

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2025

№ 70

Зарегистрирован в Федеральной службе по надзору в сфере
массовых коммуникаций, связи и охраны культурного наследия
(свидетельство о регистрации ПИ № ФС 77-29497 от 27 сентября 2007 г.)

Подписной индекс в объединённом каталоге «Пресса России» 44031

Журнал включен в «Перечень рецензируемых научных изданий,
в которых должны быть опубликованы основные научные результаты диссертаций
на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук»,
Высшей аттестационной комиссии

16+

Учредитель – Национальный исследовательский Томский государственный университет (НИ ТГУ)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Замятин Александр Владимирович – гл. редактор, проф., д-р техн. наук, НИ ТГУ. Тел. +73822529496
Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, НИ ТГУ. Тел. +73822529485
Нежелская Людмила Алексеевна – ответственный секретарь, проф., д-р физ.-мат. наук, НИ ТГУ. Тел. +73822529485
E-mail: vestnik_uvti@mail.tsu.ru

Воробейчиков Сергей Эрикович – д-р физ.-мат. наук, НИ ТГУ
Вишневский Владимир Миронович – проф., д-р техн. наук, Институт проблем управления РАН (г. Москва)
Гордов Евгений Петрович – проф., д-р физ. мат. наук, Институт мониторинга климатических и экологических систем СО РАН (г. Томск)
Кошкин Геннадий Михайлович – проф., д-р физ.-мат. наук, НИ ТГУ
Костюк Юрий Леонидович – проф., д-р техн. наук, НИ ТГУ
Матрсова Анжела Юрьевна – проф., д-р техн. наук, НИ ТГУ
Назаров Анатолий Андреевич – проф., д-р техн. наук, НИ ТГУ
Пашинская Татьяна Юрьевна – д-р физ.-мат. наук, НИ ТГУ
Перепелкин Евгений Александрович – проф., д-р техн. наук, Санкт-Петербургский государственный университет аэрокосмического приборостроения (г. Санкт-Петербург)
Пчелинцев Евгений Анатольевич – д-р физ.-мат. наук, НИ ТГУ
Самуйлов Константин Евгеньевич – проф., д-р техн. наук, РУДН (г. Москва)
Семенкин Евгений Станиславович – проф., д-р техн. наук, СибГУ им. акад. М.Ф. Решетнева (г. Красноярск)
Сушенко Сергей Петрович – проф., д-р техн. наук, НИ ТГУ
Фархадов Маис Паша оглы – д-р техн. наук, Институт проблем управления РАН (г. Москва)
Харин Юрий Семенович – академик НАНБ, проф., д-р физ.-мат. наук, БГУ, Республика Беларусь (г. Минск)
Цициашвили Гурами Шалвович – проф., д-р физ.-мат. наук, ИПМ ДВО РАН (г. Владивосток)

РЕДАКЦИОННЫЙ СОВЕТ

Александр Николаевич Дудин – проф., д-р физ.-мат. наук, БГУ, Минск, Республика Беларусь
Ана Роза Ковали (Ana Rosa Cavalli) – проф., д-р. философии (PhD), University VII Paris, Франция
Енцо Орсингер (Enco Orsinger) – проф., д-р. философии (PhD), University of Rome, Италия
Паоло Принетто (Paolo Prinetto) – проф. Politechnic Institute, Torino, Италия
Нина Владимировна Евтушенко – проф., д-р. техн. наук, ИСП РАН им. Иванникова В.П., Москва, Россия
Ервант Зориан (Yervant Zorian) – д-р. философии (PhD), Synopsys, Mountain View, США

JOURNAL INFO

Tomsk State University Journal of Control and Computer Science is an independent peer-reviewed research journal that welcomes submissions from across the world.
Tomsk State University Journal of Control and Computer Science is issued four times per year, and can be subscribed to in the Russian Press Joint Catalogue (Subscription Index 44031)
The publication in the journal is free of charge and may be in Russian or in English.
The topics of the journal are the following:
• *control of dynamical systems,*
• *mathematical modeling,*
• *data processing,*
• *informatics and programming,*
• *discrete function and automation,*
• *designing and diagnostics of computer systems.*
Rules of registration articles are given in a site:
<http://journals.tsu.ru/informatics/>
ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета. Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке
Статьи публикуются на русском и английском языках.
Тематика публикаций журнала:
• *управление динамическими системами,*
• *математическое моделирование,*
• *обработка информации,*
• *информатика и программирование,*
• *дискретные функции и автоматы,*
• *проектирование и диагностика вычислительных систем.*
Журнал входит в систему Российского Индекса Научного Цитирования (РИНЦ).
Правила оформления статей приведены на сайте:
<http://journals.tsu.ru/informatics/>
ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Антипина Е.В., Мустафина С.А., Антипин А.Ф. Численный алгоритм для поиска решения задачи оптимального быстрого действия при моделировании химических процессов	4
Мансимов К.Б. Об одной дискретной двухпараметрической задаче управления	15

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Каратаева Е.А., Старченко А.В. Параллельная реализация лагранжевой дисперсионной стохастической модели переноса примеси в атмосфере	29
Старовойтова А.В., Моисеев А.Н. Математическая модель нерегулируемого перекрестка в виде системы массового обслуживания с групповым входящим потоком и ненадежным прибором	42

ОБРАБОТКА ИНФОРМАЦИИ

Буркатовская Ю.Б., Воробейчиков С.Э. Оценивание параметров гиперэкспоненциального распределения	52
Зенкова Ж.Н., Андриевская А.А., Яковлев Н.И. Статистические свойства оценок метода измерения ценовой чувствительности PSM с учетом S^q -симметрии	62
Ким К.С., Смагин В.И. Робастная фильтрация в дискретных стохастических системах со скачкообразными параметрами и интервальной неопределенностью	73
Сперанский Д.В., Лунев С.А. Построение множеств простых путей между двумя узлами в нечетких транспортных сетях	81
Хайров М.А., Спицын В.Г. Нейросетевое тестирование алгоритма разметки данных для классификации уровней поддержки и сопротивления на финансовых рынках	91
Черникова О.С., Марарескул Т.А. Исследование точности определения расхождения шкал времени на основе нелинейных регрессионных моделей	103

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Бабанов А.М. Модель знаний оболочки экспертных систем EESS	113
Зеркаль С.М., Пешков А.В. Вычислительная диагностика объектов, позволяющих эвристическое усиление алгоритмов решения обратных задач	122

CONTENTS

CONTROL OF DYNAMICAL SYSTEMS

Antipina E.V., Mustafina S.A., Antipin A.F. Numerical algorithm for finding a solution to the problem of optimal speed in modeling chemical processes	4
Mansimov K.B. About one discrete two-parameter control problem	15

MATHEMATICAL MODELING

Karataeva E.A., Starchenko A.V. Parallel implementation of Lagrangian stochastic dispersion model of mixture transport	29
Starovoytova A.V., Moiseev A.N. Model of an unsignalized intersection in the form of queueing system with batch arrival process and unreliable server	42

DATA PROCESSING

Burkatovskaya Yu.B., Vorobeychikov S.E. Parameter estimation for hyperexponential distribution	52
Zenkova Z.N., Andrievskaya A.A., Yakovlev N.I. Statistical properties of PSM – price sensitivity meter technique with S^q -symmetry	62
Kim K.S., Smagin V.I. Robust filtering in discrete stochastic systems with jump parameters and interval uncertainty	73
Speranskiy D.V., Lunev S.A. Construction of sets of simple paths between two nodes in fuzzy transport networks	81
Khairov M.A., Spitsyn V.G. Neural network testing of a data labeling algorithm for classifying support and resistance levels in financial markets	91
Chernikova O.S., Marareskul T.A. Investigation of accuracy of determination of satellite clock bias based on nonlinear regression models	103

INFORMATICS AND PROGRAMMING

Babanov A.V. Knowledge model of the EESS expert system shell	113
Zerkal S.M., Peshkov A.V. Computational diagnostics of objects that allow the heuristic strengthening of algorithms for solving reverse problems	122

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

CONTROL OF DYNAMICAL SYSTEMS

Научная статья

УДК 519.6, 004.942

doi: 10.17223/19988605/70/1

**Численный алгоритм для поиска решения задачи оптимального быстродействия
при моделировании химических процессов****Евгения Викторовна Антипина¹, Светлана Анатольевна Мустафина²,
Андрей Федорович Антипин³**^{1,2} Уфимский университет науки и технологий, Уфа, Россия³ Стерлитамакский филиал Уфимского университета науки и технологий, Стерлитамак, Россия¹ stepashinaev@ya.ru² mustafina_sa@mail.ru³ andrejantipin@ya.ru

Аннотация. Рассматривается задача поиска оптимального по быстродействию управления химическим процессом, динамика которого описывается системой нелинейных дифференциальных уравнений. В качестве управляющего воздействия выступает температура реакционной смеси. Для решения поставленной задачи предлагается применить метод дифференциальной эволюции. Предложенный способ решения задачи оптимального быстродействия сформулирован в виде численного алгоритма, который позволяет получить решение оптимизационной задачи без необходимости выбора начального приближения. Работа алгоритма апробирована на промышленно значимом процессе синтеза фталевого ангидрида.

Ключевые слова: задача оптимального быстродействия; химический процесс; метод дифференциальной эволюции.

Благодарности: Исследование выполнено в рамках государственного задания Министерства науки и высшего образования Российской Федерации (код научной темы FZWU-2023-0002).

Для цитирования: Антипина Е.В., Мустафина С.А., Антипин А.Ф. Численный алгоритм для поиска решения задачи оптимального быстродействия при моделировании химических процессов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 4–14. doi: 10.17223/19988605/70/1

Original article

doi: 10.17223/19988605/70/1

**Numerical algorithm for finding a solution to the problem of optimal speed
in modeling chemical processes****Evgenia V. Antipina¹, Svetlana A. Mustafina², Andrey F. Antipin³**^{1,2} Ufa University of Science and Technology, Ufa, Russian Federation³ Sterlitamak branch of the Ufa University of Science and Technology, Sterlitamak, Russian Federation¹ stepashinaev@ya.ru² mustafina_sa@mail.ru³ andrejantipin@ya.ru

Abstract. The problem of finding the optimal control speed for a chemical process, the dynamics of which is described by a system of nonlinear differential equations, is considered. The temperature of the reaction mixture acts

as a control action. To solve the problem, it is proposed to apply the differential evolution method. The proposed method for solving the problem of optimal control speed is formulated as a numerical algorithm that allows obtaining a solution to the optimization problem without the need to select an initial approximation. The algorithm has been tested on an industrially significant process of phthalic anhydride synthesis.

Keywords: optimal speed problem; chemical process; differential evolution method.

Acknowledgments: This research was funded by the Ministry of Science and Higher Education of the Russian Federation (scientific code FZWU-2023-0002).

For citation: Antipina, E.V., Mustafina, S.A., Antipin, A.F. (2025) Numerical algorithm for finding a solution to the problem of optimal speed in modeling chemical processes. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 4–14. doi: 10.17223/19988605/70/1

Введение

При исследовании процессов химической технологии часто возникает необходимость в решении оптимизационных задач. Одна из таких задач – задача оптимального быстродействия, в которой требуется выбрать значения управляющих параметров в каждый момент времени, чтобы перевести процесс из заданного начального состояния в заданное конечное состояние за минимальное время. К данному классу задач относятся задачи поиска оптимального управления периодическими химико-технологическими процессами, задачи наискорейшего перевода процесса с одного режима эксплуатации на другой, задачи выбора оптимального температурного профиля в реакторе идеального вытеснения и др.

Одной из трудностей, возникающих при решении задач на основе математического описания химических процессов, является нелинейность их динамики. Наиболее часто математическую модель химического процесса представляют в виде системы обыкновенных дифференциальных уравнений, содержащую экспоненциальную зависимость константы скорости реакции от температуры [1, 2]. Нелинейность математических моделей химических процессов усложняет их исследование и управление и приводит к необходимости применения специальных подходов. Поэтому актуальной является разработка методов и алгоритмов поиска оптимального по быстродействию управления химическими процессами.

Одним из методов, применяемых для решения нелинейных задач оптимального быстродействия, является принцип максимума Понтрягина [3]. С помощью принципа максимума исходная задача сводится к краевой задаче для системы обыкновенных дифференциальных уравнений, решение которой связано с выбором начальных значений сопряженных переменных. Основная трудность применения данного метода при решении задач оптимального управления химическими процессами – необходимость определения начальных условий для решения сопряженной системы с учетом физико-химического смысла задачи.

При управлении химическими процессами в большинстве случаев достаточно иметь приближенное решение оптимизационной задачи ввиду необходимости оперативного реагирования на быстро меняющиеся условия протекания процесса. Поэтому для решения задачи быстродействия можно применить методы эволюционных вычислений, успешно зарекомендовавшие себя при решении практических задач, когда традиционные методы неэффективны [4, 5]. Один из таких методов – метод дифференциальной эволюции [6]. Данный метод является прямым и не зависит от начального приближения решения задачи. Отсутствие чувствительности метода дифференциальной эволюции к начальному приближению достигается за счет того, что на каждой итерации оптимизируется не одно возможное решение, а одновременно рассматривается их совокупность, что позволяет увеличить область поиска. Метод дифференциальной эволюции имеет меньшее количество настраиваемых параметров, в отличие от других методов эволюционного поиска, таких как генетические алгоритмы, метод искусственных иммунных систем и др. В работе [7] показано, что для решения оптимизационных задач с помощью метода дифференциальной эволюции затрачивается меньше времени по сравнению с другими стохастическими методами оптимизации.

Цель работы – разработка численного алгоритма определения решения задачи оптимального быстрогодействия для химического процесса на основе метода дифференциальной эволюции.

1. Постановка задачи

Пусть математическое описание химического процесса представляет собой систему обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = f(x(t), u(t), t), \quad (1)$$

с начальными условиями

$$x(0) = x^0, \quad (2)$$

где $x(t) = (x_1(t), \dots, x_n(t))^T$ – вектор концентраций веществ (фазовые переменные), $u(t)$ – параметр управления, $t \in [0, T]$ – время, $f(x(t), u(t), t) = (f_1(x(t), u(t), t), f_2(x(t), u(t), t), \dots, f_n(x(t), u(t), t))$ – вектор-функция, непрерывная вместе со своими частными производными.

Пусть заданы значения концентраций веществ в конечный момент времени $t = T$:

$$x(T) = x^1. \quad (3)$$

Параметром управления $u(t)$ могут выступать температура реакционной смеси, давление, скорость подачи реагентов и др. Рассмотрим в качестве управления $u(t)$ температуру реакции, на значения которой наложены ограничения

$$u_{\min} \leq u(t) \leq u_{\max}, \quad t \in [0, T]. \quad (4)$$

Будем искать температуру реакции $u(t)$ в классе кусочно-постоянных функций. Разобьем отрезок $[0, T]$ точками $t_1, t_2, \dots, t_s$ так, что $t_0 < t_1 < \dots < t_{s+1}$, $t_0 = 0$, $t_{s+1} = T$. На каждом из отрезков $t \in [t_j, t_{j+1}]$ температура принимает постоянное значение $u(t) = u_j$.

Требуется определить температурный режим $u(t)$ ($t \in [0, T]$), удовлетворяющий условию (4) и переводящий химический процесс, описываемый системой (1), из начального состояния (2) в конечное состояние (3) за минимальное время, т.е. минимизирующий целевой функционал

$$G(u) = T. \quad (5)$$

2. Алгоритм дифференциальной эволюции для решения задачи оптимального быстрогодействия

Сформулируем алгоритм поиска решения задачи (1)–(5) на основе метода дифференциальной эволюции. Работа метода основана на имитации эволюционных процессов, которым подвергаются векторы-индивиды [8, 9]. Функцией приспособленности, определяющей пригодность вектора-индивида в качестве решения задачи, выступает целевая функция. Обычно условием окончания расчетов является достижение заданного пользователем максимального количества итераций.

Введем в рассмотрение векторы-индивиды p_i ($i = \overline{1, m}$), образующие популяцию и являющиеся потенциальным решением задачи быстрогодействия. В качестве компонентов каждого вектора-индивида зададим управление $u_j = u(t_j)$, $t_j \in [0, T]$, $j = \overline{0, s}$, $t_0 = 0$, $t_{s+1} = T$, и продолжительность химической реакции:

$$p_{ij} = \begin{cases} u_{ij}, & j = \overline{0, s}, \\ T, & j = s + 1, \end{cases} \quad (6)$$

где $i = \overline{1, m}$.

Пусть функцией приспособленности является не целевой функционал (5), а функция вида:

$$Q(p_k) = \sqrt{\sum_{i=1}^n (x_i^k - x_i^1)^2}, \quad (7)$$

где $x_1^k, x_2^k, \dots, x_n^k$ – значения концентраций веществ, вычисленные путем решения системы дифференциальных уравнений (1) с начальными условиями (2) для вектора-индивида p_k в конечный момент времени $t = p_{k\ s+1}$ и кусочно-постоянного управления с узлами $(p_{i0}, p_{i1}, \dots, p_{is})$. Чем меньше значение $Q(p_k)$, тем точнее выполнены терминальные условия (3), поэтому вектор-индивид p_k является более приспособленным и больше других подходит в качестве решения задачи быстрогодействия.

Зададим в качестве критерия останова выполнение неравенства для некоторого вектора-индивида p_k текущей популяции:

$$Q(p_k) < \varepsilon, \quad (8)$$

где ε – заданный пользователем параметр.

Условие (8) может быть выполнено при нескольких возможных температурных режимах. Поэтому будем заносить возможные решения задачи быстрогодействия в отдельный массив векторов-индивидов *vectors*. В конце работы алгоритма из векторов-индивидов массива *vectors* выберем тот вектор p_k , у которого последний компонент $p_{k\ s+1}$, соответствующий продолжительности химической реакции, принимает наименьшее значение. Выбранный вектор-индивид будет приближенным решением задачи оптимального быстрогодействия (1)–(5).

Алгоритм дифференциальной эволюции для решения задачи (1)–(5) состоит из следующих шагов.

Шаг 1. Задать параметры алгоритма: m – количество векторов-индивидов в популяции, $\bar{T}$ – максимальная продолжительность химической реакции, s – количество точек разбиения отрезка $[0, \bar{T}]$, $kross \in [0, 1]$ – параметр скрещивания (кроссовера), $mut \in [0, 5; 1]$ – параметр мутации, ε – параметр окончания вычислений.

Шаг 2. Сгенерировать популяцию векторов-индивидов $p_i = (p_{i0}, p_{i1}, \dots, p_{is}, p_{i\ s+1})$, где

$$p_{ij} = \begin{cases} u_{\min} + \gamma_{ij}(u_{\max} - u_{\min}), & j = \overline{0, s}, \\ \gamma_{ij} \bar{T}, & j = s + 1. \end{cases} \quad (9)$$

Здесь $\gamma_{ij} \in [0, 1]$ – случайное число, $i = \overline{1, m}$.

Шаг 3. Рассчитать приспособленность каждого вектора-индивида p_i ($i = \overline{1, m}$). Для этого найти численное решение системы дифференциальных уравнений (1) с шагом разбиения интервала времени $[0, p_{i\ s+1}]$, равным $\frac{p_{i\ s+1}}{s}$.

Шаг 4. Проверить условие (8) для каждого вектора-индивида p_i ($i = \overline{1, m}$). Если оно выполнено, то поместить вектор p_i в массив *vectors*.

Шаг 5. Задать в качестве вектора-мишени первый вектор-индивид: $p_{mishen} = p_1$.

Шаг 6. Найти в популяции наиболее приспособленный вектор-индивид p_{prisp} .

Шаг 7. Выполнить операцию мутации. Выбрать случайным образом два различных вектора-индивида p_a , p_b , причем $p_a \neq p_{mishen}$, $p_b \neq p_{mishen}$, $p_a \neq p_{prisp}$, $p_b \neq p_{prisp}$. Создать вектор-мутант p_{mutant} :

$$p_{mutant} := p_{prisp} + mut(p_a - p_b).$$

Шаг 8. Выполнить операцию скрещивания для вектора-мутанта и вектора-мишени. Создать пробный вектор p_{prob} :

$$p_{prob\ j} = \begin{cases} p_{mutant\ j}, & \alpha_j \leq kross, \\ p_{mishen\ j}, & \alpha_j > kross, \end{cases} \quad j = \overline{0, s},$$

где $\alpha_j \in [0, 1]$ – случайное число.

Шаг 9. Вычислить значение функции приспособленности для пробного вектора $Q(p_{prob})$.

Шаг 10. Проверить условие (8) для пробного вектора p_{prob} . Если оно выполнено, то поместить пробный вектор в массив *vectors*.

Шаг 11. Обновить популяцию. Если $Q(p_{prob}) < Q(p_{mishen})$, то в популяцию вместо вектора-мишени поместить пробный вектор, т.е. $p_{mishen} = p_{prob}$. Иначе вектор-мишень p_{mishen} оставить в популяции.

Шаг 12. Если в качестве вектора-мишени рассмотрены все векторы-индивиды текущей популяции, т.е. $p_{mishen} = p_m$, то перейти к шагу 13. Иначе перейти к следующему в популяции вектору-мишени $p_{mishen} = p_{mishen+1}$ и перейти к шагу 6.

Шаг 13. Проверить условие (8) для векторов-индивидов текущей популяции. Если условие (8) выполнено для вектора p_i ($i = \overline{1, m}$), то поместить ее в массив *vectors* и перейти к шагу 14. Иначе перейти к шагу 5.

Шаг 14. Найти в массиве *vectors* вектор-индивид p_k , у которого последний компонент принимает наименьшее значение. В качестве наименьшей продолжительности химической реакции принять $T_{min} = p_{k\ s+1}$, а значениям температуры $u(t)$ в дискретные моменты времени присвоить значения элементов p_{kj} ($j = \overline{0, s}$) согласно формуле (6).

Сформулированный алгоритм дифференциальной эволюции для решения задачи оптимального быстрогодействия имеет следующие особенности:

- 1) структура вектора-индивида включает в себя дискретные значения параметра управления и времени;
- 2) функцией приспособленности вектора-индивида является не целевой функционал в форме (5), а отклонение между заданным конечным состоянием процесса (4) и вычисленным с помощью алгоритма;
- 3) введен специальный массив *vectors* для запоминания наиболее приспособленных векторов-индивидов, из которых в конце работы алгоритма выбирается решение с наименьшим значением времени;
- 4) условием окончания вычислительного процесса является отклонение от терминального состояния управляемого процесса.

3. Вычислительный эксперимент

Применим разработанный алгоритм для решения задач быстрогодействия на примере процесса получения фталевого ангидрида. На основе фталевого ангидрида производят различные красители, лекарственные препараты, инсектициды, присадки к смазочным маслам, добавки для авиационного топлива и др.

Введем следующие обозначения: X_1 – нафталин, X_2 – нафтохинон, X_3 – фталевый ангидрид, X_4 – углекислый газ, X_5 – малеиновый ангидрид. С учетом введенных обозначений совокупность химических превращений процесса получения фталевого ангидрида можно описать последовательностью стадий [10]:



Кинетику процесса можно описать с помощью системы обыкновенных дифференциальных уравнений

$$\frac{dx_i}{dt} = f_i(t, x, u) \quad (11)$$

с начальными условиями

$$x_i(0) = x_i^0, \quad i = \overline{1, 5}, \quad (12)$$

где $x = (x_1, x_2, \dots, x_5)$ – вектор концентраций веществ (моль/л), $t \in [0, T]$ – время протекания реакции (ч), u – температура реакции (К).

Согласно закону действующих масс правыми частями системы дифференциальных уравнений (10) являются функции

$$\begin{aligned} f_1(t, x, u) &= -k_1(u)x_1 - k_3(u)x_1 - k_4(u)x_1, \\ f_2(t, x, u) &= k_1(u)x_1 - k_2(u)x_2 - k_5(u)x_2, \\ f_3(t, x, u) &= k_3(u)x_1 + k_5(u)x_2 - k_6(u)x_3, \\ f_4(t, x, u) &= k_2(u)x_2 + k_4(u)x_1, \\ f_5(t, x, u) &= k_6(u)x_3, \end{aligned}$$

где $k_j(u)$ – константа скорости j -й стадии ($1/\text{ч}$, $j = \overline{1, 6}$), зависящая от температуры u исходя из уравнения Аррениуса

$$k_j(u) = k_{0j} e^{\left(-\frac{E_j}{Ru} \right)},$$

где k_{0j} – предэкспоненциальный множитель ($1/\text{ч}$), E_j – энергия активации j -й стадии (Дж/моль), R – универсальная газовая постоянная (8,31 Дж/(моль·К)).

Численные значения кинетических параметров реакции синтеза фталевого ангидрида приведены в работе [10].

Рассмотрим в качестве управления температуру реакции $u(t)$, на значения которой наложены ограничения

$$620 \text{ K} \leq u(t) \leq 644 \text{ K}, \quad t \in [0, T]. \quad (13)$$

Пусть начальные концентрации заданы значениями (моль/л)

$$x_1(0) = 1, \quad x_i(0) = 0, \quad i = \overline{2, 5}. \quad (14)$$

Целевым продуктом реакции (10) является фталевый ангидрид X_3 . Пусть требуется, чтобы его концентрация в конце реакции составляла 0,77 моль/л:

$$x_3(T) = 0,77. \quad (15)$$

Необходимо найти управление $u(t)$ с учетом ограничений (13), переводящее процесс, описываемый системой (11), из состояния (14) в состояние (15) за минимальное время $T_{\min}$.

Для решения задачи (10)–(15) на языке программирования Delphi разработан программный комплекс, реализующий алгоритм метода дифференциальной эволюции. Алгоритм применен с параметрами: размер популяции $m = 50$, максимальная продолжительность химической реакции $\bar{T} = 5$ ч, количество точек разбиения интервала времени $s = 50$, параметр скрещивания $kross = 0,7$, параметр мутации $mut = 0,8$, параметр окончания расчетов $\varepsilon = 0,01$.

В результате вычислений установлено, что наибольшее значение концентрации фталевого ангидрида, равное 0,763 моль/л, можно получить за наименьшее время $T_{\min} = 1,45$ ч, если на протяжении всего процесса поддерживать минимальную допустимую температуру (рис. 1). Относительная погрешность выполнения терминального условия составила 0,91%.

В табл. 1 представлены результаты решения системы дифференциальных уравнений (11) с начальными условиями (14) при некоторых допустимых значениях температуры и времени протекания реакции. Из табл. 1 видно, что с помощью разработанного алгоритма вычислено наименьшее время процесса синтеза фталевого ангидрида при температуре $u = 620 \text{ K}$, при этом $x_3(T) = 0,763$ моль/л.

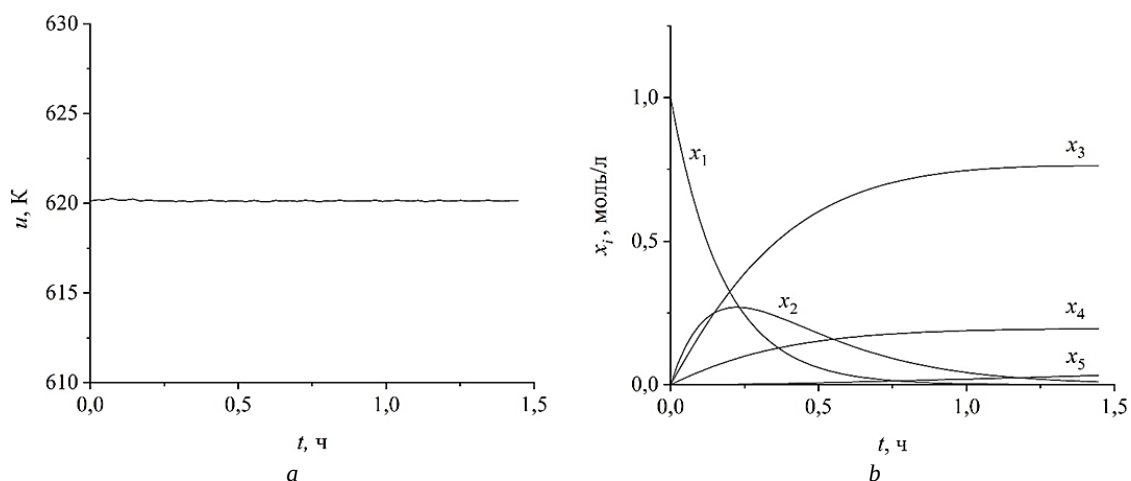


Рис. 1. Оптимальный температурный режим (а) и динамика концентраций веществ (b) в задаче (10)–(15)
Fig. 1. Optimal temperature regime (a) and dynamics of substance concentrations (b) in problem (10)–(15)

Таблица 1

Значение концентрации целевого вещества X_3 , вычисленное при различных условиях протекания реакции

u , К	T , ч	$x_3(T)$, моль/л
620	0,5	0,602
	1	0,746
	1,5	0,763
	2	0,754
	2,5	0,742
	3	0,728
628	0,5	0,632
	1	0,742
	1,5	0,742
	2	0,724
	2,5	0,703
	3	0,683
636	0,5	0,654
	1	0,728
	1,5	0,710
	2	0,680
	2,5	0,650
	3	0,621
644	0,5	0,666
	1	0,702
	1,5	0,663
	2	0,619
	2,5	0,577
	3	0,538

Рассмотрим задачу быстродействия для процесса синтеза фталевого ангидрида, в которой задана концентрация промежуточного вещества X_2 в конце реакции, равная 0,3 моль/л:

$$x_2(T) = 0,3. \quad (16)$$

С помощью разработанного алгоритма при $\varepsilon = 0,1$ вычислены оптимальный температурный режим и соответствующие концентрации веществ (рис. 2). Наименьшее время процесса составило 0,186 ч, при этом концентрация нафтохинона равна 0,286 моль/л, что на 4,7% ниже заданного значения (16).

В табл. 2 приведены значения концентрации нафтохинона, рассчитанные при варьировании продолжительности процесса для различных изотермических режимов. Из данных табл. 2 следует, что наиболее близкое значение концентрации нафтохинона к значению 0,3 моль/л (0,272 моль/л)

обеспечивается при постоянной температуре 628 К и продолжительности процесса, равной 0,2 ч. Однако погрешность выполнения условия (16) (9,3%) больше его погрешности для решения задачи, полученного с помощью алгоритма дифференциальной эволюции (4,7%). При этом время протекания процесса превышает вычисленное наименьшее значение продолжительности процесса, равное 0,186 ч.

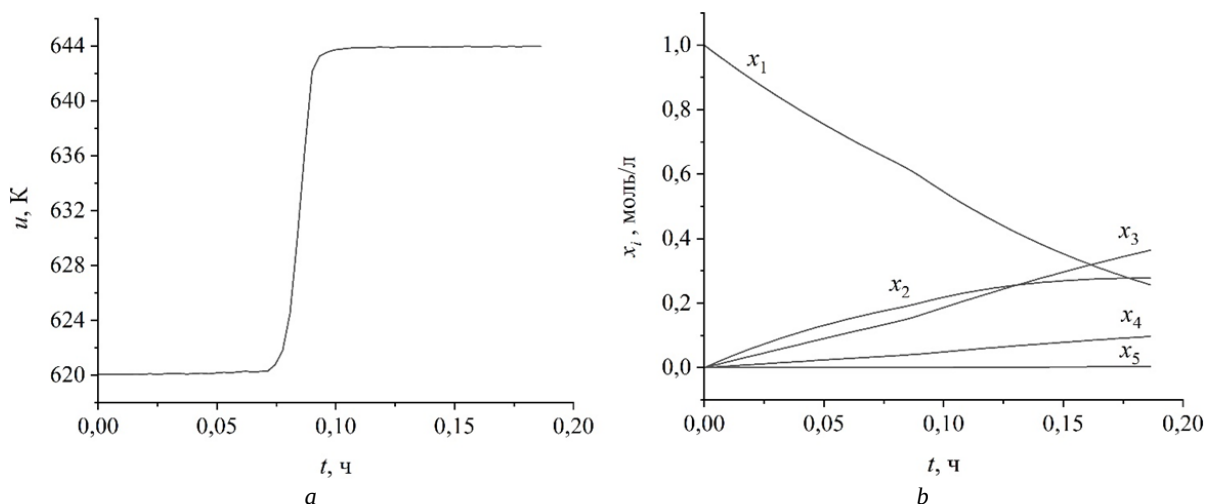


Рис. 2. Оптимальный температурный режим (а) и динамика концентраций веществ (b) в задаче (10)–(14), (16)
Fig. 2. Optimal temperature regime (a) and dynamics of substance concentrations (b) in problem (10)–(14), (16)

Таблица 2

Значение концентрации промежуточного вещества X_2 , вычисленное при различных условиях протекания реакции

u , К	T , ч	$x_2(T)$, моль/л
620	0,1	0,210
	0,2	0,268
	0,3	0,258
	0,5	0,179
	1	0,043
628	0,1	0,225
	0,2	0,272
	0,3	0,247
	0,5	0,155
	1	0,030
636	0,1	0,239
	0,2	0,271
	0,3	0,233
	0,5	0,130
	1	0,020
644	0,1	0,252
	0,2	0,266
	0,3	0,214
	0,5	0,107
	1	0,013

Пусть теперь требуется определить управление $u(t)$, при котором за наименьшее время полностью израсходуется исходное вещество – нафталин, т.е.

$$x_1(T) = 0. \quad (17)$$

Вычисления проведены с точностью $\varepsilon = 10^{-3}$. Расчеты показали, что реакцию следует проводить при максимально допустимой температуре $u = 644$ К на протяжении 1 ч (рис. 3, а). При этом в реакционной системе останется $1,61 \cdot 10^{-4}$ моль/л исходного вещества (рис. 3, b).

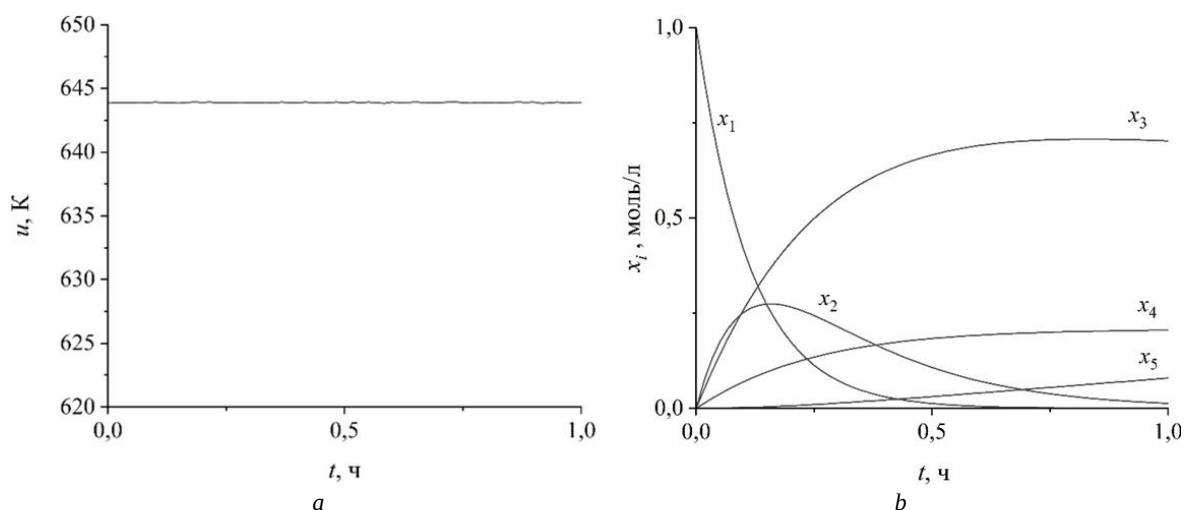


Рис. 3. Оптимальный температурный режим (a) и динамика концентраций веществ (b) в задаче (10)–(14), (17)
Fig. 3. Optimal temperature regime (a) and dynamics of substance concentrations (b) in problem (10)–(14), (17)

Значение концентрации нафталина вычислено при некоторых допустимых значениях управления u и времени T (табл. 3). Проанализировав таблицу, можно заметить, что заданной точности вычислений $\varepsilon = 10^{-3}$ с наименьшим временем протекания реакции, равной 1 ч, удовлетворяют два температурных режима: $u = 646$ К ($x_1(T) = 5,17 \cdot 10^{-4}$ моль/л) и $u = 644$ К ($x_1(T) = 1,61 \cdot 10^{-4}$ моль/л). Однако точность выполнения терминального ограничения (14) выше во втором случае. Поэтому в качестве решения задачи быстрогодействия (10)–(14), (17) можно принять значения управления u и времени $T_{\min}$, вычисленные с помощью алгоритма дифференциальной эволюции.

Таблица 3

Значение концентрации исходного вещества X_1 , вычисленное при различных условиях протекания реакции

u , К	T , ч	$x_1(T)$, моль/л
620	0,5	$0,60 \cdot 10^{-1}$
	1	$3,59 \cdot 10^{-3}$
	1,5	$2,15 \cdot 10^{-4}$
	2	$1,29 \cdot 10^{-5}$
	2,5	$7,72 \cdot 10^{-7}$
	3	$4,63 \cdot 10^{-8}$
628	0,5	$3,80 \cdot 10^{-2}$
	1	$1,45 \cdot 10^{-3}$
	1,5	$5,51 \cdot 10^{-5}$
	2	$2,09 \cdot 10^{-6}$
	2,5	$7,97 \cdot 10^{-8}$
	3	$3,03 \cdot 10^{-9}$
636	0,5	$2,30 \cdot 10^{-2}$
	1	$5,17 \cdot 10^{-4}$
	1,5	$1,17 \cdot 10^{-5}$
	2	$2,67 \cdot 10^{-7}$
	2,5	$6,06 \cdot 10^{-9}$
	3	$1,38 \cdot 10^{-10}$
644	0,5	$1,30 \cdot 10^{-2}$
	1	$1,61 \cdot 10^{-4}$
	1,5	$2,05 \cdot 10^{-6}$
	2	$2,59 \cdot 10^{-8}$
	2,5	$3,30 \cdot 10^{-9}$
	3	$4,19 \cdot 10^{-12}$

Исходя из результатов численных экспериментов, можно сделать вывод о корректной работе алгоритма поиска решения задачи оптимального быстрогодействия. Поэтому его можно применять

при исследовании закономерностей протекания химических процессов на основе их математических моделей.

Заключение

Таким образом, поиск решения задач оптимального быстродействия для химических процессов, описываемых нелинейными системами дифференциальных уравнений, можно осуществлять с помощью разработанного алгоритма дифференциальной эволюции. Особенность предложенного подхода состоит в том, что его можно применять для получения решения задач оптимального быстродействия при неизвестном начальном приближении, которое обычно задается исследователем, исходя из смысла поставленной задачи.

Работа алгоритма апробирована на процессе получения фталевого ангидрида. Вычислены оптимальный температурный режим и наименьшая продолжительность процесса при заданном значении концентрации целевого продукта реакции – фталевого ангидрида, при фиксированном содержании в реакционной смеси промежуточного вещества – нафтохинона, при условии полного израсходования исходного вещества – нафталина.

Алгоритм можно применять для решения задач быстродействия в химической технологии, в которых имеется несколько управляющих воздействий, путем модификации структуры вектора-индивида.

Список источников

1. Яблонский Г.С., Быков В.И., Горбань А.Н. Кинетические модели каталитических реакций. Новосибирск : Наука. 1983. 253 с.
2. Волков А.А., Булчевский Е.А., Лавренов А.В. Кинетика олигомеризации этилена на катализаторе $\text{NiO/B}_2\text{O}_3\text{-Al}_2\text{O}_3$ в жидкой фазе // Журнал Сибирского федерального университета. Сер. Химия. 2013. Т. 6, № 4. С. 352–360.
3. Карамзин Д.Ю. Принцип максимума Понтрягина для задачи оптимального управления с фазовыми ограничениями при ослабленных предположениях управляемости // Вопросы теории безопасности и устойчивости систем. 2018. № 20. С. 46–61.
4. Антипина Е.В., Мустафина С.И., Антипин А.Ф., Мустафина С.А. Численный алгоритм решения задачи оптимального управления с терминальными ограничениями для динамических систем // Автометрия. 2020. Т. 56, № 6. С. 132–140. doi: 10.15372/AUT20200615
5. Gulbaz R., Siddiqui A.B., Anjum N., Alotaibi A.A., Althobaiti T., Ramzan N. Balancer Genetic Algorithm–A Novel Task Scheduling Optimization Approach in Cloud Computing // Applied Sciences. 2021. V. 11, is. 14. Art. 6244. doi: 10.3390/app11146244
6. Mohamed A.W., Mohamed A.K. Adaptive Guided Differential Evolution Algorithm with Novel Mutation for Numerical Optimization // International Journal of Machine Learning and Cybernetics. 2019. V. 10. P. 253–277. doi: 10.1007/s13042-017-0711-7
7. Ковалевич А.А., Якимов А.И., Албкеират Д.М. Исследование стохастических алгоритмов оптимизации для применения в имитационном моделировании систем // Информационные технологии. 2011. № 8. С. 55–60.
8. Карпенко А.П. Эволюционные операторы популяционных алгоритмов глобальной оптимизации // Математика и математическое моделирование. 2018. № 1. С. 59–89. doi: 10.24108/mathm.0118.0000103
9. Storn R., Price K. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces // Journal of Global Optimization. 1997. V. 11. P. 341–359. doi: 10.1023/A:1008202821328
10. Антипина Е.В., Мустафина С.А., Антипин А.Ф. Численный алгоритм идентификации кинетической модели химической реакции // Вестник Технологического университета. 2019. Т. 22, № 9. С. 13–17.

References

1. Yablonsky, G.S., Bykov, V.I. & Gorban, A.N. (1983) *Kineticheskie modeli kataliticheskikh reaktsiy* [Kinetic models of catalytic reactions]. Novosibirsk: Nauka.
2. Volkov, A.A., Buluchevskiy, E.A. & Lavrenov, A.V. (2013) Kinetika oligomerizatsii etilena na katalizatore $\text{NiO/B}_2\text{O}_3\text{-Al}_2\text{O}_3$ v zhidkoy faze [Kinetics of Ethylene Oligomerization on $\text{NiO/B}_2\text{O}_3\text{-Al}_2\text{O}_3$ in Liquid Phase]. *Zhurnal Sibirskogo federal'nogo universiteta. Ser. Khimiya*. 6(4). pp. 352–360.
3. Karamzin, D.Yu. (2018) Printsip maksimuma Pontryagina dlya zadachi optimal'nogo upravleniya s fazovymi ogranicheniyami pri oslablennykh predpolozheniyakh upravlyaemosti [The Pontryagin maximum principle for state constrained optimal control problem under weakened controllability hypothesis]. *Voprosy teorii bezopasnosti i ustoychivosti sistem*. 20. pp. 46–61.
4. Antipina, E.V., Mustafina, S.I., Antipin, A.F. & Mustafina, S.A. (2020) Chislennyy algoritm resheniya zadachi optimal'nogo upravleniya s terminal'nymi ogranicheniyami dlya dinamicheskikh sistem [A numerical algorithm for solving optimal control problems with terminal constraints for dynamical systems]. *Avtometriya*. 56(6). pp. 132–140. DOI: 10.15372/AUT20200615

5. Gulbazi, R., Siddiqui, A.B., Anjum, N., Alotaibi, A.A., Althobaiti, T. & Ramzan, N. (2021) Balancer Genetic Algorithm – A Novel Task Scheduling Optimization Approach in Cloud Computing. *Applied Sciences*. 11(14). pp. 6244. DOI: 10.3390/app11146244
6. Mohamed, A.W. & Mohamed, A.K. (2019) Adaptive Guided Differential Evolution Algorithm with Novel Mutation for Numerical Optimization. *International Journal of Machine Learning and Cybernetics*. 10. pp. 253–277. DOI: 10.1007/s13042-017-0711-7
7. Kovalevich, A.A., Yakimov, A.I. & Albkeirat, D.M. (2011) Issledovanie stokhasticheskikh algoritmov optimizatsii dlya primeneniya v imitatsionnom modelirovanii sistem [Research of optimization stochastic algorithms for application in simulations of systems]. *Informatsionnye tekhnologii*. 8. pp. 55–60.
8. Karpenko, A.P. (2018) Evolyutsionnye operatory populyatsionnykh algoritmov global'noy optimizatsii [Evolutionary operators for global optimization population-based algorithms]. *Matematika i matematicheskoe modelirovanie*. 1. pp. 59–89. DOI: 10.24108/mathm.0118.0000103
9. Storn, R. & Price, K. (1997) Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*. 11. pp. 341–359. DOI: 10.1023/A:1008202821328
10. Antipina, E.V., Mustafina, S.A. & Antipin, A.F. (2019) Chislennyy algoritm identifikatsii kineticheskoy modeli khimicheskoy reaktsii [Numerical algorithm of identifying the kinetic model of a chemical reaction]. *Vestnik Tekhnologicheskogo universiteta*. 22(9). pp. 13–17.

Информация об авторах:

Антипина Евгения Викторовна – кандидат физико-математических наук, старший научный сотрудник управления научных исследований и разработок Уфимского университета науки и технологий (Уфа, Россия). E-mail: stepashinaev@ya.ru

Мустафина Светлана Анатольевна – профессор, доктор физико-математических наук, проректор по цифровой трансформации Уфимского университета науки и технологий (Уфа, Россия). E-mail: mustafina_sa@mail.ru

Антипин Андрей Федорович – доцент, кандидат технических наук, доцент кафедры прикладной информатики и программирования Стерлитамакского филиала Уфимского университета науки и технологий (Стерлитамак, Россия). E-mail: andrejantipin@ya.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Antipina Evgenia V. (Candidate of Physical and Mathematical Sciences, Senior Researcher, Ufa University of Science and Technology, Ufa, Russian Federation). E-mail: stepashinaev@ya.ru

Mustafina Svetlana A. (Doctor of Physical and Mathematical Sciences, Professor, Vice-Rector for Digital Transformation, Ufa University of Science and Technology, Ufa, Russian Federation). E-mail: mustafina_sa@mail.ru

Antipin Andrey F. (Candidate of Technical Sciences, Associate Professor, Sterlitamak branch of the Ufa University of Science and Technology, Sterlitamak, Russian Federation). E-mail: andrejantipin@ya.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 30.11.2024; принята к публикации 03.03.2025

Received 30.11.2024; accepted for publication 03.03.2025

Научная статья
УДК 517.977, 62-50
doi: 10.17223/19988605/70/2

Об одной дискретной двухпараметрической задаче управления

Камиль Байрамали оглы Мансимов

*Бакинский государственный университет, Баку, Азербайджан
Институт систем управления Министерства науки и образования Азербайджана, Баку, Азербайджан
kamilbmansimov@gmail.com*

Аннотация. Рассматривается дискретная задача оптимального управления, описываемая системой двумерных разностных уравнений типа Вольтерра с управляемым начальным условием. Предполагается, что начальная функция является решением одномерного разностного уравнения Вольтерра. Функционал качества является терминальным. Процесс управляется с двумя классами управляющих функций. Области управления являются ограниченными и открытыми множествами. Вычислены первая и вторая (в классическом смысле) вариации функционала качества. Доказан аналог уравнения Эйлера, и установлен ряд конструктивно проверяемых необходимых условий оптимальности второго порядка.

Ключевые слова: двумерное разностное уравнение типа Вольтерра; допустимое управление; область управления; открытое множество; вариация функционала; допустимая вариация управления; аналог уравнения Эйлера; необходимое условие оптимальности второго порядка.

Для цитирования: Мансимов К.Б. Об одной дискретной двухпараметрической задаче управления // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 15–28. doi: 10.17223/19988605/70/2

Original article
doi: 10.17223/19988605/70/2

About one discrete two-parameter control problem

Kamil B. Mansimov

*Baku State University, Baku, Azerbaijan
Institute of Control Systems of the Ministry of Science and Education of Azerbaijan, Baku, Azerbaijan
kamilbmansimov@gmail.com*

Abstract. We consider a discrete optimal control problem described by a system of two-dimensional Volterra-type difference equations with a controlled initial condition. It is assumed that the initial function is a solution to the one-dimensional Volterra difference equation. The quality functionality is terminal. The process is controlled with two classes of control functions. Control areas are bounded and open sets. The first and second (in the classical sense) variations of the quality functional are calculated. An analogue of the Euler equation is proved and a number of constructively verifiable necessary conditions for second-order optimality are established.

Keywords: two-dimensional difference equation of Volterra type, admissible control, control domain, open set, variation of the functional, admissible variation of the functional, analogue of the Euler equation, necessary condition for second-order optimality.

For citation: Mansimov, K.B. (2025) About one discrete two-parameter control problem. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 15–28. doi: 10.17223/19988605/70/2

Введение

К настоящему времени основательно изучен ряд аспектов задач оптимального управления, описываемых обыкновенными разностными уравнениями. Для подобных задач теория необходимых условий оптимальности первого и второго порядков достаточно полно разработана. Обзор соответствующих результатов имеется, например, в работах [1–7].

В работах [8, 9] и других изучены задачи оптимального управления, являющиеся дискретными аналогами задачи, рассмотренной в работах [10, 11] А.И. Москаленко.

В предлагаемой работе исследуется задача оптимального управления, описываемая двумерным разностным уравнением типа Вольтерра, а начальная функция определяется как решение управляемого одномерного разностного уравнения типа Вольтерра.

Таким образом, рассматриваемая задача является более общей, чем задачи, проанализированные в работах [8, 9].

В предположении открытости областей управления установлено необходимое условие оптимальности первого порядка в форме аналога уравнения Эйлера и доказан ряд необходимых условий оптимальности второго порядка.

1. Постановка задачи

Пусть $U_1 \subset R^r$, $U_2 \subset R^q$ – некоторые заданные непустые, ограниченные и открытые множества.

Предположим, что управляющий процесс описывается системой двумерных разностных уравнений типа Вольтерра

$$z(t+1, x) = \sum_{\tau=t_0}^t f(\tau, x, z(\tau, x), u_1(\tau)), \quad t \in T = \{t_0, t_0+1, \dots, t_1-1\}, \quad x \in X = \{x_0, x_0+1, \dots, x_1\} \quad (1)$$

с начальным условием

$$z(t_0, x) = a(x), \quad x \in X. \quad (2)$$

Здесь $f(t, \tau, x, z, u_1)$ – заданная n -мерная вектор-функция, дискретная по (t, τ, x) и непрерывная по (z, u_1) вместе с частными производными по (z, u_1) до второго порядка включительно, t_0, t_1, x_0, x_1 – заданные натуральные числа, $u_1(t)$ – r -мерный дискретный вектор управляющих воздействий со значениями из заданного непустого, ограниченного множества U_1 , т.е.

$$u_1(t) \in U_1 \subset R^r, \quad t \in T, \quad (3)$$

а начальная функция $a(x)$ определяется как решение аналога задачи Коши,

$$a(x+1) = \sum_{s=x_0}^x K(x, s, a(s), u_2(s)), \quad x \in X \setminus x_1, \quad (4)$$

$$a(x_0) = a_0, \quad (5)$$

где $K(x, s, a, u_2)$ – заданная n -мерная вектор-функция, дискретная по (x, s) и имеющая непрерывные по (a, u_2) производные до второго порядка включительно, a_0 – заданный постоянный вектор, $u_2(t)$ – q -мерная дискретная вектор-функция управляющих воздействий со значениями из заданного непустого, ограниченного множества U_2 , т.е.

$$u_2(x) \in U_2 \subset R^q, \quad x \in X \setminus x_1. \quad (6)$$

Управляющие функции $u_1(t)$ и $u_2(x)$ с вышеприведенными свойствами назовем допустимыми управлениями.

Как видно, уравнения (1) и (4) являются нелинейными разностными уравнениями типа Вольтерра.

Предполагается, что при каждом заданном управлении задачи Коши (1)–(2) и (4)–(5) имеют единственное дискретное решение.

На решениях задач Коши (1)–(2) и (4)–(5), порожденных всевозможными допустимыми управлениями, определим терминального типа функционал

$$J(u_1, u_2) = \sum_{x=x_0}^{x_1-1} \varphi_1(x, z(t_1, x)) + \varphi_2(a(x_1)). \quad (7)$$

Здесь $\varphi_1(x, z)$ – дискретная по x и дважды непрерывно дифференцируемая по z скалярная функция, а $\varphi_2(a)$ – заданная дважды непрерывно дифференцируемая скалярная функция.

Рассмотрим задачу о минимуме функционала (7) при ограничениях (1)–(6).

Допустимое управление $(u(t), v(t))$, доставляющее минимальное значение функционалу (7) при ограничениях (1)–(6), назовем оптимальным управлением.

Предполагается, что в рассматриваемой задаче оптимальное управление существует.

Целью предлагаемой статьи является вывод необходимых условий оптимальности первого (аналог уравнения Эйлера) и второго порядков.

2. Вычисление формулы приращения критерия качества

Пусть $(u_1(t), u_2(x))$ и $(\bar{u}_1(t) = u_1(t) + \Delta u_1(t), \bar{u}_2(x) = u_2(x) + \Delta u_2(x))$ – некоторые допустимые управления, а $(z(t, x), a(x))$ и $(\bar{z}(t, x) = z(t, x) + \Delta z(t, x), \bar{a}(x) = a(x) + \Delta a(x))$ – соответствующие им решения задач (1)–(2) и (4)–(5) соответственно.

Тогда, ясно, что $\Delta z(t, x)$ и $\Delta a(x)$ будут решениями следующих задач:

$$\Delta z(t+1, x) = \sum_{\tau=t_0}^t \left[f(t, \tau, \bar{z}(\tau, x), \bar{u}_1(\tau)) - f(t, \tau, z(\tau, x), u_1(\tau)) \right], \quad (8)$$

$$\Delta z(t_0, x) = \Delta a(x), \quad (9)$$

$$\Delta a(x+1) = \sum_{s=x_0}^x \left[K(x, s, \bar{a}(s), \bar{u}_2(s)) - K(x, s, a(s), u_2(s)) \right], \quad (10)$$

$$\Delta a(x_0) = 0. \quad (11)$$

Пусть $\psi(t, x)$ и $p(x)$ – пока произвольные дискретные n -мерные вектор-функции.

Учитывая тождества (8) и (10), получим, что

$$\sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t, x) \Delta z(t+1, x) = \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t, x) \sum_{\tau=t_0}^t \left[f(t, \tau, \bar{z}(\tau, x), \bar{u}_1(\tau)) - f(t, \tau, z(\tau, x), u_1(\tau)) \right], \quad (12)$$

$$\sum_{x=x_0}^{x_1-1} p'(x) \Delta a(x+1) = \sum_{x=x_0}^{x_1-1} p'(x) \sum_{s=x_0}^x \left[K(x, s, \bar{a}(s), \bar{u}_2(s)) - K(x, s, a(s), u_2(s)) \right]. \quad (13)$$

Учитывая начальные условия (9) и (11), доказывается, что

$$\begin{aligned} & \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t, x) \Delta z(t+1, x) = \\ & = \sum_{x=x_0}^{x_1-1} \psi'(t_1-1, x) \Delta z(t_1, x) - \sum_{x=x_0}^{x_1-1} \psi'(t_0-1, x) \Delta a(x) + \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t-1, x) \Delta z(t, x), \end{aligned} \quad (14)$$

$$\sum_{x=x_0}^{x_1-1} p'(x) \Delta a(x+1) = p'(x_1-1) \Delta a(x_1) + \sum_{x=x_0}^{x_1-1} p'(x-1) \Delta a(x). \quad (15)$$

Здесь $(')$ – операция транспонирования.

Далее, используя дискретный аналог теоремы Фубини (см., напр.: [12]), доказывается, что

$$\begin{aligned} & \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t, x) \left[\sum_{\tau=t_0}^t \left(f(t, \tau, \bar{z}(\tau, x), \bar{u}_1(\tau)) - f(t, \tau, z(\tau, x), u_1(\tau)) \right) \right] = \\ & = \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\sum_{\tau=t}^{t_1-1} \left(f(\tau, t, \bar{z}(t, x), \bar{u}_1(t)) - f(\tau, t, z(t, x), u_1(t)) \right) \right], \end{aligned} \quad (16)$$

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} p'(x) \left[\sum_{s=x_0}^x \left(K(x, s, \bar{a}(s), \bar{u}_2(s)) - K(x, s, a(s), u_2(s)) \right) \right] = \\ & = \sum_{x=x_0}^{x_1-1} \left[\sum_{s=x}^{x_1-1} \left(K(s, x, \bar{a}(x), \bar{u}_2(x)) - K(s, x, a(x), u_2(x)) \right) \right]. \end{aligned} \quad (17)$$

Введем аналоги функции Гамильтона в виде:

$$\begin{aligned} H(t, x, z(t, x), u_1(t), \psi(t, x)) &= \sum_{\tau=t}^{t_1-1} \psi'(\tau, x) f(\tau, t, x, z(t, x), u_1(t)), \\ M(x, a(x), u_2(x), p(x)) &= \sum_{s=x}^{x_1-1} p'(s) K(s, x, a(x), u_2(x)). \end{aligned}$$

Учитывая тождества (14)–(17) и введенные обозначения, формулу для приращения функционала (7) запишем в виде:

$$\begin{aligned} J(\bar{u}_1, \bar{u}_2) - J(u_1, u_2) &= \sum_{x=x_0}^{x_1-1} (\varphi_1(x, \bar{z}(t_1, x)) - \varphi_1(x, z(t_1, x))) + \varphi_2(\bar{a}(x_1)) - \varphi_2(a(x_1)) + \\ &+ \sum_{x=x_0}^{x_1-1} \psi'(t_1-1, x) \Delta z(t_1, x) - \sum_{x=x_0}^{x_1-1} \psi'(t_0-1, x) \Delta a(x) + \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t-1, x) \Delta z(t, x) - \\ &- \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} [H(t, x, \bar{z}(t, x), \bar{u}_1(t), \psi(t, x)) - H(t, x, z(t, x), u_1(t), \psi(t, x))] + \\ &+ p'(x_1-1) \Delta a(x_1) + \sum_{x=x_0}^{x_1-1} p'(x-1) \Delta a(x) - \sum_{x=x_0}^{x_1-1} [M(x, \bar{a}(x), \bar{u}_2(x), p(x)) - M(x, a(x), u_2(x), p(x))]. \end{aligned} \quad (18)$$

Учитывая условия гладкости, наложенные на параметры задачи, и применяя формулу Тейлора, получаем, что

$$\begin{aligned} & \varphi_1(x, \bar{z}(t_1, x)) - \varphi_1(x, z(t_1, x)) = \\ &= \frac{\partial \varphi_1'(x, z(t_1, x))}{\partial z} \Delta z(t_1, x) + \frac{1}{2} \Delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \Delta z(t_1, x) + o_1(\|\Delta z(t_1, x)\|^2), \end{aligned} \quad (19)$$

$$\varphi_2(\bar{a}(x_1)) - \varphi_2(a(x_1)) = \frac{\partial \varphi_2'(a(x_1))}{\partial a} \Delta a(x_1) + \frac{1}{2} \Delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \Delta a(x_1) + o_2(\|\Delta a(x_1)\|^2), \quad (20)$$

$$\begin{aligned} H(t, x, \bar{z}, \bar{u}_1, \psi) - H(t, x, z, u_1, \psi) &= \frac{\partial H'(t, x, z, u_1, \psi)}{\partial z} \Delta z + \frac{1}{2} \Delta z' \frac{\partial^2 H(t, x, z, u_1, \psi)}{\partial z^2} \Delta z + \\ &+ \frac{\partial H'(t, x, z, u_1, \psi)}{\partial u_1} \Delta u_1 + \Delta u_1' \frac{\partial^2 H(t, x, z, u_1, \psi)}{\partial u_1 \partial z} \Delta z + \frac{1}{2} \Delta u_1' \frac{\partial^2 H(t, x, z, u_1, \psi)}{\partial u_1^2} \Delta u_1 + o_3(\|\Delta z\| + \|\Delta u_1\|)^2, \end{aligned} \quad (21)$$

$$\begin{aligned} M(x, \bar{a}, \bar{u}_2, p) - M(x, a, u_2, p) &= \frac{\partial M'(x, a, u_2, p)}{\partial a} \Delta a + \frac{\partial M'(x, a, u_2, p)}{\partial u_2} \Delta u_2 + \frac{1}{2} \Delta a' \frac{\partial^2 M(x, a, u_2, p)}{\partial a^2} \Delta a + \\ &+ \Delta u_2' \frac{\partial^2 M(x, a, u_2, p)}{\partial u_2 \partial a} \Delta a + \frac{1}{2} \Delta u_2' \frac{\partial^2 M(x, a, u_2, p)}{\partial u_2^2} \Delta u_2 + o_4(\|\Delta a\| + \|\Delta u_2\|)^2. \end{aligned} \quad (22)$$

Здесь $\|\alpha\|$ есть норма вектора $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)'$, определяемая формулой $\|\alpha\| = \sum_{i=1}^n |\alpha_i|$, $o(\alpha^2)$ – величина более высокого порядка, чем α , т. е. $\frac{o(\alpha^2)}{\alpha^2} \rightarrow 0$ при $\alpha \rightarrow 0$.

Учитывая формулы (19)–(22) в формуле приращения (18), функционала (7), получаем, что

$$\begin{aligned} J(\bar{u}_1, \bar{u}_2) - J(u_1, u_2) &= \\ &= \sum_{x=x_0}^{x_1-1} \frac{\partial \varphi_1'(x, z(t_1, x))}{\partial z} \Delta z(t_1, x) + \frac{1}{2} \sum_{x=x_0}^{x_1-1} \Delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \Delta z(t_1, x) + \sum_{x=x_0}^{x_1-1} o_1(\|\Delta z(t_1, x)\|^2) + \end{aligned}$$

$$\begin{aligned}
 & + \sum_{x=x_0}^{x_1-1} \psi'(t_1-1, x) \Delta z(t_1, x) - \sum_{x=x_0}^{x_1-1} \psi'(t_0-1, x) \Delta a(x) - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \psi'(t-1, x) \Delta z(t, x) - \\
 & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\frac{\partial H'(t, x, z(t, x), u_1(t, x), \psi(t, x))}{\partial z} \Delta z(t, x) + \frac{\partial H'(t, x, z(t, x), u_1(t, x), \psi(t, x))}{\partial u_1} \Delta u_1(t, x) \right] + \\
 & + \frac{1}{2} \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\Delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \Delta z(t, x) + 2 \Delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \Delta z(t, x) + \right. \\
 & \left. + \Delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \Delta u_1(t) \right] - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} o_3 \left(\left[\|\Delta z(t, x)\| + \|\Delta u_1(t)\| \right]^2 \right) + \\
 & + p'(x_1-1) \Delta a(x_1) + \sum_{x=x_0}^{x_1-1} p'(x-1) \Delta a(x) - \\
 & - \sum_{x=x_0}^{x_1-1} \left[\frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial a} \Delta a(x) + \frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial u_2} \Delta u_2(x) \right] + \\
 & + \frac{1}{2} \sum_{x=x_0}^{x_1-1} \left[\Delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \Delta a(x) + 2 \Delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \Delta a(x) + \right. \\
 & \left. + 2 \Delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \Delta u_2(x) \right] - \sum_{x=x_0}^{x_1-1} o_4 \left(\left[\|\Delta a(x)\| + \|\Delta u_2(x)\| \right]^2 \right) + \frac{\partial \varphi_2'(a(x_1))}{\partial a} \Delta a(x_1) + \\
 & + \frac{1}{2} \Delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \Delta a(x_1) + o_2 \left(\|\Delta a(x_1)\|^2 \right). \quad (23)
 \end{aligned}$$

Теперь предположим, что вектор-функции $\psi(t, x)$ и $p(x)$ являются соответственно решениями следующих задач:

$$\begin{aligned}
 \psi(t-1, x) &= \frac{\partial H(t, x, z(t, x), u_1(t, x), \psi(t, x))}{\partial z}, \quad \psi(t_1-1, x) = -\frac{\partial \varphi_1(z(t_1, x))}{\partial z}, \\
 p(x-1) &= \frac{\partial M(x, a(x), u_2(x), p(x))}{\partial a} + \psi(t_0-1, x), \quad p(x_1-1) = -\frac{\partial \varphi_2(a(x_1))}{\partial a}.
 \end{aligned}$$

Тогда формула приращения (23) примет вид:

$$\begin{aligned}
 J(\bar{u}_1, \bar{u}_2) - J(u_1, u_2) &= - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \frac{\partial H'(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1} \Delta u_1(t, x) - \\
 &- \frac{1}{2} \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\Delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \Delta z(t, x) + 2 \Delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \Delta z(t, x) + \right. \\
 &+ \Delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \Delta u_1(t) \left. \right] + \frac{1}{2} \sum_{x=x_0}^{x_1-1} \Delta z'(t_1, x) \frac{\partial^2 \varphi_1(z(t_1, x))}{\partial z^2} \Delta z(t_1, x) - \\
 &- \sum_{x=x_0}^{x_1-1} \frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial u_2} \Delta u_2(x) - \frac{1}{2} \sum_{x=x_0}^{x_1-1} \left[\Delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \Delta a(x) + \right. \\
 &+ 2 \Delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \Delta a(x) + \Delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \Delta u_2(x) \left. \right] +
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} \Delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \Delta a(x_1) + o_2(\|\Delta a(x_1)\|^2) + \sum_{x=x_0}^{x_1-1} o_1(\|\Delta z(t_1, x)\|^2) - \\
 & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} o_3(\|\Delta z(t, x)\| + \|\Delta u_1(t, x)\|)^2 - \sum_{x=x_0}^{x_1-1} o_4(\|\Delta a(x)\| + \|\Delta u_2(t, x)\|)^2.
 \end{aligned} \quad (24)$$

Доказанная формула приращения второго порядка позволяет установить ряд конструктивно проверяемых условий оптимальности первого и второго порядков.

Легко видеть, что

$$\Delta z(t+1, x) = \sum_{\tau=t_0}^t [\Delta z(\tau+1, x) - \Delta z(\tau, x)] + \Delta z(t_0, x).$$

Поэтому с учетом формул (8), (9), применяя дискретный аналог теоремы Фубини, получаем, что

$$\Delta z(t+1, x) = \Delta a(x) + \sum_{\tau=t_0}^t \left[\sum_{s=\tau}^t (f(s, \tau, x, \bar{z}(\tau, x), \bar{u}_1(\tau)) - f(s, \tau, x, z(\tau, x), u_1(\tau))) - \Delta z(\tau, x) \right].$$

Переходя к норме в этом равенстве и используя правило треугольника, получаем

$$\begin{aligned}
 \|\Delta z(t+1, x)\| & \leq \|\Delta a(x)\| + L_1 \sum_{\tau=t_0}^t \left[\sum_{s=\tau}^t (\|\Delta z(\tau, x)\| + \|\Delta u_1(\tau)\|) + \|\Delta z(\tau, x)\| \right] \leq \\
 & \leq \|\Delta a(x)\| + L_2 \sum_{\tau=t_0}^t [\|\Delta z(\tau, x)\| + \|\Delta u_1(\tau)\|].
 \end{aligned} \quad (25)$$

Здесь $L_2 = \text{const} > 0$ – некоторая постоянная.

Далее, учитывая формулы (10) и (11), будем иметь

$$\Delta a(x+1) = \sum_{s=x_0}^x \left[\sum_{\alpha=x_0}^s [K(s, \alpha, \bar{a}(\alpha), \bar{u}_2(\alpha)) - K(s, \alpha, a(\alpha), u_2(\alpha))] - \Delta a(s) \right].$$

Отсюда на основе аналога теоремы Фубини получаем, что

$$\Delta a(x+1) = \sum_{s=x_0}^x \left[\sum_{\alpha=s}^x [K(\alpha, s, \bar{a}(s), \bar{u}_2(s)) - K(\alpha, s, a(s), u_2(s))] - \Delta a(s) \right].$$

Переходя в этой формуле к норме и используя условие Липшица, имеем

$$\|\Delta a(x+1)\| = L_3 \sum_{s=x_0}^x [\|\Delta a(s)\| + \|\Delta u_2(s)\|]. \quad (26)$$

Здесь $L_3 = \text{const} > 0$ – некоторая постоянная.

Применяя к неравенствам (25) и (26) дискретный аналог леммы Гронуолла (см., напр.: [6]), получим, что

$$\|\Delta z(t, x)\| \leq L_4 \left(\|\Delta a(x)\| + \sum_{\tau=t_0}^t \|\Delta u_1(\tau)\| \right), \quad (27)$$

$$\|\Delta a(x)\| = L_5 \sum_{s=x_0}^x \|\Delta u_2(s)\|. \quad (28)$$

Здесь L_4 и L_5 – некоторые положительные постоянные.

Далее из соотношений (8)–(11) получаем, что $\Delta z(t, x)$ и $\Delta a(x)$ являются решениями линеаризованных задач

$$\begin{aligned}
 & \Delta z(t+1, x) = \\
 & = \sum_{\tau=t_0}^t \left[\frac{\partial f(t, \tau, z(\tau, x), u_1(\tau))}{\partial z} \Delta z(\tau, x) + \frac{\partial f(t, \tau, z(\tau, x), u_1(\tau))}{\partial u_1} \Delta u_1(\tau) + o_5(\|\Delta z(\tau, x)\| + \|\Delta u_1(\tau)\|) \right],
 \end{aligned} \quad (29)$$

$$\Delta z(t_0, x) = \Delta a(x), \quad (30)$$

$$\Delta a(x+1) = \sum_{s=x_0}^x \left[\frac{\partial K(x, s, a(s), u_2(s))}{\partial a} \Delta a(s) + \frac{\partial K(x, s, a(s), u_2(s))}{\partial u_2} \Delta u_2(s) + o_6(\|\Delta a(s)\| + \|\Delta u_2(s)\|) \right], \quad (31)$$

$$\Delta a(x_0) = 0. \quad (32)$$

3. Специальные приращения управляющих функций и уравнения в вариациях

Пусть $\delta u_1(t) \in R^r, t \in T$ и $\delta u_2(x) \in R^q, x \in X \setminus x_1$ – две произвольные дискретные и ограниченные вектор-функции, а ε – достаточно малое по абсолютной величине число.

Тогда специальные приращения допустимых управлений $u_1(t)$ и $u_2(x)$ можно определить формулами

$$\begin{aligned} \Delta u_1(t; \varepsilon) &= \varepsilon \delta u_1(t), t \in T, \\ \Delta u_2(x; \varepsilon) &= \varepsilon \delta u_2(x), x \in X \setminus x_1. \end{aligned} \quad (33)$$

Через $(\Delta z(t, x; \varepsilon), \Delta a(x; \varepsilon))$ обозначим специальное приращение состояния $(z(t, x), a(x))$, отвечающее специальному приращению (33) управлений $u_1(t)$ и $u_2(x)$.

Из оценок (27) и (28) следует, что $\|\Delta a(x; \varepsilon)\|$ и $\|\Delta z(t, x; \varepsilon)\|$ имеют порядок малости ε .

Учитывая это и принимая во внимания формулу (33), с помощью формул (29)–(30), (31)–(32) доказывается следующее утверждение.

Лемма. При сделанных предположениях имеют место разложения

$$\Delta a(x; \varepsilon) = \varepsilon \delta a(x) + o(\varepsilon; x), \quad (34)$$

$$\Delta z(t, x; \varepsilon) = \varepsilon \delta z(t, x) + o(\varepsilon; t, x). \quad (35)$$

Здесь $\delta a(x)$ и $\delta z(t, x)$ являются решениями следующих задач Коши соответственно

$$\delta a(x+1) = \sum_{s=x_0}^x \left[\frac{\partial K(x, s, a(s), u_2(s))}{\partial a} \delta a(s) + \frac{\partial K(x, s, a(s), u_2(s))}{\partial u_2} \delta u_2(s) \right], \quad (36)$$

$$\delta a(x_0) = 0, \quad (37)$$

$$\delta z(t+1, x) = \sum_{\tau=t_0}^t \left[\frac{\partial f(t, \tau, z(\tau, x), u_1(\tau))}{\partial z} \delta z(\tau, x) + \frac{\partial f(t, \tau, z(\tau, x), u_1(\tau))}{\partial u_1} \delta u_1(\tau) \right], \quad (38)$$

$$\delta z(t_0, x) = \delta a(x). \quad (39)$$

Задачи (36)–(37) и (38)–(39) являются дискретными аналогами уравнений в вариациях (см., напр.: [13]).

4. Необходимые условия оптимальности

Учитывая формулы (33) и используя разложения (34), (35), из формулы приращения (24) получаем, что

$$\begin{aligned} J(u_1 + \varepsilon \delta u_1, u_2 + \varepsilon \delta u_2) - J(u_1, u_2) &= \frac{\varepsilon^2}{2} \sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) + \frac{\varepsilon^2}{2} \delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \delta a(x_1) - \\ &- \varepsilon \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \frac{\partial H'(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1} \delta u_1(t) - \varepsilon \sum_{x=x_0}^{x_1-1} \frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial u_2} \delta u_2(x) - \\ &- \frac{\varepsilon^2}{2} \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) + 2 \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \delta z(t, x) + \right. \end{aligned}$$

$$\begin{aligned}
 & +\delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \delta u_1(t) \Big] - \frac{\varepsilon^2}{2} \sum_{x=x_0}^{x_1-1} \left[\delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \delta a(x) + \right. \\
 & \left. + 2\delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \delta a(x) + \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \delta u_2(x) \right] + o(\varepsilon^2).
 \end{aligned}$$

Отсюда следует, что первая и вторая вариации функционала (7) имеют соответственно вид:

$$\begin{aligned}
 \delta^1 J(u_1, u_2, \delta u_1, \delta u_2) &= - \sum_{x=x_0}^{x_1-1} \frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial u_2} \delta u_2(x) - \\
 & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \frac{\partial H'(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1} \delta u_1(t), \\
 \delta^2 J(u_1, u_2, \delta u_1, \delta u_2) &= \sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) + \delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \delta a(x_1) - \\
 & - \sum_{x=x_0}^{x_1-1} \left[\delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \delta a(x) + \right. \\
 & \left. + 2\delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \delta a(x) + \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \delta u_2(x) \right] - \\
 & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) + \right. \\
 & \left. + 2\delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \delta z(t, x) + \right. \\
 & \left. + \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \delta u_1(t) \right].
 \end{aligned} \tag{40}$$

В силу открытости областей управления из формулы (40) следует, что вдоль оптимального управления $(u_1(t), u_2(x))$ для всех $\delta u_1(t)$ и $\delta u_2(x)$ соответственно выполняются соотношения

$$\begin{aligned}
 & \sum_{x=x_0}^{x_1-1} \frac{\partial M'(x, a(x), u_2(x), p(x))}{\partial u_2} \delta u_2(x) + \\
 & + \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \frac{\partial H'(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1} \delta u_1(t) = 0,
 \end{aligned} \tag{42}$$

Из формулы (42) в силу произвольности и независимости вариаций $\delta u_1(t)$ и $\delta u_2(x)$ управляющих функций $u_1(t)$ и $u_2(x)$ следует

Теорема 1. Для оптимальности допустимого управления $(u_1(t), u_2(x))$ необходимо, чтобы соотношения

$$\begin{aligned}
 \sum_{x=x_0}^{x_1-1} \frac{\partial H(\theta, x, z(\theta, x), u_1(\theta), \psi(\theta, x))}{\partial u_1} &= 0, \\
 \frac{\partial M(\xi, a(\xi), u_2(\xi), p(\xi))}{\partial u_2} &= 0
 \end{aligned} \tag{43}$$

выполнялись соответственно для всех $\theta \in T$ и $\xi \in X \setminus x_1$.

Это условие оптимальности (43) является аналогом уравнения Эйлера для рассматриваемой задачи. Каждое решение уравнения Эйлера назовем классической экстремалью.

В вариационном исчислении, учитывая слабость условия первого порядка – уравнения Эйлера, был установлен ряд необходимых условий оптимальности второго порядка, в частности условия оптимальности Лежандра–Клебша (см., напр.: [13]).

В задачах оптимального управления также прослеживается аналогичная ситуация.

Из формулы (41) следует, что для оптимальности классической экстремали $(u_1(t), u_2(x))$ необходимо, чтобы неравенство

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) + \delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \delta a(x_1) - \\ & - \sum_{x=x_0}^{x_1-1} \left[\delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \delta a(x) + \right. \\ & + 2\delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \delta a(x) + \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \delta u_2(x) \left. \right] - \\ & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) + \right. \\ & + 2\delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \delta z(t, x) + \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \delta u_1(t) \left. \right] \geq 0 \quad (44) \end{aligned}$$

выполнялось для всех вариаций $\delta u_1(t)$ и $\delta u_2(x)$ управляющих функций $u_1(t)$ и $u_2(x)$.

Неравенство (44) является неявным необходимым условием оптимальности второго порядка.

Возникает необходимость в получении необходимых условий оптимальности второго порядка, явно выраженных через параметры рассматриваемой задачи.

Пусть $F(t, \tau; x)$ и $\Phi(x, s)$ ($n \times n$) – матричные функции, являющиеся решениями матричных разностных уравнений типа Вольтерра

$$\begin{aligned} F(t, \tau-1; x) &= \sum_{s=\tau}^{t-1} F(t, s; x) \frac{\partial f(s, \tau, x, z(\tau, x), u_1(\tau))}{\partial z}, \\ F(t, t-1; x) &= E, \\ \Phi(x, s-1) &= \sum_{\alpha=s}^{x-1} \Phi(x, \alpha) \frac{\partial K(\alpha, s, a(s), u_2(s))}{\partial a}, \\ \Phi(x, x-1) &= E, \end{aligned}$$

где E – ($n \times n$) единичная матрица.

Из результатов работы [14] следует, что решения задач (37)–(38) и (36)–(37) могут быть представлены в виде:

$$\delta z(t, x) = F(t, t_0-1; x) \delta a(x) + \sum_{\tau=t_0}^{t-1} \sum_{s=\tau}^{t-1} F(t, s; x) \frac{\partial f(s, \tau, x, z(\tau, x), u_1(\tau))}{\partial u_1} \delta u_1(\tau), \quad (45)$$

$$\delta a(x) = \sum_{s=x_0}^{x-1} \sum_{\alpha=s}^{x-1} \Phi(x, \alpha) \frac{\partial K(\alpha, s, a(s), u_2(s))}{\partial u_2} \delta u_2(s). \quad (46)$$

Теперь, используя произвольность $\delta u_1(t)$ и $\delta u_2(x)$, положим $\delta u_2(x) = 0$.

Тогда представление (45) примет вид:

$$\delta z(t, x) = \sum_{\tau=t_0}^{t-1} \sum_{s=\tau}^{t-1} F(t, s; x) \frac{\partial f(s, \tau, x, z(\tau, x), u_1(\tau))}{\partial u_1} \delta u_1(\tau), \quad (47)$$

а из представления (46) при этом следует, что $\delta a(x) = 0$.

Поэтому из неравенства (44) получим, что вдоль оптимального управления $(u_1(t), u_2(x))$

$$\sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) + \right. \\ \left. + 2\delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \delta z(t, x) + \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \delta u_1(t) \right] \geq 0. \quad (48)$$

Введя обозначение

$$\sum_{s=\tau}^{t-1} F(t, s; x) \frac{\partial f(s, \tau, x, z(\tau, x), u_1(\tau))}{\partial u_1} = Q(t, \tau, x),$$

представление (47) можно записать в виде:

$$\delta z(t, x) = \sum_{\tau=t_0}^{t-1} Q(t, \tau, x) \delta u_1(\tau). \quad (49)$$

Используя представление (49), получим, что

$$\sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) = \\ = \sum_{\tau=t_0}^{t_1-1} \sum_{s=t_0}^{t_1-1} \delta u_1'(\tau) \left(\sum_{x=x_0}^{x_1-1} Q'(t_1, \tau, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} Q(t_1, s, x) \right) \delta u_1(s), \quad (50)$$

$$2 \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} \delta z(t, x) = \\ = 2 \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\sum_{\tau=t_0}^{t-1} \delta u_1'(\tau) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} Q(t, \tau, x) \delta u_1(\tau) \right], \quad (51)$$

$$\sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) = \\ = \sum_{\tau=t_0}^{t_1-1} \sum_{s=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta u_1'(\tau) \left(\sum_{t=\max(\tau, s)+1}^{t_1-1} Q'(t, \tau, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} Q(t, s, x) \right) \delta u_1(s). \quad (52)$$

Введем обозначение

$$K_1(\tau, s) = \sum_{x=x_0}^{x_1-1} \left(\sum_{t=\max(\tau, s)+1}^{t_1-1} Q'(t, \tau, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} Q(t, \tau, x) \right) - \\ - \sum_{x=x_0}^{x_1-1} Q'(t_1, \tau, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} Q(t_1, s, x). \quad (53)$$

С учетом обозначения (53) и тождеств (50)–(52) неравенство (44) записывается в виде:

$$\sum_{\tau=t_0}^{t_1-1} \sum_{s=t_0}^{t_1-1} \delta u_1'(\tau) K_1(\tau, s) \delta u_1(s) + 2 \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \left[\sum_{\tau=t_0}^{t-1} \delta u_1'(\tau) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1 \partial z} Q(t, \tau, x) \delta u_1(\tau) \right] + \\ + \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta u_1'(t) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} \delta u_1(t) \leq 0. \quad (54)$$

Теперь положим $\delta u_1(x) = 0$, считая, что $\delta u_2(x) \neq 0$. При этом формула представления (45) примет вид:

$$\delta z(t, x) = \sum_{s=x_0}^{x-1} \sum_{\alpha=s}^{x-1} F(t, t_0-1; x) \Phi(x, \alpha) \frac{\partial K(\alpha, s, a(s), u_2(s))}{\partial u_2} \delta u_2(s). \quad (55)$$

Введем обозначение

$$R(t, x, s) = \sum_{\alpha=s}^{x-1} F(t, t_0 - 1; x) \Phi(x, \alpha) \frac{\partial K(\alpha, s, a(s), u_2(s))}{\partial u_2},$$

из формулы (55) получим, что

$$\delta z(t, x) = \sum_{s=x_0}^{x-1} R(t, x, s) \delta u_2(s). \quad (56)$$

Далее, полагая

$$N(x, s) = \sum_{\alpha=s}^{x-1} \Phi(x, \alpha) \frac{\partial K(\alpha, s, a(s), u_2(s))}{\partial u_2},$$

представление (46) запишем в виде:

$$\delta a(x) = \sum_{s=x_0}^{x-1} N(x, s) \delta u_2(s). \quad (57)$$

При предположении $\delta u_1(x) = 0$, $\delta u_2(x) \neq 0$ неравенство (44) примет вид:

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) + \delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \delta a(x_1) - \\ & - \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) - \\ & - \sum_{x=x_0}^{x_1-1} \left[\delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \delta a(x) + \right. \\ & \left. + 2\delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \delta a(x) + \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \delta u_2(x) \right] \geq 0. \quad (58) \end{aligned}$$

Займемся преобразованием отдельных слагаемых в неравенстве (58). С использованием формул (56) и (57) доказывается, что

$$\delta a'(x_1) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} \delta a(x_1) = \sum_{\tau=x_0}^{x-1} \sum_{s=x_0}^{x-1} \delta u_2'(\tau) N'(x_1, \tau) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} N'(x_1, s) \delta u_2(s), \quad (59)$$

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} \delta z'(t_1, x) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} \delta z(t_1, x) = \\ & = \sum_{\tau=x_0}^{x-1} \sum_{s=x_0}^{x-1} \delta u_2'(\tau) \left(\sum_{x=\max(\tau, s)+1}^{x_1-1} R'(t_1, x, \tau) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} R(t_1, x, s) \right) \delta u_2(s), \quad (60) \end{aligned}$$

$$\begin{aligned} & \sum_{t=t_0}^{t_1-1} \sum_{x=x_0}^{x_1-1} \delta z'(t, x) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} \delta z(t, x) = \\ & \sum_{\tau=x_0}^{x_1-1} \sum_{s=x_0}^{x_1-1} \delta u_2'(\tau) \left[\sum_{t=t_0}^{t_1-1} \sum_{x=\max(\tau, s)+1}^{x_1-1} R'(t, x, \tau) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} R(t, x, s) \right] \delta u_2(s), \quad (61) \end{aligned}$$

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} \delta a'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} \delta a(x) = \\ & = \sum_{\tau=x_0}^{x_1-1} \sum_{s=x_0}^{x-1} \delta u_2'(\tau) \left[\sum_{x=\max(\tau, s)+1}^{x_1-1} N'(x_1, \tau) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} N(x_1, s) \right] \delta u_2(s), \quad (62) \end{aligned}$$

$$\begin{aligned} & \sum_{x=x_0}^{x_1-1} \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} \delta a(x) = \\ & = \sum_{x=x_0}^{x_1-1} \left[\sum_{s=x_0}^{x-1} \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} N(x, s) \delta u_2(s) \right]. \end{aligned} \quad (63)$$

Введя обозначение

$$\begin{aligned} K_2(\tau, s) = & -N'(x_1, \tau) \frac{\partial^2 \varphi_2(a(x_1))}{\partial a^2} N'(x_1, s) - \sum_{x=\max(\tau, s)+1}^{x_1-1} R'(t_1, x, \tau) \frac{\partial^2 \varphi_1(x, z(t_1, x))}{\partial z^2} R(t_1, x, s) + \\ & + \sum_{t=t_0}^{t_1-1} \sum_{x=\max(\tau, s)+1}^{x-1} R'(t, x, \tau) \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial z^2} R(t, x, s) + \\ & + \sum_{x=\max(\tau, s)+1}^{x_1-1} N'(x_1, \tau) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial a^2} N(x_1, s) \end{aligned}$$

и учитывая тождества (59)–(63), из неравенства (58) получаем, что

$$\begin{aligned} & \sum_{\tau=x_0}^{x_1-1} \sum_{s=x_0}^{x_1-1} \delta u_2'(\tau) K_2(\tau, s) \delta u_2(s) + 2 \sum_{x=x_0}^{x_1-1} \left[\sum_{s=x_0}^{x-1} \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2 \partial a} N(x, s) \delta u_2(s) \right] + \\ & + \sum_{x=x_0}^{x_1-1} \delta u_2'(x) \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} \delta u_2(x) \leq 0. \end{aligned} \quad (64)$$

Следовательно, доказана

Теорема 2. Для оптимальности классической экстремали $(u_1(t), u_2(x))$ необходимо, чтобы неравенства (54) и (64) выполнялись для всех допустимых вариаций $\delta u_1(t)$ и $\delta u_2(x)$ соответственно.

Замечание. Результат теоремы 2 является довольно общим. Используя произвольность допустимых вариаций $\delta u_1(t)$ и $\delta u_2(x)$, из неравенств (54) и (64) можно получить ряд относительно легко проверяемых необходимых условий оптимальности. Но они будут менее информативными.

Приведем одно из них.

Теорема 3. Для оптимальности классической экстремали $(u_1(t), u_2(x))$ необходимо, чтобы неравенства

$$\begin{aligned} & v_1' K_1(\theta, \theta) v_1 + v_1' \sum_{x=x_0}^{x_1-1} \frac{\partial^2 H(t, x, z(t, x), u_1(t), \psi(t, x))}{\partial u_1^2} v_1 \leq 0, \\ & v_2' K_2(\xi, \xi) v_2 + v_2' \sum_{x=x_0}^{x_1-1} \frac{\partial^2 M(x, a(x), u_2(x), p(x))}{\partial u_2^2} v_2 \leq 0 \end{aligned}$$

выполнялись соответственно для всех $v_1 \in R^r$ и $v_2 \in R^q$.

Заключение

В работе исследуется одна дискретная задача оптимального управления, описываемая системой двумерных и одномерных разностных уравнений типа Вольтерра.

Процесс описывается с двумя группами управляющих функций. Управляющая функция из первого класса входит в правую часть уравнения, а управляющая функция из второй группы входит в начальное условие уравнения. При этом начальная функция, являясь управляемой, определяется как решение одномерного разностного уравнения Вольтерра.

Получены аналог уравнения Эйлера и ряд необходимых условий оптимальности второго порядка.

Список источников

1. Болтянский В.Г. Оптимальное управление дискретными системами. М. : Наука, 1972. 446 с.
2. Мансимов К.Б. Дискретные системы. Баку : Изд-во БГУ, 2013. 151 с.
3. Марданов М.Дж., Меликов Т.К., Малик С.Т. К теории оптимальных процессов в дискретных системах // Математические заметки. 2019. Т. 106, вып. 3. С. 408–413.
4. Марданов М.Дж., Меликов Т.К., Малик С.Т. К необходимым условиям оптимальности в дискретных системах управления // Известия Иркутского университета. Сер. математическая. 2020. Т. 31. С. 49–61. doi: 10.26516/1997-7670.2020.31.49
5. Габасов Р., Кириллова Ф.М., Альсевич В.В. и др. Методы оптимизации. Минск : Четыре четверти, 2011. 472 с.
6. Васильев Ф.П. Методы решения экстремальных задач. М. : Наука, 1981. 400 с.
7. Габасов Р., Кириллова Ф.М. К теории необходимых условий оптимальности в дискретных системах управления // Управляемые системы : сб. тр. Новосибирск : Ин-т математики СО АН СССР, 1979. Вып. 18. С. 14–25.
8. Гараева Э.А., Мансимов К.Б. Об одной дискретной задаче оптимального управления // Вестник Бакинского университета. Сер. Физика, математика. 2014. № 1. С. 40–49.
9. Гараева Э.А. Об одной дискретной линейной задаче оптимального управления // Вестник Бакинского университета. Сер. Физика, математика. 2015. № 1. С. 70–76.
10. Москаленко А.И. Об одном классе задач оптимального регулирования // Журнал вычислительной математики и математической физики. 1969. № 1. С. 69–95.
11. Москаленко А.И. Некоторые вопросы теории оптимального управления : автореф. дис. ... канд. физ.-мат. наук. Томск, 1971. 20 с.
12. Souyousefain M., Leela S. Stability results for difference equations of Volterra type // Appl. Math. Comput. 1990. V. 36 (1). P. 51–61.
13. Габасов Р., Кириллова Ф.М. Особые оптимальные управления. М. : Либроком, 2011. 256 с.
14. Керимова А.В. О представлении решения одного класса систем линейных неоднородных разностных уравнений типа Вольтерра // Журнал Бакинского инженерного университета. Сер. Математика и компьютерные науки. 2023. № 2. С. 30–36.

References

1. Boltyansky, V.G. (1972) *Optimal'noe upravlenie diskretnymi sistemami* [Optimal control of discrete systems]. Moscow: Nauka.
2. Mansimov, K.B. (2013) *Diskretnye sistemy* [Discrete systems]. Baku: BSU.
3. Mardanov, M.J., Melikov, T.K. & Malik, S.T. (2019) K teorii optimal'nykh protsessov v diskretnykh sistemakh [On the theory of optimal processes in discrete systems]. *Matematicheskie zametki*. 106(3). pp. 408–413.
4. Mardanov, M.J., Melikov, T.K. & Malik, S.T. (2020) K neobkhodimym usloviyam optimal'nosti v diskretnykh sistemakh upravleniya [On necessary conditions for optimality in discrete control systems]. *Izvestiya Irkutskogo universitetata. Ser. matematicheskaya*. 31. pp. 49–61. DOI: 10.26516/1997-7670.2020.31.49
5. Gabasov, R., Kirillova, F.M., Alseovich, V.V. et al. (2011) *Metody optimizatsii* [Optimization Methods]. Minsk: Chetyre chetverti.
6. Vasiliev, F.P. (1981) *Metody resheniya ekstremal'nykh zadach* [Methods for Solving Extremal Problems]. Moscow: Nauka.
7. Gabasov, R. & Kirillova, F.M. (1979) K teorii neobkhodimyykh usloviy optimal'nosti v diskretnykh sistemakh upravleniya [On the theory of necessary conditions for optimality in discrete control systems]. In: *Upravlyaemye sistemy* [Controlled Systems]. Vol. 18. Novosibirsk: SB RAS. pp. 14–25.
8. Garaeva, E.A. & Mansimov, K.B. (2014) Ob odnoy diskretnoy zadache optimal'nogo upravleniya [On one discrete optimal control problem]. *Vestnik Bakinskogo universiteta. Ser. Fizika, matematika*. 1. pp. 40–49.
9. Garaeva, E.A. (2015) Ob odnoy diskretnoy lineynoy zadache optimal'nogo upravleniya [On one discrete linear optimal control problem]. *Vestnik Bakinskogo universiteta. Ser. Fizika, matematika*. 1. pp. 70–76.
10. Moskalenko, A.I. (1969) Ob odnom klasse zadach optimal'nogo regulirovaniya [On one class of optimal control problems]. *Zhurnal vychislitel'noy matematiki i matematicheskoy fiziki*. 1. pp. 69–95.
11. Moskalenko, A.I. (1971) *Nekotorye voprosy teorii optimal'nogo upravleniya* [Some issues in optimal control theory]. Abstract of Physics and Mathematics Cand. Diss. Tomsk.
12. Souyousefain, M. & Leela, S. (1990) Stability results for difference equations of Volterra type. *Applied Mathematics and Computation*. 36(1). pp. 51–61.
13. Gabasov, R. & Kirillova, F.M. (2011) *Osobyie optimal'nye upravleniya* [Special Optimal Controls]. Moscow: Librokom.
14. Kerimova, A.V. (2023) O predstavlenii resheniya odnogo klassa sistem lineynykh neodnorodnykh raznostnykh uravneniy tipa Vol'terra [On the representation of the solution of one class of systems of linear inhomogeneous difference equations of Volterra type]. *Zhurnal Bakinskogo inzhener'nogo universiteta. Ser. Matematika i komp'yuternye nauki*. 2. pp. 30–36.

Информация об авторе:

Мансимов Камиль Байрамали оглы – профессор, доктор физико-математических наук, заведующий лабораторией «Методы управления в сложных динамических системах» Института систем управления Министерства науки и образования

Азербайджана (Баку, Азербайджан); заведующий кафедрой «Математическая кибернетика» Бакинского государственного университета (Баку, Азербайджан), E-mail: kamilbmansimov@gmail.com

Автор заявляет об отсутствии конфликта интересов.

Information about the authorss:

Mansimov Kamil B. (Professor, Doctor of Physical and Mathematical Sciences, Head of the Laboratory “Control Methods in Complex Dynamic Systems” of the Institute of Control Systems of the Ministry of Science and Education of Azerbaijan, Baku, Azerbaijan); Head Department of Mathematical Cybernetics, Baku State University (Baku, Azerbaijan), E-mail: kamilbmansimov@gmail.com

The author declares no conflicts of interests.

Поступила в редакцию 30.05.2024; принята к публикации 03.03.2025

Received 30.05.2024; accepted for publication 03.03.2025

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELING

Научная статья

УДК 519.6

doi: 10.17223/19988605/70/3

Параллельная реализация лагранжевой дисперсионной
стохастической модели переноса примеси в атмосфереЕкатерина Алексеевна Каратаева¹, Александр Васильевич Старченко²^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Россия*^{1,2} *Институт оптики атмосферы им. В.Е. Зуева Сибирского отделения РАН, Томск, Россия*¹ *starch@math.tsu.ru*² *karat@iao.ru*

Аннотация. Для лагранжевой дисперсионной стохастической модели переноса примеси разработаны параллельные алгоритмы ее численной реализации для многопроцессорных многоядерных вычислительных систем с общей и распределенной памятью. Для создания параллельных версий программ использовались технологии параллельного программирования Message Passing Interface и Open MultiProcessing. Получено, что использование технологии MPI в равных условиях имеет небольшое преимущество и позволяет достичь ускорения параллельной версии программы при запуске до 1 000 000 частиц более чем в 20 раз.

Ключевые слова: лагранжева дисперсионная модель, перенос примеси, параллельная реализация, MPI, OpenMP.

Благодарности: Работа выполнена при поддержке Министерства науки и высшего образования РФ (соглашение № 075-02-2024-1437).

Для цитирования: Каратаева Е.А., Старченко А.В. Параллельная реализация лагранжевой дисперсионной стохастической модели переноса примеси в атмосфере // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 29–41. doi: 10.17223/19988605/70/3

Original article

doi: 10.17223/19988605/70/3

Parallel implementation of Lagrangian stochastic dispersion model
of mixture transportEkaterina A. Karataeva¹, Alexander V. Starchenko¹^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*^{1,2} *V.E. Zuev Institute of Atmospheric Optics of Siberian Branch of the Russian Academy of Science, Tomsk, Russian Federation*¹ *starch@math.tsu.ru*² *karat@iao.ru*

Abstract. Parallel algorithms for its numerical implementation for multiprocessor multicore computing systems with shared and distributed memory have been developed for the Lagrangian stochastic dispersion model. To create parallel versions of programs, the Message Passing Interface and Open MultiProcessing parallel programming technologies were used. It is found that MPI technology has a slight advantage and allows to accelerate the parallel version of the program when running up to 1000 000 particles by more than 20 times.

Keywords: Lagrangian stochastic dispersion model, mixture transport, parallel implementation, MPI, OpenMP.

Acknowledgments: The study was supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-02-2024-1437).

For citation: Karataeva, E.A., Starchenko, A.V. (2025) Parallel implementation of Lagrangian stochastic dispersion model of mixture transport. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 29–41. doi: 10.17223/19988605/70/3

Введение

Выбросы загрязняющих примесей в атмосферу от предприятий ведут к ухудшению экологической обстановки города, негативно влияют на здоровье его жителей, способствуют обострению имеющихся заболеваний и развитию новых. Их контроль – социально значимая задача. При наличии реальных, а когда это невозможно, расчетных данных можно давать рекомендации населению и учреждениям по профилактике и защите от вредного воздействия загрязненной атмосферы [1].

В городах и больших населенных пунктах к основным источникам загрязнений относится автотранспорт, работающий на углеводородном топливе, а также трубы промышленных предприятий и теплоэнергетических комплексов, причем последние могут играть существенную роль в загрязнении городского воздуха в условиях температурной инверсии [2, 3].

В настоящее время при численном моделировании переноса примеси в атмосфере используются эйлерова модель [4], которая содержит нестационарные дифференциальные адвективно-диффузионные уравнения для компонент примеси, и лагранжева модель [4], в которой перенос примеси представляется за счет рассмотрения большого количества траекторий частиц, увлекаемых движущимся турбулентным потоком воздуха. Для численной реализации эйлеровой модели переноса примеси применяются сеточные конечно-разностные методы, и точность численного прогноза по такой модели зависит от порядка аппроксимации разностной схемы и величины шага сетки, размер которого может меняться до нескольких десятков или сотен метров. Чем меньше шаг используемой расчетной сетки, тем более вычислительно трудоемким будет получение прогнозируемых распределений компонентов примеси, даже с использованием многопроцессорной вычислительной техники. В отличие от эйлеровой модели лагранжев подход позволяет детально описывать начальный этап поступления примеси из труб в атмосферный воздух, точность численного прогноза не зависит от шага сетки, а определяется представительными статистиками при достаточно большом количестве рассматриваемых стохастически блуждающих в воздушном потоке частиц. При численной реализации лагранжевой модели обычно используются методы численного интегрирования обыкновенных дифференциальных уравнений первого, реже второго порядка, которые можно решать одновременно для множества частиц, поскольку предполагается, что движение одной частицы не оказывает влияния на движение другой. Этот факт создает большие перспективы для параллельной реализации лагранжевых дисперсионных стохастических моделей.

В качестве источника загрязнения в данной работе рассматривается высотная труба, из которой с постоянным расходом выделяется дым, содержащий инертные частицы, способные перемещаться в атмосфере вместе с ветром, называемые примесью. Движение этих частиц описывается с помощью подхода Лагранжа, при котором перенос каждой частицы рассматривается отдельно и независимо от других. Наблюдение за частицей и фиксация ее положения прекращаются, как только она покидает рассматриваемую ограниченную область движения. К первостепенным параметрам модели относятся количество выброшенных из источника частиц, массовый или объемный расход примеси, который несет в себе каждая из них, высота источника выброса, скорость и турбулентность воздушного потока. Применяемая для расчетов модель имеет название лагранжевой дисперсионной стохастической модели. Стохастичность отражена при расчете передвижения частицы в турбулентном потоке: три компоненты актуальной скорости рассчитываются с применением датчика псевдослучайных чисел.

В настоящее время для ускорения получения результатов численного моделирования широко применяют массивные многопроцессорные вычислительные системы и рабочие станции-серверы, собранные на большом количестве связанных коммуникационной средой вычислительных узлов с многоядерными центральными (CPU) и графическими (GPU) процессорами [5]. В зависимости от количества используемых в расчетах вычислительных узлов среда для параллельных вычислений может рассматриваться как система с распределенной (несколько вычислительных узлов) или общей оперативной памятью (один вычислительный узел). В зависимости от архитектуры многопроцессорной вычислительной системы применяются различные технологии параллельного программирования [6]: библиотека Message Passing Interface – для систем с CPU и распределенной и общей памятью, компиляторы с Open Multi Processing – для систем с общей памятью, компиляторы с CUDA и OpenACC – для программирования графических процессоров. Возможно совместное применение различных технологий, позволяющее более полно использовать имеющиеся в распоряжении вычислительные ресурсы современных многопроцессорных вычислительных систем.

Распараллеливанию лагранжевых дисперсионных стохастических атмосферных моделей (ЛДСМ) переноса примеси также посвящен ряд статей в литературе. В [7] сообщается об успешной разработке в Национальном консультативном центре Министерства энергетики США параллельной версии модели атмосферной дисперсии LODI (Lagrangian Operational Dispersion Integrator), предназначенной для оперативного реагирования на чрезвычайные ситуации по выбросам в атмосферу. LODI моделирует трехмерный адвективно-диффузионный перенос примеси в рамках лагранжева стохастического метода, включающего метод Монте-Карло. Параллельная версия модели для мультимониторных систем с общей памятью реализована с помощью компиляторных директив технологии OpenMP. Для кластеров с распределенной памятью параллельная версия LODI получена на основе библиотеки интерфейса передачи сообщений MPI. Также рассматривалось совместное применение технологий MPI + OpenMP. Было получено, что параллельная версия LODI-MPI обеспечивает близкое к линейному снижению времени счета при увеличении числа используемых центральных процессоров. При снижении количества запускаемых в LODI частиц в расчете на один процессор производительность параллельной программы снижается. Параллельная версия LODI-OpenMP позволяет получить ускорение вычислений в сравнении с последовательным вариантом только в 1,4 раза на компьютере с общей памятью и четырьмя процессорами. Однако было отмечено, что при определенных условиях работы можно ожидать ускорения в 2,75 раза. Для технологии MPI этот показатель равен 3,9. Результаты достигнуты при запуске одного миллиона частиц.

В работе [8] представлены результаты для параллельной версии лагранжевой модели переноса и дисперсии частиц FLEXPART (FLEXible PARTicle), которая в первоначальном варианте разрабатывалась для расчета рассеяния опасных веществ из точечных источников, например выброшенных после аварии на атомной электростанции, в глобальном или локальном масштабе. Сегодня FLEXPART позволяет выполнять многомасштабное моделирование и анализ переноса примесей в атмосфере. Область ее применения распространилась на численное прогнозирование переноса различных компонентов атмосферных газов и аэрозолей, вулканических выбросов, микрофизики влаги в атмосфере, углеродного цикла. FLEXPART позволяет проводить и обратное моделирование, определяя по наблюдениям источники опасных выбросов в атмосферу. Вычислительный код модели был распараллелен с использованием библиотеки интерфейса передачи сообщений MPI для вычислительных систем с распределенной оперативной памятью. Было получено, что разработанная параллельная версия хорошо масштабируется вплоть до 256 процессоров, причем эффективность параллельной программы имеет значения не ниже 75% вплоть до использования 64 процессоров из нескольких вычислительных узлов. Снижение эффективности параллельного кода обусловлено наличием в программе участков кода, выполняемых последовательно.

В работе [9] рассмотрена параллельная версия лагранжевой дисперсионной модели LAMBDA, разработанной для изучения процесса переноса и рассеивания загрязняющих веществ над равнинной местностью. Авторы отмечают высокие вычислительные затраты при реализации такого сорта моделей и указывают на необходимость привлечения в расчетах многопроцессорной вычислительной

техники. В качестве технологии распараллеливания модели LAMBDA была выбрана технология передачи сообщений MPI. Рассматривались два подхода организации параллельных вычислений: распределение области исследования по активным процессам или распределение частиц по активным процессам. Первый подход показал большее нарушение балансировки загрузки между активными процессами, поэтому был выбран второй подход. В результате параллельная MPI-версия кода LAMBDA позволила получить ускорение до 9,6 раз на 10 процессорах при численной реализации траекторий 2 000 или 10 000 частиц.

В [10] представлены результаты распараллеливания кода лагранжевой дисперсионной стохастической модели PALMTREE. Используется технология MPI, и основное внимание уделяется генерации псевдослучайных чисел в параллельных потоках. Рассмотрены две объектно-ориентированные стратегии: вычисление псевдослучайных чисел «на лету» в рамках каждого процесса MPI или использование разных генераторов псевдослучайных чисел для каждой моделируемой траектории частиц. Показано преимущество второй стратегии, которая и была реализована в параллельной версии модели PALMTREE. При проведении вычислительных экспериментов достигнуты значения эффективности параллельной программы больше 70% при использовании до 120 процессоров.

В [11] рассмотрено применение технологии OpenACC при разработке параллельной версии ЛДСМ MPTRAC (Massive-Parallel Trajectory Calculations) для использования при вычислениях на графических картах. Все вычисления, за исключением получения метеорологических данных и записи результатов расчетов, были портированы на графические карты. Верификация параллельного кода и анализ его производительности в случае гибридного подхода (MPI + OpenMP + OpenACC) выполнены в Юлихском суперкомпьютерном центре Германии. Для параллельных расчетов над протяженной областью использовалось 10^8 частиц. Результаты расчетов показали, что добавление графических процессоров NVIDIA позволяет до 16 раз увеличить быстродействие вычислений по сравнению с использованием только центральных процессоров.

В заключение обзора отметим, что согласно публикации [12], в которой сообщаются результаты эффективности распараллеливания программы численного решения неоднородного нестационарного трехмерного адвективно-диффузионного уравнения (эйлерова модель переноса примеси) с помощью метода конечного объема и явно-неявных разностных схем на сетке $256 \times 256 \times 32$, возможно получение ускорения в 90 раз на 128 процессорах при использовании технологии MPI, в 10 раз на 12 процессорах при использовании OpenMP и в 30 раз на графической карте NVIDIA Tesla K80.

Таким образом, обзор литературы показывает, что наиболее используемыми технологиями разработки параллельных версий кода для лагранжевых дисперсионных стохастических моделей являются библиотека передачи сообщений MPI (для масштабных по размеру области и количеству запускаемых частиц вычислений) и компиляторы с директивами OpenMP. Возможно применение графических процессоров NVIDIA с технологией OpenACC для проведения исследований переноса примеси над небольшой по размеру областью. Ожидаемые результаты ускорения и эффективности для расчетов по модели, представленной в этой статье, не должны уступать приведенным выше.

Цель данной работы – разработка параллельных версий развиваемой лагранжевой дисперсионной стохастической модели переноса примеси [3] для многопроцессорных вычислительных систем с общей памятью (OpenMP) и распределенной памятью (MPI) и демонстрация их эффективности на задаче адвективно-диффузионного переноса примеси от высотного источника в турбулентном конвективном пограничном слое атмосферы.

1. Постановка задачи

1.1. Физическая постановка задачи

Рассматривается перенос примеси, плотность которой совпадает с плотностью воздуха, от одного или нескольких высотных точечных источников постоянной мощности, расположенных в турбулентном атмосферном пограничном слое. Известны скорость и направление ветра, параметры турбулентности, высота источников и координаты их расположения на поверхности.

1.2. Математическая постановка задачи

Рассматриваемая область пространства ограничивается некоторым объемом $[0, Lx] \times [0, Ly] \times [0, Lz]$, внутри которого располагается высотный точечный источник (рис. 1). Процесс распространения примеси от источника рассматривается с помощью Лагранжева подхода [4], в соответствии с которым движение примеси в атмосфере моделируется как последовательность запусков отдельных частиц из источников поступления примеси. В таком случае начальное положение каждой частицы загрязняющего вещества совпадает с устьем трубы:

$$x(t=0) = x_{source}, \quad y(t=0) = y_{source}, \quad z(t=0) = z_{source}.$$

Траектория движения каждой частицы вычисляется по формулам

$$\begin{aligned} x(t + \Delta t) &= x(t) + v_x(x(t), y(t), z(t), t) \Delta t, \\ y(t + \Delta t) &= y(t) + v_y(x(t), y(t), z(t), t) \Delta t, \\ z(t + \Delta t) &= z(t) + v_z(x(t), y(t), z(t), t) \Delta t, \end{aligned} \quad (1)$$

где (v_x, v_y, v_z) – вектор актуальной скорости ветра, включающий турбулентные пульсации, реализуемые статистически [4], Δt – шаг по времени. При движении частица может отражаться от верхней и нижней границ, что представляется путем замены знака у величины вертикальной компоненты скорости. Каждая частица несет в себе определенное количество вещества, которое суммируется для ячейки в зависимости от времени пребывания в ней частиц.

Актуальные компоненты скорости ветра (v_x, v_y, v_z) представляются следующим образом:

$$\begin{aligned} v_x &= \bar{v}_x + v'_x + v_{xp}, \\ v_y &= \bar{v}_y + v'_y + v_{yp}, \\ v_z &= \bar{v}_z + v'_z + v_{zp} \end{aligned} \quad (2)$$

где $\bar{v}_x, \bar{v}_y, \bar{v}_z$ – компоненты скорости ветра, рассчитываемые по модели численного прогноза погоды; v'_x, v'_y, v'_z – турбулентные пульсации компонент скорости; v_{xp}, v_{yp}, v_{zp} – возмущение скорости вследствие дополнительного подъема нагретой примеси, имеющей вертикальную скорость.

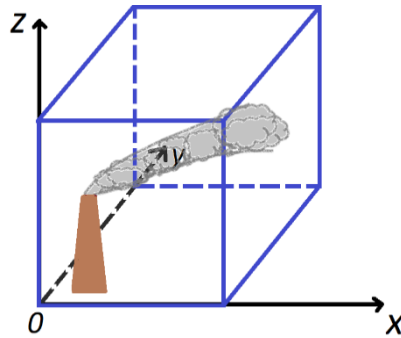


Рис. 1. Область исследования с высотным точечным источником
Fig. 1. Research area with a high-altitude point source

Для расчета турбулентных пульсаций в данной работе используются следующие соотношения:

$$v'_x = \sqrt{\frac{2}{3} k \xi_x}, \quad v'_y = \sqrt{\frac{2}{3} k \xi_y}, \quad (3)$$

где k – кинетическая энергия турбулентности, ξ_x, ξ_y – псевдослучайные числа из нормального (гаусового) распределения, имеющие нулевое среднее и дисперсию единицу, v'_z зависит от стратификации атмосферы [4, 13]:

$$dv'_z = (a_0 + a_1 v'_z + a_2 v'^2_z) dt + b_0 \xi, \quad (4)$$

где ξ – псевдослучайное число, $b_0 = \sqrt{C_0 \varepsilon \Delta t}$, где ε – диссипация энергии турбулентности k , $C_0 = 2$ – константа [14].

Формулы для коэффициентов в (4) можно найти в [14]:

$$a_2 = \frac{\frac{1}{3} \left(\frac{\partial \overline{w'^3}}{\partial t} + \frac{\partial \overline{w'^4}}{\partial z} \right) - \frac{\overline{w'^3}}{2\overline{w'^2}} \left(\frac{\partial \overline{w'^2}}{\partial t} + \frac{\partial \overline{w'^3}}{\partial z} - C_0 \varepsilon \right) - \overline{w'^2} \frac{\partial \overline{w'^2}}{\partial z}}{\overline{w'^4} - (\overline{w'^3})^2 / \overline{w'^2} - (\overline{w'^2})^2},$$

$$a_1 = \frac{1}{2\overline{w'^2}} \left(\frac{\partial \overline{w'^2}}{\partial t} + \frac{\partial \overline{w'^3}}{\partial z} - 2\overline{w'^3} a_2 - C_0 \varepsilon \right), \quad a_0 = \frac{\partial \overline{w'^2}}{\partial z} - \overline{w'^2} a_2.$$

Дисперсия вертикальной компоненты скорости $\overline{w'^2}$ может быть получена из модифицированного прогностического уравнения для $\overline{w'^2}$ в равновесном приближении, когда всеми членами адвекции и диффузии пренебрегают и делается предположение о квазиоднородном пограничном слое:

$$\overline{w'^2} = \left[\frac{2}{3} k + \frac{k}{c_{s1} \varepsilon} \left(\left(2 - c_{s2} - c_{w2} \frac{l}{kz} \right) P_s + \left(2 - c_{s3} - c_{w3} \frac{l}{kz} \right) P_b - \frac{2\varepsilon}{3} \right) \right] \left(1 + \frac{c_{w1}}{c_{s1}} \frac{l}{kz} \right)^{-1},$$

$$c_{s1} = 2,2, \quad c_{s2} = 1,63, \quad c_{s3} = 0,73, \quad c_{w1} = 1, \quad c_{w2} = 0,24, \quad c_{w3} = 0,$$

где l – интегральный масштаб турбулентности, выражения для P_s и P_b получаются из сдвига и плавучести [14].

Моменты более высокого порядка дисперсии вертикальной скорости для неустойчивой стратификации атмосферного пограничного слоя могут быть получены из дисперсии второго порядка $\overline{w'^2}$ [14]:

$$\overline{w'^3} = 0,4(\overline{w'^2})^{3/2}, \quad \overline{w'^4} = 3,5(\overline{w'^2})^2. \quad (5)$$

Для нейтральной и устойчивой стратификации атмосферы используются $\overline{w'^3} = 0$, $\overline{w'^4} = 3(\overline{w'^2})^2$ [14].

На рис. 2 приведена область исследования с высотным точечным источником. Для реализации ЛДСМ для реальных условий требуются входные параметры атмосферы, такие как скорость ветра, его направление, кинетическая энергия турбулентности, диссипация энергии турбулентности, интегральный масштаб турбулентности, температура над центром города. Для этих целей в данной работе используется мезомасштабная метеорологическая модель TSUNM3 [15], позволяющая получать необходимые метеорологические поля с периодичностью от 10 мин до 1 ч на сутки вперед.

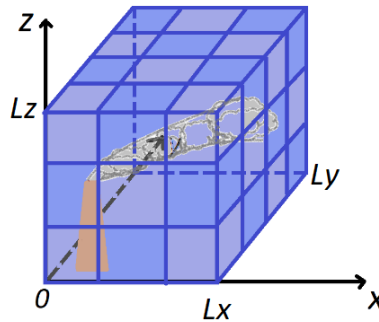


Рис. 2. Область исследования с высотным точечным источником, заключенная в трехмерную конечно-объемную сетку
Fig. 2. Research area with a high-altitude point source enclosed in a three-dimensional finite-volume grid

Получение распределения концентрации примеси осуществляется на трехмерной конечно-объемной сетке ω_h с конечными объемами (рис. 2), когда вся область делится на трехмерные подобласти, в каждой из которых фиксируются наличие и время нахождения частицы:

$$C_{i,j,k} = \sum_p \frac{Mist \cdot \Delta t}{NP \cdot Vol_{i,j,k}}, \quad (6)$$

где NP – число частиц, $Vol_{i,j,k}$ – конечный объем, через который пролетает p -я частица. Затем для каждой ячейки суммируются результаты по всем частицам, и получаются значения концентрации в каждой ячейке всей области.

Для генерации псевдослучайных чисел, необходимых для моделирования турбулентных пульсаций скорости, требуется использовать датчик случайных чисел. В данной работе для этих целей применялся модуль random C++, из которого был выбран датчик псевдослучайных чисел mt19937.

2. Выбор технологии распараллеливания

Блок-схему программы, реализующей рассмотренную выше ЛДСМ, можно представить следующим образом (рис. 3).

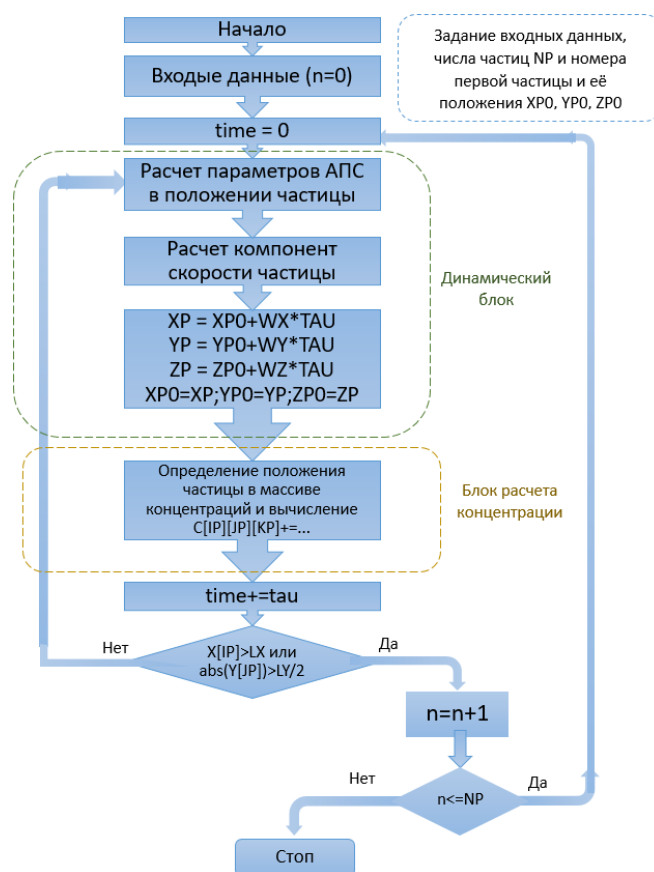


Рис. 3. Блок-схема организации расчетов в ЛДСМ
Fig. 3. Flowchart of the organization of calculations in the LDSM

Из блок-схемы видно, что расчет траекторий частиц может выполняться одновременно и независимо, поэтому представленный на рис. 3 алгоритм обладает высокой степенью параллелизма, которая в рассматриваемом случае равна количеству реализуемых в ЛДСМ частиц.

В данной работе получение параллельных версий программы для ЛДСМ осуществлялось с помощью двух технологий параллельного программирования – библиотеки интерфейса передачи сообщений MPI [16] и стандарта OpenMP [16].

Библиотека MPI предназначена для создания параллельных приложений для многопроцессорных вычислительных систем с распределенной памятью, но с успехом может быть применена и для многоядерных компьютеров с общей памятью. Интерфейс передачи сообщений MPI включает более сотни функций, позволяющих передавать информацию (данные) между параллельно выполняемыми процессами. Библиотека MPI используется при разработке параллельных программ на языках C/C++ и Fortran, при этом, однако, приходится существенно видоизменять имеющийся изначально последо-

вательный код и разработку параллельной версии вести со специально предназначенной для этого программой.

Что касается рассмотренной выше последовательной версии ЛДСМ, то основной способ ее распараллеливания заключается в распределении всего множества реализуемых частиц между активными параллельными процессами, расчетом концентрации частиц каждым параллельным процессом и последующей редукции с суммированием распределенных по параллельным процессам полей концентрации на главном процессе. Этот подход может быть представлен в следующем фрагменте кода MPI-программы на языке C/C++:

```
double start_time, C[Nx][Ny][Nz]; //массив концентраций примеси
int rank, size; //номер и общее количество параллельных процессов
MPI_Init(&argc, &argv); //инициализация MPI
MPI_Comm_size(MPI_COMM_WORLD, &size); // определяем число параллельных процессов
MPI_Comm_rank(MPI_COMM_WORLD, &rank); // определяем номер процесса
... // инициализация датчика псевдослучайных чисел на каждом параллельном процессе
...
if (rank == 0) start_time = MPI_Wtime(); //засечка начала времени счета
for (n = rank + 1; n <= NP; n += size) //распределение итераций по процессам
{
    TIME = 0; XP = XP0; YP = YP0; ZP = ZP0; //задание начального положения частицы
    do { //отслеживание траектории
        ... //расчет параметров АПС, компонент скорости и новых координат частицы
        ... //определение положения частицы (IP,JP,KP) в массиве концентрации C[Nx][Ny][Nz];
        C[IP][JP][KP] += MIST * TAU/(NP*Vol); //вычисление концентрации в объеме ячейки Vol
        TIME += TAU;
    } while (XP < LX && abs(YP) < 0.5 * LY); //пока частица внутри области
} //цикл по номерам частиц
MPI_Reduce(C, Cglobal, NX * NY * NZ, MPI_DOUBLE, MPI_SUM, 0, MPI_COMM_WORLD);
double end_time = MPI_Wtime(); //засечка окончания времени счета
```

Стандарт OpenMP предназначен для программирования многопоточных приложений на многопроцессорных системах с общей памятью. OpenMP позволяет создавать эффективные параллельные программы с помощью директив компилятора, библиотечных функций и задания переменных окружения [16]. Стандарт OpenMP встроен в современные компиляторы C/C++ и Fortran. Его важными преимуществами являются возможность распараллеливания только фрагментов последовательных программ и использования одного кода как для последовательных, так и параллельных вычислений.

В рассматриваемой версии ЛДСМ распараллеливание с помощью стандарта OpenMP также возможно провести с точки зрения распределения частиц по параллельным потокам. Тем более что для реализации этого имеется директива `#pragma omp for`. Ниже приведем фрагмент кода параллельной OpenMP-программы для ЛДСМ:

```
double start_time, end_time, C[Nx][Ny][Nz];
#pragma omp parallel //начало параллельной области
{
    ... // инициализация датчика псевдослучайных чисел для каждого параллельного потока
#pragma omp master
    start_time = omp_get_wtime();
    ...
#pragma omp for private (TIME,XP,YP,ZP,IP, JP, KP,...)
    for (n = 1; n <= NP; n++)
    {
```

```

TIME = 0; XP = XP0; YP = YP0; ZP = ZP0;
do {
    ... //расчет параметров АПС, компонент скорости и новых координат частицы
    ... //определение положения частицы (IP,JP,KP) в массиве концентрации C[Nx][Ny][Nz];
#pragma omp atomic
    C[IP][JP][KP] += MIST * TAU/(NP*Vol); //вычисление концентрации в объеме ячейки Vol
    TIME += TAU;
    } while (XP < LX && abs(YP) < 0.5 * LY); //пока частица внутри области
} //цикл по номерам частиц
#pragma omp master
    end_time = omp_get_wtime();
} //end parallel domain

```

3. Результаты расчетов

Для тестирования разработанных параллельных программ была рассмотрена задача о движении частиц в конвективном пограничном слое, который формируется в дневное время суток, от высотного точечного источника постоянной мощности $Mist = 1\,000$ условных единиц в секунду. Данная задача рассматривается в различных научных публикациях [4, 17, 18] в качестве тестовой для проверки качества моделей переноса примеси. Приведем некоторые параметры, принятые в расчетах. Горизонтальная скорость ветра 5 м/с, высота источника $Zist = 0,49Lz$, размеры области исследования $Lx = 15\,000$ м вдоль оси Ox (основное направление движения воздушного потока), $Ly = 200$ м вдоль оси Oy (горизонтальное направление, поперечное основному потоку) и $Lz = 1\,000$ м. Для получения пространственного распределения концентрации примеси использовалась обычная равномерная трехмерная декартова сетка для рассматриваемой области с общим количеством ячеек $Nx \times Ny \times Nz = 150 \times 200 \times 100 = 3\,000\,000$. Начальное положение частиц имеет следующие координаты $Xp_0 = 0$, $Yp_0 = 0$, $Zp_0 = Zist$. В процессе моделирования из источника запускался миллион частиц. Отслеживание траектории частицы заканчивалось, если она выходила за границы области ($Xp > Lx$ или $|Yp| > 0,5Ly$). При работе MPI- и OpenMP-программ на различном количестве параллельных потоков / процессов фиксировались:

- общее время работы программы;
- время работы ее динамической части, когда в рамках ЛДСМ отслеживалось ее движение и выполнялся расчет турбулентных параметров конвективного пограничного слоя;
- время работы блока программы, ответственного за определение координат частиц и расчет массивов концентрации частиц, высоты пролета и количества частиц в сечении $X = const$.

На основании полученных временных характеристик были исследованы эффективность и ускорение параллельных MPI- и OpenMP-программ.

В расчетах использовался вычислительный сервер с двумя 12-ядерными центральными процессорами Intel Xeon Intel Xeon Silver 4214 2,2 ГГц и 192 Гб оперативной памяти. При компиляции программ использовался компилятор Intel.

В таблице приведены временные характеристики работы программ, полученные в результате осреднения по трем запускам для каждого представленного случая.

Общее время Tp , время работы динамической части $Tdyn$, время работы заполнения массива концентраций $Tarr$ в зависимости от количества используемых параллельных потоков / процессов p

Tp , с	$p = 1$	$p = 3$	$p = 6$	$p = 12$	$p = 18$	$p = 24$	$p = 36$	$p = 48$
Tp MPI	541,3	188,4	96,3	49,9	35,7	26,9	28,6	21,9
Tp MPI <i>dyn</i>	217,0	73,3	38,0	19,4	13,9	10,4	8,9	7,5
Tp MPI <i>arr</i>	273,3	92,6	48,6	25,1	17,9	13,5	14,2	12,8
Tp OMP	623,4	228,1	117,6	60,2	43,2	33,0	30,2	25,8
Tp OMP <i>dyn</i>	309,6	109,1	55,5	28,7	20,5	15,5	12,5	11,6
Tp OMP <i>arr</i>	245,0	91,3	47,4	24,6	17,6	13,5	12,0	11,8

Из таблицы видно, что при увеличении числа используемых параллельных процессов значительно сокращается время счета с ~ 10 минут до 20–30 с. Также можно отметить, что использование библиотеки MPI при остальных одинаковых условиях дает меньшее (приблизительно на 20%) время выполнения параллельной программы на рассматриваемой вычислительной системе с общей памятью по сравнению с OpenMP-программой. В первую очередь это связано с расчетом динамических характеристик частиц и параметров пограничного слоя атмосферы (T_{dyn}), именно между этими значениями, полученными при применении различных параллельных технологий (MPI vs OpenMP), наблюдается заметная разница. Затраты на определение координат частиц (индексов массива) и вычисление концентрации частиц в разработанных параллельных программах *Tarr* близкие, причем для различного количества используемых параллельных процессов / потоков в OpenMP-программе они ниже, что, по-видимому, связано с прямым заполнением массива концентраций, в то время как для этого в MPI-программе применяется функция `MPI_Reduce`. Тем не менее стоит отметить, что временные затраты на наполнение информационных массивов концентраций и других рассчитываемых параметров весьма велики и составляют почти половину от общего времени работы параллельных программ.

На рис. 4 приведены графики ускорения $S_p = T_1/T_p$ и эффективности $E_p = S_p/p$ полученных параллельных программ для различного числа параллельных процессов p .

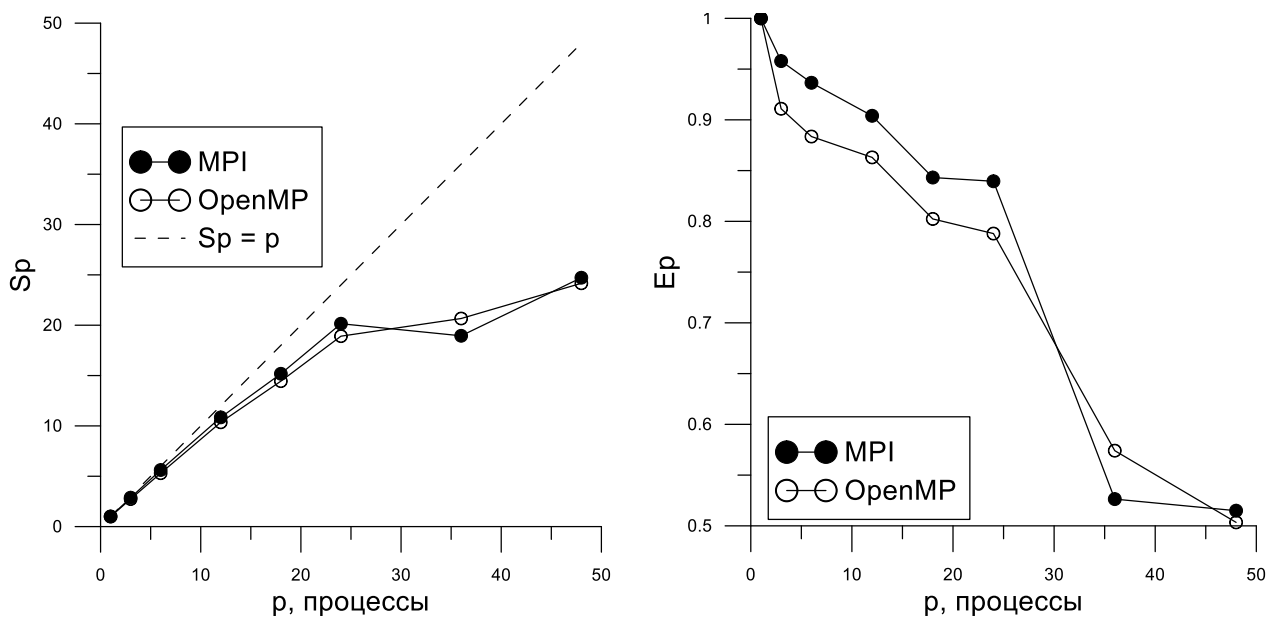


Рис. 4. Графики изменения ускорения S_p и эффективности E_p параллельных MPI- и OpenMP-программ для ЛДСМ в зависимости от количества параллельных процессов. Штриховой линией показано «идеальное» ускорение
 Fig. 4. Graphs of changes in S_p acceleration and E_p efficiency of parallel MPI and OpenMP programs for LDSM, depending on the number of parallel processes. The dashed line marks the “perfect” acceleration

На рис. 4 можно заметить, что в целом на рассматриваемой вычислительной системе обе программы до $p = 24$ показывают неплохое ускорение с достаточно близкими значениями. Далее с ростом p при использовании технологии Intel Hyper-Threading, позволяющей на одном ядре процессора обрабатывать сразу два потока данных, рост ускорения параллельных программ замедляется, однако при $p = 48$ удается достигнуть максимального ускорения в 24–25 раз. Судя по графику эффективности (см. рис. 4), MPI-версия параллельного кода является более эффективной (ее эффективность составляет более 80% при $p \leq 24$), OpenMP-программа имеет немного худшую эффективность. При увеличении числа используемых параллельных процессов получается эффективность около 50%, что связано с применением технологии Hyper-Threading.

Заключение

Для исследования распространения примеси от одного или многочисленных высотных точечных источников разработана параллельная версия лагранжевой дисперсионной стохастической модели, позволяющая с высокой эффективностью одновременно реализовывать запуски представительных частиц. При разработке параллельной версии модели использовались две технологии параллельного программирования – библиотека Message Passing Interface и встроенная в компилятор система Open Multiprocessing. Для каждой из этих технологий были подготовлены параллельные программы, которые при тестовых запусках показали их высокую эффективность: до 80% при использовании всех ядер многопроцессорной вычислительной системы. Несколько большую эффективность продемонстрировала параллельная MPI-программа.

В целом в качестве одного из направлений дальнейшего повышения эффективности параллельного программирования лагранжевой дисперсионной стохастической модели можно рассматривать снижение затрат на заполнение в процессе параллельного счета информационных массивов концентрации примеси и других представительных параметров модели. Тем не менее в рассматриваемых в работе условиях применения модели было достигнуто ускорение в 24–25 раз при запуске одного миллиона частиц.

Список источников

1. Марчук Г.И. Математическое моделирование в проблеме окружающей среды. М. : Наука, 1982. 320 с.
2. Селегей Т.С. Разработка усовершенствованного комплексного метеорологического показателя рассеивающей способности атмосферы (на примере территории Западной Сибири) : отчет о научно-исследовательской работе ФГБУ СибНИГМИ. Новосибирск, 2013.
3. Karataeva E.A. Application of the Lagrangian stochastic dispersion model for urban air pollution simulation // 29th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics : Proc. of SPIE. 2023. V. 12780. doi: 10.1117/12.2691014
4. Атмосферная турбулентность и моделирование распространения примесей / под ред. Ф.Т.М. Ньюистадта, Х. Ван Допа. Л. : Гидрометеиздат, 1985. 350 с.
5. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. СПб. : БХВ-Петербург, 2002. 608 с.
6. Starchenko A.V., Danilkin E.A. et al. A Supercomputer-Based Modeling System for Short-Term Prediction of Urban Surface Air Quality // Supercomputing Frontiers and Innovations, 2022. V. 9 (1). P. 17–31. doi: 10.14529/jsfi220102
7. Larson D.J., Nasstrom J.S. Shared- and distributed-memory parallelization of a Lagrangian atmospheric dispersion model // Atmospheric Environment. 2002. V. 36. P. 1569–1564.
8. Pisso I., Sollum E. et al. The Lagrangian particle dispersion model FLEXPART version 10.3 // Geosci. Model Dev. Discuss. 2019. doi: 10.5194/gmd-2018-333
9. Roberti D.R., Souto R.P. et al. Parallel Implementation of a Lagrangian Stochastic Model for Pollutant Dispersion // International Journal of Parallel Programming. 2005. V. 33 (5). P. 485–498. doi: 10.1007/s10766-005-7302-z
10. Lenotre L. A Strategy for Parallel Implementations of Stochastic Lagrangian Simulation // Monte Carlo and Quasi-Monte Carlo Methods : conference paper / R. Cools, D. Nuyens (eds). Springer, 2016. P. 507–520. (Springer Proceedings in Mathematics and Statistics; v. 163).
11. Hoffmann L., Baumeister P.F. et al. Massive-Parallel Trajectory Calculations version 2.2 (MPTRAC-2.2): Lagrangian transport simulations on graphics processing units (GPUs) // Geosci. Model Dev. 2022. V. 15. P. 2731–2762.
12. Семёнов Е.В., Старченко А.В. и др. Выбор технологии распараллеливания численного решения конвективно-диффузионного уравнения на гибридной многопроцессорной вычислительной системе // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2020. № 51. С. 111–120.
13. Thomson D.J. Criteria for the selection of stochastic models of particle trajectories in turbulent flows // J. Fluid Mech. 1987. V. 180. P. 529–556.
14. Hurley P.J. The air pollution model (TAPM) version 2. Part 1: Technical description // CSIRO Atmospheric Research Technical Paper. 2002. № 55. 49 p.
15. Старченко А.В., Кижнер Е.А. и др. Численное моделирование погоды и качества атмосферного воздуха в городах. Томск : Изд-во Том. ун-та, 2022. 140 с.
16. Антонов А.С. Технологии параллельного программирования MPI и OpenMP. М. : Изд-во Моск. ун-та, 2013. 344 с.
17. Luhar A.K., Britter R.E. A random walk model for dispersion in inhomogeneous turbulence in a convective boundary layer // Atmospheric Environment. 1989. V. 23 (9). P. 1911–1924.

18. Franzese P., Luhar A.K. et al. An efficient Lagrangian stochastic model of vertical dispersion in the convective boundary layer // *Atmospheric Environment*. 1999. V. 33 (15). P. 2337–2345.

References

1. Marchuk, G.I. (1982) *Matematicheskoe modelirovanie v probleme okruzhayushchey sredy* [Mathematical Modeling in the Problem of the Environment]. Moscow: Nauka.
2. Selegey, T.S. (2013) *Razrabotka usovershenstvovannogo kompleksnogo meteorologicheskogo pokazatelya rasseivayushchey spo-sobnosti atmosfery (na primere territorii Zapadnoy Sibiri) : otchet o nauchno-issledovatel'skoy rabote FGBU SibNIGMI* [Development of an improved complex meteorological indicator of the scattering capacity of the atmosphere (using the territory of Western Siberia as an example): report on the research work of the Federal State Budgetary Institution SibNIGMI]. Novosibirsk: [s.n.].
3. Karataeva, E.A. (2023) Application of the Lagrangian stochastic dispersion model for urban air pollution simulation. *29th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. Proc. of SPIE. V. 12780. DOI: 10.1117/12.2691014.
4. Newstadt, F.T.M. & Van Dope, H.L. (eds) (1985) *Atmosfernaya turbulentnost' i modelirovanie rasprostraneniya primesey* [Atmospheric Turbulence and Modeling of Impurity Propagation]. Leningrad: Gidrometeoizdat.
5. Voevodin, V.V. & Voevodin, V.I. (2002) *Parallel'nye vychisleniya* [Parallel Computing]. St. Petersburg: BKhV-Peterburg.
6. Starchenko, A.V., Danilkin, E.A. et al. (2022) A Supercomputer-Based Modeling System for Short-Term Prediction of Urban Surface Air Quality. *Supercomputing Frontiers and Innovations*. 9(1). pp. 17–31. DOI: 10.14529/jsfi220102.
7. Larson, D.J. & Nasstrom, J.S. (2002) Shared and distributed-memory parallelization of a Lagrangian atmospheric dispersion model. *Atmospheric Environment*. 36. pp. 1569–1564.
8. Pisso, I., Sollum, E. et al. (2019) The Lagrangian particle dispersion model FLEXPART version 10.3. *Geosci. Model Dev. Discuss*. DOI: 10.5194/gmd-2018-333
9. Roberti, D.R., Souto, R.P. et al. (2005) Parallel Implementation of a Lagrangian Stochastic Model for Pollutant Dispersion. *International Journal of Parallel Programming*. 33(5). pp. 485–498. DOI: 10.1007/s10766-005-7302-z
10. Lenotre, L. (2016) A Strategy for Parallel Implementations of Stochastic Lagrangian Simulation. In: Cools, R. & Nuyens, D. (eds) *Monte Carlo and Quasi-Monte Carlo Methods*. Vol. 163. Springer. pp. 507–520.
11. Hoffmann, L., Baumeister, P.F. et al. (2022) Massive-Parallel Trajectory Calculations version 2.2 (MPTRAC-2.2): Lagrangian transport simulations on graphics processing units (GPUs). *Geosci. Model Dev*. 15. pp. 2731–2762.
12. Semenov, E.V., Starchenko, A.V. et al. (2020) The choice of technology for parallelization of numerical the solution of the convective-diffusion equation on hybrid multiprocessor computer system. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 51. pp. 111–120. DOI: 10.17223/19988605/51/13
13. Thomson, D.J. (1987) Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*. 180. pp. 529–556.
14. Hurley, P.J. (2002) The air pollution model (TAPM) version 2. Part 1, Technical description. *CSIRO Atmospheric Research Technical Paper*. 55.
15. Starchenko, A.V., Kizhner, E.A. et al. (2022) *Chislennoe modelirovanie pogody i kachestva atmosfernogo vozdukha v gorodakh* [Numerical modeling of weather and atmospheric air quality in cities]. Tomsk: TSU.
16. Antonov, A.S. (2013) *Tekhnologii parallel'nogo programmirovaniya MPI i OpenMP* [MPI and OpenMP parallel programming technologies]. Moscow: MSU.
17. Luhar, A.K. & Britter, R.E. (1989) A random walk model for dispersion in inhomogeneous turbulence in a convective boundary layer. *Atmospheric Environment*. 23(9). pp. 1911–1924.
18. Franzese, P., Luhar, A.K. et al. (1999) An efficient Lagrangian stochastic model of vertical dispersion in the convective boundary layer. *Atmospheric Environment*. 33(15). pp. 2337–2345.

Информация об авторах:

Каратаева Екатерина Алексеевна – аспирант кафедры вычислительной математики и компьютерного моделирования механико-математического факультета Национального исследовательского Томского государственного университета (Томск, Россия); инженер Института оптики атмосферы им. В.Е. Зуева Сибирского отделения РАН (Томск, Россия). E-mail: karat@iao.ru

Старченко Александр Васильевич – профессор, доктор физико-математических наук, заведующий кафедрой вычислительной математики и компьютерного моделирования механико-математического факультета Национального исследовательского Томского государственного университета, ведущий научный сотрудник Регионального научно-образовательного математического центра Национального исследовательского Томского государственного университета (Томск, Россия), главный научный сотрудник Института оптики атмосферы им. В.Е. Зуева Сибирского отделения РАН (Томск, Россия). E-mail: starch@math.tsu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Karataeva Ekaterina A. (Post-graduate Student, National Research Tomsk State University, Tomsk, Russian Federation; V.E. Zuev Institute of Atmospheric Optics of Siberian Branch of the Russian Academy of Science, Tomsk, Russian Federation). E-mail: karat@iao.ru

Starchenko Alexander V. (Doctor of Physical and Mathematical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation; Senior Researcher, V.E. Zuev Institute of Atmospheric Optics of Siberian Branch of the Russian Academy of Science, Tomsk, Russian Federation). E-mail: starch@math.tsu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 23.10.2024; принята к публикации 03.03.2025

Received 23.10.2024; accepted for publication 03.03.2025

Научная статья

УДК 519.213

doi: 10.17223/19988605/70/4

Математическая модель нерегулируемого перекрестка в виде системы массового обслуживания с групповым входящим потоком и ненадежным прибором**Александра Владимировна Старовойтова¹, Александр Николаевич Моисеев²**^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Россия*¹ *starovoytova.tsu@gmail.com*² *moiseev.tsu@gmail.com*

Аннотация. Предложена математическая модель нерегулируемого перекрестка с примыканием второстепенной однополосной дороги к главной в виде однолинейной системы массового обслуживания с неординарным пуассоновским входящим потоком и ненадежным прибором. Особенностью рассматриваемой модели является то, что занятость перекрестка для транспортных средств со второстепенной дороги моделируется как ненадежный обслуживающий прибор, работоспособность которого определяется в виде цепи Маркова с двумя состояниями. Также важная отличительная особенность модели – неординарный входящий поток. В работе выполнено исследование данной модели, с применением метода производящих функций найдено стационарное распределение вероятностей числа заявок в системе. Проведены численные эксперименты.

Ключевые слова: система массового обслуживания; моделирование проезда перекрестка; ненадежный прибор; неординарный пуассоновский поток.

Для цитирования: Старовойтова А.В., Моисеев А.Н. Математическая модель нерегулируемого перекрестка в виде системы массового обслуживания с групповым входящим потоком и ненадежным прибором // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 42–51. doi: 10.17223/19988605/70/4

Original article

doi: 10.17223/19988605/70/4

Model of an Unsignalized Intersection in the Form of Queueing System with Batch Arrival Process and Unreliable Server**Alexandra V. Starovoytova¹, Alexander N. Moiseev²**^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*¹ *starovoytova.tsu@gmail.com*² *moiseev.tsu@gmail.com*

Abstract. In the paper, a mathematical model of a non-equal priority intersection of single-lane roads represented as a single-server queueing system with a batch Poisson arrival process and unreliable server. The study focuses on the passage of vehicles from the secondary road and the resulting queue. The traffic flow on the main road is modeled as the state of the server like "on/off" mode, determined by a given Markov chain. A system of Kolmogorov differential equations was constructed, and the generating function method was applied to solve the system. Additionally, the condition for the existence of a stationary regime of the system was found. As a result, the steady-state probability distribution of the number of customers is derived. Numerical experiments are conducted.

Keywords: queueing system; modeling of passing an intersection; batch Poisson process; unreliable server.

For citation: Starovoytova, A.V., Moiseev, A.N. (2025) Model of an unsignalized intersection in the form of queueing system with batch arrival process and unreliable server. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnik i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 42–51. doi: 10.17223/19988605/70/4

Введение

Реальные транспортные потоки являются стохастическими процессами, поэтому применение для них теории случайных процессов и разработка стохастических математических моделей представляют собой актуальную задачу. Математическая модель нерегулируемого перекрестка может быть описана как система массового обслуживания, где транспортные потоки взаимодействуют и создают задержки для участников движения. Исследования, такие как анализ трафика [1–2], а также теоретическая оценка задержек на нерегулируемых перекрестках Tanner'a [3], заложили основы для изучения таких систем.

В [4–5] для анализа используются клеточные автоматы и рассматривается модель «медленного старта», учитывающая время старта, которое нужно автомобилю для того, чтобы начать движение.

Большое число работ в области математического моделирования проезда перекрестков применяет методы теории массового обслуживания. Так, в статье [6] рассматривается перекресток со светофорным регулированием с одинаковой длительностью переключений красного и зеленого сигналов, а также с произвольными длительностями переключений, что соответствует нерегулируемому перекрестку. В модели неравнозначного перекрестка рассматривается система с входящим пуассоновским потоком автомобилей и постоянным временем обслуживания. Также модель регулируемого перекрестка была рассмотрена в [7]. Особенность данной работы в том, что в предложенной модели в фазе зеленого сигнала светофора интенсивность обслуживания изменяется во времени.

Модель нерегулируемого перекрестка с ограничением мест в очереди для главной и второстепенной дороги была рассмотрена в [8]. В данной модели машины со второстепенной дороги проезжают перекресток с некоторой вероятностью, которая зависит от количества машин на главной дороге.

В работах [9–11] предложена математическая модель нерегулируемого перекрестка как система массового обслуживания M/G2/1. Время ожидания в данной работе зависит от «критического интервала» – минимального безопасного промежутка для проезда, который может варьировать случайным образом. Учитывая различные типы поведения водителей, оценивается их влияние на время ожидания и длину очереди.

В исследованиях [12–14] рассматриваются задачи управления регулирующими перекрестками.

В настоящей работе предлагается модель нерегулируемого перекрестка с примыканием однополосной второстепенной дороги к главной однополосной (обе дороги односторонние). Модель описывает проезд автомобилей со второстепенной дороги и образующуюся на ней очередь. Движение по главной дороге моделируется как состояние обслуживающего прибора по типу «включено / выключено». По сути, это представлено как «поломки» прибора, режим которых определяется заданной цепью Маркова.

В отличие от [6] в данной работе рассматривается модель в виде системы массового обслуживания с неординарным (групповым) пуассоновским входящим потоком и ненадежным прибором. Групповой входящий поток позволяет моделировать поступление автомобилей на рассматриваемый перекресток «пачками», что может быть вызвано, например, наличием другого (регулируемого) перекрестка перед данным или присутствием в голове пачки медленного транспортного средства.

1. Математическая модель и постановка задачи

Рассматривается математическая модель неравнозначного нерегулируемого перекрестка с примыканием однополосной второстепенной дороги к главной однополосной в виде системы массового обслуживания с неординарным пуассоновским входящим потоком и ненадежным прибором.

Пусть в систему поступают группы машин, образующие пуассоновский поток с параметром λ , при этом в группу приходит случайное число автомобилей v , где v – дискретная случайная величина, заданная рядом распределения $P\{v = s\} = l_s, s \in \{1, 2, \dots\}$. Входящий поток в данной модели формирует только появление автомобилей на второстепенной дороге.

Проезд автомобилей по главной дороге моделируется как цепь Маркова $k(t)$ с двумя состояниями

$$k(t) = \begin{cases} 1, & \text{обслуживание транспорта на второстепенной дороге,} \\ 0, & \text{проезд перекрестка запрещен для второстепенной дороги,} \end{cases}$$

которая задается матрицей инфинитезимальных характеристик $\mathbf{Q}$ с элементами, имеющими смысл интенсивностей перехода прибора из нерабочего состояния в рабочее (элемент q_{01}) и из рабочего в нерабочее (элемент q_{10}).

Время проезда (обслуживания) автомобиля со второстепенной дороги является экспоненциальной случайной величиной с параметром μ , если перекресток свободен ($k(t) = 1$ – обслуживающий прибор работает). Если же перекресток занят ($k(t) = 0$ – обслуживающий прибор «сломан»), то автомобили со второстепенной дороги не обслуживаются. Таким образом, интенсивность их обслуживания можно определить как случайный процесс

$$\mu(t) = \begin{cases} \mu, & \text{если } k(t) = 1, \\ 0, & \text{если } k(t) = 0. \end{cases}$$

Если подъезжающий автомобиль видит, что перекресток свободен, он проезжает его без остановки. Если же перекресток занят, автомобиль становится в очередь, при этом длина очереди не ограничена. Если водитель автомобиля со второстепенной дороги неверно оценил дорожную обстановку и не успел завершить проезд перекрестка, пока тот был свободен, а с главной дороги уже подъехал другой автомобиль, то в этом случае автомобиль со второстепенной дороги будет вынужден уступить дорогу и вернуться в очередь.

Ставится задача нахождения стационарного распределения числа автомобилей в очереди на второстепенной дороге, включая автомобиль, проезжающий перекресток, если таковой имеется.

Обозначим $i(t)$ – число транспортных средств на второстепенной дороге (включая выехавшее на перекресток) в момент времени t . Таким образом, целью исследования является получение стационарного распределения вероятностей процесса $i(t)$.

Очевидно, что случайный процесс $i(t)$ не является марковским. Но марковским является двумерный процесс $\{i(t), k(t)\}$. Обозначим его распределение вероятностей как $P_k(i, t) = P\{k(t) = k, i(t) = i\}$. Тогда составим следующую систему дифференциальных уравнений Колмогорова:

$$\begin{cases} \frac{\partial P_0(0, t)}{\partial t} = -(\lambda + q_{01})P_0(0, t) + q_{10}P_1(0, t), \\ \frac{\partial P_1(0, t)}{\partial t} - (\lambda + q_{10})P_1(0, t) + q_{01}P_0(0, t) + \mu P_1(1, t) = 0, \\ \frac{\partial P_0(i, t)}{\partial t} - (\lambda + q_{01})P_0(i, t) + \lambda \sum_{s=1}^i P_0(i-s, t)l_s + q_{10}P_1(i+1, t) = 0, \text{ при } i > 0, \\ \frac{\partial P_1(i, t)}{\partial t} - (\lambda + q_{10} + \mu)P_1(i, t) + \lambda \sum_{s=1}^i P_1(i-s, t)l_s + q_{01}P_0(i, t) + \mu P_1(i+1, t) = 0, \text{ при } i > 0. \end{cases}$$

Обозначая через $\Pi_k(i)$ стационарные вероятности распределения случайного процесса $i(t)$ в произвольный момент времени, перейдем к стационарному режиму, считая t произвольным, а вероятностные характеристики системы $P_k(i, t)$ не зависящими от времени. Тогда получим следующую систему линейных алгебраических уравнений:

$$\begin{cases} -(\lambda + q_{01})\Pi_0(0) + q_{10}\Pi_1(0) = 0, \\ -(\lambda + q_{10})\Pi_1(0) + q_{01}\Pi_0(0) + \mu\Pi_1(1) = 0, \\ -(\lambda + q_{01})\Pi_0(i) + \lambda \sum_{s=1}^i \Pi_0(i-s)l_s + q_{10}\Pi_1(i+1) = 0, \text{ при } i > 0, \\ -(\lambda + q_{10} + \mu)\Pi_1(i) + \lambda \sum_{s=1}^i \Pi_1(i-s)l_s + q_{01}\Pi_0(i) + \mu\Pi_1(i+1) = 0, \text{ при } i > 0. \end{cases}$$

Для нахождения вероятностей $\Pi_k(i)$ к данной системе нужно добавить условие нормировки

$$\sum_{i=0}^{\infty} (\Pi_0(i) + \Pi_1(i)) = 1.$$

2. Уравнения для частных производящих функций и их решение

Введем частичные производящие функции

$$G_0(z) = \sum_{i=0}^{\infty} z^i \Pi_0(i),$$

$$G_1(z) = \sum_{i=0}^{\infty} z^i \Pi_1(i),$$

$$L(z) = \sum_{s=1}^{\infty} z^s l_s.$$

Заметим, что

$$\sum_{i=1}^{\infty} z^i \sum_{s=1}^i \Pi_k(i-s) l_s = \sum_{s=1}^{\infty} z^s l_s \sum_{i=s}^{\infty} z^{i-s} \Pi_k(i-s) = G_k(z) L(z).$$

Тогда получим следующую систему уравнений:

$$\begin{cases} -(\lambda + q_{01})G_0(z) + \lambda G_0(z)L(z) + q_{10}G_1(z) = 0, \\ -(\lambda + q_{10})G_1(z) - \mu[G_1(z) - \Pi_1(0)] + \frac{\mu}{z}[G_1(z) - \Pi_1(0)] + \lambda G_1(z)L(z) + q_{01}G_0(z) = 0. \end{cases}$$

Выразим $G_1(z)$ из второго уравнения:

$$\begin{cases} -(\lambda + q_{01})G_0(z) + \lambda G_0(z)L(z) + q_{10}G_1(z) = 0, \\ G_1(z) = \frac{zq_{01}G_0(z) + \Pi_1(0)\mu(1-z)}{zq_{10} + z\lambda[1-L(z)] + \mu(z-1)}. \end{cases}$$

Тогда, решив данную систему, находим

$$\begin{aligned} G_0(z) &= \frac{q_{10}(1-z)\mu\Pi_1(0)}{q_{01}[\lambda z(L(z)-1) + \mu(1-z)] - [\lambda(L(z)-1)(\mu - q_{10}z + z(\lambda L(z) - \lambda - \mu))]}, \\ G_1(z) &= \frac{(1-z)(q_{01} + \lambda - \lambda L(z))\mu\Pi_1(0)}{q_{01}[\lambda z(L(z)-1) + \mu(1-z)] - [\lambda(L(z)-1)(\mu - q_{10}z + z(\lambda L(z) - \lambda - \mu))]}, \\ G(z) = G_0(z) + G_1(z) &= \frac{(1-z)(q_{01} + q_{10} + \lambda - \lambda L(z))\mu\Pi_1(0)}{q_{01}[\lambda z(L(z)-1) + \mu(1-z)] - [\lambda(L(z)-1)(\mu - q_{10}z + z(\lambda L(z) - \lambda - \mu))]} \end{aligned}$$

В данном выражении для производящей функции осталась неизвестной вероятность $\Pi_1(0)$. Найдем ее из условия нормировки $G(1) = 1$. Прямая подстановка приводит к неопределенности типа «0/0», раскрытие которой по правилу Лопиталя дает

$$\lim_{z \rightarrow 1} G(z) = - \frac{\Pi_1(0)\mu(q_{01} + q_{10} + \lambda - \lambda L'(z))}{q_{01}[-\mu + z\lambda(L'(z)-1)] - \lambda(L(z)-1)[-q_{10} - \lambda - \mu + z\lambda L'(z) + \lambda L(z)] - \lambda\mu L'(z)[-q_{10}z + \mu + \bar{z}]},$$

где $\bar{z} = z(-\lambda - \mu + z(\lambda - \mu + \lambda L(z)))$.

Учитывая, что

$$L'(z)|_{z=1} = \sum_{s=1}^{\infty} s l_s = m, \quad L(z)|_{z=1} = \sum_{s=1}^{\infty} l_s = 1,$$

получим

$$- \frac{\Pi_1(0)\mu(q_{01} + q_{10})}{q_{01}(-\mu + \lambda m) + \lambda m q_{10}} = 1.$$

Таким образом,

$$\Pi_1(0) = \frac{q_{01}}{q_{01} + q_{10}} - m \frac{\lambda}{\mu},$$

где m – среднее число заявок в пачке. Отсюда, в частности, получаем условие существования стационарного режима

$$m \frac{\lambda}{\mu} < \frac{q_{01}}{q_{01} + q_{10}}. \quad (1)$$

Таким образом, окончательно получаем следующее выражение для производящей функции числа заявок в очереди в стационарном режиме:

$$G(z) = \frac{\left(\frac{q_{01}}{q_{01} + q_{10}} - m \frac{\lambda}{\mu} \right) \mu (1-z) [q_{01} + q_{10} + \lambda - \lambda L(z)]}{q_{01} [\lambda z (L(z) - 1) + \mu (1-z)] - \lambda (L(z) - 1) [\mu - q_{10} z + z (\lambda L(z) - \lambda - \mu)]}.$$

Отсюда можно получить стационарное распределение вероятностей, применив обратное преобразование Фурье к соответствующей характеристической функции:

$$\Pi(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-jiz} G(e^{jz}) dz,$$

где $j = \sqrt{-1}$.

3. Численные примеры

Для иллюстрации того, как в разных режимах работы системы меняется характер полученного стационарного распределения, приведем несколько иллюстративных численных примеров.

Основываясь на (1), введем понятие коэффициента загрузки системы ρ :

$$\rho = m \frac{\lambda}{\mu} \cdot \frac{q_{01} + q_{10}}{q_{01}}. \quad (2)$$

Зафиксируем $\mu = 10$, а интенсивность входящего потока λ будем вычислять из (2) на основе заданных ρ и m . Для коэффициента загрузки системы ρ будем брать три уровня значений: 0,1 (низкая загрузка), 0,5 (средняя загрузка), 0,9 (высокая загрузка). Средний размер пачки m примем равным 3 и 6. В качестве распределения размера пачки рассмотрим следующие варианты:

а) детерминированное со значением, равным 1 (ординарный входящий поток, берется для сравнения);

б) распределение с фиксированным размером пачки m ;

с) равномерное дискретное распределение от 1 до $2m - 1$;

д) геометрическое распределение с матожиданием m ;

е) смещенное на 1 распределение Пуассона с матожиданием m .

Рассмотрим два вида режима работы прибора:

– режим 1 – $q_{01} = 0,2$, $q_{10} = 0,8$, в этом режиме прибор доступен для обслуживания 20% общего времени;

– режим 2 – $q_{01} = 0,8$, $q_{10} = 0,2$, в этом режиме прибор доступен для обслуживания 80% общего времени.

На рис. 1, 2 представлены соответствующие распределения для описанных режимов работы и различных значений загрузки системы и размера пачки. На рис. 3 показано, как изменяется математическое ожидание числа заявок в системе с ростом загрузки системы. На основе проведенного численного моделирования можно заметить следующее:

1. Если размер пачки фиксирован (случай б), то в распределении вероятностей числа заявок в системе наблюдаются характерные «зубцы». Это явление можно объяснить тем, что при добавлении

заявок новой пачки к общему числу заявок в системе добавляется константа m , равная размеру пачки, а значит, вероятности значений, кратные этой величине, будут больше остальных. Такой эффект ярче проявляется в условиях средней и высокой загрузки системы ($\rho = 0,5$ и $\rho = 0,9$).

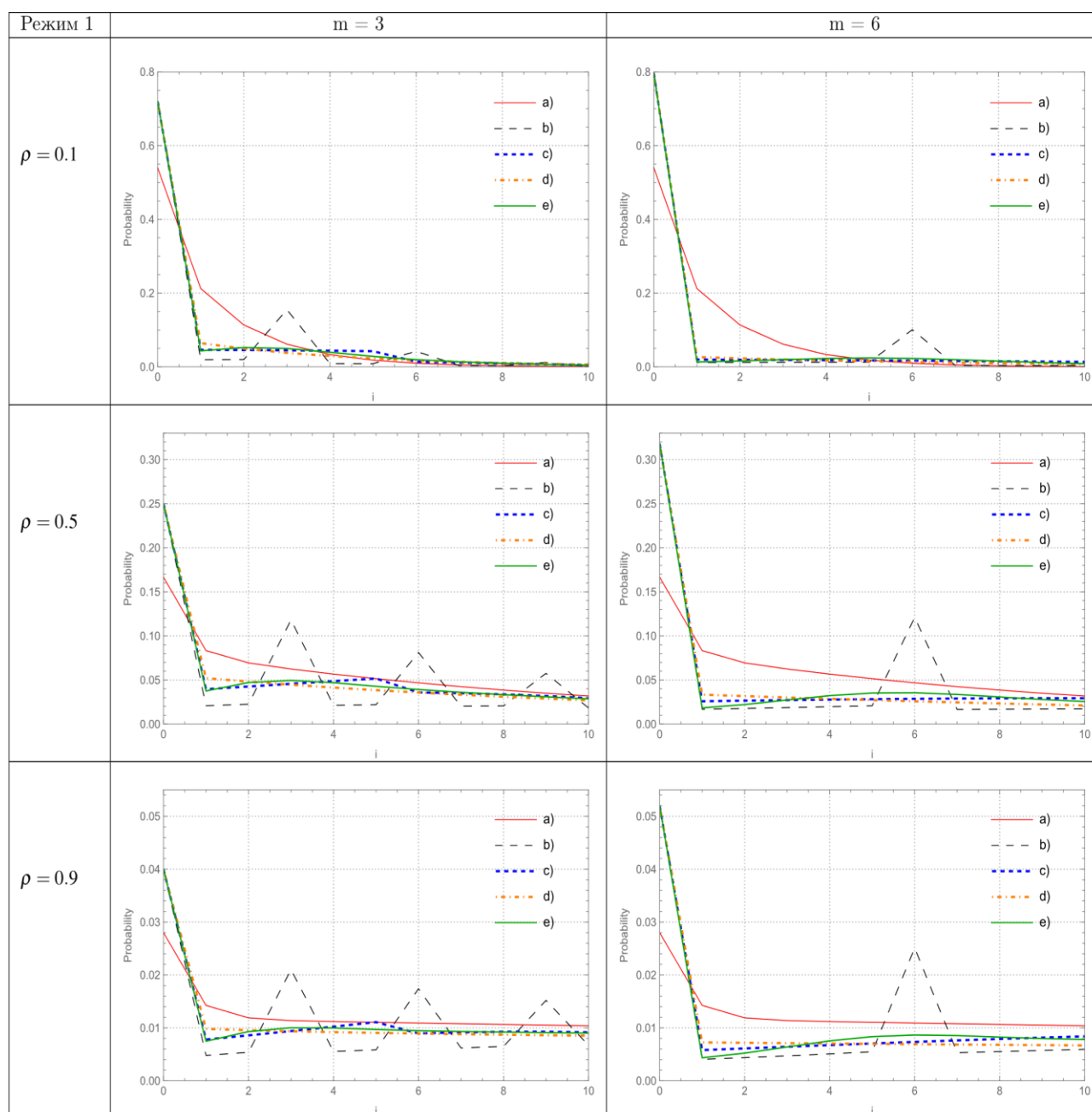


Рис. 1. Стационарные распределения вероятностей числа заявок в системе в режиме 1 работы прибора для различных значений коэффициента загрузки ρ при разных распределениях размера пачки (а–е)
 Fig. 1. Steady-state probability distributions of the number of customers in mode 1 of the server operation for various values of loading coefficient ρ and different distributions of a batch size (a–e)

2. Для некоторых типов распределений (например, равномерное, Пуассона) наблюдаются участки со сменой характера монотонности значений вероятностей числа заявок в системе, что, видимо, объясняется тем же эффектом, что и для детерминированного случая, но здесь «зубцы» полигона распределения более размыты вследствие недетерминированности размера пачки.

3. При большой загрузке распределение числа заявок в системе имеет так называемый тяжелый хвост, что иллюстрируется данными, представленными в таблице. В частности, для геометрического

распределения размера пачки ($m = 6$) даже в случае 80% доступности прибора вес хвоста для $i > 100$ (сумма всех вероятностей для $i > 100$) составляет 0,255, т.е. более четверти времени общее число заявок в системе будет превышать 100. А в случае режима 20% доступности прибора положение становится еще хуже: почти 45% общего времени число скопившихся транспортных средств будет больше 100.

4. Несмотря на то, что вероятность пустой системы для случая ординарного потока во всех случаях меньше остальных вариантов (см. рис. 1, значение при $i = 0$), среднее число заявок в системе с ординарным потоком оказывается меньше, чем в системе с групповым потоком при той же загрузке (см. рис. 3). Этот эффект только усиливается с ростом коэффициента загрузки ρ .

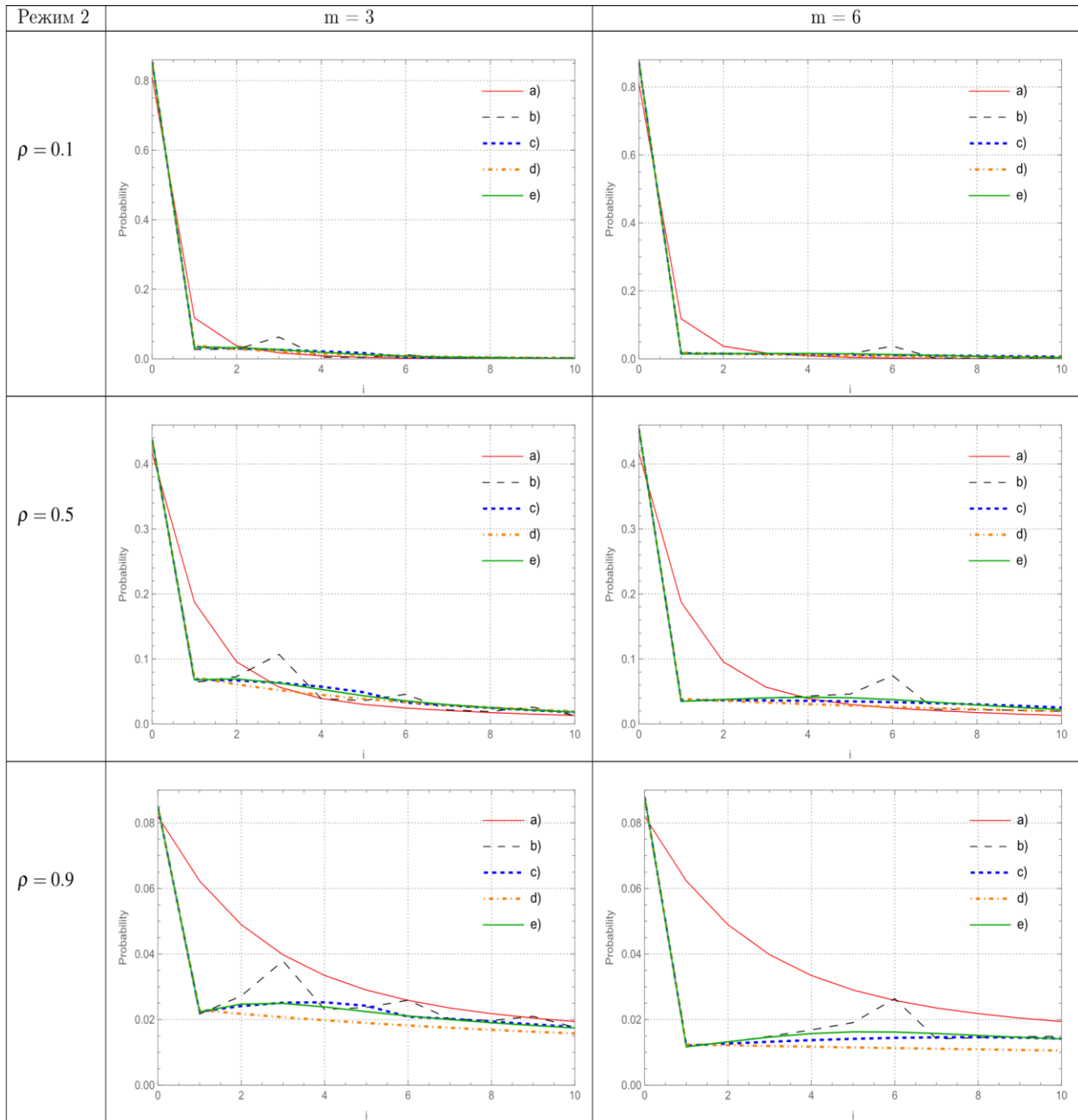


Рис. 2. Стационарные распределения вероятностей числа заявок в системе в режиме 2 работы прибора для различных значений коэффициента загрузки ρ при разных распределениях размера пачки (а–е)
 Fig. 2. Steady-state probability distributions of the number of customers in mode 2 of the server operation for various values of loading coefficient ρ and different distributions of a batch size (a–e)

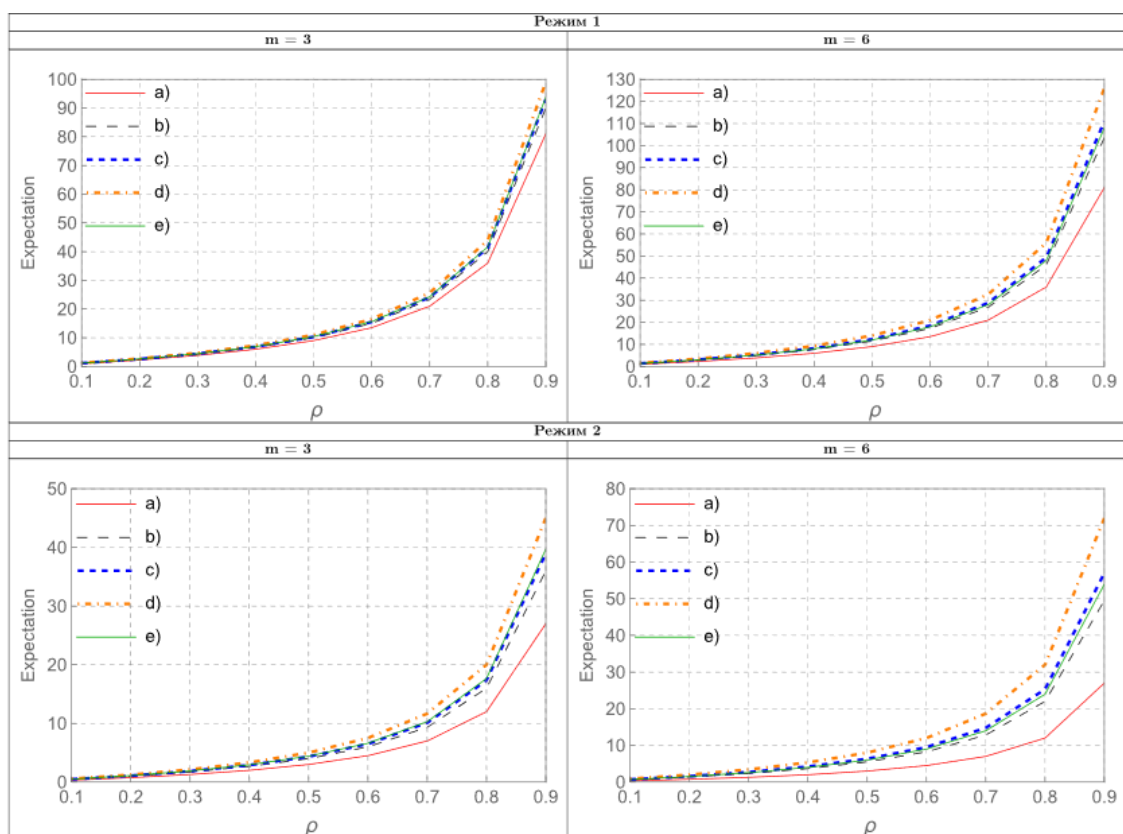


Рис. 3. Изменение математического ожидания числа заявок в системе с ростом значений коэффициента загрузки ρ
 Fig. 3. Changing of mathematical expectation of the number of customers in the system while loading coefficient ρ grows

Вес хвостов распределений для $i > 100$

Высокая нагрузка системы ($\rho = 0,9$)	$m = 3$		$m = 6$	
	Режим 1	Режим 2	Режим 1	Режим 2
Детерминированное распределение	0,287	0,038	0,287	0,038
Распределение с фиксированным размером пачки	0,328	0,076	0,378	0,144
Равномерное распределение	0,340	0,090	0,403	0,182
Геометрическое распределение	0,362	0,121	0,445	0,255
Распределение Пуассона	0,343	0,094	0,393	0,167

Заключение

В работе предложена математическая модель нерегулируемого перекрестка с примыканием однополосной второстепенной дороги к главной однополосной в виде однолинейной системы массового обслуживания с неординарным пуассоновским входящим потоком и ненадежным прибором. Особенностью рассматриваемой СМО является то, что работоспособность обслуживающего прибора определяется в виде случайного процесса (цепи Маркова с двумя состояниями), а входящий поток является неординарным. Выполнено исследование данной СМО, найдено стационарное распределение вероятностей числа заявок в системе (транспортных средств на второстепенной дороге). Полученные теоретические результаты позволяют выполнять расчеты загрузки нерегулируемого перекрестка и вычислять вероятностные характеристики числа транспортных средств, скапливающихся на второстепенной дороге.

Приведенные в работе результаты численных экспериментов позволяют представить некоторую качественную картину процессов, происходящих в соответствующих задачах реального дорожного движения. В частности, выводы 3 и 4 раздела 3 говорят о том, что если на дороге (в данном случае – второстепенной) нет условий для формирования пачек, то среднее число ожидающих автомобилей

будет ниже (при большой загрузке – значительно ниже), чем для случаев, когда имеются факторы, приводящие к формированию на дороге групп транспортных средств. Кроме того, нечасто будет скапливаться огромное число транспортных средств (например, более 100, – см. анализ хвостов распределения в выводе 3 и таблице). В случае наличия условий для формирования пачек можно порекомендовать владельцу дороги или заинтересованному лицу изменить политику дорожного движения на данном участке, например по возможности убрать регулирование на перекрестке, близко расположенном перед данным, разрешить обгон или организовать дополнительную полосу для обгона, запретить движение тихоходных транспортных средств, установить минимальное ограничение скорости и т.п.

Список источников

1. Greenberg H. An Analysis of Traffic Flow // *Operations Research*. 1959. V. 7 (1). P. 79–85.
2. Greenshields B.D. A Study of Traffic Capacity // *Highway Research Board*. 1935. V. 14. P. 448–477.
3. Tanner J.C. A theoretical analysis of delays at an uncontrolled intersection // *Biometrika*. 1962. V. 49. P. 163–170.
4. Schadschneider A., Schreckenberg M. Traffic flow models with ‘slow-to-stars’ rules // *Ann. Phys. Lpz.* 1997. V. 6. P. 541–551.
5. Castellares F.C., Ferrari P.A., Pechersky E. A slow-to-start traffic model related to a M/M/1 queue // *Journal of Statistical Mechanics: Theory and Experiment*. 2007. V. 2007 (07). Art. 07008. doi: 10.1088/1742-5468/2007/07/P07008
6. Афанасьева Л.Г., Булинская Е.В. Математические модели транспортных систем, основанные на теории очередей // *Труды Московского физико-технического института*. 2010. № 2 (4). С. 6–21.
7. Завалишин Д.С., Тимофеева Г.А. Исследование математической модели регулируемого перекрестка // *Труды Института математики и механики УрО РАН*. 2009. № 15 (4). С. 108–119
8. Семенов В.В. Математическое моделирование транспортного потока на нерегулируемом пересечении // *Математическое моделирование*. 2008. № 20 (10). С. 14–22.
9. Heidemann D., Wegmann H. Queueing at unsignalized intersections // *Transportation Research. Part B: Methodological*. 1997. V. 31 (3), P. 239–263.
10. Abhishek, Boon M., Mandjes M., Núñez-Queija R. Congestion analysis of unsignalized intersections // *International Conference on Communication Systems and Networks (COMSNETS)*. 2016. V. 8. P. 1–6. doi: 10.48550/arXiv.1802.06858
11. Abhishek, Boon M., Mandjes M. Generalized gap acceptance models for unsignalized intersections // *Math Meth Oper Res*. 2019. V. 89. P. 385–409.
12. Федоткин М.А., Зорин А.В. Стохастические модели процессов адаптивного управления конфликтными потоками неоднородных требований // *Теория вероятностей и ее применения*. 2020. № 65 (1). С. 163–164.
13. Федоткин А.М., Федоткин М.А., Кудрявцев Е.В. Динамические модели неоднородного потока транспорта на магистралях // *Автоматика и телемеханика*. 2020. № 8. С. 149–164.
14. Федоткин М.А., Кудрявцев Е.В. Управляющие системы и механизм образования транспортных пачек на магистралях с интенсивным движением // *Проблемы теоретической конференции : сб. ст. XVI Междунар. конф. (Н. Новгород, 20–25 июня 2011 г.)*. Н. Новгород : Изд-во Нижегород. гос. ун-та им. Н.И. Лобачевского, 2011. С. 503–507.

References

1. Greenberg, H. (1959) An Analysis of Traffic Flow. *Operations Research*. 7(1). pp. 79–85.
2. Greenshields, B.D. (1935) A Study of Traffic Capacity. *Highway Research Board*. 14. pp. 448–477.
3. Tanner, J.C. (1962) A theoretical analysis of delays at an uncontrolled intersection. *Biometrika*. 49. pp. 163–170.
4. Schadschneider, A. & Schreckenberg, M. (1997) Traffic flow models with ‘slow-to-stars’ rules. *Annalen der Physik*. 6. pp. 541–551.
5. Castellares, F.C., Ferrari, P.A. & Pechersky, E. (2007) A slow-to-start traffic model related to a M/M/1 queue. *Journal of Statistical Mechanics: Theory and Experiment*. 07. Art. 07008. DOI: 10.1088/1742-5468/2007/07/P07008
6. Afanasyeva, L.G. & Bulinskaya, E.V. (2010) Matematicheskie modeli transportnykh sistem, osnovannye na teorii ocherey [Mathematical models of transport systems based on queueing methods]. *Trudy Moskovskogo fiziko-tekhnicheskogo instituta*. 2(4). pp. 6–21.
7. Zavalishin, D.S. & Timofeeva, G.A. (2009) Issledovanie matematicheskoy modeli reguliruemogo perekrestka [A Study of a Mathematical Model of a Controlled Intersection]. *Trudy Instituta matematiki i mekhaniki UrO RAN*. 15(4). pp. 108–119.
8. Semenov, V.V. (2009) Matematicheskoe modelirovanie transportnogo potoka na nereguliruемом peresechenii [Mathematical Modeling of Traffic Flow at an Uncontrolled Intersection]. *Matematicheskoe modelirovanie*. 1(5). pp. 543–550.
9. Heidemann, D. & Wegmann, H. (1997) Queueing at unsignalized intersections. *Transportation Research Part B: Methodological*. 31(3). pp. 239–263.
10. Abhishek, Boon, M., Mandjes, M. & Núñez-Queija, R. (2016) Congestion analysis of unsignalized intersections. *International Conference on Communication Systems and Networks (COMSNETS)*. vol. 8, pp. 1–6. <https://doi.org/10.48550/arXiv.1802.06858>

11. Abhishek, Boon, M. & Mandjes, M. (2019) Generalized gap acceptance models for unsignalized intersections. *Mathematical Methods of Operations Research*. 89. pp. 385–409.
12. Fedotkin, M.A. & Zorin, A.V. (2020) Stokhasticheskie modeli protsessov adaptivnogo upravleniya konfliktnymi potokami neodnorodnykh trebovaniy [Stochastic Models of Adaptive Control Processes for Conflicting Flows of Heterogeneous Requests]. *Teoriya veroyatnostey i ee primeneniya*. 65(1). pp. 163–164.
13. Fedotkin, A.M., Fedotkin, M.A. & Kudryavtsev, E.V. (2020) Dinamicheskie modeli neodnorodnogo potoka transporta na magistralyakh [Dynamic Models of Heterogeneous Traffic Flow on Highways]. *Avtomatika i telemekhanika*. 8. pp. 149–164.
14. Fedotkin, M.A. & Kudryavtsev, E.V. (2011) Upravlyayushchie sistemy i mekhanizm obrazovaniya transportnykh pachek na magistralyakh s intensivnym dvizheniem [Control Systems and the Mechanism of Traffic Packet Formation on Highways with Intensive Traffic]. *Problemy teoreticheskoy konferentsii* [Problems of the Theoretical Conference]. Proc. of the 16th International Conference. Nizhny Novgorod, June 20–25, 2011. Nizhny Novgorod: N.I. Lobachevsky State University of Nizhny Novgorod. pp. 503–507.

Информация об авторах:

Старовойтова Александра Владимировна – магистрант кафедры теории вероятностей и математической статистики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: starovoytova.tsu@gmail.com

Моисеев Александр Николаевич – доктор физико-математических наук, доцент, заведующий кафедрой программной инженерии Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: moiseev.tsu@gmail.com

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Starovoytova Alexandra V. (Master's Student at the Department of Probability Theory and Mathematical Statistics, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: starovoytova.tsu@gmail.com

Moiseev Alexander N. (Doctor of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Software Engineering, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: moiseev.tsu@gmail.com

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 28.11.2024; принята к публикации 03.03.2025

Received 28.11.2024; accepted for publication 03.03.2025

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Научная статья

УДК 519.2

doi: 10.17223/19988605/70/5

Оценивание параметров гиперэкспоненциального распределения

Юлия Борисовна Буркатовская¹, Сергей Эрикович Воробейчиков²¹ *Национальный исследовательский Томский политехнический университет, Томск, Россия*^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Россия*¹ *tracey@tpu.ru*² *sev@mail.tsu.ru*

Аннотация. Рассматривается задача оценивания параметров гиперэкспоненциального распределения. Для распределения с тремя параметрами построены в явном виде оценки по методу моментов и комбинированные оценки. Получены условия применимости метода моментов и проведено его сравнение с комбинированным методом, сочетающим метод моментов и метод максимального правдоподобия. Показана возможность оценивания минимального параметра интенсивности в случае произвольного числа параметров гиперэкспоненциального распределения. Проведено численное моделирование, показывающее применимость разработанных алгоритмов оценивания.

Ключевые слова: гиперэкспоненциальное распределение; метод моментов; метод максимального правдоподобия.

Благодарности: Исследование выполнено при финансовой поддержке РФФ, грант № 24-11-00191.

Для цитирования: Буркатовская Ю.Б., Воробейчиков С.Э. Оценивание параметров гиперэкспоненциального распределения // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 52–61. doi: 10.17223/19988605/70/5

Original article

doi: 10.17223/19988605/70/5

Parameter estimation for hyperexponential distribution

Yulia B. Burkatovskaya¹, Sergey E. Vorobeychikov²¹ *National Research Tomsk Polytechnic University, Tomsk, Russian Federation*^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*¹ *tracey@tpu.ru*² *sev@mail.tsu.ru*

Abstract. The problem of the parameter estimation for the hyperexponential distribution is considered. The method of moments and the maximum likelihood method are used, explicit intensity estimators for a distribution with two intensity values are constructed. The conditions for the applicability of the method of moments are obtained and its comparison with a new method combining the method of moments and the maximum likelihood method is carried out. The possibility of estimating the minimum intensity parameter in the case of an arbitrary number of parameters of a hyperexponential distribution is shown. Numerical modeling shows the applicability of the developed estimation algorithms.

Keywords: hyperexponential distribution; method of moments; maximum likelihood method.

Acknowledgments: The study was supported by the Russian Science Foundation, grant no. 24-11-00191.

For citation: Burkatovskaya, Yu.B., Vorobeychikov, S.E. (2025) Parameter estimation for hyperexponential distribution. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 52–61. doi: 10.17223/19988605/70/5

Введение

Стохастические процессы с изменяющимися параметрами широко используются для аппроксимации реальных нелинейных процессов в различных приложениях, таких как экономика, финансы, медицина, социология, геология и т.д. С точки зрения статистики точка изменения, или разладки, – это момент времени, в котором наблюдения подчиняются одному распределению до этой точки и другому распределению после этой точки. Процессы с несколькими разладками можно определить аналогично [1]. Процессы с одной разладкой обычно используются для моделирования выхода из стабильных состояний или ошибки оборудования, тогда как процессы с несколькими разладками описывают реальные системы с несколькими состояниями, где переключения между состояниями происходят в неизвестные моменты времени. Упомянутая выше книга [1] посвящена параметрическому обнаружению точек изменения с приложениями к генетике, медицине и финансам и содержит ряд классических и недавних результатов в этой области для некоторых часто используемых распределений. В теории массового обслуживания процессы с несколькими разладками применяются для описания систем очередей с различными режимами поступления клиентов; в качестве простейшей модели можно рассмотреть процесс с двумя состояниями, соответствующими «обычному» и «пиковому» времени. Эти два состояния характеризуются разной интенсивностью наступления событий; в первом случае события происходят в целом гораздо реже, чем во втором. При этом большую роль играют модели, основанные на смеси экспоненциальных распределений.

Для оценки смесей распределений обычно используются различные модификации алгоритма ЕМ (expectation-minimization), основанные на оценках максимального правдоподобия. Каждая итерация алгоритма состоит двух шагов. На шаге ожидания (expectation) вычисляется ожидаемое значение функции правдоподобия с использованием текущей оценки параметров. На шаге максимизации (maximization) вычисляются параметры, максимизирующие ожидаемую функцию максимального правдоподобия. Впервые алгоритм с таким названием был предложен в статье [2], хотя итерационные оценки максимального правдоподобия предлагались и ранее: например, в работе [3] они использованы для анализа марковских цепей. В дальнейшем появилось большое количество публикаций, в которых предложенный подход применялся и модифицировался для разных классов случайных процессов, в том числе и для процессов, характеризующихся смесью экспоненциальных распределений. В частности, в работе [4] алгоритм со специальными сглаживающими фильтрами применен для процессов ММРР (Markov-Modulated Poisson Process), в работах [5–7] алгоритм использован для аппроксимации широкого класса случайных процессов, наблюдаемых в дискретные моменты времени, марковским процессом (Markovian Arrival Process). Обзор различных модификаций ЕМ-алгоритма проведен в работе [8].

Положительные стороны данных алгоритмов обусловлены тем, что в их основе лежит метод максимального правдоподобия, гарантирующий состоятельность и асимптотическую эффективность оценок. Однако в ряде работ отмечаются недостатки ЕМ-алгоритмов. А именно, сходимость ЕМ-алгоритма доказана при обязательном условии ограниченности логарифма функции правдоподобия. Также наличие большого числа локальных максимумов логарифма функции правдоподобия приводит к большой неустойчивости по отношению к начальному приближению и исходным данным [9]. Кроме того, итерационный характер алгоритмов приводит к большому количеству вычислений, объем которых возрастает при значительном числе параметров [10]. В связи с этим разрабатываются и альтернативные подходы к оцениванию параметров смесей распределений.

В работе [10] смесь экспоненциальных распределений используется для аппроксимации распределений с тяжелыми хвостами, которые характерны для длин файлов, времени удержания вызова,

длины сцен в видеопотоках MPEG и интервалов между запросами на подключение в интернет-трафике. Авторы показывают принципиальную возможность аппроксимации распределения с тяжелыми хвостами гиперэкспоненциальным распределением и предлагают рекурсивный алгоритм подгонки. Сначала оценивается минимальная интенсивность, которая обеспечивает наибольшие значения случайной величины, а значит, вносит наибольший вклад в хвост распределения; затем это распределение вычитается из смеси распределений и оценивается следующая интенсивность. Статья [11] развивает этот подход для аппроксимации распределения Парето при оценивании средней длительности процедуры кумулятивных сумм, для которой в результате получена формула в явном виде. В работе [12] авторы используют гиперэкспоненциальное распределение для описания сенсорных сетей. Для оценивания параметров предложено два оригинальных алгоритма, минимизирующих целевую функцию. Первый алгоритм состоит из двух стадий, на первой оцениваются параметры интенсивности, на второй – вероятности появления наблюдений с данными интенсивностями. Вероятности выражаются через интенсивности, что позволяет на каждой стадии сократить в два раза число оцениваемых параметров. Второй алгоритм оценивает все параметры одновременно.

В данной работе рассматривается задача оценивания параметров гиперэкспоненциального распределения с двумя значениями интенсивности. Получены явные формулы для оценок по методу моментов. Предложены комбинированные оценки параметров, в которых в функции правдоподобия используются два первых выборочных момента. Показана также возможность оценивания минимального параметра интенсивности в общей модели гиперэкспоненциального распределения. Отметим, что этот параметр играет решающую роль при аппроксимации распределений с тяжелыми хвостами гиперэкспоненциальным распределением. Приведены результаты численного моделирования.

1. Постановка задачи

Рассмотрим задачу оценивания параметров гиперэкспоненциального распределения. Плотность гиперэкспоненциального распределения имеет вид:

$$f(x) = \sum_{i=1}^n p_i \lambda_i e^{-\lambda_i x}. \quad (1)$$

Предполагается, что параметры удовлетворяют следующим условиям:

$$p_i > 0, \quad \sum_{i=1}^n p_i = 1, \quad 0 < \lambda_1 < \lambda_2 < \dots < \lambda_n. \quad (2)$$

Задача состоит в построении оценок оценивания параметров $\{p_i, \lambda_i\}_{1 \leq i \leq n}$ по выборке $\{\tau_i\}_{1 \leq i \leq N}$ из распределения.

2. Оценки параметров для $n = 2$ по методу моментов

Рассмотрим сначала задачу оценивания параметров для случая $n = 2$. При этом, поскольку $p_1 + p_2 = 1$, требуется оценить три параметра: $\{\lambda_1, \lambda_2, p\}$, где $p = p_1$. Получим явные формулы, используя метод моментов. Для этого требуется найти три первых момента случайной величины. Для сокращения записи введем следующие обозначения:

$$a = \frac{1}{\lambda_1}, \quad b = \frac{1}{\lambda_2}. \quad (3)$$

Общая формула для начального момента в этом случае имеет вид:

$$EX^k = k! \left(\frac{p}{\lambda_1^k} + \frac{(1-p)}{\lambda_2^k} \right) = k! (pa^k + (1-p)b^k). \quad (4)$$

Введем следующие обозначения:

$$\begin{cases} m_1 = \sum_{k=1}^N \tau_k; \\ m_2 = \frac{1}{2} \sum_{k=1}^N \tau_k^2; \\ m_3 = \frac{1}{6} \sum_{k=1}^N \tau_k^3. \end{cases} \quad (5)$$

Приравнявая величины $\{m_1, m_2, m_3\}$ к их средним значениям, приходим к системе уравнений

$$\begin{cases} m_1 = pa + (1-p)b; \\ m_2 = pa^2 + (1-p)b^2; \\ m_3 = pa^3 + (1-p)b^3. \end{cases} \quad (6)$$

Преобразовав систему, получаем:

$$\begin{cases} m_1 - a = (1-p)(b-a); \\ m_2 - m_1 a = (1-p)b(b-a); \\ m_3 - m_2 a = (1-p)b^2(b-a). \end{cases}$$

Разделив второе уравнение на первое и третье на второе, исключаем из системы $(1-p)$:

$$\begin{cases} \frac{m_2 - m_1 a}{m_1 - a} = b; \\ \frac{m_3 - m_2 a}{m_2 - m_1 a} = b. \end{cases}$$

Приравнявая левые части уравнений, получаем квадратное уравнение для нахождения a :

$$\begin{aligned} (m_2 - m_1 a)^2 &= (m_1 - a)(m_3 - m_2 a); \\ (m_2 - m_1^2)a^2 - (m_3 - m_1 m_2)a + (m_1 m_3 - m_2^2) &= 0. \end{aligned} \quad (7)$$

Последовательно исключая из системы (6) p и a , получаем аналогично уравнение для нахождения b . Таким образом, оценки параметров $\{a, b\}$ являются корнями квадратного уравнения

$$(m_2 - m_1^2)x^2 - (m_3 - m_1 m_2)x + (m_1 m_3 - m_2^2) = 0. \quad (8)$$

Учитывая обозначения (3), получаем, что оценки $\{\hat{\lambda}_1, \hat{\lambda}_2\}$ параметров $\{\lambda_1, \lambda_2\}$ являются корнями квадратного уравнения

$$(m_1 m_3 - m_2^2)x^2 - (m_3 - m_1 m_2)x + (m_2 - m_1^2) = 0. \quad (9)$$

Оценку параметра p получаем из первого уравнения системы (6):

$$\hat{p} = \frac{\hat{\lambda}_1(m_1 \hat{\lambda}_2 - 1)}{\hat{\lambda}_2 - \hat{\lambda}_1}. \quad (10)$$

Заметим, что оценки по методу моментов параметров гиперэкспоненциального распределения могут быть построены, если существуют и положительны корни квадратного уравнения (8) и оценка $\hat{p}$ в (10) также положительна.

3. Комбинированные оценки параметров для $n = 2$

Построим оценки параметров процесса (1) для $n = 2$, используя метод моментов и метод максимального правдоподобия. Сначала, используя два первых момента, выразим величины λ_1 и λ_2 через параметр p . Далее оценку параметра p найдем по методу максимального правдоподобия.

Уравнения для первых двух моментов имеют вид:

$$\begin{cases} m_1 = \frac{p}{\lambda_1} + \frac{1-p}{\lambda_2}; \\ m_2 = \frac{p}{\lambda_1^2} + \frac{1-p}{\lambda_2^2}. \end{cases} \quad (11)$$

Введем обозначения:

$$x = \frac{p}{\lambda_1}, y = \frac{1-p}{\lambda_2};$$

с их учетом уравнения (11) примут следующий вид:

$$\begin{cases} m_1 = x + y; \\ m_2 = \frac{x^2}{p} + \frac{y^2}{1-p}. \end{cases} \quad (12)$$

Выразив y из первого уравнения и подставив во второе, получим квадратное уравнение для нахождения x :

$$x^2 - 2m_1px + p(m_1^2 - (1-p)m_2) = 0. \quad (13)$$

Корни этого уравнения определяются равенствами

$$x_{1,2} = pm_1 \pm \sqrt{p(1-p)}\sqrt{m_2 - m_1^2}. \quad (14)$$

Для существования корней квадратного уравнения (14) необходимо выполнение условия

$$m_2 - m_1^2 \geq 0. \quad (15)$$

Выбрав в качестве оценки величины x больший корень, получим

$$\begin{aligned} x &= pm_1 + \sqrt{p(1-p)}\sqrt{m_2 - m_1^2}; \\ \hat{\lambda}_1 &= \frac{p}{x} = \left(m_1 + \sqrt{\frac{1-p}{p}}\sqrt{m_2 - m_1^2} \right)^{-1}; \\ y &= m_1 - x = (1-p)m_1 - \sqrt{p(1-p)}\sqrt{m_2 - m_1^2}; \\ \hat{\lambda}_2 &= \frac{1-p}{y} = \left(m_1 - \sqrt{\frac{p}{1-p}}\sqrt{m_2 - m_1^2} \right)^{-1}. \end{aligned} \quad (16)$$

Заметим, что $\hat{\lambda}_1 \leq \hat{\lambda}_2$. Если выбрать в качестве оценки для x меньший из корней (14), получим аналогичные формулы для случая $\hat{\lambda}_1 \geq \hat{\lambda}_2$.

Оценка $\hat{\lambda}_2$ в (16) является положительной, если вместе с условием (15) выполняется условие

$$m_1 - \sqrt{\frac{p}{1-p}}\sqrt{m_2 - m_1^2} > 0,$$

что возможно, если

$$m_1^2 > \frac{p}{1-p}(m_2 - m_1^2),$$

откуда следует ограничение на параметр p :

$$p < \frac{m_1^2}{m_2}. \quad (17)$$

Для оценки параметра p используем метод максимального правдоподобия. Запишем логарифм функции правдоподобия:

$$\log L(\tau; \lambda_1, \lambda_2, p) = \sum_{k=1}^N \log(p\lambda_1 e^{-\lambda_1 \tau_k} + (1-p)\lambda_2 e^{-\lambda_2 \tau_k}). \quad (18)$$

Подставляя вместо параметров λ_1, λ_2 их оценки (16) и вычисляя логарифм функции правдоподобия с приемлемым шагом для $p \in (0; m_1^2/m_2)$, получим оценку $\hat{p}$, доставляющую максимум функции (18); эту оценку затем используем в (16).

4. Оценка λ_1 для произвольного числа параметров распределения

Построим теперь оценку минимального параметра интенсивности для гиперэкспоненциального распределения с произвольным числом параметров, определяемого уравнениями (1), (2). Предлагается использовать построенные выше оценки параметров процесса с двумя интенсивностями. При этом оценка λ_1 определяется по формулам (5), (9) для метода моментов и по формулам (5), (16), (18) для комбинированной оценки.

Исследуем асимптотические (при стремлении объема наблюдений N к бесконечности) свойства оценки по методу моментов параметра λ_1 . Введем обозначения

$$a_i = \frac{1}{\lambda_i}, \quad a_1 > a_2 > \dots > a_n > 0. \quad (19)$$

При $N \rightarrow \infty$ выборочные моменты сходятся (по усиленному закону больших чисел) с вероятностью единица к соответствующим теоретическим значениям. Приравнявая три первых выборочных начальных момента теоретическим, получим систему уравнений:

$$\begin{cases} m_1 = \sum_{i=1}^n p_i a_i; \\ m_2 = \sum_{i=1}^n p_i a_i^2; \\ m_3 = \sum_{i=1}^n p_i a_i^3. \end{cases} \quad (20)$$

Найдем коэффициенты уравнения (9):

$$\begin{aligned} m_2 - m_1^2 &= \sum_{i=1}^n p_i a_i^2 \sum_{j=1}^n p_j - \left(\sum_{j=1}^n p_j a_j \right)^2 = \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i^2 - 2a_i a_j + a_j^2)^2 = \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i - a_j)^2; \\ m_3 - m_1 m_2 &= \sum_{i=1}^n p_i a_i^3 \sum_{j=1}^n p_j - \sum_{i=1}^n p_i a_i \sum_{j=1}^n p_j a_j^2 = \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i^3 - a_j^2 a_i + a_j^3 - a_i^2 a_j) = \\ &= \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i + a_j)(a_i - a_j)^2; \\ m_1 m_3 - m_2^2 &= \sum_{i=1}^n p_i a_i \sum_{j=1}^n p_j a_j^3 - \sum_{i=1}^n p_i a_i^2 \sum_{j=1}^n p_j a_j^2 = \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (-2a_i^2 a_j^2 + a_i^3 a_j + a_i a_j^3) = \\ &= \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j a_i a_j (a_i - a_j)^2. \end{aligned}$$

Уравнение (9) принимает следующий вид:

$$F(x) = \sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i - a_j)^2 a_i a_j f_{ij}(x) = 0, \quad (21)$$

$$f_{ij}(x) = x^2 - (\lambda_i + \lambda_j)x + \lambda_i \lambda_j = (x - \lambda_i)(x - \lambda_j).$$

Заметим, что корнями квадратного уравнения $f_{ij}(x) = 0$ являются величины λ_1 и λ_2 . При этом в силу условий $\lambda_1 < \lambda_i$ справедливы неравенства $f_{ij}(\lambda_1) \geq 0$, $f'_{ij}(\lambda_1) \leq 0$. Обозначим через X меньший корень уравнения (21). Найдем величину $X - \lambda_1$ смещения оценки $\hat{\lambda}_1$. Используя приближенное равенство

$$0 = F(X) \approx F(\lambda_1) + F'(\lambda_1)(X - \lambda_1),$$

получаем

$$X - \lambda_1 = -\frac{F(\lambda_1)}{F'(\lambda_1)} = -\frac{\sum_{i=2}^n \sum_{j=i+1}^n p_i p_j (a_i - a_j)^2 a_i a_j (\lambda_1 - \lambda_i)(\lambda_1 - \lambda_j)}{\sum_{i=1}^n \sum_{j=i+1}^n p_i p_j (a_i - a_j)^2 a_i a_j (\lambda_1 - \lambda_i + \lambda_1 - \lambda_j)}.$$

Поскольку величина $X - \lambda_1 > 0$, оценка $\hat{\lambda}_1$ является асимптотически смещенной, причем смещение положительно. Величина смещения будет исследована при моделировании.

5. Результаты численного эксперимента

Для изучения свойств оценок были проведены численные эксперименты, результаты которых для $n = 2$ представлены в табл. 1. Здесь $\{\lambda_1, \lambda_2\}$ – интенсивности процесса (1), $p = p_1$ – вероятность появления наблюдения с интенсивностью λ_1 ; через N обозначено число наблюдений. Далее в табл. 1 представлены результаты эксперимента: $\{d_1, d_2, d_p\}$ – среднеквадратические отклонения оценок от истинных значений параметров $\{\lambda_1, \lambda_2, p\}$ соответственно, K – количество реализаций, в которых не удалось построить оценки. Для каждого набора параметров генерировалось 1 000 реализаций процесса. Если условия существования оценок нарушались, соответствующая реализация не учитывалась при построении оценок.

Таблица 1

Результаты численного моделирования для $n = 2$

Параметры процесса			Число наблюдений	Комбинированные оценки				Оценки по методу моментов			
λ_1	λ_2	p	N	d_1	d_2	d_p	K	d_1	d_2	d_p	K
1	2	0,2	2 000	0,0539	0,2533	0,0416	1	0,0616	0,6970	0,0626	7
1	2	0,5	2 000	0,0185	0,2846	0,0381	0	0,0224	5,8567	0,0544	17
1	2	0,8	2 000	0,0235	1,2016	0,0888	20	0,0292	24,634	0,1111	247
1	4	0,2	2 000	0,0122	0,0447	0,0013	0	0,0247	0,1894	0,0043	0
1	4	0,8	2 000	0,0020	0,7153	0,0038	0	0,0067	69,520	0,0192	190
1	2	0,2	3 000	0,0367	0,0862	0,0263	0	0,0433	0,1539	0,0382	1
1	2	0,8	3 000	0,0147	1,6827	0,0627	7	0,0170	11,971	0,0780	199
1	3	0,2	3 000	0,0116	0,0260	0,0024	0	0,0203	0,0682	0,0054	0
1	3	0,5	3 000	0,0045	0,0852	0,0044	0	0,0091	0,6619	0,0121	0
1	3	0,8	3 000	0,0030	0,5943	0,0093	0	0,0080	31,696	0,0283	122
1	4	0,2	3 000	0,0076	0,0275	0,0009	0	0,0163	0,1221	0,0029	0
1	4	0,5	3 000	0,0029	0,0963	0,0017	0	0,0072	1,3881	0,0072	0
1	4	0,8	3 000	0,0014	0,5237	0,0023	0	0,0046	61,647	0,0142	124
1	2	0,2	5 000	0,0235	0,0343	0,0166	0	0,0286	0,0627	0,0244	0
1	2	0,5	5 000	0,0078	0,0809	0,0187	0	0,0105	0,7009	0,0273	0
1	2	0,8	5 000	0,0076	0,8881	0,0403	1	0,0108	8,3468	0,0581	133
1	3	0,2	5 000	0,0072	0,0145	0,0014	0	0,0132	0,0401	0,0034	0
1	3	0,5	5 000	0,0025	0,0490	0,0024	0	0,0057	0,2868	0,0077	0
1	3	0,8	5 000	0,0015	0,3540	0,0043	0	0,0036	31,621	0,0146	60
1	4	0,2	5 000	0,0044	0,0179	0,0005	0	0,0091	0,7935	0,0018	0
1	4	0,5	5 000	0,0017	0,0553	0,0010	0	0,0042	0,6718	0,0017	0
1	4	0,8	5 000	0,0009	0,3698	0,0014	0	0,0025	39,019	0,0077	74

Анализ результатов показывает, что оценки по методу моментов параметра λ_2 имеют значительное отклонение от истинного значения при величине $p = 0,8$. При этом количество реализаций,

в которых невозможно оценить параметры по методу моментов, достигает 24%. Вычисление оценок также не проводилось, если коэффициент при x^2 в уравнении (9) оказывался по модулю меньше 0,001, поскольку это приводит к большим отклонениям оценки λ_2 от истинного значения. Оценки параметра λ_1 оказываются более устойчивыми.

Комбинированная оценка не настолько сильно зависит от значения вероятности p ; число реализаций, в которых не удастся построить оценки, не превышает 2%. Следует отметить, что несмотря на использование метода максимального правдоподобия, оценка свободна от недостатков ЕМ-метода, поскольку метод максимального правдоподобия используется только для оценивания вероятности p , которая принадлежит ограниченному интервалу. Максимум функции правдоподобия нетрудно найти, вычисляя значение функции на этом интервале с желаемым шагом. При этом использования итерационных методов не требуется.

Во всех случаях точность оценки параметра λ_1 оказалась выше, чем точность оценки параметра λ_2 как для комбинированных оценок, так и для метода моментов. При этом комбинированные оценки оказываются предпочтительнее оценок по методу моментов в смысле среднего квадрата отклонения от истинных значений.

В табл. 2 представлены результаты моделирования оценки минимального параметра интенсивности λ_1 для $n = 4$, параметры интенсивностей выбирались равными $\{1, 3, 5, 8\}$; в табл. 3 представлены результаты моделирования для $n = 6$, параметры интенсивностей выбирались равными $\{1, 3, 5, 8, 15, 20\}$. Таблицы 2, 3 содержат число наблюдений в реализации и Δ – теоретическое смещение оценки. Для каждой оценки представлены d – среднеквадратическое отклонение оценки от истинного значения параметра, Δ_1 – выборочное смещение оценки. Для каждого набора параметров генерировалось 1 000 реализаций процесса.

Таблица 2

Результаты численного моделирования для $n = 4$

Параметры процесса				Число наблюдений	Смещение	Комбинированные оценки		Оценки по методу моментов	
p_1	p_2	p_3	p_4	N	Δ	d	Δ_1	d	Δ_1
0,2	0,3	0,2	0,3	2 000	0,0383	0,0244	0,1144	0,0276	0,0789
0,2	0,3	0,2	0,3	3 000	0,0383	0,0185	0,1063	0,0187	0,0573
0,2	0,3	0,2	0,3	5 000	0,0383	0,0153	0,1045	0,0119	0,0541
0,4	0,3	0,2	0,1	2 000	0,0135	0,0072	0,0398	0,0135	0,0251
0,4	0,3	0,2	0,1	3 000	0,0135	0,0053	0,0403	0,0094	0,0275
0,4	0,3	0,2	0,1	5 000	0,0135	0,0036	0,0380	0,0055	0,0200

Таблица 3

Результаты численного моделирования для $n = 6$

Параметры процесса						Число наблюдений	Смещение	Комбинированные оценки		Оценки по методу моментов	
p_1	p_2	p_3	p_4	p_5	p_6	N	Δ	d	Δ_1	d	Δ_1
0,2	0,1	0,2	0,1	0,2	0,2	2000	0,0350	0,0265	0,1333	0,0217	0,0615
0,2	0,1	0,2	0,1	0,2	0,2	3000	0,0350	0,0223	0,1310	0,0140	0,0532
0,2	0,1	0,2	0,1	0,2	0,2	5000	0,0350	0,0201	0,1285	0,0104	0,0477
0,4	0,1	0,2	0,1	0,1	0,1	2000	0,0146	0,0074	0,0592	0,0101	0,0273
0,4	0,1	0,2	0,1	0,1	0,1	3000	0,0146	0,0061	0,0583	0,0071	0,0268
0,4	0,1	0,2	0,1	0,1	0,1	5000	0,0146	0,0046	0,0550	0,0042	0,0196

Анализ результатов показывает, что комбинированные оценки характеризуются меньшим среднеквадратичным отклонением, чем оценки по методу моментов; однако их смещение больше. В целом результаты показывают хорошее качество оценивания для каждого метода. Для всех реализаций процесса удалось построить оценки.

Заключение

В работе получены явные формулы для оценок по методу моментов параметров гиперэкспоненциального распределения с двумя значениями интенсивности. Предложена комбинированная оценка параметров такого распределения, использующая только два первых выборочных момента в сочетании с методом максимального правдоподобия. Установлена возможность оценивания минимального параметра интенсивности для произвольного гиперэкспоненциального распределения. Результаты моделирования показывают, что при оценивании параметров распределения с двумя значениями интенсивности оценки по методу моментов являются неустойчивыми и во многих случаях не могут быть построены. Комбинированные оценки оказываются более предпочтительными. При оценивании минимального параметра интенсивности качество комбинированных оценок и оценок по методу моментов оказывается близким.

Список источников

1. Chen J., Gupta A.K. Parametric statistical change point analysis with applications to genetics, medicine, and finance. 2nd ed. Springer Science + Business Media, LLC. 2012.
2. Dempster A.P., Laird N.M., Rubin D.B. Maximum likelihood from incomplete data via the EM algorithm // *Journal of the Royal Statistical Society. Series B.* 1977. V. 39 (1). P. 1–38.
3. Baum L.E., Petrie T., Soules G., Weiss N. A maximization technique occurring in the statistical analysis of probabilistic function of Markov chains // *Annals of Mathematical Statistics.* 1970. V. 41 (1). P. 164–171.
4. Elliott R.J., Malcolm W.P. Discrete time expectation maximization algorithms for markov modulated poisson processes // *IEEE Transactions on Automatic Control.* 2002. V. 53. P. 247–256.
5. Breuer L., Kume A. An EM algorithm for markovian arrival processes observed at discrete times // *Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance. MMB&DFT / B. Müller-Clostermann, K. Echtle, E.P. Rathgeb (eds.). Berlin ; Heidelberg : Springer, 2010. P. 242–258. (Lecture Notes in Computer Science, v. 5987). doi: 10.1007/978-3-642-12104-3_19*
6. Horváth G., Okamura H. A fast EM algorithm for fitting marked markovian arrival processes with a new special structure // *Computer Performance Engineering. EPEW 2013 / M.S. Balsamo, W.J. Knottenbelt, A. Marin (eds.). Berlin ; Heidelberg : Springer, 2013. P. 119–133. (Lecture Notes in Computer Science; v. 8168). doi: 10.1007/978-3-642-40725-3_10*
7. Bražėnas M., Horváth G., Telek M. Efficient implementations of the EM-Algorithm for transient markovian arrival processes // *Analytical and Stochastic Modelling Techniques and Applications. ASMTA / S. Wittevrongel, T. Phung-Duc (eds.). Cham : Springer, 2016. P. 107–122. (Lecture Notes in Computer Science; v. 9845). doi: 10.1007/978-3-319-43904-4_8*
8. Королев В.Ю. EM-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений : теоретический обзор. М. : Изд-во ИПИ РАН, 2007.
9. Королев В.Ю., Назаров А.Л. Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // *Автоматика и телемеханика.* 2010. № 3. С. 98–116.
10. Feldmann A., Whitt W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models // *Performance Evaluation.* 1998. V. 31. P. 245–279.
11. Petcharat K., Areepong Yu., Sukparungsee S., Mititelu G. Fitting Pareto Distribution with Hyperexponential to Evaluate the ARL for CUSUM Chart // *International Journal of Pure and Applied Mathematics.* 2012. V. 77 (2). P. 233–244.
12. Singh L.N., Dattatreya G.R. Estimation of the Hyperexponential Density with Applications in Sensor Networks // *International Journal of Distributed Sensor Networks.* 2002. V. 3. P. 311–330.
13. Guillo A., Loisel S., Stupfler G. Estimating the parameters of a seasonal Markov-modulated Poisson process // *Statistical Methodology.* Elsevier, 2015. V. 26. P. 103–123.
14. Gunsta M., Knapika B., Mandjes B., Solleia B. Parameter estimation for a discretely observed population process under Markov-modulation // *Computational Statistics & Data Analysis.* 2019. V. 140. P. 88–103. doi: 10.1016/j.csda.2019.06.008

References

1. Chen, J. & Gupta, A.K. (2012) *Parametric Statistical Change Point Analysis With Applications to Genetics, Medicine, and Finance*. 2nd ed. Springer Science+Business Media, LLC.
2. Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B.* 39(1). pp. 1–38.
3. Baum, L. E., Petrie, T., Soules, G. & Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic function of Markov chains. *Annals of Mathematical Statistics.* 41(1). pp. 164–171.
4. Elliott, R.J. & Malcolm, W.P. (2008) Discrete Time Expectation Maximization Algorithms for Markov Modulated Poisson Processes. *IEEE Transactions on Automatic Control.* 53. pp. 247–256.

5. Breuer, L. & Kume, A. (2010) An EM Algorithm for Markovian Arrival Processes Observed at Discrete Times. In: Müller-Clostermann, B., Ehtle, K. & Rathgeb, E.P. (eds) *Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance. MMB&DFT*. Berlin; Heidelberg: Springer. pp. 242–258. DOI: 10.1007/978-3-642-12104-3_19
6. Horváth, G. & Okamura, H. (2013) A Fast EM Algorithm for Fitting Marked Markovian Arrival Processes with a New Special Structure. In: Balsamo, M.S., Knottenbelt, W.J. & Marin, A. (eds) *Computer Performance Engineering. EPEW 2013*. Berlin; Heidelberg: Springer. pp. 119–133. DOI: 10.1007/978-3-642-40725-3_10
7. Bražėnas, M., Horváth, G. & Telek, M. (2016) Efficient Implementations of the EM-Algorithm for Transient Markovian Arrival Processes. In: Wittevrongel, S. & Phung-Duc, T. (eds) *Analytical and Stochastic Modelling Techniques and Applications. ASMTA*. vol 9845. Springer, Cham. pp. 107–122. DOI: 10.1007/978-3-319-43904-4_8
8. Korolev, V.Yu. (2007) *EM-algorithm, ego modifikatsii i ikh primeneniye k zadache razdeleniya smesey veroyatnostnykh raspredeleniy: teoreticheskiy obzor* [EM-algorithm, its modifications and their application to the problem of separation of mixtures of probability distributions. Theoretical review]. Moscow: IPI RAS.
9. Korolev, V.Yu. & Nazarov, A.L. (2010) Razdeleniye smesey veroyatnostnykh raspredeleniy pri pomoshchi setochnykh metodov mo-mentov i maksimal'nogo pravdopodobiya [Separation of mixtures of probability distributions using grid methods of moments and maximum likelihood]. *Avtomatika i telemekhanika*. 3. pp. 98–116.
10. Feldmann, A. & Whitt, W. (1998) Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*. 31. pp. 245–279.
11. Petcharat, K., Areepong, Yu., Sukparungsee, S. & Mititelu, G. (2012) Fitting Pareto Distribution with Hyperexponential to Evaluate the ARL for CUSUM Chart. *International Journal of Pure and Applied Mathematics*. 77(2). pp. 233–244
12. Singh, L.N. & Dattatreya, G.R. (2002). Estimation of the Hyperexponential Density with Applications in Sensor Networks. *International Journal of Distributed Sensor Networks*. 3. pp. 311–330.
13. Guillo, A., Loisel, S. & Stupfler, G. (2015). Estimating the parameters of a seasonal Markov-modulated Poisson process. *Statistical Methodology*. 26. pp. 103–123.
14. Gunsta, M., Knapika, B., Mandjes, M. & Solleia, B. (2019) Parameter estimation for a discretely observed population process under Markov-modulation. *Computational Statistics & Data Analysis*. 140. pp. 88–103. DOI: 10.1016/j.csda.2019.06.008

Информация об авторах:

Буркатовская Юлия Борисовна – доцент, кандидат физико-математических наук, доцент отделения информационных технологий инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Томск, Россия); доцент кафедры системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: tracey@tpu.ru

Воробейчиков Сергей Эрикович – доктор физико-математических наук, профессор кафедры системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: sev@mail.tsu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Burkatovskaya Yulia B. (Candidate of Physical and Mathematical Sciences, Associate Professor, National Research Tomsk Polytechnic University, Tomsk, Russian Federation; National Research Tomsk State University, Tomsk, Russian Federation). E-mail: tracey@tpu.ru

Vorobeychikov Sergey E. (Doctor Physical and Mathematical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: sev@mail.tsu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 10.11.2024; принята к публикации 03.03.2025

Received 10.11.2024; accepted for publication 03.03.2025

Научная статья

УДК 519.2

doi: 10.17223/19988605/70/6

Статистические свойства оценок метода измерения ценовой чувствительности PSM с учетом S^a -симметрии**Жанна Николаевна Зенкова¹, Анна Андреевна Андриевская², Никита Иванович Яковлев³**^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Россия*³ *Tele2/Altel, Алматы, Казахстан*¹ *zhanna.zenkova@mail.tsu.ru*² *anna.andriev@mail.tsu.ru*³ *nyakovlev19@gmail.com*

Аннотация. Предлагается модификация метода PSM с использованием дополнительной информации об S^a -симметрии кривых ценовой чувствительности. Показаны несмещенность, состоятельность, асимптотическая нормальность новых оценок; доказано, что асимптотические дисперсии модифицированных оценок меньше, чем обычных, что говорит об улучшении точности метода при достаточно большом числе опрошенных. Получены доверительные интервалы для цен с учетом S^a -симметрии.

Ключевые слова: метод измерения ценовой чувствительности PSM – Price Sensitivity Meter; дополнительная информация; симметрия; асимптотическая нормальность; доверительный интервал.

Для цитирования: Зенкова Ж.Н., Андриевская А.А., Яковлев Н.И. Статистические свойства оценок метода измерения ценовой чувствительности PSM с учетом S^a -симметрии // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 62–72. doi: 10.17223/19988605/70/6

Original article

doi: 10.17223/19988605/70/6

Statistical properties of PSM – price sensitivity meter technique with S^a -symmetry**Zhanna N. Zenkova¹, Anna A. Andrievskaya², Nikita I. Yakovlev³**^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*³ *Tele2/Altel, Almaty, Kazakhstan*¹ *zhanna.zenkova@mail.tsu.ru*² *anna.andriev@mail.tsu.ru*³ *nyakovlev19@gmail.com*

Abstract. In the paper we modified Price Sensitivity Meter (PSM) technique using additional information on S^a -symmetry of price sensitivity curves. New price estimates are unbiased, consistent, asymptotically normal, their variances are smaller than those of the usual ones, so new estimators are more accurate. Confidence intervals for prices are obtained taking into account S^a -symmetry.

Keywords: Price Sensitivity Meter (PSM) technique; additional information; symmetry; asymptotic normality; confident interval.

For citation: Zenkova, Z.N., Andrievskaya, A.A., Yakovlev, N.I. (2025) Statistical properties of PSM – price sensitivity meter technique with S^a -symmetry. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 62–72. doi: 10.17223/19988605/70/6

Введение

Проблема установления цены на товар или услугу является фундаментальной для любого предприятия [1, 2]. Цена должна не только компенсировать понесенные предприятием затраты на производство, но и обеспечивать прибыль, она непосредственно определяет уровень спроса на товар, т.е., прежде всего, влияет на процесс принятия решения о совершении покупки [3, 4]. Именно поэтому важно, чтобы цена казалась целевому потребителю справедливой и учитывала его предпочтения [5]. В этом случае наиболее эффективными являются методы ценообразования, ориентированные на спрос. Такие методы помогают вычислить диапазоны цен, которые помогут не только найти приемлемую для целевого потребителя цену товара, но и учитывать стратегические цели фирмы [6]. Они используются для ценообразования на товары-новинки [7–9], о спросе на которые нет вообще никакой информации (например, о наблюдавшихся ранее объемах спроса при разных ценах, о реакции потребителей на маркетинговые акции, связанные с изменениями цен, и пр.). Эти методы также позволяют определить, как эффективно провести ценовую дискриминацию, и найти, по каким ценам продавать функционально один и тот же товар представителям разных целевых аудиторий. Например, стиральный порошок по составу и себестоимости может быть практически одинаков, но, будучи по-разному упакован и назван, может быть продан по разным ценам покупателям с разным уровнем дохода [10]. Именно эти разные цены помогут найти методы измерения ценовой чувствительности.

К методам ценообразования, ориентированным на спрос, относятся такие методы, как метод Штоцеля [11], метод ван Вестендорпа и метод определения ценовой чувствительности PSM (Price Sencitivity Metter) [12], метод RBR – Randomized Buying Response – ответы на вопросы о покупке по рандомизированным ценам [13], модифицированный метод ван Вестендорпа – PSM для интервально цензурированных данных [14–15]. Все эти методы основаны на опросах представителей целевой аудитории, которые могут адекватно оценить ценность определенного товара.

Так как значимость данных методов ценообразования весьма важна для осуществления качественной маркетинговой деятельности, подходы к оцениванию, а также вопросы измерения и повышения точности оценок очень актуальны. В работах Стана Липовецкого [16–18] рассматривались вопросы измерения ценовой чувствительности, в [16] были использованы системы линейных дифференциальных уравнений Чепмена–Колмогорова, которые позволили представить кривые ценовой чувствительности метода PSM как логистические функции распределения и свести задачу к оцениванию их коэффициентов с помощью регрессионных моделей. В [19] данный подход применялся для поиска ценовых значений в методе Штоцеля, точность оценивания метода Штоцеля исследовалась в [8] с помощью имитационного моделирования. В [20] анализировались статистические свойства оценок метода PSM, показаны их несмещенность, состоятельность и асимптотическая нормальность, выведена формула для асимптотических дисперсий оценок, которая позволяет судить об их точности, найдены приближенные доверительные интервалы для цен. В [21] показано, что малое число участников исследования по методу PSM негативно влияет на точность при наличии цензурирования, особенно на оценку минимально возможной цены.

В ряде работ были предприняты успешные попытки повысить точность оценивания методов за счет дополнительной информации. В [22] показано существенное влияние симметрии на повышение точности оценивания метода PSM для случая линейной зависимости кривых ценовой чувствительности; в [23, 24] привлекалась информация о знании квантилей кривых ценовой чувствительности, показано, что максимальный выигрыш дает учет медиан.

В настоящей работе рассматривалось влияние дополнительной информации об S^a -симметрии кривых ценовой чувствительности [25, 26]. На основе результатов [20] показано, что модифицированный метод дает оценки, которые являются несмещенными, состоятельными и асимптотически нормальными. Аналитически доказано, что асимптотические дисперсии новых оценок меньше, чем асимптотические дисперсии оценок, полученных стандартным методом PSM, т.е. привлечение дополнительной информации привело к увеличению точности метода. С помощью имитационного моделирования были подтверждены выводы о точности методов, показано, что сходимость к нормаль-

ному распределению наблюдается для случаев, когда число опрошенных составляет от 30 человек, при этом выигрыш в точности имеет место и для существенно меньших объемов выборок. Асимптотическая нормальность позволила найти приближенные доверительные интервалы для ценовых значений модифицированного с учетом S^a -симметрии метода PSM.

1. Метод измерения ценовой чувствительности Price Sensitivity Meter (PSM)

Для проведения исследования N целевым потребителям предварительно предлагают воспользоваться рассматриваемым товаром в течение некоторого времени, что дает им возможность объективно оценить товар и его значимость. Затем всем участникам независимо друг от друга задаются четыре вопроса [12]:

1. Ниже какого уровня цены x_1 начинают возникать подозрения о том, что товар некачественный или поддельный?
2. Какая цена x_2 является приемлемой для покупки?
3. Какая цена x_3 кажется высокой, однако вопрос о покупке еще рассматривается?
4. Начиная с какого значения цены x_4 товар кажется настолько дорогим, что вопрос о покупке даже не ставится?

В [20] исследуемые уровни цен рассматривались как случайные величины (с.в.) X_j $j = \overline{1, 4}$, с соответствующими функциями распределения (ф.р.) $F_j(x)$; результаты опроса – это четырехмерная выборка $(X_{k1}, X_{k2}, X_{k3}, X_{k4})$, $k = \overline{1, N}$. Особенности организации эксперимента обуславливают независимость выборочных элементов между собой.

Согласно методике PSM по выборочному вектору $\{X_{1j}, X_{2j}, \dots, X_{Nj}\}$ с.в. X_j $j = \overline{1, 4}$, строится эмпирическая функция распределения (э.ф.р.) $\hat{F}_j(x)$ согласно формуле [26]

$$\hat{F}_j(x) = \frac{1}{N} \sum_{k=1}^N I(X_{kj} < x), \quad (1)$$

где $I(\cdot)$ – индикаторная функция. Для $i = 1, 3$ строится эмпирическая оценка функции выживания (э.ф.в.)

$$\hat{S}_i(x) = 1 - \hat{F}_i(x), \quad (2)$$

Функции (1) и (2) в классическом варианте метода PSM сглаживаются, их называют *кривыми ценовой чувствительности*. Цены определяются как абсциссы пересечения этих кривых:

- 1) $\hat{F}_2(x)$ и $\hat{S}_1(x)$ – минимально возможная цена $\hat{x}^{(21)} = \hat{x}_{\min}$, ниже которой цену продажи назначать не рекомендуется, так как мы потеряем большое количество потребителей, поскольку слишком низкая цена вызывает сомнения в качестве товара;
- 2) $\hat{F}_4(x)$ и $\hat{S}_1(x)$ – «оптимальная цена» $\hat{x}^{(41)} = \hat{x}_{\text{опт}}$, рекомендована в случае, если стратегия фирмы – увеличение объема продаж (захват большей доли рынка);
- 3) $\hat{F}_2(x)$ и $\hat{S}_3(x)$ – «ожидаемая цена», или «точка безразличия», $\hat{x}^{(23)} = \hat{x}_{\text{безр}}$;
- 4) $\hat{F}_4(x)$ и $\hat{S}_3(x)$ – максимально возможная цена $\hat{x}^{(43)} = \hat{x}_{\max}$, выше которой цену назначать не рекомендуется, так как большинство целевой аудитории не может себе позволить этот товар из-за чрезмерной дороговизны.

В [20] цены рассчитывались как пересечения не сглаженных, а ступенчатых э.ф.р. и э.ф.в. и определялись формулами

$$\hat{x}^{(ij)} = \begin{cases} \max \{x : \hat{F}_j(x) = \hat{S}_i(x)\} \\ x : \hat{S}_i(x+0) \leq \hat{F}_j(x) < \hat{S}_i(x). \end{cases}$$

При этом для $i = 1, 3$ и $j = 2, 4$ истинное значение $x^{(ij)}$ – это абсцисса точки пересечения непрерывных теоретических кривых $F_i(x)$ и $S_j(x)$, которая в общем виде может быть получена из уравнения

$$F_j(x) = 1 - F_i(x) \quad (3)$$

что позволяет определять цены как медианы

$$\hat{x}_{0,5}^{(ij)} = \hat{G}_{ij}^{-1}(0,5) = \sup \{x : \hat{G}_{ij}(x) \leq 0,5\}$$

эмпирической оценки функции распределения

$$G_{ij}(x) = (F_i(x) + F_j(x))/2. \quad (4)$$

В итоге, для $i = 1, 3$ и $j = 2, 4$ в качестве оценки $x^{(ij)}$ берется $(N+1)$ -я порядковая статистика $X_{(N+1)}^{(ij)}$ в объединенной выборке объема $2N$, составленной из выборок $\{X_{ki}\}$ и $\{X_{kj}\}$, $k = \overline{1, N}$, т.е. $\hat{x}_{0,5}^{(ij)} = X_{(N+1)}^{(ij)}$. В [20] показаны асимптотическая нормальность и состоятельность оценок $\hat{x}_{0,5}^{(ij)}$. При этом

$$L\left(\sqrt{N}\left(\hat{x}_{0,5}^{(ij)} - x_{0,5}^{(ij)}\right)\right) \xrightarrow{N \rightarrow \infty} N\left(0; v_{ij}^2\right).$$

где

$$v_{ij}^2 = F_{ij}\left(x_{0,5}^{(ij)}, x_{0,5}^{(ij)}\right) / \left(2g_{ij}^2\left(x_{0,5}^{(ij)}\right)\right), \quad (5)$$

$F_{ij}(y, z)$ – совместная ф.р. с.в. X_i и X_j , $g_{ij}(x) = (f_i(x) + f_j(x))/2$ – плотность ф.р. $G_{ij}(x)$, при этом для $m = \overline{1, 4}$ существуют $f_m(x)$ – плотности ф.р. $F_m(x)$ и $g_m(x_{0,5}^{(ij)}) \neq 0$.

2. Учет S^a -симметрии в методе PSM

Рассмотрим случай, когда для $k = \overline{1, 4}$ непрерывные ф.р. $F_k(x)$ обладают свойством S^a -симметрии [25, 26]:

$$F_k(x) = 1 - F_k(S_k(x)), \quad (6)$$

где для $k = \overline{1, 4}$ $S_k(x)$ – непрерывные, монотонно убывающие функции, такие, что $S_k(\alpha_k) = \alpha_k$, $(S_k)^{-1}(x) = S_k(x)$, где $(S_k)^{-1}(x)$ – обратная к $S_k(x)$ функция, $F_k(\alpha_k) = 0,5$. При $S_k(x) = 2\alpha_k - x$ будем иметь обычную симметрию относительно центров α_k .

Для $k = \overline{1, 4}$ оценка ф.р. $F_k(x)$ с учетом свойства (6) строится по следующей формуле [25, 26]:

$$\hat{F}_k^S(x) = \frac{\hat{F}_k(x) + 1 - \hat{F}_k(S_k(x))}{2}. \quad (7)$$

Подставив оценки (7) в (4), получим

$$\hat{G}_{ij}^S(x) = \frac{\hat{F}_i(x) + \hat{F}_j(x) + 2 - \hat{F}_i(S_i(x)) - \hat{F}_j(S_j(x))}{4}, \quad (8)$$

следовательно, для $i = 1, 3$ и $j = 2, 4$ в качестве оценки $x^{(ij)}$ с учётом S^a -симметрии можно взять медиану $G_{ij}^S(x)$, т.е. $(2N+1)$ -ю порядковую статистику $X_{(2N+1)}^{S(ij)}$ в объединенной выборке объема $4N$, составленной из выборок $\{X_{ki}\}$, $\{X_{kj}\}$ и симметризованных выборок $\{S_i(X_{ki})\}$ и $\{S_j(X_{kj})\}$, $k = \overline{1, N}$, т.е. $\hat{x}_{0,5}^{S(ij)} = X_{(2N+1)}^{S(ij)}$.

Как и $\hat{x}_{0,5}^{(ij)}$, оценки $\hat{x}_{0,5}^{S(ij)}$ являются несмещенными, состоятельными и асимптотически нормальными, при этом

$$L\left(\sqrt{N}\left(\hat{x}_{0,5}^{S(ij)} - x_{0,5}^{S(ij)}\right)\right) \xrightarrow{N \rightarrow \infty} N\left(0; v_{ij}^{S^2}\right), \quad (9)$$

где

$$v_{ij}^{s2} = \frac{2 \left\{ 1 - F_i(\min\{t, S_i(t)\}) - F_j(\min\{t, S_j(t)\}) + F_{ij}(t, t) + F_{ij}(S_i(t), S_j(t)) - F_{ij}(t, S_j(t)) - F_{ij}(S_i(t), t) \right\}}{\left(f_i(t) + f_j(t) - f_i(S_i(t)) \cdot (S_i)'(t) - f_j(S_i(t)) \cdot (S_j)'(t) \right)^2}. \quad (10)$$

Покажем это. Рассмотрим $\text{Var } \hat{G}_{ij}^s(x) = \sigma_{ij}^2(x) / N$. Используя свойство (6), получим

$$\sigma_{ij}^2(x) = \frac{1}{16} \left(\text{Var } \hat{F}_i(x) + \text{Var } \hat{F}_j(x) + 2 \text{cov} \{ \hat{F}_i(x), \hat{F}_j(x) \} - 2 \text{cov} \{ \hat{F}_i(x), \hat{F}_i(S_i(x)) \} - 2 \text{cov} \{ \hat{F}_j(x), \hat{F}_j(S_j(x)) \} + \right. \\ \left. + \text{Var } \hat{F}_i(S_i(x)) + \text{Var } \hat{F}_j(S_j(x)) - 2 \text{cov} \{ \hat{F}_i(x), \hat{F}_j(S_j(x)) \} - 2 \text{cov} \{ \hat{F}_j(x), \hat{F}_i(S_i(x)) \} + 2 \text{cov} \{ \hat{F}_i(S_i(x)), \hat{F}_j(S_j(x)) \} \right).$$

Так как для $k = \overline{1, 4}$

$$\text{cov} \{ \hat{F}_k(x), \hat{F}_k(S_k(x)) \} = E \{ \hat{F}_k(x) \hat{F}_k(S_k(x)) \} - F(x)F(S_k(x)),$$

рассмотрим

$$E \{ \hat{F}_k(x) \hat{F}_k(S_k(x)) \} = E \left\{ \frac{1}{N^2} \sum_{s=1}^N \sum_{l=1}^N I(X_{sk} < x) \cdot I(X_{lk} < S_k(x)) \right\} = \\ = \frac{1}{N^2} \sum_{s \neq l}^N E \{ C(X_{sk} < x) \cdot C(X_{lk} < S_k(x)) \} + \frac{1}{N^2} \sum_{s=1}^N E \{ I(X_{sk} < x) \cdot I(X_{sk} < S_k(x)) \} = \\ = \left(1 - \frac{1}{N} \right) F_k(x)F_k(S_k(x)) + \frac{F_k(\min(x, S_k(x)))}{N}.$$

Откуда следует, что для $k = \overline{1, 4}$

$$\text{cov} \{ \hat{F}_k(x), \hat{F}_k(S_k(x)) \} = \frac{1}{N} (F_k(\min(x, S_k(x))) - F_k(x)F_k(S_k(x))).$$

Аналогично для $i = 1, 3$ и $j = 2, 4$ получим

$$\text{cov} \{ \hat{F}_i(x), \hat{F}_j(S_j(x)) \} = \frac{1}{N} (F_{ij}(x, S_j(x)) - F_i(x)F_j(S_j(x))).$$

Таким образом,

$$\text{Var } \hat{G}_{ij}^s(x) = \frac{1}{16N} \left(F_i(x)(1 - F_i(x)) + F_j(x)(1 - F_j(x)) + 2F_{ij}(x, x) - 2F_i(x)F_j(x) - 2F_i(\min(x, S_i(x))) + \right. \\ \left. + 2F_i(x)F_i(S_i(x)) - 2F_j(\min(x, S_j(x))) + 2F_j(x)F_j(S_j(x)) - 2F_{ij}(x, S_j(x)) + 2F_i(x)F_j(S_j(x)) + \right. \\ \left. + F_i(S_i(x))(1 - F_i(S_i(x))) + F_j(S_j(x))(1 - F_j(S_j(x))) - 2F_{ij}(S_i(x), x) + 2F_i(S_i(x))F_j(x) + \right. \\ \left. + 2F_{ij}(S_i(x), S_j(x)) - 2F_i(S_i(x))F_j(S_j(x)) \right).$$

С учетом свойства (6) получим

$$\text{Var } \hat{G}_{ij}^s(x) = \frac{1}{8N} \left(2F_i(x)(1 - F_i(x)) + 2F_j(x)(1 - F_j(x)) + F_{ij}(x, x) - F_i(\min(x, S_i(x))) - F_j(\min(x, S_j(x))) \right. \\ \left. + 2F_i(x) + 2F_j(x) - 1 - 4F_i(x)F_j(x) - F_{ij}(S_i(x), x) - F_{ij}(x, S_j(x)) + F_{ij}(S_i(x), S_j(x)) \right).$$

Таким образом,

$$\text{Var } \hat{G}_{ij}^s(x) = \frac{1}{8N} \left(2F_i(x)(1 - F_i(x)) + 2F_j(x)(1 - F_j(x)) - (1 - 2F_i(x))(1 - 2F_j(x)) - F_i(\min(x, S_i(x))) - \right. \\ \left. - F_j(\min(x, S_j(x))) + F_{ij}(x, x) - F_{ij}(x, S_j(x)) - F_{ij}(S_i(x), x) + F_{ij}(S_i(x), S_j(x)) \right).$$

Обозначим истинное значение цены $x^{(ij)}$ как t . Тогда, согласно [20], асимптотическая дисперсия с.в. $\sqrt{N} \hat{x}_{0.5}^{s(ij)}$ определяется как

$$v_{ij}^{s2} = \frac{\sigma_{ij}^{s2}(t)}{\left((G_{ij}^s)'(t) \right)^2},$$

тогда

$$\left(G_{ij}^s(t)\right)' = \frac{1}{16} \left(f_i(t) + f_j(t) - f_i(S_i(t)) \cdot (S_i)'(t) - f_j(S_j(t)) \cdot (S_j)'(t) \right)^2,$$

Так как в точке t кривые ценовой чувствительности пересекаются, т.е., $F_i(t) = 1 - F_j(t)$, будем иметь следующее:

$$\text{Var } \hat{G}_{ij}^s(t) = \frac{1}{8N} (1 - F_i(\min(t, S_i(t))) - F_j(\min(t, S_j(t))) + F_{ij}(t, t) - F_{ij}(S_i(t), t) - F_{ij}(t, S_j(t)) + F_{ij}(S_i(t), S_j(t))).$$

В итоге

$$v_{ij}^{s2} = \frac{2 \left\{ 1 - F_i(\min\{t, S_i(t)\}) - F_j(\min\{t, S_j(t)\}) + F_{ij}(t, t) - F_{ij}(t, S_j(t)) - F_{ij}(S_i(t), t) + F_{ij}(S_i(t), S_j(t)) \right\}}{\left(f_i(t) + f_j(t) - f_i(S_i(t)) \cdot (S_i)'(t) - f_j(S_j(t)) \cdot (S_j)'(t) \right)^2}. \quad (11)$$

Сравним полученное значение с v_{ij}^2 – асимптотической дисперсией немодифицированной оценки с.в. $\sqrt{N} \hat{x}_{0.5}^{(ij)}$. Так как для всех $k = \overline{1, 4}$ функции $S_k(x)$ по определению убывают, то их производные отрицательные, следовательно, знаменатель v_{ij}^s заведомо больше, чем у v_{ij}^{s2} . Осталось сравнить числители оценок. Учитывая (3), т.е. что $F_i(t) + F_j(t) = 1$, получим, что для числителя v_{ij}^{s2} верно следующее:

$$1 - F_i(\max\{t, S_i(t)\}) - F_j(\max\{t, S_j(t)\}) = F_i(t) - F_i(\max\{t, S_i(t)\}) + F_j(t) - F_j(\max\{t, S_j(t)\}) \leq 0.$$

Далее рассмотрим $P(t) = F_{ij}(S_i(t), S_j(t)) - F_{ij}(t, S_j(t)) - F_{ij}(S_i(t), t)$. Здесь возможны следующие варианты:

- 1) если для $k = i$ или j $S_k(t) \leq \alpha_k \leq t$, то очевидно, что $P(t) \leq 0$, так как всегда верно одно из двух:

$$F_{ij}(S_i(t), S_j(t)) \leq F_{ij}(t, S_j(t)) \text{ или } F_{ij}(S_i(t), S_j(t)) \leq F_{ij}(S_i(t), t);$$

- 2) если для $k = i$ и j $t \leq \alpha_k \leq S_k(t)$, то воспользуемся особенностями рассматриваемого метода ценообразования, для которого одно из значений цены всегда больше другого, например, $X_i < X_j$. Тогда искомая цена – медиана объединенных симметризованных и исходных выборок, а истинное значение цены $t = x^{(ij)}$ всегда находится между медианами, т.е. $\alpha_i \leq t \leq \alpha_j$. Следовательно, случай $t \leq \alpha_k \leq S_k(t)$ в принципе невозможен, его вероятность нулевая. Более того, в данной модели возможна реализация только такого случая, когда $S_i(t) \leq \alpha_i \leq t$ и $t \leq \alpha_j \leq S_j(t)$, что позволяет упростить формулу (11):

$$v_{ij}^{s2} = \frac{2 \left\{ F_i(t) - F_j(t) + F_{ij}(t, t) - F_{ij}(t, S_j(t)) - F_{ij}(S_i(t), t) + F_{ij}(S_i(t), S_j(t)) \right\}}{\left(f_i(t) + f_j(t) - f_i(S_i(t)) \cdot (S_i)'(t) - f_j(S_j(t)) \cdot (S_j)'(t) \right)^2}.$$

Таким образом, аналитически удалось доказать, что привлечение информации о симметрии кривых ценовой чувствительности позволяет повысить точность метода ценообразования PSM для достаточно больших N .

В качестве иллюстрации рассмотрим простейший случай, когда с.в. X_i, X_j распределены равномерно в $(0, 100)$ и $(50, 150)$ соответственно, и рассматриваются только пары $X_i < X_j$. Тогда

$$F_{ij}(x, y) = \begin{cases} 0, & x < y < 50, y > x; \\ \frac{x}{100} \cdot \frac{y-50}{100}, & x \in (0, 100), y \in (50, 150); \\ \frac{y-50}{100}, & x \geq 100, y \in (100, 150); \\ \frac{x}{100}, & x < 100, y \geq 150; \\ 1, & x \geq 100, y \geq 150, \end{cases}$$

искомое значение цены $t = \frac{50+100}{2} = 75$. Следовательно, $v_{ij}^2 = \frac{2 \cdot 75 \cdot 25}{100 \cdot 100 \cdot 0,02^2} = 937$. При этом

$S_i(t) = 2 \cdot 50 - 75 = 25$, $S_j(t) = 2 \cdot 100 - 75 = 125$. Тогда

$$v_{ij}^{S2} = \frac{2\{F_i(75) - F_j(75) + F_{ij}(75,75) + F_{ij}(25,125) - F_{ij}(75,125) - F_{ij}(25,75)\}}{(0,02 + 0,01 + 0,01)^2} =$$

$$= \frac{2\{0,75 - 0,25 + 2 \cdot 0,75 \cdot 0,25 - 0,75 \cdot 0,75 - 0,25 \cdot 0,25\}}{(0,02 + 0,01 + 0,01)^2} = 312,5,$$

то есть получили практически трехкратный выигрыш. На рис. 1 приведены графики переменных v_{ij}^2 и v_{ij}^{S2} для $F_i(x) = R_{(0,100)}(x)$ и $F_j(x) = R_{(50,150)}(x)$ в зависимости от разных объемов выборок N при $M = 10^5$. Заметим, что моделирование подтверждает верность расчета формул для асимптотических дисперсий, при этом также стоит отметить, что наблюдается достаточно высокая скорость сходимости: для достижения хорошей точности достаточно опросить 30 человек, хотя даже и при меньшем числе опрошенных наблюдается существенный выигрыш.

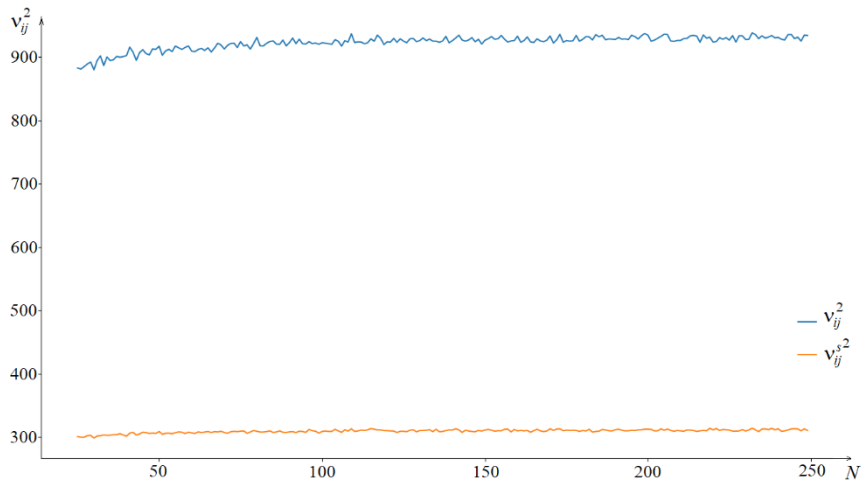


Рис. 1. Графики переменных v_{ij}^2 и v_{ij}^{S2} для $F_i(x) = R_{(0,100)}(x)$ и $F_j(x) = R_{(50,150)}(x)$

Fig. 1 Graphs of variables v_{ij}^2 and v_{ij}^{S2} for $F_i(x) = R_{(0,100)}(x)$ and $F_j(x) = R_{(50,150)}(x)$

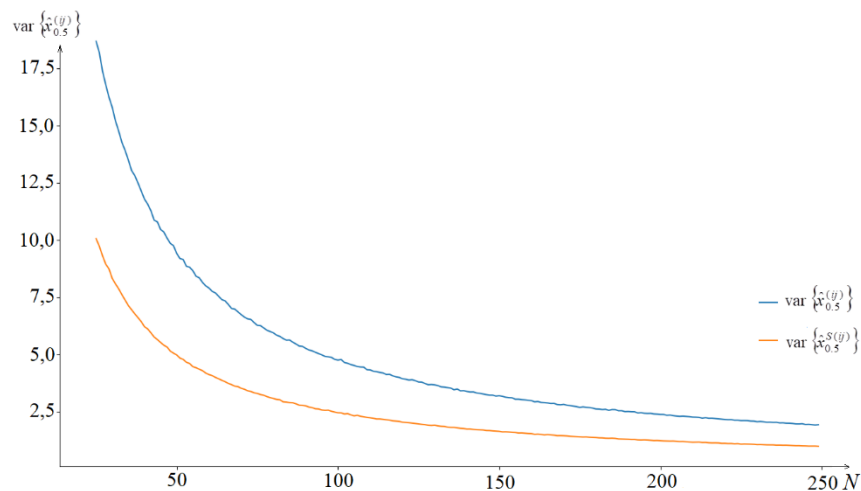


Рис. 2. Графики дисперсий $\text{var}\{\hat{z}_{0.5}^{(ij)}\}$ и $\text{var}\{\hat{z}_{0.5}^{S(ij)}\}$ для $F_i(x) = N_{(50,20^2)}(x)$ и $F_j(x) = R_{(20,100)}(x)$

Fig. 2. Variance graphs $\text{var}\{\hat{z}_{0.5}^{(ij)}\}$ and $\text{var}\{\hat{z}_{0.5}^{S(ij)}\}$ for $F_i(x) = N_{(50,20^2)}(x)$ and $F_j(x) = R_{(20,100)}(x)$

На рис. 2 приведены графики дисперсий $\text{var}\{\hat{x}_{0.5}^{(ij)}\}$ и $\text{var}\{\hat{x}_{0.5}^{S(ij)}\}$ для нормального с параметрами 50 и 400 $F_i(x) = N_{(50,20^2)}(x)$ и равномерного в (20, 100) $F_j(x) = R_{(20,100)}(x)$ распределений, на рис. 3 – для нормальных распределений $F_i(x) = N_{(50,20^2)}(x)$ и $F_j(x) = N_{(100,20^2)}(x)$ в зависимости от N при $M = 10^5$. В обоих случаях наблюдается значительный выигрыш в точности, даже в асимптотике выигрыш практически двукратный.

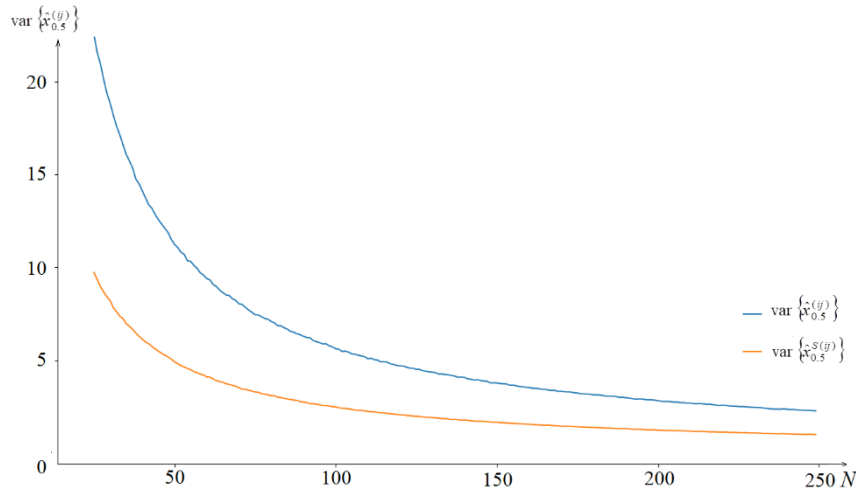


Рис. 3. Графики дисперсий $\text{var}\{\hat{x}_{0.5}^{(ij)}\}$ и $\text{var}\{\hat{x}_{0.5}^{S(ij)}\}$ для $F_i(x) = N_{(50,20^2)}(x)$ и $F_j(x) = N_{(100,20^2)}(x)$

Fig. 3. Variance graphs $\text{var}\{\hat{x}_{0.5}^{(ij)}\}$ и $\text{var}\{\hat{x}_{0.5}^{S(ij)}\}$ for $F_i(x) = N_{(50,20^2)}(x)$ and $F_j(x) = N_{(100,20^2)}(x)$

Также с помощью имитационного моделирования была подтверждена асимптотическая нормальность оценок для разных видов кривых ценовой чувствительности как с учетом, так и без учета их симметрии. При этом сходимость к нормальному распределению наблюдалась во всех рассмотренных случаях для $N \geq 30$. Следовательно, пользуясь асимптотической нормальностью, для достаточно больших N можно построить приближенные доверительные интервалы для ценовых значений с заданной доверительной вероятностью $0 < \gamma < 1$:

$$X_{(2N+1)}^{S(ij)} - \frac{z_{(1+\gamma)/2} \cdot \hat{v}_{ij}^S}{\sqrt{N}} \leq x_{0.5}^{S(ij)} \leq X_{(2N+1)}^{S(ij)} + \frac{z_{(1+\gamma)/2} \cdot \hat{v}_{ij}^S}{\sqrt{N}}, \quad (12)$$

где $z_{(1+\gamma)/2}$ – квантиль уровня $(1+\gamma)/2$ стандартного нормального распределения, $\hat{v}_{ij}^S$ – выборочная оценка v_{ij}^S .

Заключение

В работе найдены оценки ценовых значений метода измерения ценовой чувствительности PSM для случая привлечения дополнительной информации об S^α -симметрии кривых ценовой чувствительности, показаны асимптотическая нормальность и состоятельность оценок, построены приближенные доверительные интервалы для истинных значений цен. Аналитически показано, что в асимптотике точность модифицированного метода выше, чем классического. С помощью имитационного моделирования рост точности оценивания был также подтвержден для разных видов кривых ценовой чувствительности и для любого количества опрошенных, при этом нормальность наблюдалась для объемов выборок, не меньших 30.

Таким образом, дополнительная информация об S^α -симметрии кривых ценовой чувствительности позволяет получить более качественные оценки уровней цен в методе ценообразования PSM, ориентированном на спрос, даже при достаточно небольшом числе участников исследования.

Список источников

1. Simon H. Confessions of the Pricing Man: How Price Affects Everything. Springer, 2015. 236 p.
2. Нюренбергер Л.Б., Краковецкая И.В., Далибожко А.И. Функционирование МСП сферы услуг в условиях санкционного давления: вызовы и перспективы развития // Исследование проблем экономики и финансов. 2023. № 2. С. 1–7. doi: 10.31279/2782-6414-2023-2-7
3. Kotler P., Armstrong G. Principles of Marketing. Prentice-Hall, 2016. 398 p.
4. Albari A., Safitri I. The Influence of Product Price on Consumers' Purchasing // Decisions' Review of Integrative Business and Economics Research. 2018. V. 7 (2). P. 328–336.
5. Baker R.J. Implementing Value Pricing: A Radical Business Model for Professional Firms. Wiley, 2011. 400 p.
6. Nagle T.T., Müller G., Gruyaert E. The Strategy and Tactics of Pricing: A Guide to Growing More Profitably. Routledge, 2023. 360 p.
7. Chhabra S. Determining the Optimal Price Point: Using Van Westendorp's Price Sensitivity Meter // Managing in Recovering Markets. New Delhi : Springer. 2015. P. 8–20. (Springer Proceedings in Business and Economics).
8. Andrievskaya A.A., Zenkova Z.N. The Use of Stoezel Method in Pricing of New Bio-Energy Drink // Математическое и программное обеспечение информационных, технических и экономических систем : материалы VII Междунар. молодежной науч. конф. Томск : Изд. Дом ТГУ, 2019. С. 328–332.
9. Zenkova Zh.N., Cыбульникова Н.Р. Исследование цены спроса на продукцию ЗАО «Натуральное мыло» методом Штоцеля по цензурированным данным // Математическое и программное обеспечение информационных, технических и экономических систем : материалы IV Междунар. молодежной науч. конф. Томск, 20–21 мая 2016 г. Томск : Изд. Дом ТГУ, 2016. С. 62–65.
10. Docters R.G., Barzelay M., Hanson J.G., Nguyen C. Contextual Pricing: The Death of List Price and the New Market Reality. McGraw Hill, 2011. 256 p.
11. Stoezel J. Psychological / Sociological Aspects of Price // Pricing Research. Princeton, NJ : Brandon Systems Press, 1970. P. 70–74.
12. Westendorp P.H. NSS – Price Sensitivity Meter (PSM) – A New Approach to Consumer Perception of Prices // Venice Congress Main Sessions, European Marketing Research Society (ESOMAR). Amsterdam, 1976. P. 139–167.
13. Коржов М.М., Кузин Д.А. Сравнительный анализ и практическое применение методов ценообразования с ориентацией на спрос // Новости маркетинга. 2007. № 4. С. 9–17.
14. Zenkova Zh.N., Краковецкая И.В. Непараметрическая оценка Тёрнбулла для интервально-цензурированных данных в маркетинговом исследовании спроса на биоэнергетические напитки // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 3 (24). С. 6–69.
15. Zenkova Zh.N., Краковецкая И.В. Моделирование по неполным данным в логистике и маркетинге // Логистические системы в глобальной экономике. 2013. № 3 (1). С. 98–105.
16. Lipovetsky S. Van Westendorp Price Sensitivity in Statistical Modeling // International Journal of Operations and Quantitative Management. 2006. V. 12 (2). P. 141–156.
17. Lipovetsky S., Magnan S., Polzi A.Z. Pricing Models in Marketing Research // Intelligent Information Management. 2011. V. 3. P. 167–174.
18. Lipovetsky S. Statistics in Marketing Research: A Brief Review on Special Methods and Applications // Model Assisted Statistics and Applications. 2022. V. 17 (3). P. 213–216.
19. Андриевская А.А., Zenkova Zh.N. Подход Липовецкого к модификации метода ценообразования Штоцеля // Математическое и программное обеспечение информационных, технических и экономических систем : материалы VII Междунар. молодежной науч. конф. Томск : Изд. Дом ТГУ, 2019. С. 125–130.
20. Дмитриев Ю.Г., Zenkova Zh.N., Zenkov A.G. Статистические свойства оценок метода измерения ценовой чувствительности PSM // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2019. № 47. С. 30–40. doi: 10.17223/19988605/47/4
21. Журко Е.С., Zenkova Zh.N. Метод ценообразования PSM для случая цензурированных выборок малого объема // Математическое и программное обеспечение информационных, технических и экономических систем : материалы IV Междунар. молодежной науч. конф. Томск, 20–21 мая 2016 г. Томск : Изд. Дом ТГУ, 2016. С. 47–51.
22. Zenkova Zh.N., Cыбульникова Н.Р. Статистические свойства оценок ценовых значений метода ценовой чувствительности PSM для линейно зависимых наблюдений с учетом симметрии // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2021. № 56. С. 59–67.
23. Журко Е.С., Zenkova Zh.N. Метод ценообразования PSM для цензурированных данных с учетом квантиля // Наука. Технологии. Производство. 2015. № 13. С. 13–16.
24. Журко Е.С., Zenkova Zh.N. Влияние априорной информации на результаты метода ценообразования на товар-новинку PSM // Актуальные проблемы и перспективы развития государственной статистики в современных условиях : сб. материалов III Междунар. науч.-практ. конф. Саратов, 5–7 декабря 2016 г. Саратов : Саратовстат, 2017. Т. 2. С. 66–68.
25. Zenkova Zh.N. Учет информации об S_{α} -равноплечной симметрии при обработке цензурированных данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 2 (15). С. 32–45.
26. Дмитриев Ю.Г., Устинов Ю.К. Статистическое оценивание распределений вероятностей с использованием дополнительной информации. Томск : Изд-во ТГУ, 1988. 194 с.
27. Боровков А.А. Математическая статистика. Новосибирск : Наука, 1997. 772 с.

References

1. Simon, H. (2015) *Confessions of the Pricing Man: How Price Affects Everything*. Springer.
2. Nyurenberger, L.B., Krakovetskaya, I.V. & Dalibozhko, A.I. (2023) Funktsionirovanie MSP sfery uslug v usloviyakh sanktsionnogo davleniya: vyzovy i perspektivy razvitiya [Functioning of SMEs in service industry under sanction pressure: Challenges and Prospects for Development]. *Issledovanie problem ekonomiki i finansov*. 2. pp. 1–7. DOI: 10.31279/2782-6414-2023-2-7
3. Kotler, P. & Armstrong, G. (2016) *Principles of Marketing*. Prentice-Hall.
4. Albari, A. & Safitri, I. (2018) The influence of product price on consumers' purchasing. *Decisions' Review of Integrative Business and Economics Research*. 7(2). pp. 328–336.
5. Baker, R.J. (2011) *Implementing Value Pricing: A Radical Business Model for Professional Firms*. Wiley.
6. Nagle, T.T., Müller, G. & Gruyaert, E. (2023) *The Strategy and Tactics of Pricing: A Guide to Growing More Profitably*. Routledge.
7. Chhabra, S. (2015) Determining the optimal price point: using Van Westendorp's price sensitivity meter. In: *Managing in Recovering Markets. Springer Proceedings in Business and Economics*. New Delhi: Springer. pp. 8–20.
8. Andrievskaya, A.A. & Zenkova, Z.N. (2019) The use of Stotzel method in pricing of new bio-energy drink. *Matematicheskoe i programnoe obespechenie informatsionnykh, tekhnicheskikh i ekonomicheskikh sistem* [Mathematical and Software Support for Information, Technical and Economic Systems]. Proc. of the 7th International Conference. Tomsk: TSU. pp. 328–332.
9. Zenkova, Z.N. & Tsybulnikova, N.R. (2016) Issledovanie tseny sprosa na produktsiyu ZAO "Natural'noe mylo" metodom Shtotselya po tsenzurirovannym dannym [Demand price for the products of ZAO Natural'noe mylo using the Stotzel method based on censored data]. *Matematicheskoe i programnoe obespechenie informatsionnykh, tekhnicheskikh i ekonomicheskikh sistem* [Mathematical and Software Support for Information, Technical and Economic Systems]. Proc. of the Fourth Conference. Tomsk: TSU. pp. 62–65.
10. Docters, R.G., Barzelay, M., Hanson, J.G. & Nguyen, C. (2011) *Contextual Pricing: The Death of List Price and the New Market Reality*. McGraw Hill.
11. Stotzel, J. (1970) Psychological/ Sociological Aspects of Price. In: *Pricing Research*. Princeton, NJ: Brandon Systems Press. pp. 70–74.
12. Westendorp, P.H. (1976) NSS – Price sensitivity meter (PSM) – a new approach to consumer perception of prices. In: *Venice Congress Main Sessions, European Marketing Research Society (ESOMAR)*. Amsterdam. pp. 139–167.
13. Korzhov, M.M. & Kuzin, D.A. (2007) Sravnitel'nyy analiz i prakticheskoe primeneniye metodov tsenoobrazovaniya s orientatsiey na spros [Comparative Analysis and Practical Application of Demand-Driven Pricing Methods]. *Novosti marketinga*. 4. pp. 9–17.
14. Zenkova, Z.N. & Krakovetskaya, I.V. (2013) Nonparametric Turnbull estimator for interval-censored data in the marketing research of the demand of bio-energy drinks. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(24). pp. 64–69.
15. Zenkova, Z.N. & Krakovetskaya, I.V. (2013) Modelirovaniye po nepolnym dannym v logistike i marketinge [Modeling with incomplete data in logistics and marketing]. *Logisticheskie sistemy v global'noy ekonomike*. 3(1). pp. 98–105.
16. Lipovetsky, S. (2006) Van Westendorp Price sensitivity in statistical modeling. *International Journal of Operations and Quantitative Management*. 12(2). pp. 141–156.
17. Lipovetsky, S., Magnan, S. & Polzi, A.Z. (2011) Pricing models in marketing research. *Intelligent Information Management*. 3. pp. 167–174.
18. Lipovetsky, S. (2022) Statistics in marketing research: a brief review on special methods and applications. *Model Assisted Statistics and Applications*. 17(3). pp. 213–216.
19. Andrievskaya, A.A. & Zenkova, Z.N. (2019) Podkhod Lipovetskogo k modifikatsii metoda tsenoobrazovaniya Shtotselya [Lipovetsky's approach in modification of Stotzel pricing method]. *Matematicheskoe i programnoe obespechenie informatsionnykh, tekhnicheskikh i ekonomicheskikh sistem* [Mathematical and Software Support for Information, Technical and Economic Systems]. Proc. of the 7th International Conference. Tomsk: TSU. pp. 125–130.
20. Dmitriev, Yu.G., Zenkova, Z.N. & Zenkov, A.G. (2019) Statistical properties of price sensitivity meter results. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 47. pp. 30–40. DOI: 10.17223/19988605/47/4
21. Zhurko, E.S. & Zenkova, Z.N. (2016) Metod tsenoobrazovaniya PSM dlya sluchaya tsenzurirovannykh vyborok malogo ob'ema [PSM Pricing Method for Small Censored Samples]. *Matematicheskoe i programnoe obespechenie informatsionnykh, tekhnicheskikh i ekonomicheskikh sistem* [Mathematical and Software Support for Information, Technical and Economic Systems]. Proc. of the 4th International Conference. Tomsk: TSU. pp. 47–51.
22. Zenkova, Z.N. & Tsybulnikova, N.R. (2021) Statistical properties of price sensitivity meter results for linear relationship between observations with symmetry. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 56. pp. 59–67. DOI: 10.17223/19988605/56/7
23. Zhurko, E.S. & Zenkova, Z.N. (2015) Metod tsenoobrazovaniya PSM dlya tsenzurirovannykh dannyykh s uchedom kvantilya [PSM pricing method with quantile]. *Nauka. Tekhnologii. Proizvodstvo*. 13. pp. 13–16.
24. Zhurko, E.S. & Zenkova, Z.N. (2017) Vliyaniye apriornoy informatsii na rezul'taty metoda tsenoobrazovaniya na tovarnovinku PSM [The influence of a priori information on PSM pricing method for a new product]. In: *Aktual'nye problemy i perspektivy razvitiya gosudarstvennoy statistiki v sovremennykh usloviyakh* [Topical problems and prospects for the development of state statistics in modern conditions]. Vol. 2. Saratov: Saratovstat. pp. 66–68.

25. Zenkova, Z.N. (2011) Processing censored data with $S_{\square}$ -symmetry. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(15). pp. 32–45.
26. Dmitriev, Yu.G. & Ustinov, Yu.K. (1988) *Statisticheskoe otsenivanie raspredeleniy veroyatnostey s ispol'zovaniem dopolnitel'noy informatsii* [Statistical Estimation of Probability Distributions Using Additional Information]. Tomsk: TSU.
27. Borovkov, A.A. (1997) *Matematicheskaya statistika* [Mathematical Statistics]. Novosibirsk: Nauka.

Информация об авторах:

Зенкова Жанна Николаевна – доцент, кандидат физико-математических наук, MBA, доцент кафедры системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: zhanna.zenkova@mail.tsu.ru

Андриевская Анна Андреевна – ассистент кафедры системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: anna.andriev@mail.tsu.ru

Яковлев Никита Иванович – ведущий аналитик данных, Tele2/Altel (Алматы, Казахстан). E-mail: nyakovlev19@gmail.com

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Zenkova Zhanna N. (Candidate Physical and Mathematical Sciences, Associate Professor, MBA; Department of Systems Analysis and Mathematical Modeling, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: zhanna.zenkova@mail.tsu.ru

Andrievskaya Anna A. (Teaching Assistant, Department of Systems Analysis and Mathematical Modeling, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation)
E-mail: anna.andriev@mail.tsu.ru

Yakovlev Nikita I. (Senior Data Analyst, Tele2/Altel, Almaty, Kazakhstan) E-mail: nyakovlev19@gmail.com

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 31.10.2024; принята к публикации 03.03.2025

Received 31.10.2024; accepted for publication 03.03.2025

Научная статья

УДК 681.5

doi: 10.17223/19988605/70/7

**Робастная фильтрация в дискретных стохастических системах
со скачкообразными параметрами и интервальной неопределенностью****Константин Станиславович Ким¹, Валерий Иванович Смагин²**^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Российская Федерация*¹ *kks93@rambler.ru*² *vsm@mail.tsu.ru*

Аннотация. Рассматривается алгоритм робастной фильтрации для дискретной стохастической системы с интервальными параметрами, зависящей от состояния скрытой марковской цепи. Для представления интервальных параметров используется вероятностный подход, в основе которого лежит замена неопределенных параметров интервального типа независимыми случайными величинами с равномерным распределением. Используются рекуррентные схемы калмановской фильтрации и алгоритмы оценивания состояния скачкообразного процесса. Для иллюстрации предложенного подхода приводится пример.

Ключевые слова: робастная фильтрация; интервальная неопределенность; скрытая марковская цепь.

Для цитирования: Ким К.С., Смагин В.И. Робастная фильтрация в дискретных стохастических системах со скачкообразными параметрами и интервальной неопределенностью // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 73–80. doi: 10.17223/19988605/70/7

Original article

doi: 10.17223/19988605/70/7

**Robust filtering in discrete stochastic systems with jump parameters
and interval uncertainty****Konstantin S. Kim¹, Valery I. Smagin²**^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*¹ *kks93@rambler.ru*² *vsm@mail.tsu.ru*

Abstract. A robust filtering algorithm for a discrete stochastic system with interval parameters depending of the hidden Markov chain is considered. To represent the interval parameters, a probabilistic approach is used, which is based on the replacement of indeterminate interval-type parameters with independent random variables with a uniform distribution. Recurrent Kalman filtering schemes and algorithms for estimating the state of jump parameter are used. An example is provided to illustrate the proposed approach.

Keywords: robust filtering; interval uncertainty; hidden Markov chain.

For citation: Kim, K.S., Smagin, V.I. (2025) Robust filtering in discrete stochastic systems with jump parameters and interval uncertainty. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 73–80. doi: 10.17223/19988605/70/7

Введение

Задачи построения оценок и управлений для систем со случайными скачкообразными параметрами актуальны для различных реальных объектов. В качестве примера таких объектов можно рассмотреть, например, энергетические системы [1, 2], системы связи [3, 4], задачи обнаружения неисправностей [5, 6] и др.

В работах [7–11] задачи фильтрации рассматриваются для систем со случайными скачкообразными параметрами и аддитивными мультипликативными возмущениями.

В настоящей работе рассмотрена задача робастной фильтрации в условиях ненаблюдаемости марковской цепи, которая входит в описание линейной стохастической системы с интервальными параметрами. Решение получено с использованием принципа разделения, фильтрации Калмана и оценок неизвестного входного сигнала [12–17].

1. Постановка задачи

Рассмотрим линейную дискретную стохастическую систему со скачкообразными и неопределенными параметрами

$$x(k+1) = \tilde{A}_\gamma x(k) + B_\gamma q(k), \quad x(0) = x_0, \quad (1)$$

где $x(k) \in \mathbb{R}^n$ – вектор состояния системы, x_0 – случайный вектор с известным математическим ожиданием и дисперсией $\bar{x}_0 = M\{x_0\}$, $N_{0,i} = M\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T / \gamma = \gamma_i\}$, $i = \overline{1, \tau}$; $\gamma = \gamma(k)$ – марковская цепь с τ состояниями ($\gamma_1 = 1, \gamma_2 = 2, \dots, \gamma_\tau = \tau$); $\tilde{A}_\gamma$ – интервальная матрица, зависящая от процесса $\gamma(k)$; B_γ – матрица, зависящая от процесса $\gamma(k)$; $q(k) \in \mathbb{R}^m$ – гауссовские случайные векторы с характеристиками $M\{q(k)\} = 0$, $M\{q(k)q^T(j)\} = E\delta_{kj}$, ($M\{\cdot\}$ – математическое ожидание, T – символ матричного транспонирования, δ_{kj} – символ Кронекера, E – единичная матрица соответствующей размерности).

Вероятность $p_j(k) = P\{\gamma(k) = j\}$, $j = \overline{1, \tau}$, удовлетворяет уравнению

$$p_j(k+1) = \sum_{i=1}^{\tau} p_i(k) p_{i,j}, \quad p_j(0) = \bar{p}_j, \quad j = \overline{1, \tau}. \quad (2)$$

Здесь $p_{i,j}$ – вероятность перехода из состояния i в состояние j за один шаг, $\bar{p}_j$ – начальная вероятность j -го состояния.

Вектор наблюдения

$$y(k) = Sx(k) + v(k), \quad (3)$$

где $v(k)$ – гауссовская случайная последовательность, не зависящая от $q(k)$, с характеристиками $M\{v(k)\} = 0$, $M\{v(k)v^T(j)\} = V(k)\delta_{kj}$. Предполагается, что система (1), (2) наблюдаема при параметрических возмущениях матрицы динамики $\tilde{A}_\gamma$.

По наблюдениям (3) требуется найти оценку вектора состояния $\hat{x}(k)$, которую определим из условия минимума критерия

$$J(0, T, i) = M \left\{ \sum_{k=0}^T e^T(k) H e(k) / \gamma(0) = \gamma_i \right\}, \quad (4)$$

где $H > 0$ – весовая матрица, $e(k) = x(k) - \hat{x}(k)$.

2. Синтез робастного фильтра

В основу синтеза робастного фильтра положим принцип разделения, поэтому сначала для решения задачи будем использовать рекуррентный фильтр Калмана в предположении, что переменная $\gamma(k)$

доступна точному измерению. Для синтеза такого фильтра воспользуемся вероятностным подходом учета интервальной неопределенности параметров модели [18]. Суть метода заключается в том, что интервальные параметры заменяются независимыми случайными величинами, равномерно распределенными на интервале неопределенности.

Используя вероятностный подход, заменим неопределенные интервальные матрицы $\tilde{A}_\gamma$ матрицами, элементы которых зависят от случайных величин:

$$A_\gamma(\theta) = (A_\gamma + \sum_{r=1}^{m_1} \Phi_{\gamma,r} \theta_r), \quad (5)$$

где θ_r – независимые равномерно распределенные случайные величины на интервале $[-1, +1]$ ($-1 \leq \theta_r \leq 1$ ($r = \overline{1, m_1}$)). Будем считать, что случайные величины θ_r независимы от x_0 , $q(k)$ и $v(k)$. В (5) матрица $A_\gamma = \frac{1}{2}(\underline{A}_\gamma + \bar{A}_\gamma)$ является медианой интервальной матрицы $\tilde{A}_\gamma$ (здесь $\underline{A}_\gamma$ и $\bar{A}_\gamma$ – нижняя и верхняя границы интервальной матрицы). Матрицы $\Phi_{\gamma,r}$ ($r = \overline{1, m_1}$) можно задать так, чтобы один элемент, соответствующий неопределенному элементу матриц $\tilde{A}_\gamma$, оставался ненулевым (если один и тот же неопределенный элемент стоит на нескольких позициях матрицы $\tilde{A}_\gamma$, то матрицы $\Phi_{\gamma,r}$ будут иметь несколько соответствующих ненулевых элементов). Значения ненулевых элементов матриц $\Phi_{\gamma,r}$ несложно определить по ширине интервала неопределенности элементов матрицы $\tilde{A}_\gamma$.

Учитывая (5), модель системы (1) примет вид:

$$x(k+1) = (A_\gamma + \sum_{r=1}^{m_1} \Phi_{\gamma,r} \theta_r)x(k) + B_\gamma q(k), \quad x(0) = x_0. \quad (6)$$

Для построения оценки вектора состояния $\hat{x}(k)$ в предположении, что переменная $\gamma = \gamma(k)$ известна, воспользуемся рекуррентным оценивателем, совпадающим по структуре с фильтром Калмана:

$$\hat{x}(k+1) = A_i \hat{x}(k) + K_i(k)(y(k+1) - S A_i \hat{x}(k)), \quad \hat{x}(0) = \bar{x}_0, \quad (7)$$

где $A_i = A_{\gamma=\gamma_i}$ ($i = \overline{1, \tau}$), $K_i(k)$ – матрица коэффициентов передачи фильтра.

Матрицы $K_i(k)$ определяются на основе минимизации критерия (4). Запишем критерий (4) в виде суммы

$$J(0, T, i) = \sum_{k=0}^T \text{tr } N_i(k) H, \quad (8)$$

где tr – след квадратичной матрицы.

Теорема. Пусть существуют положительно определенные матрицы $N_i(k)$, удовлетворяющие следующим матричным разностным уравнениям:

$$\begin{aligned} N_i(k+1) = & (A_i - K_i(k) S A_i) (\sum_{j=1}^{\tau} p_{i,j} N_j(k)) (A_i - K_i(k) S A_i)^T + \\ & + K_i(k) V_i K_i(k)^T + (E - K_i(k) S) D_i (E - K_i(k) S)^T, \quad N_i(0) = N_0 \quad (i = \overline{1, \tau}), \end{aligned} \quad (9)$$

где

$$D_i = \frac{1}{3} \sum_{r=1}^{m_1} \Phi_{i,r} (\sum_{j=1}^{\tau} p_{i,j} N_j(k)) \Phi_{i,r}^T + \frac{1}{3} \sum_{r=1}^{m_1} \Phi_{i,r} \hat{x}(k) \hat{x}^T(k) \Phi_{i,r}^T + B_i B_i^T. \quad (10)$$

Тогда оптимальные коэффициенты передачи фильтра определяются по формуле

$$K_i(k) = (A_i (\sum_{j=1}^{\tau} p_{i,j} N_j(k)) A_i^T S^T + D_i S^T) [S A_i (\sum_{j=1}^{\tau} p_{i,j} N_j(k)) A_i^T S^T + S D_i S^T + V]^{-1}.$$

Доказательство. Уравнение (9) является уравнением для матрицы $N_i(k) = M\{e(k)e(k)^T / \gamma = \gamma_i\}$ ($i = \overline{1, \tau}$).

Введем функцию Ляпунова следующего вида:

$$W(k, N_i(k)) = \text{tr } N_i(k)H + \text{tr} \sum_{t=k}^T \bar{\Psi}_i(t)L_i(t), \quad (11)$$

где $\bar{\Psi}_i(t) = (E - K_i(k)S_i)D_i(E - K_i(k)S_i)^T + K_i(t)VK_i(t)^T + \Psi_i(t)$ ($\Psi_i(t)$ – некоторые положительно определенные матрицы). Здесь предполагается, что в (11) матрицы $L_i(t)$ положительно определены и удовлетворяют уравнению

$$\begin{aligned} L_i(t) = & (A_i - K_i(t)S_iA_i)^T \left(\sum_{j=1}^{\tau} p_{i,j}L_j(t+1) \right) (A_i - K_i(t)S_iA_i) + \\ & + H + (E - K_i(k)S_i)^T \left(\sum_{r=1}^{m_i} \Phi_{i,r}^T \left(\sum_{j=1}^{\tau} p_{i,j}L_j(t+1) \right) \Phi_{i,r} \right) (E - K_i(k)S_i), \quad L_i(T) = L_T, \end{aligned} \quad (12)$$

где L_T – некоторая положительно определенная матрица.

Суммируя по $k = \overline{t, T-1}$ конечные разности функции $W(k, N_i(k))$ и учитывая формулу (12), получим

$$\begin{aligned} \sum_{k=t}^{T-1} \Delta W(k, N_i(k)) &= \sum_{k=t}^{T-1} [W(k+1, N_i(k+1)) - W(k, N_i(k))] = \\ &= \sum_{k=t}^{T-1} \text{tr} [N_i(k+1)L_i(k+1) - N_i(k)L_i(k) - \bar{\Psi}_i(k)L_i(k)]. \end{aligned} \quad (13)$$

С другой стороны, это выражение может быть представлено следующим образом:

$$\begin{aligned} \sum_{k=t}^{T-1} \Delta W(k, N_i(k)) &= W(t+1, N_i(t+1)) - W(t, N_i(t)) + \dots + W(T, N_i(T)) - \\ &- W(T-1, N_i(T-1)) = \text{tr } N_i(T)L_T - \text{tr } N_i(t)L_i(t) - \text{tr} \sum_{k=t}^{T-1} \bar{\Psi}_i(k)L_i(k). \end{aligned} \quad (14)$$

Добавим в правую часть формулы (8) разность правых частей (13) и (14). В результате критерий (8) примет вид:

$$\begin{aligned} J(0, T, i) &= \sum_{k=0}^{T-1} \text{tr } N_i(k)H - \sum_{k=1}^{T-1} \text{tr } N_i(k)L_i(k) + \\ &+ \sum_{k=0}^{T-1} \text{tr} [(A_i - K_i(k)S_iA_i) \left(\sum_{j=1}^{\tau} p_{i,j}N_j(k) \right) (A_i - K_i(k)S_iA_i)^T + \\ &+ (E - K_i(k)S_i)D_i(E - K_i(k)S_i)^T + K_i(k)VK_i(k)^T] L_i(k+1). \end{aligned} \quad (15)$$

Применяя правила матричного дифференцирования функции tr от произведения матриц [19] и учитывая (15), получим

$$\begin{aligned} \frac{\partial J(0, T, i)}{\partial K_i(k)} &= \sum_{k=0}^{T-1} 2[-L_i(k+1)A_i \left(\sum_{j=1}^{\tau} p_{i,j}N_j(k) \right) A_i^T S^T + L_i(k+1)K_i(k)S_iA_i \times \\ &\times \left(\sum_{j=1}^{\tau} p_{i,j}N_j(k) \right) A_i^T S^T - L_i(k+1)SD_iS^T + L_i(k+1)K_i(k)SD_iS^T + L_i(k+1)K_i(k)V]. \end{aligned} \quad (16)$$

Приравнявая производную (15) нулевой матрице, получим формулу для определения матриц $K_i(k)$ ($i = \overline{1, \tau}$):

$$K_i(k) = (A_i \left(\sum_{j=1}^{\tau} p_{i,j}N_j(k) \right) A_i^T S^T + D_iS^T) [S_iA_i \left(\sum_{j=1}^{\tau} p_{i,j}N_j(k) \right) A_i^T S^T + SD_iS^T + V]^{-1}. \quad (17)$$

Учитывая (12) и (14), можно показать, что конечная разность функции Ляпунова (11) определяется по формуле

$$\begin{aligned} \Delta W(k, N_i(k)) &= W(k+1, N_i(k+1)) - W(k, N_i(k)) = \text{tr } N_i(k+1)H - \\ &- \text{tr } N_i(k)H - \text{tr} [(E - K_i(k)S_i)D_i(E - K_i(k)S_i)^T + K_i(k)VK_i(k)^T + \Psi_i(t)] L_i(k). \end{aligned} \quad (18)$$

В силу того, что решения матричных разностных уравнений (10), (12) положительно определены, функция Ляпунова (11) будет положительной. Устойчивость по Ляпунову динамики фильтра гарантируется тем, что матрицу $\Psi_i(t)$ всегда можно задать так, чтобы выражение (18) стало отрицательным.

3. Оценка скачкообразного параметра

При вычислении оценок скачкообразного параметра используется подход диагностики состояния скачкообразного параметра, предложенный в [11]. Для системы будем использовать тот факт, что при ошибочном определении значения параметра $\gamma(k)$ модель (5) можно представить в эквивалентном виде, но с дополнительным вектором входа. Действительно, когда система (5) находится в j -м состоянии ($\gamma = \gamma_j$), и это состояние ошибочно определено как i -е ($j \neq i$), уравнение (5) представляется как модель с дополнительным входом:

$$x(k+1) = (A_i + \sum_{r=1}^{m_1} \Phi_{i,r} \theta_r) x(k) + f_i(k) + B_i q(k), \quad x(0) = x_0, \quad (19)$$

где вектор дополнительного входа определяется по формуле

$$f_i(k) = (A_j - A_i) x(k) + \sum_{r=1}^{m_1} \Phi_{j,r} x(k) \theta_r(k) - \sum_{r=1}^{m_1} \Phi_{i,r} x(k) \theta_r(k) + B_j q_j(k) - B_i q_i(k). \quad (20)$$

Заметим также, так как j -е состояние нам неизвестно, то дополнительный вход $f_i(k)$ также является неизвестным.

Оценку дополнительного входа определим из условия минимума критерия

$$J^{<LMS>}(f_i(k)) = \sum_{t=1}^{k+1} \left\{ \|y(t) - S(A_i \hat{x}(t-1) + f_i(t-1))\|_{\Omega_1}^2 + \|f_i(t-1)\|_{\Omega_2}^2 \right\}. \quad (21)$$

В результате имеем следующую МНК оценку:

$$\hat{f}_i(k) = (S^T \Omega_1 S + \Omega_2)^{-1} S^T \Omega_1 [y(k+1) - S(A_i \hat{x}(k))]. \quad (22)$$

В (21) Ω_1, Ω_2 – положительно определенные весовые матрицы.

Из (20) видно, что дополнительный вход в модели (19) будет иметь нулевое математическое ожидание, если состояние скачкообразного параметра определено правильно, т.е. $j = i$. Поэтому предлагается алгоритм оценивания параметра $\gamma(k)$ (определение номера состояния системы i) строить на основе поиска такого номера i , для которого евклидова норма оценки (22) $\|\hat{f}_i(k)\|$ минимальна. Отметим, что для повышения точности определения значения оценки скачкообразного параметра $\hat{\gamma}(k)$ в [11] предлагается использовать экспоненциальное сглаживание значения нормы $\|\hat{f}_i(k)\|$.

Оценку γ_i определим по следующему алгоритму:

- 1) вычисляется норма оценки неизвестного вектора $\psi(i, k) = \|\hat{f}_i(k)\|$ для всех $\gamma = \gamma_i$ ($i = \overline{1, \tau}$);
- 2) выполняется экспоненциальное сглаживание для значений $\psi(i, k)$:

$$\omega(i, k+1) = \lambda \psi(i, k+1) + (1-\lambda) \omega(i, k), \quad \omega(i, 1) = \psi(i, 1), \quad i = \overline{1, \tau}, \quad (23)$$

где λ – коэффициент сглаживания;

3) для каждого момента времени k определяется значение i , для которого сглаженное значение нормы $\omega(i, k)$ минимально ($i = \arg \min \{\omega(i, k)\}$). В качестве оценки $\hat{\gamma}(k)$ выбирается γ_i .

Таким образом, в качестве робастной оценки фильтрации $\hat{x}(k)$ выбирается оценка (7), в которой в каждый момент k используется значение i , полученное в п. 3 описанного выше алгоритма, при этом коэффициенты передачи $K_i(k)$ определяются по формуле (17).

4. Результаты моделирования

Рассмотрим задачу фильтрации с помощью предлагаемого алгоритма для дискретной системы с интервальными параметрами и тремя состояниями марковской цепи $\gamma(k)$ со следующей матрицей вероятностей перехода:

$$[p_{i,j}] = \begin{pmatrix} 0,4 & 0,2 & 0,4 \\ 0,3 & 0,5 & 0,2 \\ 0,3 & 0,3 & 0,4 \end{pmatrix}.$$

Моделирование выполнено на интервале времени $k \in [0, T]$ для следующих исходных данных:

$$\begin{aligned} A_1 &= \begin{pmatrix} 0,85 & 0,1 \\ -0,05 & 0,91 \end{pmatrix}, A_2 = \begin{pmatrix} 0,6 & 0,05 \\ -0,02 & 0,45 \end{pmatrix}, A_3 = \begin{pmatrix} 0,8 & 0,12 \\ -0,04 & 0,62 \end{pmatrix}, \\ B_1 = B_2 = B_3 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Phi_{1,1} = \begin{pmatrix} 0,06 & 0 \\ 0 & 0 \end{pmatrix}, \Phi_{2,1} = \begin{pmatrix} 0,2 & 0 \\ 0 & 0 \end{pmatrix}, \Phi_{3,1} = \begin{pmatrix} 0,15 & 0 \\ 0 & 0 \end{pmatrix}, \\ \Phi_{1,2} &= \begin{pmatrix} 0 & 0 \\ 0 & 0,05 \end{pmatrix}, \Phi_{2,2} = \begin{pmatrix} 0 & 0 \\ 0 & 0,2 \end{pmatrix}, \Phi_{3,2} = \begin{pmatrix} 0 & 0 \\ 0 & 0,2 \end{pmatrix}, Q = \begin{pmatrix} 0,02 & 0 \\ 0 & 0,01 \end{pmatrix}, \\ S &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, V = \begin{pmatrix} 3,2 & 0 \\ 0 & 4,0 \end{pmatrix}, \Omega_1 = 1, \Omega_2 = \begin{pmatrix} 0,1 & 0 \\ 0 & 0,1 \end{pmatrix}, \\ N_1(0) = N_2(0) = N_3(0) &= \begin{pmatrix} 2,0 & 0 \\ 0 & 2,0 \end{pmatrix}, x(0) = \begin{pmatrix} 200 \\ 200 \end{pmatrix}, \hat{x}(0) = \begin{pmatrix} 200 \\ 200 \end{pmatrix}. \end{aligned} \quad (24)$$

Параметр $\gamma(k)$ оценивался с помощью алгоритма, описанного в разделе 3.

Результаты моделирования, приведенные в таблице, иллюстрируют качество оценивания алгоритма робастной фильтрации по сравнению с алгоритмом фильтрации Калмана, синтезированным для системы с заданными матрицами $A_\gamma = \frac{1}{2}(\underline{A}_\gamma + \bar{A}_\gamma)$ без учета интервальной неопределенности (при моделировании в самом объекте интервальная неопределенность параметров присутствовала). В этом случае в алгоритме расчета коэффициентов передачи фильтра матрицы $\Phi_{1,1}$, $\Phi_{2,1}$, $\Phi_{3,1}$, $\Phi_{1,2}$, $\Phi_{2,2}$, $\Phi_{3,2}$ задавались нулевыми.

Качество фильтрации оценивалось путем сравнения стандартных ошибок отклонений оценок вектора состояния

$$\sigma_s = \sqrt{\frac{\sum_{k=0}^T (x_s(k) - \hat{x}_s(k))^2}{T}}, \quad (s = \overline{1, 2})$$

для следующих алгоритмов:

- (I) робастный алгоритм фильтрации;
- (II) интервальная неопределенность в модели не учитывается (фильтр строится по медианным значениям элементов матрицы A_γ).

Стандартные ошибки отклонений оценок фильтрации состояний системы для первой и второй компоненты вектора состояний σ_1 и σ_2 приведены в таблице. Результаты моделирования получены при $T = 400$ с усреднением по 1 000 реализациям.

Результаты сравнения стандартных ошибок отклонений оценок σ_i вектора состояния

s номер компоненты	Алгоритм	
	(I)	(II)
	σ_s	σ_s
1	1,61	1,80
2	1,92	2,15

Как показали результаты моделирования, применение робастного алгоритма позволяет уменьшить стандартные ошибки отклонений σ_i на 12%.

Заключение

Получено решение задачи синтеза алгоритма робастной фильтрации для дискретной стохастической системы с интервальными параметрами, зависящей от состояния скрытой марковской цепи. Для представления интервальных параметров используется вероятностный подход, в основе которого лежит замена неопределенных параметров интервального типа независимыми случайными величинами с равномерным распределением. Робастный алгоритм фильтрации строится с использованием принципа разделения, фильтрации Калмана и оценок неизвестного дополнительного входа. Результаты моделирования показали, что применение робастного алгоритма позволяет уменьшить стандартные ошибки отклонений оценок.

Список источников

1. Ugrinovskii V.A., Pota H.R. Decentralized control of power systems via robust control of uncertain Markov jump parameter systems // *Int. J. Control*. 2005. V. 78. P. 662–677.
2. Sales-Setien E., Penarrocha-Alos I. Markovian jump system approach for the estimation and adaptive diagnosis of decreased power generation in wind farms // *Iet Control Theory and Applications*, 2019. V. 13 (18). P. 3006–3018.
3. Zhu Y., Zhong Z., Zheng W.X. et al. HMM-based H-infinity filtering for discrete-time Markov jump LPV systems over unreliable communication channels // *IEEE Transactions on Systems Man Cybernetics-Systems*. 2018. V. 48, is. 12. P. 2035–2046.
4. Wang J., Yao F., Shen H. Dissipativity-based state estimation for Markov jump discrete-time neural networks with unreliable communication links // *Neurocomputing*. 2014. V. 139/SI. P. 107–113.
5. Yao X., Wu L., Zheng W.X. Fault detection filter design for Markovian jump singular systems with intermittent measurements // *IEEE Transactions on Signal Processing*. 2011. V. 59/7. P. 3099–3109.
6. Gagliardi G., Casavola A., Famularo D.A. Fault detection and isolation filter design method for Markov jump linear parameter-varying systems // *Int. Journal of Adaptive Control and Signal. Processing*, 2012, V. 26, is. 3/SI. P. 241–257.
7. Costa O.L.V., Benites G.R.A.M. Linear minimum mean square filter for discrete-time linear systems with Markov jumps and multiplicative noises // *Automatica*. 2011. V. 47 (3). P. 466–476.
8. Liu W. On state estimation of discrete-time linear systems with multiplicative noises and markov jumps // *32nd Chinese Control Conference*. Xian, CHINA. July 26–28, 2013. P. 3744–3749.
9. Costa O.L.V., Benites G.R.A.M. Robust mode-independent filtering for discrete-time Markov jump linear systems with multiplicative noises // *Int. J. of Control*. 2013. V. 86 (5). P. 779–793.
10. Ломакина С.С., Смагин В.И. Робастная фильтрация в непрерывных системах со скачкообразными изменениями параметров в случайные моменты времени // *Автометрия*. 2005. Т. 41, № 2. С. 36–43.
11. Ким К.С., Смагин В.И. Фильтрация и диагностика в дискретных стохастических системах со скачкообразными параметрами и мультипликативными возмущениями // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2020. № 51. С. 79–86.
12. Janczak D., Grishin Y. State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming // *Control and Cybernetics*. 2006. V. 35 (4). P. 851–862.
13. Gillijns S., Moor B. Unbiased minimum-variance input and state estimation for linear discrete-time systems // *Automatica*. 2007. V. 43. P. 111–116.
14. Smagin V., Koshkin G., Udod V. State estimation for linear discrete-time systems with unknown input using nonparametric technique // *ACSR-Advances in Computer Science Research*. 2015. V. 18. P. 675–677.
15. Smagin V., Koshkin G. Kalman filtering and control algorithms for systems with unknown disturbances and parameters using nonparametric technique // *Proc. 20th International Conference on Methods and Models in Automation and Robotics (MMAR 2015)*, 24–27 August 2015, Miedzyzdroje, Poland. P. 247–251.
16. Koshkin G., Smagin V. Kalman filtering and forecasting algorithms with use of nonparametric functional estimators // *Springer Proc. in Mathematical Statistics / R. Cao et al. (Eds.)*. 2016. V. 175. P. 75–84.
17. Kim K.S., Smagin V.I. Robust filtering for discrete systems with unknown inputs and jump parameters // *Automatic Control and Computer Sciences*. 2020. V. 54 (1). P. 1–9.
18. Barmish B.R., Polyak B.T. A new approach to open robustness problems based on probabilistic predication formulae // *Proc. 13th World IFAC Congr.*, 30 June – 5 July, San Francisco, USA. 1996. V. H. P. 1–6.
19. Athans M. The matrix minimum principle // *Information and Control*. 1968. V. 11. P. 592–606.

References

1. Ugrinovskii, V.A. & Pota, H.R. (2005) Decentralized control of power systems via robust control of uncertain Markov jump parameter systems. *International Journal of Control*. 78. pp. 662–677.
2. Sales-Setien, E. & Penarrocha-Alos, I. (2019) Markovian jump system approach for the estimation and adaptive diagnosis of decreased power generation in wind farms. *Iet Control Theory and Applications*. 13(18). pp. 3006–3018.
3. Zhu, Y., Zhong, Z., Zheng, W.X. et al. (2018) HMM-based H-infinity filtering for discrete-time Markov jump LPV systems over unreliable communication channels. *IEEE Trans. on Systems Man Cybernetics-Systems*. 48(12). pp. 2035–2046.
4. Wang, J., Yao, F. & Shen, H. (2014) Dissipativity-based state estimation for Markov jump discrete-time neural networks with unreliable communication links. *Neurocomputing*. 139/SI. pp. 107–113.
5. Yao, X., Wu, L. & Zheng, W.X. (2011) Fault detection filter design for Markovian jump singular systems with intermittent measurements. *IEEE Transactions on Signal Processing*. 59/7. pp. 3099–3109.
6. Gagliardi, G., Casavola, A. & Famularo, D.A. (2012) Fault detection and isolation filter design method for Markov jump linear parameter-varying systems. *International Journal of Adaptive Control and Signal Processing*. 26(3/SI). pp. 241–257.
7. Costa, O.L.V. & Benites, G.R.A.M. (2011) Linear minimum mean square filter for discrete-time linear systems with Markov jumps and multiplicative noises. *Automatica*. 47(3). pp. 466–476.
8. Liu, W. (2013) On State Estimation of Discrete-Time Linear Systems with Multiplicative Noises and Markov Jumps. *32nd Chinese Control Conference*. Xian, China. July 26–28. pp. 3744–3749.
9. Costa, O.L.V. & Benites, G.R.A.M. (2013) Robust mode-independent filtering for discrete-time Markov jump linear systems with multiplicative noises. *International Journal of Control*. 86(5). pp. 779–793.
10. Lomakina, S.S. & Smagin, V.I. (2005) Robustnaya fil'tratsiya v nepreryvnykh sistemakh so skachkoobraznymi izmeneniyami parametrov v sluchaynye momenty vremeni [Robust filtering in continuous systems with step-like parameters changes at random time moment]. *Avtometriya – Optoelectronics, Instrumentation and Data Processing*. 41(2). pp. 36–43.
11. Kim, K.S. & Smagin, V.I. (2020) Filtration and diagnostics in discrete stochastic systems with jump parameters and multiplicative perturbations. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 51. pp. 79–86. DOI: 10.17223/19988605/51/9
12. Janczak, D. & Grishin, Y. (2006) State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming. *Control and Cybernetics*. 35(4). pp. 851–862.
13. Gillijns, S. & Moor, B. (2007) Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*. 43. pp. 111–116.
14. Smagin, V., Koshkin, G. & Udod, V. (2015) State estimation for linear discrete-time systems with unknown input using nonparametric technique. *ACSR-Advances in Computer Science Research*. 18. pp. 675–677.
15. Smagin, V. & Koshkin, G. (2015) Kalman filtering and control algorithms for systems with unknown disturbances and parameters using nonparametric technique. *Proceedings of the 20th International Conference on Methods and Models in Automation and Robotics (MMAR 2015)*. 24–27 August 2015. Miedzyzdroje, Poland. pp. 247–251.
16. Koshkin, G. & Smagin, V. (2016) Kalman filtering and forecasting algorithms with use of nonparametric functional estimators. In: Cao, R. et al. (eds) *Springer Proceeding in Mathematical Statistics*. Vol. 175. pp. 75–84.
17. Kim, K.S. & Smagin, V.I. (2020) Robust filtering for discrete systems with unknown inputs and jump parameters. *Automatic Control and Computer Sciences*. 54(1). pp. 1–9.
18. Barmish, B.R. & Polyak, B.T. (1996) A new approach to open robustness problems based on probabilistic predication formulae. *Proceedings of the 13th World IFAC Congr.* 30 June – 5 July. San Francisco. USA. vol. H. pp. 1–6.
19. Athans, M. (1968) The matrix minimum principle. *Information and Control*. 11. pp. 592–606.

Информация об авторах:

Ким Константин Станиславович – кандидат физико-математических наук, доцент Высшей IT школы Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: kks93@rambler.ru

Смагин Валерий Иванович – профессор, доктор технических наук, профессор кафедры прикладной математики Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: vsm@mail.tsu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Kim Konstantin S. (Candidate of Physical and Mathematical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: kks93@rambler.ru

Smagin Valery I. (Doctor of Technical Science, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: vsm@mail.tsu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 15.11.2024; принята к публикации 03.03.2025

Received 15.11.2024; accepted for publication 03.03.2025

Научная статья

УДК 681.516

doi: 10.17223/19988605/70/8

**Построение множеств простых путей между двумя узлами
в нечетких транспортных сетях****Дмитрий Васильевич Сперанский¹, Сергей Александрович Лунев²**^{1,2} *Российский университет транспорта (МИИТ), Москва, Россия*¹ *Speranskiy.dv@gmail.com*² *slunev@mail.ru*

Аннотация. Рассмотрена задача построения множества всех простых путей в нечетких транспортных сетях, соединяющих два заданных в ней узла. Ранее для решения этой и других аналогичных задач для четких транспортных сетей был разработан эффективный математический аппарат с использованием понятия графа простых путей. Показано, что рассматриваемая задача может быть сведена к построению множества простых путей в четкой транспортной сети. Такое сведение основано на преобразовании нечеткого графа (модели нечеткой сети) в соответствующий ему четкий граф.

Ключевые слова: четкие и нечеткие транспортные сети; неориентированные графы; простые пути в графах; методы построения простых путей.

Для цитирования: Сперанский Д.В., Лунев С.А. Построение множеств простых путей между двумя узлами в нечетких транспортных сетях // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 81–90. doi: 10.17223/19988605/70/8

Original article

doi: 10.17223/19988605/70/8

Construction of sets of simple paths between two nodes in fuzzy transport networks**Dmitriy V. Speranskiy¹, Sergey A. Lunev²**^{1,2} *Russian University of Transport (MIIT), Moscow, Russian Federation*¹ *Speranskiy.dv@gmail.com*² *slunev@mail.ru*

Abstract. The problem of constructing a set of all simple paths in fuzzy transport networks connecting two nodes specified in it is considered. Earlier, to solve this and other similar problems for clear transport networks, an effective mathematical apparatus was developed using the concept of a graph of simple paths. It is shown that the problem under consideration can be reduced to the construction of many simple paths in a clear transport network. This reduction is based on the transformation of a fuzzy graph (a fuzzy network model) into a corresponding clear graph.

Keywords: crisp and fuzzy transport networks; undirected graphs; simple paths in graphs; methods of constructing simple paths.

For citation: Speranskiy, D.V., Lunev, S.A. (2015) Construction of sets of simple paths between two nodes in fuzzy transport networks. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 81–90. doi: 10.17223/19988605/70/8

Введение

Транспортная сеть обычно понимается как совокупность линий транспортных связей внутри определенной локации, соединяющих ее транспортные узлы и пункты. Сеть состоит из отдельных транспортных линий (железных или автомобильных дорог, авиалиний, каналов связи, телекоммуникаций, линий электропередач и т.п.). Проблемы построения различных множеств простых путей представляют значительный интерес, поскольку их знание и компактное представление необходимы для оперативного управления такими сетями. Число этих путей, как правило, велико, и представлять их в виде множества затруднительно. В статьях И.А. Головинского [1, 2] предложено представлять совокупность всех простых путей в виде неориентированного графа, который назван графом путей. В этих статьях развит математический аппарат для решения четырех различных задач построения простых путей. Граф путей оказался работоспособным для масштабных сетей и может интерпретироваться как общее решение задач соединения множеств вершин простыми путями.

Описания работы сетей в [1, 2] не содержат никаких неопределенностей (нечеткостей), и потому их можно назвать четкими сетями. Их математическими моделями являются классические неориентированные графы. Однако такие модели неприменимы, если описание реально реализуемой конфигурации самой сети или ее функционирования отсутствует или получить его сложно или невозможно.

Современная наука пришла к пониманию, что некоторые наши знания и связи с внешним миром не соответствуют сложившимся ранее классическим представлениям о них. Сейчас разрабатываются новые подходы, предполагающие принципиальную невозможность обойтись без нечеткостей, которые принимаются как реальность человеческого существования. Это приводит к необходимости разработки новых понятий и методов, где такая нечеткость должна быть учтена в практических приложениях.

В статье Л. Заде [3] было введено понятие нечеткого множества, ставшего основой для создания ныне широко используемой в приложениях теории нечетких множеств и других нечетких объектов. Ныне к одному из активно развивающихся направлений относится теория нечетких графов. Интерес к ней в последнее десятилетие связан с использованием ее результатов в важных практических приложениях (проблемы живучести систем, оценки информационной надежности сложных систем и др.).

В предлагаемой статье нечеткие графы использованы как математические модели нечетких транспортных сетей (НТС). Некоторые используемые здесь понятия и термины будут напомнены. Если их определения не приведены, то они трактуются как общепринятые в теории четких графов (см., напр.: [4–6]) и в монографии по теории нечетких множеств [7].

В описываемой модели НТС феномен нечеткости проявляется в переходах по ребрам между вершинами графа при прохождении пути в сети. Этот процесс носит стохастический характер. Выбор очередной вершины для перехода в одну из вершин множества альтернативных вершин зависит от степени принадлежности соответствующих ребер нечеткого графа. Показано, что развитый в [1, 2] математический аппарат для решения задач построения простых путей в четких сетях применим и для решения рассматриваемой нами задачи построения простых путей между двумя заданными узлами в НТС. Идея ее решения для НТС состоит в преобразовании заданного исходного нечеткого графа в соответствующий ему некоторый четкий граф. Описывается способ такого преобразования, и доказывается, что множество всех простых путей в исходном нечетком графе, потенциально возникающих в процессе всех ее возможных конкретных реализаций, совпадает с множеством всех простых путей построенного четкого графа.

1. Некоторые основные понятия и определения

Напомним некоторые используемые понятия из [7]. Пусть множество A есть подмножество множества E . Принадлежность элемента x множеству A записывается с использованием характеристической функции $\mu_A(x)$, принимающей любое неотрицательное значение. Содержательно $\mu(x_i)$ определяет степень принадлежности элемента x_i подмножеству A .

Запись вида

$$A = \{x_1 \mid \mu(x_1), x_2 \mid \mu(x_2), \dots, x_n \mid \mu(x_n)\}$$

является заданием нечеткого подмножества A множества E .

Поскольку здесь используется функция принадлежности, необходимо располагать методами ее получения. Разработка методов построения функций принадлежности началась с введением понятия нечеткого множества и продолжается и ныне. Эта проблема лежит вне рамок статьи, и мы не будем на ней останавливаться. По этой тематике имеется много публикаций, в том числе монографии и обзоры соответствующих методов. Укажем некоторые из них: [8–10]. Известно, что функция принадлежности может быть выражена также и в виде значения вероятности.

Напомним определение соответствующего нечеткого графа из [7]. Пусть E_1, E_2 – два множества, и пусть элемент $x \in E_1$, а $y \in E_2$. Множество упорядоченных пар (x, y) определяет декартово произведение $E_1 \times E_2$. Нечеткое подмножество G такое, что

$$\forall (x, y) \in E_1 \times E_2 \quad \mu_G(x, y) \in M,$$

где M – множество принадлежностей множества $E_1 \times E_2$, называется нечетким графом.

Нечеткий неориентированный граф $G = (V, E)$, где V и E – множества его вершин и ребер, задается и графически изображается как классический четкий граф. Две его вершины γ, δ могут быть соединены ребром, обозначаемым как $e(\gamma, \delta)$, или e_i .

Уточним используемые далее другие понятия из классической теории графов [11]. *Маршрут* в неориентированном графе есть конечная или бесконечная последовательность ребер и вершин $(\dots, a_0, e_0, a_1, e_1, \dots, e_{n-1}, a_n, \dots)$, где каждые два соседних ребра e_{i-1} и e_i имеют общую инцидентную вершину.

Путь – это маршрут, в котором все ребра различны. *Путь* называется *простым*, если все вершины в нем различны.

В качестве математических моделей перечисленных понятий в нечетком графе G будут использованы последовательности ребер и вершин, удовлетворяющие указанным в их определениях свойствами.

В статье используется и ряд других общепринятых понятий и операций над графами из классической теории графов, поэтому приводить их все затруднительно. Если некоторые понятия используются, но их определения в статье не приводятся, то они понимаются так, как определены в [1].

2. Формулировка задачи и математическая модель нечеткой транспортной сети

Рассматриваемая далее задача аналогична первой из исследованных в [1] для четких транспортных сетей. Приведем ее формулировку для НТС. Требуется найти в заданной НТС совокупность всех простых путей, потенциально возникающих в процессе всех ее возможных конкретных реализаций, которые соединяют две заданные вершины нечеткого графа G , моделирующего НТС.

В НТС проход по пути иногда связан с переходом в одну из следующих вершин по множеству альтернативных ребер. Выбор конкретного ребра зависит от степеней их принадлежности всему множеству ребер НТС и носит стохастический (вероятностный) характер. Это служит причиной возникновения различных вариантов конкретных реализаций траекторий прохода по одному и тому же пути.

Используемая здесь математическая модель НТС представляет собой нечеткий неориентированный граф $G = (V, E)$ без петель и кратных ребер. Для краткости будем иногда называть эту модель просто графом G . В такой модели в отличие от четкого графа множество V в общем случае содержит вершины трех разных типов. Вершина первого типа – это обычная одиночная вершина (как в четком графе), далее она именуется канонической. Вершина второго типа есть множество канонических вершин с общим числом не менее двух, далее они именуются a -вершинами (альтернативными вершинами). Вершины третьего типа могут содержать в общем случае несколько множеств a -вершин (здесь каноническая вершина считается вырожденной a -вершиной).

В процессе прохода по любому пути в графе G происходят переходы между парами его вершин α и β (в общем случае это вершины разного типа) по ребру $e(\alpha, \beta)$. Эти переходы будем изображать графически как $\alpha - \beta$. Поскольку теперь в графе используются также ребра разного типа, то, казалось бы, нужно изменить приведенные выше определения маршрута, пути и простого пути, а также методы вычисления степеней принадлежности вершин и ребер. В действительности в этом нет необходимости. Во введении сказано, что решение соответствующей задачи для оперативного управления НТС сводится к решению аналогичной задачи для четких сетей. Но модель четкой сети содержит только канонические вершины. Таким образом, искомые простые пути для НТС, найденные в четком графе, содержательно интерпретируются точно так же, как и в четкой сети. Что касается разработки специальных методов вычисления степеней принадлежности, то по той же причине можно обойтись и без них.

Функция принадлежности нечеткого множества напоминает плотность распределения вероятностей. Отличие их в том, что сумма вероятностей по всем возможным значениям случайной величины всегда равна 1, а сумма S значений функции принадлежности может быть любым неотрицательным числом. Тогда можно свести функцию принадлежности к функции плотности вероятностей ее нормированием, разделив все ее значения на S .

При задании НТС (нечеткого графа) необходимо указать степени принадлежности не только для канонических, но и для вершин второго и третьего типов. Для двух последних типов их степени принадлежности, очевидно, следует положить равными сумме степеней составляющих вершину элементов. Так, например, если $w = (v_1, \dots, v_v)$ – вершина второго типа, то степень ее принадлежности полагается равной значению $\mu(w) = \sum_{i=1}^v \mu(v_i)$, где слагаемые $\mu(v_i)$ в сумме учитываются столько раз,

сколько v_i участвуют в качестве компонентов во всех вершинах множества заданного нечеткого графа. Это и есть S , упомянутое выше, которое используется при замене функции принадлежности значениями вероятностей. Знание степеней принадлежности (вероятностей) «неканонических» вершин может потребоваться при выборе продолжения траектории при построении простого пути.

Оценка степеней принадлежности ребер (вершин) полученного четкого графа (например, группой экспертов) может производиться и после завершения процесса преобразования исходного нечеткого графа с целью решения других задач, не связанных с построением простых путей.

Задание графа сводится к описанию множества его вершин, ребер и отношения инцидентности [1, 11]. Для конечных графов для этого достаточно пронумеровать множества V и E . Граф можно задать матрицей инцидентности, но в ней каждая строка имеет только два элемента, отличных от нуля, и поэтому такой способ не является экономным. Отношение инцидентности более экономно задать списком ребер, в котором для них указываются номера вершин, инцидентных ему, и степени их принадлежности.

3. Преобразование нечеткого графа G в четкий граф

Исходной информацией для рассматриваемой задачи является задание нечеткого графа $G = (V, E)$, из которого для каждой его вершины $v \in V$ находятся все ее смежные вершины. Предполагается, что для каждого ребра $e(\alpha, \beta) \in E$ заданы значения степеней принадлежности всех смежных вершин с вершинами α и β этого ребра (возможно в виде вероятности). При использовании списка ребер он может представляться в виде таблицы, имеющей два столбца. В первом столбце указываются все ребра $e(\alpha, \beta)$ графа G , а в той же строке второго столбца – смежные вершины, разрешенные для использования, и значения их степеней принадлежности.

Напомним, что сведение рассматриваемой задачи для заданной НТС к аналогичной задаче для четких сетей предлагается осуществлять путем преобразования заданного исходного нечеткого графа G в соответствующий ему некоторый четкий граф $\tilde{G}$.

Каждый этап преобразования состоит в замещении выбранного перехода графа G некоторым подграфом, вид которого определяется типом конечных вершин перехода. Для преобразования графа $G = (V, E)$ (кроме задания самого графа) для каждой вершины из V задается также множество исполь-

зуемых (разрешенных) смежных с ней вершин. Заметим, что в общем случае использоваться может только некоторое подмножество разрешенных для заданной вершины ее смежных вершин.

Опишем в целом процесс преобразования нечеткого графа G в четкий граф $\tilde{G}$. Граф G является «стартовым» для получения графа $\tilde{G}$. Преобразование выполняется поэтапно (пошагово). На очередном этапе в текущем полученном графе произвольным образом выбирается переход между двумя вершинами, по крайней мере одна из которых есть a -вершина. Выбранный переход замещается фрагментом (подграфом), определяемым видом выбранного перехода. После завершения этапа получается новый граф, в котором вершины в общем случае могут быть любого из трех типов. В полученном после этого этапа преобразования новом графе производится аналогичное замещение соответствующим фрагментом другого выбранного перехода между парой вершин, одна из которых не является канонической. Этот процесс продолжается далее до тех пор, пока после выполнения очередного этапа в текущем преобразуемом графе все его вершины не окажутся каноническими. Будет показано, что этот граф и является искомым четким графом $\tilde{G}$ для получения решения рассматриваемой задачи.

С учетом наличия разных типов вершин в графе G возможны следующие виды переходов между ними:

- 1) $v_i - v_j$, где обе вершины канонические;
- 2) $v - w$, где v – каноническая вершина, а w – a -вершина;
- 3) $w_i - w_j$, где обе вершины являются a -вершинами.

Здесь и далее буквами v, w (с индексами и без них) обозначаются соответственно канонические и a -вершины.

Переходы первого вида – это стандартные переходы между каноническими вершинами, изображаемые в виде ребра, связывающего его концевые вершины. Эти переходы не подлежат замещению.

Опишем фрагменты (подграфы), которые замещают остальные виды переходов в поэтапно возникающих новых графах.

Если a -вершину записывать как множество $\{v_1, v_2, \dots, v_k\}$, то переход второго вида представляется как $v - \{v_1, v_2, \dots, v_k\}$, а переход третьего вида в общем случае можно представить, например, в виде: $\{w_1, w_2, \dots, w_k\} - \{\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_q\}$.

При преобразовании некоторого перехода нечеткого графа (в модели НТС) участвующее в нем ребро может порождать несколько ребер, его замещающих.

Рассмотрим переход второго вида $v - \{v_1, v_2, \dots, v_k\}$. Содержательно его интерпретация зависит от направления движения по пути в графе. При направлении движения от v к $\{v_1, v_2, \dots, v_k\}$ в вершине v происходит разветвление некоторого пути на k ветвей (v_1, v_k). При этом значение степени принадлежности $\mu(v)$ вершины v должно быть распределено (например, экспертами) между степенями принадлежности смежных вершин возникающих k ветвей (v_1, v_k) так, чтобы $\mu(v) = \mu(v_1) + \dots + \mu(v_k)$. При синтезе четкого графа $\tilde{G}$ этот переход замещается фрагментом, изображенным на рис. 1.

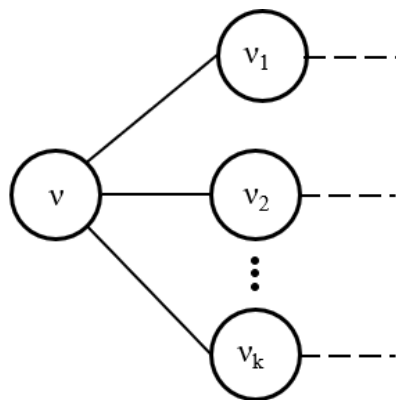


Рис. 1. Фрагмент подграфа для замещения переходов 2-го типа
Fig. 1. A fragment of a subgraph for replacing type 2 transitions

При движении по пути в противоположном направлении происходит слияние ветвей (v_k, v_l) в одну вершину v , и этот переход замещается фрагментом (подграфом), аналогичным фрагменту на рис. 1, в котором нужно поменять местами левую и правую части (зеркальное отображение рис. 1 относительно вертикальной оси). Понятно, что степень принадлежности вершины v в этом случае при заданных значениях принадлежностей для вершин v_i следует положить равной $\mu(v) = \mu(v_1) + \dots + \mu(v_k)$.

Опишем фрагменты, замещающие переходы третьего вида. Общий принцип их построения покажем на примере. Пусть имеется переход между двумя a -вершинами $\{\gamma_1, \gamma_2, \gamma_3\} - \{\delta_1, \delta_2\}$. В нашей модели НТС она будет замещаться следующими парами переходов: $\gamma_1 - \delta_1, \gamma_1 - \delta_2, \gamma_2 - \delta_1, \gamma_2 - \delta_2, \gamma_3 - \delta_1, \gamma_3 - \delta_2$. Пусть имеется ограничение на использование переходов $\gamma_3 - \delta_1, \gamma_3 - \delta_2$. Предположим, что в остальных переходах участвуют следующие a -вершины:

$$\gamma_1 = \{\tilde{v}_1, \tilde{v}_2\}, \gamma_2 = \{\tilde{v}_1, \tilde{v}_2\}, \delta_1 = \{v_1^1, v_2^1\}, \delta_2 = \{v_1^2, v_2^2\}.$$

Здесь литера v означает каноническую вершину. В этих обозначениях переход $\{\gamma_1, \gamma_2, \gamma_3\} - \{\delta_1, \delta_2\}$ будет замещен фрагментом из четырех переходов между парами a -вершин, представленным на рис. 2 (переход из вершины γ_3 не разрешен). Эти переходы содержат только a -вершины, и потому потребуются выполнять последующие промежуточные преобразования графа. Например, переход $\gamma_1 - \delta_1$ ($\{\tilde{v}_1, \tilde{v}_2\} - \{v_1^1, v_2^1\}$) на одном из следующих этапов будет замещен (по вышеописанному принципу) фрагментом из четырех переходов $\tilde{v}_1 - v_1^1, \tilde{v}_1 - v_2^1, \tilde{v}_2 - v_1^1, \tilde{v}_2 - v_2^1$. В этих переходах теперь присутствуют только канонические вершины.

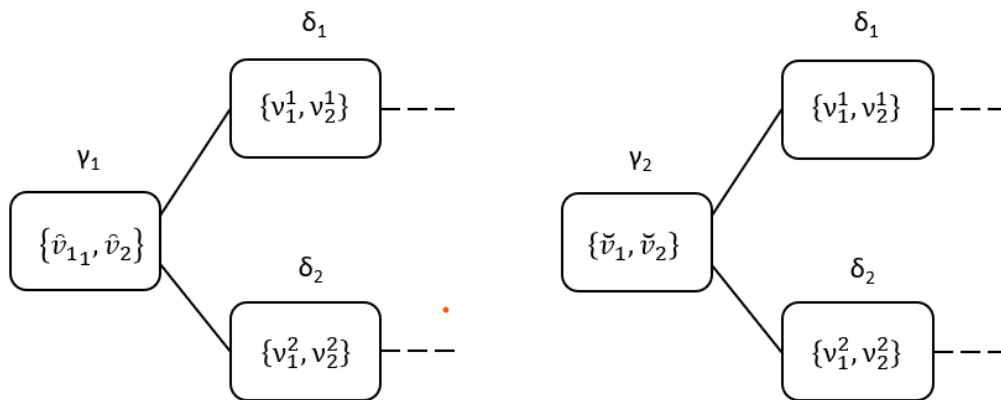


Рис. 2. Фрагменты подграфа для замещения переходов 3-го вида
Fig 2. Fragments of a subgraph to replace transitions of the 3rd type

Понятно, что подобные замещения в графе G за конечное число шагов приведут к тому, что множество вершин после завершающего этапа преобразования графа G будет содержать только канонические вершины, т.е. он является четким. Покажем далее, что этот граф и дает возможность получить решение рассматриваемой задачи.

4. Обоснование метода решения задачи и пример построения графа простых путей

Множество простых путей в классическом графе в [1] предложено представлять в компактной форме с использованием введенного там понятия графа путей. Там же показано, что решение рассматриваемой задачи построения простых путей между двумя заданными вершинами графа (и еще трех аналогичных задач) сводится к вычислениям, которые именуются анализом двусвязности графа. Хорошо известный алгоритм такого анализа был разработан в [12].

Изложенный выше алгоритм преобразования нечеткого графа связан с замещением переходов с a -вершинами на переходы между каноническими вершинами. Из структуры всех использованных фрагментов для замещения очевидно, что наличие в переходе a -вершин означает фактически продолжение пути от этой вершины по разным ветвям графа за счет разветвления. Заметим, что в каждом используемом фрагменте представлены все варианты продолжения простого пути от a -вершины.

Очевидно, что этот факт обеспечивает наличие в построенном графе $\tilde{G}$ всех возможных вариантов простых путей между двумя заданными вершинами α и β в графе G , которые потенциально могут возникнуть в процессе всех конкретных реализаций проходов по путям. Применяя далее к графу $\tilde{G}$ предложенный в [1] математический аппарат, строится граф простых путей, обеспечивающий общее решение рассматриваемой задачи.

Из изложенного вытекает справедливость следующего утверждения:

Теорема 1. Множество всех простых путей четкого графа $\tilde{G}$, полученного из заданного нечеткого графа G описанным выше алгоритмом преобразования, совпадает с множеством всех простых путей, возникающих в графе G при любых возможных конкретных реализациях прохода по пути между двумя заданными вершинами α и β .

Эта теорема является обоснованием предлагаемого метода решения рассматриваемой задачи. Он состоит из предварительного преобразования нечеткого графа в четкий и последующего построения графа простых путей. Наглядность метода и простота восприятия логики предлагаемого решения очевидны, и поэтому теорема не требует дополнительного математического обоснования.

Граф простых путей, построенный для четкого графа $\tilde{G}$, соединяющего две вершины α и β , может быть выражен через структуру двусвязного графа [1]. В рассматриваемой задаче графом путей будет объединение всех простых путей в графе G , соединяющих две вершины α и β , который обозначим $P(G, \alpha, \beta)$. Приводимая ниже дословно *теорема* из [1] показывает, что граф $P(G, \alpha, \beta)$ строится посредством декомпозиции (четкого) графа G .

Теорема 2. Пусть L – простой путь в связном графе G , соединяющий вершины α и β ; $\Theta_0, \Theta_1, \dots, \Theta_n$ – множество таких блоков графа G , которые содержат хотя бы одно из ребер пути L . Тогда:

- 1) семейство блоков $\Theta_0, \Theta_1, \dots, \Theta_n$ не зависит от выбора пути L из множества всех простых путей, соединяющих вершины α и β : оно однозначно определено парой α и β ;
- 2) семейство блоков $\Theta_0, \Theta_1, \dots, \Theta_n$ можно линейно упорядочить в такую цепочку, в которой первый и последний блоки содержат вершины α и β соответственно, а пересечение каждых двух соседних блоков в цепочке состоит из их общего шарнира;
- 3) граф простых путей, соединяющих α и β , есть объединение блоков $\Theta_0, \Theta_1, \dots, \Theta_n$:

$$P(G, \alpha, \beta) = \Theta_0 + \Theta_1 + \dots + \Theta_n$$

(в теореме знаком плюс обозначено объединение блоков).

Далее рассмотрим в качестве простого примера нечеткий граф $G = (V, E)$, изображенный на рис. 3. В нем множество V содержит 9 вершин, из которых 7 – канонические (литеры v_i), а остальные две – a -вершины (литеры w_i): $w_2 = \{v_2^1, v_2^2\}$, $w_7 = \{v_7^1, v_7^2, v_7^3\}$. Множество ребер E содержит 10 ребер (литеры e_i). Выполненный анализ двусвязности графа G показывает, что он состоит из четырех блоков: $\Theta_0 = \{v_5, v_3\}$, $\Theta_1 = \{v_1, w_2, v_4, v_3\}$, $\Theta_2 = \{v_4, v_6\}$, $\Theta_3 = \{v_6, w_7, v_8, v_9\}$. Блоки $\Theta_0 = \{v_5, v_3\}$ и $\Theta_2 = \{v_4, v_6\}$ называются вырожденными (по определению блок должен иметь более двух вершин, либо быть мостом), так как они образуют мосты.

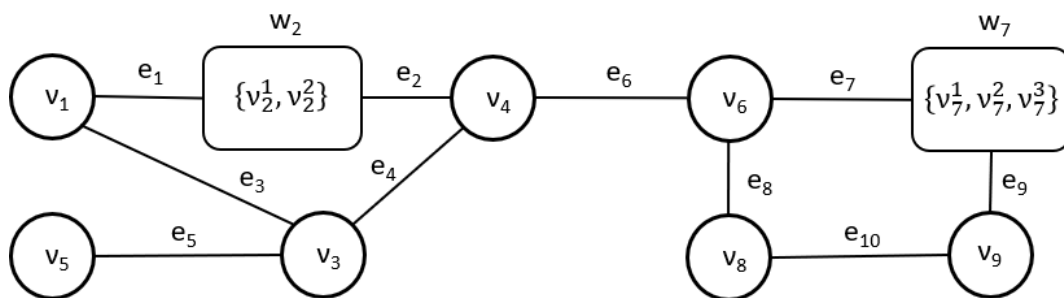


Рис. 3. Пример нечеткого графа для построения простых путей
Fig. 3. An example of a fuzzy graph for constructing simple paths

Блоки Θ_3 и Θ_1 являются по определению компонентами двусвязности графа G , так как они его максимальные двусвязные подграфы.

Вершины v_3 и v_6 являются по определению шарнирами (общая вершина двух блоков). Ребра e_5 , e_6 – мосты, так как по определению они не принадлежат никаким компонентам двусвязности графа.

Легко проверить, что выполнение преобразования графа на рис. 3 описанным выше способом дает четкий граф, изображенный на рис. 4.

Результаты анализа двусвязности показывают, что по теореме 1 четкий граф на рис. 4 представляет собой граф простых путей для рассматриваемого примера.

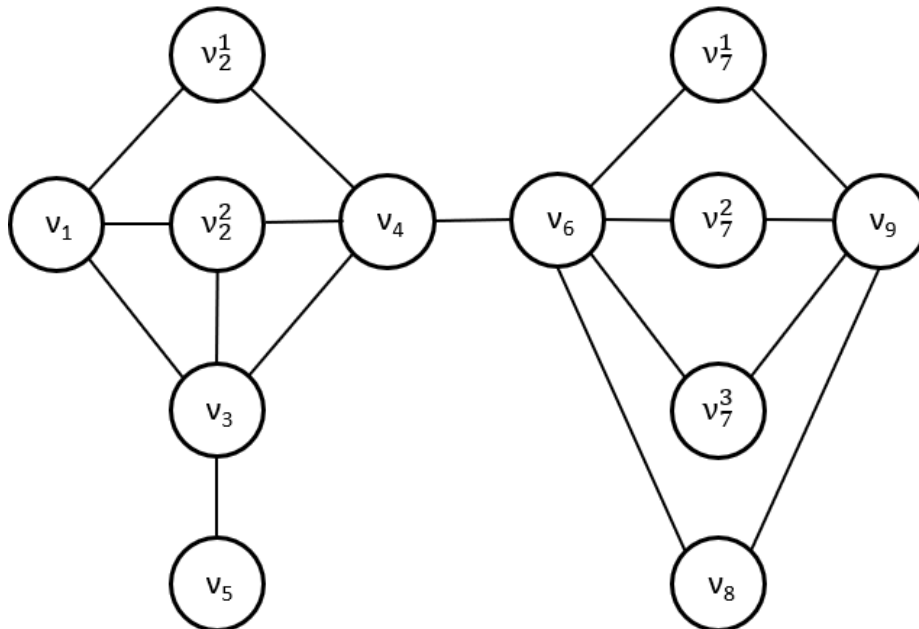


Рис. 4. Четкий граф, полученный после преобразования графа на рис. 3
Fig. 4. A clear graph obtained after the graph transformation in Fig. 3

Используя граф на рис. 4, построим, например, множество всех простых путей между вершинами v_1 и v_9 . Их поиск в графе можно осуществить путем его обхода с посещением всех вершин и исследованием всех ребер. Такой обход можно выполнить многими способами, в частности наиболее распространенными стратегиями – поисками в глубину и ширину [4]. Отметим, что из теоремы 1 в [1] вытекает, что пространство поиска может сократиться при использовании графа простых путей (по сравнению с поиском во всем исходном заданном графе в глубину или в ширину). Заметим, что сам процесс поиска может быть организован более эффективно благодаря линейной упорядоченности блоков и специфики структуры графа простых путей.

В нашем примере благодаря простоте структуры графа на рис. 4 найти все множество его простых путей очень легко. В первом (слева) блоке графа имеется пять потенциальных начальных отрезков простых путей $(v_1 - v_2^1 - v_4; v_1 - v_2^2 - v_4; v_1 - v_3 - v_4; v_1 - v_2^2 - v_3 - v_4; v_1 - v_3 - v_2^2 - v_4)$. Во втором блоке (справа) четкого графа имеется четыре потенциальных завершающих отрезка простых путей $(v_6 - v_7^1 - v_9; v_6 - v_7^2 - v_9; v_6 - v_7^3 - v_9; v_6 - v_8 - v_9)$.

Понятно, что объединенные мостом $v_4 - v_6$ все возможные пары из начальных и завершающих отрезков образуют множество всех простых путей полученного четкого графа. Таким образом, в графе на рис. 4 имеется 20 (5×4) различных простых путей.

Заключение

В статье предложен метод сведения рассмотренной задачи к аналогичной задаче для четких транспортных сетей. Он основан на преобразовании заданной НТС к четкой сети. Алгоритм такого преобразования достаточно прост, что следует из его описания. Он практически очевиден благодаря

наглядности и легко понимаемой логике построения. Отсюда понятна и простота обоснования алгоритма.

В [2] справедливо указывается, что для алгоритмов важна не только их вычислительная сложность, но также логическая простота и содержательная ясность. Два последних качества фактически определяют трудозатраты на отладку и развитие компьютерных программ реализации алгоритмов.

Важность задач построения простых путей объясняется возможностью приложений для решения некоторых актуальных проблем, например указанных в [2]: контроля надежной связности транспортных сетей, обнаружения узких мест в них, визуализации и резервирования, вычисления препятствий при движении по сети и др.

Список источников

1. Головинский И.А. Построение графов простых путей в транспортных сетях. I. Общие примеры и решения // Известия РАН. Теория и системы управления. 2021. № 1. С. 124–159. doi: 10.31857/S0002338821010066
2. Головинский И.А. Построение графов простых путей в транспортных сетях. II. Анализ двусвязности графов // Известия РАН. Теория и системы управления. 2021. № 2. С. 47–61. doi: 10.31857/S0002338821020049
3. Zadeh L. Fuzzy sets // *Information and Control*. 1965. V. 8. P. 338–353.
4. Алексеев В.Е., Таланов В.А. Графы и алгоритмы. М. : БИНОМ. Лаборатория знаний, 2011. 320 с.
5. Харари Ф. Теория графов. М. : Мир, 1973. 300 с.
6. Ахо А.В., Хопкрофт Дж.Э., Ульман Дж.Д. Структуры данных и алгоритмы. М. ; СПб. ; Киев : Вильямс, 2003. 384 с.
7. Кофман А.В. Введение в теорию нечетких множеств. М. : Радио и связь, 1982. 432 с.
8. Нечеткие множества в моделях управления и искусственного интеллекта / под ред. Д.А. Поспелова. М. : Наука, 1986. 312 с.
9. Норвич А.М., Турсон И.Б. Построение функции принадлежности и теория вероятностей : последние достижения : пер. с англ. / под ред. Р.Р. Ягера. М. : Радио и связь, 1986. 408 с.
10. Шопин А.Г. Построение функции принадлежности нечеткого множества и оценка его вероятностных характеристик // Исследовано в России : электронный многопредметный научный журнал. Шопин А.Г. Построение функции принадлежности нечеткого множества и оценка его вероятностных характеристик // Исследовано в России : электронный многопредметный научный журнал. Т. 6. С. 453–467. URL: <https://web.archive.org/web/20040320064103/http://zhurnal.ape.relarn.ru/articles/2003/040.pdf>
11. Кузнецов А.П., Адельсон-Вельский Т.М. Дискретная математика для инженера. М. : Энергоатомиздат, 1988. 480 с.
12. Hopcroft J.E., Tarjan R.E. Efficient Algorithms for Graph Manipulation // *Comm. ACM*. 1973. V. 16 (6), P. 372–378.

References

1. Golovinskiy, I.A. (2021) Postroenie grafov prostykh putey v transportnykh setyakh. I. Obshchie primery i resheniya [Construction of simple path graphs in transport networks: I. General solutions and examples]. *Izvestiya RAN. Teoriya i sistemy upravleniya*. 60(1). pp. 118–152. DOI: 10.31857/S0002338821010066
2. Golovinskiy, I.A. (2021) Postroenie grafov prostykh putey v transportnykh setyakh. II. Analiz dvusvyaznosti grafov [Construction of simple path graphs in transport networks: II. Analysis of graphs' biconnectivity]. *Izvestiya RAN. Teoriya i sistemy upravleniya*. 60(2). pp. 213–226. DOI: 10.31857/S0002338821020049
3. Zadeh, L. (1965) Fuzzy sets. *Information and Control*. 8. pp. 338–353.
4. Alekseev, V.E. & Talanov, V.A. (2011) *Grafy i algoritmy* [Graphs and Algorithms]. Moscow: BINOM. Laboratoriya znaniy.
5. Harari, F. (1973) *Teoriya grafov* [Graph Theory]. Moscow: Mir.
6. Aho, A.V., Hopcroft, G.E. & Ulman, G.D. (2003) *Struktury dannykh i algoritmy* [Data Structures and Algorithms]. Moscow; St. Petersburg; Kiev: Vil'yams.
7. Kofman, A.V. (1982) *Vvedenie v teoriyu nechetkikh mnozhestv* [Introduction to the Theory of Fuzzy Sets]. Moscow: Radio i svyaz'.
8. Pospelova, D.A. (ed.) (1986) *Nechetkie mnozhestva v modelyakh upravleniya i iskusstvennogo intellekta* [Fuzzy sets in control and artificial intelligence models]. Moscow: Nauka.
9. Norvich, A.M. & Turson, I.B. (1986) *Postroenie funktsii prinadlezhnosti i teoriya veroyatnostey: poslednie dostizheniya* [Construction of the membership function and probability theory: Recent achievements]. Translated from English. Moscow: Radio i svyaz'.
10. Shopin, A.G. (2003) Postroenie funktsii prinadlezhnosti nechetkogo mnozhestva i otsenka ego veroyatnostnykh kharakteristik [Construction of the membership function of a fuzzy set and evaluation of its probabilistic characteristics]. *Issledovano v Rossii*. 6. pp. 453–467. [Online] Available from: <https://web.archive.org/web/20040320064103/http://zhurnal.ape.relarn.ru/articles/2003/040.pdf>
11. Kuznetsov, A.P. & Adel'son-Vel'skiy, T.M. (1988) *Diskretnaya matematika dlya inzhenera* [Discrete Mathematics for Engineers]. Moscow: Energoatomizdat.
12. Hopcroft, J.E. & Tarjan, R.E. (1973) Efficient Algorithms for Graph Manipulation. *Comm. ACM*. 16(6). pp. 372–378.

Информация об авторах:

Сперанский Дмитрий Васильевич – профессор, доктор технических наук, профессор кафедры «Системы управления транспортной инфраструктурой» Российского университета транспорта (МИИТ) (Москва, Россия). E-mail: speranskiy.dv@gmail.com

Лунев Сергей Александрович – доцент, кандидат технических наук, доцент кафедры «Системы управления транспортной инфраструктурой» Российского университета транспорта (МИИТ) (Москва, Россия). E-mail: slunev@mail.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Speranskiy Dmitriy V. (Doctor of Technical Sciences, Professor of the Department of Transportation Infrastructure Management Systems of Russian University of Transport (MIIT), Moscow, Russian Federation). E-mail: speranskiy.dv@gmail.com

Lunev Sergey A. (Candidate of Technical Sciences, Associate Professor of the Department of Transportation Infrastructure Management Systems of Russian University of Transport (MIIT), Moscow, Russian Federation). E-mail: slunev@mail.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 31.10.2024; принята к публикации 03.03.2025

Received 31.10.2024; accepted for publication 03.03.2025

Научная статья
УДК 004.855.5
doi: 10.17223/19988605/70/9

Нейросетевое тестирование алгоритма разметки данных для классификации уровней поддержки и сопротивления на финансовых рынках

Марк Альбертович Хайров¹, Владимир Григорьевич Спицын²

^{1,2} *Национальный исследовательский Томский политехнический университет, Томск, Россия*

² *Национальный исследовательский Томский государственный университет, Томск, Россия*

¹ *mah9@tpu.ru,*

² *spvg@tpu.ru*

Аннотация. Представлен алгоритм разметки данных для обучения классификатора, способного определять уровни поддержки и сопротивления на данных финансовых рынков. На основе данных, полученных с помощью этого алгоритма, была обучена модель CNN-LSTM-MLP, включающая блок каузальных сверток, LSTM и полносвязные слои. Эксперимент с простой торговой стратегией продемонстрировал практическую применимость модели, сопровождавшуюся повышением доходности на 10% в сравнении с базовой стратегией, где модель не была использована.

Ключевые слова: данные рыночных котировок; глубинное обучение; причинно-следственная связь; уровни поддержки; уровни сопротивления.

Для цитирования: Хайров М.А., Спицын В.Г. Нейросетевое тестирование алгоритма разметки данных для классификации уровней поддержки и сопротивления на финансовых рынках // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 91–102. doi: 10.17223/19988605/70/9

Original article
doi: 10.17223/19988605/70/9

Neural network testing of a data labeling algorithm for classifying support and resistance levels in financial markets

Mark A. Khairov¹, Vladimir G. Spitsyn²

^{1,2} *National Research Tomsk Polytechnic University, Tomsk, Russian Federation*

² *National Research Tomsk State University, Tomsk, Russian Federation*

¹ *mah9@tpu.ru*

² *spvg@tpu.ru*

Abstract. An algorithm for labeling data is presented to train a classifier capable of identifying support and resistance levels in financial market data. Using data obtained from this algorithm, a CNN-LSTM-MLP model was trained, incorporating causal convolutions, LSTM, and fully connected layers. An experiment with a simple trading strategy demonstrated the practical applicability of the model, accompanied by a 10% increase in profitability compared to a basic strategy where the model was not used.

Keywords: market quote data; deep learning; causal convolution; support levels, resistance levels.

For citation: Khairov, M.A., Spitsyn, V.G. (2025) Neural network testing of a data labeling algorithm for classifying support and resistance levels in financial markets. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 91–102. doi: 10.17223/19988605/70/9

Введение

Уровни поддержки и сопротивления являются ключевыми концепциями в техническом анализе финансовых рынков, обеспечивающими важные ориентиры для трейдеров и инвесторов.

Уровень поддержки – это ценовой уровень, на котором наблюдается усиление спроса на актив, достаточное для предотвращения дальнейшего снижения его цены. При достижении уровня поддержки цена актива, как правило, стабилизируется или начинает повышаться, поскольку покупатели вступают в рынок, считая цену привлекательной для покупки.

Уровень сопротивления, напротив, представляет собой ценовой уровень, на котором усиливается предложение актива, что препятствует дальнейшему росту его цены. На этом уровне продавцы начинают активно продавать актив, считая текущую цену выгодной для продажи, что приводит к стабилизации или снижению цены.

Уровни поддержки и сопротивления обладают некоторой предсказательной способностью, что было показано в работах С. Osler, F. Garzarelli, K. Chung и A. Bellotti [1–3]. При том что уровни поддержки и сопротивления используются на протяжении нескольких десятков лет практикующими инвесторами и трейдерами, в академической литературе еще не было представлено алгоритма разметки данных с целью обучения классификатора для предсказания уровней поддержки и сопротивления. К тому же у других авторов нет работ по построению моделей глубокого обучения для предсказания образования уровней поддержки и сопротивления.

С другой стороны, для задачи, частично схожей с определением уровней поддержки и сопротивления предсказанием по определению направления движения цены, существуют примеры успешно обученных нейросетевых классификаторов. Так, в работе J. Sirignano и R. Cont [4] была обучена модель LSTM, качество которой не менялось со временем, также позволявшая прогнозировать динамику цен для активов, которых не было в обучающей выборке. В работе Z. Zhang и соавт. [5] представлен пример построения классификатора с высокой точностью, которая не снижается во времени. Представленный классификатор состоял из сверточного блока, блока Inception [6] и блока LSTM. Как отмечают авторы, такая архитектура позволяет модели обучаться извлекать универсальные и наиболее важные признаки.

В настоящей работе впервые представлена методика разметки данных для обучения классификатора для детектирования уровней поддержки и сопротивления. С использованием данных, полученных при помощи разработанной методики разметки, были обучены классификаторы разного типа, такие как логистическая регрессия (LR), многослойный персептрон и LSTM. Для повышения качества классификаторов использовались карты признаков, сгенерированные при помощи специального модуля генерации признаков на основе одномерных каузальных сверток.

Было продемонстрировано возможность практического применения представленных методик на примере простейшей торговой симуляции.

1. Исходные данные

В настоящей работе используются данные о лучших ценах на продажу (best ask price, $a(t)$) и покупку (best bid price, $b(t)$) активов, а также соответствующие объемы имеющихся предложений на покупку и продажу активов.

Лучшая цена на продажу в момент времени t ($a(t)$) – это цена предложения о продаже, по которой возможно наиболее выгодно продать 1 ед. актива в момент времени t . Лучшая цена на покупку в момент времени t ($b(t)$) – это цена предложения о покупке, по которой возможно наиболее выгодно купить 1 ед. актива в момент времени t .

Объемами на продажу или покупку мы называем количество акций, которые можно продать или купить по предложенной цене.

В дальнейшем будет рассматриваться средняя цена $m(t) = \frac{a(t) + b(t)}{2}$, которая лежит между best bid и best ask.

Пример исходных данных

participant_timestamp	Цена на продажу (ask price)	Объем на продажу (ask volume)	Цена на покупку (bid price)	Объем на покупку (bid volume)
1674161999987619167	135,29	1	135,27	12
1674161999988218000	135,30	44	135,28	19
1674161999988468649	135,28	43	135,28	19
...	...	...	...	...

Для эксперимента были использованы данные о best bid и best ask ценах и объемах для акций AAPL, GE, MSFT и NVDA с 19 января 2023 г. по 22 мая 2023 г. В этом периоде было 83 торговых дня. Данные были получены от провайдера Polygon.io. В табл. 1 представлен пример данных из датасета.

2. Алгоритм разметки данных для обучения классификатора

Введем формальные определения для уровней поддержки и сопротивления, которые будут использоваться в алгоритме разметки данных.

Определение. Уровень поддержки в момент времени t определяется как цена $S_t = m(t) \pm \pi$, где $m(t)$ является локальным минимумом на интервале $[t-T, t+T]$, т.е. $m(t) \leq m(\tau) \pm \pi$ для любого $\tau \in [t-T, t+T]$.

Определение. Уровень сопротивления в момент времени t как цена $R_t = m(t) \pm \pi$, где $m(t)$ является локальным максимумом на интервале $[t-T, t+T]$, т.е. $m(t) \geq m(\tau) \pm \pi$ для любого $\tau \in [t-T, t+T]$.

Алгоритм разметки представляет собой последовательность шагов, направленных на автоматическое нахождение примеров для обучения классификатора. Задача классификатора состоит в том, чтобы среди потенциальных уровней поддержки и сопротивления отделить те уровни цен, которые образуют кластеры поддержки и сопротивления

Определение. Потенциальный уровень поддержки в момент времени t определяется как цена $PS_t = m(t) \pm \pi$, где $m(t)$ является локальным минимумом на интервале $[t-T, t]$, т.е. $m(t) \leq m(\tau) \pm \pi$ для всех $\tau \in [t-T, t]$.

Определение. Потенциальный уровень сопротивления в момент времени t определяется как цена $PR_t = m(t) \pm \pi$, где $m(t)$ является локальным максимумом на интервале $[t-T, t]$, т.е. $m(t) \geq m(\tau) \pm \pi$ для всех $\tau \in [t-T, t]$.

Определение. Кластер поддержки определяется как упорядоченная последовательность уровней поддержки $CS = (S_1, S_2, \dots, S_n)$, где каждому S_i соответствует время t_i из $t_1 < t_2 < \dots < t_n$, при этом для каждого $i = 1, 2, \dots, n$, $|S_i - S_1| \leq \pi$, для каждого $t \in [t_1, t_n]$ $m(t) \geq m(t_1) - \pi$ и $t_{i+1} - t_i \leq K$ для $i = 1, 2, \dots, n-1$.

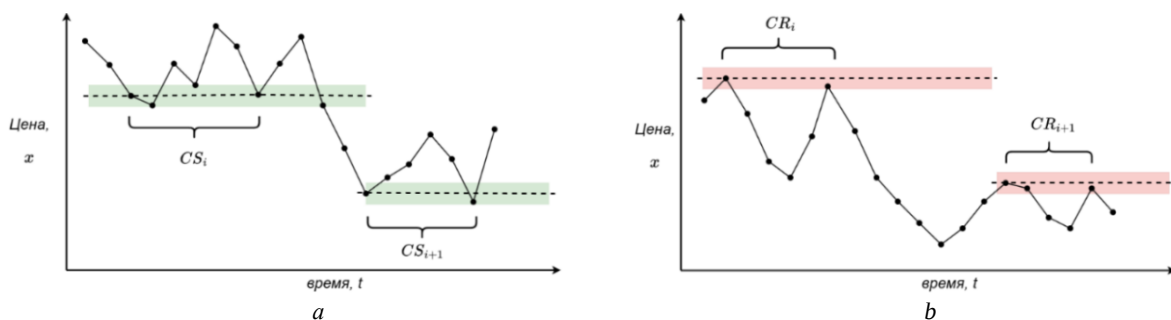


Рис. 1. Кластеры поддержки и сопротивления: а – кластеры поддержки; б – кластеры сопротивления

Fig. 1. Support and resistance clusters: a – support clusters; b – resistance clusters

Определение. Кластер сопротивления определяется как упорядоченная последовательность уровней сопротивления $CR = (R_1, R_2, \dots, R_n)$, где каждому R_i соответствует время t_i из $t_1 < t_2 < \dots < t_n$, при этом для каждого $i = 1, 2, \dots, n$, $|R_i - R_1| \leq \pi$, для каждого $t \in [t_1, t_n]$ $m(t) \leq m(t_1) + \pi$ и $t_{i+1} - t_i \leq K$ для $i = 1, 2, \dots, n-1$.

На рис. 1 представлены примеры кластеров поддержки и сопротивления.

Таким образом упорядоченную последовательность кластеров поддержки можно обозначить как $((S_{i,j})_{j=1}^k)_{i=1}^n$, а упорядоченную последовательность кластеров сопротивления как $((R_{i,j})_{j=1}^k)_{i=1}^n$.

Для каждого i -го наблюдения производятся:

1. Нахождение уровней цены:
 - а. Для уровней поддержки:
 - i. Если цена m_t является локальным минимумом на интервале $[t-T, t]$, то это потенциальный уровень поддержки PS_t .
 - ii. Если потенциальный уровень поддержки PS_t является локальным минимумом на интервале $(t, t+T]$, то это уровень поддержки S_t .
 - б. Для уровней сопротивления:
 - i. Если цена m_t является локальным максимумом на интервале $[t-T, t]$, то это потенциальный уровень сопротивления PR_t .
 - ii. Если потенциальный уровень сопротивления PR_t является локальным максимумом на интервале $(t, t+T]$, то это уровень сопротивления R_t .
2. Разбиение уровней цены на кластеры:
 - а. Для уровней поддержки:
 - i. Если для уровня поддержки S_t существует уровень поддержки $S_{t'}$, где $t' \in [t-K, t]$, которому соответствует элемент кластера поддержки $S_{n,k} \in (S_{n,j})_{j=1}^k$, то проверяется, выполняется ли условие $|S_t - S_{n,1}| \leq \pi$:
 1. Если выполняется, то S_t добавляется в кластер CS_n .
 2. Если не выполняется, то S_t образует новый кластер CS_{n+1} .
 - б. Для уровней сопротивления:
 - i. Если для уровня сопротивления R_t существует уровень поддержки $R_{t'}$, где $t' \in [t-K, t]$, которому соответствует элемент кластера сопротивления $R_{n,k} \in (R_{n,j})_{j=1}^k$, то проверяется, выполняется ли условие $|R_t - R_{n,1}| \leq \pi$:
 1. Если выполняется, то R_t добавляется в кластер CR_n .
 2. Если не выполняется, то R_t образует новый кластер CR_{n+1} .

3. Методы исследования

Далее детально рассмотрены методы исследования, в частности процесс подготовки данных, модели, используемые в экспериментах, их обучение, а также простой торговый алгоритм и его версия с использованием фильтра на основе глубокого обучения

3.1. Подготовка данных для обучения моделей

Задача детектирования уровней поддержки и сопротивления была сведена к задаче классификации по трем классам:

- нейтральный класс – множества потенциальных уровней поддержки и сопротивления, при этом не являющихся ни уровнями поддержки, ни уровнями сопротивления;

- кластер уровней сопротивления;
- кластер уровней поддержки.

Для моделирования экспериментов были отобраны данные NBBO по четырём активам AAPL, MSFT, GE и NVDA за период с 26.01.2023 по 22.05.2023 (78 дней). Такие данные содержат лучшие цены на продажу $a(t)$ и покупку $b(t)$ активов, а также их объёмы – $v_a(t)$ и $v_b(t)$.

В обучающую выборку вошли первые 40 дней (с 26.01.2023 по 23.04.2023), для валидационной выборки было выделено 15 дней (с 24.03.2023 по 14.04.2023), для тестовой был выделен 1 месяц, или 23 торговых дня (с 17.04.2023 по 17.05.2023). Соотношение дней в выборках иллюстрировано на рис. 2.

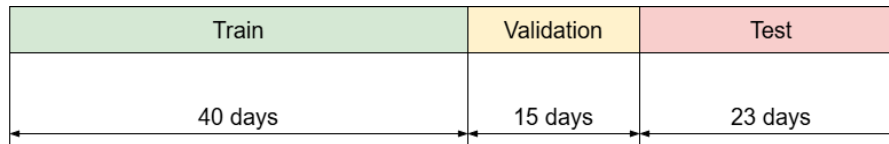


Рис. 2. Разделение исходных данных на выборки:

Train – обучающая (40 дней), Validation – валидационная (15 дней), Test – тестовая (23 дня)

Fig. 2. Dataset partitioning in:

Train – train set (40 days), Validation – validation set (15 days), Test – test set (23 days)

Для обучения тестов и валидации были отобраны исторические данные только в пределах часов осуществления торговли (с 9:30 до 16:00 GMT-5).

Параметр T был установлен как 1 600, так как из-за неоднородной плотности событий и флуктуаций цены при значениях $T < 1\,600$ во многих случаях паттерн образования локальных максимумов воспринимается человеком с трудом (по крайней мере на малых выборках). Значение параметра K было принято равным 1 600.

Обучающая, валидационная и тестовая выборки были сбалансированы при помощи приведения числа примеров в каждом из дней по каждому из классов к числу примеров наименьшего из классов в отдельно взятый день. В табл. 2 представлено количество примеров в каждой из выборок по каждому из классов.

Таблица 2

Количество примеров в каждом из классов в каждой из выборок

Выборка	Количество примеров
Обучающая	116 172
Валидационная	30 162
Тестовая	39 663

Входной тензор имел вид: $X = [x_1, x_2, \dots, x_i, \dots, x_{400}]^T \in \mathbb{R}^{400 \times 4}$, где $x_i = [a_i, v_{a,i}, b_i, v_{b,i}]$. Значение 400 соответствует количеству наблюдений, используемых для осуществления предсказаний, а 4 – количеству признаков.

К признакам к каждому из дней по отдельности была применена процедура стандартизации при помощи z-score (1), средние значения по каждому из признаков x и стандартные отклонения σ рассчитывались по пяти предшествующим дням [5, 7]:

$$z_i = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

где $\bar{x}$ – вектор средних значений; σ – вектор стандартных отклонений.

3.2. Архитектуры моделей, задействованных в эксперименте

С целью получения более полного представления о характере зависимостей уровней поддержки и сопротивления от истории цен рассматривается ряд моделей, имеющих различные функции преобразования сигнала.

В работе рассматриваются такие архитектуры глубоких нейронных сетей, как многослойный перцептрон (MLP) и гибридная искусственная нейронная сеть LSTM-MLP, которая совмещает блоки LSTM и несколько полносвязных слоев. Также рассматривается логистическая регрессия (LR) с целью проверить, позволит ли переход к нелинейным моделям осуществить прирост качества, и если позволит, то насколько существенным будет такой эффект.

Генерация дополнительных признаков может позволить улучшить производительность модели за счет извлечения каких-либо сложных зависимостей. Признаки могут быть сгенерированы с использованием знаний о предметной области, и в качестве признаков могут быть применены, например, технические и количественные индикаторы [8] или эконометрические признаки [7, 8]. Также с целью автоматического извлечения признаков могут использоваться обучаемые блоки генерации признаков. Например, известно, что для автоматического извлечения признаков подходят сверточные нейронные сети [10].

В работе в качестве блока автоматической генерации признаков используются блоки каузальных сверточных сетей [7, 11, 12]. Каузальные свертки позволяют сохранять порядок элементов последовательности.

На рис. 3 представлены архитектуры разработанных моделей с использованием CNN-блоков.

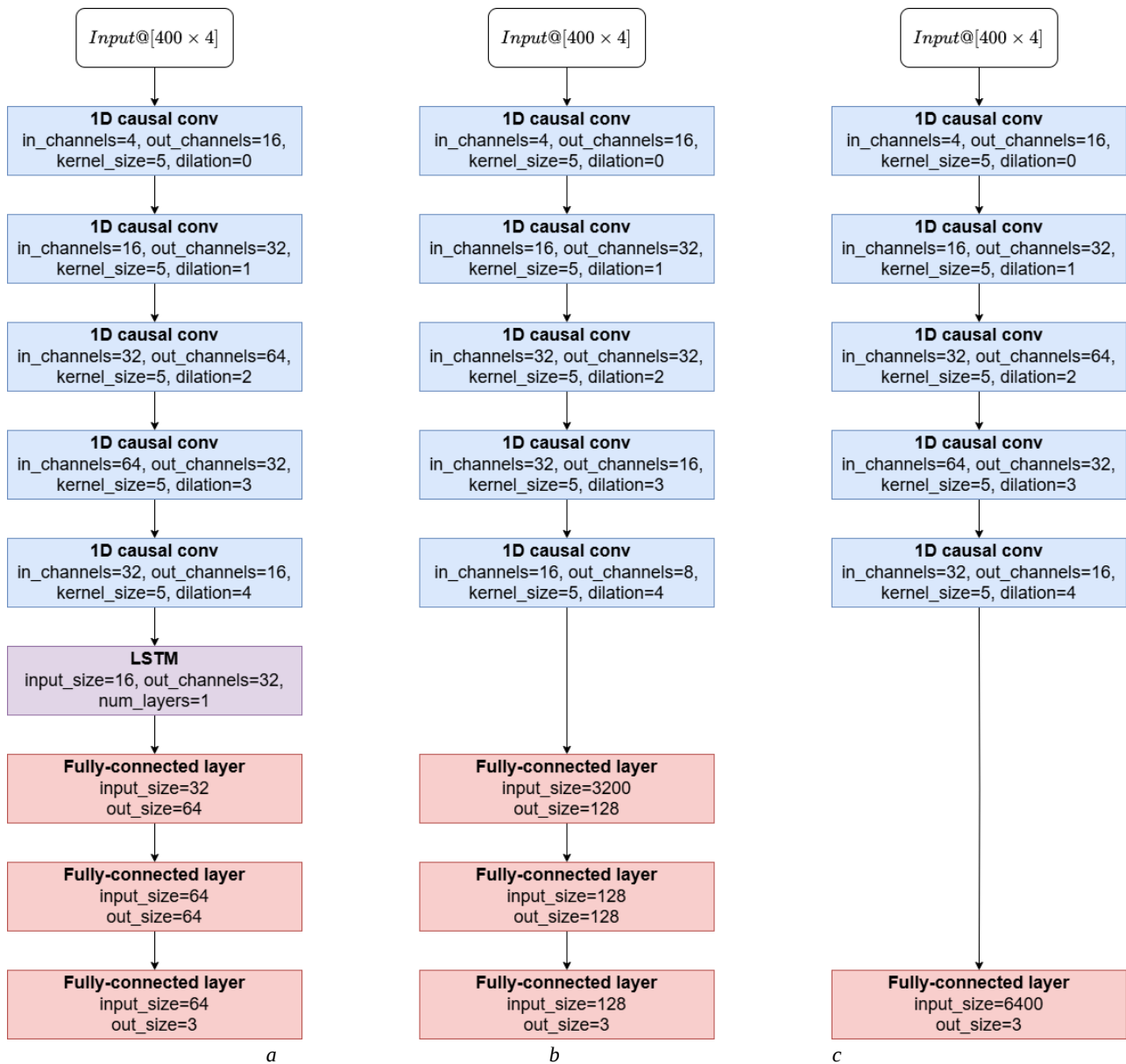


Рис 3. Архитектуры моделей: a – CNN-LSTM-MLP, b – CNN-MLP, c – CNN-LR

Fig. 3. Model's architectures: a) CNN-LSTM-MLP, b) CNN-MLP, c) CNN-LR

Таким образом, представлен ряд моделей по возрастанию их сложности, или «гибридности», с целью оценить эффект влияния каждого из блоков (CNN, MLP и LSTM) на показатели качества. В итоге производится обучение и тестирования следующих моделей:

- LR;
- MLP;
- LSTM-MLP;
- CNN-LR;
- CNN-MLP;
- CNN-LSTM-MLP.

3.3. Процесс обучения

Обучение моделей производилось при помощи стохастического градиентного спуска. В качестве функции потерь использовалась кросс-энтропия. Обучение останавливалось, когда точность (accuracy) на валидационной выборке не повышалась более чем на 0,01% в течении 20 эпох.

3.4. Торговые стратегии

С целью проверки целесообразности практического применения разработанных методик и алгоритмов в работе проводятся тестирование торговой стратегии и измерение влияния предсказаний классификатора на ее эффективность.

Алгоритм, принимающий решения о покупке / продаже активов, наблюдает за двумя событиями: появлением потенциального уровня сопротивления / поддержки и его подтверждением.

Событием подтверждения уровня поддержки мы считаем такое событие, произошедшее в момент t , когда средняя цена пересекла верхнюю границу области поддержки $m(t) > \bar{S}$.

Событием подтверждения уровня сопротивления мы считаем такое событие, произошедшее в момент t , когда средняя цена пересекла нижнюю границу области сопротивления $m(t) < \underline{R}$.

Симуляция торговли осуществляется по следующему алгоритму:

1. Определяется потенциальный уровень поддержки / сопротивления.
2. Осуществляется классификация выбранного наблюдения (к одному из 3 классов).
3. Фиксируется событие подтверждения уровня поддержки / сопротивления.
4. Если уровень поддержки / сопротивления был подтвержден, то осуществляется покупка / продажа актива соответственно.

4.1. Ожидается появление потенциального уровня поддержки и последующее его подтверждение, после чего происходит покупка актива.

4.2. После покупки актива ожидается событие появления потенциального уровня сопротивления и его подтверждения, при этом рассматриваются только уровни сопротивления, в которых цена на продажу выше цены на покупку уровня поддержки из (4.1).

В качестве расширения этой стратегии использовался дополнительный фильтр наблюдений, являющихся потенциальными уровнями поддержки / сопротивления, на основе модели глубокого обучения (Deep Learning Filter, DLF).

В случаях когда модель не классифицирует потенциальный уровень поддержки / сопротивления в соответствии с его предполагаемым классом, в дальнейшем такой уровень поддержки / сопротивления исключается из рассмотрения даже в случае появления события подтверждения для этого потенциального уровня поддержки / сопротивления.

Алгоритм симуляции торговли имеет ряд серьезных допущений:

1. Покупка и продажа осуществляются в момент, когда заявка на продажу / покупку была опубликована, по цене, указанной в заявке.

2. Действия торгового алгоритма никаким образом не влияют на цены и объемы купли / продажи.
3. Комиссия брокера составляет \$0.

3.4.1. Метрики оценки торговой стратегии

В качестве метрик качества используемой торговой стратегии используются как процент прибыли, так и коэффициент Шарпа, который позволяет оценить торговую стратегию с учетом волатильности.

Коэффициент Шарпа был рассчитан следующим образом:

1. Определяется прибыль от торговой стратегии в процентах:

$$\text{Daily Return Percentage} = \frac{\text{Daily Profit}}{\text{Starting Capital of the Day}}, \quad (2)$$

2. Пересчитывается дневная ставка инвестиционной стратегии с низким риском:

$$\text{Daily Risk - Free Rate} = (1 + \text{Anual Risk - Free Rate})^{\frac{1}{252}} - 1, \quad (3)$$

3. Рассчитывается избыточная прибыль за день:

$$\text{Excess Return} = \text{Daily Return Percentage} - \text{Daily Risk - Free Rate}, \quad (4)$$

4. Рассчитывается усредненное значение избыточной прибыли,

5. Вычисляется дневное значение коэффициента Шарпа:

$$\text{Daily Sharpe Ratio} = \frac{\text{Average Excess Return}}{\text{Standard Deviation of Return}}, \quad (5)$$

6. На основе дневного значения коэффициента Шарпа оценивается, каким может быть годовое значение метрики (предполагается, что в году 252 торговых дня):

$$\text{Anualized Sharpe Ratio} = \text{Daily Sharpe Ratio} \times \sqrt{252}. \quad (6)$$

В качестве инвестиционной стратегии с низким риском используются краткосрочные казначейские векселя США за год.

Стартовый капитал в начале каждого дня пересчитывается в соответствии с прибылями и потерями от предыдущих дней.

4. Исследование классификаторов, обученных при помощи алгоритма

В разделе представлены результаты обучения и тестов моделей, а также результаты тестирования торговой стратегии с использованием дополнительного фильтра наблюдений DLF и без его применения.

4.1. Результаты обучения моделей

В табл. 3 представлены результаты тестирования моделей. Из табл. 3 видно, что наилучшие результаты показала модель типа CNN-LSTM-MLP.

Таблица 3

Метрики, полученные на тестовой выборке

Модель	Accuracy	Precision	Recall	F1
LR	66,15	64,20	66,15	63,07
MLP	68,56	67,41	68,56	67,51
LSTM-MLP	67,59	68,07	67,59	67,76
CNN-LR	70,89	69,46	70,87	68,77
CNN-MLP	70,69	70,40	70,69	70,29
CNN-LSTM-MLP	75,98	75,62	75,98	75,73

На рис. 4 представлены матрицы ошибок для всех моделей, где можно видеть распределение предсказаний наблюдений в тестовой выборке по классам. Из рис. 4 и табл. 3 можно заключить, что модели, имевшие блок генерации признаков на основе каузальных сверток, продемонстрировали более высокие результаты в сравнении с моделями без CNN-блоков.

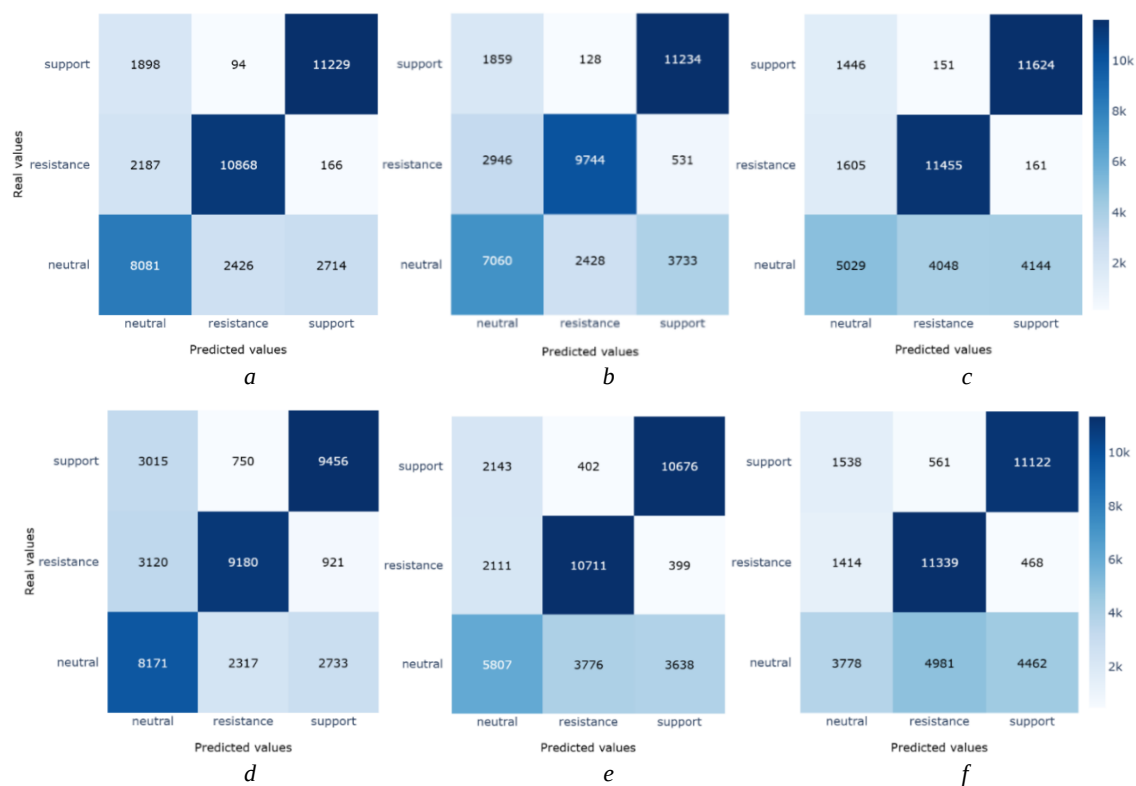


Рис 4. Матрицы ошибок, полученные на тестовой выборке:

a – CNN-LSTM-MLP, *b* – CNN-MLP, *c* – CNN-LR, *d* – LSTM-MLP, *e* – MLP, *f* – LR

Fig. 4. Confusion matrix for the training set:

a) CNN-LSTM-MLP, *b)* CNN-MLP, *c)* CNN-LR, *d)* LSTM-MLP, *e)* MLP, *f)* LR

Также из рис. 4 видно, что наибольшее количество ошибок связано с классификацией нейтрального класса. Суммарное количество ошибок первого и второго рода возрастает на матрицах ошибок как на рис. 4, *a* и 4, *b*, так и на рис. 4, *d* и 4, *f*, что может говорить о том, что глубокая архитектура, в частности несколько полносвязных слоев вместе с анализом временных зависимостей при помощи LSTM, позволяют с большей точностью классифицировать наблюдения верно. Кроме того, даже в случае CNN-LSTM-MLP (см. рис. 4, *a*) количество ложноположительных и ложноотрицательных предсказаний для нейтрального класса существенно, при этом классификаторы могут в сравнении точно отделить друг от друга классы уровней поддержки и сопротивления. Таким образом, можно предположить, что представленные определения потенциальных уровней поддержки и сопротивления не всегда обеспечивают точное разграничение между этими уровнями. Например, возможна ситуация, когда имеется два последовательных наблюдения, находящихся близко друг к другу, где одно наблюдение было определено только как потенциальный уровень поддержки или сопротивления, а последующее было определено как наблюдение, образующее кластер поддержки или сопротивления. В крайнем случае два таких наблюдения могут идти в том числе и подряд друг за другом.

4.2. Результаты тестирования торговых стратегий

Алгоритмы торговых стратегий, представленных в разд. 3.4, были применены к данным тестовой выборки. Была осуществлена торговля как с фильтром DLF, так и без него. Для фильтрации наблюдений применялась обученная модель CNN-LSTM-MLP, показавшая наиболее высокие результаты.

К активам из тестовой выборки был добавлен дополнительный актив TSLA, которого ранее не было в обучающей и валидационной выборках.

На каждый из используемых активов было выделено 6 000 долларов США стартового капитала. В качестве инвестиционной стратегии с низким риском был взят вексель казначейства США со сроком погашения в 52 недели. Предполагалось, что вексель номиналом 30 000 долларов был взят по скидке в 4,61% 17.04.2023 [13]. Таким образом, доход от векселя определяется как

$$\text{Treasury Bond Return} = \frac{\text{Face Value} - \text{Purchased Price}}{252}, \quad (7)$$

где *Face Value* – номинальная стоимость ценной бумаги; *Purchased Price* – стоимость покупки ценной бумаги.

Ежедневный доход от стратегии с низким риском рассчитывался как

$$\text{Daily Risk - Free Rate} = \frac{\text{Treasury Bond Return}}{252}. \quad (8)$$

Результаты тестирования торговой стратегии представлены в табл. 4.

Таблица 4

Результаты тестирования торговой стратегии

	Средняя дневная доходность (в долларах США)		Абсолютное значение доходности за весь период (в долларах США)		Среднесуточный коэффициент Шарпа		Годовой коэффициент Шарпа	
	Без DLF	DLF	Без DLF	DLF	Без DLF	DLF	Без DLF	DLF
Акция								
AAPL	0,46	0,48	10,57	11,44	-1,80	-1,63	-28,51	-25,82
GE	0,31	0,32	7,12	7,38	-3,70	-3,47	-58,67	-55,15
MSFT	1,20	1,28	27,58	29,40	0,10	0,20	1,53	3,23
NVDA	2,02	2,34	46,42	53,96	0,27	0,31	4,36	4,92
TSLA	1,38	1,48	31,72	34,15	0,24	0,33	3,86	5,28
Итог по всему портфелю								
	5,37	5,92	123,41	136,32	–	–	-0,74	1,22

Таким образом, использование DLF позволило увеличить общую прибыль по портфелю на 10,46%, а также получить более высокое значение коэффициента Шарпа и превратить невыгодную стратегию в выгодную с приемлемым отношением избыточной прибыли к величине риска (считается, что коэффициент Шарпа 1 и более является приемлемым для инвестиционной стратегии [14]).

Для акций AAPL и GE коэффициент Шарпа принимает отрицательные значения, что значит, что отдельно для каждого из этих двух символов торговая стратегия проигрывает по доходности инвестиционной стратегии с низким риском. Такой результат можно объяснить несовершенством выбранной торговой стратегии, так как после покупки актива ожидается появление более выгодной цены, что может случиться через значительное время или чего может не случиться вообще при общем нисходящем тренде на рынке.

Заключение

В работе предложена методика разметки для обучения классификатора с целью детектирования уровней поддержки и сопротивления. Показана возможность практического применения разработанной методики посредством использования обученных моделей в торговой симуляции.

Наиболее высоких результатов удалось достичь на основе модели типа CNN-LSTM-MLP, при помощи совмещения блоков CNN, LSTM и полносвязных слоев, т.е. глубокой искусственной нейронной сети, что может говорить о существовании нелинейной зависимости между историей цен и явлением образования уровней поддержки и сопротивления.

Можно также отметить, что логистическая регрессия продемонстрировала результат F1-метрики ~ 63%, что уже нельзя назвать случайным результатом. Из этого можно предположить, что данную методику возможно применять не только для глубокого обучения.

В качестве развития методики разметки для уровней поддержки и сопротивления можно предложить разработать понятия, которые позволят более четко разграничить уровни поддержки / сопротивления и наблюдения, которые не являются ни уровнями поддержки, ни уровнями сопротивления (нейтральный класс).

Список источников

1. Osler C. L. Support for resistance: technical analysis and intraday exchange rates // *Economic Policy Review*. 2000. V. 6 (2). P. 53–68.
2. Garzarelli F. et al. Memory effects in stock price dynamics: evidences of technical trading // *Scientific Reports*. 2014. V. 4 (1). Art. 4487. doi: 10.1038/srep04487
3. Chung K., Bellotti A. Evidence and Behaviour of Support and Resistance Levels in Financial Time Series // *arXiv preprint*. 2021. doi: 10.48550/arXiv.2101.07410. URL: <https://arxiv.org/abs/2101.07410> (accessed: 05.12.2024).
4. Sirignano J., Cont R. Universal features of price formation in financial markets: perspectives from deep learning // *Quantitative Finance*. 2019. V. 19 (9). P. 1449–1459. doi: 10.1080/14697688.2019.1622295
5. Zhang Z., Zohren S., Roberts S. Deeplob: Deep convolutional neural networks for limit order books // *IEEE Transactions on Signal Processing*. 2019. V. 67 (11). P. 3001–3012. doi: 10.1109/tsp.2019.2907260
6. Szegedy C. et al. Going deeper with convolutions // *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. P. 1–9. doi: 10.1109/cvpr.2015.7298594
7. Tsantekidis A. et al. Using deep learning for price prediction by exploiting stationary limit order book features // *Applied Soft Computing*. 2020. V. 93. Art. 106401. doi: 10.1016/j.asoc.2020.106401
8. Ntakaris A. et al. Mid-price prediction based on machine learning methods with technical and quantitative indicators // *PloS One*. 2020. V. 15 (6). e0234107. doi: 10.1371/journal.pone.0234107
9. Ntakaris A. et al. Feature engineering for mid-price prediction with deep learning // *IEEE Access*. 2019. V. 7. P. 82390–82412. doi: 10.1109/ACCESS.2019.2924353
10. LeCun Y. et al. Gradient-based learning applied to document recognition // *Proc. of the IEEE*. 1998. V. 86 (11). P. 2278–2324. doi: 10.1109/5.726791
11. Wallbridge J. Transformers for limit order books // *arXiv preprint*. 2020. doi: 10.48550/arXiv.2003.00130. URL: <https://arxiv.org/abs/2003.00130> (accessed: 05.12.2024).
12. Van Den Oord A., Dieleman S., Zen H., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K., Wavenet: A generative model for raw audio // *arXiv preprint*. 2020. doi: 10.48550/arXiv.1609.03499. URL: <https://arxiv.org/abs/1609.03499> (accessed: 05.12.2024).
13. Daily Treasury Bill Rates // U.S. Department of the Treasury. URL: https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_bill_rates&field_tdr_date_value=2023 (accessed: 05.12.2024).
14. Maverick J.B. What the Sharpe Ratio Means for Investors // *Investopedia*. URL: <https://www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp> (accessed: 05.12.2024).

References

1. Osler, C.L. (2000) Support for resistance: technical analysis and intraday exchange rates. *Economic Policy Review*. 6(2). pp. 53–68.
2. Garzarelli, F. et al. (2014) Memory effects in stock price dynamics: evidences of technical trading. *Scientific Reports*. 4(1). Art. 4487. DOI: 10.1038/srep04487
3. Chung, K. & Bellotti, A. (2021) Evidence and Behaviour of Support and Resistance Levels in Financial Time Series. *arXiv preprint arXiv:2101.07410*. DOI: 10.48550/arXiv.2101.07410
4. Sirignano, J. & Cont, R. (2019) Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*. 19(9). pp. 1449–1459. DOI: 10.1080/14697688.2019.1622295
5. Zhang, Z., Zohren, S. & Roberts, S. (2019) Deeplob: Deep convolutional neural networks for limit order books. *IEEE Transactions on Signal Processing*. 67(11). p. 3001–3012. DOI: 10.1109/tsp.2019.2907260
6. Szegedy, C. et al. (2015) Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1–9. DOI: 10.1109/cvpr.2015.7298594
7. Tsantekidis, A. et al. (2020) Using deep learning for price prediction by exploiting stationary limit order book features. *Applied Soft Computing*. 93. pp. 106401. DOI: 10.1016/j.asoc.2020.106401
8. Ntakaris, A. et al. (2020) Mid-price prediction based on machine learning methods with technical and quantitative indicators. *Plos one*. 15(6). pp. e0234107. DOI: 10.1371/journal.pone.0234107
9. Ntakaris, A. et al. (2019) Feature engineering for mid-price prediction with deep learning. *Ieee Access*. 7. pp. 82390–82412. DOI: 10.1109/ACCESS.2019.2924353
10. LeCun, Y. et al. (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 86(11). pp. 2278–2324. DOI: 10.1109/5.726791
11. Wallbridge, J. (2020) Transformers for limit order books. *arXiv preprint arXiv:2003.00130*. DOI: 10.48550/arXiv.2003.00130

12. Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. & Kavukcuoglu, K. (2016) Wavenet A generative model for raw audio. *arXiv preprint* arXiv:1609.03499. DOI: 10.48550/arXiv.1609.03499
13. U.S. Department of the Treasury. (n.d.) *Daily Treasury Bill Rates*. [Online] Available from: https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_bill_rates&field_tdr_date_value=2023 (Accessed: 5th December 2024).
14. Maverick, J.B. (2024) *What the Sharpe Ratio Means for Investors*. [Online]. Available from: <https://www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp> (Accessed: 5th December 2024).

Информация об авторах:

Хайров Марк Альбертович – аспирант отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского томского политехнического университета (Томск, Россия). E-mail: mah9@tpu.ru

Спицын Владимир Григорьевич – доктор технических наук, профессор отделения информационных технологий Инженерной школы информационных технологий и робототехники Национального исследовательского томского политехнического университета; профессор Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: spvg@tpu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Khairov Mark A. (Post-graduate Student, Information Technology Department, School of Information Technology and Robotics Engineering, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: mah9@tpu.ru

Spitsyn Vladimir G. (Doctor of Technical Sciences, Professor, Information Technology Department, School of Information Technology and Robotics Engineering, National Research Tomsk Polytechnic University; National Research Tomsk State University, Tomsk, Russian Federation, Tomsk, Russian Federation). E-mail: spvg@tpu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 29.10.2024; принята к публикации 03.03.2025

Received 29.10.2024; accepted for publication 03.03.2025

Научная статья
УДК 53.088, 521.3
doi: 10.17223/19988605/70/10

Исследование точности определения расхождения шкал времени на основе нелинейных регрессионных моделей

Оксана Сергеевна Черникова¹, Татьяна Александровна Марарескул²

¹ Новосибирский государственный технический университет, Новосибирск, Россия,

² АО «Информационные спутниковые системы» имени академика М.Ф. Решетнёва, Железногорск, Россия,

¹ chernikova@corp.nstu.ru

² mta@iss-reshetnev.ru

Аннотация. Рассмотрена задача описания расхождения бортовых часов, установленных на спутниках ГЛОНАСС, относительно системной шкалы времени линейной комбинацией полиномов Чебышева, нелинейными регрессионными моделями, включающими трендовую (полиномиальную, кусочно-линейную функции, функцию логистического роста), периодическую и индикаторную составляющие. Показано, что поведение расхождения шкал времени с наименьшим средним значением среднеквадратической ошибки, для большинства спутников не превышающего 0,2 нс, описывается нелинейной регрессионной моделью с трендовой, периодической и индикаторной составляющими в комбинации с авторегрессионной моделью.

Ключевые слова: расхождение шкал времени; бортовая шкала времени; синхронизация; частотно-временные поправки; регрессионная модель.

Для цитирования: Черникова О.С., Марарескул Т.А. Исследование точности определения расхождения шкал времени на основе нелинейных регрессионных моделей // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 103–112. doi: 10.17223/19988605/70/10

Original article
doi: 10.17223/19988605/70/10

Investigation of accuracy of determination of satellite clock bias based on nonlinear regression models

Oksana S. Chernikova¹, Tatiana A. Marareskul²

¹ Novosibirsk State Technical University, Novosibirsk, Russian Federation

² JSC Academician M.F. Reshetnev «Information Satellite System», Zheleznogorsk, Russian Federation

¹ chernikova@corp.nstu.ru

² mta@iss-reshetnev.ru

Abstract. The problem of determining the discrepancy of the onboard clock installed on GLONASS satellites relative to the system time scale based on regression models is being considered. It has been shown that the divergence behavior of time scales is best described by an additive nonlinear regression model in combination with an autoregressive model, with the mean square error for most satellites not exceeding 0.2 ns.

Keywords: satellite clock bias; onboard time scale; synchronization; frequency-time corrections; regression model.

For citation: Chernikova, O.S., Marareskul, T.A. (2025) Investigation of accuracy of determination of satellite clock bias based on nonlinear regression models. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 103–112. doi: 10.17223/19988605/70/10

Введение

Задача синхронизации разнесенных в пространстве стандартов частоты является одной из актуальных задач для развития глобальных навигационных спутниковых систем (ГНСС). В большинстве случаев достаточно обеспечить высокоточную математическую привязку временных шкал космических аппаратов к выбранной опорной шкале времени без физического приведения их в состояние синхронных. Для этого необходимо определять величину расхождения шкал времени (РШВ) космических аппаратов относительно опорной шкалы времени. Надежное определение поведения часов в режиме реального времени необходимо для обеспечения точных навигационных характеристик и оптимизации интервала между загрузкой поправок к значениям опорной шкалы.

Поправки рассчитываются в виде прогнозов уходов бортовых шкал времени (БШВ) относительно эталонной шкалы центрального синхронизатора системы. Причинами ухода бортовых шкал являются собственная нестабильность квантового перехода стандарта частоты и регулярные воздействия в виде гравитационного влияния на частоту часов совокупности факторов, таких как смещение шкалы собственного времени часов спутника вследствие воздействия неоднородности гравитационного поля Земли, смещение шкалы собственного времени наземных / бортовых часов, вызванное приливными потенциалами Луны и Солнца, смещение времени и частоты задающего генератора наземных часов, вызванное неравномерностью вращения Земли [1, 2].

На современном этапе развития ГНСС ГЛОНАСС одним из актуальных вопросов является выбор математической модели, описывающей поведение РШВ и обеспечивающей наилучшие точностные характеристики (значение среднеквадратической ошибки не более 0,3–0,5 нс с доверительной вероятностью 0,95).

Для описания поведения РШВ применяются различные подходы: экстраполяция степенным полиномом [3–6], построение моделей авторегрессии и интегрированного скользящего среднего (AR, ARIMA) [7–10], моделей «серого ящика» (GMs) [11–14], моделей спектрального анализа (SA models) [15, 16]. Указанные модели обладают недостатками. Так, увеличение длины мерного интервала приводит к снижению точности экстраполяции степенными полиномами; модели спектрального анализа, авторегрессии и интегрированного скользящего среднего, «серого ящика» более точно определяют поведение РШВ, но зависят от оптимизации ряда параметров, а также наличия выбросов и пропусков в измерительных данных.

В работе авторов [6] приводятся результаты построения линейной регрессионной модели и коррекции ее постоянного члена на основе измерений, полученных на последнем сеансе. Установлено, что средняя погрешность описания РШВ на мерном интервале в зависимости от номера спутника находится в интервале от 0,24 до 0,69 нс, и требуется пересчет постоянного члена линейной модели для получения более высокой точности прогноза.

Для учета колебательных и периодических составляющих в измерительных данных рассмотрим нелинейные модели применительно к описанию РШВ на интервале наблюдения (мерных интервалах). Построение нелинейных модельных структур, с одной стороны, предполагает усложнение используемого математического аппарата и увеличение временных затрат на обработку данных, а с другой – позволит учесть процессы, вызванные геофизическими и (или) космическими факторами, а также наличие скачков в измерительных данных и идентифицировать наилучшую модельную структуру, описывающую поведение РШВ.

В работе исследуется точность описания поведения РШВ космических аппаратов ГЛОНАСС относительно системной шкалы времени на основе нелинейных моделей. Предлагается двухэтапный подход, включающий построение нелинейной модели и учет возможной автокорреляции остатков.

1. Постановка задачи и методы исследования

Апостериорные данные о поведении БШВ космических аппаратов ГЛОНАСС относительно системной шкалы предоставляются системой высокоточного определения эфемерид и временных

поправок (СВОЭВП) [17], размещаются в открытом доступе и могут применяться исследователями для решения широкого круга задач, связанных с анализом поведения и прогнозированием частотно-временных поправок.

Проведенный предварительный анализ апостериорных данных о поведении БШВ на годовом интервале наблюдения с 00:00 01.01.2021 до 23:59 31.12. 2021 показал, что для всех аппаратов измерительные данные содержат пропуски [6] и на большей части 6-часовых мерных интервалов, на которые разбивается годовой интервал наблюдения, исследуемые временные ряды являются нестационарными (для проверки стационарности применялись расширенный тест Дикки–Фуллера (ADF-тест) и тест Квятковского–Филлипса–Шмидта–Шина (KPSS-тест)) [18]. Декомпозиция ряда на эмпирические компоненты методом сингулярного спектрального анализа (Singular Spectrum Analysis, SSA) позволила выявить, что временной ряд значений РШВ в основном характеризуется линейным трендом, но присутствуют также и гармонические процессы. Так, например, первой значимой эмпирической компонентой разложения ряда РШВ для аппарата R01 является линейная функция с вероятностью 0,99, остальные компоненты представляют собой гармонические колебания (см. подробнее: [18]).

Линейный тренд, описывающий временной ряд значений РШВ y_k , характеризуется математическим ожиданием закона распределения первых разностей $\Delta y_k = y_{k+1} - y_k$. Величина $10^{10}\Delta y$ описывается обобщенным нормальным законом с функцией плотности

$$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \exp \left\{ - \left(\frac{|x - \theta_0|}{\theta_1} \right)^{\theta_2} \right\},$$

где параметр сдвига $\theta_0 = 0,168657644$, параметр масштаба $\theta_1 = 0,916965722$, параметр формы $\theta_2 = 1,218194917$, полная гамма-функция $\Gamma(1/\theta_2) = \int_0^{\infty} y^{\frac{1}{\theta_2}-1} e^{-y} dy$ (рис. 1, 2).

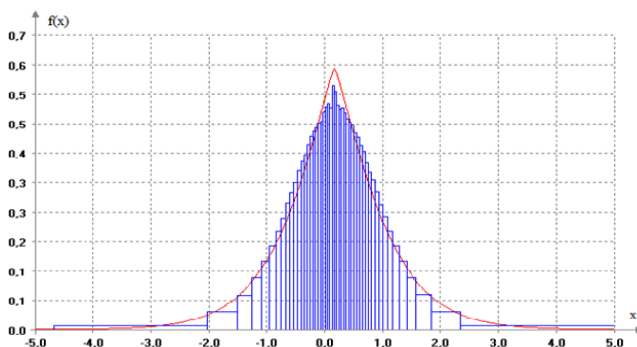


Рис. 1. Гистограмма первых разностей $10^{10}\Delta y$ и функция плотности обобщенного нормального закона

Fig. 1. Histogram of the first differences $10^{10}\Delta y$ and the generalized density function

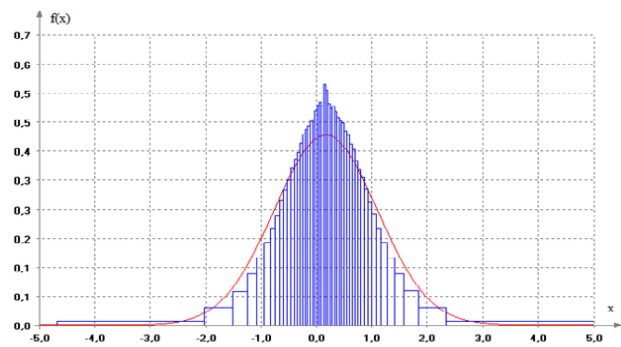


Рис. 2. Гистограмма первых разностей $10^{10}\Delta y$ и функция плотности нормального закона

Fig. 2. Histogram of the first differences $10^{10}\Delta y$ and the normal density function

Нельзя утверждать, что найденный обобщенный нормальный закон наилучшим образом описывает ряд Δy_k , так как в рядах первых разностей обнаруживаются неслучайные последовательности, искажающие реальный (неизвестный) закон распределения Δy_k . Кроме линейной составляющей, обусловленной отклонением частоты генератора от номинального значения, присутствуют колебательные и периодические компоненты, которые могут быть связаны с влиянием на бортовой стандарт частоты и времени как внутренних процессов, протекающих в космическом аппарате, так и внешних, таких как геомагнитная активность, изменения солнечной активности и др.

Для описания РШВ рассмотрим следующие модельные структуры.

Модель 1. Опишем линейной комбинацией полиномов Чебышева РШВ на выбранном мерном интервале:

$$y^M(t) = \alpha_1 T_1(t) + \alpha_2 T_2(t) + \dots + \alpha_m T_m(t) + \varepsilon(t), \quad t \in I_{\text{опред}}, \quad (1)$$

где $T_0(t)=1$, $T_1(t)=t$, $T_2(t)=2t^2-1$, ..., $T_m(t)=2tT_{m-1}(t)-T_{m-2}(t)$, $\varepsilon(t)$ – белый гауссовский шум с нулевым математическим ожиданием и конечной дисперсией.

Коэффициенты α_i в соотношении (1) найдем с помощью взвешенного метода наименьших квадратов (МНК) с весовой матрицей, позволяющей учесть последние серии измерений с наибольшим весом:

$$\Omega = \begin{pmatrix} \Omega_1 & O & O \\ O & \Omega_2 & O \\ O & O & \Omega_3 \end{pmatrix}^k, \quad \Omega_1 = \text{diag}\left(\frac{1}{3}, \frac{1}{3}, \dots, \frac{1}{3}\right), \quad \Omega_2 = \text{diag}\left(\frac{2}{3}, \frac{2}{3}, \dots, \frac{2}{3}\right), \quad \Omega_3 = \text{diag}(1, 1, \dots, 1). \quad (2)$$

Количество полиномов m в (1) может оцениваться одновременно с коэффициентами линейной комбинации для достижения требуемой точности.

Модель 2. Рассмотрим нелинейную регрессионную модель вида:

$$y^M(t) = a_0 + a_1 t + a_2 t^2 + \sum_{k=1}^n b_k \sin\left(\frac{2\pi t}{p_k} + \varphi_k\right) + \varepsilon(t), \quad (3)$$

где коэффициенты a_k, b_k, p_k, φ_k находятся на основе метода взвешенного МНК с весовой матрицей (2), $\varepsilon(t)$ – белый гауссовский шум с нулевым математическим ожиданием и конечной дисперсией.

Модель 3. Опишем временной ряд РШВ нелинейной регрессионной моделью, представляющей собой сумму трех основных составляющих [19]:

$$y^M(t) = g(t) + s(t) + h(t) + \varepsilon(t), \quad (4)$$

где $g(t)$ – кусочно-линейная функция или логистическая функция роста, моделирующие непериодические изменения значений РШВ (аппроксимация трендовой составляющей ряда значений РШВ):

$$g(t) = (k + a^T(t)\delta)t + \bar{m} + a^T(t)\gamma \quad \text{или} \quad g(t) = \frac{C(t)}{1 + \exp(-(k + a^T(t)\delta)(t - (\bar{m} + a^T(t)\gamma)))},$$

где k – скорость роста, $\bar{m}$ – параметр смещения, δ – вектор, корректирующий скорость роста,

$a_j(t) = \begin{cases} 1, & t > s_j, \\ 0, & \text{в противном случае} \end{cases}$ – функция, отражающая изменения трендовой составляющей (s_j – множество точек излома тренда).

Отметим, что включение в функцию $g(t)$ точек изменения трендовой составляющей позволяет адаптировать модель к скачкам в измерительных данных РШВ. Так, скорость роста в любой момент t равна сумме скорости k и ее корректировок до этой точки $k + \sum_{j:t>s_j} \delta_j$, где δ_j – изменение скорости в момент s_j .

При изменении скорости k для обеспечения гладкости кривой тренда $g(t)$ вводится функция γ :

$$\gamma_j = \left(s_j - \bar{m} - \sum_{l < j} \gamma_l \right) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l} \right).$$

В (4) $s(t)$ – функция, позволяющая учесть периодические колебания, описываемые рядом Фурье:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right),$$

$h(t)$ – функция-индикатор, учитывающая аномальные значения в измерительных данных:

$$h(t) = Z(t)q, q \in \text{Norm}(0, v^2),$$

где $Z(t) = \begin{cases} 1, & t \in L, \\ 0, & t \notin L, \end{cases}$ L – набор точек, в которых ожидаются аномальные измерения (выбросы или пропущенные значения); $\varepsilon(t)$ – белый гауссовский шум с нулевым математическим ожиданием и конечной дисперсией.

Модель (4), построение которой относят к методам машинного обучения, основывается на одномомерном сглаживании измерительных данных с последующим построением непараметрической регрессии. Данный подход был впервые предложен в [20] и зарекомендовал себя как более гибкий и удобный в сравнении с традиционной линейной регрессией.

Перспективным направлением по повышению точности модельных структур является комбинированный подход [18, 21], сочетающий модели разных типов. В работе [18] авторами показана целесообразность двухэтапного подхода, предполагающего после определения трендовой составляющей построение альтернативной модели остаточной составляющей. В предлагаемом двухэтапном подходе к построению модели РШВ (рис. 3) на первом этапе процедуры, в отличие от [18], по результатам измерений РШВ на выбранном мерном интервале строится нелинейная модель тренда (выбирается одна из приведенных модельных структур 1–3).

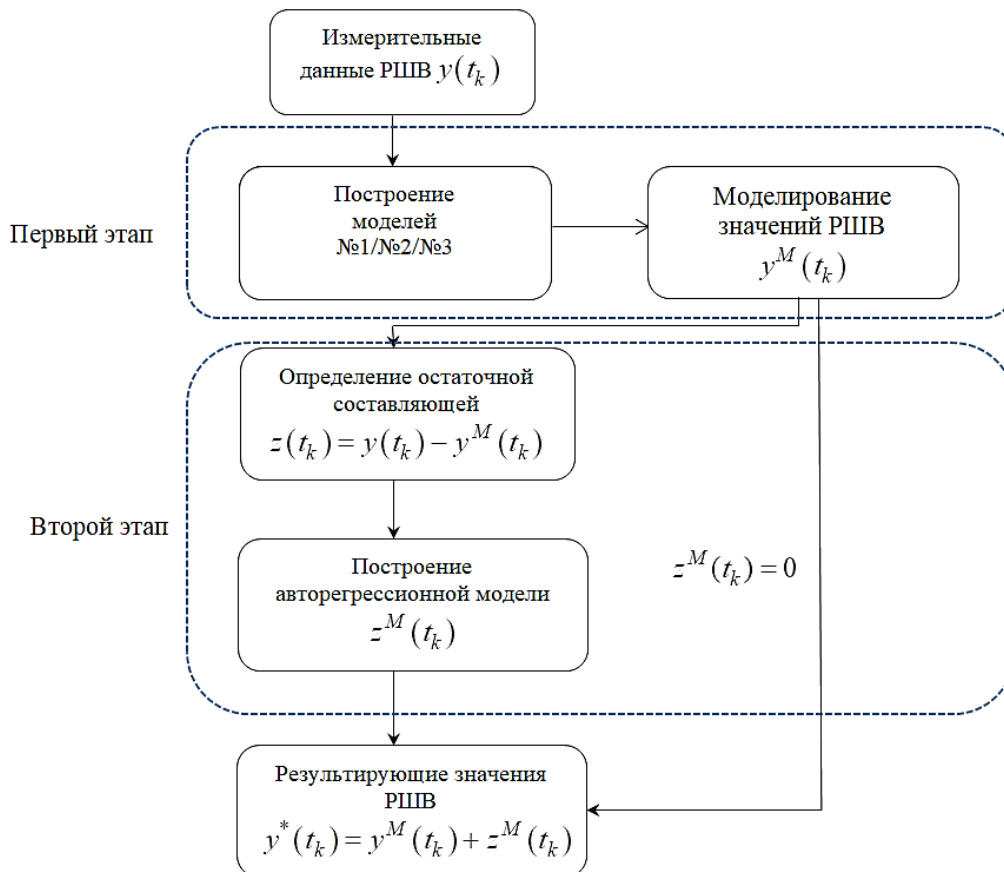


Рис. 3. Схема двухэтапного подхода построения модели РШВ

Fig. 3. Scheme of the two-stage approach to constructing the model of the satellite clock bias

На втором этапе определяется неучтенная остаточная составляющая ряда значений РШВ:

$$z(t_k) = y(t_k) - y^M(t_k), k = 0, 1, \dots, N,$$

где $y(t_k)$ – величины расхождения шкалы времени спутника относительно системной шкалы времени, $y^M(t_k)$ – значения, вычисленные с помощью выбранной модели, и строится описывающая ее модель, отличная от модели тренда. В качестве такой модели используется авторегрессионная модель с параметрами $\alpha_1, \dots, \alpha_p$ вида:

$$z^M(t_k) = \sum_{j=1}^p \alpha_j z^M(t_{k-j}) + \varepsilon(t_k), \quad (5)$$

где p – порядок модели, определяемый на основе информационного критерия Акаике, исходя из выбранной точности, $\varepsilon(t_k)$ – последовательность независимых одинаково распределенных случайных

величин с нулевым математическим ожиданием и конечной дисперсией. Для определения параметров модели может применяться как классический метод наименьших квадратов, так и метод взвешенных наименьших квадратов, учитывающий в том числе и наличие аддитивных ошибок измерений [22].

Следует отметить, что точность идентификации модельных структур повышается при проведении предварительной обработки исходных апостериорных данных, а именно восстановлении пропущенных данных, возникших, например, в результате аппаратного сбоя или передачи данных. Преимуществом использования на первом этапе модели вида (4) является отсутствие необходимости предварительной интерполяции пропущенных значений РШВ.

3. Численные результаты и обсуждение

Для численного исследования эффективности двухэтапного подхода построения моделей РШВ использовались апостериорные данные СВОЭВП для аппаратов R01–R24 на годовом интервале с 00:00 01.01.2021 до 23:59 31.12. 2021 (исключение составили R06, R09, R10, R23).

Для построения моделей годовой интервал наблюдения разбивался на k некоррелированных 6-часовых мерных интервалов, на каждом из которых для определения трендовой составляющей строились модели 1–3, а для остаточной составляющей ряда значений РШВ – авторегрессионная модель (5).

В качестве характеристик точности построения модели на интервале наблюдения использованы минимальное, среднее и максимальное среднеквадратические отклонения (СКО) по уровню доверительной вероятности 0,95, определяемые по всем k -мерным интервалам.

В таблице приведены минимальное, среднее и максимальное СКО РШВ, полученные для описанных модельных структур по годовым апостериорным данным. На рис. 4 представлена соответствующая результатам таблицы графическая интерпретация.

Минимальное, среднее и максимальное среднеквадратическое отклонения РШВ, полученные для различных модельных структур по годовым апостериорным данным РШВ, нс¹

Спутник	Модель №1			Модель №2			Модель №3		
	Минимальное СКО	Среднее СКО	Максимальное СКО	Минимальное СКО	Среднее СКО	Максимальное СКО	Минимальное СКО	Среднее СКО	Максимальное СКО
R01	2,816E-08	0,661	1,078	2,787E-08	0,346	0,571	0,006	0,142	0,248
R02	2,811E-07	1,315	1,648	2,784E-07	0,687	1,089	0,086	0,278	0,425
R03	2,827E-08	0,609	0,815	2,786E-08	0,312	0,512	0,001	0,126	0,194
R04	1,214E-10	0,593	1,116	2,848E-08	0,264	0,475	0,001	0,114	0,177
R05	2,885E-08	0,621	1,187	2,778E-08	0,311	0,546	0,001	0,132	0,204
R07	2,812E-08	0,684	1,143	2,742E-08	0,363	0,622	0,002	0,150	0,243
R08	2,876E-08	0,841	1,129	2,879E-08	0,446	0,752	0,009	0,182	0,275
R11	6,482E-11	0,687	1,122	2,879E-09	0,334	0,756	3,034E-05	0,124	0,368
R12	2,788E-07	0,574	0,832	2,761E-07	0,279	0,484	0,033	0,141	0,211
R13	2,781E-08	1,476	2,441	2,773E-08	0,803	1,366	0,008	0,298	0,475
R14	2,882E-08	0,548	1,187	2,781E-08	0,277	0,476	0,002	0,121	0,181
R15	2,877E-07	0,546	1,133	2,785E-07	0,267	0,461	4,115E-04	0,121	0,194
R16	8,697E-09	0,812	1,775	8,657E-09	0,474	0,748	2,161E-04	0,163	0,319
R17	2,867E-07	0,533	0,954	2,855E-07	0,263	0,462	0,091	0,164	0,254
R18	2,689E-08	0,654	1,171	2,687E-08	0,336	0,565	0,004	0,137	0,200
R19	2,873E-07	0,709	1,801	2,682E-07	0,462	0,741	0,006	0,189	0,295
R20	2,686E-08	0,733	1,454	3,049E-08	0,376	0,688	0,005	0,166	0,288
R21	2,702E-07	0,520	1,155	2,684E-07	0,271	0,465	0,023	0,136	0,211
R22	2,885E-08	1,139	2,302	2,890E-08	0,478	0,917	0,006	0,211	0,542
R24	2,877E-08	0,737	1,127	2,568E-08	0,316	0,664	0,001	0,146	0,229

¹ Серым цветом выделены наименьшие минимальные, средние и максимальные СКО, найденные по приведенным модельным структурам.

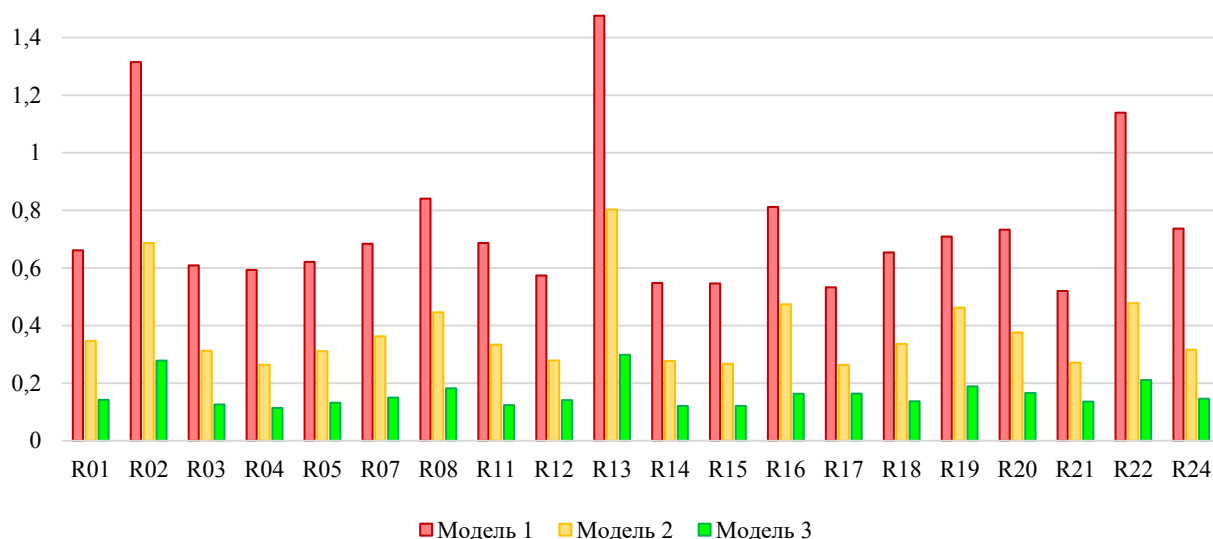


Рис.4. Среднее значение СКО для различных моделей, нс
Fig.4. Average RMSE for different models, ns

Проведенные исследования позволяют сделать следующие выводы:

- Для всех спутников наибольшие среднее и максимальное СКО были получены при использовании на первом этапе модели 1; минимальное СКО также оказалось наибольшим, за исключением спутников R04, R08, R11, R20.
- Наименьшее минимальное СКО для большинства спутников, за исключением R04, R08, R11, R20, было получено при использовании модели 2; среднее СКО в этом случае у 14 спутников превысило 0,3 нс и варьировало от 0,263 до 0,803 нс, максимальное СКО у всех аппаратов было выше 0,3 нс и варьировало от 0,461 до 1,366 нс.
- Для всех спутников наименьшие среднее и максимальное СКО были получены при использовании на первом этапе модели 3, при этом для всех спутников среднее СКО оказалось меньше 0,3 нс и варьировало от 0,114 до 0,298 нс; максимальное СКО у 5 спутников превысило 0,3 нс и варьировало от 0,177 до 0,542 нс.
- В группу спутников, для которых были получены наибольшие средние и максимальные значения СКО, можно включить R02, R13, R16, R22.

Заключение

Для описания поведения РШВ спутников ГЛОНАСС рассмотрены линейная комбинация полиномов Чебышева и нелинейные регрессионные модели, включающие полиномиальную, кусочно-линейную функцию и функцию логистического роста, периодическую и индикаторную составляющие.

Показано, что поведение РШВ с наименьшим СКО описывается нелинейной моделью, включающей трендовую, периодическую и индикаторную составляющие (модель 3). Применение двухэтапного подхода, сочетающего построение данной модели с авторегрессией, позволяет для всех спутников описать поведения РШВ на интервале наблюдения со средним и максимальным СКО меньше 0,3 нс.

Отметим, что предложенный двухэтапный подход к идентификации поведения РШВ является устойчивым к отсутствию измерительных данных и достаточно хорошо учитывает выбросы в измерительных данных, однако в качестве недостатка следует отметить существенные временные затраты, необходимые для построения модельных структур при использовании годовых измерительных данных для каждого спутника, которые могут быть сокращены при использовании параллельных вычислительных технологий.

Список источников

1. Фатеев В.Ф. К вопросу о релятивистских эффектах смещения шкал времени и частоты наземных и спутниковых квантовых часов в глобальных навигационных спутниковых системах // Альманах современной метрологии. 2023. № 2 (34). С. 43–64.
2. Сальцберг А.В., Тимошенко Е.В., Шупен К.Г. Влияние космической среды на ход бортовых часов // Труды Института прикладной астрономии РАН. 2020. Вып. 52. С. 51–56.
3. Bidikar B., Rao G.S., Laveti G., Kumar M.S. Satellite Clock Error and Orbital Solution Error Estimation for Precise Navigation Applications // Positioning. 2014. Is. 05 (01). P. 22–26. doi: 10.4236/pos.2014.51003
4. Liu X., Wu X., Tian Y. Study on atomic prediction of time based on interpolation model with tchebychev polynomials // Journal of Geodesy and Geodynamics. 2010. Is. 30 (1). P. 77–82.
5. Wang Y., Lu Z., Sun D., Wang N. A new navigation satellite clock bias prediction method based on modified clock-bias quadratic polynomial model // Acta Astronomica Sinica. 2016. Is. 57 (1). P. 78–90.
6. Черникова О.С., Марарескул Т.А. Прогнозирование расхождения шкал времени на основе скорректированной линейной модели // Системы анализа и обработки данных. 2022. Т. 87, № 3. С. 37–58. doi: 10.17212/2782-2001-2022-3-37-58
7. Xu X., Hu X., Zhou Y., Song Y. Research on high accuracy prediction model of satellite clock bias // China Satellite Navigation Conference (CSNC): Lecture Notes in Electrical Engineering. 2014. Is. 3 (1). P. 155–164. doi: 10.1007/978-3-642-54740-9_14
8. Liu Q., Sun J., Chen X., Liu J., Zhang Q. Application analysis of CPSO-LSSVM algorithm in AR clock error prediction // Journal of Jilin University (Engineering and Technology Edition). 2014. Is. 44 (3). P. 807–811. doi: 10.13229/j.cnki.jdxbgxb201403036
9. Zhao L., Lan X.Q., Sheng J.Y. Application of ARIMA model in satellite clock error forecasting // J. Water Resour Archit. Eng. 2014. V. 10 (1). P. 135–137.
10. Han S., Gong Y., Li J., Li G., Li X., Guo J. The AR model with trend item and the corresponding algorithm of AO detection and satellite clock error prediction // Mapping Intimacies. 2021. P. 584–598. doi: 10.1007/978-981-16-3138-2_54
11. Cui X.Q., Jiao W.H. Grey system model for the satellite clock error predicting // Geomatics and Information Science of Wuhan University. 2005. Is. 30 (5). P. 447–450.
12. Liang Y., Ren C., Yang X., Pang G., Lan L. A grey model based on first differences in the application of satellite clock bias prediction // Chinese Astronomy and Astrophysics. 2015. Is. 40 (1). P. 79–93. doi: 10.1016/j.chinastron.2016.01.008
13. Yu Y., Huang M., Duan T., Wang C., Hu R. Enhancing satellite clock bias prediction accuracy in the case of jumps with an improved grey model // Hindawi Mathematical Problems in Engineering. 2020. Is. 3. P. 1–11. doi: 10.1155/2020/8186568
14. Yu Y., Huang M., Wang C., Hu R., Duan T. A new BDS-2 satellite clock bias prediction algorithm with an improved exponential smoothing method // Applied Sciences. 2020. V. 10 (21). Art. 7456. doi: 10.3390/app10217456
15. Zhang B., Qu J.K., Yuan Y.B. et al. Fitting method for GPS satellites clock errors using wavelet and spectrum analysis // Geomatics and Information Science of Wuhan University. 2007. Is. 32 (8). P. 715–718.
16. Zheng Z.Y., Dang Y.M., Lu X.S., Xu W.M. Prediction model with periodic item and its application to the prediction of GPS satellite clock bias // Acta Astronomica Sinica. 2010. Is. 51 (1). P. 95–102.
17. Система высокоточного определения эфемерид и временных поправок (СВОЭВП) // Филиал «Прецизионного навигационно-баллистического обеспечения» АО «НПК «СПП». URL: <http://www.glonass-svoevp.ru/> (дата обращения: 20.12.2023).
18. Черникова О.С., Марарескул Т.А. Двухэтапный подход к прогнозированию расхождения шкал времени на основе скорректированной линейной модели // Журнал Белорусского государственного университета. Математика. Информатика. 2023. № 2. С. 80–93. doi: 10.33581/2520-6508-2023-2-80-93
19. Harvey A., Peters S. Estimation procedures for structural time series models // Journal of Forecasting. 1990. Is. 9. P. 89–108.
20. Friedman J.H., Stuetzle W. Projection Pursuit Regression // Journal of the American Statistical Association. 1981. Is. 76. P. 817–823.
21. Chernikova O.S. Two-stage parametric identification procedure for a satellite motion model based on adaptive unscented Kalman filters // Bulletin of the South Ural State University, Ser. Mathematical Modelling, Programming and Computer Software. 2022. V. 15 (4). P. 32–43. doi: 10.14529/mmp220403
22. Жданов А.И., Кацюба О.А. Идентификация по методу наименьших квадратов параметров уравнений авторегрессии при аддитивных ошибках измерений // Автоматика и телемеханика. 1982. Вып. 2. С. 29–38.

References

1. Fateev, V.F. (2023) K voprosu o relyativistskikh effektakh smeshcheniya shkal vremeni i chastoty nazemnykh i sputnikovyykh kvantovyykh chasov v global'nykh navigatsionnykh sputnikovyykh sistemakh [On the relativistic effects of time scale and frequency shift of terrestrial and satellite quantum clocks in global navigation satellite systems]. *Al'manakh sovremennoy metrologii*. 34(2). pp. 43–64.
2. Saltsberg, A.V., Timoshenkova, E.V. & Shupen K.G. (2020) Vliyanie kosmicheskoy sredy na khod bortovyykh chasov [The influence of the space environment on the course of the onboard clock]. *Trudy Instituta prikladnoy astronomii RAN*. 52. pp. 51–56.

3. Bidikar, B., Rao, G.S., Laveti, G. & Kumar, M.S. (2014) Satellite Clock Error and Orbital Solution Error Estimation for Precise Navigation Applications. *Positioning*. 05(01). pp. 22–26. DOI: 10.4236/pos.2014.51003
4. Liu, X., Wu, X. & Tian, Y. (2010) Study on atomic prediction of time based on interpolation model with Tchebychev polynomials. *Journal of Geodesy and Geodynamics*. 30(1). pp. 77–82.
5. Wang, Y., Lu, Z., Sun, D. & Wang, N. (2016) A new navigation satellite clock bias prediction method based on modified clock-bias quadratic polynomial model. *Acta Astronomica Sinica*. 57(1). pp. 78–90.
6. Chernikova, O.S. & Marareskul, T.A. (2022) Prognozirovaniye raskhozhdeniya shkal vremeni na osnove skorrektirovannoy lineynoy modeli [Prediction of time scale divergence based on an adjusted linear model]. *Sistemy analiza i obrabotki dannykh*. 87(3). pp. 37–58. DOI: 10.17212/2782-2001-2022-3-37-58.
7. Xu, X., Hu, X., Zhou, Y. & Song, Y. (2014) Research on high accuracy prediction model of satellite clock bias. *China Satellite Navigation Conference (CSNC): Lecture Notes in Electrical Engineering*. 3(1). pp. 155–164. DOI: 10.1007/978-3-642-54740-9_14
8. Liu, Q., Sun, J., Chen, X., Liu, J. & Zhang, Q. (2014) Application analysis of CPSO-LSSVM algorithm in AR clock error prediction. *Journal of Jilin University (Engineering and Technology Edition)*. 44(3). pp. 807–811. DOI: 10.13229/j.cnki.jdxbgxb201403036
9. Zhao, L., Lan, X.Q. & Sheng, J.Y. (2014) Application of ARIMA model in satellite clock error forecasting. *Journal of Water Resources and Architectural Engineering*. 10(1). pp. 135–137.
10. Han, S., Gong, Y., Li, J., Li, G., Li, X. & Guo, J. (2021) The AR model with trend item and the corresponding algorithm of AO detection and satellite clock error prediction. *Mapping Intimacies*. pp. 584–598. DOI: 10.1007/978-981-16-3138-2_54.
11. Cui, X.Q. & Jiao, W.H. (2005) Grey system model for the satellite clock error predicting. *Geomatics and Information Science of Wuhan University*. 30(5). pp. 447–450.
12. Liang, Y., Ren, C., Yang, X., Pang, G. & Lan, L. (2015) A grey model based on first differences in the application of satellite clock bias prediction. *Chinese Astronomy and Astrophysics*. 40(1). pp. 79–93. DOI: 10.1016/j.chinastron.2016.01.008.
13. Yu, Y., Huang, M., Duan, T., Wang, C. & Hu, R. (2020) Enhancing satellite clock bias prediction accuracy in the case of jumps with an improved grey model. *Hindawi Mathematical Problems in Engineering*. 3. pp. 1–11. DOI: 10.1155/2020/8186568.
14. Yu, Y., Huang, M., Wang, C., Hu, R. & Duan, T. (2020) A new BDS-2 satellite clock bias prediction algorithm with an improved exponential smoothing method. *Applied Sciences*. 10(21). pp. 7456. DOI: 10.3390/app10217456.
15. Zhang, B., Qu, J.K., Yuan, Y.B. et al. (2007) Fitting method for GPS satellites clock errors using wavelet and spectrum analysis. *Geomatics and Information Science of Wuhan University*. 32(8). pp. 715–718.
16. Zheng, Z.Y., Dang, Y.M., Lu, X.S. & Xu, W.M. (2010) Prediction model with periodic item and its application to the prediction of GPS satellite clock bias. *Acta Astronomica Sinica*. 51(1). pp. 95–102.
17. AO NPK SPP. (n.d.) *Sistema vysokotochnogo opredeleniya efemerid i vremennykh popravok (SVOEVP)* [Point of Metrological Control SVOEVP (PMC SVOEVP)]. [Online] Available from: <http://www.glonass-svoevp.ru>. (Accessed: 12th December 2023).
18. Chernikova, O.S. & Marareskul, T.A. (2023) Dvukhetapnyy podkhod k prognozirovaniyu raskhozhdeniya shkal vremeni na osnove skorrektirovannoy lineynoy modeli [A two-stage approach to forecasting the divergence of time scales based on an adjusted linear model]. *Zhurnal Belorusskogo gosudarstvennogo universiteta. Matematika. Informatika*. 2. pp. 80–93. DOI: 10.33581/2520-6508-2023-2-80-93.
19. Harvey, A. & Peters, S. (1990) Estimation procedures for structural time series models. *Journal of Forecasting*. 9. pp. 89–108.
20. Friedman, J.H. & Stuetzle, W. (1981) Projection Pursuit Regression. *Journal of the American Statistical Association*. 76. pp. 817–823.
21. Chernikova, O.S. (2022) Two-stage parametric identification procedure for a satellite motion model based on adaptive unscented Kalman filters. *Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software*. 15(4). pp. 32–43. DOI: 10.14529/mmp220403
22. Zhdanov, A.I. & Katsuba, O.A. (1982) Identifikatsiya po metodu naimen'shikh kvadratov parametrov uravneniy avtoregressii pri additivnykh oshibkakh izmereniy [Identification by the method of least squares of the parameters of autoregression equations with additive measurement errors]. *Avtomatika i telemekhanika*. 2. pp. 29–38.

Информация об авторах:

Черникова Оксана Сергеевна – доцент, кандидат технических наук, доцент кафедры теоретической и прикладной информатики Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: chernikova@corp.nstu.ru

Марарескул Татьяна Александровна – кандидат технических наук, начальник отдела АО «Информационные спутниковые системы» имени академика М.Ф. Решетнёва» (Железногорск, Россия). E-mail: mta@iss-reshetnev.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Chernikova Oksana S. (Candidate of Technical Sciences, Associate Professor at the department of theoretical and applied informatics Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: chernikova@corp.nstu.ru

Marareskul Tatiana A. (Candidate of Technical Sciences, Head of the Department of JSC Academician M.F. Reshetnev «Information Satellite System», Zheleznogorsk, Russian Federation). E-mail: mta@iss-reshetnev.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 30.05.2024; принята к публикации 03.03.2025

Received 30.05.2024; accepted for publication 03.03.2025

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

INFORMATICS AND PROGRAMMING

Научная статья

УДК 004.891.2, 004.827

doi: 10.17223/19988605/70/11

Модель знаний оболочки экспертных систем EESS

Алексей Михайлович Бабанов

*Национальный исследовательский Томский государственный университет, Томск, Россия
babanov@mail.tsu.ru*

Аннотация. Предлагается оригинальный подход к построению инструментов для тестирования знаний с использованием продукционных экспертных систем. Наличие механизма вывода и аппарата для работы с неточно представленной информацией обеспечивает недетерминированный характер работы вопросно-ответной системы. Возможность применения подхода в различных областях обеспечивает оболочка экспертных систем. Задачи тестирования знаний требуют развития традиционных представлений о продукционных экспертных системах. В статье формулируются требования к таким системам и определяется модель знаний расширенной оболочки EESS (Extended Expert System Shell). В основном идеи разработки родились в применении к проведению тестирования знаний студентов, но предлагаемые возможности будут полезны и в других областях, например в медицинской диагностике – традиционной предметной области экспертных систем.

Ключевые слова: оболочка экспертных систем; продукционная модель знаний; неопределенность информации; тестирование знаний.

Для цитирования: Бабанов А.М. Модель знаний оболочки экспертных систем EESS // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 113–121. doi: 10.17223/19988605/70/11

Original article

doi: 10.17223/19988605/70/11

Knowledge model of the EESS expert system shell

Alexey M. Babanov

Tomsk State University, Tomsk, Russian Federation, babanov@mail.tsu.ru

Abstract. The article proposes an original approach to building tools for knowledge testing using production expert systems. The presence of an inference mechanism and an apparatus for working with inaccurately presented information ensures the non-deterministic nature of the question-answer system. The ability to apply the approach in various fields is provided by the expert system shell. Knowledge testing tasks require the extension of traditional ideas about production expert systems. The article formulates requirements for such systems and defines a knowledge model of EESS (Extended Expert System Shell). The development ideas were mainly born in application to testing students' knowledge, but the proposed capabilities will also be useful in other areas, for example, in medical diagnostics – a traditional subject area of expert systems.

Keywords: expert system shell; production knowledge model; information uncertainty; knowledge testing.

For citation: Babanov, A.V. (2024) Knowledge model of the EESS expert system shell. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 113–121. doi: 10.17223/19988605/70/11

Введение

Экспертные системы (ЭС), основанные на продукциях [1–3], позволяют представить знания экспертов в виде множества правил «если..., то...». С их помощью даже неподготовленный человек способен проанализировать типичную ситуацию и сделать профессиональный вывод. В ходе консультации ЭС определяет очередной вопрос, задает его человеку, принимает ответ, агрегирует его с предыдущими ответами и в ходе логического вывода определяет следующую текущую цель, которая приведет к очередному вопросу. На основании полученных ответов ЭС придет к определенному решению, выраженному в виде некоторого значения целевой переменной. Причем ЭС не бездумно задает всегда одни и те же вопросы в фиксированном порядке, а определяет разумную траекторию консультации в зависимости от полученных ею ответов.

К положительным характеристикам ЭС относятся:

- мощные механизмы прямого и обратного логического вывода;
- многочисленные настройки логического вывода, позволяющие тонко определить поведение системы;
- возможность использования аппарата для работы с нечеткими значениями и правилами, позволяющего оперировать неточными, размытыми знаниями.

Наряду с ЭС, раз и навсегда жестко фиксирующими набор правил в исполняемом коде, существуют так называемые оболочки экспертных систем (ОЭС) [4]. Они предлагают проектировщику ЭС некоторую модель представления знаний и язык, на котором опытный эксперт должен представить свой опыт анализа ситуаций предметной области. После определения ЭС пользователи могут проводить консультации. Таким образом, эти универсальные программные продукты позволяют использовать их в самых различных предметных областях.

В ранних ЭС пользователь, запустив консультацию, отвечал на вопросы в течение одного сеанса и в конце получал на экране результат – значение целевой переменной. Ответы и результат не сохранялись. В простейших случаях этой функциональности было достаточно.

Кроме однопользовательского сценария экспертизы с простыми ответами продукционную модель представления знаний и механизмы логического вывода можно использовать и в более сложных ситуациях, например при тестировании знаний. Для этого в первую очередь и создается расширенная оболочка экспертных систем EESS.

Целью данной разработки является обеспечение применения идей и моделей представления знаний ЭС для создания интеллектуальных систем тестирования знаний студентов. Использование экспертных систем в этом случае требует развития форм ответов пользователей, распределения консультации по нескольким сеансам и между несколькими пользователями, хранения данных экспертизы между сеансами и многого другого. О том, как при этом меняется модель представления знаний, и идет речь в статье.

1. Задачи расширенной оболочки экспертных систем

Опишем использование системы для тестирования знаний. Одна ЭС предназначена для тестирования по одному предмету (в примерах будет использоваться курс «Базы данных»). В течение периода обучения последовательно изучаются и тестируются разделы курса. Перед началом освоения курса преподаватель создает ЭС, включающую множество тестовых заданий различной сложности по разделам курса и правила определения окончательной оценки за тестирование. В начале курса он запускает глобальные консультации для каждого студента, и далее система по мере необходимости сама инициирует частные консультации со студентами по разделам курса. После подготовки к тестированию очередного раздела курса студент отвечает на тестовые задания консультации, а система оценивает его ответы и последовательно агрегирует эти оценки в итоговую оценку. Таким образом, глобальная консультация продолжается весь период обучения, а студенты выполняют множество мелких консультаций (тестирований) по окончании каждого раздела курса.

С момента появления ЭС и ОЭС предпринимались многочисленные попытки их совершенствования [5–7]. В каждом случае преследовались свои определенные цели. В нашем проекте развитие модели знаний оболочки ЭС идет по нескольким направлениям.

Построение базы знаний комплексной консультации – очень непростая интеллектуальная задача, часто неподвластная одному человеку. Таким образом, ЭС может представлять собой совокупность нескольких взаимосвязанных частей (экспертных подсистем), подготовленных различными узкими специалистами, но в целом решающих одну задачу. Это напоминает ситуацию, когда врач-терапевт направляет пациента на консультации к специалистам и впоследствии учитывает их мнения при постановке окончательного диагноза. Значит, система должна предусматривать соавторство экспертов и обеспечивать в ходе консультации обращение к другим консультациям. Этого же требует предлагаемая процедура тестирования. Экспертизы по разным разделам часто определяют различные педагоги.

В сложных случаях, когда одному человеку затруднительно отвечать на вопросы ЭС, требующие специальных знаний, можно устраивать что-то, напоминающее консилиум. При этом система при необходимости посылает вопрос нужному специалисту и продолжает консультацию, получив его ответ. Таким образом, консультация становится распределенной и многопользовательской. В нашем примере с тестированием начинает глобальную консультацию один пользователь (преподаватель), и далее подключаются другие пользователи (студенты).

Одно из применений расширенной оболочки – создание тестирующих систем, в которых традиционно используются самые разнообразные варианты тестовых заданий. Напрашивается расширение спектра возможных видов ответов. Помимо ввода одного четкого значения, неплохо было бы предусмотреть:

- ввод нечеткого значения (в случае сомнений пользователей);
- выбор из предложенных альтернатив (простой или множественный);
- ввод нескольких значений;
- упорядоченный ввод нескольких значений;
- упорядочивание предложенных альтернатив.

И, наконец, хорошо было бы иметь возможность откладывать консультацию, сохранять полученные ответы, промежуточные и окончательные результаты в долговременном хранилище для последующего возобновления консультации, анализа результатов и, возможно, изменения ответов и хода вывода. Это абсолютно необходимо для задачи тестирования знаний по всему курсу.

2. Общение с пользователем

Вся информация, касающаяся одной конкретной консультации (или экспертизы), представляется в виде значений переменных. Назовем их переменными экспертизы. Они могут быть четкими (иметь одно значение) или нечеткими. Каждое из значений нечеткой переменной имеет свой фактор уверенности – ФУ (число от 0 до 100, отражающее уверенность в том, что переменная имеет данное значение). Наряду с русскоязычной аббревиатурой ФУ часто используется англоязычная CF (Certainty Factor).

Кроме переменных экспертизы в модели знаний имеются системные переменные, которые предназначены для настройки ЭС на определенную стратегию поведения. Эти переменные будем называть переменными среды (их имена имеют префикс «Е.»). Некоторые из них, связанные с выводом и нечеткой информацией, будут представлены в дальнейшем. Остальные останутся за рамками статьи.

Если значения переменных среды устанавливаются разработчиком ЭС, хранятся в ней и преимущественно не изменяются в ходе консультации, то значения переменных экспертизы определяются либо ответами пользователя, либо механизмом вывода автоматически с использованием правил. Причем способ означивания система выбирает на основании значений переменных среды.

Проектируя работу с ответами пользователя, разработчики анализировали два варианта значений переменных экспертизы.

1-й вариант. Значение переменной – сам введенный объект (строка, номер альтернативы, множество, последовательность). Тогда надо расширять набор типов переменных, включая в него множества и последовательности. В этом случае сравнение с эталоном ответа должно осуществляться в посылке правил, что приведет к серьезному расширению грамматики и необходимости реализации сравнений множеств и последовательностей. К тому же этот вариант неудобен при массовом вводе тестов (а это неизбежно) и потребует большого количества правил, что необоснованно перегрузит ЭС.

2-й вариант. Реализовать специальные функции ввода ответов (строка, номер альтернативы, множество, последовательность) и в них осуществлять проверку с эталоном, представленным в параметрах функций, а возвращать агрегированное значение оценки выполнения задания (в истинностных значениях или в баллах). Тогда в посылках правил будет проще проверять любые ответы единообразно. При этом можно определять задания заранее и хранить их вместе с эталонами ответов в таблицах БД, а в определении переменной ссылаться на элементы заданий одного типа. Его можно выбирать и случайно.

Второй вариант проще и перспективней.

Покажем, например, как может выглядеть его реализация для ввода одиночного значения (функция GETVAL) простого (функция GETIND) и множественного (функция GETSET) выбора из предложенных альтернатив, а также выбора и упорядочивания (функция GETSEQ) предложенных альтернатив.

Заранее (в момент создания ЭС) для каждой функции вводятся таблицы с заданиями (табл. 1–4), что позволяет не указывать параметры тестового задания при обращении к функции.

Таблица 1

Таблица базы данных ЭС с заданиями, использующими функцию GETVAL

Переменная	Параметр QWE	Параметр ANSW
TEST1	Как называется комплекс, все элементы которого лежат в одной позиции?	Множество
TEST1	Другой вопрос	Другой ответ

Определение переменной TEST1 может выглядеть так. В первом варианте задание указывается в параметрах функции, во втором – ссылкой на первую строку соответствующей таблицы БД с заданным именем переменной. Для остальных примеров используется тот же принцип.

VAR: TEST1

GETVAL QWE: «Как называется комплекс, все элементы которого лежат в одной позиции?»

ANSW: «Множество».

Или обращаемся к ранее введенной таблице GETVAL (1).

Переменная принимает четкое значение TRUE или FALSE в зависимости от того, совпал ответ или нет.

Таблица 2

Таблица базы данных ЭС с заданиями, использующими функцию GETIND

Переменная	Параметр QWE	Параметр ALT	Параметр ANSW
TEST2	Что не является множеством?	1) отношение, 2) домен, 3) кортеж, 4) атрибут	3
TEST2	Другой вопрос		Другой ответ

VAR: TEST2

GETIND QWE: «Что не является множеством?» ALT: 1) отношение, 2) домен, 3) кортеж, 4) атрибут. ANSW: 3.

Или обращаемся к ранее введенной таблице GETIND (1).

Переменная принимает четкое значение TRUE или FALSE в зависимости от того, совпал ответ или нет.

Таблица 3

Таблица базы данных ЭС с заданиями, использующими функцию GETSET

Переменная	Параметр QWE	Параметр ALT	Параметр ANSW
TEST3	Укажите действия с данными, которые есть в любой модели данных	1) удаление, 2) установка текущей, 3) изменение, 4) включение, 5) выборка, 6) селекция, 7) поиск	{1, 3, 4, 5}
TEST3	Другой вопрос		Другой ответ

VAR: TEST3

GETSET QWE: «Укажите действия с данными, которые есть в любой модели данных» ALT: 1) удаление, 2) установка текущей, 3) изменение, 4) включение, 5) выборка, 6) селекция, 7) поиск. ANSW: {1, 3, 4, 5}.

Или обращаемся к ранее введенной таблице GETSET (1).

В этом случае для вычисления значения переменной возможны различные варианты манипуляции с количеством совпавших (С), количеством недостающих (Н) и количеством лишних (Л) элементов. Например, С, С-Н-Л и т.д.

Переменная принимает нечеткое значение TRUE с соответствующим ФУ. Если результат (С или С-Н-Л) совпал с мощностью эталона, значение = TRUE с ФУ 100. Если результат ≤ 0 , значение = TRUE с ФУ 0. Таким образом, ФУ для TRUE равен $\text{MAX}((\text{С или С-Н-Л})/|\text{ЭТАЛОН}|, 0) \times 100$.

Например, если ответ будет {1, 3, 4, 6, 7}, С = 3, Н = 1, Л = 2, С-Н-Л = 0. Значение = TRUE с ФУ 0. Если ответ будет {1, 2, 4, 6, 7}, С = 2, Н = 2, Л = 3, С-Н-Л = -3. Значение = TRUE с ФУ 0. Если ответ будет {1, 2, 3, 4, 5}, С = 4, Н = 0, Л = 1, С-Н-Л = 3. Значение = TRUE с ФУ 75.

Таблица 4

Таблица базы данных ЭС с заданиями, использующими функцию GETSEQ

Переменная	Параметр QWE	Параметр ALT	Параметр ANSW
TEST4	Укажите формы данных, соответствующие элементам восприятия мира человеком: связь, характеристика, тип связей, тип объектов, объект	1) отношение, 2) домен, 3) атрибут, 4) комплекс, 5) кортеж	<5, 3, 1, 1, 5>
TEST4	Другой вопрос		Другой ответ

VAR: TEST4

GETSEQ QWE: «Укажите формы данных, соответствующие элементам восприятия мира человеком: связь, характеристика, тип связей, тип объектов, объект» ALT: 1) отношение, 2) домен, 3) атрибут, 4) комплекс, 5) кортеж. ANSW: <5, 3, 1, 1, 5>.

Или обращаемся к ранее введенной таблице GETSEQ (1).

В этом случае для вычисления значения переменной возможны различные варианты манипуляции с количеством совпавших (С) и количеством стоящих на своих позициях (П) элементов. Например, С + П и т.д.

Переменная принимает нечеткое значение TRUE с соответствующим ФУ. Если результат (С + П) совпал с удвоенной мощностью эталона, значение = TRUE с ФУ 100. Если результат = 0, значение = TRUE с ФУ 0. Таким образом, ФУ для TRUE равен $(\text{С} + \text{П}) / (2 \times |\text{ЭТАЛОН}|) \times 100$.

Например, если ответ будет <1, 3, 4, 5, 5>, С = 4, П = 2, С + П = 6. Значение = TRUE с ФУ 60. Если ответ будет <1, 2, 4, 6, 7>, С = 1, П = 0, С + П = 1. Значение = TRUE с ФУ 10. Если ответ будет <5, 4, 3, 2, 1>, С = 3, П = 1, С + П = 4. Значение = TRUE с ФУ 40.

3. Правила вывода

Другой способ означивания переменных экспертизы применяется механизмом вывода. Каждая ЭС имеет целевую переменную экспертизы, и задачей системы является получение значения этой переменной. Это значение не вводится пользователем (иначе как бы система проявляла свой интеллект), а выводится на основании правила (или нескольких правил). Поэтому набор правил вывода составляет ядро знаний системы [8].

Основные элементы любого правила – посылка (IF) и заключение (THEN). Посылка представляет собой некоторое условие; если оно верно, то выполняются действия, указанные в заключении (в основном присвоения значений переменных экспертизы).

Условие посылки – это логическое выражение, состоящее из операций сравнения значений переменных экспертизы с точными константами и логических операций (NOT, AND, OR). Заключение составляет одна или несколько операций присвоения переменной экспертизы четких или нечетких значений. Операция присвоения “=” замещает значение, если оно существовало. К переменной также можно добавлять новые значения (с сохранением имеющихся). Для этого используется операция “+=”. Можно также удалять значения. Для этого применяется операция “-=”. Каждое добавление того же значения увеличивает его ФУ, каждое удаление значения уменьшает.

IF TEST3 >= 80

THEN MARK = {4 CF 90, 5 CF 100}

Здесь каждое значение нечеткой переменной имеет фактор уверенности (он указывается после ключевого слова CF).

В качестве источника значения в операции присвоения может быть указана консультация, значение целевой переменной которой и будет использовано в этом случае; например, THEN MARK = CONSULT(THEORY, “Иванов”). Здесь THEORY – имя экспертной системы, а “Иванов” – имя пользователя, который будет отвечать на вопросы во время консультации.

4. Работа с неточной информацией

Одним из обязательных требований к ЭС является наличие механизмов представления и обработки нечетких и недостоверных знаний [9–10]. Основными способами представления таких знаний являются:

- факторы уверенности, отражающие степень уверенности в некотором элементе знаний – значении переменной, посылке правила;
- нечеткие переменные – переменные, принимающие одновременно несколько значений, каждое со своим фактором уверенности.

Операции по обработке факторов уверенности, выполняемые в ходе сеанса работы с ЭС, сводятся к следующим:

- определение фактора уверенности посылки на основании значений переменных, используемых в ней;
- определение фактора уверенности значений переменных заключения после применения правила;
- определение обобщенного фактора уверенности значения переменной, полученного применением нескольких правил.

Способ вычисления фактора уверенности посылки заключается в поэтапном применении формул преобразования ФУ каждого из двух операндов бинарной логической операции в их общий ФУ. Для операции AND алгоритм преобразования ФУ задается переменной среды E.CFA (Certainty Factor of AND), а для операции OR – переменной E.CFO (Certainty Factor of OR). При использовании логической связки AND общий фактор уверенности не превышает минимального из факторов уверенности операндов (так как для истинности операции необходима истинность обоих условий). При использовании логической связки OR фактор уверенности посылки не может быть меньше, чем максимальный из факторов уверенности отдельных операндов (так как для истинности операции достаточно, чтобы истинным было хотя бы одно из условий).

Понятно, что для такого алгоритма подсчета ФУ посылки ее логическое выражение должно быть представлено в соответствующем виде, например в КНФ или в ДНФ. Этого можно потребовать от эксперта при создании ЭС, а можно строить их автоматически и сохранять в БД уже готовыми для вычислений ФУ.

Возможные значения этих переменных среды и соответствующие им формулы для вычисления фактора уверенности операции приведены в табл. 5, где CF₁, CF₂ – факторы уверенности операндов бинарной операции, CF_{оп} – фактор уверенности операции в целом).

Возможные значения переменных среды и формулы для вычисления фактора уверенности бинарной логической операции

Значения переменных	E.CFA	E.CFO
М	$CF_{оп} = \min(CF_1, CF_2)$	$CF_{оп} = \max(CF_1, CF_2)$
Р	$CF_{оп} = CF_1 \times CF_2 / 100$	$CF_{оп} = CF_1 + CF_2 - CF_1 \times CF_2 / 100$

Формулы, применяемые для вычисления факторов уверенности в экспертных системах, не имеют строгого математического обоснования. Формулы, соответствующие значениям “М”, близки к интуитивным правилам оценки уверенности, применяемым в рассуждениях человека и теории нечетких множеств. Формулы, соответствующие значениям “Р”, применяются в теории вероятностей.

Способ вычисления фактора уверенности значений переменных заключения, получаемых в результате применения правила, задается значением переменной среды E.CFSV (Certainty Factor of Single Value). Способ вычисления обобщенного фактора уверенности для совпадающих значений одной и той же переменной, полученных в нескольких правилах, задается значением переменной E.CFMV (Certainty Factor of Multiple Values).

Возможные значения этих переменных среды и соответствующие им формулы для вычисления факторов уверенности приведены в табл. 6. В таблице использованы следующие обозначения: $CF_{пос}$ – фактор уверенности посылки; $CF_{ук}$ – фактор уверенности значения, указанный в заключении; $CF_{закл}$ – фактор уверенности значения, полученный в результате применения правила; $CF_{закл1}$, $CF_{закл2}$ – факторы уверенности одного и того же значения некоторой переменной, полученного в заключениях разных правил; $CF_{об}$ – обобщенный фактор уверенности значения переменной, полученного в результате применения обоих правил.

Таблица 6

Возможные значения переменных среды и формулы для вычисления фактора уверенности значений переменных заключения

Значения	E.CFSV	E.CFMV
М	$CF_{закл} = \min(CF_{пос}, CF_{ук})$	$CF_{об} = \max(CF_{закл1}, CF_{закл2})$
Р	$CF_{закл} = CF_{пос} \times CF_{ук} / 100$	$CF_{об} = CF_{закл1} + CF_{закл2} - CF_{закл1} \times CF_{закл2} / 100$

Обобщенный фактор уверенности повышается по сравнению с факторами уверенности, полученными в отдельных правилах, так как значение переменной подтверждается несколькими правилами.

Покажем, как происходит работа ЭС в условиях неполной информации.

Пусть $TEST3 = \{TRUE \text{ CF } 75\}$, $TEST4 = \{TRUE \text{ CF } 60\}$, $E.CFA = \text{“М”}$, $E.CFO = \text{“М”}$, $E.CFSV = \text{“Р”}$, $E.CFMV = \text{“Р”}$.

В примере используется системная функция $CFV(<переменная>, <значение>)$. Она возвращает фактор уверенности указанного значения переменной; если такого значения нет, то возвращается фактор уверенности, равный нулю.

Предположим, что механизм вывода последовательно применяет правила R1 и R2.

Сначала рассматривается правило R1.

IF $TEST3 = TRUE \text{ AND } CFV(TEST3, TRUE) \geq 70$

THEN $MARK += \{3 \text{ CF } 60, 4 \text{ CF } 90, 5 \text{ CF } 100\}$.

Фактор уверенности условия $TEST3 = TRUE$ равен 75. Условие $CFV(TEST3, TRUE) \geq 70$ равно TRUE с ФУ 100. Фактор уверенности посылки в целом находится как минимальный из факторов уверенности отдельных условий (в соответствии с переменной среды E.CFA, так как в правиле используется связка AND): $\min(75, 100) = 75$.

Далее находятся факторы уверенности значений переменной заключения. Вычисление выполняется в соответствии со значением переменной среды E.CFSV. Для значения 3 фактор уверенности находится как $75 \times 60 / 100 = 45$ (здесь 75 – фактор уверенности посылки, 60 – фактор уверенности значения 3, указанный в заключении правила). Аналогично находятся факторы уверенности для значений 4 ($75 \times 90 / 100 = 68$) и 5 ($75 \times 100 / 100 = 75$). Таким образом, переменная MARK после рассмот-

рения правила R1 принимает следующие значения: {3 CF 45, 4 CF 68, 5 CF 75}. ФУ здесь представляет собой уверенность в соответствующем значении оценки за определенный раздел курса.

Далее обрабатывается правило R2.

IF TEST4 = TRUE AND CFV(TEST4, TRUE) >= 60

THEN MARK += {3 CF 70, 4 CF 90, 5 CF 80}.

Фактор уверенности условия TEST4 = TRUE равен 60. Условие CFV(TEST4, TRUE) >= 60 равно TRUE с ФУ 100. Фактор уверенности посылки в целом находится как минимальный из факторов уверенности отдельных условий (в соответствии с переменной среды E.CFA, так как в правиле используется связка AND): $\min(60, 100) = 60$.

Далее находятся факторы уверенности значений переменной заключения. Вычисление выполняется в соответствии со значением переменной среды E.CFSV. Для значения 3 фактор уверенности находится как $60 \times 70/100 = 42$. Аналогично находятся факторы уверенности для значений 4 ($60 \times 90/100 = 54$) и 5 ($60 \times 80/100 = 48$). Таким образом, переменная MARK в заключении правила R2 принимает следующие значения: {3 CF 42, 4 CF 54, 5 CF 48}.

Поскольку в заключении используется операция "+=", эти значения добавляются к значениям переменной MARK, полученным ранее (в правиле R1). К совпадающим значениям 3, 4 и 5 применяется алгоритм обобщения в соответствии со значением переменной E.CFMV. Для значения 3 фактор уверенности находится как $45 + 42 - 45 \times 42/100 = 68$. Аналогично находятся факторы уверенности для значений 4 ($68 + 54 - 68 \times 54/100 = 85$) и 5 ($75 + 48 - 75 \times 48/100 = 87$).

Таким образом, в результате выполнения правил R1 и R2 переменная MARK принимает следующие значения {3 CF 68, 4 CF 85, 5 CF 87}, и ЭС может сделать вывод, что по результатам первых двух тестовых заданий студент заслуживает оценки «отлично». Действительно, значение 5 имеет наивысший фактор уверенности.

Заключение

В статье представлены основные нетрадиционные элементы модели расширенной оболочки экспертных систем EESS, позволяющие применить технологию ЭС для тестирования знаний. К ним относятся:

- развитые способы получения ответов пользователей;
- использование в заключениях правил значений целевых переменных других ЭС;
- разнообразные возможности для учета неполноты информации.

Эти способности модели продиктованы требованиями, предъявляемыми со стороны экспертных систем для тестирования знаний студентов. Однако оболочка имеет универсальный характер и с успехом может использоваться во многих предметных областях.

Список источников

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach (Pearson Series in Artificial Intelligence). 4th ed. Pearson, 2020. 1136 p.
2. Vigo R., Zeigler D.E., Wimsatt J. Uncharted Aspects of Human Intelligence in Knowledge Based “Intelligent” Systems // Philosophies. 2022. V. 7 (46). P. 1–19. doi: 10.3390/philosophies7030046
3. Krivoulya G.F., Shkil A.S., Kucherenko D.Y. Analysis of production rules in expert systems of diagnosis // Automatic Control and Computer Sciences. 2013. Is. 47. P. 331–341.
4. Knowledge Engineering Shells / N.G. Bourbakis (ed.). Singapore ; River Edge, NJ : World Scientific Publishing Company, 1993. 536 p. (Advanced Series On Artificial Intelligence; v. 2).
5. Николайчук О.А., Павлов А.И., Юрин А.Ю. Компонентный подход: модуль производственной экспертной системы // Программные продукты и системы. 2010. № 3. С. 41–44.
6. Грищенко М.А., Николайчук О.А., Павлов А.И., Юрин А.Ю. Инструментальное средство создания производственных экспертных систем на основе MDA // Образовательные ресурсы и технологии. 2016. № 2 (14). С. 144–151.
7. Еремеев А.П. Проектирование экспертных систем средствами инструментальной системы GURU. М. : МЭИ, 1996. 52 с.
8. Nalepa G.J. Rules as a Knowledge Representation Paradigm // Nalepa G.J. Modeling with Rules Using Semantic Knowledge Engineering. Springer, 2018. P. 3–25. doi: 10.1007/978-3-319-66655-6

9. Walley P. Measures of uncertainty in expert systems // *Artificial Intelligence*. 1996. V. 83. P. 1–58.
10. Dubey S., Pandey R.K., Gautam S.S. Dealing with Uncertainty in Expert Systems // *International Journal of Soft Computing and Engineering (IJSCE)*. 2014. V. 4 (3). P. 105–111.

References

1. Russell, S. & Norvig, P. (2020) *Artificial Intelligence: A Modern Approach (Pearson Series in Artificial Intelligence)*. 4th ed. Pearson.
2. Vigo, R., Zeigler, D.E. & Wimsatt, J. (2022) Uncharted Aspects of Human Intelligence in Knowledge Based “Intelligent” Systems. *Philosophies*. 7(46). pp. 1–19. DOI: 10.3390/philosophies7030046
3. Krivoulya, G.F., Shkil, A.S. & Kucherenko, D.Y. (2013) Analysis of production rules in expert systems of diagnosis. *Automatic Control and Computer Sciences*. 47. pp. 331–341.
4. Bourbakis, N.G. (ed.) (1993) *Knowledge Engineering Shells*. Singapore; River Edge, NJ: World Scientific Publishing Company.
5. Nikolaychuk, O.A., Pavlov, A.I. & Yurin, A.Yu. (2010) Komponentnyy podkhod: modul' produktsionnoy ekspertnoy sistemy [Component approach: Module of production expert system]. *Programmnye produkty i sistemy*. 3. pp. 41–44.
6. Grishchenko, M.A., Nikolaychuk, O.A., Pavlov, A.I. & Yurin, A.Yu. (2016) Instrumental'noe sredstvo sozdaniya produktsionnykh ekspertnykh sistem na osnove MDA [Tool for creating production expert systems based on MDA]. *Obrazovatel'nye resursy i tekhnologii*. 2(14). pp. 144–151.
7. Eremeev, A.P. (1996) *Proektirovanie ekspertnykh sistem sredstvami instrumental'noy sistemy GURU* [Design of expert systems by means of the GURU instrumental system]. Moscow: MEI.
8. Nalepa, G.J. (2018) *Modeling with Rules Using Semantic Knowledge Engineering*. Springer. pp. 3–25. DOI: 10.1007/978-3-319-66655-6
9. Walley, P. (1996) Measures of uncertainty in expert systems. *Artificial Intelligence*. 83. pp. 1–58.
10. Dubey, S., Pandey, R.K. & Gautam, S.S. (2014) Dealing with Uncertainty in Expert Systems. *International Journal of Soft Computing and Engineering (IJSCE)* 4(3). pp. 105–111.

Информация об авторе:

Бабанов Алексей Михайлович – доцент, кандидат технических наук, доцент кафедры программной инженерии Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: babanov2000@mail.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Babanov Alexey M. (Associate Professor, Candidate of Technical Sciences, Associate Professor of the Department of Software Engineering at the Institute of Applied Mathematics and Computer Science of the National Research Tomsk State University, Tomsk, Russian Federation). E-mail: babanov2000@mail.ru

The author declares no conflicts of interests.

Поступила в редакцию 26.08.2024; принята к публикации 03.03.2025

Received 26.08.2024; accepted for publication 03.03.2025

Научная статья

УДК 620.179.1, 004.415.2

doi: 10.17223/19988605/70/12

Вычислительная диагностика объектов, позволяющих эвристическое усиление алгоритмов решения обратных задач**Сергей Михайлович Зеркаль¹, Александр Викторович Пешков²**^{1,2} *Новосибирский государственный технический университет, Новосибирск, Россия*¹ *zermal@corp.nstu.ru*² *mupeskov1997@mail.ru*

Аннотация. Рассматривается задача диагностики промышленных изделий на наличие в них тонких протяженных дефектов типа трещин и отслоений. Данная задача решается в условиях томографической реконструкции и принадлежит к классу обратных задач. Классический алгоритм компьютерной томографии имеет высокую трудоемкость при решении данной задачи, а в случае малой толщины дефекта может и вообще не восстановить его. В данной работе разработано программное обеспечение, позволяющее решить специальную задачу дефектоскопии на основе модифицированного алгоритма, использующего априорную информацию об эталонном образце изделия в качестве эвристического усиления. Полученные результаты вычислительных экспериментов демонстрируют эффективность предложенного подхода при решении задачи обнаружения трещин в исследуемом объекте и позволяют сделать вывод о его перспективности в задачах промышленной дефектоскопии.

Ключевые слова: программное обеспечение; компьютерная томография; промышленная дефектоскопия; эвристический алгоритм; эталонный образец; многопоточная обработка.

Для цитирования: Зеркаль С.М., Пешков А.В. Вычислительная диагностика объектов, позволяющих эвристическое усиление алгоритмов решения обратных задач // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 70. С. 122–133. doi: 10.17223/19988605/70/12

Original article

doi: 10.17223/19988605/70/12

Computational diagnostics of objects that allow the heuristic strengthening of algorithms for solving reverse problems**Sergey M. Zerkal¹, Alexander V. Peshkov²**^{1,2} *Novosibirsk State Technical University, Novosibirsk, Russian Federation*¹ *zermal@corp.nstu.ru*² *mupeskov1997@mail.ru*

Abstract. The task of diagnosing industrial products for the presence of thin-length defects such as cracks and abrasions is considered. This problem is solved in the conditions of tomographic reconstruction and belongs to the class of reverse tasks. The classic algorithm of computed tomography has a high complexity in solving this problem, and in the case of a small thickness of the defect, it may not restore it at all. In this paper, software has been developed to solve a special problem of flaw detection based on a modified algorithm that uses a priori information about the reference of the product as a heuristic amplification. The results of computational experiments demonstrate the effectiveness of the proposed approach in solving the problem of detecting cracks in the object under study and allow us to conclude that its prospects in the tasks of industrial flaw detector.

Keywords: software; computed tomography; industrial flaw detection; heuristic algorithm; reference sample; multi-thread processing.

For citation: Zerkal, S.M., Peshkov, A.V. (2025) Computational diagnostics of objects that allow the heuristic strengthening of algorithms for solving reverse problems. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 70. pp. 122–133. doi: 10.17223/19988605/70/12

Введение

Исследования, о которых ниже пойдет речь, относятся к обширному и важному научно-техническому направлению *интроскопия*, которое следует понимать как «внутреннее видение и описание» по аналогии с термином *рентгеноскопия*. Одним из важных инструментов здесь является компьютерная (вычислительная) томография (КТ). КТ – научное направление, имеющее широкое распространение в своих приложениях. Она появилась в середине XX столетия и пережила бурное развитие во второй его половине, за создание медицинского томографа были присуждены Нобелевские премии (1979, 2003).

Суть метода КТ заключается в определении внутреннего устройства, геометрической конфигурации, физического состояния, химического состава и других свойств исследуемого объекта по результатам измерения функционалов зондирующего излучения, прошедшего через объект исследования (трансмиссионная томография) [1], или излучения, генерируемого самим объектом (эмиссионная томография) [2]. Очевидно, что такие возможности КТ востребованы во многих областях человеческого знания и деятельности. Успехи КТ как метода вычислительной диагностики определяются во многом способностью использовать весьма значительный объем измеренной информации об исследуемом объекте по сравнению с другими известными методами [3].

Одно из важных приложений КТ – неразрушающий контроль промышленных изделий, или промышленная дефектоскопия. При этом габариты используемой томографической системы определяются в зависимости от размеров изделия. К настоящему времени кроме традиционной (классической) КТ возникли и нередко успешно развиваются проявления КТ, существенно отличающиеся от классического варианта [4–7]. Именно к такой неклассической КТ относится *специальная задача дефектоскопии*. Ее постановка, построение решения и вычислительный алгоритм, реализующий это решение, излагаются ниже.

Промышленные томографы, использующие более жесткое рентгеновское излучение (в отличие от медицинских), с реализацией измерений по методикам классической КТ успешно справляются. Следует отметить, что успехи КТ в промышленности несколько скромнее, чем в медицине [8]. Это, конечно, объясняется объектами исследования. Промышленная КТ применяется прежде всего в дефектоскопии, где возникают именно специфические задачи, затрудняющие непосредственное использование КТ или вообще исключающие реализацию классических вычислительных алгоритмов КТ [9, 10]. Одной из подобных задач, имеющей важное прикладное значение, является диагностика таких объектов, как тонкие протяженные трещины, отслоения, нарушения сплошности поверхностей склеивания и т.п.

Геометрия этих объектов такова, что порождает их избирательную чувствительность к зондирующему сигналу, что приводит к неполноте проекционных данных (результатов измерений) и, как следствие, к невозможности использовать метод КТ в его классическом варианте [9]. Задача КТ с неполными проекционными данными возникает достаточно часто, если, например, выполнить измерения по тем или иным причинам с определенных ракурсов невозможно. Для этого к настоящему времени разработаны определенные технологии «пополнения» проекционных данных [11–13]. В нашем случае ситуация неполноты данных связана со спецификой исследуемого объекта и чувствительностью измерительного тракта, т.е. приемников. Процедура «пополнения» здесь не имеет смысла, ибо все требуемые измерения выполнены, и дело в том, что часть из них не является информативной. Потеря информативности, конечно, мнимая, но использовать эти измерения нельзя. Действительно информативными будут лучи, которые имеют в своем распространении достаточный путь внутри дефекта – так называемые «тангенциальные» лучи. А лучи, проходящие дефект в крест его прости-

рания (называемые «нормальными»), из-за малой ширины раскрытия дефекта его не почувствуют (рис. 1).

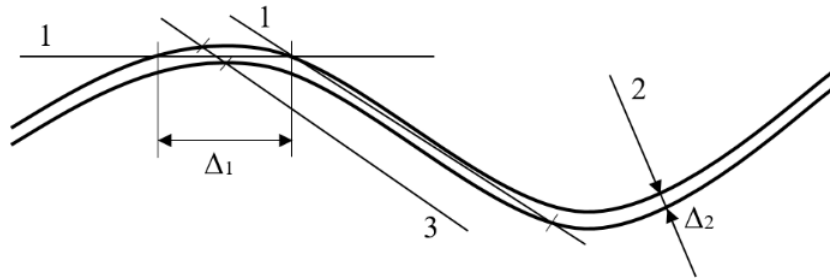


Рис. 1. Классификация типов лучей (1 – тангенциальный луч, Δ_1 – его путь по дефекту; 2 – нормальный луч, Δ_2 – его путь по дефекту, $\Delta_1 \gg \Delta_2$; 3 – промежуточный луч)
 Fig. 1. Classification of types of rays (1 – tangential ray, Δ_1 – its path along the defect; 2 – normal ray, Δ_2 – its path along the defect, $\Delta_1 \gg \Delta_2$; 3 – intermediate ray)

Изначально отмеченная ситуация в рентгеновской интроскопии в применении к промышленной дефектоскопии появилась как вычислительная задача и как проблема прикладной математики в научной школе академика М.М. Лаврентьева в конце 80-х гг. прошлого столетия. Возникла настоятельная потребность диагностики отслоений и тещин в ответственных промышленных изделиях для обеспечения безопасности их использования, в том числе и в условиях эксплуатационного мониторинга. Здесь логично выделить традиционно два уровня исследований:

1. На уровне теорем, необходимых для обратной задачи математической физики, исследовать теоретически:

а) существование и единственность решения;

б) степень математической некорректности данной задачи, зависимость ошибки решения от погрешностей в исходных данных [14].

2. Разработать вычислительный алгоритм для возникшей в условиях КТ задачи промышленной дефектоскопии [15].

Рассматриваемая задача получила название *специальной задачи дефектоскопии*. При построении вычислительного алгоритма удалось использовать априорную информацию, заключающуюся в том, что об испытуемом изделии известно практически все, поскольку имеются проектная документация и возможность путем экспериментов получить проекционную матрицу так называемого *эталонного образца*. Действительно, есть возможность после снятия проекционных данных разобрать испытуемое изделие и визуально его исследовать, а в случае отсутствия дефектов принять результаты измерений за проекционную матрицу эталонного образца. Кроме того, поскольку законы взаимодействия зондирующего излучения описываются математическими формулами, есть возможность рассчитать синтетическую проекционную матрицу эталонного изделия. Таким образом, в распоряжении исследователя находятся две проекционные матрицы, а именно эталонного образца и испытуемого промышленного изделия (объекта исследования). Это позволило составить вычислительный алгоритм геометрического «экспресс»-построения огибающей дефекта типа трещины или отслоения [16]. Алгоритм оказался достаточно информативным и послужил аносом для дальнейшего более глубокого исследования с использованием аппарата искусственного интеллекта (ИИ) в части целенаправленного эвристического поиска решения исследуемой обратной задачи.

В настоящее время наиболее распространено несколько различных алгоритмов, используемых в коммерческой томографии: алгебраическая реконструкция, итеративная реконструкция, фильтрованная обратная проекция, сверточная обратная проекция и реконструкция, основанная на глубоком обучении [17, 18]. Однако все эти методы и их модификации не учитывают специфику задач промышленной дефектоскопии, например, возможную неполноту проекционных данных. Для преодоления этой проблемы предлагается использовать некую априорную информацию, которая является важным компонентом при решении обратной задачи томографической реконструкции.

Цель данной работы – разработка эффективной реализации алгоритма, использующего априорную информацию, для вычислительной диагностики объектов с целью выявления в них тонких протяженных дефектов.

1. Постановка задачи

Задача компьютерной томографии заключается в нахождении функции коэффициента поглощения внутри объекта по известным значениям интегралов вдоль лучей, проходящих через объект. Вращая всю систему из источника излучения и детекторов вокруг объекта (при этом оставаясь в одной плоскости) или вращая сам объект вокруг оси, перпендикулярной плоскости системы наблюдений, получают достаточно полную информацию о преобразовании Радона оптической плотности в данном срезе объекта. Используя обратное преобразование Радона, можно восстановить поглощение излучения в любой точке данной плоскости объекта [19]. Однако из-за неполноты проекционных данных использовать интегральные преобразования в данной задаче не представляется возможным, поэтому используется их алгебраизация [20]. То есть решение задачи томографической реконструкции сводится к решению системы линейных алгебраических уравнений (СЛАУ) вида:

$$A \times X = B, \quad (1)$$

где A – матрица длин лучей в пикселях сетки, X – вектор значений плотности объекта в каждом пикселе, B – вектор значений интенсивности лучей на приемнике.

Таким образом, найденные корни системы (1) будут равны коэффициентам поглощения в соответствующих пикселях сетки, наложенной на объект исследования. На основе найденных коэффициентов и известной модели поглощения излучения можно перейти к искомым величинам плотностей внутри объекта.

Однако при использовании классического алгоритма КТ для диагностики тонких дефектов необходимо использовать очень мелкую сетку и большое количество измерений, что приводит к резкому увеличению вычислительных затрат (сетка 50×50 порождает систему из 2 500 уравнений, 80×80 – уже 6 400 уравнений и т.д.). Для возможности решения данной задачи за приемлемое время и без необходимости увеличения количества измерений используется эвристическое усиление, суть которого описана ниже.

2. Использование априорной информации в томографической реконструкции

В условиях промышленной дефектоскопии есть одно важное обстоятельство, которое позволяет отойти от исчислимой обратной задачи математической физики (каковой и является классическая томографическая задача) и включить в рассмотрение свойственный ИИ эвристический подход, основанный на доминировании априорной информации. Такой подход способствует построению «сильных» алгоритмов, способных решить усложненные задачи, недоступные в условиях исчислимости. В данном случае такая априорная информация есть – это проектная документация о промышленном изделии, который является объектом исследования.

Очевидно, что методы решения тех или иных задач имеют границы применимости, определяющие их возможности. Для исчислимых задач весьма редко эти границы удастся даже незначительно расширить. Это становится возможным при появлении дополнительной информации об объекте исследования. Отмеченная информация суть дополнительные знания, порожденные этой информацией. Если такие знания удастся встроить в вычислительный алгоритм, то границы применимости метода расширятся. Сказанное выше соответствует основному принципу инженерии знаний в том смысле, что возможности «решателя» в первую очередь определяются его информационной базой и лишь во вторую – используемым методом. В случае специальной задачи дефектоскопии, имеющей исчислимую основу, а именно метод КТ, все-таки движение идет на первом шаге от метода, где задача предъявляет свои требования к усилению используемого метода.

С использованием исключительного объема информации об исследуемом объекте такое усиление становится возможным. Строится проекционная матрица эталонного образца A_R , и выполняется

измерение проекционной матрицы для испытуемого изделия. Особенность задачи заключается в том, что не требуется получить информацию об исследуемом объекте как таковом, ибо это известно, а интерес представляют его изменения, причем незначительные, которые могут быть или нет. Построение разности проекционных матриц

$$A^* = A_T - A_R, \quad (2)$$

где A_R – проекционная матрица эталонного образца, A_T – проекционная матрица испытуемого изделия, с последующим использованием ненулевых элементов A^* является вторым шагом.

Этот шаг можно считать началом процесса целенаправленного «эвристического поиска» диагностируемого дефекта. Действительно, появляется информация о наличии того, что мы ищем, т.е. возникает цель, и она расположена внутри известного объекта, остается локализовать эту цель и определить ее геометрию. В статье [21] описан вычислительно-эвристический алгоритм, позволяющий осуществить поиск дефекта в исследуемом объекте. Эвристика заключается в том, что поиск дефекта осуществляется не во всех пикселях наложенной сетки, а только в «потенциальных». Таким образом, на основе априорной информации об эталонном образце изделия удастся заметно сократить число вариантов потенциальных решений.

3. Разработка программного обеспечения

Для решения задачи диагностики объектов на основе эвристического усиления было разработано оконное приложение с использованием языков программирования *Java* и *Matlab*. За основу было взято ранее разработанное программное обеспечение [22], которое позволяет строить модель объекта с круглым сечением и дефектом заданной формы, генерировать проекционные матрицы сканирования, а также производить томографическую реконструкцию объекта классическим и модифицированным алгоритмами. Подробное описание реализаций данных алгоритмов томографической реконструкции приведено в [21].

В рамках доработок в ядро программы был добавлен алгоритм для генерации дефекта в виде реальной трещины. Также реализовано сохранение всех изображений в рабочей директории программы с возможностью перехода на следующее и предыдущее изображение. Для отслеживания текущих операций программы был реализован механизм логирования на основе текстовых записей в область вывода. На рис. 2 представлен графический интерфейс разработанной программы с инициализированной моделью.

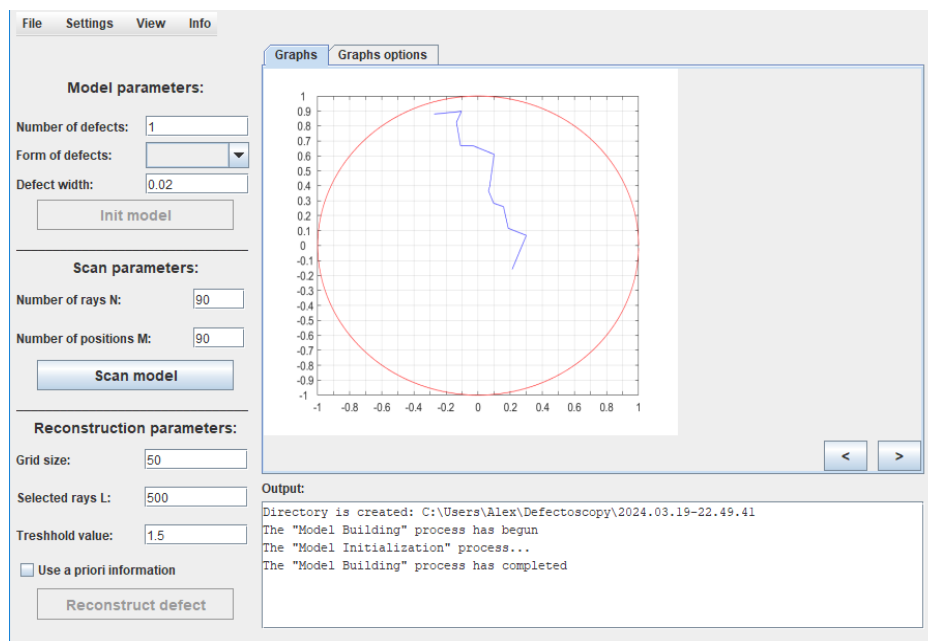


Рис. 2. Графический интерфейс доработанного программного обеспечения

Fig. 2. Graphic interface of improved software

Для повышения быстродействия используемого алгоритма была реализована многопоточная обработка с использованием паттерна «пул потоков» (Thread Pool) [23]. Данный подход применен для всех операций алгоритма, обрабатывающих большие циклы. В табл. 1 представлены результаты применения многопоточного подхода при распараллеливании процесса генерации проекционных данных для сетки 80×80 пикселей. Вычислительные эксперименты проведены на компьютере следующей конфигурации: процессор – Intel Core i7-8700K 4.7 ГГц (6 физических ядер, 12 логических ядер), оперативная память – 32GB DDR4 3 200 МГц.

Таблица 1

Результаты применения многопоточной обработки

Количество потоков	Время выполнения операции «Вычисление значений на детекторе для образца», с	Средняя загрузка CPU, %
1	99,953	11,0
2	50,553	21,1
3	33,272	30,5
4	25,289	39,9
5	20,331	49,7
6	17,392	58,9
7	15,467	68,4
8	13,228	77,9
9	11,519	87,5
10	10,365	97,2
11	9,663	98,8
12	9,191	99,8
13	9,225	99,8
14	9,269	99,8

Показанные результаты демонстрируют, что применение многопоточного подхода позволило в разы сократить время генерации проекционных данных. Причем оптимальное число потоков в данном случае составляет 12. Это совпадает с рекомендациями по выбору оптимального числа потоков, которые гласят, что для задач, выполняющих интенсивные вычисления без блокирующих операций, оптимальная производительность достигается при количестве потоков, равном количеству вычислительных ядер [24].

4. Результаты вычислительных экспериментов

В качестве объекта исследования выбрано круглое сечение промышленного изделия с единственным радиусом, коэффициент поглощения излучения в котором описывается функцией поверхности перевернутого параболоида

$$\mu(x, y) = -x^2 - y^2 + 2, \quad (3)$$

где $\mu(x, y)$ – функция коэффициента поглощения излучения, зависящая от плотности материала объекта, x, y – декартовы координаты.

На рис. 3 представлено графическое представление функции плотности для исследуемого объекта, а также проекция данного отображения на плоскость XOY , так как в условиях томографической реконструкции исследуются именно такие плоские срезы. На всех следующих рисунках числа по осям x и y соответствуют узлам сетки, по оси z (или цветом) – значению коэффициента поглощения.

При снятии измерений используется веерная система наблюдений, состоящая из 90 лучей и 90 положений источника. Для алгебраической реконструкции используется сетка 50×50 пикселей. Критериями точности восстановления модели являются средняя абсолютная ошибка (Median Absolute Error, MAE) и корень среднеквадратического отклонения (Root Mean Square Error, RMSE).

При анализе выходных характеристик следует учитывать, что метрики точности рассчитываются по пикселям всего объекта. Соответственно, показатель точности восстановления может быть высоким, даже если дефект вообще не был обнаружен. То есть при дефектоскопии исследуемого объ-

екта интегральные показатели точности не являются объективными индикаторами наличия или отсутствия дефекта. Поэтому важный этап анализа – рассмотрение графического представления результатов томографической реконструкции.

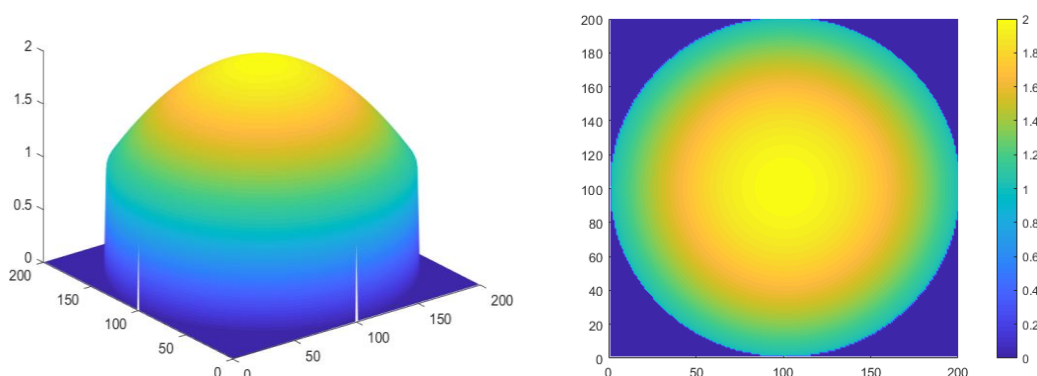


Рис. 3. Графическое представление функции плотности объекта исследования
Fig. 3. Graphic representation of the density function of the object of the study

На рис. 4 показаны результаты томографической реконструкции исследуемого объекта с помощью классического подхода, при котором выполняется алгебраическая реконструкция с использованием всех измерений. Выходные характеристики всех проведенных экспериментов приведены в табл. 2.

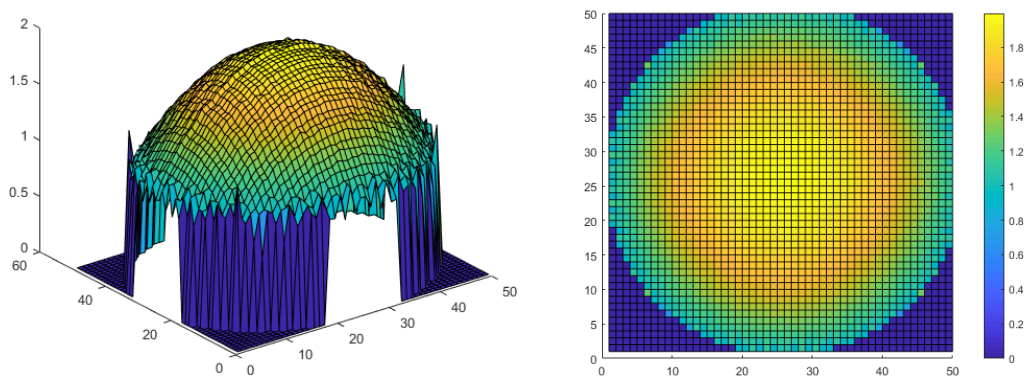


Рис. 4. Результат томографической реконструкции изделия без дефекта
Fig. 4. The result of the tomographic reconstruction of the product without a defect

Таблица 2

Результаты вычислительных экспериментов по томографической реконструкции объекта

Ширина раскрытия дефекта	Размер сетки	Кол-во лучей, участвующих в реконструкции	Точность восстановления MAE	Точность восстановления RMSE	Время вычислений, мс
—	50 × 50	8 100	0,0263	0,0914	131 600
2,5%	50 × 50	8 100	0,0414	0,0989	132 332
0,5%	50 × 50	8 100	0,0434	0,0985	139 058
0,5%	80 × 80	8 100	0,0607	0,0993	391 407
0,5%	80 × 80	2 000	0,0217	0,0702	271 285
0,5%	80 × 80	1 000	0,0157	0,0694	166 191
0,5%	80 × 80	500	0,0130	0,0792	111 535

Приведенные выше результаты свидетельствуют об успешности решения задачи томографической реконструкции, так как внутренняя структура объекта и его плотность хорошо восстановились.

Для следующего эксперимента в модель изделия был добавлен дефект в виде трещины протяженной формы. Ширина раскрытия трещины в самом широком месте составляет 2,5% от диаметра изделия. Данная форма моделируется областью между двух парабол, при этом плотность материала в данной области принимается нулевой. На рис. 5 приведены исходная модель с дефектом и результат ее томографической реконструкции при помощи классического подхода.

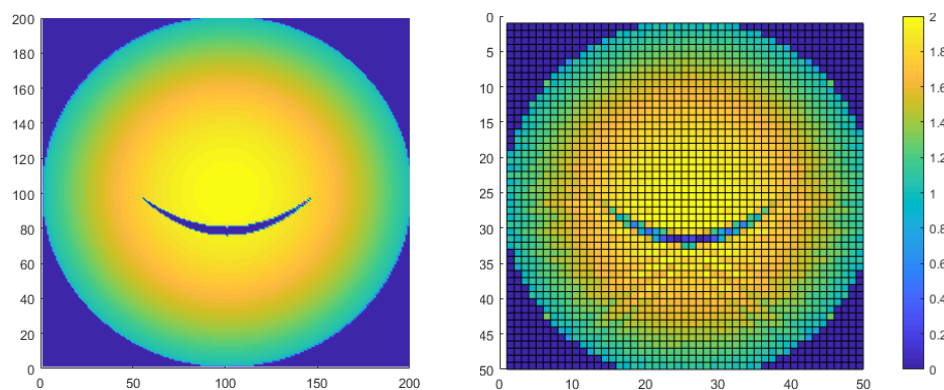


Рис. 5. Модель изделия с большим дефектом и результат ее томографической реконструкции
 Fig. 5. Product model with a large defect and the result of its tomographic reconstruction

Приведенные выше результаты свидетельствуют об успешности решения задачи томографической реконструкции, так как внутренняя структура объекта, его плотность и внутренний дефект хорошо восстановились.

Однако на практике обычно приходится диагностировать дефекты намного меньшей ширины и более сложной формы. Поэтому для исследования более реального случая в модель изделия была добавлена другая трещина, ширина раскрытия которой составляет 0,5% от диаметра системы наблюдений. На рис. 6 приведены исходная модель со сложным дефектом и результаты ее томографической реконструкции при помощи классического и модифицированного подходов (для более объективной оценки отображение линий сетки было отключено).

По рис. 6, *b* можно сделать вывод, что классический алгоритм не справляется с реконструкцией исходной модели с тонким дефектом, так как дефект размывается при восстановлении. Чтобы повысить точность алгебраической реконструкции, классический подход предлагает уменьшить сетку [9]. Поэтому в следующем эксперименте была взята сетка 80×80 пикселей, при этом остальные параметры были оставлены без изменений. На рис. 6, *c* приведены результаты томографической реконструкции исследуемого объекта со сложным дефектом при помощи классического подхода на сетке 80×80 пикселей. В этом случае дефект восстанавливается уже лучше, однако уменьшение сетки приводит к резкому увеличению времени реконструкции (в 2,8 раза).

Такой рост вычислительных затрат закономерен, так как в классическом алгоритме КТ уравнения строятся на основе всех имеющихся измерений по всем неизвестным пикселям. Соответственно, при системе измерений из 90 лучей и 90 положений источника и сетки 50×50 пикселей формируется СЛАУ из 8 100 уравнений и 2 500 неизвестных. А при сетке 80×80 пикселей количество неизвестных составляет уже 6 400. Для диагностики такого типа дефектов и уменьшения вычислительных затрат в работе [21] был описан вычислительно-эвристический алгоритм томографической реконструкции на основе использования априорной информации. В качестве априорной информации используется проекционная матрица эталонного изделия (не содержащего дефектов). Предложенный алгоритм использует при алгебраической реконструкции только часть лучей, отобранных по критерию максимального отличия уровня интенсивности от соответствующих им лучей для эталонного образца. На рис. 6, *d*, *e*, *f* представлены результаты томографической реконструкции с использованием вычислительно-эвристического алгоритма для отобранных 2 000, 1 000 и 500 лучей соответственно. В этих случаях результаты томографической реконструкции можно считать пригодными для решения специальной задачи дефектоскопии, так как трещина четко выделяется в восстановленном изображении. В данной серии экспериментов наилучшие результаты по восстановлению дефекта были получены при использовании 1 000 отобранных лучей, так как в этом случае достигаются высокие показатели точности восстановления, а графический результат реконструкции позволяет выявить трещину. При использовании 500 лучей трещина становится еще более контрастной, однако она прерывается в некоторых местах, что может привести к ошибочным интерпретациям при дальнейшем анализе данных изображений.

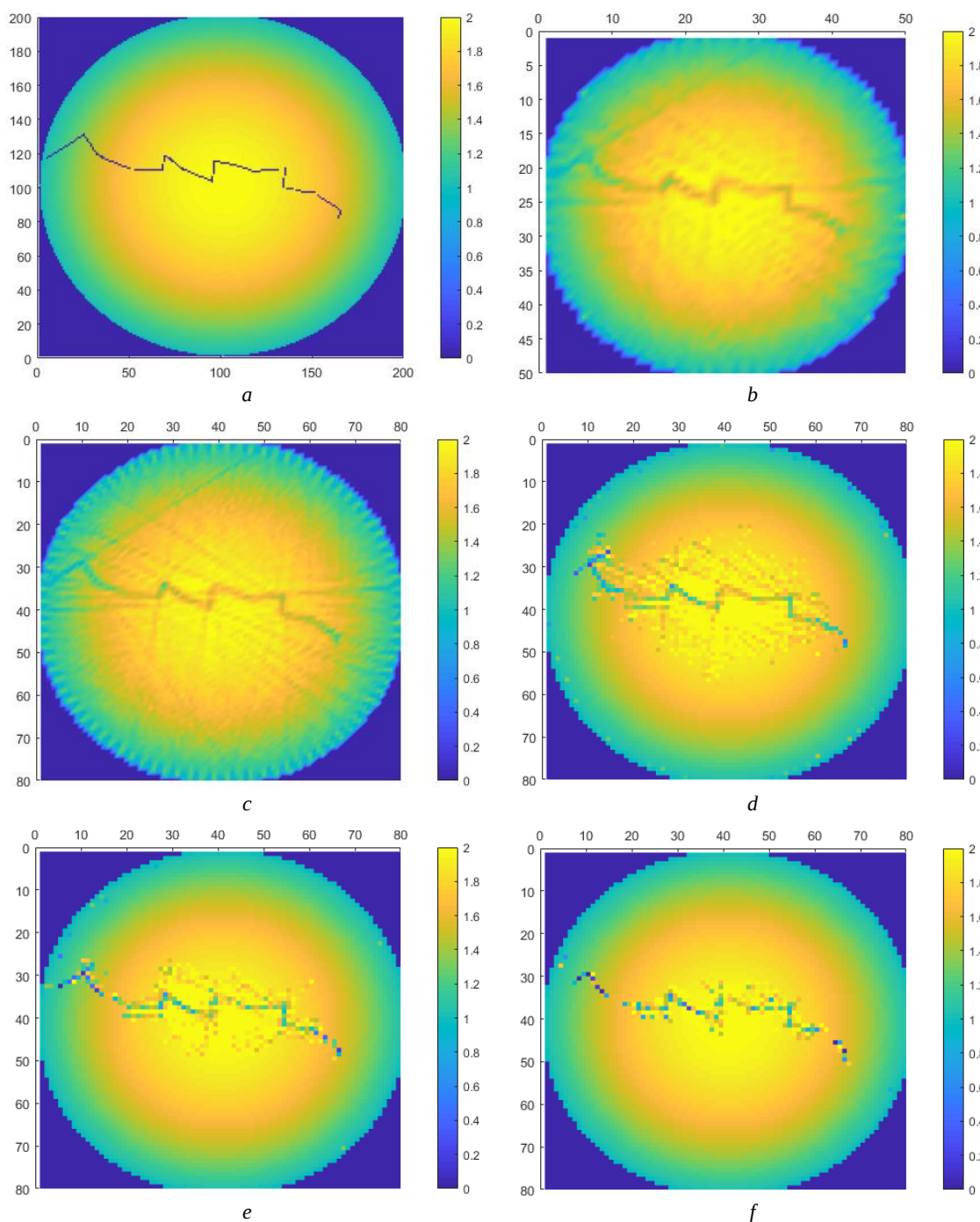


Рис. 6. Результат томографической реконструкции изделия с тонким дефектом: *a* – исходная модель с трещиной; *b* – результат реконструкции для параметров: 8 100 лучей, сетка 50×50 ; *c* – результат реконструкции для параметров: 8 100 лучей, сетка 80×80 ; *d* – результат реконструкции для параметров: 2 000 отобранных лучей, сетка 80×80 ; *e* – результат реконструкции для параметров: 1 000 отобранных лучей, сетка 80×80 ; *f* – результат реконструкции для параметров: 500 отобранных лучей, сетка 80×80

Fig. 6. The result of tomographic reconstruction of a product with a subtle defect: *a* – original model with crack; *b* – reconstruction result for parameters: 8100 rays, 50×50 grid; *c* – reconstruction result for parameters: 8100 rays, 80×80 grid; *d* – reconstruction result for parameters: 2000 selected rays, 80×80 grid; *e* – reconstruction result for parameters: 1000 selected rays, 80×80 grid; *f* – reconstruction result for parameters: 500 selected rays, 80×80 grid

В таблице 2 представлены численные результаты всех проведенных вычислительных экспериментов.

В результате проведенной серии экспериментов можно сделать следующие выводы:

– классический алгоритм КТ имеет высокие вычислительные затраты;

- классический алгоритм КТ плохо справляется с восстановлением дефектов малой толщины из-за неполноты проекционных данных;
- использование модифицированного алгоритма позволяет с высокой точностью реконструировать тонкий дефект на основе априорной информации об эталонном образце исследуемого изделия;
- число отобранных информативных лучей влияет на точность восстановления, оптимальные результаты реконструкции были получены при использовании 1 000 рабочих лучей.

Заключение

В данной работе рассмотрено разработанное модифицированное программное обеспечение, позволяющее использовать ранее предложенный вычислительный алгоритм для решения задач диагностики объектов. Предложенный подход позволяет обнаруживать дефекты малой толщины типа трещин в случаях, когда классический алгоритм компьютерной томографии уже не справляется из-за неполноты проекционных данных. Приведенные результаты экспериментов демонстрируют, что эвристическое усиление дает возможность решить задачу обнаружения дефекта, не меняя систему наблюдений и используя тот же набор измерений. Также необходимо отметить значительный (в 2,4 раза) выигрыш по времени работы модифицированного алгоритма по сравнению с классическим.

В результате дальнейших исследований необходимо предложить критерий для определения оптимального числа информативных лучей, с помощью которых возможно реконструировать дефект при минимуме вычислительных затрат. Также необходимо разработать обширное тестовое покрытие для данного модифицированного алгоритма, чтобы определить границы его применимости. В рамках дальнейшей разработки планируется модифицировать созданное программное обеспечение, добавив в его ядро блок искусственного интеллекта. Предполагается, что это будет сверточная нейронная сеть, способная определять тип дефекта. Следующий этап архитектурной модификации данного программного обеспечения будет заключаться в переходе с монолитного приложения на микросервисное для декомпозиции слабосвязанных вычислительных компонент.

Список источников

1. Bai H., Su M., Pang C., Xiong Z., Binyuan Xia, Zhao D., Li C., Mo Z., Gao F. An image reconstruction method for transmission computed tomography with the constraint of the linear attenuation coefficients // *Applied Radiation and Isotopes*. 2023. V. 202. Art. 111062. doi: 10.1016/j.apradiso.2023.111062
2. Choi H.-J., Jeong J., Lee H., Kim S.G., Ahn J.J., Lee H.C., Min C.H. Image reconstruction method of gamma emission tomography based on prior-aware information and machine learning for partial-defect detection of PWR-type spent nuclear fuel // *Nuclear Engineering and Technology*. 2024. V. 56 (11). P. 4770–4781. doi: 10.1016/j.net.2024.06.040
3. Serrano E., Garcia-Blas J., Carretero J., Desco M., Abella M. Accelerated iterative image reconstruction for cone-beam computed tomography through Big Data frameworks // *Future Generation Computer Systems*. 2020. V. 106. P. 534–544. doi: 10.1016/j.future.2019.12.042
4. Shyamala Bharathi P., Shalini C. Advanced hybrid attention-based deep learning network with heuristic algorithm for adaptive CT and PET image fusion in lung cancer detection // *Medical Engineering & Physics*. 2024. V. 126. Art. 104138. doi: 10.1016/j.medengphy.2024.104138
5. Bellens S., Guerrero P., Vandewalle P., Dewulf W. Machine learning in industrial X-ray computed tomography – a review // *CIRP Journal of Manufacturing Science and Technology*. 2024. V. 51. P. 324–341. doi: 10.1016/j.cirpj.2024.05.004
6. Xu P.-P., Liu T.-Y., Zhou F., Chen Q., Rowe J., Tesche C., Zhang L.-J. Artificial intelligence in coronary computed tomography angiography // *Medicine Plus*. 2024. V. 1, is. 1. doi: 10.1016/j.medp.2023.100001
7. Zhu G., Fu J. A lightweight solution of industrial computed tomography with convolutional neural network // *NDT & E International*. 2020. V. 116. Art. 102347. doi: 10.1016/j.ndteint.2020.102347.
8. Plessis A., Roux S.G., Guelpa A. Comparison of medical and industrial X-ray computed tomography for non-destructive testing // *Case Studies in Nondestructive Testing and Evaluation*. 2016. V. 6, pt. A. P. 17–25. doi: 10.1016/j.csndt.2016.07.001
9. Тихонов А.Н., Арсенин В.Я., Тимонов А.А. Математические задачи компьютерной томографии. М. : Наука, 1987. 160 с.
10. Ito K., Jin B. Inverse Problems: Tikhonov theory and algorithms. Singapore : World Scientific Publishing Company, 2014. 332 p.
11. Важенцева Н.В., Лихачев А.В. Новый метод трехмерной томографии для неполных траекторий источника // *Труды Международной конференции «Современные проблемы прикладной математики и механики: теория, эксперимент и практика», посвященной 90-летию со дня рождения академика Н.Н. Яненко (Новосибирск, Россия, 30 мая – 4 июня 2011)*.

- Новосибирск : ИБТ СО РАН, 2011. URL: http://conf.nsc.ru/files/conferences/niknik-90/fulltext/38813/46493/Likhachov_2.pdf (дата обращения: 04.10.2024).
12. Liu Y., Beyer A., Schuetz P., Hofmann J., Flisch A., Sennhauser U. Cooperative data fusion of transmission and surface scan for improving limited-angle computed tomography reconstruction // *NDT & E International*. 2016. V. 83. P. 24–31. doi: 10.1016/j.ndteint.2016.05.003
 13. Piau P., King A., Henry L., Rathore J.S., Guignot N., Deslandes J.-P., Itié J.-P. A thresholding based iterative reconstruction method for limited-angle tomography data // *Tomography of Materials and Structures*. 2023. V. 2. Art. 100008. doi: 10.1016/j.tmater.2023.100008.
 14. Хахлютин В.П. Об одной задаче интегральной геометрии на плоскости // *Доклады Академии наук СССР*. 1991. Т. 320, № 4. С. 832–834.
 15. Зеркаль С.М. Решение на ЭВМ одного класса задач дефектоскопии // *Доклады Академии наук СССР*. 1990. Т. 314, № 1. С. 180–182.
 16. Зеркаль С.М. Локальная томографическая реконструкция огибающей тонкого дефекта с использованием эталонного образца // *Сибирский журнал индустриальной математики*. 2000. Т. 3, № 1 (5). С. 110–115.
 17. Image reconstruction (CT) // *Radiopaedia*. 2023. 23 Mar. URL: <https://radiopaedia.org/articles/image-reconstruction-ct?lang=us> (accessed: 04.10.2024).
 18. Takase A. How Does CT Reconstruction Work? // *Rigaku*. 2023. Jan. 11. URL: <https://rigaku.com/products/imaging-ndt/x-ray-ct/learning/blog/how-does-ct-reconstruction-work> (accessed: 04.10.2024).
 19. Преобразование Радона // *Википедия*. URL: https://ru.wikipedia.org/wiki/Преобразование_Радона (дата обращения: 04.10.2024).
 20. Бронников А.В., Воскобойников Ю.Е., Преображенский Н.Г. Итерационные алгоритмы в задачах томографии полупрозрачных сред. Новосибирск : ИТПМ, 1989. 43 с. (Препринт; № 18).
 21. Пешков А.В. Вычислительная диагностика трещин и отслоений с использованием томографического подхода при наличии эталона исследуемого объекта // *Доклады Томского государственного университета систем управления и радиоэлектроники*. 2023. Т. 26, № 4. С. 95–101. doi: 10.21293/1818-0442-2023-26-4-95-101
 22. Peshkov A.V., Zerkal S.M. Computational-heuristic algorithm for tomographic solution of industrial flaw detection problems // *16 International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE-2023): Proc.* Novosibirsk, 10–12 Nov. 2023. IEEE, 2023. P. 910–915. doi: 10.1109/APEIE59731.2023.10347879
 23. The Java Tutorials. Thread Pools // *Oracle. Java Documentation*. URL: <https://docs.oracle.com/javase/tutorial/essential/concurrency/pools.html> (accessed: 04.10.2024).
 24. The optimal thread-pool size in Java: Explaining the formula // *Backendhance*. 2023. Aug. 9. URL: <https://backendhance.com/en/blog/2023/optimal-thread-pool-size/> (accessed: 04.10.2024).

References

1. Bai, H., Su, M., Pang, C., Xiong, Z., Binyuan, Xia, Zhao, D., Li, C., Mo, Z. & Gao, F. (2023) An image reconstruction method for transmission computed tomography with the constraint of the linear attenuation coefficients. *Applied Radiation and Isotopes*. 202. Art. 111062. DOI: 10.1016/j.apradiso.2023.111062
2. Choi, H.-J., Jeong, J., Lee, H., Kim, S.G., Ahn, J.J., Lee, H.C. & Min, C.H. (2024) Image reconstruction method of gamma emission tomography based on prior-aware information and machine learning for partial-defect detection of PWR-type spent nuclear fuel. *Nuclear Engineering and Technology*. 56(11). pp. 4770–4781. DOI: 10.1016/j.net.2024.06.040
3. Serrano, E., Garcia-Blas, J., Carretero, J., Desco, M. & Abella, M. (2020) Accelerated iterative image reconstruction for cone-beam computed tomography through Big Data frameworks. *Future Generation Computer Systems*. 106. pp. 534–544. DOI: 10.1016/j.future.2019.12.042
4. Shyamala Bharathi, P. & Shalini, C. (2024) Advanced hybrid attention-based deep learning network with heuristic algorithm for adaptive CT and PET image fusion in lung cancer detection. *Medical Engineering & Physics*. 126. Art. 104138. DOI: 10.1016/j.medengphy.2024.104138
5. Bellens, S., Guerrero, P., Vandewalle, P. & Dewulf, W. (2024) Machine learning in industrial X-ray computed tomography – a review. *CIRP Journal of Manufacturing Science and Technology*. 51. pp. 324–341. DOI: 10.1016/j.cirpj.2024.05.004
6. Xu, P.-P., Liu, T.-Y., Zhou, F., Chen, Q., Rowe, J., Tesche, C. & Zhang, L.-J. (2024) Artificial intelligence in coronary computed tomography angiography. *Medicine Plus*. 1(1). DOI: 10.1016/j.medp.2023.100001
7. Zhu, G. & Fu, J. (2020) A lightweight solution of industrial computed tomography with convolutional neural network. *NDT & E International*. 116. Art. 102347. DOI: 10.1016/j.ndteint.2020.102347
8. Plessis, A., Roux, S.G. & Guelpa, A. (2016) Comparison of medical and industrial X-ray computed tomography for non-destructive testing. *Case Studies in Nondestructive Testing and Evaluation*. 6(A). pp. 17–25. DOI: 10.1016/j.csndt.2016.07.001
9. Tikhonov, A.N., Arsenin, V.Y. & Timonov, A.A. (1987) *Matematicheskie zadachi komp'yuternoy tomografii* [Mathematical tasks of computed tomography]. Moscow: Nauka.
10. Ito, K. & Jin, B. (2014) *Inverse Problems: Tikhonov theory and algorithms*. Singapore: World Scientific Publishing Company.
11. Vazhentseva, N.V. & Likhachev, A.V. (2011) Novyy metod trekhmernoy tomografii dlya nepolnykh traektoriy istochnika [The new method of three-dimensional tomography for incomplete source trajectories]. *Sovremennye problemy prikladnoy matematiki i mekhaniki: teoriya, eksperiment i praktika* [Modern Problems of Applied Mathematics and Mechanics: Theory,

- Experiment and Practice]. Proc. of the International Conference. Novosibirsk: SB RAS. [Online] Available from: http://conf.nsc.ru/files/conferences/niknik-90/fulltext/38813/46493/Likhachov_2.pdf (Accessed: 4th October 2024).
12. Liu, Y., Beyer, A., Schuetz, P., Hofmann, J., Flisch, A. & Sennhauser, U. (2016) Cooperative data fusion of transmission and surface scan for improving limited-angle computed tomography reconstruction. *NDT & E International*. 83. pp. 24–31. DOI: 10.1016/j.ndteint.2016.05.003
 13. Piault, P., King, A., Henry, L., Rathore, J.S., Guignot, N., Deslandes, J.-P. & Itié, J.-P. (2023) A thresholding based iterative reconstruction method for limited-angle tomography data. *Tomography of Materials and Structures*. 2. Art. 100008. DOI: 10.1016/j.tmater.2023.100008
 14. Khakhlyutin, V.P. (1991) Ob odnoy zadache integral'noy geometrii na ploskosti [About one task of integral geometry on the plane]. *Doklady Akademii nauk SSSR*. 320(4). pp. 832–834.
 15. Zerkal, S.M. (1990) Reshenie na EVM odnogo klassa zadach defektoskopii [Computer solution of one class of flaw detection problems]. *Doklady Akademii nauk SSSR*. 314(1). pp. 180–182.
 16. Zerkal, S.M. (2000) Lokal'naya tomograficheskaya rekonstruktsiya ogibayushchey tonkogo defekta s ispol'zovaniem etalonnogo obraztsa [Local tomographic reconstruction of the envelope of a thin defect using a reference sample]. *Sibirskiy zhurnal industrial'noy matematiki*. 3(1). pp. 110–115.
 17. Radiopaedia. (2023) Image reconstruction (CT). 23rd March. [Online] Available from: <https://radiopaedia.org/articles/image-reconstruction-ct?lang=us> (Accessed: 4th October 2024).
 18. Takase, A. (2023) How Does CT Reconstruction Work? *Rigaku*. 11th January. [Online] Available from: <https://rigaku.com/products/imaging-ndt/x-ray-ct/learning/blog/how-does-ct-reconstruction-work> (Accessed: 4th October 2024).
 19. Wikipedia. (n.d.) *Преобразование Радона* [Radon transformation]. [Online] Available from: https://ru.wikipedia.org/wiki/Преобразование_Радона (Accessed: 4th October 2024).
 20. Bronnikov, A.V., Voskoboynikov, Yu.E. & Preobrazhenskiy, N.G. (1989) *Iteratsionnye algoritmy v zadachakh tomografii poluprozrachnykh sred* [Iterative algorithms in problems of tomography of semi-transparent media]. Novosibirsk: ITPM.
 21. Peshkov, A.V. (2023) Vychislitel'naya diagnostika treshchin i otsloeniy s ispol'zovaniem tomograficheskogo podkhoda pri nalichii etalona issleduemogo ob"ekta [Computational diagnostics of cracks and delaminations using a tomographic approach in the presence of a standard of the object under study]. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki*. 26(4). pp. 95–101. DOI: 10.21293/1818-0442-2023-26-4-95-101
 22. Peshkov, A.V. & Zerkal, S.M. (2023) Computational-heuristic algorithm for tomographic solution of industrial flaw detection problems. *16 International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE–2023)*. Novosibirsk, 10–12 Nov. 2023. pp. 910–915. DOI: 10.1109/APEIE59731.2023.10347879
 23. *The Java Tutorials. Thread Pools*. [Online] Available from: <https://docs.oracle.com/javase/tutorial/essential/concurrency/pools.html> (Accessed: 4th October 2024).
 24. *Backendhance*. (2023) The optimal thread-pool size in Java: Explaining the formula. 9th August. [Online] Available from: <https://backendhance.com/en/blog/2023/optimal-thread-pool-size/> (Accessed: 4th October 2024).

Информация об авторах:

Зеркаль Сергей Михайлович – профессор, доктор технических наук, профессор кафедры вычислительной техники факультета автоматики и вычислительной техники Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: zerkal@corp.nstu.ru

Пешков Александр Викторович – аспирант кафедры вычислительной техники факультета автоматики и вычислительной техники Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: mupeskov1997@mail.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Zerkal Sergey M. (Doctor of Technical Sciences, Professor, Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: zerkal@corp.nstu.ru

Peshkov Alexander V. (Post-graduate Student, Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: mupeskov1997@mail.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 16.10.2024; принята к публикации 03.03.2025

Received 16.10.2024; accepted for publication 03.03.2025

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2025. № 70

Редактор Е.Г. Шумская
Оригинал-макет Е.Г. Шумской
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцовой

Подписано к печати 28.03.2025 г. Формат 60x84¹/₈.
Гарнитура Times. Усл. печ. л. 15,8.
Тираж 250 экз. Заказ № 6251. Цена свободная.

Дата выхода в свет 01.04.2025 г.

Журнал отпечатан на полиграфическом оборудовании
Издательства Томского государственного университета
634050, г. Томск, Ленина, 36
Тел. 8(382-2)–52-98-49; 8(382-2)–52-96-75
Сайт: <http://publish.tsu.ru>; E-mail: rio.tsu@mail.ru