

ISSN 0002-3388

Номер 6

Ноябрь-Декабрь 2024

ИЗВЕСТИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК

# ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ



# СОДЕРЖАНИЕ

---

## Номер 6, 2024

---

### УПРАВЛЕНИЕ В ДЕТЕРМИНИРОВАННЫХ СИСТЕМАХ

Об одном подходе к аналитическому синтезу модального управления по выходу линеинными стационарными системами четвертого порядка с двумя входами и двумя выходами

*Н. Е. Зубов, А. В. Лапин*

3

Матричное уравнение Сильвестра в задаче прямого синтеза стабилизирующей обратной связи линейной дискретной стационарной системы на основе данных о ее состояниях

*Е. А. Перепелкин*

20

### ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

Максимизация дальности полета для упрощенной модели летательного аппарата

*Е. В. Малых, О. Ю. Черкасов*

28

### УПРАВЛЕНИЕ СИСТЕМАМИ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ

Задача распределенного управления колебаниями эмульсии слабовязких сжимаемых жидкостей

*А. А. Егорова, А. С. Шамаев*

41

### МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Влияние аэродинамической асимметрии на критические режимы полета дозвуковых самолетов и управление для вывода из них

*М. Е. Сидорюк, А. Н. Храбров*

46

Стационарные режимы ветроэнергетической установки на базе ротора Савониуса при наличии стабилизатора напряжения

*В. М. Буданов, А. П. Голуб, М. З. Досаев, К. В. Климов, Ю. Д. Селюцкий*

63

### СИСТЕМНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Дифференцированные оценки многопользовательской сети при повреждении узлов

*Ю. Е. Малашенко, И. А. Назарова*

79

### ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Применение бустинга в рекомендательных системах

*М. А. Жарова, В. И. Цурков*

91

Обнаружение и классификация объектов на трехмерных изображениях с использованием методов глубокого обучения

*И. А. Матвеев, А. А. Юрченко*

111

Модели машинного обучения для автоматического выявления новых функциональных паттернов в многомерных временных рядах

*А. И. Майсурадзе, Л. С. Сидоров*

135

Морфология момента и структура динамических процессов в коммуникативных

*В. П. Кальян*

144

### РОБОТОТЕХНИКА

Релейные управление для безынерционного двухзвенного манипуляционного робота

*Ю. Ф. Долгий, И. А. Чупин*

158

# CONTENTS

---

---

Number 6, 2024

---

## MANAGEMENT IN DETERMINISTIC SYSTEMS

On one approach to analytic synthesis of modal control by output  
for fourth-order dynamic systems with two inputs and two outputs

*N. E. Zubov, A. V. Lapin* 3

Matrix sylvester equation in problem of direct design of stabilizing feedback of linear discrete-time  
stationary system on basis of data on its states

*E. A. Perepelkin* 20

---

## OPTIMAL CONTROL

Flight range maximization problem for a simplified aircraft model

*E. V. Malykh, O. Yu. Cherkasov* 28

---

## MANAGEMENT OF SYSTEMS WITH DISTRIBUTED PARAMETERS

The problem of distributed control of emulsion vibrations of weakly viscous compressible liquids

*A. A. Egorova, A. S. Shamaev* 41

---

## MATHEMATICAL MODELING

Influence of aerodynamic asymmetry on critical flight regimes  
of subsonic aircraft and recovery control

*M. E. Sidoryuk, A. N. Khrabrov* 46

Steady regimes of a savonius rotor-based wind power generator with voltage stabilizer

*V. M. Budanov, A. P. Holub, M. Z. Dosaev, K. V. Klimov, Y. D. Selyutskiy* 63

---

## SYSTEM ANALYSIS AND OPERATIONS RESEARCH

Differentiated estimates of a multiuser network in case of node damage

*Yu. E. Malashenko, I. A. Nazarova* 79

---

## ARTIFICIAL INTELLIGENCE

Application of boosting in recommender systems

*M. A. Zharova, V. I. Tsurkov* 91

Detection and classification of objects in three-dimensional images using deep learning methods

*I. A. Matveev, A. A. Yurchenko* 111

Machine learning models to automatically discover novel functional patterns  
in multidimensional time series

*A. I. Maysuradze, L. S. Sidorov* 135

Morphology of the moment and structure of dynamic processes in communicative situations

*V. Kalyan* 144

---

## ROBOTICS

Relay controls for an inertialess two-link manipulation robot

*Yu. F. Dolgii, I. A. Chupin* 158

# ОБ ОДНОМ ПОДХОДЕ К АНАЛИТИЧЕСКОМУ СИНТЕЗУ МОДАЛЬНОГО УПРАВЛЕНИЯ ПО ВЫХОДУ ЛИНЕЙНЫМИ СТАЦИОНАРНЫМИ СИСТЕМАМИ ЧЕТВЕРТОГО ПОРЯДКА С ДВУМЯ ВХОДАМИ И ДВУМЯ ВЫХОДАМИ

© 2024 г. Н. Е. Зубов<sup>a</sup>, \*, А. В. Лапин<sup>a</sup>

<sup>a</sup>МГТУ им. Н. Э. Баумана, Москва, Россия

\*e-mail: nik.zubov@gmail.com

Поступила в редакцию 24.02.2024 г.

После доработки 02.04.2024 г.

Принята к публикации 19.07.2024 г.

Рассматривается задача модального управления по выходу линейной стационарной системой четвертого порядка с двумя управляющими входами и двумя наблюдаемыми выходами при условии, что индекс управляемости не равен индексу наблюдаемости. Показано, что эта задача разрешима, несмотря на то, что суммарная размерность входного и выходного векторов не превышает размерность вектора состояния. Предложены компактные аналитические решения данной задачи с использованием матричных аннуляторов и условий разрешимости односторонних линейных матричных уравнений. Доказаны теоремы, реализующие прямой (по управлению) и дуальный (по наблюдению) подходы к решению задачи для случаев, когда индекс управляемости соответственно больше и меньше индекса наблюдаемости. Приведены примеры, показывающие корректность работы каждого из подходов.

**Ключевые слова:** модальное управление по выходу, система управления четвертого порядка, индекс управляемости, индекс наблюдаемости, матричный аннулятор, одностороннее матричное уравнение.

DOI: 10.31857/S0002338824060012, EDN: SVUZCA

## ON ONE APPROACH TO ANALYTIC SYNTHESIS OF MODAL CONTROL BY OUTPUT FOR FOURTH-ORDER DYNAMIC SYSTEMS WITH TWO INPUTS AND TWO OUTPUTS

Н. Е. Zubov<sup>a</sup>, \*, А. В. Lapin<sup>a</sup>

<sup>a</sup>Bauman Moscow State Technical University, Moscow, Russian Federation

\*e-mail: nik.zubov@gmail.com

The problem of modal control by output for a fourth-order dynamic system with two control inputs and two measured outputs is considered, provided that the controllability index is not equal to the observability index. It is shown that this problem is solvable, despite the fact that the total dimension of the input and output vectors does not exceed the dimension of the state vector. Compact analytic solutions to this problem using matrix zero-divisors and solvability conditions of one-sided linear matrix equations are proposed. Theorems are proved that implement direct (control) and dual (observation) approaches to solving the problem for cases when the controllability index is correspondingly greater and less than the observability index. Examples are given confirming the efficiency of each approach.

**Keywords:** modal control by output, fourth-order control system, controllability index, observability index, matrix zero-divisor, one-sided matrix equation.

**Введение.** В практике синтеза законов управления стабилизацией динамические системы четвертого порядка являются достаточно распространенными. Здесь можно привести примеры из авиации, когда продольное и боковое движение летательного аппарата рассматриваются раздельно и соответственно каждое из них описывается системой четвертого порядка [1–3].

Аналогичная ситуация характерна и для орбитальной стабилизации космического аппарата [4], где движение во взаимосвязанных каналах управления крен-рысканье также представляется системой четвертого порядка.

При построении аналитических законов управления по выходу в настоящее время в основном используется декомпозиционный метод, базирующийся на подходах Ван дер Вауда [5] и описанный авторами в работах [1, 6–8]. Однако этот метод содержит требование о том, что суммарная размерность входного и выходного векторов должна строго превышать размерность вектора состояния, поэтому его нельзя применить напрямую к системам четвертого порядка с двумя входами и двумя выходами.

Несмотря на то, что задача управления по выходу системой четвертого порядка с двумя входами и двумя выходами является классической в теории автоматического управления, ее исчерпывающего решения на данный момент не существует. Были предприняты попытки решения этой задачи для некоторых частных случаев [9, 10]. Так, в [9] описано приведение рассматриваемой задачи к управлению по состоянию системой с одним входом, весьма ограниченное условиями приводимости. А в [10] разобран конкретный практический пример управления боковым движением воздушного судна при отсутствии измерений углов скольжения и крена без изложения общего подхода к аналитическому синтезу.

В настоящей работе получены и строго доказаны новые аналитические формулы решения задачи модального управления по выходу, основанные на использовании матричных аннуляторов и условий разрешимости односторонних линейных матричных уравнений [11]. Формулы могут применяться не только в задачах, приводимых к управлению со скалярным входом, но и в любых других задачах, если в системе четвертого порядка с двумя входами и двумя выходами индекс управляемости не равен индексу наблюдаемости. Таким образом, предлагаемое решение существенно расширяет класс охватываемых систем управления.

**1. Постановка задачи.** Рассматривается полностью управляемая и полностью наблюдаемая четырехмерная линейная стационарная система управления с двумя входами и двумя выходами:

$$\begin{cases} \sigma \mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t), \end{cases} \quad (1.1)$$

где  $\sigma$  – оператор, соответствующий дифференцированию по времени  $t$  в непрерывном случае  $t \in \mathbb{R}$  или сдвигу на шаг вперед  $t + 1$  в дискретном случае  $t \in \mathbb{Z}$ ;  $\mathbf{x}(t) \in \mathbb{R}^4$ ,  $\mathbf{u}(t) \in \mathbb{R}^2$  и  $\mathbf{y}(t) \in \mathbb{R}^2$  – вектора состояния, управления (входов) и наблюдения (выходов),  $\mathbf{A} \in \mathbb{R}^{4 \times 4}$ ,  $\mathbf{B} \in \mathbb{R}^{4 \times 2}$  и  $\mathbf{C} \in \mathbb{R}^{2 \times 4}$  – соответствующие им постоянные матрицы; запись вида  $\mathbb{R}^{p \times q}$  обозначает множество вещественных матриц, имеющих размерность  $p \times q$ . Далее очевидная зависимость векторов от времени опускается.

Предполагается, что матрицы  $\mathbf{B}$  и  $\mathbf{C}$  имеют полный ранг:

$$\text{rank } \mathbf{B} = \text{rank } \mathbf{C} = 2.$$

Управление строится в соответствии с законом:

$$\mathbf{u}(t) = -\mathbf{F}\mathbf{y}(t), \quad (1.2)$$

где  $\mathbf{F}$  – постоянная матрица регулятора по выходу.

Требуется определить матрицу  $\mathbf{F}$ , обеспечивающую матрице  $\mathbf{A}^* = \mathbf{A} - \mathbf{B}\mathbf{F}\mathbf{C}$  замкнутой системы (1.1), (1.2) заданный спектр  $\Lambda^* = \{\phi_{B1}, \phi_{B2}, \phi_{C1}, \phi_{C2}\}$ :

$$\text{eig } \mathbf{A}^* = \Lambda^*.$$

Введем следующие обозначения для матриц, использующих желаемые полюса:

$$\mathbf{D}_B = (\mathbf{A} - \phi_{B1}\mathbf{I}_4)(\mathbf{A} - \phi_{B2}\mathbf{I}_4), \quad \mathbf{D}_C = (\mathbf{A} - \phi_{C1}\mathbf{I}_4)(\mathbf{A} - \phi_{C2}\mathbf{I}_4).$$

Здесь и далее  $\mathbf{I}_n$  – единичная матрица размерности  $n \times n$ .

Поскольку в рассматриваемой задаче размерности управления и наблюдения в сумме не превосходят размерность состояния, задача не может быть решена с помощью известного декомпозиционного метода модального управления по выходу [1, 6–8].

**2. Достаточное условие модальной управляемости по выходу.** Примем за основу декомпозиционный метод решения соответствующих задач с обратной связью по состоянию [12, 13]: прямой задачи (модального управления):

$$\begin{aligned} \mathbf{A}_{B1} &= \mathbf{B}^L \mathbf{A} \mathbf{B}^{L+}, \quad \mathbf{B}_1 = \mathbf{B}^L \mathbf{A} \mathbf{B}, \\ \mathbf{B}^- &= \mathbf{B}^+ + \mathbf{K}_1 \mathbf{B}^L, \quad \mathbf{K} = \mathbf{B}^- \mathbf{A} - \Phi_{B0} \mathbf{B}^-, \\ \text{eig}(\mathbf{A} - \mathbf{B} \mathbf{K}) &= \text{eig}(\mathbf{A}_{B1} - \mathbf{B}_1 \mathbf{K}_1) \cup \text{eig} \Phi_{B0} = \Lambda^*, \end{aligned}$$

дуальной задачи (модального наблюдения):

$$\begin{aligned} \mathbf{A}_{C1} &= \mathbf{C}^{R+} \mathbf{A} \mathbf{C}^R, \quad \mathbf{C}_1 = \mathbf{C} \mathbf{A} \mathbf{C}^R, \\ \mathbf{C}^- &= \mathbf{C}^+ + \mathbf{C}^R \mathbf{L}_1, \quad \mathbf{L} = \mathbf{A} \mathbf{C}^- - \mathbf{C}^- \Phi_{C0}, \\ \text{eig}(\mathbf{A} - \mathbf{L} \mathbf{C}) &= \text{eig}(\mathbf{A}_{C1} - \mathbf{L}_1 \mathbf{C}_1) \cup \text{eig} \Phi_{C0} = \Lambda^*. \end{aligned}$$

Здесь и далее верхними индексами “ $R$ ”, “ $L$ ” и “ $+$ ” обозначаются соответственно правые и левые матричные аннуляторы максимального ранга и псевдообратные матрицы [11]. Кроме того, для матриц полного ранга будем использовать известные соотношения [11]:

$$\begin{aligned} \left[ \begin{array}{c|c} \mathbf{B} & \mathbf{B}^{L+} \end{array} \right] &= \left[ \begin{array}{c} \mathbf{B}^+ \\ \mathbf{B}^L \end{array} \right]^{-1}, \quad \left[ \begin{array}{c} \mathbf{C} \\ \mathbf{C}^{R+} \end{array} \right] = \left[ \begin{array}{c|c} \mathbf{C}^+ & \mathbf{C}^R \end{array} \right]^{-1}, \\ \mathbf{B}^+ &= (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T, \quad \mathbf{C}^+ = \mathbf{C}^T (\mathbf{C} \mathbf{C}^T)^{-1}. \end{aligned}$$

Запишем *прямое* условие существования матрицы  $\mathbf{F}$  (разрешимости уравнения  $\mathbf{F}\mathbf{C} = \mathbf{K}$  [11]):

$$\underbrace{(\mathbf{B}^- \mathbf{A} - \Phi_{B0} \mathbf{B}^-)}_{\mathbf{K}} \mathbf{C}^R = \mathbf{0}_{2 \times 2}. \quad (2.1)$$

Предположим, что матрица  $\mathbf{G}_B = \mathbf{B}^- \mathbf{C}^R$  неособенная, т.е.

$$\text{rank} \left[ \begin{array}{c} \mathbf{B}^- \\ \mathbf{C} \end{array} \right] = 4. \quad (2.2)$$

Тогда условие (2.1) принимает вид

$$\mathbf{B}^- (\mathbf{A} \mathbf{C}^R - \mathbf{C}^R \tilde{\Phi}_{B0}) = \mathbf{0}_{2 \times 2}, \quad (2.3)$$

где  $\tilde{\Phi}_{B0} = \mathbf{G}_B^{-1} \Phi_{B0} \mathbf{G}_B$ . Второй множитель в выражении (2.3) можно представить иначе:

$$\mathbf{A} \mathbf{C}^R - \mathbf{C}^R \tilde{\Phi}_{B0} = (\mathbf{C}^+ \mathbf{C} + \mathbf{C}^R \mathbf{C}^{R+}) (\mathbf{A} \mathbf{C}^R - \mathbf{C}^R \tilde{\Phi}_{B0}) = \mathbf{C}^+ \mathbf{C}_1 + \mathbf{C}^R (\mathbf{A}_{C1} - \tilde{\Phi}_{B0}).$$

Если положить  $\tilde{\Phi}_{B0} = \mathbf{A}_{C1} - \mathbf{L}_1 \mathbf{C}_1$ , то условие упростится следующим образом:

$$\mathbf{B}^- \mathbf{C}^- \mathbf{C}_1 = \mathbf{0}_{2 \times 2}. \quad (2.4)$$

Запишем *дуальное* условие существования матрицы  $\mathbf{F}$  (разрешимости уравнения  $\mathbf{B}\mathbf{F} = \mathbf{L}$  [11]):

$$\underbrace{\mathbf{B}^L (\mathbf{A} \mathbf{C}^- - \mathbf{C}^- \Phi_{C0})}_{\mathbf{L}} = \mathbf{0}_{2 \times 2}. \quad (2.5)$$

Предположим, что матрица  $\mathbf{G}_C = \mathbf{B}^L \mathbf{C}^-$  неособенная, т.е.

$$\text{rank} \left[ \begin{array}{c} \mathbf{C}^- \\ \mathbf{B} \end{array} \right] = 4. \quad (2.6)$$

Тогда условие (2.5) принимает вид

$$(\mathbf{B}^L \mathbf{A} - \tilde{\Phi}_{C0} \mathbf{B}^L) \mathbf{C}^- = \mathbf{0}_{2 \times 2}, \quad (2.7)$$

где  $\tilde{\Phi}_{C0} = \mathbf{G}_C \Phi_{C0} \mathbf{G}_C^{-1}$ . Первый множитель в выражении (2.7) можно представить иначе:

$$\mathbf{B}^L \mathbf{A} - \tilde{\Phi}_{C0} \mathbf{B}^L = (\mathbf{B}^L \mathbf{A} - \tilde{\Phi}_{C0} \mathbf{B}^L) (\mathbf{B} \mathbf{B}^+ + \mathbf{B}^{L+} \mathbf{B}^L) = \mathbf{B}_1 \mathbf{B}^+ + (\mathbf{A}_{B1} - \tilde{\Phi}_{C0}) \mathbf{B}^L.$$

Если положить  $\tilde{\Phi}_{C0} = \mathbf{A}_{B1} - \mathbf{B}_1 \mathbf{K}_1$ , то условие (2.7) упростится следующим образом:

$$\mathbf{B}_1 \mathbf{B}^- \mathbf{C}^- = \mathbf{0}_{2 \times 2}. \quad (2.8)$$

При выполнении условий (2.1) и (2.5) матрица регулятора по выходу рассчитывается из соотношений:

$$\mathbf{F} = \mathbf{K} \mathbf{C}^+ = \mathbf{B}^+ \mathbf{L}.$$

Подставив сюда значения

$$\mathbf{C}^+ = \mathbf{C}^- - \mathbf{C}^R \mathbf{L}_1, \quad \mathbf{K} \mathbf{C}^R = \mathbf{0}_{2 \times 2}, \quad \mathbf{B}^+ = \mathbf{B}^- - \mathbf{K}_1 \mathbf{B}^L, \quad \mathbf{B}^L \mathbf{L} = \mathbf{0}_{2 \times 2},$$

получим

$$\mathbf{F} = \underbrace{(\mathbf{B}^- \mathbf{A} - \Phi_{B0} \mathbf{B}^-)}_{\mathbf{K}} \mathbf{C}^- = \mathbf{B}^- \underbrace{(\mathbf{A} \mathbf{C}^- - \mathbf{C}^- \Phi_{C0})}_{\mathbf{L}}.$$

Таким образом,

$$\mathbf{F} = \mathbf{B}^- \mathbf{A} \mathbf{C}^- - \Delta, \quad (2.9)$$

где  $\Delta = \Phi_{B0} \mathbf{B}^- \mathbf{C}^- = \mathbf{B}^- \mathbf{C}^- \Phi_{C0}$ .

Из прямого условия (2.4) и дуального условия (2.8) видно, что для существования матрицы регулятора по выходу  $\mathbf{F}$  достаточно выполнения обобщенного условия

$$\mathbf{B}^- \mathbf{C}^- = \mathbf{0}_{2 \times 2}, \quad (2.10)$$

которое обеспечивается путем выбора матриц регулятора  $\mathbf{K}_1$  и наблюдателя  $\mathbf{L}_1$  на первом уровне декомпозиции с сохранением заданного спектра:

$$\text{eig}(\mathbf{A}_{B1} - \mathbf{B}_1 \mathbf{K}_1) \cup \text{eig}(\mathbf{A}_{C1} - \mathbf{L}_1 \mathbf{C}_1) = \Lambda^*. \quad (2.11)$$

При этом формула для расчета матрицы  $\mathbf{F}$  (2.9) упрощается и принимает вид

$$\mathbf{F} = \mathbf{B}^- \mathbf{A} \mathbf{C}^-. \quad (2.12)$$

Основной сложностью применения расчетной формулы (2.12) является получение общего аналитического решения системы условий (2.10), (2.11) относительно матриц  $\mathbf{K}_1$  и  $\mathbf{L}_1$ . В данной работе найдены аналитические решения для случаев, когда система (1.1) имеет неравные индексы управляемости и наблюдаемости [14].

Индексы управляемости и наблюдаемости системы могут принимать только два возможных значения: 2 или 3. Введем в рассмотрение матрицы, содержащие соответственно первые два блока матриц управляемости и наблюдаемости:

$$\tilde{\mathbf{U}} = [\mathbf{B} \mid \mathbf{AB}], \quad \tilde{\mathbf{N}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix}.$$

Тогда индекс 2 будет характеризоваться полным рангом 4 соответствующей матрицы  $\tilde{\mathbf{U}}$  или  $\tilde{\mathbf{N}}$ , а индекс 3 – неполным рангом 3. При ранге 2 теряется соответствующее свойство полной управляемости или полной наблюдаемости.

**3. Случай большего индекса управляемости.** Рассмотрим случай, когда индекс управляемости равен 3, а индекса наблюдаемости – 2:

$$\text{rank } \tilde{\mathbf{U}} = 3, \quad \text{rank } \tilde{\mathbf{N}} = 4. \quad (3.1)$$

**Л е м м а 1.** Любой левый аннулятор матрицы  $\mathbf{B}$  является левым аннулятором для левого множителя ее скелетного разложения  $\mathbf{B} = \hat{\mathbf{B}} \mathbf{S}_B$ , т.е. для матрицы  $\hat{\mathbf{B}}$ .

Доказательство. Согласно определению левого аннулятора,  $\mathbf{B}^L \hat{\mathbf{B}} \mathbf{S}_B = \mathbf{0}$ . Предположим, что при этом  $\mathbf{B}^L \hat{\mathbf{B}} \neq \mathbf{0}$ . Но тогда матрица  $\mathbf{B}^L \hat{\mathbf{B}}$  является левым аннулятором для матрицы  $\mathbf{S}_B$ , что невозможно, так как последняя в силу определения скелетного разложения имеет полный ранг, равный числу строк. Лемма доказана.

Лемма 2. Общее решение задачи модального управления для пары матриц ( $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{n \times m}$ ) и спектра  $\Lambda^*$  при вырождении ранга  $\text{rank } \mathbf{B} \neq m$  (скелетное разложение  $\mathbf{B} = \mathbf{B} \mathbf{S}_B$ ) имеет вид:

$$\mathbf{K} = \left[ \mathbf{S}_B^+ \mid \mathbf{S}_B^R \right] \begin{bmatrix} \hat{\mathbf{K}} \\ \Omega_B \end{bmatrix}, \quad (3.2)$$

где  $\hat{\mathbf{K}}$  – общее решение соответствующей задачи для пары матриц ( $\mathbf{A}, \hat{\mathbf{B}}$ ), а  $\Omega_B$  – произвольная матрица подходящей размерности.

Доказательство. Введем в рассмотрение матрицу  $\Phi_B$  с желаемым спектром  $\text{eig } \Phi_B = \Lambda^*$  и составим уравнение

$$\mathbf{A} - \mathbf{B} \mathbf{K} = \Phi_B \Leftrightarrow \mathbf{B} \mathbf{K} = \mathbf{A} - \Phi_B \quad (3.3)$$

Это уравнение разрешимо относительно матрицы  $\mathbf{K}$  при условии  $\hat{\mathbf{B}}^L \Phi_B = \hat{\mathbf{B}}^L \mathbf{A}$  (с учетом леммы 1). Значит,

$$\Phi_B = \hat{\mathbf{B}}^{L+} \hat{\mathbf{B}}^L \mathbf{A} - \hat{\mathbf{B}} \Omega_{\Phi_B} = \mathbf{A} - \hat{\mathbf{B}} \hat{\mathbf{K}}, \quad (3.4)$$

где  $\Omega_{\Phi_B}$  – произвольная матрица, а за счет выбора множителя  $\hat{\mathbf{K}} = \hat{\mathbf{B}}^+ \mathbf{A} + \Omega_{\Phi_B}$  матрице  $\Phi_B$  обеспечивается заданный спектр. Если  $\hat{\mathbf{B}}$  – квадратная матрица (левого аннулятора  $\mathbf{B}^L$  не существует), то условие  $\Phi_B = \mathbf{A} - \mathbf{B} \mathbf{K}$  не накладывает ограничений на матрицу  $\Phi_B$ .

Общее решение уравнения (3.3) с подстановкой значения (3.4) принимает вид:

$$\mathbf{K} = \underbrace{\mathbf{S}_B^+ \hat{\mathbf{B}}^+}_{\mathbf{B}^+} (\mathbf{A} - \Phi_B) + \mathbf{S}_B^R \Omega_B = \mathbf{S}_B^+ \hat{\mathbf{B}}^+ \hat{\mathbf{B}} \hat{\mathbf{K}} + \mathbf{S}_B^R \Omega_B = \left[ \mathbf{S}_B^+ \mid \mathbf{S}_B^R \right] \begin{bmatrix} \hat{\mathbf{K}} \\ \Omega_B \end{bmatrix},$$

что и требовалось показать. Лемма доказана.

Теорема 1. Для системы с индексом управляемости 3 и индексом наблюдаемости 2 при условии  $\det(\tilde{\mathbf{B}}^- \mathbf{B}) \neq 0$  регулятор по выходу описывается матрицей

$$\mathbf{F} = (\tilde{\mathbf{B}}^- \mathbf{B})^{-1} \tilde{\mathbf{B}}^- \tilde{\mathbf{L}}, \quad (3.5)$$

где

$$\begin{aligned} \tilde{\mathbf{B}}^- &= \begin{bmatrix} \mathbf{b}^T \\ \mathbf{b}^T (\mathbf{A} - \tilde{\mathbf{L}} \mathbf{C}) \end{bmatrix}, \quad \mathbf{b}^T = \tilde{\mathbf{U}}^L \mathbf{D}_B, \quad \tilde{\mathbf{L}} = \mathbf{D}_C \mathbf{C}^R, \quad \mathbf{C}^R = \tilde{\mathbf{N}}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}, \\ \tilde{\mathbf{U}} &= [\mathbf{B} \mid \mathbf{AB}], \quad \tilde{\mathbf{N}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix}, \end{aligned}$$

независимо от конкретного значения левого аннулятора  $\tilde{\mathbf{U}}^L$ .

Доказательство. Теорема доказывается в четыре шага.

Шаг 1. Расчет регулятора на 1-м уровне ( $\mathbf{K}_1$  и  $\mathbf{B}^-$ ),  $\text{eig}(\mathbf{A}_{B1} - \mathbf{B}_1 \mathbf{K}_1) = \{\phi_{B1}, \phi_{B2}\}$ .

В силу условия (3.1)  $\det \tilde{\mathbf{U}} = 0$ , а также равенства

$$\det \left( \begin{bmatrix} \mathbf{B}^+ \\ \mathbf{B}^L \end{bmatrix} \tilde{\mathbf{U}} \right) = \det \left[ \begin{array}{c|c} \mathbf{I}_2 & \mathbf{B}^+ \mathbf{AB} \\ \mathbf{0}_{2 \times 2} & \mathbf{B}_1 \end{array} \right] = \det \mathbf{B}_1$$

выполним скелетное разложение вырожденной матрицы  $\mathbf{B}_1 = \hat{\mathbf{B}}_1 \mathbf{S}_{B1}$  и воспользуемся леммой 2:

$$\mathbf{K}_1 = \left[ \mathbf{S}_{B1}^+ \mid \mathbf{S}_{B1}^R \right] \begin{bmatrix} \hat{\mathbf{K}}_1 \\ -\Omega_{B1} \end{bmatrix},$$

где  $\hat{\mathbf{K}}_1$  – матрица регулятора для пары матриц  $\mathbf{A}_{B1}$  и  $\hat{\mathbf{B}}_1$ ,  $\Omega_{B1}$  – произвольная матрица. Таким образом,

$$\mathbf{B}^- = \mathbf{B}^+ + \mathbf{K}_1 \mathbf{B}^L = \left[ \mathbf{S}_{B1}^+ \mid \mathbf{S}_{B1}^R \right] \left( \begin{bmatrix} \mathbf{S}_{B1} \\ \mathbf{S}_{B1}^{R+} \end{bmatrix} \mathbf{B}^+ + \begin{bmatrix} \hat{\mathbf{K}}_1 \\ -\Omega_{B1} \end{bmatrix} \mathbf{B}^L \right) = \left[ \mathbf{S}_{B1}^+ \mid \mathbf{S}_{B1}^R \right] \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \end{bmatrix}, \quad (3.6)$$

где  $\mathbf{b}_1^T = \mathbf{S}_{B1} \mathbf{B}^+ + \hat{\mathbf{K}}_1 \mathbf{B}^L$ ,  $\mathbf{b}_2^T = \mathbf{S}_{B1}^{R+} \mathbf{B}^+ - \Omega_{B1} \mathbf{B}^L$ .

По формуле Аккермана [15] получаем

$$\hat{\mathbf{K}}_1 = \mathbf{B}_1^L \left( \mathbf{A}_{B1}^2 - (\phi_{B1} + \phi_{B2}) \mathbf{A}_{B1} + \phi_{B1} \phi_{B2} \mathbf{I}_2 \right) = \mathbf{B}_1^L \mathbf{B}^L \mathbf{D}_B \mathbf{B}^{L+}, \quad (3.7)$$

где

$$\mathbf{B}_1^L = [0 \ 1] \left[ \hat{\mathbf{B}}_1 \mid \mathbf{A}_{B1} \hat{\mathbf{B}}_1 \right]^{-1},$$

$$\mathbf{A}_{B1}^2 = \mathbf{B}^L \mathbf{A} \mathbf{B}^{L+} \mathbf{B}^L \mathbf{A} \mathbf{B}^{L+} = \mathbf{B}^L \mathbf{A} \left( \mathbf{I}_4 - \mathbf{B} \mathbf{B}^+ \right) \mathbf{A} \mathbf{B}^{L+} = \mathbf{B}^L \mathbf{A}^2 \mathbf{B}^{L+} - \mathbf{B}_1 \mathbf{B}^+ \mathbf{A} \mathbf{B}^{L+}.$$

Рассчитаем вектор  $\mathbf{b}_1^T$ , используемый в формуле (3.6), с учетом значения (3.7):

$$\mathbf{b}_1^T = \mathbf{S}_{B1} \mathbf{B}^+ + \mathbf{B}_1^L \mathbf{B}^L \mathbf{D}_B \left( \mathbf{I}_4 - \mathbf{B} \mathbf{B}^+ \right) = \mathbf{B}_1^L \mathbf{B}^L \mathbf{D}_B, \quad (3.8)$$

так как

$$\mathbf{B}_1^L \mathbf{B}^L \mathbf{D}_B \mathbf{B} = \mathbf{B}_1^L \mathbf{B}^L \underbrace{\mathbf{A} \left( \mathbf{B} \mathbf{B}^+ + \mathbf{B}^{L+} \mathbf{B}^L \right) \mathbf{A} \mathbf{B}}_{\mathbf{A}^2} = [0 \ 1] \left[ \hat{\mathbf{B}}_1 \mid \mathbf{A}_{B1} \hat{\mathbf{B}}_1 \right]^{-1} \mathbf{A}_{B1} \hat{\mathbf{B}}_1 \mathbf{S}_{B1} = \mathbf{S}_{B1}.$$

Шаг 2. Расчет наблюдателя на 1-м уровне ( $\mathbf{L}_1$  и  $\mathbf{C}^-$ ),  $\text{eig}(\mathbf{A}_{C1} - \mathbf{L}_1 \mathbf{C}_1) = \{\phi_{C1}, \phi_{C2}\}$ .

В силу условия (3.1)  $\det \tilde{\mathbf{N}} \neq 0$ , правый аннулятор  $\mathbf{C}^R$  и матрицы  $\mathbf{C}_1$  и  $\mathbf{L}_1$  могут быть записаны в виде

$$\mathbf{C}^R = \tilde{\mathbf{N}}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}, \quad \mathbf{C}_1 = \mathbf{C} \mathbf{A} \mathbf{C}^R = \mathbf{I}_2, \quad \mathbf{L}_1 = \mathbf{A}_{C1} - \Phi_{C1},$$

где  $\text{eig} \Phi_{C1} = \{\phi_{C1}, \phi_{C2}\}$ . Таким образом,

$$\mathbf{C}^- = \mathbf{C}^+ + \mathbf{C}^R \mathbf{L}_1 = \mathbf{C}^+ + \mathbf{C}^R (\mathbf{A}_{C1} - \Phi_{C1}) = \mathbf{A} \mathbf{C}^R - \mathbf{C}^R \Phi_{C1}. \quad (3.9)$$

Шаг 3. Обеспечение условия существования (равенство  $\mathbf{B}^- \mathbf{C}^- = \mathbf{0}_{2 \times 2}$ ) (2.10).

Обнулим произведение матриц (3.6) и (3.9) в первой строке за счет матрицы  $\Phi_{C1}$  с желаемым спектром  $\{\phi_{C1}, \phi_{C2}\}$ , а во второй строке – за счет матрицы параметров  $\Omega_{B1}$ :

$$\mathbf{b}_1^T \mathbf{C}^- = \mathbf{0}_{1 \times 2} \Leftrightarrow \mathbf{b}_1^T \mathbf{C}^R \Phi_{C1} = \mathbf{b}_1^T \mathbf{A} \mathbf{C}^R, \quad (3.10)$$

$$\mathbf{b}_2^T \mathbf{C}^- = \mathbf{0}_{1 \times 2} \Leftrightarrow \Omega_{B1} = \mathbf{S}_{B1}^{R+} \mathbf{B}^+ \mathbf{C}^- \mathbf{G}_C^{-1}.$$

Уравнение (3.10) разрешимо относительно матрицы  $\Phi_{C1}$  с любым заданным спектром при условии

$$\text{rank} \left( \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix} \mathbf{C}^R \right) = 2 \Leftrightarrow \text{rank} \begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix} = 4. \quad (3.11)$$

С учетом равенств  $\mathbf{C} \mathbf{C}^- = \mathbf{C} (\mathbf{C}^+ + \mathbf{C}^R \mathbf{L}_1) = \mathbf{I}_2$  и (3.10) можно записать соотношение

$$\begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix} \mathbf{C}^- = \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \\ \mathbf{b}_1^T \mathbf{A} \mathbf{C}^- \end{bmatrix}. \quad (3.12)$$

Рассчитаем произведение  $\mathbf{A} \mathbf{C}^-$ , используя значение (3.9), а также тождество  $\Phi_{C1}^2 + s_C \Phi_{C1} + m_C \mathbf{I}_2 = \mathbf{0}_{2 \times 2}$  (теорема Кэли – Гамильтона [11]):

$$\begin{aligned}
\mathbf{AC}^- &= \mathbf{A}(\mathbf{AC}^R - \mathbf{C}^R \Phi_{C1}) = \mathbf{A}^2 \mathbf{C}^R - (\mathbf{C}^- + \mathbf{C}^R \Phi_{C1}) \Phi_{C1} = \mathbf{A}^2 \mathbf{C}^R - \mathbf{C}^R \Phi_{C1}^2 - \mathbf{C}^- \Phi_{C1} = \\
&= \mathbf{A}^2 \mathbf{C}^R + \mathbf{C}^R (s_C \Phi_{C1} + m_C \mathbf{I}_2) - \mathbf{C}^- \Phi_{C1} = \mathbf{A}^2 \mathbf{C}^R + s_C (\mathbf{AC}^R - \mathbf{C}^-) + m_C \mathbf{C}^R - \mathbf{C}^- \Phi_{C1} = \\
&= (\mathbf{A}^2 + s_C \mathbf{A} + m_C \mathbf{I}_4) \mathbf{C}^R - \mathbf{C}^- (\Phi_{C1} + s_C \mathbf{I}_2) = \mathbf{D}_C \mathbf{C}^R - \mathbf{C}^- (\Phi_{C1} + s_C \mathbf{I}_2).
\end{aligned}$$

Таким образом,

$$\mathbf{AC}^- = \mathbf{D}_C \mathbf{C}^R - \mathbf{C}^- (\Phi_{C1} + s_C \mathbf{I}_2). \quad (3.13)$$

Так как в силу равенств (3.10) и (3.13)  $\mathbf{b}_1^T \mathbf{AC}^- = \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R$ , из соотношения (3.12) при условии (3.11) получим

$$\mathbf{C}^- = \begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R \end{bmatrix}. \quad (3.14)$$

Из условия (2.10) и равенства  $\mathbf{B}^- \mathbf{B} = (\mathbf{B}^+ + \mathbf{K}_1 \mathbf{B}^L) \mathbf{B} = \mathbf{I}_2$  следует

$$\mathbf{B}^- \begin{bmatrix} \mathbf{C}^- & | & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix}.$$

Поэтому с учетом условия (2.6) и значения (3.14) запишем

$$\mathbf{B}^- = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{C}^- & | & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & | & \mathbf{CB} \\ \mathbf{0}_{1 \times 2} & | & \mathbf{b}_1^T \mathbf{B} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & | & \mathbf{b}_1^T \mathbf{AB} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix}. \quad (3.15)$$

Шаг 4. Расчет регулятора по выходу (матрица  $\mathbf{F}$ ).

Определим искомую матрицу регулятора по выходу, используя равенства (2.10), (3.13) и (3.15):

$$\mathbf{F} = \mathbf{B}^- \mathbf{AC}^- = \mathbf{B}^- \mathbf{D}_C \mathbf{C}^R = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & | & \mathbf{CB} \\ \mathbf{0}_{1 \times 2} & | & \mathbf{b}_1^T \mathbf{B} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & | & \mathbf{b}_1^T \mathbf{AB} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix} \mathbf{D}_C \mathbf{C}^R.$$

Принимая во внимание тот факт, что

$$\begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & | & \mathbf{CB} \\ \mathbf{0}_{1 \times 2} & | & \mathbf{b}_1^T \mathbf{B} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & | & \mathbf{b}_1^T \mathbf{AB} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{2 \times 2} \end{bmatrix} = \mathbf{0}_{2 \times 2},$$

перепишем полученное выражение в виде

$$\begin{aligned}
\mathbf{F} &= \mathbf{B}^- \mathbf{AC}^- = \mathbf{B}^- \mathbf{D}_C \mathbf{C}^R = \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & | & \mathbf{CB} \\ \mathbf{0}_{1 \times 2} & | & \mathbf{b}_1^T \mathbf{B} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & | & \mathbf{b}_1^T \mathbf{AB} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C} \\ \mathbf{b}_1^T \\ \mathbf{b}_1^T \mathbf{A} \end{bmatrix} \mathbf{D}_C \mathbf{C}^R - \begin{bmatrix} \mathbf{I}_2 \\ \mathbf{0}_{1 \times 2} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R \end{bmatrix} \mathbf{CD}_C \mathbf{C}^R = \\
&= \begin{bmatrix} \mathbf{0}_{2 \times 2} & | & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_2 & | & \mathbf{CB} \\ \mathbf{0}_{1 \times 2} & | & \mathbf{b}_1^T \mathbf{B} \\ \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & | & \mathbf{b}_1^T \mathbf{AB} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_1^T (\mathbf{A} - \mathbf{D}_C \mathbf{C}^R \mathbf{C}) \end{bmatrix} \mathbf{D}_C \mathbf{C}^R.
\end{aligned}$$

Таким образом, после инвертирования указанной блочной матрицы используется только правый нижний блок. Рассчитаем его с помощью дополнения Шура [11]:

$$\begin{aligned}
& [\mathbf{0}_{2 \times 2} \mid \mathbf{I}_2] \left[ \begin{array}{c|c} \mathbf{I}_2 & \mathbf{C}\mathbf{B} \\ \hline \mathbf{0}_{1 \times 2} & \mathbf{b}_1^T \mathbf{B} \\ \hline \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & \mathbf{b}_1^T \mathbf{A}\mathbf{B} \end{array} \right]^{-1} \left[ \begin{array}{c} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{array} \right] = \\
& = \left( \left[ \begin{array}{c} \mathbf{b}_1^T \mathbf{B} \\ \hline \mathbf{b}_1^T \mathbf{A}\mathbf{B} \end{array} \right] - \left[ \begin{array}{c} \mathbf{0}_{1 \times 2} \\ \hline \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R \end{array} \right] (\mathbf{I}_2)^{-1} \mathbf{C}\mathbf{B} \right)^{-1} = \left( \left[ \begin{array}{c} \mathbf{b}_1^T \\ \hline \mathbf{b}_1^T (\mathbf{A} - \mathbf{D}_C \mathbf{C}^R \mathbf{C}) \end{array} \right] \mathbf{B} \right)^{-1}.
\end{aligned}$$

Подставив это значение в расчетную формулу, получим

$$\mathbf{F} = \left( \left[ \begin{array}{c} \mathbf{b}_1^T \\ \hline \mathbf{b}_1^T (\mathbf{A} - \mathbf{D}_C \mathbf{C}^R \mathbf{C}) \end{array} \right] \mathbf{B} \right)^{-1} \left[ \begin{array}{c} \mathbf{b}_1^T \\ \hline \mathbf{b}_1^T (\mathbf{A} - \mathbf{D}_C \mathbf{C}^R \mathbf{C}) \end{array} \right] \mathbf{D}_C \mathbf{C}^R.$$

Здесь  $\mathbf{b}_1^T \in \mathbb{R}^{1 \times 4}$  – вектор-строка, принимающий значение (3.8). Заметим, что в найденной формуле он использован в двух множителях, один из которых инвертируется. Поэтому вместо вектора  $\mathbf{b}_1^T$  может быть записан вектор  $\mathbf{b}^T = \kappa_b \mathbf{b}_1^T$  с произвольным коэффициентом  $\kappa_b \neq 0$ . Это позволяет применять в записи

$$\mathbf{b}^T = \mathbf{B}_1^L \mathbf{B}^L \mathbf{D}_B = \tilde{\mathbf{U}}^L \mathbf{D}_B$$

любые левые аннуляторы  $\mathbf{B}_1^L \in \mathbb{R}^{1 \times 2}$  и  $\tilde{\mathbf{U}}^L \in \mathbb{R}^{1 \times 4}$  в силу равенства

$$\underbrace{\mathbf{B}_1^L \mathbf{B}^L}_{\tilde{\mathbf{U}}^L} \underbrace{[\mathbf{B} \mid \mathbf{A}\mathbf{B}]}_{\tilde{\mathbf{U}}} = \left[ \begin{array}{c|c} 0 & \mathbf{B}_1^L \mathbf{B}_1 \end{array} \right] = \left[ \begin{array}{c|c} 0 & 0 \end{array} \right].$$

Таким образом, окончательная формула для расчета матрицы регулятора по выходу принимает вид

$$\mathbf{F} = (\tilde{\mathbf{B}}^T \mathbf{B})^{-1} \tilde{\mathbf{B}}^T \tilde{\mathbf{L}},$$

где

$$\tilde{\mathbf{B}}^T = \left[ \begin{array}{c} \mathbf{b}^T \\ \hline \mathbf{b}^T (\mathbf{A} - \tilde{\mathbf{L}} \mathbf{C}) \end{array} \right], \quad \mathbf{b}^T = \tilde{\mathbf{U}}^L \mathbf{D}_B, \quad \tilde{\mathbf{L}} = \mathbf{D}_C \mathbf{C}^R, \quad \mathbf{C}^R = \tilde{\mathbf{N}}^{-1} \left[ \begin{array}{c} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{array} \right],$$

что совпадает с результатом (3.5), представленным в теореме.

Условие применимости формулы (3.5)  $\det(\tilde{\mathbf{B}}^T \mathbf{B}) \neq 0$  вытекает из условий (2.6) и (3.11) в силу формулы Шура для блочных определителей [11]:

$$\det \left( \left[ \begin{array}{c|c} \mathbf{C} & \\ \hline \mathbf{b}_1^T & \mathbf{C}^T \mid \mathbf{B} \end{array} \right] \right) = \det \left[ \begin{array}{c|c|c} \mathbf{I}_2 & \mathbf{C}\mathbf{B} & \\ \hline \mathbf{0}_{1 \times 2} & \mathbf{b}_1^T \mathbf{B} & \\ \hline \mathbf{b}_1^T \mathbf{D}_C \mathbf{C}^R & \mathbf{b}_1^T \mathbf{A}\mathbf{B} & \end{array} \right] = \det \left( \left[ \begin{array}{c} \mathbf{b}_1^T \\ \hline \mathbf{b}_1^T (\mathbf{A} - \mathbf{D}_C \mathbf{C}^R \mathbf{C}) \end{array} \right] \mathbf{B} \right) = \det(\tilde{\mathbf{B}}^T \mathbf{B}).$$

Теорема доказана.

**4. Случай большего индекса наблюдаемости.** Рассмотрим случай, когда индекс управляемости равен 2, а индекс наблюдаемости – 3:

$$\text{rank } \tilde{\mathbf{U}} = 4, \quad \text{rank } \tilde{\mathbf{N}} = 3. \quad (4.11)$$

**Л е м м а 3.** Любой правый аннулятор матрицы  $\mathbf{C}$  является правым аннулятором для правого множителя ее скелетного разложения  $\mathbf{C} = \mathbf{S}_C \mathbf{C}$ , т.е. для матрицы  $\hat{\mathbf{C}}$ .

**Д о к а з а т е л ь с т в о.** Согласно определению правого аннулятора,  $\mathbf{S}_C \hat{\mathbf{C}} \mathbf{C}^R = \mathbf{0}$ . Предположим, что при этом  $\hat{\mathbf{C}} \mathbf{C}^R \neq \mathbf{0}$ . Но тогда матрица  $\hat{\mathbf{C}} \mathbf{C}^R$  является правым аннулятором для матрицы  $\mathbf{S}_C$ , что невозможно, так как последняя в силу определения скелетного разложения имеет полный ранг, равный числу столбцов. Лемма доказана.

Лемма 4. Общее решение задачи модального наблюдения для пары матриц  $(A \in \mathbb{R}^{n \times n}, C \in \mathbb{R}^{l \times n})$  и спектра  $\Lambda^*$  при вырождении ранга  $\text{rank } C \neq l$  (скелетное разложение  $C = S_C \hat{C}$ ) имеет вид

$$L = [\hat{L} \mid \Omega_C] \begin{bmatrix} S_C^+ \\ S_C^L \end{bmatrix}, \quad (4.2)$$

где  $\hat{L}$  – общее решение соответствующей задачи для пары матриц  $(A, \hat{C})$ , а  $\Omega_C$  – произвольная матрица подходящей размерности.

Доказательство. Введем в рассмотрение матрицу  $\Phi_C$  с желаемым спектром  $\text{eig } \Phi_C = \Lambda^*$  и составим уравнение

$$A - LC = \Phi_C \Leftrightarrow LC = A - \Phi_C. \quad (4.3)$$

Это уравнение разрешимо относительно матрицы  $L$  при условии  $\Phi_C \hat{C}^R = A \hat{C}^R$  (с учетом леммы 3). Значит,

$$\Phi_C = A \hat{C}^R \hat{C}^{R+} - \Omega_{\Phi_C} \hat{C} = A - \hat{L} \hat{C}, \quad (4.4)$$

где  $\Omega_{\Phi_C}$  – произвольная матрица, а за счет выбора множителя  $\hat{L} = A \hat{C}^+ + \Omega_{\Phi_C}$  матрице  $\Phi_C$  обеспечивается заданный спектр. Если  $\hat{C}^+$  – квадратная матрица (правого аннулятора  $\hat{C}^R$  не существует), то условие  $\Phi_C = A - \hat{L} \hat{C}$  не накладывает ограничений на матрицу  $\Phi_C$ .

Общее решение уравнения (4.3) с подстановкой значения (4.4) принимает вид

$$L = (A - \Phi_C) \underbrace{\hat{C}^+ S_C^+}_{C^+} + \Omega_C S_C^L = \hat{L} \hat{C} \hat{C}^+ S_C^+ + \Omega_C S_C^L = [\hat{L} \mid \Omega_C] \begin{bmatrix} S_C^+ \\ S_C^L \end{bmatrix},$$

что и требовалось показать. Лемма доказана.

Теорема 2. Для системы (1.1) с индексом управляемости 2 и индексом наблюдаемости 3 при условии  $\det(\tilde{C} \tilde{C}^-) \neq 0$  регулятор по выходу (1.2) описывается матрицей

$$F = \tilde{K} \tilde{C}^- (\tilde{C} \tilde{C}^-)^{-1}, \quad (4.5)$$

где

$$\begin{aligned} \tilde{C}^- &= [c \mid (A - B \tilde{K})c], \quad \tilde{K} = B^L D_B, \quad c = D_C \tilde{N}^R, \quad B^L = [\mathbf{0}_{2 \times 2} \mid I_2] \tilde{U}^{-1}, \\ \tilde{U} &= [B \mid AB], \quad \tilde{N} = \begin{bmatrix} C \\ CA \end{bmatrix}, \end{aligned}$$

независимо от конкретного значения правого аннулятора  $\tilde{N}^R$ .

Доказательство. Теорема доказывается в четыре шага.

Шаг 1. Расчет регулятора на 1-м уровне ( $K_1$  и  $B^-$ ),  $\text{eig}(A_{B1} - B_1 K_1) = \{\phi_{B1}, \phi_{B2}\}$ .

В силу условия (4.1)  $\det \tilde{U} \neq 0$ , левый аннулятор  $B^L$  и матрицы  $B_1$  и  $K_1$  могут быть записаны в виде

$$B^L = [\mathbf{0}_{2 \times 2} \mid I_2] \tilde{U}^{-1}, \quad B_1 = B^L A B = I_2, \quad K_1 = A_{B1} - \Phi_{B1},$$

где  $\text{eig } \Phi_{B1} = \{\phi_{B1}, \phi_{B2}\}$ . Таким образом,

$$B^- = B^+ + K_1 B^L = B^+ + (A_{B1} - \Phi_{B1}) B^L = B^L A - \Phi_{B1} B^L. \quad (4.6)$$

Шаг 2. Расчет наблюдателя на 1-м уровне ( $L_1$  и  $C^-$ ),  $\text{eig}(A_{C1} - L_1 C_1) = \{\phi_{C1}, \phi_{C2}\}$ .

В силу условия (4.1)  $\det \tilde{N} = 0$ , а также равенства

$$\det(\tilde{N} [C^+ \mid C^R]) = \det \begin{bmatrix} I_2 & \mathbf{0}_{2 \times 2} \\ C A C^+ & C_1 \end{bmatrix} = \det C_1,$$

выполним скелетное разложение вырожденной матрицы  $C_1 = S_{C1} \hat{C}_1$  и воспользуемся леммой 4:

$$L_1 = [\hat{L}_1 \mid -\Omega_{C1}] \begin{bmatrix} S_{C1}^+ \\ S_{C1}^L \end{bmatrix},$$

где  $\hat{\mathbf{L}}_1$  – матрица наблюдателя для пары матриц  $\mathbf{A}_{C1}$  и  $\hat{\mathbf{C}}_1$ ,  $\Omega_{C1}$  – произвольная матрица. Таким образом,

$$\mathbf{C}^- = \mathbf{C}^+ + \mathbf{C}^R \mathbf{L}_1 = \left( \mathbf{C}^+ \begin{bmatrix} \mathbf{S}_{C1} & \mathbf{S}_{C1}^{L+} \end{bmatrix} + \mathbf{C}^R \begin{bmatrix} \hat{\mathbf{L}}_1 & -\Omega_{C1} \end{bmatrix} \right) \begin{bmatrix} \mathbf{S}_{C1}^+ \\ \mathbf{S}_{C1}^L \end{bmatrix} = [\mathbf{c}_1 \mid \mathbf{c}_2] \begin{bmatrix} \mathbf{S}_{C1}^+ \\ \mathbf{S}_{C1}^L \end{bmatrix}, \quad (4.7)$$

где  $\mathbf{c}_1 = \mathbf{C}^+ \mathbf{S}_{C1} + \mathbf{C}^R \hat{\mathbf{L}}_1$ ,  $\mathbf{c}_2 = \mathbf{C}^+ \mathbf{S}_{C1}^{L+} - \mathbf{C}^R \Omega_{C1}$ .

По дуальной формуле Аккермана [15] получаем

$$\hat{\mathbf{L}}_1 = \left( \mathbf{A}_{C1}^2 - (\phi_{C1} + \phi_{C2}) \mathbf{A}_{C1} + \phi_{C1} \phi_{C2} \mathbf{I}_2 \right) \mathbf{C}_1^R = \mathbf{C}^{R+} \mathbf{D}_C \mathbf{C}^R \mathbf{C}_1^R, \quad (4.8)$$

где

$$\mathbf{A}_{C1}^2 = \mathbf{C}^{R+} \mathbf{A} \mathbf{C}^R \mathbf{C}^{R+} \mathbf{A} \mathbf{C}^R = \mathbf{C}^{R+} \mathbf{A} \left( \mathbf{I}_4 - \mathbf{C}^+ \mathbf{C} \right) \mathbf{A} \mathbf{C}^R = \mathbf{C}^{R+} \mathbf{A}^2 \mathbf{C}^R - \mathbf{C}^{R+} \mathbf{A} \mathbf{C}^+ \mathbf{C}_1,$$

$$\mathbf{C}_1^R = \begin{bmatrix} \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_1 \mathbf{A}_{C1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Рассчитаем вектор  $\mathbf{c}_1$ , используемый в формуле (4.7), с учетом значения (4.8):

$$\mathbf{c}_1 = \mathbf{C}^+ \mathbf{S}_{C1} + (\mathbf{I}_4 - \mathbf{C}^+ \mathbf{C}) \mathbf{D}_C \mathbf{C}^R \mathbf{C}_1^R = \mathbf{D}_C \mathbf{C}^R \mathbf{C}_1^R, \quad (4.9)$$

так как

$$\mathbf{C} \mathbf{D}_C \mathbf{C}^R \mathbf{C}_1^R = \mathbf{C} \underbrace{\mathbf{A} \left( \mathbf{C}^+ \mathbf{C} + \mathbf{C}^R \mathbf{C}^{R+} \right) \mathbf{A} \mathbf{C}^R \mathbf{C}_1^R}_{\mathbf{A}^2} = \mathbf{S}_{C1} \hat{\mathbf{C}}_1 \mathbf{A}_{C1} \begin{bmatrix} \hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_1 \mathbf{A}_{C1} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \mathbf{S}_{C1}.$$

Шаг 3. Обеспечение условия существования (равенство  $\mathbf{B}^- \mathbf{C}^- = \mathbf{0}_{2 \times 2}$ ) (2.10).

Обнулим произведение матриц (4.6) и (4.7) в первом столбце за счет матрицы  $\Phi_{B1}$  с желаемым спектром  $\{\phi_{B1}, \phi_{B2}\}$ , а во втором столбце – за счет матрицы параметров  $\Omega_{C1}$ :

$$\mathbf{B}^- \mathbf{c}_1 = \mathbf{0}_{2 \times 1} \Leftrightarrow \Phi_{B1} \mathbf{B}^L \mathbf{c}_1 = \mathbf{B}^L \mathbf{A} \mathbf{c}_1, \quad (4.10)$$

$$\mathbf{B}^- \mathbf{c}_2 = \mathbf{0}_{2 \times 1} \Leftrightarrow \Omega_{C1} = \mathbf{G}_B^{-1} \mathbf{B}^- \mathbf{C}^+ \mathbf{S}_{C1}^{L+}.$$

Уравнение (4.10) разрешимо относительно матрицы  $\Phi_{B1}$  с любым заданным спектром при условии

$$\text{rank}(\mathbf{B}^L [\mathbf{c}_1 \mid \mathbf{A} \mathbf{c}_1]) = 2 \Leftrightarrow \text{rank}[\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{A} \mathbf{c}_1] = 4. \quad (4.11)$$

С учетом равенств  $\mathbf{B}^- \mathbf{B} = (\mathbf{B}^+ + \mathbf{K}_1 \mathbf{B}^L) \mathbf{B} = \mathbf{I}_2$  и (4.10) можно записать соотношение

$$\mathbf{B}^- [\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{A} \mathbf{c}_1] = [\mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^- \mathbf{A} \mathbf{c}_1]. \quad (4.12)$$

Рассчитаем произведение  $\mathbf{B}^- \mathbf{A}$ , используя значение (4.6), а также тождество  $\Phi_{B1}^2 + s_B \Phi_{B1} + m_B \mathbf{I}_2 = \mathbf{0}_{2 \times 2}$  (теорема Кэли – Гамильтона):

$$\begin{aligned} \mathbf{B}^- \mathbf{A} &= (\mathbf{B}^L \mathbf{A} - \Phi_{B1} \mathbf{B}^L) \mathbf{A} = \mathbf{B}^L \mathbf{A}^2 - \Phi_{B1} (\mathbf{B}^- + \Phi_{B1} \mathbf{B}^L) = \mathbf{B}^L \mathbf{A}^2 - \Phi_{B1}^2 \mathbf{B}^L - \Phi_{B1} \mathbf{B}^- = \\ &= \mathbf{B}^L \mathbf{A}^2 + (s_B \Phi_{B1} + m_B \mathbf{I}_2) \mathbf{B}^L - \Phi_{B1} \mathbf{B}^- = \mathbf{B}^L \mathbf{A}^2 + s_B (\mathbf{B}^L \mathbf{A} - \mathbf{B}^-) + m_B \mathbf{B}^L - \Phi_{B1} \mathbf{B}^- = \\ &= \mathbf{B}^L (A^2 + s_B A + m_B \mathbf{I}_4) - (\Phi_{B1} + s_B \mathbf{I}_2) \mathbf{B}^- = \mathbf{B}^L \mathbf{D}_B - (\Phi_{B1} + s_B \mathbf{I}_2) \mathbf{B}^-. \end{aligned}$$

Таким образом,

$$\mathbf{B}^- \mathbf{A} = \mathbf{B}^L \mathbf{D}_B - (\Phi_{B1} + s_B \mathbf{I}_2) \mathbf{B}^-. \quad (4.13)$$

Так как в силу равенств (4.10) и (4.13)  $\mathbf{B}^- \mathbf{A} \mathbf{c}_1 = \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1$ , из соотношения (4.12) при условии (4.11) получим

$$\mathbf{B}^- = [\mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1] [\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{A} \mathbf{c}_1]^{-1}. \quad (4.14)$$

Из условия (2.10) и равенства  $\mathbf{C}\mathbf{C}^- = \mathbf{C}(\mathbf{C}^+ + \mathbf{C}^R\mathbf{L}_1) = \mathbf{I}_2$  следует

$$\begin{bmatrix} \mathbf{B}^- \\ \mathbf{C} \end{bmatrix} \mathbf{C}^- = \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}.$$

Поэтому с учетом условия (2.2) и значения (4.14) запишем

$$\mathbf{C}^- = \begin{bmatrix} \mathbf{B}^- \\ \mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = [\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{Ac}_1] \begin{bmatrix} \mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1 \\ \mathbf{CB} \mid \mathbf{Cc}_1 \mid \mathbf{CAC}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}.$$

Шаг 4. Расчет регулятора по выходу (матрица  $\mathbf{F}$ ) (2.12).

Определим искомую матрицу регулятора по выходу, используя равенства (2.10), (4.13) и (4.15):

$$\mathbf{F} = \mathbf{B}^- \mathbf{AC}^- = \mathbf{B}^L \mathbf{D}_B \mathbf{C}^- = \mathbf{B}^L \mathbf{D}_B [\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{Ac}_1] \begin{bmatrix} \mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1 \\ \mathbf{CB} \mid \mathbf{Cc}_1 \mid \mathbf{CAC}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}.$$

Принимая во внимание тот факт, что

$$[\mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1] \begin{bmatrix} \mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1 \\ \mathbf{CB} \mid \mathbf{S}_{\mathbf{C}_1} \mid \mathbf{CAC}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = [\mathbf{I}_2 \mid \mathbf{0}_{2 \times 2}] \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = \mathbf{0}_{2 \times 2},$$

перепишем полученное выражение в виде

$$\begin{aligned} \mathbf{F} &= \left( \mathbf{B}^L \mathbf{D}_B [\mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{Ac}_1] - \mathbf{B}^L \mathbf{D}_B \mathbf{B} [\mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1] \right) \begin{bmatrix} \mathbf{I}_2 \mid \mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1 \\ \mathbf{CB} \mid \mathbf{Cc}_1 \mid \mathbf{CAC}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = \\ &= \mathbf{B}^L \mathbf{D}_B [\mathbf{c}_1 \mid (\mathbf{A} - \mathbf{B}\mathbf{B}^L \mathbf{D}_B) \mathbf{c}_1] [\mathbf{0}_{2 \times 2} \mid \mathbf{I}_2] \begin{bmatrix} \mathbf{I}_2 \mid [\mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1] \\ \mathbf{CB} \mid [\mathbf{Cc}_1 \mid \mathbf{CAC}_1] \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix}. \end{aligned}$$

Таким образом, после инвертирования указанной блочной матрицы используется только правый нижний блок. Рассчитаем его с помощью дополнения Шура [11]:

$$\begin{aligned} [\mathbf{0}_{2 \times 2} \mid \mathbf{I}_2] \begin{bmatrix} \mathbf{I}_2 \mid [\mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1] \\ \mathbf{CB} \mid [\mathbf{Cc}_1 \mid \mathbf{CAC}_1] \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = \\ = ([\mathbf{Cc}_1 \mid \mathbf{CAC}_1] - \mathbf{CB}(\mathbf{I}_2)^{-1} [\mathbf{0}_{2 \times 1} \mid \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1])^{-1} = (\mathbf{C}[\mathbf{c}_1 \mid (\mathbf{A} - \mathbf{B}\mathbf{B}^L \mathbf{D}_B) \mathbf{c}_1])^{-1}. \end{aligned}$$

Подставив это значение в расчетную формулу, получим

$$\mathbf{F} = \mathbf{B}^L \mathbf{D}_B [\mathbf{c}_1 \mid (\mathbf{A} - \mathbf{B}\mathbf{B}^L \mathbf{D}_B) \mathbf{c}_1] (\mathbf{C}[\mathbf{c}_1 \mid (\mathbf{A} - \mathbf{B}\mathbf{B}^L \mathbf{D}_B) \mathbf{c}_1])^{-1}.$$

Здесь  $\mathbf{c}_1 \in \mathbb{R}^{4 \times 1}$  — вектор-столбец, принимающий значение (4.9). Заметим, что в найденной формуле он использован в двух множителях, один из которых инвертируется. Поэтому вместо вектора  $\mathbf{c}_1$  может быть записан вектор  $\mathbf{c} = \kappa_c \mathbf{c}_1$  с произвольным коэффициентом  $\kappa_c \neq 0$ . Это позволяет применять в записи

$$\mathbf{c} = \mathbf{D}_C \mathbf{C}^R \mathbf{C}_1^R = \mathbf{D}_C \tilde{\mathbf{N}}^R$$

любые правые аннуляторы  $\mathbf{C}_1^R \in \mathbb{R}^{2 \times 1}$  и  $\tilde{\mathbf{N}}^R \in \mathbb{R}^{4 \times 1}$  в силу равенства

$$\underbrace{\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix}}_{\mathbf{N}} \underbrace{\mathbf{C}^R \mathbf{C}_1^R}_{\tilde{\mathbf{N}}^R} = \begin{bmatrix} 0 \\ \mathbf{C}_1 \mathbf{C}_1^R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Таким образом, окончательная формула для расчета матрицы регулятора по выходу принимает вид

$$\mathbf{F} = \tilde{\mathbf{K}} \tilde{\mathbf{C}}^- (\mathbf{C} \tilde{\mathbf{C}}^-)^{-1}$$

где

$$\tilde{\mathbf{C}}^- = [\mathbf{c} \mid (\mathbf{A} - \mathbf{B}\tilde{\mathbf{K}}) \mathbf{c}], \quad \tilde{\mathbf{K}} = \mathbf{B}^L \mathbf{D}_B, \quad \mathbf{c} = \mathbf{D}_C \tilde{\mathbf{N}}^R, \quad \mathbf{B}^L = [\mathbf{0}_{2 \times 2} \mid \mathbf{I}_2] \tilde{\mathbf{U}}^{-1},$$

что совпадает с результатом (4.5), представленным в теореме.

Условие применимости формулы  $\det(\mathbf{C}\tilde{\mathbf{C}}^-) \neq 0$  вытекает из условий (2.2) и (4.11) в силу формулы Шура для блочных определителей [11]:

$$\det\left(\begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \left[ \mathbf{B} \mid \mathbf{c}_1 \mid \mathbf{A}\mathbf{c}_1 \right] \right) = \det\left(\begin{array}{c|c|c} \mathbf{I}_2 & \mathbf{0}_{2 \times 1} & \mathbf{B}^L \mathbf{D}_B \mathbf{c}_1 \\ \mathbf{C}\mathbf{B} & \mathbf{C}\mathbf{c}_1 & \mathbf{C}\mathbf{A}\mathbf{c}_1 \end{array} \right) = \det\left(\mathbf{C} \left[ \mathbf{c}_1 \mid (\mathbf{A} - \mathbf{B}\tilde{\mathbf{K}})\mathbf{c}_1 \right] \right) = \det(\mathbf{C}\tilde{\mathbf{C}}^-).$$

Теорема доказана.

**5. Примеры.** Рассмотрим два примера, соответствующих разд. 3 (индекс управляемости больше индекса наблюдаемости) и 4 (индекс наблюдаемости больше индекса управляемости).

Требуется для заданных матриц состояния  $\mathbf{A}$ , управления  $\mathbf{B}$  и наблюдения  $\mathbf{C}$  (пара матриц  $\mathbf{A}$  и  $\mathbf{B}$  полностью управляема, а пара матриц  $\mathbf{A}$  и  $\mathbf{C}$  полностью наблюдаема) синтезировать регулятор по выходу (определить матрицу  $\mathbf{F}$ ), обеспечивающий матрице замкнутой системы  $\mathbf{A}^* = \mathbf{A} - \mathbf{B}\mathbf{F}\mathbf{C}$  желаемый спектр  $\Lambda^* = \{\phi_{B1}, \phi_{B2}, \phi_{C1}, \phi_{C2}\}$ .

Сформируем из желаемых полюсов коэффициенты следующих полиномов:

$$(\lambda - \phi_{B1})(\lambda - \phi_{B2}) = \lambda^2 - \underbrace{(\phi_{B1} + \phi_{B2})}_{s_B} \lambda + \underbrace{\phi_{B1}\phi_{B2}}_{m_B},$$

$$(\lambda - \phi_{C1})(\lambda - \phi_{C2}) = \lambda^2 - \underbrace{(\phi_{C1} + \phi_{C2})}_{s_C} \lambda + \underbrace{\phi_{C1}\phi_{C2}}_{m_C},$$

$$(\lambda^2 + s_B\lambda + m_B)(\lambda^2 + s_C\lambda + m_C) = \lambda^4 + p_3\lambda^3 + p_2\lambda^2 + p_1\lambda + p_0,$$

где

$$p_0 = m_B m_C = \phi_{B1}\phi_{B2}\phi_{C1}\phi_{C2},$$

$$p_1 = s_B m_C + s_C m_B = -\phi_{B1}\phi_{B2}\phi_{C1} - \phi_{B1}\phi_{B2}\phi_{C2} - \phi_{B1}\phi_{C1}\phi_{C2} - \phi_{B2}\phi_{C1}\phi_{C2},$$

$$p_2 = s_B s_C + m_B + m_C = \phi_{B1}\phi_{B2} + \phi_{B1}\phi_{C1} + \phi_{B1}\phi_{C2} + \phi_{B2}\phi_{C1} + \phi_{B2}\phi_{C2} + \phi_{C1}\phi_{C2},$$

$$p_3 = s_B + s_C = -\phi_{B1} - \phi_{B2} - \phi_{C1} - \phi_{C2}.$$

**Пример 1.** Заданы матрицы состояния, управления и наблюдения соответственно:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & 0 & 0 & a_{1,4} \\ 0 & 0 & a_{2,3} & 0 \\ 0 & a_{3,2} & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ b_{2,1} & 0 \\ 0 & b_{3,2} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Рассчитаем первые два блочных столбца матрицы управляемости (матрицу  $\tilde{\mathbf{U}}$ ) и первые две блочных строки матрицы наблюдаемости (матрицу  $\tilde{\mathbf{N}}$ ):

$$\tilde{\mathbf{U}} = [\mathbf{B} \mid \mathbf{AB}] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ b_{2,1} & 0 & 0 & a_{2,3}b_{3,2} \\ 0 & b_{3,2} & a_{3,2}b_{2,1} & 0 \\ 0 & 0 & b_{2,1} & 0 \end{bmatrix}, \quad \tilde{\mathbf{N}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a_{1,1} & 0 & 0 & a_{1,4} \\ 0 & a_{3,2} & 0 & 0 \end{bmatrix}.$$

Поскольку  $\text{rank } \tilde{\mathbf{U}} = 3$ , а  $\text{rank } \tilde{\mathbf{N}} = 4$ , имеет место случай, когда индекс управляемости 3 больше индекса наблюдаемости 2, т.е. можно воспользоваться теоремой 1.

Запишем левый аннулятор первых двух блочных столбцов матрицы управляемости (аннулятор матрицы  $\tilde{\mathbf{U}}$ ) и правый аннулятор первой блочной строки матрицы наблюдаемости (аннулятор матрицы  $\mathbf{C}$ ):

$$\tilde{\mathbf{U}}^L = [1 \ 0 \ 0 \ 0], \quad \mathbf{C}^R = \tilde{\mathbf{N}}^{-1} \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ \mathbf{I}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{a_{3,2}} \\ 0 & 0 \\ \frac{1}{a_{1,4}} & 0 \end{bmatrix}.$$

Найдем выражения с матричными полиномами  $\mathbf{D}_B = \mathbf{A}^2 + s_B \mathbf{A} + m_B \mathbf{I}_4$  и  $\mathbf{D}_C = \mathbf{A}^2 + s_C \mathbf{A} + m_C \mathbf{I}_4$ :

$$\mathbf{b}^T = \tilde{\mathbf{U}}^L \mathbf{D}_B = [a_{1,1}^2 + s_B a_{1,1} + m_B \ a_{1,4} \ 0 \ a_{1,4} (a_{1,1} + s_B)],$$

$$\tilde{\mathbf{L}} = \mathbf{D}_C \mathbf{C}^R = \begin{bmatrix} a_{1,1} + s_C & \frac{a_{1,4}}{a_{3,2}} \\ 0 & a_{2,3} + \frac{m_C}{a_{3,2}} \\ 0 & s_C \\ \frac{m_C}{a_{1,4}} & \frac{s_C}{a_{3,2}} \end{bmatrix}.$$

С помощью вспомогательной матрицы

$$\tilde{\mathbf{B}}^- = \begin{bmatrix} \mathbf{b}^T \\ \mathbf{b}^T (\mathbf{A} - \tilde{\mathbf{L}} \mathbf{C}) \end{bmatrix} = \begin{bmatrix} a_{1,1}^2 + s_B a_{1,1} + m_B & -s_C a_{1,1}^2 - (p_2 - m_B) a_{1,1} - p_1 \\ a_{1,4} & a_{1,4} (a_{1,1} + s_B) \\ 0 & -\frac{a_{1,4}}{a_{3,2}} (a_{1,1}^2 + p_3 a_{1,1} + p_2) \\ a_{1,4} (a_{1,1} + s_B) & a_{1,4} (a_{1,1}^2 + s_B a_{1,1} + m_B) \end{bmatrix}^T$$

определим первый множитель в расчетной формуле (3.5):

$$\tilde{\mathbf{B}}^- \mathbf{B} = a_{1,4} \begin{bmatrix} b_{2,1} & 0 \\ (a_{1,1} + s_B) b_{2,1} & -\left(a_{1,1}^2 + p_3 a_{1,1} + p_2\right) \frac{b_{3,2}}{a_{3,2}} \end{bmatrix},$$

$$(\tilde{\mathbf{B}}^- \mathbf{B})^{-1} = \frac{1}{a_{1,4}} \begin{bmatrix} \frac{1}{b_{2,1}} & 0 \\ \frac{a_{3,2}}{b_{3,2}} \frac{a_{1,1} + s_B}{a_{1,1}^2 + p_3 a_{1,1} + p_2} & -\frac{a_{3,2}}{b_{3,2}} \frac{1}{a_{1,1}^2 + p_3 a_{1,1} + p_2} \end{bmatrix}.$$

и второй множитель в той же формуле:

$$\tilde{\mathbf{B}}^- \tilde{\mathbf{L}} = \begin{bmatrix} a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1 & a_{1,4} \left( a_{2,3} + \frac{a_{1,1}^2 + p_3 a_{1,1} + p_2}{a_{3,2}} \right) \\ p_0 - s_C (a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1) & a_{1,4} \left( (a_{1,1} + s_B) a_{2,3} - s_C \frac{a_{1,1}^2 + p_3 a_{1,1} + p_2}{a_{3,2}} \right) \end{bmatrix}.$$

Окончательно по формуле (3.5) рассчитаем матрицу регулятора по выходу:

$$\mathbf{F} = \left( \tilde{\mathbf{B}}^{-1} \mathbf{B} \right)^{-1} \tilde{\mathbf{B}}^{-1} \tilde{\mathbf{L}} =$$

$$= \begin{bmatrix} \frac{a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1}{a_{1,4} b_{2,1}} & \frac{1}{b_{2,1}} \left( a_{2,3} + \frac{a_{1,1}^2 + p_3 a_{1,1} + p_2}{a_{3,2}} \right) \\ \frac{a_{3,2}}{a_{1,4} b_{3,2}} \frac{(a_{1,1} + p_3)(a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1) - p_0}{a_{1,1}^2 + p_3 a_{1,1} + p_2} & \frac{a_{1,1} + p_3}{b_{3,2}} \end{bmatrix}.$$

Выполним проверку характеристического полинома замкнутой системы:

$$\mathbf{A}^* = \mathbf{A} - \mathbf{B} \mathbf{F} \mathbf{C} =$$

$$= \begin{bmatrix} a_{1,1} & 0 & 0 & a_{1,4} \\ -\frac{a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1}{a_{1,4}} & 0 & -\frac{a_{1,1}^2 + p_3 a_{1,1} + p_2}{a_{3,2}} & 0 \\ -\frac{a_{3,2}}{a_{1,4}} \frac{(a_{1,1} + p_3)(a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1) - p_0}{a_{1,1}^2 + p_3 a_{1,1} + p_2} & a_{3,2} & -a_{1,1} - p_3 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\det(\lambda \mathbf{I}_4 - \mathbf{A}^*) = (\lambda - a_{1,1}) \lambda (\lambda(\lambda + a_{1,1} + p_3) + a_{1,1}^2 + p_3 a_{1,1} + p_2) +$$

$$+ (a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1)(\lambda + a_{1,1} + p_3) - (a_{1,1} + p_3)(a_{1,1}^3 + p_3 a_{1,1}^2 + p_2 a_{1,1} + p_1) + p_0 =$$

$$= \lambda^4 + p_3 \lambda^3 + p_2 \lambda^2 + p_1 \lambda + p_0.$$

Полученный полином совпадает с желаемым характеристическим полиномом.

Пример 2. Заданы матрицы состояния, управления и наблюдения соответственно:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & a_{1,3} & a_{1,4} \\ 0 & a_{2,2} & 0 & 0 \\ a_{3,1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ b_{2,1} & 0 \\ 0 & b_{3,2} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Рассчитаем первые два блочных столбца матрицы управляемости (матрицу  $\tilde{\mathbf{U}}$ ) и первые две блочных строки матрицы наблюдаемости (матрицу  $\tilde{\mathbf{N}}$ ):

$$\tilde{\mathbf{U}} = [\mathbf{B} \mid \mathbf{AB}] = \begin{bmatrix} 0 & 0 & 0 & a_{1,3} b_{3,2} \\ b_{2,1} & 0 & a_{2,2} b_{2,1} & 0 \\ 0 & b_{3,2} & 0 & 0 \\ 0 & 0 & b_{2,1} & 0 \end{bmatrix}, \quad \tilde{\mathbf{N}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & a_{2,2} & 0 & 0 \\ a_{3,1} & 0 & 0 & 0 \end{bmatrix}.$$

Поскольку  $\text{rank } \tilde{\mathbf{U}} = 4$ , а  $\text{rank } \tilde{\mathbf{N}} = 3$ , имеет место случай, когда индекс наблюдаемости 3 больше индекса управляемости 2, т.е. можно воспользоваться теоремой 2.

Запишем левый аннулятор первого блочного столбца матрицы управляемости (аннулятор матрицы  $\mathbf{B}$ ) и правый аннулятор первых двух блочных строк матрицы наблюдаемости (аннулятор матрицы  $\tilde{\mathbf{N}}$ ):

$$\mathbf{B}^L = [\mathbf{0}_{2 \times 2} \mid \mathbf{I}_2] \tilde{\mathbf{U}}^{-1} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{b_{2,1}} \\ \frac{1}{a_{1,3} b_{3,2}} & 0 & 0 & 0 \end{bmatrix}, \quad \tilde{\mathbf{N}}^R = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Найдем выражения с матричными полиномами  $\mathbf{D}_B = \mathbf{A}^2 + s_B \mathbf{A} + m_B \mathbf{I}_4$  и  $\mathbf{D}_C = \mathbf{A}^2 + s_C \mathbf{A} + m_C \mathbf{I}_4$ :

$$\tilde{\mathbf{K}} = \mathbf{B}^L \mathbf{D}_B = \begin{bmatrix} 0 & \frac{a_{2,2} + s_B}{b_{2,1}} & 0 & \frac{m_B}{b_{2,1}} \\ \frac{1}{b_{3,2}} \left( a_{3,1} + \frac{m_B}{a_{1,3}} \right) & \frac{a_{1,4}}{a_{1,3} b_{3,2}} & \frac{s_B}{b_{3,2}} & \frac{a_{1,4} s_B}{a_{1,3} b_{3,2}} \end{bmatrix},$$

$$\mathbf{c} = \mathbf{D}_C \tilde{\mathbf{N}}^R = \begin{bmatrix} a_{1,4} s_C \\ 0 \\ a_{1,4} a_{3,1} \\ m_C \end{bmatrix}.$$

С помощью вспомогательной матрицы

$$\tilde{\mathbf{C}}^- = \left[ \mathbf{c} \mid (\mathbf{A} - \mathbf{B} \tilde{\mathbf{K}}) \mathbf{c} \right] = \begin{bmatrix} a_{1,4} s_C & a_{1,4} (a_{1,3} a_{3,1} + m_C) \\ 0 & -p_0 \\ a_{1,4} a_{3,1} & -a_{1,4} \left( a_{3,1} s_B + \frac{p_1}{a_{1,3}} \right) \\ m_C & 0 \end{bmatrix}$$

определим первый множитель в расчетной формуле (4.5):

$$\tilde{\mathbf{K}} \tilde{\mathbf{C}}^- = \begin{bmatrix} \frac{p_0}{b_{2,1}} & -\frac{p_0}{b_{2,1}} (a_{2,2} + s_B) \\ \frac{a_{1,4}}{b_{3,2}} \left( \frac{p_1}{a_{1,3}} + a_{3,1} p_3 \right) & -\frac{a_{1,4}}{b_{3,2}} \left( \frac{p_1 s_B}{a_{1,3}} - a_{3,1} (a_{1,3} a_{3,1} + p_2 - p_3 s_B) \right) \end{bmatrix}$$

и второй множитель в той же формуле:

$$\mathbf{C} \tilde{\mathbf{C}}^- = \begin{bmatrix} 0 & -p_0 \\ a_{1,4} a_{3,1} & -a_{1,4} \left( \frac{p_1}{a_{1,3}} + a_{3,1} s_B \right) \end{bmatrix},$$

$$(\mathbf{C} \tilde{\mathbf{C}}^-)^{-1} = \begin{bmatrix} -\frac{1}{p_0} \left( \frac{p_1}{a_{1,3} a_{3,1}} + s_B \right) & \frac{1}{a_{1,4} a_{3,1}} \\ -\frac{1}{p_0} & 0 \end{bmatrix}.$$

Окончательно по формуле (4.5) рассчитаем матрицу регулятора по выходу:

$$\mathbf{F} = \tilde{\mathbf{K}} \tilde{\mathbf{C}}^- (\mathbf{C} \tilde{\mathbf{C}}^-)^{-1} =$$

$$= \begin{bmatrix} \frac{1}{b_{2,1}} \left( a_{2,2} - \frac{p_1}{a_{1,3} a_{3,1}} \right) & \frac{p_0}{a_{1,4} a_{3,1} b_{2,1}} \\ -\frac{a_{1,4}}{b_{3,2} p_0} \left( a_{3,1} (a_{1,3} a_{3,1} + p_2) + \frac{p_1}{a_{1,3}} \left( \frac{p_1}{a_{1,3} a_{3,1}} + p_3 \right) \right) & \frac{1}{b_{3,2}} \left( \frac{p_1}{a_{1,3} a_{3,1}} + p_3 \right) \end{bmatrix}.$$

Выполним проверку характеристического полинома замкнутой системы:

$$\begin{aligned}
 \mathbf{A}^* &= \mathbf{A} - \mathbf{B}\mathbf{F}\mathbf{C} = \\
 &= \begin{bmatrix} 0 & 0 & a_{1,3} & a_{1,4} \\ 0 & \frac{p_1}{a_{1,3}a_{3,1}} & -\frac{p_0}{a_{1,4}a_{3,1}} & 0 \\ a_{3,1} & \frac{a_{1,4}}{p_0} \left( a_{3,1} \left( a_{1,3}a_{3,1} + p_2 \right) + \frac{p_1}{a_{1,3}} \left( \frac{p_1}{a_{1,3}a_{3,1}} + p_3 \right) \right) & -\frac{p_1}{a_{1,3}a_{3,1}} - p_3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\
 \det(\lambda\mathbf{I}_4 - \mathbf{A}^*) &= \lambda^2 \left( \left( \lambda - \frac{p_1}{a_{1,3}a_{3,1}} \right) \left( \lambda + \frac{p_1}{a_{1,3}a_{3,1}} + p_3 \right) + a_{1,3}a_{3,1} + p_2 + \frac{p_1}{a_{1,3}a_{3,1}} \left( \frac{p_1}{a_{1,3}a_{3,1}} + p_3 \right) \right) + \\
 &\quad + \lambda(p_1 - a_{1,3}a_{3,1}\lambda) + p_0 = \lambda^4 + p_3\lambda^3 + p_2\lambda^2 + p_1\lambda + p_0.
 \end{aligned}$$

Полученный полином совпадает с желаемым характеристическим полиномом.

**Заключение.** Определено аналитическое решение задачи модального управления по выходу для широкого класса динамических систем с суммарной размерностью входного и выходного векторов, не превышающей размерность вектора состояния. Решение применимо к системам четвертого порядка с двумя входами и двумя выходами при неравных индексах управляемости и наблюдаемости. Для каждого назначаемого спектра решение единствено и представляется компактной аналитической формулой, не требующей дополнительных матричных разложений или декомпозиции. Рассмотрены случаи, в которых индекс управляемости принимает значение как больше, так и меньше значения индекса наблюдаемости. Для каждого случая на основе двухуровневой декомпозиции с параметризацией обоих уровней доказаны соответствующие взаимно дуальные теоремы об управлении по выходу и приведены конкретные примеры аналитического расчета регуляторов. Примеры подтверждают совпадение спектров замкнутых систем управления по выходу с желаемыми спектрами.

#### СПИСОК ЛИТЕРАТУРЫ

1. Zubov N.E., Mikrin E.A., Ryabchenko V.N. Controlling a Linear MIMO-System by a Measurement Vector Using Multilevel Decomposition // J. Comput. Syst. Sci. Int. 2020. V. 59. Iss. 2. P. 151–160. <https://doi.org/10.1134/S1064230720020136>
2. Zubov N.E., Mikrin E.A., Ryabchenko V.N., Fomichev A.V. Synthesis of Control Laws for Aircraft Lateral Motion at the Lack of Data on the Slip Angle: Analytical Solution // Russian Aeronautics. 2017. V. 60. Iss. 1. P. 64–73. <https://doi.org/10.3103/S106879981701010X>
3. Zubov N.E., Ryabchenko V.N., Sorokin I.V. Output Control of the Longitudinal Motion Spectrum of a Single-Rotor Helicopter // Russian Aeronautics. 2020. V. 63. Iss. 2. P. 249–259. <https://doi.org/10.3103/S1068799820030319>
4. Zubov N.E., Mikrin E.A., Oleynik A.S., Ryabchenko V.N., Efyanov D.E. The Spacecraft Angular Velocity Estimation in the Orbital Stabilization Mode by the Results of the Local Vertical Sensor Measurements // Herald of the Bauman Moscow State Technical University, Series Instrument Engineering. 2014. Iss. 5. P. 3–13.
5. Van der Woude J.W. A Note on Pole Placement by Static Output Feedback for Single-Input Systems // Systems & Control Letters. 1988. V. 11. Iss. 4. P. 285–287. [https://doi.org/10.1016/0167-6911\(88\)90072-2](https://doi.org/10.1016/0167-6911(88)90072-2)
6. Zubov N.E., Zybin E.Y., Mikrin E.A., Misrikhanov M.Sh., Proletarskii A.V., Ryabchenko V.N. Output Control of a Spacecraft Motion Spectrum // J. Comput. Syst. Sci. Int. 2014. V. 53. Iss. 4. P. 576–586. <https://doi.org/10.1134/S1064230714040170>
7. Lapin A.V., Zubov N.E. Parametric Synthesis of Modal Control with Output Feedback for Descent Module Attitude Stabilization // Int. Russian Autom. Conf. 2019. P. 1–6. <https://doi.org/10.1109/RusAutoCon.2019.8867744>
8. Zubov N.E., Lapin A.V., Mikrin E.A., Ryabchenko V.N. Output Control of the Spectrum of a Linear Dynamic System in Terms of the Van der Woude Method // Doklady Mathematics. 2017. V. 96. Iss. 2. P. 457–460. <https://doi.org/10.1134/S1064562417050179>
9. Zubov N.E., Lapin A.V. Reducing the Problem of the Modal Control by Output for Stationary Forth-Order Systems with Two Inputs and Two Outputs to the Control by State for a System with a Single Input // J. Comput. Syst. Sci. Int. 2023. V. 62. Iss. 1. P. 43–60. <https://doi.org/10.1134/S1064230723010124>

10. *Zubov N.E., Zybin E.Yu., Lapin A.V.* Analytical Synthesis of an Aircraft's Lateral Motion Control by Output at the Lack of Measurements of Slip and Roll Angles // *J. Comput. Syst. Sci. Int.* 2023. V. 62. Iss. 2. P. 354–361.  
<https://doi.org/10.1134/S1064230723020193>
11. *Gantmacher F.R.* The Theory of Matrices. Providence: AMS Chelsea Publishing, 2000.
12. *Ryabchenko V.N., Zubov N.E., Sorokin I.V., Proletarskii A.V.* Complete Pole Placement Method for Linear MIMO Systems // *Mechatronics, Automation, Control.* 2018. V. 19. Iss. 1. P. 11–18.  
<https://doi.org/10.17587/mau.19.11-18>.
13. *Zubov N.E., Mikrin E.A., Misrikhanov M.Sh., Ryabchenko V.N.* Modification of the Exact Pole Placement Method and its Application for the Control of Spacecraft Motion // *J. Comput. Syst. Sci. Int.* 2013. V. 52. Iss. 2. P. 279–292.  
<https://doi.org/10.1134/S1064230713020135>
14. *Haidar I., Barbot J.-P., Rapaport A.* A Multi Observers Approach when Observability Index is Higher than the State Dimension – a Case Study // *IEEE 58th Conf. on Decision and Control.* 2019. P. 1571–1576.  
<https://doi.org/10.1109/CDC40024.2019.9029675>
15. *Ackermann J.* Der Entwurf linearer Regelungssysteme im Zustandraum // *Regeltech, Proz.-Datenverarb.* 1972. V. 20. Iss. 1. P. 297–300.  
<https://doi.org/10.1524/auto.1972.20.112.297>

## МАТРИЧНОЕ УРАВНЕНИЕ СИЛЬВЕСТРА В ЗАДАЧЕ ПРЯМОГО СИНТЕЗА СТАБИЛИЗИРУЮЩЕЙ ОБРАТНОЙ СВЯЗИ ЛИНЕЙНОЙ ДИСКРЕТНОЙ СТАЦИОНАРНОЙ СИСТЕМЫ НА ОСНОВЕ ДАННЫХ О ЕЕ СОСТОЯНИЯХ

© 2024 г. Е. А. Перепелкин<sup>a,\*</sup>

<sup>a</sup>Санкт-Петербургский государственный ун-т  
аэрокосмического приборостроения, Санкт-Петербург, Россия  
\*e-mail: perepelkin@guap.ru

Поступила в редакцию 15.04.2024 г.

После доработки 10.08.2024 г.

Принята к публикации 16.10.2024 г.

Работа посвящена решению задачи синтеза стабилизирующей обратной связи по состоянию для линейной многосвязной дискретной стационарной системы на основе данных о поведении системы. Предполагается, что матрицы системы неизвестны. Рассматривается алгоритм прямого синтеза матрицы обратной связи на базе матричного уравнения Сильвестра без решения задачи идентификации. Получены условия существования решения задачи синтеза. Рассмотрен численный пример.

**Ключевые слова:** линейная дискретная система, стабилизирующая обратная связь, синтез на основе данных, матричное уравнение Сильвестра

**DOI:** 10.31857/S0002338824060022, **EDN:** SVUDRI

## MATRIX SYLVESTER EQUATION IN PROBLEM OF DIRECT DESIGN OF STABILIZING FEEDBACK OF LINEAR DISCRETE-TIME STATIONARY SYSTEM ON BASIS OF DATA ON ITS STATES

Е. А. Perepelkin<sup>a,\*</sup>

<sup>a</sup>Saint-Petersburg, Saint-Petersburg State University of Aerospace Instrumentation  
\*e-mail: perepelkin@guap.ru

The paper presents a solution to the problem of designing a stabilizing state feedback for a linear multivariable discrete-time stationary system based on data of the system's behavior. It is assumed that the system matrices are unknown. An algorithm for directly designing a feedback matrix based on the Sylvester matrix equation without solving the identification problem is considered. Conditions for the existence of a solution to the design problem are obtained. A numerical example is considered.

**Keywords:** linear discrete-time system, stabilizing state feedback, data-driven design, matrix Sylvester equation

**Введение.** Прямой синтез систем управления на основе данных о поведении объекта управления можно рассматривать как альтернативу синтеза с помощью решения задачи идентификации [1, 2]. Данному подходу, получившему наименование data-driven control, посвящены многочисленные публикации последних лет. Современное состояние теории синтеза систем автоматического управления на основе данных представлено в работах [3–5]. Предложены решения задач оптимального управления [6–10], прогнозирующего управления [11, 12], модального управления [13]. Управление на базе данных применяется для линейных и нелинейных систем, систем с непрерывным и дискретным временем, для систем с запаздыванием [14–18].

В данной работе описывается алгоритм решения задачи синтеза стабилизирующей обратной связи для линейной многосвязной дискретной системы, матрицы которой неизвестны. Особенность алгоритма заключается в том, что он не предусматривает построение модели объекта управления. Выполняется прямой синтез матрицы обратной связи на базе данных о поведении системы. Алгоритм основан на решении матричного уравнения Сильвестра. Матрицы уравнения Сильвестра составляются с использованием данных о значениях входа и состояния системы. Исследуются условия существования и единственности решения уравнения Сильвестра. Приводится численный пример, подтверждающий теоретические результаты.

**1. Постановка задачи.** Рассмотрим линейную стационарную систему с дискретным временем, поведение которой описывается уравнением:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \quad k = 0, 1, 2, \dots, \quad (1.1)$$

где  $\mathbf{x}(k) \in \mathbb{R}^n$  – вектор состояния,  $\mathbf{u}(k) \in \mathbb{R}^m$  – вектор управления, матрицы системы  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{n \times m}$ . Будем считать, что пара  $(\mathbf{A}, \mathbf{B})$  управляема:

$$\text{rank}[\mathbf{B}, \mathbf{AB}, \dots, \mathbf{A}^{n-1}\mathbf{B}] = n.$$

В дальнейшем нам потребуется критерий управляемости [19, с. 225], связанный со спектральными свойствами матрицы  $\mathbf{A}$ .

Утверждение 1. Пара  $(\mathbf{A}, \mathbf{B})$  управляема тогда и только тогда, когда  $\text{rank}[\lambda\mathbf{E} - \mathbf{A}, \mathbf{B}] = n$  для любого собственного числа  $\lambda$  матрицы  $\mathbf{A}$  или, что эквивалентно,  $\mathbf{x}^T \mathbf{B} \neq 0$  для любого левого собственного вектора  $\mathbf{x}^T$  матрицы  $\mathbf{A}$ .

Управление в системе (1.1) будем искать в виде обратной связи по состоянию  $\mathbf{u} = \mathbf{Fx}$ , где  $\mathbf{F} \in \mathbb{R}^{m \times n}$  – матрица обратной связи. Замкнутая система

$$\mathbf{x}(k+1) = (\mathbf{A} + \mathbf{BF})\mathbf{x}(k), \quad k = 0, 1, 2, \dots$$

асимптотически устойчива тогда и только тогда, когда собственные числа матрицы замкнутой системы  $\mathbf{A} + \mathbf{BF}$  по модулю меньше единицы.

Далее через  $\lambda_i(\cdot)$  обозначим собственные числа матрицы. Например,  $\lambda_i(\mathbf{A})$ ,  $i = \overline{1, n}$ , есть собственные числа матрицы  $\mathbf{A}$ . Задача заключается в нахождении матрицы обратной связи, при которой собственные числа матрицы замкнутой системы удовлетворяют условию

$$|\lambda_i(\mathbf{A} + \mathbf{BF})| \leq \rho, \quad i = \overline{1, n}, \quad 0 < \rho < 1. \quad (1.2)$$

Предположим, что матрицы  $\mathbf{A}$  и  $\mathbf{B}$  неизвестны. Известны результаты наблюдения за поведением системы для моментов времени  $k = 0, N$ ,  $N \geq n + m$ . Эти данные запишем в виде блочных матриц:

$$\begin{aligned} \mathbf{X}_0 &= [\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(N-1)] \in \mathbb{R}^{n \times N}, \\ \mathbf{X}_1 &= [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)] \in \mathbb{R}^{n \times N}, \\ \mathbf{U}_0 &= [\mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(N-1)] \in \mathbb{R}^{m \times N}. \end{aligned} \quad (1.3)$$

Справедливо равенство  $\mathbf{X}_1 = \mathbf{AX}_0 + \mathbf{BU}_0$ . Будем считать, что

$$\text{rank} \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{U}_0 \end{bmatrix} = n + m. \quad (1.4)$$

Заметим, что из (1.4) следует:  $N \geq n + m$ ,  $\text{rank} \mathbf{X}_0 = n$ ,  $\text{rank} \mathbf{U}_0 = m$ .

**2. Алгоритм синтеза матрицы обратной связи.** Рассмотрим матричное уравнение

$$\mathbf{X}_0 \mathbf{D} = \mathbf{E}_n, \quad (2.1)$$

где  $\mathbf{E}_n$  – единичная матрица порядка  $n$ ,  $\mathbf{D} \in \mathbb{R}^{N \times n}$  – искомая матрица. Общее решение уравнения (2.1) равно

$$\mathbf{D} = \mathbf{Y} + \mathbf{ZR}, \quad (2.2)$$

где  $\mathbf{Y} = \mathbf{X}_0^T \left( \mathbf{X}_0 \mathbf{X}_0^T \right)^{-1} \in \mathbb{R}^{N \times n}$  – частное решение,  $\mathbf{R} \in \mathbb{R}^{(N-n) \times n}$  – произвольная матрица,  $\mathbf{Z} \in \mathbb{R}^{N \times (N-n)}$  – матрица, столбцы которой образуют базис ядра  $\mathbf{X}_0$ ,

$$\ker(\mathbf{X}_0) = \left\{ \mathbf{z} \in \mathbb{R}^N \mid \mathbf{X}_0 \mathbf{z} = 0 \right\}.$$

Матрицу обратной связи будем искать в виде  $\mathbf{F} = \mathbf{U}_0 \mathbf{D}$ . При этом матрица замкнутой системы

$$\mathbf{A} + \mathbf{B}\mathbf{F} = \mathbf{A}\mathbf{X}_0 \mathbf{D} + \mathbf{B}\mathbf{U}_0 \mathbf{D} = (\mathbf{A}\mathbf{X}_0 + \mathbf{B}\mathbf{U}_0)\mathbf{D} = \mathbf{X}_1 \mathbf{D}.$$

Учитывая (2.2), получим:

$$\mathbf{A} + \mathbf{B}\mathbf{F} = \mathbf{X}_1 \mathbf{D} = \mathbf{X}_1(\mathbf{Y} + \mathbf{Z}\mathbf{R}) = \mathbf{X}_1 \mathbf{Y} + \mathbf{X}_1 \mathbf{Z}\mathbf{R} = \mathbf{V} + \mathbf{W}\mathbf{R},$$

где  $\mathbf{V} = \mathbf{X}_1 \mathbf{Y} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{W} = \mathbf{X}_1 \mathbf{Z} \in \mathbb{R}^{n \times (N-n)}$ . Таким образом, задача заключается в нахождении матрицы  $\mathbf{R}$ , при которой

$$|\lambda_i(\mathbf{V} + \mathbf{W}\mathbf{R})| \leq \rho, \quad i = 1, n. \quad (2.3)$$

Достаточным условием существования матрицы  $\mathbf{R}$ , при которой справедливы неравенства (2.3), является условие управляемости пары  $(\mathbf{V}, \mathbf{W})$ . Если пара  $(\mathbf{V}, \mathbf{W})$  управляема, то все собственные числа матрицы  $\mathbf{V} + \mathbf{W}\mathbf{R}$  можно произвольно задать, выбирая матрицу  $\mathbf{R}$  [19, с. 330]. Следовательно, эти собственные значения можно задать так, чтобы выполнялись неравенства (2.3).

Следующее утверждение формулирует необходимое и достаточное условие управляемости пары  $(\mathbf{V}, \mathbf{W})$ .

Утверждение 2. Пусть данные о поведении системы удовлетворяют условию (1.4). Тогда пара  $(\mathbf{V}, \mathbf{W})$  управляема, если и только если управляема пара  $(\mathbf{A}, \mathbf{B})$ .

Доказательство. Необходимость. Пусть пара  $(\mathbf{A}, \mathbf{B})$  неуправляема. Согласно утверждению 1, существует левый собственный вектор  $\mathbf{x}^T$  матрицы  $\mathbf{A}$ , отвечающий собственному числу  $\lambda$ , такой, что  $\mathbf{x}^T \mathbf{A} = \lambda \mathbf{x}^T$ ,  $\mathbf{x}^T \mathbf{B} = \mathbf{0}$ . Учитывая, что  $\mathbf{X}_0 \mathbf{Z} = \mathbf{0}$ ,  $\mathbf{X}_0 \mathbf{Y} = \mathbf{E}_n$ , получим:

$$\mathbf{W} = \mathbf{X}_1 \mathbf{Z} = (\mathbf{A}\mathbf{X}_0 + \mathbf{B}\mathbf{U}_0)\mathbf{Z} = \mathbf{B}\mathbf{U}_0 \mathbf{Z},$$

$$\mathbf{V} = \mathbf{X}_1 \mathbf{Y} = (\mathbf{A}\mathbf{X}_0 + \mathbf{B}\mathbf{U}_0)\mathbf{Y} = \mathbf{A} + \mathbf{B}\mathbf{U}_0 \mathbf{Y}.$$

Следовательно,  $\mathbf{x}^T \mathbf{V} = \mathbf{x}^T (\mathbf{A} + \mathbf{B}\mathbf{U}_0 \mathbf{Y}) = \lambda \mathbf{x}^T$ ,  $\mathbf{x}^T \mathbf{W} = \mathbf{x}^T \mathbf{B}\mathbf{U}_0 \mathbf{Z} = \mathbf{0}$ . Таким образом, если пара  $(\mathbf{A}, \mathbf{B})$  неуправляема, то неуправляема и пара  $(\mathbf{V}, \mathbf{W})$ .

Достаточность. Пусть пара  $(\mathbf{A}, \mathbf{B})$  управляема. Предположим, что пара  $(\mathbf{V}, \mathbf{W})$  неуправляема. Тогда существует левый собственный вектор  $\mathbf{x}^T$  матрицы  $\mathbf{V}$ , отвечающий собственному числу  $\lambda$ , такой, что

$$\mathbf{x}^T \mathbf{V} = \mathbf{x}^T (\mathbf{A} + \mathbf{B}\mathbf{U}_0 \mathbf{Y}) = \lambda \mathbf{x}^T, \quad (2.4)$$

$$\mathbf{x}^T \mathbf{W} = \mathbf{x}^T \mathbf{B}\mathbf{U}_0 \mathbf{Z} = \mathbf{0}. \quad (2.5)$$

Равенство (2.5) является необходимым и достаточным условием существования решения уравнения

$$\mathbf{x}^T \mathbf{B}\mathbf{U}_0 = \mathbf{q}^T \mathbf{X}_0 \quad (2.6)$$

относительно вектора  $\mathbf{q}^T$ . Запишем соотношение (2.6) в виде:

$$[\mathbf{q}^T, -\mathbf{x}^T \mathbf{B}] \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{U}_0 \end{bmatrix} = \mathbf{0}.$$

Из условия (1.4) следует, что  $\mathbf{q}^T = \mathbf{0}$  и  $\mathbf{x}^T \mathbf{B} = \mathbf{0}$ . Равенство (2.4) принимает вид  $\mathbf{x}^T \mathbf{V} = \mathbf{x}^T (\mathbf{A} + \mathbf{B}\mathbf{U}_0 \mathbf{Y}) = \mathbf{x}^T \mathbf{A} = \lambda \mathbf{x}^T$ . Таким образом, вектор  $\mathbf{x}^T$  является левым собственным вектором матрицы  $\mathbf{A}$  и при этом  $\mathbf{x}^T \mathbf{B} = \mathbf{0}$ . Следовательно, пара  $(\mathbf{A}, \mathbf{B})$  неуправляема, что противоречит исходному предположению. Утверждение 2 доказано.

Рассмотрим алгоритм нахождения матрицы  $\mathbf{R}$ , при которой выполняется условие (2.3). Матрицу  $\mathbf{R}$  будем искать на основе решения матричного уравнения Сильвестра:

$$\mathbf{V}\mathbf{P} - r\mathbf{P}\mathbf{V}^T = -\mathbf{W}\mathbf{W}^T, \quad (2.7)$$

где  $r$  – некоторый параметр,  $\mathbf{P} \in \mathbb{R}^{n \times n}$  – искомая матрица. Пусть уравнение (2.7) имеет невырожденное решение  $\mathbf{P}$ . Положим

$$\mathbf{R} = \mathbf{W}^T \mathbf{P}^{-1}. \quad (2.8)$$

Из (2.7) и (2.8) получим равенство  $\mathbf{V} + \mathbf{WR} = r\mathbf{PV}^T \mathbf{P}^{-1}$ , из которого следует, что матрицы  $\mathbf{V} + \mathbf{WR}$  и  $r\mathbf{V}^T$  подобны. Собственные числа этих матриц совпадают. При значениях параметра  $0 < r \max_i |\lambda_i(\mathbf{V})| \leq \rho$  собственные числа матрицы  $r\mathbf{V}^T$  удовлетворяют условию  $|\lambda_i(r\mathbf{V}^T)| \leq \rho$ ,  $i = 1, n$ . Следовательно, собственные числа матрицы  $\mathbf{V} + \mathbf{WR}$  удовлетворяют условию (2.3).

Рассмотрим условия существования невырожденного решения матричного уравнения Сильвестра (2.7). Известно [20, с. 200], что линейное матричное уравнение

$$\mathbf{AX} - \mathbf{XB} = \mathbf{C}, \quad (2.9)$$

где  $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{n \times n}$ , имеет единственное решение  $\mathbf{X} \in \mathbb{R}^{n \times n}$  тогда и только тогда, когда у матриц  $\mathbf{A}$  и  $\mathbf{B}$  нет общих собственных чисел. Следовательно, уравнение (2.7) имеет единственное решение тогда и только тогда, когда  $\lambda_i(\mathbf{V}) \neq r\lambda_j(\mathbf{V})$ ,  $i, j = 1, n$ .

Условия невырожденности решений уравнения вида (2.9) детально исследованы в работах [21, 22]. Применительно к уравнению (2.7) эти условия можно сформулировать в следующем виде.

У т е р ж д е н и е 3. Управляемость пары  $(\mathbf{V}, \mathbf{W})$  является необходимым условием невырожденности единственного решения уравнения (2.7). Это условие будет достаточным в случае одноранговой матрицы  $\mathbf{W}$ .

Д о к а з а т е л ь с т в о. Докажем, учитывая специфику уравнения (2.7). Пусть пара  $(\mathbf{V}, \mathbf{W})$  неуправляема. Согласно утверждению 1, существует левый собственный вектор  $\mathbf{x}^T$  матрицы  $\mathbf{V}$ , отвечающий собственному числу  $\lambda$ , такой, что  $\mathbf{x}^T \mathbf{W} = \mathbf{0}$ . Умножим левую и правую части уравнения (2.7) слева на  $\mathbf{x}^T$ . Получим равенство  $\mathbf{x}^T \mathbf{P}(\lambda \mathbf{E}_n - r\mathbf{V}^T) = \mathbf{0}$ . Матрица  $\lambda \mathbf{E}_n - r\mathbf{V}^T$  невырожденная, поскольку  $\lambda_i(\mathbf{V}) \neq r\lambda_j(\mathbf{V})$ ,  $i, j = 1, n$ . Следовательно,  $\mathbf{x}^T \mathbf{P} = \mathbf{0}$ , что означает вырожденность матрицы  $\mathbf{P}$ .

Пусть  $\text{rank } \mathbf{W} = 1$ . Матрица  $\mathbf{WW}^T$  симметричная и одноранговая. Эту матрицу можно записать в виде  $\mathbf{WW}^T = \mathbf{qq}^T$ , где  $\mathbf{q} \in \mathbb{R}^n$  [20, с. 30]. При этом пара  $(\mathbf{V}, \mathbf{W})$  управляема тогда и только тогда, когда управляема пара  $(\mathbf{V}, \mathbf{q})$ . Действительно,  $\mathbf{x}^T \mathbf{W} = \mathbf{0}$  тогда и только тогда, когда  $\mathbf{x}^T \mathbf{q} = \mathbf{0}$ . Это справедливо и для любого левого собственного вектора  $\mathbf{x}^T$  матрицы  $\mathbf{V}$ .

Уравнение (2.7) принимает вид:

$$\mathbf{VP} - r\mathbf{PV}^T = -\mathbf{qq}^T. \quad (2.10)$$

Пусть матрица  $\mathbf{P}$  является решением уравнения (2.10) и при этом вырождена. Тогда из (2.10) следует, что  $\mathbf{x}^T \mathbf{qq}^T \mathbf{y} = 0$  для любых  $\mathbf{x} \in \ker(\mathbf{P}^T)$ ,  $\mathbf{y} \in \ker(\mathbf{P})$ . Рассмотрим два случая:  $\mathbf{q}^T \mathbf{y} \neq \mathbf{0}$  для некоторого  $\mathbf{y} \in \ker(\mathbf{P})$  и  $\mathbf{q}^T \mathbf{y} = \mathbf{0}$  для любого  $\mathbf{y} \in \ker(\mathbf{P})$ .

В первом случае получим:  $\mathbf{x}^T \mathbf{q} = \mathbf{0}$ ,  $\mathbf{x}^T \mathbf{VP} = \mathbf{0}$ . Умножим левую и правую часть уравнения (2.10) слева на  $\mathbf{V}$ . Из равенства  $\mathbf{V}^2 \mathbf{P} - r\mathbf{VPV}^T = -\mathbf{Vqq}^T$  получим:  $\mathbf{x}^T \mathbf{Vq} = \mathbf{0}$ ,  $\mathbf{x}^T \mathbf{V}^2 \mathbf{P} = \mathbf{0}$ . Продолжая таким образом, приходим к заключению, что

$$\mathbf{x}^T [\mathbf{q}, \mathbf{Vq}, \dots, \mathbf{V}^{n-1} \mathbf{q}] = \mathbf{0}.$$

Это противоречит предположению об управляемости пары  $(\mathbf{V}, \mathbf{q})$ .

Во втором случае рассмотрим уравнение  $\mathbf{q}^T = \mathbf{h}^T \mathbf{P}$  относительно неизвестного вектора  $\mathbf{h}^T$ . Это уравнение имеет решение, поскольку  $\mathbf{q}^T \mathbf{y} = \mathbf{0}$  для любого  $\mathbf{y} \in \ker(\mathbf{P})$ . Уравнение (2.10) можно записать в виде:

$$(\mathbf{V} + \mathbf{qh}^T) \mathbf{P} = r\mathbf{PV}^T. \quad (2.11)$$

Из (2.11) следует, что  $\mathbf{PV}^T \mathbf{y} = \mathbf{0}$ . Из равенства  $\mathbf{q}^T \mathbf{V}^T = \mathbf{h}^T \mathbf{PV}^T$  получим  $\mathbf{q}^T \mathbf{V}^T \mathbf{y} = \mathbf{0}$ . Далее умножим левую и правую части уравнения (2.10) справа на  $\mathbf{V}^T$ . Из равенства

$$\mathbf{VPV}^T - r\mathbf{P}(\mathbf{V}^T)^2 = -\mathbf{qq}^T \mathbf{V}^T$$

получим  $\mathbf{P}(\mathbf{V}^T)^2 \mathbf{y} = \mathbf{0}$ . Следовательно,

$$\mathbf{q}^T (\mathbf{V}^T)^2 \mathbf{y} = \mathbf{h}^T \mathbf{P}(\mathbf{V}^T)^2 \mathbf{y} = \mathbf{0}.$$

Продолжая таким образом, приходим к заключению, что

$$\mathbf{y}^T[\mathbf{q}, \mathbf{Vq}, \dots, \mathbf{V}^{n-1}\mathbf{q}] = \mathbf{0}.$$

Это противоречит предположению об управляемости пары  $(\mathbf{V}, \mathbf{q})$ . Утверждение 3 доказано.

Таким образом, алгоритм прямого синтеза стабилизирующей обратной связи  $\mathbf{u}(k) = \mathbf{Fx}(k)$  в системе (1.1), при которой выполняется условие (1.2), заключается в следующем. На основе данных о поведении системы (1.3) находим матрицу  $\mathbf{Y} = \mathbf{X}_0^T (\mathbf{X}_0 \mathbf{X}_0^T)^{-1}$  и матрицу  $\mathbf{Z}$ , столбцы которой образуют базис  $\ker(\mathbf{X}_0)$ . Вычисляем матрицы  $\mathbf{V} = \mathbf{X}_1 \mathbf{Y}$  и  $\mathbf{W} = \mathbf{X}_1 \mathbf{Z}$ . Проверяем управляемость пары  $(\mathbf{V}, \mathbf{W})$ . Задаем параметр  $r$  из условия  $0 < r \max_i |\lambda_i(\mathbf{V})| \leq \rho$ . Решаем матричное уравнение Сильвестра (2.7). Вычисляем матрицу  $\mathbf{R}$  по формуле (2.8). При этом собственные числа матрицы  $\mathbf{V} + \mathbf{WR}$  будут находиться в области  $|\lambda_i(\mathbf{V} + \mathbf{WR})| \leq \rho$ ,  $i = 1, n$ . Далее вычисляем матрицу  $\mathbf{D} = \mathbf{Y} + \mathbf{ZR}$  и матрицу обратной связи  $\mathbf{F} = \mathbf{U}_0 \mathbf{D}$ . Собственные числа матрицы  $\mathbf{A} + \mathbf{BF}$  удовлетворяют условию (1.2). Замкнутая система асимптотически устойчива.

**3. Численный пример.** Рассмотрим систему (1.1) с матрицами:

$$\mathbf{A} = \begin{bmatrix} -0.7137 & -0.3071 & 1.8129 & 0.4256 & 0.2648 \\ -1.4595 & 0.1654 & -2.2034 & -0.1081 & -1.1426 \\ 0.2885 & 0.0123 & -1.3014 & -0.6484 & -0.5537 \\ 0.3935 & -2.2585 & 0.9719 & -0.9385 & -0.7729 \\ -0.7640 & -0.3322 & 0.6615 & 0.5816 & -0.1766 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -1.1010 & 1.3071 \\ -0.4901 & -2.1556 \\ 0.2701 & 0.6810 \\ 1.0599 & -0.5786 \\ 1.7190 & -2.0371 \end{bmatrix}.$$

Система является управляемой, но не является асимптотически устойчивой, поскольку  $\max_i |\lambda_i(\mathbf{A})| = 2.0183$ . Необходимо найти матрицу обратной связи  $\mathbf{F}$ , при которой собственные числа матрицы замкнутой системы удовлетворяют условию  $\max_i |\lambda_i(\mathbf{A} + \mathbf{BF})| \leq \rho$ ,  $\rho = 0.25$ .

Результаты моделирования, на основе которых выполнялся синтез матрицы обратной связи, получены при  $\mathbf{x}(0) = \mathbf{0}$  и значениях  $\mathbf{u}(k)$ ,  $k = 0, 9$  в виде случайного процесса с нормальным распределением с нулевым средним и единичной матрицей ковариаций. Матрица значений входа

$$\mathbf{U}_0 = [\mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(9)] = \\ = \begin{bmatrix} -0.7138 & 0.2885 & -0.7640 & 0.1654 & -2.2585 & 1.8129 & -1.3014 & 0.6615 & -0.1081 \\ -1.4595 & 0.3935 & -0.3071 & 0.0123 & -0.3322 & -2.2034 & 0.9719 & 0.4256 & -0.6484 \end{bmatrix}.$$

Матрицы

$$\mathbf{X}_0 = \begin{bmatrix} 0 & -1.1218 & -1.7279 & -2.8818 & 12.879 & -25.864 & 39.366 & -37.586 & -56.423 \\ 0 & 3.4959 & 1.8361 & 5.6312 & -3.1402 & -5.2546 & 33.208 & -72.29 & 142.24 \\ 0 & -1.1867 & 0.5858 & 6.2925 & -10.051 & 16.72 & -17.059 & -16.663 & 131.05 \\ 0 & 0.0879 & -10.844 & 6.565 & -9.397 & 1.6191 & 39.562 & -150.46 & 306.15 \\ 0 & 1.7461 & -1.6522 & -5.6054 & 9.5608 & -25.804 & 45.67 & -41.665 & -38.171 \end{bmatrix},$$

$$\mathbf{X}_1 = \begin{bmatrix} -1.1218 & -1.7279 & -2.8818 & 12.879 & -25.864 & 39.366 & -37.586 & -56.423 & 353.62 \\ 3.4959 & 1.8361 & 5.6312 & -3.1402 & -5.2546 & 33.208 & -72.29 & 142.24 & -170.91 \\ -1.1867 & 0.5858 & 6.2925 & -10.051 & 16.72 & -17.059 & -16.663 & 131.05 & -362.92 \\ 0.0879 & -10.844 & 6.565 & -9.397 & 1.6191 & 39.562 & -150.46 & 306.15 & -473.65 \\ 1.7461 & -1.6522 & -5.6054 & 9.5608 & -25.804 & 45.67 & -41.665 & -38.171 & 268.4 \end{bmatrix}.$$

В соответствии с описанным выше алгоритмом последовательно вычисляем матрицы  $\mathbf{Y}, \mathbf{Z}, \mathbf{V}, \mathbf{W}$ :

$$\mathbf{Y} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -0.0645 & 0.082 & -0.2109 & 0.0277 & -0.1006 \\ 0.0533 & 0.0495 & 0.0304 & -0.031 & -0.04 \\ 0.0964 & 0.1107 & -0.1352 & -0.0025 & -0.2137 \\ 0.0724 & 0.004 & 0.0516 & -0.0152 & -0.0347 \\ -0.0379 & 0.0828 & -0.2615 & 0.0455 & -0.1688 \\ -0.0527 & 0.0374 & -0.0372 & -0.0077 & 0.0281 \\ -0.0302 & -0.0121 & 0.0899 & -0.0274 & 0.0897 \\ -0.0058 & -0.0117 & 0.056 & -0.0109 & 0.0435 \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} 0.7388 & 0.4613 & -0.4487 & 0.2001 \\ -0.4414 & 0.2712 & -0.5195 & -0.1603 \\ 0.1225 & -0.6203 & -0.426 & 0.0223 \\ 0.0831 & 0.3246 & 0.3259 & -0.3242 \\ -0.4177 & 0.2389 & -0.1136 & 0.7369 \\ 0.196 & -0.3226 & 0.1982 & 0.4646 \\ 0.1372 & 0.0541 & 0.384 & 0.2297 \\ -0.0516 & 0.2282 & 0.2021 & 0.1177 \\ -0.0543 & 0.0852 & 0.0232 & 0.0561 \end{bmatrix},$$

$$\mathbf{V} = \begin{bmatrix} -0.5186 & -0.571 & 3.0307 & 0.1568 & 1.0431 \\ -1.2592 & 0.3941 & -2.841 & 0.0101 & -1.866 \\ 0.1992 & -0.0496 & -1.1686 & -0.6695 & -0.3533 \\ 0.1775 & -2.0901 & 0.0913 & -0.7386 & -1.2573 \\ -1.0687 & 0.0798 & -1.24 & 1.0012 & -1.3916 \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} -2.2686 & 1.317 & -0.0533 & -0.3854 \\ 2.9184 & -0.3834 & -1.7508 & 3.6537 \\ -0.8807 & 0.0315 & 0.6449 & -1.3054 \\ 1.2094 & -1.0286 & 0.4799 & -0.5789 \\ 3.5379 & -2.0568 & 0.0859 & 0.597 \end{bmatrix}.$$

Далее решаем уравнение Сильвестра (2.7). Параметр  $r$  выбираем из условия  $0 < r \max_i |\lambda_i(\mathbf{V})| \leq \rho$ . В нашем случае  $\rho = 0.25$ ,  $\max_i |\lambda_i(\mathbf{V})| = 2.0265$ . Следовательно, значения параметра  $r$  можно выбрать в интервале  $(0, 0.1234)$ .

Например, при  $r = 0.1$  решение уравнения Сильвестра (2.7) равно

$$\mathbf{P} = \begin{bmatrix} 9.8573 & -60.979 & 23.988 & 10.599 & -18.951 \\ 59.215 & -180.84 & 27.174 & 121.54 & -19.161 \\ 38.279 & -123.22 & 23.347 & 76.399 & -16.837 \\ -47.264 & 104.02 & 1.7165 & -112.94 & -4.4827 \\ -65.457 & 214.54 & -45.943 & -115.33 & 39.417 \end{bmatrix}.$$

Вычисляем матрицу  $\mathbf{R}$  по формуле (2.8):

$$\mathbf{R} = \begin{bmatrix} 0.0518 & 0.1229 & 0.8046 & 0.1269 & 0.5325 \\ 0.0869 & -0.1193 & -0.7022 & -0.1882 & -0.3897 \\ -0.16 & 0.0697 & 0.343 & 0.1612 & 0.124 \\ 0.2876 & -0.0949 & -0.425 & -0.2524 & -0.1029 \end{bmatrix}.$$

Находим матрицы  $\mathbf{D} = \mathbf{Y} + \mathbf{Z}\mathbf{R}$  и  $\mathbf{F} = \mathbf{U}_0\mathbf{D}$ . Искомая матрица обратной связи

$$\mathbf{F} = \begin{bmatrix} -0.7794 & 0.4061 & 1.0673 & 0.7915 & 0.483 \\ -0.5878 & -0.1686 & -0.1621 & 0.1187 & -0.2894 \end{bmatrix}.$$

Собственные числа матрицы замкнутой системы  $\mathbf{A} + \mathbf{BF}$  равны:  $-0.1728 + 0.1017i$ ;  $-0.1728 - 0.1017i$ ;  $0.0181 + 0.0695i$ ;  $0.0181 - 0.0695i$ ;  $-0.0333$ . Максимальное значение модуля собственного числа  $\max_i |\lambda_i(\mathbf{A} + \mathbf{BF})| = 0.2005 < \rho$ . Замкнутая система асимптотически устойчива.

Для сравнения приведем матрицу обратной связи, полученную при  $r = 0.05$ :

$$\mathbf{F} = \begin{bmatrix} -0.8583 & 0.4128 & 1.1526 & 0.8406 & 0.4914 \\ -0.6032 & -0.1728 & -0.1411 & 0.1269 & -0.2892 \end{bmatrix}.$$

Собственные числа матрицы замкнутой системы равны:  $0.039 + 0.0698i$ ;  $0.039 - 0.0698i$ ;  $-0.0928$ ;  $-0.0784 + 0.05i$ ;  $-0.0784 - 0.05i$ . Значение  $\max_i |\lambda_i(\mathbf{A} + \mathbf{BF})| = 0.093 < \rho$ . Замкнутая система асимптотически устойчива. Заметим, что меньшему значению параметра  $r$  соответствует меньшее значение  $\max_i |\lambda_i(\mathbf{A} + \mathbf{BF})|$ .

Все вычисления проводились в системе компьютерной математики Scilab с применением функций для нахождения ядра матрицы и решения уравнения Сильвестра.

**Заключение.** Предложен алгоритм решения задачи синтеза стабилизирующей обратной связи для линейной дискретной многосвязной системы с неизвестными матрицами объекта управления. Особенность алгоритма заключается в том, что алгоритм не предусматривает решения задачи идентификации. Матрица обратной связи находится на основе решения матричного уравнения Сильвестра с матрицами, составленными по результатам наблюдения за поведением системы. Получены условия существования решения задачи синтеза. Рассмотрен численный пример.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Hou Z.S., Wang Z.* From Model-Based Control To Data-Driven Control: Survey. Classification and Perspective // Information Sciences. 2013. V. 235. P. 3–35.
2. *Bazanella A.S., Campestrini C., Eckhard D.* The Data-Driven Approach To Classical Control Theory // Annual Reviews in Control. 2023. V. 56. Article 100906.
3. *Waarde H.J., Eising J., Trentelman H.L., Camlibel M.K.* Data Informativity: A New Perspective on Data-Driven Analysis and Control // IEEE Transactions on Automatic Control. 2020. V. 65. № 11. P. 4753–4768.
4. *Khaki-Sedigh A.* An Introduction to Data-Driven Control Systems. John Wiley & Sons, 2023.
5. *Novara C., Formentin S.* Data-Driven Modeling, Filtering and Control: Methods and Applications. London: IET, 2023.
6. *Dai T., Sznaier M.* Data-driven Quadratic Stabilization and LQR Control of LTI Systems // Automatica. 2023. V. 153. Article 111041.
7. *Pellegrino F., Blanchini F., Fenu G., Salvato E.* Data-Driven Dynamic Relatively Optimal Control // European Journal of Control. 2023. V. 74. Article 100839.
8. *Коган М.М., Степанов А.В.* Синтез субоптимальных робастных регуляторов на основе априорных и экспериментальных данных // А и Т. 2023. № 8. С. 24–42.
9. *Коган М.М., Степанов А.В.* Синтез обобщенного  $H_\infty$ -субоптимального управления по экспериментальным и априорным данным // А и Т. 2024. № 1. С. 3–20.
10. *Дмитрук Н.М., Манжуллина Е.А.* Оптимальное управление линейными стационарными дискретными системами без предварительной параметрической идентификации // А и Т. 2022. № 2. С. 3–21.
11. *Verheijen P.C.N., Breschi V., Lazar M.* Handbook of Linear Data-driven Predictive Control: Theory, Implementation and Design // Annual Reviews in Control. 2023. V. 56. 2023. Article 100914.
12. *Berberich J., Köhler J., Müller M.A., Allgöwer F.* Data-Driven Model Predictive Control With Stability and Robustness Guarantees // IEEE Transactions on Automatic Control. 2021. V. 66. № 4. P. 1702–1717.
13. *Mukherjee S., Hossain R.R.* Data-Driven Pole Placement in LMI Regions with Robustness Guarantees // IEEE61st Conf. on Decision and Control (CDC). Cancun, Mexico, 2022. P. 4010–4015.
14. *Persis C., Tesi P.* Formulas for Data-Driven Control: Stabilization, Optimality, and Robustness // IEEE Transactions on Automatic Control. 2020. V. 65. № 3. P. 909–924.

15. *Baggio G, Bassett D.S., Pasqualetti F.* Data-Driven Control of Complex Networks // *Nat. Commun.* 2021. V. 12. Article 1429.
16. *Martin T., Schön T.B., Allgöwer F.* Guarantees for Data-driven Control of Nonlinear Systems Using Semidefinite Programming: A Survey // *Annual Reviews in Control.* 2023. V. 56. Article 100911.
17. *Strässer R., Berberich J., Allgöwer F.* Robust Data-Driven Control for Nonlinear Systems Using the Koopman Operator // *IFAC-PapersOnLine.* 2023. V. 56. № 2. P. 2257–2262.
18. *Rueda-Escobedo J.G., Fridman E., Schiffer J.* Data-Driven Control for Linear Discrete-Time Delay Systems // *IEEE Transactions on Automatic Control.* 2022. V. 67. № 7. P. 3321–3336.
19. *Кюо Б.* Теория и проектирование цифровых систем управления. М.: Машиностроение, 1986.
20. *Гантмacher Ф.Р.* Теория матриц. М.: Физматлит, 2010.
21. *Hearon J.Z.* Nonsingular Solution of  $TA - BT + C$  // *Linear Algebra and Its Applications.* 1977. V. 16. P. 57–63.
22. *Souza E.* Controllability, Observability and the Solution  $AX - XB = C$  // *Linear Algebra and Its Applications.* 1981. V. 39. P. 167–188.

## МАКСИМИЗАЦИЯ ДАЛЬНОСТИ ПОЛЕТА ДЛЯ УПРОЩЕННОЙ МОДЕЛИ ЛЕТАТЕЛЬНОГО АППАРАТА<sup>1</sup>

© 2024 г. Е. В. Малых<sup>a</sup>, \*, О. Ю. Черкасов<sup>a, b, \*\*</sup>

<sup>a</sup>МГУ им. М.В. Ломоносова, Москва, Россия

<sup>b</sup>Университет МГУ-ППИ, Шэнъянь, Китай

\*e-mail: wyvling@gmail.com

\*\*e-mail: oyuche@yandex.ru

Поступила в редакцию 14.05.2024 г.

После доработки 15.05.2024 г.

Принята к публикации 22.07.2024 г.

Рассматривается задача максимизации дальности полета для упрощенной модели летательного аппарата с учетом влияния количества топлива на динамику центра масс. Предполагается, что движение происходит в вертикальной плоскости под действием однородных сил тяжести и сопротивления среды. Кроме того, имеются активная сила тяги и возможность изменять угол наклона траектории. Эти параметры приняты в качестве управлений. Построена область в пространстве исходных переменных, для которой решена задача оптимального синтеза. Показано, что в этой области тяга может быть максимальной, нулевой или особой. Установлено количество и порядок следования участков траектории с соответствующей тягой.

**Ключевые слова:** двумерная задача Годдарда, обобщение задачи о брахистохроне, особое управление, оптимальный синтез

**DOI:** 10.31857/S0002338824060037, **EDN:** SVSBYV

## FLIGHT RANGE MAXIMIZATION PROBLEM FOR A SIMPLIFIED AIRCRAFT MODEL

Е. В. Малых<sup>a</sup>, \*, О. Ю. Черкасов<sup>a, b, \*\*</sup>

<sup>a</sup>Lomonosov Moscow State University

<sup>b</sup>Shenzhen MSU-BIT University

\*e-mail: wyvling@gmail.com

\*\*e-mail: oyuche@yandex.ru

The flight range maximization problem for a simplified aircraft model is considered, considering the influence of the amount of fuel on the dynamics of the center of mass. It is assumed that the motion occurs in a vertical plane under the influence of homogeneous gravity forces and homogeneous resistance of the medium. In addition, there are an active thrust force and the ability to change the angle of inclination of the trajectory. These parameters are accepted as controls. An area in the space of initial variables is constructed, for which the problem of optimal synthesis is solved. It is shown that in this area the thrust can be of maximum value, zero value or singular value. The number and consequence of the trajectory arcs with the corresponding thrust have been established.

**Keywords:** Two-Dimensional Goddard Problem, Brachistochrone Problem Generalization, Singular control, Optimal Synthesis

**Введение.** Одной из первых задач оптимизации формы траектории движения точки в вертикальной плоскости под действием силы тяжести является задача о брахистохроне [1]. В 1697 г. И. Бернулли сформулировал следующую задачу: найти форму кривой, двигаясь по которой

<sup>1</sup> Работа поддержана Министерством науки и Высшего образования Российской Федерации по программе "Сверхзвук" (договор №075-15-2020-923).

в вертикальной плоскости под действием только силы тяжести материальная точка переместится из одной заданной точки в другую заданную точку за минимальное время. Обобщения задачи о брахистохроне, в которых учитывалось влияние сухого трения, представлены в работах [2–9]. Влияние вязкого трения изучено, например, в [10, 11]. Задача с ограничениями на силу реакции опоры кривой или на кривизну траектории рассматривалась в публикациях [12, 13] соответственно. В [14–16] исследовались фазовые ограничения различного вида. Обобщения задачи о брахистохроне при наличии разгоняющей силы описаны в [10, 17–19]. В [10] приведены результаты моделирования для случая разгоняющей силы, пропорциональной скорости. В [17] задача изучалась при наличии постоянной разгоняющей силы и силы линейного вязкого трения. В [18] исследован случай квазипостоянной разгоняющей силы, а в [19] установлены качественные свойства траекторий при наличии нелинейного сопротивления и постоянной разгоняющей силы. При этом масса точки предполагалась постоянной. В [20] рассматривалась задача о брахистохроне при переменной тяге, являющейся управлением, и штрафе за ее расход. Задача о брахистохроне для случая, когда масса частицы меняется по известному закону, приведена в [21] при действии сухого трения и в [22] при действии вязкого трения. В [23] исследовалась задача оптимизации формы траектории движения центра масс летательного аппарата в вертикальной плоскости в случае переменной массы, меняющейся в зависимости от заданного закона тяги двигателя. Модель летательного аппарата, в которой в качестве управления принят угол наклона траектории, в [23] называется «промежуточной», мы будем использовать для нее термин «упрощенная». В рамках этой модели предполагается, что подъемная сила, необходимая для изменения формы траектории, не влияет на силу сопротивления. Данное допущение справедливо при достаточно малых углах атаки. В такой постановке подъемная сила в задаче оптимизации формы траектории точечной модели летательного аппарата играет роль реакции опорной кривой в задаче о брахистохроне для точечной массы. Обсуждение границ применимости такой модели приводится, например, в работах [23–25].

Задача программирования тяги летательного аппарата вдоль траектории, в частности максимизации высоты полета с заданным количеством топлива, была сформулирована Р. Годдардом в 1919 г. [26]. Эта задача исследовалась методами вариационного исчисления (например, [27–29]) и стала одной из первых, решаемых методами теории оптимального управления. Два частных случая, а именно один с линейной зависимостью сопротивления от скорости, а другой с квадратичной зависимостью от скорости, были рассмотрены в [28]. Было установлено, что оптимальная программа изменения тяги состоит, как правило, из дуги максимальной тяги, затем промежуточной тяги и заканчивается нулевой тягой. Такая структура имеет место в случае достаточно большого верхнего предела тяги, когда промежуточная тяга удовлетворяет ограничениям на протяжении всего процесса и остается допустимой. Позже результаты были обобщены на случай экспоненциальной зависимости плотности воздуха от высоты и переменного ускорения силы тяжести (например, [30–32]). В [33] для задачи максимизации дальности горизонтального полета было обнаружено, что оптимальная тяга имеет ту же структуру, что и для задачи о вертикальном подъеме ракеты. Упомянутые работы инициировали развитие методов решения задач оптимального управления, содержащих особые участки. Другие обобщения задачи Годдарда при ограничениях на время полета и с учетом зависимости силы лобового сопротивления от высоты представлены, например, в статье [34]. Там также было показано, что при определенных условиях оптимальное управление может иметь более сложную структуру переключения, главным образом из-за возможности появления второй дуги полной тяги после дуги особой тяги. В [35] решена задача выбора оптимального режима расходования топлива и оптимальной программы для направления тяги при движении в вертикальной плоскости при отсутствии сопротивления воздуха. Для решения задачи сформулирован и применен метод первой вариации функционала, известный также как метод множителей Лагранжа для задач управления. В дальнейшем этот метод получил название метода Охочимского-Понtryагина [36].

В [24] была рассмотрена двумерная задача Годдарда, в которой искался оптимальный закон изменения тяги при условии, что закон изменения угла наклона траектории задан. Задача одновременного управления углом наклона траектории и расходом массы была рассмотрена в [37] для простейшего случая линейного вязкого трения.

В работе исследуется задача максимизации дальности полета с учетом влияния количества топлива на динамику точечной массы. В этом состоит главное отличие данной статьи от публикаций [17–20], где предполагалось, что сила тяги не влияет на массу точки. В отличие от работ [21, 22] скорость изменения массы точки рассматривается как управляющая переменная наряду с управлением углом наклона. По сравнению со статьями [23–25] закон изменения угла наклона траектории рассматривается как управление наряду с управлением

тягой. В [32] представлено исследование задачи Годдарда для невертикальных траекторий для достаточно общей модели аэродинамического сопротивления и центрального поля сил тяжести. Численным моделированием подтверждено существование оптимальной траектории с особой дугой. В настоящей работе результаты, полученные в [37], обобщаются на случай нелинейного закона сопротивления среды. В частности, для определенной области исходных переменных аналитически строится оптимальный синтез управления тягой.

**1. Постановка задачи.** Рассматривается движение материальной точки в вертикальной плоскости в однородном поле сил тяжести под действием сил тяги и сопротивления среды. Задача состоит в определении формы траектории, обеспечивающей максимизацию горизонтальной координаты точки при движении ее из заданного начального состояния за фиксированный промежуток времени при определенном количестве топлива. Математическая модель движения центра масс твердого тела в атмосфере, включающая предположение о тяге вдоль траектории, широко применяется в различных исследованиях (например, [23]). Модель, используемая в данной работе, отличается от модели [37] нелинейной зависимостью силы сопротивления от скорости достаточно общего вида. Действующие силы представлены на рис. 1. Через  $Q(V)$  обозначена сила сопротивления, где  $V$  – модуль скорости точки. Предполагается, что  $Q(V)$  – дважды дифференцируемая выпуклая вниз функция скорости. Зависимостью силы сопротивления от высоты будем пренебречь, полагая, что движение происходит в таком диапазоне высот, что плотность среды можно считать постоянной далее,  $M$  – масса точки,  $g$  – ускорение свободного падения,  $L$  – подъемная сила. Если вместо движения центра масс летательного аппарата исследовать движение точки вдоль заданной кривой, как в классической

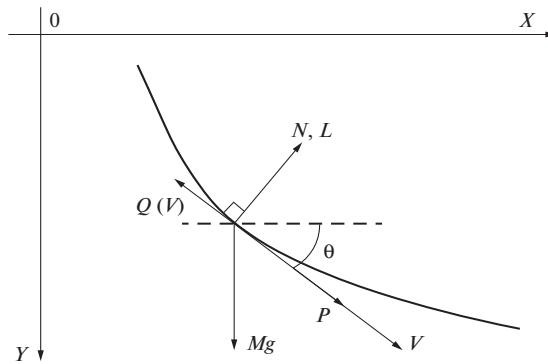


Рис. 1. Действующие силы и переменные

задаче о брахистохроне, то вместо подъемной силы следует рассматривать силу реакции опорной кривой  $N$ . Подъемная сила  $L$  или сила реакции опоры  $N$  являются управлением. Угол наклона траектории обозначен через  $\theta$ . В рамках упрощенной модели летательного аппарата угол  $\theta$  может рассматриваться в качестве управления вместо  $N$  и  $L$ . Второе управление, сила тяги  $P$ ,  $P = cU$ , где  $U$  – скорость расхода массы, а  $c$  – скорость истечения рабочего тела.

Уравнения движения имеют следующий вид [23]:

$$\dot{X} = V \cos \theta, \dot{Y} = V \sin \theta, \dot{V} = (-Q(V) + cU) / M - g \sin \theta, \dot{M} = -U.$$

Здесь через  $X$  и  $Y$  обозначены горизонтальная и вертикальная координаты точки, Переменные  $\theta$  и  $U$  рассматриваются в качестве управлений, принадлежащих классу кусочно-непрерывных функций времени. Дифференцирование уравнений движения осуществляется относительно размерного времени  $\tau$ . На управление  $\theta$  ограничения не наложены, а управление  $U(\tau)$  удовлетворяет неравенству  $0 \leq U(\tau) \leq \bar{U}$ , где  $\bar{U} > 0$  – заданная константа. Для дальнейшей работы перейдем к безразмерным переменным по следующим формулам:

$$X = \tilde{X}x, Y = \tilde{Y}y, U = \tilde{U}u, V = \tilde{V}v, \tau = \tilde{\tau}t, M = \tilde{M}m,$$

где  $\tilde{X} = c^2 / g$ ,  $\tilde{Y} = c^2 / g$ ,  $\tilde{U} = \bar{U} / \bar{u}$ ,  $\tilde{V} = c$ ,  $\tilde{\tau} = c / g$ ,  $\tilde{M} = \bar{U}c / (\bar{g}\bar{u})$  – масштабы горизонтальной и вертикальной координат точки, скорости расхода массы, скорости точки, времени и массы

соответственно,  $x, y$  – безразмерные горизонтальная и вертикальная координаты точки соответственно,  $u$  – безразмерная скорость расхода массы,  $u > 0$  – верхнее ограничение на безразмерную скорость расхода массы,  $v$  – безразмерная скорость точки,  $t$  – безразмерное время,  $m$  – безразмерная масса точки. Далее производную по безразмерному времени  $t$  будем обозначать точкой. Уравнения движения в безразмерных переменных имеют следующий вид:

$$\dot{x} = v \cos \theta, \dot{y} = v \sin \theta, \dot{v} = (-q(v) + u) / m - \sin \theta, \dot{m} = -u, \quad (1.1)$$

$$0 \leq u \leq \bar{u} \quad (1.2)$$

где  $\bar{u} > 0$  – заданная константа.

Начальные и конечные условия для системы (1.1) запишем как

$$x(0) = x_0, y(0) = y_0, v(0) = v_0, m(0) = m_0, m(T) = m_T > 0, \quad (1.3)$$

$x(T), y(T), v(T)$  – свободны,  $x_0, y_0, v_0$  – заданные параметры  $T$  – заданный момент окончания процесса, определяемый относительно безразмерного времени  $t$ .

Целью управления является минимизация функционала

$$J = -x(T) \rightarrow \min_{\theta, u} \quad (1.4)$$

на траекториях динамической системы (1.1) с краевыми условиями (1.3) при помощи управлений  $\theta$ ,  $u$  с учетом ограничений (1.2). Иначе говоря, нужно максимизировать горизонтальную координату точки в заданный момент  $T$  окончания процесса. Поскольку условия на конечное значение  $y(T)$  отсутствуют, решение задачи (1.1)–(1.4) дает достижимую верхнюю оценку максимально возможной дальности.

**2. Условия оптимальности.** Для исследования поставленной задачи применим принцип максимума Понtryгина [38]. Функция Гамильтона для задачи (1.1)–(1.4) имеет следующий вид:

$$H = \psi_x v \cos \theta + \psi_y v \sin \theta + \psi_v ((-q + u) / m - \sin \theta) - \psi_m u. \quad (2.1)$$

Уравнения относительно сопряженных переменных запишем как

$$\dot{\psi}_x = 0, \dot{\psi}_y = 0, \dot{\psi}_v = -\psi_x \cos \theta - \psi_y \sin \theta + \frac{dq}{dv} \cdot \psi_v / m, \dot{\psi}_m = (u - q) \psi_v / (m^2). \quad (2.2)$$

Вдоль оптимальной траектории выполняется соотношение  $H = C$ , где  $C$  – неизвестная константа,  $C \neq 0$  в задаче с фиксированным временем окончания процесса. Из условий трансверсальности получаем конечные значения для сопряженных переменных:

$$\psi_x(T) = 1, \psi_y(T) = 0, \psi_v(T) = 0, \psi_m(T) = a, \quad (2.3)$$

где  $a$  – неизвестная константа.

Из (2.2), (2.3) следует  $\psi_x(t) = 1, \psi_y(t) = 0 \forall t \in [0, T]$ .

Максимизация функции  $H$  по управлению  $\theta$  приводит к соотношениям:

$$\frac{\partial H}{\partial \theta} = 0 \Leftrightarrow -\psi_v \cos \theta - v \sin \theta = 0. \quad (2.4)$$

Поскольку движение с углом  $\theta = \pm\pi/2$  представляется неоптимальным в смысле достижения максимальной дальности по горизонтали, будем считать, что  $\cos \theta \neq 0$ . Тогда из уравнения (2.4) следует

$$\psi_v = -v \tan \theta. \quad (2.5)$$

Заметим, что в соответствии с соотношением (2.5) экстремальное управление углом  $\theta$  является функцией одного класса гладкости с сопряженной переменной  $\psi_v$ . С учетом (2.5) получаем

$$\frac{\partial^2 H}{\partial \theta^2} \leq 0 \Leftrightarrow \psi_v \sin \theta - v \cos \theta = -v / \cos \theta \leq 0 \Leftrightarrow \cos \theta > 0. \quad (2.6)$$

Из (2.6) следует тривиальный вывод: для максимизации дальности нужно двигаться в сторону увеличения дальности.

Представим функцию (2.1) в виде

$$H = H_0 + H_1 u,$$

где

$$H_0 = v \cos \theta + (-q / m - \sin \theta) \psi_v, \quad H_1 = \psi_v / m - \psi_m.$$

Функция  $H_1$  называется функцией переключения.

Максимизация функции  $H$  по управлению  $u$  приводит к следующему правилу выбора экстремальной тяги:

$$u = \begin{cases} \bar{u}, & H_1 > 0, \\ u_s, & H_1 \equiv 0, \\ 0, & H_1 < 0. \end{cases} \quad (2.7)$$

Случаи  $H_1 > 0$  и  $H_1 < 0$  отвечают регулярному управлению. Если существует отрезок времени  $\sigma$   $[0, T]$ , такой, что  $t \in \sigma$  функция переключения обращается в нуль тождественно, т.е. выполнено  $H_1 = 0$ , то на отрезке времени  $\sigma$  возникает сингулярное или особое управление [39]. В формуле (2.7) через  $u_s$  обозначено искомое особое управление. На отрезке особого управления выполнены следующие соотношения:

$$H_0 \equiv C, \quad H_1 \equiv 0, \quad \frac{dH_1}{dt} \equiv 0, \quad \frac{d^2H_1}{dt^2} \equiv 0.$$

Если эти соотношения линейно независимы, то из уравнения  $d^2H_1 / dt^2 \equiv 0$  может быть найдено особое управление. В таком случае оно называется особым управлением первого порядка.

Дифференцируя соотношение  $H_1 = \psi_v / m - \psi_m$  по времени в силу систем (1.1) и (2.2) и учитывая выражение (2.4), получим

$$\dot{H}_1 = \left( \left( \dot{q}_v + q \right) (-\operatorname{vtg} \theta) - m \cos \theta \right) / m^2 = 0, \quad \forall t \in \sigma \subset [0, T]. \quad (2.8)$$

Здесь и далее через  $\dot{q}_v$ ,  $\ddot{q}_{vv}$  обозначены первая и вторая производные от силы сопротивления по скорости. Для вычисления  $d^2H_1 / dt^2$  вначале продифференцируем соотношение (2.4) с учетом уравнений (1.1) и (2.2) и находим дифференциальное уравнение для управляющей переменной  $\theta$ :

$$\dot{\theta} = \left( 1 + \left( v \dot{q}_v + q - u \right) \sin \theta / m \right) \cos \theta / v. \quad (2.9)$$

Конечное условие для уравнения (2.9) получается из (2.3), (2.4):

$$\theta(T) = -\operatorname{arctg}(\psi_v / v) |_{t=T} = 0. \quad (2.10)$$

Дифференцируя соотношение (2.8) в силу систем (1.1), (2.2) и учитывая (2.9), запишем

$$\frac{d}{dt}(\dot{H}_1) = \frac{d}{dt} \left( \left( \left( \dot{q}_v + q \right) (-\operatorname{vtg} \theta) - m \cos \theta \right) / m^2 \right) = 0,$$

или

$$\left( 1 - \sin^2 \theta / v + \left( \ddot{q}_{vv} + \dot{q}_v \right) / \left( \dot{q}_v + q \right) \right) (u - q) + \dot{q}_v \sin^2 \theta - m \sin \theta \left( \left( \ddot{q}_{vv} + \dot{q}_v \right) / \left( \dot{q}_v + q \right) - 1 / v \right) = 0,$$

откуда, подставив выражение для массы из соотношения (2.8), найдем выражение для особого управления как функции от скорости и угла наклона траектории:

$$u_s(v, \theta) = q - \operatorname{tg}^2 \theta \left( v \ddot{q}_{vv} + \dot{q}_v \left( v - \sin^2 \theta \right) - q \right) / \left( 1 - \sin^2 \theta / v + \left( \ddot{q}_{vv} + \dot{q}_v \right) / \left( \dot{q}_v + q \right) \right). \quad (2.11)$$

**3. Уравнение особой поверхности.** Соотношения  $H_0 \equiv C$ ,  $H_1 \equiv 0$ ,  $dH_1 / dt \equiv 0$  представляют собой семейство особых поверхностей задачи (1.1)–(1.4). Выражая из формулы (2.8) массу, получим

$$m = -v \left( \dot{q}_v + q \right) \sin \theta / \cos^2 \theta. \quad (3.1)$$

Значение  $m$  в (3.1) положительно, если  $\theta \in (-\pi / 2, 0)$ . Значит, особый участок возможен только на нисходящей ветви траектории. Кроме того, как следует из (2.10), (3.1) и условия положительности массы, участок особой, промежуточной тяги не может примыкать к концу оптимальной траектории. Поверхность особого управления в пространстве  $(\theta, v, m)$  представлена на рис. 2.

Вычислим частные производные от  $m(v, \theta)$  по скорости и углу. Из неравенства

$$\frac{\partial m(v, \theta)}{\partial v} = -\left(vq_{vv}'' + q_v'(v+1) + q\right)\sin\theta / \cos^2\theta > 0, \theta \in (-\pi/2; 0)$$

следует, что масса при движении по особой поверхности возрастает с увеличением скорости. Из неравенства

$$\frac{\partial m(v, \theta)}{\partial \theta} = -v\left(q_v' + q\right)\left(1 + \sin^2\theta\right) / \cos^3\theta < 0, \theta \in (-\pi/2; 0)$$

заключаем, что масса при движении по особой поверхности убывает с увеличением угла. Эти свойства будут использованы при построении синтеза оптимального управления тягой.

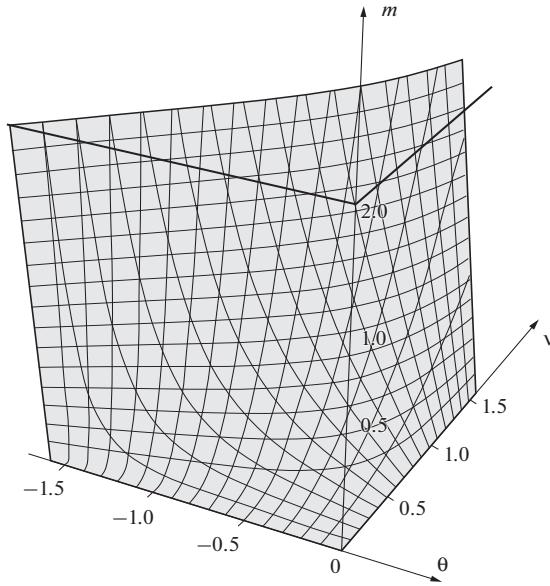


Рис. 2. Поверхность особого управления

**4. Синтез оптимального управления.** В результате применения принципа максимума задача оптимального управления (1.1)–(1.4) сведена к краевой задаче для системы уравнений (1.1)–(1.3), (2.9), (2.10), не содержащей сопряженных переменных. При решении задачи управление тягой определяется в соответствии с правилом (2.7).

Пусть выполнено неравенство

$$m > -v\left(q_v' + q\right)\sin\theta / \cos^2\theta. \quad (4.1)$$

В этом случае точка в пространстве  $(\theta, v, m)$  лежит выше особой поверхности и из выражения (2.8) следует  $\dot{H}_1 < 0$ . Следовательно, принципу максимума удовлетворяет только переключение с управления  $u = \bar{u}$  на управление  $u = 0$ . Если же точка лежит ниже особой поверхности, т.е. выполнено неравенство

$$m < -v\left(q_v' + q\right)\sin\theta / \cos^2\theta, \quad (4.2)$$

то  $\dot{H}_1 > 0$  и оптимальным может быть только переключение с управления  $u = 0$  на управление  $u = \bar{u}$ . Задача состоит в построении оптимального синтеза управления тягой, т.е. нахождение тяги как функции исходных переменных системы (1.1).

Примем следующие предположения.

П р е д п о л о ж е н и е 1. Параметры системы (1.1) и граничные условия (1.2) таковы, что при движении с управлением  $u = 0$  точка находится в области, в которой выполнены неравенства  $\dot{v} < 0, \dot{\theta} > 0$ .

П р е д п о л о ж е н и е 2. Движение происходит в области пространства  $(\theta, v, m)$ , где особое управление (2.11) является допустимым, т.е. верно условие  $u_s \in [0, \bar{u}]$ .

При предположениях 1 и 2 справедливы следующие утверждения 1–4.

Утверждение 1. Если начальная точка в пространстве  $(\theta, v, m)$  лежит выше особой поверхности, то  $u(0) = \bar{u}$ .

Доказательство. Предположим, что условие (4.1) выполнено в начальный момент времени  $t=0$  и пусть  $u(0) = 0$ . Тогда масса остается постоянной и в соответствии с условиями оптимальности переключение с управления  $u = 0$  на управление  $u(0) = \bar{u}$  не оптимально. На рис. 3 изображен фазовый портрет системы, составленной из третьего уравнения системы (1.1) и уравнения (2.9) в случае сопротивления, пропорционального второй степени скорости  $q(v) = v^2$ :

$$\dot{v} = (-v^2 + u) / m - \sin \theta, \dot{\theta} = (1 + 3v^2 \sin \theta / m) \cos \theta / v \quad (4.3)$$

инулевой тяги ( $u=0, m=\text{const}$ ). Подробный анализ системы (4.3) приведен, например, в работе [11]. Серым цветом на рис. 3 отмечена область, для которой возможно удовлетворение конечного условия (2.10). Область, удовлетворяющая предположению 1, отмечена горизонтальной штриховкой. Эта область ограничена кривыми  $\dot{v} = 0, \theta = 0$  и осью ординат. Как видно из рис. 3, вдоль траекторий в заштрихованной области скорость монотонно уменьшается, а угол монотонно увеличивается, и движение с нулевой тягой происходит выше особой поверхности и не пересекает ее. Поэтому в момент окончания процесса конечное условие по массе не может быть выполнено. Следовательно, при указанных начальных условиях движение с тягой  $u = 0$  над особой поверхностью не оптимально. Значит,  $u(0) = \bar{u}$ . Утверждение 1 доказано.

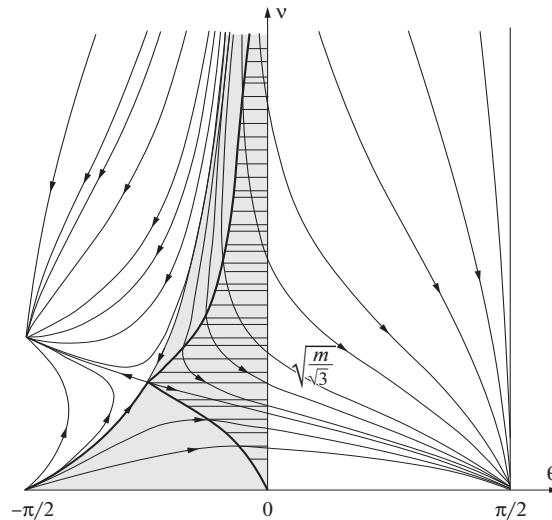


Рис. 3. Фазовый портрет системы (4.3)

Утверждение 2. Если начальная точка в пространстве  $(\theta, v, m)$  лежит выше особой поверхности, то движение с  $u(t) = \bar{u}$  продолжается до достижения особой поверхности или до полного выгорания топлива.

Доказательство. Над особой поверхностью необходимым условиям оптимальности удовлетворяет переключение с управления  $u(t) = \bar{u}$  на управление  $u(t) = 0$ . Пусть конечное условие по массе (достижение плоскости  $m(t) = m_T$ ) выполняется раньше достижения особой поверхности. Тогда необходимо переключиться с максимальной тяги на нулевую тягу и двигаться с ней до выполнения условия  $\theta(T) = 0$ . Если же при достижении поверхности особого управления конечное условие по массе не верно, то аналогично доказательству утверждения 1 получаем, что переключение на нулевую тягу также не оптимально, поскольку конечное условие по массе не будет выполнено. Значит, движение с максимальной тягой должно продолжаться до достижения особой поверхности. Утверждение 2 доказано.

Утверждение 3. Если начальная точка в пространстве  $(\theta, v, m)$  лежит выше особой поверхности, то:

а) если конечная масса достигается раньше достижения особой поверхности, то оптимальная траектория задачи (1.1)–(1.4) состоит из двух дуг: дуги с управлением  $u(t) = \bar{u}$  в начале, затем дуги с управлением  $u(t) = 0$  до окончания процесса;

б) если особая поверхность достигается раньше, чем будет достигнуто конечное условие по массе, тогда оптимальная траектория задачи (1.1)–(1.4) состоит из трех дуг: дуги с управлением  $u(t) = \bar{u}$  в начале, затем дуги с промежуточной тягой  $u(t) = u_S(t)$  и, наконец, дуги с нулевой тягой  $u(t) = 0$  до окончания процесса.

Доказательство. Предположим, что траектория, начинающаяся с дуги максимальной тяги  $u(t) = \bar{u}$  пересекает особую поверхность и оказывается ниже нее. Тогда дальнейшее движение должно продолжаться также с управлением  $u(t) = \bar{u}$ . В этом случае масса убывает быстрее, чем при движении по особой поверхности. Поэтому движение с управлением  $u(t) = \bar{u}$  продолжается до полного выгорания топлива. Значит, в момент достижения конечной массы происходит переключение на режим нулевой тяги  $u(t) = 0$ , что неоптимально, если точка находится ниже особой поверхности. Следовательно, траектория не может пересечь особую поверхность, и в момент прихода на нее следует переключиться на режим промежуточной тяги  $u(t) = u_S$ . При этом сходить с особой поверхности на дугу с максимальной тягой  $u(t) = \bar{u}$  также не оптимально. Значит, движение вдоль особой поверхности продолжается до полного выгорания топлива, после чего оптимальная траектория продолжается до выполнения краевого условия по углу наклона траектории. Утверждение 3 доказано.

Утверждение 4. Если начальная точка лежит ниже особой поверхности, то оптимальная траектория задачи (1.1)–(1.4) состоит из трех дуг: дуги с нулевой тягой  $u(t) = 0$  в начале, затем следует особая дуга с управлением  $u(t) = u_S(t)$  и, наконец, дуга с нулевой тягой  $u(t) = 0$ .

Доказательство аналогично доказательству утверждений 1–3.

Замечание 1. Если предположения 1 и 2 не выполнены, то оптимальная программа изменения тяги может иметь более сложную структуру. Этот случай требует дополнительного исследования.

Замечание 2. Строгое построение оптимального синтеза в одномерной задаче Годдарда также возможно только в определенной области значений переменных и параметров.

**5. Движение в отсутствии сопротивления.** При отсутствии сопротивления из соотношения (2.7) вытекает неравенство  $\dot{H}_1 = -\cos\theta / m < 0$ . Следовательно, особого управления тягой не возникает и только переключение с управления  $u(t) = \bar{u}$  на управление  $u(t) = 0$  удовлетворяет необходимым условиям оптимальности.

Утверждение 5. При отсутствии сопротивления оптимальная траектория задачи (1.1)–(1.4) состоит из двух дуг: дуги максимальной тяги  $u(t) = \bar{u}$  в начале и дуги нулевой тяги  $u(t) = 0$  в конце.

**6. Численное моделирование.** Представлены результаты численного моделирования краевой задачи (1.1)–(1.3), (2.9), (2.10) с найденным законом управления тягой (2.11). Приведенные результаты призваны проиллюстрировать аналитические выводы разд. 4. Для решения краевой задачи применялся метод стрельбы. Краевые значения для переменных были подобраны таким образом, чтобы продемонстрировать желаемые структуры управления. Функция сопротивления  $q$  принимается квадратично зависящей от скорости  $q(v) = v^2$ . Для всех рассмотренных случаев в начальный момент времени принимаются следующие начальные условия:  $x_0(0) = 0, y_0(0) = 0$ . Численное моделирование краевой задачи для функции сопротивления  $q$ , линейно зависящей от скорости  $q(v) = v$  и случая отсутствия сопротивления среды, представлено в [37].

1. Начальная точка лежит выше особой поверхности. Набор краевых значений принимается следующий:  $v(0) = 0.4, m(0) = 2, m(T) = 1, \bar{u} = 2, T = 1.5$ .

Результаты моделирования приведены на рис. 4–6. Для этого набора условий на рис. 4 рассмотрена траектория движения в пространстве  $(\theta, v, m)$  и особая поверхность (3.1), на рис. 5 – траектория движения в плоскости  $(x, y)$ . Точками отмечены моменты переключения управления  $u$ . На рис. 6 изображена структура изменения экстремальной тяги во времени  $u(t) = \bar{u} \rightarrow u(t) = u_S(t) \rightarrow u(t) = 0$ .

2. Начальная точка лежит ниже особой поверхности. Набор краевых значений принимается следующий:  $v(0) = 0.75, m(0) = 2, m(T) = 1.25, \bar{u} = 2, T = 1.25$ .

Для данного набора параметров на рис. 7, 8 представлены соответствующие траектории. На рис. 9 изображена структура изменения экстремальной тяги во времени  $u(t) = 0 \rightarrow u(t) = u_S(t) \rightarrow u(t) = 0$ .

3. Начальная точка лежит на особой поверхности. Набор краевых значений принимается следующий:  $v(0) = 1, m(0) = 2, m(T) = 1.25, \bar{u} = 2, T = 1.25$ . Структура изменения тяги выглядит так:  $u(t) = u_S(t) \rightarrow u(t) = 0$  (рис. 10–12).

4. Начальная точка лежит выше особой поверхности. Набор краевых значений принимается следующий:  $v(0) = 0.2, m(0) = 2, m(T) = 1.5, \bar{u} = 2, T = 0.4$ . Структура тяги выглядит так:  $u(t) = \bar{u} \rightarrow u(t) = 0$  (особый участок отсутствует) (рис. 13–15).

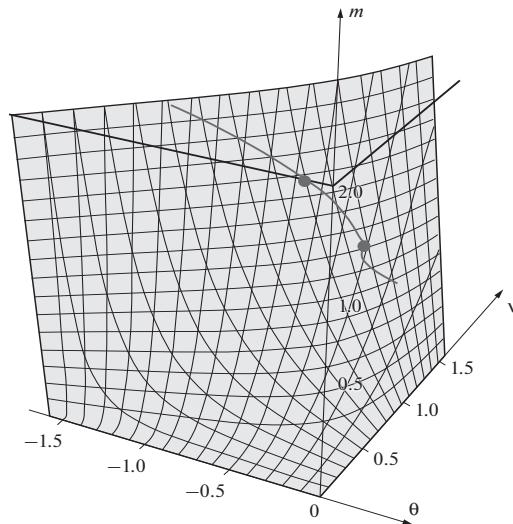


Рис. 4. Экстремальная траектория в пространстве  $(\theta, v, m)$  для  $v(0) = 0.4, m(0) = 2, m(T) = 1, \bar{u} = 2, T = 1.5$

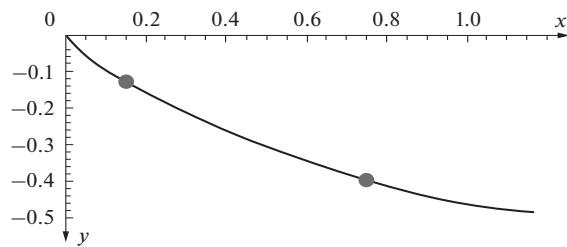


Рис. 5. Экстремальная траектория в плоскости  $(x, y)$  для  $v(0) = 0.4, m(0) = 2, m(T) = 1, \bar{u} = 2, T = 1.5$

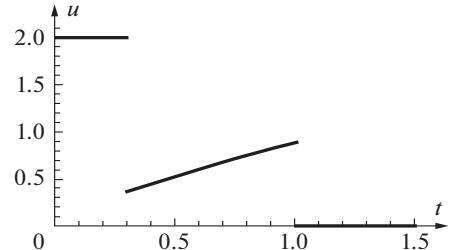


Рис. 6. Изменение тяги во времени  $(u, t)$  на экстремальной траектории для  $v(0) = 0.4, m(0) = 2, m(T) = 1, \bar{u} = 2, T = 1.5$

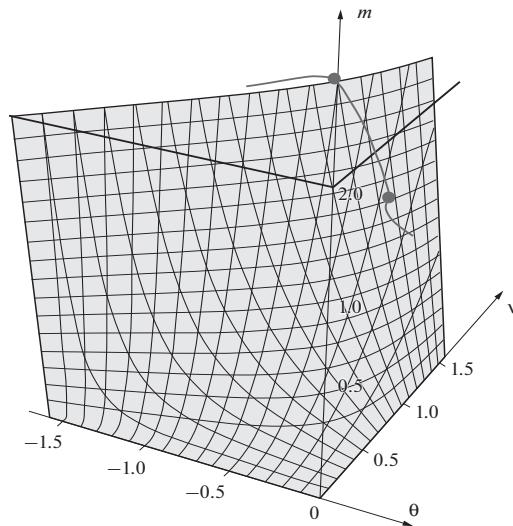


Рис. 7. Экстремальная траектория в пространстве  $(\theta, v, m)$  для  $v(0) = 0.75, m(0) = 2, m(T) = 1.25, \bar{u} = 2, T = 3$

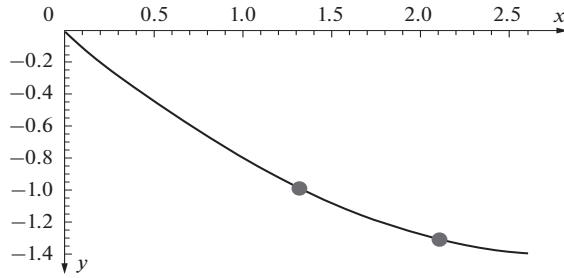


Рис. 8. Экстремальная траектория в плоскости  $(x, y)$  для  $v(0) = 0.75$ ,  $m(0) = 2$ ,  $m(T) = 1.25$ ,  $\bar{u} = 2$ ,  $T = 3$

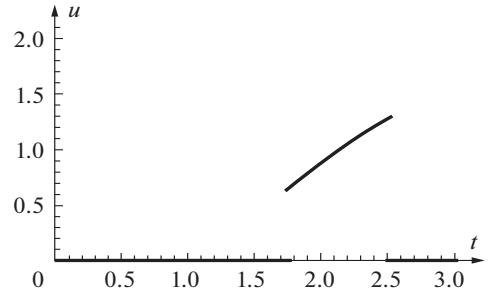


Рис. 9. Изменение тяги во времени  $(u, t)$  на экстремальной траектории для  $v(0) = 0.75$ ,  $m(0) = 2$ ,  $m(T) = 1.25$ ,  $\bar{u} = 2$ ,  $T = 3$

Найденные параметры из численного моделирования сведены в таблицу, где  $T_1$  и  $T_2$  – моменты времени прихода и ухода с особой поверхности.

Отметим, что оптимальный закон изменения угла наклона, представленный на рис. 4, 7, 10, 13, может быть пассивно реализован с помощью стационарных идеальных голономных механических связей. Для приведенных траекторий проверено выполнение необходимого условия Келли оптимальности особых управлений [39].

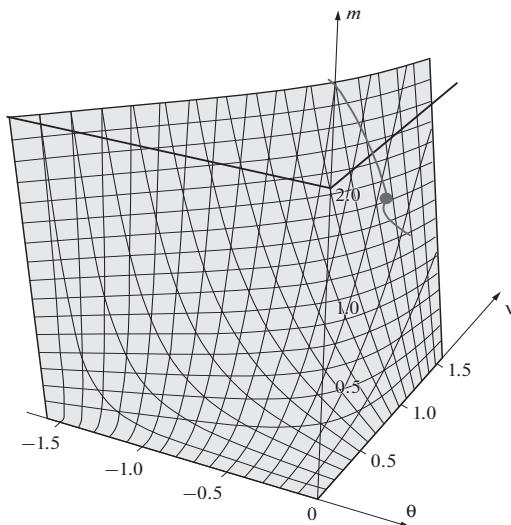


Рис. 10. Экстремальная траектория в пространстве  $(\theta, v, m)$  для  $v(0) = 1$ ,  $m(0) = 2$ ,  $m(T) = 1.25$ ,  $\bar{u} = 2$ ,  $T = 1.25$

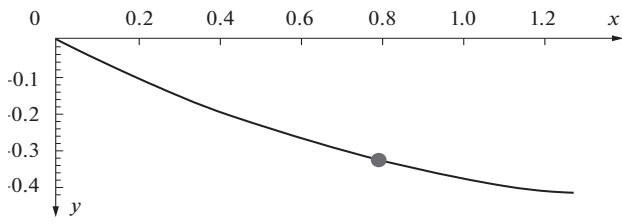


Рис. 11. Экстремальная траектория в плоскости  $(x, y)$  для  $v(0) = 1$ ,  $m(0) = 2$ ,  $m(T) = 1.25$ ,  $\bar{u} = 2$ ,  $T = 1.25$

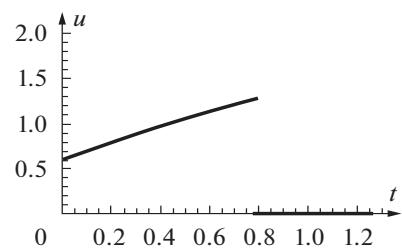


Рис. 12. Изменение тяги во времени  $(u, t)$  на экстремальной траектории для  $v(0) = 1$ ,  $m(0) = 2$ ,  $m(T) = 1.25$ ,  $\bar{u} = 2$ ,  $T = 1.25$

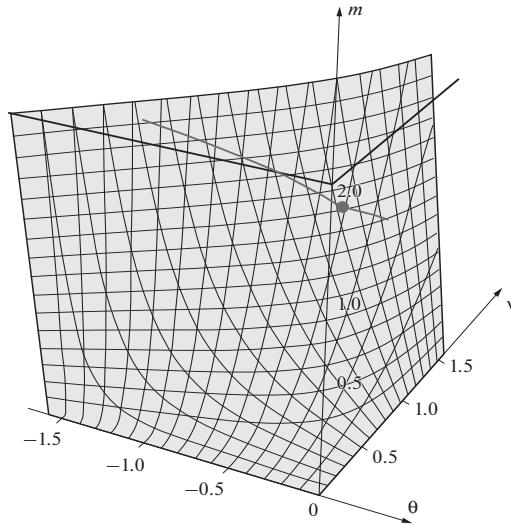


Рис. 13. Экстремальная траектория в пространстве  $(\theta, v, m)$  для  $v(0) = 0.2, m(0) = 2, m(T) = 1.5, \bar{u} = 2, T = 4$

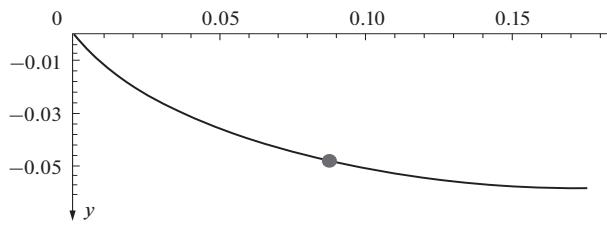


Рис. 14. Экстремальная траектория в плоскости  $(x, y)$  для  $v(0) = 0.2, m(0) = 2, m(T) = 1.5, \bar{u} = 2, T = 4$

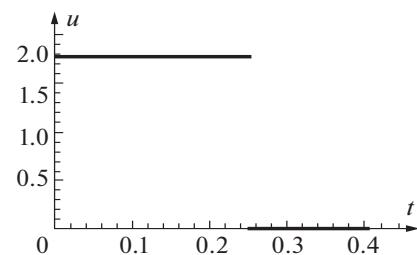


Рис. 15. Изменение тяги во времени  $(u, t)$  на экстремальной траектории для  $v(0) = 0.2, m(0) = 2, m(T) = 1.5, \bar{u} = 2, T = 4$

Таблица 1. Результаты численного моделирования

Набор параметров	Найденное значение $\theta(0)$	$T_1$	$T_2$	Конечное время
1	-0.991232	0.299497	1.007302	1.5
2	-0.815502	1.751761	2.515850	3
3	-0.523601	0	0.784194	1.25
4	-1.012674	—	—	0.4

**Заключение.** Определено аналитическое решение задачи модального управления по выходу для широкого класса динамических систем с суммарной размерностью входного и выходного векторов, не превышающей размерность вектора состояния. Решение применимо к системам четвертого порядка с двумя входами и двумя выходами при неравных индексах управляемости и наблюдаемости. Для каждого назначаемого спектра решение единственно и представляется компактной аналитической формулой, не требующей дополнительных матричных разложений или декомпозиции. Рассмотрены случаи, в которых индекс управляемости принимает

значение как больше, так и меньше значения индекса наблюдаемости. Для каждого случая на основе двухуровневой декомпозиции с параметризацией обоих уровней доказаны соответствующие взаимно дуальные теоремы об управлении по выходу и приведены конкретные примеры аналитического расчета регуляторов. Примеры подтверждают совпадение спектров замкнутых систем управления по выходу с желаемыми спектрами.

### СПИСОК ЛИТЕРАТУРЫ

1. *Goldstine H.H.* A History of the Calculus of Variations from the 17 Th Through the 19 Th Century, Studies in the History of Mathematics and Physical Sciences. V.5. New York-Heidelberg-Berlin, Springer-Verlag, 1980. P. 410.
2. *Ashby N., Britten W.E., Love W.F., Wyss W.* Brachistochrone with Coulomb Friction // Amer. J. Phys. 1975. V. 43. № 10. P. 902–905.
3. *Гершман М.Д., Нагаев Р.Ф.* О фрикционной брахистохроне // МТТ 1976. № 4. С. 85–88.
4. *Lipp S.C.* Brachistochrone with Coulomb Friction // SIAM J. Control Optim. 1997. V. 35. № 2. P. 562–584.
5. *Van der Heijden A.M.A., Diepstraten J.D.* On the Brachistochrone with Dry Friction // Intern. J. Non-Linear Mech. 1975. V. 10. № 2. P. 97–112.
6. *Šalinić S.* Contribution to the Brachistochrone Problem with Coulomb Friction // Acta Mech. 2009. V. 208. P. 97–115.
7. *Sumbatov A.S.* Brachistochrone with Coulomb friction as the Solution of an Isoperimetric Variational Problem // Intern. J. Non-Linear Mech. 2017. V. 88. P. 135–141.
8. *Hayen J.C.* Brachistochrone with Coulomb Friction // Int. J. Non-Linear Mech. 2005. V. 40. P. 1057–1075.
9. *Голубев Ю.Ф.* Брахистохрона с трением // Изв. РАН. ТиСУ. 2010. № 5. С. 41–52.  
<https://doi.org/10.1134/S1064230710050060>
10. *Vratnar B., Saje M.* On the Analytical Solution of the Brachistochrone Problem in a Non-conservative Field // Intern. J. Non-Linear Mechanics. 1998. V. 33. № 3. P. 489–505.
11. *Зароднюк А.В., Черкасов О.Ю.* Качественный анализ оптимальных траекторий движения материальной точки в сопротивляющейся среде и задача о брахистохроне // Изв. РАН. ТиСУ. 2015. № 1. С. 41–49.
12. *Šalinić S., Obradović A., Mitrović Z., Rusov S.* Brachistochrone with Limited Reaction of Constraint in an Arbitrary Force Field // Nonlinear Dynamics. 2012. V. 69. P. 211–222.
13. *Lemak S.S., Belousova M.D.* The Brachistochrone Problem with Constraints on the Curvature of the Trajectory // IFAC PapersOnLine. Moscow. V. 54. P. 437–442.
14. *Брайсон А., Хо Ю.Ши.* Прикладная теория оптимального управления. М.: Мир. 1972. С. 544.
15. *Feehery W.F.* Dynamic Optimization with Path Constraints (Ph. D. Thesis) Massachusetts Institute of Technology. Cambridge, 1998.
16. *Cherkasov O.Yu., Smirnova N.V.* On the Brachistochrone Problem with State Constraints on the Slope Angle // Intern. J. Non-Linear Mech. 2022. V. 139.
17. *Drummond J.E., Downes G.L.* The Brachistochrone with Acceleration: A Running Track // J. Optimization Theory and Applications. 1971. V. 7. № 6. P. 444–449.
18. *Вондрухов А.С., Голубев Ю.Ф.* Брахистохрона с разгоняющей силой // Изв. РАН. ТиСУ. 2014. № 6. С.42–57.
19. *Зароднюк А.В., Черкасов О.Ю.* О максимизации горизонтальной дальности и брахистохроне с разгоняющей силой и вязким трением // Изв. РАН. ТиСУ. 2017. № 4. С. 3–10.
20. *Smirnova N.V., Cherkasov O.Yu.* Range Maximization Problem with a Penalty on Fuel Consumption in the Modified Brachistochrone Problem // Applied Mathematical Modelling. 2021. V. 91. P. 581–589.  
<https://doi.org/10.1016/j.apm.2020.10.001>
21. *Руссаловская А.В., Иванов Г.И., Иванов А.И.* О брахистохроне точки переменной массы с трением и экспоненциальным законом истечения массы // Докл. АН УССР. Сер. А. 1973. С. 1024–1026.
22. *Jeremić O., Šalinić S., Obradović A., Mitrović Z.* On the Brachistochrone of a Variable Mass Particle in General Force Fields // Mathematical and Computer Modelling. 2011. V. 54. P. 2900–2912.
23. *Menon P.K.A., Kelley H.J., Cliff E.M.* Optimal Symmetric Flight with an Intermediate Vehicle Model // J. GUIDANCE. 1984. V. 8. № 3. P. 312–319.
24. *Indig N., Ben-Asher J.Z., Sigal E.* Singular Control for Two-Dimensional Goddard Problems Under Various Trajectory Bending Laws // J. Guidance, Control and Dynamics 2018. V. 42. № 3. P. 1–15.  
<https://doi.org/10.2514/1.G003670>
25. *Indig N., Ben-Asher J.Z., Sigal E.* Optimal Guidance with Additional Thrust Control for Various Flight Tasks // AIAA Guidance, Navigation and Control. Conf. Texas AIAA, 2017. P. 1737.  
<https://doi.org/10.2514/6.2017-1737>
26. *Goddard R.H.* A Method of Reaching Extreme Altitudes. Washington, Smithsonian Institute Miscellaneous Collections, 1919. V. 7. P. 71. (Reprinted by American Rocket Society. 1946.)
27. *Охочимский Д.Е.* К теории движения ракет // ПММ. 1946, Т. 10. № 2. С. 251–272.
28. *Tsien H.S., Evans R.C.* Optimum Thrust Programming for a Sounding Rocket // J. American Rocket Society. 1951. V. 21. № 5. P. 99–107.
29. *Leitmann G.A.* Calculus of Variations Solution of Goddard's Problem // Astronautica Acta. 1956. V.2. № 2. P. 55–62.
30. *Seywald H., Cliff E.M.* Goddard Problem in Presence of a Dynamic Pressure Limit // J. Guid. Control Dyn. 1993. V. 6. № 4. P. 776–781.

<https://doi.org/10.2514/3.21080>

31. *Graichen K., Kugi A., Petit N., Chaplain F.* Handling Constraints in Optimal Control with Saturation Functions and System Extension // *Systems & Control Letters*. 2010. V. 59. № 11. P. 671–679.  
<https://doi.org/10.1016/j.sysconle.2010.08.003>

32. *Bonnans F., Martinon P., Trélat E.* Singular Arcs in the Generalized Goddard's Problem // *J. Optim Theory Appl.* 2008. V. 139. P. 439–461.  
<https://doi.org/10.1007/s10957-008-9387-1>

33. *Miele A.* Extremization of Linear Integrals by Green's Theorem // *Mathematics in Science and Engineering*. 1962. V. 5. P. 69–98  
[https://doi.org/10.1016/S0076-5392\(08\)62091-3](https://doi.org/10.1016/S0076-5392(08)62091-3)

34. *Tsiotras P., Kelley H.J.* Goddard Problem with Constrained Time of Flight // *J. Guidance, Control and Dynamics*. 1992. V. 15. № 2. P. 289–296.  
<https://doi.org/10.23919/ACC.1988.4789942>

35. *Охочимский Д.Е., Энеев Т.М.* Некоторые вариационные задачи, связанные с запуском искусственного спутника Земли // УФН. 1957. № 1а. С. 5–32.

36. *Голубев Ю.Ф.* Метод Охочимского-Понтрягина в теории управления и аналитической механике. Ч. 1. Метод Охочимского-Понтрягина в теории управления // Вестн. МГУ. Сер.1. Математика, механика. 2008. № 6. С. 49–55.

37. *Cherkasov O.Y., Malykh E.V., Smirnova N.V.* Brachistochrone Problem and Two-dimensional Goddard Problem // *Nonlinear Dyn.* 2023. V.111. P. 243–254.  
<https://doi.org/10.1007/s11071-022-07857-x>

38. *Понтрягин Л.С., Болтянский В.Г., Гамкrelidze Р.В., Мищенко Е.Ф.* Математическая теория оптимальных процессов. М.: Наука, 1983. 393с.

39. *Габасов Р., Кириллова Ф.М.* Особые оптимальные управления. М.: Наука, 1973. 256с.

---

**УПРАВЛЕНИЕ СИСТЕМАМИ  
С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ**

---

УДК 517.958

**ЗАДАЧА РАСПРЕДЕЛЕННОГО УПРАВЛЕНИЯ  
КОЛЕБАНИЯМИ ЭМУЛЬСИИ СЛАБОВЯЗКИХ  
СЖИМАЕМЫХ ЖИДКОСТЕЙ<sup>1</sup>**

© 2024 г. А. А. Егорова<sup>a</sup>, \*, А. С. Шамаев<sup>b</sup>, \*\*

<sup>a</sup>Российский технологический университет (МИРЭА), Москва, Россия

<sup>b</sup>Институт проблем механики им. А.Ю. Ишлинского РАН, Москва, Россия

\*e-mail: alena.andreevna.egorova@yandex.ru

\*\*e-mail: sham@rambler.ru

Поступила в редакцию 05.08.2024 г.

После доработки 13.08.2024 г.

Принята к публикации 16.10.2024 г.

Рассматривается задача распределенного управления колебаниями эффективной (усредненной) среды, соответствующей двухфазной среде из слабовязких жидкостей. Усредненная модель описывается краевой задачей для интегродифференциального уравнения. Показано, что для этой модели силовым воздействием на всю область невозможно привести за конечное время колебания в состояние покоя.

*Ключевые слова:* управление колебаниями, асимптотические методы, усреднение

DOI: 10.31857/S0002338824060048, EDN: SVRBOO

**THE PROBLEM OF DISTRIBUTED CONTROL OF EMULSION  
VIBRATIONS OF WEAKLY VISCOUS COMPRESSIBLE LIQUIDS**

А. А. Егорова<sup>a</sup>, \*, А. С. Шамаев<sup>b</sup>, \*\*

<sup>a</sup>MIREA – Russian Technological University, Moscow, Russia

<sup>b</sup>IPMech RAS, Moscow, Russia

\*e-mail: alena.andreevna.egorova@yandex.ru

\*\*e-mail: sham@rambler.ru

The problem of distributed control of oscillations of an effective (averaged) medium corresponding to a two-phase medium of slightly viscous liquids is considered. The averaged model is described by a boundary value problem for an integro-differential equation. It is shown that for this model, it is impossible to bring vibrations to a state of rest in a finite time by force action on the entire region.

*Keywords:* control of oscillations, asymptotic methods, averaging

**Введение.** В современном мире потребности в новых материалах, исследование акустики морского дна, проблемы геологической разведки вызывают интерес к разработке моделей, описывающих физические процессы в неоднородных средах. Например, задачи распространения акустических волн в комбинированных средах, состоящих из упругого каркаса и вязкой жидкости либо из упругих и вязкоупругих материалов и т.п., изучались многими исследователями в течение последних 70 лет [1, 2]. Для анализа в подобных средах часто полагают включение фаз (двух или нескольких) с быстро чередующейся периодической структурой, при этом характерный размер чередования  $\epsilon$  принимают за малый параметр. Тогда можно постро-

---

<sup>1</sup> Работа выполнена в рамках госзадания №124012500443–0.

ить усредненную (эффективную) модель – краевые задачи для уравнений с медленно меняющимися коэффициентами, к решениям которых сходятся решения исходных задач при  $\varepsilon \rightarrow 0$ . При этом тип уравнений в усредненной модели определяется характеристиками (вязкость, упругость, сжимаемость) каждой из фаз, геометрическими особенностями включений одной фазы в другую, а также поведением параметров задачи относительно  $\varepsilon$  [1, 3].

Задача управления состоит в том, чтобы понять, возможно ли привести колебания среды в покой за конечное время посредством ограниченного по абсолютной величине силового воздействия на всю область, занимаемой образцом комбинированной среды. В данной работе рассматривается проблема приведения в состояние покоя за конечное время малых колебаний в эффективной модели, соответствующей колебаниям в двухфазной среде, которая состоит из двух слабосжимаемых жидкостей. Показано, что для этой модели силовым воздействием на всю область невозможно привести за конечное время колебания в состояние покоя. В то же время для дифференциальных уравнений теплопроводности для струны, мембранны, пластины в работе [4] установлена возможность приведения в состояние полного покоя за конечное время.

**1. Математическая модель эмульсии двух жидкостей.** Рассматривается движение эмульсии в ограниченной области  $\Omega$  с гладкой границей  $\partial\Omega$  при внешних звуковых возмущениях  $f(t, x)$  малой амплитуды. Периодически расположенные одинаковые включения имеют достаточно гладкую форму (рисунок).

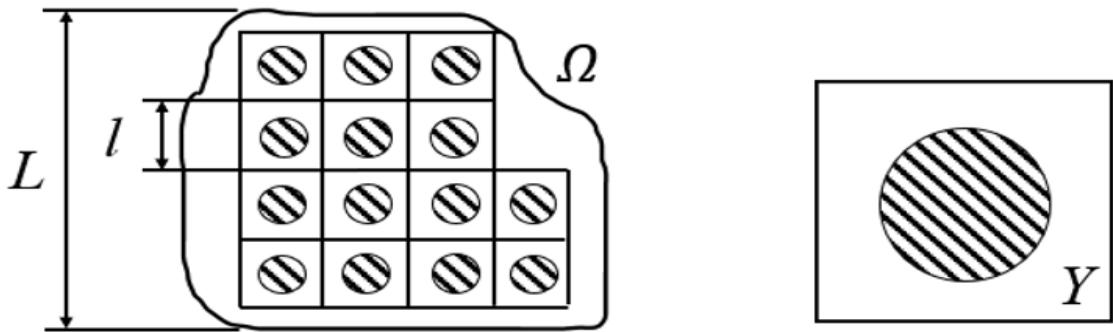


Рис. 1. Область  $\Omega$  и ячейка периода  $Y$ .

Характеристики жидкостей в состоянии покоя (вязкости  $\mu$  и плотности  $\rho$ , скорости звука  $c$ ) считаются сравнимыми, отношение стороны  $l$  ячейки периодичности  $Y$  и размера  $L$  области  $\Omega$  мало и равно  $\varepsilon \ll 1$ . Закон движения в эмульсии задается уравнением Навье-Стокса:

$$\frac{\partial \mathbf{v}}{\partial t} = -\nabla p + \varepsilon^2 \mu \Delta \mathbf{v} + \nabla \left( \frac{\varepsilon^2 \mu}{3} \right) \operatorname{div} \mathbf{v} + \mathbf{f}, \quad (1.1)$$

где  $\mathbf{v}(x, t, \varepsilon)$  – скорость дисперсионной среды,  $p(x, t, \varepsilon)$  – давление,  $\mathbf{f}(x, t)$  – функция воздействия звукового поля на эмульсию, удовлетворяющая неравенству:

$$|\mathbf{f}(x, t)| \leq M,$$

$M > 0$  – заданная константа. На границах фаз предполагается непрерывность скорости и тензора напряжений. Перемещение  $\mathbf{u}(x, t, \varepsilon) \in (H_0^1(\Omega))^3$  связано со скоростью  $\mathbf{v}$  и давлением  $p$  следующими соотношениями:

$$\mathbf{u}(x, t, \varepsilon) = \int_0^t \mathbf{v}(x, s, \varepsilon) ds, \quad p(x, t, \varepsilon) = -\gamma \left( \frac{x}{\varepsilon} \right) \operatorname{div} \mathbf{u}, \quad (1.2)$$

где  $\gamma(\xi) = c^2(\xi) \rho(\xi) > 0$  – модуль объемной упругости. Начальное перемещение и скорость задаются вектор-функциями:

$$\mathbf{u}|_{t=0} = \Phi_0(x), \quad \mathbf{v}|_{t=0} = \Phi_1(x). \quad (1.3)$$

**2. Эффективная модель.** Из работы [5] следует, что краевая задача для усредненной модели микстуры из двух слабовязких жидкостей удовлетворяет динамическому закону Дарси, который описывается следующей системой:

$$\begin{cases} \hat{\psi}^0 + \operatorname{div}_x (\mathbf{K}(t) * (f(t, x) - \nabla p^0)) = 0, & x \in \Omega \\ (\mathbf{K}(t) * \nabla p^0, \mathbf{n}) = (\mathbf{K}(t) * f, \mathbf{n}), & x \in \partial\Omega, \\ f(t, x)|_{t<0} \equiv 0, p(t, x)|_{t<0} \equiv 0, \end{cases} \quad (2.1)$$

где  $p^0(t, x)$  – усредненное акустическое давление;  $\mathbf{K}(t)$  – некоторая матрица, элементы которой выражаются через решения краевых задач, заданных на ячейке периодичности  $Y$ , для системы типа Стокса с периодическими условиями, знаком \* обозначена свертка матричной функции  $G_1(t)$  и вектор-функции  $g_2(t)$  по переменной  $t$ :

$$G_1(t) * g_2(t) = \int_0^t G_1(t-\tau) g_2(\tau) d\tau.$$

В работе изучается случай, когда эту матрицу можно представить в виде произведения единичной матрицы  $E$  и ряда из убывающих экспонент:

$$\mathbf{K}(t) = E \sum_{i=1}^{\infty} c_i e^{-\lambda_i t}, \quad (2.2)$$

где  $c_i, \lambda_i$  – некоторые положительные постоянные, такие, что

$$0 < \lambda_1 < \lambda_2 < \dots, \quad \sum_{i=1}^{\infty} c_i < +\infty.$$

При этом для предельных функций перемещения  $\mathbf{u}^0\left(t, x, \frac{x}{\varepsilon}\right)$  и скорости  $\mathbf{v}^0\left(t, x, \frac{x}{\varepsilon}\right)$ , имеющих значения в  $\left(H^1\left(\Omega, (H_{per}^1(Y))^3\right)\right)^3$ , справедливы равенства:

$$\mathbf{v}^0(t, x, \xi) = \frac{\partial \mathbf{u}^0(t, x, \xi)}{\partial t}, \quad (2.3)$$

$$\mathbf{v}^0(t, x, \xi)_Y = \int_0^t \mathbf{K}(t-s) (f(s, x) - \nabla p^0(s, x)) ds, \quad (2.4)$$

где

$$g(\xi)_Y = \frac{1}{|Y|} \int_Y g(\xi) d\xi$$

– среднее по ячейке периодичности  $Y$ .

В [6] было показано  $L_2$ -сходимость решений исходной задачи (1.1)–(1.3) к решениям усредненной задачи (2.1)–(2.4).

**Теорема 1.** Для решений задач (1.1)–(1.3) и (2.1)–(2.4) справедливы равенства:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \left\| \mathbf{v}(x; t; \varepsilon) - \mathbf{v}^0\left(x; \frac{x}{\varepsilon}; t\right) \right\|_{L_2(\Omega)} &= 0, \\ \lim_{\varepsilon \rightarrow 0} \left\| \varepsilon \nabla_x \mathbf{v}(x; t; \varepsilon) - \nabla_x \mathbf{v}^0\left(x; \frac{x}{\varepsilon}; t\right) \right\|_{L_2(\Omega)} &= 0, \\ \lim_{\varepsilon \rightarrow 0} \left\| p(x; t; \varepsilon) - p^0(x; t) \right\|_{L_2(\Omega)} &= 0. \end{aligned}$$

### 3. Невозможность приведения системы в покой за конечное время.

**Предложение 1.** Система (2.1)–(2.4) приводима в покой, если для любого начального условия  $\mathbf{u}^0|_{t=0} = \Phi_0(x)$  можно найти функцию управления  $f(t, x)$  и время  $T > 0$  такое, что  $f(t, x) \equiv 0 \forall t > T$  и соответствующее решение  $\mathbf{u}^0(t, x)$  задачи (2.1)–(2.4) также тождественно равно нулю для любого  $t > T$ .

Теорема 2. Существуют начальные условия  $\Phi_0 \in (H_0^1(\Omega))^3$ , для которых систему (2.1)–(2.4) невозможно привести в покой за конечное время  $T > 0$  через управление всей областью  $\Omega$ .

Доказательство. В доказательстве используются теорема Винера–Пели об образе преобразования Лапласа и оценка типа Ли–Яу для спектра оператора Стокса [7].

Допустим, что для начального условия  $\mathbf{u}^0|_{t=0} = \Phi_0(x)$  можно найти функцию управления  $f(t, x)$  и время  $T > 0$ , такое, что  $f(t, x) \equiv 0 \forall t > T$  и соответствующее решение  $\mathbf{u}^0(t, x)$  задачи (2.1)–(2.4) также тождественно равно нулю для любого  $t > T$ . Применив к уравнениям (2.1)–(2.4) преобразование Лапласа по переменной времени  $t$ , получим:

$$\lambda \langle \hat{\mathbf{u}}^0 \rangle_Y - \Phi_0(x) = \hat{\mathbf{K}}(\lambda) \left( \hat{f} - \nabla \hat{p}^0 \right),$$

где вектор-функции  $\hat{\mathbf{u}}^0(\lambda, x, \xi)$ ,  $\hat{f}(\lambda, x)$  – преобразования Лапласа вектор-функций  $\mathbf{u}^0(t, x, \xi)$ ,  $f(t, x)$  соответственно,  $\hat{p}^0(\lambda, x)$  – преобразование Лапласа для функции  $p^0(t, x)$ . Из (2.2) следует, что преобразование Лапласа для матрицы  $\mathbf{K}(t)$  имеет следующий вид:

$$\hat{\mathbf{K}}(\lambda) = E \sum_{i=1}^{\infty} \frac{c_i}{\lambda + \lambda_i}. \quad (3.1)$$

Здесь положительные числа  $\lambda_i$  совпадают с собственными значениями оператора Стокса. Аналогично утверждениям из работы [7] можно доказать справедливость следующей леммы.

Лемма. Каждое собственное значение оператора Стокса в трехмерной периодической области  $\Omega$  удовлетворяет неравенству:

$$\alpha_1 k^{\frac{2}{3}} \leq \lambda_k \leq \alpha_2 k^{\frac{2}{3}}. \quad (3.2)$$

Если система управляема в покой, то по теореме Винера–Пэли все компоненты вектор-функций  $\hat{\mathbf{u}}^0(\lambda, x, \xi)$ ,  $\hat{f}(\lambda, x)$ ,  $\nabla \hat{p}^0$  должны быть целыми функциями экспоненциального типа, как преобразования Лапласа финитных по  $t$  функций. Из равенства (3.1) получим, что

$$\lambda \langle \hat{\mathbf{u}}^0 \rangle_Y = \Phi_0(x) + \sum_{i=1}^{\infty} \frac{c_i}{\lambda + \lambda_i} \left( \hat{f} - \nabla \hat{p}^0 \right).$$

Пусть  $\Phi_0(x) = a_1 e_1$ , где  $a_1$  – ненулевая постоянная,  $e_1$  – базисный вектор. Тогда для первой компоненты получаем равенство:

$$\lambda \langle \hat{\mathbf{u}}^0 \rangle_Y = a_1 + \sum_{i=1}^{\infty} \frac{c_i}{\lambda + \lambda_i} \left( \hat{f}_1 - \frac{\partial \hat{p}^0}{\partial x_1} \right).$$

Так как  $\hat{u}_1^0(\lambda)$ ,  $\hat{f}_1(\lambda) - \frac{\partial \hat{p}^0}{\partial x_1}(\lambda)$  – целые функции, то нули функции  $\hat{f}_1 - \frac{\partial \hat{p}^0}{\partial x_1}$  должны совпадать с  $-\lambda_i$ ,  $i = 1, 2, \dots$ .

Определение 2. Показателем сходимости последовательности комплексных чисел  $\{z_k\}$  называется число [8]

$$\tau(\{z_k\}) = \inf \left\{ \alpha > 0 : \sum_{k=1}^{+\infty} \frac{1}{|z_k|^\alpha} < +\infty \right\}.$$

Из (3.2) следует, что  $\tau(\{\lambda_k\}) > 1$ . По определению из работы [8] порядком роста целой функции  $\varphi(\lambda)$  называется число

$$\rho = \inf \left\{ \mu > 0 : \max_{|\lambda|=r} |\varphi(\lambda)| < e^{r^\mu} \right\}.$$

Очевидно, что для целой функции экспоненциального типа порядок роста  $\rho = 1$ . В работе [8] доказано, что показатель сходимости последовательности всех нулей целой функции не превосходит ее порядка роста ( $\tau \leq \rho$ ). Тогда для целой функции  $\hat{f}_1 - \frac{\partial \hat{p}^0}{\partial x_1}$  получаем противоречие, так как последовательность его корней  $\{-\lambda_i\}$  имеет показатель сходимости больше единицы. Теорема 2 доказана.

**Заключение.** Полученный результат показывает, что ряд механических моделей, которые описываются уравнениями и системами интегродифференциальных уравнений, обладают

принципиально иными свойствами для задач управления. Так, рассмотренная в работе задача остановки колебаний распределенной ограниченной силой, приложенной к системе, оказывается неразрешимой, в отличие от задач управления классическими системами, задаваемыми дифференциальными уравнениями [4]. Интегродифференциальные системы требуют иных подходов для задач управления, чем дифференциальные системы, – это большое новое поле для исследований.

#### СПИСОК ЛИТЕРАТУРЫ

1. Санчес-Паленсия Э. Неоднородные среды и теория колебаний. М.: Мир, 1984.
2. Пятницкий А.Л., Чечкин Г.А., Шамаев А.С. Усреднение. Методы и некоторые приложения. Новосибирск: Изд-во “Тамара Рожковская”, 2007.
3. Nguetseng G. Asymptotic Analysis for a Staff Variational Problem Arising in Mathematics // SIAMJ. Math. Anal. 1990. V. 21. № 6. P. 1396–1414.
4. Черноусько Ф.Л. Ограничные управление в системах с распределенными параметрами // ПММ. 1992 Т. 56. Вып. 2. С. 179–191.
5. Vlasov V.V., Gavrikov A.A., Ivanov S.A., Knyazkov D.Yu., Samarin V.A., Shamaev A.S. Spectral Properties of Combined Media // Mathematical Sciences. 2010. V. 164. № 6. P. 948–963.
6. Гавриков А.А., Шамаев А.С. Некоторые вопросы акустики эмульсий // ДАН. 2010. Т. 434. № 1. С. 42–46.
7. Ильин А.А. О спектре оператора Стокса // Функц. анализ и его приложения. 2009. Т. 43. Вып. 4. С. 14–25.
8. Седлецкий А.М. Целые функции в анализе. [https://math.msu.ru/sites/default/files/entire\\_functions.pdf](https://math.msu.ru/sites/default/files/entire_functions.pdf)

## ВЛИЯНИЕ АЭРОДИНАМИЧЕСКОЙ АСИММЕТРИИ НА КРИТИЧЕСКИЕ РЕЖИМЫ ПОЛЕТА ДОЗВУКОВЫХ САМОЛЕТОВ И УПРАВЛЕНИЕ ДЛЯ ВЫВОДА ИЗ НИХ<sup>1</sup>

© 2024 г. М. Е. Сидорюк<sup>a</sup>, \*, А. Н. Храбров<sup>a</sup>, \*\*

<sup>a</sup>ФАУ “Центральный аэрогидродинамический институт  
им. проф. Н.Е. Жуковского”, Жуковский, Россия

\*e-mail: mariya.sidoryuk@tsagi.ru

\*\*e-mail: aleksandr.khrabrov@tsagi.ru

Поступила в редакцию 14.08.2023г.

После доработки 28.03.2024 г.

Принята к публикации 16.10.2024 г.

Рассмотрены стационарные и устойчивые периодические режимы пространственного движения самолета. Исследовано влияние аэродинамической асимметрии на параметры критических режимов с помощью прямого расчета методом непрерывного продолжения по параметру. Описаны причины появления аэродинамической асимметрии при симметричных условиях обтекания. Анализируется влияние возникающей аэродинамической асимметрии на возможность попадания самолета в штопор и вывод из этого критического режима. С помощью инструментов робастного анализа исследована чувствительность параметров штопора к неопределенности математической модели аэродинамики.

**Ключевые слова:** аэродинамическая асимметрия, штопор самолета, продолжение по параметру, неопределенность, робастный анализ.

DOI: 10.31857/S0002338824060051, EDN: SVIPPG

## INFLUENCE OF AERODYNAMIC ASYMMETRY ON CRITICAL FLIGHT REGIMES OF SUBSONIC AIRCRAFT AND RECOVERY CONTROL

М. Е. Сидорюк<sup>a</sup>, \*, А. Н. Храбров<sup>a</sup>, \*\*

<sup>a</sup>Central Aerohydrodynamic Institute, Zhukovsky, Russia

\*e-mail: mariya.sidoryuk@tsagi.ru

\*\*e-mail: aleksandr.khrabrov@tsagi.ru

Stationary and stable periodic modes of spatial motion of the aircraft are considered. The influence of aerodynamic asymmetry on critical flight regimes is studied using direct calculation of their parameters using continuation on a parameter technique. The reasons for the appearance of aerodynamic asymmetry under symmetric flow conditions are described. The influence of the aerodynamic asymmetry on the possibility of entry into a spin and recovery control are analyzed. The sensitivity of the spin parameters to uncertainties of the mathematical model of aerodynamics is studied using robust analysis tools.

**Keywords:** aerodynamic asymmetry, aircraft spin, continuation on a parameter, uncertainty, robust analysis.

---

<sup>1</sup> Работа подготовлена в рамках реализации Программы создания и развития научного центра мирового уровня «Сверхзвук» на 2020–2050 годы при финансовой поддержке Минобрнауки России (соглашение о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научных центров мирового уровня, выполняющих исследование и разработки по приоритетам научно-технологического развития от 17 мая 2022 г. № 075-15-2022-1023).

**Введение.** Надежное предсказание критических режимов, необходимое для обеспечения безопасности полета, полностью связано с адекватностью модели аэродинамики самолета, используемой для анализа динамики. Наиболее известными критическими режимами являются сваливание и штопор. Под сваливанием понимается самопроизвольное движение самолета относительно любой из осей, вызванное изменением его характеристик на больших углах атаки вследствие изменения обтекания при возникновении явлений отрыва потока. Штопор — это вертикальное движение самолета на больших углах атаки с интенсивным вращением по крену и рысканию. Известно, что аэродинамическая асимметрия играет важную роль в сваливании и определяет характер штопора сверхзвуковых самолетов [1, 2]. Из-за асимметрии момента рыскания правый штопор может существенно отличаться от левого, а режимы плоского штопора могут стать невыводимыми. В работе рассматривается несколько аспектов влияния аэродинамической асимметрии на критические режимы дозвуковых самолетов на примере двух моделей. Первый объект исследований (модель 1) — компоновка типичного магистрального самолета со стреловидным крылом, установленным по схеме «низкоплан» и палубным горизонтальным оперением [3]. Второй объект (модель 2) — конфигурация самолета с прямым крылом большого удлинения, установленным по схеме «верхнеплан» с Т-образным горизонтальным оперением, размещенным на конце вертикального киля (рис. 1).



Рис. 1. Общий вид используемых моделей самолетов.

Аэродинамическая асимметрия рассматривается как характерное свойство летательных аппаратов, обладающих высокой геометрической и массовой симметрией, развивающееся в определенном диапазоне параметров полета. Практически все летательные аппараты имеют вертикальную плоскость симметрии, за исключением небольшого числа научно-исследовательских проектов [4–6] и реализованных летающих моделей (например, «Премьер I» [4]), которые ныне хранятся в музеях. Приводятся источники аэродинамической асимметрии симметричных летательных аппаратов и ее влияние на безопасность полета посредством анализа зависимости параметров критических режимов полета, в особенности штопора, от величины наблюдаемой асимметрии.

Данное исследование основано на большом количестве экспериментальных результатов, полученных в нескольких аэродинамических трубах (АДТ) ЦАГИ, и на прямом расчете параметров критических режимов на базе разработанной аэродинамической модели с использованием метода непрерывного продолжения по параметру [2, 7, 8] и связанных с ним методов анализа бифуркаций и устойчивости. Сравнение аэродинамических характеристик, полученных в двух разных аэродинамических трубах, демонстрирует случайность знака аэродинамической асимметрии, возникающей при нулевом угле скольжения и нулевой угловой скорости из-за несимметричного срыва потока в некоторых диапазонах углов атаки. Величина этой асимметрии, как и все измеряемые параметры, имеет неопределенность, особенно с учетом того, что необходимо переносить аэродинамические характеристики, определенные в АДТ при низких значениях числа Рейнольдса, на реальный самолет. Существенное влияние неопределенности некоторых аэродинамических характеристик на параметры штопора было показано в [9]. Величина аэродинамической асимметрии зависит от особенностей отрыва потока, и в данной работе при исследовании штопора предполагается, что она может варьироваться от нуля до

максимальных значений, наблюдаемых в экспериментах в АДТ. Учитывается также возможное влияние неопределенностей других параметров математической модели аэrodинамики. Влияние этих неопределенностей на параметры штопора изучается с помощью робастного ( $\mu$ ) анализа. Анализируется также влияние аэродинамической асимметрии на восприимчивость самолета к штопору и возможность выхода из него.

В разд. 1 изложены причины аэродинамической асимметрии и область ее проявления. В разд. 2 представлена структура аэродинамической модели и основные аэродинамические характеристики для исследования критических режимов полета. В разд. 3 исследуется влияние аэродинамической асимметрии на параметры критических режимов полета на примере модели типичного магистрального самолета. В разд. 4 рассмотрено влияние аэродинамической асимметрии на размер области притяжения режимов штопора. Оценки размера области притяжения для различных режимов штопора могут рассматриваться как мера восприимчивости самолета к штопору и относительной опасности отдельных режимов штопора. Наконец, в разд. 5 описывается возможное влияние боковой аэродинамической асимметрии на выход самолета из штопора.

**1. Причины возникновения аэродинамической асимметрии.** Для многих компоновок современных самолетов в некоторых диапазонах углов атаки (различных для самолетов разных типов и конфигураций) можно наблюдать явление аэродинамической асимметрии. В условиях симметричного набегающего потока (при нулевом угле скольжения) возникает несимметричное отрывное или вихревое обтекание компоновки и, следовательно, появляются несимметричные аэродинамические нагрузки на нее.

Природа этого явления связана с понятием устойчивости симметричного отрывного обтекания. Проблема напоминает задачу Эйлера о потере устойчивости упругого сжатого стержня: если упругий прямолинейный стержень с шарнирными опорами на концах сжимать строго вертикальной силой, то начиная с некоторого критического значения происходит бифуркация. Кроме стационарного прямолинейного симметричного решения возникают два несимметричных решения с прогибом вправо или влево, причем симметричное решение становится неустойчивым, т.е. малые возмущения приводят к перешелкиванию стержня в ту или другую сторону.

Аналогично в аэродинамике, когда степень устойчивости симметричного отрывного обтекания снижается, оно становится чувствительным к малым возмущениям, например, к малым искажениям геометрической формы самолета или несимметричным воздействиям набегающего потока. В работах [10, 11] был рассмотрен пример модельного обтекания комбинации удлиненного конусовидного тела с треугольным крылом большой стреловидности. С передней кромки треугольного крыла сходили спиральные вихри (рис. 2). Решение задачи обте-

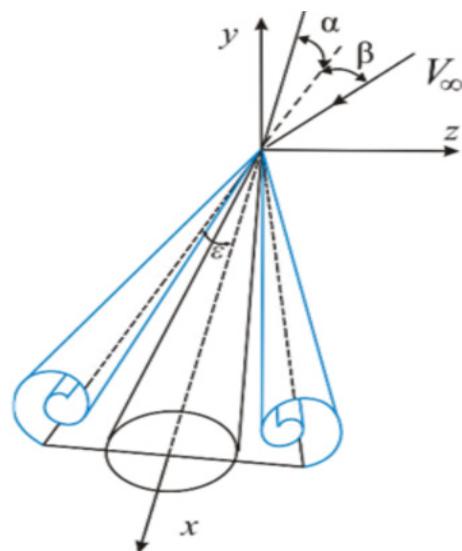


Рис. 2. Модельная задача исследования возникновения аэродинамической асимметрии.

кания, включающее математическое моделирование распределенной спиральной вихревой пелены без учета разрушения вихрей (конечный отрезок пелены соединен математическим «питающим» разрезом с ядром — точечным дискретным вихрем), показало, что при симметричном набегающем потоке (угол скольжения  $\beta$  равен нулю) до некоторого значения приведенного критического угла атаки  $\bar{\alpha}_{kp}$  сходящие с передней кромки треугольного крыла вихри симметричны (рис. 3).

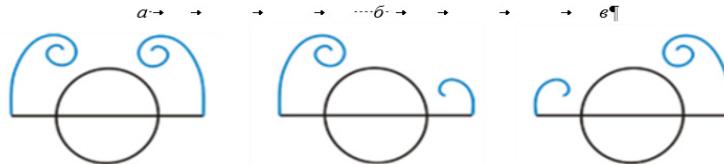


Рис. 3. Симметричное (а) и несимметричные (б, в) решения при  $\bar{\alpha} > \bar{\alpha}_{kp}$ .

Равновесное состояние спиральной вихревой пелены получается путем итерационного перестроения ее геометрии вплоть до удовлетворения условиям стационарности и непрерывности давления на поверхности тангенциального разрыва скорости. При таком моделировании вихревой пелены расчетное значение критического угла атаки возникновения асимметрии близко к значению, получаемому в эксперименте.

Когда угол атаки превышает некоторое критическое значение  $\bar{\alpha}_{kp}$ , симметричное решение (рис. 3, а) становится неустойчивым и появляются два новых, устойчивых, «зеркально отраженных» несимметричных решения, определяющих различное положение и интенсивность правого и левого вихрей (рис. 3, б, в). То, какое из несимметричных решений реализуется, зависит от малых возмущений нелинейной динамической системы. Такими возмущениями в модельной задаче могут служить наличие небольшого угла скольжения. Для реального самолета возмущениями могут быть также малые отклонения от идеально симметричной конструкции, например, выступающие элементы фюзеляжа, приемник воздушного давления и т.п. Возникающая асимметрия обтекания может приводить к появлению значительных моментов по крену и рысканию, величина которых может меняться вплоть до смены направления их действия.

Описанная модель поясняет причины образования и развития аэродинамической асимметрии, хотя для случая обтекания на больших углах атаки модели дозвукового самолета физические причины могут быть различными и не связанными с потерей устойчивости взаимодействующих вихрей.

**2. Структура аэродинамической модели и основные аэродинамические характеристики.** Обычным способом анализа динамики самолета на критических режимах является рассмотрение автономной системы уравнений движения 8-го порядка, полученной из полных уравнений движения в предположении о постоянстве высоты [1, 12–14]:

$$\begin{aligned}
 \frac{d\alpha}{dt} &= \omega_z - \left[ (a_x - \omega_y \sin \beta) \sin \alpha + (a_y + \omega_x \sin \beta) \cos \alpha \right] \cos^{-1} \beta, \\
 \frac{d\beta}{dt} &= a_z \cos \beta - (a_x \sin \beta - \omega_y) \cos \alpha + (a_y \sin \beta + \omega_x) \sin \alpha, \\
 \frac{dV}{dt} &= V (a_x \cos \beta \cos \alpha - a_y \cos \beta \sin \alpha + a_z \sin \beta), \\
 \frac{d\vartheta}{dt} &= \omega_y \sin \gamma + \omega_z \cos \gamma, \\
 \frac{d\gamma}{dt} &= \omega_x - \operatorname{tg} \vartheta (\omega_y \cos \gamma - \omega_z \sin \gamma), \\
 \frac{d\omega}{dt} &= J^{-1} (-\omega \times J \omega + \mathbf{M}_a),
 \end{aligned} \tag{2.1}$$

где

$$a_x = -\frac{\rho VS}{2m} c_x - \frac{g}{V} \sin \vartheta, a_y = \frac{\rho VS}{2m} c_y - \frac{g}{V} \cos \vartheta \cos \gamma, a_z = \frac{\rho VS}{2m} c_z + \frac{g}{V} \cos \vartheta \sin \gamma,$$

$$\mathbf{R}_a = (c_x \frac{\rho V^2}{2} S, c_y \frac{\rho V^2}{2} S, c_z \frac{\rho V^2}{2} S)^T,$$

$$\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^T, \mathbf{M}_a = \left( \frac{\rho V^2 S l}{2} m_x, \frac{\rho V^2 S l}{2} m_y, \frac{\rho V^2 S b_a}{2} m_z \right)^T.$$

Здесь  $\vartheta$  – угол тангажа,  $\gamma$  – угол крена,  $J, m$  – тензор инерции и масса самолета или модели самолета,  $S$  – площадь крыла,  $l, b_a$  – длина и средняя хорда крыла,  $\rho$  – плотность воздуха,  $g$  – ускорение свободного падения,  $\omega_x, \omega_y, \omega_z$  – проекции вектора угловой скорости вращения самолета на оси связанный системы координат,  $V$  – истинная скорость самолета,  $\mathbf{R}_a$  – вектор аэродинамических сил, действующих на самолет,  $c_x, c_y, c_z$  – безразмерные коэффициенты аэродинамических сил, действующих на самолет,  $\mathbf{M}_a$  – вектор момента от аэродинамических сил, действующих на самолет,  $m_x, m_y, m_z$  – безразмерные коэффициенты момента от аэродинамических сил, действующих на самолет. Тяга считается равной нулю. Координата курса, показывающая вращение самолета в штопоре, определяется независимым интегрированием уравнения  $\dot{\psi} = (\omega_x \cos \gamma + \omega_z \sin \gamma) / \cos \vartheta$ .

Для адекватного расчета штопора и других критических режимов полета необходима существенно нелинейная математическая модель аэродинамических характеристик самолета, действительная в широком диапазоне параметров движения (углов атаки и скольжения, угловых скоростей вращения). Математическая модель аэродинамических характеристик на больших углах атаки при наличии интенсивного вращения базируется на результатах двух основных экспериментальных методов: вынужденных колебаний с малой амплитудой по тангажу, крену и рысканию и установившегося конического вращения около вектора скорости набегающего потока. При больших углах атаки движение самолета с интенсивным вращением может значительно влиять на отрыв потока, и аэродинамические коэффициенты становятся существенно зависимыми от угловой скорости самолета. Безразмерные коэффициенты аэродинамических сил и моментов, используемых для расчета критических режимов полета, имеют следующую структуру [9]:

$$\begin{aligned} c_x &= c_x(\alpha) + \Delta c_x(\alpha, \beta), \\ c_y &= c_y(\alpha) + \Delta c_y(\alpha, \beta) + \Delta c_y(\alpha, \delta_B), \\ c_z &= k_a c_{z0}(\alpha) + \Delta c_z(\alpha, \beta) + \Delta c_z(\alpha, \beta, \bar{\omega}_{xa}) + \Delta c_z(\alpha, \beta, \delta_H), \\ m_x &= k_a m_{x0}(\alpha) + \Delta m_x(\alpha, \beta) + \Delta m_x(\alpha, \beta, \bar{\omega}_{xa}) + \tilde{m}_x^{\bar{\omega}_{ya}} \bar{\omega}_{ya} + \tilde{m}_x^{\bar{\omega}_{za}} \bar{\omega}_{za} + \Delta m_x(\alpha, \beta, \delta_E) + \Delta m_x(\alpha, \beta, \delta_H), \\ m_y &= k_a m_{y0}(\alpha) + \Delta m_y(\alpha, \beta) + \Delta m_y(\alpha, \beta, \bar{\omega}_{xa}) + \tilde{m}_y^{\bar{\omega}_{ya}} \bar{\omega}_{ya} + \tilde{m}_y^{\bar{\omega}_{za}} \bar{\omega}_{za} + \Delta m_y(\alpha, \beta, \delta_H), \\ m_z &= m_z(\alpha) + \Delta m_z(\alpha, \beta) + \tilde{m}_z^{\bar{\omega}_{za}} \bar{\omega}_{za} + \Delta m_z(\alpha, \beta, \delta_B). \end{aligned} \quad (2.2)$$

Здесь  $\alpha, \beta$  – углы атаки и скольжения,  $\delta = (\delta_B, \delta_E, \delta_H)$  – вектор отклонений руля высоты, элеронов и руля направления,  $\bar{\omega}_{xa}, \bar{\omega}_{ya}, \bar{\omega}_{za}$  – безразмерные проекции вектора угловой скорости на оси скоростной системы координат:  $\bar{\omega}_{xa} = \omega_{xa} L / (2V)$ ,  $\bar{\omega}_{ya} = \omega_{ya} L / (2V)$ ,  $\bar{\omega}_{za} = \omega_{za} b_a / V$ . Члены вида  $\Delta m_i(\alpha, \beta, \bar{\omega}_{xa})$ ,  $i = x, y$  и  $\Delta c_z(\alpha, \beta, \bar{\omega}_{xa})$  в соотношениях (2.2) описывают вклад интенсивного вращения в моментные характеристики и боковую силу, который измеряется в экспериментах с установившимся коническим вращением. Коэффициенты боковой силы  $k_a c_{z0}(\alpha)$  и моментов  $k_a m_{x0}(\alpha), k_a m_{y0}(\alpha)$  в соотношениях (2.2) описывают аэродинамическую асимметрию, вклад от которой определяется мультипликативным параметром  $k_a \in [-1, 1]$ , позволяющим менять ее величину и знак.

На рис. 4 показаны зависимости этих коэффициентов от угла атаки, полученные в двух различных аэродинамических трубах для первой модели при нулевом скольжении. Наблюдаются три области асимметрии:  $\alpha = 10 \dots 14^\circ$  – несимметричный отрыв потока на крыле,  $\alpha = 24 \dots 28^\circ$  – несимметричное отрывное обтекание мотогондол, и самая заметная асимметрия при  $\alpha = 50 \dots 70^\circ$ , связанная, по-видимому, с асимметрией отрывного обтекания фюзеляжа. Заметим, что последняя имеет разные знаки в разных аэродинамических трубах.

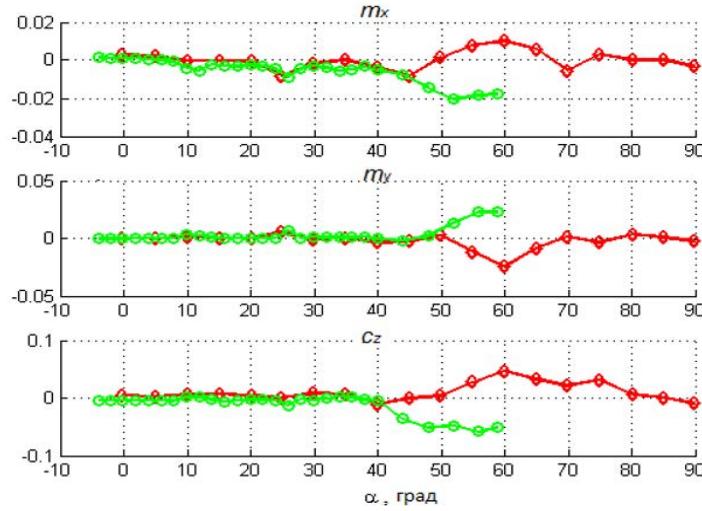


Рис. 4. Статические зависимости боковых аэродинамических коэффициентов модели 1, полученные в двух разных аэродинамических трубах при нулевом скольжении.

Зависимость коэффициентов момента крена  $\Delta m_x$  и рыскания  $\Delta m_y$  модели 1 от углов атаки и скольжения представлена на рис. 5. Видно, что асимметрия значительно проявляется себя в районе углов атаки  $\alpha = 25$  и  $60^\circ$ . Зависимость коэффициентов момента крена и рыскания от угла атаки и безразмерной угловой конического вращения для той же модели приведена на рис. 6. Видно, что влияние асимметрии заметно в тех же диапазонах углов атаки.

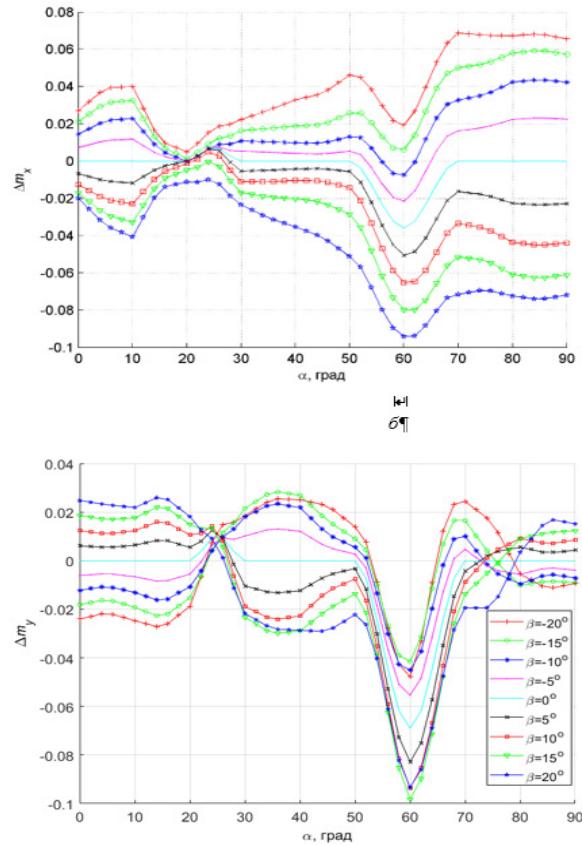


Рис. 5. Зависимость коэффициентов момента крена и рыскания модели 1 от угла атаки и скольжения.

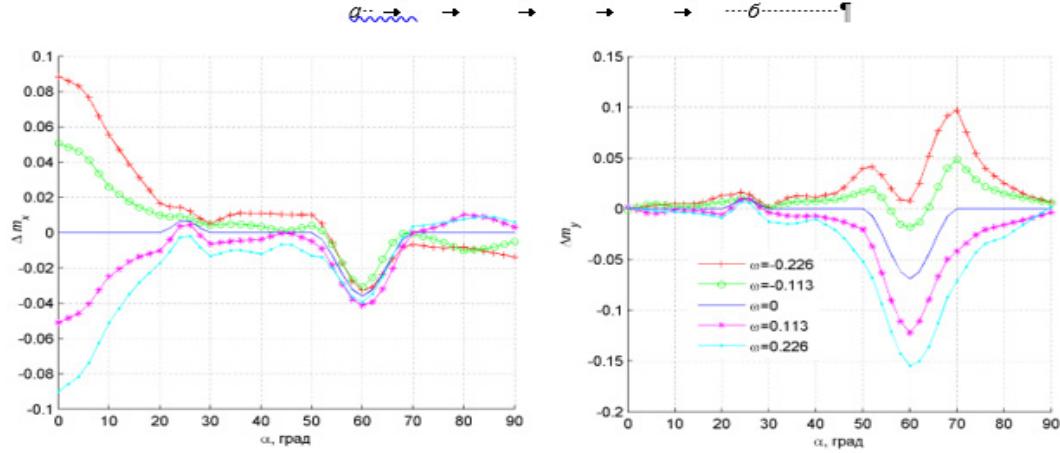


Рис. 6. Зависимость коэффициентов момента крена и рыскания модели 1 от угла атаки и безразмерной угловой скорости.

Зависимость коэффициента момента крена  $\Delta m_x$  от углов атаки и скольжения для модели 2 показана на рис. 7. Видно, что здесь максимальное влияние асимметрии наблюдается в районе  $\alpha = 15^\circ$  и связано с несимметричным отрывом на прямом крыле большого удлинения.

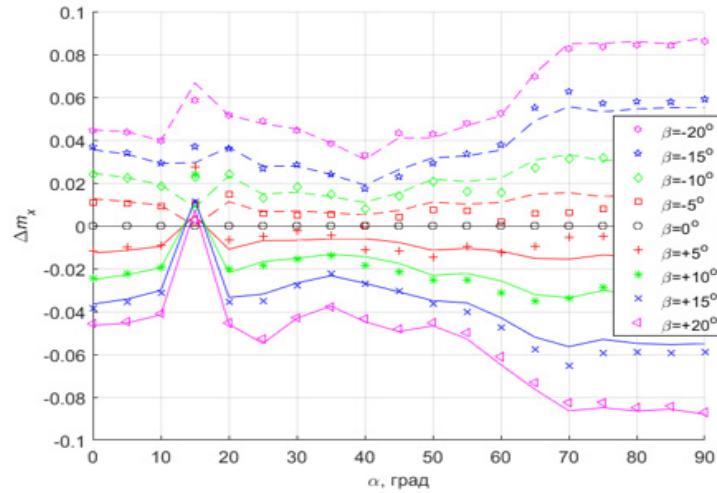


Рис. 7. Зависимость коэффициента момента крена модели 2 от углов атаки и скольжения.

**3. Влияние аэродинамической асимметрии на параметры критических режимов.** 3.1. Чувствительность параметров штопора к боковой аэродинамической асимметрии. Расчет параметров штопора и других критических режимов и их зависимости от аэродинамической асимметрии выполняется методом численного продолжения по параметру стационарных и периодических решений [7, 8]. Запишем уравнения (2.1) в виде

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \boldsymbol{\delta}),$$

где  $\mathbf{x} = (\alpha, \beta, V, \vartheta, \gamma, \omega_x, \omega_y, \omega_z)^T \in R^8$ ,  $\boldsymbol{\delta} = (\delta_B, \delta_\Theta, \delta_H)$ . Состояния равновесия определяются системой алгебраических уравнений:

$$\mathbf{F}(\mathbf{x}, \boldsymbol{\delta}) = 0. \quad (3.1)$$

В процедуре продолжения по параметру равновесных решений в качестве параметра используется отклонение одного из органов управления, а два других фиксируются.

Начальные точки кривых (3.1) в  $R^8$  можно найти приближенными методами [1]. При расчете равновесий также анализируется их локальная устойчивость с помощью линеаризован-

ной системы уравнений. Продолжение предельных циклов или периодических решений с неизвестным периодом  $T$  выполняется путем решения зависящей от параметра краевой задачи:

$$x(T, \delta) - x(0, \delta) = 0.$$

На рис. 8 и 9 приведены примеры параметров стационарных решений и амплитуд предельных циклов для модели 1 в зависимости от отклонения элеронов при  $\delta_b = -15^\circ$ ,  $\delta_h = 0$  для значения параметра асимметрии, совпадающего с наблюдаемым в АДТ ( $k_a = 1$ ), и для симметризованной модели аэродинамики ( $k_a = 0$ ). Разные маркеры (цвета) обозначают различные свойства устойчивости (качественно различные распределения собственных значений). Наблюдаемые в полете критические режимы соответствуют устойчивым равновесным решениям (равномерный штопор, инерционное вращение и др.) или устойчивым периодическим (предельные циклы) решениям (автоколебания, колебательный штопор).

Как видно из рис. 8, 9, имеются устойчивые равновесия и предельные циклы при  $\alpha \sim 20^\circ$ , которые практически не зависят от аэродинамической асимметрии. Предельные циклы в диапазо-

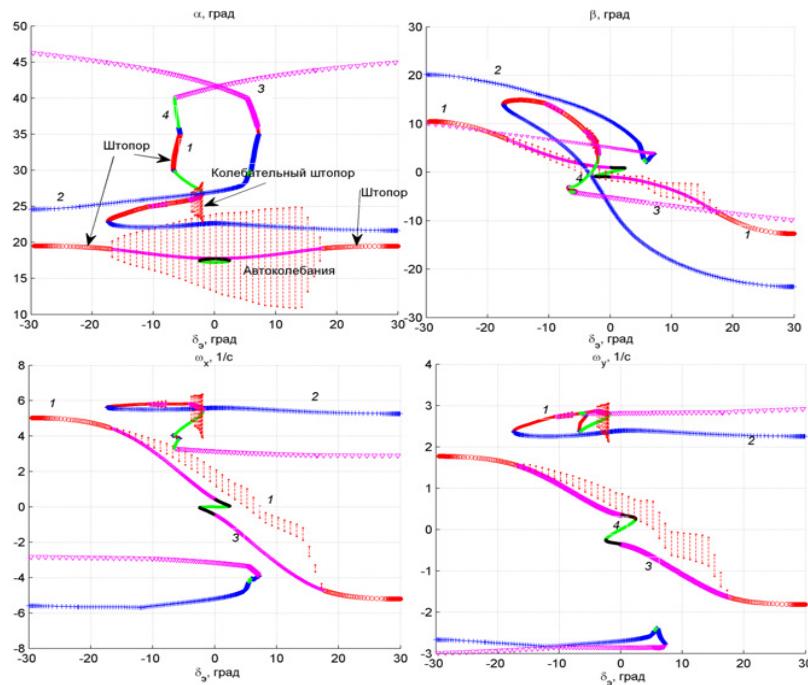


Рис. 8. Стационарные решения и предельные циклы для модели 1 в зависимости от отклонения элеронов ( $\delta_b = -15^\circ$ ,  $\delta_h = 0$ ), модель аэродинамики с асимметрией:

1 – устойчивые решения, 2 – апериодически неустойчивые, 3 – колебательно неустойчивые, 4 – три или более собственных значений в правой полуплоскости.

не  $\delta_g \in (-18; 18^\circ)$  описывают автоколебания с большой амплитудой по тангажу, которые переходят в режим авторотации при увеличении отклонения элеронов. Амплитуда автоколебаний, как видно из приведенных рисунков, в случае асимметрии зависит от знака отклонения элеронов.

Полученные диаграммы носят выраженный бифуркационный характер [2], различия между ними лежат главным образом в диапазоне углов атаки  $\alpha = 24\dots35^\circ$  (рис. 10). В случае асимметрии существует диапазон очень крутых режимов штопора при  $\delta_g = -17\dots0^\circ$ ,  $\alpha = 23\dots25^\circ$  и  $\delta_g = -4\dots0^\circ$ ,  $\alpha = 30\dots25^\circ$ , которые отсутствуют в симметричном случае. Первый диапазон этих режимов вызван исключительно аэродинамической асимметрией, второй диапазон зависит от неопределенности других аэродинамическими параметров. Он может появляться или исчезать при незначительном изменении симметричной части аэродинамики. Пример траектории в устойчивом колебательном штопоре, вызванном аэродинамической асимметрией, показан на рис. 11.

По результатам расчетов в рассматриваемом примере у модели 1 нет режимов штопора или других устойчивых критических режимов при углах атаки  $\alpha > 35^\circ$ , поэтому значительная

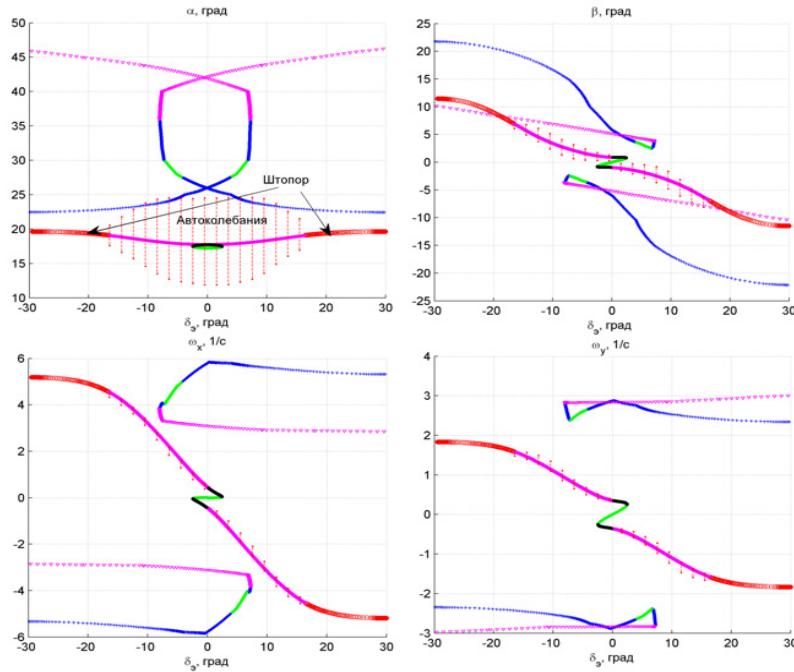


Рис. 9. Стационарные решения и предельные циклы для модели 1 в зависимости от отклонения элеронов ( $\delta_B = -15^\circ$ ,  $\delta_H = 0$ ), нет аэродинамической асимметрии.

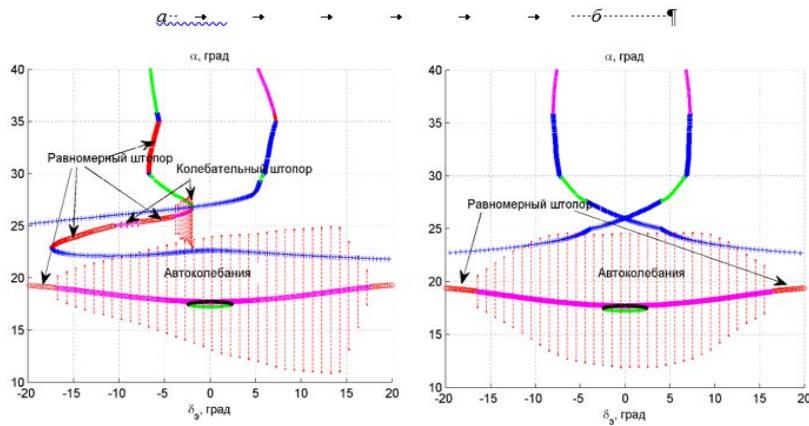


Рис. 10. Сравнение стационарных решений и предельных циклов для модели 1 в зависимости от отклонения элеронов ( $\delta_B = -15^\circ$ ,  $\delta_H = 0$ ): а – с асимметрией, б – без асимметрии.

аэродинамическая асимметрия в диапазоне углов атаки  $\alpha = 50\ldots70^\circ$  не оказывает заметного влияния на динамику самолета на этих углах атаки.

Поскольку величина аэродинамической асимметрии является весьма неопределенным параметром, необходимо проанализировать зависимость параметров штопора от ее величины. На рис. 12 и 13 приведены стационарные режимы штопора для различных значений параметра  $k_a$ , определяющего уровень аэродинамической асимметрии (здесь представлены только три компоненты:  $\alpha$ ,  $\beta$ ,  $\omega_x$  из всего вектора состояния). Видно, что диапазон отклонения элеронов, при котором возможен штопор, значительно уменьшается с уменьшением уровня аэродинамической асимметрии  $k_a$  и при  $k_a = 0.25$  штопора не существует. Таким образом, для надежного прогнозирования штопора самолета в полете важно точно оценивать величину аэродинамической асимметрии и уточнять ее при переносе данных, полученных в аэродинамической трубе для динамически подобной модели, на реальный полет самолета.

3.2. Чувствительность параметров штопора, вызванного аэродинамической асимметрией, к неопределенности других аэро-

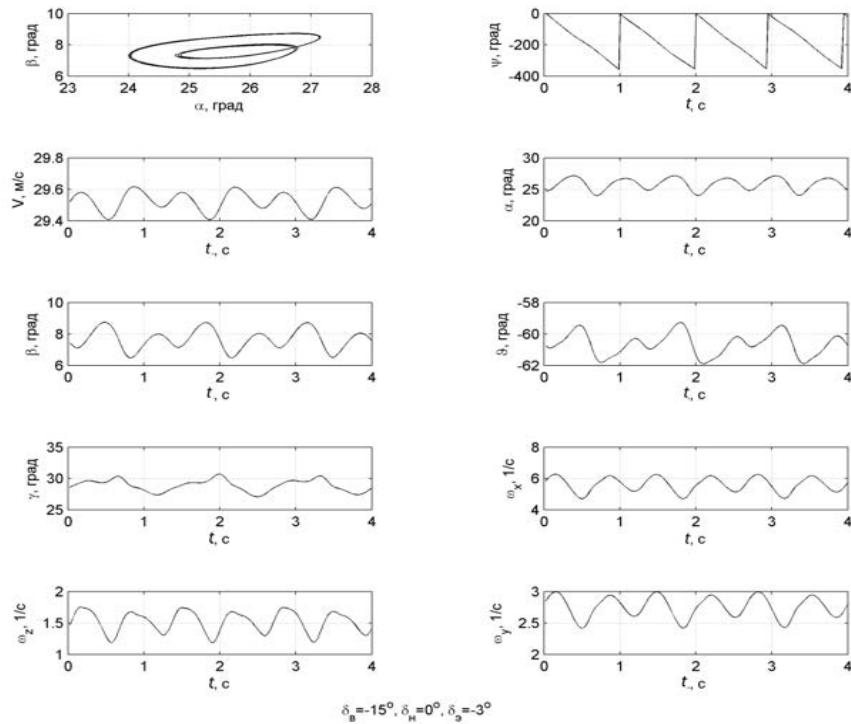


Рис. 11. Пример траектории в колебательном штопоре, обусловленном аэродинамической асимметрией.

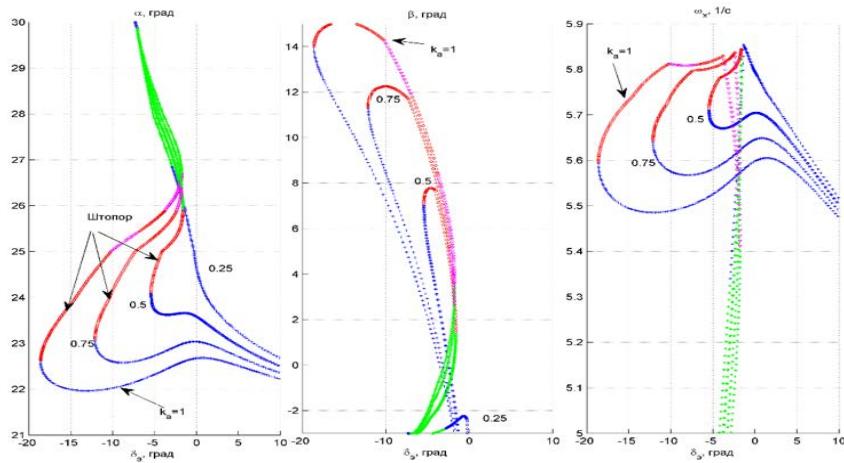


Рис. 12. Режимы штопора при различных значениях параметра аэродинамической асимметрии,  $\omega > 0$ .

динамических параметров. В работе [9] было обнаружено, что неопределенности аэродинамической модели могут существенно влиять на бифуркационные диаграммы, рассчитанные для анализа штопора, и параметры штопора. Значения неопределенностей, включенных в аэродинамическую модель в [9], были получены в результате многократных измерений статических характеристик самолета при различных комбинациях угла атаки и скольжения, различных методах и условиях экспериментов в АДТ. Анализ проводился путем прямого расчета параметров штопора с помощью метода продолжения по параметру.

В данной работе используется другой подход к анализу чувствительности параметров штопора к неопределенностям аэродинамической модели. Рассмотрим любой стационарный ре-

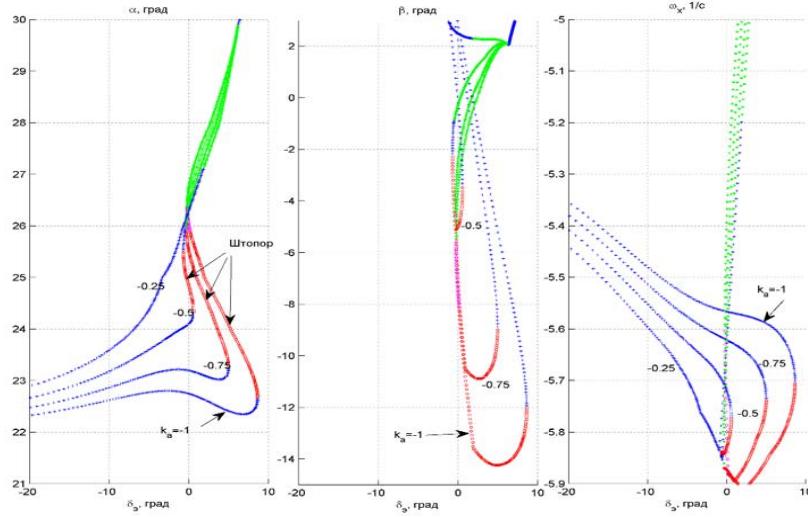


Рис. 13. Режимы штопора при различных значениях параметра аэродинамической асимметрии,  $\omega < 0$ .

жим штопора, вызванный аэродинамической асимметрией (рис. 8, 11, *a*). Его устойчивость определяется путем вычисления собственных значений системы 8-го порядка, линеаризованной вблизи этого устойчивого положения равновесия:

$$\Delta \dot{\mathbf{x}} = \mathbf{A} \Delta \mathbf{x}, \mathbf{A} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \Big|_{\mathbf{X}_S}, \quad (3.2)$$

где  $\mathbf{X}_S$  – какой-либо стационарный режим штопора. Определить, остается ли положение равновесия устойчивым, когда аэродинамические параметры изменяются вблизи их номинальных значений, возможно с применением методов анализа робастной устойчивости [15]. В частности, анализа с помощью структурированных сингулярных чисел –  $\mu$ -анализа, который гарантирует робастную устойчивость системы (3.2) с неопределенной матрицей  $\mathbf{A}$  при любом разбросе значений элементов матрицы  $\mathbf{A}$  внутри заданных границ, если  $\mu$  (верхняя граница структурированного сингулярного числа) для заданной линейной неопределенной системы меньше 1.

Поскольку рассматриваемая модель аэродинамики является эвристической, все параметры будут неопределенными. Было установлено [9], что неопределенности коэффициентов момента крена и рыскания оказывают особенно большое влияние на параметры штопора. По этой причине неопределенными считаем аэродинамические производные  $m_x^\alpha, m_x^\beta, m_x^{\omega_x}, m_x^{\omega_y}, m_y^\alpha, m_y^\beta, m_y^{\omega_x}, m_y^{\omega_y}$  все вместе или по отдельности, рассчитанные вблизи нескольких режимов штопора, которые обусловлены аэродинамической асимметрией. Это означает, что элементы  $A_{6,2}, A_{6,3}, A_{6,6}, A_{6,7}, A_{7,2}, A_{7,3}, A_{7,6}, A_{7,7}$  матрицы  $\mathbf{A}$  в уравнении (3.2) являются неопределенными. Считаем, что значение неопределенности каждой из этих производных равно 10%. На рис. 14 приведены значения  $\mu$ , рассчитанные для случаев одной из перечисленных неопределенностей или всех вместе. Видно, что система может потерять устойчивость в случае всех неопределенных производных для первого режима штопора ( $\alpha = 23.06^\circ$ ), но остается робастно устойчивой в трех других случаях. Отметим, что если равновесие становится неустойчивым, это не означает исчезновения штопора, так как установившийся штопор может стать колебательным и ответ может быть дан с помощью бифуркационного анализа.

Величина  $\text{margin} = 1 / \mu$  дает запас устойчивости неопределенной системы: система устойчива, если норма  $i$ -й неопределенности меньше  $\text{margin} \beta_i$ . Рисунок 14 позволяет оценить, неопределенность какой производной больше влияет на робастную устойчивость штопора: в данном случае это производные  $m_x^{\omega_x}$  и  $m_y^\alpha$ . В случае робастного анализа устойчивости режимов штопора чем больше запас устойчивости, тем более опасным является этот режим. Другими словами, этот режим штопора существует, несмотря на неточное знание аэродинамических характеристик самолета, и это важный вывод для эвристических моделей аэродинамики.

На рис. 15 показаны бифуркационные диаграммы, аналогичные части рис. 8, 10, *a*, при различных неопределенностях моментов рыскания и крена в рассматриваемых пределах. Видно, что крутой штопор при  $\alpha = 30\dots35^\circ$  изменяется незначительно, но область очень крутого штопора при  $\alpha = 23\dots28^\circ$ , обусловленного аэродинамической асимметрией, может значительно измениться.

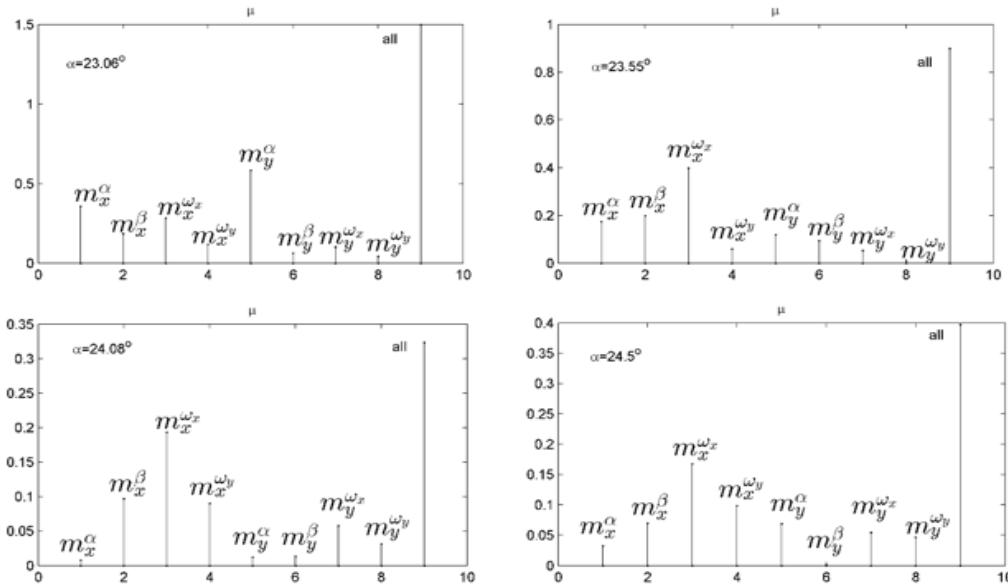


Рис. 14. Робастная устойчивость режимов штопора, обусловленного аэродинамической асимметрией при различных неопределенностях аэродинамических производных в моментах рыскания и крена.

**4. Влияние аэродинамической асимметрии на восприимчивость самолета к штопору.** Влияние аэродинамической асимметрии на параметры критических режимов и восприимчивость к штопору магистрального самолета может существенно зависеть от типа самолета. Для сравнительного анализа рассмотрим критические режимы для модели самолета 2, для которого уровень асимметрии боковых аэродинамических характеристик выше, чем для модели 1. Зависимость коэффициента момента крена от углов атаки и скольжения для модели 2 приведена на рис. 7. Параметры стационарных критических режимов, вычисленные методом продолжения по параметру при имеющейся аэродинамической асимметрии, в зависимости от отклонения руля направления при  $\delta_B = -5^\circ$ ,  $\delta_e = 0$  представлены на рис. 16. В отличие от ранее рассмотренного типа самолета, в данном случае во всем диапазоне отклонения руля направления имеются режимы устойчивого штопора, диапазон углов атаки которых составляет от 20 до  $55^\circ$ , причем при

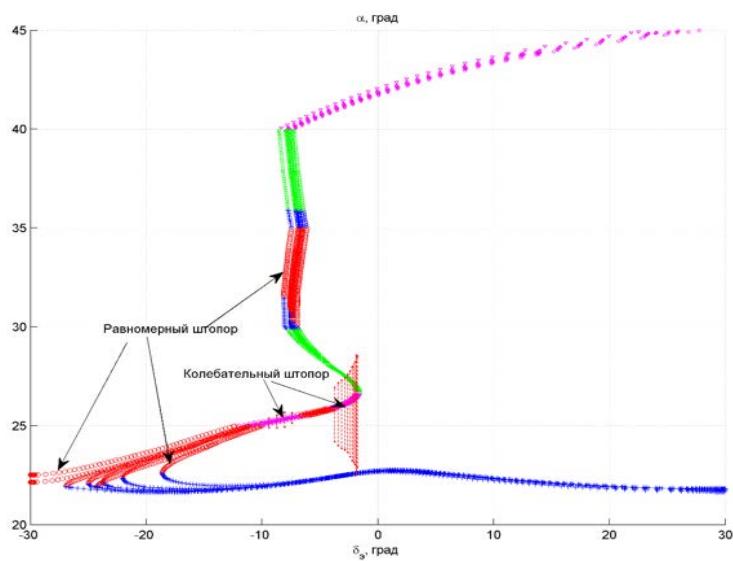


Рис. 15. Параметры штопора, вызванного аэродинамической асимметрией, в зависимости от отклонения элеронов, при различных неопределенностях моментов рыскания и крена.

некоторых отклонениях руля направления имеется несколько устойчивых режимов штопора. Параметры правого и левого штопора различны. Чтобы оценить, в какой из режимов штопора может попасть самолет, желательно ввести меру восприимчивости самолета к штопору.

Эффективный подход к оценке восприимчивости самолета к штопору был предложен в [16], где для этой цели использовался размер области притяжения (ОП) устойчивого режима штопора, рассматриваемый как метрика для оценки относительной опасности различных режимов штопора и вероятность попадания в конкретный режим в случае нескольких устойчивых режимов для одного и того же набора параметров. Оценка размера ОП проведена с помощью разработанных в последние годы количественных методов анализа нелинейных систем, применяющихся до этого лишь для анализа нормальных режимов полета. Количественные методы анализа нелинейных систем основаны на SOS-оптимизации (sum-of-squares – сумма квадратов) [17–19].

Поскольку количественные методы анализа [17–19] применимы лишь к полиномиальным системам, получены полиномиальные аппроксимации редуцированной системы 5-го порядка, аппроксимирующей систему 8-го порядка (2.1) с учетом нелинейности динамики и всех нелинейностей аэродинамических характеристик самолета вблизи ряда устойчивых режимов штопора [16], подтвержденные сравнительным моделированием. Для полученных прибли-

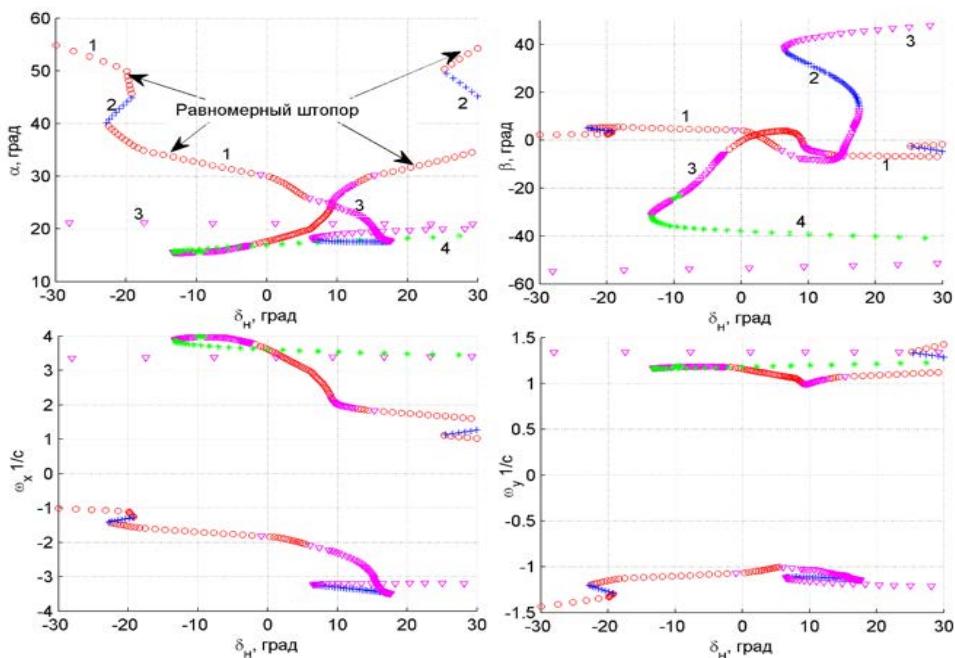


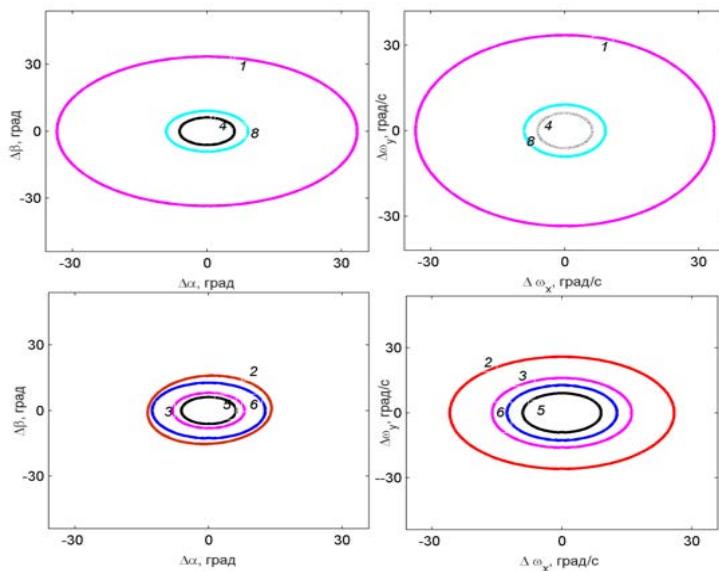
Рис. 16. Параметры стационарных критических режимов для модели 2 в зависимости от отклонения руля направления при  $\delta_b = 5^\circ$ ,  $\delta_s = 0$ : 1 – устойчивые решения, 2 – апериодически неустойчивые, 3 – колебательно неустойчивые, 4 – три или более собственных значений в правой полуплоскости.

женных полиномиальных уравнений 5-го порядка оценены нижние границы эллипсоидальных аппроксимаций областей притяжения с использованием метода оптимизации на основе функций Ляпунова, дающих оценку области притяжения. Размеры области притяжения для различных режимов штопора могут рассматриваться как мера восприимчивости самолета к штопору и относительной опасности отдельных штопоров.

Результаты оценок ОП для режимов штопора, перечисленных в таблице, приведены на рис. 17. Видно, что из-за боковой аэродинамической асимметрии правый и левый режимы штопора имеют разные размеры области притяжения, т.е. разные вероятности попадания в эти режимы штопора. Сравнение ОП для двух режимов, существующих при одних и тех же значениях параметров (режимы № 2:  $\delta_h = -20^\circ$ ,  $\alpha = 50^\circ$  и № 3:  $\delta_h = -20^\circ$ ,  $\alpha = 37^\circ$  в таблице), показывает, что второй режим имеет несколько больший размер ОП и, следовательно, более вероятен. Сравнение ОП для двух разных режимов штопора при  $\delta_h = 30^\circ$  (режимы № 7:  $\alpha = 35^\circ$  и № 8:  $\alpha = 54^\circ$ ) показывает, что в этом случае размеры ОП практически равны, т.е. эти режимы равновероятны. Сравнение ОП правого и левого штопора (№ 1:  $\delta_h = -30^\circ$ ,  $\alpha = 55^\circ$  и № 8:

$\delta_h = 30^\circ$ ,  $\alpha = 55^\circ$ , а также № 3:  $\delta_h = -20^\circ$ ,  $\alpha = 37^\circ$  и № 6:  $\delta_h = 20^\circ$ ,  $\alpha = 31.5^\circ$ ) показывает, что для левого штопора (отрицательные отклонения руля направления) размер области притяжения значительно больше, чем для правого. Таким образом, аэродинамическая асимметрия существенно влияет на размер области притяжения штопора и соответственно на восприимчивость самолета к правому или левому штопору.

**5. Влияние уровня аэродинамической асимметрии на управление при выводе самолета из штопора для двух типов самолета.** Известно, что аэродинамическая асимметрия играет важную роль в характере штопора истребителей, делая некоторые режимы штопора невыводимыми [1,2]. Как показывает анализ, для модели 1 величина аэродинамической асимметрии практически не влияет на вывод из штопора, который может быть осуществлен приведением органов управления в нейтральное положение. Для модели 2 это не так, и аэродинамическая асимметрия влияет на вывод из самолета штопора. Это имеет место для штопора при нейтральном положении органов управления. Отклонение руля направления против штопора на несколько секунд приводит к выходу из штопора в случае симметричной модели аэродинамики (рис. 18, *a*) и не приводит к выходу из штопора при учете аэродинамической асимметрии (рис. 18, *b*).



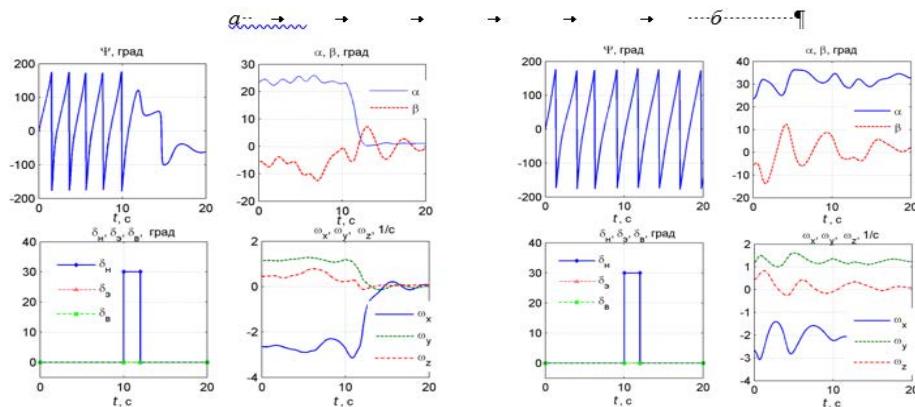
**Рис. 17.** Проекции областей притяжения для нескольких режимов штопора, цифры соответствуют номерам режимов в таблице: 1 –  $\delta_h = -30^\circ$ ,  $\alpha = 55^\circ$ ; 8 –  $\delta_h = 30^\circ$ ,  $\alpha = 55^\circ$ ; 4 –  $\delta_h = 0^\circ$ ,  $\alpha = 30^\circ$ ; 2 –  $\delta_h = -20^\circ$ ,  $\alpha = 50^\circ$ ; 3 –  $\delta_h = -20^\circ$ ,  $\alpha = 37^\circ$ ; 5 –  $\delta_h = 0^\circ$ ,  $\alpha = 18^\circ$ ; 6 –  $\delta_h = 20^\circ$ ,  $\alpha = 32^\circ$ ; проекции для режима 7 ( $\delta_h = -30^\circ$ ,  $\alpha = 35^\circ$ ) практически совпадают с проекциями для режима 8.

**Таблица 1.** Список оцениваемых режимов штопора

Номер режима	$\delta_h$ , град	V, м/с	$\alpha$	$\beta$	$\omega_x$	$\omega_y$	$\omega_z$
			град			$\text{с}^{-1}$	
1	-30	77.8	54.8	2.34	-1.01	-1.43	0.029
2	-20	83.8	50.1	2.43	-1.11	-1.33	0.037
3	-20	101.8	36.8	5.34	-1.53	-1.15	-0.046
4	0	119.8	30.0	4.23	-1.83	-1.07	-0.012
5	0	174.3	17.6	-0.39	3.62	1.16	0.155
6	20	114.4	31.5	-6.59	1.75	1.09	-0.097
7	30	106.4	34.8	-6.87	1.59	1.12	-0.099
8	30	78.4	54.3	-1.78	1.023	1.43	0.047

Штопор при нейтральном положении органов управления может произойти, например, в результате интенсивного управления рулем высоты и рулем направления, как на рис. 19, 20. Приведение органов управления в нейтральное положение не прекращает штопора, но кратковременное отклонение руля направления против штопора при отсутствии асимметрии приводит к успеху (рис. 18, *a*).

При наличии аэродинамической асимметрии такой способ оказывается неприменим. Однако в этом случае самолет может быть выведен из штопора так называемым усиленным методом: одновременное отклонение руля направления против штопора, (элеронов по штопору или в нейтральное положение, если они отклонены), руля высоты на кабрирование до прекращения вращения, затем перевод всех органов управления в нейтральное положение (рис. 20). Этот способ применим для вывода из штопора рассматриваемого самолета при любом уровне аэродинамической асимметрии. Таким образом, приведенный пример показывает возможность влияния аэродинамической асимметрии на параметры и выход магистрального самолета из штопора и важность учета этого фактора при разработке рекомендаций летчикам по способам пилотирования для вывода самолета из штопора.



**Рис. 18.** а – симметричная модель аэродинамики: выход из штопора, б – модель аэродинамики с учетом асимметрии: невыход.

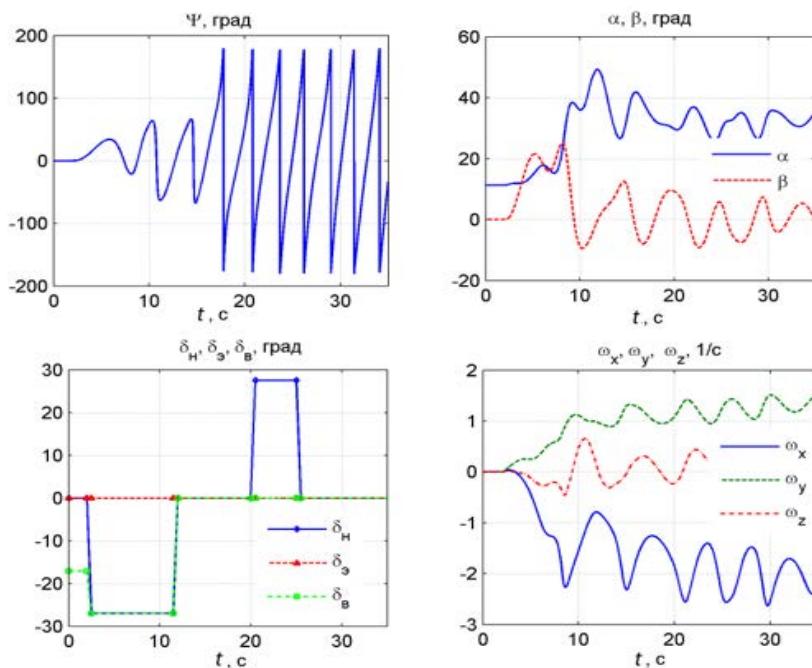


Рис. 19. Модель аэродинамики с учетом асимметрии: невыход из штопора.

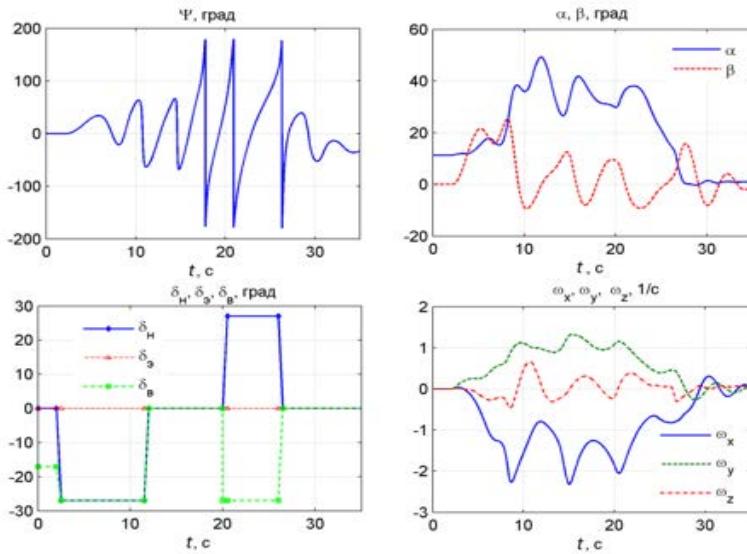


Рис. 20. Модель аэродинамики с учетом асимметрии, выход усиленным методом пилотирования.

**Заключение.** Исследованы вопросы влияния развития аэродинамической асимметрии на особенности критических режимов полета и, в частности, штопора, для двух моделей дозвуковых самолетов. Показано, что это влияние зависит от величины боковых моментов, вызываемых аэродинамической асимметрией, и может быть значительным, обусловливая появление устойчивых режимов штопора, не существующих при ее отсутствии или малой величине, а также различие параметров правого и левого штопора. Оценка размера области притяжения различных режимов штопора, рассматриваемая как мера восприимчивости самолета к штопору, показывает различие относительной опасности правого и левого штопора при наличии аэродинамической асимметрии. Продемонстрировано также влияние аэродинамической асимметрии на возможный способ выхода самолета из штопора и важность учета этого фактора при разработке рекомендаций летчикам по способам вывода самолета из штопора.

#### СПИСОК ЛИТЕРАТУРЫ

1. Аэродинамика, устойчивость и управляемость сверхзвуковых самолетов / Под. ред. Г. С. Бюшгенса. М.: Наука. Физматлит, 1998.
2. Goman M.G., Khrabrov A.V. Application of Continuation and Bifurcation Methods to the Design of Control Systems // Philosophical Transactions of the Royal Society of London. Ser. A: Mathematical, Physical, and Engineering Sciences. 1998. V. 356. P. 2277–2295.
3. <https://commonresearchmodel.larc.nasa.gov/geometry/>.
4. Jackson P., Munson K., Peacock L. Jane's All the Worlds Aircraft 2005–2006. Surrey: Jane's Information Group. 2006.
5. Richardson T.S., McFarlane C., Isikveren A., Badcock K.J., Da Ronch A. Analysis of Conventional and Asymmetric aircraft Configurations using CEASIOM // Progress in Aerospace Sciences. 2011. V. 47. № 8. P. 647–659.
6. Khrabrov A., Sidoryuk M. Flight Control Law Design for Asymmetrical General Aviation Aircraft // Frontiers in Aerospace Engineering. 2013. V.2. № 4. P. 258–266.
7. Goman M.G., Zagaynov G.I., Khrabrov A.V. Application of Bifurcation Methods to Nonlinear Flight Dynamics Problems // Progress in Aerospace Sciences. 1997. V. 33. № 59. P. 539–586.
8. Guicheteau P. Bifurcation Theory: a Tool for Nonlinear flight Dynamics // Philosophical Transactions of the Royal Society of London. Ser. A. 1998. № 356. P. 2181–2201.
9. Farcy D., Khrabrov A., Sidoryuk M. Sensitivity of Spin Parameters to Uncertainties of the Aircraft Aerodynamic Model // Journal of Aircraft. 2020. V. 57. № 5. P. 1–16.
10. Гоман М.Г., Захаров С.Б., Храбров А.Н. Симметричное и несимметричное отрывное обтекание крыла малого удлинения с фюзеляжем // Уч. зап. ЦАГИ. 1985. № 6. С. 1–8.
11. Гоман М.Г., Захаров С.Б., Храбров А.Н. Аэродинамический гистерезис при стационарном отрывном обтекании удлиненных тел // Докл. АН СССР. 1985. № 1. С. 28–31.
12. ГОСТ 20058-80. Динамика летательных аппаратов в атмосфере. Термины, определения и обозначения. М., 1981.
13. ГОСТ 21890-76. Фюзеляж, крылья и оперение самолетов и вертолетов. Термины и определения. М., 1976.
14. ГОСТ 22833-77. Характеристики самолета геометрические. Термины, определения и буквенные обозначения. М., 1987.

15. *Packard A., Balas G., Safonov M., Chiang R., Gahinet P., Nemirovski A., Apkarian P.* Robust Control Toolbox // The MathWorks Inc. Natick. 2016.
16. *Sidoryuk M., Khrabrov A.* Estimation of Regions of Attraction of Aircraft Spin Modes // Journal of Aircraft. 2019. V. 56. № 1. P. 205–216.
17. *Topcu U., Packard A., Seiler P., Wheeler T.* Stability Region Analysis Using Simulations and Sum-Of-Squares Programming // Proc. American Control Conf. N.Y., 2007. P. 6009–6014.
18. *Topcu U., Packard A., Seiler P.* Local Stability Analysis Using Simulations and Sum-Of-Squares Programming // Automatica. 2008. V. 44. № 10. P. 2669–2675.
19. *Tibken B.* Estimation of the Domain of Attraction for Polynomial Systems via LMIS // Proc. IEEE Conf. on Decision and Control. Sydney, 2000. P. 3860–3864.

## СТАЦИОНАРНЫЕ РЕЖИМЫ ВЕТРОЭНЕРГЕТИЧЕСКОЙ УСТАНОВКИ НА БАЗЕ РОТОРА САВОНИУСА ПРИ НАЛИЧИИ СТАБИЛИЗАТОРА НАПРЯЖЕНИЯ<sup>1</sup>

© 2024 г. В. М. Буданов<sup>а</sup>, А. П. Голуб<sup>а</sup>, М. З. Досаев<sup>а</sup>,  
К. В. Климов<sup>а</sup>, Ю. Д. Сельютский<sup>а</sup>, \*

<sup>а</sup>НИИ механики МГУ, Москва, Россия

\*e-mail: seliutski@imec.msu.ru

Поступила в редакцию 02.04.2024 г.

После доработки 17.04.2024 г.

Принята к публикации 16.10.2024 г.

Исследуется малая ветроэнергетическая установка, рабочим элементом которой является ротор Савониуса. В электрическую цепь установки включен стабилизатор напряжения и нагрузочное сопротивление. На основе феноменологических соображений построена математическая модель рассматриваемой электромеханической системы. В результате аналитического исследования этой модели показано, что при определенных условиях на параметры в системе возникает явление гистерезиса. Необходимо отметить, что такой гистерезис обусловлен именно наличием стабилизатора напряжения. Проведена серия экспериментов. С использованием полученных экспериментальных данных проведена идентификация параметров предложенной модели. Результаты экспериментов достаточно хорошо согласуются с результатами аналитического исследования. Предложен алгоритм регулирования нагрузочного сопротивления, позволяющий обеспечить максимальную выходную мощность.

**Ключевые слова:** математическое моделирование, ветроэнергетическая установка, стабилизатор напряжения, идентификация параметров, гистерезис, управление.

**DOI:** 10.31857/S0002338824060063, **EDN:** SUPMAV

## STEADY REGIMES OF A SAVONIUS ROTOR-BASED WIND POWER GENERATOR WITH VOLTAGE STABILIZER

V. M. Budanov<sup>а</sup>, A. P. Holub<sup>а</sup>, M. Z. Dosaev<sup>а</sup>, K. V. Klimov<sup>а</sup>,  
Y. D. Selyutskiy<sup>а</sup>, \*

<sup>а</sup>Institute of Mechanics of Lomonosov Moscow State University, Moscow, Russia

\*e-mail: seliutski@imec.msu.ru

A small-scale wind power generator is studied, where the operating element is a Savonius rotor. The electric circuit of the generator comprises a voltage stabilizer and a load resistance. Based on phenomenological considerations, a mathematical model of this electro-mechanical system is constructed. In the result of analytical study of this model, it is shown that the hysteresis phenomenon occurs in this system under certain conditions imposed on parameters. It should be noted that this hysteresis is due to the presence of the voltage stabilizer. A series of experiments is performed. Based on the obtained experimental data, parameters of the proposed model are identified. Results of experiments are in good agreement with results of the analytical study. An algorithm of regulation of the load resistance is proposed aimed at reaching the maximum output power.

**Keywords:** mathematical modeling, voltage stabilizer, wind power generator, parameter identification, hysteresis, control.

<sup>1</sup> Исследование выполнено при финансовой поддержке НОШ МГУ «Фундаментальные и прикладные исследования космоса» (проект № 23-Ш01-11).

**Введение.** Одним из наиболее распространенных типов малогабаритных ветротурбин, используемых для преобразования энергии ветра в электроэнергию, является ротор Савониуса, который был разработан и запатентован С. Савониусом в 1926 г. «Классический» ротор Савониуса состоит из двух лопастей, имеющих форму полуцилиндра. Лопасти закреплены на оси, параллельной их образующей, и могут частично перекрываться. Ветротурбина устанавливается таким образом, что ее ось вращения перпендикулярна потоку среды (воздуха или воды). В результате воздействия потока на лопасти возникает аэродинамический момент, приводящий турбину во вращение. При этом ведущую роль в формировании этого момента играет сила лобового сопротивления.

Ротор Савониуса вращается заметно медленнее и вырабатывает меньшую мощность, чем ветротурбины других типов. Однако он начинает вращаться при достаточно низкой скорости потока (существенно меньшей, чем турбины типа Дарье) и не зависит от направления потока (в отличие от горизонтально-осевых турбин, которым необходима система ориентации на ветер). Способность достаточно эффективно функционировать в диапазоне малых скоростей потока обуславливает интерес к этому ротору в ситуациях, когда требуется обеспечить автономное питание малых потребителей [1]. Помимо этого ротор Савониуса создает достаточно большой момент на оси, что открывает перспективы его использования в качестве силового привода.

Одним из основных направлений исследований роторов Савониуса является изучение влияния различных геометрических характеристик ротора на его поведение (достаточно подробный обзор таких работ приведен, в частности, в [2, 3]). Зависимости мощности и момента, создаваемых ветротурбиной указанного выше типа, от числа лопастей (как правило, от двух до четырех) анализируются в целом ряде публикаций (например, [4, 5]). Отмечается, что для конфигурации с двумя лопастями максимальная мощность выше, но для нее имеются положения, из которых она не может стартовать, и угловая скорость сильнее изменяется на обороте.

В [6–8] изучается влияние перекрытия лопастей на выходную мощность и отмечается, что имеется некоторое оптимальное соотношение между размером этого перекрытия и радиусом лопасти. Ряд работ посвящен анализу аэродинамики ветротурбин, состоящих из нескольких роторов Савониуса, закрепленных друг над другом на общей оси [9–11]. Такая конфигурация обеспечивает отсутствие положений, из которых ротор не может стартовать, и уменьшает колебания угловой скорости на обороте.

Активно исследуются возможности увеличения выходной мощности путем изменения формы лопастей. В частности, в [12–16] рассматриваются турбины, лопасти которых в поперечном сечении представляют собой не полуокружности, а более сложные кривые (комбинации дуг окружности с прямолинейными отрезками, дуги эллипсов и т.д.). В [17, 18] для оптимизации формы лопастей используются генетические алгоритмы. Проведенные эксперименты и расчеты показывают, что выигрыш в мощности за счет оптимизации формы лопасти может достигать 35%.

Еще одним способом повысить эффективность роторов Савониуса с точки зрения мощности является установка различных приспособлений, специальным образом отклоняющих поток, который набегает на лопасти (например, [17, 19, 20]). В работе [21] построена замкнутая математическая модель ветроэнергетической установки, рабочий элемент которой представляет собой ротор Савониуса. Был предложен феноменологический подход для описания зависимости аэродинамических сил и момента, действующих на лопасти, от угловой скорости и угла поворота ротора.

Естественной задачей, которую должна решать малая ветроэнергетическая установка, является зарядка аккумулятора. Один из типичных режимов зарядки – зарядка постоянным напряжением. Поскольку напряжение, создаваемое генератором, который приводится во вращение ветротурбиной, как правило, не совпадает с оптимальным напряжением для зарядки, необходимо добавить в электрическую цепь стабилизатор напряжения. Исследованию таких стабилизаторов посвящены работы [22, 23]. Однако в них анализируется только функционирование стабилизатора в нормальном режиме. В то же время при исследовании динамики ветроэнергетической установки необходимо учитывать, что мощности, отбираемой у потока, может оказаться недостаточно для обеспечения работы стабилизатора в нормальном режиме. Таким образом, возникает задача исследования стационарных режимов ветроэнергетической установки, в цепь которой включен стабилизатор напряжения. Решению этой задачи и посвящена настоящая статья.

**1. Постановка задачи и математическая модель системы.** Рассмотрим электромеханическую систему (рис. 1), состоящую из ветротурбины 1 в форме ротора Савониуса, ось которой жестко соединена с валом электрогенератора 2 переменного тока. В цепь генератора включен вы-

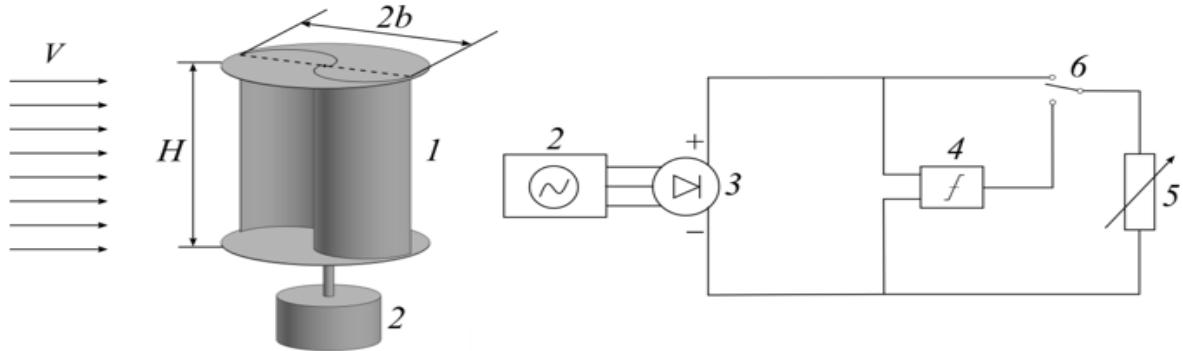


Рис. 1. Схема рассматриваемой системы

прямитель 3 и стабилизатор напряжения 4, а также нагрузочное сопротивление 5. Кроме того, в цепи имеется ключ 6, позволяющий подключать сопротивление к генератору либо через стабилизатор, либо напрямую. Ветротурбина помещена в стационарный поток среды.

Пренебрегая потерями на трение, уравнение динамики ветротурбины можно записать в следующем виде:

$$J\dot{\omega} = M_a + M_{el}. \quad (1.1)$$

Здесь  $\omega$  — угловая скорость ветротурбины,  $J$  — момент инерции ветротурбины вместе с ротором генератора,  $M_a$  — аэродинамический момент, действующий на лопасти ветротурбины,  $M_{el}$  — момент электромагнитных сил, действующих на вал ветротурбины со стороны генератора. Аэродинамический момент традиционно представляется в следующем виде:

$$M_a = \rho b^2 H V^2 C_m. \quad (1.2)$$

Здесь  $\rho$  — плотность среды,  $b$  — радиус ротора,  $H$  — высота ротора,  $V$  — скорость набегающего потока,  $C_m$  — безразмерный коэффициент аэродинамического момента. Для описания аэродинамики воспользуемся квазистатическим подходом. В его рамках коэффициент момента зависит только от безразмерной угловой скорости (так называемой быстроходности)  $\Omega$ :

$$\Omega = \frac{b\omega}{V}. \quad (1.3)$$

Рассмотрим генератор и выпрямитель как единый блок (в некотором смысле, генератор постоянного тока) и смоделируем его работу в целом. Как и в [24], будем считать, что электромеханическое взаимодействие является линейным. Тогда электродвижущая сила (ЭДС), вырабатываемая генератором, и электромагнитный момент, действующий на вал его ротора, определяются следующими формулами:

$$E = c\omega, \quad M_{el} = -cI_{in}, \quad (1.4)$$

где  $c$  — коэффициент электромагнитного взаимодействия,  $I_{in}$  — ток в цепи якоря (и ток на входе стабилизатора). Таким образом, с учетом соотношений (1.2)–(1.4) уравнение (1.1) принимает вид

$$J\dot{\omega} = \frac{\rho S}{2} V^2 b C_m(\Omega) - cI_{in}. \quad (1.5)$$

Предположим, что характерное время протекания процессов в электрической цепи существенно меньше, чем характерное время переходных процессов в механической части системы. Тогда можно считать, что ток  $I_{in}$  определяется мгновенной угловой скоростью  $\omega$  якоря. Для того чтобы найти эту зависимость, необходимо описать процессы, проходящие в стабилизаторе. Ввиду отсутствия общепринятых достаточно простых моделей стабилизатора, описывающих его функционирование на режимах, отличных от номинального, воспользуемся феноменологическим подходом. Запишем уравнение баланса мощности для стабилизатора:

$$P_{in} = P_{out} + P_s + P_l. \quad (1.6)$$

Здесь  $P_{in}$  – мощность на входе стабилизатора,  $P_{out}$  – мощность на его выходе  $P_s$  – мощность, расходуемая внутри стабилизатора (на функционирование системы управления и т.д.),  $P_l$  – мощность потерь. Будем считать, что  $P_l = \xi' P_{in} = \xi' U_{in} I_{in}$ , где  $0 < \xi' < 1$  – некоторый постоянный коэффициент, а  $U_{in}$  и  $I_{in}$  – напряжение и ток на входе стабилизатора. Предположим также, что  $P_s = U_{in} I_s$ , где  $I_s$  – некоторый постоянный коэффициент («внутренний» ток в стабилизаторе). Выходная мощность, очевидно, связана с выходным напряжением  $U_{out}$  следующей формулой:  $P_{out} = U_{out}^2 / R$ .

С учетом изложенного выше из (1.6) получаем

$$(1 - \xi') U_{in} I_{in} = \frac{U_{out}^2}{R} + U_{in} I_s. \quad (1.7)$$

Далее для сокращения записи будем обозначать  $\xi = 1 - \xi'$  (эта величина, разумеется, положительна). Отметим, что напряжение на входе стабилизатора зависит от режима работы стабилизатора и определяется следующим выражением:

$$U_{in} = c\omega - rI_{in}, \quad (1.8)$$

где  $r$  – эффективное внутреннее сопротивление генератора (вместе с выпрямителем).

Для того чтобы система уравнений (1.7), (1.8) была замкнутой, необходимо связать выходное напряжение с входным. Воспользуемся следующими соображениями. Естественно ожидать, что если входное напряжение слишком мало, то стабилизатор отключается, при этом цепь разрывается и выходное напряжение, а также входной и выходной токи становятся равными нулю. Если же входное напряжение достаточно велико, то реализуется номинальный режим, на котором выходное напряжение почти постоянно и близко к номинальному значению  $U_n$ . Вообще говоря, пороговое значение  $U_{in}^*$  отключения стабилизатора и пороговое значение  $U_{in}^{**}$  выхода на номинальный режим не совпадают. Будем считать, что на интервале  $U_{in}^* < U_{in} < U_{in}^{**}$  величина  $U_{out}$  зависит от  $U_{in}$  линейно, причем при  $U_{in} = U_{in}^*$  возможен разрыв, а при  $U_{in} = U_{in}^{**}$  зависимость является непрерывной. Таким образом, получаем следующую формулу:

$$U_{out} = \begin{cases} 0, & 0 < U_{in} < U_{in}^*, \\ kU_{in} - U_0, & U_{in}^* < U_{in} < U_{in}^{**}, \\ U_n, & U_{in}^{**} < U_{in}. \end{cases} \quad (1.9)$$

Здесь величины  $U_{in}^*$ ,  $k$  и  $U_0$  – некоторые положительные параметры. Величина  $U_{in}^{**}$  определяется так, чтобы обеспечить непрерывность зависимости  $U_{in}$  от  $U_{out}$ :

$$U_{in}^{**} = \frac{U_n + U_0}{k}.$$

Зависимость (1.9) качественно изображена на рис. 2. Заметим, что она, в сущности, представляет собой характеристику стабилизатора. Ниже будет продемонстрировано, что для реального стабилизатора соответствующая характеристика выглядит так, как показано на рис. 2.

С учетом вышеизложенного, соотношения (1.5), (1.7)–(1.9) образуют замкнутую систему, описывающую динамику рассматриваемой ветроэнергетической установки. Выразим входной ток из (1.8) и подставим полученное выражение вместе с (1.9) в уравнение (1.7). Получится уравнение относительно  $U_{in}$ . Рассмотрим его решения на каждом из указанных выше интервалов.

1.1. И н т е р в а л  $0 < U_{in} < U_{in}^*$  (р е ж и м о т к л ю ч е н и я). Здесь соотношение (1.7) принимает следующий вид:

$$\xi U_{in} I_{in} = U_{in} I_s. \quad (1.10)$$

На этом интервале стабилизатор разрывает цепь, так что входной ток также равен нулю. Соответственно из (1.8) имеем  $U_{in} = c\omega$ . Для того, чтобы это выражение являлось решением уравнения (1.10), следует положить на этом интервале  $I_s = 0$ . Очевидно, значение  $U_{in}^*$  достигается при

$$\omega = \omega_s = \frac{U_{in}^*}{c}. \quad (1.11)$$

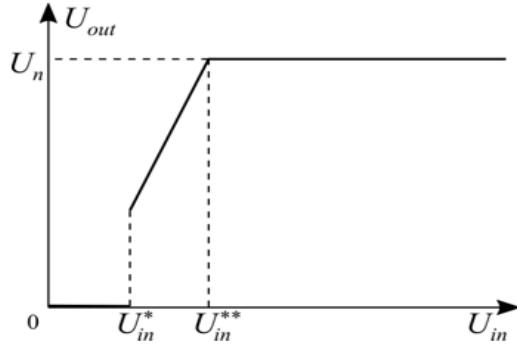


Рис. 2. Зависимость напряжения на выходе стабилизатора от напряжения на входе

Таким образом, этот режим реализуется при  $0 < \omega < \omega_s$ .

1.2. И н т е р в а л  $U_{in}^* < U_{in} < U_{in}^{**}$  (п е р е х о д н ы й р е ж и м). Здесь уравнение (1.7) принимает следующий вид:

$$\xi U_{in} \frac{c\omega - U_{in}}{r} = \frac{(kU_{in} - U_0)^2}{R} + U_{in} I_s.$$

Отсюда получаем:

$$\left( \frac{\xi}{r} + \frac{k^2}{R} \right) U_{in}^2 - \left( \frac{\xi c\omega}{r} + \frac{2kU_0}{R} - I_s \right) U_{in} + \frac{U_0^2}{R} = 0. \quad (1.12)$$

Уравнение (1.12) имеет два решения:

$$\hat{U}_{1,2} = \frac{\frac{\xi c\omega}{r} + \frac{2kU_0}{R} - I_s \pm \sqrt{\left( \frac{\xi c\omega}{r} + \frac{2kU_0}{R} - I_s \right)^2 - 4 \left( \frac{\xi}{r} + \frac{k^2}{R} \right) \frac{U_0^2}{R}}}{2 \left( \frac{\xi}{r} + \frac{k^2}{R} \right)}. \quad (1.13)$$

Величины  $\hat{U}_{1,2}$  действительны, если значение  $\omega$  достаточно велико.

Теперь необходимо выяснить, какое из решений (1.13) может быть больше  $U_{in}^*$ , и при каких значениях  $\omega$ . Как было отмечено выше, когда напряжение на входе стабилизатора проходит значение  $U_{in}^*$ , величина  $U_{out}$  скачком увеличивается от нуля до некоторого положительного значения. Поэтому

$$kU_{in}^* - U_0 > 0. \quad (1.14)$$

С учетом неравенства (1.14) нетрудно показать, что соотношение  $\hat{U}_2 = U_{in}^*$ , рассматриваемое как уравнение относительно  $\omega$ , не имеет решений. В то же время, уравнение  $\hat{U}_1 = U_{in}^*$  имеет единственное решение:

$$\omega^* = \frac{U_{in}^*}{c} + \frac{rI_s}{c\xi} + \frac{r(kU_{in}^* - U_0)^2}{Rc\xi U_{in}^*}. \quad (1.15)$$

Из (1.13) и (1.15) вытекает, что

$$\frac{d\hat{U}_1}{d\omega} > 0 \text{ при } \omega > \omega^*.$$

Следовательно, с учетом того обстоятельства, что  $\hat{U}_1 > \hat{U}_2$ , получаем, что из (1.13) следует выбрать ветвь  $\hat{U}_1$ .

Таким образом, изложенное выше означает, что «переходный» режим может реализоваться при

$$\omega^* \leq \omega \leq \omega^{**},$$

где  $\omega^{**}$  – значение угловой скорости, при которой  $\hat{U}_1 = U_{in}^{**}$ :

$$\omega = \omega^{**} = \frac{U_{in}^{**}}{c} + \frac{rI_s}{c\xi} + \frac{r(kU_{in}^{**} - U_0)^2}{Rc\xi U_{in}^{**}} > 0. \quad (1.16)$$

Из (1.11) и (1.15) видно, что  $\omega_s < \omega^*$ . Это значит, что на интервале  $\omega_s < \omega < \omega^*$  стационарный режим отсутствует. Можно ожидать, что при угловых скоростях из этого интервала стабилизатор будет постоянно «переключаться» между режимом отключения и переходным режимом (т.е. имеет место режим типа скользящего). Более детальное описание этой ситуации требует тщательного моделирования динамики стабилизатора, что выходит за рамки настоящей работы.

1.3. И н т е р в а л  $U_{in}^{**} < U_{in}$  (н о м и н а л ь н ы й р е ж и м). Здесь уравнение (1.7) принимает следующий вид:

$$\xi U_{in} \frac{c\omega - U_{in}}{r} = \frac{U_n^2}{R} + U_{in} I_s.$$

Отсюда получаем:

$$\frac{\xi}{r} U_{in}^2 - \left( \frac{\xi c\omega}{r} - I_s \right) U_{in} + \frac{U_n^2}{R} = 0. \quad (1.17)$$

Если угловая скорость ротора достаточно велика, так что  $\omega > \omega_0$ , где

$$\omega_0 = 2 \frac{U_n}{c} \sqrt{\frac{r}{\xi R}} + \frac{I_s r}{c \xi}, \quad (1.18)$$

то у уравнения (1.17) существует две ветви решений:

$$\tilde{U}_{1,2} = \frac{\xi c\omega - rI_s \pm \sqrt{(\xi c\omega - rI_s)^2 - 4\xi U_n^2 \frac{r}{R}}}{2\xi}. \quad (1.19)$$

Нетрудно видеть, что  $\lim_{r \rightarrow 0} \tilde{U}_1 = c\omega$  и  $\lim_{r \rightarrow 0} \tilde{U}_2 = 0$ . Однако из (1.8) следует, что когда внутреннее сопротивление генератора стремится к нулю, напряжение на его выходе (которое в нашем случае равно напряжению на входе стабилизатора) стремится к  $c\omega$ . Поэтому нужно выбрать в качестве решения первую ветвь (1.19). Из (1.19) вытекает, что функция  $\tilde{U}_1$  монотонно растет с ростом  $\omega$ . Обозначим ее минимальное значение через  $U_*$ . Оно, очевидно, равно  $\tilde{U}_1(\omega_0)$ . Таким образом, номинальный режим может реализоваться только при  $U_{in} > U_*$ .

Пусть

$$U_{in}^{**} = \frac{U_n + U_0}{k} > U_* = U_n \sqrt{\frac{r}{\xi R}}. \quad (1.20)$$

Нетрудно показать, что тогда кривая  $\hat{U}_1$  пересекается с кривой  $\tilde{U}_1$ . В этом случае при изменении  $\omega$  в интервале от  $\omega^*$  до бесконечности входное напряжение изменяется непрерывно. Если же выполнено неравенство, противоположное (1.20), то  $\hat{U}_1$  не пересекается с кривой  $\tilde{U}_{in}$ . Соответственно входное напряжение при изменении  $\omega$  претерпевает разрыв: если угловая скорость растет, то  $U_{in}$  скачком возрастает при  $\omega = \omega^{**}$ , переходя с кривой  $\hat{U}_1$  на кривую  $\tilde{U}_1$ ; если же угловая скорость убывает, то  $U_{in}$  скачком убывает при  $\omega = \omega_0$ , переходя с кривой  $\tilde{U}_1$  на кривую  $\hat{U}_1$ . Таким образом, в интервале  $(\omega_0, \omega^{**})$  имеет место гистерезис входного напряжения (и, разумеется, входного тока).

Уравнение (1.5), описывающее динамику рассматриваемой системы, удобнее исследовать в терминах входного тока. Зависимость  $I_{in}$  от угловой скорости получается из формулы (1.8) с учетом соотношений (1.13) и (1.19). Нетрудно показать, что из (1.13) следует, что при  $\omega > \omega^*$  имеет место неравенство  $d^2 I_{in} / d\omega^2 > 0$ , причем

$$\left. \frac{dI_{in}}{d\omega} \right|_{\omega=\omega^*} = c \frac{\left( kU_{in}^* \right)^2 - U_0^2}{\left( \left( kU_{in}^* \right)^2 - U_0^2 \right) r + \left( kU_{in}^* \right)^2 R \xi}.$$

В силу (1.14), это выражение больше нуля. Таким образом, на интервале  $U_{in}^* < U_{in} < U_{in}^{**}$  входной ток монотонно растет с ростом угловой скорости. На номинальном режиме имеем

$$\begin{aligned} \frac{dI_{in}}{d\omega} &= c \frac{-R(c\omega\xi - I_s r) + \sqrt{R^2(c\omega\xi - I_s r)^2 - 4U_n^2 r R \xi}}{2r \sqrt{R^2(c\omega\xi - I_s r)^2 - 4U_n^2 r R \xi}}, \\ \frac{d^2 I_{in}}{d\omega^2} &= \frac{2c^2 \xi^2 R U_n^2}{\sqrt{R} \left( (c\omega\xi - I_s r)^2 R - 4U_n^2 r \xi \right)^{3/2}}. \end{aligned} \quad (1.21)$$

Поскольку при выполнении (1.18) первое из выражений (1.21) отрицательно,  $I_{in}$  монотонно убывает с ростом  $\omega$ , стремясь к величине  $I_s/\xi$ . Второе выражение (1.21) положительно, так что кривая  $I_{in}(\omega)$  имеет на рассматриваемом интервале положительную кривизну.

Рассмотренные ситуации качественно проиллюстрированы на рис. 3, *a*, *б*, на котором изображены зависимости входного напряжения и тока от угловой скорости.

Жирными черными линиями отмечены ветви, отвечающие номинальным режимам, серыми линиями – ветви, отвечающие переходным режимам, и тонкими черными линиями – ветви, отвечающие ситуации, когда стабилизатор отключен.

Отметим, что предложенная математическая модель содержит ряд параметров –  $c$ ,  $r$ ,  $\xi$ ,  $I_s$ ,  $U_{in}^*$ ,  $k$  и  $U_0$ . Они подлежат идентификации с помощью экспериментальных данных.

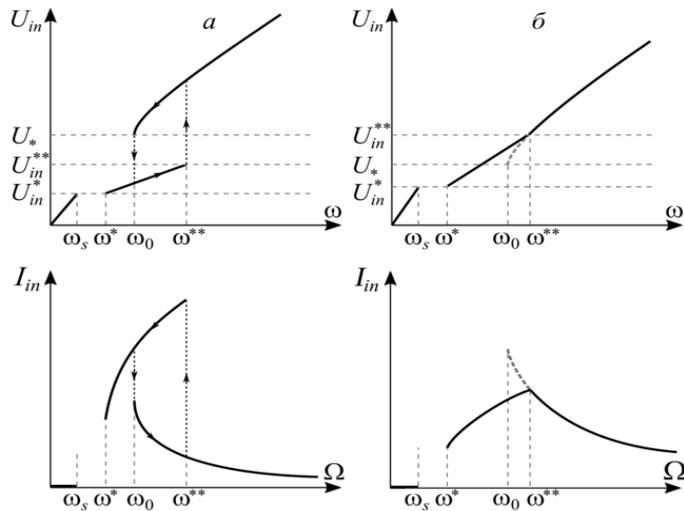


Рис. 3. Зависимости входного напряжения и тока от угловой скорости: а –  $U_{in}^{**} < U_*$ ; б –  $U_{in}^{**} > U_*$

**2. Установившиеся режимы.** Переидем к исследованию установившихся режимов, которые могут возникать в рассматриваемой электромеханической системе. Отметим, что с точки зрения практической эксплуатации желательными являются только те из них, на которых стабилизатор находится в номинальном режиме (т.е. выдает напряжение  $U_n$ ).

Прежде всего, приведем уравнение (1.5) к безразмерному виду. Введем безразмерное время  $\tau = Vt/b$  и следующие безразмерные величины:

$$\begin{aligned} \iota_{in} &= \frac{rI_{in}}{U_n}, \quad \bar{J} = \frac{2J}{\rho S b^3}, \quad \zeta = \frac{2cU_n}{\rho S b r V^2}, \quad u_{in} = \frac{U_{in}}{U_n}, \\ u_0 &= \frac{U_0}{U_n}, \quad u_* = \frac{U_*}{U_n} = 2\sqrt{\frac{r}{\xi R}}, \quad u_{in}^{**} = \frac{U_{in}^{**}}{U_n} = \frac{1+u_0}{k}, \\ \bar{R} &= \frac{R}{r}, \quad \iota_s = \frac{I_s r}{U_n}, \quad \kappa = \frac{cV}{bU_n}. \end{aligned} \quad (2.1)$$

С учетом (2.1) уравнение динамики системы примет следующий вид:

$$\bar{J}\Omega' = C_m(\Omega) - \zeta \iota_{in}(\Omega) \quad (2.2)$$

На интервале  $u_{in}^* < u_{in} < u_{in}^{**}$  имеем:

$$\iota_{in} = \iota_1 = \frac{\bar{R}(\xi\Omega\kappa + \iota_s) + 2\Omega k^2\kappa - 2ku_0 - \sqrt{(\bar{R}\xi\Omega\kappa + 2ku_0 - \iota_s\bar{R})^2 - 4u_0^2(\bar{R}\xi + k^2)}}{2(\bar{R}\xi + k^2)}. \quad (2.3)$$

На интервале  $u_{in}^{**} < u_{in}$  (номинальный режим) имеем:

$$\iota_{in} = \iota_2 = \frac{\xi\Omega\kappa + \iota_s - \sqrt{(\xi\Omega\kappa - \iota_s)^2 - 4\xi\bar{R}^{-1}}}{2\xi}. \quad (2.4)$$

Нетрудно показать, что

$$\frac{d\iota_1}{d\bar{R}} < 0, \quad \frac{d\iota_2}{d\bar{R}} < 0, \quad \lim_{\bar{R} \rightarrow \infty} \iota_{1,2} = \frac{\iota_s}{\xi}.$$

Отметим, что из формул (1.15), (1.16) и (1.18) следует, что величины  $\Omega^* = b\omega^*/V$ ,  $\Omega_0 = b\omega_0/V$  и  $\Omega^{**} = b\omega^{**}/V$  монотонно убывают с ростом  $\bar{R}$ . Кроме того, при достаточно малых значениях нагрузочного сопротивления, таких, что

$$\bar{R} < \bar{R}_* = \frac{k^2}{\xi(1+u_0)^2}.$$

неравенство (1.20) не выполняется и имеет место гистерезис, описанный выше; если же  $\bar{R} > \bar{R}_*$ , то условие (1.20) удовлетворяется и гистерезис для стабилизатора отсутствует.

Многочисленные эксперименты, проведенные с роторами Савониуса, показывают, что функция момента удовлетворяет следующим условиям:

$$\begin{aligned} C_m(0) &> 0; \quad \exists! \Omega_m > 0 : C_m(\Omega_m) = 0, \\ \exists! \Omega_M &\in (0, \Omega_m) : \left. \frac{dC_m(\Omega)}{d\Omega} \right|_{\Omega=\Omega_M} = 0; \quad \forall \Omega \in (0, \Omega_m) : C_m''(\Omega) < 0. \end{aligned} \quad (2.5)$$

Стационарные режимы системы (2.2) находятся из уравнения

$$C_m(\Omega) = \zeta \iota_{in}(\Omega). \quad (2.6)$$

Здесь безразмерный ток на входе стабилизатора определяется формулами (2.3)–(2.4).

Достаточное условие асимптотической устойчивости стационарных режимов, очевидно, выглядит так:

$$\frac{dC_m}{d\Omega} - \zeta \frac{d\iota_{in}}{d\Omega} < 0. \quad (2.7)$$

Здесь производные, разумеется, вычисляются для  $\Omega$ , отвечающего стационарному режиму. Заметим, что из (1.15), (1.16) и (2.5) следует, что если

$$\frac{u_{in}^*}{\kappa} + \frac{\iota_s}{\kappa\xi} \geq \Omega_m, \quad (2.8)$$

то при всех  $\bar{R}$  выполняются неравенства  $\Omega^* \geq \Omega_m$ ,  $\Omega^{**} \geq \Omega_m$  и уравнение (2.6) не имеет решений, отличных от  $\Omega = \Omega_m$ ,  $\iota_{in} = 0$ . Такая ситуация возникает при слишком низкой скорости потока. В дальнейшем будем предполагать, что неравенство (2.8) не выполняется. Тогда существует такое значение  $\bar{R}_{**}$ , что при  $\bar{R} > \bar{R}_{**}$  имеет место условие  $\Omega^{**} < \Omega_m$  (заметим, что из (1.16), (2.1) и (2.5) следует  $\lim_{V \rightarrow \infty} \bar{R}_{**} = 0$ ).

Графически стационарные режимы (2.6) представляют собой точки пересечения кривых  $C_m(\Omega)\zeta^{-1}$  и  $\iota_{in}(\Omega)$ . В зависимости от значений параметров могут реализовываться различные варианты взаимного расположения этих кривых. Предположим, что выполнены следующие условия:  $\bar{R}_{**} < \bar{R}_*$ ,  $\exists \bar{R}_l \in (\bar{R}_{**}, \bar{R}_*) : \min(\iota_1(\Omega^*), \iota_2(\Omega_m)) > \zeta^{-1}C_m(\Omega_M) > \iota_s/\xi$ . Заметим, что обеспечить выполнение второго неравенства можно за счет надлежащего выбора параметра  $\zeta$ . Тогда при  $\bar{R}_l$  уравнение (2.6) не имеет решений. В то же время при достаточно большом нагрузочном сопротивлении решение существует. Четыре варианта взаимного расположе-

ния рассматриваемых кривых для разных значений нагрузочного сопротивления ( $\bar{R}_1$ ,  $\bar{R}_2$ ,  $\bar{R}_3$  и  $\bar{R}_4$ , причем  $\bar{R}_{**} < \bar{R}_1 < \bar{R}_2 < \bar{R}_3 < \bar{R}_* < \bar{R}_4$ ) качественно изображены для этой ситуации на рис. 4. Жирными черными линиями отмечены ветви, на которых стабилизатор находится в номинальном режиме, а серыми – ветви, на которых он находится в переходном режиме.

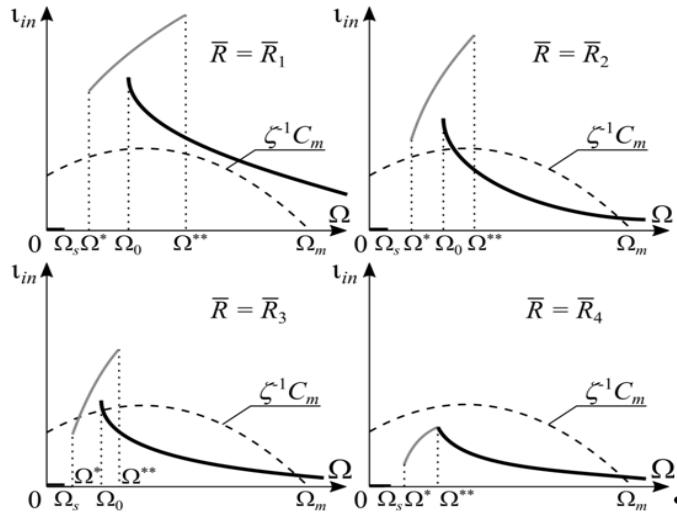


Рис. 4. Зависимости  $C_m$  и  $i_{in}$  от быстроходности при разных значениях нагрузочного сопротивления

В первом случае ( $\bar{R} = \bar{R}_1$ ) стационарных режимов нет. Если начальная быстроходность больше  $\Omega^*$ , то  $\Omega$  будет убывать, пока она не достигнет величины  $\Omega^*$ . При дальнейшем уменьшении  $\Omega$  стабилизатор перейдет в режим типа скользящего, на котором он будет отключаться и снова включаться. Если начальная быстроходность меньше  $\Omega_s$ , то  $\Omega$  будет расти, а после того, как она достигнет  $\Omega_s$ , стабилизатор также перейдет в режим типа скользящего. Можно ожидать, что в результате в системе сформируется некоторый «особый» режим, на котором быстроходность турбины остается на интервале  $\Omega \in (\Omega_s, \Omega^*)$ , а стабилизатор постоянно включается/отключается. При этом в среднем достигается баланс между мощностью, отбираемой у потока, и мощностью, расходуемой в электрической цепи.

Во втором случае ( $\bar{R} = \bar{R}_2$ ) существует два стационарных режима. Один из них реализуется при «большой» быстроходности и отвечает номинальному режиму работы стабилизатора. Как видно из (2.7), он асимптотически устойчив. Второй режим существует в зоне гистерезиса входного тока и является неустойчивым. И, как и в предыдущем случае, можно ожидать наличия «особого режима», на котором быстроходность остается в интервале  $(\Omega_s, \Omega^*)$ .

В третьем случае ( $\bar{R} = \bar{R}_3$ ) существует три стационарных режима. Один из них «высокоскоростной», как и в предыдущем случае. Второй попадает в зону «переходного» режима работы стабилизатора. Оба этих стационарных режима, очевидно, асимптотически устойчивы. Третий стационарный режим находится в зоне гистерезиса и является отталкивающим.

Наконец, в четвертом случае ( $\bar{R} = \bar{R}_4$ ) в системе имеется только один «высокоскоростной» стационарный режим, и он является асимптотически устойчивым. При этом стабилизатор выдает номинальное напряжение.

Учитывая изложенное выше, бифуркационная диаграмма стационарных режимов ветроэнергетической установки качественно выглядит так, как изображено на рис. 5. Сплошной черной линией показаны устойчивые режимы, соответствующие номинальному режиму работы стабилизатора, сплошной серой линией – устойчивые режимы, на которых стабилизатор находится в переходном режиме. Жирной пунктирной линией изображены неустойчивые режимы. Область, в которой реализуются «особые» режимы, выделена серым цветом, тонкие пунктирные линии – зависимости  $\Omega_s$  и  $\Omega^*$  от сопротивления.

Таким образом, при изменении нагрузочного сопротивления в системе имеет место гистерезис установившихся режимов функционирования установки (который нужно отличать от гистерезиса стационарных режимов стабилизатора напряжения, описанного выше). Следует отметить, что, хотя динамика системы описывается дифференциальным уравнением первого

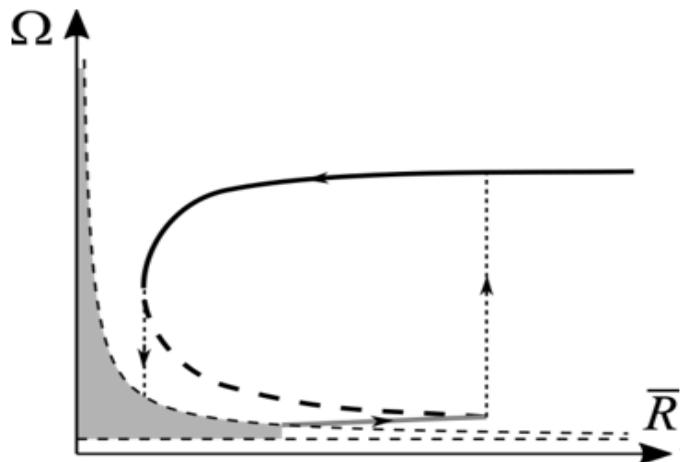


Рис. 5. Бифуркационная диаграмма стационарных режимов

порядка, задать одно начальное условие на  $\Omega$  может оказаться недостаточным для определения последующего поведения системы. Необходимо определить, какой ветви входного тока отвечает состояние стабилизатора. Для решения этой задачи, по-видимому, нужно более детально описать динамику стабилизатора, увеличив порядок динамической системы. Это остается предметом последующего исследования.

Прежде чем переходить к результатам экспериментов, обсудим коротко ситуацию, когда стабилизатор выключен из цепи и ток от генератора идет напрямую на нагрузочное сопротивление. Нетрудно видеть, что при этом из закона Ома для контура получаем

$$i_{out} = \frac{\kappa\Omega}{\bar{R} + 1}. \quad (2.9)$$

С учетом (2.9) уравнение (2.6) стационарных режимов принимает такой вид:

$$C_m(\Omega) = \frac{\zeta\kappa\Omega}{\bar{R} + 1}. \quad (2.10)$$

Из соотношений (2.5) следует, что уравнение (2.10) имеет единственное решение. Это означает, в частности, что в случае, когда нагрузкой в ветроэнергетической установке с рабочим элементом в виде ротора Савониуса является просто активное сопротивление, гистерезис стационарных режимов отсутствует.

**3. Экспериментальная верификация и идентификация параметров.** С целью верификации сформированной математической модели системы и идентификации параметров было проведено экспериментальное исследование. В его рамках были изготовлены две лабораторных установки. Одна из них состоит из электродвигателя, вал которого жестко соединен с ротором генератора (генератор 1). В цепь ротора генератора включен стабилизатор напряжения на 5В постоянного тока и нагрузочное сопротивление. Вторая установка состоит из вертикально-осевой ветротурбины (ротор Савониуса), ось которой жестко соединена с ротором электрогенератора (генератор 2). В цепь ротора генератора также включен стабилизатор напряжения на 5В постоянного тока и нагрузочное сопротивление. Установка помещена в дозвуковую аэrodинамическую трубу А10 НИИ механики МГУ.

В обеих установках имеются устройства для измерения тока и напряжения на входе стабилизатора, а также тока на выходе стабилизатора. В электрических цепях обеих установок предусмотрен ключ, позволяющий отключить стабилизатор (так, чтобы ток от генератора шел напрямую на нагрузочное сопротивление). Нагрузочное сопротивление задается с помощью магазина сопротивлений. Первая установка предназначена для исследования характеристик стабилизатора. Вторая – для изучения динамики ветроэнергетической установки.

**3.1. Идентификация параметров стабилизатора и генератора 1.** Вначале были проведены эксперименты, направленные на определение характеристик генератора 1. Для этого вал генератора с помощью двигателя приводился во вращение с некоторой угловой скоростью. Стабилизатор был отключен с помощью упомянутого выше ключа. При различных значениях нагрузочного сопротивления измерялись угловая скорость и ток на нагрузочном сопротивлении.

В рамках модели (1.4) уравнение Кирхгофа для электрической цепи без стабилизатора выглядит так:

$$c\omega = (R + r)I, \quad (3.1)$$

где  $I$  – ток, текущий через нагрузочное сопротивление. Отсюда имеем:

$$\frac{\omega}{I} = \frac{R}{c} + \frac{r}{c}. \quad (3.2)$$

Результаты экспериментов представлены на рис. 6 точками. Сплошной линией изображена зависимость (3.2) при значениях параметров, определенных с помощью метода наименьших квадратов:  $c = 1.3$  В·с и  $r = 65$  Ом. Видно, что расчет достаточно хорошо согласуется с экспериментальными данными.

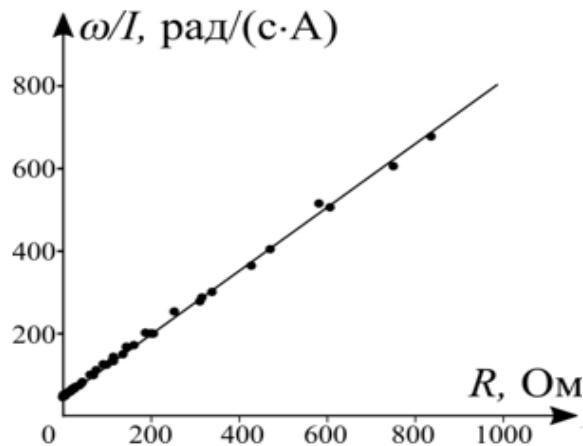


Рис. 6. Зависимость  $\omega/I$  от нагрузочного сопротивления для генератора 1:  
точки – экспериментальные данные, прямая – аппроксимация (3.2)

Для идентификации параметров стабилизатора была проведена серия экспериментов при включенном стабилизаторе. В их ходе при различных значениях нагрузочного сопротивления и угловой скорости ротора генератора измерялся ток и напряжение на входе стабилизатора, а также напряжение на нагрузочном сопротивлении.

На рис. 7 точками представлена экспериментальная зависимость  $U_{out}$  от  $U_{in}$ , пунктирной линией – зависимость (1.9) при  $k = 1$ ,  $U_0 = 1.8$  В и  $U_{in}^* = 4.7$  В. Видно, что предложенная аппроксимационная кривая достаточно хорошо приближает экспериментальные данные.

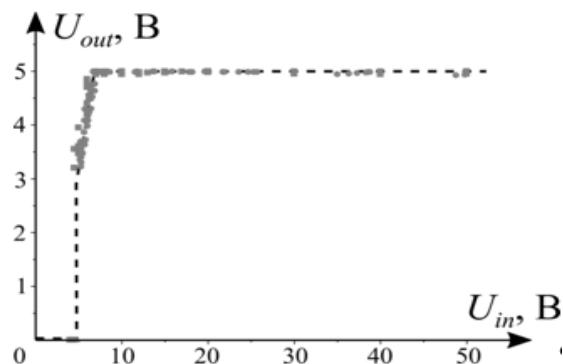


Рис. 7. Зависимости напряжения на выходе стабилизатора от напряжения на входе

Зависимость входного тока от угловой скорости вала генератора при разных значениях нагрузочного сопротивления изображена на рис. 8 точками (эксперимент) и сплошными линиями (расчет в рамках модели). Из рисунка видно, что расчет достаточно хорошо согласуется с результатами измерений. Следует отметить, что, как и предсказывает модель, при сравнительно небольших значениях  $R$  в экспериментах наблюдается гистерезис, который пропадает с ростом  $R$  (границы зон гистерезиса отмечены пунктирными линиями).

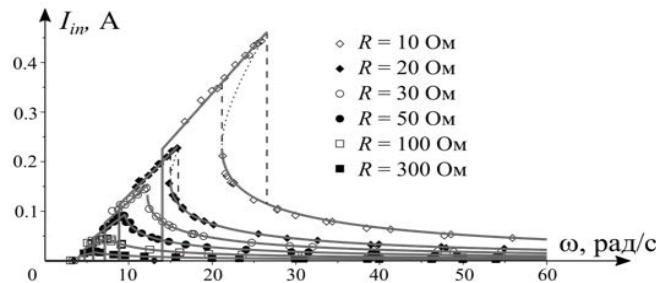


Рис. 8. Зависимости тока на входе стабилизатора от угловой скорости вала генератора при разных значениях нагрузочного сопротивления

Таким образом, предложенная модель достаточно хорошо согласуется с экспериментальными данными и может использоваться для исследования динамики электромеханических систем, в состав которых входит стабилизатор напряжения. Необходимо отметить, что при достаточно низких значениях угловой скорости в эксперименте наблюдались значительные колебания тока (в некоторых случаях от 0 до 70 мА) и напряжения на входе стабилизатора. Напряжение на выходе стабилизатора при этом было в среднем менее 3 В. Это соответствует «скользящему» режиму, описанному в предыдущем разделе. При дальнейшем уменьшении  $\omega$  входной ток и выходное напряжение становились равными нулю. Это, по-видимому, означает, что стабилизатор размыкал цепь.

3.2. Эксперименты с ветротурбиной. После того, как параметры стабилизатора были идентифицированы с помощью первой установки, стабилизатор был перенесен во вторую установку, с которой затем были проведены эксперименты в дозвуковой аэродинамической трубе А10 НИИ механики МГУ.

Эта труба может использоваться с закрытой или с открытой рабочей частью (внутри камеры Эйфеля). Сечение трубы на входе в рабочую часть имеет форму правильного восьмиугольника со стороной 33 см. Длина рабочей части — 1.35 м. Диапазон скоростей потока — от 5 до 50 м/с (с открытой рабочей частью) и до 70 м/с (с закрытой рабочей частью). В описываемых экспериментах рабочая часть была открыта.

Лопасти ветротурбины представляли собой пластиковые полуцилиндры диаметром 130 мм и высотой 250 мм, толщина лопастей — 2 мм. Сверху и снизу лопасти были прикреплены к пластиковым дискам толщиной 6 мм. Диски вместе с лопастями были закреплены на металлическом валу диаметром 14 мм. Лопасти были установлены с перекрытием, причем между краем лопасти и валом был оставлен зазор 15 мм. Таким образом, итоговый радиус ротора (т.е. расстояние от центра вала до дальнего края лопасти) составлял 108 мм. Вал был жестко соединен с валом ротора генератора 2. Турбина была закреплена в раме и установлена в рабочей части аэродинамической трубы (см. рис. 9).

Вначале для определения характеристик генератора 2 была проведена серия экспериментов с отключенным стабилизатором. В ходе испытаний при разных скоростях потока и разных значениях нагрузочного сопротивления измерялась угловая скорость ветротурбины и ток на сопротивлении. На рис. 10 представлена экспериментальная зависимость отношения  $\omega/I_{in}$  от нагрузочного сопротивления (точки) и зависимость (3.2) при  $c = 1.5$  В·с и  $r = 107$  Ом (величины параметров определены с помощью метода наименьших квадратов).

Из рисунка видно, что в рассматриваемом диапазоне угловых скоростей и величин сопротивления модель линейного электромеханического взаимодействия достаточно хорошо согласуется с экспериментальными данными.

На рис. 11 представлены изображены зависимость величины быстродействия на стационарных режимах от безразмерного нагрузочного сопротивления (отнесенного к идентифици-

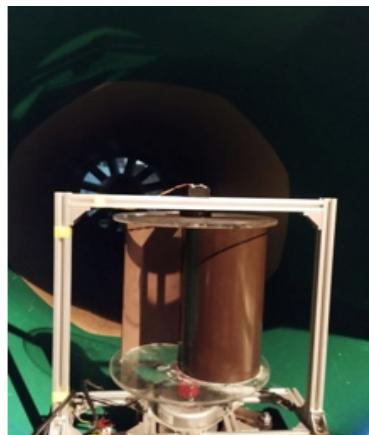


Рис. 9. Лабораторная установка с ротором Савониуса в аэродинамической трубе

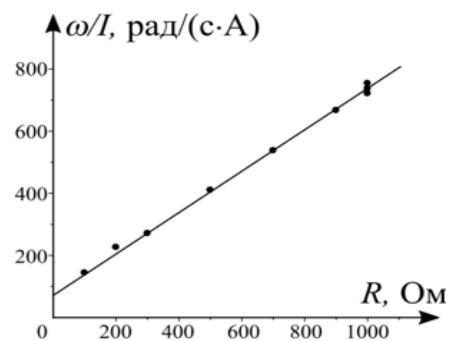


Рис. 10. Зависимость  $\omega/I$  от нагрузочного сопротивления для генератора 2: точки – экспериментальные данные, прямая – аппроксимация (3.2)

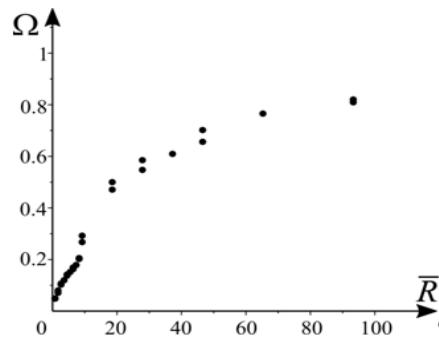


Рис. 11. Зависимость быстроходности на стационарных режимах от безразмерного нагрузочного сопротивления (стабилизатор отключен)

рованному внутреннему сопротивлению генератора 2). Отметим, что каждому значению  $R$  соответствует один стационарный режим и гистерезис отсутствует.

Затем была проведена серия экспериментов с подключенным стабилизатором. Скорость потока составляла 8 м/с. В ходе экспериментов фиксировалось нагрузочное сопротивление и измерялись ток и напряжение на входе стабилизатора, напряжение на выходе стабилизатора и угловая скорость турбины.

На рис. 12 представлены значения безразмерной угловой скорости  $\Omega$  ротора Савониуса на стационарных и «особых» режимах (естественно, в экспериментах наблюдалась только притягивающие режимы) в зависимости от безразмерного нагрузочного сопротивления  $R$ . Стационарные режимы обозначены черными точками, «особые» – серыми. Стационарные

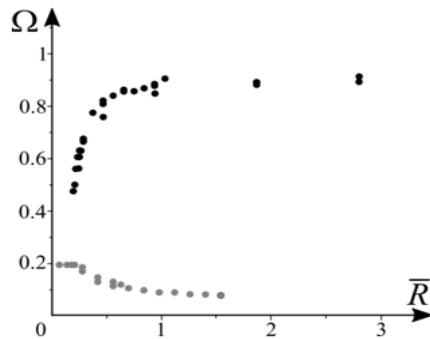


Рис. 12. Зависимость быстроходности на стационарных режимах от безразмерного нагруженного сопротивления (стабилизатор включен)

режимы, соответствующие переходному режиму работы стабилизатора, не были обнаружены. Это, по-видимому, связано с недостаточно мелким шагом по  $\bar{R}$ .

Видно, что эта диаграмма качественно соответствует результатам, полученным в рамках предложенной модели (см. рис. 6). В частности, в эксперименте наблюдается эффект гистерезиса режимов функционирования установки, предсказанный моделью.

**4. О возможностях регулирования работы установки с помощью нагрузочного сопротивления.** Основной характеристикой работы ветроэнергетической установки, предназначенной для выработки электроэнергии, является выходная мощность. В нормальном режиме функционирования установки, когда стабилизатор находится в номинальном режиме, она определяется следующей формулой:

$$P_{out} = \frac{U_n^2}{R}. \quad (4.1)$$

Ясно, что для максимизации этой величины необходимо уменьшать сопротивление. В то же время при маленьких сопротивлениях, как было показано выше, единственным притягивающим режимом функционирования установки является «особый» режим, на котором мощность практически равна нулю. Соответственно сопротивление не должно быть меньше некоторого критического значения. Это значение, очевидно, зависит от скорости потока. Поэтому, для того чтобы система могла перейти на номинальный режим в случае, если в результате падения скорости потока она окажется на «особом» режиме, необходимо регулировать нагрузку.

Учитывая, что при  $\Omega_M < \Omega < \Omega_m$  имеем  $C_m'' < 0$ , и  $d^2\iota_{in2}/d\Omega^2 > 0$ , стационарные решения, на которых стабилизатор находится в номинальном режиме, существуют только при  $\bar{R} > \bar{R}_m$ , причем величина  $\bar{R}_m$  определяется из следующей системы алгебраических уравнений:

$$C_m(\Omega) = \zeta \frac{\xi \Omega \kappa + \iota_s - \sqrt{(\xi \Omega \kappa - \iota_s)^2 - 4\xi \bar{R}_m^{-1}}}{2\xi},$$

$$\frac{dC_m}{d\Omega} = \zeta \frac{\kappa}{2} - \zeta \frac{\kappa(\xi \Omega \kappa - \iota_s)}{2\sqrt{(\xi \Omega \kappa - \iota_s)^2 - 4\xi \bar{R}_m^{-1}}}.$$

Эту систему следует рассматривать как уравнения относительно  $\bar{R}_m$  и  $\Omega$ . Соответственно максимальная достижимая мощность равна

$$P_{out}^{\max} = \frac{U_n^2}{r\bar{R}_m}. \quad (4.2)$$

Кроме того, если скорость потока не слишком мала, то существует такое сопротивление  $\bar{R}_M$ , что если  $R > \bar{R}_M$ , то существует только стационарное решение, на котором стабилизатор находится в номинальном режиме. С учетом соотношений (2.6) и того обстоятельства, что  $d\iota_{in1}/d\Omega > 0$ , оценить величину  $\bar{R}_M$  можно следующим образом:  $\bar{R}_M \leq \bar{R}_M^{est}$ , где  $\bar{R}_M^{est}$  определяется из следующего уравнения:

$$C = \zeta \frac{\bar{R}_M^{est} (\xi \Omega^{**} \kappa + \iota_s) + 2\Omega^{**} k^2 \kappa - 2ku_0 - \sqrt{(\bar{R}_M^{est} \xi \Omega^{**} \kappa + 2ku_0 - \iota_s \bar{R}_M)^2 - 4u_0^2 (\bar{R}_M^{est} \xi + k^2)}}{2(\bar{R}_M^{est} \xi + k^2)},$$

где

$$C = \begin{cases} C_m(\Omega^{**}), & \Omega_M = 0, \\ \min_{\Omega \in [0, \Omega_M]} C_m(\Omega), & \Omega_M > 0. \end{cases}$$

Очевидно,  $\bar{R}_M > \bar{R}_m$ .

Таким образом, если установка переходит в нежелательный режим (на котором стабилизатор не находится вnomинальном режиме), то следует увеличить нагрузочное сопротивление, чтобы оно стало больше, чем  $R_M = r\bar{R}_M$ . После того, как система перейдет на стационарный режим, нужно уменьшить сопротивление до величины  $R_m = r\bar{R}_m$  (на практике целесообразно выбирать значение, несколько большее  $R_m$ ). Если требуется затормозить вращение ветротурбины (например, когда скорость ветра становится слишком большой), то следует уменьшить нагрузочное сопротивление до величины, меньшей  $R_m$ .

**Заключение.** Проведено исследование динамики ветроэнергетической установки, рабочим элементом которой является ротор Савониуса, а в цепь генератора включен стабилизатор напряжения. Предложена математическая модель, описывающая поведение стабилизатора в зависимости от входного напряжения и подключенной к нему нагрузки. Исследованы стационарные режимы, существующие в этой системе при разных значениях параметров. Установлено, что при изменении нагрузочного сопротивления в системе возникает гистерезис. Проведена серия экспериментов по идентификации параметров электрической части системы, а также испытания ветроэнергетической установки в дозвуковой аэродинамической трубе. Показано, что данные, полученные в ходе экспериментов, достаточно хорошо согласуются с результатами аналитического исследования. Предложен алгоритм регулирования нагрузочного сопротивления, позволяющий обеспечить максимальную выходную мощность.

#### СПИСОК ЛИТЕРАТУРЫ

1. Wang H., Xiong B., Zhang Z., Zhang H., Azam A. Small Wind Turbines and Their Potential for Internet of Things Applications // iScience. 2023. V. 26. № 9. P. 107674. <https://doi.org/10.1016/j.isci.2023.107674>.
2. Al-Kayiem H.H., Bhayo B.A., Assadi M. Comparative Critique on the Design Parameters and Their Effect on the Performance of S-Rotors // Renewable Energy. 2016. V. 99. P. 1306–1317. <https://doi.org/10.1016/j.renene.2016.07.015>.
3. Cuevas-Carvajal N., Cortes-Ramirez J.S., Norato J.A., Hernandez C., Montoya-Vallejo M.F. Effect of Geometrical Parameters on the Performance of Conventional Savonius VAWT: A Review // Renewable & Sustainable Energy Reviews. 2022. V. 161. № 3. P. 112314.
4. Sheldahl R.E., Blackwell B.F., Feltz L.V. Wind Tunnel Performance Data for Two- and Three-Bucket Savonius Rotors // J. Energy. 1978. V. 2. № 3. P. 160–164.
5. Saha U.K., Thotla S., Maity D. Optimum Design Configuration of Savonius Rotor Through Wind Tunnel Experiments // J. Wind Eng. Ind. Aerodyn. 2008. V. 96. P. 1359–1375. <https://doi.org/10.1016/j.jweia.2008.03.005>.
6. Fujisawa N. On the Torque Mechanism of Savonius Rotors // J. Wind Eng. & Ind. Aerodyn. 1992. V. 40. P. 277–292.
7. Jian C., Kumbernuss J., Linhua Z., Lin L., Hongxing Y. Influence of Phase-Shift and Overlap Ratio on Savonius Wind Turbine's Performance // ASME. J. Sol. Energy Eng. 2012. V. 134. № 1. P. 011016. <https://doi.org/10.1115/1.4004980>.
8. Nasef M.H., El-Askary W.A., Abdel-Hamid A.A., Gad H.E. Evaluation of Savonius Rotor Performance: Static and Dynamic Studies // J. Wind Eng. & Ind. Aerodyn. 2013. V. 123. P. 1–11. <https://doi.org/10.1016/j.jweia.2013.09.009>.
9. Bach G. Untersuchungen über Savonius-Rotoren und Verwandte Strömungsmaschinen // Forschung auf dem Gebiete des Ingenieurwesens. 1931. Bd. 2. S. 218–231.
10. Kamoji M.A., Kedare S.B., Prabhu S.V. Experimental Investigations on single Stage, Two Stage and Three Stage Conventional Savonius Rotor // Intern. J. Energy Res. 2008. V. 32. P. 877–895. <https://doi.org/10.1002/er.1399>.
11. Frikha S., Driss Z., Ayadi E., Masmoudi Z., Abid M.S. Numerical and Experimental Characterization of Multi-Stage Savonius Rotors // Energy. 2016. V. 114. P. 382–404. <https://doi.org/10.1016/j.energy.2016.08.017>.

12. *Kacprzak K., Liskiewicz G., Sobczak K.* Numerical Investigation of Conventional and Modified Savonius Wind Turbines // *Renewable Energy*. 2013. V. 60. P. 578–585.  
<https://doi.org/10.1016/j.renene.2013.06.009>.
13. *Roy S., Saha U.K.* Wind Tunnel Experiments of a Newly Developed Two Bladed Savonius-Style Wind Turbine // *Applied Energy*. 2015. V. 137. P. 117–125.  
<https://doi.org/10.1016/j.apenergy.2014.10.022>.
14. *Roy S., Ducoin A.* Unsteady Analysis on the Instantaneous Forces and Moment Arms Acting on a Novel Savonius-Style Wind Turbine // *Energy Conversion & Management*. 2016. V. 121. P. 281–296.  
<https://doi.org/10.1016/j.enconman.2016.05.044>.
15. *Scheaua F.D.* Comparative Numerical Analysis on Vertical Wind Turbine Rotor Pattern of Bach and Benesh Type // *Energies*. 2020. V. 13. P. 2311.  
<https://doi.org/10.3390/en13092311>.
16. *Rizk M., Nasr K.* Computational Fluid Dynamics Investigations Over Conventional and Modified Savonius Wind Turbines // *Heliyon*. 2023. V.9. № 6. P. 16876.  
<https://doi.org/10.1016/j.heliyon.2023.e16876>.
17. *Mohamed M.H., Janiga G., Pap E., Thévenin D.* Optimal Blade Shape of a Modified Savonius Turbine Using an Obstacle Shielding the Returning Blade // *Energy Conversion Management*. 2011. V. 52. № 1. P. 236–242.  
<https://doi.org/10.1016/j.enconman.2010.06.070>.
18. *Chan C.M., Bai H.L., He D.Q.* Blade Shape Optimization of the Savonius Wind Turbine Using a Genetic Algorithm // *Applied Energy*. 2018. V. 213. P. 148–157.  
<https://doi.org/10.1016/j.apenergy.2018.01.02>.
19. *Tartuferi M., D'Alessandro B., Montelpare S., Ricci R.* Enhancement of Savonius Wind Rotor Aerodynamic Performance: a Computational Study of New Blade Shapes and Curtain Systems // *Energy*. 2015. V. 79. P. 371–384.  
<https://doi.org/10.1016/j.energy.2014.11.023>.
20. *Tomar S.S., Dewan A., Singh T.P.* Effects of Axisymmetric-Omnidirectional Deflector on Aerodynamics of Modified Bach Savonius Rotor for Power Enhancement // *Energy Convers. Manag.* 2023. V. 297. P. 117720.  
<https://doi.org/10.1016/j.enconman.2023.117720>.
21. *Selyutskiy Y.D., Klimina L.A., Masterova A.A., Hwang S.S., Lin C.H.* Savonius Rotor as a Part of Complex Systems // *J. Sound & Vibr.* 2019. V. 442. P. 1–10.  
<https://doi.org/10.1016/j.jsv.2018.10.020>.
22. *Zolotov I.I., Shevtsov A.A., Mkrtchyan S.V.* Simulation Model of Dynamic Voltage Stabilizer for Autonomous Power Supply Systems // *J. Physics: Conference Series*. 2019. V. 1333, N. 6. P. 062034.  
<https://doi.org/10.1088/1742-6596/1333/6/062034>.
23. *Коршунов А.И.* Два подхода к анализу устойчивости стабилизаторов напряжения постоянного тока с переменной структурой силовой части // *Практическая силовая электроника*. 2017. № 2(66). С. 12–19.
24. *Dosaev M.Z., Lin C.H., Lu W.L., Samsonov V.A., Selyutskii Yu.D.* A Qualitative Analysis of the Steady Modes of Operation of Small Wind Power Generators // *J. Appl. Math. Mech.* 2009. V. 73. № 3. P. 259–263.  
<https://doi.org/10.1016/j.jappmathmech.2009.07.015>.

## ДИФФЕРЕНЦИРОВАННЫЕ ОЦЕНКИ МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СЕТИ ПРИ ПОВРЕЖДЕНИИ УЗЛОВ

© 2024 г. Ю. Е. Малашенко<sup>a</sup>, \*, И. А. Назарова<sup>a</sup>, \*\*

<sup>a</sup>ФИЦ ИУ РАН, Москва, Россия

\*e-mail: mala-yur@yandex.ru

\*\*e-mail: irina-nazar@yandex.ru

Поступила в редакцию 08.08.2024 г.

После доработки 15.08.2024 г.

Принята к публикации 16.10.2024 г.

В рамках вычислительного эксперимента исследуется модель управления многопользовательской системой связи. Схема управления состоит в одновременной передаче равных межузловых потоков по всем кратчайшим маршрутам. В ходе эксперимента последовательно рассматривается выход из строя каждого узла при сохранении правил диспетчеризации. Создаются массивы данных, отражающие работоспособность поврежденной сети. Полученные показатели анализируются и сравниваются с аналогичными в исходной сети. Построены гарантированные и усредненные оценки основных функциональных характеристик сети при разрушении отдельных узлов с учетом действующих правил управления. Приводятся итоговые диаграммы для сетей с различными структурными особенностями.

**Ключевые слова:** потоковая модель сети связи; оценка повреждений узлов; загрузка ребер

**DOI:** 10.31857/S0002338824060072, **EDN:** SUFLJM

## DIFFERENTIATED ESTIMATES OF A MULTIUSER NETWORK IN CASE OF NODE DAMAGE

Yu. E. Malashenko<sup>a</sup>, \*, I. A. Nazarova<sup>a</sup>, \*\*

<sup>a</sup>Federal Research Center "Computer Science and Control"  
of the Russian Academy of Sciences, Moscow, 119333 Russia

\*e-mail: mala-yur@yandex.ru

\*\*e-mail: irina-nazar@yandex.ru

In the framework of computational experiments a management model of the multiuser communication system is being investigated. The control scheme consists of simultaneous transmission of equal internodal flows along all shortest routes. During the experiment, the failure of each node is consistently considered while maintaining the rules of dispatching. Arrays of data are being created that reflect the system state of the damaged network. The obtained indicators are analyzed and compared with those in the original network. Guaranteed and averaged estimates of the main functional characteristics of the network during the destruction of individual nodes are constructed, taking into account the current management rules. Summary diagrams for networks with various structural features are provided.

**Keywords:** multicommodity flow model of network, peak load of the damaged network, structural damage response vector

**Введение.** В рамках вычислительных экспериментов исследуется стационарный режим управления многопользовательской пространственно-распределенной сетью связи [1] при разрушении узлов. При моделировании предполагается, что как в исходной, так и в поврежденной сетях используется нормативно-закрепленная схема управления. Согласно установленным диспетчерским правилам, равные входные межузловые информационные потоки

передаются одновременно между всеми парами узлов-корреспондентов по всем возможным существующим кратчайшим маршрутам их передачи.

В ходе эксперимента для каждой поврежденной сети формируются массивы данных о маршрутах, реберных нагрузках и удельных затратах. Сравнительный анализ полученных характеристик позволяет проследить зависимость показателей функционирования от уменьшения пропускной способности поврежденной сети. На основе эталонных данных для исходной сети строятся дифференцированные оценки, характеризующие поврежденную сеть: увеличение транзитных потоков и перегрузку на ребрах, дефицит ресурсов в сетевых узлах.

По результатам экспериментов вычисляются максимальные оценки изменения загрузки ребер и узлов при повреждениях. Определяются усредненные значения загрузки ребер и удельных затрат при передаче межузловых потоков. На основе агрегированных расчетных показателей строятся итоговые диаграммы для сетей с различными структурными особенностями.

При построении моделей реальных систем, в которых ресурсы ограничены, возникают сложности при описании их функционирования и получении численных результатов. В работе [2] приводится обзор работ и моделей беспроводных систем связи, основанных на оценке вероятностно-временных характеристик и алгоритмах, пригодных для получения численных результатов. В работе [3] предложен метод, позволяющий оценить минимальную плотность развертывания базовых станций для обеспечения заданной производительности при поддержке ресурсоемких приложений, требующих чрезвычайно высоких скоростей на уровне радиоинтерфейса. В работах [4, 5] рассматриваются задачи синтеза изменения нагрузки на узлы сетевой вычислительной инфраструктуры. Рост нагрузки в реальных системах ведет к необходимости перераспределения ресурсов и перенаправления потоков данных. Предложенные подходы [4, 5] могут быть использованы для оценки потребности в ресурсах при изучении различных телекоммуникационных систем.

Данная статья продолжает изучение многопользовательской сети [6–8] и опирается на методологию исследования операций [9–11]. Предложенную в настоящей работе процедуру можно рассматривать как один из возможных подходов к принятию решений в условиях реально существующей неопределенности о месте и цели повреждения коммутационных узлов систем связи и управления специального назначения [1].

**1. Математическая модель.** Для описания многопользовательской сетевой системы связи воспользуемся следующей математической записью модели передачи многопродуктового потока. Сеть  $G$  задается множествами  $\{V, R, U, P\}$ : узлов (вершин) сети  $V = \{v_1, v_2, \dots, v_n, \dots, v_N\}$ ; неориентированных ребер  $R = \{r_1, r_2, \dots, r_k, \dots, r_E\}$ ; ориентированных дуг  $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$ ; пар узлов-корреспондентов  $P = \{p_1, p_2, \dots, p_M\}$ . Предполагается, что в сети отсутствуют петли и сдвоенные ребра.

Ребро  $r_k \in R$  соединяет смежные вершины  $v_{n_k}, v_{j_k}$ . Каждому ребру  $r_k$  ставится в соответствие две ориентированные дуги  $u_k, u_{k+E}$  из множества  $U$ . Дуги  $\{u_k, u_{k+E}\}$  определяют прямое и обратное направление передачи потока по ребру  $r_k$  между концевыми вершинами  $v_{n_k}, v_{j_k}$ . Для каждой вершины  $v_n$ , формируется список  $K(n)$  номеров инцидентных ей ребер:  $K(n) = \{k_1^n, k_2^n, \dots, k_{a(n)}^n\}$ , где  $a(n)$  – число инцидентных ребер для  $v_n$ .

В многопользовательской сети  $G$  рассматривается  $M = N(N - 1)$  независимых, невзаимозаменяемых и равноправных межузловых потоков различных видов. Каждой паре узлов-корреспондентов  $p_m$  из множества  $P$  соответствуют: вершина-источник с номером  $s_m$ , из  $s_m$  входной поток  $m$ -го вида поступает в сеть; вершина-приемник с номером  $t_m$ , из  $t_m$  поток  $m$ -го вида покидает сеть. Для каждой вершины  $v_n \in V, n = 1, N$ , выделяется подмножество  $P(v_n)$  всех пар-корреспондентов, для которых вершина  $v_n$  служит узлом-источником:

$$P(v_n) = \{p_m | s_m = n, t_m \neq n, t_m = \overline{1, N}\},$$

а для каждого  $P(v_n)$  – список номеров  $M(n)$  пар  $p_m$ , входящих в подмножество  $P(v_n)$ :

$$M(n) = \{m_1(n), m_2(n), \dots, m_{N-1}(n)\}.$$

Обозначим через  $z_m$  величину межузлового потока  $m$ -го вида, поступающего в сеть через узел с номером  $s_m$  и покидающего сеть из узла с номером  $t_m$ ;  $x_{mk}, x_{m(k+E)}$  – поток  $m$ -го вида, который передается по дугам  $u_k$  и  $u_{k+E}$ , согласно направлению передачи,  $x_{mk} \geq 0, x_{m(k+E)} \geq 0$ ,  $m = \overline{1, M}, k = \overline{1, E}; S(v_n)$  – множество номеров исходящих дуг, по ним поток покидает узел  $v_n$ ;  $T(v_n)$  – множество номеров входящих дуг, по ним поток поступает в узел  $v_n$ . Состав множеств  $S(v_n), T(v_n)$  однозначно определяется в ходе выполнения следующей процедуры. Пусть неко-

торое ребро  $r_k \in R$  соединяет вершины с номерами  $n$  и  $j$ , такими, что  $n < j$ . Тогда ориентированная дуга  $u_k = (v_n, v_j)$ , направленная из вершины  $v_n$  в  $v_j$ , считается *исходящей* из вершины  $v_n$ , и ее номер  $k$  заносится в множество  $S(v_n)$ , а дуга  $u_{k+E}$ , направленная из  $v_j$  в  $v_n$  – *входящей* для  $v_n$ , и ее номер  $k+E$  помещается в список  $T(v_n)$ . Дуга  $u_k$  является *входящей* для  $v_j$ , и ее номер  $k$  попадает в  $T(v_j)$ , а дуга  $u_{k+E}$  – *исходящей*, и номер  $k+E$  вносится в список исходящих дуг  $S(v_j)$ .

Во всех узлах сети  $v_n \in V$ ,  $n = \overline{1, N}$ , для каждого вида потока должны выполняться условия сохранения потоков:

$$\sum_{i \in S(v_n)} x_{mi} - \sum_{i \in T(v_n)} x_{mi} = \begin{cases} z_m, & \text{если } v_n = v_{s_m}, \\ -z_m, & \text{если } v_n = v_{t_m}, \\ 0 & \text{в остальных случаях,} \end{cases}$$

$$n = \overline{1, N}, m = \overline{1, M}, x_{mi} \geq 0, z_m \geq 0,$$

Величина  $z_m$  равна входному межузловому потоку  $m$ -го вида, проходящему от источника  $s_m$  к приемнику  $t_m$  пары  $p_m$  при распределении потоков  $\{x_{mi}\} = \mathbf{x}$  по дугам сети.

Каждому ребру  $r_k \in R$  приписывается неотрицательное число  $d_k$  – суммарный предельно допустимый поток, который можно передать по ребру  $r_k$  в обоих направлениях. В исходной сети компоненты вектора пропускных способностей  $\mathbf{d} = (d_1, d_2, \dots, d_k, \dots, d_E)$  – положительные числа  $d_k > 0$ .

Вектор  $\mathbf{d}$  определяет следующие ограничения на сумму потоков всех видов, передаваемых по ребру  $r_k$  одновременно:

$$\sum_{m=1}^M (x_{mk} + x_{m(k+E)}) \leq d_k, \quad x_{mk} \geq 0, x_{m(k+E)} \geq 0, k = \overline{1, E}. \quad (1.2)$$

Ограничения (1.1), (1.2) задают множество допустимых значений вектора межузловых потоков  $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$ :

$$Z(\mathbf{d}) = \{ \mathbf{z} \geq 0 \mid \exists \mathbf{x} \geq 0: (\mathbf{z}, \mathbf{x}) \text{ удовлетворяют (1.1), (1.2)} \}$$

**2. Схема вычислительного эксперимента.** Вычислительный эксперимент проводился на моделях сетевых систем, представленных на рис. 1, 2. В каждой сети имеется 69 узлов. Пропускные способности ребер  $d_k(0)$  равномерно распределены на отрезке  $[900, 999]$ . В ходе исследования была сделана нормировка, и суммарная пропускная способность в обеих сетях одинакова:

$$\sum_{k=1}^E d_k(0) = D(0) = 68256.$$

При моделировании вначале подсчитываются межузловые потоки в неповрежденной сети  $G(0)$  с пропускными способностями  $d_k(0)$ . Предполагается, что поток одного вида передается одновременно по всем маршрутам, содержащим минимальное число ребер (далее – MER-маршруты от англ. minimum edge route). Для оценки величины “расщепленного” потока

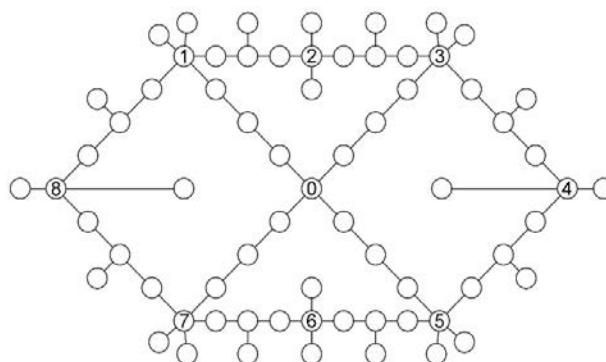


Рис. 1. Базовая сеть.

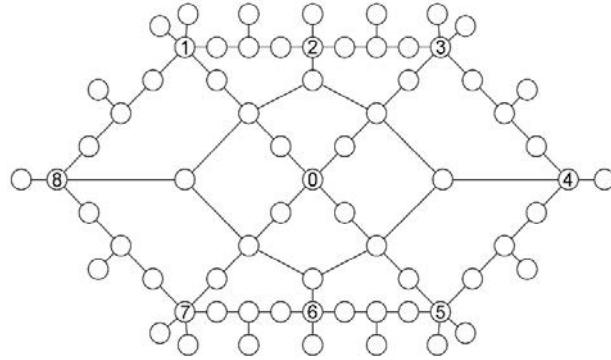


Рис. 2. Кольцевая сеть.

для каждой пары узлов  $p_m = (s_m, t_m)$  в сети  $G(0)$  формируется набор  $H_m(1)$  путей, которые далее рассматриваются как MER-маршруты передачи  $m$ -го вида потока:

$$H_m(1) = \left\{ h_m^1(1), h_m^2(1), \dots, h_m^j(1), \dots, h_m^{J_m(1)}(1) \right\},$$

где  $h_m^j(1)$  – список номеров дуг в  $j$ -м пути в сети  $G(0)$  между узлами;  $s_m$  – число ребер в MER-маршруте  $t_m$ ;  $J_m(1)$  – число MER-маршрутов для  $m$ -й пары.

Для каждой пары  $p_m \in P$  по всем MER-маршрутам из  $H_m(1)$  передается единичный межузловый поток  $z_m$  и вычисляются значения индикаторной функции:

$$\eta_k^j(m) = \begin{cases} 1, & \text{если } k \in h_m^j(1), \\ 0 & \text{в остальных случаях} \end{cases}$$

Определяются дуговые потоки для пары  $p_m$ :

$$x_{mk}^0(1) = \sum_{j=1}^{J_m(1)} \eta_k^j(m), m = \overline{1, M}, k = \overline{1, 2E}. \quad (2.1)$$

Межузловой поток по MER-маршрутам (далее – MER-поток)  $z_m^0(1)$  между узлами  $s_m$  и  $t_m$  вычисляется по формулам (1.1), (2.1). Рассчитывается нормирующий коэффициент:

$$\omega_m^0(1) = \frac{1}{z_m^0(1)}, z_m^0(1) \neq 0, m = \overline{1, M},$$

и дуговые потоки

$$x_{mk}^0 = \omega_m^0(1) x_{mk}^0(1), m = \overline{1, M}, k = \overline{1, 2E}. \quad (2.2)$$

При передаче всех потоков  $x_{mk}^0$  по ребрам сети межузловой поток из узла  $s_m$  в узел  $t_m$  равен единице для всех  $p_m \in P$ .

В сети  $G(0)$  для каждого ребра  $r_k$  подсчитывается загрузка:

$$\Delta_k(0) = \sum_{m=1}^M \left[ x_{mk}^0 + x_{m(k+E)}^0 \right],$$

т. е. величина  $\Delta_k(0)$  находится на основании значений  $x_{mk}^0$  при одновременной передаче всех межузловых потоков, таких, что

$$z_m^0 = 1, \text{ для всех } p_m \in P.$$

**3. Оценки изменения показателей функционирования.** Для оценки последствий повреждения узла  $v_j$  на основе исходной сети  $G(0)$  формируется сеть  $G(v_j)$ , в которой пропускная способность ребер, инцидентных вершине  $v_j$ , полагается равной нулю:

$$d_k(v_j) = \begin{cases} 0, & \text{если } k \in K(v_j), \\ d_k(0) & \text{в остальных случаях,} \end{cases}$$

$$D(v_j) = D(0) - \sum_{k \in K(v_j)} d_k(0),$$

здесь и далее  $K(v_j)$  – множество номеров ребер, инцидентных вершине  $v_j$ .

Формально для сети  $G(v_j)$  применяется схема построения MER-маршрутов и, согласно (2.2), строятся дуговые потоки  $\{x_{mk}^0\}$ ,  $k = 1, 2E$ , для межузловых потоков  $z_m^j$ . Вводятся обозначения:  $P^-(v_j) = \{m_1, m_2, \dots\}$  – список пар, для которых в сети  $G(v_j)$  не существует пути соединения;  $M^-(v_j)$  – число таких пар. Для всех пар  $p_m \in P^-(v_j)$  соответствующие потоки  $z_m^j$  равны нулю. Обозначим через  $P^+(v_j)$  список пар узлов, таких, что поток между ними в сети  $G(v_j)$  не равен нулю, т.е. для всех пар-корреспондентов выполняется

$$z_m^j = 0, m \in P^-(v_j), z_m^j = 1, m \in P^+(v_j).$$

Пусть  $M^+(v_j)$  – число пар в  $P^+(v_j)$ , а  $M^-(v_j)$  – соответственно в  $P^-(v_j)$ . При этом

$$P = P^+(v_j) \cup P^-(v_j), M = M^+(v_j) + M^-(v_j)$$

В каждой сети  $G(v_j)$ ,  $j = \overline{1, N}$ , относительное изменение (превышение/уменьшение) загрузки ребра  $r_k$  рассчитывается относительно начальных показателей:

$$\xi_k^j = \left[ \frac{\Delta_k(v_j)}{\Delta_k(0)} - 1 \right], k = \overline{1, E}.$$

Для оценки относительного изменения пропускной способности сети при удалении узла  $v_j$  вычисляется

$$\delta^j = \frac{\sum_{k \in K(v_j)} d_k(0)}{D(0)}, j = \overline{1, N},$$

где  $K(v_j)$  – множество номеров ребер, инцидентных узлу  $v_j$ . Для них пропускная способность в сети  $G(v_j)$  становится равной нулю. Для получения количественных оценок изменения функциональных показателей системы при удалении узла  $v_j$  подсчитываются величины:

$$\theta_k^j = \frac{\xi_k^j}{\delta^j}, k = \overline{1, E}.$$

Фактически, если значения  $\xi_k^j$  и  $\delta^j$  записать в процентах, то величина  $\theta_k^j$  показывает, на сколько процентов увеличится загрузка  $k$ -го ребра при уменьшении пропускной способности сети на 1% при удалении  $j$ -го узла. Значения  $\theta_k^j$  упорядочиваются по величине от большего к меньшему (по невозрастанию) и перенумеровываются согласно порядку:

$$\{\theta_i^j\}: \theta_i^j \geq \theta_{i+1}^j, i = \overline{1, E-1}.$$

Полученные лексикографически упорядоченные величины  $\{\theta_i^j\}$  являются  $i$ -ми компонентами вектор-отклика  $\theta(j)$ .

Для каждой фиксированной  $i$ -й компоненты вектор-отклика  $\theta(j)$  среди всех повреждений во всех сетях  $G(v_j)$  определяются

$$\Theta_i^* = \max_j \theta_i^j, \Theta_i^{**} = \min_j \theta_i^j, i = \overline{1, E}.$$

Последовательность  $\{\Theta_i^*\}$ ,  $i = \overline{1, E}$ , представляет собой гарантированные оценки компонент векторов-отклика системы на повреждение узла.

На диаграммах рис. 3, 4 для базовой и кольцевой сетей представлены оценки  $\Theta_i^*$ , которые упорядочены по величине от большего к меньшему, а номера указаны по горизонтальной оси. Кривые, описывающие  $\Theta_i^*$ , служат верхними огибающими для всех распределений вектор-откликов  $\theta(j)$  при повреждениях узлов  $v_j$ . По построению  $\Theta_i^*$  можно рассматривать как гарантированные верхние оценки откликов при разрушении любой вершины графа сети. При повреждении любого узла  $v_j$  точки, описывающие лексикографически упорядоченное распределение вектор-откликов  $\theta_i^j$ , лежат ниже  $\Theta_i^*$ .

Анализ диаграмм рис. 3, 4 показывает, что при уменьшении пропускной способности сети на 1% рост загрузки ребер в базовой сети не превышает 30%, однако в кольцевой сети при некоторых повреждениях может достигать 60%. В ходе экспериментов с выходом из строя узла пропускная способность сети  $G(v_j)$  в среднем уменьшается на 3–4%. Таким образом в базовой сети

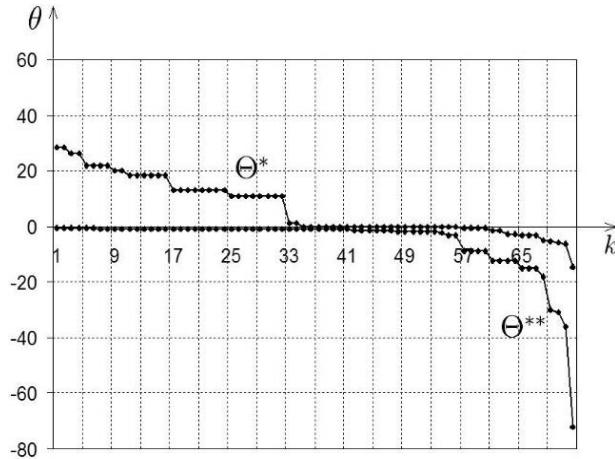


Рис. 3. Оценки компонент векторов отклика в базовой сети.

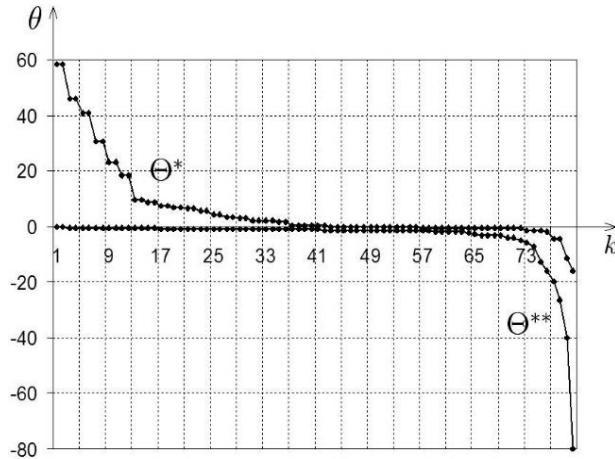


Рис. 4. Оценки компонент векторов отклика в кольцевой сети.

реберная нагрузка может удваиваться, а в кольцевой – возрасти в 3 раза. Разрушение отдельного узла требует изменения маршрутов передачи потока для значительного числа пар-корреспондентов, вызывает перегрузки на определенных участках сети и ухудшает ее работоспособность. Диаграммы на рис. 3, 4 иллюстрируют количественные оценки последствий разрушений.

**4. Оценка загрузки узлов при повреждениях.** На основании (2.2) для получения оценки максимального значения равных межузловых потоков решается следующая задача.

Задача. Найти  $\tilde{\alpha} = \max_{\alpha} \alpha$

$$\text{при условиях: } \alpha \sum_{m=1}^M \left[ x_{mk}^0 + x_{m(k+E)}^0 \right] \leq d_k(0), \alpha \geq 0, k = \overline{1, E}.$$

С помощью решения задачи для всех  $p_m \in P$  определяется вектор  $\tilde{z}(\tilde{\alpha})$ , все компоненты которого равны  $\tilde{\alpha}$ , т.е.

$$\tilde{z}_m = \tilde{\alpha}, m = \overline{1, M}, \tilde{x}_{mk} = \tilde{\alpha} x_{mk}^0, m = \overline{1, M}, k = \overline{1, 2E}.$$

При одновременной передаче всех потоков в сети  $G(0)$  для каждого ребра  $r_k$  загрузка  $\tilde{\Delta}_k$  подсчитывается исходя из дуговых потоков  $\tilde{x}_{mk}$ :

$$\tilde{\Delta}_k = \sum_{m=1}^M \left[ \tilde{x}_{mk} + \tilde{x}_{m(k+E)} \right].$$

Для каждой пары узлов-корреспондентов  $p_m \in P$ , для полученного допустимого межузлового потока  $\tilde{z}_m$  и соответствующих значений дуговых потоков  $\tilde{x}_{mk}$ ,  $k = \overline{1, 2E}$ , величина

$$\tilde{Y}_m = \sum_{k=1}^{2E} \tilde{x}_{mk}, \quad m = \overline{1, M},$$

характеризует результирующую межузловую *нагрузку* на ребра сети  $G(0)$  при передаче межузлового потока  $\tilde{z}_m$  из узла-источника  $s_m$  в узел-приемник  $t_m$ . Величина  $\tilde{Y}_m$  показывает, какая суммарная пропускная способность сети потребуется для передачи дуговых потоков  $\tilde{x}_{mk}$ .

В рамках модели отношение результирующей межузловой нагрузки и межузлового потока

$$w_m = \frac{\tilde{Y}_m}{\tilde{z}_m}, \quad m = \overline{1, M},$$

можно трактовать как *удельные затраты* ресурсов сети при передаче единичного потока  $m$ -го вида между узлами  $s_m$  и  $t_m$  при дуговых потоках  $\tilde{x}_{mk}$ .

В каждой поврежденной сети  $G(v_j)$ ,  $j = \overline{1, N}$ , определяется набор кратчайших путей, согласно (2.2), потоки

$$z_m^j = 1, \quad m \in P^+(v_j), \quad x_{mk}^0, \quad m = \overline{1, M},$$

и далее формируется набор:

$$\tilde{z}_m(v_j) = \tilde{\alpha}, \quad m \in P^+(v_j), \quad \tilde{x}_{mk}(v_j) = \tilde{\alpha} x_{mk}^0, \quad m = \overline{1, M}, \quad k = \overline{1, 2E}.$$

В сети  $G(v_j)$  для каждого ребра  $r_k$  подсчитывается загрузка:

$$\tilde{\Delta}_k(v_j) = \sum_{m=1}^M \left[ \tilde{x}_{mk}(v_j) + \tilde{x}_{m(k+E)}(v_j) \right].$$

При одновременной передаче всех межузловых потоков из множества  $P^+(v_j)$  их значения

$$\tilde{z}_m(v_j) = \tilde{\alpha} \text{ для всех } m \in P^+(v_j).$$

Для каждой неповрежденной вершины  $v_n$  в сети  $G(v)$  загрузка всех инцидентных ей ребер равна

$$\gamma_n(v_j) = \sum_{k \in K(v_n)} \tilde{\Delta}_k(v_j),$$

пропускная способность всех инцидентных ребер составляет

$$D_n(v_j) = \sum_{k \in K(v_n)} d_k(v_j).$$

Относительная величина

$$\rho_n(v_j) = \frac{\gamma_n(v_j)}{D_n(v_j)}, \quad n = \overline{1, N}, \quad n \neq j, \quad (4.1)$$

показывает, как меняется загрузка узла с номером  $n$  при сохранении передачи межузловых потоков  $\tilde{z}_n$ , равных  $\tilde{\alpha}$ , в поврежденной сети  $G(v_j)$ .

Для получения гарантированных оценок функционирования сети все  $\rho_n(v_j)$  при фиксированном  $v_j$  переупорядочиваются от большего к меньшему (по неубыванию):

$$\{\rho_k(v_j)\}: \rho_i(v_j) \geq \rho_{i+1}(v_j), \quad i = \overline{1, N-1},$$

и определяются верхняя и нижняя оценки:

$$\rho_i^* = \max_j \rho_i(v_j), \quad \rho_i^{**} = \min_j \rho_i(v_j).$$

На рис. 5, 6 представлены диаграммы значений  $\rho_i$  и  $\rho_i^*$  для базовой и кольцевой сетей. Из диаграмм следует, что в базовой сети величины перегрузок больше и возникают они чаще, чем в кольцевой. В базовой сети удаление узла, расположенного, например, на “кольцевой” линии, приводит к удлинению маршрутов передачи (соединений) для большого числа корреспондентов и, как следствие, увеличению транзитной нагрузки на оставшиеся узлы. В кольцевой сети на 10% больше ребер и короче маршруты соединения, что уменьшает суммарную

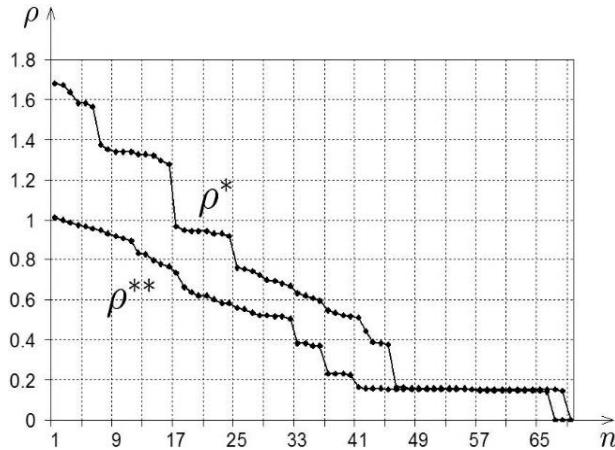


Рис. 5. Показатели перегрузки узлов в базовой сети.

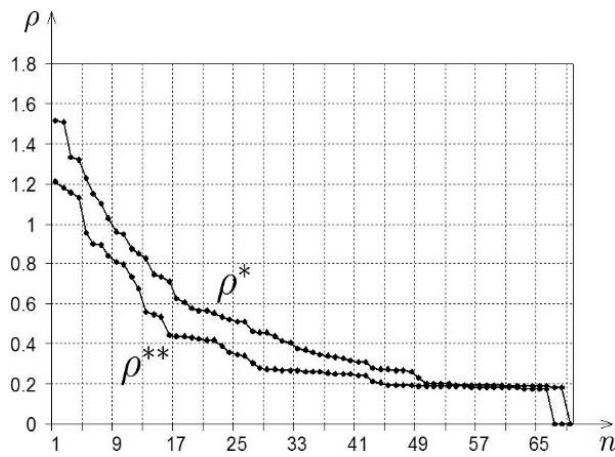


Рис. 6. Показатели перегрузки узлов в кольцевой сети.

загрузку сети. Однако высокая загрузка ребер на внутреннем кольце приводит к перегрузкам в узлах, смежных с разрушенным. Указанные узлы в этом случае становятся транзитными для тех потоков, которые передавались в неповрежденной сети по другим маршрутам.

**5. Оценка показателей работоспособности при повреждении узла.** Повреждение отдельного узла вынуждает систему управления использовать другие маршруты передачи и перенаправлять потоки, что в свою очередь ведет к перегрузкам в узлах. Будем считать, что показатели *работоспособности* узла находятся в норме, если загрузка узла не превышает его суммарной пропускной способности. Далее исследуем случай, когда узел функционирует не в штатном режиме.

Для оценки показателей *работоспособности* при повреждении некоторого узла  $v_j$  используется массив расчетных данных (4.1). Для каждого узла  $v_j, j = \overline{1, N}$ , в поврежденной сети  $G(v_j)$  среди всех неповрежденных узлов вычисляется максимально возможное превышение реберными потоками (загрузкой) суммарной пропускной способности:

$$\Psi_1(v_j) = \max_n \left\{ \left[ \rho_n(v_j) - 1 \right] \mid n = \overline{1, N}, n \neq j \right\}.$$

Обозначим через  $\varkappa_n(v_j)$  индикаторную функцию, показывающую изменение загрузки  $n$ -й вершины в сети:

$$\varkappa_n(v_j) = \begin{cases} 1, & \text{если } \rho_n(v_j) > 1, \\ 0 & \text{в остальных случаях;} \end{cases}$$

$\varkappa^+(v_j)$  – число узлов, для которых изменение загрузки больше единицы:

$$\varkappa^+(v_j) = \sum_{n=1}^N \varkappa_n(v_j).$$

В каждой сети  $G(v_j), j = \overline{1, N}$  подсчитываются  $\varkappa^+(v_j)$  и доля узлов с изменением загрузки больше единицы:

$$\Psi_2(v_j) = \frac{\varkappa^+(v_j)}{N}.$$

На рис. 7, 8 представлены диаграммы значений  $\{\Psi_1(v_j), \Psi_2(v_j)\}$  для всех  $j = \overline{1, N}$ .

Анализ диаграмм на рис. 7, 8 показывает, что выход из строя узлов базовой сети приводит к более значительным изменениям показателей функционирования по сравнению с кольцевой. Равенство  $\Psi_2(v_j) = 0.25$  означает, что при некотором повреждении перегрузка будет наблюдаться в каждом четвертом узле базовой сети, а в кольцевой – только в каждом десятом. В кольцевой сети максимальная перегрузка  $\Psi_2(v_j) \leq 0.5$  – не более 50%. В базовой сети максимальные значения  $0.6 \leq \Psi_2(v_j) \leq 0.7$  лежат в диапазоне 60–70%. В кольцевой сети на 10% больше ребер, есть два кольца, большая связность сети, поэтому целый ряд показателей *работоспособности* при разрушении отдельного узла меняется меньше, чем в базовой.

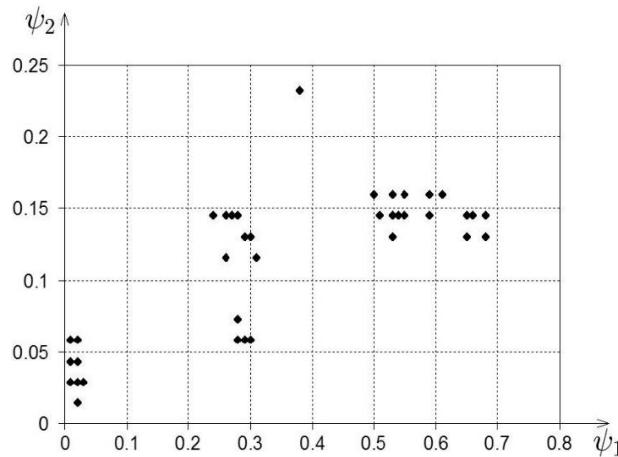


Рис. 7. Сравнительные показатели работоспособности в базовой сети.

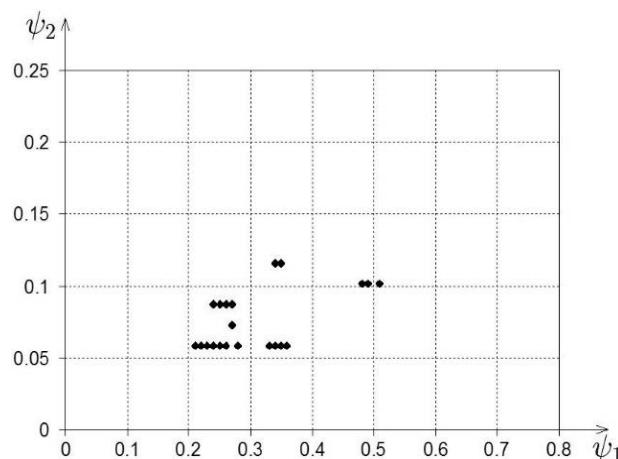


Рис. 8. Сравнительные показатели работоспособности в кольцевой сети.

**6. Усредненные показатели.** Для оценки средних значений загрузки ребер относительно пропускной способности для каждой сети  $G(v_j), j = 1, N$ , вычисляется

$$v_k^j = \frac{\Delta_k(v_j)}{d_k(v_j)} \text{ для всех } k = \overline{1, E} \text{ таких, что } d_k(v_j) = d_k(0) > 0,$$

и средняя (относительная) загрузка ребра при различных повреждениях:

$$v_k^+ = \frac{1}{N} \sum_{j=1}^N v_k^j, k = \overline{1, E}.$$

Для оценки отклонений  $v_k^j$  в поврежденной сети  $G(v_j)$ , исходя из начальных  $v_k(0)$ , определяется среднее значение:

$$\delta_k^+ = \frac{1}{\sqrt{N}} \left[ \sum_{j=1}^N (v_k^j - v_k(0))^2 \right]^{1/2}, \text{ где } v_k(0) = \frac{\Delta_k}{d_k(0)}, k = \overline{1, E}.$$

На рис. 9, 10 для базовой и кольцевой сетей представлены диаграммы значений  $v_k^+$ ,  $\delta_k^+$ , упорядоченных от большего к меньшему, номера указаны по горизонтальной оси. Значения  $v_k^+$  служат оценками средней загрузки ребра при различных повреждениях, а  $\delta_k^+$  используются как оценки отклонения загрузок в поврежденной сети от исходных.

Правые части графиков на рис. 9, 10 практически параллельны оси абсцисс и соответствуют ребрам, инцидентным висячим вершинам. Как уже отмечалось выше, загрузка ребер

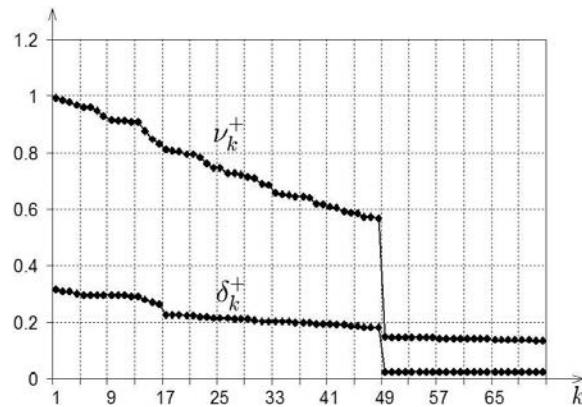


Рис. 9. Средняя оценка загрузки ребер при повреждениях в базовой сети.

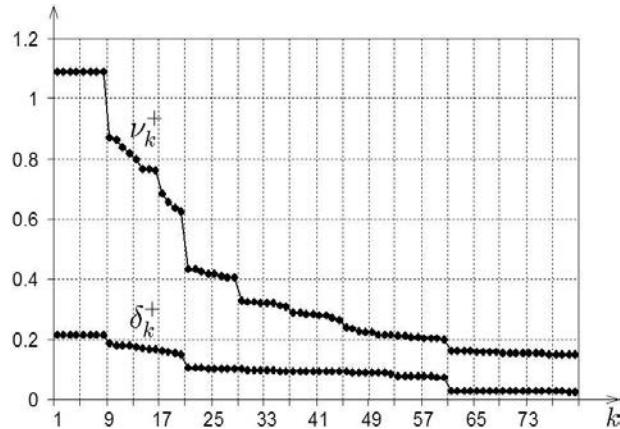


Рис. 10. Средняя оценка загрузки ребер при повреждениях в кольцевой сети.

к висячим вершинам совпадает с исходной при повреждении большинства узлов и становит-  
ся равной нулю при разрушении инцидентных висячих вершин. В базовой сети средняя за-  
грузка ребер при повреждениях меньше исходной, но отклонения невелики. Для кольцевой  
сети даже усредненные показатели для некоторых ребер превышают загрузки в исходной сети  
и разброс значений больше, чем в базовой. В кольцевой сети через центральные узлы прохо-  
дит много кратчайших маршрутов, и удаление этих вершин приводит к резкому увеличению  
транзитных потоков за счет возрастания длины путей соединения.

**7. Оценка удельных затрат.** Формально для сети  $G(v_j)$ , применяемая схема построения  
MER-маршрутов и, согласно (2.2), строятся дуговые потоки  $\{x_{mk}^0\}$ ,  $k = 1, 2E$ , для межузловых  
потоков  $z_m^j$ .

В разд. 3 через  $P^+(v_j)$  был обозначен список пар, поток между которыми в сети  $G(v_j)$  не ра-  
вен нулю, т.е.

$$z_m^j = 1, w_m^j \geq w_m(0), m \in P^+(v_j),$$

где  $w_m(0)$

— удельные затраты в исходной сети  $G(0)$  при условии (2.2). Для всех  $m \in P^+(v_j)$  вычисляются

$$\varphi_m^j = \frac{w_m^j}{w_m(0)}, w_m(0) \neq 0,$$

и подсчитывается среднее значение относительного увеличения удельных затрат на передачу  
межузлового потока  $z_m$  при повреждении узла  $v_j$ :

$$\varphi_+^j = \frac{1}{M^+(v_j)} \sum_{p_m \in P^+(v_j)} \varphi_m^j.$$

Определяется доля пар  $p_m$ , для которых путь соединения в сети  $G(v_j)$  отсутствует:

$$\Psi_-^j = \frac{M^-(v_j)}{M}.$$

Показатели  $\varphi_+^j$ ,  $\Psi_-^j$  используются далее в качестве оценки ущерба при повреждении узла  $v_j$ .  
Значения  $\varphi_+^j$ ,  $\Psi_-^j$  находятся последовательно для всех узлов  $v_j$ ,  $j = \overline{1, N}$ .

На рис. 11, 12 представлены диаграммы изменения удельных затрат  $\varphi_+^j$  и доли пар  $\Psi_-^j$  при  
повреждении узла  $v_j$ . Величины  $\varphi_+^j$  и  $\Psi_-^j$  переупорядочены от большего к меньшему и откладыва-  
ются по вертикальной оси. По горизонтальной оси указаны порядковые номера поврежден-  
ных узлов. Верхние кривые на рис. 11, 12 относятся к  $\varphi_+^j$  и соответствуют средним значениям  
удельных затрат на передачу межузловых потоков при повреждении узла  $v_j$ . Нижние указывают  
на  $\Psi_-^j$  — долю общего числа корреспондентов, для которых не существует пути передачи. При  
повреждении одного узла пропускная способность от одного до пяти ребер становится равной  
нулю. В среднем повреждение 3% ребер “разделяет” 3–4% пар-корреспондентов.

Из рис. 12 для кольцевой сети следует, что число повреждений около 20% приводит к ро-  
сту удельных затрат на 10%. Число “разъединенных” пар-корреспондентов не превышает 3%  
при 75% повреждений. Для базовой сети (рис. 11) наличие более 75% повреждений увеличи-  
вает удельные затраты более чем на 10%, а более 25% “разъединяют” более 5% пар. В обеих

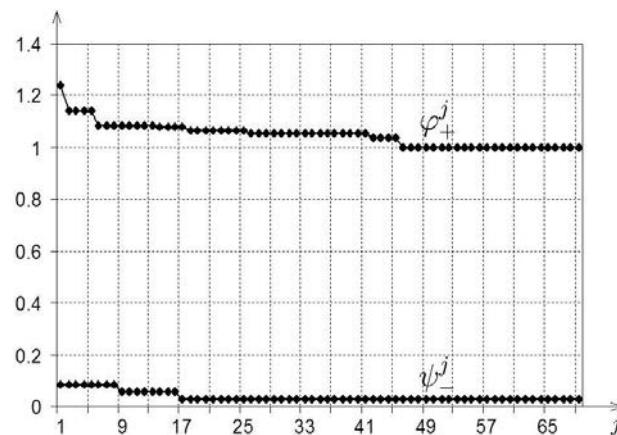


Рис. 11. Средние удельные затраты при повреждении узла в базовой сети.

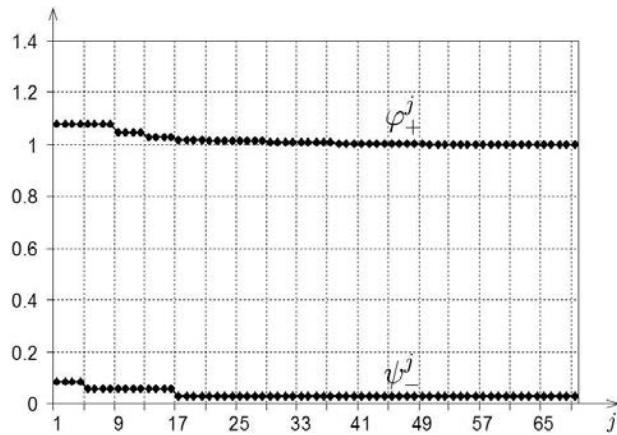


Рис. 12. Средние удельные затраты при повреждении узла в кольцевой сети.

сетях при повреждении висячих узлов кратчайшие пути для всех корреспондентов остаются неизменными. В результате правые части диаграмм на рис. 11, 12 практически совпадают для номеров узлов более 45 как для кольцевой, так и для базовой сетей.

**Заключение.** Вычислительные эксперименты и последующий анализ проводились в рамках формализма многопродуктовой потоковой модели. Использование указанной математической модели позволяет получить агрегированный срез данных о взаимосвязи между жесткой комбинаторно-графовой природой сети и множеством векторов-откликов. Полученные количественные оценки фактически характеризуют изменения потоковой нагрузки на сеть в случае выхода из строя узлов и ребер.

Представленную выше модель можно расценивать как один из вариантов постановки традиционных задач о поиске и определении критических элементов сети, тогда предложенная агрегированная гарантированная оценка ущерба будет служить вариантом постановки [12]. Для поиска критически опасных повреждений [13], который сводится к NP-трудной задаче, в качестве эффективных эвристик для метода ветвей-и-границ подойдут изложенные в данной статье варианты вычисления векторных оценок.

#### СПИСОК ЛИТЕРАТУРЫ

1. Пехтерев С.В., Макаренко С.И., Ковальский А.А. Описательная модель системы спутниковой связи Starlink // Системы управления, связи и безопасности. 2022. № 4. С. 190–255.
2. Горбунова А.В., Наумов В.А., Гайдамака Ю.В., Самуйлов К.Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи // Информатика и ее применения. 2018. Т. 12. Вып. 3. С. 48–55.
3. Бесчастный В.А., Острикова Д.Ю., Шоргин С.Я., Молчанов Д.А., Гайдамака Ю.В. Анализ плотности базовых станций 5G NR для предоставления услуг виртуальной и дополненной реальности // Информатика и ее применения. 2022. Т. 16. Вып. 2. С. 102–108.
4. Горшенин А.К., Горбунов С.А., Волканов Д.Ю. О кластеризации объектов сетевой вычислительной инфраструктуры на основе анализа статистических аномалий в трафике // Информатика и ее применения. 2023. Т. 17. Вып. 3. С. 76–87.
5. Горшенин А.К. О некоторых математических и программных методах построения структурных моделей информационных потоков // Информатика и ее применения. 2017. Т. 11. Вып. 1. С. 58–68.
6. Малашенко Ю.Е., Назарова И.А. Сравнительный анализ узловых мультипотоков в многопользовательской сетевой системе // Информатика и ее применения. 2024. Т. 18. Вып. 1. С. 40–45.
7. Малашенко Ю.Е., Назарова И.А. Анализ загрузки многопользовательской сети при расщеплении потоков по кратчайшим маршрутам // Информатика и ее применения. 2023. Т. 17. Вып. 3. С. 19–24.
8. Малашенко Ю.Е., Назарова И.А. Анализ критически опасных повреждений сети связи. IV. Многокритериальные оценки уязвимости кластеров // Изв. РАН. ТиСУ. 2022. № 1. С. 56–66.
9. Фрэнк Г., Фриши М. Сети, связь и потоки. М.: Связь, 1978. 448 с.
10. Гермейер Ю.Б. Введение в теорию исследования операций. М.: Наука, 1971. 384 с.
11. Данскин Дж.М. Теория максимина и ее приложение к задачам распределения вооружения. М.: Сов. радио, 1970. 200 с.
12. Ponton J., Wei P., Sun D. Weighted Clustering Coefficient Maximization for Air Transportation Networks // Control Conference (ECC). Zurich, 2013. P. 866–871.
13. Kuhnle A., Nguyen N.P., Dinh T.N., Thai M.T. Vulnerability of Clustering under Nodes Failure in Complex Networks // Social Network Analysis and Mining. 2017. V. 7. Iss. 1. P. 8–24.

## ПРИМЕНЕНИЕ БУСТИНГА В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ

© 2024 г. М. А. Жарова<sup>a, \*</sup>, В. И. Щурков<sup>b, \*\*</sup>

<sup>a</sup>МФТИ, Долгопрудный, Россия

<sup>b</sup>ФИЦ ИУ РАН, Москва, Россия

\*e-mail: zharova.ma@phystech.edu

\*\*e-mail: v.tsurkov@frccsc.ru

Поступила в редакцию 07.08.2024 г.

После доработки 01.09.2024 г.

Принята к публикации 16.10.2024 г.

В современной цифровой эпохе рекомендательные системы прочно укрепились, став важным инструментом для эффективного управления информационными потоками. Их востребованность в значительной степени обусловлена динамикой нынешнего общества, а именно информационным переизбытком и необходимостью персонализации данных. С расширением сфер применения рекомендательных алгоритмов появляется и немало нестандартных случаев, для которых использование классических подходов не столь эффективно. Рассматривается один из таких: малое число объектов при относительно большом числе пользователей в условиях высокой корреляции между некоторыми объектами. Для моделирования предлагается градиентный бустинг — алгоритм машинного обучения на основе ансамбля решающих деревьев.

**Ключевые слова:** рекомендательные системы, бустинг, пользователи и объекты, корреляция, калибровка

**DOI:** 10.31857/S0002338824060083, **EDN:** SUDEVR

## APPLICATION OF BOOSTING IN RECOMMENDER SYSTEMS

М. А. Жарова<sup>a, \*</sup>, В. И. Щурков<sup>b, \*\*</sup>

<sup>a</sup>*Moscow Institute of Physics and Technology (MIPT),  
Dolgoprudny, Moscow oblast, Russia*

<sup>b</sup>*Federal Research Center “Computer Science and Control”, Russian  
Academy of Sciences, Moscow, Russia*

\*e-mail: zharova.ma@phystech.edu

\*\*e-mail: v.tsurkov@frccsc.ru

In today's digital era, recommender systems have gained a strong foothold, becoming an important tool for effectively managing information flows. Their demand is largely due to the dynamics of current society, namely information overload and the need to personalize data. With the expansion of the scope of application of recommendation algorithms, many non-standard cases appear, for which the use of classical approaches is not so effective. This paper examines one of these: a small number of objects with a relatively large number of users in conditions of high correlation between some objects. For modeling, it is proposed to use gradient boosting, a machine learning algorithm based on an ensemble of decision trees.

**Keywords:** recommender systems, boosting, users and items, correlation, calibration

**Введение.** В настоящий момент самым частым подходом к построению рекомендательных систем является гибридный, который, как правило, представляет собой комбинацию модели матричной факторизации и какого-либо контентного метода [1]. На первом этапе происходит начальный отбор кандидатов, здесь коллаборативные модели в отдельности способны давать более точные и быстрые прогнозы [2]. При помощи контентных данных решается проблема

холодного старта, свойственная всем алгоритмам коллаборативной фильтрации, а также происходит дополнительное уточнение рекомендаций на основе метаданных.

Таким образом, для получения качественного решения необходимы коллаборативные данные – о взаимодействии пользователей и объектов. Обычно они представляются в виде матрицы и в большинстве классических случаев являются разреженными из-за того, что пользователи взаимодействуют далеко не со всеми объектами. Эта проблема, как известно, решается на этапе построения модели матричной факторизации: через разложение матриц и поиск скрытых факторов [3].

Но что если число объектов в данных много меньше числа пользователей – например, на 1 млн уникальных пользователей приходится лишь 10 уникальных объектов? Во-первых, в такой ситуации матрицы уже нельзя считать достаточно разреженными, поэтому подобную задачу эффективнее решать методами формирования последовательностей, используя теорию информации [4]. Более того, в приведенном порядке соотношений 1:100 000 векторы для каждого пользователя будут иметь настолько низкую размерность, что среди них появится много дубликатов. Например, в простейшем случае, когда не учитывается вес взаимодействия пользователя с объектом, все числа в матрице будут равняться либо 0, либо 1. Тогда для 10 объектов суммарно можно получить лишь  $2^{10} = 1024$  уникальных строк.

Можно было бы это исправить, взяв лишь часть из имеющихся пользователей, которые уникальны по поведению, далее рассчитать для них первичные рекомендации, а затем сделать переранжирование с учетом метаданных. Но в условиях поставленной задачи добавляется проблема корреляции некоторых объектов друг с другом. Так как число объектов крайне ограничено, нужно удалить часть из них, что означает еще большее уменьшение числа признаков или, другими словами, понижение размерности “коллаборативных” векторов для пользователей, из которых состоит матрица.

В качестве альтернативы можно также рассмотреть чистый контентный подход [5]. Он довольно тривиальный и в классическом виде учитывает только метаданные пользователей или объектов в отдельности, поэтому проблема сильной разницы в их числе пропадает. Если рассматривать вышеупомянутый метод на основе характеристик объектов, высокая корреляция между некоторыми из них сохраняется, из-за чего алгоритм поиска похожих просто деградирует до выдачи одинаковых предсказаний для определенных пар объектов. В случае применения контентного алгоритма на основе характеристик пользователей главной проблемой может оказаться ограниченность данных: технически взять их можно из анкетирования, проведения опросов, поиска в сторонних источниках и т.п. Обычно подобные мероприятия не носят обязательного характера, соответственно информацию и достоверную картину про всех пользователей получить сложно. Более того, немаловажную роль здесь играют различные законы о персональных данных.

В ходе анализа получается, что использование классических подходов неуместно для имеющихся данных главным образом из-за малого числа объектов и корреляции между некоторыми из них. Чтобы решить последнюю проблему, можно независимо обучить несколько отдельных моделей под каждый объект. Эта идея была бы трудно реализуемой при большом количестве разнообразных объектов, как это чаще всего бывает в задачах рекомендаций; но в рассматриваемых условиях их количество как раз невелико.

Таким образом, каждая из независимых моделей будет представлять собой обыкновенный классификационный алгоритм. Для его обучения можно использовать метаданные без каких-либо дополнительных преобразований, а коллаборативные перевести в одномерные признаки путем агрегации признаков действий на разных временных периодах: например, посчитать количество просмотров, среднюю сумму покупок, число звонков и т.п. за последние неделю/месяц/квартал. Это позволит учитывать в модели наиболее ценные данные о взаимодействиях пользователей и объектов, чтобы идея не сводилась к обычному контентному методу на основе только метаинформации.

Итак, реализовав описанный подход, можно решить сразу ряд проблем: малого числа объектов, высокой корреляции между некоторыми из них, а также учесть в одноуровневом каскаде моделей и коллаборативные, и метаданные. Затем полученные предсказания обязательно необходимо откалибровать, чтобы получить сравнимые вероятности заинтересованности пользователей во всех объектах.

**1. Постановка задачи.** Формализуем данный подход для математической модели. Пусть имеется множество пользователей  $U$  и их характеристики; множество объектов  $P$  и их характеристики; информация о взаимодействиях  $R = U \times P$  за некоторый временной промежуток.

Составим  $n = |P|$  отдельных наборов данных по пользователям  $U$ , каждый из них будет выглядеть следующим образом:

- 1) характеристики пользователей  $U_i$ , где  $U_i \in U$  – подмножество пользователей, у которых нет соответствующего  $i$ -го объекта  $i = (1, n)$ ;
- 2) рассчитанные агрегированные признаки по выбранному объекту для каждого пользователя (количество взаимодействий различного типа: потраченная сумма, средние показатели и статистики);
- 3) характеристики выбранного объекта  $p$ .

В данной задаче, как и в любой рекомендательной системе, важно обратить внимание на временные особенности – за какой промежуток использовать признаковое пространство и на каком собирать целевую переменную. Это поможет сконструировать правдоподобный процесс работы системы на реальных пользователях, определить примерное время переобучения модели в совокупности с возможностями доступных вычислительных ресурсов и корректно оценить качество, исключив проблему утечки данных [6].

Модели матричной факторизации основываются на заполнении пропусков в матрице user-item, что позволяет спрогнозировать заинтересованность пользователей в объектах, с которыми они еще не взаимодействовали [7]. При формировании списка итоговых рекомендаций те объекты, с которыми целевое взаимодействие уже было, не попадают в него. Это аналогичным образом необходимо учитывать в предложенной архитектуре из независимых классификаторов, а именно на этапе формирования множества пользователей для обучения и тестирования: выборки для каждой модели должны быть отфильтрованы так, чтобы в них попадали только пользователи, у которых нет взаимодействий с соответствующим объектом за определенный промежуток времени  $T_1$ .

При работе с моделями на колаборативных данных одним из немаловажных аспектов является также проблема холодного старта. В рассматриваемой задаче на каждый месяц приходится не более 5% новых пользователей, для которых имеется ряд контентных характеристик. Не будем исключать их из выборки, а признаки, рассчитанные на основе колаборативных, просто заполним нулями.

Еще одна важная особенность, связанная со временем – период созревания целевой переменной после предоставления новых рекомендаций пользователям. Для надежной оценки качества между начальной и конечной датами тестирования должно пройти не менее  $T_2$  времени, чтобы пользователи успели отреагировать. Также при выборе  $T_2$  можно отталкиваться и от цели использования системы: на каком горизонте нужно спрогнозировать поведение пользователей – коротком или более долгосрочном.

С учетом вышеперечисленных требований можно определить схему формирования тренировочной и тестовой выборок (рис. 1):  $T_1$  – промежуток времени для взятия признакового пространства,  $T_2$  – период, за который необходимо рассматривать целевое событие для достоверной оценки работы модели. Другими словами, будем считать, что взаимодействие произошло, если пользователь в течение промежутка  $T_2$  хотя бы раз взаимодействовал с соответствующим объектом.

Возвращаясь к архитектуре системы, можно формализовать понятие целевой переменной. Она представляет собой бинарный признак, который равен 1, если пользователь взаимодействовал с соответствующим объектом, и 0 – если нет, взаимодействия рассматриваются на временном промежутке  $T_2$ . Далее необходимо собрать все предсказания воедино. Здесь важно учесть, что значения, полученные от каждой модели, не являются чистыми вероятностями склонности к тому или иному объекту. Это лишь степень уверенности модели в том, что конкретная запись принадлежит положительному классу. Поэтому чтобы можно было сравнивать



Рис. 1. Формирование обучающих и тестовых наборов данных во времени.

предсказания отдельных моделей друг с другом (это понадобится для составления итогового ранжированного списка рекомендаций), необходимо дополнительно провести калибровку для всех классификаторов. Главные условия, которые должны при этом соблюдаться, – отсутствие переранжирования между преобразованными вероятностями и исходными значениями, а также соответствие среднего значения целевой переменной итоговым вероятностям [8].

**2. Методы обучения: бустинговые алгоритмы.** После того, как схема рекомендательной системы зафиксирована, необходимо выбрать наиболее подходящие алгоритмы для каждой ее отдельной части. В первую очередь происходит обучение классификаторов, и при выборе модели важно учитывать тот факт, что размеры датасетов невелики (порядка нескольких сотен тысяч записей). Интерпретировать это можно следующим образом: чем популярнее объект, тем больше им пользуются и тем меньшему количеству пользователей его можно рекомендовать. Соответственно в обучающую выборку модели для такого объекта попадет мало пользователей. С одной стороны, это можно было бы решить, взяв больший промежуток  $T_1$  для обучения. Но это может повлечь ряд проблем, первая из которых – не совсем актуальные рекомендации, так как модель обучится на старых данных. Также может уменьшиться число уникальных пользователей, так как по определению необходимо обучать классификаторы только на тех, кто не взаимодействовал с соответствующим объектом на  $T_1$ ; соответственно увеличится количество записей для одних и тех же пользователей, что повышает риск переобучения. И желательно, чтобы все классификаторы, даже те, которые не для популярных объектов, были обучены на примерно одинаковых по размеру выборках, чтобы их результаты можно было сравнить друг с другом.

Все алгоритмы искусственного интеллекта верхнеуровнево можно разделить на две группы: классические (машинного обучения) и с применением нейронных сетей (глубокого обучения). В условиях малых датасетов последние демонстрируют менее эффективные результаты, это обусловлено требованием большого объема данных, необходимых для обучения нейросетей, поскольку их сложная архитектура и значительное количество параметров увеличивают риск переобучения на небольших выборках. Классические методы показывают более стабильные и точные результаты в таких условиях [9, 10]. Более того, для работы с табличными структурированными данными такие алгоритмы в принципе более эффективны [11, 12].

Далее при выборе конкретной модели необходимо учитывать, что доступных признаков порядка 500 единиц, среди которых встречаются категориальные и со сложной структурой. При работе с табличными данными, содержащими большое количество сложных признаков, эффективнее использовать нелинейные алгоритмы. Лучше всего с этой задачей справляется градиентный бустинг. Более того, существует ряд современных библиотек, которые позволяют автоматически решать многие проблемы – заполнение пропусков, обработка категориальных признаков, экстраполяции предсказаний на диапазонах данных вне обучающей выборки [13, 14].

Классический градиентный бустинг представляет собой ансамблевую технику, которая объединяет слабые модели, а именно деревья решений для последовательного улучшения предсказаний путем минимизации функции потерь. Модель дерева решений представляет собой бинарное дерево: вершине на каждом уровне соответствует один из признаков данных (без повторений по уровням), разбиение на двух потомков происходит по некоторому значению порога, которое подбирается путем перебора. Последовательность признаков также не случайна, они располагаются согласно критериям информативности, например по дисперсии выборки для задачи регрессии или по формуле кроссэнтропии для классификации. Основная идея бустинга состоит в пошаговом обучении деревьев таким образом, чтобы каждое следующее компенсировало ошибки предыдущей. Так, итоговая модель обновляется на каждом шаге с учетом градиента функции потерь, что позволяет постепенно улучшать предсказания.

Основные шаги метода следующие:

1. Инициализация. На этом этапе подбирается некая функция  $a_0(x)$ :  $X \rightarrow \mathbb{R}$ , где  $X$  – признаковое пространство, состоящее из действительных чисел, размерность которого равна количеству признаков (столбцов) в данных. Обычно эта функция представляет собой среднее значение целевой переменной для задачи регрессии или логарифм отношения шансов для задачи классификации. На каждом следующем шаге также будет подбираться некоторая функция уже путем решения оптимизационной задачи, а итоговая модель будет являться их композицией. В последующем описании каждая отдельная подбираемая функция именуется “базовый алгоритм”, а композиция – просто “алгоритм” с указанием номера шага. В этих терминах на этапе инициализации создается базовый алгоритм  $a_0(x)$ . Соответственно алгоритмом на  $m$ -м шаге будет  $a_m(x)$ , где  $x$  – произвольный вектор из признакового пространства  $X$ ; в обучающей выборке содержится  $n$  таких векторов  $x_1, \dots, x_n$ .

2. Вычисление вектора сдвигов. Каждый следующий базовый алгоритм строится таким образом, чтобы уменьшить суммарную ошибку уже имеющейся системы. Для этого на каждом шаге необходимо найти оптимальный вектор значений, на которые нужно сдвинуть предсказания. Другими словами, на некотором  $m$ -м шаге решается следующая задача:

$$\sum_{i=1}^n L(y_i, a_{m-1}(x_i) + \gamma_i) \rightarrow \min_{\gamma} \frac{dy}{dx},$$

где  $L$  – функция потерь,  $y_i$  – истинные значения целевой переменной,  $n$  – количество наблюдений в обучающей выборке, а  $\gamma_i$ ,  $i = 1, n$ , – компоненты вектора сдвига  $\gamma$ , его можно вычислить следующим образом:

$$\gamma = -\nabla F_m = (s_1, \dots, s_n)^T, s_i = -\frac{\partial L(y_i, z_i)}{\partial z_i}, z_i = a_{m-1}(x_i),$$

$$F_m(\gamma) = \sum_{i=1}^n L(y_i, a_{m-1}(x_i) + \gamma_i).$$

3. Построение нового дерева. Здесь  $m$ -й базовый алгоритм фактически предсказывает производные функции потерь с отрицательным знаком, вычисленные в точках, соответствующих ответам композиции на обучающей выборке на  $(m-1)$ -м шаге построения алгоритма:

$$f_m(x) = \operatorname{argmin}_f \frac{1}{n} \sum_{i=1}^n (f(x_i) - \gamma_i)^2.$$

Информация об исходной функции потерь  $L(y, z)$  уже содержится в выражении для вектора оптимального сдвига  $\gamma$ . В ходе оптимизации подбирается такой базовый алгоритм  $f$  решаящего дерева, который минимизирует суммарное значение ошибки по всем объектам обучающей выборки  $x_1, \dots, x_n$ .

4. Обновление модели. Построенное на предыдущем шаге дерево добавляется в итоговую модель с некоторым коэффициентом  $\eta$ :

$$a_m(x) = a_{m-1}(x) + \eta f_m(x),$$

где  $\eta$  – коэффициент обучения, контролирующий вклад каждого нового дерева в общую модель.

5. Повторение шагов 2–4. Процесс повторяется до тех пор, пока не будет достигнуто заданное число итераций или пока улучшения в модели на каждом шаге не станут незначительными.

На текущий момент существует множество модификаций градиентного бустинга, самые известные представлены в библиотеках Light Gradient Boosting Machine (LightGBM) [15], Extreme Gradient Boosting (XGBoost) [16] и Categorical Boosting (CatBoost) [17]. Рассмотрим каждый из них подробнее.

2.1. L i g h t G B M – это эффективная реализация градиентного бустинга, разработанная для повышения скорости обучения и уменьшения потребления памяти при работе с большими объемами данных [18]. В его основе лежат несколько ключевых идей оптимизации.

1. LightGBM разбивает непрерывные признаки по бинам, образуя гистограммы. По ним алгоритм ищет оптимальные точки разбиения, что позволяет прежде всего снизить вычислительную сложность, так как тогда необходимо обработать ограниченный набор дискретных значений бинов. Данное упрощение также помогает задействовать меньше памяти. Формально это можно описать следующим образом: сначала непрерывный признак  $l$  разбивается на  $k$  бинов:

$$l \rightarrow \{b_1, b_2, \dots, b_k\}.$$

Затем для каждого бина  $b_j$  вычисляется гистограмма:

$$B_j = \sum_{i \in b_j} g_i, g_i = \frac{\partial L(y_i, z_i)}{\partial z_i},$$

здесь  $g_i$  – градиенты для  $i$ -го объекта обучающей выборки, в  $B_j$  суммируются только попавшие в бин  $b_j$ . Далее оптимальное разбиение находится путем поиска максимального прироста информации:

$$Gain(s) = \frac{1}{2} \left( \frac{G_L^2(s)}{H_L(s) + \lambda} + \frac{G_R^2(s)}{H_R(s) + \lambda} - \frac{(G_L(s) + G_R(s))^2}{H_L(s) + H_R(s) + \lambda} \right) - \alpha,$$

$$G_L = \sum_{i \in D_L} g_i, G_R = \sum_{i \in D_R} g_i, g_i = \frac{\partial L(y_i, z_i)}{\partial z_i},$$

$$H_L = \sum_{i \in D_L} h_i, H_R = \sum_{i \in D_R} h_i, h_i = \frac{\partial L^2(y_i, z_i)}{\partial z_i^2},$$

где  $D_L$  и  $D_R$  – подмножества данных, попадающих в левое и правое поддеревья после разбиения. Таким образом,  $G_L$  и  $G_R$  можно трактовать как соответствующие суммы элементов вектора градиента функции потерь, а  $H_L$  и  $H_R$  – суммы диагональных элементов матрицы Гессе функции потерь в левых и правых поддеревьях,  $\lambda$  – коэффициент  $L2$ -регуляризации,  $\alpha$  – параметр регуляризации, контролирующий минимальный прирост информации необходимый для осуществления разбиения.

2. В обычном градиентном бустинге используется подход level-wise, при котором дерево растет по уровням, т.е. все листья на текущем уровне добавляются одновременно. LightGBM же применяет метод leaf-wise, при котором дерево растет по листьям: на каждом шаге построения дерева выбирается лист с наибольшей ошибкой для разбиения. Такой подход более эффективен для поиска сложных зависимостей, но есть риск, что дерево будет расти в глубину только по одной ветке, что повлечет переобучение. Такая проблема распространена при обучении на малых объемах данных [19].

3. LightGBM поддерживает параллельное обучение, может эффективно использовать многопроцессорные системы и распределенные вычислительные кластеры, что ускоряет его работу, особенно на больших объемах данных.

2.2. X G B o o s t – это усовершенствованный алгоритм градиентного бустинга, в котором применяется оптимизация второго порядка и два типа регуляризации [20].

1. XGBoost использует вторые производные функции потерь для более точного подбора оптимального решения. Формально нужно подобрать алгоритм  $f(x)$  – скалярную функцию, решающую следующую задачу:

$$\sum_{i=1}^n L(y_i, a_{m-1}(x_i) + f(x_i)) \rightarrow \min_f.$$

Для этого функция  $L$  раскладывается в ряд Тейлора до второго члена с центром в ответе композиции  $a_{m-1}$ :

$$L(y_i, a_{m-1}(x_i) + f(x_i)) \approx L(y_i, a_{m-1}(x_i)) + s_i f(x_i) + \frac{1}{2} h_i f^2(x_i),$$

$$s_i = \frac{\partial L(y_i, z_i)}{\partial z_i}, h_i = \frac{\partial L^2(y_i, z_i)}{\partial z_i^2}, z_i = a_{m-1}(x_i), \quad i = \overline{1, n}.$$

Первое слагаемое не зависит от нового базового алгоритма, поэтому можно его не учитывать.

2. XGBoost включает регуляризационные члены, которые предотвращают численные проблемы, связанные с малым значением второй производной, а также добавляют штрафы за количество листьев и норму коэффициентов, помогая избежать переобучения:

$$\Omega = \beta T + \frac{1}{2} \delta \sum_{j=1}^T w_j^2.$$

Здесь  $T$  – количество листьев в дереве,  $w_j$  – вес узла,  $\beta$  и  $\delta$  – гиперпараметры регуляризации, задаются вручную или путем экспериментального подбора в зависимости от конкретной задачи. Полная функция потерь с регуляризацией и отбрасыванием первого слагаемого имеет вид:

$$\sum_{i=1}^n \left[ s_i f(x_i) + \frac{1}{2} h_i f^2(x_i) \right] + \beta T + \frac{1}{2} \delta \sum_{j=1}^T w_j^2.$$

3. При построении дерева XGBoost использует, как и LightGBM, критерий информативности, зависящий от оптимального вектора сдвига:

$$Gain(s) = \frac{1}{2} \left( \frac{G^2(s)}{H(s) + \lambda} - \frac{(G_L^2(s) + G_R^2(s))}{H_L(s) + H_R(s) + \lambda} \right) - \alpha.$$

4. Критерий останова в XGBoost также зависит от оптимального сдвига: дерево перестает расти, когда дальнейшее разбиение не приводит к значительному улучшению критерия информативности.

2.3. С а t B o o s t – библиотека градиентного бустинга, которая специально оптимизирована для работы с категориальными признаками [21].

1. Обычный бустинг требует предварительной обработки категориальных признаков, например с помощью one-hot encoding. Этот метод не создает ложного ранжирования между категориями признака, но может значительно увеличить размерность данных. Существует альтернативный алгоритм кодирования target encoding, который отражает влияние отдельных категорий на целевую переменную [22]. Этот достаточно эффективный метод автоматически применяется в CatBoost:

$$Encoded\_Feature(x_i) = \sum_{j \neq i} I(x_j = x_i) y_j / \sum_{j \neq i} I(x_j = x_i),$$

где  $y_j$  – целевая переменная,  $I(x_j = x_i)$  – индикатор функции, равный 1, если  $x_j = x_i$ . Также можно использовать схему порядковых кодов: она основывается на порядке появления данных. Подобный подход уменьшает риск утечки информации и снижает риск переобучения. Формула порядкового кодирования выглядит следующим образом:

$$Ordered\_Target\_Mean(x_i) = \sum_{j < i} I(x_j = x_i) y_j / \sum_{j < i} I(x_j = x_i).$$

2. CatBoost применяет новую метрику совокупного убывания (ordered boosting), которая снижает риск переобучения и позволяет создать более устойчивую модель. Идея заключается в применении градиентного бустинга на случайных подвыборках данных в каждой итерации:

$$L(y, \hat{y}) = \sum_{i=1}^n l(y_i, \hat{y}_{(i)}),$$

где  $\hat{y}_{(i)}$  – предсказание для  $i$ -го примера, основанное на модели, обученной на примерах до  $i$ .

3. CatBoost поддерживает также многомерный градиентный бустинг, что позволяет одновременно решать несколько задач регрессии или классификации.

Таким образом, LightGBM рекомендуется использовать в задачах, где важна скорость обучения и низкое потребление памяти. XGBoost наиболее эффективен там, где нужна высокая точность и стабильность модели, особенно на больших объемах данных. CatBoost идеально подходит для датасетов с категориальными признаками благодаря встроенным механизмам кодирования, а также в условиях ограниченного количества информации (табл. 1, рис. 2).

**Таблица 1.** Сравнение моделей LightGBM, XGBoost, CatBoost

Характеристика	LightGBM	XGBoost	CatBoost
Разработчик	Microsoft	DMLC	Yandex
Год выпуска	2016	2014	2017
Базовые деревья	Несимметричные, рост по листьям	Несимметричные, рост по уровням	Симметричные
Метод разбиения	GOSS	Гистограммный на отсортированных данных	Жадный
Поддержка категориальных признаков	Может интерпретировать порядковые признаки	Необходима предобработка вручную	Может обрабатывать самостоятельно
Текстовые неструктурированные признаки	Не поддерживаются	Не поддерживаются	Несколько встроенных методов, включая Наивного Байеса и мешок слов

Окончание таблицы 1

Борьба с дисбалансом классов	Встроенные методы Weighted Loss и Focal Loss	Встроенные методы Weighted Loss и Focal Loss	Автоматическая балансировка классов и Balanced Cross-Entropy
Скорость обучения	Высокая благодаря росту по листьям и большому числу оптимизаций на разных шагах алгоритма	Медленнее, чем LGBM, но может использовать оптимизации column block и sparsity-aware split finding	Высокая скорость благодаря поддержке категориальных данных и упрощенному перебору гиперпараметров

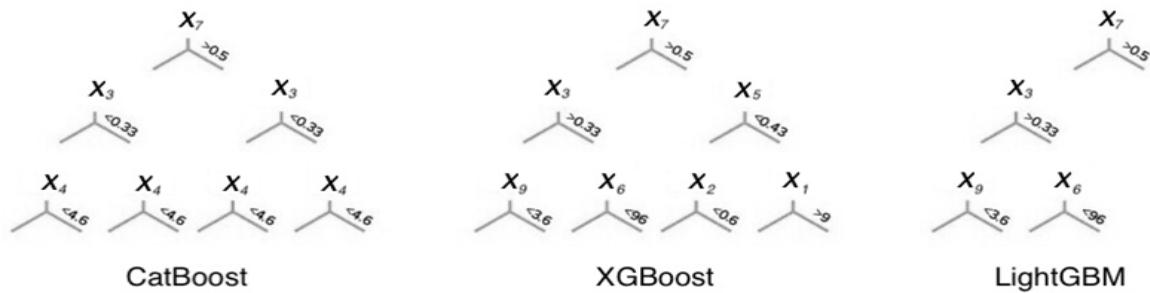


Рис. 2. Сравнение базовых деревьев LightGBM, XGBoost, CatBoost.

Возвращаясь к особенностям решаемой задачи с учетом сравнения алгоритмов бустинга, наиболее подходящим для моделирования в данной задаче будет CatBoost, так как наборы данных для каждого классификатора небольшие, скорость обучения моделей не столь важна, качество стоит на первом месте, в датасете достаточно много категориальных признаков. Таким образом, для обучения каскада моделей классификации в дальнейшем будет использоваться CatBoost.

**3. Методы калибровки.** Калибровка моделей машинного обучения важна для получения точных вероятностных предсказаний. Числа, которые выдают большинство моделей классификации, CatBoost в частности, отождествляются со степенью уверенности алгоритма в принадлежности того или иного объекта какому-либо классу. Чтобы привести их к чистым интерпретируемым вероятностям, нужна калибровка. В данной задаче это особенно необходимо, так как для формирования списка рекомендаций для каждого пользователя необходимо будет сравнивать вероятности от каждой модели друг с другом.

Цель процедуры калибровки – преобразовать выходные числа модели так, чтобы они соответствовали среднему значению целевой переменной. Для этого необходимо построить калибровочную кривую: разбить отрезок  $[0, 1]$  на части (бины) и в каждой рассчитать среднее значение таргета. Полученная кривая в координатах “уверенности модели – калиброванные вероятности” аппроксимируется некоторой функцией (рис. 3), которая затем применяется к значениям уверенности модели; ее результат – искомые вероятности. Остается только выбрать метод аппроксимации и количество частей, на которые следует разбить диапазон  $[0, 1]$ . Более того, важно учитывать следующие требования [23, 24]:

1) функция должна быть монотонной, чтобы не было переранжирования вероятностей после калибровки;

2) откалиброванные вероятности должны соответствовать среднему значению целевой переменной внутри бинов;

3) желательно, чтобы после калибровки распределение вероятностей не приобретало разрывный вид;

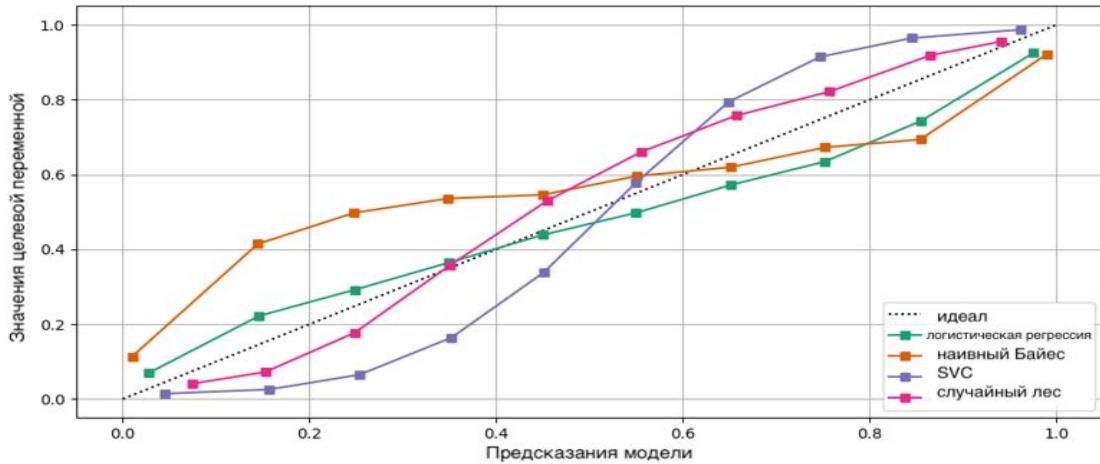


Рис. 3. Примеры калибровочных функций.

4) внутри каждого бина также важно учитывать индивидуальный рост среднего значения целевой переменной.

Если разбить отрезок  $[0, 1]$  на несколько частей  $b_i, i = \overline{1, n}$  (обычно берется  $n = 10$ ), в каждом посчитать усредненную частоту положительного класса  $target\_mean_i$  и среднее значение уверенности модели  $conf\_mean_i$ , то можно вычислить ошибку калибровки ECE (expected calibration error):

$$target_{mean_i} = \frac{1}{|b_i|} \sum_{j \in b_i} y_j,$$

$$conf\_mean_i = \frac{1}{|b_i|} \sum_{j \in b_i} p_j,$$

$$ECE = \sum_{i=1}^n \frac{|b_i|}{N} |target_{mean_i} - conf_{mean_i}|,$$

где  $y_j$  — истинные значения целевой переменной,  $p_j$  — предсказания модели,  $N$  — размерность данных. Рассмотрим основные алгоритмы калибровки.

3.1. И з о т о н и ч е с к а я р е г р е с с и я. Это непараметрический метод, который особенно эффективен, когда предсказания модели показывают сильные отклонения и не следуют линейной зависимости. Калибровочная кривая подбирается как кусочно-постоянная неубывающая функция (рис. 4). На вход алгоритму подаются исходные предсказания, в качестве целевой переменной выступает исходная, а в ходе обучения подбираются интервалы постоянства и значения функции на них путем минимизации среднеквадратичной ошибки [25, 26].

Формализуем подход: пусть имеется набор данных  $(x_i, y_i)$ , где  $i = 1, n$ ,  $x_i$  — предсказанные значения, а  $y_i$  — истинные метки классов. Изотоническая регрессия ищет монотонно неубывающую функцию  $f$ , такую, что

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_n).$$

Она минимизирует следующую функцию потерь:

$$\sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \min.$$

Наиболее часто используемый алгоритм для решения задачи изотонической регрессии — PAVA (pool adjacent violators algorithm). Его принцип работы заключается в следующем:

- 1) инициализация: сначала предполагается, что  $f(x_i) = y_i, i = 1, n$ ;
- 2) проверяются пары смежных значений  $y_i$  и  $y_{i+1}$ : если обнаружено, что  $y_i > y_{i+1}$ , эти значения объединяются в один отрезок и заменяются средним:

$$y_{new} = \frac{y_i n_i + y_{i+1} n_{i+1}}{n_i + n_{i+1}},$$

где  $n_i$  и  $n_{i+1}$  — количество элементов в бинах до объединения;

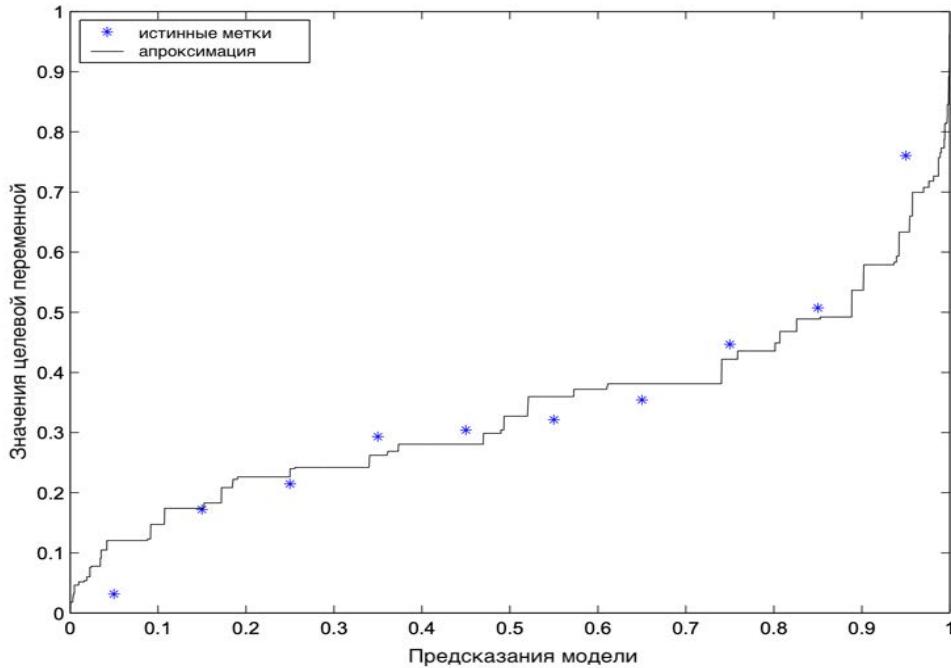


Рис. 4. Пример калибровки изотонической регрессией.

3) процесс повторяется до тех пор, пока все значения не будут следовать монотонно неубывающей последовательности.

Главные преимущества изотонической регрессии в том, что не нужно никаких предположений о распределении данных, алгоритм хорошо работает со сложными нелинейными зависимостями. Основные недостатки – склонность к переобучению, чувствительность к выбросам, нарушение ранжирования внутри бинов и необходимость большого объема данных для стабильной оценки.

3.2. Калибровка Плэтта. Данный метод был предложен Джоном Плэттом для калибровки SVM (support vector machine) [27]. Идея заключается в применении логистической регрессии к выходам модели:

$$P(y = 1|x) = \frac{1}{1 + \exp(Af(x) + B)},$$

где  $A$  и  $B$  – параметры, которые необходимо подобрать через максимизацию логарифмического правдоподобия:

$$\min_{A, B} \sum_{i=1}^n \left[ y_i \log(P(y_i = 1 \# f(x_i))) + (1 - y_i) \log(1 - P(y_i = 1 \# f(x_i))) \right],$$

$$P(y_i = 1 \# f(x_i)) = \frac{1}{1 + \exp(Af(x_i) + B)}.$$

Далее можно подставить в формулу найденные коэффициенты и использовать ее для преобразования оценок [28]:

$$P(y = 1|x) = \frac{1}{1 + \exp(Af(x) + B)}.$$

Данный метод популярен из-за его простоты и эффективности для калибровочных функций, которые хорошо описываются сигмоидой (рис. 5). Это одновременно и недостаток, так как для сложных кривых он плохо применим. Также стоит отметить, что для получения качественных результатов необходимо значительное количество данных.

3.3. Температурный шкалирование. Этот метод основывается на попытке аппроксимировать калибровочную кривую сигмоидальной функцией, только логиты – действительные числа или вектора, которые формирует модель, – предварительно делятся на

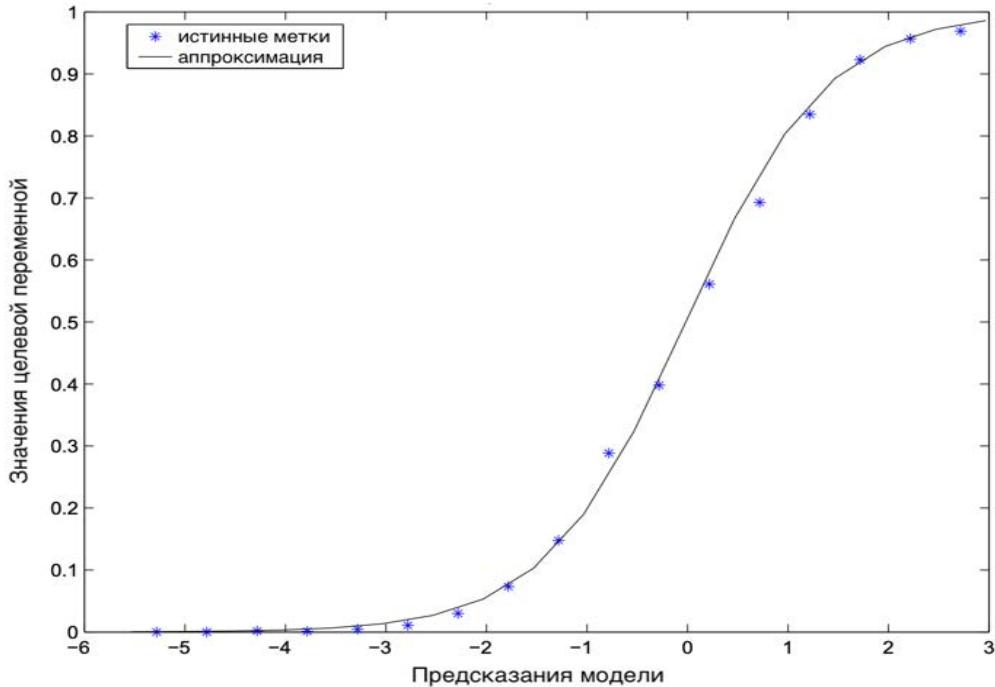


Рис. 5. Пример калибровки Плэтта.

общий коэффициент температуры  $T$  [29]. Он подбирается так, чтобы минимизировать логарифмическую функцию потерь на валидационном наборе данных:

$$\min_T -\frac{1}{n} \sum_{i=1}^n \left( \hat{y}_i \log(p_i) + (1 - \hat{y}_i) \log(1 - p_i) \right),$$

где  $\hat{y}_i$  – истинные значения целевой переменной,  $\hat{p}_i$  – предсказанные моделью значения.

Далее вычисляются корректированные логиты  $z'$  путем деления исходных логитов  $z$  на значение температуры  $T$ , и с их помощью рассчитываются откалиброванные вероятности:

$$z' = \frac{z}{T}, \quad \hat{p}_i = \exp\left(\frac{z_i}{T}\right) / \sum_{j=1}^n \exp\left(\frac{z_j}{T}\right).$$

Алгоритм температурного шкалирования часто применяется для современных глубоких нейронных сетей, он прост в использовании и мало переобучается, так как зависит только от одного параметра – температуры. Но опять же, он качественно аппроксимирует только калибровочные кривые, которые близки по виду к сигмоиде или softmax (рис. 6).

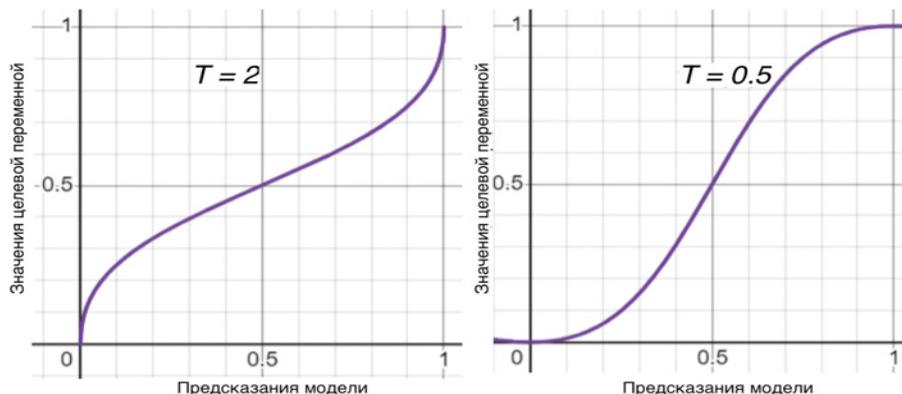


Рис. 6. Примеры калибровки температурным шкалированием.

3.4. Гистограммная калибровка. Данный метод отчасти похож на изотоническую регрессию: отрезок  $[0, 1]$  разбивается на части, внутри каждой из которых происходит аппроксимация калибровочной кривой. Отличие в том, что количество и длина этих подотрезков задаются вручную. Можно применять различные модификации для подбора функции на каждом из них, решая таким образом важные проблемы, которые свойственны классическим подходам.

Первым делом необходимо разбить отрезок  $[0, 1]$  на бины  $b_1, \dots, b_n$ : можно сделать это так, чтобы они были одинаковой длины или же одинаковые по количеству наблюдений (равномощные) [30]. Простейшая постановка задачи — это подобрать для каждого бина некое число  $\theta_i$ , которое и будет искомой вероятностью. Подбираются они так, чтобы  $\theta_i, i = 1, \dots, n$ , как можно лучше приближали средние метки классов на соответствующих подотрезках:

$$\sum_{i=1}^n \left| \sum_{j=1}^N I\{q(x_j) \in b_i\} y_i / |b_i| - \theta_i \right| \rightarrow \min_{(\theta_1, \dots, \theta_n)},$$

где  $N$  — размерность данных.

Недостаток такого подхода в том, что в каждом бине предсказания аппроксимируются одним числом, из-за чего теряется их внутреннее ранжирование. Это можно исправить, подбрав не просто числа  $\theta_i$ , а также коэффициенты наклона  $k_i$ . Таким образом, внутри каждого подотрезка приближение будет не константной функцией, а линейной (рис. 7):

$$\frac{1}{|b_i|} \sum_{n \in b_i} y_n = k_i \frac{1}{|b_i|} \sum_{n \in b_i} q(x_n) \Rightarrow k_i = \frac{\text{mean}_{n \in b_i}(y_n)}{\text{mean}_{n \in b_i}(q(x_n))}.$$

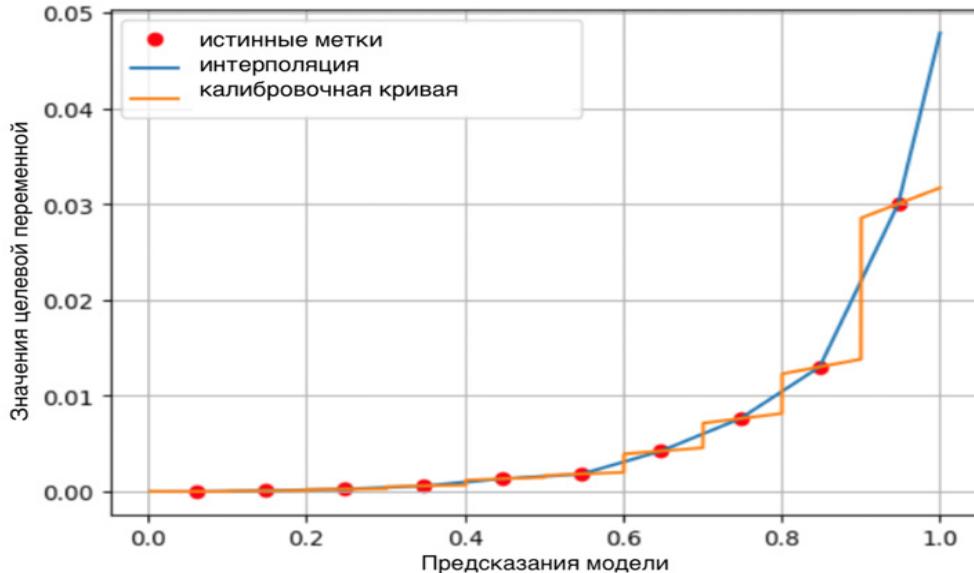


Рис. 7. Пример гистограммной калибровки с одним коэффициентом.

Но тогда может появиться еще одна весомая проблема: нарушения общей монотонности функции, особенно на участках медленного роста калибровочной кривой (рис. 8). Это происходит из-за того, что найденные для каждого бина линейные зависимости обязательно проходят через точку  $(0, 0)$ . Устранить данный недостаток можно простым добавлением коэффициента смещения  $b$  в уравнение (рис. 9). Тогда при поиске оптимальной прямой необходимо руководствоваться также условием, что общая функция калибровки должна быть неубывающей. Упрощенно можно было бы решать эту задачу как построение линейной регрессии на каждом из участков [31].

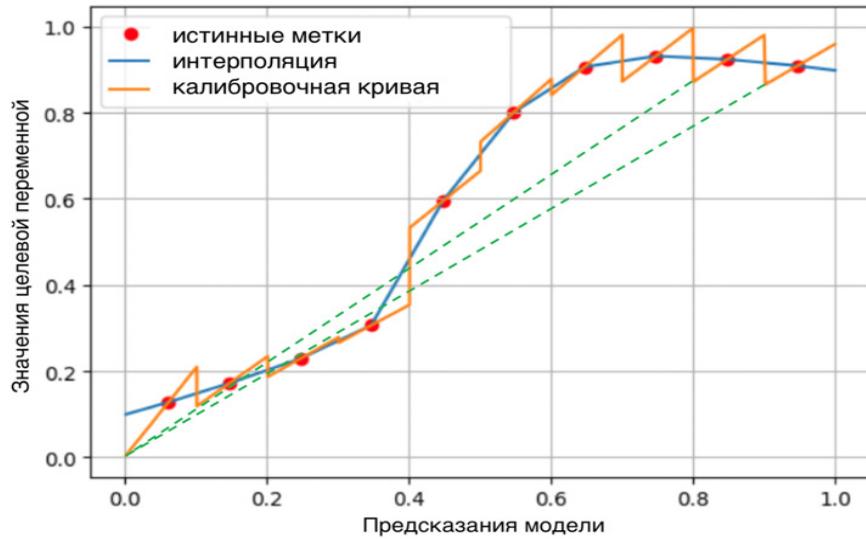


Рис. 8. Пример нарушения ранжирования при гистограммной калибровке с одним коэффициентом.

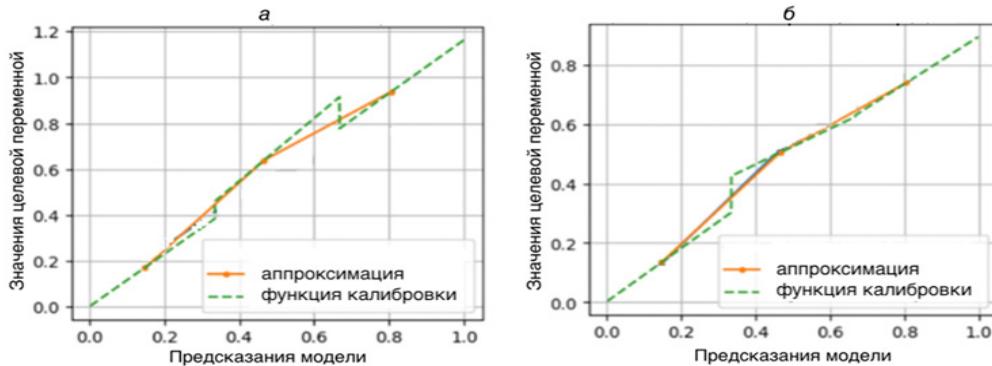


Рис. 9. Сравнение гистограммной калибровки с одним (а) и двумя (б) коэффициентами в критическом случае.

Рассмотрим пример поиска такой калибровочной кривой для случая деления  $[0, 1]$  на три части. Формализуем каждое из условий для нахождения решения.

1. Средняя преобразованная вероятность должна быть равна среднему значению целевой переменной в каждом бине:

$$\begin{cases} k_1 s_1 + b_1 = t_1, \\ k_2 s_2 + b_2 = t_2, \\ k_3 s_3 + b_3 = t_3. \end{cases}$$

2. Общая функция калибровки должна быть неубывающей и кусочно непрерывной на  $[0, 1]$ :

$$\begin{cases} 0 \leq k_1 l_1 + b_1, \\ k_1 r_1 + b_1 = k_2 r_2 + b_2, \\ k_2 r_2 + b_2 = k_3 r_3 + b_3, \\ k_3 r_3 + b_3 \leq 1, \\ k_1, k_2, k_3 > 0. \end{cases}$$

В уравнениях выше  $s_i$  — средние значения исходных предсказаний модели в  $i$ -м подотрезке,  $t_i$  — средние значения целевой переменной,  $l_i$  и  $r_i$  — левая и правая границы исходных предсказаний.

Таким образом, количество неизвестных в первой системе равно шести, а уравнений всего три. В совокупности с требованиями второй системы возможных решений будет либо бесконечно много, либо вообще не будет (в случае, если средние значения целевой переменной убывают по бинам “слева направо”). Если посмотреть на задачу геометрически, становится понятно, что чем меньше коэффициент  $k$ , тем более сжатыми будут распределения вероятностей в каждом бине. Это нежелательно, так как тогда преобразованные предсказания разделятся на дискретные части; лучше, если они будут образовывать равномерное распределение.

Поэтому дополнительно потребуем, чтобы ширина преобразованных вероятностей была максимальна:  $\Delta y = k\Delta x \rightarrow \max$ ; технически это означает максимизацию коэффициента  $k$ , а также чтобы расстояния между соответствующими крайними точками в соседних бинах были минимальными [32].

Самый тривиальный способ решения такой задачи — одновременно минимизировать два выражения: разницу между  $k_i s_i + b_i$  и  $t_i$  (например, по формуле среднеквадратичной ошибки) и средний квадрат расстояний между бинами; оставшиеся неравенства из систем выше будут являться ограничениями. На практике такой способ подбирает коэффициенты с достаточно малыми расстояниями между бинами, но ошибка калибровки получается больше, так как минимизируется разница между средним таргетом и средней вероятностью, их равенства не требуется.

Альтернативный вариант — использовать следующий алгоритм.

1. Выразим коэффициенты  $b_i$  из уравнений первой системы, тогда искомые калибровочные функции для каждого бина преобразуются в  $y_i(x) = k_i(x - s_i) + t_i$ .

2. Для каждого бина находим потенциальных кандидатов на роль коэффициента  $k_i : k_l = 0.5(t_{i-1} + t_i)$  и  $k_r = 0.5(t_i + t_{i+1})$ . Геометрически эти условия означают поиск середины между средними значениями целевых переменных в соседних бинах.

3. Далее выбираем из кандидатов один итоговый коэффициент как  $k_i = \min(k_l, k_r)$ . Это условие гарантирует, что преобразованные вероятности в соседних бинах не будут пересекаться.

4. На данном этапе между заполненными диапазонами могут быть значительные расстояния. Чтобы это исправить, необходимо увеличивать  $k_i$ , пока крайние значения вероятностей не начнут пересекаться с ближайшим соседом. Единственное, что могут быть ограничения на расширение бинов из-за соседних, а также из-за того, что нельзя выходить за пределы  $[0, 1]$ .

Преимущество этого подхода в том, что ошибка калибровки в нем отсутствует. Но могут быть расстояния между крайними значениями вероятностей в соседних бинах.

Из перечисленных алгоритмов калибровка Плэтта и температурное шкалирование не будут эффективными в рассматриваемой задаче, так как не для всех классификаторов калибровочная кривая хорошо описывается сигмоидой. Ввиду широких возможностей для тонкой настройки алгоритма остановимся на методе гистограммной калибровки, а именно последнем описанном подходе: с подбором двух коэффициентов прямой для каждого диапазона и максимизацией  $k$  наряду с минимизацией расстояния между соседними бинами. По сравнению с простейшей реализацией этого алгоритма и изотонической регрессией, модифицированный подход не так сильно переобучается и решает проблему потери ранжирования между предсказаниями внутри отдельных бинов, оставляя при этом распределение преобразованных вероятностей максимально равномерным (табл. 2).

Таблица 2. Сравнение различных алгоритмов калибровки

Алгоритм	Преимущества	Недостатки
Изотоническая регрессия	Достаточно точно приближает калибровочную кривую любой формы	Нарушается ранжирование внутри участков постоянства, склонность к переобучению
Калибровка Плэтта	Подбирается только два коэффициента, почти не переобучается	Эффективна только в случаях, когда калибровочная кривая хорошо описывается сигмоидой
Температурное шкалирование	Подбирается только два коэффициента, почти не переобучается, не требуется дополнительных моделей	Качественно приближает лишь ограниченный класс функций

Окончание таблицы 2

Гистограммная калибровка	Можно достаточно точно приблизить калибровочную кривую любой формы, возможности для модификации решают важные проблемы переранжирования внутри бинов, непрерывности функции, нестабильности целевой переменной	Сильно зависит от выбора разбиения по бинам, также необходимо подбирать количество коэффициентов в зависимости от сложности задачи
--------------------------	--	--

Примем во внимание еще некоторые особенности решаемой задачи и добавим к улучшенному алгоритму гистограммной калибровки несколько необходимых модификаций.

1. В дальнейшем необходимо будет сравнивать выходы моделей друг с другом. Но если их предсказания в среднем сильно отличались друг от друга до калибровки, эта же тенденция может сохраняться и после. Например, у одной модели все уверенности могут быть расположены на отрезке от 0 до 0.5, а у другой – от 0.5 до 1. Тогда при формировании итоговых рекомендаций объект, соответствующий второй модели, будет всегда выше первого (рис. 10). Это косвенно может быть связано с так называемой проблемой “популярности”: чем больше пользователей взаимодействуют с некоторым объектом, тем выше для него будет среднее значение целевой переменной и тем выше значения предсказаний как до, так и после калибровки. Исправить ситуацию можно нормализацией выходных значений модели, например привести их к логнормальному распределению:

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), x > 0.$$

Его особенность состоит в том, что логарифм от него дает нормальное распределение:  $X \sim \log N(\mu, \sigma^2) \Rightarrow \ln(X) \sim N(\mu, \sigma^2)$ . После применения такого преобразования распределения предсказаний моделей становятся более схожими, можно регулировать их также при помощи коэффициента сдвига по оси  $Ox$ .

2. Другой немаловажный момент заключается в том, что для некоторых объектов средние значения целевой переменной ведут себя нестабильно во времени. Поэтому если ориентироваться на одно число внутри каждого бина, повышается риск переобучения и снижается обобщающая

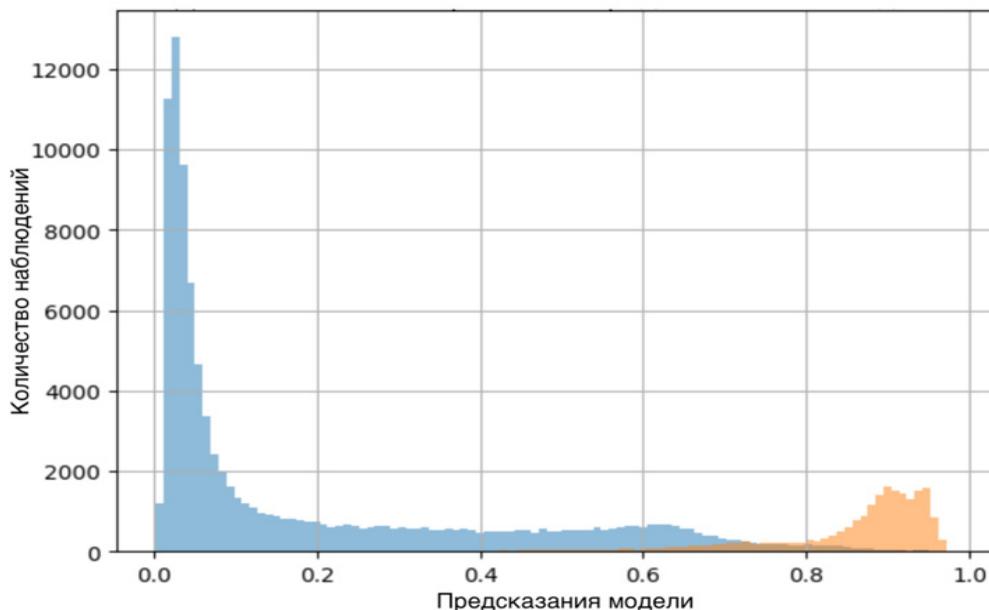


Рис. 10. Пример неравномерно распределенных предсказаний двух моделей.

способность модели. Попробовать исправить эту проблему можно через аппроксимацию кумулятивного среднего значения целевой переменной во времени. Это позволит сгладить выбросы и выявить тренд, что сделает модель более устойчивой. Зависимости целевых переменных во времени достаточно хорошо описываются степенной функцией со сдвигом вида  $(kx + b)^n$ . Другими словами, необходимо подобрать коэффициенты  $k$ ,  $b$  и показатель степени  $n$  так, чтобы итоговая функция наилучшим образом проходила через имеющиеся точки значений таргета (рис. 11). Далее коэффициент наклона для участка калибровочной кривой будет вычисляться как отношение предсказанного тренда и фактического значения целевой переменной (рис. 12):

$$\text{linear\_coef} = \frac{\text{expected\_mean\_target}}{\text{fact\_mean\_target}} = \frac{\min(1, (kx + b)^n)}{\text{fact\_mean\_target}}.$$

Теперь, когда все методы решения и особенности задачи определены, можно приступать к применению их на практике.

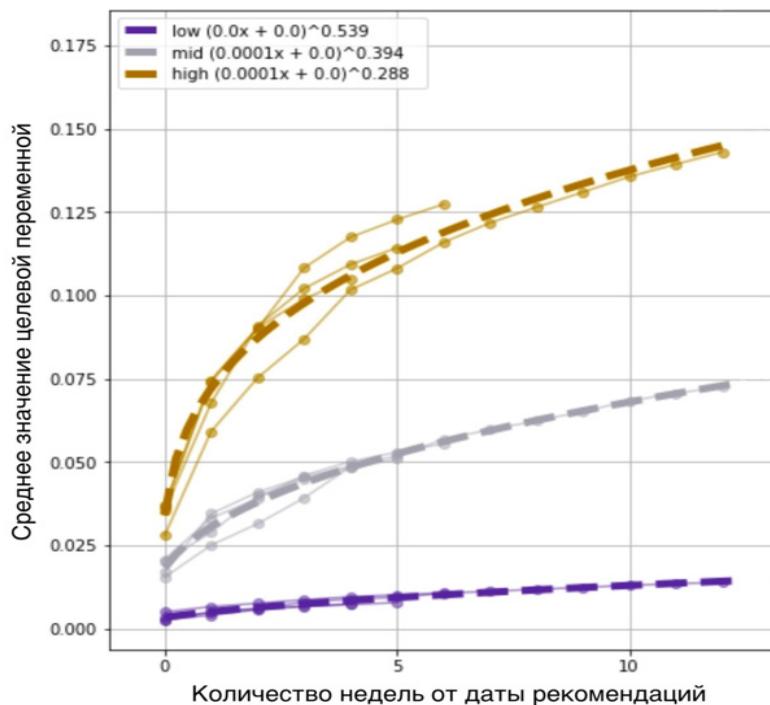


Рис. 11. Аппроксимация кумулятивной доли значений целевой переменной в разрезе групп по величине склонности.

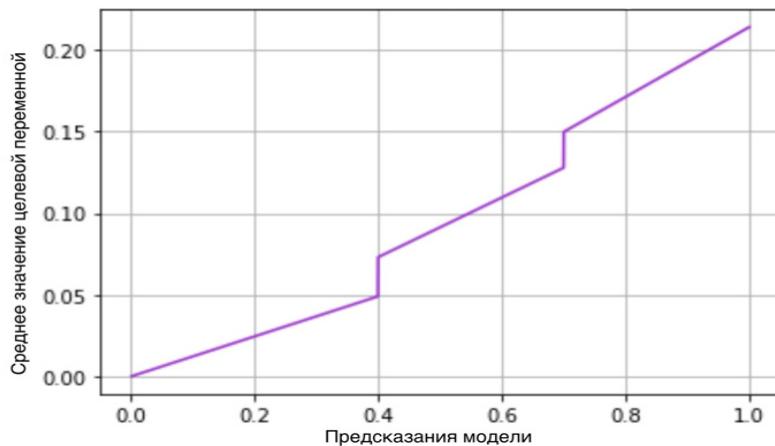


Рис. 12. Калибровочная функция для трех бинов с аппроксимацией таргета.

**4. Проведение эксперимента.** На основе исследования осталось собрать данные, обучить модели, откалибровать их и собрать в единую систему. Последовательность основных действий и результаты следующие.

1. Сбор данных. Для каждого классификатора необходимо сформировать отдельный набор данных, суммарно их получится 10 штук. В состав пользователей должны входить только те, которые не имеют соответствующего объекта на дату выдачи рекомендаций моделью: мета-информация по пользователям и объектам и агрегированные колаборативные факторы (для новых пользователей заполняются нулями). В качестве периода формирования обучающего датасета и срока готовности целевой переменной возьмем значения  $T_1 = 3$  мес.,  $T_2 = 1$  мес. (рис. 1); размер выборки для out-of-time тестирования – 1 мес. (рис. 13). Для некоторых объектов наблюдается дисбаланс в целевой переменной (рис. 14), в дальнейшем это можно исправить коэффициентами весов для классов при обучении.

2. Предобработка данных. Стандартно предлагается удалить неинформативные и уникальные признаки с высокой долей пропущенных значений (более 70%), а затем отобрать наиболее значимые по SHAP-значениям (shapley additive explanations). Последнее необходимо для снижения риска переобучения и ускорения процесса обучения [33, 34], в качестве инструмента используется встроенный алгоритм CatBoost Feature Selector [21].

3. Обучение моделей и подбор гиперпараметров. Для всех 10 экспериментальных объектов обучаются отдельные модели классификации CatBoostClassifier [21]. Для подбора гиперпараметров выделяется 30% от обучающего набора данных, для оптимизации процесса используется optuna [35]. Часть гиперпараметров фиксируется вручную (табл. 3), на них в совокупности с наилучшими подобранными значениями обучаются финальные модели (табл. 4).

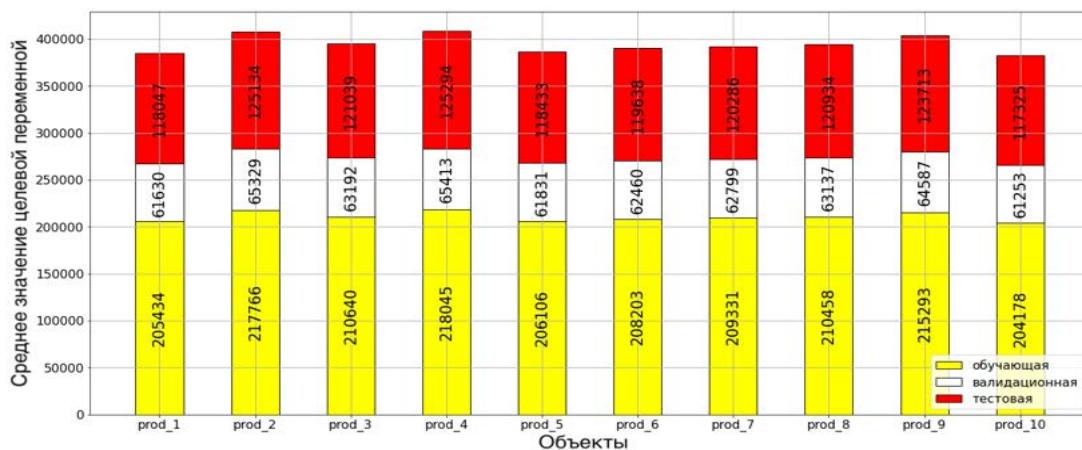


Рис. 13. Размеры выборок для классификаторов.

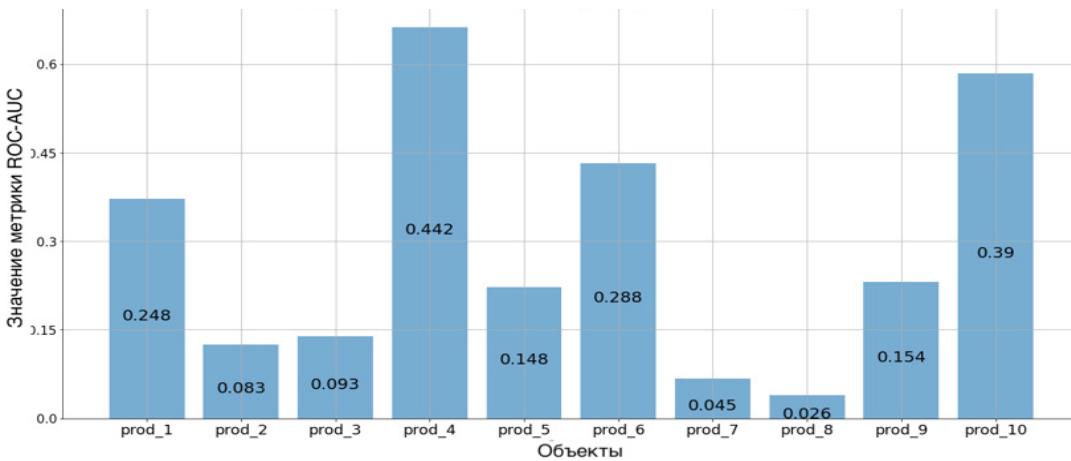


Рис. 14. Доли целевых событий в тренировочных выборках.

Таблица 3. Фиксированные гиперпараметры моделей

Гиперпараметр	Значение
iterations	1000
loss_function	‘Logloss’
one_hot_max_size	5
boosting_type	‘Ordered’
n_estimators	50
scale_pos_weight	w*

\* Задается для каждой модели вручную как отношения размера отрицательного класса к положительному.

Таблица 4. Оптимизируемые гиперпараметры моделей

Гиперпараметр	Диапазон для подбора и шаг
learning_rate	[0.001, 0.1], 0.001
depth	[2, 8], 1
l2_leaf_reg	[1, 10], 1
bagging_temperature	[0.01, 0.9], 0.05
border_count	[32, 255], 5
min_data_in_leaf	[5, 100], 5
min_child_samples	[10, 200], 1
subsample	[0.2, 1], 0.1
rsm	[0.1, 1], 0.1
random_strength	[0, 10], 0.1

4. Промежуточная оценка качества. Для полученных классификаторов можно уже рассчитать некоторые метрики: в качестве таковой рассматривается ROC-AUC (receiver operating characteristic area under curve), так как она не зависит от порога и позволяет оценить общую ранжирующую способность обученных алгоритмов (рис. 15).

5. Калибровка предсказаний. Для перехода от степени уверенности модели к интерпретируемым значениям вероятности производится калибровка улучшенным гистограммным методом, описанным ранее.

6. Составление списка рекомендаций. На данном этапе остается соединить все вероятности в единую рекомендательную систему: для каждого пользователя отсортировать все имеющиеся для него значения по убыванию, это и будет итоговый список рекомендаций. Оценка качества полученной системы проводится по метрикам Hit Rate, Precision@3 и Precision@5 (табл. 5).

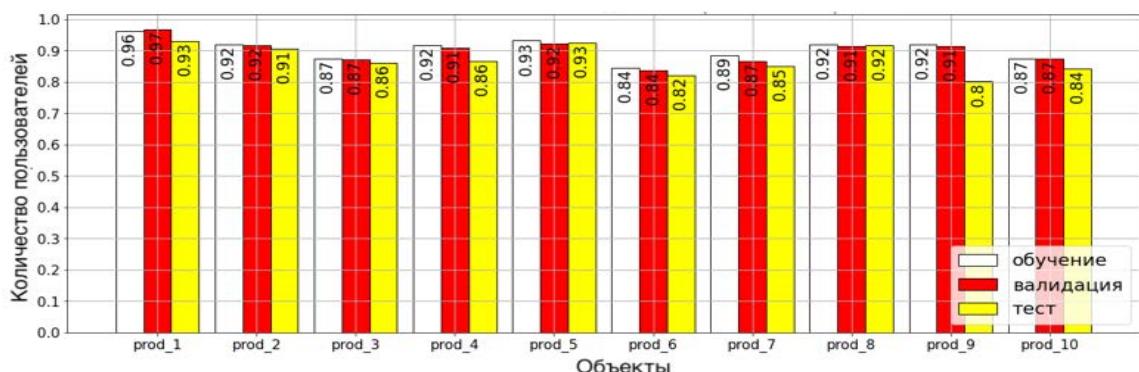


Рис. 15. Значения ROC-AUC всех моделей на разных выборках.

Таблица 5. Значения метрик итоговой рекомендательной системы

Названия метрик	Значение
Hit Rate	0.0859
Precision@3	0.0462
Precision@5	0.0352

**Заключение.** Построена рекомендательная система, которая может быть использована для нестандартного случая малого числа уникальных объектов в условиях высокой корреляции между некоторыми из них. Для этого предложен способ обучения отдельных классификаторов по каждому объекту. В данных представлены как метапризнаки, так и агрегаты колаборативных, модели обучены на предварительно подобранных гиперпараметрах. В конце производится калибровка для перехода к вероятностям и формируются списки итоговых рекомендаций по пользователям.

В качестве алгоритма для обучения выбран бустинг, так как наборы данных по каждому объекту небольшие по размеру и хорошо структурированы. Конкретная выбранная реализация – модель CatBoost из-за ее способности гибко обрабатывать категориальные признаки, а также высокого качества на открытых данных. В процессе выбора метода калибровки обнаружен ряд недостатков стандартных методов, которые могут быть решены дополнительными надстройками для легко адаптируемого алгоритма гистограммной калибровки.

Промежуточное качество классификаторов достаточное, отдельные модели показывают значительную ранжирующую способность на соответствующих выборках. Меры качества итоговой рекомендательной системы также высокие, причем наблюдается нелинейное повышение точности на  $k$  первых объектах с ростом  $k$ , что свидетельствует о том, что финальный алгоритм способен правильно ранжировать подходящие объекты по релевантности.

Результат может быть применен для улучшения качества работы рекомендательных алгоритмов при малом числе уникальных объектов, часть из которых взаимосвязана друг с другом. Описанный подход позволяет независимо получить вероятности заинтересованности пользователей во всех объектах, которые могут быть использованы для любых других целей. Технически метод гибкий в плане добавления новых объектов в общий набор и предоставляет широкие возможности для быстрого переобучения и повышения качества отдельных классификаторов.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Cano E., Morisio M.* Hybrid Recommender Systems: A Systematic Literature Review // Intelligent Data Analysis. 2017. V. 21. P. 1487–1524.
2. *Al-bashiri H., Abdulhak M., Romli A., Hujainah F.* Collaborative Filtering Recommender System: Overview and Challenges // J. Computational and Theoretical Nanoscience. 2017. V. 23. P. 9045–9049.
3. *Jahrer M., Toscher A.* Collaborative Filtering Ensemble // J. Machine Learning Research. 2012. V. 18. P. 61–74.
4. *Ahn H., Kang H., Lee J.* Selecting a Small Number of Products for Effective User Profiling in Collaborative Filtering // Expert Systems with Applications. 2010. V. 37. P. 3055–3062.
5. *Zharova M., Tsurkov V.* Neural Network Approaches for Recommender Systems // J. Computer and Systems Sciences International. 2024. V. 62. P. 1048–1062.
6. *Castells P., Moffat A.* Offline Recommender System Evaluation: Challenges and New Directions // AI Magazine. 2022. V. 43. P. 225–238.
7. *Bokde D., Girase S., Mukhopadhyay D.* Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey // Procedia Computer Science. 2015. V. 49. P. 136–146.
8. *Filho T., Song H., Perello-Nieto M.* Classifier Calibration: a Survey on How to Assess and Improve Predicted Class Probabilities // Machine Learning. 2023. P. 3211–3260.
9. *Alzubaidi L., Bai J., Al-Sabaawi A.* A Survey on Deep Learning Tools Dealing with Data Scarcity: Definitions, Challenges, Solutions, Tips, and Applications // J. Big Data. 2023. V. 10. № 46.
10. *Grinsztajn L., Oyallon E., Varoquaux G.* Why do Tree-based Models Still Outperform Deep Learning on Tabular Data? // arXiv:2207.08815v1, 2022.
11. *Alzubaidi L., Zhang J., Humaidi A.* Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions // arXiv:2207.08815v1, 2022.
12. *Borisov V., Leemann T., Sebler K.* Deep Neural Networks and Tabular Data: A Survey // arXiv:2110.01889v3, 2022.
13. *Bentejac C., Csorgo A., Martinez-Munoz G.* A Comparative Analysis of Gradient Boosting Algorithms // Artificial Intelligence Review. 2020. V. 54. P. 1937–1967.
14. *Sahour H., Gholami V., Torkaman J.* Random Forest and Extreme Gradient Boosting Algorithms for Streamflow Modeling Using Vessel Features and Tree-rings // Environmental Earth Sciences. 2021. V. 80. № 747.

15. Имплементация модели LightGBM на Python // GitHub. Microsoft LightGBM: website <https://github.com/microsoft/LightGBM> (accessed: 10.07.2024).
16. Имплементация модели XGBoost на Python // GitHub. Distributed (Deep) Machine Learning Community XGBoost: website <https://github.com/dmlc/xgboost> (accessed: 10.07.2024).
17. Имплементация модели CatBoost на Python // GitHub. CatBoost: website <https://github.com/catboost/catboost> (accessed: 10.07.2024).
18. *Ke1 G., Meng Q., Finley T.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Advances in Neural Information Processing Systems. 2017. P. 3146–3154.
19. Эксперименты с моделью LightGBM // Kaggle. LightGBM experiments: website <https://www.kaggle.com/code/prashant111/lightgbm-classifier-in-python> (accessed: 10.07.2024).
20. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // arXiv:1603.02754v3, 2016.
21. *Dorogush A., Prokhorenkova L., Gusev G.* CatBoost: Unbiased Boosting with Categorical Features // arXiv:1706.09516v5, 2019.
22. *Pargament F., Pfisterer F., Thomas J., Bischl D.* Regularized Target Encoding Outperforms Traditional Methods in Supervised Machine Learning with High Cardinality Features // Computational Statistics. 2022. V. 37. P. 2671–2692.
23. *Niculescu-Mizil A., Caruana R.* Predicting Good Probabilities with Supervised Learning // Machine Learning, Proc. 22nd Intern. Conf. (ICML). Bonn, Germany, 2005. P. 625–632.
24. *Guo C., Pleiss G., Sun Y., Weinberger K.* On Calibration of Modern Neural Networks // arXiv:1706.04599v2, 2017.
25. *Barlow R., Bartholomew D., Bremner J., Brunk H.* Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression // Royal Statistical Society. Series A: General. 1974. V. 137. P. 92–93.
26. *Zadrozny B., Elkan C.* Transforming Classifier Scores into Accurate Multiclass Probability Estimates // The Eighth ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining. Edmonton, 2002.
27. *Platt J.* Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods // Advances in Large Margin Classifiers. Cambridge: MIT Press, 2000. P. 61–74.
28. *Zadrozny B., Elkan C.* Transforming Classifier Scores into Accurate Multiclass Probability Estimates // Proc. 8th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining. N. Y., 2002. P. 694–699.
29. *Guo C., Pleiss G., Sun Y., Weinberger K.* On Calibration of Modern Neural Networks // Machine Learning, Proc. 34th Intern. Conf. (ICML). Sydney, 2017.
30. *Gupta C., Ramdas A.* Distribution-free Calibration Guarantees for Histogram Binning without Sample Splitting // arXiv:2105.04656v2, 2021.
31. *Naeini M., Cooper G.* Binary Classifier Calibration Using an Ensemble of Piecewise Linear Regression Models // Knowledge and Information Systems. 2018. V. 54. P. 151–170.
32. *Filho T., Song H., Perello-Nieto M.* Classifier Calibration: a Survey on How to Assess and Improve Predicted Class Probabilities // Machine Learning. 2023. V. 112. P. 3211–3260.
33. *Wang H., Liang Q., Hancock J., Khoshgoftaar T.* Feature Selection Strategies: a Comparative Analysis of SHAP-value and Importance-based Methods // J. Big Data. 2024. V. 11. № 44.
34. *Gebreyesus Y., Dalton D., Nixon S., Chiara D., Chinnic M.* Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP) // Future Internet. 2023. V. 15. № 88.
35. Имплементация библиотеки для подбора гиперпараметров Optuna на Python // GitHub. Optuna: website <https://github.com/optuna/> (accessed: 20.07.2024).

## ОБНАРУЖЕНИЕ И КЛАССИФИКАЦИЯ ОБЪЕКТОВ НА ТРЕХМЕРНЫХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

© 2024 г. И. А. Матвеев<sup>b, \*</sup>, А. А. Юрченко<sup>a, b, \*\*</sup>

<sup>a</sup>Федеральное государственное автономное образовательное учреждение высшего образования “Московский физико-технический институт (национальный исследовательский ун-т)” МФТИ, Москва, Россия

<sup>b</sup>ФИЦ ИУ РАН, Москва, Россия

\*e-mail: matveev@frccsc.ru

\*\*e-mail: yurchenko.aa@phystech.edu

Поступила в редакцию 13.09.2024 г.

После доработки 27.08.2024 г.

Принята к публикации 16.10.2024 г.

Рассматривается задача автоматического обнаружения и классификации объектов на трехмерных одноканальных растровых изображениях. Объекты могут иметь малую контрастность и значительную изменчивость формы, что затрудняет построение модели в явном виде. Для решения используются методы машинного обучения по размеченному базе прецедентов. Предлагается двухэтапная обработка, на первом шаге производится обнаружение объектов на изображении, на втором – уменьшение доли ложноположительных результатов и классификация. Применяется подход глубокого обучения с единственным входом и тренировкой на одновременное решение нескольких задач. Решается актуальная практическая задача построения клинически применимой автоматической системы поддержки принятия решений при обнаружении и определении категорий переломов ребер по данным снимков компьютерной томографии. Проведены вычислительные эксперименты на общедоступном наборе данных RibFrac. Показано, что предложенная система обеспечивает значение чувствительности обнаружения 0.935 при среднем количестве ложноположительных предсказаний на одно изображение 4.7. Приводится сравнение полученного алгоритма с существующими методами на основании количественных мер.

**Ключевые слова:** компьютерная томография, детекция переломов ребер, классификация переломов ребер, глубокое обучение

DOI: 10.31857/S0002338824060091, EDN: SUBUGN

## DETECTION AND CLASSIFICATION OF OBJECTS IN THREE-DIMENSIONAL IMAGES USING DEEP LEARNING METHODS

И. А. Матвеев<sup>b, \*</sup>, А. А. Юрченко<sup>a, b, \*\*</sup>

<sup>a</sup>Moscow Institute of Physics and Technology  
(National Research University), Moscow, Russia

<sup>b</sup>Federal Research Center “Computer Science and Control”  
of the Russian Academy of Sciences, Moscow, Russia

\*e-mail: matveev@frccsc.ru

\*\*e-mail: yurchenko.aa@phystech.edu

The issue of automatic object detection and categorization in three-dimensional, single-channel, raster images is considered. Objects may have low contrast and substantial shape variability, making it challenging to explicitly construct a model. The proposed solution employs machine learning techniques based on a labeled database of use scenarios. A two-step algorithm is presented, with the first stage being the detection of objects within the image and the second being the reduction of false positives and object categorization. Deep

learning approach is applied with a single input and trained for the simultaneous solution of multiple tasks. The practical goal of developing a clinically viable automatic decision support system to detect and classify rib fractures based on CT scans is being solved. Computational experiments were conducted on the publicly available RibFrac dataset. The proposed system was shown to achieve a detection sensitivity of 0.935, with an average number of false positive predictions per image of 4.7. The resulting algorithm was compared with existing methods using quantitative measures.

*Keywords:* computed tomography, rib fracture detection, rib fracture classification, deep learning

**Введение.** Задачи обнаружения, локализации и классификации объемных объектов на трехмерных изображениях значительно сложнее своих двумерных аналогов как с точки зрения формализации и алгоритмизации, так и в вычислительном отношении. Человек воспринимает большинство трехмерных объектов лишь как их поверхности, но не как действительно трехмерную структуру. В связи с этим большинство “классических” (не основанных на машинном обучении) решений при обработке трехмерных данных сразу сводятся к построению поверхностей. Кроме того, сама алгоритмизация работы с трехмерными данными существенно сложнее. В результате разнообразие и качество решения задач для трехмерных изображений существенно уступают таковым для двухмерных. Методы машинного обучения, опирающиеся на извлечение взаимосвязей на базе обучающих данных (набора прецедентов), свободны от таких ограничений. По этой причине методы машинного обучения даже на основании небольшого набора прецедентов достигают результатов, сравнимых или превосходящих по качеству получаемые “классическими” подходами. Все сказанное в полной мере относится к медицинской компьютерной томографии (КТ).

КТ дает возможность получить трехмерное изображение структур тела человека с высокой детализацией и является одним из наиболее эффективных методов обнаружения переломов ребер, позволяя находить значительно больше травм в области грудной клетки в сравнении с рентгенографией. Однако идентификация переломов ребер на тонкосрезовых КТ-изображениях с аксиальным пространственным разрешением 0.5–1.5 мм – трудоемкая задача для медиков. В силу наличия у ребер сложной формы и их диагонального расположения на срезах КТ-изображений, некоторые виды переломов тяжело диагностируются и имеют низкий процент обнаружения в клинической практике.

На практике для анализа КТ-изображений грудной клетки при обнаружении переломов ребер используются алгоритмы криволинейной развертки ребер, основанные на извлечении осевой линии каждого ребра с их последующим проецированием в двухмерную плоскость [1] или преобразованиях каждого аксиального среза изображения через цилиндрическую проекцию [2]. Данные методы позволяют снизить время анализа одного исследования врачом-диагностом, однако имеют более низкую чувствительность по сравнению с исходным трехмерным КТ-изображением [3, 4]. Также было отмечено [4], что применение двухмерной развертки приводит к возникновению ложноположительных результатов, некорректной оценке величины смещения ребер при переломах с диастазом, а также снижению коэффициента согласованности мнений экспертов.

Последние достижения в области искусственного интеллекта и компьютерного зрения привели к быстрому внедрению данных технологий в клиническую диагностику и анализ медицинских изображений. Глубокое обучение позволило достичь высоких результатов во многих задачах медицинской диагностики. В последние годы активно развивался нейросетевой подход к проблеме семантической сегментации переломов ребер на основании снимков КТ. Большая часть данных методов базируется на использовании глубоких сверточных нейросетей 3D U-Net [5], DeepLab v3+ [6], 3D ResNet [7] и их комбинаций, например FracNet [8] и FracNet+ [9].

Привлечение дополнительной информации о сегментационных масках ребер для обнаружения переломов дает значительное улучшение качества работы алгоритмов. Так, в работе [8] рассматривается метод предсказания переломов ребер с помощью скользящего окна по предварительно вычисленным областям расположения ребер. Продолжая данную идею, авторы статьи [9] предлагают добавлять сегментационные маски ребер, предварительно приведенные к необходимому пространственному разрешению и преобразованные набором сверточных модулей, на каждом слое энкодера и декодера сети U-Net. Для выделения области ребер могут применяться алгоритмы глубокого обучения, основанные на сегментации трехмерных воксельных изображений или разреженных облаков точек, например моделями PointNet [10] или DGCNN [11].

Знание о расположении ребер может быть привлечено при обучении модели для решения нескольких задач. Например, в SA-FracNet [12] предлагается одновременное решение задачи сегментации переломов и определение их границы с учетом формы ребер. В работе [13] приводится схема обучения, сочетающая решение задачи детекции переломов на основании двухмерных аксиальных срезов и локализации ребер с использованием трехмерного КТ-изображения. Данный подход позволяет добиться повышения эффективности обучения и точности прогнозирования по сравнению с обучением моделей по отдельности.

С увеличением разнообразия решаемых методами глубокого обучения задач начинает развиваться подход, представляющий из себя автоматический поиск подходящей архитектуры в решении конкретной задачи. Одним из наиболее популярных фреймворков (наборов инструментов), используемых при решении задач обработки двухмерных и трехмерных медицинских изображений, является nnU-Net [14]. Фреймворк представляет из себя метод семантической сегментации, который автоматически адаптируется к заданному набору данных. На основании анализа статистик предоставленных обучающих примеров производится автоматическая настройка соответствующего алгоритма сегментации на базе нейросетевой модели U-Net [5]. Настройка подразумевает определение оптимальной конфигурации сети, алгоритмов предварительной обработки изображений и аугментаций данных, используемых в процессе обучения. Инструмент Auto3Dseg [15] также реализует методы поиска оптимальной архитектуры среди заданного класса сегментационных моделей, включающих в себя Swin UNETR [16], UNet [5] и SegResNet [17]. Алгоритмы были апробированы на широком наборе медицинских тестовых данных и продемонстрировали высокие показатели мер качества сегментации и детекции во многих соревнованиях по локализации структур организма и обнаружению патологий.

С ростом объемов открытых наборов данных с медицинскими изображениями различных модальностей становится распространенной практика применения предобученных моделей в решении более узких медицинских задач, например STU-Net [18], которая построена на основании фреймворка nnU-Net. Предварительно обученная на крупномасштабном наборе данных TotalSegmentator [19] модель STU-Net была применена авторами работы [9] в решении задачи сегментации переломов ребер и продемонстрировала более высокие результаты в сравнении с подходом, реализующим обучение аналогичной архитектуры с нуля.

Предварительное обучение базовых моделей для последующего решения целевой задачи может быть выполнено с применением техник обучения с частичным привлечением учителя (англ. semi-supervised learning). Использование данного подхода позволяет включить в обучение значительно больше КТ-изображений, привлекая объемы немаркированных снимков, число которых зачастую значительно превосходит количество размеченных сканов. В работе [13] предлагается применение техники, основанной на контрастном обучении (англ. contrastive learning) на уровне пикселей, для предварительной настройки весов базовой модели с последующей тонкой настройкой параметров сети для решения задачи сегментации переломов ребер.

**1. Постановка задачи.** На вход подается КТ-изображение области  $I$ , являющееся трехмерной матрицей размерами  $W \times H \times D$ , содержащей целые числа из диапазона от  $-1024$  до  $3071$ , которые задают значение плотности каждого вокселя изображения по шкале денситометрических показателей Хаунсфилда (англ. hounsfield units (HU)). Требуется выделить области изображения, на которых присутствуют объекты (переломы ребер), и определить класс каждого объекта из заданного списка возможных клинических категорий переломов: консолидированный перелом, перелом с диастазом (смещением), перелом без диастаза.

С точки зрения решения задачи сегментации результат работы алгоритма может быть описан в виде сегментационной маски объектов, представляющей собой трехмерную матрицу  $F$ , которая соответствует размерам исходного изображения  $I - W \times H \times D$ , и списка классов объектов  $C$ . Значения вокселей сегментационной маски  $F$  являются целыми числами в диапазоне от  $0$  до  $N$ , где  $N$  – количество обнаруженных на изображении объектов. Воксели маски, имеющие значение, равное  $i \in [1, N]$ , описывают область локализации объекта  $i$ , принадлежащего к классу  $C[i]$ . Значение  $O$  вокселей маски соответствует заднему плану и не описывает объекты. Обозначая через  $F_i$  набор координат, соответствующих области сегментационной маски  $F_i = \{(x, y, z), \text{ если } F(x, y, z) = i\}$ , получаем область локализации объекта на исходном изображении  $I_{F_i} = \{I(x, y, z) : (x, y, z) \in F_i\}$  и ее класс  $C[i]$ .

**1.1. Задача детекции.** В большинстве работ, посвященных решению задачи локализации и классификации переломов ребер, рассматриваются методы семантической сегментации, использующие в качестве целевых объектов предсказания воксельные сегментационные маски.

Однако начальная постановка задачи сегментации как выделения области исходного КТ-изображения имеет несколько недостатков, усложняющих получение результатов. Во-первых, внешний вид переломов сильно варьируется между различными классами. Переломы с диастазом имеют различной степени выраженности разрывы между двумя частями одного ребра. Сегментация патологии такого вида может интерпретироваться неоднозначно – к сегментационной маске перелома могут быть отнесены различные объемы кости на концах частей ребра и окружающего вещества, находящегося в разрыве. В то же время консолидированные переломы, являющиеся полноценным восстановлением кости после перелома, не имеют четких границ. Во-вторых, на данный момент не существует единой методологии, регламентирующей правила нанесения разметки для различных классов переломов. Мaska области перелома (ее форма, объем и ориентация) сильно зависит от специалиста, выполняющего разметку. Разметка существующих наборов данных КТ-снимков пациентов с переломами составляется широким кругом рентгенологов и врачей-диагностов, каждый из которых придерживается собственного метода определения объема и формы области перелома. Так, в открытых наборах данных переломы одного и того же вида, размеченные разными специалистами, могут иметь объемы сегментационных масок, отличающиеся в 3 и более раз. В некоторых случаях экспертом может быть размечена вся область ребра, содержащая перелом, в других – только непосредственная область перелома. Часто эксперты делают разный отступ разметки от границ ребра. Описанные различия в сегментационной разметке переломов со смещениями представлены на рис. 1. На изображениях 1, 3 и 7 маска выступает за границы кости самого ребра. Для изображений 5, 6 и 8 маска перекрывает всю или практически всю область ребра, в то время как для переломов 2 и 4 к сегментационной маске отнесены меньшие объемы костной ткани.

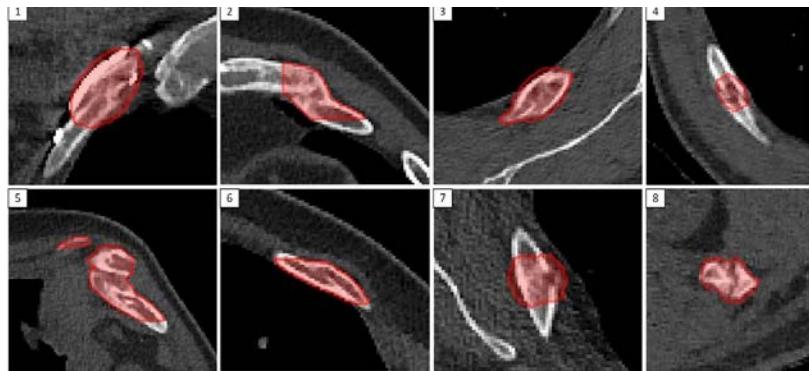


Рис. 1. Примеры, демонстрирующие различие в сегментационной разметке переломов со смещениями.

Исходя из этого, задача переформулирована как задача определения параллелепипеда, ограничивающего область интереса (bounding box), что принято называть *задачей детекции*. При необходимости можно перейти к сегментационным маскам путем выделения области объекта, но уже внутри ограничивающего параллелепипеда, а не на всем изображении.

В терминах задачи детекции результат представляется в виде набора ограничивающих параллелепипедов  $B$ , каждый элемент которого задается шестью координатами – минимальными и максимальными координатами вершин ограничивающего параллелепипеда по трем осям:

$$B[i] = \left( x \in \left[ x_{\min}^{(i)}, x_{\max}^{(i)} \right], y \in \left[ y_{\min}^{(i)}, y_{\max}^{(i)} \right], z \in \left[ z_{\min}^{(i)}, z_{\max}^{(i)} \right] \right), i \in \overline{1, N}, \quad (1.1)$$

и списка классов объектов  $C$ . Количество элементов списков  $B$  и  $C$  совпадает и равно количеству обнаруженных объектов  $N$ . Обозначим как  $I_{B[i]}$  область изображения  $I$ , заключенную внутри параллелепипеда  $B[i]$ . Формально

$$I_{B[i]} = \{I(x, y, z) : (x, y, z) \in B[i]\}. \quad (1.2)$$

Область  $I_{B[i]}$  является локализацией объекта  $i$ , относящегося к классу  $C[i]$ .

1.2. Переход к задаче детекции. Преобразование от сегментационных масок к ограничивающим параллелепипедам достаточно просто и сводится к вычислению минимальных и максимальных координат вокселей маски по трем осям. Пусть  $F$  – сегментационная маска объектов,  $B$  – список ограничивающих параллелепипедов, задаваемый изначально

как пустой. Пусть  $F_i = \{(x, y, z) : M(x, y, z) = i\}$  – сегментационная маска перелома  $i$  в координатах  $M$ . Тогда прямоугольные границы маски перелома  $i$  с учетом отступа в 1voxель могут быть записаны как

$$x_{\min}^{(i)} = \max(0, \min\{x : (x, y, z) \in F_i\} + 1),$$

$$x_{\max}^{(i)} = \min(W', \max\{x : (x, y, z) \in F_i\} + 1).$$

Аналогично определяются остальные границы параллелепипеда (1.1) –  $y_{\min}^{(i)}, y_{\max}^{(i)}, z_{\min}^{(i)}, z_{\max}^{(i)}$ .

**2. Архитектура системы.** Представленная система методов решения задачи локализации и классификации переломов ребер на КТ-изображениях имеет следующие основные блоки:

1) предварительная обработка КТ-изображения, включающая выделение области ребер методом сегментации разреженного облака точек с применением модели PointNet++, приведение к целевому пространственному разрешению и нормирование интенсивности;

2) определение областей локализаций переломов ребер в виде набора ограничивающих параллелепипедов с применением детекционной модели;

3) формирование итоговых детекционных предположений с учетом определенных экспериментально оптимальных гиперпараметров отбора ограничивающих параллелепипедов;

4) разделение полученных локализаций переломов на клинические категории и удаление ложноположительных предсказаний с помощью классификационной модели.

Отдельно стоит отметить, что дополнительный выход детекционной сети, реализующий решение задачи семантической сегментации ребер, используется только в процессе обучения и может быть исключен на этапе применения модели при диагностировании переломов.

Общая архитектурная схема предлагаемой системы приведена на рис. 2.

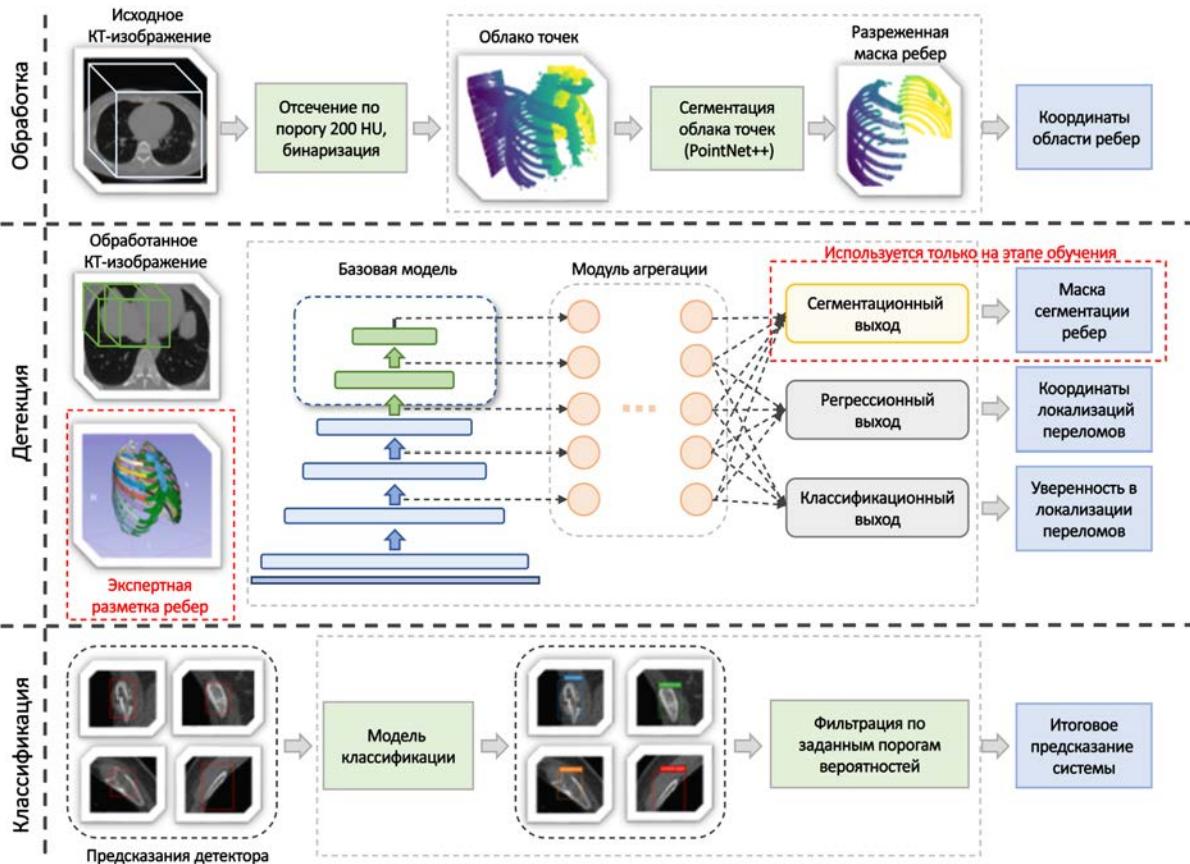


Рис. 2. Архитектурная схема предлагаемого двухэтапного метода, включающего детектирование переломов с последующей их классификацией.

2.1. П р е д о б р а б о т к а д а н н ы х. Первоначальным этапом обработки КТ-исследований является приведение всех изображений к единому пространственному разрешению. Данное преобразование необходимо для выравнивания расстояний расположения анатомических структур относительно друг друга. В качестве целевого пространственного разрешения было выбрано медианное значение спейсинга для КТ-сканов из отобранной части исследований датасета RibFrac – (1.25, 0.75, 0.75). Преобразование КТ-изображений к указанному пространственному разрешению осуществлялось методом трилинейной интерполяции, преобразование соответствующих сегментационных масок путем интерполяции методом ближайших соседей.

Также важным этапом обработки исходных КТ-изображений, непосредственно влияющим на качество и скорость работы алгоритмов обнаружения целевых патологий, является предварительное выделение области изображения, в котором анатомически предполагается локализация искомой патологии. В задаче обнаружения переломов ребер естественным ограничением области поиска выступает граница внутри грудной клетки, включающая в себя ребра. Непосредственное выделение области ребер позволяет исключить другие костные структуры (такие как ключицы, нижние элементы черепа и части позвоночника) из рассмотрения, что уменьшает общую ошибку обнаружения переломов и снижает количество ложноположительных предсказаний. Более того, это дает возможность уменьшить размер исходного изображения, тем самым сокращая время работы и количество используемой памяти, что особенно важно при обработке трехмерных изображений методами глубокого обучения с помощью сверточных нейросетей.

Для выделения целевой области на КТ-изображении может быть использован ряд методов, основанных на априорных знаниях о плотности структур и тканей, эвристических правилах, морфологических операциях, а также алгоритмах компьютерного зрения. Зачастую при сегментации области интереса не требуется субвоксельная точность работы алгоритма. В силу данного факта при применении методов компьютерного зрения в решении данной задачи могут быть использованы изображения, предварительно переведенные в более низкое пространственное разрешение методами интерполяции. Таким образом, сегментация области грудной клетки, содержащей в себе ребра, может быть выполнена посредством применения сегментационных нейросетевых методов, например U-Net-подобных архитектур.

В статье для выделения области ребер предлагается использовать подход, основанный на сегментации облака точек алгоритмами глубокого обучения. Данный метод широко применяется в задачах, связанных с сегментацией ребер, выделением центр-линий ребер и обнаружением переломов. Авторы работ [22, 23] рассматривают модель сегментации разреженного облака точек PointNet++ [10] непосредственно для решения задачи определения маски каждого ребра и последующего выделения центральных линий. В работах [8, 9] применяется данный метод как вспомогательный при решении задачи сегментации переломов ребер.

Преобразование исходного КТ-изображения в облако точек осуществляется путем отсечения вокселей по порогу 200 HU с последующей бинаризацией. Предварительно производится понижение линейного пространственного разрешения КТ-изображения в 2 раза в целях ускорения работы алгоритма и сглаживания отдельных вокселей высокой плотности. Полученное облако точек сегментируется на два класса (ребра и фон) с помощью модели PointNet++, обученной на случайных подвыборках размером 30 тыс. элементов из исходного облака точек. В процессе обучения модели минимизировалась функция потерь перекрестной энтропии.

Обучение модели происходило с использованием сегментационных масок ребер, представленных в датасете RibSeg v2 [22]. По аналогии с преобразованием исходных КТ-сканов осуществлялось приведение разметки ребер к целевому пространственному разрешению с последующим понижением разрешения в 2 раза путем интерполяции методом ближайших соседей.

Предсказание модели на этапе использования для всего облака точек проводилось путем последовательного выбора некоторого количества непересекающихся наборов по 30 тыс. точек в каждом с последующим применением обученной модели для каждого из них. В силу отсутствия требования классификации каждой точки количество выборок не обязательно должно покрывать все облако точек – для достижения необходимой точности локализации области ребер может быть использовано небольшое число случайных выборок. Эмпирически установлено, что количество выборок, равное трем, обеспечивает баланс между скоростью работы и точностью локализации области ребер. Общая структура алгоритма сегментации ребер на базе разреженного облака точек приведена на рис. 3.

На основе полученной маски сегментации ребер вычисляются координаты ограничивающего параллелепипеда с добавлением отступа в 5 вокселей по каждому направлению и производится обрезание исходного изображения. Использование отступов позволяет уменьшить возможные погрешности модели сегментации и не терять информацию об анатомических

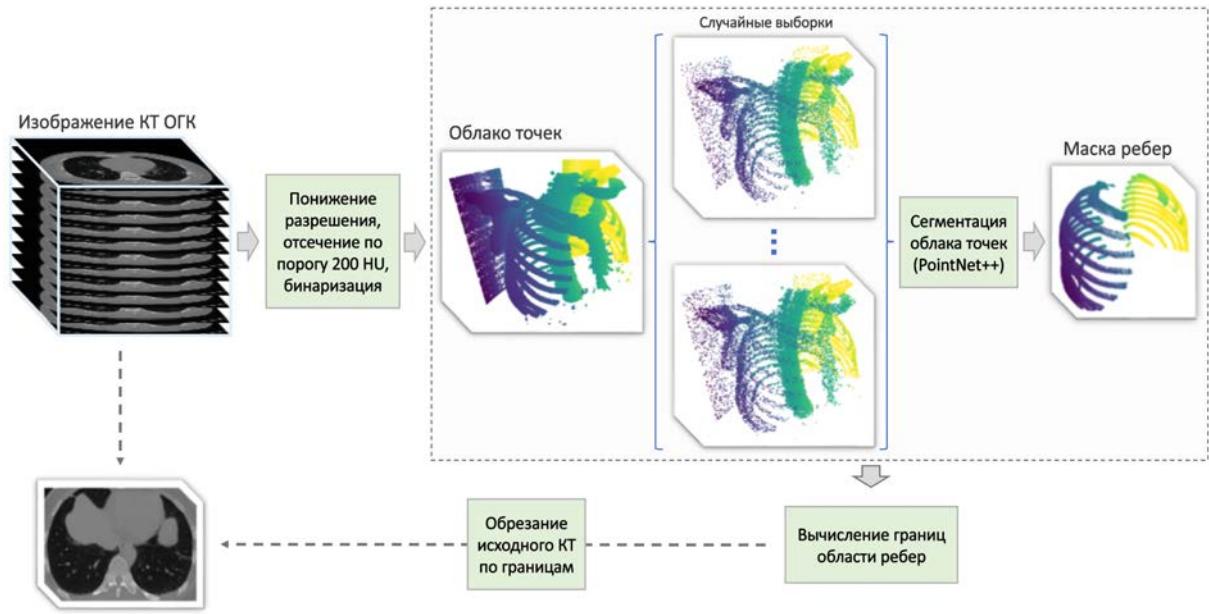


Рис. 3. Сегментация ребер с использованием преобразования исходного КТ-изображения в разреженное облако точек и модели PointNet++.

структурах на краях обрезанного изображения. Аналогичные преобразования применяются к соответствующим сегментационным маскам переломов и ребер.

В рамках разработки модели семантической сегментации ребер также был проанализирован метод сегментации воксельного КТ-изображения с помощью нейросетевой модели SegResNet, предварительно обученной на наборе данных TotalSegmentator [19]. Среднее время предсказания алгоритма с учетом предварительной обработки исследования, работы модели и финальной обработки результата, включающей фильтрацию малых областей сегментации и применение морфологических операций, составило 15–20 с, что в рамках разрабатываемой системы является превышением допустимого времени работы. Использование архитектур на основании сети U-Net при аналогичном качестве требует большего времени на формирование предсказания.

Таким образом, предварительная обработка КТ-изображения и сегментационных масок переломов может быть представлена в виде следующего алгоритма:

1. Исходное изображение  $I$  и сегментационные маски переломов  $F$  размерами  $W \times H \times D$  преобразуются к пространственному разрешению (1.25, 0.75, 0.75) методами трилинейной интерполяции и ближайших соседей соответственно. Результат данного шага – изображение  $I'$  и маски переломов  $F'$  размерами  $W' \times H' \times D'$ .

2. Методом трилинейной интерполяции разрешение изображения  $I'$  понижается в 2 раза и выполняется его преобразование в облако точек  $O$  путем отсечения вокселей по порогу 200 HU;  $O = \{(x, y, z) : I'_{1/2}(x, y, z) > 200\}$ .

3. К трем случайным выборкам размерами 30 тыс. элементов в каждом из результирующего облака точек применяется обученная модель сегментации PointNet++, классифицирующая каждую точку изображения на ребра ( $c_r = 1$ ) и фон ( $c_{bg} = 0$ ). Формируется набор меток классов  $S = \{s_i = \text{PointNet}(o_i) \in [0, 1] : o_i \in O\}$ . Тогда маска ребер может быть определена как  $R = \{o_i \in O : s_i = 1\}$ .

4. Задаются границы маски ребер с учетом отступа 5 вокселей по каждому из пространственных направлений:

$$x_{\min} = \max(0, 2 \cdot \min\{x : (x, y, z) \in R\} - 5),$$

$$x_{\max} = \min(W', 2 \cdot \max\{x : (x, y, z) \in R\} + 5).$$

Аналогично находятся координаты  $y_{\min}, y_{\max}, z_{\min}, z_{\max}$ .

5. Изображение  $I'$  и маски переломов  $F'$  обрезаются в соответствии с полученными координатами:

$$I'' = \{I'(x, y, z) : x_{\min} \leq x \leq x_{\max}, y_{\min} \leq y \leq y_{\max}, z_{\min} \leq z \leq z_{\max}\},$$

$$F'' = \{F'(x, y, z) : x_{\min} \leq x \leq x_{\max}, y_{\min} \leq y \leq y_{\max}, z_{\min} \leq z \leq z_{\max}\}.$$

2.2. М етод сегментации. В силу того, что на первом этапе работы предлагаемого алгоритма обнаружения переломов ребер требуется только их локализация, рассматривается задача одноклассовой семантической сегментации методами глубокого обучения.

В качестве начальной точки разработки сегментационной модели представлено использование алгоритмов автоматического определения наиболее подходящей архитектуры для заданного набора данных. В решении задачи сегментации переломов ребер приведены 3D-реализации моделей SegResNet с механизмом обучения Deep Supervision [24], U-Net и Swin UNETR v2 [25]. Оптимальные гиперпараметры сети U-Net были подобраны с помощью методов фреймворка nnU-Net [14], модель SegResNet показала лучшие результаты на основании алгоритма выбора модели методами Auto3DSeg [15].

Конфигурация модели SegResNet: число начальных фильтров (каналов на первом слое) – 16, количество блоков энкодера и декодера – (1, 2, 2, 4) и (1, 1, 1) соответственно, глубина используемых карт признаков для механизма Deep Supervision – 3. На каждом блоке энкодера пространственное разрешение изображения уменьшается в 2 раза, количество каналов соответственно увеличивается в 2 раза.

Конфигурация модели SwinUNETR: количество модулей на слоях – (2, 2, 2, 2), число выходов механизма внимания на каждом слое – (3, 6, 12, 24), размер признакового описания скрытого состояния – 24. Применялась вторая версия данной сети.

Конфигурация модели U-Net: количество каналов на слоях энкодера и декодера – (16, 32, 64, 128, 256, 384), функция активации LeakyReLU с наклоном в области отрицательных значений, равным 0.01, размеры сверток –  $3 \times 3 \times 3$ , шаг ядер свертки на первом слое энкодера –  $1 \times 1 \times 1$ , на последующих –  $2 \times 2 \times 2$ . Использовалась модификация сети с применением механизма Skip Connection [26].

Основываясь на результатах исследований, демонстрирующих прирост качества методов сегментации переломов ребер при привлечении дополнительной информации о ребрах [8, 9], был предложен вариант обучения моделей с добавлением второго канала с информацией о voxелях изображения в областях, соответствующих сегментационным маскам ребер. Пусть  $I$  – исходное КТ-изображение размером  $W \times H \times D$ ,  $R$  – соответствующая изображению маска семантической сегментации ребер. Тогда дополнительный канал  $I^c$  может быть представлен в виде трехмерной матрицы размером  $W \times H \times D$  с элементами  $I^c(x, y, z)$  следующего вида:

$$I^c(x, y, z) = \begin{cases} I(x, y, z), & \text{если } R(x, y, z) \neq 0, \\ -1024 & \text{иначе.} \end{cases} \quad (2.1)$$

В процессе обучения дополнительный канал конкатенировался с исходным изображением и использовался в качестве входа в модель.

При обучении сегментационных моделей функциями потерь выступала сумма функции потерь Focal Loss (2.7) и функции потерь Dice (2.6). Вычисление функции потерь на дополнительных выходах сети SegResNet осуществлялось с предварительным повышением разрешения соответствующих карт признаков до размера целевой сегментационной маски переломов. Компонента  $i$  функции потерь, вычисленная для выхода глубины  $i$ , нормировалась на значение  $1/2^i$ . Итоговая функция потерь определялась в виде суммы компонент с различных выходов декодера:

$$L_{total}(y_{true}, y_{pred}) = \sum_{i=0}^{d-1} \frac{1}{2^i} L_i(y_{true}, y_{pred}^{(i)}). \quad (2.2)$$

Здесь  $\{y_{pred}^{(i)} : i \in [1, \dots, d]\}$  – выходы модели с последних  $d$  слоев декодера.

2.3. А рх и т е к т у р ы д е т е к т о р о в. Первым этапом работы алгоритма обнаружения и классификации переломов является локализация областей, содержащих данную травму. Результат работы данного шага в терминах задачи детекции может быть описан в виде набора ограничивающих параллелепипедов  $B$  (1.1) и степени уверенности алгоритма  $P$  в наличии перелома в каждой из предсказанных областей. С учетом нормировки уверенность модели  $P[i]$

для предсказания с номером  $i$  в нахождении перелома в области изображения  $I_{B[i]}$  (1.2) может быть интерпретирована как вероятность нахождения патологии в данной области. На этом этапе требуется обнаружение как можно большего числа переломов, т.е. высокое значение чувствительности алгоритма.

Классическим подходом к построению нейросетевой модели детекции является представление архитектуры в виде комбинации трех основных составляющих, включающих в себя базовую модель (англ. backbone), модуль извлечения и агрегации признаковых описаний с различными пространственными разрешениями (англ. neck) и выходы для предсказания координат ограничивающих параллелепипедов и уверенности модели в локализации объекта в заданной области (англ. regression and classification heads). Задача базовой модели заключается в извлечении информативных признаковых описаний различных пространственных разрешений из исходного изображения. Модуль извлечения и агрегации преобразует полученные выходы нижних и верхних уровней базовой модели путем комбинации карт признаков различного пространственного разрешения между собой. Классификационный и регрессионный выходы на основании объединенных карт признаков формируют финальные предсказания, извлекая информацию о классах и координатах объектов на изображении.

В качестве базовой детекционной модели предлагается применять anchor-based [27] детектор, концептуальная схема которого продемонстрирована на рис. 4. Выбор модели, использующей предварительно заданный набор ограничивающих параллелепипедов, основан на эмпирических выводах и результатах сравнения anchor-based и anchor-free детекционных архитектур с точки зрения мер качества полноты и чувствительности [28, 29].

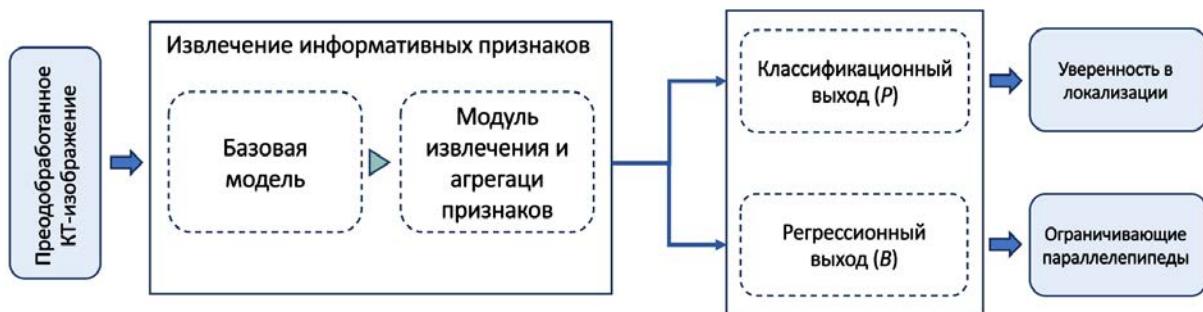


Рис. 4. Концептуальная схема модели детекции переломов.

В рамках разработки и тестирования методов детекции были реализованы 3D-модификации архитектур RetinaNet [30] и EfficientDet [31] с применением в качестве базовых моделей ResNet (residual network) [7], SEResNeXt (squeeze-and-excitation residual network with external transformations) [32, 33], HRNet (high-resolution net) [34] и EfficientNet [35] (B1-B5) соответственно. Сверточные ядра первого слоя в моделях ResNet, SEResNeXt и HRNet были заменены на свертки размером  $3 \times 3 \times 3$  и шагом  $1 \times 1 \times 1$  с целью увеличения пространственного разрешения начальной и последующих карт признаков в 2 раза. Рассматривались модули агрегации признаков FPN (feature pyramid network) [36] и BiFPN (bidirectional feature pyramid network) [31] с применением различного числа выходных слоев с базовыми моделями, количество которых варьировалось от двух до четырех. Количество карт признаков, используемых в выходных слоях (модулях регрессии и классификации), также варьировалось в диапазоне от трех до пяти. Финальными модулями, преобразующими полученные карты признаков различных пространственных разрешений в логиты классификации и регрессии на ограничивающих параллелепипедах, выступали сверточные блоки, состоящие из четырех последовательных слоев. Выход с последнего пятого слоя в данных модулях не применяется в силу малой пространственной размерности карт признаков данного слоя, получающихся при использовании сегмента КТ-изображения с разрешением  $W \times H \times D = 192 \times 192 \times 192$ .

Формирование дополнительных карт признаков, используемых в модулях агрегации, осуществлялось с помощью добавления дополнительных блоков к последнему слою базовой модели. Увеличение количества признаковых описаний различных пространственных разрешений на единицу производилось с применением операции MaxPooling с размером ядра  $3 \times 3 \times 3$  и шагом  $2 \times 2 \times 2$  к последнему выходу сети, на два — путем последовательного применения двух сверточных модулей с финальной операцией MaxPooling. Архитектурная схема детекционной модели приведена на рис. 5.

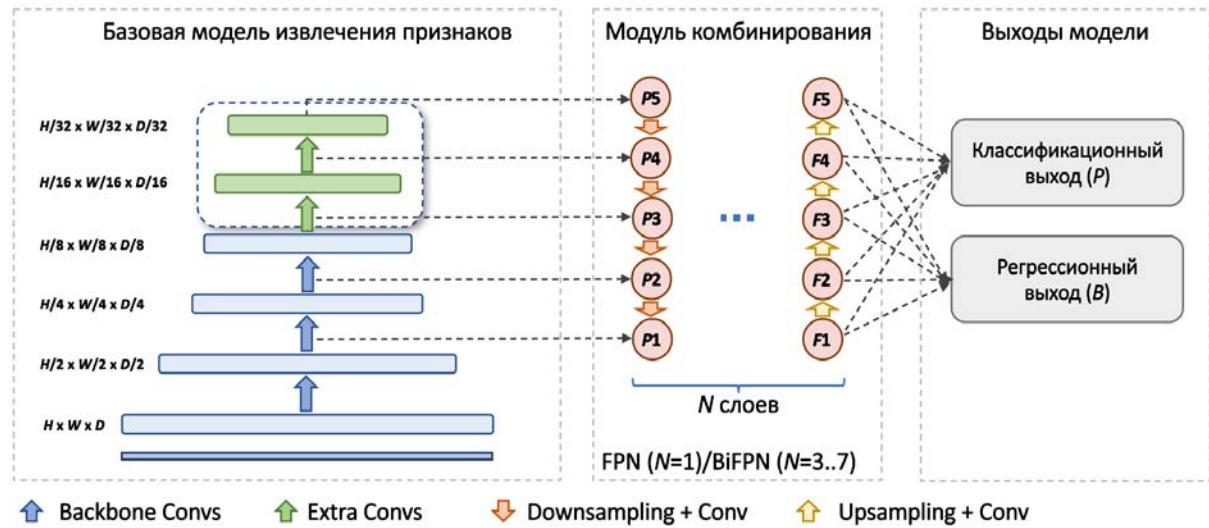


Рис. 5. Общая архитектура модели детекции переломов: базовая модель, модуль комбинирования карт признаков и модули классификации и регрессии.

2.4. Использование дополнительного выхода для сегментации областей интереса. Применение подхода обучения нейросетей на несколько задач в некоторых случаях позволяет добиться более высокого качества решения целевой задачи, а также получить более устойчивую модель [37] за счет выделения более информативного и общего признакового описания исходных данных. Как было отмечено ранее, интуитивным шагом в решении задачи обнаружения переломов ребер является привлечение дополнительной информации в виде масок семантической сегментации ребер [9]. Однако подходы, использующие данную информацию непосредственно на шаге предсказания модели, имеют существенный недостаток. Качество итогового предсказания модели будет зависеть от качества работы алгоритма сегментации ребер, и, соответственно, ошибки на данном этапе могут приводить к увеличению итоговой ошибки работы системы. Более того, увеличивается суммарное время работы алгоритма. Таким образом, предлагается использовать информацию о семантических масках ребер только на этапе обучения модели. В исследовании рассматривались сегментационные маски ребер, представленные в датасете RibSeg v2 [22].

В качестве дополнительного выхода детекционной модели предлагается применять модуль сегментации ребер на изображениях пониженного разрешения. Понижение разрешения осуществляется в 2 раза для соответствия пространственному разрешению первой карты признаков, формирующейся на выходе базовой модели и модуля агрегации. Среди вариантов реализации дополнительного сегментационного выхода были рассмотрены простые реализации декодеров, представляющие из себя объединение карт признаков с различными пространственными разрешениями в финальную сегментационную маску, и U-Net-подобные архитектуры, такие как 3D U-Net, 3D U-Net++ [38] и 3D Attention U-Net [39]. Картами признаков энкодеров данных модулей выступают выходы с модуля извлечения и комбинирования признаков детекционной модели, предварительно преобразованные сверточными модулями.

Особое внимание в данном случае будет уделено модификации сети Attention U-Net, схема которой представлена на рис. 6. Карты признаков каждого слоя декодера получаются путем конкатенации экстраполированных карт признаков декодера с предыдущего слоя и карт признаков энкодера, прошедших через модуль Attention Gate. Механизм пространственного внимания позволяет сети сфокусироваться на важных регионах изображения, игнорируя нерелевантные или шумные участки. С точки зрения подхода к построению архитектуры на решение нескольких задач использование данного приема позволит модели выучивать важные признаки как для задачи детекции переломов ребер, так и для задачи сегментации костных структур ребер. Более того, информация о расположении ребер теоретически может заметно снизить число ложноположительных предсказаний, локализованных вне ребер, так как у модели будет информация о структуре костной системы грудной клетки.

Положим, что  $\{f^l\}_{l=1}^n$  — карты признаков, получаемые на выходе из модуля извлечения признаков детекционной модели, например BiFPN. Здесь  $n$  — количество выходов модуля

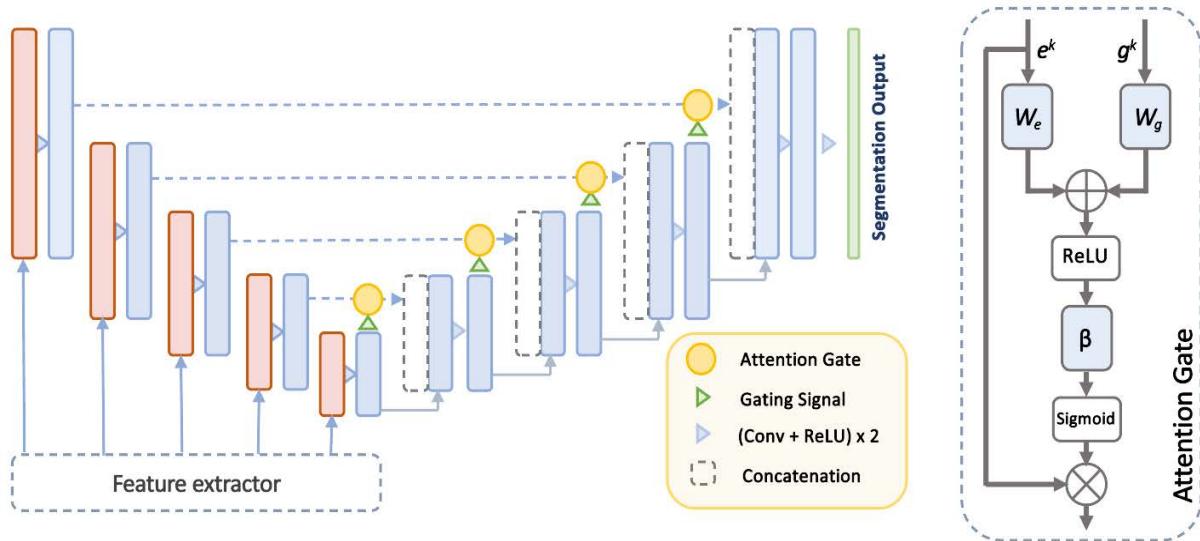


Рис. 6. Архитектура модифицированного варианта сети Attention U-Net с механизмом пространственного внимания Attention Gate.

извлечения признаков. Тогда картами признаков энкодера будет являться набор  $\{e^l\}_{l=1}^n$ , в котором  $e^l$  получается из  $f^l$  путем последовательного применения модулей свертки, батч-нормализации [40] и функций активации.

Механизм внимания Attention Gate может быть формализован в следующем виде. Пусть  $e^l$  – карта признаков энкодера на слое  $l$  размером  $C_e^l \times W_e^l \times H_e^l \times D_e^l$ ,  $d^l$  – карта признаков на слое декодера  $l$  размером  $C_d^l \times W_d^l \times H_d^l \times D_d^l$  (нумерация в обратном порядке по мере уменьшения пространственного размера карт признаков). Входной сигнал  $g^l$  (gating signal) для модуля внимания формируется на основании карты признаков декодера  $d^l$  путем последовательного выполнения операций интерполяции, свертки, батч-нормализации и функции активации. Размер  $g^l$  равен  $C_g^l \times W_g^l \times H_g^l \times D_g^l$  и совпадает с пространственным разрешением соответствующей карты признаков энкодера. С помощью свертки  $1 \times 1 \times 1$  полученные карты признаков выравниваются по числу каналов. Данные карты признаков  $g^l$  размера  $C_{inner}^l \times W_e^l \times H_e^l \times D_e^l$  и  $e^l$  размера  $C_{inner}^l \times W_e^l \times H_e^l \times D_e^l$  складываются с последующим применением функции активации ReLU. С помощью последовательного использования операций свертки, батч-нормализации и функции активации-сигмоиды формируется карта внимания размерами  $1 \times W_e^l \times H_e^l \times D_e^l$ , которая поэлементно умножается на карту признаков энкодера  $e^l$ . Формирование карты признаков  $d^{l-1}$  следующего слоя декодера осуществляется путем конкатенации полученного результата с картой признаков  $g^l$ :

$$\begin{aligned}
 Q_a &= \beta^T \left( \sigma_2 \left( \mathbb{W}_e^T e^l + \mathbb{W}_g^T g^l + b_g \right) \right) + b_\beta, \\
 d^{l-1} &= \left[ \sigma_1(Q_a) * e^l, g^l \right], \\
 \mathbb{W}_e &\in \mathbb{R}^{C_e \times C_{inner}}, \mathbb{W}_g \in \mathbb{R}^{C_g \times C_{inner}}, \beta \in \mathbb{R}^{C_{inner} \times 1}, b_\beta \in \mathbb{R}, b_g \in \mathbb{R}^{C_{inner}}.
 \end{aligned} \tag{2.3}$$

Здесь  $\sigma_1$  – функция активации ReLU,  $\sigma_2$  – функция активации сигмоида, “\*” – операция поэлементного произведения.

2.5. Функция потерь общей модели включает в себя три компоненты в соответствии с тремя различными выходами сети: классификационным, регрессионным (на ограничивающих параллелепипедах) и сегментационным. В силу того, что на данном шаге алгоритма не производится классификация переломов ребер по клиническим категориям, классификационный выход модели предсказывает вероятности только для одного класса, которые могут интерпретироваться как уверенность модели в локализации перелома в рамках

соответствующего ограничивающего параллелепипеда. На основании этого в классификационном выходе сети была применена функция потерь бинарной перекрестной энтропии:

$$BCELoss(y_{true}, y_{pred}) = -\frac{1}{N} \sum_{i=1}^N [y_{true} \log y_{pred} + (1 - y_{true}) \log (1 - y_{pred})]. \quad (2.4)$$

В качестве функции потерь, использующейся в регрессионном выходе сети, был применен сглаженный  $L1$  лосс [41] с коэффициентом  $\beta$ , равным  $1/9$ :

$$SmoothL1Loss(x_n, y_n, \beta) = \begin{cases} 0.5(x_n - y_n)^2 / \beta, & \text{если } |x_n - y_n| < \beta, \\ |x_n - y_n| - 0.5\beta & \text{иначе.} \end{cases} \quad (2.5)$$

Агрегирование данной функции потерь производилось путем усреднения значений по всем предсказаниям и по шести измерениям каждого предсказания.

Целевой функцией оптимизации для сегментационного выхода модели выступала равновесная сумма функций потерь Dice Loss [42] и Focal Loss [30] с коэффициентом  $\gamma$ , равным двум:

$$DiceLoss(y_{true}, y_{pred}) = 1 - \frac{2 \sum (y_{true} * y_{pred}) + \epsilon}{\sum y_{true} + \sum y_{pred} + \epsilon}. \quad (2.6)$$

Здесь “ $*$ ” в числителе – операция поэлементного произведения. Суммы рассчитываются по всем вокселям предсказанной и действительной масок сегментации объектов. Коэффициент сглаживания  $\epsilon = 1$  используется для обеспечения численной стабильности вычисления функции потерь Dice:

$$FocalLoss(y_{true}, y_{pred}, \gamma) = -\alpha_{\tau} (1 - p_{\tau})^{\gamma} \log(p_{\tau}), \text{ где } p_{\tau} = \begin{cases} y_{pred}, & y_{true} = 1, \\ 1 - y_{pred} & \text{иначе.} \end{cases} \quad (2.7)$$

Коэффициент  $\alpha_{\tau}$  обеспечивает возможность балансирования классов.

В случае решения задачи многоклассовой сегментации производится усреднение функций потерь по каждому из классов.

2.6. Классификация обнаженных объектов. В некоторых случаях для решения задач многоклассовой сегментации или детекции структур, образований или патологий, имеющих схожую природу и разделенных пространственно друг от друга (например, переломы ребер, узелки в легких или патологии лимфатических узлов), используются двухстадийные алгоритмы [43, 44]. Результатом работы первого этапа являются локализованные цеплевые структуры, не дифференцированные по различным классам. На втором этапе выполняется решение задачи классификации предсказаний, полученных на первом шаге. В ряде задач такой подход позволяет повысить полноту обнаружения патологий и в то же время увеличить меры качества классификации.

В статье предлагается использовать модель классификации итоговых предсказаний алгоритма детекции переломов ребер не только для разделения переломов на требуемые клинические категории, но и с целью уменьшения количества ложноположительных предсказаний первого шага алгоритма. Выходом такого классификатора для одного предсказания детектора будет являться вектор  $\tilde{p} = [\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4]$  с компонентами, характеризующими вероятности следующих классов: отсутствие перелома, перелом со смещением, перелом без смещения, консолидированный перелом.

Результат работы модели классификации и примеры различных классов переломов с учетом возможного варианта ложноположительного предсказания детектора приведены на рис. 7. Изображение 1 – консолидированный перелом, 2, 3 – переломы без смещения, 5, 6 – переломы со смещением, 4 – ложноположительное предсказание детекционной модели.

2.7. Меры качества детекции и классификации. На различных этапах работы алгоритма присутствуют разные требования к мерам качества предсказаний. Так, на первом шаге локализации патологий необходимо обнаружение как можно большего числа переломов, т.е. максимизация полноты AR (average recall). В силу того, что решается задача детекции, данная мера качества рассчитывается для определенного уровня значения  $IoU$  пересечения предсказаний с реальными локализациями объектов. При вычислении AR используются  $N$  (гиперпараметр алгоритма) предсказаний с наибольшей уверенностью модели:

$$Recall@(\text{IoU} = t) = \frac{TP}{TP + FN}. \quad (2.8)$$

Здесь  $TP$  – количество верно предсказанных локализаций переломов, для которых  $IoU$  с ограничивающим параллелепипедом какого-либо реального перелома больше порога  $t$ ,  $FN$  – коли-



Рис. 7. Примеры различных классов переломов с учетом возможного варианта ложноположительного предсказания детектора.

чество переломов, не обнаруженных на уровне порога  $t$  по величине  $IoU$ . Также рассматривалась мера качества  $mAR$ , являющаяся усреднением  $AR$  для различных порогов пересечения по  $IoU$ :

$$mAR = \sum_{t \in [t_1, \dots, t_n]} \text{Recall}@(\text{IoU} = t). \quad (2.9)$$

Список порогов  $t$ , по которым производится усреднение меры  $AR$ , выбирается исходя из требований к желаемой чувствительности локализации переломов. В данной статье целевыми мерами качества работы первого этапа алгоритма были выбраны мера  $AR$  для  $N = 100$  предсказаний с наибольшей уверенностью алгоритма на уровне значения  $IoU = 0.3$  и мера  $mAR$  с порогами от 0.1 до 0.5 и шагом 0.05. Выбор данных значений порогов для расчета мер полноты обусловлен результатами визуального анализа предсказаний, для которых значение меры  $IoU$  с каким-либо ограничивающим параллелепипедом действительной локализации перелома находилось на уровне от 0.1 до 0.5.

На втором этапе работы алгоритма требуются высокие значения мер точности (precision) и полноты (recall) классификации переломов и объектов, не являющихся переломами, т.е. максимизация макроусредненного по всем классам (отсутствие перелома, перелом со смещением, перелом без смещения, консолидированный перелом) значений данных мер качества:

$$\text{Precision} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FP_i}, \quad (2.10)$$

$$\text{Recall} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FN_i}. \quad (2.11)$$

Также рассматривалась макроусредненная мера качества  $F1$ :

$$F1 = \frac{1}{4} \sum_{i=1}^4 2 \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}. \quad (2.12)$$

Для определения качества детекции применялись меры  $IoU$  и  $Dice$ , вычисленные для  $N$  детекционных предположений с наибольшей уверенностью модели:

$$IoU(y_{true}, y_{pred}) = \frac{|y_{true} \cap y_{pred}|}{|y_{true} \cup y_{pred}|}, \quad (2.13)$$

$$Dice(y_{true}, y_{pred}) = \frac{|y_{true} \cap y_{pred}|}{|y_{true}| + |y_{pred}|}. \quad (2.14)$$

Основной мерой качества работы алгоритмов обнаружения патологий, в том числе переломов ребер, принято считать меру FROC (англ. free-response receiver operating characteristic), оценивающую свободный ответ при различных порогах принятия решения. В качестве порогов рассматриваются значения 0.5, 1, 2, 4 и 8 среднего количества ложноположительных

предсказаний переломов алгоритма на одном КТ-изображении, в качестве ответа – значение чувствительности  $AR$  (2.8) при  $N = 100$  и  $IoU = 0.3$ .

2.8. **Ф о р м и р о в а н и е и т о г о в о г о п р е д с к а з а н и я.** Выбор гиперпараметров алгоритма обнаружения и классификации переломов для формирования финального предсказания системы включает в себя определение следующих величин:  $T_{conf}$  – степень уверенности алгоритма (confidence), по которой происходит отсечение предсказанных локализаций переломов;  $N_{map}$  – максимальное количество детекций, приходящихся на одну карту признаков на каждом уровне с различными пространственными разрешениями;  $T_{nms}$  – уровень порога по  $IoU$  для работы алгоритма non-maximum suppression;  $N_{img}$  – максимальное количество детекций на одно изображение;  $T_{non-frac}$  – значение уровня вероятности не принадлежности ни одному из клинических категорий переломов, предсказанный классификатором.

При заданных гиперпараметрах формирование финального предсказания системы может быть представлено в виде следующего алгоритма:

1. Для каждого уровня карт признаков, на которых осуществляется предсказание переломов ребер, удаляются предсказания со значением уверенности модели меньше  $T_{conf}$ .
2. Для каждого уровня карт признаков сохраняются только  $N_{map}$  предсказаний с наибольшим значением уверенности модели детекции.
3. Для каждого изображения применяется алгоритм non-maximum suppression со значением порога по  $IoU = T_{nms}$ .
4. Для каждого изображения оставляются только  $N_{img}$  предсказаний детектора.

5. Оставляются только те предсказания, для которых уровень вероятности принадлежности к классу, описывающему отсутствие перелома, меньше  $T_{non-frac}$ .

Данные параметры определяются с помощью методов поиска по сетке, генетических алгоритмов или Байесовской оптимизации. Для ускорения поиска формируются предварительные наборы предсказаний, полученные при эмпирически определенных краевых значениях порогов, которые обеспечивают формирование наиболее широкого набора предсказаний. В качестве целевой оптимизируемой меры качества предлагается рассматривать меру FROC.

**3. Набор данных.** Основным открытым набором данных, используемых исследователями при разработке и тестировании методов решения задачи обнаружения и классификации переломов ребер, является датасет RibFrac [20]. Датасет содержит 660 снимков компьютерной томографии области грудной клетки пациентов с различными классами переломов ребер. Каждый КТ-скан RibFrac сопоставлен соответствующим воксельным маскам сегментации переломов ребер, аннотированным одним из четырех клинических категорий переломов: перелом-пряжка, перелом с диастазом, перелом без диастаза или сегментарный перелом. Отдельной категорией в данном наборе исследований выделяются сопутствующие находки и переломы, класс которых не может быть обозначен из-за неопределенности или трудностей диагностики.

Предварительно набор данных разделен авторами на три части – обучающая, валидационная и тестовая. Обучающая часть содержит 420 КТ-изображений, валидационная – 80 и тестовая – 120. Маски переломов ребер для тестовой части датасета являются закрытыми и используются для оценки качества алгоритмов, предложенных участниками соревнования RibFrac Challange [21].

Исходные КТ-изображения из набора данных RibFrac были получены с применением различных КТ-сканеров и настроек их работы и имеют различные значения пространственного разрешения и количества аксиальных срезов. Медианный спейсинг (пространственное расстояние между вокселями в миллиметрах) КТ-изображений в датасете равен, пиксельное разрешение одного среза исследования – 512 на 512. Медианное количество аксиальных срезов на одно исследование равно 357.

В работе была использована частичная выборка изображений из набора данных RibFrac, содержащая 250 исследований, повторно проанализированных и исправленных специалистами-медиками. В процессе валидации были скорректированы и дополнены существующие маски сегментации переломов ребер и добавлена информация о величине диастаза в миллиметрах для переломов со смещением. Более того, было произведено изменение меток классификации переломов на три следующих класса: переломы с диастазом, переломы без диастаза и консолидированные переломы. Корректирование исходной разметки набора данных RibFrac обусловлено присутствием неточностей изначальной диагностики патологий ребер и требованиями, предъявляемыми к разрабатываемой системе.

В клинической практике выделяют три основных типа переломов: переломы с диастазом, переломы без диастаза и консолидированные переломы. Переломы без диастаза – это переломы, которые не вызывают смещения костей. Их бывает сложно идентифицировать при

рентгенографии. Часто переломы ребер без смещения можно обнаружить только при последующих осмотрах, когда уже проявились признаки заживления. Переломы с диастазом – это переломы со значительным отклонением в положении двух частей поврежденного ребра. При возникновении таких переломов могут возникнуть травмы окружающих тканей и структур. Величиной диастаза для данной категории переломов называется расстояние максимального расхождения частей ребра или его фрагментов. Консолидированные переломы – это полное или частичное восстановление целостности кости после ее перелома с формированием сначала первичной, а затем и вторичной костной мозоли.

Дальнейший анализ методов и подходов будет осуществляться с использованием описанной выше выборки данных. На рис. 8. приведены гистограммы распределения количества КТ-изображений для отобранный части исследований из набора данных RibFrac по коронарному и сагиттальному пространственным разрешениям, количеству аксиальных срезов и общему количеству переломов на одно исследование. Статистики рассчитаны до преобразований.

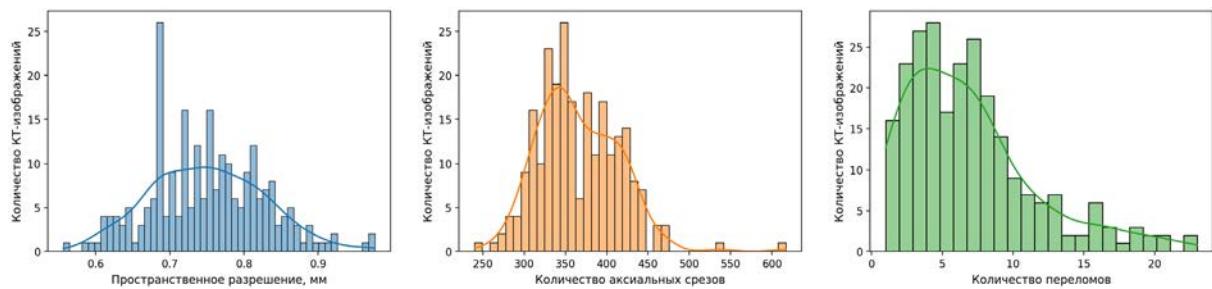


Рис. 7. Гистограммы распределения количества КТ-изображений для отобранный части исследований из набора данных RibFrac.

Для решения задач предварительной обработки данных и исследования моделей локализации, задействующих информацию о структуре ребер, рассматривался датасет RibSeg v2 [22], содержащей воксельные сегментационные маски ребер, разделенные на 24 класса, и информацию о центральных линиях каждого ребра. Разметка в наборе данных RibSeg v2 формировалась на основании исследований датасета RibFrac. Таким образом, каждое КТ-изображение набора данных RibFrac имеет соответствующую разметку ребер из датасета RibSeg v2.

3.1. Формирование отрицательных примеров. Важным аспектом при обучении модели снижения числа ложноположительных предсказаний является генерация отрицательных примеров, представляющих из себя сегменты изображения, не локализующие целевую патологию. Существуют различные способы сэмплирования таких примеров. Некоторые из методов описаны, например, в работе [45].

В качестве основной составляющей числа отрицательных примеров уместно рассматривать ложноположительные предсказания алгоритма детекции переломов ребер на первом шаге. Предсказание разумно считать ложным, если значение  $IoU$  (intersection over union) предсказания меньше 0.01 для любого истинного перелома на изображении. В силу того, что ложноположительные предсказания модели детекции почти всегда выделяют область высокой плотности, локализующую какую-либо часть костной ткани (чаще всего ребер), естественным будет включение в число отрицательных примеров случайно выбранных областей ребер, не локализующих никакой из переломов на изображении ( $IoU < 0.01$ ). Размеры выбираемых из изображения областей могут быть определены случайным образом или через набор базовых параллелепипедов, использующихся в процессе обучения детекционной модели. Дополнительно добавляется некоторая доля от общего числа отрицательных примеров случайно выбранных из изображения произвольных регионов, не содержащих переломы.

Формирование отрицательных примеров может выполняться в динамическом режиме в процессе обучения для каждого изображения обучающей выборки. В совокупности с использованием пространственных аугментаций такой метод повышает разнообразие и вариативность набора отрицательных регионов. Динамическое формирование примеров позволяет учитывать ложноположительные предсказания, возникновение которых обусловлено пространственными аугментациями или изменениями интенсивности, в силу того, что модель детекции применяется к каждому преобразованному изображению заново. Однако данный

подход имеет существенный недостаток, заключающийся в низкой скорости обучения модели в силу необходимости обрабатывать все КТ-исследование целиком на каждой итерации обучения с применением предсказания модели локализации переломов.

В качестве альтернативного варианта генерации отрицательных сэмплов предлагается использовать предварительное выделение определенного набора областей изображения. Для обеспечения вариативности формируемого набора примеров рекомендуется вырезать области с учетом добавления дополнительного случайного отступа к границам исходных параллелепипедов. Данный шаг позволит модели классификации учитывать больше информации об окружающих анатомических структурах относительно применения фиксированного размера ограничивающих параллелепипедов, что в совокупности с использованием пространственных аугментаций увеличивает обобщающую способность модели. При таком подходе предсказание модели детекции переломов осуществляется для каждого изображения только один раз, что в теории может снижать вариативность ложноположительных предсказаний, вызываемых трансформациями исходного КТ-исследования.

Пусть  $I^{(i)}$  – изображение обучающего набора с номером  $i$ , являющееся трехмерной матрицей размерами  $W^{(i)} \times H^{(i)} \times D^{(i)}$ ,  $R_{\text{ribs}}^{(i)}$  – сегментационная маска ребер, сопоставленная данному изображению,  $B_{\text{anchors}}$  – базовый набор ограничивающих параллелепипедов, используемый при обучении модели детекции,  $B_{\text{real}}^{(i)}$  – набор ограничивающих параллелепипедов, описывающий действительные локализации переломов ребер на изображении  $I^{(i)}$ ,  $B_{\text{pred}}^{(i)}$  – предсказанный детектором набор ограничивающих параллелепипедов переломов ребер для изображения  $I^{(i)}$ , отсортированный в порядке убывания уверенности модели.

Коэффициенты  $\alpha = [\alpha_1, \alpha_2]$ ,  $\alpha_1 + \alpha_2 = 1$  определяются как доли количества формируемых отрицательных примеров, являющихся ложноположительными предсказаниями детектора и случайными элементами в областях ребер, и число  $N$  – как общее количество генерируемых отрицательных примеров на одно изображение. Тогда формирование набора отрицательных примеров для обучения модели классификации может быть описано в виде следующего алгоритма:

для изображений  $I^{(i)}$  в обучающей выборке

$$\begin{aligned}
 B &\leftarrow B + \left\{ B_{\text{pred},j}^{(i)} : j \leq \lfloor \pm_1 N \rfloor \right\}, k = 0; \\
 \text{до тех пор, пока } k < \max\left(\left[\alpha_2 N\right], N - \left|B_{\text{pred}}^{(i)}\right|\right), \\
 (x_c, y_c, z_c) &= \mathcal{RS}\left(\left\{(x, y, z) : (x, y, z) \in R_{\text{ribs}}^{(i)}\right\}\right), \\
 (b_x, b_y, b_z) &= \text{Size}(\mathcal{RS}(B_{\text{anchors}})) + \mathcal{U}_{\{x, y, z\}}[0, 10], \\
 \tilde{B} &= \left( x \in \left[ x_c - \frac{b_x}{2}, x_c + \frac{b_x}{2} \right], y \in \left[ y_c - \frac{b_y}{2}, y_c + \frac{b_y}{2} \right], z \in \left[ z_c - \frac{b_z}{2}, z_c + \frac{b_z}{2} \right] \right); \\
 \text{если } \max\left(\left\{ \text{IoU}(\tilde{B}, B_j) : B_j \in B_{\text{real}}^{(i)} \right\} \right) &< 0.01, \text{ тогда } B \leftarrow B + \tilde{B}, k \leftarrow k + 1.
 \end{aligned} \tag{3.1}$$

Здесь  $[x]$  – операция целочисленного округления  $x$  вниз,  $[x]$  – операция целочисленного округления  $x$  вверх,  $\mathcal{RS}$  – операция взятия случайного элемента множества,  $\text{Size}$  – операция вычисления размера параллелепипеда по трем осям, результатом которой является трехмерный вектор,  $\mathcal{U}_{\{x, y, z\}}[a, b]$  – трехмерный вектор, каждый элемент которого взят из равномерного распределения на заданном отрезке  $[a, b]$ .

Помимо положительных примеров из числа размеченных специалистами масок переломов, ограниченных параллелепипедами описанным в разд. 1.2 методом, могут быть использованы предсказания детектора, для которых существует разметка с  $\text{IoU} < 0.5$ . Дополнительное расширение ограничивающих параллелепипедов может быть выполнено по аналогии с методом, применяемым для обработки генерируемых отрицательных регионов. Соотношение общего числа отрицательных и положительных сэмплов является гиперпараметром модели классификации.

**4. Экспериментальная часть.** Основной акцент в данном исследовании сделан на разработке двухстадийной системы детектирования переломов ребер с помощью КТ-изображений органов грудной клетки (ОГК). В процессе исследования рассматривались следующие задачи:

1. Построение архитектуры обнаружения переломов ребер в терминах решения задачи семантической сегментации.

2. Разработка и определение оптимальной конфигурации модели детекции, использующей заданный набор ограничивающих параллелепипедов (anchor-based):

- подбор оптимальных нейросетевых архитектур, примеряемых в качестве базовой модели и модуля комбинирования признаковых описаний различных пространственных разрешений;
- добавление и конфигурирование дополнительного выхода детекционной модели, использующегося для решения задачи семантической сегментации ребер на этапе обучения;
- определение оптимальной конфигурации гиперпараметров выбора ограничивающих параллелепипедов на этапах обучения и предсказания.

3. Построение модели классификации переломов и уменьшения числа ложноположительных предсказаний, полученных на первом шаге. Выбор оптимальной стратегии формирования отрицательных примеров, применяемых при обучении модели классификации.

4. Подбор оптимальных гиперпараметров всей системы локализации и классификации переломов ребер.

5. Сравнение результатов полученной системы с существующими методами решения задачи сегментации переломов ребер с точки зрения мер качества чувствительности и FROC.

4.1. П р е д о б р а б о т к а и а у г м е н т а ц и я д а н н ы х . Для обучения моделей использовался выделенный и провалидированный врачами-рентгенологами набор из 250 КТ-исследований датасета RibFrac, предварительно обработанный алгоритмом, который описан в разд. 2.1. В качестве обучающего набора были случайным образом отобраны 80% изображений. Оставшиеся 20% КТ-сканов применялись для валидации. Пространственное разрешение исследований данного набора – (1.25, 0.75, 0.75), медианный размер одного изображения в осях,  $Oz$ ,  $Oy$ ,  $Ox$  равен (392, 277, 284). Предварительно каждое изображение нормировалось в диапазоне от –240 до 1040 HU, что являлось 0.025 и 0.975 квантилями распределения плотности в деснитометрических показателях, рассчитанной в областях переломов по всем изображениям. При обучении моделей рассматривались случайные сегменты изображения размерами  $192 \times 192 \times 192$  вокселей, содержащие сегментационные маски переломов ребер.

В качестве аугментаций исходного набора данных использовались случайные изменения интенсивности методами сдвига и шкалирования, добавление случайного Гауссова шума и Гауссова сглаживания. Пространственными аугментациями выступали случайные аффинные преобразования с максимальным углом поворота по каждой оси  $\pi/15$  и максимальным относительным растяжением (сжатием) 0.15, а также случайное приближение или отдаление с амплитудой в диапазоне [0.9, 1.1]. Для изображений валидационного набора применялось только нормирование плотности по шкале Хаунсфилда.

Для обучающей части набора данных нормировка КТ-изображения по диапазону  $HU \in [-240, 1040]$  осуществлялась после применения аугментаций интенсивности. Выделение сегмента размерами  $192 \times 192 \times 192$  вокселей происходило после использования пространственных трансформаций. Аналогичные пространственные преобразования применялись для сегментационных масок ребер и переломов ребер. При помощи трансформаций данного вида на КТ-изображениях использовался метод трилинейной интерполяции, на сегментационных масках – интерполяция методом ближайших соседей. Вычисление ограничивающих параллелепипедов для преобразованных в процессе применения аугментаций сегментационных масок производилось с помощью алгоритма, который представлен в разд. 1.2.

4.2. Р е ш е н и е з а д а ч и с е г м е н т а ц и и . Изначальный подход к решению задачи локализации переломов ребер был реализован с помощью методов семантической сегментации. В качестве базовых архитектур рассматривались модели, описанные в разд. 2.2.

Обучение моделей производилось в течение 300 эпох на видеокарте Nvidia RTX 3090 с 24 Гб оперативной памяти. Оптимизация выполнялась с помощью градиентного спуска методом Adam с параметрами  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  и коэффициентом регуляризации весов  $3e - 5$ . Скорость обучения  $1e - 3$ . Обучение производилось с использованием смешанной точности вычислений с плавающей точкой, исходные данные преобразовывались в формат половинной точности. Размер пакета (батча) – 1. Формирование результата на этапе предсказания модели осуществлялось с применением метода скользящего окна, соответствующего размерам регионов изображения, используемых в процессе обучения – вокселей  $192 \times 192 \times 192$ . Доля линейного перекрытия между различными окнами – 0.625.

Полученные предсказания моделей сегментации обрабатывались в два шага: на первом этапе предсказания корректировались с помощью морфологических операций замыкания и размыкания, на втором этапе удалялись сегментационные маски объемом менее 50 вокселей с целью исключения шумов. В процессе эксперимента были рассмотрены варианты обучения моделей с использованием дополнительного канала, содержащего информацию о вокселях КТ-изображения, которые соответствуют сегментационной маске ребер.

Вычисление мер качества FROC и Recall производилось на уровне значения IoU, равном 0.2. Результаты соответствующих экспериментов приведены в табл. 1.

**Таблица 1.** Результаты сравнения сегментационных моделей с учетом и без учета использования дополнительного канала с масками сегментации ребер

Канал для ребер	Модель	FROC	Recall	Dice
Не использовался	SegResNet	0.647	0.771	0.567
	Swin UNETR	0.445	0.56	0.319
	nnU-Net	0.601	0.691	0.54
Использовался	SegResNet	<b>0.683</b>	<b>0.795</b>	<b>0.594</b>
	Swin UNETR	0.489	0.61	0.413
	nnU-Net	0.624	0.725	0.553

Полученные меры качества детекции были расценены как неудовлетворительные в рамках решения данной задачи. В связи с этим было принято решение изменить подход к формализации задачи и переформулировать ее в терминах решения задачи детекции. Подробно переход описан в разд. 1.2.

**4.3. О бу чен ие д ет ек ци он н ой м од ели.** Начальным шагом разработки системы обнаружения переломов ребер являлось экспериментальное определение оптимальной архитектуры модели детекции, использующей заданный набор ограничивающих параллелепипедов.

Сопоставление реальных примеров переломов с предсказаниями модели в процессе обучения детектора производилось с помощью алгоритма адаптивного выбора обучающей выборки ATSS (англ. adaptive training sample selection) [29], который представляет собой метод автоматического определения положительных и отрицательных выборок в соответствии со статистическими характеристиками ограничивающих параллелепипедов. Рассматривался вариант алгоритма ATSS с дополнительным требованием на нахождение центра предсказанного параллелепипеда внутри ограничивающего параллелепипеда реального перелома для отнесения объекта к числу положительных примеров.

Определение набора положительных и отрицательных примеров, по которым в дальнейшем будет производиться расчет классификационной и регрессионной функций потерь, производится с помощью стратегии жесткой отрицательной выборки [46]. Размер пакета (батча) на изображение был выбран равным 64, минимальное количество отрицательных примеров, выбираемых случайным образом из 160 отрицательных примеров с наибольшим значением уверенности модели, равен 8, соотношение количества положительных и отрицательных примеров – 0.3.

Вычисление оптимального набора ограничивающих параллелепипедов было выполнено с помощью алгоритма, предложенного авторами nnDetection [47]. Данный алгоритм максимизирует значение *IoU* между набором заданных параллелепипедов и реальными локализациями переломов ребер по всем изображениям обучающей выборки данных. Базовым набором выбрали список из шести параллелепипедов.

Обучение моделей производилось в течение 250 эпох на видеоускорителе A100 с 80 Гб оперативной памяти. Оптимизация выполнялась с помощью стохастического градиентного спуска SGD (англ. stochastic gradient descent) совместно с использованием накопления момента с коэффициентом при экспоненциальном скользящем среднем  $\beta = 0.9$  и коэффициентом регуляризации весов  $3e - 5$ . Скорость обучения  $1e - 3$ . Все расчеты производились с применением смешанной точности вычислений с плавающей точкой. Исходные данные предварительно преобразовывались в формат половинной точности *fp16*. Размер пакета (батча) варьировался от 1 до 2 в зависимости от модели.

Формирование результата на этапе предсказания модели осуществлялось с применением метода скользящего окна, соответствующего размерам регионов изображения, используемых в процессе обучения –  $192 \times 192 \times 192$  вокселей. Доля линейного перекрытия между различными окнами составляла 0.5.

**4.4. О п ред ел ен ие оптимальной архитектуры.** На первом этапе выбора детекционной модели были реализованы 3D-модификации архитектур RetinaNet и EfficientDet с использованием в качестве базовых моделей ResNet, SEResNeXt, High-Resolution Net

и EfficientDet (B1-B5) с базовой сетью EfficientNet (B1-B5) соответственно. Сверточные ядра первого слоя в моделях ResNet, SEResNeXt и EfficientNet были заменены на ядра размером  $3 \times 3 \times 3$  и шагом  $1 \times 1 \times 1$  с целью повышения пространственного разрешения карт признаков. Эмпирически было определено, что данная замена оказывает значимое влияние на качество работы детекционной модели. В качестве модулей агрегации признаков различных пространственных разрешений исследовались FPN и BiFPN с привлечением различного числа выходных слоев с базовых моделей, количество которых варьировалось от 2 до 4. Количество карт признаков, применяемых в выходных модулях классификации и регрессии на ограничивающих параллелепипедах, варьировалось в диапазоне от 3 до 5. В исследовании рассматривались лишь некоторые наиболее осмысленные и относительно различные варианты комбинаций модулей и используемых слоев из возможных. Основной мерой качества, по которой происходил выбор оптимальной архитектуры, являлось значение полноты (Recall) на уровне  $IoU = 0.3$  – сокращенно  $Recall@(IoU = 0.3)$ .

Изначально были зафиксированы значения количества выходных слоев из базовых моделей и из модулей извлечения признаков на величинах 3 и 4 соответственно, анализировались только архитектуры. Данный выбор обусловлен анализом размера карт признаков на этих слоях. Для входного изображения размером  $192 \times 192 \times 192$  вокселей размеры карт признаков с выходных слоев 1–4 модулей извлечения признаков равны  $96 \times 96 \times 96$ ,  $48 \times 48 \times 48$ ,  $24 \times 24 \times 24$  и  $12 \times 12 \times 12$  соответственно. Рассмотрение карт признаков с более низким пространственным разрешением или использование более глубоких слоев базовых моделей на данном этапе исследования не было целесообразно.

Сравнение в данном эксперименте выполнялось с помощью детекционных мер FROC и  $Recall@(IoU = 0.3)$ , а также мер качества локализации IoU и Dice. Результаты эксперимента приведены в табл. 2.

**Таблица 2.** Результаты сравнения базовых архитектур и модулей комбинирования карт признаков различного пространственного разрешения

Модуль агрегации	Базовая модель	FROC	Recall	Dice	IoU
FPN	ResNet-50	<b>0.815</b>	0.887	0.687	0.523
	SEResNeXt-50	0.801	0.903	0.683	0.519
	HRNet-50	0.785	0.892	0.678	0.513
BiFPN	ResNet-50	0.798	0.8933	0.694	0.531
	<b>SEResNeXt-50</b>	0.809	<b>0.906</b>	<b>0.703</b>	<b>0.542</b>
	EfficienNet-B1	0.765	0.887	0.664	0.497
	EfficienNet-B3	0.771	0.9024	0.667	0.501
	EfficienNet-B5	0.733	0.854	0.657	0.489

Исследование влияния количества выходных слоев из базовых моделей и количества слоев модуля комбинирования карт признаков различных пространственных разрешений было проведено для сети ResNet с модулем FPN, SEResNeXt с модулем BiFPN и архитектуры EfficientDet-B3 (EfficientNet-B3 с BiFPN). Для блока FPN различие между числом входных и выходных признаков равно 1, так как в качестве дополнительного используется только 1-й слой, получаемый применением операции MaxPooling к последнему выходу базовой модели. Нумерация слоев производилась от выхода с наибольшим пространственным разрешением карт признаков. Сравнение в данном эксперименте проводилось только с точки зрения меры  $Recall@(IoU = 0.3)$ . Результаты приведены в табл. 3.

4.5. Использование дополнительного выхода сегментации. Второй этап исследования архитектуры детекционной модели включал определение эффекта применения дополнительного выхода модели для решения задачи семантической сегментации ребер. Сравнение производилось для архитектур ResNet-50 с модулем FPN и SEResNeXt-50 с модулем BiFPN. Количества выходных слоев базовой модели и модулей извлечения признаков были выбраны равными 3 и 4 соответственно на основании результатов предыдущего эксперимента. В случае использования модуля BiFPN для сегментационного выхода сети применялись выходы со всех 5 его слоев. Результаты эксперимента приведены в табл. 4.

**Таблица 3.** Результаты сравнения различного числа выходных слоев из базовых моделей и количества признаковых описаний из модулей FPN и BiFPN по мере  $Recall @ (IoU = 0.3)$

Слои базовой модели, слои модуля комбинирования					
Архитектура	(2, 3)	(2, 4)	(2, 5)	(3, 4)	(3, 5)
ResNet + FPN	0.8902	—	—	0.8872	—
SEResNeXt + BiFPN	<b>0.8971</b>	0.8993	0.8734	<b>0.9063</b>	0.8913
EfficientNet + BiFPN	0.8811	—	—	<b>0.9024</b>	0.8689

**Таблица 4.** Результаты сравнения архитектур различных конфигураций дополнительного выхода сегментации ребер

Дополнительный выход	Архитектура	FROC	Recall	Dice	IoU
Не используется	ResNet	0.815	0.887	0.687	0.523
	SEResNeXt	0.809	0.906	0.703	0.542
Upsampling	ResNet	0.829	0.934	0.697	0.535
	SEResNeXt	0.816	0.921	0.711	0.551
U-Net	ResNet	0.83	0.939	0.698	0.536
	SEResNeXt	0.818	0.927	0.709	0.549
Attention U-Net	ResNet	<b>0.839</b>	0.944	0.701	0.54
	SEResNeXt	0.837	<b>0.948</b>	<b>0.714</b>	<b>0.556</b>

Можно видеть, что при добавлении дополнительного выхода для сегментации ребер наблюдается существенное повышение меры полноты детекции переломов. Наилучший результат получается при применении модификации сети Attention U-Net.

4.6. Финальная архитектура детектора. В качестве финальной детекционной модели предлагается использовать anchor-based детектор, основанный на 3D-реализации сети SEResNeXt-50 с модулем извлечения признаков BiFPN и дополнительным сегментационным выходом, представляющим из себя модификацию сети Attention U-Net. В качестве выхода из модели SEResNeXt выбираются карты признаков со слоев 1, 2 и 3 с количеством каналов в слоях 256, 512 и 1024 соответственно. Дополнительные карты признаков для входа в модуль BiFPN получаются путем добавления сверточных блоков на выходе последнего слоя базовой модели. В модулях классификации и регрессии на ограничивающих параллелепипедах на вход подаются карты признаков с 1-го по 4-й слой модуля BiFPN. Модулями, преобразующими карты признаков в логиты классификации и регрессии на ограничивающих параллелепипедах, выступают сверточные блоки, состоящие из 4 слоев. В модуле сегментации ребер используются выходы со всех 5 слоев BiFPN.

4.7. Генерация положительных и отрицательных примеров. В рамках разработки и обучения модели классификации предсказаний переломов ребер, полученных на предыдущем шаге работы алгоритма, рассматривались различные стратегии формирования отрицательных и положительных примеров (разд. 3.1). В эксперименте рассматривались два варианта генерации отрицательных обучающих примеров. В обоих вариантах в качестве положительных примеров были выбраны реальные локализации переломов ребер и корректные предсказания детекционной модели с  $IoU > 0.5$ . Варианты генерации отрицательных примеров, применяемых в процессе обучения классификатора, могут быть описаны следующим образом:

1) формирование отрицательных примеров только из числа ложноположительных предсказаний модели, т.е. таких, для которых значение  $IoU$  с любым истинным параллелепипедом, ограничивающим перелом ребра, составляет менее 0.05;

2) формирование отрицательных примеров в виде комбинации ложноположительных предсказаний детекционной модели, случайных областей ребер, не локализующих никакой

из переломов на изображении (для которых значение  $IoU$  с любой истинной локализацией равно 0), а также некоторой доли случайно выбранных областей изображений.

В случае, когда для генерации примеров не использовались предсказания детекционной модели, размер выбираемой из изображения области определялся с помощью алгоритма из разд. 3.1. с применением набора шести предварительно заданных параллелепипедов, применяемых при обучении модели локализации переломов.

4.8. О б у ч е н и е м о д е л и к л а с с и ф и к а ц и и. В качестве классификационной модели использовалась сверточная сеть ResNet-18. Выбор данной архитектуры обусловлен малыми размерами ограничивающих параллелепипедов, по которым производилось выделение регионов изображения. Рассматривались две стратегии обучения модели-классификатора.

1. Динамическая – обучение с помощью пакетов (батчей), формируемых на основании положительных и отрицательных примеров, генерируемых по изображению в моменте обучения на каждой итерации.

2. Статическая – обучение с использованием случайного сэмплирования из заранее сформированного набора положительных и отрицательных примеров по всем изображениям обучающей выборки. На каждое изображение формировалось до 100 отрицательных примеров при применении варианта генерации с помощью случайных областей ребер.

Положительные и отрицательные примеры объединялись в 1 батч размером 16, каждый элемент данного набора дополнялся до максимального размера изображения в батче (с учетом делимости на 16) симметричным нулевым отступом. Соотношение положительных и отрицательных примеров равно 0.3, минимальное число отрицательных примеров в одном батче – 8. В случае, когда количество изначально полученных отрицательных или положительных примеров было недостаточно для удовлетворения поставленных условий, они дополнялись аугментированными вариантами изображений соответствующего класса.

В качестве функции потерь использовалась перекрестная энтропия с весами классов, обратно пропорциональных мощности классов в обучающих примерах. Меры качества, представленные в табл. 5, найдены путем макроусреднения соответствующих мер по всем классам.

**Таблица 5.** Результаты сравнения методов обучения классификационной модели для различных вариантов формирования положительных и отрицательных примеров

Метод генерации	Метод генерации	Precision	Recall	F1
Регионы ребер	Динамический	0.744	0.785	0.763
	Статический	0.732	0.761	0.748
Предсказания детектора	Динамический	—	—	—
	Статический	0.753	0.797	0.773
Предсказания + ребра	Динамический	—	—	—
	Статический	<b>0.756</b>	<b>0.801</b>	<b>0.778</b>

Применение ложноположительных предсказаний детектора при формировании отрицательных примеров на каждой итерации обучения (динамический метод) не осуществлялось из-за необходимости использования детекционной модели на каждом шаге, что приводило к значительному увеличению времени обучения модели классификации.

По результатам эксперимента наиболее оптимальной стратегией формирования отрицательных примеров для обучения классификатора является применение ложноположительных предсказаний детектора в совокупности со случайными регионами с центром в точках ребер. Меры качества классификации для различных клинических категорий переломов с использованием стратегии смешанной генерации отрицательных примеров во время обучения приведены в табл. 6.

4.9. П о д б о р г и п е р п а р а м е т р о в с и с т е м ы. Выбор гиперпараметров моделей детекции и классификации предполагает определение оптимальных значений порогов (разд. 2.8). В целях ускорения проведения вычислительных экспериментов предварительно производилась генерация предсказаний для граничных значений  $T_{conf} = 0.05$ ,  $N_{map} = 10\,000$ ,  $N_{img} = 1000$  при различных значениях порога для алгоритма non maximum suppression  $T_{nms} \in [0.3, 0.4, 0.5, 0.6]$ . Дальнейший подбор оптимальных гиперпараметров выполнялся ме-

**Таблица 6.** Меры качества классификации для различных клинических категорий переломов и ложноположительных предсказаний

Клиническая категория	Precision	Recall	F1
С диастазом	0.905	0.849	0.876
Без диастаза	0.643	0.67	0.656
Консолидированный	0.552	0.671	0.606
Не перелом	0.986	0.981	0.983

тодом поиска по сетке с целевой функцией оптимизации FROC. Значение порога  $T_{non-frac}$  определялось эвристически. В ходе эксперимента были выбраны следующие значения соответствующих параметров:

$$T_{conf} = 0.83, N_{map} = 1000, T_{nms} = 0.5, N_{img} = 100, T_{non-frac} = 0.4.$$

4.10. Сравнение результатов. Большинство существующих подходов ориентированы на решение задачи сегментации переломов ребер, поэтому сравнение данных методов с предлагаемым подходом с точки зрения мер *IoU* или *Dice* не информативно. Общее качество системы обнаружения переломов наилучшим образом описывается мерами качества FROC, максимальной чувствительностью при заданном минимально допустимом значении *IoU* и средним количеством ложноположительных предсказаний на изображение. Пороги, на основании которых производится усреднение меры FROC, были выбраны равными 0.5, 1, 2, 4 и 8 в соответствии со значениями данных порогов, предлагаемых участникам соревнования RibFrac Challange. Порог по значению *IoU* был выбран на уровне 0.3. Для сравнения авторами [9] предлагается использовать порог  $IoU = 0.2$ . Более высокое значение порога по *IoU* при решении задачи детекции относительно задачи сегментации обусловлено различиями в объемах перекрытия прямоугольных и произвольных криволинейных сегментационных областей.

Важно отметить, что прямое сравнение методов в данном случае не будет корректным, так как обучение и тестирование производились для различных частей набора данных RibFrac. Более того, разметка данных при разработке FracNet++ [9] и предлагаемого метода дополнительно корректировалась медиками. Наиболее показательным являлось бы сравнение методов на закрытой части датасета RibFrac, однако соревнование на данный момент времени закрыто для участия.

Результаты сравнения предлагаемого метода с алгоритмами сегментации переломов ребер FracNet++ [9] и SA-FracNet [12] с точки зрения детекционных мер качества приведены в табл. 7.

**Таблица 7.** Результаты сравнения предложенного метода с методами сегментации переломов FracNet++ [8] и SA-FracNet [12] (сравнение было выполнено на различных частях набора данных RibFrac с учетом дополнительного корректирования разметки и не может быть рассмотрено как прямое сравнение методов)

Архитектура	Detection FROC	Max Sensitivity	Average FP
Предлагаемый метод (1 шаг)	0.809	<b>0.948</b>	24.3
Предлагаемый метод (2 шага)	<b>0.837</b>	0.935	<b>4.7</b>
FracNet++	0.83	0.898	17.73
SA-FracNet	—	0.926	—

**Заключение.** Предложен и реализован двухстадийный метод решения задачи сегментации переломов ребер, в котором обнаружение и классификация разделены по различным стадиям. На первом этапе производится детектирование локализаций переломов ребер с помощью anchor-based модели, основанной на 3D-модификации нейросети SEResNeXt с модулем комбинирования признаков различных пространственных разрешений BiFPN и дополнительным выходом в виде 3D Attention U-Net, используемым для предсказания семантических масок

ребер на этапе обучения модели. Результатом работы первого шага алгоритма являются предварительные предсказания переломов, локализованные параллелепипедами в трехмерном пространстве КТ-изображения. На втором этапе предсказанные локализации переломов классифицируются с целью определения клинической категории перелома ребра и уменьшения числа ложноположительных предсказаний детектора. Более того, предлагается метод формирования отрицательных примеров для обучения модели классификации, обеспечивающий оптимальную скорость работы и повышающий точность обнаружения ложноположительных предсказаний детекционной модели.

Экспериментально демонстрируется, что разработанная система обеспечивает высокое качество обнаружения переломов ребер на отобранной и дополнительно скорректированной части набора данных RibFrac с максимальным значением чувствительности обнаружения 0.935 при среднем количестве ложноположительных предсказаний на одно изображение, равном 4.7.

### СПИСОК ЛИТЕРАТУРЫ

1. *Topker M., Ringl H., Lazar M.* The Ribs Unfolded – a CT Visualization Algorithm for Fast Detection of Rib Fractures: Effect on Sensitivity and Specificity in Trauma Patients // *Eur Radiol.* 2015. V. 25. P. 1865–1874.
2. *Urbaneja A., De Verbijer J., Formery A. S. et al.* Automatic Rib Cage Unfolding with CT Cylindrical Projection Reformat in Polytraumatized Patients for Rib Fracture Detection and Characterization: Feasibility and Clinical Application // *European J. Radiology.* 2019. V. 110. P. 121–127.
3. *Jang Y.C., Lee C.W., Huang C.C.* Diagnostic Accuracy for Acute Rib Fractures: A Cross-sectional Study Utilizing Automatic Rib Unfolding and 3D Volume-Rendered Reformation // *Acad. Radiol.* 2024. V. 31. № 4. P. 1538–1547.
4. *Erdemir A., Onur M.R., Idilman I. et al.* Pros and Cons of Rib Unfolding Software: a Reliability and Reproducibility Study on Trauma Patients // *Turkish J. Trauma and Emergency Surgery.* 2023. V. 29. № 6. P. 717–723.
5. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation // *arXiv preprint.* 2015. *arXiv:1505.04597.*
6. *Chen L.C., Zhu Y.* Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // *Computer Vision – ECCV.* 2018. V. 11211. P. 833–851.
7. *Kaiming H.* Deep Residual Learning for Image Recognition // *arXiv preprint.* 2015. *arXiv:1512.03385.*
8. *Jin L., Yang J., Kuang K. et al.* Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet // *EBioMedicine.* 2020. V. 62. № 12. P. 103106.
9. *Yang J., Shi R., Jin. L. et al.* Deep Rib Fracture Instance Segmentation and Classification from CT on the RibFrac Challenge // *arXiv preprint.* 2024. *arXiv:2402.09372.*
10. *Charles R.Q.* PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space // *arXiv preprint.* 2017. *arXiv:1706.02413.*
11. *Wang Y., Sun. Y., Liu Z. et al.* Dynamic Graph CNN for Learning on Point Clouds // *ACM Transactions on Graphics.* 2019. V. 38. № 5. P. 146.
12. *Cao Z., Xu L., Chen D. et al.* A Robust Shape-Aware Rib Fracture Detection and Segmentation Framework With Contrastive Learning // *IEEE Transactions on Multimedia.* 2023. V. 1. P. 1–8.
13. *Wu M., Chai Z., Qian G. et al.* Development and Evaluation of a Deep Learning Algorithm for Rib Segmentation and Fracture Detection from Multicenter Chest CT Images // *Radiology: Artificial Intelligence.* 2021. V. 3. № 9. P.e200248.
14. *Isensee F., Wald T., Ulrich C. et al.* nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation // *arXiv preprint.* 2024. *arXiv:2404.09556.*
15. *Cardoso M.J., Li W., Brown R. et al.* MONAI: An Open-source Framework for Deep Learning in Healthcare // *arXiv preprint.* 2022. *arXiv:2211.02701.*
16. *Hatamizadeh A., Nath V., Tang Y. et al.* Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images // *arXiv preprint.* 2022. *arXiv:2201.01266.*
17. *Myronenko A.* 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization // *arXiv preprint.* 2018. *arXiv:1810.11654.*
18. *Huang Z., Wang H., Deng Z. et al.* STU-Net: Scalable and Transferable Medical Image Segmentation Models Empowered by Large-Scale Supervised Pre-training // *arXiv preprint.* 2023. *arXiv:2304.06716.*
19. *D'Antonoli T.A., Berger L.K., Indrakanti A.K. et al.* TotalSegmentator MRI: Sequence-Independent Segmentation of 59 Anatomical Structures in MR Images // *arXiv preprint.* 2024. *arXiv:2405.19492.*
20. *Jin L., Yang J., Kuang K. et al.* Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet // *EBioMedicine.* 2020. V. 62. № 12. P. 103106.
21. *Yang J., Shi R., Jin L. et al.* Deep Rib Fracture Instance Segmentation and Classification from CT on the RibFrac Challenge // *arXiv preprint.* 2024. *arXiv:2402.09372.*
22. *Jin L., Gu S., Wei D. et al.* RibSeg v2: A Large-scale Benchmark for Rib Labeling and Anatomical Centerline Extraction // *IEEE Transactions on Medical Imaging (TMI).* 2024. V. 43. № 1. P. 570–581.
23. *Yang J., Gu S., Wei D. et al.* RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans // *Intern. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI).* Strasbourg, 2021. P. 611–621.
24. *Li C., Zia M.Z., Tran Q.H. et al.* Deep Supervision with Intermediate Concepts // *arXiv preprint.* 2018. *arXiv:1801.03399.*

25. *He Z., Nath V., Yang D. et al.* SwinUNETR-V2: Stronger Swin Transformers with Stage-wise Convolutions for 3D Medical Image Segmentation // Medical Image Computing and Computer Assisted Intervention (MICCAI). Vancouver, 2023. P. 416–426.
26. *Drozdal M., Vorontsov E., Chartrand G. et al.* The Importance of Skip Connections in Biomedical Image Segmentation // arXiv preprint. 2016. arXiv:1608.04117.
27. *Redmon J., Divvala S., Girshick R. et al.* You Only Look Once: Unified, Real-Time Object Detection // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016. P. 779–788.
28. *Tian Z., Shen C., Chen H. et al.* FCOS: Fully Convolutional One-Stage Object Detection // IEEE/CVF Intern. Conf. on Computer Vision (ICCV). Seoul, 2019. P. 9626–9635.
29. *Zhang S., Chi C., Yao Y. et al.* Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020. P. 9756–9765.
30. *Lin T.Y., Goyal P., Girshick R. et al.* Focal Loss for Dense Object Detection // IEEE Intern. Conf. on Computer Vision (ICCV). Venice, 2017. P. 2999–3007.
31. *Tan M., Pang R., Quoc V.L.* EfficientDet: Scalable and Efficient Object Detection // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020. P. 10778–10787.
32. *Hu J., Shen L., Albanie S. et al.* Squeeze-and-Excitation Networks // IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City, 2018. P. 7132–7141.
33. *Xie S., Girshick R., Dollar P. et al.* Aggregated Residual Transformations for Deep Neural Networks // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017. P. 5987–5995.
34. *Wang J., Sun K., Cheng T. et al.* Deep High-Resolution Representation Learning for Visual Recognition // arXiv preprint. 2019. arXiv: 1908.07919.
35. *Tan M., Le Q.V.* EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // arXiv preprint. 2019. arXiv: 1905.11946.
36. *Lin T.Y., Dollar P., Girshick R.B. et al.* Feature Pyramid Networks for Object Detection // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017. P. 936–944.
37. *Liu S., Johns E., Davison A.J.* End-To-End Multi-Task Learning With Attention // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019. P. 1871–1880.
38. *Zhou Z., Siddiquee M., Tajbakhsh N. et al.* UNet++: A Nested U-Net Architecture for Medical Image Segmentation // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 2018. V. 11045. P. 3–11.
39. *Oktay O., Schlemper J., Folgoc L.L. et al.* Attention U-Net: Learning Where to Look for the Pancreas // arXiv preprint. 2018. arXiv: 1804.03999.
40. *Ioffe S., Szegedy C.* Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv preprint. 2015. arXiv: 1502.03167.
41. *Fu C.Y., Shvets M., Berg A.C.* RetinaMask: Learning to Predict Masks Improves State-of-the-art Single-Shot Detection for Free // arXiv preprint. 2019. arXiv: 1901.03353.
42. *Sudre C.H., Li W., Vercauteren T. et al.* Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations // Lecture Notes in Computer Science. 2017. V. 2017. P. 240–248.
43. *Wang C., MacGillivray T., Macnaught G. et al.* A Two-stage 3D Unet Framework for Multi-class Segmentation on Full Resolution Image // arXiv preprint. 2018. arXiv: 1804.04314.
44. *Jeon U., Kim H., Hong H., Wang J.H.* Two-stage Meniscus Segmentation Framework Integrating Multiclass Localization Network and Adversarial Learning-based Segmentation Network in Knee MR Images // Medical Imaging 2021: Computer-Aided Diagnosis. 2021. V. 11597. P. 1159714.
45. *Jin D., Ma X., Zhang C. et al.* Towards Overcoming False Positives in Visual Relationship Detection // arXiv preprint. 2020. arXiv:2012.12510.
46. *Shi W., Chen J., Feng F. et al.* On the Theories Behind Hard Negative Sampling for Recommendation // Proc. ACM Web Conf. Austin, 2023. P. 812–822.
47. *Baumgartner M., Jaeger P.F., Isensee F. et al.* NnDetection: A Self-configuring Method for Medical Object Detection // Lecture Notes in Computer Science. 2021. V. 12905. P. 530–539.

## МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЧЕСКОГО ВЫЯВЛЕНИЯ НОВЫХ ФУНКЦИОНАЛЬНЫХ ПАТТЕРНОВ В МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДАХ

© 2024 г. А. И. Майсурадзе<sup>a, \*</sup>, Л. С. Сидоров<sup>a, \*\*</sup>

<sup>a</sup>МГУ им. М.В. Ломоносова, Москва, Россия

\*e-mail: maysuradze@cs.msu.ru

\*\*e-mail: leon.sidorov@gmail.com

Поступила в редакцию 16.05.2024 г.

После доработки 09.06.2024 г.

Принята к публикации 16.10.2024 г.

Предлагается метод и конкретная архитектура модели машинного обучения, которая помогает исследователям в различных областях автоматически выявлять функциональные закономерности в многомерных временных рядах из серии экспериментов. Исходная задача была формализована в терминах машинного обучения, поэтому у исследователей нет необходимости быть профессионалами в рассматриваемой предметной области. Эффективность метода была продемонстрирована в области нейрофизиологии для данных, где уже известно о существовании приведенного паттерна Р300. Для дальнейших исследований представляет интерес обобщить предложенный метод на другие области, например данные с датчиков на производственных линиях или банковские транзакции.

**Ключевые слова:** многомерные временные ряды, функциональные паттерны, электроэнцефалограмма, волна Р300

**DOI:** 10.31857/S0002338824060106, **EDN:** SUBTIL

## MACHINE LEARNING MODELS TO AUTOMATICALLY DISCOVER NOVEL FUNCTIONAL PATTERNS IN MULTIVARIATE TIME SERIES

А. И. Майсурадзе<sup>a, \*</sup>, Л. С. Сидоров<sup>a, \*\*</sup>

<sup>a</sup>Lomonosov Moscow State University, Moscow, Russia

\*e-mail: maysuradze@cs.msu.ru

\*\*e-mail: leon.sidorov@gmail.com

In this paper, we propose a method and a specific architecture for a machine learning model that assists researchers across various fields in automatically identifying functional patterns in multivariate time series from a series of experiments. The initial problem was formalized in terms of machine learning, eliminating the need for researchers to be experts in the specific subject matter under examination. The effectiveness of the method has been demonstrated in the field of neurophysiology with data where the existence of the P300 pattern is already known. For further research, it would be beneficial to generalize the proposed method to other areas, such as sensor data from production lines or banking transactions.

**Keywords:** multivariate time series, functional patterns, electroencephalogram, P300 wave

**Введение.** Во многих предметных областях исследуемые данные имеют вид многомерных временных рядов. Это могут быть показания датчиков на производственных линиях, акции на фондовом рынке или истории денежных транзакций. Иногда такие временные ряды возникают как результат дополнительной обработки, например видео может преобразовываться в пу-

чок траекторий специфических точек. При этом на многомерных временных рядах решаются как традиционные задачи классификации, так и специфические задачи поиска аномалий или точки изменения тренда. Соответственно, для всех этих прикладных областей и задач разрабатывают различные модели машинного обучения, которые принимают на вход многомерные временные ряды.

В наши дни исследователи ждут от моделей машинного обучения не просто решения прикладных задач, но и интерпретируемости полученных результатов. Модели должны явно демонстрировать пользователям закономерности, которые они нашли. При этом анализ многомерных временных рядов во многих случаях сводится к выявлению так называемых функциональных паттернов [1], т.е. особенностей поведения временного ряда, соответствующих некоторым интересующим исследователей состояниям системы. На протяжении многих лет эксперты из различных предметных областей уже выявили некоторое количество функциональных паттернов. Основная задача данной работы – разработка методики автоматического поиска подобных паттернов в многомерных временных рядах без обладания какими-либо априорными знаниями о предметной области. Мы хотим предложить модель, которая как частный случай найдет уже известный паттерн, что подтвердит выдвинутую гипотезу. Развивая подобный подход, в будущем исследователи смогут использовать модели машинного обучения, чтобы быстрее находить новые виды паттернов и закономерностей в исходных данных многомерных временных рядов.

В работе в качестве предметной области взята нейрофизиология. Соответственно, работа методики будет продемонстрирована для выявления одного из самых известных функциональных паттернов в анализе электроэнцефалограммы (ЭЭГ), а именно волны Р300. Функциональные паттерны, на которые сделан упор в исследовании, служат обобщением данного понятия для случая произвольного многомерного временного ряда.

Паттерн ЭЭГ – это целостная картина биоэлектрической активности по всем исследуемым областям мозга, отражающая особенности распределения различных компонентов ЭЭГ. Он является качественной характеристикой ЭЭГ, отражающей функциональное состояние головного мозга человека.

Существуют нормальные паттерны, которые сопровождают нас в состоянии покоя или во сне. Они имеют тенденцию к мономорфности: состоят из повторяющихся волн со схожей формой и длиной волны. Всякое внешнее раздражение, примененное впервые, вызывает десинхронизацию и обнаружение частых колебаний.

В работе осуществляется анализ непостоянных паттернов, которые возникают в качестве реакции мозга на какие-либо внешние раздражители. В отличие от нормальных паттернов зачастую они состоят из одной волны и поэтому почти не подходят для визуального анализа. Однако подобные паттерны активно используются для взаимодействия человека с компьютером через BCI (*brain computer interface*).

Одна волна – это одиночное колебание потенциала любой амплитуды и формы. Амплитуда волны  $\lambda$  – величина колебания потенциала от пика до пика, измеряется в микровольтах (мкВ) и милливольтах (мВ). Период  $T$  – длительность интервала между началом и концом одиночной волны или комплекса волн. Эти величины связаны соотношением

$$T = \frac{\lambda}{c},$$

где  $c$  – это скорость распространения волн.

**1. Формальная постановка задачи.** У паттерна волны Р300 есть два определения: положительный импульс ( $\lambda > 0$ ), возникающий в интервале от 250 до 500 мс, или же третий положительный пик после возникновения стимула (отсюда альтернативное название Р3).

Оба эти определения говорят о том, что данный паттерн характеризуется количеством и расположением пиков с положительной амплитудой на ЭЭГ, поэтому для дальнейшего анализа ключевое значение имеют операции локализации и измерения максимума ( $\max$  и  $\arg\max$ ).

В данной работе эксперименты проводятся на наборе данных, где целевая переменная характеризует стимул, который должен вызывать волну Р300. Подробнее набор данных описан в разд. 1. Решается задача бинарной классификации, где 1 означает наличие воздействия, а 0 – его отсутствие. Эксперимент по постановке не гарантирует, что во время воздействия возникнет волна Р300, однако шанс этого события крайне велик.

В таком случае модель машинного обучения не получает информацию о наличие волны Р300 напрямую, однако может выделить ее в качестве метапризнака. Основная задача данной работы заключается в проверке этой гипотезы. Объектом выборки является многомерный временной ряд фиксированной длины. На каналах задано отношение соседства по цепи, поэтому

каждое наблюдение  $X_i \in \mathbb{R}^{N \times T}$  можно задать матрицей размера  $N \times T$ , где  $N$  – количество каналов (электродов), а  $T$  – количество промежутков времени, в которые были осуществлены снятия показаний с прибора.

**2. Анализ литературы.** Отметим, что в нейрофизиологии базовым методом определения паттерна выступает усреднение временных рядов по каждому классу. Даже для ярко выраженных паттернов такой подход требует привлечения десятков респондентов. В наших данных присутствует только два респондента, соответственно, базовый метод не использовался ни в работе [1], ни в нашем исследовании.

Рассмотрим основные подходы интерпретации моделей машинного обучения в смежной области – медицине, где подобное свойство играет ключевое значение. Как и в разработанном нами подходе, возможность интерпретации результатов в этих сетях была заложена на этапе проектирования.

В работе [2] авторы использовали декодер сети U-Net, чтобы кодировать внутренние органы человека в разных масштабах. Последующее применение механизма внимания позволяет модели фокусироваться на ключевых участках исходного изображения и в явном виде визуализировать их вклад в итоговый прогноз. Эта техника напоминает анализ функциональных паттернов в многомерных временных рядах. Аналогичный подход, но для задачи сегментации изображений, был применен в работе [3].

Алгоритм глубокого обучения из работы [4] учитывает текущие и предыдущие снимки компьютерной томографии пациента для прогнозирования риска развития рака легких. Авторы использовали метод интегрированного градиента [5] для получения карт выраженности и пригласили опытных радиологов для проверки точности этих карт. Эксперты полностью согласились с тем, что модель действительно фокусировалась на опухолях.

В работе [6] предложили глубокую сверточную нейронную сеть для прогнозирования смертности, основанную на измерении 22 различных показателей в отделении интенсивной терапии, таких как индекс натрия, диурез и т.д. В данной работе интерпретация производилась посредством визуализации сверточных фильтров нейронной сети на разных масштабах.

Авторы [7] представили сеть, которая объединяет извлеченные из ResNet признаки и предварительные клинические знания для прогнозирования риска смертности пациентов на основе компьютерной томографии. Чтобы дополнительно подтвердить эффективность предложенной модели, они использовали метод t-SNE [8] для уменьшения размерности карт признаков злокачественных и доброкачественных новообразований и обнаружили, что злокачественные и доброкачественные наблюдения были хорошо разделены.

Как мы видим из приведенных примеров, в области медицины уже предпринимались многочисленные попытки интерпретации полученных моделью машинного обучения прогнозов, однако в основном они касались моделей для анализа компьютерных снимков и других задач компьютерного зрения. Наш же подход предполагает интерпретацию многомерных временных рядов и поиск новых типов закономерностей в анализируемых сигналах.

**3. Предложенная архитектура.** Основной подход в изучении многомерных временных рядов заключается в использовании архитектуры сверточной нейронной сети, которая применяет отдельные фильтры для анализа временных и пространственных зависимостей. Пространственные сверточные слои используются для извлечения признаков из данных ЭЭГ, отражающих активность различных электродов, в то время как временной сверточный слой извлекает признаки временных рядов. Затем полносвязный слой агрегирует всю полученную информацию для решения поставленной задачи. Этот подход был применен на рассматриваемом наборе данных в работе [9].

Подобное семейство архитектур нейронных сетей весьма богато и обладает хорошей интерпретируемостью. Модели такого типа состоят из двух блоков обработки, которые представлены пространственными и временными свертками соответственно. Анализируя каждый из них, можно выявлять пространственные и временные характеристики функциональных паттернов. Эти блоки могут быть применены в любом порядке, однако основной подход заключается в том, чтобы сначала обработать пространственную информацию, а только потом перейти к анализу временных зависимостей. Наша архитектура также будет придерживаться данного принципа.

Рассматриваемый набор данных состоит из относительно небольшого количества наблюдений с ярко выраженным дисбалансом классов, а специфика многомерных временных рядов не позволяет нам использовать большие предобученные модели. Поэтому цель этой работы заключается в разработке модели с относительно небольшим количеством параметров, кото-

рая потенциально была бы применима к многомерным временным рядам произвольной природы. По этой причине пространственный слой представляет собой свертку  $1 \times 1$  по аналогии с [9], где выход слоя  $z^s(j)$  формируется следующим образом:

$$z^s(j) = b^s + \sum_{i=0}^{N-1} w_i^s X_m(i, j).$$

Здесь  $i = \overline{0, N-1}$ ,  $j = \overline{0, T-1}$ ,  $N$  – количество электродов, а  $T$  – количество временных отсчетов,  $b^s$  – сдвиг пространственного слоя,  $w_i^s$  – вес  $i$ -го электрода в пространственном слое, а  $X_m(i, j)$  – элемент в  $i$ -й строке и  $j$ -м столбце исходной матрицы наблюдения  $X_m$ .

В качестве слоя для обработки времени нами был выбран механизм внимания из [10] (*self attention*). Голова внимания (*attention head*) была взята лишь одна, чтобы добиться наименьшего количества параметров в итоговой модели.

Выход временного слоя  $z^t(X) \in \mathbb{R}^{T \times 1}$  считается по следующей формуле:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D}}\right)V,$$

$$z^t(X) = \text{Attention}(W^Q X, W^K X, W^V X) W^{OT},$$

где  $\text{Attention}(Q, K, V) \in \mathbb{R}^{T \times D}$  – оператор внимания, зависящий от матриц  $Q, K, V \in \mathbb{R}^{T \times D}$ ,  $X \in \mathbb{R}^{T \times 1}$  – вход слоя, а  $W^Q, W^K, W^V, W^O \in \mathbb{R}^{1 \times D}$  – его матрицы весов,  $T$  – количество временных отсчетов, а  $D$  – внутренняя размерность механизма внимания, которая подбирается экспериментально. Если  $Y \in \mathbb{R}^{A \times B}$ , то обозначим

$$\text{softmax}(Y)_{ij} = e^{Y_{ij}} / \sum_{k=1}^B e^{Y_{ik}}, \quad i = \overline{1, A}, \quad j = \overline{1, B}.$$

Механизм внимания способен учитывать контекст каждого наблюдения, т.е. он также способен воспроизвести операции *max* и *argmax*. Согласно разд. 1, этого достаточно, чтобы определять наличие волны P300.

Предложенная нами модель машинного обучения способна определять амплитуду волны  $\lambda$  и оценивать ее форму посредством поиска минимума и максимума, что в контексте анализа волны P300 имеет ключевое значение. Нейронная сеть поочередно применяет фильтры к каналам и временными промежуткам на исходных данных, выявляя тем самым важность тех или иных электродов на голове человека или моментов времени во время записи.

Итоговая модель (рис. 1) состоит из свертки  $1 \times 1$  для обработки пространственной информации и слоя внимания [10] для поиска функциональных паттернов во времени. На выходе расположен один линейный слой для решения задачи классификации на основе полученных представлений. Кроме того, каждый блок, кроме линейного слоя на выходе, сопровождается нелинейной функцией ReLU (*rectified linear unit*) [11].

Приведенная модель использует идеи визуализации фильтров в сверточных слоях и анализ карт внимания для интерпретации полученных результатов и локализации функционального паттерна. При решении исходной задачи модели предлагается сравнить состояние покоя пациента с реакцией на внешнее воздействие.

**4. Рассматриваемый набор данных.** Данные были собраны с помощью парадигмы странностей (англ. *oddball paradigm*). *Oddball paradigm* – это дизайн эксперимента, используемый в психологических и нейрофизиологических исследованиях. Испытуемому демонстрируют последовательности повторяющихся стимулов, которые изредка прерываются нестандартным воздействием. Реакция испытуемого на этот “странный” стимул впоследствии записывается и анализируется, подробнее этот процесс описан в работе [12].

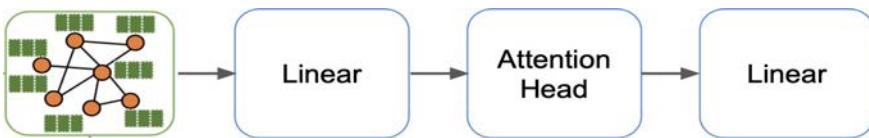


Рис. 1. Предложенная архитектура.

Пользовательские интерфейсы для сбора данных были представлены в работе [13]. Всего в этой работе описывались семь подобных интерфейсов, однако мы решили сфокусироваться на одном функциональном паттерне и рассматривать лишь те парадигмы, что успешно применялись для изучения волны Р300. Такие подходы генерируют большое количество наблюдений, что позволяет успешно обучать искусственные нейронные сети.

4.1. П р о ц е с с с б о р а д а н н ы х . Представленный набор данных был записан с использованием символьной матрицы, с помощью которой человек может писать слова, концентрируя свое внимание на отдельных символах. Этот набор данных является частью соревнования [1] и содержит множество экспериментальных результатов, полученных с применением большого разнообразия моделей. Исходно данные были собраны с целью понять, насколько хорошо алгоритмы могут предсказать целевой символ из данных о мозговой активности. Соответственно, в работе [1] также описаны существующие подходы из классического машинного обучения, которые уже достигают достаточно высокого уровня качества на представленном наборе данных. Но в то время исследователи еще не рассматривали, что именно находят их модели.

Продолжительность одного фрагмента не фиксирована и зависит от постановки задачи. Данные могут быть разделены на фрагменты разной длины. Однако волна Р300 может быть обнаружена в масштабе 1 с, поэтому нет смысла использовать более длительные временные интервалы. Мы работали с фрагментами по 600 мс.

Набор данных содержит ЭЭГ двух здоровых добровольцев (А и В). Графический пользовательский интерфейс для проведения эксперимента показан на рис. 2. Он представляет собой матрицу  $6 \times 6$  символов. В нейтральном состоянии все символы имеют одинаковую низкую яркость. Стока или столбец могут быть временно подсвечены.



Рис. 2. Интерфейс для сбора данных из [13].

Для очередного целевого символа проводят 15 блоков подсвечиваний матрицы, которые в совокупности называются эпохой. В каждом блоке каждая из строк и столбцов подсвечивается по одному разу, т.е. всего 12 подсвечиваний. В блоке столбцы и строки загорались в случайном порядке, т.е. в разных блоках одной эпохи порядок подсвечиваний разный. Испытуемому предлагалось сосредоточиться на целевом символе и молча считать количество мигающих повторений строки и столбца, содержащих целевой символ. Символ определяется парой из строки и столбца. Таким образом, в каждом блоке целевой символ подсвечивается дважды, всего в эпохе – 30 раз. Для каждого подсвечивания легко определить, участвовал ли в нем целевой символ.

Все сигналы ЭЭГ были собраны с помощью 64-электродного скальпа, отфильтрованы в диапазоне от 0.1 до 60 Гц и оцифрованы с частотой 240 Гц. Более подробно прочитать об этом можно в работе [1].

4.2. П р е о б р а з о в а н и е ц е л е в о г о п р и з н а к а . Согласно исследованиям нейрофизиологов, предполагается, но формально не гарантируется, что подсвечивание целевого символа дает волну Р300. Хотя мы знаем, когда стоит ожидать сигнал Р300, его появление зависит от испытуемого, причем сам испытуемый тоже не контролирует этот процесс. Производство волны Р300 не является феноменом сознания, оно возникает из-за внешних стимулов (мерцание строк и столбцов). Кроме того, волна Р300 интерферирует со многими другими волнами, соответственно, ее выраженность может сильно варьироваться.

Для рассматриваемого набора данных часто задача ставится следующим образом. Проверяется, содержит ли целевой символ в подсвечиваемых строке или столбце. В таком случае задача может рассматриваться как бинарная классификация. На вход подается ЭЭГ, соответствующая подсвечиванию одной строки или столбца, а выходом служит бинарная метка класса (является ли текущее подсвечивание целевым). Как уже говорилось выше, на каждое подсвечивание отводится запись длиною в 600 мс, что больше продолжительности самого подсвечивания, однако последний объект эпохи не обрежется, потому что каждая запись заканчивается паузой в 2.5 с, во время которой матрица гаснет.

Такая постановка задачи позволяет нам сфокусироваться на выявлении функциональных паттернов и интерпретации полученных результатов. Итоговые размеры обучающей и тестовой выборок для каждого из пациентов приведены в табл. 1.

Нейрофизиологи выявили, что во время подсвечивания целевого символа в человеческом мозге возникает волна Р300. Предполагается, что символ соответствует пересечению накопле-

**Таблица 1.** Размеры выборок для каждого из пациентов

Целевой класс	Обучающая выборка	Тестовая выборка
Есть подсвечивание	2550	3000
Нет подсвечивания	12 750	15 000

ния нескольких волн Р300. Наибольшее накопление волн Р300 для вертикальных мигающих огней определяет столбец желаемого символа. То же справедливо и для горизонтальных мигающих огней и строк.

Хотя из обучающей разметки мы явно знаем, когда загорался целевой символ, эксперимент по постановке не гарантирует, что во время воздействия возникнет именно волна Р300, хотя шанс этого события и крайне велик. Соответственно, модель не получает информацию о наличии волны Р300 напрямую. Мы хотим убедиться, что модель машинного обучения действительно сможет самостоятельно найти функциональный паттерн, который нейрофизиологам уже известен.

**5. Эксперименты.** Для проведения экспериментов будем использовать предоставленное авторами из работы [1] разделение выборки на обучение и тест. Таким образом, можно сравнить наши алгоритмы с другими подходами, в том числе с теми, которые были разработаны во время оригинального соревнования на используемых данных.

Как отмечалось ранее, обучающая выборка состоит из 85 символов, а тестовая – из 100. При переходе от классификации символов к классификации строк и столбцов получаем, соответственно, обучающую выборку из 15 300 наблюдений и тестовую выборку из 18 000 наблюдений. При обучении важен порядок поступления данных, лучших результатов мы добились при соответствии постановке эксперимента, который проводился при записи исходных данных. Мы установили размер мини-батча  $K = 180$ , именно столько наблюдений описывает один целевой символ из исходной задачи распознавания целевых символов.

Согласно *oddball paradigm*, на каждые 12 объектов приходится 2 наблюдения положительного класса и 10 отрицательного. Следовательно, в задаче существует значительный дисбаланс классов (1 к 6). Кроме того, каждый батч содержит ровно 30 объектов положительного класса и 150 объектов отрицательного, что гарантируется из постановки эксперимента. Для подобной классической проблемы в качестве функции ошибки  $L$  была выбрана бинарная кросс-энтропия:

$$L = - \sum_{n=1}^K \left[ y_n \log x_n + (1 - y_n) \log(1 - x_n) \right],$$

где  $y_n \in \{0,1\}$  – метки классов, а  $x_n \in [0,1]$  – вероятность положительной метки, полученная из модели.

5.1. П р о с т р а н с т в е н н ы й а н а л и з. Теперь, когда процедура обучения определена, можно перейти непосредственно к анализу полученных результатов. Для начала визуализируем веса свертки  $1 \times 1$ , которая является первым этапом обработки входных данных. Благодаря своей структуре она выучивает некоторый вес в соответствие каждому электроду на голове человека. Таким образом, на выходе получаем карту важности электродов с точки зрения модели (рис. 3), другими словами, локализацию функционального паттерна в пространстве.

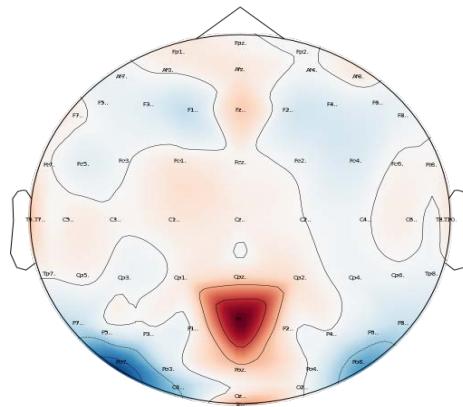


Рис. 3. Карта важности каналов в линейном слое.

Заметим, что модель самостоятельно выявила оптимальную линейную комбинацию исходных электродов. Рс7 и Рс8 были взяты с отрицательным знаком, чтобы учесть негативные значения потенциала, предшествующие Р300, а электрод Рz, наоборот, – с положительным, чтобы уловить пик самой волны. Эту гипотезу подтверждают значения потенциала на выделенных электродах (рис. 4), для построения графика было взято произвольное наблюдение из тестовой выборки, на котором присутствует волна Р300. Для Рz характерен провал в значении потенциала непосредственно перед пиком Р300, а Рс7 и Рс8 вместе образуют самый высокий пик на данных ЭЭГ.

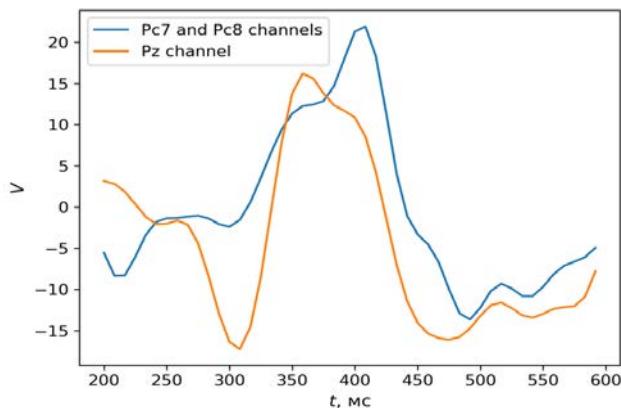


Рис. 4. Потенциал на выделенных электродах.

Также важно отметить, что модель предложила использовать всего три канала ЭЭГ, выделив их как точки концентрации информации о наблюдаемом явлении. Этот результат не противоречит нашим знаниям о человеческом мозге, ведь именно затылочная его доля отвечает за распознавание визуальных образов, которыми и являются целевые символы на экране компьютера.

5.2. Временной анализ. Также немаловажную роль в интерпретации результатов сыграл и слой с механизмом внимания (рис. 5). На графике с картой внимания для одного конкретного наблюдения волны Р300 можно заметить, что модель концентрирует свое внимание на максимальном и минимальном значениях промежуточного представления ЭЭГ. Это говорит о том, что нейронная сеть действительно самостоятельно приходит к преобразованиям вида *max* и *argmax*, что служит необходимым признаком для выявления функционального паттерна Р300, согласно разд. 1.

Взаимное расположение положительных и отрицательных пиков играет ключевую роль в обоих определениях волны Р300. Таким образом, предложенная модель может не только локализовать возникновение функционального паттерна в пространстве, но и определять наличие паттерна по его характерным особенностям. Другими словами, алгоритм выявил вол-

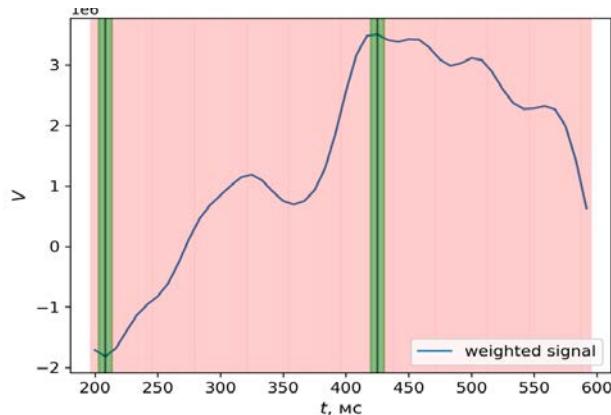


Рис. 5. Карта внимания модели для преобразованной волны P300.

ну P300 из данных, хотя изначальная постановка задачи говорила о его существовании лишь косвенно, через этап сбора данных. При этом данные наших экспериментов (табл. 2) подтверждают, что качество решения исходной задачи бинарной классификации сразу же оказалось на уровне лучших алгоритмов из работы [1] и нейросетевого подхода из [9].

Таблица 2. Сравнение точности бинарной классификации с классическими алгоритмами

Модель	Пациент А, %	Пациент В, %
Логистическая регрессия	65.9	68.4
Градиентный бустинг	67.6	69.0
Сверточная нейросеть	72.4	75.7
Предложенная модель	72.7	75.2

**Заключение.** Предложена методика, помогающая исследователям по серии экспериментов автоматически выявить функциональный паттерн в многомерных временных рядах. При этом было достаточно формализовать исходную задачу в терминах машинного обучения и не требовалось углубляться в предметную область.

Работоспособность методики продемонстрирована на данных ЭЭГ. Модель выявляет паттерн волны P300 с высокой точностью, а также способна явно выделить его отличительные признаки. Вся эта информация была неявно извлечена из постановки задачи.

Данный подход может быть доработан и расширен для других предметных областей, что и представляет интерес для дальнейших исследований.

#### СПИСОК ЛИТЕРАТУРЫ

1. Blankertz B., Muller K.R., Krusienski D.J., Schalk G., Wolpaw J.R., Schlogl A., et al. The BCI Competition iii: Validating Alternative Approaches to Actual BCI Problems // IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2006. V. 14 (2). P. 153–159.
2. Oktay O., Schlemper J., Folgoc L.L., Lee M., Heinrich M., Misawa K. et. al. Attention U-Net: Learning Where to Look for the Pancreas // arXiv preprint arXiv:1804.03999, 2018.
3. Lee H., Yune S., Mansouri M., Kim M., Tajmir S.H., Guerrier C.E. et. al. An Explainable Deep-learning Algorithm for the Detection of Acute Intracranial Haemorrhage from Small Datasets // Nature Biomedical Engineering. 2019. V. 3 (3). P. 173–182.
4. Ardila D., Kiraly A.P., Bharadwaj S., Choi B., Reicher J.J., Peng L. et. al. End-to-end Lung Cancer Screening with Three-dimensional Deep Learning on Low-dose Chest Computed Tomography // Nature Medicine. 2019. V. 25 (6). P. 954–961.
5. Sundararajan M., Taly A., Yan Q. Axiomatic Attribution for Deep Networks // Intern. Conf. on Machine Learning. PMLR. Sydney, Australia, 2017. P. 3319–3328.
6. Caicedo-Torres W., Gutierrez J. ISeeU: Visually Interpretable Deep Learning for Mortality Prediction Inside the ICU // J. Biomedical Informatics. 2019. V. 98. P. 103269.

7. *Guo H., Kruger U., Wang G., Kalra M.K., Yan P.* Knowledge-based Analysis for Mortality Prediction from CT Images // IEEE Biomedical and Health Informatics. 2019. V. 24 (2). P. 457–464.
8. *Van der Maaten L., Hinton G.* Visualizing Data Using t-SNE // J. Machine Learning Research. 2008. V. 9(11).
9. *Cecotti H., Graser A.* Convolutional Neural Networks for P300 Detection with Application to Brain-computer Interfaces // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010. V. 33 (3). P. 433–445.
10. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L. et al.* Attention is All you Need // Advances in Neural Information Processing Systems. 2017. V. 30.
11. *Krizhevsky A., Sutskever I., Hinton G. E.* Imagenet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems. 2012. V. 25.
12. *Squires N.K., Squires K.C., Hillyard S.A.* Two Varieties of Long-latency Positive Waves Evoked by Unpredictable Auditory Stimuli in Man // Electroencephalography and Clinical Neurophysiology. 1975. V. 38 (4). P. 387–401.
13. *Schalk G., McFarland D.J., Hinterberger T., Birbaumer N., Wolpaw J.R.* BCI2000: A General-purpose Brain-computer Interface (BCI) System // IEEE Transactions on Biomedical Engineering. 2004. V. 51 (6). P. 1034–1043.

## МОРФОЛОГИЯ МОМЕНТА И СТРУКТУРА ДИНАМИЧЕСКИХ ПРОЦЕССОВ В КОММУНИКАТИВНЫХ СИТУАЦИЯХ

© 2024 г. В. П. Кальян<sup>a</sup>, \*

<sup>a</sup>ФИЦ ИУ РАН, Москва, Россия

\*e-mail: [vkalyan@mail.ru](mailto:vkalyan@mail.ru)

Поступила в редакцию 07.05.2024 г.

После доработки 28.06.2024 г.

Принята к публикации 16.10.2024 г.

Описан подход к решению проблемы прогнозирования развития коммуникативных ситуаций на основе информации, поступающей от экспертов. Инструментом структуризации процессов служит разработанная автором базовая модель — морфология широкого класса коммуникативных ситуаций. Для построения и обучения имитационной модели изучаемой ситуации используется язык визуального программирования. На первом этапе создания и обучения имитатора экспертом вводятся в систему значения параметров нескольких начальных фаз. Расчет дальнейшего развития событий, оценки и классификации состояний системы осуществляется с помощью математических методов распознавания, комбинаторных методов анализа, исследования операций, оптимального управления. Подтверждение расчетов достигается сравнением полученных результатов от имитатора с реальными изменениями параметров изучаемых процессов во времени.

**Ключевые слова:** морфология ситуации; событие; фазы состояния системы; динамика параметров; обучение с учителем; интенции; роль; статус; качество.

**DOI:** 10.31857/S0002338824060116, **EDN:** SUBBFM

## PMORPHOLOGY OF THE MOMENT AND STRUCTURE OF DYNAMIC PROCESSES IN COMMUNICATIVE SITUATIONS

V. P. Kalyan<sup>a</sup>, \*

<sup>a</sup>Federal Research Center “Computer Science and Control,”  
Russian Academy of Sciences, Moscow, 119333 Russia

\*e-mail: [vkalyan@mail.ru](mailto:vkalyan@mail.ru)

The paper describes an approach to solving the problem of predicting the development of communicative situations based on information received from experts. The tool for structuring processes is the basic model developed by the author — the morphology of a wide class of communicative situations. A visual programming language is used to build and train a simulation model of the situation being studied. At the first stage of creating and training the simulator, the expert enters the parameter values of several initial phases into the system. Calculation of further developments of events, assessment and classification of system states is carried out using mathematical recognition methods, combinatorial methods of analysis, operations research, and optimal control. Confirmation of calculations is achieved by comparing the results obtained from the simulator with real changes in the parameters of the processes being studied over time.

**Keywords:** morphology of the situation; event; phases of the system state; dynamics of parameters; supervised learning; intentions; role; status; quality.

**0. Введение.** Условия моделирования коммуникативных ситуаций (КС) заметно отличаются от условий моделирования ситуаций технических (ТС), поскольку КС включает в себя Человека — «НОМО» — отдельных людей, интенциональные группы (ИГ) — группы людей с когерентными взглядами или аудиторию из людей с самыми разными интенциями.

Все базовые элементы ситуации — именованные сущности (термин, заимствованный из семантического анализа текстов, вероятностного тематического моделирования и т.п.), обла-

дают разными потенциалами влияния на ее развитие, и мы, углубляясь в структуру ситуации, выявляем ее морфемы и представляем в модели тектоническими частями (именами) и флексиями – ролями, статусами и качествами по отношению к целям развития ситуации. Положение усложняется тем, что участники ситуации могут иметь как разные, так и согласующиеся цели, могут менять их и под воздействием обстоятельств – изменяться сами.

Наблюдение за любыми динамическими ситуациями показывает нам, что динамика ситуации включает статические участки, плавные динамические участки и участки, где плавное развитие процессов чередуется со скачками, разрывами. Они появляются при преодолении дисбалансов, противоречий, которые разрешаются через изменение параметров базовых элементов ситуации и выливаются в некоторое *событие*, которое представляет собой узловые точки развития процессов и выступает по сути триггером, переводящим плавно или скачком ситуации в следующую фазу, которая отличается от прежней фазы изменившейся морфологией (рис. 1).

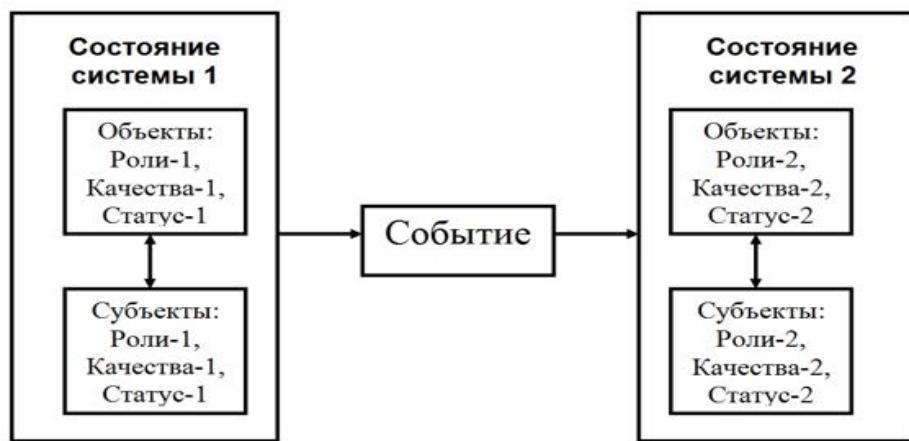


Рис. 1. Трехфазное представление одного звена в развитии ситуации

Схема элементарного движения ситуации изображенная на рис. 1 состоит из трех фаз. Две из них – состояния системы и между ними – *событие*, которое играет роль триггера, переключателя фаз. Однако в таком описании событие не рассматривается как процесс или результат развития некоторого процесса, например противоборства субъектов или их согласованная деятельность. В реальной жизни событие является не только триггером, переключающим систему из одного состояния в другое, но процессом и результатом протекания процесса, которые на схеме рис. 1 никак не отражены.

Для построения модели реального развития ситуации нам нужно отобразить внутреннюю структуру момента (статической фазы ситуации), чтобы из нее можно было вывести схему развивающегося процесса. Подход к описанию ситуаций был заложен еще в [1, 2], технических ситуаций – в [3], для коммуникативных ситуаций в этнологии – в [4], чей подход был развит в [5], где описывалась структура мифа именно как динамика изменения морфологии ситуации.

Автором настоящей работы на обширном материале были экспериментально выявлены основные морфемы коммуникативных ситуаций, определено их значение для представления ситуативной статики, динамики и решения широкого круга задач моделирования ситуативных процессов (рис. 2).

Именованные сущности составляют морфологическую основу ситуации в экспертной разметке, которая предусматривает не только неизменность значений имени сущности *N* (name), но и возможность фазовых изменений значений природы сущности *σ* (*οὐσία*); статические и динамические свойства каждой сущности могут быть вписаны в предикатно-актантную форму отображения как текущей *j*-й стихийно сложившейся фазы развития ситуации, так и каждой возможной фазы развития ситуации из предустановленного экспертом множества состояний системы. Эти изменения свойств сущностей в дальнейшем реализуются в виде изменения *взаимоотношений*, которые обусловлены сменой *ролей, качеств, статусов* участников событий.

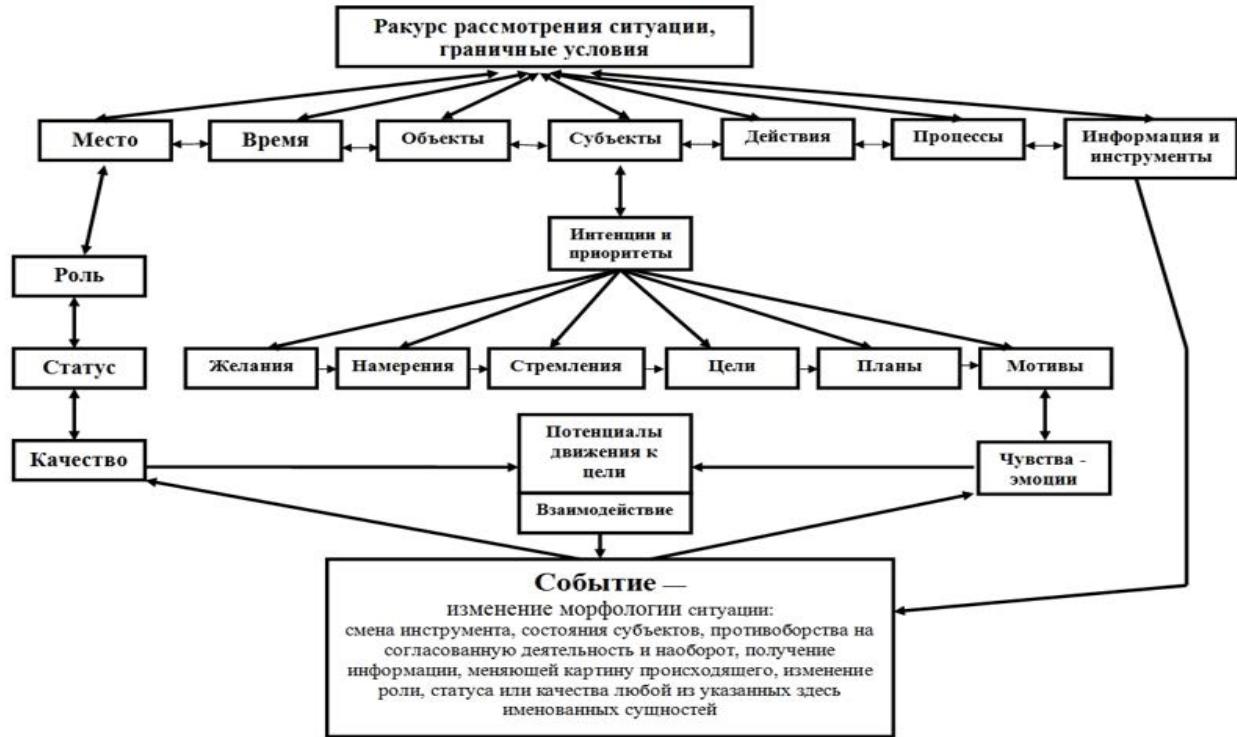


Рис. 2. Структура момента – основа динамики. Базовая морфология коммуникативных ситуаций. Именованные сущности и их параметры

Фрейм состояния системы в каждой фазе ее динамического развития в предикатно-актантной форме запишется как

$$c_j = (n_i, \sigma_{k,j,i}, r_{k,j,i}, s_{k,j,i}, q_{k,j,i}, I_{k,j,i}), \quad (0.1)$$

где  $j$ ,  $j = \overline{1, j}$  – номер фазы развития рассматриваемой ситуации, в котором система принимает одно из множества  $L^f$  возможных конечных ( $C^f$ ) или множества  $L^I$  промежуточных  $C^I$  состояний ( $C^f \subset C$ ,  $C^I \subset C$ ,  $C^f \cup C^I \equiv C$ ),  $L^f + L^I = L$ ,  $n_i$  – имя  $i$ -й сущности,  $k$  – идентифицирующий номер характеристики  $i$ -й сущности,  $\sigma_{k,j,i}$ ,  $k = \overline{1, m}$ , – текущее натуральное состояние – тип сущности (в данной фазе  $j$  и в данном контексте рассмотрения – это объект, субъект, действие, процесс, время, место),  $r_{k,j,i}$  – роль,  $s_{k,j,i}$  – статус/стратегии,  $q_{k,j,i}$  – качество сущности по отношению к рассматриваемой цели,  $I_{k,j,i}$  – интенция  $i$ -го субъекта, определяющая актуальную цель в данной фазе  $j$ .

**1. Постановка задачи.** Для построения полноценной динамической модели ситуации нам нужно прояснить смысл и структуру предстоящего исследования.

1. Определить цель исследования и найти, какие ракурсы развития ситуации нас интересуют.  
 2. Описать каждую  $i$ -ю сущность из множества  $N$  именованных сущностей с помощью вектора  $F$ , характеризующего ее потенциал в статических фазах  $j$  –  $F_{j,i}(x_1, x_2, \dots, x_k, \dots, x_m)$  и там, где это возможно – его поведение в динамических фазах  $F_{j,i}(x_1(t), x_2(t), \dots, x_k(t), \dots, x_m(t))$ . Здесь  $x$  – спроектированное на доминирующую направление развития ситуации для  $j$ -й фазы силовое значение каждого параметра  $x_k$   $i$ -й именованной сущности, где  $k = \overline{1, m}$ ,  $m$  – количество параметров, через которые сущности проявляют себя и происходит воздействие на ситуативные процессы от каждой именованной сущности.

3. Задать (априорно) множество всех наиболее очевидных (вероятных) конечных и промежуточных состояний системы  $C$  исходя из опыта наблюдения за развитием данной и подобных ситуаций. Таким образом, здесь и в дальнейшем  $C$  – множество конечных и промежуточных состояний системы, элемент этого множества имеет наименование  $c_h$ , где  $h = \overline{1, L}$ .

4. Установить, преодоление каких пороговых значений параметров меняют значимые для ситуации свойства именованных сущностей и, следовательно, меняют и состояние всей системы. Фрагменты динамических процессов, приводящих к таким изменениям, определяем как *событие*. Множество событий обозначаем через  $E$ . Нас здесь интересуют изменения значений каждого  $k$ -го параметра  $i$ -й именованной сущности из  $i = 1, N$ , где  $N$  – количество именованных сущностей, составляющих основу морфологии ситуации, и  $m$  – количество параметров  $x_k$  (функциональных характеристик) этих именованных сущностей.

На основании этой экспертной разметки можно априорно оценить в  $j$ -й фазе сумму параметров каждой сущности, определяющих потенциал ее возможного воздействия на развитие ситуации:

$$F_{j,i} = \sum_{k=1}^m x_{k,i,j}, \quad (1.1)$$

и на результат взаимодействия сущностей, приняв во внимание данную выше в (1.1) предварительную оценку силовому потенциалу векторов:

$$\pi_j = \prod_{i=1}^N F_{i,j}. \quad (1.2)$$

Следовательно, можно рассчитать первую вероятную  $j$ -последовательность событий  $E$  и состояний системы  $C$ , выразив ее через секвенцию:

$$\mathfrak{R} = \{c_1 \rightarrow E_1(\pi_1) \rightarrow c_2, \dots, c_j \rightarrow E_j(\pi_j) \rightarrow c_{j+1}, \dots, E_{N-1}(\pi_{N-1}) \rightarrow c_N\}. \quad (1.3)$$

Определим, что предписанный каждой сущности в данной ситуации нормативный набор действий/(реакций на воздействие)  $S_i$  является ее *статусом*. Выход за пределы норматива (не совершение обязательных и совершение не предписанных нормативом действий) во время  $j$ -го события отображается параметром *качество*  $Q_{j,i}(x_k)$ .

Роль, которую сущность  $n_i$  фактически играла в некотором событии  $E_j$ , отнесенная к ее статусу, показывает нам ее качество:

$$Q_{j,i} = \frac{R_i(E_j)}{S_i(E_j)}. \quad (1.4)$$

Как мы теперь понимаем, характеристики сущностей  $x_{i,k}$  в динамике реальной ситуации от фазы к фазе (и внутри фазы как функция физического времени  $t$ ) могут менять значения, включая тип (природу) своей сущности. Динамика системы в зависимости от совпадения/противоборства интенций разворачивается либо по законам совместной деятельности, либо противоборства.

Оценку потенциала противоборствующих субъектов и вероятности достижения желаемого  $i$ -м субъектом конечного состояния  $c_i^f$  при соотнесении всех используемых субъектами  $k$ -х свойств для осуществления своих стратегий  $s_{i,k}$  будем вводить в систему с помощью модели обучения с учителем. Процесс построения вероятной последовательности событий  $E_j$ , где учитываются силовые потенциалы на основании характеристик противоборствующих сторон, выражющиеся в интенсивности воздействий на текущее состояние системы  $\lambda_{j,x}, \mu_{x,j}$  на конкретном примере, рассмотрим ниже.

Разработка этого подхода открывает широкие возможности экстренного построения, верификации и использования модельного инструментария в решении сложных задач прогнозирования, исследования причин возникновения проблем, особенностей протекания процессов при планировании, управлении, поиске решений в таких областях, как следственная практика, здравоохранение, формирование общественного мнения, пути разрешения конфликтов, взаимосвязь процессов в экономике и геополитике, планирование военных операций и т.д.

**2. Между согласием и конфликтом: динамика ситуации в пространстве взаимоотношений.** Характер смены фаз динамической ситуации в определенной мере зависит от направления интенций субъектов, их целей в данной локальных фазе, соотношения их потенциалов, влияющих на силу, инструментальную оснащенность, мотивированность, информированность, т.е. значений всех характеристик, дающих вклад в значение потенциала субъектов по отношению к глобальной и локальной целям  $\{c_1, c_2, \dots, c_i, \dots, c_F, c_F \in C^F, C^F \subset C^L\}$ , здесь  $\overline{1, F}$  – номера списка глобальных целей субъектов. Множество желательных для них конечных состояний системы  $C^F$  договариваемся определять из состава их интенций. Это задает нам алфавит из  $F$  вероятных

конечных состояний  $c_f \in C^F$  системы и определяет для нас набор из  $I$  (intermediate) промежуточных – эти локальные цели вычисляем исходя из вероятных морфологических изменений параметров в результате противоборства  $N$  участников ситуации при движении каждого  $n_i \in N$  к желаемому результату из  $c_I \subset C^I$  состояний  $\{c_1, c_2, \dots, c_h, \dots, c_I, c_I \in C^I, C^I \subset C^L\}$ ,  $C^F$  и фиксации промежуточных (intermediate) состояний  $C^I$  из всего множества состояний  $C^L$ .

Отметим, что потенциал развития ситуации заложен как в негативном характере отношений (сомнение, несогласие, противоречие, возражение, протест, вражда, провокация, война на уничтожение противника) и связанных с ними интенциями, так и в позитивном характере, предполагающем когерентность взглядов, совпадение интенций, целей. В последнем случае субъекты, как правило, предпринимают совместные усилия по преодолению внешних препятствий, согласованию действий для работы по общему плану, объединяясь в ИГ.

Обозначим через  $\pi_{j,i,k}$  потенциал  $k$ -го фактора каждой  $i$ -й сущности (времени, места, объектов, субъектов, интенций, действий/процессов, инструмента, информации) для той или иной роли, которую они могут играть по отношению к текущему  $j$ -му событию и доминирующему направлению движения из состава целей, которые преследуют его участники. Тогда суммарный потенциал силовых характеристик каждой сущности при стремлении к одному из возможных следующих промежуточных состояний определим, вычисляя сумму:

$$\gamma_{j,i}(t) = \sum_{k=1}^m \pi_{j,i,k}(x_{j,i,k}(t)). \quad (2.1)$$

Время развития каждого события  $E_j$  договоримся обозначать как  $T_j$ . А оценку результата динамического взаимодействия сущностей зададим с помощью интеграла функции динамики векторов:

$$\gamma_{j,i} = \int_0^{T_j} y_{j,i}(t) dt. \quad (2.2)$$

Таким образом,  $\gamma$  мы определяем как функцию результатов события  $E_j$ , а возможное следующее за  $j$ -м  $j+1$ -е состояние системы из всего набора состояний можно оценить как ближайшее к  $\gamma$ :

$$C_{j+1} = \arg \min_{h=1, L} \left\{ C_h - \gamma_{j,i}(E_j) \right\}. \quad (2.3)$$

**3. От причины к следствию. Анатомия события.** Обозначим текущее событие через  $E_j$ . Это  $j$ -е событие соответствует переходу состояния системы  $C_h$  из  $j$ -й фазы в  $j+1$ -ю фазу (где состояния системы будет уже некоторым  $c_x$ ), параметры которого вычисляются по итогам процесса внутри события  $E_j$  и принимают наиболее близкое значение из набора целей, выявленных по интенциям субъектов  $c_x \subset C$ . (Для простоты описания договоримся здесь и далее текущее состояние системы  $c_h(j)$  обозначать через  $c_j$ , помня, что  $c_j = c_h(j)$ .) Таким образом,  $j$ -е событие переводит систему из  $j$ -й фазы в следующую  $j+1$ -ю фазу, значение которого  $c_h \in C$  нужно установить из реального хода событий или вычислить по параметрам, на основании текстового или речевого описания момента.

Если перед началом события  $E_j$  в состоянии системы  $c_j$  априорно или по результатам предыдущих событий, как показано для примера на схеме рис. 2, статус  $S_{i,j}$  является заданными параметром, то при начале динамических изменений роль каждой сущности  $R_{i,j}$  и ее качество  $Q_{i,j}$  понимаем как отношение роли к статусу:

$$Q_{i,j} = \frac{R_{i,j}(C_j)}{S_{i,j}(C_j)}, \quad (3.1)$$

а по свершению события  $E_j$  наблюдаемыми апостериорно параметрами данной сущности в  $j+1$ -й фазе становятся роль и качество. Тогда можно ожидать, что если наблюдаемые статусы в событии и вычисленные по прошествии события

$$S_{i,j+1}(C_{j+1}) = \frac{R_{i,j+1}(C_{j+1})}{Q_{i,j+1}(C_{j+1})}, \quad (3.2)$$

если статусы до и после события различаются, то при оптимальном управлении ситуацией статус данной сущности стоящими над ситуацией инстанциями (политическим руководством, руководителями производства, медицины, образования, СМИ, военным командова-

нием и т.д.) должен быть приведен в соответствии с реальным положением вещей, и это действие охарактеризует качество уже руководящих инстанций.

Теперь мы должны внутри каждого перехода из состояния  $c_j$  в  $c_{j+1}$  рассмотреть анатомию самого  $j$ -го события  $E_j$ , переводящего  $c_j$  в  $c_{j+1}$ , а именно три последовательные «подфазы» состояния системы:  $c_j$ ,  $E_j$ ,  $c_{j+1}$ . Здесь  $c_j$  – фаза состояния системы в момент перед началом события включает в себя обстоятельства, являющиеся триггером начавшихся изменений,  $E_j(\pi_j)$  – ядро события, а сами воздействия, интенсивность потока которых обозначаем как  $\lambda_{j,x}, \mu_{x,j}, \lambda_{j,x}$  – поток внешних воздействий, меняющих морфологию системы и приводящих ее в данное состояние,  $\mu_{x,j}$  – поток воздействий, выводящих ее из данного состояния и переводящий в следующую после его совершения фазу – «апостериорное» состояние системы:

$$c_{j+1} = (n_i, \sigma_{k,j+1,i}, r_{k,j+1,i}, s_{k,j+1,i}, q_{k,j+1,i}, I_{k,j+1,i}), \quad (3.3)$$

фиксирует произошедшие изменения.

Преимущество силового потенциала того или иного субъекта/интенциональной группы казалось бы показывает, в каком направлении будут происходить изменения состояний системы  $c_j \in C$  и какова будет следующая  $j+1$ -я фаза состояния системы (см. (3.3)). Однако при вероятностном характере срабатывания параметров ситуации в реальности (возможно, неполном использовании потенциала базовых элементов ситуации, учите не всех участников ситуации, появлении дополнительных неучтенных факторов) о процессе развития событий можно говорить лишь как о вероятности их свершения.

Рассмотрим подробно, через какую цепь состояний-фаз  $\{c_j, c_{j+1}, \dots, c_f\}$  будет проходить данная система от текущего момента  $j$ -й фазы, где последующие актуализирующиеся в текущих фазах состояния системы  $c_j, c_{j+1}, \dots, c_f \subset C$  – суть подмножество всех возможных состояний системы  $C$ . Актуальные параметры  $c_j$  в каждом звене из цепи событий  $E_j$  вычисляются на основании значений параметров в предыдущем состоянии системы из множества  $C_L$  и наблюдаемого потока воздействий  $\lambda_{(j-1)j}$  из состояния  $c_{(j-1)}$  в  $c_j$ .

Поток воздействий от  $i$ -й сущности как функция от времени в  $j$ -м событии после  $j$ -й фазы состояния системы определяется развертыванием ее силового ресурса, который, как мы теперь понимаем, должен вычисляться по формуле

$$\alpha_{j,i}(t) = \sum_{k=1}^m \pi_{i,k}(x_{j,i,k}(t)), \quad (3.4)$$

а максимально возможное интегральное воздействие от него в  $j$ -й фазе можно оценить как

$$\delta_{j,i}(j) = \int_0^{T_j} \alpha_{j,i}(t) dt. \quad (3.5)$$

**4. Программирование взаимодействия элементов модели.** Произведенную экспертную оценку параметров и взаимоотношений именованных сущностей для некоторой  $j$ -й фазы динамики ситуации используем для построения управляемой обучающей модели. Патчи и подпатчи – графические интерактивные схемы, которые создаем с помощью инструментов библиотеки фреймворка и операционных окон типа патчера. Другими словами, окна, разработанные в языке программирования комплекса программ MAX/MSP/PureData, позволяют создавать наглядные графические схемы управления динамикой сущностей. Эти интерактивные графические схемы применяются вместо принятых в математике выражений и делают визуальное управление взаимодействием сущностей подобным управлению инструментами в электронном оркестре.

Данная работа начиналась с создания собственного фреймворка – библиотеки графических примитивов – интерактивных паттернов нижнего уровня. Паттерны нижнего уровня – работающие части общей модели, патчи и подпатчи. Они состоят из схем расчета функционирования именованных сущностей рассматриваемой ситуации и схем их взаимодействия в модели динамики системы.

**5. Описание языка и процесса программирования модели ситуации.** Патчи конструируем из подпатчей с помощью виртуальных проводов, ими соединяем патчи и подпатчи между собой, а также элементы внутреннего устройства подпатчей – вычислительные схемы из объектов, которые, как правило, выполняют простейшие функции типа арифметических операций, со-

ртировки, обработки параметров сущностей, операций управления динамикой типа *старт/стоп, метроном*, композиции функций их динамики и т.д.

Важными элементами управления динамикой модельных конструкций всех уровней — патчей и подпатчей являются такие объекты графического языка, как *слайдер, тогл, банг, старт, стоп, метро, функция/процесс — модуль* (рис. 3).



Рис. 3. Базовые элементы визуального языка программирования моделей ситуации

*Слайдер* внешне напоминает ползунковый потенциометр. Он меняет в реальном времени числа, ограниченные указанным диапазоном в объекте, который производит вычисления — константы и коэффициенты функций. Первоначально заданное в объекте число меняет свое значение с помощью смещения ползунка на величину сдвига. *Слайдер* управляет параметрами подсоединенных к нему объектов, например, частотой управляющих временных отсчетов объекта *метро*, дискретизации времени процессов и временной метрикой развертывания событий, коэффициенты в содержащихся в них выражениях и т.д. (рис. 4)



Рис. 4. Объект *слайдер*

*Число* — объект, в который можно ввести любое число извне — из другого объекта, с клавиатуры компьютера, нажатием стрелочки в его боксе. Значение будет меняться на предустановленные извне величины. Содержимое объекта *число* может меняться из объекта *слайдер* перемещением ползунка, также по заданной *слайдеру* извне шкале. Выход значения числа может быть направлен и передаваться через соединение со входом другого объекта, подпатча или патча (рис. 5).



Рис. 5. Объект *число*

*Банг* — это сообщения-триггер, он активизирует процесс в подсоединенном к нему боксе, старт его запускает в сети патча, *стоп* — останавливает (рис. 6).



Рис. 6. Объекты *банг, старт, стоп*

*Тогл* — чередует эти функции при нажатии его кликом мышки последовательно (рис. 7).



Рис. 7. Объект *тогл*

Управление всеми функциями во времени осуществляется через временные отсчеты, которые генерирует элемент управления временной метрикой – *метро*. Он управляет частотой управляющих временных отсчетов параметров подсоединенных к нему объектов – *слайдер*, *тогл*, *банг/старт/стоп*, *число*, временной метрикой развертывания событий (как метром в музыке – темпом, ритмом), а также развертыванием во времени действия объекта *функция* – процесса, описывающего механизм его действия в патче (рис. 8).



Рис. 8. Объект *метро*

Объект *функция* представляет собой предустановленные для текущей ситуации функции от простейших тригонометрических функций  $\sin(t)$ ,  $\cos(t)$  до вычисленных зависимостей по методу Стронгина, Пауэлла, Колмогорова, Бурбаки, функций распределения Больцмана-Гиббса и т.д. и управляется во времени связкой объектов *метро*, *слайдер*, *тогл*, *банг*, *старт*, *стоп*, *число*, временной метрикой развертывания событий (физическими несущими частотой дискретизации времени на общей оси времени – TimeLine), темпо-ритмом движения процессов внутри состояний и событий, порождаемых объектом *функция* (рис. 9).



Рис. 9. *Функция/процесс-модуль*

Экземпляры объекта *операция* выполняют простейшие действия типа арифметических операций сложения, вычитания, умножения, деления, возведения в степень, извлечения корня, сравнения, сортировки, логического следования (импликации – если – то) и других операций обработки параметров сущностей (рис. 10).



Рис. 10. Объект *операция*

Объект *принт* выводит полученные/вычисленные данные на консоль или в указанное место (вход другого модуля – объекта/подпатча, патча) для дальнейших вычислений (рис. 11).

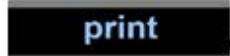


Рис. 11. Объект *принт*

Соединение элементарных объектов в подпатчи (далее – подпатчи в патчи и патчи между собой) конструируем с помощью виртуальных проводов. Элементарные объекты выполняют простейшие функции типа арифметических операций, сортировки, обработки параметров сущностей, операций управления динамикой типа *старт/стоп*, метроном, композицией функций, их динамики и т.д. На рис. 12 представлены конструкция, управляющая процессом с помощью объектов *тогл*, *слайдер*, *метро*, который отвечает за изменение темпа отсчетов для управления объектом *функция/процесс* во времени. Изменение характеристик процесса во времени выводится в консоль и другие смежные патчи системы объектом *принт*.

**6. Последовательность действий при конструировании и обучении имитатора ситуации.** Первоначально построим первые подпатчи из элементов библиотеки элементарных вычислитель-

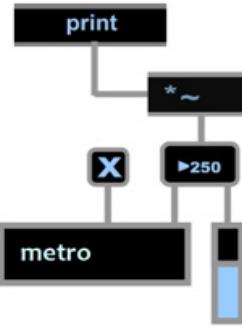


Рис. 12. Простейший патч – схема запуска динамики процесса в объекте функция/процесс связкой объектов тогл, метро, слайдер и вывода данных из объекта функция/процесс с помощью объекта print

ных операций и готовых объектов в виде классических тригонометрических функций. Далее мы строим модели для расчета нужных нам в данной задаче динамических процессов по методам оптимального управления Стронгина, Пауэлла, Колмогорова, Бурбаки, функций распределения Больцмана-Гиббса и т.д.

Патчи конструируем из подпатчей помошью виртуальных проводов. Ими соединяем подпатчи, которые, как правило, выполняют простейшие функции типа арифметических операций, сортировки, обработки параметров сущностей, операций управления динамикой типа старт/стоп, метроном, композиции функций их динамики и т.д. Модель каждой сущности имеет входы и выходы – это каналы загрузки данных в подпатчи и патчи сущностей, в схемы обмена данными внутри их моделей и схемы взаимодействия сущностей через связь модулей и их инкапсуляцию с помощью объекта *bpatcher*.

В окне под названием *patcher* создаем программы с помощью интерфейса, подобного программам для рисования. Выбираем объект из палитры (библиотеки объектов) и затем щелкаем мышью там, где хотим разместить объект. Для того, чтобы добавить модуль в рабочий патч, просто конструируем объект *bpatcher*, а в инспекторе его свойств указываем путь к файлу модуля (патча, подпатча). Весь процесс написания программы сводится к многочисленным инкапсуляциям и добавлениям субпатчей в работающий патч. По сути, просто закапываем рабочие модули программы по группам все глубже и глубже, а на выходе будем иметь элементарную структуру всего из нескольких объектов (рис. 13).



Рис. 13. Структура программы в визуальном языке программирования

**7. Пример построения имитатора ситуации с морфодинамикой из четырех событий и четырех состояний.** Разберем частный пример имитации противоборства двух субъектов: коллективного (интенциональной группы) и индивидуального – некоего некогерентного участника ситуации. При этом договоримся об условии, что финальное состояние, при котором единое всеми согласованное решение не принято, исключается.

Именно такие ситуации неоднократно имели место в действительности или были описаны в литературе. Можно их формально представить с помощью следующих рассуждений: пусть некоторая система имеет  $L$  возможных состояний (число состояний в нашем примере  $L = 4$ ) и пусть алфавит этих состояний  $C_1, C_2, C_3, C_4$ , из которых два ( $C_1$  и  $C_4$ ) являются вариантами конечного состояния, а два ( $C_2$  и  $C_3$ ) – промежуточными результатами событий  $E_1, E_2, E_3, E_4$ , описывающих острые фазы взаимодействия двух противоборствующих сторон. Итак, мы строим конструкцию из четырех состояний и четырех переходов.

Опишем кратко содержание состояний системы и связанных с ними событий на примере двух широко известных ситуаций противоборства двух мнений, которые обсуждаются на собрании уполномоченных представителей из числа субъектов-участников ситуации для вынесения некоторого важного решения: краткое конфиденциальное совещание перед заседанием кабинета министров Австро-Венгрии о немедленном начале войны с Сербией (фактически давшее старт началу Первой мировой войны) и совещания присяжных в фильме по мотивам пьесы «Двенадцать рассерженных мужчин» Р. Роуза и его ремейка – фильма Н. Михалкова «12».

Состояние  $C_1$  – участники ситуации имеют мнения, совпадающие с мнением председательствующего на совещании; не имеющие альтернативной позиции (или такая к началу совещания позиция подавлена) просто соглашаются с господствующим мнением.

Событие  $E_1$  (действие/процесс) – в обоих случаях появляется участник ситуации с особым мнением и высказывает его, оказывая воздействие  $\mu_{12}$ , которое переводит состояние системы из  $C_1$  в  $C_2$ , или воздействие  $\mu_{13}$ , переводящее систему из состояния  $C_1$  в  $C_3$ .

Состояние  $C_2$  – зафиксировано особое мнение одного или нескольких участников и это требует дальнейшего обсуждения, выяснения причин разных позиций.

Событие  $E_2$  (действие/процесс) – начинается обсуждение, где все участники совещания высказывают свое мнение:

а) совершаются согласованные коллективные попытки убедить участника, имеющего особое мнение, в ошибочности такой позиции, оказываются воздействия  $\lambda_{21}$  на вольнодумца по склонению его к единой точки зрения, т.е. заранее намеченному состоянию, предусматривающему единственно допустимое развитие событий,

б) действия участника с особым мнением  $\lambda_{24}$  по глубоко аргументированному изложению своей точки зрения и вовлечению нескольких участников совещания в процесс осмыслиения обсуждаемой ситуации и убеждения всех присяжных в правильности особого мнения.

Состояние  $C_3$  – произошло осмысление лишь частью участников совещания неоднозначности ситуации и возросла готовность присяжных к открытой дискуссии.

Событие  $E_3$  (действие/процесс) – последовательность воздействий участника с особым мнением  $\lambda_{34}$  по информированию каждого участника с неустойчивым мнением все новыми свидетельствами несостоительности обвинений и применение им приемов психологического воздействия на каждого участника промежуточных состояний системы, возникающих из однотипных событий – действий по обмену мнениями участников с утверждением правильности особого мнения и переводом состояния системы к состоянию  $C_4$ , когда особое мнение постепенно становится общим.

Состояние  $C_4$  – подавляющее большинство или все участники ситуации присоединились к особому мнению; имеющий особое мнение участник убедил критически важное для принятия решения число участников совещания в верности своей позиции, и итогом совещания становится присоединение всех участников к особому мнению.

Событие  $E_4$  (действие/процесс) – подавляющее большинство или все участники ситуации вроде бы присоединились к особому мнению, но появление дополнительной информации рушит созданную конструкцию и один или несколько участников совещания (присяжный или руководитель государственного совета) возвращают совещание в состояние  $C_2$  или  $C_3$ .

Интенсивности потока воздействий  $\lambda_{31}$  определяющих совершение события, которое переводит систему из состояния  $C_3$  в состояние  $C_1$ , и  $\lambda_{21}$  переводящего систему из состояния  $C_2$  в состояние  $C_1$ , потока  $\lambda_{43}$ , переводящего состояние системы из  $C_4$  в  $C_3$ , потока  $\lambda_{42}$ , переводящего состояние системы из  $C_4$  в  $C_2$ , будем вычислять, предполагая, что среднее время дискурса зависит от глубины аргументации, интенсивности и остроты спора, ответственности каждого из участников за принятное решение. Априорно схему вероятных переходов внутри этих четырех состояний отобразим размеченным графиком на рис. 14.

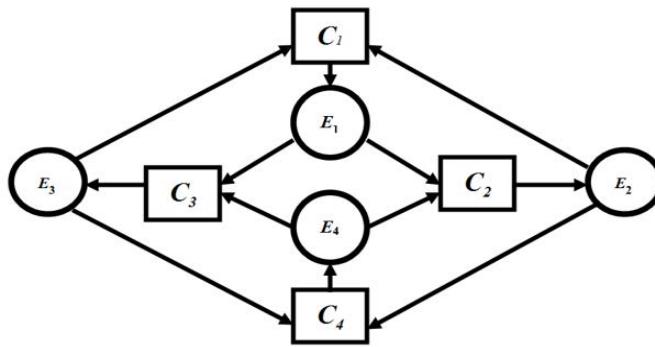


Рис. 14. Схема возможных переходов между четырьмя ожидаемыми состояниями

Имея в распоряжении размеченный граф переходов, используем уравнения Колмогорова для вероятностей состояний, можно написать и решить алгебраические уравнения для финальных вероятностей. Вероятность того, что система будет находиться в одном из промежуточных состояний  $c_h$  в  $j$ -й фазе обозначаем  $p_{j,h}$ ; вероятность этого и последующих состояний зависит от динамических характеристик события  $E_j(\lambda(t))$ , наличия прямых и обратных переходов в развитии ситуации. Если в качестве одного из аргументов присутствует интенсивность потока действий  $\lambda_{jx}(t)$ , а другого —  $\mu_{xy}$ , можно рассчитать вероятность состояния, следующего за событием, где имеют место эти воздействия, в частности, с помощью уравнений Колмогорова для нашего примера с четырьмя состояниями и четырьмя событиями.

Общее правило составления уравнений Колмогорова — в левой части каждого из них стоит производная вероятности какого-то ( $j$ -го) состояния. В правой части — сумма произведений вероятностей всех состояний, из которых идут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного ( $j$ -го) состояния.

Пользуясь этим правилом, запишем уравнения Колмогорова для системы  $S$ , размеченный граф состояний которой приведен на рис. 14. Общее правило составления уравнений Колмогорова — в левой части каждого из них стоит производная вероятности попадания какого-то состояния из  $h = 1, N$  в  $j$ -ю фазу. В правой части — сумма произведений вероятностей всех состояний, из которых идут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность попадания данного состояния в  $j$ -ю фазу.

Какой поток событий переводит систему обратно, например из  $C_{j+1}$  в  $C_j$ ? Очевидно, подавляющий поток воздействий при стремлении сущности (наиболее решительного субъекта) к цели (см. событие  $E_4$  на Рис. 14). Его интенсивность ( $\mu_{j+1,j}$ ) равна единице, деленной на среднее время события, достигающего состояния  $C_j$ . Аналогично вычисляются интенсивности потоков событий, переводящих систему по всем стрелкам графа рис. 14, где  $t$  — физическое время внутри события.

Для рассматриваемой ситуации из четырех событий и четырех состояний система уравнений запишется следующим образом:

$$\begin{aligned}
 \frac{dp_1}{dt} &= \mu_{21}p_2 + \mu_{31}p_3 - (\lambda_{12} + \lambda_{13})p_1, \\
 \frac{dp_2}{dt} &= \mu_{12}p_1 + \mu_{42}p_4 - (\lambda_{21} + \lambda_{24})p_2, \\
 \frac{dp_3}{dt} &= \mu_{13}p_1 + \mu_{43}p_4 - (\lambda_{31} + \lambda_{34})p_3, \\
 \frac{dp_4}{dt} &= \mu_{34}p_3 + \mu_{24}p_2 - (\lambda_{42} + \lambda_{43})p_4.
 \end{aligned} \tag{7.1}$$

Чтобы решить уравнения Колмогорова и найти вероятности состояний, прежде всего надо задать начальные условия. Если мы точно знаем начальное состояние системы  $C_0$  в началь-

ный момент (при  $t = 0$ ), то  $p_0(0) = 1$ , а все остальные начальные вероятности равны нулю. Так, например, уравнения, описывающие вероятности состояний в ситуации приведенные графом на рис. 14, естественно решать при начальных условиях  $p_0(0) = 1$ ,  $p_1(0) = p_2(0) = p_3(0) = 0$ . При  $t \rightarrow \infty$  в системе  $S$  устанавливается предельный стационарный режим, в ходе которого система случайным образом меняет свои состояния, но их вероятности уже не зависят от времени. Финальную вероятность состояния  $C_j$  можно истолковать как среднее относительное время пребывания системы в этом состоянии. Например, если система  $S$  имеет четыре состояния  $C_1$ ,  $C_2$ ,  $C_3$  и  $C_4$ , а их финальные вероятности равны 0.1; 0.2; 0.7 и 0.0, то это значит, что в предельном, стационарном режиме система в среднем 0.1 времени проводит в состоянии  $C_1$ , 0.2 – в состоянии  $C_2$ , большую часть времени – в состоянии  $C_3$ , а в состояние  $C_4$  она не переходит.

При приближении состояния системы, стремящейся к финалу, все производные становятся равны нулю, и уравнения приобретают вид:

$$\begin{aligned} (\lambda_{12} + \lambda_{13})p_1 &= \mu_{21}p_2 + \mu_{31}p_3, \\ (\lambda_{21} + \lambda_{24})p_2 &= \mu_{12}p_1 + \mu_{42}p_4, \\ (\lambda_{31} + \lambda_{34})p_3 &= \mu_{13}p_1 + \mu_{43}p_4, \\ (\lambda_{42} + \lambda_{43})p_4 &= \mu_{34}p_3 + \mu_{24}p_2. \end{aligned} \quad (7.2)$$

Учитывая, что сумма всех вероятностей в любом состоянии  $j$  равна единице – ( $p_1 + p_2 + p_3 + p_4 = 1$ ), и подставляя наблюдаемые значения интенсивности в это уравнение, можно получить значения вероятности состояния системы на следующем шаге.

Таким образом, уравнения Колмогорова дают возможность найти все вероятности перехода из состояния  $c_j$  в следующее состояние  $c_{j+1}$  как функцию интенсивности воздействий от времени  $E_j(\lambda_{jx}(t))$  и  $E_j(\mu_{jx}(t))$  внутри события и наглядно представить схему вероятных сценариев развития ситуации с помощью ориентированного размеченнего графа.

Однако вероятностная картина переходов, связанная в основном с критерием интенсивности действий внутри события, имеет сильный корректирующий фактор глубинной природы происходящего процесса, который определяет состав намерений и первоначально заявленных стратегий. Это оказывается как на динамике интенсивности воздействий  $\lambda_{jx}(t)$  и  $\mu_{jx}(t)$  во время свершения события  $E_j$ , так и на алфавите (множестве) состояний системы. В этом случае более надежного результата можно добиться, учитывая не столько поддающуюся измерению интенсивность воздействий в текущий момент времени, сколько оценку потенциала основных сущностей, которые являются морфологической базой рассматриваемой ситуации. Следовательно, мы получим схему высоковероятных переходов в виде линейного графа реализации двух разных сценариев.

Такие действующие модели можно реализовать с помощью языка агентов с реактивным поведением [6–8, 9] или языков визуального программирования с акцентом на обратные связи элементов многоуровневого патчевого описания модулей и субмодулей программируемой системы [10]. Здесь можно упорядочить выражение (1.1) и перейти от индексации  $F_{i,j}$  к развертыванию во временную последовательность, записав как функцию от времени  $F_{i,j}(t)$ . В этом случае динамическая модель ситуации представляет собой цепь чередования участков –  $j$ -х фаз плавно развивающихся процессов:

$$f_j(t) = \prod_{i=1}^N F_{j,i}(t) \quad (7.3)$$

с интенсивными потоками воздействий, которые оформляются в события-триггеры  $E$ , определяющими правила их переключения от  $j$ -й фазы к  $j+1$ -й фазе по предустановленным состояниям, которые диктуются мерой близости и порогами изменения  $sw_j$  морфологии системы.

Состояние системы в  $j+1$ -й фазе в результате процесса взаимодействия сущностей в рамках события  $E_j$  может быть определено выбором ближайшего значения из множества состояний  $C_h$ :

$$C_{j+1}(E_j) = \arg \min_{h=1, L} \left\{ C_h - \sum_i^N \delta_{j,i}(j) \right\}. \quad (7.4)$$

Для некоторых случаев удается решить заранее уравнения в буквенном виде, в частности, если граф состояний системы (рис. 14) представляет собой известную из математической биологии «схему гибели или размножения», его управляемую модель можно представить в виде патча динамики (рис. 15), а последовательность событий/состояний – в виде графа ветвления сценария (рис. 16).

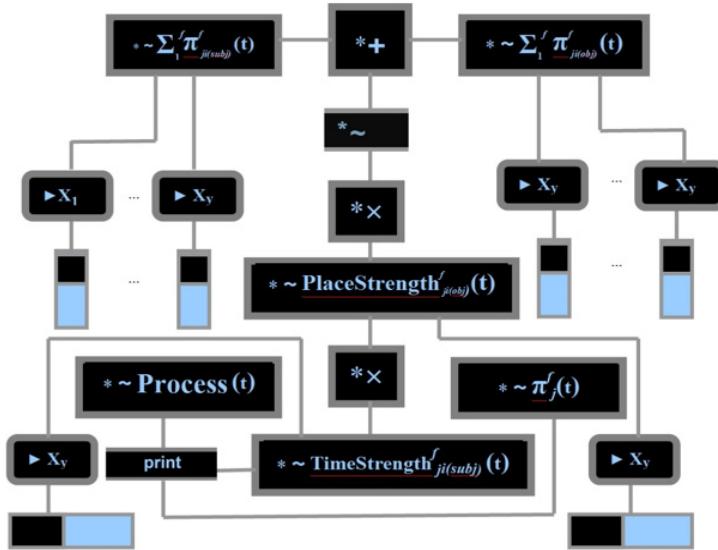


Рис. 15. Патч динамики суммы потенциалов при взаимодействии всех сущностей в текущей  $j$ -й фазе

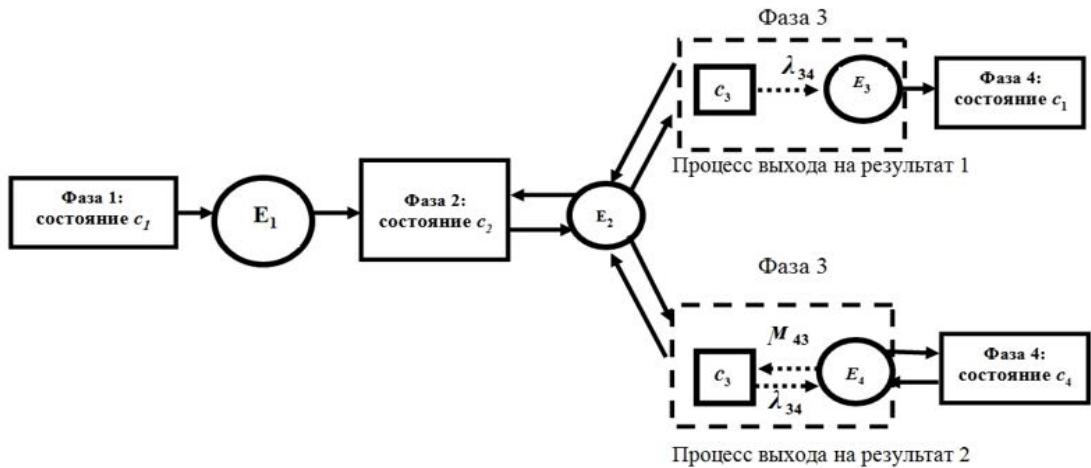


Рис. 16. Граф ветвления сценария на ранней фазе обучения системы и дальнейшего хода развития событий

**Заключение.** Рассмотрена задача имитации развития коммуникативных ситуаций по методу обучения с учителем на основании экспертных разметок. Применен подход автоматизированного выявления морфологии конкретной ситуации с помощью разработанной автором базовой морфологии для широкого класса таких ситуаций. Учитывались особенности морфологии коммуникативных ситуаций в отношении как самой природы именованных сущностей, так и основных динамических свойств базовых элементов — статуса, роли, качества, целевых приоритетов. На аналитическом материале исследователей из других научных областей (этнологии, юриспруденции, политологии) обобщены законы перехода или трансформации субъектов в объекты, в ролевые функции, а качества — в роли, статусы и т.д.

Параметризация именованных сущностей, обозначение связей между ними, конкретизация значения параметров нескольких первых фаз динамического развития вводились в систему экспертом на первом этапе создания и обучения модели. Расчет дальнейшего развития событий, оценка и классификации состояний системы осуществлялись с помощью математических методов распознавания, комбинаторных методов анализа, исследования операций и оптимального управления. Была сформирована концепция глубокого анализа структуры си-

туации для статичных состояний. Моделирование динамики параметров именованных сущностей внутри событий и обмен данными между моделями были поддержаны возможностями языка визуального программирования.

Методика построения динамических расчетных моделей — имитаторов коммуникативных ситуаций, заимствованная из комплекса программ MAX/MSP/PureData, была дополнена авторской концепцией математического описания структуры ситуации. Управляемые вычислительные модели были построены в виде графических схем, состоящих из интерактивных элементов, которые дают возможность наглядного управления динамикой параметров базовых сущностей и мониторинга происходящих процессов.

Модульная структура отображения ситуаций позволяет осуществлять широкий охват контекстных зависимостей, обеспечивает обмен данными между подключаемыми расчетными модулями, делает подобные ситуации сравнимыми с разных ракурсов, разной степени связности и детализации. Данный инструментальный подход может быть применен к описанию коммуникативных процессов не только по заданным экспертом, но и текстовым, графическим, фото/видео и аудиоданным. Он открывает путь к сравнению внешне схожих ситуаций, их классификации, включению в более широкий ситуативный контекст.

### СПИСОК ЛИТЕРАТУРЫ

1. *Barwise J., Perry J.* The Situation Underground // Stanford Working Papers in Semantics. V.1/Eds. J. Barwise, I. Sag. Stanford: Stanford Cognitive Science Group, 1980, Section D. P. 1–55.
2. *Devlin K.* Logic and Information. Cambridge: Cambridge University Press, 1991.
3. *Постолов Д.А.* Ситуационное управление // Теория и практика. М.: Наука, Физматлит, 1986.
4. *Пропп В.Я.* Морфология «волшебной» сказки. Исторические корни волшебной сказки// Собр.пр. В. Я. Проппа. М.: Лабиринт, 1998. 512 с.
5. *Леви-Строс К.* Структурная антропология. М.: Изд-во ЭКСМО-Пресс, 2001. 512 с.
6. *Бродский Ю.И.* Модельный синтез и модельно-ориентированное программирование // М.: ВЦ им. А. А. Дородницына РАН, 2013.
7. *Бродский Ю.И.* Роды структур Н. Бурбаки в задаче синтеза имитационных моделей сложных систем и модельно-ориентированное программирование // ЖВМ и МФ. 2015. Т. 55. № 1. С. 153–164.
8. *Бродский Ю.И.* О математическом моделировании поведения мультиагентных систем. М.: ФИЦ ИУ РАН, 2018.
9. *Бурбаки Н.* Теория множеств. М.: Мир, 1965. Springer: <https://www.springer.com/gp/book/9783540340348> (дата обращения: 23.09.2019).
10. MSP Tutorials // Table of Contents – [https://docs.cycling74.com/max8/tutorials/00\\_msplibindex](https://docs.cycling74.com/max8/tutorials/00_msplibindex) (дата обращения: 11.09.2023).

## РЕЛЕЙНЫЕ УПРАВЛЕНИЯ ДЛЯ БЕЗЫНЕРЦИОННОГО ДВУХЗВЕННОГО МАНИПУЛЯЦИОННОГО РОБОТА

© 2024 г. Ю. Ф. Долгий<sup>a, b, \*</sup>, И. А. Чупин<sup>b, \*\*</sup>

<sup>a</sup>Институт математики и механики УрО РАН, Екатеринбург, Россия

<sup>b</sup>Уральский федеральный ун-т, Екатеринбург, Россия

\*e-mail: jury.dolgy@urfu.ru

\*\*e-mail: mr.tchupin@yandex.ru

Поступила в редакцию 14.04.2024 г.

После доработки 29.07.2024 г.

Принята к публикации 16.10.2024 г.

Применяются релейные управление при решении задачи перевода безынерционного манипуляционного робота из заданного начального положения равновесия в заданное конечное положение равновесия. Предлагается метод нахождения точек переключения релейных управлений. Задача сводится к нахождению условий существования решения нелинейной краевой задачи для системы дифференциальных уравнений, определяющих траекторию движения манипулятора.

*Ключевые слова:* оптимальное управление, релейное управление, краевая задача, манипуляционный робот

**DOI:** 10.31857/S0002338824060128, **EDN:** STXELX

## RELAY CONTROLS FOR AN INERTIALESS TWO-LINK MANIPULATION ROBOT

Yu. F. Dolgii<sup>a, b, \*</sup>, I. A. Chupin<sup>b, \*\*</sup>

<sup>a</sup>Krasovskii Institute of Mathematics and Mechanics Ural Branch

of the Russian Academy of Sciences, Yekaterinburg, Russia

<sup>b</sup>Ural Federal University, Yekaterinburg, Russia

\*e-mail: jury.dolgy@urfu.ru

\*\*e-mail: mr.tchupin@yandex.ru

Relay controls are used to solve the problem of moving an inertialess manipulator from a given starting position to a desired final equilibrium position. A method for finding relay control switching points has been proposed. The main issue is to find the conditions for a solution to a nonlinear boundary value problem of a system of differential equations that determines the trajectory of a manipulator.

*Keywords:* optimal control, relay control, boundary problem, manipulation robot

**Введение.** Математические модели механических систем для манипуляционных роботов описываются управляемыми нелинейными системами обыкновенных дифференциальных уравнений. Управления выбираются таким образом, чтобы перевести манипуляционного робота из заданного начального положения равновесия в заданное конечное положение равновесия. Качество управления оценивается временем и энергией управления, которые затрачиваются на этот перевод. Движения управляемой механической системы являются решениями специальных краевых задач для обыкновенных дифференциальных уравнений. Аналитические решения рассматриваемых задач получены, когда математические модели манипуляционных роботов описываются интегрируемыми системами обыкновенных дифференциальных уравнений. Для построения интегрируемых математических моделей активно использовались методы декомпозиции [1]. Если неуправляемая механическая система с  $n$  степенями свободы описывается интегрируемой канонической системой уравнений Гамильтона, т.е. имеет  $n$

независимых первых интегралов, находящихся в инволюции, то, вычисляя полный интеграл уравнения Гамильтона–Якоби и применяя теорему Якоби, можно определить независимую систему  $2n$  первых интегралов канонической системы уравнений [2, 3]. С их помощью задается траектория свободного движения манипуляционного робота в фазовом пространстве, проекция которой на координатное пространство соединяет начальную и конечную точки движения манипулятора, а также вычисляется время движения манипулятора по траектории. Выход робота из начального положения равновесия на траекторию свободного движения можно обеспечить идеальным импульсным управлением. Аналогично с помощью идеального импульсного управления можно обеспечить торможение робота в конечном положении равновесия [4, 5]. При реализации идеальных импульсных управлений функцию Дирака можно моделировать прямоугольным импульсом. При оптимизации движений манипуляционного робота активно использовался принцип максимума Понтрягина [6]. Большое количество работ посвящено оптимизации движений двухзвенного манипулятора [7–11]. При нахождении оптимальных управлений применялись различные модификации численного метода для принципа максимума Понтрягина [12–14], а также краевые задачи для обыкновенных дифференциальных уравнений при построении оптимальных движений манипуляторов [1, 14, 15]. Для задачи оптимального быстродействия можно ограничиться программными релейными управлениями и находить только их точки переключения. При этом можно игнорировать уравнения для сопряженных переменных принципа максимума [15]. Для безынерционного двухзвенного манипуляционного робота при решении задачи оптимального быстродействия в работе [14] используется метод последовательных приближений для принципа максимума Понтрягина. В настоящей работе изучается задача нахождения точек переключения релейных управлений для безынерционного двухзвенного манипуляционного робота. Соответствующая механическая система имеет две степени свободы, и времена прихода манипулятора в конечное положение по различным обобщенным координатам могут отличаться [14]. В статье [16] с помощью метода декомпозиции были построены управления, которые обеспечивают равенство оценок времен движения по разным степеням свободы. При решении поставленной задачи используется специальная краевая задача для траектории движения манипулятора. В настоящей работе получены условия, для которых времена прихода манипулятора в конечное положение по различным координатам совпадают.

**1. Математическая модель манипулятора и постановка задачи.** Рассматривается двухзвенный манипуляционный робот (рис. 1), состоящий из двух абсолютно твердых тел и схвата с грузом, которые движутся параллельно горизонтальной плоскости. На рис. 2 приведена схема манипулятора в его проекции на горизонтальную плоскость. Первое тело цилиндрическим шарниром  $O$  связано с неподвижным основанием, а второе тело цилиндрическим шарниром  $O_1$  связано с первым телом. На конце второго тела в точке  $O_2$  укреплен схват, в котором находится перемещаемый объект (груз). Будем предполагать, что линейные размеры схвата и груза много меньше длины звеньев манипулятора, и при исследовании транспортных движений считать схват с грузом материальной точкой. Ограничимся рассмотрением безынерционного манипулятора, когда масса звеньев много меньше массы перемещаемого груза, считая, что вся его масса сосредоточена в точке  $O_2$  и равна массе груза  $m$ . Предположение о малости массы звеньев по сравнению с массой груза выполняется для некоторых исследовательских роботов [14].

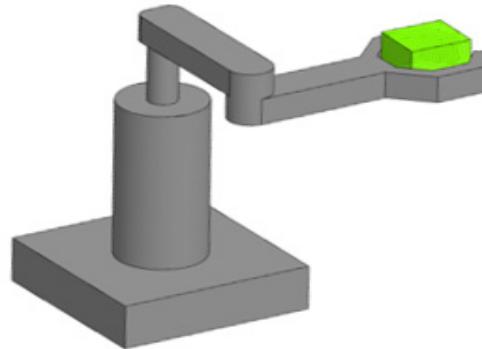


Рис. 1. Двухзвенный манипулятор.

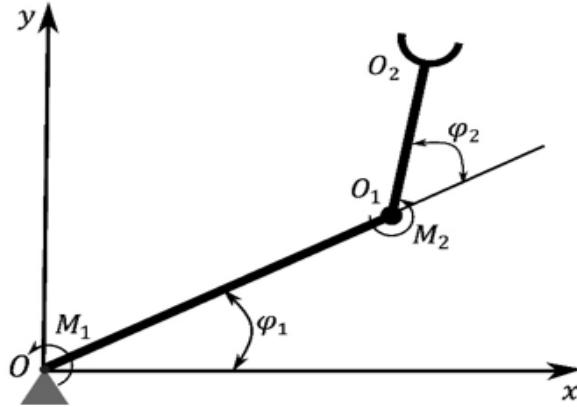


Рис. 2. Схема двухзвенного манипулятора.

Манипулятор управляется при помощи двух независимых приводов, расположенных в шарнирах  $O$  и  $O_1$  соответственно. Главные моменты сил, создаваемых приводами, равны  $M_1$  и  $M_2$ . Для описания движения схватка вводится в горизонтальной плоскости движения звеньев манипулятора неподвижная декартова система координат  $Oxy$  с началом в точке  $O$ . Пусть  $x$  и  $y$  – декартовые координаты схватка (груза). Закон движения схватка определяется формулами  $x(t)$ ,  $y(t)$ . Предполагается, что в начальный момент времени  $t = 0$  схват находится в начальном положении  $x(0) = x_0$ ,  $y(0) = y_0$  и имеет нулевые проекции скорости  $\dot{x}(0) = 0$ ,  $\dot{y}(0) = 0$ . Требуется привести его в момент времени  $t = T$  в заданное конечное положение  $x(T) = x_T$ ,  $y(T) = y_T$  с нулевой скоростью  $\dot{x}(T) = 0$ ,  $\dot{y}(T) = 0$ . Задача управления состоит в нахождении программных законов изменения управляемых моментов  $M_1(t)$ ,  $M_2(t)$ ,  $0 \leq t \leq T$ , обеспечивающих приведение манипулятора из заданного начального состояния равновесия в заданное конечное положение равновесия.

На рис. 2 угол  $\varphi_1$  между осью  $Ox$  неподвижной системы координат и прямой  $OO_1$ , соединяющей цилиндрические шарниры, определяет положение первого звена манипулятора, а угол  $\varphi_2$  между прямой  $OO_1$  и  $O_1O_2$ , соединяющей второй шарнир со схватом, находит относительное положение второго звена манипулятора относительно первого. Декартовы координаты схватка запишем как

$$x = L_1 \cos \varphi_1 + L_2 \cos (\varphi_1 + \varphi_2), \quad y = L_1 \sin \varphi_1 + L_2 \sin (\varphi_1 + \varphi_2),$$

где  $L_1 = |OO_1|$  – длина первого звена,  $L_2 = |O_1O_2|$  – длина второго звена. Будем полагать, что  $L_2 \leq L_1$ . Следовательно, рабочая зона манипулятора определяется неравенствами:

$$L_1 - L_2 \leq \sqrt{x^2 + y^2} \leq L_1 + L_2.$$

Рассматривается ситуация, когда значение угла  $\varphi_2$  положительно во время движения манипулятора и схват не достигает границ рабочей зоны.

Для описания движения манипулятора будем использовать уравнения Лагранжа 2-го рода. Их вид зависит от выбора обобщенных координат. Следуя [14], в качестве обобщенных координат возьмем полярные координаты точки  $O_2$ . Имеем:

$$x = r \cos \varphi, \quad y = \sin \varphi.$$

Зависимость углов  $\varphi_1$ ,  $\varphi_2$  от полярных координат определяется формулами:

$$\varphi_1 = \varphi - \arccos \frac{r^2 + L_1^2 - L_2^2}{2L_1 r},$$

$$\varphi_1 + \varphi_2 = \varphi + \arccos \frac{r^2 + L_2^2 - L_1^2}{2L_2 r}.$$

Находим вариации углов:

$$\delta\varphi_1 = \delta\varphi + \frac{(r^2 + L_2^2 - L_1^2)\delta r}{r \sqrt{(r^2 - (L_1 - L_2)^2)((L_1 + L_2)^2 - r^2)}},$$

$$\delta\varphi_1 + \delta\varphi_2 = \delta\varphi - \frac{(r^2 + L_1^2 - L_2^2)\delta r}{r\sqrt{(r^2 - (L_1 - L_2)^2)((L_1 + L_2)^2 - r^2)}}.$$

Кинетическая энергия манипулятора определяется как

$$K = \frac{m}{2}(\dot{r}^2 + r^2\dot{\varphi}^2),$$

а элементарная работа сил на виртуальных перемещениях —

$$\delta A = M_1\delta\varphi_1 + M_2\delta\varphi_2 = M_\varphi\delta\varphi + M_r\delta r$$

с обобщенными силами:

$$M_\varphi = M_1, M_r = \frac{r^2(M_1 - 2M_2) - (L_1^2 - L_2^2)M_1}{r\sqrt{(r^2 - (L_1 - L_2)^2)((L_1 + L_2)^2 - r^2)}}.$$

Дифференциальные уравнения Лагранжа 2-го рода имеют вид:

$$\frac{d(r^2\dot{\varphi})}{dt} = \frac{M_1(t)}{m}, \quad (1.1)$$

$$\ddot{r} - r\dot{\varphi}^2 = \frac{M_1(t)}{m}F_1(r) - \frac{M_2(t)}{m}F_2(r), \quad 0 \leq t \leq T, \quad (1.2)$$

где функции  $F_1(\cdot)$ ,  $F_2(\cdot)$  определяются следующим образом:

$$F_1(r) = \frac{r^2 - L_1^2 + L_2^2}{r\sqrt{(r^2 - (L_1 - L_2)^2)((L_1 + L_2)^2 - r^2)}},$$

$$F_2(r) = \frac{2r}{\sqrt{(r^2 - (L_1 - L_2)^2)((L_1 + L_2)^2 - r^2)}}, \quad L_1 - L_2 < r < L_1 + L_2.$$

Требуется найти управлении  $M_1(t)$ ,  $M_2(t)$ ,  $t \in [0, T]$ , переводящие систему (1.1), (1.2) из заданного начального положения равновесия:

$$\varphi(0) = \varphi_0, \dot{\varphi}(0) = 0, r(0) = r_0, \dot{r}(0) = 0,$$

в заданное конечное положение равновесия:

$$\varphi(T) = \varphi_T, \dot{\varphi}(T) = 0, r(T) = r_T, \dot{r}(T) = 0,$$

и время  $T$ , необходимое для этой операции. Здесь  $L_1 - L_2 < r_0$ ,  $r_T < L_1 + L_2$ ,  $\varphi_T \neq \varphi_0$ ,  $r_T \neq r_0$ . Без ограничения общности, используя специальный выбор направления оси  $x$ , можно положить  $\varphi_0 = 0$ .

При реализации рассматриваемой операции будем применять специальные релейные управлений:

$$M_1(t) = \begin{cases} M_1^0, & 0 \leq t < t_\varphi, \\ -M_1^0, & t_\varphi \leq t \leq T, \end{cases}$$

$$M_2(t) = \begin{cases} M_2^0, & 0 \leq t < t_r, \\ -M_2^0, & t_r \leq t \leq T, \end{cases}$$

с заданными модулями величин  $M_1^0$  и  $M_2^0$ . Знаки этих величин задают направления управляющих моментов сил и определяются специальным образом. В результате требуется найти

условия, когда решение поставленной выше задачи существует в классе выбранных релейных управлений, а также моменты  $t_\varphi$ ,  $t_r$  переключения релейных управлений и время  $T$  перехода манипулятора из начального положения равновесия в конечное.

**2. Краевая задача для траектории манипулятора.** Сформулированная задача имеет решение, если существуют такие значения параметров  $t_\varphi$ ,  $t_r$ ,  $T$ ,  $M_1^0$ ,  $M_2^0$ , для которых система дифференциальных уравнений

$$\frac{d(r^2\dot{\varphi})}{dt} = \frac{M_1^0}{m} \operatorname{sgn}(t_\varphi - t), \quad (2.1)$$

$$\ddot{r} - r\dot{\varphi}^2 = \frac{M_1^0 \operatorname{sgn}(t_\varphi - t)}{m} F_1(r) - \frac{M_2^0 \operatorname{sgn}(t_r - t)}{m} F_2(r) \quad (2.2)$$

с краевыми условиями

$$\varphi(0) = 0, \dot{\varphi}(0) = 0, \varphi(T) = \varphi_T, \dot{\varphi}(T) = 0, \quad (2.3)$$

$$r(0) = r_0, \dot{r}(0) = 0, r(T) = r_T, \dot{r}(T) = 0 \quad (2.4)$$

имеет решение. Ограничимся изучением случая, когда  $\varphi_T > 0$ ,  $r_T \neq r_0$ .

Преобразуем указанную краевую задачу. Пусть требуемые значения параметров  $M_1^0$ ,  $M_2^0$ ,  $t_\varphi$ ,  $t_r$  и  $T$  существуют и им отвечает решение  $\varphi(t)$ ,  $r(t)$ ,  $t \in [0, T]$ , краевой задачи (2.1)–(2.4). Тогда из (2.1), (2.3) находим:

$$\dot{\varphi}(t) = \frac{M_1^0 \alpha(t)}{mr^2(t)}, \quad \varphi(t) = \frac{M_1^0}{m} \int_0^t \frac{\alpha(s)}{r^2(s)} ds, \quad (2.5)$$

где функция  $\alpha(\cdot)$  определяется формулами:

$$\alpha(t) = \begin{cases} t, & 0 \leq t < t_\varphi, \\ 2t_\varphi - t, & t \leq t_\varphi < T, \end{cases} \quad t \in [0, T].$$

С учетом условия  $\dot{\varphi}(T) = 0$  имеем  $\alpha(T) = 0$ . Следовательно, должно выполняться равенство  $T = 2t_\varphi$ .

Из (2.5) следует, что на отрезке  $[0, T]$  функция  $\varphi(\cdot)$  возрастает и  $M_1^0 > 0$ . Тогда существует обратная функция  $t(\varphi)$ ,  $\varphi \in [0, \varphi_T]$ ,  $t(0) = 0$ ,  $t(\varphi_T) = T$  и углы  $\varphi_\varphi$ ,  $\varphi_r$ , такие, что  $t(\varphi_\varphi) = t_\varphi$ ,  $t(\varphi_r) = t_r$ . Эта функция неявно определяется уравнением:

$$\varphi = \frac{M_1^0}{m} \int_0^{t(\varphi)} \frac{\alpha(s)}{r^2(s)} ds, \quad \varphi \in [0, \varphi_T],$$

дифференцируя которое, имеем:

$$\frac{M_1^0}{m} \frac{\alpha(t(\varphi))}{\dot{r}^2(\varphi)} t'(\varphi) = 1, \quad \varphi \in [0, \varphi_T]. \quad (2.6)$$

Здесь  $\dot{r}(\varphi) = r(t(\varphi))$ ,  $\varphi \in [0, \varphi_T]$ ,

$$\alpha(t(\varphi)) = \begin{cases} t(\varphi), & \varphi \in [0, \varphi_\varphi], \\ T - t(\varphi), & \varphi \in [\varphi_\varphi, \varphi_T]. \end{cases} \quad (2.7)$$

Из (2.6) и (2.7) получим:

$$t(\varphi) t'(\varphi) = \frac{m}{M_1^0} \dot{r}^2(\varphi), \quad \varphi \in [0, \varphi_\varphi],$$

$$(T - t(\varphi)) t'(\varphi) = \frac{m}{M_1^0} \dot{r}^2(\varphi), \quad \varphi \in [\varphi_\varphi, \varphi_T].$$

Отсюда находим:

$$\begin{aligned} t^2(\varphi) &= \frac{2m}{M_1^0} \int_0^{\varphi} \tilde{r}^2(s) ds, \varphi \in [0, \varphi_\varphi], \\ (T - t(\varphi))^2 &= \frac{2m}{M_1^0} \int_{\varphi}^{\varphi_T} \tilde{r}^2(s) ds, \varphi \in [\varphi_\varphi, \varphi_T]. \end{aligned} \quad (2.8)$$

В уравнении (2.2) независимую переменную  $t$  заменим на  $\varphi$ . Имеем:

$$\begin{aligned} \dot{r}(t) \Big|_{t=t(\varphi)} &= \frac{M_1^{0^2} \alpha^2(t(\varphi))}{m^2 \tilde{r}^3(\varphi)} + \\ &+ \frac{M_1^0 \operatorname{sgn}(\varphi_\varphi - \varphi)}{m} F_1(\tilde{r}(\varphi)) - \frac{M_2^0 \operatorname{sgn}(\varphi_r - \varphi)}{m} F_2(\tilde{r}(\varphi)). \end{aligned} \quad (2.9)$$

Учитывая определение функции  $\tilde{r}(\cdot)$  и формулы (2.5), получим:

$$\begin{aligned} \dot{r}(t) \Big|_{t=t(\varphi)} &= \frac{d\tilde{r}(\varphi(t))}{dt} \Big|_{t=t(\varphi)} = \tilde{r}'(\varphi) \dot{\varphi}(t) \Big|_{t=t(\varphi)} = \tilde{r}'(\varphi) \frac{M_1^0 \alpha(t(\varphi))}{m \tilde{r}^2(\varphi)}, \\ \dot{r}(t) \Big|_{t=t(\varphi)} &= \frac{d^2\tilde{r}(\varphi(t))}{dt^2} \Big|_{t=t(\varphi)} = \frac{M_1^0}{m} \frac{d}{d\varphi} \left( \frac{\tilde{r}'(\varphi) \alpha(t(\varphi))}{\tilde{r}^2(\varphi)} \right) \dot{\varphi}(t) \Big|_{t=t(\varphi)} = \\ &= \frac{M_1^{0^2}}{m^2} \frac{d}{d\varphi} \left( \frac{\tilde{r}'(\varphi) \alpha(t(\varphi))}{\tilde{r}^2(\varphi)} \right) \frac{\alpha(t(\varphi))}{\tilde{r}^2(\varphi)} = \\ &= \frac{M_1^{0^2}}{m^2} \left( \frac{d}{d\varphi} \left( \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} \right) \alpha^2(t(\varphi)) + \frac{\tilde{r}'(\varphi)}{2\tilde{r}^4(\varphi)} \frac{d\alpha^2(t(\varphi))}{d\varphi} \right). \end{aligned}$$

Используя (2.7) и (2.8), находим:

$$\dot{r}(t) \Big|_{t=t(\varphi)} = \frac{M_1^{0^2}}{m^2} \frac{d}{d\varphi} \left( \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} \right) \alpha^2(t(\varphi)) + \frac{M_1^0}{m} \operatorname{sgn}(\varphi_\varphi - \varphi) \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)}, \varphi \in [0, \varphi_T].$$

В результате уравнение (2.9) примет вид:

$$2R(\varphi) \frac{d}{d\varphi} \left( \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} \right) = \frac{2R(\varphi)}{\tilde{r}(\varphi)} + \operatorname{sgn}(\varphi_\varphi - \varphi) \left( F_1(\tilde{r}(\varphi)) - \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} \right) - \operatorname{sgn}(\varphi_r - \varphi) k F_2(\tilde{r}(\varphi)),$$

где  $k, R(\varphi)$  определяются формулами:

$$k = \frac{M_2^0}{M_1^0}, R(\varphi) = \frac{M_1^0}{2m} \frac{\alpha^2(t(\varphi))}{\tilde{r}^2(\varphi)}, \varphi \in [0, \varphi_T].$$

Используя формулы (2.7) и (2.8), имеем:

$$R(\varphi) = \frac{1}{\tilde{r}^2(\varphi)} \begin{cases} \int_0^{\varphi} \tilde{r}^2(s) ds, \varphi \in [0, \varphi_\varphi], \\ \int_{\varphi}^{\varphi_T} \tilde{r}^2(s) ds, \varphi \in [\varphi_\varphi, \varphi_T]. \end{cases}$$

Введем новую функцию:

$$\rho(\varphi) = \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)}, \quad \varphi \in [0, \varphi_T].$$

Из определения функций  $\tilde{r}(\cdot)$ ,  $R(\cdot)$ ,  $\rho(\cdot)$  следует, что они удовлетворяют следующей системе дифференциальных уравнений:

$$\tilde{r}' = \tilde{r}^2 \rho, \quad (2.11)$$

$$R' = -2\tilde{r}R\rho + \operatorname{sgn}(\varphi_\varphi - \varphi), \quad (2.12)$$

$$\rho' = \frac{1}{\tilde{r}} + \frac{1}{2R} \left( \operatorname{sgn}(\varphi_\varphi - \varphi) (F_1(\tilde{r}) - \rho) - \operatorname{sgn}(\varphi_r - \varphi) k F_2(\tilde{r}) \right). \quad (2.13)$$

Из условий (2.4) следует, что  $\tilde{r}(0) = r_0$ ,  $\tilde{r}(\varphi_T) = r_T$ . Из формулы (2.10) имеем  $R(0) = R(\varphi_T) = 0$ . Из определения функции  $\rho(\cdot)$  находим:

$$\begin{aligned} \rho(0) &= \lim_{\varphi \rightarrow 0^+} \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} = \frac{m}{M_1^0} \lim_{t \rightarrow 0^+} \frac{\dot{r}(t)}{\alpha(t)} = \frac{m}{M_1^0} \ddot{r}(+0) = F_1(r_0) - kF_2(r_0), \\ \rho(\varphi_T) &= \lim_{\varphi \rightarrow \varphi_T^-} \frac{\tilde{r}'(\varphi)}{\tilde{r}^2(\varphi)} = \frac{m}{M_1^0} \lim_{t \rightarrow T^-} \frac{\dot{r}(t)}{\alpha(t)} = -\frac{m}{M_1^0} \ddot{r}(T-0) = F_1(r_T) - kF_2(r_T). \end{aligned}$$

В результате манипулятор можно перевести из заданного начального положения равновесия в заданное конечное положение равновесия, если существуют такие значения  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$ , в которых для системы дифференциальных уравнений (2.12), (2.13) с краевыми условиями:

$$\tilde{r}(0) = r_0, \quad R(0) = 0, \quad \rho(0) = F_1(r_0) - kF_2(r_0), \quad (2.14)$$

$$\tilde{r}(\varphi_T) = r_T, \quad R(\varphi_T) = 0, \quad \rho(\varphi_T) = F_1(r_T) - kF_2(r_T), \quad (2.15)$$

соответствующая краевая задача имеет решение.

Для реализации операции одновременного по каждой координате прихода манипулятора в конечное положение равновесия требуется, чтобы отношение параметров управления соответствовало значению параметра  $k$ . Следовательно, значения релейных управлений не могут быть выбраны произвольными.

Если найдены параметры  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$ , то для определения времени перехода манипулятора из начального положения в конечное, используя равенство  $T = 2t_\varphi$  и формулы (2.8), имеем:

$$\frac{1}{4}T^2 = t_\varphi^2 = \frac{2m}{M_1^0} \int_0^{\varphi_\varphi} \tilde{r}^2(s) ds = \frac{2m}{M_1^0} \int_{\varphi_\varphi}^{\varphi_T} \tilde{r}^2(s) ds.$$

Откуда получаем формулу:

$$T^2 = \frac{4m}{M_1^0} \int_0^{\varphi_T} \tilde{r}^2(s) ds.$$

**3. Численное моделирование движений манипулятора.** При численном решении краевой задачи (2.11)–(2.15) используем модифицированный метод стрельбы. Пусть  $\tilde{r}^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\varphi \in [0, \varphi_T]$ , – компоненты решения системы (2.11)–(2.13) с начальными условиями  $\tilde{r}^+(0, \varphi_\varphi, \varphi_r, k) = r_0$ ,  $R^+(0, \varphi_\varphi, \varphi_r, k) = 0$ ,  $\rho^+(0, \varphi_\varphi, \varphi_r, k) = F_1(r_0) - kF_2(r_0)$ , а  $\tilde{r}^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\varphi \in [0, \varphi_T]$ , – компоненты решения системы (2.11)–(2.13) с начальными условиями  $\tilde{r}^-(\varphi_T, \varphi_\varphi, \varphi_r, k) = r_T$ ,  $R^-(\varphi_T, \varphi_\varphi, \varphi_r, k) = 0$ ,  $\rho^-(\varphi_T, \varphi_\varphi, \varphi_r, k) = F_1(r_T) - kF_2(r_T)$ . Если существует решение краевой задачи (2.11)–(2.15), то определенные выше решения начальных задач Коши совпадают. При фиксированном значении параметра  $\varphi_T$  находим значения параметров  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$  из условий существования решения краевой задачи (2.11)–(2.15).

Рассмотрим следующие параметры манипуляционного робота:  $L_1 = 5$ ,  $L_2 = 3$ ,  $m = 1$ ,  $r_0 = 6.5$ ,  $r_T = 3$ . При реализации условий  $0 < \varphi_r < \varphi_\varphi < \varphi_T$  система уравнений для нахождения параметров  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$  имеет вид:

$$\begin{aligned}\tilde{r}^{++}(\varphi_\varphi - 0, k) &= \tilde{r}^{--}(\varphi_\varphi + 0, k), \\ R^{++}(\varphi_\varphi - 0, k) &= R^{--}(\varphi_\varphi + 0, k), \\ \rho^{++}(\varphi_\varphi - 0, k) &= \rho^{--}(\varphi_\varphi + 0, k).\end{aligned}$$

Здесь  $\tilde{r}^{++}(\varphi, k) = \tilde{r}^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^{++}(\varphi, k) = R^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^{++}(\varphi, k) = \rho^+(\varphi, \varphi_\varphi, \varphi_r, k)$ , если  $\varphi \in [0, \varphi_\varphi]$ ,  $\tilde{r}^{--}(\varphi, k) = \tilde{r}^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^{--}(\varphi, k) = R^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^{--}(\varphi, k) = \rho^-(\varphi, \varphi_\varphi, \varphi_r, k)$ , если  $\varphi \in [\varphi_\varphi, \varphi_T]$ .

Методом стрельбы для заданного значения  $\varphi_T$  находятся значения параметров  $\varphi_r$ ,  $k$ , для которых существует такое значение  $\varphi_\varphi$ , что выполняется следующее неравенство:

$$\begin{aligned}(\tilde{r}^{++}(\varphi_\varphi - 0, k) - \tilde{r}^{--}(\varphi_\varphi + 0, k))^2 + (R^{++}(\varphi_\varphi - 0, k) - R^{--}(\varphi_\varphi + 0, k))^2 + \\ + (\rho^{++}(\varphi_\varphi - 0, k) - \rho^{--}(\varphi_\varphi + 0, k))^2 \leq \xi,\end{aligned}$$

где  $\xi$  – заданная погрешность.

При реализации условий  $0 < \varphi_\varphi < \varphi_r < \varphi_T$  система уравнений для нахождения параметров  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$  имеет вид:

$$\begin{aligned}\tilde{r}^{++}(\varphi_r - 0, k) &= \tilde{r}^{--}(\varphi_r + 0, k), \\ R^{++}(\varphi_r - 0, k) &= R^{--}(\varphi_r + 0, k), \\ \rho^{++}(\varphi_r - 0, k) &= \rho^{--}(\varphi_r + 0, k).\end{aligned}$$

Здесь  $\tilde{r}^{++}(\varphi, k) = \tilde{r}^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^{++}(\varphi, k) = R^+(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^{++}(\varphi, k) = \rho^+(\varphi, \varphi_\varphi, \varphi_r, k)$ , если  $\varphi \in [0, \varphi_\varphi]$ ,  $\tilde{r}^{--}(\varphi, k) = \tilde{r}^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $R^{--}(\varphi, k) = R^-(\varphi, \varphi_\varphi, \varphi_r, k)$ ,  $\rho^{--}(\varphi, k) = \rho^-(\varphi, \varphi_\varphi, \varphi_r, k)$ , если  $\varphi \in [\varphi_\varphi, \varphi_T]$ .

Методом стрельбы для заданного значения  $\varphi_T$  находятся значения параметров  $\varphi_\varphi$ ,  $k$ , для которых существует такое значение  $\varphi_r$ , что выполняется следующее неравенство:

$$\begin{aligned}(\tilde{r}^{++}(\varphi_r - 0, k) - \tilde{r}^{--}(\varphi_r + 0, k))^2 + (R^{++}(\varphi_r - 0, k) - R^{--}(\varphi_r + 0, k))^2 + \\ + (\rho^{++}(\varphi_r - 0, k) - \rho^{--}(\varphi_r + 0, k))^2 \leq \xi,\end{aligned}$$

где  $\xi$  – заданная погрешность.

**Таблица 1.** Значения параметров  $\varphi_\varphi$ ,  $\varphi_r$ ,  $k$ ,  $t_r$ ,  $t_\varphi$ ,  $T$  при различных значениях  $\varphi_T$

$\varphi_T$	$k$	$\varphi_r$	$\varphi_\varphi$	$t_r$	$t_\varphi$	$T$
0.05	9.575	0.014	0.016	0.919	0.974	1.953
0.08	5.941	0.023	0.026	1.193	1.240	2.480
0.1	4.732	0.030	0.032	1.350	1.390	2.780
0.13	3.616	0.040	0.041	1.574	1.591	3.182
0.15	3.121	0.046	0.046	1.695	1.695	3.426
0.16	2.920	0.051	0.052	1.786	1.770	3.544
0.2	2.318	0.064	0.070	2.061	1.992	3.983

Окончание таблицы 1

$\varphi_T$	$k$	$\varphi_r$	$\varphi_\varphi$	$t_r$	$t_\varphi$	$T$
0.3	1.524	0.096	0.123	2.733	2.473	4.948
0.5	0.919	0.164	0.272	4.138	3.290	6.579
0.8	0.640	0.274	0.575	6.310	4.313	8.627
1	0.579	0.352	0.804	7.661	4.883	9.764
1.2	0.559	0.431	1.041	8.854	5.351	10.698
1.3	0.561	0.470	1.158	9.355	5.541	11.084
1.5	0.582	0.543	1.400	10.200	5.806	11.613
1.7	0.637	0.600	1.647	10.682	5.847	11.694
1.9	0.744	0.618	1.885	10.635	5.580	11.159
1.95	0.754	0.641	1.949	11.104	5.619	11.245

Основные результаты численного моделирования представлены в таблице. Она показывает, что, начиная с некоторого значения  $\varphi_T$ , неравенство  $\varphi_\varphi > \varphi_r$  сменяется на  $\varphi_\varphi < \varphi_r$ . С ростом параметра  $\varphi_T$  до значения 1.2 параметр  $k$  убывает, после чего начинает возрастать.

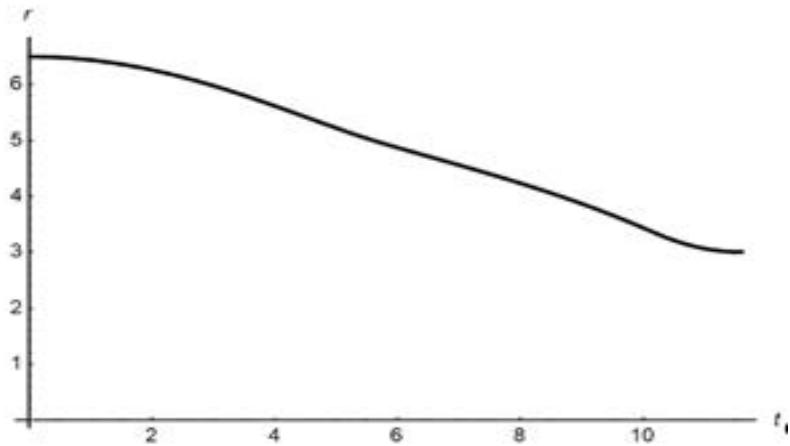


Рис. 3. График функции  $r(t)$ ,  $0 \leq t \leq T$ .

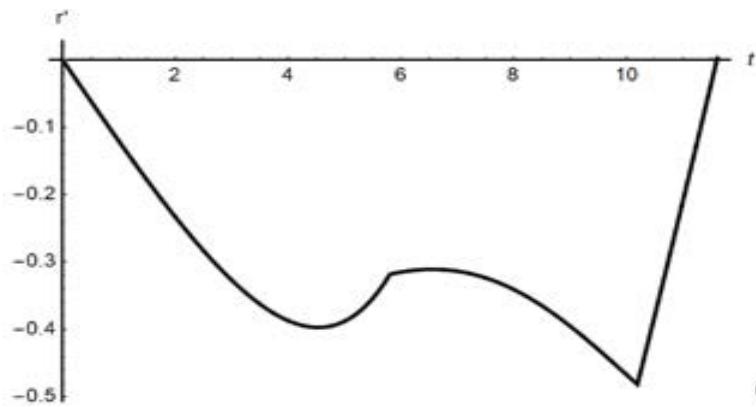


Рис. 4. График функции  $r'(t)$ ,  $0 \leq t \leq T$ .

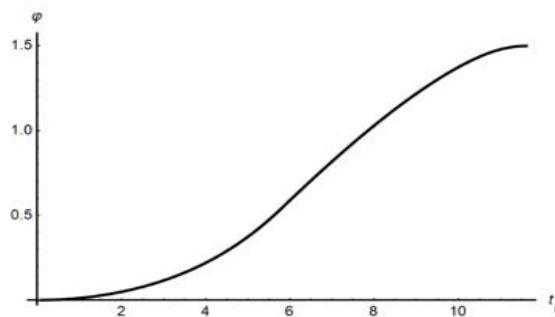
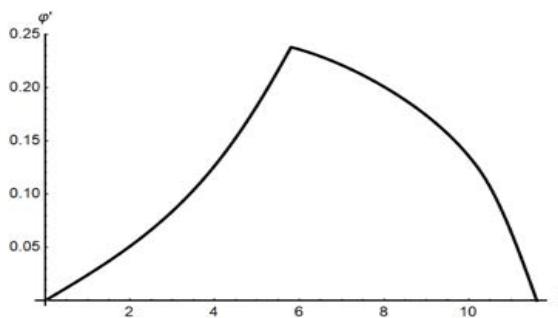
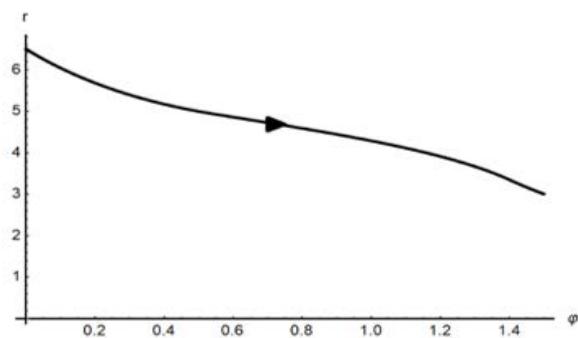
Рис. 5. График функции  $\varphi(t)$ ,  $0 \leq t \leq T$ .Рис. 6. График функции  $\varphi'(t)$ ,  $0 \leq t \leq T$ .

Рис. 7. Траектория схвата в полярной системе координат.

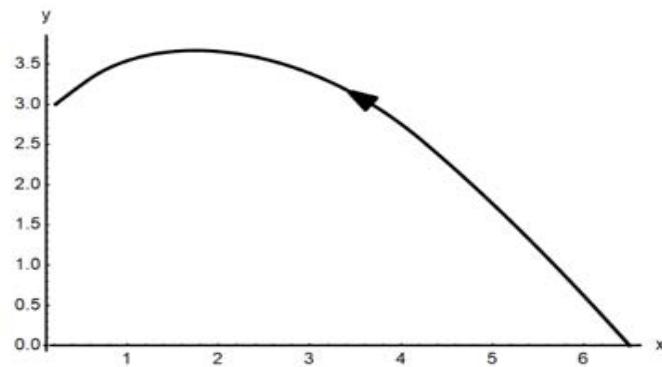


Рис. 8. Траектория схвата в декартовой системе координат.

Ниже на рис. 3–8 приведены графики компонент движения манипуляционного робота для значений параметров  $\varphi_T = 1.5$ ,  $\varphi_\phi = 0.582$ ,  $\varphi_r = 0.1400$ ,  $k = 0.582$ .

**Заключение.** Получено условие для релейных управлений, которое обеспечивает приход схвата в конечную заданную точку одновременно по обеим координатам. Предложен метод нахождения точек переключения релейных управлений для указанного случая движения манипуляционного робота. Метод может быть использован также, когда времена прихода схвата в заданную конечную точку по разным координатам не совпадают.

### СПИСОК ЛИТЕРАТУРЫ

- Черноусько Ф.Л., Ананьевский И.М., Решмин С.А. Методы управления нелинейными механическими системами. М.: Физматлит, 2006. 328 с.
- Маркеев А.П. Теоретическая механика. М.: Наука, 1990. 416 с.
- Архангельский Ю.А. Аналитическая динамика твердого тела. М.: Наука, 1977. 328 с.
- Долгий Ю.Ф., Чупин И.А. Импульсные управления двухзвенным манипуляционным роботом // Изв. Института математики и информатики Удмуртского государственного ун-та. 2021. № 57. С. 77–90.
- Чупин И.А. Нахождение импульсных управлений для многозвенных манипуляционных роботов // Вестн. Бурятского государственного ун-та. Математика. Информатика. 2023. № 4. С. 53–65.
- Понtryгин Л.С., Болтянский В.Г., Гамкrelidze Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. М.: Наука, 1969. 384 с.
- Куценок Л.Б., Лутманов С.В. Решение задачи предельного быстродействия управления движением плоского двухзвенного манипулятора // Вестн. Пермского ун-та. Математика. Механика. Информатика. 2014. № 3 (26). С. 28–33.
- Болотник Н.Н., Капунов А.А. Оптимизация управления и конфигураций двухзвенного манипулятора // Изв. АН СССР. Техн. кибернетика. 1983. № 4. С. 144–150.
- Meier E.B., Bryson A.E. Efficient Algorithm for Time-optimal of a Two-link Manipulator // J. Guidance, Control and Dynamics. 1990. V. 13. Iss. 5. P. 859–866.
- Аветисян В.В. Оптимальное по быстродействию управление перемещением схвата двухзвенного манипулятора с учетом типа конечной конфигурации // Аи Т. 2021. № 2. С. 3–15.
- Загороднев М.Ю., Афанасьев А.Ю. Оптимальное управление движением двухзвенного манипулятора с моментными двигателями // Вестн. Казанского государственного ун-та им. А. М. Туполева. 2014. № 1. С. 104–108.
- Крылов И.А., Черноусько Ф.Л. О методе последовательных приближений для решения задач оптимального управления // ЖВМ и МФ. 1962. Т. 2. № 6. С. 1132–1139.
- Любушин А.А. О применении метода последовательных приближений для решения задач оптимального управления // ЖВМ и МФ. 1982. Т. 22. № 1. С. 30–35.
- Черноусько Ф.Л., Болотник Н.Н., Градецкий В.Г. Манипуляционные роботы. Динамика, управление, оптимизация. М.: Наука, 1989. 368 с.
- Dolgii Yu.F., Chupin I.A. Optimal Control of Manipulator // The Bulletin of Irkutsk University. Series Mathematics. 2023. V. 43. P. 3–18.
- Решмин С.А. Синтез управления двухзвенным манипулятором // Изв. РАН. ТиСУ. 1997. № 2. С. 146–150.