носитель функции распределения. Отсюда следует, что выходное отверстие будет оказывать влияние на области вниз по потоку. На фиг. 1 это – область, где $z \ge 0$.

Характеристики, приходящие в точку $\mathbf{x}$, представим в виде

$$\boldsymbol{\xi}(s) = \boldsymbol{\xi} - \int_0^s \mathbf{E}(\tilde{\mathbf{x}}(\tau))d\tau, \quad \tilde{\mathbf{x}}(s) = \mathbf{x} - \boldsymbol{\xi}s + \int_0^s \left(\int_0^\tau \mathbf{E}(\tilde{\mathbf{x}}(\sigma))d\sigma\right)d\tau. \tag{3.1}$$

Для характеристик (3.1), приходящих в точку $\mathbf{x}$ из отверстия, значения скорости $\boldsymbol{\xi}$ таково, что имеют место соотношения $D = \{\tilde{\xi}_z(t_b) \ge 0, R_2 \le r \le R_1, r = \sqrt{\tilde{x}^2(t_b) + \tilde{y}^2(t_b)}\}$, где t_b определяется соотношением $\tilde{z}(t_b) = 0$. Из представления для f^k следует, что $n_i^k = n_1 + n_2$. Используя введенные выше обозначения, будем иметь

$$n_1 = (B_1/\pi)^{3/2} \int_D \bar{n}(r)\pi^{-3/2} \exp\left\{-B_1(\boldsymbol{\xi}(t_b) - \bar{\mathbf{u}}(r))^2 - \int_0^{t_b} \nu(\tilde{\mathbf{x}}(s))ds\right\}d\boldsymbol{\xi}. \tag{3.2}$$

Введем обозначения

$$\tilde{\mathbf{x}}(t_b) = \mathbf{x}^* - \boldsymbol{\xi}t_b, \quad \mathbf{x}^* = x + \int_0^{t_b}\left(\int_0^\tau \mathbf{E}(\tilde{\mathbf{x}}(\sigma))d\sigma\right)d\tau, \quad \mathbf{x}^* = \{x^*, y^*, z^*\},$$

$$\tilde{\boldsymbol{\xi}}(t_b) = \boldsymbol{\xi} - \mathbf{A}, \quad \mathbf{A} = \int_0^{t_b} \mathbf{E}(\tilde{\mathbf{x}}(s))ds, \quad \mathbf{A} = \{A_x, A_y, A_z\}.$$

Тогда $t_b = z^*/\xi_z$, а $D = \{\xi_z - A_z \ge 0, R_2 \le \sqrt{(x^* - \xi_x t_b)^2 + (y^* - \xi_y)^2} \le R_2\}$. В интеграле (3.2) перейдем к новым переменным интегрирования переменных:

$$\xi_x = \lambda(x^* - r\cos\theta), \quad \xi_y = \lambda(y^* - r\sin\theta), \quad \xi_z = p, \quad \lambda = p/z^*. \tag{3.3}$$

В результате преобразования (3.3) область D перейдет в $\bar{D} = \{A_z \le p \le \infty, R_2 \le r \le R_1, 0 \le \theta \le 2\pi\}$. Якобиан преобразования (3.3) запишем в виде $J = \lambda^2 r\bar{J}$, где

$$\bar{J} = 1 + r^{-1}\left(\sin\theta\frac{\partial x^*}{\partial\theta} - \cos\theta\frac{\partial y^*}{\partial\theta}\right) - \cos\theta\frac{\partial x^*}{\partial r} - \sin\theta\frac{\partial y^*}{\partial r} + r^{-1}\frac{\partial(x^*, y^*)}{\partial(r, \theta)} +$$

$$+ \frac{y^* - r\sin\theta}{z^* r}\left(\cos\theta\frac{\partial z^*}{\partial\theta} + r\sin\theta\frac{\partial z^*}{\partial r} - \frac{\partial(x^*, z^*)}{\partial(r, \theta)}\right) + \frac{x^* - r\cos\theta}{z^* r}\left(r\cos\theta\frac{\partial z^*}{\partial r} - \sin\theta\frac{\partial z^*}{\partial\theta} + \frac{\partial(y^*, z^*)}{\partial(r, \theta)}\right),$$

$$\frac{\partial(Q, G)}{\partial(r, \theta)} = \frac{\partial Q}{\partial r}\frac{\partial G}{\partial\theta} - \frac{\partial Q}{\partial\theta}\frac{\partial G}{\partial r}.$$

В результате перехода к новым переменным интеграл (2.2) теперь будет вычисляться по площади выходного отверстия и одной скоростной переменной. Это дает возможность для его вычисления построить численную схему, учитывающую дельтообразный характер функции распределения.

Разобьем отрезки интегрирования по r и θ на m_0 и n_0 частей соответственно, т.е.

$$\int_{\bar{R}_1}^{1}\int_{0}^{2\pi}(\cdot)d\theta dr = \sum_{m=1}^{m_0}\sum_{n=1}^{n_0}\int_{r_{m-1}}^{r_m}\int_{\theta_{n-1}}^{\theta_n}(\cdot)d\theta dr,$$

где $r_m = \bar{R}_1 + (m-1)\Delta r$, $m = 1, 2, \ldots, m_0$, $\theta_n = (n-1)\Delta\theta$, $n = 1, 2, \ldots, n_0$, $\Delta r = (R_1 - R_2)/m_0$, $\Delta\theta = 2\pi/n_0$. В каждом двойном интеграле по области $[r_{m-1}, r_m]\times[\theta_{n-1}, \theta_n]$ величины $\mathbf{x}$, $\mathbf{A}$, $\bar{n}(r)$, $\bar{\mathbf{u}}(r)$, $\bar{J}$ считаются постоянными и вычисленными в каждом таком интеграле в точках $\bar{r}_m = (r_{m-1} + r_m)/2$, $\bar{\theta}_n = (\theta_{n-1} + \theta_n)/2$. Сделаем замену переменной p: $p = A_z + \bar{u}_z + cB_1^{-1/2}$. Тогда имеем

$$n_1^i \approx \sum_{m=1}^{m_0}\sum_{n=1}^{n_0}D_{mn}, \quad D_{mn} = \int_{r_{m-1}}^{r_m}\int_{\theta_{m-1}}^{\theta_m}\int_{-\bar{u}_z\sqrt{B_1}}^{\infty}\bar{n}\lambda^2 r|\bar{J}|B_1 e^{-\bar{\nu}-p^2}\left(\left(1-\frac{\partial A_z}{\partial p}\right)\pi^{3/2}\right)^{-1}d\mathbf{G},$$

$$d\mathbf{G} = e^{-\lambda^2 B_1 g^2}dc d\theta dr, \quad g^2 = [r - \rho\cos(\theta-\tilde{\theta}) + q\bar{u}_r(r)]^2 + [\rho\sin(\theta-\tilde{\theta}) + \bar{u}_\varphi]^2, \quad \bar{\nu} = \int_0^{t_b}\nu(\mathbf{x}(s))ds, \tag{3.4}$$

$$\rho = \sqrt{(x^* - qA_x)^2 + (y^* - qA_y)^2}, \quad \tilde{\theta} = \operatorname{arctg}(y^* - qA_y)/(x^* - qA_x), \quad q = \lambda^{-1}.$$

В (3.4) интеграл по переменной c вычисляется при помощи квадратурных формул Гаусса вида

$$Q = \int_{-\bar{u}_z\sqrt{B_1}}^{\infty}h(c)e^{-c^2}dc \cong \sum_{k=1}^{k_0}C_k h(c_k),$$

где C_k – коэффициенты, а c_k – узлы соответствующей квадратурной формулы. Если $\bar{u}_z\sqrt{B_1} \geq 3$, то интеграл вычисляется квадратурной формулой Гаусса для интегралов типа $\int_{-\infty}^{\infty}h(c)e^{-c^2}dc$. При $\bar{u}_z\sqrt{B_1} < 3$ интеграл разбивается на два, а именно:

$$Q = \int_{-\bar{u}_z\sqrt{B_1}}^{0}h(c)e^{-c^2}dc + \int_0^{\infty}(2t^{1/2})^{-1}e^{-t}h(t)dt.$$

Первый интеграл вычисляется по Гауссу для интегралов вида $\int_{-a}^{a}h(u)du$ (с помощью линейной замены переменных можно сделать симметричные пределы интегрирования). Второй интеграл, в котором предварительно сделана замена $c^2 = t$, вычисляется с помощью формул Гаусса для интегралов $\int_0^{\infty}h(t)e^{-t}dt$. Введем в (3.4) новые переменные $t = r - \rho_k\cos(\theta - \tilde{\theta}) + \bar{u}_r q_k$, $\gamma = \sin(\theta - \tilde{\theta}) + \bar{u}_\varphi q_k/\rho_k$. Выполнив интегрирование в двойных интегралах, получим

$$n_1 = \sum_{m=1}^{m_0}\sum_{k=1}^{k_0}\sum_{n=1}^{n_0}C_k S_k/4(q_k\bar{\Theta}_n/\rho_k(Q_m - R_m\bar{u}_r) + \Theta_n R_m), \text{ где } S_k = \left|J_k/\left(1-\frac{\partial A_z}{\partial p}\right)_k\right|e^{-\nu_k}\pi^{-1/2},$$

$$\Theta_n = [\operatorname{sign}(\gamma_n)\operatorname{Erf}(\omega_k\rho_k|\gamma_n|) - \operatorname{sign}(\gamma_{n-1})\operatorname{Erf}(\omega_k\rho_k|\gamma_{n-1}|)], \quad \bar{\Theta}_n = \Theta_n/\mu, \quad \mu = \cos(\bar{\theta} - \tilde{\theta}_k),$$

$$Q_m = (B_1\pi)^{-1/2}[\exp(-\omega_k^2 u_{m-1}^2) - \exp(-\omega_k^2 u_m^2)],$$

$$R_m = [\operatorname{sign}(u_m)\operatorname{Erf}(\omega_k|u_m|) - \operatorname{sign}(u_{m-1})\operatorname{Erf}(\omega_k|u_{m-1}|)],$$

$$u_l = r_l - \rho_k\cos(\bar{\theta} - \tilde{\theta}_k) + \bar{u}_r q_k, \quad l = m-1, \quad m, \omega_l = \lambda_k\sqrt{B_1}.$$

Трудности использования построенной выше схемы состоят в том, что для величин $\mathbf{x}^*$, A_z, J, $\partial A_z/\partial p$ и производных, входящих в J, отсутствуют аналитические выражения и их приходится определять численно. Для этого нужно найти такие характеристики (2.2), которые, будучи выпущенными из отверстия со скоростями, принадлежащими носителю заданной на границе функции распределения, попадали бы в заданную точку $\mathbf{x}$ физического пространства. На таких характеристиках имеют место следующие соотношения:

$$\tilde{\mathbf{x}}(0) = \mathbf{x}, \quad \tilde{z}(t_b) = 0, \quad \tilde{x}(t_b) = \bar{r}_m\cos\bar{\theta}_n, \quad \tilde{y}(t_b) = \bar{r}_m\sin\bar{\theta}_n, \quad \tilde{\boldsymbol{\xi}}_z(\mathbf{x}) = p_k = c_k B_1^{-1/2} + \bar{u}_z(\bar{r}_m) + A_z.$$

Так как на этих траекториях $\tilde{\xi}_z \ge 0$, то они могут быть найдены из решения задачи

$$L\mathbf{q} = \frac{d^2\mathbf{q}}{d\tilde{z}^2} + \frac{1}{\tilde{\xi}_z^2}\left(\frac{d\mathbf{q}}{d\tilde{z}} - \mathbf{E}_q\right) = 0, \quad \mathbf{q} = \{\tilde{x}(\tilde{z}), \tilde{y}(\tilde{z})\}, \quad \mathbf{E}_q = \{E_x, E_y\}, \quad 0 \le \tilde{z} \le z,$$

$$\tilde{\xi}_z^2 = p_k^2 - 2\int_{\tilde{z}}^{z} E_z(\mathbf{q}((s), s))ds, \quad \mathbf{q}(0) = \{\bar{r}_m\cos\bar{\theta}_n, \bar{r}_m\sin\bar{\theta}_n\} = \mathbf{q}_0, \quad \mathbf{q}(z) = \{x, y\} = \mathbf{q}_1. \tag{3.5}$$

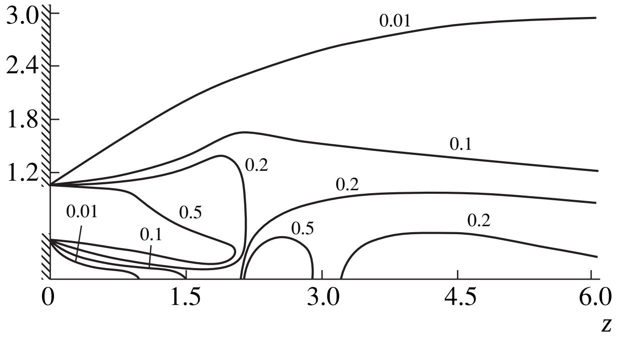
Краевая задача (3.5) решалась методом установления по схеме $\partial\mathbf{q}/\partial t = L\mathbf{q}$, $\mathbf{q}(0) = \mathbf{q}_0$, $\mathbf{q}(z) = \mathbf{q}_1$. При этом система уравнений решалась численно при помощи трехточечной разностной схемы:

$$\frac{\mathbf{q}_i^{l+1} - \mathbf{q}_i^l}{\Delta t} = \frac{\mathbf{q}_{i+1}^{l+1} - 2\mathbf{q}_i^{l+1} + \mathbf{q}_{i-1}^{l+1}}{\Delta z^2} + C_i\frac{\mathbf{q}_{i+1}^{l+1} - \mathbf{q}_{i-1}^{l+1}}{2\Delta z} - \mathbf{F}_i, \quad C_i = \left(\frac{E_z}{\tilde{\xi}_z^2}\right)_i^l, \quad \mathbf{F}_i = \left(\frac{\mathbf{E}_q}{\tilde{\xi}_z^2}\right)_i^l,$$

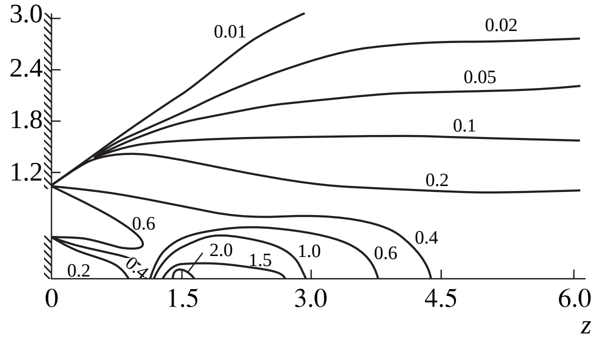
где все нелинейные члены брались с предыдущего шага по времени. Решение разностной задачи осуществлялось обычным методом прогонки (см. [9]). Из приведенного в [9] условия корректности и устойчивости метода в нашем случае вытекает следующее: если $e_i = \Delta z|C_i|/2 < 1$, то шаг по времени Δt может быть любым, в противном случае Δt связан с Δz соотношением $2\Delta t \le \Delta z^2/(e_i - 1)$. Считалось, что режим установления решения наступил, если было выполнено неравенство $\max\limits_{0 \le \tilde{z} \le z}\left(\sqrt{(\tilde{x}^{l+1} - \tilde{x}^l)^2 + (\tilde{y}^{l+1} - \tilde{y}^l)^2}\right) < 0.5p$, где p есть минимальный размер ячейки разбиения физического пространства. Если правая часть в выражении для $\tilde{\xi}_z^2$ становилась отрицательной, то соответствующий коэффициент C_k обнулялся.

Описанные выше схемы использовались во всех расчетах задачи о струе. Надо отметить, что на работу этого блока тратилась бóльшая часть времени счета одной итерации, но это необходимо, так как корректный учет влияния отверстия позволяет определить всю структуру струйного течения вниз по потоку.

На фиг. 2 и фиг. 3 показаны линии равной плотности ионов (см. [1]), полученные при различных методах расчетов. На фиг. 2 изображены изоконцентрали ионов, полученные в расчетах, когда не проводилось разбиения функции распределения, а характеристика тянулась к границе от точки $\mathbf{x}$. На фиг. 3 такие же расчеты были сделаны с разбиением функции распределения и для определения вкладов от f_1 использовалась описанная выше процедура. Различие в картинах распределения плотности существенно. Видно, что на фиг. 2 отсутствует “кроссовер” (располагающаяся на оси симметрии область повышенной плотности ионов, см. [1]–[3]). Так что хотя по-



Фиг. 2.



Фиг. 3.

строенная процедура требует довольно значительных затрат машинного времени, она позволяет адекватно воспроизвести картину возникающего течения плазменной струи.

При $z \longrightarrow 0$ дельтообразный характер подынтегральной функции для r и θ в интегралах (3.4) имеет место и для $B_1 \approx 1$. Точное значение вклада от f_1 будет равно $n_1 = \bar{n}/\{2[1 + \mathrm{Erf}(\bar{u}_z\sqrt{B_1})]\}$. При $z \longrightarrow 0$ из приведенной выше формулы для численного расчета n_1 можно получить, что

$$n_1^i \cong \bar{n}(\bar{r}_m)\sum_{k=1}^{k_0} C_k h(p_k).$$

Отсюда следует, что построенные выше формулы в случае $z \longrightarrow 0$ имеют по Δr точность не ниже первого порядка. Точность вычисления интеграла по ξ соответствует точности квадратурной формулы Гаусса. В невырожденных случаях построенные формулы обладают вторым порядком точности как по Δr, так и по $\Delta\theta$. В этом нетрудно убедиться, если разложить все экспоненты и "эрфы" по формуле Тейлора.

В случае $\rho_k \cong 0$ значение плотности и других макропараметров может сильно возрастать. Этот эффект особенно сильно проявляется в осесимметричном случае, когда на оси симметрии плотность ионов становится порядка $\sqrt{B_1}$. Нетрудно видеть, что в этом случае построенная численная схема дает правильный результат со вторым порядком точности по $\Delta\theta$ и Δr.

Таким образом, построенная выше схема вычислений является равномерно пригодной, т.е. дает правильный главный асимптотический член без специального выбора шага интегрирования. Некоторые общие идеи построения таких схем имеются в [10].

Введенная ранее функция g_1, соответствующие интегралы от которой определяют вклады в макропараметры нейтралов от границ, в k-м приближении представляется в виде

$$g_1(\mathbf{x}, \mathbf{w}) = \pi^{-3/2} e^{-[\mathbf{w} - \tilde{\mathbf{u}}(\mathbf{x}_w^0)]^2 - \tilde{\nu}_0} + \sum_{l=1}^{6} n_w^l(\mathbf{x}_w^l)(B_w/\pi)^{3/2} e^{-B_w \mathbf{w}^2 - \tilde{\nu}_l}, \tag{3.6}$$

где

$$\mathbf{x}_w = \mathbf{x} - wt_b^l, \quad \tilde{\nu}_l = \int_0^{t_b^l} \nu_2^{k-1}(\mathbf{x} - \mathbf{w}s)ds, \quad l = 0, \ldots, 6.$$

Первое слагаемое в (3.6) есть член, отвечающий за влияние выходного отверстия. Эти части g_1 определяются с помощью процедуры, построенной ранее для ионов. В этом случае $\mathbf{x}^* = \mathbf{x}$, $A_z = 0$, $\bar{J} = 1$, что существенно сокращает время, необходимое для работы этого блока. Отметим, что, в отличие от ионной компоненты, вклад в макропараметры нейтралов от этого члена сравнительно мал. Это связано с тем, что функция распределения нейтралов, выходящих из отверстия, не является дельтообразной, а площадь самого отверстия мала по сравнению с площадью грани СПД, с которой осуществляется выпуск струи.

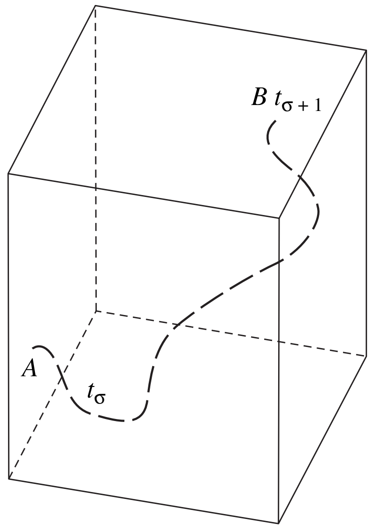
Второе слагаемое в (3.6) определяет вклады твердых поверхностей в функцию распределения нейтралов. В трехмерной постановке задачи о струе этими поверхностями будут шесть граней самого плазменного двигателя.

Численная схема для определения вкладов в макропараметры от какого-либо слагаемого (3.6) строится так же, как в рассмотренном выше случае, т.е. осуществляется переход к новым переменным интегрирования, в результате чего интегрирование ведется по площади соответствующей грани и одной скоростной переменной. Соответствующие формулы довольно просты и их нетрудно получать, следуя уже описанному методу.

4. ЧИСЛЕННАЯ СХЕМА ДЛЯ НАХОЖДЕНИЯ f_2

Введенная ранее при разбиении f^k функция f_2 подчиняется второму уравнению (2.4). Решение этого уравнения при нулевых граничных условиях записывается в виде

$$f(\mathbf{x}, \boldsymbol{\xi}) = \int_0^{t_b} F^+(\tilde{\mathbf{x}}(\tau), \boldsymbol{\xi}(\tau)) \exp\left(-\int_0^{\tau} \nu_1(\tilde{\mathbf{x}}(s))ds\right) d\tau, \tag{4.1}$$



Фиг. 4.

где через f обозначено f_2. Вычисление f согласно (4.1) осуществляется следующей процедурой:

$$f(\mathbf{x}^k, \boldsymbol{\xi}^i) = \sum_{\sigma=1}^{\sigma_0} D_\sigma, \quad D_\sigma = \exp\left(-\int_0^{t_\sigma} \nu_1(s)ds\right) \int_{tt_\sigma}^{t_{\sigma+1}} F^+(\tau)\exp\left(-\int_{t_\sigma}^{\tau} \nu_1(s)ds\right)d\tau. \tag{4.2}$$

Здесь $F^+(\tau)$, $\nu_1(s)$ есть $F^+(\tilde{\mathbf{x}}(\tau), \tilde{\boldsymbol{\xi}}(\tau))$, $\nu_1(\tilde{\mathbf{x}}(s))$, $\mathbf{x}^k = \{x_k, y_l, z_m\}$, $\boldsymbol{\xi}^i = \{\xi_{xi}, \xi_{yj}, \xi_{zn}\}$ суть узлы фазового пространства. Они должны быть заданы. При $\tau = t_\sigma$ характеристика входит в показанную на фиг. 4 счетную ячейку (t_σ известна), а при $\tau = t_{\sigma+1}$ покидает ее (это значение τ должно быть найдено). Электрическое поле считалось постоянным внутри ячейки и определялось по ее центру. Это делалось при помощи численного дифференцирования вычисленного в узлах физического пространства потенциала электрического поля с последующим усреднением. Это предположение позволяет внутри ячейки записать выражения для $\tilde{\mathbf{x}}(\tau)$ и $\tilde{\boldsymbol{\xi}}(\tau)$ в виде

$$\tilde{\mathbf{x}}(\tau) = \mathbf{x}_1 - \boldsymbol{\xi}_1(\tau - t_\sigma) + \bar{\mathbf{E}}(\tau - t_\sigma)^2/2, \quad \tilde{\boldsymbol{\xi}}(\tau) = \boldsymbol{\xi}_1 - \bar{\mathbf{E}}(\tau - t_\sigma), \tag{4.3}$$

где $\mathbf{x}_1 = \tilde{\mathbf{x}}(t_\sigma)$, $\boldsymbol{\xi}_1 = \tilde{\boldsymbol{\xi}}(t_\sigma)$, $\bar{\mathbf{E}}$ – значение напряженности электрического поля внутри ячейки. Обозначим через $\mathbf{x}_b = \{x_b, y_b, z\}$ грани счетной ячейки, на которые могла бы прийти характеристика, если бы движение осуществлялось с постоянной скоростью $\boldsymbol{\xi}_1$. Полагая в (4.3) $\tilde{\mathbf{x}}(\tau) = \mathbf{x}_b$, находим значение $\Delta t_\gamma = t_\gamma - t_\sigma$, $\gamma = 1, 2, 3$, при котором ион, двигаясь в постоянном электрическом поле $\mathbf{E}$, достигает соответствующей грани x_b^γ ($x_1^b = x_b$, $x_2^b = y_b$, $x_3^b = z_b$):

$$\Delta t_\gamma = 2(x_b^\gamma - x_1^\gamma)/(\xi_1^\gamma + \operatorname{sign}(\xi_1^\gamma)\sqrt{D}), \quad D = (\xi_1^\gamma)^2 - 2(x_b^\gamma - x_1^\gamma). \tag{4.4}$$

Положительность D в (4.4) является критерием того, что линия AB доходит до x_b^γ. Если $D < 0$, то внутри счетной ячейки существует точка, в которой соответствующая компонента $\tilde{\boldsymbol{\xi}}$ меняет свой знак. При этом характеристика уже не достигает грани x_b^γ, а поворачивает к грани $x_h^\gamma = x_b^\gamma +$ $+ \operatorname{sign}(\xi_1^\gamma)\Delta x_\gamma$ (Δx_γ – шаг разбиения соответствующей координаты физического пространства). В этом случае Δt_γ определяется из условия попадания иона после поворота на грань x_h^γ. Оно равно

$$\Delta t_\gamma = \xi_1^\gamma/\bar{E}_\gamma + \sqrt{2(x_h^\gamma - x_*^\gamma)/\bar{E}_\gamma}, \quad x_*^\gamma = x_1^\gamma - (\xi_1^\gamma)^2/(2\bar{E}_\gamma).$$

Значение $t_{\sigma+1}$ определяется по формуле $t_{\sigma+1} = \Delta t + t_\sigma$, где $\Delta t = \min\{\Delta t_1, \Delta t_2, \Delta t_3\}$.

После определения $t_{\sigma+1}$ можно вычислить D_σ. Это делается с помощью модификации схемы с экспонентами. В нашем случае имеем

$$D_\sigma = \exp\left(-\int_0^{t_\sigma} \nu_1(q)dq\right)(F_\sigma^+(1-e^{-\Delta s}) + (F_{\sigma+1}^+ - F_\sigma^+)(-e^{-\Delta s} + (1-e^{-\Delta s})/\Delta s), \tag{4.5}$$

где

$$\Delta s = \int_{t_\sigma}^{t_{\sigma+1}} \nu_1(q)dq \cong (\nu_{1\sigma} + \nu_{1\sigma+1})\Delta t/2, \quad F_l^+(\tilde{\mathbf{x}}(t_l), \tilde{\boldsymbol{\xi}}(t_l)), \quad l = \sigma, \sigma+1.$$

Напомним, что эта схема обладает вторым порядком точности по Δs. Если $\Delta s \ll 1$, то (4.5) переходит в обычную формулу трапеции. При $\Delta s > 1$ схема (4.5) дает правильную асимптотику до членов $(\Delta s)^{-2}$. В отличие от [6], в нашем случае $\tilde{\boldsymbol{\xi}}$ меняется вдоль характеристики. Если электрическое поле потенциально, то $|\tilde{\boldsymbol{\xi}}|$ меняется по ячейке так же, как и макропараметры, поэтому использование схемы (4.5) является естественным.

Отличием задач с силовым полем от задач разреженного газа является тот факт, что траектория частицы может никогда не выходить на границу счетной области. В этом случае принималось, что $t_b = \infty$, причем интегралы в (4.1) будут сходиться, что и использовалось в расчетах.

В настоящей работе, так же как и в [6], [7], расчет осуществлялся без запоминания функции распределения (только недавно появились ЭВМ, позволяющие запоминать шестимерные массивы от функций распределения). Вклады в макропараметры от f_2 осуществлялись при помощи квадратурных формул численного интегрирования вида

$$\begin{pmatrix} n \\ n\mathbf{u} \\ n(3/(2T) + \mathbf{u}^2) \end{pmatrix}_n^k = \sum_{m=1}^{m_0}\sum_{n=1}^{n_0}\sum_{\sigma=1}^{\sigma_0} A_{mn\sigma}\begin{pmatrix} 1 \\ \mathbf{w}_{mn\sigma} \\ \mathbf{w}_{mn\sigma}^2 \end{pmatrix} g(\mathbf{x}_{ijl}, \mathbf{w}_{mn\sigma}), \tag{4.6}$$

где $A_{nm\sigma}$ – коэффициенты квадратурной формулы, $m_0 \times n_0 \times \sigma_0$ – число узлов скоростного пространства. Зафиксировав $\mathbf{x}_{ijl}$, вычислим f для каждого $\boldsymbol{\xi}_{mn\sigma}$ при помощи (4.4), (4.5), сделав при этом вклад в правую часть (4.6), после чего перейдем к вычислению f_2 в следующей точке физического пространства. Такая процедура вычисления не требует запоминания функции распределения в узлах шестимерного фазового пространства. В (4.1) имеем $F^+ \sim (B_2/B_1)^{3/2}\exp[-B_2/T_n^{k-1}(\boldsymbol{\xi} - \mathbf{u}_n^k/\sqrt{B_2})^2]$. При вычислении вкладов в макропараметры ионов все интегралы надо умножать на $B_1^{3/2}$, а $B_2 \cong 1000$, так что под интегралом в (4.1) стоит почти дельта-функция, поэтому при вычислении интегралов по скоростному пространству осуществлялся переход к следующей переменной $\boldsymbol{\xi} = (\mathbf{u}_n^{k-1} + \mathbf{c})B_2^{-1/2}$. Тогда вклад в плотность ионов, например, будет определяться следующим образом:

$$n_2^k = B_1^{3/2}\int f_2(\mathbf{x}, \boldsymbol{\xi})d\boldsymbol{\xi} = (B_1/B_2)^{3/2}\int f_2(\mathbf{x}, (\mathbf{u}_n^{k-1} + \mathbf{c})B_2^{-1/2})d\mathbf{c} \approx \sum_{k=1}^{6} C_k e^{y_k^2}\bar{f}_2(\mathbf{x}, \boldsymbol{\xi}_k),$$

$$\boldsymbol{\xi}_k = (\mathbf{u}_n^{k-1} + \mathbf{y}_k)B_2^{-1/2}, \quad \bar{f}_2 = (B_1/B_2)^{3/2}f_2,$$

где y_k – узел, а C_k – коэффициент соответствующей формулы Гаусса ($k_0 = 6$). Построенная процедура осуществляет вычисление $\bar{f}_2$, величина которой порядка единицы. Аналогично вычислялись вклады в макропараметры от g_2 и g_3. Как известно, квадратурные формулы Гаусса чувствительны к гладкости подынтегральной функции (см. [10]). В описанном способе вычисления f_2, g_2, g_3 будут непрерывны в пространстве скоростей, поэтому применение формул Гаусса является оправданным.

5. РЕЗУЛЬТАТЫ ВЫЧИСЛЕНИЙ

Проведенное в [1]–[3] в осесимметричной постановке моделирование струйного движения, возникающего от работы СПД, выявило наличие обратных ионных токов, т.е. токов, направленных на срез двигателя (грань CC_1D_1C на фиг. 1). В рамках проведенного в названных работах моделирования все расчеты велись вниз по потоку, так что сказать что-либо конкретное о масштабе влияния струи на области вверх по потоку, кроме констатации его наличия, было нельзя. Так как теперь расчеты проводятся как вниз, так и вверх по струе, то можно ответить на вопросы, касающиеся влияния ионных токов на названные выше области.

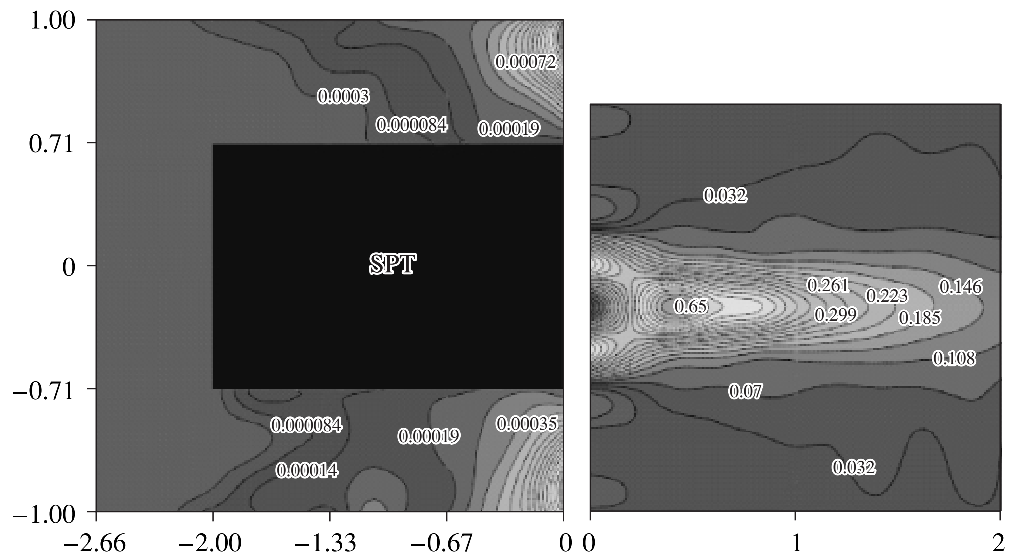
Первоначально была найдена тяга с учетом ионов, попадающих на корпус двигателя. Сила воздействия потока на аппарат равна

$$\mathbf{F} = -\oint\limits_S\int m_i \int \boldsymbol{\xi}\xi_n f d\boldsymbol{\xi} d\sigma. \tag{5.1}$$

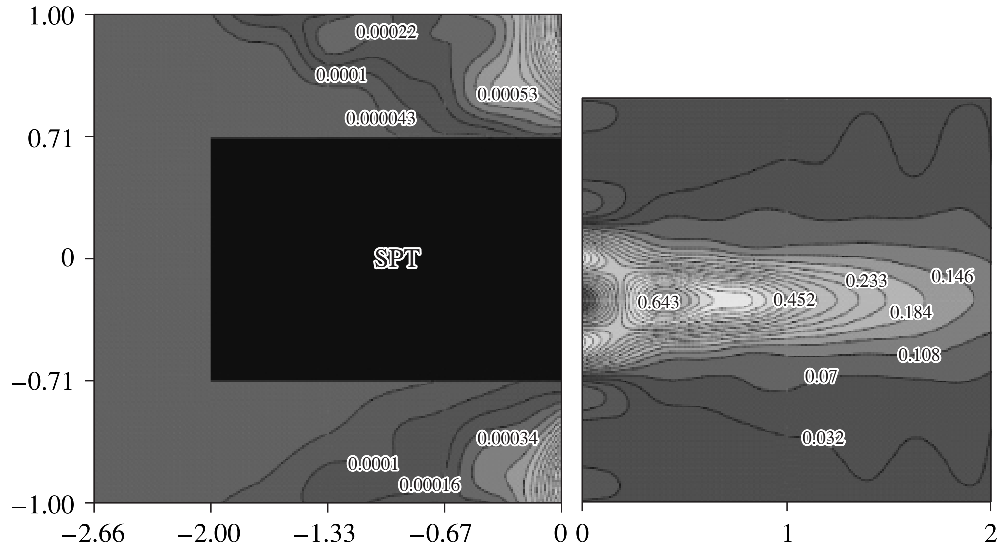
Вообще говоря, в правой части (5.1) нужно было бы учесть импульс, передаваемый телу нейтральной компонентой, но он существенно меньше импульса, который передают телу ионы. Расчеты величины F_z показали, что это есть тяга, основная часть которой создается потоком ионов, выходящим из отверстия. Вклады остальных граней не превышают 0.3% от этого значения и все дают тягу, кроме грани BB_1C_1C, которая дает сопротивление.

На фиг. 5 и фиг. 6 представлены линии пересечения поверхностей уровня плотности ионов с плоскостями $y = x$ и $y = -x$ соответственно при $\bar{u}_\varphi = 0$. На правых рисунках изображены линии для области $z \geq 0$, а на левых – для области $-l \leq z \leq 0$, где l – длина корпуса двигателя. Сам двигатель показан черным цветом. Из-за симметрии геометрии двигателя распределение линий уровня в этих плоскостях должны быть идентичны, поэтому имеющиеся отличия дают представление о точности применяемого численного метода. Как видно из сравнения рисунков, линии уровня со значениями, большими или равными одной тысячной, воспроизводятся в численном счете достаточно хорошо. На фиг. 7 представлены линии уровня в плоскости XOZ ($y = 0$). Сравнение с изображениями на фиг. 5 и фиг. 6 показывает, как влияет неосесимметричность постановки задачи на картину течения. Видно, что вниз по потоку струя будет практически осесимметричной. Для $z < 0$ (в области вверх по потоку) имеется небольшое отличие в распределении линий уровня плотности ионов, обусловленное неосесимметричностью постановки задачи.

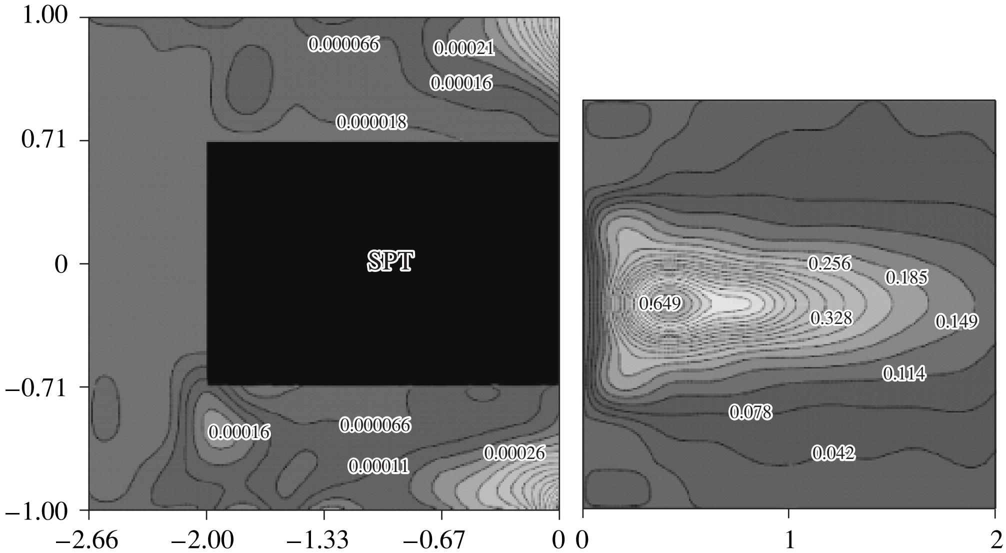
Анализ движения плазмы в канале СПД показывает, что выходящая из двигателя струя должна быть закрученной. Оценки показывают, что величина $\bar{u}_\varphi$ составляет ~5–7% от скорости $\bar{u}_z$. Возможность проведения расчетов в произвольной геометрии позволяет определить влияние эффектов закрученности на распределение параметров в струе.



Фиг. 5.



Фиг. 6.

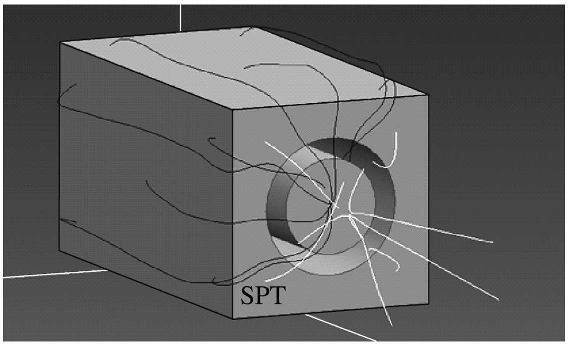


Фиг. 7.

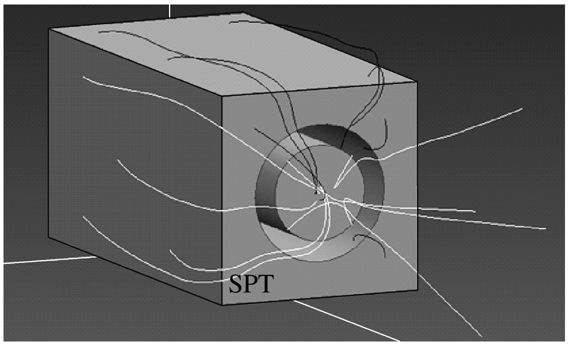
Были проведены расчеты, в которых фигурирующая в граничном условии (1.4) $\bar{u}_\varphi$ полагалась равной $\bar{u}_\varphi = 0.1(1 + \cos\varphi)$.

На фиг. 8, 9 представлены линии ионного тока для случаев $\bar{u}_\varphi = 0$ и $\bar{u}_\varphi \neq 0$ соответственно, т.е. линии, являющиеся интегральными кривыми системы $d\mathbf{x}/dt = q\mathbf{u}^i$, $\mathbf{x}(0) = \mathbf{x}_w$, $\mathbf{x}_w \in ABCDA_1B_1C_1D_1$, причем $q = 1$ (светлые линии), если $u_n^i(\mathbf{x}_w) \geq 0$, в противном случае $q = -1$ (темные линии). В целом картина линий тока с $\bar{u}_\varphi = 0$ и с $\bar{u}_\varphi \neq 0$ одинаковая. Отличие состоит в том, что при $\bar{u}_\varphi = 0$ на грани $ABCD$ и на BB_1C_1C приходят обратные токи, тогда как для струи с закруткой обратные токи приходят только на грань BB_1C_1C, а с грани $ABCD$ линии тока сходят и заканчиваются на торцевой поверхности, где расположено выходное отверстие.

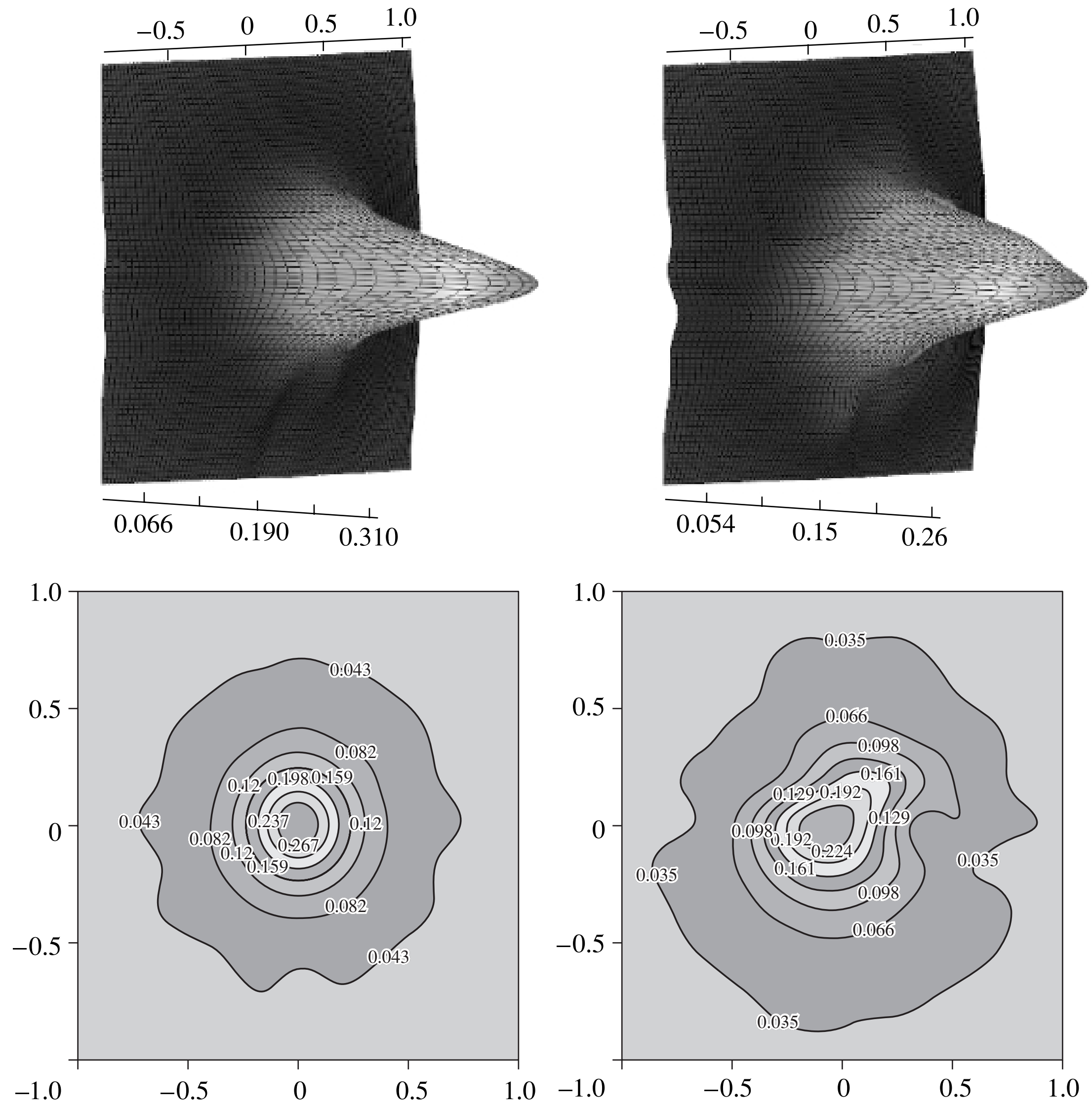
На фиг. 10–12 представлено: на верхних рисунках – графики функций $n^i = n^i(x, y, z_0)$, а на нижних – соответствующие им линии уровня этих функций в плоскости $z = z_0$. Слева – для $\bar{u}_\varphi = 0$, а



Фиг. 8.

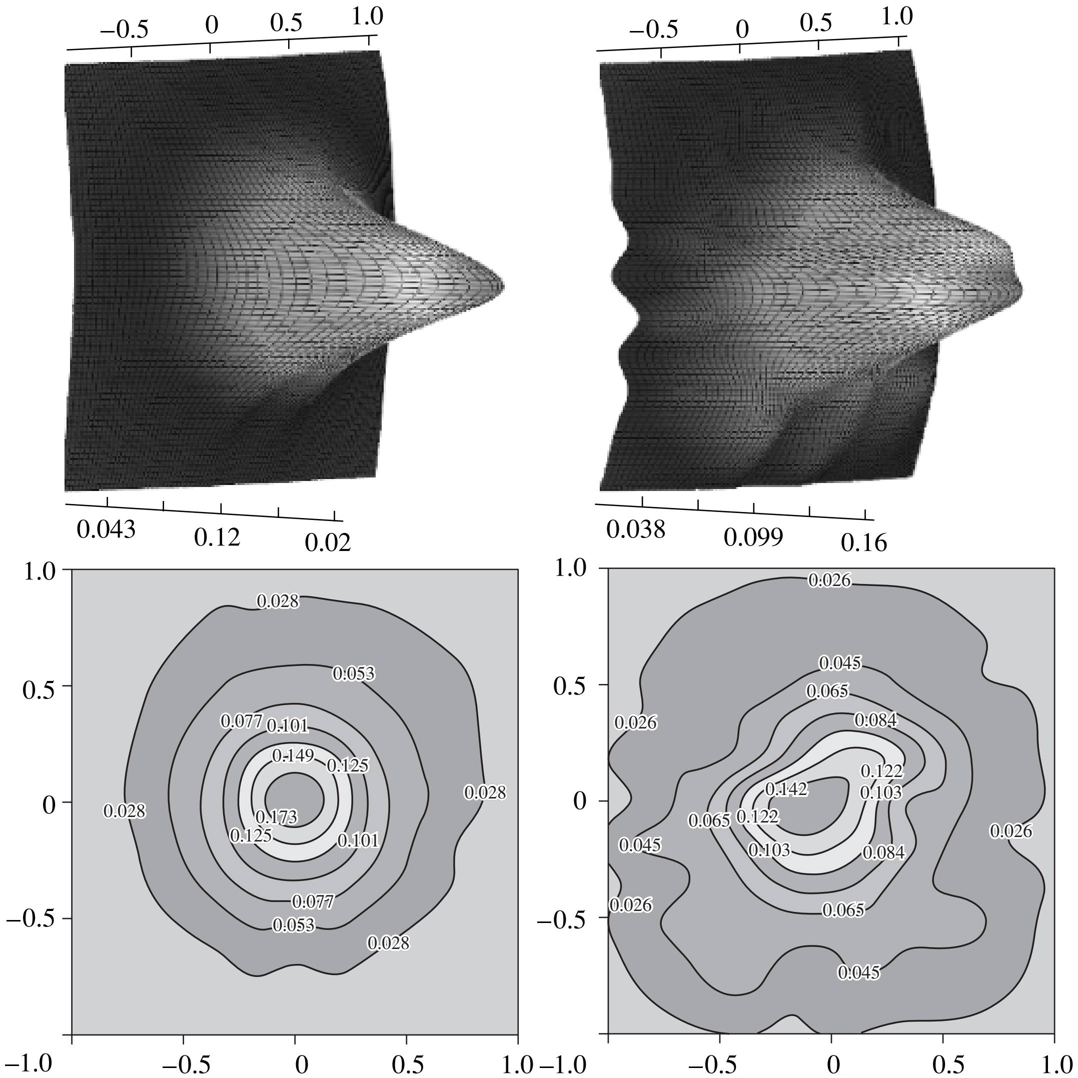


Фиг. 9.



Фиг. 10.

справа – для $\bar{u}_\varphi \neq 0$. При $z_0 = 1$ (фиг. 10) видно, что закрученность струи сказывается на положении максимума плотности и его величине. Если для $\bar{u}_\varphi = 0$ этот максимум расположен на оси z, т.е.

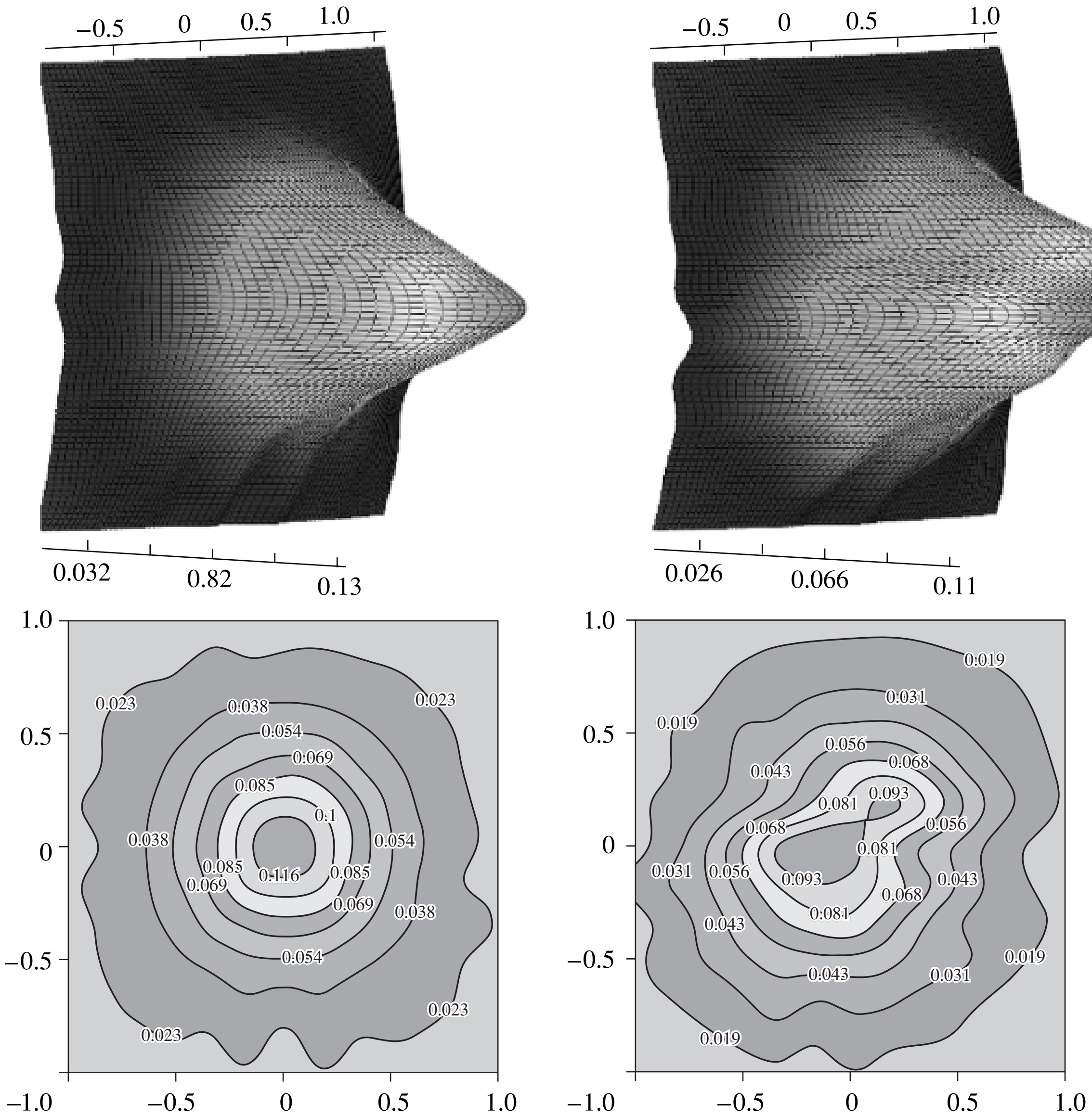


Фиг. 11.

есть "кроссовер", то, когда $\bar{u}_\varphi \neq 0$, максимум заметно меньше, а если судить по линиям уровня – соответствует отверстию.

При $z_0 = 1.5$ (см. фиг. 11) наблюдается уже заметное отличие в поведении плотности ионов. При $\bar{u}_\varphi \neq 0$ начинают образовываться два пика в плотности ионов. При $z_0 = 2$ (см. фиг. 12) эти пики окончательно расходятся и отличие в поведении плотности ионов в струе с закруткой и без нее проявляется наиболее отчетливо. При этом видно, что кроссовера в этом случае нет, так как пики соответствуют влиянию отверстия. Таким образом, показано, что закрученность струи может являться причиной того, что иногда в экспериментах кроссовер не наблюдается. Остается неясной зависимость $\bar{u}_\varphi$ от φ. Возможно, что двугорбость в распределении плотности вызвана спецификой заданной зависимости $\bar{u}_\varphi$ от азимутального угла. Здесь последнее слово должно принадлежать экспериментаторам.

В заключение надо отметить, что возможность запоминать многомерные массивы от функции распределения создала возможность для интенсивного развития нестационарных численных методов для решения задач разреженных газов и кинетической теории. Поэтому следующий шаг в развитии направления, связанного с решением проблемы моделирования течений, возникающих от работы электрических реактивных двигателей, должен быть направлен на создание нестационарных методов решения задачи (1.1)–(1.6). Основная трудность, которую необходимо



Фиг. 12.

преодолеть, это создание нестационарного численного метода, учитывающего дельтообразность граничной функции распределения ионов и большую разницу в масштабах скоростных пространств у ионов и нейтралов. Стандартные варианты метода частиц здесь не годятся.

СПИСОК ЛИТЕРАТУРЫ

1. *Бишаев А.М.* Численное моделирование струи разреженного слабо ионизованного газа, выходящего из кольцевого отверстия // Ж. вычисл. матем. и матем. физ. 1993. Т. 33. № 7. С. 1109–1118.

2. *Бишаев А.М., Калашников В.К., Ким В.* Численное исследование струи разреженной плазмы стационарного ускорителя с замкнутым дрейфом электронов (НЗДП) // Физ. плазмы. 1992. Т. 18. Вып. 6. С. 698–708.

3. *Бишаев А.М., Калашников В.К., Ким В., Шавыкина А.В.* Численное моделирование плазменной струи стационарного плазменного двигателя, распространяющейся в среде низкого давления // Физ. плазмы. 1998. Т. 24. № 11. С. 989–995.

4. *Волков Б.И., Морозов А.И., Свешников А.Г., Якунин С.А.* Численное моделирование ионов в системе с замкнутым дрейфом // Физ. плазмы. 1981. Т. 7. Вып. 2. С. 245–253.
5. *Жевандров П.И., Морозов А.И., Якунин С.А.* Динамика плазмы, образующейся при ионизации разреженного газа // Физ. плазмы. 1984. Т. 10. Вып. 2. С. 353–365.
6. *Шахов Е.М.* Метод исследования движений разреженного газа. М.: Наука, 1974.
7. *Ларина И.Н.* Сопротивление сферы в сильно разреженном газе // Числ. методы в динамике разреженных газов. М.: ВЦ АН СССР, 1973. С. 22–30.
8. *Жук В.И.* Решение кинетического уравнения для газа в поле тяготения планеты // Докл. АН СССР. 1977. Т. 233. № 3. С. 325–328.
9. *Самарский А.А., Николаев Е.С.* Методы решения сеточных уравнений. М.: Физматгиз, 1978.
10. *Никифоров А.Ф., Уваров В.Б.* Специальные функции математической физики. М.: Физматгиз, 1978.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 506–529

УДК 519.634

МЕТОД РЕШЕНИЯ УРАВНЕНИЙ ТЕОРИИ ВЗАИМОДЕЙСТВИЯ ПРОСТРАНСТВЕННОГО ПОГРАНИЧНОГО СЛОЯ И ВНЕШНЕГО НЕВЯЗКОГО ПОТОКА[1)]

© 2007 г. Г. Л. Королев

(140180 Жуковский, М.о., ул. Жуковского, 1, ФГУП ЦАГИ)

e-mail: glk777@mail.ru

Поступила в редакцию 29.03.2006 г.

Разработан новый подход к решению пространственных стационарных уравнений теории взаимодействия ламинарного пограничного слоя с внешним невязким потоком. Метод с успехом может использоваться также для исследования взаимодействия плоских течений. На основе этого метода впервые решена задача пространственного обтекания неровности (бугорок и впадина) вязким сверхзвуковым потоком газа в режиме классического взаимодействия, определены асимптотические значения параметра высоты неровности, соответствующие безотрывному обтеканию потока, построены картины отрывного обтекания. Библ. 27. Фиг. 9.

Ключевые слова: численные методы, сверхзвуковое течение, пространственный пограничный слой, взаимодействие, отрывное обтекание.

1. ВВЕДЕНИЕ

Численное исследование пространственных течений в области взаимодействия пограничного слоя с внешним невязким потоком является одной из сложнейших проблем вычислительной аэрогидродинамики. Исследования такого типа находятся еще в стадии развития. Исторически разработка численных методов решения уравнений теории взаимодействия пограничного слоя с внешним невязким потоком начиналась с разработки методов решения плоских двумерных течений со сверхзвуковым типом взаимодействия (см. [1], [2]). Уже позже, в продолжение развития этих методов, они разрабатывались для решения уравнений плоских течений с дозвуковым, гиперзвуковым и трансзвуковым типом взаимодействия. Подробно об этих методах см. [3]–[7]. Что касается пространственных трехмерных течений, то ситуация на сегодняшний день складывается с точностью до наоборот. Несмотря на то что постановка задачи для области классического пространственного сверхзвукового взаимодействия, в которой наклоны линий тока, вызванные обтекаемым телом и толщиной вытеснения вязкого подслоя, являются величинами одного порядка во внешнем невязком потоке, была сформулирована 25 лет назад (см. [8], [9]), до сих пор не было получено решения в простейшей линейной постановке. Хотя уже разработаны методы и решен ряд задач для дозвукового (см. [10]) и сильного гиперзвукового (см. [11]) типов взаимодействия, для компенсационного режима течения (см. [12])[2)]. По-видимому, это связано со сложным видом условия взаимодействия для пространственного сверхзвукового течения. В частности, решение линейной задачи для течений несжимаемой жидкости или слабого гиперзвукового взаимодействия (в отличие от сверхзвукового пространственного течения) может быть выписано через интегралы Фурье в аналитическом виде. Данная работа ограничивается рассмотрением задач теории взаимодействия стационарных уравнений движения. Постановка задач для нестационарных уравнений теории взаимодействия, а также вопросы исследования гидродинамической устойчивости этих решений подробно представлены в [13].

Наиболее полно разработаны численные методы и исследованы течения несжимаемой жидкости. В [14] рассмотрена проблема трехмерного взаимодействия течений несжимаемой жидкости в трубах. Более того, в этой же работе указана еще одна фундаментальная проблема, возни-

[1)] Работа выполнена при финансовой поддержке РФФИ (код проекта 04-01-00765) и Государственной поддержке ведущих научных школ (номер гранта НШ-2001.2003.1).

[2)] Постановка задачи о слабом пространственном гиперзвуковом взаимодействии была доложена на EUROMECH 384 Colloquium on Steady and Unsteady Separated Flows в Манчестере в июле 1998 г. Боссом, Кравцовой, Рубаном. Также там были представлены некоторые решения линейной и нелинейной задачи, но результаты этой работы опубликованы не были.

кающая при решении уравнений пространственного пограничного слоя обратным методом, т.е. решение уравнений пространственного пограничного слоя по заданной величине толщины вытеснения. Анализ этих уравнений указывает на их эллиптический характер, следовательно, приводит к отклонению от решения в результате использования маршевого метода при решении уравнений Прандтля.

В [15] развит обратный конечно-разностный метод решения уравнений пограничного слоя, который использовался для анализа течения несжимаемой жидкости около трехмерного крыла.

Один из наиболее надежных и быстрых численных методов для расчета трехмерных течений является спектральный метод, разработанный в [10] для расчета трехмерного до звукового течения около малых препятствий. Спектральный метод является эффективным для исследования течений с возвратными линиями тока. Однако применение этого метода ограничивается обычными условиями сходимости преобразований Фурье, так что формы тела с разрывами наклона поверхности являются более сложными для исследования, при этом проявляются осцилляции как результат решения.

В [16] применялся квазиодновременный метод для вычисления трехмерного отрывного течения в пограничном слое. В [17] этот метод использовался для анализа течения около трехмерного крыла. В [18] рассмотрено периодическое трехмерное течение вокруг лопатки вертолета в дозвуковом потоке с использованием полунеявного метода предиктор-корректор с локальной линеаризацией и маршевым интегрированием уравнений пограничного слоя в периодическом направлении.

Цель данной работы – представить быстрый и надежный численный метод, позволяющий рассчитывать пространственные течения жидкости в области взаимодействия ламинарного пограничного слоя со сверхзвуковым невязким потоком, а также течения с другими типами взаимодействия, включая режимы отрывного течения.

2. ПОСТАНОВКА ЗАДАЧИ

Несмотря на множество физических ситуаций, которые приводят к взаимодействию ламинарного пограничного слоя с внешним невязким потоком, формулировка математической постановки задачи в безразмерных переменных является в определенной степени универсальной (см. [8], [19], [20]). Для описания течения вблизи пристеночного слоя используются уравнения Прандтля:

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = -\frac{\partial p}{\partial x} + \frac{\partial^2 u}{\partial y^2}, \tag{2.1}$$

$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} = -\frac{\partial p}{\partial z} + \frac{\partial^2 w}{\partial y^2}, \tag{2.2}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{2.3}$$

Условия прилипания на поверхности обтекаемого тела записываются в форме

$$u = v = w = 0 \text{ или } y = 0, \tag{2.4}$$

где в используемой ортогональной системе координат x – расстояние, измеряемое вдоль поверхности тела, y – расстояние, нормальное к нему, z – расстояние в поперечном направлении. Форма других граничных условий зависит от конкретной модели взаимодействия. Если используется классическая модель взаимодействия пограничного слоя с внешним невязким потоком, когда величина поверхностного трения на поверхности тела является величиной конечной вверх по течению от области взаимодействия, тогда уравнения (2.1)–(2.3) должны удовлетворять условиям сращивания с решением вверх по течению и в поперечном направлении вдоль оси z в виде

$$u = y, \quad w = 0 \text{ при } x \longrightarrow -\infty, \tag{2.5}$$

$$u = y, \quad w = 0 \text{ при } z \longrightarrow \pm\infty \tag{2.6}$$

и условию сращивания с решением для основной части пограничного слоя

$$u = y - \delta(x, z) + \ldots \text{ при } y \longrightarrow \infty, \tag{2.7}$$

$$w = O(y^{-1}), \tag{2.8}$$

где $\delta(x, z)$ – приращение толщины вытеснения пограничного слоя, обусловленной влиянием области взаимодействия относительно его величины в невозмущенном набегающем потоке. Если функция $\delta(x, z)$ известна и известна форма обтекаемой поверхности $f(x, z)$, то можно найти распределение давления $p(x, z)$ вдоль области взаимодействия, используя соответствующий закон взаимодействия. В сверхзвуковом течении он дается формулой (см. [21])

$$p = \frac{1}{\pi}\int\limits_{-\infty}^{x}\int\limits_{z' = z-(x'-x)}^{z' = z+(x'-x)} \frac{\delta''_{xx}(x', z') + f''_{xx}(x', z')}{\sqrt{(x-x')^2-(z-z')^2}}dx'dz', \tag{2.9}$$

в то время как в дозвуковом потоке интеграл Гильберта из теории линейных потенциальных течений дает соотношение p и δ (см. [19]) в виде

$$p = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \frac{\delta''_{x'x'} + f''_{x'x'}}{\sqrt{(x-x')^2+(z-z')^2}}dx'dz', \tag{2.10}$$

где $f(x, z)$ – форма обтекаемой поверхности.

Некоторые формы трансзвукового взаимодействия, в которых давление и толщина вытеснения связаны в явном виде, можно найти в [22]. В общем случае для трансзвукового течения необходимо одновременно решать нелинейные трехмерные уравнения невязкого трансзвукового течения с уравнениями пространственного пограничного слоя.

В гиперзвуковом течении закон слабого взаимодействия может быть представлен в форме (см. [23])

$$p = \frac{\partial\delta}{\partial x} + \frac{\partial f}{\partial x}. \tag{2.11}$$

Формула (2.11) также описывает в плоском двумерном течении закон сверхзвукового взаимодействия.

Для компенсационного режима обтекания неровности закон взаимодействия выглядит как отсутствие изменения суммарной величины толщины вытеснения и формы тела при прохождении над препятствием (см. [12]):

$$\frac{\partial\delta}{\partial x} + \frac{\partial f}{\partial x} = 0. \tag{2.12}$$

Более полный обзор возможных типов взаимодействия пространственного пограничного слоя с внешним невязким дозвуковым и сверхзвуковым потоком и соответствующие этим типам взаимодействия краевые условия см. в библиографиях работ [24]–[26].

В общем виде, включающем представленные выше типы взаимодействия и ряд других, запишем условие взаимодействия в виде

$$Q\left(p(x, z), \frac{\partial\delta(x, z)}{\partial x}, \frac{\partial f(x, z)}{\partial x}\right) = 0. \tag{2.13}$$

3. СИСТЕМА ВЕКТОРНЫХ УРАВНЕНИЙ

Дифференцируя уравнения (2.1), (2.2) по y и складывая их, получаем

$$u\frac{\partial\omega}{\partial x} + v\frac{\partial\omega}{\partial y} + w\frac{\partial\omega}{\partial z} - \omega\frac{\partial w}{\partial z} + \tau\frac{\partial u}{\partial z} = \frac{\partial^2\omega}{\partial y^2}, \tag{3.1}$$

$$u\frac{\partial\tau}{\partial x} + v\frac{\partial\tau}{\partial y} + w\frac{\partial\tau}{\partial z} + \omega\frac{\partial w}{\partial x} - \tau\frac{\partial u}{\partial x} = \frac{\partial^2\tau}{\partial y^2}, \tag{3.2}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{3.3}$$

$$\frac{\partial u}{\partial y} = \omega, \tag{3.4}$$

$$\frac{\partial w}{\partial y} = \tau. \tag{3.5}$$

Граничные условия на поверхности твердого тела при $y = 0$ могут быть записаны в форме

$$\frac{\partial \omega}{\partial y} = \frac{\partial p}{\partial x}, \quad \frac{\partial \tau}{\partial y} = \frac{\partial p}{\partial z}, \tag{3.6}$$

$$u = 0, \quad v = 0, \quad w = 0. \tag{3.7}$$

Первое и второе условия могут быть выведены из уравнений момента количества движения (2.1), (2.2) при $y = 0$ с использованием условия прилипания.

Рассмотрим теперь постановку краевых условий при больших y. Будем искать решение в форме

$$u = y + A(x, z) + c(x, z)y^{-1} + \ldots, \tag{3.8}$$

$$w = b(x, z)y^{-1} + \ldots. \tag{3.9}$$

Здесь $A = -\delta$. Из уравнения неразрывности (2.3) получаем

$$v = -\frac{\partial A(x, z)}{\partial x}y - \left(\frac{\partial c(x, z)}{\partial x} + \frac{\partial b(x, z)}{\partial z}\right)\ln y + B(x, z) + \ldots. \tag{3.10}$$

Подставляя асимптотическое разложение (3.8)–(3.10) в (2.1), (2.2) и приравнивая члены с одинаковым порядком степени y, получаем следующие уравнения. Из уравнения (2.1) имеем

$$O(\ln y): \frac{\partial c}{\partial x} + \frac{\partial b}{\partial z} = 0, \tag{3.11}$$

$$O(1): \frac{\partial c}{\partial x} + A\frac{\partial A}{\partial x} + B + \frac{\partial p}{\partial x} = 0, \tag{3.12}$$

и из (2.2)

$$O(1): \frac{\partial b}{\partial x} + \frac{\partial p}{\partial z} = 0. \tag{3.13}$$

Чтобы исключить логарифмический рост v при больших y, что следует из (3.11), запишем, используя (3.9) и (2.2), краевые условия для завихренности при больших y в виде

$$\left.\begin{aligned} &\frac{\partial \omega}{\partial x} + \frac{\partial \tau}{\partial z} = 0, \\ &\frac{\partial \tau}{\partial x} - \frac{\partial p}{\partial z}\frac{\omega}{u^2} = 0, \end{aligned}\right. \quad \text{при } y \longrightarrow \infty. \tag{3.14}$$

Два граничных условия при $x \longrightarrow -\infty$

$$\omega = 1, \quad \tau = 0 \text{ при } x = -\infty, \tag{3.15}$$

и

$$\omega = 1, \quad \tau = 0 \text{ при } z = \pm\infty \tag{3.16}$$

непосредственно следуют из (2.5), (2.6).

Остается выразить закон взаимодействия (2.13) через v. Дифференцируя уравнение (3.10) по y и подставляя результат в (2.13), можно прийти к выводу, что закон взаимодействия может быть записан в виде

$$Q\left(p(x, z), \lim_{y \to \infty}\frac{\partial v(x, y, z)}{\partial y}, \frac{\partial f(x, z)}{\partial x}\right) = 0. \tag{3.17}$$

Чтобы разработать общий подход к численному решению трехмерных уравнений пограничного слоя со взаимодействием и отрывом, предположим, что в общем случае проблема может быть записана в векторном виде в следующей форме:

$$\mathbf{A}\frac{\partial\mathbf{\Omega}}{\partial x} + \mathbf{B}\frac{\partial\mathbf{\Omega}}{\partial y} + \mathbf{C}\frac{\partial\mathbf{\Omega}}{\partial z} + \mathbf{D}\frac{\partial^2\mathbf{\Omega}}{\partial y^2} + \mathbf{HP} = 0, \tag{3.18}$$

где вектор $\mathbf{\Omega}$ представляет собой неизвестный компонент векторной функции I_Ω в точке (x, y, z), вектор $\mathbf{P}$ – неизвестный компонент I_p векторной функции градиента давления в точке (x, z).

Граничные условия на поверхности тела имеют вид

$$\mathbf{E\Omega} + \mathbf{G}\frac{\partial\mathbf{\Omega}}{\partial y} + \mathbf{H}_w\mathbf{P} + \mathbf{F} = 0 \text{ при } y = 0, \tag{3.19}$$

и граничные условия сращивания с основным пограничным слоем в виде

$$\mathbf{A}_t\frac{\partial\mathbf{\Omega}}{\partial x} + \mathbf{B}_t\frac{\partial\mathbf{\Omega}}{\partial y} + \mathbf{C}_t\frac{\partial\mathbf{\Omega}}{\partial z} + \mathbf{E}_t\mathbf{\Omega} + \mathbf{H}_t\mathbf{P} + \mathbf{F}_t = 0 \text{ при } y \longrightarrow \infty. \tag{3.20}$$

Условия сращивания с невозмущенным течением впереди области взаимодействия и в поперечном направлении вдоль оси z имеет вид

$$\mathbf{A}_{\text{in}}\frac{\partial\mathbf{\Omega}}{\partial x} + \mathbf{B}_{\text{in}}\frac{\partial\mathbf{\Omega}}{\partial y} + \mathbf{C}_{\text{in}}\frac{\partial\mathbf{\Omega}}{\partial z} + \mathbf{E}_{\text{in}}\mathbf{\Omega} + \mathbf{H}_{\text{in}}\mathbf{P} + \mathbf{F}_{\text{in}} = 0 \text{ при } x \longrightarrow -\infty, \tag{3.21}$$

$$\mathbf{A}_{\text{un}}\frac{\partial\mathbf{\Omega}}{\partial x} + \mathbf{B}_{\text{un}}\frac{\partial\mathbf{\Omega}}{\partial y} + \mathbf{C}_{\text{un}}\frac{\partial\mathbf{\Omega}}{\partial z} + \mathbf{E}_{\text{un}}\mathbf{\Omega} + \mathbf{H}_{\text{un}}\mathbf{P} + \mathbf{F}_{\text{un}} = 0 \text{ при } z \longrightarrow \pm\infty. \tag{3.22}$$

Условие взаимодействия представляется в виде

$$Q(\lim_{y\to\infty}\mathbf{\Omega}, p, f'_x) = 0. \tag{3.23}$$

В частности, для уравнений (3.1)–(3.5) векторы $\mathbf{\Omega}$ и $\mathbf{P}$ могут быть записаны таким образом:

$$\mathbf{\Omega} = \begin{bmatrix} \omega \\ \tau \\ v \\ u \\ w \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} p'_x \\ p'_z \end{bmatrix}, \quad I_\Omega = 5, \quad I_p = 2. \tag{3.24}$$

Тогда матрицы $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ и $\mathbf{D}$, полученные из уравнений (3.1)–(3.5), имеют вид

$$\mathbf{A} = \begin{bmatrix} u & 0 & 0 & \omega & 0 \\ 0 & u & 0 & -\tau & \omega \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} v & 0 & \omega & 0 & 0 \\ 0 & v & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} w & 0 & 0 & \tau & 0 \\ 0 & w & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Матрица $\mathbf{H}$ нулевая; из уравнений (3.6), (3.7) находим

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{H}_w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Матрицы $\mathbf{H}_w$ и $\mathbf{F}$ нулевые; из уравнений (3.14) и (3.3)–(3.5) находим

$$\mathbf{A}_t = \begin{bmatrix} 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&0&1&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{bmatrix}, \quad \mathbf{B}_t = \begin{bmatrix} 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&0&1&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{bmatrix}, \quad \mathbf{C}_t = \begin{bmatrix} 0&1&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&1\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{bmatrix}, \quad \mathbf{E}_t = \begin{bmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1 \end{bmatrix}, \quad \mathbf{H}_t = \begin{bmatrix} 0&0\\ 0&-\omega/u^2\\ 0&0\\ 0&0\\ 0&0 \end{bmatrix}.$$

Матрица $\mathbf{F}_t$ нулевая; из уравнений (3.15), (3.16) получаем

$$\mathbf{E}_{\text{in}} = \mathbf{E}_{\text{un}} = \begin{bmatrix} 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1 \end{bmatrix}, \quad \mathbf{F}_{\text{in}} = \mathbf{F}_{\text{un}} = \begin{bmatrix} -1\\ 0\\ 0\\ -y_k\\ 0 \end{bmatrix}.$$

Матрицы $\mathbf{A}_{\text{in}}$, $\mathbf{B}_{\text{in}}$, $\mathbf{C}_{\text{in}}$, $\mathbf{H}_{\text{in}}$, а также $\mathbf{A}_{\text{un}}$, $\mathbf{B}_{\text{un}}$, $\mathbf{C}_{\text{un}}$, $\mathbf{H}_{\text{un}}$ нулевые.

Введем конечно-разностную сетку

$$(y_k, x_j, z_l), \quad k = 1, 2, \ldots, M, \quad j = 1, 2, \ldots, N, \quad l = 1, 2, \ldots, I,$$

и обозначим величины $\mathbf{\Omega}(y_k, x_j, z_l)$ в узлах этой сетки через $\mathbf{\Omega}_{k,j,l}$, а величины $\mathbf{P}(x_j, z_l)$ вектора $\mathbf{P}$ в узлах конечно-разностной сетки – через $\mathbf{P}_{j,l}$.

Сделаем некоторое общее предположение относительно конечно-разностных схем, используемых при аппроксимации системы векторных уравнений (3.18)–(3.22).

Уравнения (3.18)–(3.22) являются уравнениями параболического типа. Вектор направления параболичности зависит от знаков компоненты векторов скорости u и w. Если знак u или w изменить на противоположный, направление распространения возмущений также меняется. Поэтому, для того чтобы иметь устойчивую конечно-разностную схему, метод аппроксимации нелинейных членов уравнений (3.18) должен зависеть от знака u и w. Обычно для положительных величин u используется метод аппроксимации вверх по потоку, в области отрыва для отрицательных значений u используется метод аппроксимации вниз по потоку. Чтобы получить аппроксимацию первых производных со вторым порядком точности, необходимо использовать два или три значения величин функции в узлах сетки. Однако если используется изменение метода аппроксимации в зависимости от знака скорости, в общем случае имеется пять неизвестных величин функции в точках сетки в каждом направлении при аппроксимации первых производных в уравнениях. Пяти величин достаточно также, чтобы получить второй порядок точности при аппроксимации второй производной функции. В дальнейшем это предположение будет использоваться при построении численного решения.

В этом случае каждой точке k, j, l конечно-разностной сетки ставим в соответствие конечно-разностное векторное уравнение $\mathbf{L}_{k,j,l} = 0$, которое представляет собой векторное представление конечно-разностной аппроксимации уравнений (3.18)–(3.22) вблизи точки (y_k, x_j, z_l).

Внутри сетки $\{2 \le k \le M-1;\ 2 \le j \le N;\ 2 \le l \le I-1\}$, где уравнения пограничного слоя (3.18) аппроксимируются конечно-разностными схемами, вектор $\mathbf{L}_{k,j,l}$ зависит от векторов $\mathbf{\Omega}_{k-2,j-2,l-2}, \ldots, \mathbf{\Omega}_{k,j,l}, \ldots, \mathbf{\Omega}_{k+2,j+2,l+2}$ и векторов $\mathbf{P}_{j-2,l-2}\ldots, \mathbf{P}_{j,l}, \ldots, \mathbf{P}_{j+2,l+2}$:

$$\mathbf{L}_{k,j,l} = \mathbf{L}_{k,j,l}(\mathbf{\Omega}_{k-2,j-2,l-2}, \ldots, \mathbf{\Omega}_{k,j,l}, \ldots, \mathbf{\Omega}_{k+2,j+2,l+2}, \mathbf{P}_{j-2,l-2}, \ldots, \mathbf{P}_{j,l}, \ldots, \mathbf{P}_{j+2,l+2}) = 0, \tag{3.25}$$

$$k = 2, 3, \ldots, M-1, \quad j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1. \tag{3.26}$$

(Здесь и далее если величины индексов k, j, l функций в формулах становятся равными нулю или отрицательными либо больше, чем соответствующие величины N, M, I, то это значит, что здесь нет зависимости функций от величин с такими индексами.)

Выбирая левую граничную сторону расчетной области достаточно далеко вверх по потоку от препятствия, записываем граничные условия (3.21) в форме

$$\mathbf{L}_{k,1,l} = \mathbf{L}_{k,1,l}(\mathbf{\Omega}_{k-2,1,l-2}, \ldots, \mathbf{\Omega}_{k,1,l}, \ldots, \mathbf{\Omega}_{k+2,3,l+2}) = 0, \tag{3.27}$$

$$k = 1, 2, \ldots, M, \quad l = 1, 2, \ldots, I. \tag{3.28}$$

Граничные условия на верхней границе расчетной сетки могут быть представлены в виде

$$\mathbf{L}_{M,j,l} = \mathbf{L}_{M,j,l}(\mathbf{\Omega}_{M-2,j-2,l-2}, \ldots, \mathbf{\Omega}_{M,j,l}, \ldots, \mathbf{\Omega}_{M,j+2,l+2}, \mathbf{P}_{j-2,l-2}, \ldots, \mathbf{P}_{j,l}, \ldots, \mathbf{P}_{j+2,l+2}) = 0, \tag{3.29}$$

$$j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1, \tag{3.30}$$

и, соответственно, на нижней границе расчетной сетки – в виде

$$\mathbf{L}_{1,j,l} = \mathbf{L}_{1,j,l}(\mathbf{\Omega}_{1,j-2,l-2}, \ldots, \mathbf{\Omega}_{1,j,l}, \ldots, \mathbf{\Omega}_{3,j+2,l+2}, \mathbf{P}_{j-2,l-2}, \ldots, \mathbf{P}_{j,l}, \ldots, \mathbf{P}_{j+2,l+2}) = 0, \tag{3.31}$$

$$j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1. \tag{3.32}$$

Условие взаимодействия (3.23) может быть записано в виде

$$Q_{j,l} = Q_{j,l}(\mathbf{\Omega}_{M,1,1}, \ldots, \mathbf{\Omega}_{M,j,l}, \ldots, \mathbf{\Omega}_{M,N,I}, p_{1,1}, \ldots, p_{j,l}, \ldots, p_{N,I}) = 0, \tag{3.33}$$

$$j = 1, 2, \ldots, N, \quad l = 1, 2, \ldots, I, \tag{3.34}$$

$$\mathbf{P}_{j,l} = \mathbf{P}_{j,l}(p_{j-2,l-2}, \ldots, p_{j,l}, \ldots, p_{j+2,l+2}). \tag{3.35}$$

Это завершает построение системы конечно-разностных уравнений.

4. МЕТОД РЕШЕНИЯ

Предположим, что имеется некоторое приближение для векторов $\mathbf{\Omega}^n_{k,j,l}$, $\mathbf{P}^n_{j,l}$ и $p^n_{j,l}$, полученных на предыдущей n-й итерации или взятой в качестве нулевого приближения. Улучшенное приближение

$$\begin{gathered}\mathbf{\Omega}^{n+1}_{k,j,l} = \mathbf{\Omega}^n_{k,j,l} + \Delta\mathbf{\Omega}_{k,j,l},\\ \mathbf{P}^{n+1}_{j,l} = \mathbf{P}^n_{j,l} + \Delta\mathbf{P}_{j,l},\\ p^{n+1}_{j,l} = p^n_{j,l} + \Delta p_{j,l}\end{gathered} \tag{4.1}$$

может быть найдено подстановкой (4.1) в уравнения (3.25)–(3.33) при использовании разложения Тейлора. В результате получается следующая система линейных уравнений для векторов поправки $\Delta\mathbf{\Omega}_{k,j,l}$ и $\Delta\mathbf{P}_{j,l}$:

$$\sum_{s=j-2}^{j+2}\sum_{t=l-2}^{l+2}\left(\sum_{r=k-2}^{k+2}\frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t} + \frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{k,j,l}, \tag{4.2}$$
$$k = 2, 3, \ldots, M-1, \quad j = 2, 3, \ldots, N-2, \quad l = 2, 3, \ldots, I-1,$$

во внутренних точках вычислительной области. На левой стороне, верхней и нижней границах имеем

$$\sum_{s=1}^{3}\sum_{t=l-2}^{l+2}\left(\sum_{r=k-2}^{k+2}\frac{\partial\mathbf{L}_{k,1,l}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t} + \frac{\partial\mathbf{L}_{k,1,l}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{k,1,l}, \tag{4.3}$$
$$k = 2, 3, \ldots, M-1, \quad l = 2, 3, \ldots, I-1,$$

$$\sum_{s=j-2}^{j+2}\sum_{t=l-2}^{l+2}\left(\sum_{r=M-2}^{M}\frac{\partial\mathbf{L}_{M,j,l}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t} + \frac{\partial\mathbf{L}_{M,j,l}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{M,j,l}, \tag{4.4}$$
$$j = 1, 2, \ldots, N, \quad l = 1, 2, \ldots, I,$$

$$\sum_{s=j-2}^{j+2}\sum_{t=l-2}^{l+2}\left(\sum_{r=1}^{3}\frac{\partial\mathbf{L}_{1,j,l}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t} + \frac{\partial\mathbf{L}_{1,j,l}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{1,j,l}, \tag{4.5}$$
$$j = 1, 2, \ldots, N, \quad l = 1, 2, \ldots, I.$$

На передней и задней границах также имеем

$$\sum_{s=j-2}^{j+2}\sum_{t=1}^{3}\left(\sum_{r=k-2}^{k+2}\frac{\partial\mathbf{L}_{k,j,1}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t}+\frac{\partial\mathbf{L}_{k,j,1}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{k,j,1}, \tag{4.6}$$

$$k = 2, \ldots, M-1, \quad j = 2, \ldots, N-2,$$

$$\sum_{s=j-2}^{j+2}\sum_{t=I-2}^{I}\left(\sum_{r=k-2}^{k+2}\frac{\partial\mathbf{L}_{k,j,I}}{\partial\mathbf{\Omega}_{r,s,t}}\Delta\mathbf{\Omega}_{r,s,t}+\frac{\partial\mathbf{L}_{k,j,I}}{\partial\mathbf{P}_{s,t}}\Delta\mathbf{P}_{s,t}\right) = -\mathbf{L}_{k,j,I}, \tag{4.7}$$

$$k = 2, 3, \ldots, M-1, \quad j = 2, 3, \ldots, N-2. \tag{4.8}$$

Чтобы найти приближенное решение этой системы линейных уравнений, используем метод матричной прогонки, основанный на следующих рекуррентных формулах:

$$\begin{gathered}\Delta\mathbf{\Omega}_{k,j,l} = \mathcal{R}_{k,j,l}\Delta\mathbf{\Omega}_{k+1,j,l}+\mathcal{T}_{k,j,l}\Delta\mathbf{\Omega}_{k+2,j,l}+\mathcal{R}'_{k,j,l}\Delta\mathbf{\Omega}_{k,j+1,l}+\mathcal{T}'_{k,j,l}\Delta\mathbf{\Omega}_{k,j+2,l}+\\ +\mathcal{R}''_{k,j,l}\Delta\mathbf{\Omega}_{k,j,l+1}+\mathcal{T}''_{k,j,l}\Delta\mathbf{\Omega}_{k,j,l+2}+\mathcal{X}_{k,j,l}\Delta\mathcal{P}_{j,l}+\mathbf{S}_{k,j,l},\\ k = 2, 3, \ldots, M-1, \quad j = 1, 2, \ldots, N-2, \quad l = 2, 3, \ldots, I-1.\end{gathered} \tag{4.9}$$

Здесь $\mathcal{R}_{k,j,l}$, $\mathcal{R}'_{k,j,l}$, $\mathcal{R}''_{k,j,l}$ и $\mathcal{T}_{k,j,l}$, $\mathcal{T}'_{k,j,l}$, $\mathcal{T}''_{k,j,l}$ – матрицы размеров $I_\Omega \times I_\Omega$, $\mathbf{S}_{k,j,l}$ есть I_Ω-компонентный вектор. Таким образом, используются неявные схемы аппроксимации при решении системы уравнений по координатам x, y z и условию взаимодействия. Однако в случае отрыва потока ($u < 0$) в уравнениях пограничного слоя в области возвратного течения по x фактически реализуется явная схема, что, конечно, должно повлиять на скорость сходимости итерационного процесса. Для проблемы взаимодействия трехмерного пограничного слоя с невязким потоком очень важно, как выбрать вектор $\Delta\mathcal{P}_{j,l}$, чтобы получить устойчивое решение. Было найдено, что достаточно надежные и быстрые численные методы могут быть реализованы в случае, если вектор $\Delta\mathcal{P}_{j,l}$ определить в виде

$$\Delta\mathcal{P}_{j,l} = \begin{bmatrix}\Delta\mathbf{P}_{j,l}\\ \Delta\mathbf{P}_{j-1,l}\\ \Delta\mathbf{P}_{j-2,l}\\ \Delta\mathbf{P}_{j,l-1}\\ \Delta\mathbf{P}_{j,l+1}\end{bmatrix}, \quad j = 3, 4, \ldots, N, \quad \Delta\mathcal{P}_{1,l} = \begin{bmatrix}\Delta\mathbf{P}_{1,l}\\ 0\\ 0\\ 0\\ 0\end{bmatrix}, \quad \Delta\mathcal{P}_{2,l} = \begin{bmatrix}\Delta\mathbf{P}_{2,l}\\ \Delta\mathbf{P}_{1,l}\\ 0\\ 0\\ 0\end{bmatrix}; \tag{4.10}$$

тогда $\Delta\mathcal{P}_{j,l}$ есть $5I_p$-компонентный вектор и $\mathcal{X}_{k,j,l}$ есть $I_\Omega \times 5I_p$-матрица. Для $j = N-1$ и $j = N$ формула (4.8) приводит к выражениям

$$\begin{gathered}\Delta\mathbf{\Omega}_{k,N-1,l} = \mathcal{R}_{k,N-1,l}\Delta\mathbf{\Omega}_{k+1,N-1,l}+\mathcal{T}_{k,N-1,l}\Delta\mathbf{\Omega}_{k+2,N-1,l}+\mathcal{R}'_{k,N-1,l}\Delta\mathbf{\Omega}_{k,N,l}+\\ +\mathcal{R}''_{k,N-1,l}\Delta\mathbf{\Omega}_{k,N,l+1}+\mathcal{T}''_{k,N-1,l}\Delta\mathbf{\Omega}_{k,N,l+2}+\mathcal{X}_{k,N-1,l}\Delta\mathcal{P}_{N-1,l}+\mathbf{S}_{k,N-1,l},\end{gathered} \tag{4.11}$$

$$\begin{gathered}\Delta\mathbf{\Omega}_{k,N,l} = \mathcal{R}_{k,N,l}\Delta\mathbf{\Omega}_{k+1,N,l}+\mathcal{T}_{k,N,l}\Delta\mathbf{\Omega}_{k+2,N,l}+\\ +\mathcal{R}''_{k,N,l}\Delta\mathbf{\Omega}_{k,N,l+1}+\mathcal{T}''_{k,N,l}\Delta\mathbf{\Omega}_{k,N,l+2}+\mathcal{X}_{k,N,l}\Delta\mathcal{P}_{N,l}+\mathbf{S}_{k,N,l},\\ k = 2, 3, \ldots, M-1, \quad l = 2, 3, \ldots, I-1.\end{gathered} \tag{4.12}$$

Комбинируя (4.8) и (4.9) с уравнениями (4.2), можно ввести следующие рекурсивные формулы для вычисления прогоночных матриц:

$$\mathcal{R}_{k,j,l} = -\mathcal{D}^{-1}_{k,j,l}\left(\mathcal{B}_{k,j,l}\mathcal{T}_{k-1,j,l}+\frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{\Omega}_{k+1,j,l}}\right), \quad \mathcal{T}_{k,j,l} = -\mathcal{D}^{-1}_{k,j,l}\frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{\Omega}_{k+2,j,l}},$$

$$\mathcal{R}'_{k,j,l} = -\mathcal{D}^{-1}_{k,j,l}\left(\mathcal{B}'_{k,j,l}\mathcal{T}'_{k,j-1,l}+\frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{\Omega}_{k,j+1,l}}\right), \quad \mathcal{T}'_{k,j,l} = -\mathcal{D}^{-1}_{k,j,l}\frac{\partial\mathbf{L}_{k,j,l}}{\partial\mathbf{\Omega}_{k,j+2,l}},$$

$$\mathcal{R}''_{k,j,l} = -\mathcal{D}_{k,j,l}^{-1}\left(\mathcal{B}''_{k,j,l}\mathcal{T}''_{k,j,l-1} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l+1}}\right), \quad \mathcal{T}''_{k,j,l} = -\mathcal{D}_{k,j,l}^{-1}\frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l+2}},$$

$$\mathcal{Z}_{k,j,l} = -\mathcal{D}_{k,j,l}^{-1}\left(\mathcal{B}_{k,j,l}\mathcal{Z}_{k-1,j,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k-2,j,l}}\mathcal{Z}_{k-2,j,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathcal{P}_{j,l}} + \mathcal{H}'_{j-1}\mathcal{B}'_{k,j,l}\mathcal{Z}_{k,j-1,l} + \right.$$
$$\left. + \mathcal{H}'_{j-2}\frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-2,l}}\mathcal{Z}_{k,j-2,l} + \mathcal{H}''_{l-1}\mathcal{B}''_{k,j,l}\mathcal{Z}_{k,j,l-1} + \mathcal{H}''_{l-2}\frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l-2}}\mathcal{Z}_{k,j,l-2}\right),$$

$$\mathbf{S}_{k,j,l} = -\mathcal{D}_{k,j,l}^{-1}\left(\mathcal{B}_{k,j,l}\mathbf{S}_{k-1,j,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k-2,j,l}}\mathbf{S}_{k-2,j,l} + \mathbf{L}_{k,j,l} + \mathcal{B}'_{k,j,l}\mathbf{S}_{k,j-1,l} + \right.$$
$$\left. + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-2,l}}\mathbf{S}_{k,j-2,l} + \mathcal{B}''_{k,j,l}\mathbf{S}_{k,j,l-1} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l-2}}\mathbf{S}_{k,j,l-2}\right),$$

$$k = 2, 3, \ldots, M-1, \quad j = 2, 3, \ldots, N-2, \quad l = 2, 3, \ldots, I-1.$$

где матрицы $\mathcal{B}_{k,j,l}$, $\mathcal{B}'_{k,j,l}$, $\mathcal{B}''_{k,j,l}$ и $\mathcal{D}_{k,j,l}$ имеют вид

$$\mathcal{B}_{k,j,l} = \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k-2,j,l}}\mathcal{R}_{k-2,j,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k-1,j,l}}, \quad \mathcal{B}'_{k,j,l} = \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-2,l}}\mathcal{R}'_{k,j-2,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-1,l}},$$

$$\mathcal{B}''_{k,j,l} = \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l-2}}\mathcal{R}''_{k,j,l-2} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l-1}},$$

$$\mathcal{D}_{k,j,l} = \mathcal{B}_{k,j,l}\mathcal{R}_{k-1,j,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k-2,j,l}}\mathcal{T}_{k-2,j,l} + \mathcal{B}'_{k,j,l}\mathcal{R}'_{k,j-1,l} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-2,l}}\mathcal{T}'_{k,j-2,l} +$$
$$+ \mathcal{B}''_{k,j,l}\mathcal{R}''_{k,j,l-1} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l-2}}\mathcal{T}''_{k,j,l-2} + \frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j,l}}.$$

Преобразование матрицы $\mathcal{H}'_{j-1}$ вырезает последние I_p векторов матрицы $\mathcal{B}'_{k,j,l}\mathcal{Z}_{k,j-1,l}$, делает сдвиг вправо на величину I_p всех элементов и делает нулевыми первые I_p векторов этой матрицы; соответственно, преобразование матрицы $\mathcal{H}'_{j-2}$ удаляет последние $2I_p$ векторов матрицы $\frac{\partial \mathbf{L}_{k,j,l}}{\partial \mathbf{\Omega}_{k,j-2,l}}\mathcal{Z}_{k,j-2,l}$, совершает сдвиг вправо на величину $2I_p$ всех элементов и делает нулевыми первые $2I_p$ векторов этой матрицы. Матрицы $\mathcal{H}''_{l-1}$, $\mathcal{H}''_{l-1}$ равны нулю.

Чтобы начать вычисления, использовали следующие начальные условия внизу, слева и на передней границе расчетной сетки:

$$\mathcal{R}_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\left(\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{2,j,l}}\right), \quad \mathcal{T}_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1+2,j,l}},$$

$$\mathcal{R}'_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\left(\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j+1,l}}\right), \quad \mathcal{T}'_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j+2,l}},$$

$$\mathcal{R}''_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\left(\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l+1}}\right), \quad \mathcal{T}''_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l+2}},$$

$$\mathcal{Z}_{1,j,l} = -\mathcal{D}_{1,j,l}^{-1}\left(\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathcal{P}_{j,l}} + \mathcal{H}'_{j-1}\mathcal{B}'_{1,j,l}\mathcal{Z}_{1,j-1,l} + \right.$$
$$\left. + \mathcal{H}'_{j-2}\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j-2,l}}\mathcal{Z}_{1,j-2,l} + \mathcal{H}''_{l-1}\mathcal{B}''_{1,j,l}\mathcal{Z}_{1,j,l-1} + \mathcal{H}''_{l-2}\frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l-2}}\mathcal{Z}_{1,j,l-2}\right),$$

$$\mathbf{S}_{1,j,l} = -\mathcal{D}^{-1}_{1,j,l}\left(\mathbf{L}_{1,j,l} + \mathcal{B}'_{1,j,l}\mathbf{S}_{1,j-1,l} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j-2,l}}\mathbf{S}_{1,j-2,l} + \mathcal{B}''_{1,j,l}\mathbf{S}_{1,j,l-1} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l-2}}\mathbf{S}_{1,j,l-2}\right),$$

где матрицы $\mathcal{B}_{1,j,l}$, $\mathcal{B}'_{1,j,l}$, $\mathcal{B}''_{1,j,l}$ и $\mathcal{D}_{1,j,l}$ даются с помощью формул

$$\mathcal{B}_{1,j,l} = 0, \quad \mathcal{B}'_{1,j,l} = \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j-2,l}}\mathcal{R}'_{1,j-2,l} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j-1,l}}, \quad \mathcal{B}''_{1,j,l} = \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l-2}}\mathcal{R}''_{1,j,l-2} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l-1}},$$

$$\mathcal{D}_{1,j,l} = \mathcal{B}'_{1,j,l}\mathcal{R}'_{1,j-1,l} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j-2,l}}\mathcal{T}'_{1,j-2,l} + \mathcal{B}''_{1,j,l}\mathcal{R}''_{1,j,l-1} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l-2}}\mathcal{T}'_{1,j,l-2} + \frac{\partial \mathbf{L}_{1,j,l}}{\partial \mathbf{\Omega}_{1,j,l}},$$

$$\mathcal{R}_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\left(\mathcal{B}_{k,1,l}\mathcal{T}_{k-1,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k+1,1,l}}\right), \quad \mathcal{T}_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k+2,1,l}},$$

$$\mathcal{R}'_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1+1,l}}, \quad \mathcal{T}''_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1+2,l}},$$

$$\mathcal{R}''_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\left(\mathcal{B}''_{k,1,l}\mathcal{T}''_{k,1,l-1} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l+1}}\right), \quad \mathcal{T}''_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l+2}},$$

$$\mathfrak{X}_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\left(\mathcal{B}_{k,1,l}\mathfrak{X}_{k-1,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k-2,1,l}}\mathfrak{X}_{k-2,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathcal{P}_{1,l}} + \right.$$

$$\left. + \mathcal{H}''_{l-1}\mathcal{B}''_{k,1,l}\mathfrak{X}_{k,1,l-1} + \mathcal{H}''_{l-2}\frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l-2}}\mathfrak{X}_{k,1,l-2}\right),$$

$$\mathbf{S}_{k,1,l} = -\mathcal{D}^{-1}_{k,1,l}\left(\mathcal{B}_{k,1,l}\mathbf{S}_{k-1,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k-2,1,l}}\mathbf{S}_{k-2,1,l} + \mathbf{L}_{k,1,l} + \mathcal{B}''_{k,1,l}\mathbf{S}_{k,1,l-1} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l-2}}\mathbf{S}_{k,1,l-2}\right);$$

матрицы $\mathcal{B}_{k,1,l}$, $\mathcal{B}'_{k,1,l}$, $\mathcal{B}''_{k,1,l}$ и $\mathcal{D}_{k,1,l}$ имеют вид

$$\mathcal{B}_{k,1,l} = \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k-2,1,l}}\mathcal{R}_{k-2,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k-1,1,l}}, \quad \mathcal{B}'_{k,1,l} = 0, \quad \mathcal{B}''_{k,1,l} = \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l-2}}\mathcal{R}''_{k,1,l-2} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l-1}},$$

$$\mathcal{D}_{k,1,l} = \mathcal{B}_{k,1,l}\mathcal{R}_{k-1,1,l} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k-2,1,l}}\mathcal{T}_{k-2,1,l} + \mathcal{B}''_{k,1,l}\mathcal{R}''_{k,1,l-1} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l-2}}\mathcal{T}_{k,1,l-2} + \frac{\partial \mathbf{L}_{k,1,l}}{\partial \mathbf{\Omega}_{k,1,l}},$$

$$\mathcal{R}_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\left(\mathcal{B}_{k,j,1}\mathcal{T}_{k-1,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k+1,j,1}}\right), \quad \mathcal{T}''_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k+2,j,1}},$$

$$\mathcal{R}'_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\left(\mathcal{B}'_{k,j,1}\mathcal{T}'_{k,j-1,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j+1,1}}\right), \quad \mathcal{T}'_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j+2,1}},$$

$$\mathcal{R}''_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\left(\frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j,2}}\right), \quad \mathcal{T}''_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j,3}},$$

$$\mathfrak{X}_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\left(\mathcal{B}_{k,j,1}\mathfrak{X}_{k-1,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k-2,j,1}}\mathfrak{X}_{k-2,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathcal{P}_{j,1}} + \right.$$

$$\left. + \mathcal{H}'_{j-1}\mathcal{B}'_{k,j,1}\mathfrak{X}_{k,j-1,1} + \mathcal{H}'_{j-2}\frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j-2,1}}\mathfrak{X}_{k,j-2,1}\right),$$

$$\mathbf{S}_{k,j,1} = -\mathcal{D}^{-1}_{k,j,1}\left(\mathcal{B}_{k,j,1}\mathbf{S}_{k-1,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k-2,j,1}}\mathbf{S}_{k-2,j,1} + \mathbf{L}_{k,j,1} + \mathcal{B}'_{k,j,1}\mathbf{S}_{k,j-1,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j-2,1}}\mathbf{S}_{k,j-2,1}\right),$$

а матрицы $\mathcal{B}_{k,j,1}$, $\mathcal{B}'_{k,j,1}$, $\mathcal{B}''_{k,j,1}$ и $\mathcal{D}_{k,j,1}$ имеют вид

$$\mathcal{B}_{k,j,1} = \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k-2,j,1}}\mathcal{R}_{k-2,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k-1,j,1}}, \quad \mathcal{B}'_{k,j,1} = \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j-2,1}}\mathcal{R}'_{k,j-2,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j-1,1}}, \quad \mathcal{B}''_{k,j,1} = 0,$$

$$\mathcal{D}_{k,j,1} = \mathcal{B}_{k,j,1}\mathcal{R}_{k-1,j,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k-2,j,1}}\mathcal{T}_{k-2,j,1} + \mathcal{B}'_{k,j,1}\mathcal{R}'_{k,j-1,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j-2,1}}\mathcal{T}'_{k,j-2,1} + \frac{\partial \mathbf{L}_{k,j,1}}{\partial \mathbf{\Omega}_{k,j,1}}.$$

Эти формулы получены подстановкой (4.8) в (4.3), (4.5) и (4.6). Как только прогоночные матрицы найдены, можно свести уравнение (4.8) к системе

$$\Delta \mathbf{\Omega}_{k,j,l} = \mathcal{F}_{k,j,l}\Delta \mathcal{P}_{j,l} + \mathbf{G}_{k,j,l}, \quad j = 1, 2, \ldots, N. \quad k = 1, 2, \ldots, M, \quad l = 1, 2, \ldots, I. \tag{4.13}$$

Здесь матрицы $\mathcal{F}_{k,j,l}$ и векторы $\mathbf{G}_{k,j,l}$ вычисляются рекурсивно с использованием формул

$$\mathcal{F}_{k,j,l} = \mathcal{Z}_{k,j,l} + \mathcal{R}_{k,j,l}\mathcal{F}_{k+1,j,l} + \mathcal{T}_{k,j,l}\mathcal{F}_{k+2,j,l} + \mathcal{H}'_{j+1}\mathcal{R}'_{k,j,l}\mathcal{F}_{k,j+1,l} + \mathcal{H}'_{j+2}\mathcal{T}'_{k,j,l}\mathcal{F}_{k,j+2,l} +$$
$$+ \mathcal{H}''_{l+1}\mathcal{R}''_{k,j,l}\mathcal{F}_{k,j,l+1} + \mathcal{H}''_{l+2}\mathcal{T}''_{k,j,l}\mathcal{F}_{k,j,l+2},$$

$$\mathbf{G}_{k,j,l} = \mathbf{S}_{k,j,l} + \mathcal{R}_{k,j,l}\mathbf{G}_{k+1,j,l} + \mathcal{T}_{k,j,l}\mathbf{G}_{k+2,j,l} + \mathcal{R}'_{k,j,l}\mathbf{G}_{k,j+1,l} + \mathcal{T}'_{k,j,l}\mathbf{G}_{k,j+2,l} +$$
$$+ \mathcal{R}''_{k,j,l}\mathbf{G}_{k,j,l+1} + \mathcal{T}''_{k,j,l}\mathbf{G}_{k,j,l+2}.$$

Преобразование матрицы $\mathcal{H}'_{j+1}$ удаляет последние I_p векторов матрицы $\mathcal{R}'_{k,j,l}\mathcal{F}_{k,j+1,l}$, сдвигает влево на величину I_p все элементы и делает нулевыми последние I_p векторов этой матрицы, соответственно. Преобразование матрицы $\mathcal{H}'_{j+2}$ удаляет последние $2I_p$ векторов матрицы $\mathcal{T}'_{k,j,l}\mathcal{F}_{k,j+2,l}$, сдвигает влево на величину $2I_p$ все элементы и делает нулевыми последние $2I_p$ векторов этой матрицы. Преобразование матриц $\mathcal{H}''_{l+1}$ удаляет первые I_p и последние I_p векторов матрицы $\mathcal{R}''_{k,j,l}\mathcal{F}_{k,j,l+1}$, сдвигает влево на величину I_p все элементы матрицы и делает нулевыми последние $2I_p$ векторов этой матрицы соответственно. Преобразование матриц $\mathcal{H}''_{l+2}$ удаляет первые $2I_p$ векторов матрицы $\mathcal{T}''_{k,j,l}\mathcal{F}_{k,j,l+2}$, сдвигает влево на величину $2I_p$ все элементы матрицы и делает нулевыми последние $2I_p$ векторов этой матрицы. Тогда начальные условия должны быть сформулированы на верхней, левой и задней стороне вычислительной области. Из (4.4), (4.10) и (4.7) видим, что

$$\mathcal{F}_{M,j,l} = \mathcal{Z}_{M,j,l} + \mathcal{H}'_{j+1}\mathcal{R}'_{M,j,l}\mathcal{F}_{M,j+1,l} + \mathcal{H}'_{j+2}\mathcal{T}'_{M,j,l}\mathcal{F}_{M,j+2,l} +$$
$$+ \mathcal{H}''_{l+1}\mathcal{R}''_{M,j,l}\mathcal{F}_{M,j,l+1} + \mathcal{H}''_{l+2}\mathcal{T}''_{M,j,l}\mathcal{F}_{M,j,l+2},$$

$$\mathbf{G}_{M,j,l} = \mathbf{S}_{M,j,l} + \mathcal{R}'_{M,j,l}\mathbf{G}_{M,j+1,l} + \mathcal{T}'_{M,j,l}\mathbf{G}_{M,j+2,l} + \mathcal{R}''_{M,j,l}\mathbf{G}_{M,j,l+1} + \mathcal{T}''_{M,j,l}\mathbf{G}_{M,j,l+2};$$

$$\mathcal{F}_{k,N,l} = \mathcal{Z}_{k,N,l} + \mathcal{R}_{k,N,l}\mathcal{F}_{k+1,N,l} + \mathcal{T}_{k,N,l}\mathcal{F}_{k+2,N,l} + \mathcal{H}''_{l+1}\mathcal{R}''_{k,N,l}\mathcal{F}_{k,N,l+1} + \mathcal{H}''_{l+2}\mathcal{T}''_{k,N,l}\mathcal{F}_{k,N,l+2},$$

$$\mathbf{G}_{k,N,l} = \mathbf{S}_{k,N,l} + \mathcal{R}_{k,N,l}\mathbf{G}_{k+1,N,l} + \mathcal{T}_{k,N,l}\mathbf{G}_{k+2,N,l} + \mathcal{R}''_{k,N,l}\mathbf{G}_{k,N,l+1} + \mathcal{T}''_{k,N,l}\mathbf{G}_{k,N,l+2};$$

$$\mathcal{F}_{k,j,I} = \mathcal{Z}_{k,j,I} + \mathcal{R}_{k,j,I}\mathcal{F}_{k+1,j,I} + \mathcal{T}_{k,j,I}\mathcal{F}_{k+2,j,I} + \mathcal{H}'_{j+1}\mathcal{R}'_{k,j,I}\mathcal{F}_{k,j+1,I} + \mathcal{H}'_{j+2}\mathcal{T}'_{k,j,I}\mathcal{F}_{k,j+2,I},$$

$$\mathbf{G}_{k,j,I} = \mathbf{S}_{k,j,I} + \mathcal{R}_{k,j,I}\mathbf{G}_{k+1,j,I} + \mathcal{T}_{k,j,I}\mathbf{G}_{k+2,j,I} + \mathcal{R}'_{k,j,I}\mathbf{G}_{k,j+1,I} + \mathcal{T}'_{k,j,I}\mathbf{G}_{k,j+2,I}.$$

Чтобы определить неизвестный вектор $\mathcal{P}_{j,l}$, используем систему уравнений (3.33). Из этих уравнений получим

$$\sum_{s=1}^{N}\sum_{t=1}^{I}\left(\frac{\partial Q_{j,l}}{\partial \mathbf{\Omega}_{M,s,t}}\Delta \mathbf{\Omega}_{M,s,t} + \frac{\partial Q_{j,l}}{\partial p_{s,t}}\Delta p_{s,t}\right) = -Q_{j,l}. \tag{4.14}$$

Совместно с (4.10), (4.13) имеет систему уравнений для нахождения $\Delta p_{j,\,l}$ и затем $\Delta\mathcal{P}_{j,\,l}$. Эта система решается с использованием метода матричной прогонки. Введем вектор $\Delta\mathbf{p}_j$ размерности I:

$$\Delta\mathbf{p}_j = \begin{bmatrix} \Delta p_{j,\,1} \\ \dots \\ \Delta p_{j,\,l} \\ \dots \\ \Delta p_{j,\,I} \end{bmatrix}.$$

Для этой векторной функции в общем случае имеем систему уравнений

$$\sum_{s=1}^{N} \mathbf{A}_{j,\,s}\Delta\mathbf{p}_s + \mathbf{F}_j = 0, \quad j = 1, 2, \dots, N, \tag{4.15}$$

где матрицы $\mathbf{A}_{j,\,s}$ и векторы $\mathbf{F}_j$ находятся из (4.13), (4.10), (4.14). Будем искать приближенное решение этих уравнений на каждой n-й итерации и будем принимать в рассмотрение для каждого вектора уравнения с индексом j только пять матриц, ближайших к диагональной матрице $\mathbf{A}$ с индексом j, j, т.е. матрицы $\mathbf{A}_{j,\,j-2}, \dots, \mathbf{A}_{j,\,j+2}$. Для решения этих уравнений использовался метод матричной прогонки, основанный на рекуррентном использовании формулы

$$\Delta\mathbf{p}_j = \mathbf{R}_j\Delta\mathbf{p}_{j+1} + \mathbf{T}_j\Delta\mathbf{p}_{j+2} + \mathbf{S}_j, \quad j = 1, 2, \dots, N-2. \tag{4.16}$$

Здесь $\mathbf{R}_j$ и $\mathbf{T}$ являются $I \times I$-матрицами, а $\mathbf{S}_j$ есть I-компонентный вектор.

Комбинируя (4.16) с уравнениями (4.14), можно вывести следующую рекурсивную формулу для вычисления прогоночных матриц:

$$\mathbf{R}_j = -\mathbf{D}_j^{-1}(\mathbf{B}_j\mathbf{T}_{j-1} + \mathbf{A}_{j,\,j+1}), \quad j = 3, 4, \dots, N-1,$$

$$\mathbf{T}_j = -\mathbf{D}_j^{-1}\mathbf{A}_{j,\,j+2}, \quad j = 3, 4, \dots, N-2,$$

$$\mathbf{S}_j = -\mathbf{D}_j^{-1}(\mathbf{B}_j\mathbf{S}_{j-1} + \mathbf{S}_{j-2} + \mathbf{F}_j), \quad j = 3, 4, \dots, N,$$

где матрицы $\mathbf{B}_j$ и $\mathbf{D}_j$ даются выражениями

$$\mathbf{B}_j = \mathbf{A}_{j,\,j-2}\mathbf{R}_{j-2} + \mathbf{A}_{j,\,j-1}, \quad \mathbf{D}_j = \mathbf{B}_j\mathbf{R}_{j-1} + \mathbf{A}_{j,\,j-2}\mathbf{T}_{j-2} + \mathbf{A}_{j,\,j}.$$

Чтобы начать вычисления, мы использовали следующие начальные условия:

$$\mathbf{R}_1 = -\mathbf{A}_{1,\,1}^{-1}\mathbf{A}_{1,\,2}, \quad \mathbf{T}_1 = -\mathbf{A}_{1,\,1}^{-1}\mathbf{A}_{1,\,3}, \quad \mathbf{S}_1 = \mathbf{A}_{1,\,1}\mathbf{F}_1$$

и

$$\mathbf{R}_2 = \mathbf{A}'(\mathbf{A}_{2,\,1}\mathbf{T}_1 + \mathbf{A}_{2,\,3}), \quad \mathbf{A}' = -(\mathbf{A}_{2,\,1}\mathbf{R}_1 + \mathbf{A}_{2,\,2})^{-1},$$

$$\mathbf{S}_2 = \mathbf{A}'(\mathbf{A}_{2,\,1}\mathbf{F}_1 + \mathbf{F}_2), \quad \mathbf{T}_2 = \mathbf{A}'\mathbf{A}_{2,\,4}.$$

При $j = N-1$ матрицы $\mathbf{A}_{j,\,j+2}$ и при $j = N$ матрицы $\mathbf{A}_{j,\,j+1}$, $\mathbf{A}_{j,\,j+2}$ отсутствуют, т.е. формально эти матрицы равны нулю. Таким образом, матрицы $\mathbf{R}_N$, $\mathbf{T}_N$, $\mathbf{T}_{N-1}$ равны нулю и при $j = N$ находим $\Delta\mathbf{p}_N$:

$$\Delta\mathbf{p}_N = \mathbf{S}_N. \tag{4.17}$$

Далее находим $\Delta\mathbf{p}_j$ из (4.16) для $j = N-1, \dots, 1$ и $\Delta\mathcal{P}_{j,\,l}^n$ из (4.10). Векторы $\Delta\mathbf{\Omega}_{k,\,j,\,l}$ находятся из (4.13). Новое приближение для векторов $\mathbf{\Omega}_{k,\,j,\,l}^{n+1}$ и $\mathcal{P}_{j,\,l}^{n+1}$ берется в форме

$$\mathbf{\Omega}_{k,\,j,\,l}^{n+1} = \mathbf{\Omega}_{k,\,j,\,l}^{n} + r\Delta\mathbf{\Omega}_{k,\,j,\,l}, \quad \mathcal{P}_{j,\,l}^{n+1} = \mathcal{P}_{j,\,l}^{n} + r\Delta\mathcal{P}_{j,\,l}, \quad p_{j,\,l}^{n+1} = p_{j,\,l}^{n} + r\Delta p_{j,\,l}, \tag{4.18}$$

где r – релаксационный параметр.

5. МЕТОД А

На основе данного алгоритма были разработаны методы решения трехмерных уравнений пограничного слоя с отрывом и взаимодействием. В методе А решение находится для вектора $\mathbf{\Omega}$, определенного в (3.24):

$$\mathbf{\Omega}_{k,j,l} = \begin{bmatrix} \omega_{k,j,l} \\ \tau_{k,j,l} \\ v_{k,j,l} \\ u_{k,j,l} \\ w_{k,j,l} \end{bmatrix}. \tag{5.1}$$

Для этого вектора система граничных условий имеет следующую форму. С левой стороны вычислительной области имеем

$$\begin{gathered} L^1_{k,1,l} = \omega_{k,1,l} - 1 = 0, \quad L^2_{k,1,l} = \tau_{k,1,l} = 0, \quad L^3_{k,1,l} = v_{k,1,l} = 0, \\ L^4_{k,1,l} = u_{k,1,l} - y_k = 0, \quad L^5_{k,1,l} = w_{k,1,l} = 0, \quad k = 1, 2, \dots, M, \quad l = 1, 2, \dots, I. \end{gathered} \tag{5.2}$$

Здесь и далее $L^s_{k,j,l}$, $s = 1, \dots, 5$, определяет конечно-разностный оператор, действующий на сеточную функцию $\{\omega_{k,j,l}\}$, $\{\tau_{k,1,l}\}$, $\{v_{k,1,l}\}$, $\{u_{k,1,l}\}$, $\{w_{k,j,l}\}$. Тогда граничные условия (5.2) могут быть представлены в виде

$$\mathbf{L}_{k,1,l} = \begin{bmatrix} L^1_{k,1,l} \\ \dots \\ L^3_{k,1,l} \\ \dots \\ L^5_{k,1,l} \end{bmatrix} = 0. \tag{5.3}$$

Возвратимся к уравнениям (3.1), (3.2). Для узловых точек линии сетки, следующей вверх по потоку от границы, использовалась аппроксимация Крэнка–Николсон для $\partial\omega/\partial x$, $\partial\tau/\partial x$, приводя к следующей конечно-разностной схеме второго порядка для (3.1), (3.2), которая используется для $j = 2$ и всех $k = 2, 3, \dots, M-1$, $l = 2, 3, \dots, I-1$:

$$\begin{aligned} L^1_{k,j,l} = {} & \frac{1}{2}(u_{k,j,l} + u_{k,j-1,l})\frac{\omega_{k,j,l} - \omega_{k,j-1,l}}{x_j - x_{j-1}} + \frac{1}{2}(v_{k,j,l} + v_{k,j-1,l})\frac{\lambda_y\omega_{k,j,l} + \lambda_y\omega_{k,j-1,l}}{2} + \\ & + \frac{1}{2}(w_{k,j,l} + w_{k,j-1,l})\frac{\omega_{k,j,l-1} - \omega_{k,j,l+1} + \omega_{k,j-1,l-1} - \omega_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})} - \\ & - \frac{1}{2}(\omega_{k,j,l} + \omega_{k,j-1,l})\frac{w_{k,j,l-1} - w_{k,j,l+1} + w_{k,j-1,l-1} - w_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})} + \\ & + \frac{1}{2}(\tau_{k,j,l} + \tau_{k,j-1,l})\frac{u_{k,j,l-1} - u_{k,j,l+1} + u_{k,j-1,l-1} - u_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})} - \frac{\lambda_{yy}\omega_{k,j,l} + \lambda_{yy}\omega_{k,j-1,l}}{2} = 0, \end{aligned} \tag{5.4}$$

$$\begin{aligned} L^2_{k,j,l} = {} & \frac{1}{2}(u_{k,j,l} + u_{k,j-1,l})\frac{\tau_{k,j,l} - \tau_{k,j-1,l}}{x_j - x_{j-1}} + \frac{1}{2}(v_{k,j,l} + v_{k,j-1,l})\frac{\lambda_y\tau_{k,j,l} + \lambda_y\tau_{k,j-1,l}}{2} + \\ & + \frac{1}{2}(w_{k,j,l} + w_{k,j-1,l})\frac{\tau_{k,j,l-1} - \tau_{k,j,l+1} + \tau_{k,j-1,l-1} - \tau_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})} - \\ & - \frac{1}{2}(\omega_{k,j,l} + \omega_{k,j-1,l})\frac{w_{k,j,l} - w_{k,j-1,l}}{x_j - x_{j-1}} + \frac{1}{2}(\tau_{k,j,l} + \tau_{k,j-1,l})\frac{u_{k,j,l} - u_{k,j-1,l}}{x_j - x_{j-1}} - \frac{\lambda_{yy}\tau_{k,j,l} + \lambda_{yy}\tau_{k,j-1,l}}{2} = 0. \end{aligned} \tag{5.5}$$

$$L^3_{k,j,l} = \frac{v_{k,j,l} + v_{k,j-1,l} - v_{k-1,j,l} - v_{k-1,j-1,l}}{2(y_k - y_{k-1})} + \frac{u_{k,j,l} - u_{k,j-1,l}}{x_j - x_{j-1}} + \\ + \frac{w_{k,j,l-1} - w_{k,j,l+1} + w_{k,j-1,l-1} - w_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})}, \tag{5.6}$$

$$L^4_{k,j,l} = \frac{u_{k,j,l} - u_{k-1,j,l}}{y_k - y_{k-1}} - 0.5(\omega_{k,j,l} + \omega_{k-1,j,l}), \tag{5.7}$$

$$L^5_{k,j,l} = \frac{w_{k,j,l} - w_{k-1,j,l}}{y_k - y_{k-1}} - 0.5(\tau_{k,j,l} + \tau_{k-1,j,l}). \tag{5.8}$$

Символы λ_y и λ_{yy} в (5.4), (5.5) обозначают конечно-разностные операторы, действующие на сеточную функцию, чтобы представить первую и вторую производные по y. Будучи применены к завихренности, они могут быть записаны в виде

$$\lambda_y\omega_{k,j,l} = \frac{-\xi^2\omega_{k-2\beta,j,l} + \omega_{k-\beta,j,l} - (1-\xi^2)\omega_{k,j,l}}{(y_{k-2\beta} - y_k)(\xi - \xi^2)}, \tag{5.9}$$

$$\lambda_{yy}\omega_{k,j,l} = 2\frac{\nu\omega_{k-2\beta,j,l} + \bar{\delta}\omega_{k-\beta,j,l} + \omega_{k+\beta,j,l} - (1+\bar{\delta}+\nu)\omega_{k,j,l}}{(\nu + \bar{\delta}\xi^2 + \gamma^2)(y_{k-2\beta} - y_k)^2}. \tag{5.10}$$

Здесь

$$\beta = \begin{cases} \operatorname{sign}(v_{k,j,l} + v_{k,j-1,l}), & k \ne 2, M-1, \\ -1, & k = 2, \\ 1, & k = M-1 \end{cases}$$

и

$$\gamma = \frac{y_{k+\beta} - y_k}{y_{k-2\beta} - y_k}, \quad \xi = \frac{y_{k-\beta} - y_k}{y_{k-2\beta} - y_k}, \quad \bar{\delta} = \frac{\gamma - \gamma^3}{\xi^3 - \xi}, \quad \nu = \frac{\gamma^3 - \gamma\xi^2}{\xi^2 - 1}.$$

Выражения (5.4), (5.5), будучи подставлены в (3.1), (3.2), позволяют вычислить величины $L_{k,j,l}$ для любого распределения $\{\omega_{k,j,l}\}$ в потоке. Эта схема устойчива при условии, что продольная компонента скорости u положительна, что справедливо для левой границы расчетной сетки ($j = 1$) и для линий сетки, ближайшей к ней ($j = 2$). Однако далее вниз по потоку продольная компонента скорости изменяет знак, если отрыв имеет место. Чтобы управлять этой ситуацией, необходимо сделать схему устойчивой независимо от знака u. Это было достигнуто использованием в (3.1), (3.2) вместо аппроксимации Крэнка–Николсон аппроксимации против потока для $\partial/\partial x$, представленной в виде

$$\lambda_x\omega_{k,j,l} = \frac{1}{x_j - x_{j-\alpha}}\left[\frac{\omega_{k,j-2\alpha,l}}{\mu(\mu-1)} - \frac{\mu}{\mu-1}\omega_{k,j-\alpha,l} + \frac{1+\mu}{\mu}\omega_{k,j,l}\right], \tag{5.11}$$

где

$$\alpha = \operatorname{sign}u_{k,j,l}, \quad \mu = \frac{x_j - x_{j-2\alpha}}{x_j - x_{j-\alpha}}.$$

Тогда уравнения (3.1), (3.2) принимают вид

$$L^1_{k,j,l} = u_{k,j,l}\lambda_x\omega_{k,j,l} + v_{k,j,l}\lambda_y\omega_{k,j,l} + w_{k,j,l}\lambda_z\omega_{k,j,l} - \omega_{k,j,l}\lambda_z w_{k,j,l} + \tau_{k,j,l}\lambda_z u_{k,j,l} - \lambda_{yy}\omega_{k,j,l} = 0, \tag{5.12}$$

$$L^2_{k,j,l} = u_{k,j,l}\lambda_x\tau_{k,j,l} + v_{k,j,l}\lambda_y\tau_{k,j,l} + w_{k,j,l}\lambda_z\tau_{k,j,l} - \tau_{k,j,l}\lambda_x u_{k,j,l} + \omega_{k,j,l}\lambda_x w_{k,j,l} - \lambda_{yy}\tau_{k,j,l} = 0, \tag{5.13}$$

что было использовано для $j = 3, 4, \ldots, N$ и $k = 2, 3, \ldots, M-1$. Снова используем формулы (5.9), (5.10) для производных ω и τ по y. Кроме того, параметр β зависит от знака $v_{k,j,l}$, т.е.

$$\beta = \begin{cases} \operatorname{sign} v_{k,j,l}, & k \neq 2, M-1, \\ -1, & k = 2, \\ 1, & k = M-1. \end{cases} \tag{5.14}$$

Для вычисления последней компоненты скорости $v_{k,j,l}$ в (5.12) и (5.13) используется следующая аппроксимация:

$$L^3_{k,j,l} = \frac{v_{k,j,l} - v_{k-1,j,l}}{y_k - y_{k-1}} + \lambda_x u_{k,j,l} + \lambda_z w_{k,j,l} = 0. \tag{5.15}$$

Здесь для $\lambda_x u_{k,j}$ формула (5.11) используется с $\alpha = 1$, которая соответствует аппроксимации вверх по потоку производной $\partial u/\partial x$. Для оператора λ_z в уравнениях (5.12), (5.13) и (5.15) используется центральная разность. Аппроксимация $\partial/\partial z$ дается формулой

$$\lambda_z \omega_{k,j,l} = \frac{1}{z_l - z_{l+1}} \left[\frac{\omega_{k,j,l-1}}{\mu'(\mu'-1)} - \frac{\mu'}{\mu'-1}\omega_{k,j,l+1} + \frac{1+\mu'}{\mu'}\omega_{k,j,l} \right], \tag{5.16}$$

где

$$\mu' = \frac{z_l - z_{l-1}}{z_l - z_{l+1}}.$$

Компоненты продольной $u_{k,j,l}$ и поперечной скорости $w_{k,j,l}$ в (5.12) и (5.13) вычисляются с помощью (5.7) и (5.8). На нижней границе вычислительной области граничные условия (3.7) представляются в виде

$$\mathbf{L}_{1,j,l} = \begin{bmatrix} \lambda_y^{(\beta=-1)} \omega_{1,j,l} - p'_{x,j,l} \\ \lambda_y^{(\beta=-1)} \tau_{1,j,l} - p'_{z,j,l} \\ v_{1,j,l} \\ u_{1,j,l} \\ w_{1,j,l} \end{bmatrix} = 0, \quad j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1,$$

где оператор $\lambda_y^{(\beta=-1)}$ дается формулой (5.9) с $\beta = -1$. Граничные условия (3.14) на верхней поверхности расчетной сетки аналогичны уравнению для завихренности (3.1). Для $j = 2$ используется аппроксимация Крэнка–Николсон:

$$L^1_{M,j,l} = \frac{\omega_{M,j,l} - \omega_{M,j-1,l}}{x_j - x_{j-1}} + \frac{\tau_{k,j,l-1} - \tau_{k,j,l+1} + \tau_{k,j-1,l-1} - \tau_{k,j-1,l+1}}{2(z_{l+1} - z_{l-1})} = 0, \tag{5.17}$$

$$L^2_{M,j,l} = \frac{\tau_{M,j,l} - \tau_{M,j-1,l}}{x_j - x_{j-1}} - \frac{1}{2}(p'_{zj,l}\omega_{M,j,l}/u^2_{M,j,l} + p'_{zj-1,l}\omega_{M,j-1,l}/u^2_{M,j-1,l}) = 0. \tag{5.18}$$

Для остальных точек расчетной области, т.е. для $j = 3, 4, \ldots, N$, используется аппроксимация вверх по потоку для представления $\partial/\partial x$ из (3.14), приводящая к виду

$$L^1_{M,j,l} = \lambda_x^{(\alpha=1)} \omega_{M,j,l} + \lambda_z \tau_{k,j,l} = 0, \tag{5.19}$$

$$L^2_{M,j,l} = \lambda_x^{(\alpha=1)} \tau_{M,j,l} - p'_{zj,l}\omega_{M,j,l}/u^2_{M,j,l} = 0. \tag{5.20}$$

Здесь $\lambda_x^{(\alpha=1)}$ взята из (5.11) с $\alpha = 1$. Операторы $L^3_{M,j,l}$, $L^4_{M,j,l}$, $L^5_{M,j,l}$ те же самые, что и $L^3_{k,j,l}$, $L^4_{k,j,l}$,

$L^5_{k,j,l}$ из уравнений (5.6)–(5.8) при $k = M$. На передней и задней границах также имеем для $l = 1$ и $l = I$:

$$L^1_{k,j,l} = \omega_{k,1,l} - 1 = 0,\quad L^2_{k,j,l} = \tau_{k,1,l} = 0,\quad L^3_{k,j,l} = v_{k,1,l} = 0,$$
$$L^4_{k,j,l} = u_{k,1,l} - y_k = 0,\quad L^5_{k,j,l} = w_{k,1,l} = 0,\quad k = 1, 2, \ldots, M,\quad l = 2, 3, \ldots, N. \tag{5.21}$$

Этот общий подход был применен для решения трехмерных уравнений со сверхзвуковым условием взаимодействия. Возвращаясь к условию взаимодействия (2.9), надо принять во внимание, что взаимодействие ведет к влиянию вверх по потоку через весь пограничный слой, даже когда пограничный слой на теле является всюду безотрывным. Это значит, что требуется дополнительное граничное условие вниз по потоку от области взаимодействия. Будем предполагать, что форма обтекаемой неровности стремится к нулю при $x^2 + z^2 \longrightarrow \infty$ и отрыв потока, если имеется, является локальным, поэтому

$$p = 0 \ \text{ при } \ x = \infty, \tag{5.22}$$

$$p = 0 \ \text{ при } \ z = \pm\infty. \tag{5.23}$$

Это значит, что уравнение (2.9) должно быть аппроксимировано так, чтобы привносить зависимость решения в вычислительной области от состояния потока вниз по течению. Это достигается использованием аппроксимации вниз по потоку для представления функции p в (2.9). В результате получаем следующий ряд операторных уравнений:

$$Q_{j,l}\left(\frac{1}{2}(p_{j,l} + p_{j+1,l}), \frac{\lambda_y^{(\beta=1)}(v_{M,j,l} + v_{M,j+1,l})}{2}, \frac{1}{2}\left(\frac{\partial f_{j,l}}{\partial x} + \frac{\partial f_{j+1,l}}{\partial x}\right)\right) = 0, \tag{5.24}$$

который должен быть использован для $j = 1, 2, \ldots, N-1$, $l = 2, 3, \ldots, I-1$. В итоге условие вниз по потоку (5.22) дает

$$Q_{N,l} = p_{N,l} = 0,\quad l = 1, 2, \ldots, I, \tag{5.25}$$

и условие затухания давления при больших $|z|$ имеет вид

$$Q_{j,l} = p_{j,l} = 0,\quad j = 2, 3, \ldots, N,\quad l = 1 \ \text{ и } \ l = I. \tag{5.26}$$

Таким образом, распределение $p_{1,l}$ ($l = 2, 3, \ldots, I-1$) не задается заранее, а определяется через систему уравнений (5.24), аппроксимирующих условие взаимодействия, и систему краевых условий (5.25), (5.26).

Необходимо определить, как аппроксимировать градиент давления в формуле (3.24):

$$\frac{\partial p_{j,l}}{\partial x} = 2\frac{\alpha' p_{j+1,l} + p_{j,l} + \beta' p_{j-1,l}}{(x_{j+1} - x_j)(\alpha' + \eta + \xi\beta')}, \tag{5.27}$$

$$\frac{\partial p_{j,l}}{\partial z} = -\frac{\eta}{1-\eta}\lambda_z p_{j+1,l} + \frac{1}{1-\eta}\lambda_z p_{j,l},$$
$$j = 2, 3, \ldots, N-1,\quad l = 2, 3, \ldots, I-1, \tag{5.28}$$

где

$$\alpha' = -\frac{\xi^2 - \eta^2}{1 - \xi^2},\quad \beta' = \frac{\eta^2 - 1}{1 - \xi^2},\quad \eta = \frac{x_{j-1} - x_1}{x_{j+1} - x_j},$$

$$\xi = -\frac{2x_j - x_{j-1} - x_{j-2}}{x_{j+1} - x_j},\quad j = 3, 4, \ldots, N-1,\quad l = 2, 3, \ldots, I-1,\quad \xi = 3\frac{x_j - x_{j-1}}{x_{j+1} - x_j},\quad j = 2.$$

Это заканчивает формулирование конечно-разностных уравнений.

6. МЕТОД B

Этот метод основан на использовании функции σ, определенной в [19] по формуле

$$\sigma = \frac{\partial\omega}{\partial x} + \frac{\partial\tau}{\partial z}. \tag{6.1}$$

Дифференцируя уравнения (3.1) по x и уравнения (3.2) по z и складывая их вместе, получаем

$$u\frac{\partial\sigma}{\partial x} + v\frac{\partial\sigma}{\partial y} + w\frac{\partial\sigma}{\partial z} + \Phi = \frac{\partial^2\sigma}{\partial y^2}, \tag{6.2}$$

где

$$\Phi = 2\left(\frac{\partial u}{\partial z}\frac{\partial\tau}{\partial x} + \frac{\partial w}{\partial x}\frac{\partial\omega}{\partial z}\right) + \left(\frac{\partial u}{\partial x} - \frac{\partial w}{\partial z}\right)\left(\frac{\partial\omega}{\partial x} - \frac{\partial\tau}{\partial z}\right) + \frac{\partial v}{\partial x}\frac{\partial\omega}{\partial y} + \frac{\partial v}{\partial z}\frac{\partial\tau}{\partial y}.$$

Уравнения (6.2) должны удовлетворять граничным условиям

$$\sigma = 0 \quad \text{при} \quad x \longrightarrow -\infty, \tag{6.3}$$

$$\sigma = 0 \quad \text{при} \quad z \longrightarrow \pm\infty, \tag{6.4}$$

$$\sigma = 0 \quad \text{при} \quad y \longrightarrow \infty, \tag{6.5}$$

$$\frac{\partial\sigma}{\partial y} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} \quad \text{при} \quad y = 0. \tag{6.6}$$

Условия (6.3) и (6.4) выводятся с помощью дифференцирования уравнений (2.5) и (2.6) по x и z и сложением результатов. Условие (6.5) выводится из (3.14). Условие (6.6) получается в результате дифференцирования (2.1) и (2.2) по x и z и сложением результатов при $y = 0$. Дифференцируя уравнение неразрывности (2.3) по y, находим, что вторая производная вертикальной компоненты скорости определяется по формуле σ, а именно

$$\frac{\partial^2 v}{\partial y^2} = -\sigma, \tag{6.7}$$

с граничными условиями

$$\frac{\partial v}{\partial y} = 0 \quad \text{при} \quad y = 0, \tag{6.8}$$

что следует из уравнения (2.3) при $y = 0$. С целью переформулирования задачи определим вектор переменных $\mathbf{\Omega}$ и $\mathbf{P}$:

$$\mathbf{\Omega} = \begin{bmatrix} \omega \\ \tau \\ v \\ u \\ w \\ \sigma \\ v'_y \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} p'_x \\ p'_z \\ p''_{xx} + p''_{zz} \end{bmatrix}, \quad I_\Omega = 7, \quad I_p = 3. \tag{6.9}$$

Обозначим значения $\mathbf{\Omega}(y_k, x_j, z_l)$ вектора $\mathbf{\Omega}$ в узлах расчетной сетки через $\mathbf{\Omega}_{k, j, l}$, а значения $\mathbf{P}(x_j, z_l)$ вектора $\mathbf{P}$ в узлах расчетной сетки – через $\mathbf{P}_{j, l}$. Для этих векторов система граничных условий имеет следующий вид. Сначала на левой границе расчетной области используем уравнения

$$\mathbf{L}_{k, 1, l} = \begin{bmatrix} \omega_{k, 1, l} - 1 \\ \tau_{k, 1, l} \\ v_{k, 1, l} \\ u_{k, 1, l} - y_k \\ w_{k, 1, l} \\ \sigma_{k, 1, l} \\ v'_{yk, 1, l} \end{bmatrix} = 0, \quad k = 1, 2, \dots, M, \quad l = 1, 2, \dots, I. \tag{6.10}$$

Для узловых точек на сетке, следующих после этой границы вниз по потоку, используется аппроксимация Крэнка–Николсон для первых производных по x $\partial\sigma/\partial x$, что приводит к следующей конечно-разностной схеме второго порядка точности для уравнения (6.2), используемой при $j = 2$ и всех $k = 2, 3, \ldots, M-1$, $l = 2, 3, \ldots, I-1$:

$$L^6_{k,j,l} = \frac{1}{2}(u_{k,j,l} + u_{k,j-1,l})\lambda^c_x\sigma_{k,j,l} + \frac{1}{2}(v_{k,j,l} + v_{k,j-1,l})\frac{\lambda_y\sigma_{k,j,l} + \lambda_y\sigma_{k,j-1,l}}{2} + \\ + \frac{1}{2}(w_{k,j,l} + w_{k,j-1,l})\lambda^c_z\sigma_{k,j,l} + \Phi_{k,j,l} - \frac{\lambda_{yy}\sigma_{k,j,l} + \lambda_{yy}\sigma_{k,j-1,l}}{2} = 0, \tag{6.11}$$

$$\Phi_{k,j,l} = 2(\lambda^c_z u_{k,j,l}\lambda^c_x\tau_{k,j,l} + \lambda^c_x w_{k,j,l}\lambda^c_z\omega_{k,j,l}) + (\lambda^c_x u_{k,j,l} - \lambda^c_z w_{k,j,l})(\lambda^c_x\omega_{k,j,l} - \lambda^c_z\tau_{k,j,l}) + \\ + \lambda^c_x v_{k,j,l}\frac{\lambda_y(\omega_{k,j,l} + \omega_{k,j-1,l})}{2} + \lambda^c_z v_{k,j,l}\frac{\lambda_y(\tau_{k,j,l} + \tau_{k,j-1,l})}{2}, \tag{6.12}$$

где операторы λ^c_x и λ^c_z имеют вид

$$\lambda^c_x\sigma_{k,j,l} = \frac{\sigma_{k,j,l} - \sigma_{k,j-1,l}}{x_j - x_{j-1}}, \quad \lambda^c_z\sigma_{k,j,l} = \frac{\lambda_z(\sigma_{k,j,l} + \sigma_{k,j-1,l})}{2}$$

и λ_z дается выражением (5.16). Конечно-разностные операторы уравнения неразрывности представляются в виде

$$L^3_{k,j,l} = \frac{v_{k,j,l} + v_{k,j-1,l} - v_{k-1,j,l} - v_{k-1,j-1,l}}{2(y_k - y_{k-1})} - 0.5(v'_{yk,j,l} - v'_{yk,j-1,l}) = 0, \tag{6.13}$$

$$L^7_{k,j,l} = \frac{v'_{yk,j,l} - v'_{yk-1,j,l}}{y_k - y_{k-1}} + 0.5(\sigma_{k,j,l} + \sigma_{k-1,j,l}) = 0. \tag{6.14}$$

Во внутренних точках вычислительной сетки при $j = 3, 4, \ldots, N$, $k = 2, 3, \ldots, M$, $l = 2, 3, \ldots, I-1$ вместо схемы Крэнка–Николсон, используя аппроксимацию против потока для представления $\partial/\partial x$, получаем

$$L^6_{k,j,l} = u_{k,j,l}\lambda_x\sigma_{k,j,l} + v_{k,j,l}\lambda_y\sigma_{k,j,l} + w_{k,j,l}\lambda_z\sigma_{k,j,l} + \Phi_{k,j,l} - \lambda_{yy}\sigma_{k,j,l} = 0, \tag{6.15}$$

где

$$\Phi_{k,j,l} = 2(\lambda_z u_{k,j,l}\lambda_x\tau_{k,j,l} + \lambda_x w_{k,j,l}\lambda_z\omega_{k,j,l}) + (\lambda_x u_{k,j,l} - \lambda_z w_{k,j,l})(\lambda_x\omega_{k,j,l} - \lambda_z\tau_{k,j,l}) + \\ + \lambda_x v_{k,j,l}\lambda_y\omega_{k,j,l} + \lambda_z v_{k,j,l}\lambda_y\tau_{k,j,l}, \tag{6.16}$$

операторы λ_x и λ_z определены в (5.11), (5.16).

Конечно-разностные операторы $L^1_{k,j,l}$, $L^2_{k,j,l}$, $L^5_{k,j,l}$, $L^6_{k,j,l}$ для $j = 3, 4, \ldots, N$, $k = 2, 3, \ldots, M$, $l = 2, 3, \ldots, I-1$ те же самые, что и в (5.12), (5.13), (5.7), (5.8); операторы $L^4_{k,j,l}$ и $L^7_{k,j,l}$ также те же самые, что и в (6.13), (6.14).

Граничные условия на верхней границе расчетной области могут быть представлены в виде

$$L^6_{M,j,l} = \sigma_{M,j,l} = 0, \quad j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1. \tag{6.17}$$

Конечно-разностные операторы $L^1_{M,j,l}$, $L^2_{M,j,l}$, $L^5_{M,j,l}$, $L^6_{M,j,l}$ для $j = 2, 3, \ldots, N$, $l = 2, 3, \ldots, I-1$ те же самые, что и в (5.12), (5.13), (5.7), (5.8); операторы $L^4_{k,j,l}$ и $L^7_{k,j,l}$ также те же самые, что и в (6.13), (6.14).

На нижней границе расчетной области граничные условия имеют вид

$$\mathbf{L}_{1,j,l} = \begin{bmatrix} \lambda_y^{(\beta=-1)}\omega_{1,j,l} - p'_{x,j,l} \\ \lambda_y^{(\beta=-1)}\tau_{1,j,l} - p'_{z,j,l} \\ v_{1,j,l} \\ u_{1,j,l} \\ w_{1,j,l} \\ \lambda_y^{(\beta=-1)}(\sigma_{1,j,l} + \sigma_{1,j-1,l}) - \nabla^2(p_{j,l} + p_{j-1,l}) \\ v'_{y1,j,l} \end{bmatrix} = 0, \quad j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1,$$

где оператор $\lambda_y^{(\beta=-1)}$ определяется из (5.9) с $\beta = -1$. Конечно-разностные операторы на верхней границе расчетной области $L^1_{M,j,l}$, $L^2_{M,j,l}$, $L^4_{M,j,l}$, $L^5_{M,j,l}$ те же самые, что и в методе А. Операторы $L^3_{M,j,l}$, $L^7_{M,j,l}$ находятся из (6.13), (6.14) при $k = M$. На передней и задней границах расчетной области для $l = 1$ и $l = I$ также имеем

$$\mathbf{L}_{k,j,l} = \begin{bmatrix} \omega_{k,1,l} - 1 \\ \tau_{k,1,l} \\ v_{k,1,l} \\ u_{k,1,l} - y_k \\ w_{k,1,l} \\ \sigma_{k,1,l} \\ v'_{yk,1,l} \end{bmatrix} = 0, \quad k = 1, 2, \ldots, M, \quad j = 2, 2, \ldots, N.$$

Величина $\nabla^2 p_{j,l} = p''_{xxj,l} + p''_{zzj,l}$ вычисляется с использованием следующей конечно-разностной схемы:

$$\begin{aligned} p''_{xxj,l} + p''_{zzj,l} &= 2\frac{\hat{\nu}p_{j-2\hat{\alpha},l} + \hat{\delta}p_{j-\hat{\alpha},l} + p_{j+\hat{\alpha},l} - (1+\hat{\delta}+\hat{\nu})p_{j,l}}{(\hat{\nu} + \hat{\delta}\hat{\xi}^2 + \hat{\gamma}^2)(x_{j-2\hat{\alpha}} - x_j)^2} + \\ &+ 2\frac{\tilde{\nu}p_{j,l-2\tilde{\beta}} + \tilde{\delta}p_{j,l-\tilde{\beta}} + p_{j,l+\tilde{\beta}} - (1+\hat{\delta}+\hat{\nu})p_{j,l}}{(\tilde{\nu} + \tilde{\delta}\tilde{\xi}^2 + \tilde{\gamma}^2)(z_{j-2\tilde{\beta}} - z_l)^2}. \end{aligned} \qquad (6.18)$$

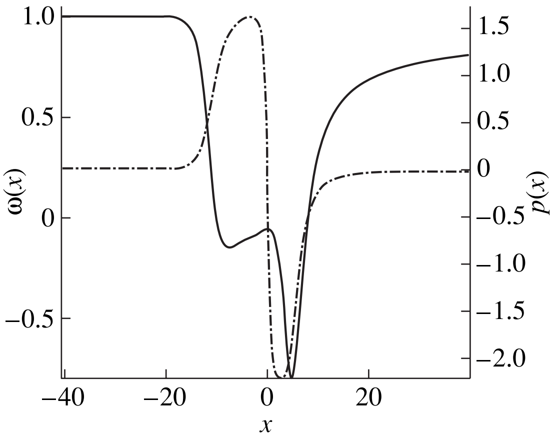
Здесь

$$\hat{\alpha} = \begin{cases} -1, & j = 2, \\ 1, & j = 3, 4, \ldots, N-1, \end{cases} \qquad \tilde{\beta} = \begin{cases} -1, & z_l \le 0, \\ 1, & z_l > 0 \end{cases}$$

и

$$\hat{\gamma} = \frac{x_{j+\hat{\alpha}} - x_j}{x_{j-2\hat{\alpha}} - x_j}, \quad \hat{\xi} = \frac{x_{j-\hat{\alpha}} - x_j}{x_{j-2\hat{\alpha}} - x_j}, \quad \hat{\delta} = \frac{\hat{\gamma} - \hat{\gamma}^3}{\hat{\xi}^3 - \hat{\xi}}, \quad \hat{\nu} = \frac{\hat{\gamma}^3 - \hat{\gamma}\hat{\xi}^2}{\hat{\xi}^2 - 1},$$

$$\tilde{\gamma} = \frac{z_{l+\tilde{\beta}} - z_l}{z_{l-2\tilde{\beta}} - z_l}, \quad \tilde{\xi} = \frac{z_{l-\tilde{\beta}} - z_l}{z_{l-2\tilde{\beta}} - z_l}, \quad \tilde{\delta} = \frac{\tilde{\gamma} - \tilde{\gamma}^3}{\tilde{\xi}^3 - \tilde{\xi}}, \quad \tilde{\nu} = \frac{\tilde{\gamma}^3 - \tilde{\gamma}\tilde{\xi}^2}{\tilde{\xi}^2 - 1}.$$



Фиг. 1. Сплошная линия – для $\tau(x)$, штриховая – для $p(x)$, сетка 701 × 101.

7. СРАВНЕНИЕ СКОРОСТИ СХОДИМОСТИ МЕТОДОВ А И В С ДРУГИМИ МЕТОДАМИ

Чтобы провести сравнение скорости сходимости предложенных методов с другими, желательно рассчитать какую-либо “тестовую проблему”. Так как в настоящее время нам не удалось обнаружить в научной литературе какого-нибудь трехмерного расчета уравнения пограничного слоя с условием сверхзвукового взаимодействия, то мы протестировали методы на примере решения двумерной задачи обтекания вязким сверхзвуковым потоком со взаимодействием и отрывом малой полости. Формулировка задачи для такого случая дана выше при условиях $\partial/\partial z = 0$, $w = \tau = 0$. Условие взаимодействия для этого случая в безразмерных переменных описывается уравнением (2.11). Форма полости дается формулой $f(x) = h/(1 + x^2)$, где h – параметр задачи. Размер расчетной сетки был 701 × 101 с постоянным шагом $\Delta x = 0.1$, $\Delta y = 0.1$. Течение рассчитывалось от величины $h = 0$ до $h = -4$ с шагом $\Delta h = -0.25$. Величина r во всех вычислениях была равной $r = 0.9$. Метод А и метод В показали практически одинаковую скорость сходимости итерационного процесса. В результате требовалось ~12 итераций, чтобы получить решение с точностью порядка 10^{-4} для безотрывных течений и ~15 итераций для течений с отрывом потока для $-2 > h \geq -4$:

$$\max_{k,\, j}\left|\omega_{k,\, j}^{n+1} - \omega_{k,\, j}^{n};\ p_j^{n+1} - p_j^{n}\right| < 10^{-4}.$$

Распределение трения и давления на поверхности тела представлены для этой задачи на фиг. 1 для $h = -4$. Для решения с помощью метода Фурье, полученного Даком, требуется ~50 итераций. Результаты показывают, что для численного решения задач двумерного течения представленные методы уступают по скорости сходимости итерационного процесса только прямому методу (см. [27]), при использовании которого требуется ~5–6 итераций.

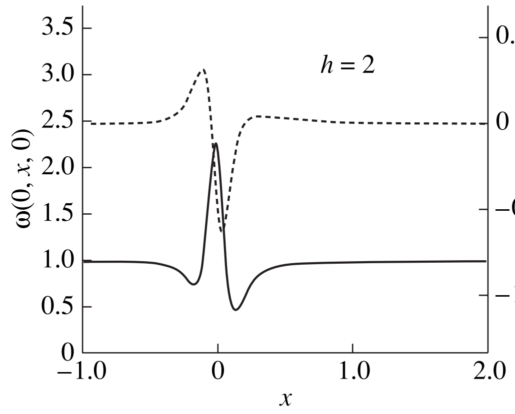
8. ТРЕХМЕРНЫЕ СВЕРХЗВУКОВЫЕ ТЕЧЕНИЯ

Для расчета пространственных сверхзвуковых течений использовалась следующая схема аппроксимации интеграла (2.9). Сначала перепишем уравнения (2.9) в виде

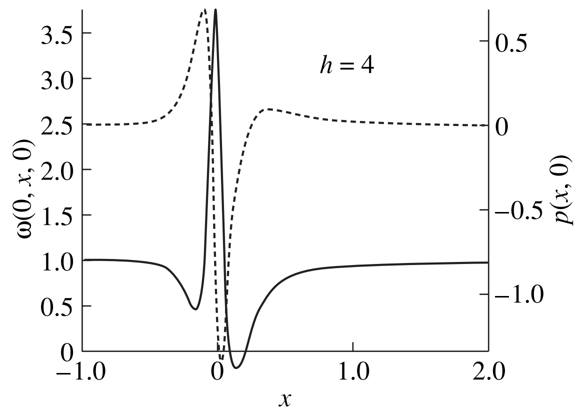
$$p = \frac{1}{\pi}\int_{-\infty}^{x}\int_{z' = z - (x' - x)}^{z' = z + (x' - x)}(-[\delta''_{xx}(x', z') + f''_{xx}(x', z')])\left(\mathrm{d}\arcsin\left(\frac{z - z'}{x - x'}\right)\right)dx'. \tag{8.1}$$

Конечно-разностное представление этого условия взаимодействия будем удовлетворять в точке $(x_{j-0.5}, z_l)$. Будем аппроксимировать в этом выражении δ''_{xx} и f''_{xx} по формуле

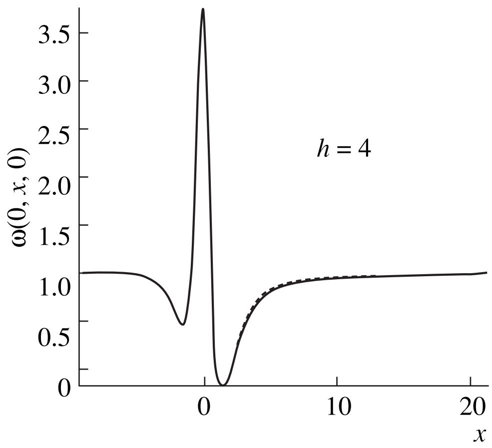
$$\delta''_{xx} + f''_{xx} = \left(\frac{\partial v}{\partial y}_{M,\, j,\, l} - \frac{\partial v}{\partial y}_{M,\, j-1,\, l} + \frac{\partial f}{\partial x}_{j,\, l} - \frac{\partial f}{\partial x}_{j-1,\, l}\right)(x_j - x_{j-1})^{-1}.$$



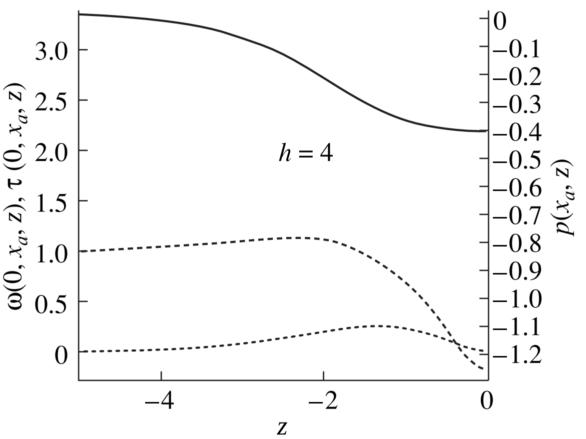
Фиг. 2.



Фиг. 3.



Фиг. 4.



Фиг. 5.

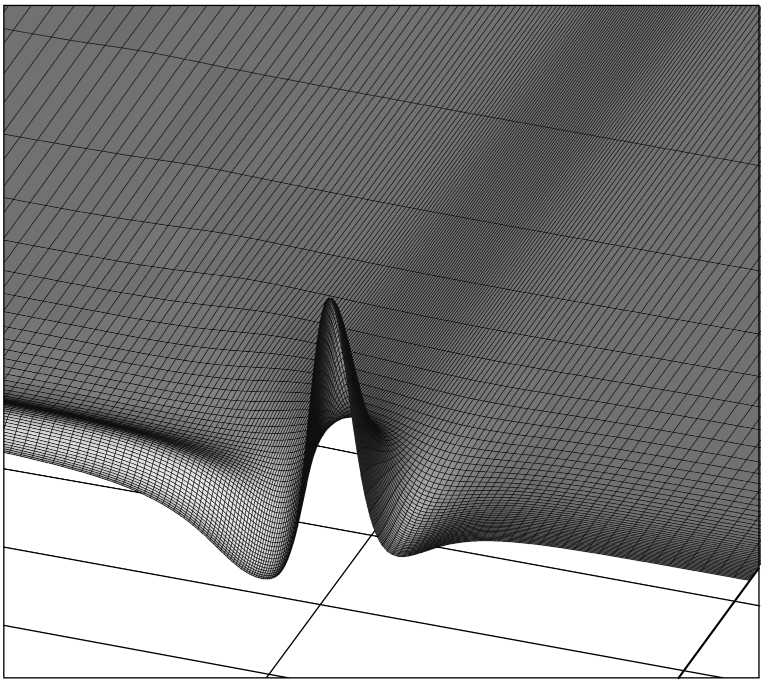
Тогда вблизи особой точки x_j, z_l часть интеграла (8.1) от малой области $x_{j-1} < x < x_j$ и $z_l + (x_j - x_{j-1}) <$ $< z < z_l - (x_j - x_{j-1})$ можно оценить через $\frac{\partial v}{\partial y}_{M,j,l} - \frac{\partial v}{\partial y}_{M,j-1,l} + \frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_{j-1}}$. Оставшаяся часть интеграла вычисляется уже обычным методом трапеций, где подынтегральное выражение $\delta''_{xx}(x', z') +$ $+ f''_{xx}(x', z')$ оценивается в центре каждой интегральной ячейки, а член $\mathrm{d}\arcsin\left(\frac{z-z'}{x-x'}\right)$ определяется как разность величин $\arcsin\left(\frac{z-z'}{x-x'}\right)$, вычисленных на середине верхней и нижней границ интегральной ячейки. В результате имеем

$$Q_{j,l} = \frac{p_{j,l} + p_{j-1,l}}{2} - \frac{\partial v}{\partial y}_{M,j,l} + \frac{\partial v}{\partial y}_{M,j-1,l} - \frac{\partial f}{\partial x_j} + \frac{\partial f}{\partial x_{j-1}} + \sum_{s'=1}^{j-1}\sum_{r'=1}^{I}\left(a_{j,l,s',r'}\frac{\partial v}{\partial y}_{M,s',r'} + b_{j,l,s',r'}\frac{\partial f}{\partial x_{s',r'}}\right) = 0,$$

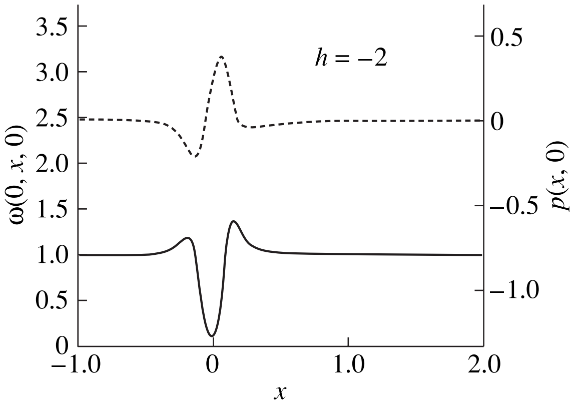
$$j = 2, 3, \ldots, N, \quad l = 2, 3, \ldots, I-1,$$

где коэффициенты $a_{j,l,s',r'}$, $b_{j,l,s',r'}$ зависят от конечно-разностной сетки и из-за громоздкости изложения не приводятся.

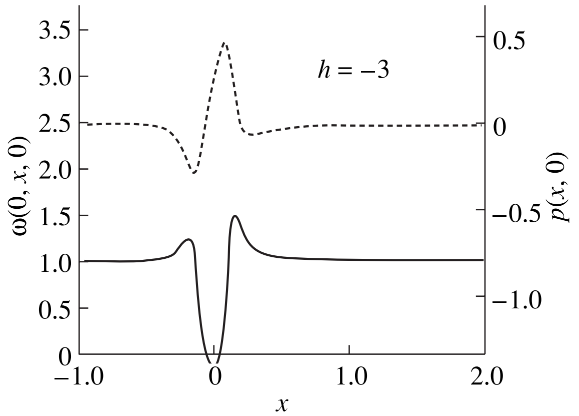
Если используется такой тип аппроксимации, то в случае, когда решение не зависит от z, получается с точностью $\frac{\partial v}{\partial y}_{M,1,l} + \frac{\partial f}{\partial x_{1,l}}$ условие взаимодействия для двумерного случая.



Фиг. 6.



Фиг. 7



Фиг. 8.

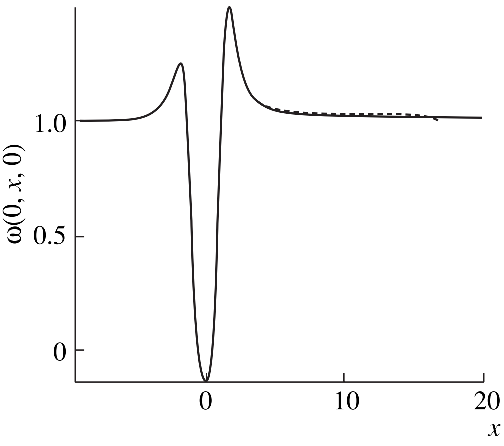
Была проведена проверка точности аппроксимации интеграла (2.9) для некоторых простых случаев, когда возможно найти аналитическое выражение для интеграла (2.9). В частности, если член $\delta''_{xx}(x, z) + f''_{xx}(x, z)$ заменяется выражением $2r\exp(-r^2)$; $r^2 = x^2 + z^2$, то можно получить аналитическую величину для p в точке $x = 0$, $z = 0$. Разность между аналитической величиной и вычисленной для разностной сетки $N = 101$, $L = 41$ достаточно мала. Для величины давления $p(0, 0)$ аналитическое значение есть 0.8473, а вычисленное 0.8323.

Было проведено исследование течения около малой неровности, описываемое уравнением

$$f(x, z) = h\exp(-r^2), \quad r^2 = x^2 + z^2,$$

где h – параметр задачи. Положительные значения h соответствуют форме бугорка, отрицательные – лунки.

Сначала вычисления были выполнены на однородной сетке размерностью $M \times N \times I = 101 \times \times 61 \times 41$. Шаги сетки $\Delta y = 0.2$, $\Delta x = 0.2$, $\Delta z = 0.4$. В диапазоне безотрывного течения для $3 > h > -2$



Фиг. 9.

требуется только 18–20 итераций, чтобы получить решение с точностью 10^{-4}:

$$\max_{k,\,j,\,l}\left|\Delta\mathbf{\Omega}_{k,\,j,\,l};\ \Delta\mathbf{P}_{j,\,l}\right| < 10^{-4}.$$

Величина параметра релаксации во всех этих расчетах была равна $r = 0.9$. Однако в области отрыва течения для величин параметра $h < -2$ и $h > 3$, чтобы получить сходимость итерационного процесса, необходимо уменьшить параметр r до величины $r = 0.4$. Число необходимых итераций также увеличивается до $n \approx 70$–80 в диапазоне параметра h, соответствующего отрывным течениям.

Результаты решения представлены на фиг. 2, 3 для параметров $h = 2$ и соответственно на оси симметрии течения $z = 0$. Размер расчетной сетки $161 \times 101 \times 101$. Сравнение результатов, сделанных на различных сетках, представлены на фиг. 4: штриховая линия – результат на сетке $101 \times 61 \times 41$, сплошная – сетка $161 \times 101 \times 101$ для $z = 0$. Характеристики течения в поперечном направлении на поверхности тела представлены на фиг. 5 для $x_a = 1.3$, где величина поверхностного трения $\omega(0, x, z)$ принимает минимальное значение. Здесь сплошной линией показано распределение давления, штриховая линия – компонента величины поверхностного трения вдоль оси x (функция ω), точечная линия есть компонента величины поверхностного трения вдоль оси z (функция τ). Пространственное распределение компоненты величины трения ω на поверхности тела представлено на фиг. 6 для $h = 4$.

Решения для отрицательных значений h представлены на фиг. 7, 8 на оси симметрии течения $z = 0$. Размер конечно-разностной сетки $161 \times 101 \times 101$. Сравнение результатов, полученных на различных сетках, дано на фиг. 9, для параметра $h = -3$. Здесь штриховая линия – результаты на сетке размерности $101 \times 61 \times 41$, сплошная линия – на сетке $161 \times 101 \times 101$ для $z = 0$.

Заметим, что представленные численные решения ограничены диапазоном $-3 \le h \le 4$. Итерационная сходимость численного решения была получена и в диапазоне, превышающем данный почти в два раза. Однако ограниченные размеры максимальной конечно-разностной сетки, используемой в данных расчетах, не позволяют получить решение с необходимой степенью точности в областях с большими градиентами функций. Вблизи них решение начинает зависеть от размера сеток. Поэтому здесь эти решения не приводятся.

9. ВЫВОДЫ

Разработан эффективный метод решения уравнений теории взаимодействия, позволяющий рассчитывать пространственные течения жидкости в области взаимодействия ламинарного пограничного слоя со сверхзвуковым невязким потоком, а также течения с другими типами взаимодействия, включая режимы отрывного течения. Этот метод без особых сложностей может быть использован для расчета турбулентного пограничного слоя со взаимодействием, а также течений жидкости и газа, где существенно влияние температуры, плотности, энтальпии и др. Также можно ввести нестационарные члены в уравнения и рассмотреть нестационарные процессы.

СПИСОК ЛИТЕРАТУРЫ

1. *Нейланд В.Я.* К теории отрыва ламинарного пограничного слоя в сверхзвуковом потоке // Изв. АН СССР. Механ. жидкости и газа. 1969. № 4. С. 53–57.
2. *Stewartson K., Williams P.G.* Self-induced separation // Proc. Roy. Soc. London. Ser. A. 1969. V. 312. № 1509. P. 181–206.
3. *Сычёв В.В., Рубан А.И., Сычёв Вик.В., Королёв Г.Л.* Асимптотическая теория отрывных течений. М.: Нака, 1987.
4. *Нейланд В.Я., Боголепов В.В., Дудин Г.Н., Липатов И.И.* Асимптотическая теория сверхзвуковых течений вязкого газа. М.: Физматлит, 2004.
5. *Bodonyi R.J., Kluwick A.* Transonic trailing-edge flow // Quart. J. Mech. and Appl. Math. 1998. V. 51. P. 297–310.

6. *Богданов А.Н., Диесперов В.Н., Жук В.И. и др.* Асимптотические модели вязких трансзвуковых течений // Анн. докл. Всес. съезда по теор. и прикл. механ. 23–29 августа. Пермь, 2003. С. 106–107.
7. *Диесперов В.Н., Королев Г.Л.* Возникновение сверхзвуковых зон и зон локального отрыва при трансзвуковом стационарном обтекании неровности поверхности в режиме свободного взаимодействия // Изв. РАН. Механ. жидкости и газа. 2003. № 1. С. 50–59.
8. *Рыжов О.С.* О нестационарном пространственном пограничном слое, свободно взаимодействующем с внешним потоком // Прикл. матем. и механ. 1980. Т. 44. Вып. 6. С. 1035–1052.
9. *Smith F.T.* On the high Reynolds number theory of laminar flows // IMA J. Appl. Math. 1982. V. 28. № 3. P. 207–281.
10. *Duck P.W., Burgraf O.R.* Spectral solutions for the three-dimensional triple-deck flow over surface topography // J. Fluid Mech. 1986. V. 162. P. 1–22.
11. *Лыжин В.Д., Дудин Г.Н.* Об одном методе расчета режима сильного вязкого взаимодействия на треугольном крыле // Изв. АН СССР. Механ. жидкости и газа. 1983. № 4. С. 119–124.
12. *Lipatov I.I., Vinogradov I.V.* Three-dimensional flow near surface distortions for the compensation regime // Philos. Trans. Roy. Soc. London. Ser. A. 2000. V. 358. № 777. P. 3143–3153.
13. *Жук В.И.* Волны Толлмина–Шлихтинга и солитоны. М.: Наука, 2001.
14. *Smith F.T.* A three-dimensional boundary layer separation // J. Fluid Mech. 1991. V. 99. P. 185–201.
15. *Delery J.M., Formery M.J.* A finite difference method for inverse solutions of 3-d turbulent boundary-layer flow: AIAA 83-0301. AIAA 21-st Aerospace Science Meeting. Jan. 10–13, 1983, Reno, Nevada.
16. *Bodonyi R.J., Duck P.W.* A numerical method for treating strongly interactive three-dimensional viscous-inviscid flows // Internat. J. Comput. Fluids. 1988. V. 16. № 3. P. 279–290.
17. *Карась О.В., Ковалев В.Е.* Применение обратного метода расчета трехмерного пограничного слоя к задаче обтекания крыла с учетом влияния вязкости // Уч. зап. ЦАГИ, 1989. Т. 20. № 5. С. 1–11.
18. *Timoshin S.N., Smith F.T.* Singularities encountered in three-dimensional boundary layers under adverse or favourable pressure gradient // Philos. Trans. Roy. Soc. London. Ser. A. 1995. V. 352. № 1698. P. 45–87.
19. *Smith F.T., Sykes R.I., Brighton P.W.M.* A two-dimensional boundary layer encountering a three-dimensional hump // J. Fluid Mech. 1977. V. 83. № 1. P. 163–176.
20. *Боголепов В.В.* Исследование малых пространственных возмущений ламинарного пограничного слоя // Прикл. механ. и техн. физ. 1987. № 5. С. 69–79.
21. *Красильщикова Е.А.* Крыло конечного размаха в сжимаемом потоке. М.-Л.: Гостехтеориздат, 1952.
22. *Bodonyi R.J., Kluwick A.* Supercritical transonic trailing-edge flow // Quart. J. Mech. and Appl. Math. 1982. V. 35. № 2. P. 265–277.
23. *Сычёв В.В., Рубан А.И.* Гиперзвуковое вязкое течение газа около крыла малого удлинения // Уч. зап. ЦАГИ. 1973. Т. 5. № 5. С. 42–51.
24. *Smith F.T.* Steady and unsteady 3-D interactive boundary layers // Comput. and Fluids. 1991. V. 20. № 3. P. 243–268.
25. *Сычёв Вик.В.* О пространственных течениях около неровностей на поверхности осесимметричного тела // Уч. зап. ЦАГИ. 1993. Т. 24. № 1. С. 12–28.
26. *Жук В.И., Рыжов О.С.* О локально-невязких возмущениях в пограничном слое с самоиндуцированным давлением // Докл. АН СССР. 1982. Т. 263. № 1. С. 56–59.
27. *Королёв Г.Л.* Об одном методе решения задач асимптотической теории взаимодействия пограничного слоя с внешним потоком // Ж. вычисл. матем. и матем. физ. 1987. Т. 27. № 8. С. 1224–1232.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 530–537

УДК 519.634

О ГЕНЕРАЦИИ ЗВУКОМ ТРЕХМЕРНЫХ ВОЛН ТОЛЛМИНА–ШЛИХТИНГА В ПОГРАНИЧНОМ СЛОЕ НА УПРУГОЙ ПОВЕРХНОСТИ ПРИ ТРАНСЗВУКОВЫХ СКОРОСТЯХ ВНЕШНЕГО ПОТОКА[1)]

© 2007 г. И. В. Савенков

(119991 Москва, ул. Вавилова, 40, ВЦ РАН)

e-mail: isavenkov@mail.ru

Поступила в редакцию 29.09.2006 г.

В рамках асимптотической теории свободного взаимодействия изучено влияние упругости обтекаемой поверхности на трехмерные пакеты волн Толлмина–Шлихтинга, генерируемые акустическими возмущениями, возникающими вблизи пограничного слоя при трансзвуковых скоростях набегающего потока. Показано, что упругость обтекаемой поверхности значительно ослабляет наиболее неустойчивые косые волны, но не меняет характерную подковообразную форму волновых пакетов с двумя максимумами возмущенного движения, распространяющимися под углом к набегающему потоку. Библ. 13. Фиг. 7.

Ключевые слова: акустические возмущения, трехмерные волны Толлмина–Шлихтинга, волновые пакеты, трансзвуковая гидродинамика, упругость поверхности.

ВВЕДЕНИЕ

Акустические возмущения, вызывающие рост волн Толлмина–Шлихтинга, являются одной из возможных причин потери устойчивости пограничного слоя с последующим переходом ламинарного течения в турбулентное. С помощью теории свободного взаимодействия (см. [1]–[3]) в [4], [5] было показано, что акустические возмущения, возникающие в потенциальной области трансзвукового потока вблизи пограничного слоя, могут непосредственно преобразовываться в волны Толлмина–Шлихтинга без наличия любых других локальных неоднородностей течения.

Одним из эффективных факторов, сдерживающих рост волн Толлмина–Шлихтинга, является упругость обтекаемой поверхности, что было показано еще в ранних экспериментальных (см. [6], [7]) и теоретических (см. [8], [9]) работах. В рамках теории свободного взаимодействия в [10] было изучено влияние упругости обтекаемой поверхности на характеристики отдельных плоских волн Толлмина–Шлихтинга, а в [11] показано, что упругость поверхности может существенно замедлять развитие пакетов волн Толлмина–Шлихтинга, генерируемых двумерными акустическими возмущениями, проникающими в пограничный слой из внешнего потока. Данная работа является обобщением [11] на трехмерный случай.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим обтекание плоской пластинки равномерным потоком сжимаемого газа со скоростью U^*_∞, мало отличающейся от скорости c^*_∞ распространения звуковых волн (число Маха $M_\infty = U^*_\infty/c^*_\infty \approx 1$). Пусть на расстоянии L^* от передней кромки пластины во внешнем потоке на расстоянии $y^* \sim \varepsilon^{-2}R^{-1/2}L^*$ от поверхности пластины возникло трехмерное безвихревое возмущение с характерным поперечным размером $\Delta y^* \sim \varepsilon^{-2}R^{-1/2}L^*$, продольным $\Delta x^* \sim \varepsilon^{-3/2}R^{-1/2}L^*$ и боковым $\Delta z^* \sim \varepsilon^{-2}R^{-1/2}L^*$ ($\varepsilon \sim R^{-1/9}$ – малый параметр теории, см. [12], [13], где число Рейнольдса $R = \mu^*_\infty U^*_\infty L^*/\rho^*_\infty \longrightarrow \infty$). Пусть все гидродинамические функции этого возмущения (плотность, давление и скорости) отклонились на величины порядка ε^2. Тогда возмущенное движение во внешней области (с продольной координатой $X = \varepsilon^{3/2}R^{1/2}(x^* - L^*)/L^*$, боковой $z = \varepsilon^2R^{1/2}z^*/L^*$ и по-

[1)] Работа выполнена при финансовой поддержке РФФИ (код проекта 04-01-00807).

перечной $y = \varepsilon^2 R^{1/2} y^*/L^*$) описывается следующим уравнением для потенциала φ (см. [13], [5]):

$$\frac{\partial^2\varphi}{\partial T\partial X} + c_\infty \frac{\partial^2\varphi}{\partial X^2} - \frac{\partial^2\varphi}{\partial y^2} - \frac{\partial^2\varphi}{\partial z^2} = 0, \tag{1.1}$$

где $c_\infty \sim (\mathrm{M}_\infty^2 - 1)\mathrm{R}^{1/9} = O(1)$ – параметр, связанный с малым отклонением числа Маха от единицы (T – время).

Возмущенное движение внешней области порождает давление $P(T, X, Z) = -\partial\varphi/\partial X(T, X, y = 0, Z)$, передающееся без изменения как в основную толщу пограничного слоя (где $y^* \sim \mathrm{R}^{-1/2}L^*$), так и в узкий пристеночный слой (с поперечной координатой $Y = \varepsilon^{-1}\mathrm{R}^{1/2}y^*/L^*$), вызывая в нем движение, описываемое системой уравнений Прандтля (см. [13], [5])

$$\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y} = 0, \quad \frac{\partial P}{\partial Y} = 0,$$

$$\frac{\partial U}{\partial T} + U\frac{\partial U}{\partial X} + V\frac{\partial U}{\partial Y} = -\frac{\partial P}{\partial X} + \frac{\partial^2 U}{\partial Y^2}, \tag{1.2}$$

$$\frac{\partial W}{\partial T} + U\frac{\partial W}{\partial X} + V\frac{\partial W}{\partial Y} = -\frac{\partial P}{\partial Z} + \frac{\partial^2 W}{\partial Y^2}.$$

Условия сращивания решений в пристеночном подслое и внешней области (через основную толщу пограничного слоя) имеют вид

$$U - Y \longrightarrow A(T, X, Z) + F(T, X, Z), \quad W \longrightarrow 0 \ \text{при} \ Y \longrightarrow \infty, \tag{1.3}$$

$$\partial\varphi/\partial X = -P, \quad \partial\varphi/\partial y = -\partial A/\partial X \ \text{при} \ y = 0, \tag{1.4}$$

где функция A имеет физический смысл мгновенного смещения линий тока в основной толще пограничного слоя. (В записи (1.1), (1.3) также использовалось преобразование Прандтля, благодаря чему обтекаемой поверхности, деформированной на величину F, в новых переменных соответствует $Y = 0$.)

Осталось еще выставить условие затухания возмущений при выходе из области внешнего потока

$$\partial\varphi/\partial y \longrightarrow 0 \ \text{при} \ y \longrightarrow \infty \tag{1.5}$$

и задать начальное условие

$$\varphi(T = 0, X, Y, Z) = \varphi_0(X, Y, Z), \tag{1.6}$$

где $\varphi_0(X, Y = 0, Z) \equiv 0$ (поскольку мы предполагаем, что в начальный момент возмущение зародилось вне пограничного слоя).

И, конечно, надо не забывать про условия прилипания на стенке

$$U = V = W = 0 \ \text{при} \ Y = 0 \tag{1.7}$$

вкупе со связью между деформацией поверхности и прикладываемым давлением

$$P = -KF, \tag{1.8}$$

где K – коэффициент упругости.

Тем самым мы завершили постановку нашей задачи. Давление P в ней является самоиндуцированным и ищется наряду с другими функциями течения.

2. ЛИНЕЙНОЕ РЕШЕНИЕ

Считая возмущения малыми, линеаризуем задачу (1.1)–(1.8) по малому амплитудному параметру $\delta \longrightarrow 0$

$$(U - Y, V, W, P, A, \varphi) = \delta(U', V', W', P', A', \varphi'),$$

а затем разложим искомые функции в интегралы Фурье–Лапласа. Например, разложение для давления имеет вид

$$P'(T, X, Z) = \mathrm{Re}\left(\frac{1}{2\pi i}\int_{-\infty}^{\infty} dm \exp(imZ)\int_0^{\infty} dk \int_{b-i\infty}^{b+i\infty} \bar{P}(k, m, \omega)\exp(\omega T + ikX)d\omega\right). \quad (2.1)$$

Тогда для образов искомых функций получается система дифференциальных уравнений (по Y), сводящаяся к решению уравнения Эйри. Окончательно имеем (см. [12], [13])

$$\bar{P} = -(ik)^{-1/3}Q(k, \omega, m)\frac{\Phi(\Omega)\int_0^{\infty}\overline{\varphi_0}(k, m, Y)dY}{\Phi(\Omega) - Q(k, \omega, m)}, \quad (2.2)$$

где

$$Q(k, \omega, m) = (ik)^{1/3}k\mu/(K - \mu), \quad \mu = k^2/\lambda, \quad \lambda = \sqrt{ik(\omega + ikc_\infty) + m^2}, \quad \mathrm{Re}\lambda > 0,$$

$$\Phi(\Omega) = \frac{dAi(\zeta)}{d\zeta}\left[\int_{\Omega}^{\infty}\mathrm{Ai}(\zeta)d\zeta\right]^{-1}, \quad \Omega = \omega(ik)^{-2/3},$$

где $\mathrm{Ai}(\zeta)$ – функция Эйри, экспоненциально затухающая в секторе $|\arg\zeta| < \pi/3$.

Приравнивание знаменателя в (2.2) к нулю приводит к дисперсионному соотношению

$$\Phi(\Omega) = Q(k, \omega, m; K), \quad (2.3)$$

описывающему собственные колебания пограничного слоя (волны Толлмина–Шлихтинга). В плоском случае ($m = 0$) это дисперсионное соотношение подробно изучалось в [10]. Было показано, что оно имеет счетный набор корней, неустойчивым из которых является только первый: $\mathrm{Re}\,\omega_1(k, m = 0) > 0$ при превышении некоторого порогового, “нейтрального” значения $k = k_*$. Анализ показал, что при $m \neq 0$ дисперсионное соотношение (2.3) ведет себя аналогичным образом: из всего счетного набора корней неустойчивым по-прежнему остается только первый.

В случае жесткой поверхности ($K = 0$) дисперсионное соотношение (2.3) изучалось в [5]. Было показано, что максимальный инкремент нарастания $\sigma_{\max} = \max\sigma$ ($0 \le k < \infty$, $m = \mathrm{const}$) неограниченно возрастает с ростом m, точнее,

$$\sigma_{\max} \sim \sigma_{0,\max}(k/k_{0,\max})^{2/3} \text{ при } k \longrightarrow +\infty \text{ и } m \sim k^{7/3}/|\Phi(\Omega_{0,\max})|. \quad (2.4)$$

Можно показать, что оценка (2.4) остается справедливой и для упругой поверхности (зависимость от K появляется лишь в следующих членах разложения). Таким образом, в трансзвуковом пограничном слое наиболее неустойчивы косые волны (распространяющиеся под углом к набегающему потоку), независимо от свойств упругости обтекаемой поверхности.

Типичная дисперсионная зависимость инкремента нарастания возмущений $\sigma = \mathrm{Re}\,\omega_1$ от k и m приведена на фиг. 1 при $c_\infty = 1$ и $K = 1$. Такой вид изолирований инкремента нарастания возмущений (как и изолиний $\mathrm{Im}\,\omega_1(k, m)$ на фиг. 2) характерен и в случае жесткой поверхности ($1/K = 0$, см. [5]).

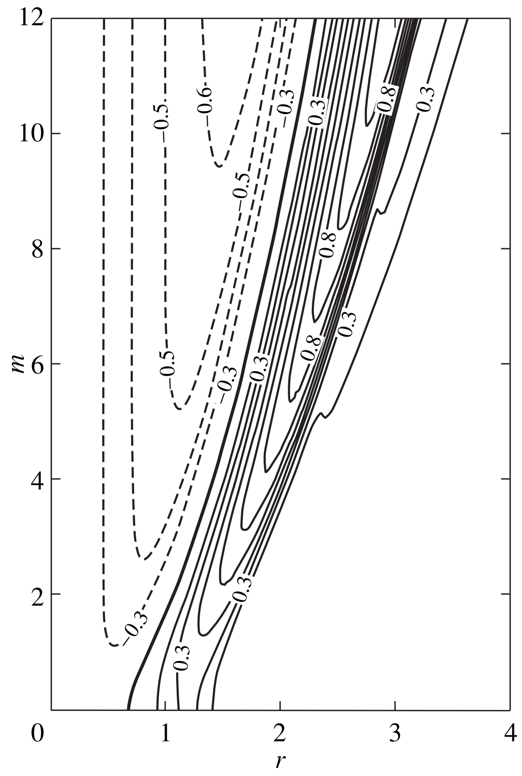
Перепишем (2.1) в виде

$$P'(T, X, Z) = \mathrm{Re}\left(\int_{-\infty}^{\infty} I(m, T, X)\exp(imZ)dm\right), \quad (2.5)$$

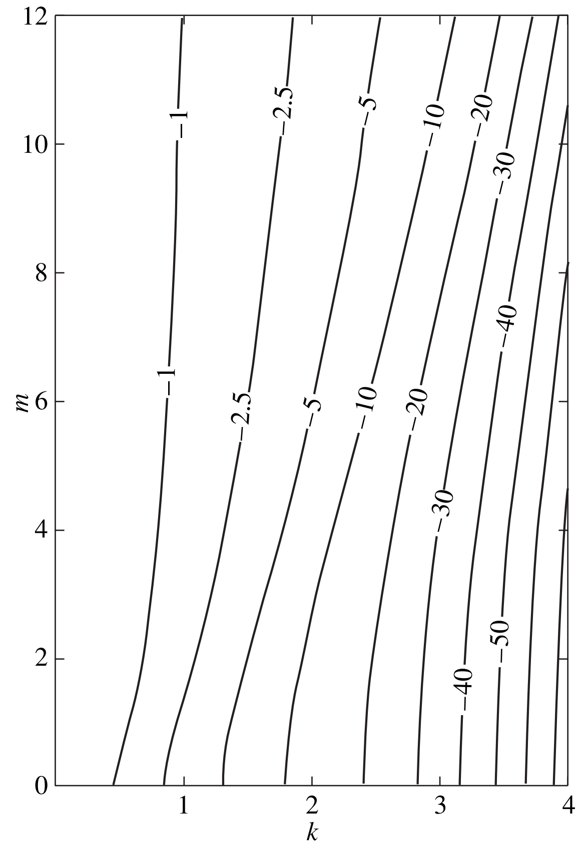
$$I = \frac{1}{2\pi i}\int_0^{\infty} dk \int_{b-i\infty}^{b+i\infty} \bar{P}(k, m, \omega)\exp(\omega T + ikX)d\omega. \quad (2.6)$$

Интеграл (2.6) изучался ранее в [11] при $m = 0$. Распространяя описанную в [11] методику на общий случай $m \neq 0$, имеем асимптотическую оценку для больших времен T (см. также [5])

$$I(m, T, X) \sim \int_0^{\infty} dk(ik)^{1/3}\int_0^{\infty}\overline{\varphi_0}(k, m, Y)\exp(-\lambda_1 Y)dY\frac{\Phi(\Omega_1)Q(\Omega_1; k, m)}{\frac{d\Phi}{d\Omega}(\Omega_1) - \frac{dQ}{d\Omega}(\Omega_1; k, m)}\exp[\omega_1(k, m)T + ikX],$$



Фиг. 1.



Фиг. 2.

где $\lambda_1 = \sqrt{ik(\omega_1 + ikc_\infty) + m^2}$, $\omega_1 = (ik)^{2/3}\Omega_1(k, m)$, Ω_1 – первый корень дисперсионного соотношения (2.3) (а под λ_1 подразумевается регулярная ветвь корня с $\mathrm{Re}\lambda_1 > 0$).

3. ВОЛНОВОЙ ПАКЕТ

Рассмотрим для простоты точечный источник возмущений, находящийся на расстоянии $Y_0 \neq 0$ от поверхности пластины: $\varphi_0 = 4\pi^2\delta(X)\delta(Z)\delta(Y - Y_0)$. Тогда

$$\int_0^\infty \overline{\varphi_0}(k, m, Y)\exp(-\lambda_1 Y)dY = \exp(-\lambda_1 Y_0)$$

и все дело сводится к вычислению двойного интеграла

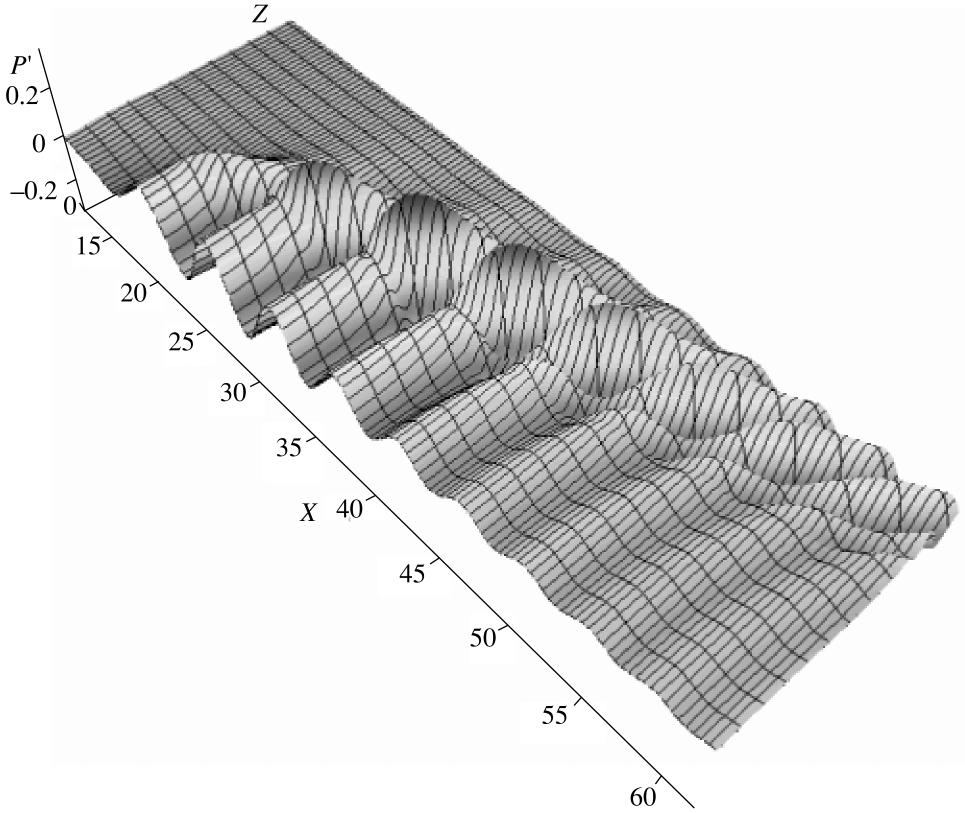
$$P_c = \int_{-\infty}^{\infty} dm \int_0^{\infty} dk (ik)^{1/3} \frac{\Phi(\Omega_1)Q(\Omega_1; k, m)}{\frac{d\Phi}{d\Omega}(\Omega_1) - \frac{dQ}{d\Omega}(\Omega_1; k, m)} \exp[\omega_1(k, m)T + ikX + imZ - \lambda_1(k, m)Y_0], \tag{3.1}$$

$$P'(T, X, Z) = \mathrm{Re}(P_c),$$

где, как уже было сказано, ω_1 является первым корнем дисперсионного соотношения (2.3).

Заметим, что, несмотря на неограниченный рост инкремента нарастания $\mathrm{Re}\omega_1 \sim \sigma_{0,\,\max}(k/k_{0,\,\max})^{2/3}$ с увеличением k, интеграл (3.1) сходится за счет асимптотики $\lambda_1 \sim k^{7/3}/|\Phi(\Omega_{0,\,\max})|$ при $k \longrightarrow +\infty$.

Были проведены численные расчеты интеграла (3.1) по методу трапеций с шагами $\Delta k = 0.003$ и $\Delta m = 0.1$ (в качестве верхнего предела достаточно было взять $k_\infty = 6.0$ и $m_\infty = 12.0$), что гарантировало 1%-ю точность вычислений.



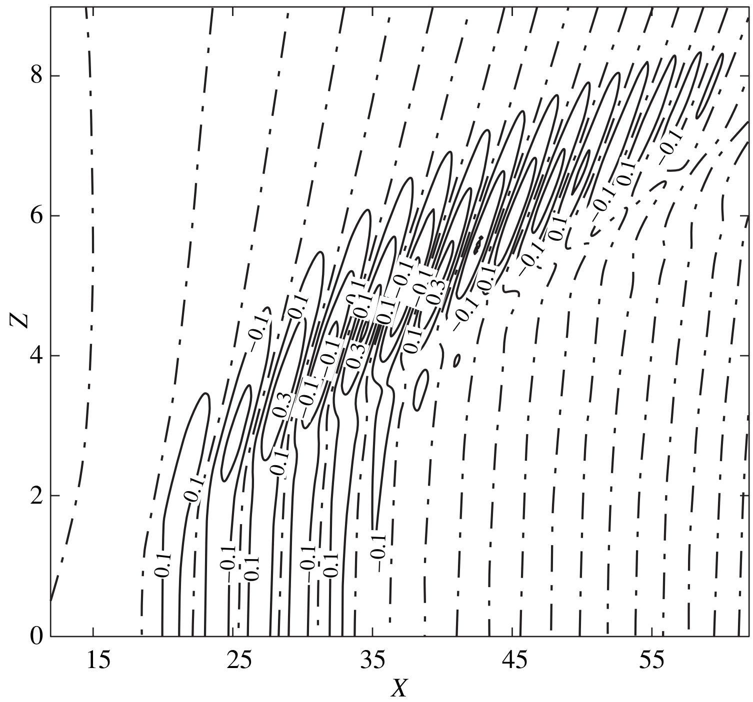
Фиг. 3.

Расчеты показали, что к моменту времени $T = 5$ (см. фиг. 3 и фиг. 4) распределение давления P' принимает характерную форму волнового пакета, растущего с течением времени (в силу симметрии по Z, приводится только половина картинки с $Z \geq 0$). Амплитудные характеристики волнового пакета удобнее проследить по фиг. 5, на которой представлены изолинии функции $|P_c|$ (огибающей P') на тот же момент времени $T = 5$. Видно, что область с максимальными амплитудами возмущений в окрестности $P_{\max} = \max|P_c(-\infty < X < \infty, 0 \leq Z < \infty)| = |P_c(X_{\max}, Z_{\max})|$ смещена с оси симметрии $Z = 0$, что свидетельствует о раздвоении возмущений (поскольку картина возмущенного движения симметрична относительно $Z = 0$). Точно такая же картина наблюдается в случае жесткой поверхности (см. [5]) и объясняется тем же фактом наибольшей неустойчивости косых волн.

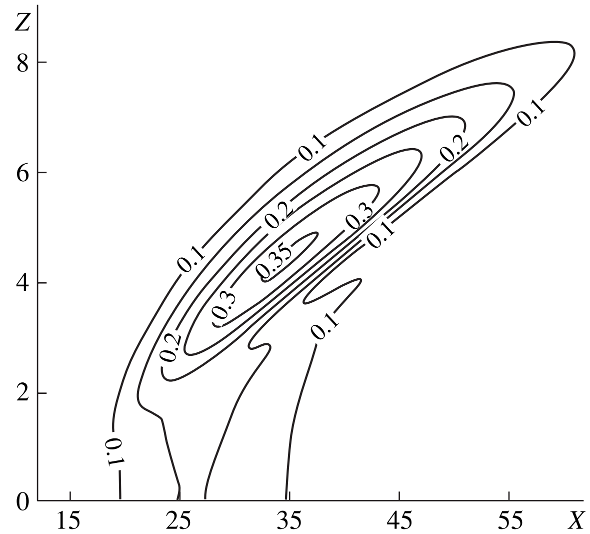
В самом деле, численный анализ подынтегральной функции из (3.1) показывает, что в момент времени $T = 5$ ее модуль достигает максимума при $k_{\max} = 1.350$ и $m_{\max} = 1.60$. Групповая скорость движения этих наиболее неустойчивых волн составляет $W_{\max} = |\partial \mathrm{Im}(\omega_1(k_{\max}, m_{\max}))/\partial m| = 0.95$ и $U_{\max} = |\partial \mathrm{Im}(\omega_1(k_{\max}, m_{\max}))/\partial k| = 6.7$, что дает оценку $X_{\max} = U_{\max}T = 33.5$ и $Z_{\max} = W_{\max}T = 4.95$, хорошо согласующуюся с вычисленными значениями $X_{\max} = 34.7$ и $Z_{\max} = 4.40$ (фиг. 5). Далее, фронт этих волн повернут под углом $\beta_{\max} = \arctan(m_{\max}/k_{\max}) = 50°$ по отношению к оси X, что также хорошо согласуется с представленными на фиг. 4 результатами.

Найденное поведение волнового пакета (вытянутая вперед подковообразная форма с максимумами возмущений на “ножках”) является характерной практически для всего трансзвукового диапазона, поскольку определяется лишь характером поведения дисперсионных кривых $\mathrm{Re}\,\omega_1(k, m)$ и $\mathrm{Im}\,\omega_1(k, m)$ (см. фиг. 1 и фиг. 2), ведущих себя подобным образом в широком диапазоне чисел c_∞ и K (выявлено из численных расчетов). Таким образом, упругость обтекаемой поверхности не влияет на качественную картину развития возмущений (образование подковообразной формы), но приводит к значительному ослаблению волновых пакетов и их более быстрому сносу вниз по течению (ср. $P_{\max} = 0.694$, $X_{\max} = 34.7$, $Z_{\max} = 4.40$ в случае $K = 1$ с $P_{\max} = 10.7$, $X_{\max} = 17.9$, $Z_{\max} = 1.70$ в случае $1/K = 0$). Тем самым основной вывод работы [11] обобщен на трехмерный случай.

Посмотрим теперь, сколь далеко во внешнюю область проникает возмущенное движение. Для образа давления $P'_{\text{out}}(T, X, Y, Z)$ во внешней области легко получить простое выражение, свя-



Фиг. 4.



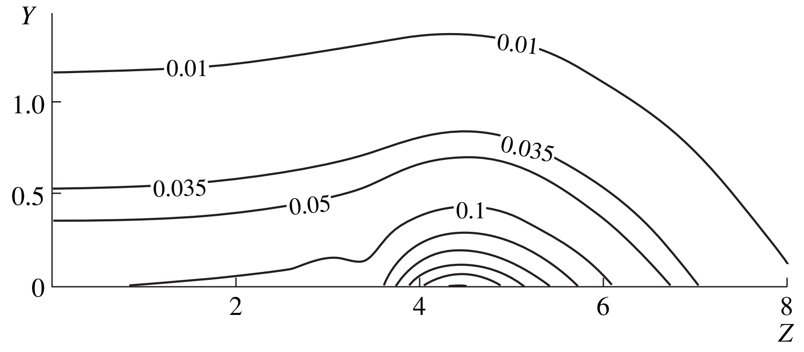
Фиг. 5.

зывающее его с образом уже изученного давления $P'(T, X, Z)$ на поверхности пластины:

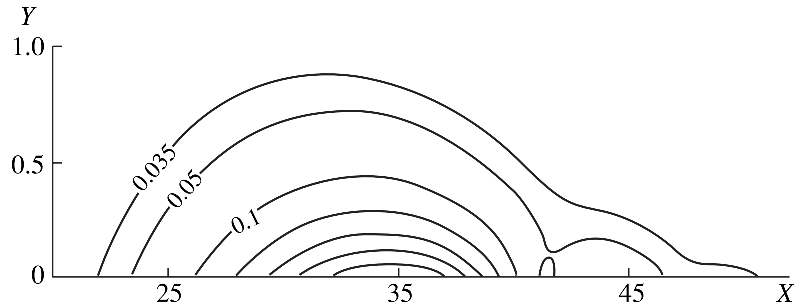
$$\bar{P}'_{\text{out}}(k, m, \omega, Y) = \bar{P}(k, m, \omega)\exp[-\lambda(k, m, \omega)Y], \quad \lambda = \sqrt{ik(\omega + ikc_\infty) + m^2}, \quad \text{Re}\lambda > 0.$$

Отсюда, учитывая только первый корень дисперсионного соотношения (2.3), сразу же получаем $P'_{\text{out}}(T, X, Y, Z) \sim \text{Re}(P_{\text{out},\,c}(T, X, Y, Z))$ с формулой

$$P_{\text{out},\,c} = \int\limits_{-\infty}^{\infty} dm \int\limits_{0}^{\infty} dk (ik)^{1/3} \frac{\Phi(\Omega_1)Q(\Omega_1;\, k, m)}{\dfrac{d\Phi}{d\Omega}(\Omega_1) - \dfrac{dQ}{d\Omega}(\Omega_1;\, k, m)} \exp[\omega_1(k, m)T + ikX + imZ - \lambda_1(k, m)[Y + Y_0]], \quad (3.2)$$



Фиг. 6.



Фиг. 7.

совпадающей с (3.1), если в ней заменить Y на $Y + Y_0$. Изолинии амплитуды давления $|P_{\text{out}, c}|$ в сечении плоскостями $X = X_{\max} = 34.7$ и $Z = Z_{\max} = 4.40$ представлены на фиг. 6 и фиг. 7 соответственно. Видно, что давление резко затухает с увеличением расстояния от стенки, уменьшаясь уже примерно в 10 раз при отходе от нее всего лишь на $Y = 0.75$.

В заключение зададимся вопросом, как удаленность изначального возмущения от стенки влияет на амплитуду генерируемого волнового пакета, т.е. как P' зависит от Y_0 при фиксированных T, X и Z. Из сопоставления формул (3.1) и (3.2) легко видеть, что

$$P'(T, X, Z; Y_0 + \Delta Y) = P'_{\text{out}}(T, X, Z, \Delta Y).$$

Следовательно, ответ на поставленный вопрос дают все те же фиг. 6 и 7, откуда заключаем, что амплитуда генерируемых волновых пакетов резко (практически по экспоненте) затухает с отдалением изначального возмущения от стенки.

ЗАКЛЮЧЕНИЕ

Таким образом, в рамках асимптотической теории показано, что упругость обтекаемой поверхности значительно ослабляет трехмерные волновые пакеты, генерируемые акустическими возмущениями. При этом не меняется характерная подковообразная форма пакета с двумя максимумами возмущений, распространяющимися под углом к внешнему течению.

СПИСОК ЛИТЕРАТУРЫ

1. *Нейланд В.Я.* К теории ламинарного пограничного слоя в сверхзвуковом потоке // Изв. АН СССР. Механ. жидкости и газа. 1969. № 4. С. 53–58.
2. *Stewartson K., Williams P.G.* Self-induced separation // Proc. Roy. Soc. A. 1969. V. 312. № 1509. P. 181–206.
3. *Messiter A.F.* Boundary-layer flow near the trailing edge of a flat plate // SIAM J. Appl. Math. 1970. V. 18. № 1. P. 241–257.

4. *Савенков И.В.* О преобразовании акустических возмущений в волны Толлмина–Шлихтинга в пограничном слое при трансзвуковых скоростях внешнего потока // Ж. вычисл. матем. и матем. физ. 2004. Т. 44. № 8. С. 1505–1510.
5. *Савенков И.В.* О пространственных эффектах при генерации трехмерных волн Толлмина–Шлихтинга в пограничном слое при трансзвуковых скоростях внешнего потока // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. № 10. С. 1886–1892.
6. *Kramer M.O.* Boundary-layer stabilization by disturbed damping // J. Aeronaut. Sci. 1957. V. 24. P. 459.
7. *Kramer M.O.* Boundary-layer stabilization by disturbed damping // J. Amer. Soc. Naval Engrs. 1960. V. 72. P. 25–33.
8. *Carpenter P.W., Garrad A.D.* The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 1. Tollmien-Schlichting instabilities // J. Fluid Mech. 1985. V. 155. P. 465–510.
9. *Carpenter P.W., Garrad A.D.* The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 1. Flow-induced surface instabilities // J. Fluid Mech. 1986. V. 170. P. 199–232.
10. *Савенков И.В.* О влиянии упругости обтекаемой поверхности на устойчивость пограничного слоя при трансзвуковых скоростях внешнего потока // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. № 1. С. 135–140.
11. *Савенков И.В.* О влиянии упругости обтекаемой поверхности на преобразование акустических возмущений в волны Толлмина–Шлихтинга в пограничном слое при трансзвуковых скоростях внешнего потока // Ж. вычисл. матем. и матем. физ. 2006. Т. 46. № 5. С. 948–954.
12. *Рыжов О.С., Савенков И.В.* Об устойчивости пограничного слоя при трансзвуковых скоростях внешнего потока // Ж. прикл. механ. и техн. физ. 1990. № 2. С. 65–71.
13. *Савенков И.В.* Трансзвуковое уравнение, описывающее распространение трехмерных солитонов в пограничном слое // Ж. вычисл. матем. и матем. физ. 2002. Т. 42. № 11. С. 1738–1743.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 538–546

УДК 519.712

О ПОСТРОЕНИИ ТУПИКОВЫХ ПОКРЫТИЙ ЦЕЛОЧИСЛЕННОЙ МАТРИЦЫ[1)]

© 2007 г. Е. А. Демьянов*, Е. В. Дюкова**

(* 119421 Москва, ул. Обручева, 26, корп. 2, Фонд "Общественное мнение";

** 119991 Москва, ул. Вавилова, 40, ВЦ РАН)

e-mail: egor_d13@mail.ru; djukova@ccas.ru

Поступила в редакцию 26.06.2006 г.

Получены новые оценки вычислительной сложности задачи построения тупиковых покрытий целочисленной матрицы (поиска максимальных конъюнкций логической функции специального вида). Библ. 7. Табл. 1.

Ключевые слова: дискретные процедуры распознавания и классификации, тупиковое покрытие целочисленной матрицы, асимптотически оптимальный алгоритм, метрические свойства множества покрытий, метрические свойства дизъюнктивных нормальных форм.

Комбинаторный (логический) анализ данных в задачах распознавания и классификации основан на использовании аппарата дискретной математики, в частности методов построения покрытий булевых и целочисленных матриц (нормальных форм логических функций), что оказывается сложным в силу чисто вычислительных трудностей переборного характера. При этом особую сложность представляет построение тупиковых покрытий целочисленной матрицы (специальной нормальной формы двузначной логической функции, заданной перечислением нулей). Поиски эффективных алгоритмов построения тупиковых покрытий ведутся с середины 1960-х годов.

В [1]–[4] рассмотрен случай, когда число строк m в матрице имеет более низкий порядок роста, чем число столбцов n, при условии, что $n \longrightarrow \infty$. Для этого случая построен асимптотически оптимальный алгоритм поиска тупиковых покрытий. Алгоритм работает с полиномиальной задержкой относительно m и n (число элементарных операций, выполняемых алгоритмом на каждом шаге, ограничено сверху полиномом от m и n) и число его шагов при $n \longrightarrow \infty$ почти всегда (для почти всех матриц данного размера) асимптотически равно числу тупиковых покрытий.

В настоящей работе рассмотрен случай, когда $n \le m$. Построен алгоритм поиска тупиковых покрытий целочисленной матрицы с элементами из $\{0, 1, \ldots, k-1\}$, $k \ge 2$, с полиномиальной задержкой на каждом шаге и такой, что логарифм по основанию k числа его шагов при $n \longrightarrow \infty$ почти всегда асимптотически равен логарифму по основанию k числа тупиковых покрытий. Обоснование алгоритма базируется на изучении метрических (количественных) свойств множества тупиковых покрытий и так называемых σ-подматриц целочисленной матрицы. Получены новые асимптотические оценки типичных значений числа тупиковых покрытий и длины тупикового покрытия и такие же оценки для количественных характеристик множества σ-подматриц. Аналогичные результаты приведены для задач построения нормальных форм двузначной логической функции, заданной перечислением нулевых точек.

1. МЕТРИЧЕСКИЕ СВОЙСТВА ТУПИКОВЫХ ПОКРЫТИЙ И σ-ПОДМАТРИЦ В СЛУЧАЕ $n \le m \le k^{n^\beta}$, $\beta < 1/2$

Пусть M_{mn}^k – множество всех матриц размера $m \times n$ с элементами из $\{0, 1, \ldots, k-1\}$, $k \ge 2$; E_k^r, $k \ge 2$, $r \le n$, – множество всех наборов вида $(\sigma_1, \ldots, \sigma_r)$, где $\sigma_i \in \{0, 1, \ldots, k-1\}$.

1) Работа выполнена при частичной финансовой поддержке РФФИ (код проекта 04-01-00795), гранта президента РФ по поддержке ведущих научных школ НШ № 5833.2006.1 и программы ОМН РАН "Алгебраические и комбинаторные методы математической кибернетики".

Рассмотрим $\sigma \in E_k^r$, $\sigma = (\sigma_1, \ldots, \sigma_r)$. Через $Q_p(\sigma)$, $p \in \{1, 2, \ldots, r\}$, обозначим множество всех наборов $(\beta_1, \ldots, \beta_r)$ в E_k^r таких, что $\beta_p \neq \sigma_p$ и $\beta_j = \sigma_j$ при $j \in \{1, 2, \ldots, r\} \backslash \{p\}$.

Пусть $L \in M_{mn}^k$. *Тупиковым σ-покрытием* матрицы L называется набор столбцов H этой матрицы такой, что подматрица L^H матрицы L, образованная столбцами набора H, обладает следующими двумя свойствами: 1) L^H не содержит строку σ; 2) если $p \in \{1, 2, \ldots, r\}$, то L^H содержит хотя бы одну строку из множества $Q_p(\sigma)$. Подматрица матрицы L, имеющая с точностью до перестановки строк вид

$$\begin{bmatrix} \beta_1 & \sigma_2 & \sigma_3 & \ldots & \sigma_{r-1} & \sigma_r \\ \sigma_1 & \beta_2 & \sigma_3 & \ldots & \sigma_{r-1} & \sigma_r \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \sigma_1 & \sigma_2 & \sigma_3 & \ldots & \sigma_{r-1} & \beta_r \end{bmatrix},$$

где $\beta_p \neq \sigma_p$ для $p = 1, 2, \ldots, r$, называется *σ-подматрицей.*

Таким образом, H является тупиковым σ-покрытием тогда и только тогда, когда L^H не содержит строку σ и содержит σ-подматрицу.

Понятие тупикового $(0, \ldots, 0)$-покрытия булевой матрицы совпадает с хорошо известным понятием неприводимого покрытия булевой матрицы. Отметим, что $(0, \ldots, 0)$-подматрица булевой матрицы является единичной подматрицей.

Положим, что $S(L, \sigma)$ – множество всех σ-подматриц матрицы L, $B(L, \sigma)$ – множество всех тупиковых σ-покрытий матрицы L. Пусть, далее,

$$S_r(L) = \bigcup_{\sigma \in E_k^r} S(L, \sigma), \quad B_r(L) = \bigcup_{\sigma \in E_k^r} B(L, \sigma),$$

$$S(L) = \bigcup_{r=1}^{n} S_r(L), \quad B(L) = \bigcup_{r=1}^{n} S_r(L),$$

$r_1 = [\log_k m - \log_k \ln \log_k m - 1]$, $|V|$ – мощность множества V.

Пусть $n \le m \le k^{n^\beta}$, $\beta < 1/2$. В рассматриваемом случае получена асимптотика логарифма типичного числа подматриц из $S(L)$ с порядком не меньшим r_1 (при $n \longrightarrow \infty$). Показано, что эта асимптотика совпадает с асимптотикой логарифма типичного числа покрытий из $B(L)$ и совпадает с асимптотикой логарифма типичного числа тех покрытий из $B(L)$, у которых длины не меньше r_1. Получена оценка типичной длины покрытия из $B(L)$. Указанные оценки приведены в сформулированных ниже теоремах 1–3.

Замечание 1. Если в матрице L из M_{mn}^k есть две одинаковые строки, то, очевидно, при удалении одной из них множество $B(L)$ не меняется. Следовательно, при подсчете числа тупиковых покрытий матрицы L имеет смысл рассматривать случай, когда в L нет одинаковых строк. При $m \le k^{n^\beta}$, $\beta < 1/2$, требуемым свойством обладают почти все матрицы M_{mn}^k (см. ниже утверждение 4, разд. 3).

Теорема 1. *Если $n \le m \le k^{n^\beta}$, $\beta < 1/2$, то при $n \longrightarrow \infty$ для почти всех матриц L из M_{mn}^k логарифм по основанию k числа всех подматриц из $S(L)$ с порядком не меньшим r_1 асимптотически равен при $n \longrightarrow \infty$ логарифму по основанию k числа всех покрытий из $B(L)$ с длиной не меньшей r_1 и асимптотически равен* $\log_k C_n^{r_1} + r_1$.

Теорема 2. *Если $n \le m \le k^{n^\beta}$, $\beta < 1/2$, то при $n \longrightarrow \infty$ для почти всех матриц L из M_{mn}^k логарифм по основанию k числа всех покрытий из $B(L)$ с длиной не меньшей r_1 асимптотически равен логарифму по основанию k числа всех покрытий из $B(L)$ и асимптотически равен* $\log_k C_n^{r_1} + r_1$.

Теорема 3. *Если* $n \le m \le k^{n^\beta}$, $\beta < 1/2$, *то при* $n \longrightarrow \infty$ *у почти всех матриц* L *из* M^k_{mn} *длины почти всех покрытий из* $B(L)$ *принадлежат интервалу* $[r_1, \log_k mn]$.

Доказательства теорем 1–3 опираются на леммы 1–6, приведенные далее. При доказательстве лемм 1–6 использовались результаты работ [5] и [6].

Будем считать $M^k_{mn} = \{L\}$ пространством элементарных событий, в котором каждое событие L происходит с вероятностью $1/|M^k_{mn}|$. Математическое ожидание случайной величины $X(L)$, определенной на множестве M^k_{mn}, будем обозначать через $\mathbf{M}X(L)$, дисперсию – через $\mathbf{D}X(L)$.

Легко доказывается

Лемма 1. *Пусть* $X(L) \ge 0$, $\theta > 0$, $\nu_\theta(n)$ – *доля тех матриц* L *из* M^k_{mn}, *для которых* $X(L) \ge \theta\mathbf{M}X(L)$. *Тогда справедливо неравенство* $\nu_\theta(n) \le 1/\theta$.

Далее записи $a_n \approx b_n$, $n \longrightarrow \infty$, и $a_n \le_n b_n$, $n \longrightarrow \infty$, соответственно означают, что $\lim a_n/b_n = 1$, $n \longrightarrow \infty$, и $\lim a_n/b_n \le 1$, $n \longrightarrow \infty$.

Пусть $a_r = C^r_n C^r_m r!(k-1)^r k^{r-r^2}$. Так как $a_{r-1} = \bar{o}(a_r)$ при $n \le m$, $r \le r_1$, то справедлива

Лемма 2. *При* $n \le m$ *имеет место соотношение*

$$\sum_{r \ge r_1} C^r_n C^r_m r!(k-1)^r k^{r-r^2} \approx C^{r_1}_n C^{r_1}_m r_1!(k-1)^{r_1} k^{r_1 - r_1^2}, \quad n \longrightarrow \infty.$$

Лемма 3. *При* $m \le k^{n^\beta}$, $\beta < 1/2$, *имеет место соотношение*

$$\log_k(C^{r_1}_m r_1!(k-1)^{r_1} k^{-r_1^2}) = o(\log_k(C^{r_1}_n k^{r_1})), \quad n \longrightarrow \infty.$$

Доказательство. В справедливости леммы можно убедиться непосредственной проверкой. Действительно, из очевидного неравенства $C^r_n \ge ((n-r)/r)^r$ следует, что

$$\log_k C^{r_1}_n \ge_n (1-\beta) r_1 \log_k n, \quad n \longrightarrow \infty. \tag{1}$$

С другой стороны,

$$b = C^{r_1}_m r_1!(k-1)^{r_1} k^{-r_1^2} \le (mk)^{r_1}/k^{r_1} \le (k^2 \ln\log_k m)^{r_1}.$$

Следовательно, имеем

$$\log_k b \le_n r_1 \log_k \ln n, \quad n \longrightarrow \infty. \tag{2}$$

Из (1) и (2) следует утверждение леммы 3.

На $M^k_{mn} = \{L\}$ рассмотрим случайные величины $\eta_r(L) = |B_r(L)|$, $\zeta_r(L) = |S_r(L)|$. Нетрудно подсчитать (см. [5]), что имеет место формула

$$\mathbf{M}\zeta_r(L) = C^r_n C^r_m r!(k-1)^r k^{r-r^2}.$$

Лемма 4. *При* $n \le m \le k^{n^\beta}$, $\beta < 1/2$, *имеет место соотношениее*

$$\log_k \sum_{r \ge r_1} \eta_r(L) \le \log_k \sum_{r \ge r_1} \zeta_r(L) \le_n \log_k C^{r_1}_n + r_1, \quad n \longrightarrow \infty.$$

Доказательство. Из лемм 2 и 3 сразу следует, что

$$\log_k \sum_{r \ge r_1} \mathbf{M}\eta_r(L) \le \log_k \sum_{r \ge r_1} \mathbf{M}\zeta_r(L) \approx \log_k C^{r_1}_n + r_1, \quad n \longrightarrow \infty.$$

Отсюда, применяя лемму 1 с $\theta = \log_k \log_k n$, получаем утверждение леммы 4.

Лемма 5. *При* $m \le k^{n^\beta}$, $\beta < 1/2$, *имеет место соотношение*

$$\log_k \sum_{r \ge r_1} \zeta_r(L) \ge \log_k \sum_{r \ge r_1} \eta_r(L) \ge_n \log_k C_n^{r_1} + r_1, \quad n \longrightarrow \infty.$$

Доказательство. В [6] показано, что

$$\eta_{r_1}(L) \approx C_n^{r_1} 2^{r_1} (1 - k^{-r_1})^m, \quad n \longrightarrow \infty.$$

Оценим $(1 - k^{-r_1})^m$:

$$(1 - k^{-r_1})^m \ge \exp(-2m/k^{r_1}) \ge (\log_k m)^{-2k}.$$

Отсюда, учитывая, что $\log_k \log_k m = \bar{o}(\log_k C_n^{r_1})$, получаем утверждение леммы.

Из лемм 4 и 5 следует утверждение теоремы 1.

Лемма 6. *При* $n \le m \le k^{n^\beta}$, $\beta < 1/2$, *имеет место соотношение*

$$\sum_{r < r_1} \eta_r(L) = o\left(\sum_{r \ge r_1} \eta_r(L)\right), \quad n \longrightarrow \infty.$$

Доказательство. В [6] показано, что при $m \le k^{n^\beta}$, $\beta < 1/2$,

$$\sum_{r < r_1} \eta_r(L) = o(C_n^{r_1} k^{r_1}), \quad n \longrightarrow \infty.$$

С другой стороны, согласно теореме 1, имеем

$$\sum_{r \ge r_1} \eta_r(L) = (C_n^{r_1} k^{r_1})^{1 + \delta(n)},$$

где $\delta(n) \ge 0$, $\delta(n) \longrightarrow 0$ при $n \longrightarrow \infty$.

Следовательно,

$$\sum_{r < r_1} \eta_r(L) = o\left(\sum_{r \ge r_1} \eta_r(L)\right), \quad n \longrightarrow \infty.$$

Отсюда получаем утверждение доказываемой леммы.

Из теоремы 1 и леммы 6 следует утверждение теоремы 2. Из леммы 6 следует утверждение теоремы 3.

Замечание 2. Оценка для $|B(L)|$, приведенная в теореме 2, ранее была получена более сложным способом в [7].

2. ПОСТРОЕНИЕ ТУПИКОВЫХ ПОКРЫТИЙ В СЛУЧАЕ $n \le m \le k^{n^\gamma}$, $\gamma < 1/(2k + 1)$

Пусть $G(L)$ – конечная совокупность наборов столбцов матрицы L, содержащая $B(L)$. Предполагается, что каждый набор из $G(L)$ не содержит одинаковых столбцов и некоторые наборы столбцов в $G(L)$ могут встречаться более одного раза.

Пусть алгоритм A строит покрытия из $B(L)$ путем последовательного просмотра всех наборов из $G(L)$. При этом каждый набор из $G(L)$ просматривается столько раз, сколько раз он встречается в $G(L)$. Таким образом, на каждом шаге алгоритма A строится некоторый набор столбцов H из $G(L)$ и проверяется принадлежность H к $B(L)$. Число шагов алгоритма A обозначим через $N_A(L)$.

Нас будет интересовать вычислительная сложность алгоритма A в типичном случае (для почти всех булевых матриц размера $m \times n$ при $n \longrightarrow \infty$).

Будем говорить, что алгоритм A строит $G(L)$ с полиномиальной задержкой, если на каждом шаге выполняется не более $d(m, n)$ элементарных операций и $d(m, n)$ ограничено сверху полиномом от m, n. При этом под элементарной операцией понимается просмотр одного элемента матрицы L.

Алгоритм А назовем *асимптотически оптимальным*, если он строит $G(L)$ с полиномиальной задержкой и для почти всех матриц L из M_{mn} при $n \longrightarrow \infty$ величина $N_A(L)$ асимптотически равна числу покрытий из $B(L)$.

Как уже было отмечено выше, в [1]–[4] был построен асимптотически оптимальный алгоритм поиска тупиковых покрытий целочисленной матрицы (для случая, когда число строк в матрице существенно меньше числа столбцов, а именно когда $m^\alpha \le n \le k^{m^\beta}$, $\alpha > 1$, $\beta < 1$). Алгоритм основан на переборе с полиномиальной задержкой $O(mn)$ наборов столбцов матрицы L, порождаемых подматрицами из $S(L)$, и имеет повторяющиеся шаги (некоторые наборы столбцов строятся неоднократно). Повторения возникают из-за того, что набор столбцов матрицы L длины r может содержать несколько одинаковых σ-подматриц порядка r. Указанный недостаток не является существенным, если число строк в матрице L достаточно мало́ по сравнению с числом столбцов. В этом случае число шагов алгоритма, равное $|S(L)|$, почти всегда при $n \longrightarrow \infty$ асимптотически равно $|B(L)|$.

Ситуация меняется, когда m начинает расти, а n остается постоянным. Повторений становится слишком много. Особенно много их среди тех шагов, на которых строятся σ-подматрицы с небольшим порядком. Такие подматрицы порождают мало тупиковых покрытий (их число существенно меньше числа всех тупиковых покрытий). Алгоритм становится неэффективным. В данном случае представляется целесообразным организовать процедуру поиска тупиковых покрытий следующим образом.

(i) Просматривать все “короткие” наборы столбцов (их длины не должны превосходить некоторого порога p), отбирая среди них тупиковые покрытия и строя одновременно все те σ-подматрицы, порядки которых равны p и которые не порождают тупиковые покрытия. (ii) Построив σ-подматрицу с указанными свойствами, достраивать ее, применяя алгоритм из [1]–[4], до подматриц из $S(L)$, порождающих тупиковые покрытия.

Данная схема легла в основу предлагаемого в настоящей работе асимптотически оптимального алгоритма. В качестве порога p взято число $r_1 - 1$. При обосновании асимптотической оптимальности алгоритма использованы оценки, приведенные в теоремах 1 и 2.

Дадим описание работы алгоритма в случае поиска тупиковых покрытий целочисленной матрицы $L = (a_{ij})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$.

Пусть набор H состоит из столбцов матрицы L с номерами $j_1, \ldots, j_r$ и $j_1 < \ldots < j_r$. Набор $\{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$ элементов матрицы L назовем *совместимым* относительно σ, $\sigma \in E_k^r$, если в подматрице L^H строка с номером i_p принадлежит $Q_p(\sigma)$ при $p \in \{1, 2, \ldots, r\}$.

Обозначим через $Q(L)$ совокупность всех совместимых наборов из элементов матрицы L.

Каждому элементу a_{ij} в L присвоим номер $N[i, j] = (j - 1)m + i$.

Пусть $R(L)$ – множество всех элементов в L. Если $R \subseteq R(L)$, то через $e_1(R)$ и $e_2(R)$ будем обозначать, соответственно, элементы с наименьшим и наибольшим номерами в R.

Пусть $Q \in Q(L)$, $Q = \{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$, Q совместим относительно σ, $\sigma \in E_k^r$. Будем считать, что $N[i_{t+1}, j_{t+1}] > N[i_t, j_t]$ при $t = 1, 2, \ldots, r - 1$. Число $N[i_1, j_1] + \ldots + N[i_r, j_r]$ назовем *весом* набора Q. Нетрудно видеть, что набор Q определяет σ-подматрицу матрицы L. Набор Q назовем *верхним*, если из того, что набор $\{a_{p_1 j_1}, \ldots, a_{p_r j_r}\}$ является совместимым относительно σ, следует, что $i_u \le p_u$ при $u = 1, 2, \ldots, r$.

Пусть $Q \in Q(L)$, $Q = \{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$, $t \in \{1, 2, \ldots, r\}$. Обозначим через $G_t(Q)$ совокупность всех элементов матрицы L, построенную следующим образом. Для каждого ν, $\nu \in \{1, 2, \ldots, t\}$, из L вычеркивается столбец с номером j_ν и строки, дающие $a_{i_\nu j_\nu}$ в пересечении со столбцом с номером j_ν. Из каждого не вычеркнутого столбца с номером j удаляются элементы, совпадающие хотя бы с одним из элементов $a_{i_1 j}, \ldots, a_{i_t j}$. Тогда $G_t(Q)$ – совокупность всех не вычеркнутых элементов матрицы L. Нетрудно видеть, что $\{a_{i_1 j_1}, \ldots, a_{i_t j_t}, a_{ij}\} \in Q(L)$ тогда и только тогда, когда $a_{ij} \in G_t(Q)$. Обозначим через $R_t(Q)$ множество элементов в $R(L)$, номера которых больше $N[i_t, j_t]$, через $\lambda_t(Q)$ – число вычеркнутых строк матрицы L. Если $\lambda_r(Q) = m$, то, очевидно, набор столбцов с номерами $j_1, \ldots, j_r$ принадлежит $B(L)$.

Пусть $\pi(L)$ – множество всех пар вида (H, σ), где H – набор столбцов матрицы L длины r, $r \le r_1 - 1$, и $\sigma \in E^r$.

Символом $\geq$ обозначим отношения лексикографического порядка для слов в алфавитах $\{1, 2, \ldots, n\}$ и $\{0, 1\}$. Пусть (H, σ) и (H', σ') – две различные пары из $\pi(L)$, и пусть H состоит из столбцов с номерами $j_1, \ldots, j_r, j_1 < \ldots < j_r$, H' состоит из столбцов с номерами $j'_1, \ldots, j'_t$, $j'_1 < \ldots < j'_t$, $\sigma = (\sigma_1, \ldots, \sigma_r)$, $\sigma' = (\sigma'_1, \ldots, \sigma'_t)$. Запись $(H', \sigma') > (H, \sigma)$ означает, что либо $H' \neq H$, $j'_1 \ldots j'_t \geq j_1 \ldots j_r$, либо $H' = H$, $\sigma' \neq \sigma$, $\sigma'_1 \ldots \sigma'_t \geq \sigma_1 \ldots \sigma_r$.

Будем говорить, что (H', σ') следует за (H, σ) в $\pi(L)$, если $(H', \sigma') > (H, \sigma)$ и не существует в $\pi(L)$ пары (H'', σ'') такой, что $(H', \sigma') > (H'', \sigma'') > (H, \sigma)$.

Через $\min\pi(L)$ и $\max\pi(L)$ обозначим, соответственно, минимальный и максимальный элементы в $\pi(L)$. Очевидно, $\min\pi(L) = \{\{1\}, (0)\}$, $\max\pi(L) = \{\{n\}, (1)\}$.

Пусть $p = r_1 - 1$, $(H, \sigma) \in \pi(L)$ и H состоит из столбцов с номерами $j_1, \ldots, j_p, j_1 < \ldots < j_p$.

Через $P_0(H, \sigma)$ обозначим совокупность наборов Q из $Q(L)$ вида $\{a_{i_1 j_1}, \ldots, a_{i_p j_p}\}$ таких, что набор Q совместим относительно σ и $G_p(Q) \cap R_p(Q) \neq \emptyset$. Наборы из $P_0(H, \sigma)$, имеющие наименьший и наибольший веса, обозначим, соответственно, через $Q_{\min}(H)$ и $Q_{\max}(H)$.

Через $P(H, \sigma)$ обозначим совокупность всех наборов из $Q(L)$ вида $\{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$ таких, что $r \geq r_1$ и $\{a_{i_1 j_1}, \ldots, a_{i_p j_p}\} \in P_0(H, \sigma)$.

Положим $\min P(H, \sigma) = Q_1 \cup \{e_1(G_p(Q_1) \cap R_p(Q_1))\}$, $Q_1 = Q_{\min}(H)$, и $\max P(H, \sigma) = Q_2 \cup \{e_2(G_p(Q_2) \cap \cap R_p(Q_2))\}$, $Q_2 = Q_{\max}(H)$.

Упорядочим $P(H, \sigma)$, $P(H, \sigma) \neq \emptyset$. Для каждого Q из $P(H, \sigma)$, $Q = \{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$, $Q \neq \max P(H, \sigma)$, построим следующий за ним набор $>Q$. Пусть Q' – набор из $P_0(L, H)$, ближайший по весу к набору $\{a_{i_1 j_1}, \ldots, a_{i_p j_p}\}$. Перечислим следующие возможные случаи:

1) $G_r(Q) \cap R_r(Q) \neq \emptyset$; тогда $>Q = Q \cup \{e_1(G_r(Q) \cap R_r(Q))\}$;

2) $G_r(Q) \cap R_r(Q) = \emptyset$;

а) $G_{r-1}(Q) \cap R_r(Q) \neq \emptyset$; тогда $>Q = Q\backslash\{a_{i_r j_r}\} \cup \{e_1(G_{r-1}(Q) \cap R_r(Q))\}$;

б) $r \geq r_1$ и $G_{r-1}(Q) \cap R_r(Q) = \emptyset$; тогда если $r = r_1$, то $>Q = Q' \cup \{e_1(G_p(Q') \cap R_p(Q'))\}$, если же $r > r_1$, то $>Q = (Q\backslash\{a_{i_{r-1} j_{r-1}}, a_{i_r j_r}\}) \cup \{e_1(G_{r-2}(Q) \cap R_{r-1}(Q))\}$.

Заметим, что $\{e_1(G_{r-2}(Q) \cap R_{r-1}(Q))\} \neq \emptyset$ при $r > r_1$, так как $a_{i_r j_r} \in \{e_1(G_{r-2}(Q) \cap R_{r-1}(Q))\}$.

Приведем схему работы алгоритма.

В установленном порядке, начиная с $\min\pi(L)$, просматриваются пары из $\pi(L)$. Пусть на шаге i рассматривается пара (H_i, σ_i). Возможны следующие варианты: а) $|H_i| < r_1 - 1$; б) $|H_i| = r_1 - 1$, $P(H_i, \sigma_i) = \emptyset$; в) $|H_i| = r_1 - 1$, $P(H_i, \sigma_i) \neq \emptyset$.

В вариантах а) и б) сначала проверяется условие $H_i \notin B(L, \sigma_i)$. Если это условие выполнено, то набор H_i заносится в искомое множество тупиковых покрытий. Далее алгоритм проверяет условие $(H_i, \sigma_i) \neq \max\pi(L)$. Если указанное условие выполнено, алгоритм переходит к шагу $i + 1$, на котором рассматривается следующая за (H_i, σ_i) пара из $\pi(L)$, если же $(H_i, \sigma_i) = \max\pi(L)$, то алгоритм заканчивает работу.

Если имеет место вариант в), то, очевидно, $H_i \notin B(L, \sigma_i)$. В этом случае на каждом из последующих $|P(H_i, \sigma_i)|$ шагов строится с помощью алгоритма из [1]–[4] набор из $P(H_i, \sigma_i)$. Затем проверяется условие $(H_i, \sigma_i) \neq \max\pi(L)$. Если указанное условие выполнено, то алгоритм переходит к шагу $i + |P(H_i, \sigma_i)| + 1$, на котором рассматривается следующая за (H_i, σ_i) пара из $\pi(L)$. В противном случае алгоритм заканчивает работу.

Пусть имеет место $|H_i| = r_1 - 1$ и $P(H_i, \sigma_i) \neq \emptyset$, и пусть на шаге $i + j$, $1 \le j \le |P(H_i, \sigma_i)|$, построен набор $Q[i + j]$ из $P(H_i, \sigma_i)$ вида $\{a_{i_1 j_1}, \ldots, a_{i_r j_r}\}$, совместимый относительно некоторого σ' из E_k^r.

Если набор $Q[i + j]$ является верхним и подматрица, образованная столбцами с номерами $j_1, \ldots, j_r$, не содержит строку σ', то набор столбцов с номерами $j_1, \ldots, j_r$ заносится в искомое множество тупиковых покрытий. В противном случае множество тупиковых покрытий, построенное на предыдущих шагах, не пополняется. Построение наборов из $P(H_i, \sigma_i)$ происходит по следующим правилам:

1) $Q[i + 1] = \min P(H, \sigma)$;

2) если $Q[i + j] \neq \max P(H, \sigma)$, то $Q[i + j + 1] = \succ Q[i + j]$;

3) если $Q[i + j] = \max P(H, \sigma)$, то алгоритм переходит к шагу $i + j + 1$, на котором рассматривается следующая за (H_i, σ_i) пара из $\pi(L)$.

Таким образом, алгоритм строит $B(L)$ за $N(L) = N_1(L) + N_2(L)$ шагов, где $N_1(L) = \sum_{r < r_1} k^r C_n^r$, $N_2(L) = |P(L)|$, $P(L)$ – объединение множеств $P(H, \sigma)$ по всем парам (H, σ) из $\pi(L)$. Нетрудно видеть, что $N_2(L)$ – число всех подматриц из $S(L)$ с порядком, не меньшим r_1.

Нетрудно подсчитать, что задержка алгоритма не превосходит $O(mn^2)$.

Теорема 4. *Если* $n \leq m \leq k^{n^\gamma}$, $\gamma < 1/(2k + 1)$, *то при* $n \longrightarrow \infty$ *для почти всех матриц* L *из* M_{mn}^k *имеет место соотношение*

$$\log_k N(L) \approx \log_k N_2(L) \approx \log_k |B(L)| \approx \log_k C_n^{r_1} + r_1.$$

Доказательство. Пусть $n \leq m \leq k^{n^\gamma}$, $\gamma < 1/(2k + 1)$. При $r < r_1$ имеем $C_n^r k^r/(C_n^{r+1} k^{r+1}) < 1/(kn^{1-\gamma})$. Следовательно, $N_1(L) \leq_n C_n^{r_1} k^{r_1}/(kn^{1-\gamma})$, $n \longrightarrow \infty$. С другой стороны, в [7] показано, что при указанных ограничениях на m и n для почти всех матриц L из M_{mn}^k имеет место $|B(L)| \geq_n \log_k^{-2k} m C_n^{r_1} k^{r_1}$, $n \longrightarrow \infty$. Следовательно, при $n \longrightarrow \infty$ для почти всех матриц L из M_{mn}^k справедливо $N_1(L) = o(|B(L)|)$ и $N(L) \approx N_2(L)$. Отсюда в силу теорем 1 и 2 получаем утверждение доказываемой теоремы.

Из приведенных рассуждений следует, что в типичном случае описанный нами алгоритм является логарифмически оптимальным.

3. СЛОЖНОСТЬ ПОСТРОЕНИЯ МАКСИМАЛЬНЫХ КОНЪЮНКЦИЙ ЛОГИЧЕСКОЙ ФУНКЦИИ

Оценки, полученные в теоремах 1–3, могут быть использованы для оценки типичных значений аналогичных количественных характеристик множеств максимальных и неприводимых конъюнкций двузначной логической функции специального вида.

Действительно, пусть функция $F(x_1, \ldots, x_n\}$, определена на E_k^n, принимает значения из $\{0, 1\}$, B_F – множество наборов, на которых $F = 0$, $|B_F| = m$. Обозначим через N_{mn}^k множество всех таких функций. Положим

$$x^\sigma = \begin{cases} 1, & \text{если } x = \sigma, \\ 0, & \text{если } x \neq \sigma, \end{cases}$$

$$x, \sigma \in \{0, 1, \ldots, k - 1\}.$$

Обычным образом введем понятие элементарной конъюнкции. Элементарной конъюнкцией (ЭК) над переменными $x_1, \ldots, x_n$ назовем выражение вида $x_{j_1}^{\sigma_1} \cdot \ldots \cdot x_{j_r}^{\sigma_r}$, где $x_{j_i} \in \{x_1, \ldots, x_n\}$ при $i = 1, 2, \ldots, r$ и $x_{j_q} \neq x_{j_t}$ при $t, q \in \{1, 2, \ldots, r\}$, $t \neq q$. ЭК принимает значение 1 тогда и только тогда, когда каждый ее множитель равен 1.

Пусть N_B – интервал истинности ЭК B. ЭК B назовем *допустимой* для F, если $N_B \cap B_F = \varnothing$. ЭК B назовем *неприводимой* для F, если не существует допустимой ЭК B' такой, что $N_{B'} \supset N_B$. ЭК B назовем *максимальной* для F, если она является допустимой и не существует допустимой ЭК B' такой, что $N_{B'} \supset N_B$.

Из приведенных определений следует, что ЭК является максимальной для F тогда и только тогда, когда она является допустимой и неприводимой.

Пусть L_F – матрица, строками которой являются наборы из B_F. Очевидными являются приводимые ниже утверждения 1–3.

Утверждение 1. *ЭК* $x_{j_1}^{\sigma_1}\dots x_{j_r}^{\sigma_r}$ *является допустимой для F тогда и только тогда, когда набор столбцов матрицы L с номерами* $j_1, \dots, j_r$ *является* $(\sigma_1, \dots, \sigma_r)$*-покрытием.*

Утверждение 2. *ЭК* $x_{j_1}^{\sigma_1}\dots x_{j_r}^{\sigma_r}$ *является неприводимой для F тогда и только тогда, когда набор столбцов матрицы L с номерами* $j_1, \dots, j_r$ *содержит* $(\sigma_1, \dots, \sigma_r)$*-подматрицу.*

Утверждение 3. *ЭК* $x_{j_1}^{\sigma_1}\dots x_{j_r}^{\sigma_r}$ *является максимальной для F тогда и только тогда, когда набор столбцов матрицы L с номерами* $j_1, \dots, j_r$ *является тупиковым* $(\sigma_1, \dots, \sigma_r)$*-покрытием.*

Нетрудно также доказывается

Утверждение 4. *Если* $m^2 = o(k^n)$, *то почти все матрицы из* M_{mn}^k *при* $n \longrightarrow \infty$ *являются матрицами с попарно различными строками.*

Пусть $n \le m \le k^{n^\beta}$, $\beta < 1/2$. Из теорем 1–3 и утверждений 1–4 следуют приводимые ниже теоремы 4–6.

Теорема 4. *При* $n \longrightarrow \infty$ *для почти всех функций F из* N_{mn}^k *логарифм по основанию k числа всех неприводимых конъюнкций с рангом не меньшим* r_1 *асимптотически равен логарифму по основанию k числа всех максимальных конъюнкций с рангом не меньшим* r_1 *и асимптотически равен* $\log_k C_n^{r_1} + r_1$.

Теорема 5. *При* $n \longrightarrow \infty$ *для почти всех функций F из* N_{mn}^k *логарифм по основанию k числа всех максимальных конъюнкций с рангом не меньшим* r_1 *асимптотически равен логарифму по основанию k числа всех максимальных конъюнкций и асимптотически равен* $\log_k C_n^{r_1} + r_1$.

Теорема 6. *При* $n \longrightarrow \infty$ *у почти всех функций F из* N_{mn}^k *ранги почти максимальных конъюнкций принадлежат интервалу* $[r_1, \log_k mn]$.

Модифицируя алгоритм, описанный в разд. 2, для построения максимальных конъюнкций функции F из N_{mn}^k, получаем в случае $n \le m \le k^{n^\gamma}$, $\gamma < 1/(2k + 1)$, алгоритм с полиномиальной задержкой $O(mn^2)$ и такой, что при $n \longrightarrow \infty$ логарифм по основанию k числа его шагов почти всегда асимптотически равен логарифму по основанию k числа максимальных конъюнкций функции F.

4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Для оценки практической применимости построенного в настоящей работе алгоритма были проведены две серии экспериментов на случайных матрицах разного размера. Данный алгоритм, обозначаемый далее как COMB, был модифицирован для поиска неприводимых покрытий (тупиковых (0, …, 0)-покрытий) булевой матрицы L. Было проведено его сравнение с алгоритмом из [1]–[4] и алгоритмом из [3], обозначаемыми далее, соответственно, через АО1 и АО2. Как уже было отмечено в разд. 2, алгоритм АО1 основан на переборе с полиномиальной задержкой $O(mn)$ наборов столбцов матрицы L, порождаемых единичными подматрицами. Алгоритм АО2 основан на переборе с полиномиальной задержкой $O(m^3n)$ наборов столбцов, образующих неприводимые покрытия, и имеет повторяющиеся шаги.

Результаты экспериментов приведены в таблице. Для каждого алгоритма указано среднее время счета. Наилучшее время счета алгоритма COMB, приведенное в таблице, достигается не при теоретической оценке порога $p = r_1 - 1$, а при некотором пороге d, незначительно отличающемся от p. Для алгоритма COMB также указано $N_1(L)$ – число шагов, выполненных на первом этапе работы алгоритма, $N_2(L)$ – число шагов, выполненных на втором этапе работы алгоритма, $V_d(L)$ – число единичных подматриц с порядком, не превосходящим d, $|P_1(L)|$ – число неприводимых покрытий длины, не превосходящей d, $|P_2(L)|$ – число неприводимых покрытий длины, превосходящей d. Также в таблице приведено число шагов алгоритма АО1, равное $|S(L)|$.

Первая серия экспериментов проводилась на случайных матрицах размера 50×50 с равновероятным появлением нулей и единиц. Выборка состояла из 15 матриц. Среднее время поиска

Таблица

Размер матриц	COMB						AO1		AO2
	время счета	$V_d(L)$	$N_1(L)$	$N_2(L)$	$\|P_1(L)\|$	$\|P_2(L)\|$	время счета	$\|S(L)\|$	время счета
50×50	67.8	34.4 млн.	2516853	1833604	156336	157561	104	36.2 млн.	60
100×30	16.4	79.91 млн.	2956206	670	46866	173	–	79.96 млн.	28

всех неприводимых покрытий алгоритмом AO1 составило 104 с. Алгоритмы AO2 и COMB имеют значительно меньше повторяющихся шагов и требуют меньших временны́х затрат. Алгоритм AO2 на тех же случайных матрицах показал среднее время счета 60 с, а алгоритм COMB показал 68 с. Указанное для алгоритма COMB время достигается при пороге $d = r_1 + 2$. При этом на первом и втором этапе работы алгоритма было построено примерно одинаковое число покрытий ($|P_1(L)| \approx |P_2(L)|$), а $V_d(L)$ оказалось равным 34.4 млн. (при $S(L)$ примерно равном 36 млн.).

Вторая серия экспериментов проводилась на случайных матрицах размера 100×30 с равновероятным появлением нулей и единиц. Выборка состояла из 15 матриц. Среднее время поиска неприводимых покрытий алгоритмом AO2 составило 28 с. Алгоритм COMB на тех же матрицах показал результат 16 с при пороге $d = r_1 + 3$. При этом почти все неприводимые покрытия ($|P_1(L)| \approx$ ≈ 47000) были построены на первом этапе работы алгоритма и лишь малая часть ($|P_2(L)| \approx 170$) – на втором. Почти все единичные подматрицы (около 80 млн.) имели порядок меньше либо равный d. Алгоритм AO1 в данной серии не исследовался ввиду слишком большого времени счета.

Результаты экспериментов подтверждают, что предлагаемый в работе алгоритм является эффективным для применения на практике. На квадратных матрицах он оказался значительно эффективнее алгоритма AO1 и незначительно уступает алгоритму AO2. На вертикально вытянутых матрицах время работы предлагаемого алгоритма оказалось почти в 2 раза меньше, чем время работы алгоритма AO2.

Также результаты экспериментов подтверждают, что в случае $m < n$ основной причиной эффективности предлагаемого алгоритма по сравнению с AO1 и AO2 является отказ от просмотра большого числа единичных подматриц небольшого порядка, которые либо не порождают покрытий, либо порождают повторяющиеся неприводимые покрытия. Число таких подматриц составляет десятки миллионов, в то время как число соответствующих им покрытий – лишь десятки тысяч.

СПИСОК ЛИТЕРАТУРЫ

1. *Дюкова Е.В.* О сложности реализации некоторых процедур распознавания // Ж. вычисл. матем. и матем. физ. 1987. Т. 27. № 1. С. 114–127.
2. *Дюкова Е.В.* Алгоритмы распознавания типа “Кора”: сложность реализации и метрические свойства // Распознавание, классификация, прогноз (матем. методы и их применение). М.: Наука, 1989. Вып. 2. С. 99–125.
3. *Дюкова Е.В.* О сложности реализации дискретных (логических) процедур распознавания // Ж. вычисл. матем. и матем. физ. 2004. Т. 44. № 3. С. 562–572.
4. *Дюкова Е.В., Журавлёв Ю.И.* Дискретный анализ признаковых описаний в задачах распознавания большой размерности // Ж. вычисл. матем. и матем. физ. 2000. Т. 40. №8. С. 1264–1278.
5. *Дюкова Е.В., Песков Н.В.* Поиск информативных фрагментов описаний объектов в дискретных процедурах распознавания // Ж. вычисл. матем. и матем. физ. 2002. № 5. Т. 42. С. 741–753.
6. *Дюкова Е.В., Инякин А.С.* О процедурах классификации, основанных на построении покрытий классов // Ж. вычисл. матем. и матем. физ. 2003. Т. 43. № 12. С. 1884–1895.
7. *Дюкова Е.В.* О числе тупиковых покрытий целочисленной матрицы // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. № 5. С. 935–940.

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2007, том 47, № 3, с. 547–552

УДК 519.714

ПРЕДИКАТНОЕ ЗАДАНИЕ УНИВЕРСАЛЬНЫХ ОГРАНИЧЕНИЙ В АЛГЕБРАИЧЕСКОМ ПОДХОДЕ К ЗАДАЧАМ РАСПОЗНАВАНИЯ

© 2007 г. Р. С. Таханов
(119991 Москва, у. Вавилова, 40, ВЦ РАН)
e-mail: takhanov@mail.ru

Поступила в редакцию 09.06.2005 г.
Переработанный вариант 17.07.2006 г.

В рамках алгебраического подхода к проблеме распознавания и на основании теории локальных и универсальных ограничений предлагается описывать универсальные ограничения как множества отображений, сохраняющих m-местные предикаты. Показано, что симметрические и функциональные ограничения допускают подобное описание. Библ. 3.

Ключевые слова: обучение по прецедентам, распознавание образов, алгебраический подход, теория локальных и универсальных ограничений.

1. ВВЕДЕНИЕ

Согласно [1], проблема распознавания заключается в том, чтобы построить отображение из множества объектов (описаний объектов) в конечное множество ответов, удовлетворяющее локальным и универсальным ограничениям. Локальные ограничения, как правило, формулируются в следующем виде: задано конечное множество прецедентов – пар вида “объект–ответ”, и требуется построить отображение, которое на каждом из этих объектов выдает соответствующий ему ответ. Универсальные ограничения в общем случае задаются как некоторое подмножество всех отображений (например, множество всех монотонных относительно некоторого частичного порядка отображений).

В алгебраическом подходе (см. [2]) требуемое отображение строится в виде суперпозиции алгоритмических операторов, корректирующих операций и решающего правила таким образом, чтобы универсальные ограничения выполнялись “по построению”. Необходимые и достаточные условия существования таких отображений в наиболее общем виде, на языке теории категорий, получены в теории универсальных и локальных ограничений.

В работах [1], [3] рассматривались различные виды универсальных ограничений: симметрические, функциональные и монотонные. В данной работе предлагается новый, более общий способ описания универсальных ограничений, формулируемый как условие сохранения заданных m-местных предикатов. Показывается, как построить эти предикаты, чтобы описать симметрические, функциональные и монотонные универсальные ограничения в рамках предлагаемого формализма.

Пусть задано $\mathfrak{I}_i$ – множество объектов и $\mathfrak{I}_f$ – множество возможных ответов. Универсальные ограничения I_s^u задаются, как подмножество $M[I_s^u]$ множества отображений из $\mathfrak{I}_i$ в $\mathfrak{I}_f$, т.е.

$$M[I_s^u] \subset M = \{a : \mathfrak{I}_i \longrightarrow \mathfrak{I}_f\}.$$

Локальные ограничения $M[I_s^l]$ задаются, как множество отображений

$$M[I_s^l] = \{a : \mathfrak{I}_i \longrightarrow \mathfrak{I}_f;\ a(x_i) = y_i,\ i = \overline{1, n}\},$$

где $(x_i, y_i)_{i=1}^n$ – конечное множество прецедентов, $x_i \in \mathfrak{I}_i$, $y_i \in \mathfrak{I}_f$.

Требуется найти отображение (если оно существует) $a^* \in M[I_s^u] \cap M[I_s^l]$. Любое такое a^* принимается как решение задачи.

В алгебраическом подходе наряду с множествами $\mathfrak{I}_i$ и $\mathfrak{I}_f$ вводится пространство оценок $\mathfrak{I}_e$. Затем выбирается множество алгоритмических операторов $\mathfrak{M}_0 \subset \mathfrak{M}_0^* = \{b : \mathfrak{I}_i \longrightarrow \mathfrak{I}_e\}$, семейство решающих правил $\mathfrak{M}_1 \subset \bigcup_{p=1}^{\infty}\{c : \mathfrak{I}_e^p \longrightarrow \mathfrak{I}_f\}$ и семейство корректирующих операций $\mathfrak{F} \subset \bigcup_{p=1}^{\infty}\{f : \mathfrak{I}_e^p \longrightarrow \mathfrak{I}_e\}$. Все семейства выбираются так, чтобы выполнялось включение $\mathfrak{M}_1 \circ \mathfrak{F}(\mathfrak{M}_0) \subseteq M[I_s^u]$.

2. ПРЕДИКАТНОЕ ЗАДАНИЕ УНИВЕРСАЛЬНЫХ ОГРАНИЧЕНИЙ

В общей теории универсальных и локальных ограничений (см. [1]) универсальные ограничения рассматриваются как некоторые специальные категории, называемые допустимыми. Допустимые категории характеризуются следующими свойствами.

1. Объекты категории – некоторые множества и всевозможные декартовы степени этих множеств, морфизмы – отображения.

2. Если $f : A^n \longrightarrow B^m$ – морфизм из объекта A^n в объект B^m, то $f_\Delta(x) = f(x, \ldots, x)$ – морфизм из объекта A в объект B^m.

3. Если $f : A^n \longrightarrow B^m$ и $g : A^p \longrightarrow B^q$ – морфизмы, то $f \times g : A^{n+p} \longrightarrow B^{m+q}$, определяемый в виде $f \times g(x_1, \ldots, x_n, y_1, \ldots, y_p) = (f(x_1, \ldots, x_n), g(y_1, \ldots, y_p))$, – тоже морфизм.

Покажем, как с помощью m-местных предикатов можно задавать категории такого типа.

Определение 1. Пусть на множествах A и B заданы m-местные предикаты ρ_A и ρ_B, т.е. $\rho_A \subset A^m$ и $\rho_B \subset B^m$. Тогда будем говорить, что функция $\varphi : A \longrightarrow B$ сохраняет пару (ρ_A, ρ_B), если $\forall (x_1, \ldots, x_m) \in \rho_A \longrightarrow \langle \varphi(x_1), \ldots, \varphi(x_m)\rangle \in \rho_B$. Множество функций, сохраняющих пару (ρ_A, ρ_B), будем обозначать через $H(\rho_A, \rho_B)$.

Следующее очевидное свойство предикатного задания и определяет его выбор в качестве языка описания соответствующих универсальных ограничений.

Предложение 1. *Если $f : A \longrightarrow B$ сохраняет пару (ρ_A, ρ_B), а $g : B \longrightarrow C$ сохраняет пару (ρ_B, ρ_C), то $g \circ f$ сохраняет пару (ρ_A, ρ_C).*

Определение 2. Пусть на множестве A задан m-местный предикат ρ_A. Тогда назовем s-й степенью ρ_A и через ρ_A^s обозначим m-местный предикат на A^s такой, что $\langle (a_1^1, \ldots, a_1^s), \ldots, (a_1^m, \ldots, a_s^m)\rangle \in \rho_A^s \Leftrightarrow \forall i = \overline{1, s}\ (a_i^1, \ldots, a_i^m) \in \rho_A$.

Рассмотрим некоторый класс множеств I. Пусть на каждом множестве $A \in I$ задан m-местный предикат ρ_A, и пусть дана категория $\Psi(I, \rho)$, объектами которой являются множества A и их декартовы степени A^s, $s = 1, 2, 3, \ldots$, а множества морфизмов для любых $A, B \in I$ и $s, r \in N$ определяются равенством $\mathrm{Hom}(A^s, B^r) = H(\rho_A^s, \rho_B^r)$. То, что Ψ – категория, следует из предложения 1 и из того, что $I_A^s : A^s \longrightarrow A^s$ (единичный оператор) сохраняет (ρ_A^s, ρ_A^s), т.е. $I_A^s \in \mathrm{Hom}(A^s, A^s)$. Допустимость этой категории очевидна из построения. Скажем, что категория $\Psi(I, \rho)$ порождена классом всех таких пар $\langle A, \rho_A\rangle$. Обозначим этот класс через $\mathfrak{K}(I, \rho)$.

Рассмотрим примеры категорий, порождаемых классами $\mathfrak{K}(I, \rho)$.

2.1. Ограничения монотонности

Пусть каждому множеству A из класса I поставлен в соответствие частичный порядок $\geq^A$ на этом множестве. Категория $(I, \geq)$, порожденная классом $\mathfrak{K}(I, \geq)$, формализует универсальные ограничения монотонности.

2.2. Симметрические ограничения

Следуя [3], рассмотрим множество $\mathfrak{C}_{q,l}(A)$ – множество матриц $q \times l$ над множеством A, и пусть I – класс всех множеств $\mathfrak{C}_{q,l}(A)$. Пусть нам дана некоторая подгруппа $S = \{e, s_1, \ldots, s_k\}$ группы подстановок $S_{q,l}$ над $\{(1, 1), (1, 2), \ldots, (q, l)\}$.

Для $U \in \mathfrak{C}_{q,l}(A)$ и $s \in S_{q,l}$ положим $s(\|U_{ij}\|_{q,l}) = \|U_{s(i,j)}\|_{q,l}$ и для $U_1, \ldots, U_d \in \mathfrak{C}_{q,l}(A)$ и $s \in S_{q,l}$ положим $s(U_1, \ldots, U_d) = (s(U_1), \ldots, s(U_d))$.

Напомним, что симметрической категорией, соответствующей группе S, называется категория Ψ_S, классом объектов которой служат множества матриц $\mathfrak{C}_{q,l}(A)$ над произвольными множествами A и их декартовы степени $\mathfrak{C}^s_{q,l}(A)$, а множества морфизмов определяются в виде

$$\mathrm{Hom}_{\Psi_S}(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \{\varphi : \mathfrak{C}^r_{q,l}(A) \longrightarrow \mathfrak{C}^s_{q,l}(B) \mid \forall i = \overline{1,k},\ \forall U \in \mathfrak{C}^r_{q,l}(A)\ \varphi(s_i(U)) = s_i(\varphi(U))\}.$$

Теперь для каждого множества $\mathfrak{C}_{q,l}(A)$ определим $k+1$-местный предикат $\rho_A = \{(U, s_1(U), \ldots$ $\ldots, s_k(U)) \mid U \in \mathfrak{C}_{q,l}(A)\} \subseteq \mathfrak{C}^{k+1}_{q,l}(A)$. Легко видеть, что r-я степень предиката ρ_A есть $\rho^r_A = \{(U,$ $s_1(U), \ldots, s_k(U)) \mid U \in \mathfrak{C}^r_{q,l}(A)\} \subseteq (\mathfrak{C}^r_{q,l}(A))^{k+1}$. Таким образом, класс $\mathfrak{K}(I, \rho)$ определяет категорию $\Psi(I, \rho)$. Тогда справедливо

Предложение 2. *Справедливо соотношение*

$$\mathrm{Hom}_{\Psi(I,\rho)}(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) =$$
$$= \{\varphi : \mathfrak{C}^r_{q,l}(A) \longrightarrow \mathfrak{C}^s_{q,l}(B) \mid \forall i = \overline{1,k},\ \forall U \in \mathfrak{C}^r_{q,l}(A)\ \varphi(s_i(U)) = s_i(\varphi(U))\}.$$

Доказательство. По определению,

$$\mathrm{Hom}_{\Psi(I,\rho)}(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) =$$
$$= \{\varphi : \mathfrak{C}^r_{q,l}(A) \longrightarrow \mathfrak{C}^s_{q,l}(B) \mid \forall U \in \mathfrak{C}^r_{q,l}(A)\ \exists V \in \mathfrak{C}^s_{q,l}(B)\ (\varphi(U) = V) \wedge (\forall i = \overline{1,k}\ \varphi(s_i(U)) = s_i(V))\} =$$
$$= \{\varphi : \mathfrak{C}^r_{q,l}(A) \longrightarrow \mathfrak{C}^s_{q,l}(B) \mid \forall U \in \mathfrak{C}^r_{q,l}(A),\ \forall i = \overline{1,k}\ \varphi(s_i(U)) = s_i(\varphi(U))\}.$$

Утверждение доказано.

Таким образом, $\Psi(I, \rho) = \Psi_S$, т.е. категория $\Psi(I, \rho)$ определяет симметрические ограничения, введенные в [3].

2.3. Функциональные ограничения

Пусть дана некоторая сигнатура $\varphi = \langle \mathbb{S}_{(1,1)}, \mathbb{S}_{(1,2)}, \ldots, \mathbb{S}_{(q,l)}, \lambda \rangle$, где $\mathbb{S}_{(i,j)}$ – линейно упорядоченные подмножества множества $\mathbb{S} = \{(1, 1), (1, 2), \ldots, (q, l)\}$, а $\lambda : \mathbb{S} \longrightarrow \{1, 2, \ldots, t\}$, $t < ql$, причем $[\lambda(i_1, j_1) = \lambda(i_2, j_2)] \longrightarrow |\mathbb{S}_{(i_1, j_1)}| = |\mathbb{S}_{(i_2, j_2)}|$. Следуя [3], обозначаем $z(i, j) = |\mathbb{S}_{(i,j)}|$ и k-й элемент $\mathbb{S}_{(i,j)}$ – как $\xi(i, j, k)$, а также $\forall r \in \{1, 2, \ldots, t\}$ $\ z(r) \triangleq z(i(r), j(r))$, если $\lambda(i(r), j(r)) = r$.

Оставим объекты теми же и определим множество отображений в виде

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \{u : \mathfrak{C}^r_{q,l}(A) \longrightarrow \mathfrak{C}^s_{q,l}(B) \mid \exists f^k_n : A^{z(n)r} \longrightarrow B, k = \overline{1,s}, n = \overline{1,t},$$
$$u(\|U^1_{ij}\|_{q,l}, \ldots, \|U^r_{ij}\|_{q,l}) = (\|F^1_{ij}\|_{q,l}, \ldots, \|F^r_{ij}\|_{q,l}), F^u_{ij} = f^u_{\lambda(i,j)}(U^1_{\xi(i,j,1)}, \ldots, U^1_{\xi(i,j,z(i,j))}, \ldots, U^r_{\xi(i,j,z(i,j))})\}.$$

В [3] было показано, что полученная структура будет категорией (ее обозначают через Ψ_φ), если и только если будут выполнены следующие условия:

1) $\forall (i, j) \in \mathbb{S} \longrightarrow (i, j) \in \mathbb{S}_{(i,j)}$;

2) $[\lambda(i_1, j_1) = \lambda(i_2, j_2)] \wedge [(i_1, j_1) = \xi(i_1, j_1, k)] \longrightarrow [(i_2, j_2) = \xi(i_2, j_2, k)]$;

3) $[\lambda(i_1, j_1) = \lambda(i_2, j_2)] \longrightarrow [\lambda(\xi(i_1, j_1, k)) = \lambda(\xi(i_2, j_2, k))]$;

4) $(i, j) \in \mathbb{S}_{(i_2, j_2)} \longrightarrow \mathbb{S}_{(i_1, j_1)} \subseteq \mathbb{S}_{(i_2, j_2)}$;

5) $[\lambda(i_1, j_1) = \lambda(i_2, j_2)] \longrightarrow [\xi(\xi(i_1, j_1, k), k_1) = \xi(i_1, j_1, k_2) \equiv \xi(\xi(i_2, j_2, k), k_1) = \xi(i_2, j_2, k_2)]$.

Предложение 3. *При выполнении условий* 1) *и* 4), *а также при* $\lambda(i_1, j_1) = \lambda(i_2, j_2) \Leftrightarrow (i_1, j_2) = (i_2, j_2)$ *множество* Φ *можно переписать в следующем виде*:

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \bigcap_{i,j} \Phi_{i,j},$$

где $\Phi_{i,j}$ *– множество функций таких, что элементы с индексами из* $\mathbb{S}_{(i,j)}$ *матриц образа зави-*

сят только от элементов с индексами из $\mathbb{S}_{(i,\,j)}$ *прообраза, т.е.*

$$\Phi_{i,\,j} = \{u : \mathfrak{C}^r_{q,\,l}(A) \longrightarrow \mathfrak{C}^s_{q,\,l}(B) \mid \forall \|U^1_{tv}\|_{q,\,l}, \|V^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}, \|V^r_{tv}\|_{q,\,l} : \forall p = \overline{1, r}$$
$$U^p_{\xi(i,\,j,\,1)} = V^p_{\xi(i,\,j,\,1)}, \ldots, U^p_{\xi(i,\,j,\,z(i,\,j))} = V^p_{\xi(i,\,j,\,z(i,\,j))} \longrightarrow \forall k = \overline{1, z(i, j)}$$
$$(u(\|U^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}))_{\xi(i,\,j,\,k)} = (u(\|V^1_{tv}\|_{q,\,l}, \ldots, \|V^r_{tv}\|_{q,\,l}))_{\xi(i,\,j,\,k)}\}.$$

Доказательство. Пусть $\Phi'_{i,\,j}$ – множество функций таких, что элементы с индексом (i, j) матриц образа зависят только от элементов с индексами из $\mathbb{S}_{(i,\,j)}$ матриц прообраза, т.е.

$$\Phi'_{i,\,j} = \{u : \mathfrak{C}^r_{q,\,l}(A) \longrightarrow \mathfrak{C}^s_{q,\,l}(B) \mid \exists f_{i,\,j}\ (u(\|U^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}))_{i,\,j} =$$
$$= f_{i,\,j}(U^1_{\xi(i,\,j,\,1)}, \ldots, U^1_{\xi(i,\,j,\,z(i,\,j))}, \ldots, U^r_{\xi(i,\,j,\,z(i,\,j))})\}.$$

Очевидно, что $\Phi_{i,\,j} \subseteq \Phi'_{i,\,j}$, так как $(i, j) \in \mathbb{S}_{(i,\,j)}$ по свойству 1). Отсюда имеем

$$\bigcap_{i,\,j} \Phi_{i,\,j} \subseteq \bigcap_{i,\,j} \Phi'_{i,\,j} = \Phi(\mathfrak{C}^r_{q,\,l}(A), \mathfrak{C}^s_{q,\,l}(B)).$$

Обратное вложение также верно, так как, по свойству 4), имеем $(i_1, j_1) \in \mathbb{S}_{(i_2,\,j_2)} \longrightarrow \mathbb{S}_{(i_1,\,j_1)} \subseteq \mathbb{S}_{(i_2,\,j_2)}$, т.е. элементы с индексами из $\mathbb{S}_{(i_2,\,j_2)}$ матриц образа зависят только от элементов с индексами из $\mathbb{S}_{(i_2,\,j_2)}$ матриц прообраза. Утверждение доказано.

Предложение 4. *При удовлетворении условий*, 1)–5) *множество* Φ *можно описать в следующем виде*:

$$\Phi(\mathfrak{C}^r_{q,\,l}(A), \mathfrak{C}^s_{q,\,l}(B)) = \bigcap_{i,\,j,\,i',\,j' : \lambda(i,\,j) = \lambda(i',\,j')} \Phi_{i,\,j,\,i',\,j'},$$

где $\Phi_{i,\,j,\,i',\,j'}$ – *множество функций таких, что функция зависимости элементов с индексами из* $\mathbb{S}_{(i,\,j)}$ *матриц образа от элементов с индексами из* $\mathbb{S}_{(i,\,j)}$ *матриц прообраза та же*, *что и функция зависимости элементов с индексами из* $\mathbb{S}_{(i',\,j')}$ *матриц образа от элементов с индексами из* $\mathbb{S}_{(i',\,j')}$ *матриц прообраза*, *т.е.*

$$\Phi_{i,\,j,\,i',\,j'} = \{u : \mathfrak{C}^r_{q,\,l}(A) \longrightarrow \mathfrak{C}^s_{q,\,l}(B) \mid \forall \|U^1_{tv}\|_{q,\,l}, \|V^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}, \|V^r_{tv}\|_{q,\,l} : \forall p = \overline{1, r}$$
$$U^p_{\xi(i,\,j,\,1)} = V^p_{\xi(i',\,j',\,1)}, \ldots, U^p_{\xi(i,\,j,\,z(i,\,j))} = V^p_{\xi(i',\,j',\,z(i',\,j'))} \longrightarrow \forall k = \overline{1, z(i, j)}$$
$$(u(\|U^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}))_{\xi(i,\,j,\,k)} = (u(\|V^1_{tv}\|_{q,\,l}, \ldots, \|V^r_{tv}\|_{q,\,l}))_{\xi(i',\,j',\,k)}\}.$$

Доказательство. Так как $\Phi_{i,\,j,\,i,\,j} = \Phi_{i,\,j}$, то, очевидно, $\Phi_{i,\,j,\,i,\,j} \subseteq \Phi'_{i,\,j}$ согласно свойству 1). И если при $\lambda(i, j) = \lambda(i', j')$ имеет место

$$\Phi'_{i,\,j,\,i',\,j'} = \{u : \mathfrak{C}^r_{q,\,l}(A) \longrightarrow \mathfrak{C}^s_{q,\,l}(B) \mid \forall \|U^1_{tv}\|_{q,\,l}, \|V^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}, \|V^r_{tv}\|_{q,\,l} : \forall i = \overline{1, r}$$
$$U^i_{\xi(i,\,j,\,1)} = V^i_{\xi(i',\,j',\,1)}, \ldots, U^i_{\xi(i,\,j,\,z(i,\,j))} = V^i_{\xi(i',\,j',\,z(i',\,j'))} \longrightarrow$$
$$\longrightarrow (u(\|U^1_{tv}\|_{q,\,l}, \ldots, \|U^r_{tv}\|_{q,\,l}))_{i,\,j} = (u(\|V^1_{tv}\|_{q,\,l}, \ldots, \|V^r_{tv}\|_{q,\,l}))_{i',\,j'}\},$$

то $\Phi_{i,\,j,\,i',\,j'} \subseteq \Phi'_{i,\,j,\,i',\,j'}$, так как если $(i, j) = \xi(i, j, k)$, то $(i', j') = \xi(i', j', k)$ (свойство 2)); $\Phi'_{i,\,j,\,i',\,j'}$ – множество функций таких, что функция зависимости элементов с индексом i, j матриц образа от элементов с индексами из $\mathbb{S}_{(i,\,j)}$ матриц прообраза та же, что и функция зависимости элементов с индексами i', j' матриц образа от элементов с индексами из $\mathbb{S}_{(i',\,j')}$ матриц прообраза. Тогда очевидно, что

$$\bigcap_{i,\,j,\,i',\,j' : \lambda(i,\,j) = \lambda(i',\,j')} \Phi_{i,\,j,\,i',\,j'} \subseteq \bigcap_{i,\,j} \Phi'_{i,\,j} \bigcap_{i,\,j,\,i',\,j' : \lambda(i,\,j) = \lambda(i',\,j')} \Phi'_{i,\,j,\,i',\,j'} = \Phi(\mathfrak{C}^r_{q,\,l}(A), \mathfrak{C}^s_{q,\,l}(B)).$$

Но, с другой стороны, при $\lambda(i, j) = \lambda(i', j')$ выполняется включение $\Phi(\mathfrak{C}^r_{q,\,l}(A), \mathfrak{C}^s_{q,\,l}(B)) \subseteq \Phi_{i,\,j,\,i',\,j'}$, так как имеем

$$\forall u \in \Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) \Rightarrow \forall \|U^1_{tv}\|_{q,l}, \|V^1_{tv}\|_{q,l}, \ldots, \|U^r_{tv}\|_{q,l}, \|V^r_{tv}\|_{q,l} : \forall i = \overline{1, r}$$

$$U^i_{\xi(i,j,1)} = V^i_{\xi(i',j',1)}, \ldots, U^i_{\xi(i,j,z(i,j))} = V^i_{\xi(i',j',z(i',j'))} \longrightarrow \forall k = \overline{1, z(i,j)}$$

$$(u(\|U^1_{tv}\|_{q,l}, \ldots, \|U^r_{tv}\|_{q,l}))_{\xi(i,j,k)} =$$

$$= \langle f^1_{\lambda(\xi(i,j,k))}(U^1_{\xi(\xi(i,j,k),1)}, \ldots, U^r_{\xi(\xi(i,j,k),z(\xi(i,j,k)))}), \ldots, f^s_{\lambda(\xi(i,j,k))}(U^1_{\xi(\xi(i,j,k),1)}, \ldots, U^r_{\xi(\xi(i,j,k),z(\xi(i,j,k)))})\rangle =$$

$$= \langle f^1_{\lambda(\xi(i',j',k))}(V^1_{\xi(\xi(i',j',k),1)}, \ldots, V^r_{\xi(\xi(i',j',k),z(\xi(i',j',k)))}), \ldots, f^s_{\lambda(\xi(i',j',k))}(V^1_{\xi(\xi(i',j',k),1)}, \ldots, V^r_{\xi(\xi(i',j',k),z(\xi(i',j',k)))})\rangle =$$

$$= (u(\|V^1_{tv}\|_{q,l}, \ldots, \|V^r_{tv}\|_{q,l}))_{\xi(i',j',k)}.$$

Здесь было использовано то, что $\lambda(\xi(i, j, k)) = \lambda(\xi(i', j', k)) \Rightarrow f^x_{\lambda(\xi(i,j,k))} = f^x_{\lambda(\xi(i',j',k))}$, и то, что из $\xi(\xi(i, j, k), k_1) = \xi(i, j, k_2)$ и $\xi(\xi(i', j', k), k_1) = \xi(i', j', k_2)$ следует $U^x_{\xi(\xi(i,j,k),k_1)} = U^x_{\xi(i,j,k_2)} = U^x_{\xi(i',j',k_2)} = V^x_{\xi(\xi(i',j',k),k_1)}$ (свойства 3) и 5)).

Итак, получаем

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) \subseteq \bigcap_{i,j,i',j' : \lambda(i,j) = \lambda(i',j')} \Phi_{i,j,i',j'}.$$

С учетом того, что обратное вложение было доказано выше, получаем

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \bigcap_{i,j,i',j' : \lambda(i,j) = \lambda(i',j')} \Phi_{i,j,i',j'}.$$

Утверждение доказано.

Напомним, что $H(\rho_A, \rho_B)$ для произвольных предикатов ρ_A, ρ_B – множество функций, сохраняющих пару (ρ_A, ρ_B). Справедливо

Предложение 5. *Пусть даны множества* X, Y *и пары предикатов* $(\rho'_1, \rho''_1), \ldots, (\rho'_l, \rho''_l)$, $\rho'_k \subseteq X^{m_k}$, $\rho''_k \subseteq Y^{m_k}$, $k = \overline{1, l}$. *Тогда имеем*

$$\bigcap_{i=1}^{l} H(\rho'_i, \rho''_i) = H(\rho'_1 \times \ldots \times \rho'_l, \rho''_1 \times \ldots \times \rho''_l).$$

Из определения $\Phi_{i,j,i',j'}$ видно, что $\Phi_{i,j,i',j'} = H(\rho^{A,r}_{i,j,i',j'}, \rho^{B,s}_{i,j,i',j'})$, где

$$\rho^{A,r}_{i,j,i',j'} = \{(\langle \|U^1_{tv}\|_{q,l}, \ldots, \|U^r_{tv}\|_{q,l}\rangle, \langle \|V^1_{tv}\|_{q,l}, \ldots, \|V^r_{tv}\|_{q,l}\rangle) \mid U^p_{tv}, V^p_{tv} \in \mathfrak{C}_{q,l}(A);$$
$$\forall k = \overline{1, z(i,j)}, p = \overline{1, r} \quad U^p_{\xi(i,j,k)} = V^p_{\xi(i',j',k)}\},$$

причем $\rho^{A,r}_{i,j,i',j'} = (\rho^{A}_{i,j,i',j'})^r$.

Предложение 6. *При выполнении условий* 1) *и* 4), *а также при* $\lambda(i_1, j_1) = \lambda(i_2, j_2) \Leftrightarrow (i_1, j_1) = (i_2, j_2)$ *получаем*

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = H\left(\left(\prod_{i,j} \rho^A_{i,j,i,j}\right)^r, \left(\prod_{i,j} \rho^B_{i,j,i,j}\right)^s\right).$$

Доказательство. Согласно предложениям 3 и 5 имеем

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \bigcap_{i,j} \Phi_{i,j} = \bigcap_{i,j} H(\rho^{A,r}_{i,j,i,j}, \rho^{B,s}_{i,j,i,j}) = H\left(\prod_{i,j} \rho^{A,r}_{i,j,i,j}, \prod_{i,j} \rho^{B,s}_{i,j,i,j}\right) =$$

$$= H\left(\prod_{i,j} (\rho^{A}_{i,j,i,j})^r, \prod_{i,j} (\rho^{B}_{i,j,i,j})^r\right) = H\left(\left(\prod_{i,j} \rho^{A}_{i,j,i,j}\right)^r, \left(\prod_{i,j} \rho^{B}_{i,j,i,j}\right)^s\right).$$

Предложение 7. *При выполнении условий* 1)–5) *получаем*

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = H\left(\left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{A}_{i,j,i',j'}\right)^r, \left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{B}_{i,j,i',j'}\right)^s\right).$$

Доказательство. Согласно предложениям 4 и 5 имеем

$$\Phi(\mathfrak{C}^r_{q,l}(A), \mathfrak{C}^s_{q,l}(B)) = \bigcap_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \Phi_{i,j,i',j'} = \bigcap_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} H(\rho^{A,r}_{i,j,i',j'}, \rho^{B,s}_{i,j,i',j'}) =$$

$$= H\left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{A,r}_{i,j,i',j'}, \prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{B,s}_{i,j,i',j'}\right) =$$

$$= H\left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} (\rho^{A}_{i,j,i',j'})^r, \prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} (\rho^{B}_{i,j,i',j'})^r\right) =$$

$$= H\left(\left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{A}_{i,j,i',j'}\right)^r, \left(\prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^{B}_{i,j,i',j'}\right)^s\right).$$

Отсюда получаем два утверждения.

Теорема 1. *При условиях* 1) *и* 4), *а также при* $\lambda(i_1, j_1) = \lambda(i_2, j_2) \Leftrightarrow (i_1, j_1) = (i_2, j_2)$ *класс пар* $\langle \mathfrak{C}_{q,l}(A), \prod_{i,j} \rho^A_{i,j,i,j} \rangle$ *порождает категорию* Ψ_φ.

Теорема 2. *При условиях* 1)–5) *класс пар* $\langle \mathfrak{C}_{q,l}(A), \prod_{i,j,i',j'\,:\,\lambda(i,j)=\lambda(i',j')} \rho^A_{i,j,i',j'} \rangle$ *порождает категорию* Ψ_φ.

В заключение автор выражает глубокую признательность К.В. Рудакову за постановку задачи, внимание к работе и ценные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. *Рудаков К.В.* Об алгебраической теории универсальных и локальных ограничений для задач классификации // Распознавание, классификация, прогноз. Вып. 1. М.: Наука, 1989. С. 176–201.
2. *Журавлев Ю.И.* Об алгебраическом подходе к решению задач распознавания или классификации // Пробл. кибернетики. 1979. Вып. 33. С. 5–68.
3. *Рудаков К.В.* О симметрических и функциональных ограничениях для алгоритмов классификации // Докл. АН СССР. 1987. Т. 297. № 1. С. 43–46.

Сдано в набор 17.11.2006 г. Подписано к печати 02.02.2007 г. Формат бумаги $60 \times 88^1/_8$
Цифровая печать Усл. печ. л. 25.0 Усл. кр.-отт. 9.2 тыс. Уч.-изд. л. 25.0 Бум. л. 12.5
Тираж 361 экз. Зак. 2149

Учредители: Российская академия наук, Вычислительный центр им. А.А. Дородницына РАН

Издатель: Академиздатцентр “Наука”, 117997 Москва, Профсоюзная ул., 90
Оригинал-макет подготовлен МАИК “Наука/Интерпериодика”
Отпечатано в ППП “Типография “Наука”, 121099 Москва, Шубинский пер., 6